王华忠文集

地震波反演成像方法理论

王华忠

波现象与智能反演成像研究组（WPI）

同济大学海洋与地球科学学院

2025 年 1 月

王华忠文集

地震波反演成像方法理论

同济大学

WPI简介

波现象与智能反演成像研究组（WPI）由在勘探地震学领域有20余年研究经验的王华忠教授担任学术带头人，致力于探索研究勘探地震学领域的前沿问题。目前研究团队有教授1人，研究员2人，博士生13人，硕士生6人，常任秘书2人。

波现象与智能反演成像研究组的组建目的

·培养地震数据分析领域中的领军人才；

·提出地震波反演成像方面的新理论与新方法；

·针对中国各大油气战略探区遇到的地震波反演成像问题发展出新方法和新技术。

波现象与智能反演成像研究组的总体学术思想

在Bayes估计理论的框架下，针对非线性、非高斯和非平稳数据情形，发展更为精确的信号与图像建模方法；基于波场特征表达、参数场特征表达，用一组凸的参数估计反问题逼近强非线性反问题（FWI）的解；用机器学习（ML）和人工智能（AI）算法改进地震数据分析中问题的解决方案。

WPI网站：

http://www.wpi-tongji.cn/

王华忠教授邮箱：

herbhuak@vip.163.com

WPI微信公众号

目 录

前言…… I

2012 年 WPI 论文集

1 地震波反演的框架性认识…… 1
2 地震波反演问题理论分析…… 11
3 全波形反演与波形层析成像的数值计算问题…… 39

2013 年 WPI 论文集

4 地震波传播的散射场表达与逆散射成像…… 67
5 地震波反演成像中的现代图像处理方法及应用…… 85
6 地震波反演成像中的线性与非线性方程组的解法…… 105
7 面向岩性油气藏的地震波反演成像理论…… 123
8 面向岩性油气藏的最小二乘叠前深度偏移成像…… 147
9 透射波与反射波层析、反射波最小二乘偏移…… 169
10 地震波反演成像方法与技术核心问题分析…… 187
11 地震反演及其在储层刻画中的应用…… 199

2014 年 WPI 论文集

12 勘探地震学中可分解的速度反演…… 213
13 地震波反演成像正则化思想与方法…… 221

2015 年 WPI 论文集

14 状态变量独立的地震波反演成像…… 235
15 反射波成像与散射波成像的对比分析…… 249

2016 年 WPI 论文集

16 勘探地震中应用泛函分析基础知识总结…… 263
17 扩散方程和椭圆方程在地震波反演成像中的应用…… 301
18 特征波反演成像理论框架…… 319
19 最小二乘叠前深度偏移成像理论与方法…… 333

2017 年 WPI 论文集

20 地震数据成像处理中若干重要问题的思考…… 351
21 勘探地震学技术发展方向分析…… 377

2018 年 WPI 论文集

22 从线性高斯反演成像到非线性非高斯反演成像的演变…… 383
23 基于二阶统计量的信号分析方法存在问题及解决方案…… 397

2019 年 WPI 论文集

24 PCA/RPCA/CCA/ICA 与地震信号建模…… 405

25 反问题提法与解法的本质逻辑探讨…….. 423
26 线性与非线性地震波反演成像理论问题分析……. 439

2020 年 WPI 论文集

27 勘探地震中地震信号的相位问题分析…….. 471
28 勘探地震中强非线性反问题的解法…… 481
29 人体勘探与油气勘探方法技术的对比分析…… 507

2021 年 WPI 论文集

30 AI 和 ML 在勘探地震中的应用前景分析…… 525
31 L0 与 L1 正则化约束下反问题解的性质分析…….. 539
32 地震数据分析中的计算模式分析…….. 553
33 反褶积问题的基本理论与在勘探地震中应用的认识…….. 563
34 复杂地表地震勘探的认识及复杂地表探区地震数据成像处理的观点…….. 575
35 高维局部数据体中线性信号预测基本理论与方法…… 591
36 基于信息理论的推断原则和学习算法…… 609

2022 年 WPI 论文集

37 Bayes 参数估计与 Bayes 决策理论方法及勘探地震数据分析中的应用…….. 619
38 几何在地震数据分析中的应用…… 631
39 约束非线性优化问题的提出与求解…… 641
40 再议地震波成像道集物理实质…… 655
41 针对岩性储层描述的定量地震波成像…… 663

2023 年 WPI 论文集

42 OBN 地震数据成像处理基本逻辑与关键方法技术…… 693
43 迭代法解线性系统的基本思想…….. 711
44 多源（元）信息融合及空间数据分析的基本逻辑…….. 721
45 反射波 FWI 在陆上地震数据成像处理中的实用化途径探讨…… 733
46 高维数据分析的本质问题分析…….. 743
47 随机模拟及随机反演的思想与方法…….. 749

2024 年 WPI 论文集

48 典型的逆矩阵及在统计推断理论与方法中的应用…… 761
49 勘探地震中非线性问题及求解方法分析…… 783
50 勘探地震中需要六分量（包括旋转分量的）观测吗？…… 793
51 深层超深层油气勘探技术路线及针对性地震波成像处理方法技术…….. 807
52 再论随机模拟及在统计估计中的应用…… 825
53 再议空间数据分析及多信息融合思想与方法…….. 841

结语…… 857

致谢…… 859

前言

勘探地震学从一开始就是个以数据采集和数据分析为基本任务的学科。数据采集和数据分析也是信号分析学科的基本研究内容。我们一直认为，21 世纪的勘探地震学进入了一个以现代信号分析为核心理论框架的技术发展时代。信号的预测理论和 Bayes 框架下的参数估计理论是现代信号分析学科的两项基本任务。勘探地震学中地震数据分析的核心任务也是这两项。

油气地震勘探的根本目标是利用叠前数据体通过反演成像方法获取宽波数带的弹性参数场，与岩石物理结合，对含油气储层实现尽可能精确的描述。其中的关键问题是如何获取宽波数带的弹性参数场。

我们一直认为应该用系统的观点看待信号分析、图像分析和弹性参数估计问题。信号是随机的，信号总是系统的输出。把三维探区视为一个要研究的系统，在地表（或空中、或井中）接收到的信号和探区内弹性参数场构成了信号与系统的关系。研究它们之间的关系是现代信号分析学科，也是勘探地震学科，的根本任务。现代信号分析本质上包含两部分研究内容：信号的表达（或建模）；系统参数估计。信号的建模是基础，期望建立的模型能对信号进行准确的预测。若把波动方程视为波场预测器，信号分析和弹性参数反演就完全统一在了一个理论框架下，可以认为就是 Bayes 估计理论框架。当前，所有的现代信号/图像分析的新思想和新进展都可以借用来进行地震波反演成像。

在信号和波场能被准确预测表达的基础上，估计理论框架下要解决的问题就是假设地震数据是符合某种分布的随机信号（就是说随机噪音的分布是可以用某种统计模型描述的），然后建立合适的误差泛函，并假设它是在局部上具有一定凸性的，接着求解一个局部最优问题，得到介质参数的某种估计。信号分析中的去噪音、插值等等也是按上述逻辑进行的。在这个过程中，把反演问题提成一个拟凸问题或凸问题是最关键的；拟凸问题或凸问题的求解方法是第二位的。多尺度特征表达数据和模型，并建立分尺度的二者之间的比较线性的关系是把反演问题提成一个凸问题的核心思想。现代图像处理逐步地在勘探地震中得到重要应用，从成像中自动地提取反射结构信息逐渐成为了反演成像过程中的核心步骤。

Tarantola（2005）的代表作“模型参数估计的反问题理论与方法（Inverse Problem Theory and methods for Model Parameter Estimation）”非常明确地指出地震波反演要估计的是波动理论正问题模型中所包含的模型参数。模型参数和地下介质弹性参数之间的关系由波动理论正问题能预测实际观测数据而建立起来。正是基于此，估计模型参数也就能在某种程度上“等价于”估计地下介质弹性参数。可以说我们直接地估计模型参数，间接地估计地下介质的弹性参数。从这一点可充分地看出波动理论正问题模型在地震波反演中到底有多重要！正问题描述与刻画的合理性才是有效地解决反问题的核心！

数据，实测数据，是估计地下介质弹性参数真正的信息源，波在地下介质中的传播过程受到地下介质弹性参数的改造，从而使得在地表接收到的数据（或波场）中携带了地下介质弹性参数的信息。勘探地震中的实测数据总是不完备的，所蕴含的信息不足以高精度地估计地下介质弹性参数。

因此，Bayes 框架下的弹性参数估计是当然的选择。Bayes 估计理论真是一个完美的理论架构！

很遗憾，勘探地震中高维的弹性参数估计过程中并不能计算出后验概率密度函数，也不能据此算出估计结果的期望值。因此，实测数据与模型参数之间的先验概率分布被假设成 Gauss 的或广义 Gauss 的。同样地，必须要引入的模型参数的先验概率分布也被假设成满足 Gauss 的或广义 Gauss 的。引入如此假设的目的是试图利用 Gauss 或广义 Gauss 情形下后验概率密度函数的最大化来对模型参数估计结果的优劣做出判决。众所周知，Gauss 或广义 Gauss 情形下后验概率密度函数的最大化等价于对应的代价函数的最小化，从而把高维弹性参数估计问题退化为一个对代价函数求极小的最优化问题。

这还没有完！Gauss 或广义 Gauss 情形下后验概率密度函数本质上要求波动理论正问题模型是线性的或

弱非线性的。强非线性性的波动理论正问题模型与后验概率密度函数满足 Gauss 或广义 Gauss 分布假设是矛盾的！因此，引入线性化的正问题进行数据预测就成了必须！

为了发展出有效的模型参数估计方法，我们不得不在各种假设下建立各种“模型”。这些“模型”与要解决的实际问题符合吗？显然差距甚大。

Gauss 线性反演理论是又一个完美的理论架构！在对波场预测算子进行线性化、预测误差满足 Gauss 分布（尤其是 Gauss 白噪分布）、L2 范数定义预测误差的前提下，Gauss 线性反演问题所有内涵全部是清楚的。最理想情况下，模型参数估计结果是无偏的和方差最小的。理论是完美的！现实却是差距甚大的！当前，勘探地震学中的地震波线性反演方法与技术都在解决复杂的地下介质情况、复杂的地震波场、不完全的地震观测、不准确的地震波数值模拟情况下，如何把线性反演理论用在实际数据的反演过程中，以得到更好的成像结果，各种正则化思想与方法也备受重视。地震波线性反演方法与技术的进步是有目共睹的！但是，代表性的方法 FWI 和 LS_RTM 还是不能广泛地应用与实际数据成像处理中。

下一步应该怎么办？我们认为在非 Gauss 和非线性情况下发展出有效的模型参数估计算法是问题的关键！

总之，勘探地震学中的地震数据采集、成像处理和参数反演、乃至地震地质解释逐步统一在了现代信号/图像分析的理论框架下。在此总框架下，基于“两宽一高”的叠前地震数据，得到宽波数带的弹性参数估计结果，结合机器学习实现高精度的储层描述，是波现象与反演成像研究组确定的学术和方法技术发展导向。

地震波反演的框架性认识

王华忠，王雄文

①同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】Bayes 理论框架下的地震波全波形反演是油气勘探中的导引性技术，其基本思想在去噪音、反褶积、地震数据规则化、一维波阻抗反演、AVA（叠前）弹性参数反演、叠前偏移成像、速度分析及层析成像中占据核心位置。但由于观测数据与反演模型参数之间的高度非线性性，导致地震波全波形反演（FWI）在实际地震数据处理中的难度较大。据此，首先从概率论的观点说明了地震波反演的本质，指出在假设观测噪音为高斯白噪的情况下，Bayes 估计可以在最小二乘意义下实现；接着分析了数据空间向参数空间映射的数学物理含义，指出映射的非线性性强弱取决于数据和模型之间的非线性关系，更准确地说取决于介质模型的复杂性和描述地震波物理传播过程的正算子的复杂性；最后在分析陆上和海上地震数据特点的基础上，指出了地震波反演走向实用化的策略。通过以上分析，提出：①速度场的反演是利用特征波场的反演（CWI），而不是全波形的反演；②合理地增加波场的相位信息在范函中所占的比例；③尽量充分考虑初始模型的先验信息。只有满足以上 3 个条件，才能使地震波全波形反演逐步走向实用化。

【关键词】地震波反演；线性与非线性；数据空间与模型空间；正则化；多层次与多尺度反演

1 引言

地震波反演成像包含 2 个层次，即反射系数估计（或称真振幅成像，其简化形式为一般的叠前偏移成像，它仅仅定位反射系数出现的位置）；全波形反演（Full Waveform Inversion，简写为 FWI）估计速度和密度（也可估计各向异性参数和吸收衰减参数等，但多参数反演的难度更大）。

地震波反演的基本思想是 Bayes 估计理论。Tarantola[1]详细而深入地分析了地震波全波形反演的理论问题。在整个地震数据处理与反演成像过程中，去噪音、反褶积、数据规则化、一维波阻抗反演、AVA（叠前）弹性参数反演、叠前偏移成像、速度分析、层析成像等都是在此框架下进行的。根据 Bayes 估计理论，首先假设存在一个正演算子，其可以预测观测数据，预测噪音为高斯白噪，此时的 Bayes 估计可以在最小二乘意义下实现；在一定的范数意义下建立一个预测误差，数据误差一般用 L2 范数，模型约束一般用 L1 或 Cauchy 范数等；根据正算子的线性与否，选择用线性最小二乘方法或非线性最小二乘方法求解反演问题。

预测观测数据的正算子的作用十分重要。褶积算子用于反褶积；一维波动方程用于预测自激自收的地震道（波阻抗反演）；AR 模型算子常用于线性信号预测（压制噪音）；旅行时计算算子用于预测实测数据中的旅行时（层析成像）；Zoeppritz 方程及各种简化形式预测角度反射系数（AVA 弹性参数反演）；Fourier 变换算子用于预测规则无假频数据的谱；动校正时距关系预测实测 CMP 道集中的双曲时距关系（估计均方根速度）；各种形式的波动方程预测实际观测的炮记录（全波形反演）。

反演问题是典型的不适定问题，包括解的存在性、唯一性和稳定性问题。在地球物理反演中，解的存在性是有物理保证的，主要是解的不唯一性和计算过程中的不稳定性问题。从最小二乘解法来看，二者具有关联性，计算的不稳定性更加剧了解的不唯一性，但最根本的还是解的不唯一性问题。

反演问题解的不唯一性本质上是由数据空间向模型空间映射的非线性性引起的，更进一步地，是由模型参数变化的复杂性和描述波场和参数场关系的正算子决定的。不唯一性只可以减弱，不可以完全消除。减弱反演解的不唯一性的关键是减弱由数据空间向模型空间映射的非线性性，以及增加关于反演解的先验信息。

20 世纪 80 年代，地震波反演处于热潮时期，多位学者[2-6]对全波形反演理论做了十分扎实的奠基工作。但受当时野外地震观测技术和高性能计算机技术的局限，无法将提出的理论方法转为实用。此后，一维波阻抗反演、AVA 叠前弹性参数反演得到不断发展，但从理论上看，这些反演方法都不具备高维叠前反演的潜力。

近几年，地震波高维叠前反演又成为勘探地球物理方面的核心议题，这首先得益于地震数据采集技术和大规模高性能计算机技术的进步，很多研究者[7-24]对地震波高维叠前反演研究的兴起也做了大量铺垫工作。

笔者首先从概率论的观点说明了地震波反演的本质，指出在假设观测噪音为高斯白噪的情况下，Bayes 估计可以在最小二乘意义下实现；分析了数据空间向参数空间映射的数学物理含义，指出映射的非

线性性强弱取决于数据和模型之间的非线性关系，更准确地说取决于介质模型的复杂性和描述地震波物理传播过程的正算子的复杂性；在分析陆上和海上地震数据特点的基础上，指出了地震波反演走向实用化的策略。

2 地震波反演的概率论观点

2.1 实测地震信号（地震波场）的随机性

一般认为观测到的地震数据是一个确定性信号，但如果从概率论的角度看，将采集地震数据这一过程视为一个随机过程，则观测到的地震数据只是该随机过程的一次具体实现。因此，地震数据是一个随机的信号，具有均值、方差等统计特征，而不是一个确定性的信号。

反演的结果在很大程度上取决于所用的观测数据，当观测数据不同时，所得到的反演结果也肯定不同，因此当所用的观测数据具有随机性时，反演的结果也肯定是随机的。由于反演结果的方差描述了反演结果的可靠性，所以方差的值不能太大，否则所得到结果的可靠性就会很低，反演问题就会失去其意义和价值。所以，对反演问题的求解过程也可以理解为利用尽可能多的先验信息和约束条件来降低反演结果方差的一个过程。

综上所述，基于随机观测数据的反演结果也应该是随机的，其主要统计特征可以用均值和方差来衡量。所得估计结果的统计特征应该是无偏、方差最小。无偏是为了保证估计结果与真实情况保持一致，方差最小是为了保证估计结果的可靠性达到最高。为了得到这样的估计结果，可以在 Bayes 理论框架下求解反演问题。

2.2 Bayes 估计基本思想

Bayes 估计的思想可以简单概括为：估计的结果 $\hat{\mathbf{m}}$ 是由后验概率密度函数 $P(\mathbf{m}/\mathbf{y})$ 所决定的均值 $E\{\mathbf{m}/\mathbf{y}\}$，其中 $\mathbf{y}$ 代表观测数据。

$$\hat{\mathbf{m}} = E\{\mathbf{m}/\mathbf{y}\} = \int \mathbf{m} P(\mathbf{m}/\mathbf{y}) d\mathbf{m} \tag{1}$$

式(1)表示参数估计的结果是由观测数据所对应的模型参数出现的概率所决定的。

由于后验概率密度 $P(\mathbf{m}/\mathbf{y})$ 很难估计，所以将其转化为较容易估计的模型空间的先验概率密度函数 $P(\mathbf{m})$、数据空间的概率密度函数 $P(\mathbf{y})$ 和条件概率密度函数 $P(\mathbf{y}/\mathbf{m})$ 来表示，其中 $P(\mathbf{y}/\mathbf{m})$ 体现了正算子的不确定性，描述了从模型空间到数据空间的映射关系。该 Bayes 公式为

$$P(\mathbf{m}/\mathbf{y}) = \frac{P(\mathbf{y}/\mathbf{m})P(\mathbf{m})}{P(\mathbf{y})} \tag{2}$$

2.3 地震波反演的概率论认识

关于模型空间的概率密度函数 $P(\mathbf{m})$，其先验信息就是模型参数的均值和方差，$P(\mathbf{m})$ 的取值反映了对模型的认识情况。$P(\mathbf{m})$ 的均值和方差在反演求解过程中主要起到约束解的取值范围的作用。如果 $P(\mathbf{m})$ 某个分量的方差较小，代表由先验信息所决定的该分量的值的可靠性就较高，因此该分量的估计结果主要由先验信息所决定；如果 $P(\mathbf{m})$ 某个分量的方差较大，代表由先验信息所决定的该分量的值的可靠性就较低，因此该分量的估计结果主要由观测数据所决定。$P(\mathbf{m})$ 初步定义了反演问题的解在整个模型空间的概率密度函数，而反演问题的求解过程就是利用所观测到的数据更新这一概率密度函数的过程。

关于数据空间的概率密度函数 $P(\mathbf{y})$，可采用以下方式定义其均值和方差。地震数据可认为是随机过程的一次具体实现，由于该随机过程只实现一次，所以可将该次实现看作该次随机过程的均值，即 $P(\mathbf{y})$

的均值由采集到的地震数据所决定。而$P(\mathbf{y})$的方差主要由地震数据的观测方式所决定，地震数据在某个空间范围内的采样密度越高，该空间范围内的地震数据所对应的方差就越大。如果假设数据空间的概率密度函数为高斯函数，则数据空间的概率密度函数可以用上述估计到的均值和方差来决定。

关于条件概率密度函数$P(\mathbf{y}/\mathbf{m})$，它表示在给定模型$\mathbf{m}$的条件下，采集到数据$\mathbf{y}$的可能性的大小。实际上，$P(\mathbf{y}/\mathbf{m})$也体现了用来描述实际物理现象的正算子的精确程度，当所用的正算子可以准确无误地描述实际物理过程时，$P(\mathbf{y}/\mathbf{m})$应该是一个δ类函数，否则它也应该有自己的均值和方差。因此，正演问题描述方法的差异会显著地影响$P(\mathbf{y}/\mathbf{m})$的取值。一般情况下，假设正算子在某些假设条件下可以比较准确地描述地震波的物理传播过程，并且其概率密度函数也是高斯函数。但实际情况并非如此，由于使正算子成立所要求的那些假设条件在实际情况中不一定都能得到满足，因此用正算子预测的数据与观测数据会存在较大的差异，并且正算子的概率密度函数也不一定是高斯函数，再加上先验信息的不充分，反演问题实际上是一个非常复杂的问题，如果不加上合理的约束条件，则很难直接得到一个具有地质意义的结果。

根据 Bayes 公式可以看出，$P(\mathbf{m}/\mathbf{y})$的均值和方差主要由$P(\mathbf{m})$，$P(\mathbf{y})$和$P(\mathbf{y}/\mathbf{m})$来决定。从这个角度看，对反演问题的求解可以认为是在$P(\mathbf{y})$，$P(\mathbf{m})$和$P(\mathbf{y}/\mathbf{m})$的约束下，使$P(\mathbf{m}/\mathbf{y})$最大化，并且期望均值为$\hat{\mathbf{m}}=\int \mathbf{m}P(\mathbf{m}/\mathbf{y})d\mathbf{m}$时，模型空间的方差最小。目前，常规反演问题的基本做法是假设$P(\mathbf{m})$和$P(\mathbf{y}/\mathbf{m})$均为高斯函数，在忽略观测系统不规则性的前提下，假设$P(\mathbf{y})$是标准的正态分布（当地震数据是规则采样时，其各个分量的方差应均相等），把$P(\mathbf{m}/\mathbf{y})$的最大化问题转化为代价函数$S(\mathbf{m})$的最小化问题。一般地，如果假设$P(\mathbf{m})$和$P(\mathbf{y})$均为高斯分布，并且忽略正算子的不确定性，则代价函数$S(\mathbf{m})$可以简单地定义为

$$S(\mathbf{m})=\frac{1}{2}\begin{bmatrix}(g(\mathbf{m})-\mathbf{d}_{obs})^H C_D^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})\\+\lambda(\mathbf{m}-\mathbf{m}_{prior})^H C_M^{-1}(\mathbf{m}-\mathbf{m}_{prior})\end{bmatrix} \tag{3}$$

式中：C_D为数据协方差矩阵；C_M为模型的协方差矩阵。

如果正演问题$g(\mathbf{m})$是线性函数，即$g(\mathbf{m})=G\mathbf{m}$，并且矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）是正定的，则$S(\mathbf{m})$是一个正定二次函数，有唯一的极小值点。求解$S(\mathbf{m})$极小值点的收敛速度和解的稳定性取决于矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）条件数的大小，因此可以用 Newton 类的方法快速得到其最优解。如果矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）是不定的，即其特征值有正有负，则要先对$S(\mathbf{m})$做适当的正则化处理（在$S(\mathbf{m})$的构造过程中加入适当的约束条件），再求取其极小值点。

如果正演问题$g(\mathbf{m})$是非线性问题，但代价函数$S(\mathbf{m})$是个严格凸的函数，那么$S(\mathbf{m})$的极小值点也是唯一的。求解$S(\mathbf{m})$极小值点的收敛速度和解的稳定性，取决于$S(\mathbf{m})$的二阶导数在模型空间各个方向上的比值的大小，此时各种最优化方法均可以用来求解其极小值点。如果初始值在极小值点附近，可以用一个二次函数来近似$S(\mathbf{m})$在该局部范围内的值，这时也可以用 Newton 类方法加快收敛速度。如果$S(\mathbf{m})$不是严格凸的函数，则$S(\mathbf{m})$的极小值点不再唯一，解决此类问题可以用信赖域方法、Monte Carlo 方法等。这些方法都可以用来求解非线性最优化问题，但由于 Monte Carlo 方法计算量大，如果没有一个比较好的策略则很难得以实现，因此信赖域方法在这种情况下是一个比较好的选择。信赖域方法的具体实现方式有很多种，其主要思想就是通过相关的先验信息来约束解的范围，从而降低解空间的大小，使$S(\mathbf{m})$在约束后的解空间里为一个严格凸的函数，再用相应的最优化方法即可求取其极小值点。

通常也可以把反演过程视为在所有可能的参数模型形成的模型空间（也是方程的解空间）中进行采样。反演结果对应的采样点应该是能够使式(3)达到极小值，或是使后验概率密度 $P(\mathbf{m}/\mathbf{y})$ 达到最大值的采样点。基于上述用采样的观点来理解反演问题，可以认为，在确定性的观点下利用梯度导引类的方法求解非线性反演问题仅仅是在初始值附近的局部范围内对可能的模型参数进行采样。但勘探地震学中模型空间规模巨大，初始值虽很重要，但其选择也很困难。如果初始值的选择不能接近真实值，梯度导引类的算法将不能得到有意义的反演结果。另一方面也说明所有可能的参数模型依赖于建模过程中对参数所选择的描述方式，其会直接影响到模型空间的采样方式。因此，在正演过程的建模中要尽可能选择最佳的描述方式以使得最终的模型空间足够小。

基于此观点，Monte Carlo 采样反演方法就显得非常直观。与梯度导引类方法不同，Monte Carlo 反演方法是基于一定的概率密度函数在所有可能的参数模型所形成的模型空间中进行采样，而不是基于梯度导引类方法在初始值附近对模型空间进行采样。Monte Carlo 方法对模型空间的采样所依据的概率密度函数体现了对模型的认识，这种认识主要来源于模型空间的先验信息。与梯度导引类方法相比，Monte Carlo 方法对初始值的依赖性要小很多，但无论是 Monte Carlo 随机采样方法还是梯度导引类采样方法，都应该是在 Bayes 估计的思想框架下进行的[25-26]。

因此，对反演问题最后结果的评价原则上应该是统计决策问题，主要评价依据是反演结果的均值和方差，其中方差的意义尤其重要，它反映了反演结果的可靠程度。

3 数据空间向参数空间的投影

3.1 地震波反演中的数据空间与模型空间

地震波反演的正演过程可以简单地描述为：由震源激发出来的地震波在由地球介质所组成的系统中进行传播，最后被检波器接收。在这一物理过程中，任何与地震波传播现象有关的介质参数都可以是地震波反演所估计的模型参数（如纵波速度 v_p，横波速度 v_s，吸收衰减系数 Q 等），这些参数就构成了反演问题的模型空间。这里用 $\mathbf{M}$ 表示模型空间，用 $\mathbf{m}$ 表示模型空间中的元素，那么在地球介质中传播后被检波器接收到的波场可以理解为震源激发的子波经过地球介质系统后的系统响应，它包含了地球介质系统的参数信息。因此，在接收到的波场中，与模型参数有联系的地震波场就构成了反演问题的数据空间，这里用 $\mathbf{D}$ 表示数据空间，用 $\mathbf{d}$ 表示数据空间中的元素。

在地震波反演问题中可以将模型空间 $\mathbf{M}$ 和数据空间 $\mathbf{D}$ 都视为线性空间，并且元素 $\mathbf{m}$ 和 $\mathbf{d}$ 的表达形式取决于线性空间中对基函数的选择。线性空间中对基函数的选择应该尽量遵循的原则是：线性空间中的元素 $\mathbf{m}$ 和 $\mathbf{d}$ 可以被稀疏地表达；联系元素 $\mathbf{m}$ 和 $\mathbf{d}$ 的算子可以被稀疏地表达，并且是稳定的。其中，第一个原则是为了降低地震波反演的规模。由于地震波反演问题中的模型空间经常是一个维数很高的线性空间，为了提高计算效率、降低地震波反演的规模，研究模型与数据的稀疏表达就显得非常必要；第二个原则是为了提高地震波反演的计算效率，由于梯度类的反演方法需要计算目标函数的梯度和步长，其中梯度的计算需要实现一次波场正演和一次波场反传播（本质上也是一次正演），步长的计算需要至少一次的波场正演，每一次模型的更新需要至少三次的正演计算，因此正算子的稀疏性会影响地震波反演每一次迭代的效率。又由于算子的稳定性会影响反演问题的适定性，最终体现在求解反演问题所需要的迭代次数。算子的稀疏性和稳定性决定了反演问题的计算效率，而算子的这 2 个原则同模型与数据的表达方式有很大关系，所以这也是模型空间 $\mathbf{M}$、数据空间 $\mathbf{D}$ 在选择基函数过程中必须考虑的一个因素。

3.2 数据空间与参数空间的联系

很显然，观测数据与模型参数的关系涉及到实际物理现象和数学建模这 2 个层次。

从物理层面上看，炮点激发的地震波经过地球介质系统后在地表被检波器所接收，在此物理过程中，观测数据是介质系统对激发的响应结果。在给定介质系统以及忽略观测数据中的随机噪音的前提下，观测数据是唯一的。此时从地球介质系统 $\mathbf{m}$ 到观测数据 $\mathbf{d}$ 的映射关系 $f:\mathbf{m}\mapsto\mathbf{d}$ 可视为一个单映射关系，但并不能保证这种映射关系的可逆性，这也是 FWI 不适定的最根本原因。通过上述物理过程可知，由地球介质系统组成的模型空间和由观测数据组成的数据空间联系紧密，只有联合二者才能够完整地描述地震波的传播过程。因此，FWI 既不能只靠在数据空间的逼近也不能只靠在模型空间的约束条件，而应该是在模型空

间和数据空间组成的联合空间$(\mathbf{m},\mathbf{d})$中根据对模型合理的约束条件，利用数据空间的逼近关系来求解反演问题的解。

但仅靠物理过程无法得到介质参数的分布结果，为此需要在数学层面上人为地引入正演算子把实测数据和介质参数系统联系在一起，当该正演算子满足一定的假设条件时，就可以通过正演算子的逆算子或伴随算子来估计介质参数模型。但这些正演算子在大部分情况下是不可逆的，因此只能根据模型空间的先验信息获得反演问题的一个初始解，再利用正演算子获得该初始解所对应的正演数据，并根据正演数据与观测数据的差异修改初始解。当正演数据与观测数据非常逼近时，可认为所用的初始解就是反演问题可以被接受的解。但这在概念上也存在问题，因为一个观测数据可能会对应无数的模型参数，而在一定约束条件下，只存在一个模型与观测数据相对应。前已述及，正演算子可以用各种褶积模型或利用第一类 Fredholm 积分方程或偏微分方程来表达，并建立介质参数（或是介质参数的扰动）与预测数据（或是观测数据）之间的关系，这种映射关系可以认为是实际物理现象的一种数学表达方式。在此数学建模过程中，需要对实际地球介质系统和物理过程做相应的假设，最后得到的模型空间是实际地球介质系统的某种抽象表达，数据空间的元素也是对实际观测数据的某种近似表达。这种抽象的、近似的表达与实际情况的差异主要由正演算子所提出的假设条件是否被满足所决定，正演算子的随机性在这里表现为实际情况随机地满足正演算子所要求的假设条件。

3.3 数据空间向参数空间的投影及投影的非线性分析

FWI 的根本问题还是观测的地震数据与模型参数之间的非线性映射关系，这导致将实测数据逆映射回模型空间时不是一一对应的，而是一对多，且极不稳定，观测数据中的微小扰动经常会导致逆映射结果的巨大变化。

已知实测数据（包括数据空间的元素）与实际介质模型（包括模型空间的元素）之间的关系是非线性的。非线性性的强弱不仅取决于介质参数的复杂性，还取决于描述这一物理过程的正算子。因此，非线性关系既取决于实际物理问题，又取决于描述该物理问题的数学形式，即非线性性的强弱在某种意义下是可以人为改变的。因为线性关系是一对一的，比较容易求解，所以常把一个非线性性很强的问题逐步线性化，用一系列线性问题来逼近原来的非线性问题。

3.4 反演问题解的不唯一性分析

反演问题解唯一存在的必要条件是代价函数$S(m)$在模型空间（或是解空间）中为严格凸，而$S(m)$的凸性主要由正算子的性质所决定。正算子定义了从模型空间到数据空间的映射关系，当这种映射关系为线性时，代价函数$S(m)$是一个非负的二次型函数（二次项系数非负），此时保证了$S(m)$的凸性。当正算子为线性时，正演过程$G(m)=d$可以简化为$G(m)=\mathbf{G}m=d$，其中$\mathbf{G}$是离散成矩阵形式的正演算子。正算子矩阵$\mathbf{G}$条件数的大小决定了代价函数$S(m)$凸性的好坏：矩阵$\mathbf{G}$的条件数越小（正算子越稳定），代价函数$S(m)$的凸性越好。当正算子被一线性函数近似时，其稳定性主要是指正算子矩阵$\mathbf{G}$的核空间为空集$A=\{m\,|\,\mathbf{G}m=0\}=\varnothing$，并且条件数比较小，这一性质保证了矩阵$\mathbf{G}$的逆是稳定存在的。因此，正算子的线性性保证了代价函数的凸性；正算子的稳定性保证了代价函数具有好的凸性，而正算子的线性性和稳定性决定了反演问题解的唯一性。

实际反演问题所对应的正算子经常是不能同时满足线性性和稳定性这 2 个条件的，这也是反演问题解不唯一或是不稳定的主要原因。以地震数据规则化为例，其对应的正算子（傅立叶算子）虽然是线性的，但由于其不稳定性（算子矩阵的条件数比较大），使得该反演问题的解也不唯一。对于全波形反演，正算子不仅是不稳定的，而且是非线性的，这就使得代价函数$S(m)$在解空间中的凸性更难得到保证。解决这一问题的有效方法是在原来模型空间的某个子空间M_s中研究$S(m)$，使$S(m)$在该子空间里具有比较好的凸性。如果$S(m)$在子空间M_s中具有比较好的凸性，那么$S(m)$可以用一个二次型函数来逼近，而相应的正算子也可以用一个线性函数来表示。这就保证了反问题在子空间M_s中的解是唯一的。之后再根据

相关的先验信息合理地扩展子空间 M_s 的范围，使求得的解是最初模型空间中合理的解。选择子空间 M_s 的方法很多，如加约束条件、分尺度反演、修改误差函数的形式、修改正算子等都可以理解为选择子空间的过程，但其原则只有一个，即保证 $S(m)$ 在子空间 M_s 中具有较好的凸性。

综上所述，反演问题解的不唯一性的最本质原因是代价函数 $S(m)$ 在模型空间 M 中不为严格凸或是凸性不好，$S(m)$ 该性质主要由正算子的非线性性、不稳定性所决定。解决这一问题的有效方法是限制解空间的范围，使得 $S(m)$ 在新的解空间中具有较好的凸性。分尺度反演、加约束条件、逐步线性化等方法都可以理解为运用不同的手段选择符合先验信息的子空间，它们获得的解只是反演问题在某个子空间中最合理的解。反演问题求解过程中运用的各种正则化方法和手段其实并未消除解的不唯一性，只是通过限制反演问题解空间的范围来一定程度地减弱反演问题解的不唯一性，此时求得的解应该理解为某种意义下的最佳解。

4 海上和陆上地震数据的特点分析

地震波反演需要将计算机的正演结果与野外实际采集到的地震数据进行比较，然后根据数据残差来更新地下模型参数的值。当计算机的正演结果可以重现野外的大部分波现象时，说明此做法是合理的。但如果野外采集的地震记录中有很多能量在正演算子中无法重现，或者由于简化正演算子而被忽略掉，那么仅靠数据残差进行模型参数的更正是不合理的。这是因为在这种情况下，导致数据残差的因素主要有模型参数的误差和正演算子的误差，而且这 2 个因素所引入的误差几乎是在同一个数量级上。因此，当野外的波现象比较复杂，正演算子无法重现野外大部分波现象时，地震波反演就得不到好的反演结果。

海上地震数据由于在均匀的海水中进行激发和接收，首先是各炮的气枪震源的子波特征相对比较一致，而且可以较容易地记录下来。同时，与海水面有关的噪音相对比较简单，包括鸣震、源鬼波和检鬼波、与海水面有关的长程多次波，对这些噪音的消除或者正演预测都相对比较容易。另外，目前海上宽方位、长拖缆采集方式也基本保证了高维全波形反演的空间采样需求。

陆上地震数据与海上地震数据相比，地表条件要复杂得多，导致观测的波场很难用一个比较简单的波动方程来进行预测。首先，陆上采集用的震源是炸药震源或可控震源，由炸药震源激发的子波会受很多因素影响，因此在陆上采集时很难得到一个各炮一致性比较好的子波，而子波的不一致性在地震波反演中会体现在数据残差上；可控震源的子波形态由扫描信号和经过地层滤波作用后反射的扫描信号之间的相关结果决定，其在地震波反演中的数学物理含义是不清楚的。其次，陆上探区近地表的高程变化和岩性变化导致了极其复杂的波现象，这些波现象是正算子无法重现的，这种误差会最终体现在数据残差上而影响反演迭代的收敛情况。这些无法重现的波现象虽然含有地表介质的信息，但很难被利用，在反演过程中反而成了噪音。由于传播距离较近，这部分波现象的能量经常较大，会在反演中影响每一次迭代过程。这些不能预测或不能较好预测的波现象也难以较好地消除。可以说，与地表有关的复杂波现象和各炮子波的不一致性是陆上地震波全波形反演的最大障碍。

综上所述，陆上数据在做速度反演前还有很多预处理工作需要完成。最主要的工作可以概括为去噪音和表层介质参数建模。在地震波反演中，有效信号应该是指那些包含了地层介质信息并且能够被正算子所描述的那部分波现象。其他观测到的反射能量在地震波反演中都应该被视为噪音，包括那些虽然包含地层介质信息，但却无法被正算子所描述的波现象。表层建模是为了消除复杂地表对反演的影响，避免反演的迭代过程陷入某个局部极值。

因此，海上数据的全波形反演相对于陆上数据而言要简单得多，这也是目前海上地震数据全波形反演有若干应用实例的原因。

5 地震波反演的多层次多尺度实现

到目前为止，地震波全波形反演 FWI 技术很难用于实际数据以得到符合地质实际的参数模型，或者用于逆时深度偏移成像（RTM）或进一步的储层油气预测过程中，其主要原因是实测数据波现象过于复杂、激发震源函数未知、正演模拟出的预测波场与实测波场差异过大、观测数据的照明角度不足。

5.1 梯度导引类反演方法的局部寻优特征

梯度导引类反演方法的主要步骤可以表述为：根据某种条件（如常见的共轭梯度或最速下降条件）计算模型更新的方向；计算模型沿更新方向的最佳更新量；根据模型的更新方向和更新量对模型进行更新，并将更新后的模型作为初始值以进行下一次迭代。从该类反演方法的思想可以看出，最终反演结果的好坏完全取决于每一次迭代过程中所选择的模型更新方向和更新量的大小。当反演问题建立的范函是一个严格凸函数时，梯度导引类方法的思想可以保证最终的迭代结果为全局极小值点。寻优的收敛速度由范函的凸性（或是算子矩阵的条件数）决定，尤其是当初始值特别接近真实值时，可以用 Newton 类方法加快寻优的收敛速度。但如果反演问题的范函不是一个严格凸函数，梯度导引类的反演方法可能无法找到合适的模型更新方向或是更新量，从而使迭代过程陷入某个局部极值范围，最终得到的迭代结果只能是范函的某个局部极值点。

地震波反演问题的误差范函是一个凸性非常差的函数，直接根据数据逼近以及用梯度导引类方法很难得到合理的反演结果。为了避免地震波反演在迭代过程中陷入某个局部极值范围，可以在建立范函的过程中加入合适的约束条件来限制解空间的范围，或者用分尺度、分层次的方法增加范函的凸性，如频率域的 FWI 就是一种分尺度反演方法。频率域 FWI 利用模型中线性性、凸性较好的低波数成分建立误差范函，先估计出该低波数成分，然后再估计模型中的高波数成分。其他各种对模型、数据成分分解的方法，只要能够有效地利用模型、数据的多尺度、多层次信息建立范函，增加范函的凸性和线性性，都可以帮助梯度导引类方法获得合理的反演结果。

5.2 地震波反演多层次实现的必要性

综上所述，完全依赖于数据逼近，反演问题的非唯一性会很强；完全依赖先验信息，就没有了反演的必要。非线性最小二乘泛函反演理论指出：只有泛函的凸性比较好时，利用梯度导引的最优化算法就可以得到比较可靠的反演结果。

实际上，全波形反演的泛函凸性是相当差的，只有给出比较合适的初始模型才能改善泛函的凸性。另外，还可以通过修改泛函形式，如用旅行时误差泛函来改善泛函凸性。但是此做法就不是理论意义下的全波形反演。

地震波反演的目的是得到一个宽频带的介质参数场，既包括比较低频的光滑部分，也包括中高频的快速变化部分。Mora [6]指出 Inversion = Tomography + PSDM，这是有很明确的地球物理含义的。PSDM 估计角度反射系数，反映了反射界面两边速度或波阻抗的跃变；Tomography 估计光滑的背景速度场 [27]。Inversion 可以给出宽频带的介质参数场。理论上，上述说法毋庸置疑，但对于频带有限、观测孔径有限的实际地震数据，宽频带的模型参数场的估计还应考虑反演策略（其中包括很多正则化思想与方法）。

实际上，无论是 Tomography 还是 PSDM 都需要一个比较正确的、有一定背景的介质参数场，但通常用比较粗糙的方法来估计介质参数的背景变化。这说明全波形反演不是一种简单的算法，而是一个综合性的全频带介质参数场的反演流程，需要多层次地实现 [28]。

笔者提出的宽频带速度场估计的多层次实现方案为：①各种速度分析方法提供初始速度场；②地表初至波 Tomography（层析成像）+角度道集 Tomography（层析成像）建立更精细的初始速度场；③全波形反演估计更宽频带的比较精确的速度场（如果继续进行弹性参数估计，可以继续下面层次的步骤）：④在观测数据比较理想的情况下，RTM 提供保真的角度成像道集；⑤零角度或小角度的成像剖面结合测井数据进行 1 维或 1.5 维粘声介质波阻抗反演；⑥中、大角度成像道集结合测井数据进行 AVA 弹性参数反演；⑦与岩石物理学结合进行储层描述。

目前，全波形反演的定位仅仅是为 RTM 提供一个更为精确的偏移速度场，还没有达到通过 FWI 解决储层参数描述的程度，但这并非是说 FWI 理论方法没有这样的能力，而是说各种因素限制了 FWI，使其不能直接用于储层描述。

5.3 地震波反演多尺度实现的必要性

为了把 FWI 这种勘探地震学中的领先技术应用于实际，我们提出的方法技术的核心是特征波场反演（Characteristic Waveform Inversion，简写为 CWI）。这种方法的核心思想是根据正算子重新筛选模型空间、

数据空间的基函数，用此基函数重新表达模型空间、数据空间和正演算子，然后在新的基函数下研究地震波的传播现象和反演问题，其主要目的是降低数据空间与模型空间之间的非线性性。

降低数据与模型参数之间的非线性性的主要方法可以归纳为：

（1）利用特征波场。特征波场是根据正算子的形式对波场新的描述方式，特征波场应该是波场中对模型参数的扰动反应比较敏感的那部分波场信息，因此在特征波场中求解逆算子可以降低逆算子核空间的维数，从而降低反演问题的不适定性和非线性性。

（2）修改误差泛函。不同的泛函形式定义了不同的空间度量关系，因此也会改变代价函数$S(\mathbf{m})$相应的几何特征，从而改变反演问题非线性性的强弱。

（3）利用关于模型的先验信息。这其实是根据模型的先验信息来约束解空间的范围，在模型空间的一个子空间中研究反演问题，并且希望正算子在该子空间里是线性、可逆的。

根据上述方法，首先在反演的初始阶段，不宜用基于全波形逼近的误差泛函，而应该用基于特征波场逼近的误差泛函；另外要充分利用模型的先验信息，不能希望仅依靠预测数据和叠前数据的逼近就可以得到宽波数带的参数模型，在噪音较大、正算子误差较大的情况下，仅基于数据拟合给出的反演结果的地质含义可能不大。

所谓的特征波场可以是（单）低频数据、空间局部数据、单向波数据等。值得注意的是，旅行时场也是一种特征数据。不同的特征波场需要用不同的正算子来描述，如（单）低频数据对应频率空间域的正演算子，空间局部数据（局部平面波场）对应 Beam 波算子，单向波数据对应单向波算子等等。这些特征波场实质上对应了不同参数模型的描述方式，用来表达波场对参数模型在不同描述方式下的扰动的响应，如低频数据对应低波数的、光滑的参数变化（如速度场变化）。特征波反演首先要把大尺度的、可靠性比较高的参数反演出来（或确定下来），然后再根据实际情况，如数据的观测情况、数据信噪比、模型的复杂程度等，再决定是否能进行更高波数成分的参数变化成分的反演。

高波数的参数变化与数据之间的关系是高度非线性的，需先把低波数参数变化反演出来再充分利用关于参数模型的先验信息外，否则难以准确反演出来。偏移成像给出的反射界面的几何结构信息和反射系数信息是应该充分利用的关于参数模型的先验信息，很多参数模型的高波数变化信息都包含在其中。为此，笔者提出把零角度反射系数成像剖面上的一维波阻抗反演和（或）基于反射界面的几何结构信息的AVA反演结果与CWI/FWI反演结果结合在一起，以得到尽可能宽波数带的参数模型。期望这样的参数模型能有助于刻画储层的性质，这将是勘探地震学进入储层地震学的较优途径。

6 结论与讨论

全波形反演是油气勘探领域中的引领性思想、方法和技术。在地震数据处理和成像整个流程中，压噪音、数据规则化、反褶积、一维波阻抗反演（一维全波形反演）、速度分析、层析成像等关键技术环节都可以表达在 Bayes 估计理论反演框架下。

得益于目前野外宽方位、高密度、宽频带地震数据观测技术的进步和高性能计算机技术的进步，高维全波形反演又一次成为勘探地震领域的热点技术。但是，高维全波形反演不会像逆时深度偏移成像技术那样迅速地发展成熟并应用于实际，其实际实用会是一个漫长的过程。在此过程中，勘探地震学中地震数据处理与成像的理念、储层描述的理念等都会随着对高维全波形反演认识的深化而发生质的变化。

笔者从概率论的观点对反演问题的本质进行了分析，指出了全波形反演可用于海上和陆上观测数据，并认为仅依靠地震数据全波形的逼近建立误差泛函很难得到可满足于 RTM 偏移成像的背景速度场，其原因在于数据与模型参数之间的关系是高度非线性的，而要得到比较可靠的反演结果必须降低这种非线性性。因此，比较合理的观点应该是：利用特征波场进行反演；修改误差范函定义式、充分利用地震波的相位信息；充分考虑初始模型的先验信息。这些观点应用于实际工作时即为：做特征波场反演（CWI 反演）而不是全波形反演（FWI 反演）；把地震数据和参数模型视为随机信号，在建立误差范函的过程中考虑其均值、方差等统计学特征；不是仅仅依赖观测数据的反演而是充分考虑反演参数的先验信息的综合反演，利用先验信息约束解空间的范围。

7 参考文献

[1]. Tarantola，A.Inverse problem theory and methods for model parameter estimation[M].SIAM，2005. Society for Industrial and Applied Mathematics

[2]. Tarantola A.Inversion of seismic reflection data in the acoustic approximation[J].Geophysics，1984，49(8)：1259-1266.

[3]. Bleistein N，Cohen J，Hagin F.Two and one-half dimensional Born inversion with an arbitrary reference[J].Geophysics，1987，52（1）：26-36.

[4]. Clayton R W，Stolt R H.A Born-WKBJ inversion method for acoustic reflection data[J].Geophysics，1981，46（11）：1559-1567.

[5]. Cohen J K，Hagin F G.Velocity inversion using a stratified reference[J].Geophysics，1985，50（11）：1689-1700.

[6]. Mora P R.Nonlinear two-dimensional elastic inversion of multi-offset seismic data[J].Geophysics，1987，52（9）：1211-1228.

[7]. Pratt R G.Inverse theory applied to multisource cross-hole tomography，Part II: Elastic wave-equation method[J].Geophysical Prospecting，1990，38（3）：311–330.

[8]. Pratt R G，Shin C，Hicks G J.Gauss-Newton and full-Newton methods in frequency-space seismic waveform inversion[J].Geophysical Journal International，1998，133（2）：341-362.

[9]. Pratt R G.Seismic waveform inversion in the frequency domain，Part I: Theory and verification in a physical scale model[J].Geophysics，1999,64（3）：888–901.

[10]. Bunks C，Saleck F M，Zaleski S，et al.Multiscale seismic waveform inversion[J].Geophysics，1995，60（5）：1457-1473.

[11]. Shin C，Min D J.Waveform inversion using a logarithmic wavefield[J].Geophysics，2006，71（3）：R31-R42.

[12]. Shin C，Cha Y H.Waveform inversion in the Laplace domain[J].Geophysical Journal International，2008，173(3)：922–931.

[13]. Shin C，Cha Y H.Waveform inversion in the Laplace-Fourier domain[J].Geophysical Journal International，2009，177（3）：1067–1079.

[14]. Plessix R E，Chavent G，Roeck Y H，et al.Waveform inversion of reflection seismic data for kinematic parameters by local Optimization[J].SIAM Journal of Scientific Computing, 1999，20（3）：1033-1052.

[15]. Plessix R E. Areview of the adjoint-state method for computing the gradient of a functional with geophysical applications[J].Geophysical Journal International，2006，167（2）：495-503.

[16]. Symes W W.A differential semblance criterion for inversion of multioffset seismic reflection data[J].Journal of Geophysical Reservoirs，1993，98（B2）：2061-2073.

[17]. Symes W W.High frequency asymptotics, differential semblance, and velocity estimation[J].SEG 68th Annual Meeting Expanded Abstracts，1998，68th,1616-1619.

[18]. Symes W W.Approximate linearized inversion by optimal scaling of prestack depth migration[J].Geophysics，2008，73（2）：R23-R35.

[19]. Symes W W.Migration velocity analysis and waveform inversion[J]. Geophysical Prospecting，2008，56（6）：765-790.

[20]. de Hoop，M V，Uhlmann G.Characterization and‘source-receiver’continuation of seismic reflection data[J].Communications on Mathematical Physics，2006,263(1):1-19.

[21]. de Hoop M V，van der Hilst R D. On sensitivity kernels for ‘wave-equation’ transmission tomography[J].Geophysical Journal International，2005，160(2):621-633.

[22]. de Hoop M V，van der Hilst R D，Shen P.Wave-equation reflection tomography： annihilators and sensitivity kernels[J].Geophysical Journal International，2006，167(3)：1332-1352.

[23]. de Hoop M V，Smith H，Uhlmann G，et al.Seismic imaging with the generalized Radon transform：a curvelet transform perspective[J].Inverse Problems，2009，25：025005.

[24]. Tromp J，Tape C，Liu Q.Seismic tomography, adjoint methods，time reversal and banana-doughnut kernels[J].Geophysical Journal International，2005，160（1）：195-216.

[25]. Jin S，Madariaga R.Background velocity inversion with a genetic algorithm[J]. GEOPHYSICAL RESEARCH LETTERS.，1993，20(2)：93-96.

[26]. Jin S，Madariaga R.Nonlinear velocity inversion by a two-step Monte Carlo method[J].Geophysics，1994，59（4）：577-590.
[27]. Nolet G.Seismic tomography:with applications in global seismology and exploration geophysics[M]. Geophysics Reidel Publishing Company，Dordrecht Holland，1987.
[28]. Varela C L，Stoffa P L，Sen M K.Background velocity estimation using nonlinear optimization for reflection tomography and migration misfit[J].Geophyscal Prospecting，1998，46(1)：51-78.

地震波反演问题理论分析

王华忠 王雄文

波现象与反演成像（WPI），同济大学海洋与地球科学学院，上海，200092

【摘要】全波形反演是目前勘探地震学中的引领性技术。它涉及到波传播的模拟预测问题、地震波反演理论框架问题、线性化反演问题、非线性反演问题、梯度计算问题、正则化问题。考虑到地震信号的随机性，它的高斯性与线性性、非高斯性与非线性性也是值得关注的问题。本文对上述问题进行了针对地震波全波形反演的简要分析。最后提出了我们认为应该遵循的路线，使地震波全波形反演逐步走向实用化。本文的理论分析工作目的是为开展地震波全波形反演提供理论基础。

【关键词】地震波全波形反演，波场预测，参数估计，线性性与非线性性，高斯与非高斯性，正则化

引言

地震波反演理论、方法与技术是勘探地震学，也是整个地球物理学的核心问题，它具有一整套完整的理论和方法。尽管勘探地震学并没有完全按照地震波反演理论的框架进行地震数据的反演成像，而是发展了一套由各个处理环节形成的流程来对地震数据进行成像处理，这一点也没有降低地震波反演理论的重要性。

地震波反演理论的核心内容包括如下几个层次。第一个层次的问题是利用概率论的观点在 Bayesian 估计理论框架下建立评价反演结果的准则，回答用什么信息进行地震波反演以及地震波反演的理念是什么等问题；另外，误差泛函建立的理论与方法解决了 Bayesian 框架的实用化问题。第二个层次的问题是误差泛函的最优化求解，包括线性优化问题和非线性优化问题的求解。前者主要解一个法方程，后者解一个梯度导引的迭代优化问题。第三个层次的问题是正则化问题。由于地震波反演问题的解是非唯一的，因此回答如何得到我们所期望的解，或如何得到有实际意义的解是很重要的。地震波反演理论的这三个层次是互相影响、互相关联的。

撇开理念的层次，更具体的讲，地震波反演是利用模型空间、数据空间以及两个空间的映射关系，根据实测数据（位于数据空间），利用映射的共轭算子把数据空间中的元素投影到模型空间实现对模型参数的估计。很显然，这一过程首先涉及到的是对模型的描述，通过合理的模型描述方式建立起模型空间（所有可能的模型组成的空间），合理的模型描述方式应该是在变换域中用特征值所对应的特征向量来稀疏表达模型。通过网格剖分来表达模型是比较差的（但是是最简单、最常用的）模型描述方法，这种对模型的描述方式会形成一个巨大的模型空间，对反演问题的求解非常不利。对数据的描述方式在地震波反演理论中也占有重要的地位，不同的数据描述方式代表着不同的逼近方式。目前用得比较多的数据描述方式有网格化方法、平面波法、局部平面波法、Blended-data 等。然后是两个空间的投影算子的描述，通俗地讲就是地震波正演算子的描述，不同的正演算子代表了对波现象不同的描述方式。地震波正演假设已知震源函数、模型，通过数值方以求解一个以波动方程为控制方程的定解问题，得到数据空间的一个投影值。与模型的网格剖分相对应的有限差分方法是常用的地震波正演方法。实际上，我们常常把模型空间中的元素与数据空间中的元素之间的映射关系假设成线性的，用 Fredholm 第一类积分方程来描述。很显然，模型的描述方法与地震波正演方法是紧密联系在一起的。至于数据空间中的元素，主要是由采集到的随机信号组成。一般认为地震数据是由确定性的反射信号加上随机噪音构成的。反射信号是可以由正问题预测的，随机噪音是不可以预测的。实际上，由于波传播过程的复杂性，任何正问题都仅仅是物理问题的一种近似，不可能完全模拟复杂的波现象。不能模拟的部分，实际上也归入了噪音部分。而噪音的概率分布特征决定了建立误差泛函的形式，这正是反演的核心部分，也是被经常忽略的部分。在很多反演实践中，无论是一维波阻抗反演，还是高维的 FWI 反演，我们很少关心噪音到底是符合什么的分布规律。原因在于我们默认噪音是满足高斯分布的。为了简单，我们不对后面的很多参考文献进行逐个评价，但所有参考文献的思想都没有脱离上面的描述。

在上述具体层次的基础上，我们在后面各节讨论了地震波正演的理论表达；地震波反演问题的理论表达；线性与非线性情形下的梯度计算问题；线性反演问题；非线性反演问题；正则化参数的选择问题；以及高斯型和线性性以及非高斯性及非线性性在反演理论中的地位。总之，试图把地震波反演理论用具体的数学语言来表达，为地震波反演的具体化奠定基础。

1 地震波正演的理论表达

1.1 基本声波方程

一般地,我们假设波在地下介质中的传播可以由如下的声波方程控制，

$$\left[\frac{1}{c(\mathbf{x})^2}\frac{\partial^2}{\partial t^2}-\rho(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla\right)\right]u(\mathbf{x},t)=S(\mathbf{x},t) \tag{1.1}$$

其中，$u(\mathbf{x},t)$为声场势，它是时间$t\in R$和空间坐标$\mathbf{x}\in\Omega\subseteq R$的函数。$c(\mathbf{x})$代表速度场；$\rho(\mathbf{x})$代表密度场；$S(\mathbf{x},t)$代表体积场源。我们也可以用声压来表示声场，它也满足（1）式定义的波动方程，此时声源函数变为$\partial S/\partial t$。注意，质点速度为$\rho^{-1}\nabla u$。

1.2 基本声波方程的线性化

定义背景介质中的声波传播满足如下方程，

$$\begin{aligned}L_0u_0&\equiv\left(\nabla\cdot\frac{1}{\rho_0}\nabla+\frac{\omega^2}{\kappa_0}\right)u_0(\vec{x},\vec{x}_s,\omega)\\&=\delta(\vec{x}-\vec{x}_s)S(\omega)\end{aligned} \tag{1.2}$$

借助于恒等式$L=L_0+L_0\left(L_0^{-1}-L^{-1}\right)L$和定义$V=L_0^{-1}-L^{-1}$可以得到

$$L=L_0+L_0VL \tag{1.3}$$

在频率域中，方程（1.2）右端的震源函数是一个脉冲。因此，算子L是一个 Green 函数。它描述所有的波传播现象。L^{-1}和L_0^{-1}是“逆”Green 函数，它们经常被各自的共轭转置来代替。物理上，它们描述的是波的反传播现象。Green 函数可以描述从源到散射点，再到接收点的波传播；同样地，也可以描述从接收点到散射点，再到源点的波传播过程。假如波传播算子L用 Green 函数G表示，方程（1.3）可以重写为

$$G=G_0+G_0VG \tag{1.4}$$

方程（1.4）是一个非线性的，它能被进一步重写为

$$G=\left(I-G_0V\right)^{-1}G_0 \tag{1.5}$$

对上式的右端项进行 Taylor 展开，

$$G=\left[\sum_{j=0}^{\infty}\left(G_0V\right)^j\right]G_0. \tag{1.6}$$

方程（1.4）、（1.6）被称为 Lippmann-Schwinger 方程。很显然，假如$j\le 2$,方程（1.6）描述了一个忽略了二阶以上散射被的波传播过程。因此，线性化的波传播算子仅仅描述波传播过程中的一阶散射。这就是

$$G=G_0+G_0VG_0 \tag{1.7}$$

上式就是 Born 近似下的波传播算子，其中方程右端的第一项描述了背景场的传播现象，第二项描述了一阶散射波的传播现象。方程（1.6）和（1.7）清楚地说明了 Born 近似的物理含义。为了具体地定义散射势V，根据G及G_0的表达式可以得到如下式子，

$$V=\left(\nabla\cdot\frac{1}{\rho}\nabla+\frac{\omega^2}{\kappa}\right)-\left(\nabla\cdot\frac{1}{\rho_0}\nabla+\frac{\omega^2}{\kappa_0}\right)$$

$$
\begin{aligned}
&= \nabla \cdot \left(\frac{1}{\rho} - \frac{1}{\rho_0} \right) \nabla + \omega^2 \left(\frac{1}{\kappa} - \frac{1}{\kappa_0} \right) \\
&= \nabla \cdot \frac{1}{\rho_0} \left(\frac{\rho_0}{\rho} - 1 \right) \nabla + \omega^2 \frac{1}{\kappa_0} \left(\frac{\kappa_0}{\kappa} - 1 \right) \\
&= \nabla \cdot \frac{a_1}{\rho_0} \nabla + \omega^2 \frac{a_2}{\kappa_0}
\end{aligned} \tag{1.8}
$$

其中 $a_1 = \dfrac{\rho_0}{\rho} - 1 = \dfrac{\Delta\rho}{\rho}$; $\Delta\rho = \rho_0 - \rho$; $a_2 = \dfrac{\kappa_0}{\kappa} - 1 = \dfrac{\Delta\kappa_0}{\kappa}$ 和 $\Delta\kappa = \kappa_0 - \kappa$。从（1.7）式可以看出，Born 近似线性化模拟的波场包括两部分：第一部分是由背景 Green 函数所描述的背景波场；第二部分是由散射势引起的方程（1.7）中的第二项所描述的散射场。

根据方程（1.7），线性化的总波场可以写成

$$
\begin{aligned}
&u\left(\vec{x}_r, \vec{x}_s, \omega\right) = u_0\left(\vec{x}_r, \vec{x}_s, \omega\right) + \\
&\omega^2 \int d^3x G_0\left(\vec{x}_r, \vec{x}, \omega\right) a\left(\vec{x}\right) u\left(\vec{x}_r, \vec{x}_s, \omega\right)
\end{aligned} \tag{1.9}
$$

Born 近似后的散射波场为

$$
u_s\left(\vec{x}_r, \vec{x}_s, \omega\right) = \omega^2 \int d^3x G_0\left(\vec{x}_r, \vec{x}, \omega\right) a\left(\vec{x}\right) u_0\left(\vec{x}_r, \vec{x}_s, \omega\right) \tag{1.10}
$$

Born 近似的物理实质是，在背景场 $u_0\left(\mathbf{x}_r, \mathbf{x}_s, \omega\right)$ 中没有散射场，在扰动场中，只存在一次散射场。尽管用这种方式描述波传播与实际的波传播不同，但当二阶及二阶以上的散射场能量比较弱时，用 Born 近似进行 PSDM 和 Tomography 还是比较有效的。

1.3 高频近似下的 Born 近似解释

Born 近似下的散射场可以表示为，

$$
u_s\left(\vec{\mathbf{x}}_r, \vec{\mathbf{x}}_s, \omega\right) = \omega^2 \int d\mathbf{x} G_0\left(\mathbf{x}_r, \mathbf{x}, \omega\right) a\left(\mathbf{x}\right) G_0\left(\mathbf{x}_r, \mathbf{x}_s, \omega\right) \tag{1.11}
$$

其中，速度扰动为 $a\left(\mathbf{x}\right) = 2c_0^{-3}\left(\mathbf{x}\right)\delta c\left(\mathbf{x}\right)$。用渐进 Green 函数（即 Green 函数的 WKBJ 近似解）代入上式，

$$
u_s\left(\mathbf{x}_r, \mathbf{x}_s, \omega\right) = \omega^2 \int d\mathbf{x} \begin{pmatrix} a\left(\mathbf{x}\right) A\left(\mathbf{x}, \mathbf{x}_s\right) A\left(\mathbf{x}_r, \mathbf{x}\right) \\ \exp\left[-i\omega\left(T\left(\mathbf{x}, \mathbf{x}_s\right) + T\left(\mathbf{x}_r, \mathbf{x}\right)\right)\right] \end{pmatrix} \tag{1.12}
$$

在散射点周围，对旅行时进行线性化，即围绕中心射线进行 Taylor 展开，

$$
T\left(\mathbf{x}, \mathbf{x}'\right) \approx T\left(\mathbf{x}_0, \mathbf{x}'\right) + \left(\mathbf{x} - \mathbf{x}_0\right) \cdot \mathbf{p}\left(\mathbf{x}_0, \mathbf{x}'\right) \tag{1.13}
$$

其中 $\mathbf{p}\left(\mathbf{x}_0, \mathbf{x}'\right)$ 表示旅行时场 $T\left(\mathbf{x}, \mathbf{x}'\right)$ 在点 $\mathbf{x}_0$ 处的梯度，其方向为中心射线的传播方向。因此，（1.12）式中的指数部分可以化为，

$$
\begin{aligned}
&\exp\left[-i\omega\left(T\left(\mathbf{x}, \mathbf{x}_s\right) + T\left(\mathbf{x}_r, \mathbf{x}\right)\right)\right] \\
&= \exp\left[-i\omega\left(T\left(\mathbf{x}_0, \mathbf{x}_s\right) + T\left(\mathbf{x}_r, \mathbf{x}_0\right)\right)\right] \\
&\exp\left[-i\omega\left(\mathbf{x} - \mathbf{x}_0\right)\left(\mathbf{p}\left(\mathbf{x}_0, \mathbf{x}_s\right) + \mathbf{p}\left(\mathbf{x}_r, \mathbf{x}_0\right)\right)\right] \\
&= \exp\left[-i\omega\left(T\left(\mathbf{x}_0, \mathbf{x}_s\right) + T\left(\mathbf{x}_r, \mathbf{x}_0\right)\right)\right] \\
&\exp\left[-i\left(\mathbf{x} - \mathbf{x}_0\right)\mathbf{k}\right]
\end{aligned} \tag{1.14}
$$

其中，

$$\mathbf{k}=\omega\left(\mathbf{p}\left(\mathbf{x}_0,\mathbf{x}_s\right)+\mathbf{p}\left(\mathbf{x}_0,\mathbf{x}_r\right)\right) \tag{1.15}$$

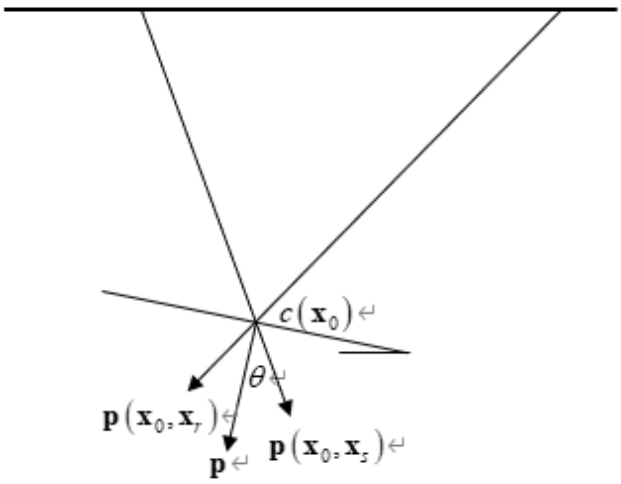


图 1 地震波传播过程中波矢量在反射点处的几何关系

如图可知：$\mathbf{p}=\omega\left(\mathbf{p}\left(\mathbf{x}_0,\mathbf{x}_s\right)+\mathbf{p}\left(\mathbf{x}_0,\mathbf{x}_r\right)\right)=2\mathbf{n}\cos\theta/c\left(\mathbf{x}_0\right)$，其中 $\mathbf{n}$ 为界面的外法线单位方向向量。θ 为射线的入射角。$c\left(\mathbf{x}_0\right)$ 为 $\mathbf{x}_0$ 处的速度。根据上面的几何关系则有，

$$\mathbf{k}=\omega\mathbf{p}=\frac{2\omega\cos\theta}{c\left(\mathbf{x}_0\right)}\mathbf{n} \tag{1.16}$$

从 $\mathbf{k}$ 的表达式中可以看出，确定 $\mathbf{k}$ 的低波数部分，需要大角度散射波和低频数据；确定 $\mathbf{k}$ 的高波数部分，需要小角度散射波和高频数据。这与全波形速度反演得到认识是一致的。

1.4 地震波反演中正问题的简洁表达

非线性正问题的数学描述可以抽象为

$$A\left(\mathbf{m}\right)\mathbf{u}=\mathbf{f} \tag{1.17}$$

其中，A 为正演算子。在地震波反演中，A 的形式为

$$A=\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)=\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\Delta \tag{1.18}$$

A 的形式受所选择的波动方程的形式的影响。可以选常密度声波，变密度声波方程，各种拟声波方程，弹性波方程等。$\mathbf{u}$ 为方程 1.17 要求解的未知量，表示正演合成的波场，$\mathbf{f}$ 为震源函数。

野外观测到的地震数据可以表述为

$$\mathbf{d}=C\mathbf{u}+\boldsymbol{\eta} \tag{1.19}$$

此处，C 表示地震数据的采样算子，由观测系统决定。它使得连续的波场 $u\left(x,y,z,t\right)$ 变得规则离散化（或非规则离散化）及孔径有限化。采样算子的状态受实际条件的制约。η 表示噪音函数，在大部分情况下假设它是高斯白噪的。

连续波场受采样算子制约后，采样后的数据可以记为

$$\left[C\mathbf{u}\right]\approx\mathbf{u}\left(\mathbf{x}_i\right),\quad i=1,2,\cdots,n \tag{1.20}$$

其中 $\mathbf{x}_i$ 表示第 i 个采样点的坐标。根据方程 1.17，可以将连续波场的正演计算形式地表达为

$$\mathbf{u}=\left(A(m)\right)^{-1}\mathbf{f} \tag{1.21}$$

2 地震波反演问题的数学表达

2.1 最佳数据逼近意义下的反演问题

基于前面对正问题的定义，根据数据残差最小的原则可以建立如下的优化问题：

$$\begin{aligned}&\min:\quad J(\mathbf{m})\\&S.T.:\quad C\mathbf{u}=\mathbf{d}\end{aligned} \tag{2.1}$$

可以用如下的方式定义上面优化问题所对应的误差泛函，

$$\min_{\mathbf{u}\in\mathcal{H}_2,\mathbf{m}\in\mathcal{H}_1}\frac{1}{2}\left\|C\mathbf{u}-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{2.2}$$

上式中的第一项代表正演波场与观测波场的波场残差，后一项代表对模型的先验约束，α 是正则化系数。一般地，我们可以通过假设模型中的元素满足某种概率分布条件来施加对模型的先验限制。把（1.21）式代入上式，得到

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\Im(\mathbf{m})=\frac{1}{2}\left\|F(\mathbf{m})-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{2.3}$$

其中，与观测数据逼近的正演计算结果可以表达为

$$F(\mathbf{m})=C\left(A(\mathbf{m})\right)^{-1}\mathbf{f} \tag{2.4}$$

将上述结果代入（2.3）式后，得到最后的泛函定义式，

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\Im(\mathbf{m})=\frac{1}{2}\left\|C\left(A(\mathbf{m})\right)^{-1}\mathbf{f}-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{2.5}$$

通过此式，可以通过各种优化方法求解参数估计问题。假设上述定义的泛函至少是 Fréchet 二阶可微的，那么，可以对该泛函做如下的 Taylor 展开，

$$\begin{aligned}\Im(\mathbf{m}+\delta\mathbf{m})&=\Im(\mathbf{m})+\left\langle \mathrm{Grad}\Im(\mathbf{m}),\delta\mathbf{m}\right\rangle\\&+\frac{1}{2}\left\langle \mathrm{Hess}\Im(\mathbf{m})\delta\mathbf{m},\delta\mathbf{m}\right\rangle+o\left(\left\|\delta\mathbf{m}\right\|^3\right)\end{aligned} \tag{2.6}$$

如果泛函 $\Im(\mathbf{m}):\mathbb{R}^n\to\mathbb{R}$ 有连续的二阶偏导数，那么 $\mathrm{Grad}\Im(\mathbf{m})$ 是一个 n 维向量，表示泛函的梯度方向，$\mathrm{Hess}\Im(\mathbf{m})$ 是一个 $n\times n$ 的方阵，称为 Hessian 矩阵。其中，n 为模型参数的个数，m 为观测数据的个数。

定理：如果泛函 $\Im(\mathbf{m}):\mathbb{R}^n\to\mathbb{R}$ 在 $\mathbf{m}_0$ 点的某个邻域 $\Omega(\mathbf{m}_0,\delta)$ 内同时满足以下几个条件：1. $\Im(\mathbf{m})$ 在该邻域内是 Fréchet 二阶可微的；2. 梯度 $\mathrm{Grad}\Im(\mathbf{m}_0)=0$；3. Hessian 矩阵 $\mathrm{Hess}\Im(\mathbf{m}_0)$ 是强正定的，那么 $\mathbf{m}_0$ 是泛函 $\Im(\mathbf{m}):\mathbb{R}^n\to\mathbb{R}$ 的严格极小点。

2.2 最大后验概率意义下的反演问题

上节通过最佳数据逼近的方式推导出反演问题所对应的误差范函。另外一种获得反问题误差范函的方式是利用概率论的观点求解观测数据所对应的后验概率分布，通过求取最大后验概率密度的方式推导得到误差范函的表达形式。

假设模型空间和数据空间的元素均满足高斯概率分布条件，则它们的概率分布函数可以由它们的均值和方差唯一决定。用 $\mathbf{m}_0$、$\mathbf{d}_0$ 和 $C_{\mathbf{M}}$、$C_{\mathbf{D}}$ 分别表示它们的均值和方差，则模型空间和数据空间任意元素的

概率密度可以表示为：

$$\begin{aligned}\rho_M(\mathbf{m}) &= const_M \exp\left\{-\frac{1}{2}(\mathbf{m}-\mathbf{m}_0)^H C_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_0)\right\}\\ \rho_D(\mathbf{d}) &= const_D \exp\left\{-\frac{1}{2}(\mathbf{d}-\mathbf{d}_0)^H C_{\mathbf{D}}^{-1}(\mathbf{d}-\mathbf{d}_0)\right\}\end{aligned} \tag{2.7}$$

其中 $const_M$ 、 $const_D$ 分别表示两个空间概率密度函数的归一化系数。模型空间和数据空间的元素并不是相互无关的，它们通过正算子联系在一起。正算子是对实际物理现象的一种近似描述，并不等同于物理现象本身，因此也具有一定的随机性。但由于正算子的随机性可以和数据空间元素的随机性一起讨论，所以一般情况下认为正算子是没有误差的，即正算子是一个确定的算子，模型空间、数据空间的元素满足下面的映射关系：

$$A(\mathbf{m}) = \mathbf{d} \tag{2.8}$$

根据式子（2.7）、（2.8），由模型空间和数据空间所组成的联合空间 $(\mathbf{m},\mathbf{d})$ 的边缘概率密度可以表示为：

$$\begin{aligned}\sigma_M(\mathbf{m}) &= \frac{\rho_M(\mathbf{m})\rho_D(A(\mathbf{m}))}{\int_M d\mathbf{m}'\rho_M(\mathbf{m}')\rho_D(A(\mathbf{m}'))} = \frac{\exp\{-S(\mathbf{m})\}}{Const}\\ S(\mathbf{m}) &= \frac{1}{2}\begin{pmatrix}(A(\mathbf{m})-\mathbf{d}_0)^H C_{\mathbf{D}}^{-1}(A(\mathbf{m})-\mathbf{d}_0)+\\ (\mathbf{m}-\mathbf{m}_0)^H C_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_0)\end{pmatrix}\end{aligned} \tag{2.9}$$

上面式子中右端项的分母是一个常数。（2.9）式将反问题的求解转化为边缘概率密度函数 $\sigma_M(\mathbf{m})$ 最大值的求解。 $\sigma_M(\mathbf{m})$ 取得最大值的充要条件为 $S(\mathbf{m})$ 取得最小值。如果正算子 $A(\mathbf{m}) = A\mathbf{m}$ 是一个线性算子且满秩，则函数 $S(\mathbf{m})$ 是一个正定的二次型函数，则该函数只存在一个极小值点，且在极小值点处满足条件： $\partial S(\mathbf{m})/\partial\mathbf{m} = 0$ 。根据该条件可以得到 $S(\mathbf{m})$ 极小值点的计算公式如下，

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + \left(A^H C_D^{-1} A + C_M^{-1}\right)^{-1} A^H C_D^{-1}(\mathbf{d} - A\mathbf{m}_0) \tag{2.10}$$

如果模型空间、数据空间的元素是相互独立的，则协方差矩阵 C_M 、 C_D 退化为一个对角矩阵；如果模型空间、数据空间的每个元素的方差相等，则协方差矩阵可以表示为：

$$C_M = \varepsilon_M I, \quad C_D = \varepsilon_D I \tag{2.11}$$

将（2.11）代入到（2.10）可以得到极小值点的表达如下，

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + \left(A^H A + \alpha I\right)^{-1} A^H(\mathbf{d} - A\mathbf{m}_0) \tag{2.12}$$

在地震波反演中，该方程右端第一项为初始速度，第二项为残留波场的最小二乘偏移结果，其中 A^H 为偏移算子， $\left(A^H A + \alpha I\right)^{-1}$ 为 Hessian 矩阵的逆。

如果正算子 $A(\mathbf{m})$ 不是一个线性算子，但 $S(\mathbf{m})$ 是个凸函数，那么 $S(\mathbf{m})$ 的极小值点也是唯一的，可以通过迭代算法求解该极小值点，该极小值点没有如（2.10）那样的解析表达式。

从上面的分析可以看出，如果模型空间、数据空间和正算子满足以下的条件：1. 模型空间、数据空间的元素满足高斯概率分布条件，且不同元素之间相互独立；2. 不同元素的方差值相等；3. 正算子是线性的，即 $A(\mathbf{m}) = A\mathbf{m}$ ；那么最小二乘反演得到的结果 $\tilde{\mathbf{m}}$ 使得边缘概率密度 $\sigma(\tilde{\mathbf{m}})$ 取得极大值。可见，通过概率论的观点建立反问题的误差范函更具一般性，可以描述更为复杂的系统。

3 梯度计算问题

3.1 线性化问题的梯度计算

在线性问题中，利用泛函变分求解极值问题的表述如下。假设 $K:\mathrm{H}_1 \to \mathrm{H}_2$ 是一个有界的线性算子，定义如下的 Tikhonov 泛函，

$$\mathfrak{I}(\mathbf{m})=\frac{1}{2}\|K\mathbf{m}-\mathbf{d}\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\|\mathbf{m}\|_{\mathrm{H}_1}^2 \tag{3.1}$$

对任意的介质扰动 $\delta\mathbf{m}\in\mathrm{H}_1$ 和扰动步长 $\tau\in\mathbb{R}$，总存在如下的泛函变分

$$\begin{aligned}
&\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})=\frac{1}{2}\|K(\mathbf{m}+\tau\delta\mathbf{m})-\mathbf{d}\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\|\mathbf{m}+\tau\delta\mathbf{m}\|_{\mathrm{H}_1}^2\\
&=\frac{1}{2}\langle K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m}\rangle_{\mathrm{H}_2}\\
&+\frac{\alpha}{2}\langle\mathbf{m}+\tau\delta\mathbf{m},\mathbf{m}+\tau\delta\mathbf{m}\rangle_{\mathrm{H}_1}\\
&=\frac{1}{2}\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}+\frac{\tau}{2}\left(\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}\right)+\frac{\tau^2}{2}\langle K\delta\mathbf{m},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}\\
&+\frac{\alpha}{2}\langle\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}+\frac{\alpha}{2}\tau\left(\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}+\langle\delta\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}\right)+\frac{\alpha\tau^2}{2}\langle\delta\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}\\
&=\frac{1}{2}\left(\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}+\alpha\langle\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau}{2}\left(\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}+\langle\delta\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\langle K\delta\mathbf{m},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\alpha\langle\delta\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\
&=\frac{1}{2}\left(\|K\mathbf{m}-\mathbf{d}\|_{\mathrm{H}_2}^2+\alpha\|\mathbf{m}\|_{\mathrm{H}_1}^2\right)\\
&+\frac{\tau}{2}\left(\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}+\langle\delta\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\|K\delta\mathbf{m}\|_{\mathrm{H}_2}^2+\alpha\|\delta\mathbf{m}\|_{\mathrm{H}_1}^2\right)
\end{aligned} \tag{3.2}$$

对上式定义的变分求偏导数，导数等于零，泛函可能达到极值点。

$$
\begin{aligned}
&\frac{\partial}{\partial\tau}\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right)\Big|_{\tau=0}\\
&=\left\{\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)+\tau\left(\left\|K\delta\mathbf{m}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\right\}\Big|_{\tau=0}\\
&=\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\left(\left\langle K^*\left(K\mathbf{m}-\mathbf{d}\right),\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\alpha\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\left\langle K^*\left(K\mathbf{m}-\mathbf{d}\right)+\alpha\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\left\langle\left(K^*K+\alpha I\right)\mathbf{m}-K^*\mathbf{d},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\left\langle\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right),\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}
\end{aligned}
\tag{3.3}
$$

其中 $\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)=\left(K^*K+\alpha\mathbf{I}\right)\mathbf{m}-K^*\mathbf{d}$

当梯度等于零时，可以得到线性系统的最小二乘解：

$$\mathbf{m}=\left(K^*K+\alpha\mathbf{I}\right)^{-1}K^*\mathbf{d} \tag{3.4}$$

实际上，我们也可以用迭代法来实现最小二乘反演，迭代格式为，

$$
\begin{aligned}
\mathbf{m}^{k+1}&=\mathbf{m}^k-\mu\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)\\
&=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}-\mathbf{d}^{obs}\right)\right]\\
&=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}^k-\mathbf{d}^{obs}\right)\right]\\
&=\mathbf{m}^k-\mu\left[K^*\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)\right]
\end{aligned}
\tag{3.5}
$$

迭代的最小二乘方法中，每次模型更新后，都要重新计算核函数 K，并计算波场 $\mathbf{d}^{cal}$。在勘探地球物理的实践中，我们常常用反偏移（Demigration）来近似估计 $\mathbf{d}^{cal}$。在做数据空间的波场逼近时，还要考虑 $\mathbf{d}^{obs}$ 的规则化问题，或者统一两个数据的空间采样方式。

3.2 非线性问题的梯度计算问题

假设泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 处是 Fréchet 可微的，由 Reisz 表示定理，存在梯度 $\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)\in\mathcal{H}$，称为泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 点的梯度，满足

$$\mathfrak{I}'\left(\mathbf{m}\right)\delta\mathbf{m}=\left\langle\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right),\delta\mathbf{m}\right\rangle_{\mathcal{H}} \tag{3.6}$$

若泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 处是 Fréchet 可微的，则对任意的 $\delta\mathbf{m}\in\mathcal{H}$，则关于变量 τ 的映射 $\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right):\mathbb{R}\mapsto\mathbb{R}$ 在 $\tau=0$ 处也是可微的，

$$\frac{d}{d\tau}\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right)\Big|_{\tau=0}=\left\langle\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right),\delta\mathbf{m}\right\rangle_{\mathcal{H}} \tag{3.7}$$

因此，我们可以从上述 Gateaux 导数的定义计算泛函的微分。

3.2.1 Jacobi 方法

计算非线性问题梯度的最简单方法是利用范函的差分（差商）值来代替微分值。该方法通过差分的方式计算范函在任意点 $\mathbf{m}$ 处的梯度值，其精度取决于差分的步长，表达式如下，

$$\left[\mathrm{Grad}\mathfrak{J}_n(\mathbf{m})\right]_i=\frac{\mathfrak{J}_n(\mathbf{m}+\tau_i\mathbf{e}_i)-\mathfrak{J}_n(\mathbf{m})}{\tau_i}+O(\tau_i^2),\quad i=1,2,\cdots,n \tag{3.8}$$

上面的式子表示范函沿任意一个模型空间方向$\mathbf{e}_i$的导数值可以表示为一个差分的值加上一个二阶误差项，其中$\mathbf{e}_i$表示模型空间中的第i个单位向量，n表示模型空间的维数。所有导数值形成一个向量，构成了范函在点$\mathbf{m}$处的梯度值。

3.2.2. Adjoint-State 方法（伴随状态方法）

已知算子$A(\mathbf{m})$，它可以理解为一个投影算子，建立了模型空间和数据空间之间的映射关系。$\partial A/\partial\mathbf{m}$表示算子的微分，它定义了一个 Jacobi 矩阵。在一定的定义下，它是一个Fréchet微商。假设算子$A(\mathbf{m})$存在逆$A(\mathbf{m})^{-1}$，则有，

$$A(\mathbf{m})A(\mathbf{m})^{-1}=\mathbf{I} \tag{3.9}$$

上式两边对$\mathbf{m}$求导可得，

$$\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}A(\mathbf{m})^{-1}+A(\mathbf{m})\frac{\partial A(\mathbf{m})^{-1}}{\partial\mathbf{m}}=0 \tag{3.10}$$

上式可改写为，

$$\frac{\partial A(\mathbf{m})^{-1}}{\partial\mathbf{m}}=-A(\mathbf{m})^{-1}\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}A(\mathbf{m})^{-1} \tag{3.11}$$

定义数据拟合的残差矢量为

$$\mathbf{r}(\mathbf{m})=F(\mathbf{m})-\mathbf{d}_{obs} \tag{3.12}$$

其中$\mathbf{r}(\mathbf{m})$表示波场残差，$F(\mathbf{m})$表示正算子根据模型$\mathbf{m}$得到的数据，$\mathbf{d}_{obs}$表示观测到的数据。Gateaux导数可以用如下的方式进行定义，

$$\left.\frac{\partial A(\mathbf{m}+\tau\mathbf{e}_i)}{\partial\tau}\right|_{\tau=0}=\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\cdot\mathbf{e}_i \tag{3.13}$$

$$\begin{aligned}
[G_{LS}]_i&=\frac{\partial}{\partial\tau}\mathfrak{J}_{LS}(\mathbf{m}+\tau\mathbf{e}_i)\Big|_{\tau=0}=\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}\cdot\mathbf{r}(\mathbf{m})\\
&=\left\langle\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0},\mathbf{r}(\mathbf{m})\right\rangle\\
&=\left\langle C\frac{\partial}{\partial\mathbf{m}}\left(A(\mathbf{m})^{-1}\mathbf{e}_i\right)f,\mathbf{r}(\mathbf{m})\right\rangle_Y\\
&=-\left\langle CA(\mathbf{m})^{-1}\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)A(\mathbf{m})^{-1}f,\mathbf{r}(\mathbf{m})\right\rangle_Y\\
&=-\left\langle\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)A(\mathbf{m})^{-1}f,\left(A(\mathbf{m})^{-1}\right)^*C^*\mathbf{r}(\mathbf{m})\right\rangle_U\\
&=\left\langle\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u},\mathbf{z}\right\rangle_U
\end{aligned} \tag{3.14}$$

其中，$\mathbf{u}=A(\mathbf{m})^{-1}f$为正演波场；$\mathbf{z}=-\left(A(\mathbf{m})^{-1}\right)^*C^*\mathbf{r}(\mathbf{m})$为残差波场的反传播，可以理解为偏移结果，内积可以视为零相位的互相关。

可以看出，波场$\mathbf{z}$的求解需要解如下的微分方程

$$\left(A(\mathbf{m})\right)^{*}\mathbf{z}=-C^{*}\mathbf{r}(\mathbf{m}) \tag{3.15}$$

前已述及，$A(\mathbf{m})$是一个共轭算子，满足条件：$\left(A(m)^{-1}\right)^{*}=\left(A(m)^{*}\right)^{-1}$。

利用 Adjoint-State 方法求解梯度的地球物理意义可以解释为：误差函数对模型扰动的一阶导数值为震源函数激发的波场与残差波场的反传波得到的场的内积结果。这种计算梯度的方法需要求解方程$\mathbf{u}=A(\mathbf{m})^{-1}f$和$A(\mathbf{m})^{*}\mathbf{z}=-C^{*}\mathbf{r}(\mathbf{m})$，这是两次波场正演计算。可以看出，反问题的计算量主要集中在梯度的计算上，每次迭代原则上至少需要两次正演计算。

相对于 Jacobi 方法而言，Adjoint-State 方法可以方便、更高效地计算出泛函的梯度值，用于非线性迭代公式的计算。这是目前全波形反演中最常用的梯度计算方法。

4 线性反演问题

4.1 线性反演基本问题的描述

线性反演问题可以归结为，假设已知背景模型$\mathbf{m}_0$，利用线性预测器（正演模拟算子）$G(\mathbf{m}_0)\delta\mathbf{m}=\delta\hat{\mathbf{d}}$，预测期望数据$\delta\mathbf{d}(\mathbf{x}_s,\mathbf{x}_r,t)$。当预测误差能量达到最小时，即得到估计的$\delta\mathbf{m}$。

下式表示地下所有散射点在震源激发下产生的二次震源场的叠加，用如下的 Born 近似后的散射场计算公式计算

$$\delta d(\mathbf{x}_r,\mathbf{x}_s,\omega)=\omega^2\int d^3xG_0(\mathbf{x},\mathbf{x}_s,\omega)\delta m(\mathbf{x})G_0(\mathbf{x}_r,\mathbf{x},\omega) \tag{4.1}$$

在地震勘探中，期望数据是地表观测到的地下所有散射点的散射场的叠加，记为$\delta d(\mathbf{x}_s,\mathbf{x}_r,t)$。可以认为该数据是随机信号的一次实现。利用该数据进行系统参数估计得到的结果也是一个随机信号。但是一般地，我们并非如此看待地震观测数据及反演结果。地震观测数据经常被认为是确定性的，反演结果也是确定的。这样的观点会严重阻遏我们对反问题认识的深化，特别是在对反演结果可靠性的认识上。

（4.1）式定义的线性化的正演模型可以用如下的矩阵方程进行描述，

$$A(\mathbf{m}_0)\delta\mathbf{m}=\delta\mathbf{d} \tag{4.2}$$

它是一个波场预测器，利用已知的背景模型来预测实测的数据。利用该预测器，相对于实测数据，数据预测误差定义为，

$$\mathbf{a}_N=\delta\mathbf{d}^{obs}-\delta\mathbf{d}^{cal}=\delta\mathbf{d}^{obs}-A\delta\mathbf{m} \tag{4.3}$$

系统参数估计问题一般地归结为 Bayesian 意义下的均方误差最小时模型参数的估计问题。实际地球物理勘探中，由于观测数据只有一次，反演问题归结为 Bayesian 意义下预测误差能量最小时估计值的求取问题。

在线性预测器$A\delta\mathbf{m}=\delta\mathbf{d}$下，假设预测误差$\mathbf{a}_N$是高斯分布的，在线性 Bayesian 估计框架下，得到的最佳估计结果是线性最优估计，没有比此估计结果更好的结果。附录中的 Gauss-Markov 定理说明了这一点。但如果所用的预测器是非线性的，并且该预测器可以更准确地描述正过程，则有可能得到更好的模型参数估计结果。

在线性预测器$A\delta\mathbf{m}=\delta\mathbf{d}$存在并假设预测误差$\mathbf{a}_N$是高斯分布的情况下，分布参数$\delta\mathbf{m}$的估计可以表示为对下述二次型误差泛函的求极小值问题，

$$\begin{aligned}\min_{\delta\mathbf{m}}\|a_N\|^2&=\left(\delta\mathbf{d}^{obs}-\delta\mathbf{d}^{cal}\right)^{H}C_{\mathbf{d}}^{-1}\left(\delta\mathbf{d}^{obs}-\delta\mathbf{d}^{cal}\right)\\&=\left(\delta\mathbf{d}^{obs}-A\delta\mathbf{m}\right)^{H}C_{\mathbf{d}}^{-1}\left(\delta\mathbf{d}^{obs}-A\delta\mathbf{m}\right)\end{aligned} \tag{4.4}$$

在最小二乘意义下，上式的解可以写为

$$\delta\mathbf{m}_e = \left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1} A^H C_{\mathbf{d}}^{-1} \delta\mathbf{d}^{obs} \tag{4.5}$$

其中，$\delta\mathbf{m}_e$ 代表利用线性的波场预测器和已知观测数据时对分布参数的估计结果。在地震波反演问题中，正算子矩阵 A 为 Green 函数矩阵，A^H 为偏移算子，$A^H C_d^{-1} A$ 为考虑了数据协方差的 Hessian 矩阵，$C_{\mathbf{d}}$ 是数据的协方差矩阵，它代表数据分量的可靠程度。在该意义下，模型的更新量即为最小二乘偏移结果。该解对应的最小数据误差为，

$$\begin{aligned}\mathbf{a}_N &= \delta\mathbf{d}^{obs} - A\delta\mathbf{m}_e = \delta\mathbf{d}^{obs} - A\left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1} A^H C_{\mathbf{d}}^{-1}\delta\mathbf{d}^{obs} \\ &= \left[\mathbf{I} - A\left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1} A^H C_{\mathbf{d}}^{-1}\right]\delta\mathbf{d}^{obs} = \left[\mathbf{I} - R_{\mathbf{d}}\right]\delta\mathbf{d}^{obs}\end{aligned} \tag{4.6}$$

其中，数据分辨率矩阵定义为 $R_{\mathbf{d}} = A\left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1} A^H C_{\mathbf{d}}^{-1}$。估计数据和实测数据之间的关系为，

$$\delta\hat{\mathbf{d}}^{cal} = A\left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1} A^H C_{\mathbf{d}}^{-1}\delta\mathbf{d}^{obs} \tag{4.7}$$

数据分辨率矩阵表明了估计出的数据与实测数据之间的差异。当数据分辨率矩阵为单位矩阵时，说明估计出的数据等于实测数据。因此，它可以用来评估数据估计的精度。在地震数据规则化处理时，可以用数据分辨率矩阵来评价数据规则化方法的好坏。该解对应的最小模型误差为，

$$\begin{aligned}\mathbf{a}_{\mathbf{m}} &= \delta\mathbf{m}_t - \delta\mathbf{m}_e = \delta\mathbf{m}_t - \left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1} A^H C_{\mathbf{d}}^{-1}\delta\mathbf{d}^{obs} \\ &= \delta\mathbf{m}_t - \left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1}\left(A^H C_{\mathbf{d}}^{-1} A\right)\delta\mathbf{m}_t \\ &= \left[\mathbf{I} - \left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1}\left(A^H C_{\mathbf{d}}^{-1} A\right)\right]\delta\mathbf{m}_t = \left[\mathbf{I} - R_{\mathbf{m}}\right]\delta\mathbf{m}_t\end{aligned} \tag{4.8}$$

其中，$\delta\mathbf{m}_t$ 表示真实的模型误差，模型分辨率矩阵定义为，

$$R_{\mathbf{m}} = \left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1}\left(A^H C_{\mathbf{d}}^{-1} A\right) \tag{4.9}$$

估计模型与真模型之间的关系为，

$$\delta\mathbf{m}_e = \left(A^H C_{\mathbf{d}}^{-1} A\right)^{-1}\left(A^H C_{\mathbf{d}}^{-1} A\right)\delta\mathbf{m}_t = R_{\mathbf{m}}\delta\mathbf{m}_t \tag{4.10}$$

很显然，当 $R_{\mathbf{m}} = \mathbf{I}$ 时，模型分辨率达到最高。否则，我们估计出的模型是真实模型的某种光滑结果。从模型分辨率矩阵的形式可以发现，Hessian 算子的逆和 Hessian 算子本身相乘即可得到模型分辨率矩阵 $R_{\mathbf{m}}$。模型分辨率矩阵可以评估参数估计的精度。

4.2 Hessian 算子的地球物理含义

在线性化反演过程中，方程（4.5）定义的 Hessian 具有特别重要的意义，有必要把它的地球物理含义弄清楚。

在反演过程中，正问题描述了地下所有散射点的介质摄动所引起的扰动场经过叠加后形成了地表观测波场的过程；反问题是利用地表观测数据对地下每个点的扰动量进行求解的参数估计过程。

（4.1）式是积分方程形式的频率空间域正问题表达，可以将其写成如下的矩阵方程形式

$$\delta\mathbf{d} = \boldsymbol{G}\left(\mathbf{m}_0\right)\delta\mathbf{m} \tag{4.11}$$

上式的具体形式可以表示如下

$$
\begin{matrix} & & & & & NZ=1 & & & & & & NZ=2 & & & & & & & & NZ=\mathrm{NZ} & & & \\ \begin{bmatrix} \tilde{U}_{S_1R_1} \\ \tilde{U}_{S_1R_2} \\ \vdots \\ \tilde{U}_{S_1R_N} \\ \vdots \\ \tilde{U}_{S_NR_1} \\ \tilde{U}_{S_NR_2} \\ \vdots \\ \tilde{U}_{S_NR_N} \end{bmatrix} & = & \begin{bmatrix} G_{S_1R_1}^{X_1Z_1} & G_{S_1R_1}^{X_2Z_1} & \cdots & G_{S_1R_1}^{X_{NX}Z_1} & G_{S_1R_1}^{X_1Z_2} & G_{S_1R_1}^{X_2Z_2} & \cdots & G_{S_1R_1}^{X_{NX}Z_2} & \cdots & \cdots & G_{S_1R_1}^{X_1Z_{NZ}} & G_{S_1R_1}^{X_2Z_{NZ}} & \cdots & G_{S_1R_1}^{X_{NX}Z_{NZ}} \\ G_{S_1R_2}^{X_1Z_1} & G_{S_1R_2}^{X_2Z_1} & \cdots & G_{S_1R_2}^{X_{NX}Z_1} & G_{S_1R_2}^{X_1Z_2} & G_{S_1R_2}^{X_2Z_2} & \cdots & G_{S_1R_2}^{X_{NX}Z_2} & \cdots & \cdots & G_{S_1R_2}^{X_1Z_{NZ}} & G_{S_1R_2}^{X_2Z_{NZ}} & \cdots & G_{S_1R_2}^{X_{NX}Z_{NZ}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ G_{S_1R_N}^{X_1Z_1} & G_{S_1R_N}^{X_2Z_2} & \cdots & G_{S_1R_N}^{X_{NX}Z_1} & G_{S_1R_N}^{X_1Z_2} & G_{S_1R_N}^{X_2Z_2} & \cdots & G_{S_1R_N}^{X_{NX}Z_2} & \cdots & \cdots & G_{S_1R_N}^{X_1Z_{NZ}} & G_{S_1R_N}^{X_2Z_{NZ}} & \cdots & G_{S_1R_N}^{X_{NX}Z_{NZ}} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ G_{S_NR_1}^{X_1Z_1} & G_{S_NR_1}^{X_2Z_2} & \cdots & G_{S_NR_1}^{X_{NX}Z_1} & G_{S_NR_1}^{X_1Z_2} & G_{S_NR_1}^{X_2Z_2} & \cdots & G_{S_NR_1}^{X_{NX}Z_2} & \cdots & \cdots & G_{S_NR_1}^{X_1Z_{NZ}} & G_{S_NR_1}^{X_2Z_{NZ}} & \cdots & G_{S_NR_1}^{X_{NX}Z_{NZ}} \\ G_{S_NR_2}^{X_1Z_1} & G_{S_NR_2}^{X_2Z_2} & \cdots & G_{S_NR_2}^{X_{NX}Z} & G_{S_NR_2}^{X_1Z_2} & G_{S_NR_2}^{X_2Z_2} & \cdots & G_{S_NR_2}^{X_{NX}Z_2} & \cdots & \cdots & G_{S_NR_2}^{X_1Z_{NZ}} & G_{S_NR_2}^{X_2Z_{NZ}} & \cdots & G_{S_NR_2}^{X_{NX}Z_{NZ}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ G_{S_NR_N}^{X_1Z_1} & G_{S_NR_N}^{X_2Z_2} & \cdots & G_{S_NR_N}^{X_{NX}Z_1} & G_{S_NR_N}^{X_1Z_2} & G_{S_NR_N}^{X_2Z_2} & \cdots & G_{S_NR_N}^{X_{NX}Z_2} & \cdots & \cdots & G_{S_NR_N}^{X_1Z_{NZ}} & G_{S_NR_N}^{X_2Z_{NZ}} & \cdots & G_{S_NR_N}^{X_{NX}Z_{NZ}} \end{bmatrix} & \begin{bmatrix} \delta m_{1,1} \\ \delta m_{2,1} \\ \vdots \\ \delta m_{NX,1} \\ \vdots \\ \delta m_{1,NZ} \\ \delta m_{2,NZ} \\ \vdots \\ \delta m_{NX,NZ} \end{bmatrix} \end{matrix}
$$

很显然，矩阵方程左端项中每一个元素对应一个地震道的一个频谱分量，它是由地下所有绕射点产生的绕射波叠加形成的。因此，传播矩阵 $\boldsymbol{L}$ 的每一行可以分成 NZ 份（NZ 代表成像深度离散点数），每一份中有 NX 个离散点（NX 代表横向成像范围的离散点数）对应的 Green 函数。每点对应的 Green 函数与介质扰动的（频率域）乘积就产生了地表一个地震道观测的能量，它是由地下所有绕射点产生的绕射波叠加形成的。用该观点描述的物理过程是一个线性过程。

介质扰动向量 $\delta\mathbf{m}$ 以深度层作为主序号，横向点作为次序号排列形成。相当于把每一层每一个空间点的扰动量首尾相连形成一个向量。这就把本来二维表示的扰动矩阵 $\delta\mathbf{m}(\mathrm{NX},\ \mathrm{NZ})$ 排成一个向量的形式。即：

$$
\delta\mathbf{m}=\begin{pmatrix} \delta m_{11,}\delta m_{12,}\delta m_{13,}\cdots, \\ \delta m_{1n};\delta m_{21,}\delta m_{22,}\cdots, \\ \delta m_{2n};\cdots;\delta m_{m1,}\delta m_{m2,}\cdots,\delta m_{mn} \end{pmatrix}^{\mathrm{T}} \tag{4.12}
$$

其中，n 代表成像深度离散点数下标，m 代表横向成像范围的离散点数下标。矩阵方程（4.11）一般为超定方程，即地震道的个数多于要确定的扰动参数 $\delta\mathbf{m}$ 的个数。但在某些实际情况中，该方程也可能是欠定的。这也是地下网格点之间的间隔不能太细的原因之一。因为网格点的间隔越细，模型空间的维数就会越高，会人为地扩大参数空间的维度，导致反问题的不稳定性更加严重。因此，在反演过程中研究模型的表达方式，降低模型空间的维数是很重要的。

在最小二乘意义下，可以导出如下的法方程：

$$
\delta\mathbf{m}=\left(\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{G}\right)^{-1}\boldsymbol{G}^{\boldsymbol{T}}\delta\mathbf{d}^{\boldsymbol{obs}}=\boldsymbol{H}^{-1}\boldsymbol{G}^{\boldsymbol{T}}\delta\mathbf{d}^{\boldsymbol{obs}} \tag{4.13}
$$

其中，$\boldsymbol{H}=\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{G}$ 构成 Hessian 矩阵。假若传播矩阵 $\boldsymbol{G}$ 是酉矩阵，则有：

$$
\delta\mathbf{m}=\boldsymbol{I}^{-1}\boldsymbol{G}^{\boldsymbol{T}}\delta\mathbf{d}^{\boldsymbol{obs}}=\boldsymbol{G}^{\boldsymbol{T}}\delta\mathbf{d}^{\boldsymbol{obs}} \tag{4.14}
$$

其中，$\boldsymbol{H}$ 的具体形式为（4.16）式所示。分析该矩阵可知，二维情况下 Hessian 矩阵是一个 $(NX\times NZ)\times(NX\times NZ)$ 维度的大规模带状稀疏矩阵。对角线上的元素 $H(i,i)$ 代表所有炮检点对特定散射点的地震波照明“能量”之和。即先对特定炮检对与特定散射点形成的 Green 函数求自相关，然后把所有炮检对所有频率与该特定散射点对应的 Green 函数的自相关求和的值。由于 Hessian 矩阵的庞大，对其求解与求逆的过程变得十分复杂。鉴于其主对角占优的特征（Plessix 和 Mulder，2004；Ren 等，2011），较为简单的近似是只取其对角元素参与运算：

$$
\delta\mathbf{m}=\boldsymbol{D}^{-1}\boldsymbol{G}^{\boldsymbol{T}}\delta\mathbf{d}^{\boldsymbol{obs}} \tag{4.15}
$$

其中，$D=H(i,i)$ 为 Hessian 矩阵的对角线元素。（4.15）式的偏移结果与常规叠前偏移结果的区别在于用各点的照明能量对常规叠前偏移结果进行点对点的能量校正，即对偏移结果进行照明能量的补偿。

$$
\begin{array}{c}
NZ=1 \qquad\qquad\qquad\qquad\qquad\qquad NZ=2 \qquad\qquad\qquad\qquad\qquad\qquad NZ=NZ \\
\left[\begin{array}{cccccccccccccccc}
\sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_1},G_{SR_i}^{X_{nx}z_{nz}}) \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_1},G_{SR_i}^{X_{nx}z_{nz}}) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_1},G_{SR_i}^{X_{nx}z_{nz}}) \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_2},G_{SR_i}^{X_{nx}z_{nz}}) \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_2},G_{SR_i}^{X_{nx}z_{nz}}) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_2},G_{SR_i}^{X_{nx}z_{nz}}) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_1z_{nz}},G_{SR_i}^{X_{nx}z_{nz}}) \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_2z_{nz}},G_{SR_i}^{X_{nx}z_{nz}}) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_1z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_2z_1}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_{nx}z_1}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_1z_2}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_2z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_{nx}z_2}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_1z_{nz}}) & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_2z_{nz}}) & \cdots & \sum_{i=1}^{NT}(G_{SR_i}^{X_{nx}z_{nz}},G_{SR_i}^{X_{nx}z_{nz}})
\end{array}\right]
\end{array}
\tag{4.16}
$$

Hessian 矩阵中任意元素的地球物理含义可以通过分析任意一行元素的意义进行理解。由（4.16）式知，Hessian 矩阵的每一行元素可以分为 NZ 份，分别对应着每一个深度，每一份又包含 NX 个横向散射点的贡献。对角线周围的那一组 NX 个点，对应当前散射点所在深度层的 NX 个点。其中 $H(i,i)$ 的含义已经说明，其他点 $H(i,j)(i\neq j)$ 的数学含义为：先对特定炮检对与特定散射点及该深度层上其他散射点形成的 Green 函数求互相关，然后把所有炮检对与该特定散射点及与该深度层上其他散射点对应的 Green 函数的互相关求和。同样的，对应其他深度层上 NX 个散射点的含义为：对特定炮检对与特定散射点及其他深度层上其他散射点形成的 Green 函数求互相关，然后把所有炮检对与该特定散射点及与其他深度层上其他散射点对应的 Green 函数的互相关求和。因此，Hessian 矩阵的物理含义可以理解为：对地下一个成像点的能量正传播，由于受到观测孔径和算子精度等因素的影响在反传过程中对其他成像点的照明。显然，在理想情况下（观测孔径无限大、地震信号子波频带无限宽、速度正确），对一个成像点反射能量的反传播就应当在该点成像，而不应当反传播到其他的成像点。即理想情况下的 Hessian 矩阵的非对角元素为零。因此，Hessian 矩阵的非对角元素不为零的情况恰恰反映了整个系统照明的误差或是对波场观测的不完整。对全 Hessian 矩阵的求逆应当可以弥补这种误差，但由于 Hessian 矩阵的庞大，对其直接求逆面临难以承受的计算量。

4.3 勘探地震学中线性化反问题的认识

线性化的反问题，尤其是勘探地震学中的线性化的反问题，是在线性化处理后建立起的场的扰动和参数扰动之间的线性关系。

利用该线性关系建立起的正问题进行反演，有两个典型的实例，一个是线性化的层析反演；一个是最小二乘偏移成像。

线性化的层析反演目的是估计背景速度上的速度扰动，在已知的初始模型中进行射线追踪计算核函数矩阵，用该核函数矩阵的转置来估计速度扰动，用估计的速度扰动更新初始速度模型，并不断地进行迭代，直至收敛。线性化层析的每一次迭代过程是一个线性问题，但整个迭代过程是非线性的。它利用一系列线性问题来逼近原来的非线性问题。层析可以建立一个比较光滑的背景速度场，作为叠前深度偏移成像和 FWI 的初始速度模型。速度合适的标准一般以静校正（也是一种成像）质量是否满足要求和地震深度偏移道集是否拉平为准。当然，也可以辅以其他的标准，譬如模型修正量是否小到不必再修改的程度等。

最小二乘反演成像，目的也是估计背景速度上的速度扰动。它与层析反演的差异在于描述波传播现象

所用的正算子，即描述波传播现象的方式不同。最小二乘反演的核函数矩阵的构造是用（4.11）式，其中矩阵中的元素不再是射线在一个网格内的长度，而是（4.11）式中定义的 Green 函数。因此，该核函数的转置矩阵反传播的不再是旅行时误差，而是波场误差。它不仅利用了旅行时误差的信息，还利用了振幅的信息。这就是是为什么用最小二乘偏移成像估计出的速度扰动比用层析反演估计出的速度扰动的分辨率更高的原因。另外，如果用迭代的方法进行最小二乘偏移成像，每个迭代步所用的正算子应该是反偏移算子（Demigration），而不是波场正演算子（Modeling）。但是，即便是迭代的最小二乘，它估计的依然是速度场扰动场的高频成份，而不是全频带的速度扰动场。因此，如果偏移速度误差较大，最小二乘偏移的误差也会很大。实际上，最小二乘偏移应该是在已经比较理想地完成了偏移速度分析或角度成像道集层析速度反演的基础上进行。二者结合在一起进行全频带速度场的估计是可行的，这也是 FWI 速度估计的一种实用化的替代方案。

类似于前面提及的层析速度反演的非线性迭代过程，最小二乘偏移成像也可以构成非线性的迭代过程。在两步线性化的估计速度扰动量的迭代最小二乘偏移之间增加角度成像道集层析速度反演步骤。这样可以把速度场低波数成分的估计与速度场高波数成分的估计结合在一起，实现全波数范围的速度场的估计。

5 非线性反演问题

非线性最小二乘反演问题可以定义为

$$\underset{\mathbf{m}\in R^n}{Min}\, f(\mathbf{m})=\frac{1}{2}R(\mathbf{m})^T R(\mathbf{m})=\frac{1}{2}\sum_{i=1}^{m} r_i^2(\mathbf{m}) \tag{5.1}$$

其中，$R(\mathbf{m})=g(\mathbf{m})-\mathbf{d}_{obs}$ 代表剩余波场，是一个关于 $\mathbf{m}$ 的非线性函数，因此误差范函 $f(\mathbf{m})$ 也是非线性函数。$r_i^2(\mathbf{m})$ 代表剩余波场的分量。可以用牛顿（类）方法求解（5.1）式的极小值，因此需要计算 $f(\mathbf{m})$ 的一阶、二阶导数。（5.2）式给出了 $f(\mathbf{m})$ 的一阶、二阶导数

$$\begin{aligned}\nabla f(\mathbf{m})&=\sum_{i=1}^{m} r_i(\mathbf{m})\nabla r_i(\mathbf{m})=G^T(\mathbf{m})R(\mathbf{m})\\ \nabla^2 f(\mathbf{m})&=\sum_{i=1}^{m}\nabla r_i(\mathbf{m})\nabla r_i(\mathbf{m})^T+r_i(\mathbf{m})\nabla^2 r_i(\mathbf{m})\\ &=G^T(\mathbf{m})G(\mathbf{m})+S(\mathbf{m})\end{aligned} \tag{5.2}$$

上式中的 $S(\mathbf{m})$ 表达如下

$$S(\mathbf{m})=\sum_{i=1}^{m} r_i(\mathbf{m})\nabla^2 r_i(\mathbf{m}) \tag{5.3}$$

根据 $f(\mathbf{m})$ 的一阶、二阶导，可以在初始模型 $\mathbf{m}_0$ 周围对 $f(\mathbf{m})$ 进行二阶 Taylor 展开，其表达式如下

$$\begin{aligned}f(\delta\mathbf{m}+\mathbf{m}_0)&=f(\mathbf{m}_0)+\nabla f^T(\mathbf{m}_0)\delta\mathbf{m}\\ &\quad+\frac{1}{2}\delta\mathbf{m}^T\nabla^2 f(\mathbf{m}_0)\delta\mathbf{m}\\ &=\frac{1}{2}R^T(\mathbf{m}_0)R(\mathbf{m}_0)+R^T(\mathbf{m}_0)G(\mathbf{m}_0)\delta\mathbf{m}\\ &\quad+\frac{1}{2}\delta\mathbf{m}^T(G^TG+S(\mathbf{m}_0))\delta\mathbf{m}\end{aligned} \tag{5.4}$$

上式的极值点构成了方程 $\partial f(\delta\mathbf{m}+\mathbf{m}_0)/\partial\delta\mathbf{m}=0$ 的解。根据该方程可以获得 $\delta\mathbf{m}$ 的表达式如下

$$\delta\mathbf{m}=-(G^TG+S(\mathbf{m}_0))^{-1}G^TR(\mathbf{m}_0) \tag{5.5a}$$

因此，最终的模型可以表示为

$$\mathbf{m}^*=\mathbf{m}_0-(G^TG+S(\mathbf{m}_0))^{-1}G^TR(\mathbf{m}_0) \tag{5.5b}$$

（5.5b）称非线性问题的全牛顿算法。由于 $\nabla^2 r_i(\mathbf{m})$ 很难计算，一般地，可以舍弃此高阶项，得到 Gauss-

Newton 算法，

$$\mathbf{m}^* = \mathbf{m}_0 - (G^T G)^{-1} G^T R(\mathbf{m}_0) \tag{5.6}$$

其中，$G^T G$ 定义为 Hessian 矩阵。由于 Hessian 矩阵的逆比较难于计算，又有一系列拟牛顿算法和正则化方法等用来求解该 Hessian 矩阵的近似逆。至此，我们建立了非线性最小二乘方法的计算框架。

对比（5.5a）式和（4.5）式可见，线性与非线性最小二乘法之间是有紧密联系的，完全可以把（5.5a）式化为每一步的线性问题来求解，关键是如何能把每一步尽可能化为线性问题。

在无约束非线性优化理论中，（5.1）式的求解有很多不同的方法，譬如最速下降方法、共轭梯度方法、全 Newton 方法、Gauss-Newton 方法、拟 Newton 方法等等。一般地，（5.6）式的迭代格式可以写为

$$\mathbf{m}^{k+1} = \mathbf{m}^k - \mu_k \mathbf{g}_k \tag{5.7}$$

其中，μ_k 为第 k 个迭代步的步长，$\mathbf{g}_k$ 为第 k 个迭代步的梯度，由（5.2）式定义。在不同的迭代方法中，$\mathbf{g}_k$ 也可以认为是修正后的梯度方向，譬如共轭梯度方向等。步长的选择对迭代法的收敛性和收敛效率都影响甚大，具体地参考 Dennis 和 Schnabel（1996）以及 Nocedal 和 Wright（1999）。

6 地震波反演的正则化问题

6.1 概述

地球物理学，尤其是勘探地震学中的反问题描述中，正问题是居于核心位置的。一维褶积模型是常用的，但已远远满足不了工业界生产的需求，因此目前比较有意义的是高维正、反演。高维正演同样可以视为高维褶积，在 Born 近似下具有如下的积分形式

$$u_s\left(\mathbf{x}_r, \mathbf{x}_s, \omega\right) = \omega^2 \int d^3\mathbf{x} G_{or}\left(\mathbf{x}_r, \mathbf{x}, \omega\right) \cdot a\left(\mathbf{x}, \mathbf{x}_s, \omega\right) G_{so}\left(\mathbf{x}, \mathbf{x}_s, \omega\right) \tag{6.1}$$

其中 $G_{or}\left(\mathbf{x}_r, \mathbf{x}_s, \omega\right)$ 代表炮点到散射点的 Green 函数；$G_{so}\left(\mathbf{x}, \mathbf{x}_s, \omega\right)$ 代表散射点到检波点的 Green 函数，$a\left(\mathbf{x}\right)$ 是要反演的参数扰动量，$u_s\left(\mathbf{x}_r, \mathbf{x}_s, \omega\right)$ 是在炮点 $\mathbf{x}_s$ 和检波点 $\mathbf{x}_r$ 观测到的散射场。（6.1）式可抽象地表达为，

$$g\left(\mathbf{x}\right) = \int K\left(\mathbf{x} - \mathbf{x}'\right) f\left(\mathbf{x}'\right) d\mathbf{x}' \tag{6.2}$$

其中积分核 $K\left(\mathbf{x} - \mathbf{x}'\right)$ 与正问题的描述方式有关，与介质参数的分布有关，与观测方式有关，与激发源有关。将（6.2）式离散化后可得到如下的矩阵方程

$$\mathbf{d} = \mathbf{K}\mathbf{m}_{true} + \boldsymbol{\eta} \tag{6.3}$$

从上面的方程中可以看出，$\mathbf{K}$ 是否满秩，$\mathbf{K}$ 的条件数大小分别决定了解的唯一性与稳定性问题，$\boldsymbol{\eta}$ 是观测数据中的噪音。一般地，$\mathbf{K}$ 既非满秩，条件数也比较大，因此上面的矩阵方程是极其不稳定的。因此，引入对解的约束，即正规化方法，以降低解空间的大小是必须的。据 Hadamard(1902)年给出的如下定义：若一个问题的解，假定是存在的，但它并不是唯一的，不是连续地依赖于数据的（即数据中的一个小扰动引起解的大扰动），则该问题是病态的。

地球物理反问题是典型的病态问题。理论上讲，病态问题是不可解的，虽然很多具有应用价值的问题都是病态问题。很显然仅仅依据观测的不完整的数据去估计（6.3）式中的 $\mathbf{m}_{true}$ 是不可能的。但我们可以从概率论的观点来评价（6.3）式的解估计，就是得到均值无偏的方差最小的解估计。若这种解估计结果能满足应用的要求，就可以在某种程度上认为解决了该反问题。同样，对估计得到的解的可靠性同样可以从概率论的观点加以评价。因此，正规化思想的最高层次是在 Bayes 思想下，假定估计结果是随机的，它的协方差是可以预知的，或联合概率密度是已知的。把这样的关于解的约束加到求解估计的过程中，在数据及解估计随机向量都是 Gauss 分布情况下，Bayes 方程的 MAP 估计与 Tikhonov 正则化估计的表达式是一致

的。但 Tikhonov 正规化的思想相对而言要朴素得多，它所得到的反问题的解是 Bayes 估计理论的一种特殊情况下所对应的反问题的解。

直接利用 $\underset{\mathbf{m}^*}{Min}\left\|\mathbf{d}^{obs}-\mathbf{d}^{cal}\right\|^2$ 来估计模型参数是严重病态的地球物理问题。$\mathbf{d}^{cal}=\mathbf{K}\mathbf{m}_{true}$ 代表算出的波场。仅仅利用计算（模拟）波场与观测波场的逼近来估计 $\mathbf{m}_{true}$ 是非常不可靠。为此需对估计结果做限制，一般地把反问题描述为

$$\underset{\mathbf{m}^*}{Min}\left\|\mathbf{m}\right\|^2,\ \text{Subject to}\ \left\|\mathbf{K}\mathbf{m}-\mathbf{d}\right\|_2\le\delta \tag{6.4}$$

或者

$$\underset{\mathbf{m}^*}{Min}\left\|\mathbf{K}\mathbf{m}-\mathbf{d}^{obs}\right\|^2,\ \text{Subject to}\ \left\|\mathbf{m}\right\|^2\le\varepsilon \tag{6.5}$$

也可表达为

$$\underset{\mathbf{m}^*}{Min}\left\|\mathbf{K}\mathbf{m}-\mathbf{d}^{obs}\right\|^2+\alpha\left\|\mathbf{m}\right\|^2 \tag{6.6}$$

其中，α 称为 *Superparameter*，即正规化参数。这是 *Tikhonov* 正规化的基本思想。实质上，正则化思想应理解为对解的约束。在此意义下各种约束解的思想方法都可以加入。后面会明确地说明我们对正规化思想的具体实现措施。

6.2 线性系统的 SVD 分析

在地球物理学领域（其它领域也是如此），正问题的研究，尤其是反演框架下的正问题研究，都希望所建立的系统是线性的，希望能用线性模型的估计解来预测数据本身。即使正、反问题都是非线性的，迭代反演问题的每一步也希望用线性模型来表达。

对离散线性模型 $\mathbf{d}=\mathbf{K}\mathbf{m}_{true}+\boldsymbol{\eta}$ 构成的矩阵方程进行研究，可以分析核函数的特性对估计解的影响，也会对正则化思想与方法有更深入的理解。假设核函数（系数矩阵、或变换矩阵）存在如下的 SVD 分解

$$\mathbf{K}=\mathbf{U}diag\left(s_i\right)\mathbf{V}^T \tag{6.7}$$

也可写为，

$$\mathbf{K}\mathbf{V}=diag\left(s_i\right)\mathbf{U} \tag{6.8}$$

其中，$\mathbf{V}$ 的列向量称右奇异特征向量，$\mathbf{U}$ 的列向量称左奇异特征向量。并且有，

$$\mathbf{u}_i^T\mathbf{u}_j=\delta_{ij}\ \text{和}\ \mathbf{v}_i^T\mathbf{v}_j=\delta_{ij} \tag{6.9}$$

由此，（6.8）式可写为，

$$\mathbf{K}\mathbf{v}_i=\delta_i\mathbf{u}_i\ \text{或}\ \mathbf{K}^T\mathbf{u}_i=\delta_i\mathbf{v}_i \tag{6.10}$$

则线性方程的解可以表示为，

$$\mathbf{K}^{-1}\mathbf{d}=\mathbf{V}diag\left(s_i^{-1}\right)\mathbf{U}^T\mathbf{d}=\mathbf{m}_{true}+\sum_{i=1}^{n}s_i^{-1}\left(\mathbf{u}_i^T\boldsymbol{\eta}\right)\mathbf{v}_i \tag{6.11}$$

从解的形式（6.11）式中可以看出，不稳定性是由小奇异值引起，不唯一性是由 $\mathbf{K}$ 不满秩而引起。

定义 1：紧线性算子 $\mathbf{K}$：$H_1\to H_2$ 的奇异系统是一个三项 $\left\{\mathbf{u}_j,s_j,\mathbf{v}_j\right\}$ 可数集合。它具有下列特点：

①右奇异向量 $\mathbf{v}_i$ 形成零空间 $\left(\mathbf{K}\right)^{\perp}$ 的正交基；

②左奇异向量$\mathbf{u}_j$形成$\mathbf{K}$的定义域的闭包的正交基；

③奇异值是正实数，并有$s_1 \geq s_2 \geq \cdots > 0$；

④对每个j成立$\mathbf{K}\mathbf{v}_j = s_j\mathbf{u}_j$；

⑤对每个j成立$\mathbf{K}^H\mathbf{u}_j = s_j\mathbf{v}_j$。

分析（6.11）式知，若$\mathbf{K}$是满秩的，估计$\mathbf{m}_{true}$的结果，受噪音水平的影响，也受$\mathbf{K}$的影响，即s_i的影响。若矩阵$\mathbf{K}$的条件数比较大，则受s_i的影响更多一些，在这种情况下，噪音的存在并不是解非唯一的主要原因。

6.3 最小二乘解的几何意义

地球物理学中的很多物理现象的描述可以归结为广义*Radon*变换，也可以把该广义*Radon*变换视为第一类Fredholm积分方程，如式（6.1）或式（6.2）所示

积分核把观测系统、介质参数分布、波传播的描述方式及震源特性（包括检波器特性）都关联在一起。若把线性化的第一类Fredholm积分方程表达成矩阵形式就成为了（6.3）式描述的离散线性系统。积分核是从模型空间向数据空间的投影算子，它本质上就是Fredholm微商，也称敏感度矩阵。地球物理意义受观测系统几何、介质参数分布、波传播的描述方式及源特性决定，正是这些因素决定了它所对应的系数矩阵的性质，譬如特征值、条件数等。

反演问题正则化的本质操作可以认为就是改变系数矩阵的性质，目的是提高反问题求解的稳定性和精度。在Hilbert空间中，假设$\mathbf{m}^*$是$\mathbf{m}_{true} \in M$的最佳逼近，即

$$\mathbf{m}^* = \arg\min_{\mathbf{m}\in M} \left\| \mathbf{m} - \mathbf{m}_{true} \right\| \tag{6.12}$$

则存在如下的投影定理。

投影定理：假若$\mathbf{m}^*$是$\mathbf{m}$的子空间中的元素$\mathbf{m}_{true}$的最佳逼近，则

$$\left\langle \mathbf{m}^* - \mathbf{m}_{true}, \mathbf{m} \right\rangle = 0, \mathbf{m} \in M \tag{6.13}$$

成立。

由投影定理及（6.3）式定义的线性系统，假若线性系统K的列向量是不相关的，则他们构成一组基函数$\left\{\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_N\right\}$，由此得

$$\mathbf{m}^* = \left(\mathbf{K}^T\mathbf{K}\right)^{-1}\mathbf{K}^T\mathbf{d}_{obs} \tag{6.14}$$

6.4 Tikhonov正则化思想

由（6.11）式知，反问题解的不稳定，不唯一性主要由K的性质及观测噪音决定。解的误差定义为

$$\mathbf{m}_\alpha = \sum_{i=1}^{n} s_i^{-1}\left(\mathbf{u}_i^T\boldsymbol{\eta}\right)\mathbf{v}_i \tag{6.15}$$

可见，小的奇异值可能导致巨大的解估差，甚至不可解。为此，*Tikhonov*正规化方法引入滤波函数，

$$w_\alpha\left(s^2\right) = \frac{s^2}{s^2 + \alpha} \tag{6.16}$$

利用上面的滤波函数可以将估计解写成如下的形式

$$\begin{aligned}\mathbf{m}_\alpha &= \mathbf{V}diag\left(w_\alpha\left(s_i^2\right)s_i^{-1}\right)\mathbf{U}^T\mathbf{d}\\ &=\sum_{i=1}^{n} w_\alpha\left(s_i^2\right)s_i^{-1}\left(\mathbf{u}_i^T\mathbf{d}\right)\mathbf{v}_i\end{aligned} \tag{6.17}$$

（6.16）式是典型的*Tikhonov*正则化滤波函数，下面是另外一种比较常用的滤波函数

$$w_\alpha\left(s^2\right)=\begin{cases}1, & s^2>\alpha\\ 0, & s^2\le\alpha\end{cases} \tag{6.18}$$

此时近似估计解为，

$$\mathbf{m}_\alpha = \sum_{s_i^2>\alpha} s_i^{-1}\left(\mathbf{u}_i^T\mathbf{d}\right)\mathbf{v}_i \tag{6.19}$$

利用上式求得的解称为 TSVD 解（截断 SVD 解）。（6.16）式的正则化滤波函数对应的近似估计解为

$$\mathbf{m}_\alpha=\sum_{i=1}^{n}\frac{s_i\left(\mathbf{u}_i\mathbf{d}\right)}{s_i^2+\alpha}=\left(\mathbf{K}^T\mathbf{K}+\alpha\mathbf{I}\right)^{-1}\mathbf{K}^T\mathbf{d} \tag{6.20}$$

从（6.20）可更明显地看出*Tikhonov*正则化的含义，从该表达式可知：*Tikhonov*正则化思想与最大后验概率方法（Bayes方法）是完全一致的。但不同的观点带来不同的认识，从Bayes估计的角度看α可认为是信噪比。从地球物理的角度看，由于K的性质受到复杂因素的影响，仅靠α这个正则化因子不一定能大幅度缩减解空间的范围。因此，关于解的先验信息的发掘和利用将构成反问题实用化的核心。

6.5 Tikhonov 正则化化参数选择

既然正则化参数在反演过程中起到很关键的作用，需要讨论数学上常常推荐的几种方法。我认为正则化问题还是要从约束解空间展布的角度处理，不仅仅是α的选择，不过这超出了本节的讨论。

6.5.1 UPRE 方法

UPRE（The Unbiased Predictive Risk Estimator）是正则化参数选择的基本方法。它是按如下方式定义风险函数，

$$\frac{1}{n}\left\|\mathbf{p}_\alpha\right\|^2=\frac{1}{n}\left\|\mathbf{K}\mathbf{m}_\alpha-\mathbf{K}\mathbf{m}_{true}\right\|^2 \tag{6.21}$$

然后通过该风险函数来估计正则化参数的，即用预测数据和实测数据的残差情况来评价正则化参数选择是否合适。原则上，正则化参数额选择合适与否应该看反演结果是否逼近真的反演结果。但是，由于正反演结果均不可知，转而用上述思想评价正则化参数选择的合适程度。上式中$\mathbf{m}_\alpha=\mathbf{R}_\alpha\mathbf{d}$，$R_\alpha=\left(\mathbf{K}^T\mathrm{K}+\alpha\mathbf{I}\right)^{-1}\mathbf{K}^T$。定义$\mathbf{A}_{resolution}=\mathbf{R}_\alpha\mathbf{K}$为分辨率矩阵。

$$\mathbf{m}_\alpha=\mathbf{R}_\alpha\mathbf{d}=\mathbf{R}_\alpha\mathbf{K}\mathbf{m}_{true}=\mathbf{A}\mathbf{m}_{true} \tag{6.22}$$

定义$\mathbf{A}_{influence}=\mathbf{K}\mathbf{R}_\alpha$为影响矩阵（也称数据分辨率矩阵），则

$$\mathbf{d}_\alpha=\mathbf{K}\mathbf{m}_\alpha=\mathbf{K}\mathbf{R}_\alpha\mathbf{d}=\mathbf{A}_{\mathrm{inf}\,luence}\mathbf{d} \tag{6.23}$$

利用上面结果，根据定义可得

$$\begin{aligned}\mathbf{p}_\alpha&=\mathbf{K}\mathbf{m}_\alpha-\mathbf{K}\mathbf{m}_{true}=\mathbf{K}\mathbf{R}_\alpha\mathbf{d}-\mathbf{K}\mathbf{m}_{true}\\ &=\mathbf{A}_{\mathrm{inf}\,luence}\left(\mathbf{K}\mathbf{m}_{true}-\boldsymbol{\eta}\right)-\mathbf{K}\mathbf{m}_{true}\\ &=\left(\mathbf{A}_{\mathrm{inf}\,luence}-\mathbf{I}\right)K\mathbf{m}_{true}+\mathbf{A}_{\mathrm{inf}\,luence}\boldsymbol{\eta}\end{aligned} \tag{6.24}$$

不失一般性，假设$\boldsymbol{\eta}$为高斯白噪声，$\mathbf{A}_{influence}$为对称阵，则

$$E\left(\frac{1}{n}\left\|\mathbf{p}_\alpha\right\|^2\right)=\frac{1}{n}\left\|\left(\mathbf{A}_\alpha-\mathbf{I}\right)\mathbf{K}\mathbf{m}_{true}\right\|^2 +\frac{\sigma^2}{n}Trace(\mathbf{A}_{influence}^2) \tag{6.25}$$

定义正则化后的数据残差为

$$\begin{aligned}&\mathbf{r}_\alpha\cong\mathbf{K}\mathbf{m}_\alpha-\mathbf{d}=\mathbf{K}\mathbf{R}_\alpha\mathbf{m}_\alpha-\mathbf{d}=\left(\mathbf{A}_{influence}-\mathbf{I}\right)\mathbf{d}=\left(\mathbf{A}_{influence}-\mathbf{I}\right)\left(\mathbf{K}\mathbf{m}_{true}+\boldsymbol{\eta}\right)\\&=\left(\mathbf{A}_{influence}-\mathbf{I}\right)\mathbf{K}\mathbf{m}_{true}+\left(\mathbf{A}_{influence}-\mathbf{I}\right)\boldsymbol{\eta}\end{aligned} \tag{6.26}$$

则

$$\begin{aligned}&E\left(\frac{1}{n}\left\|\mathbf{r}_\alpha\right\|^2\right)=\frac{1}{n}\left\|\left(\mathbf{A}_\alpha-\mathbf{I}\right)\mathbf{K}\mathbf{m}_{true}\right\|^2\\&+\frac{1}{n}\left\|\left(\mathbf{A}_\alpha-\mathbf{I}\right)\right\|E(\left\|\boldsymbol{\eta}\right\|^2)\\&=\frac{1}{n}\left\|\left(\mathbf{A}_\alpha-\mathbf{I}\right)\mathbf{K}\mathbf{m}_{true}\right\|^2\\&+\frac{2\sigma^2}{n}Trace\left(\mathbf{A}_\alpha^2\right)+\sigma^2\end{aligned} \tag{6.27}$$

将（6.27）和（6.25）式结合，可以得到下面结果

$$\begin{aligned}&E\left(\frac{1}{n}\left\|\mathbf{p}_\alpha\right\|^2\right)=E\left(\frac{1}{n}\left\|\mathbf{r}_\alpha\right\|^2\right)\\&+\frac{2\sigma^2}{n}Trace(\mathbf{A}_\alpha^2)-\sigma^2\end{aligned} \tag{6.28}$$

最后，定义 UPRE 估计α为

$$\alpha_u=\arg\min_\alpha U\left(\alpha\right) \tag{6.29}$$

其中

$$U\left(\alpha\right)=\left(\frac{1}{n}\left\|\mathbf{r}_\alpha\right\|^2\right)+\frac{2\sigma^2}{n}Trace(\mathbf{A}_\alpha)-\sigma^2 \tag{6.30}$$

式子（6.29）和（6.30）是所有正则化参数α选择方法的基础。分析可知用（6.29）和（6.30）选择α需付出巨大的计算代价，$U\left(\alpha\right)$本身是一个非线性函数，极小值往往不止一个。该方法用于线性化的小规模的参数模型反演比较合适，此时可以用较小的计算代价得到$Trace(\mathbf{A}_\alpha)$。

6.5.2 常用的正则化参数选择方法

① Generalized cross Validation （广义交叉校验方法）

该方法的思想是利用排除部分数据的反演结果来进行交叉检验，看反演结果能否预测被排除的数据，若能，说明反演结果比较好。定义排除数据d_k的反演（d_k被丢弃）泛函为：

$$\min\sum_{i\neq k}\left(\left(\mathbf{K}\mathbf{m}\right)_i-\mathbf{d}_i\right)^2+\alpha^2\left\|L\mathbf{m}\right\|_2^2 \tag{6.31}$$

$\mathbf{m}_{\alpha.L}^{[k]}$为该反问题的解，$L$的选择依赖于罚函数的定义。因此，选择$\alpha$使

$$\min V_0(\alpha)=\frac{1}{m}\sum_{k=1}^{m}\left(\left(\mathbf{K}\mathbf{m}_{\alpha.L}^{[k]}\right)_k-\mathbf{d}_k\right)^2 \tag{6.32}$$

成立，该α就是 GCV 方法要找的。该方法逻辑上很合理，不过需付出的计算太大，要解（6.31）定义的 m 次反问题。为此需对（6.32）式的计算进行简化。首先定义

$$\tilde{\mathrm{d}}_i=\begin{cases}\left(K\mathrm{m}_{\alpha.L}^{[k]}\right)_{\mathrm{k}} & i=k\\ d_i & i\neq k\end{cases} \tag{6.33}$$

则$\mathbf{m}_{\alpha,l}^{[k]}$使得

$$\min\left((G\mathbf{m})_k-\tilde{\mathbf{d}}_k\right)^2+\sum_{i\neq k}\begin{pmatrix}\left((K\mathbf{m})_i-\mathbf{d}_i\right)^2\\ +\alpha^2\left\|L\mathbf{m}\right\|_2^2\end{pmatrix} \tag{6.34}$$

成立，即

$$Min\left\|G\mathbf{m}-\tilde{\mathbf{d}}\right\|_2^2+\alpha^2\left\|L\mathbf{m}\right\|_2^2 \tag{6.35}$$

成立。由此可得

$$\mathbf{m}_{\alpha,l}^{[k]}=R_\alpha\tilde{\mathbf{d}} \tag{6.36}$$

易知

$$\frac{\left(\mathbf{A}_\alpha\tilde{\mathbf{d}}\right)_k-\left(\mathbf{A}_\alpha\mathbf{d}\right)_k}{\tilde{\mathbf{d}}_k-\mathbf{d}_k}=\mathbf{A}_{k,k} \tag{6.37}$$

上式两边同时减 1，得

$$\frac{\tilde{\mathbf{d}}_k-\mathbf{d}_k-\left(\mathbf{A}_\alpha\mathbf{d}\right)_k-\left(\mathbf{A}_\alpha\tilde{\mathbf{d}}\right)_k}{\tilde{\mathbf{d}}_k-\mathbf{d}_k}=1-\mathbf{A}_{k.k} \tag{6.38}$$

由于$\left(\mathbf{d}\right)_k=\left(\mathbf{K}\mathbf{m}_{\alpha.L}\right)_k$，$\left(A_\alpha\tilde{\mathbf{d}}\right)_k=\tilde{\mathbf{d}}_k$，$K\mathbf{m}_{\alpha.L}^{[k]}=\tilde{\mathbf{d}}_R$，则

$$\frac{\left(\mathbf{K}\mathbf{m}_{\alpha.L}\right)-\mathbf{d}_k}{\left(\mathbf{K}\mathbf{m}_{\alpha.L}^{[k]}\right)-\mathbf{d}_k}=\mathbf{I}-\mathbf{A}_{k,k} \tag{6.39}$$

把（6.39）代入（6.32）得

$$V_0(\alpha)=\frac{1}{m}\sum_{k=1}^{m}\left(\frac{\left(\mathbf{K}\mathbf{m}_{\alpha,L}\right)_k-\mathbf{d}_k}{\mathbf{I}-\mathbf{A}_{k,k}}\right)^2 \tag{6.40}$$

进一步的有

$$\mathbf{A}_{k,k}=\frac{1}{m}Trace(\mathbf{A}) \tag{6.41}$$

最后

$$V_0(\alpha)=\frac{m\left\|K\mathbf{m}_{\alpha,L}-\mathbf{d}\right\|_2^2}{Trace\left(\mathbf{I}-\mathbf{A}\right)} \tag{6.42}$$

该方法应用于小规模的参数模型情况下是可以的，譬如 Radon 变换反演 τ-p 域局部平面波。对于 FWI 这一类的大规模的参数模型反演不能用此方法。

②The L-curve Method

L 曲线方法的本质含义是选择 α，使得数据误差的能量的权重和模型误差能量的权重在总的误差泛函中达到平衡。在实际反演过程中，二者的误差能量一致化非常重要，因二者的物理量纲根本不同。一般地，我们划一个以两误差能量对数为纵横坐标的误差泛函曲线图。曲线转角处对应的 α 是我们所要的。

定义两坐标为：$X(\alpha)\equiv\log\left\|\boldsymbol{\gamma}_\alpha\right\|^2$ 和 $Y(\alpha)\equiv\log\left\|\mathbf{m}_\alpha\right\|^2$。曲线的拐角是曲率最小的地方，因此，我们选 α 使如下曲率函数最小

$$K(\alpha)\equiv\frac{X''(\alpha)Y'(\alpha)-X'(\alpha)Y''(\alpha)}{\left(X'(\alpha)+Y'(\alpha)\right)^{3/2}} \tag{6.43}$$

（6.43）式达到最小的思想选择正则化参数是比较常用的方法，但是在大规模地震数据的 FWI 反演中，不一定照搬（6.43）式，只要利用其思想就好了。譬如用抛物线拟合法决定正则化参数就更实用一下。

③ The Discrepancy Principle

一般情况下可以认为当反演结果比较正确时，合成数据与观测数据之差应非常接近白噪声，即 $\left\|\boldsymbol{\gamma}_\alpha\right\|/n=\sigma^2$。也就是，

$$E\left(\frac{1}{n}\left\|K\mathbf{m}_\alpha-\mathbf{d}\right\|^2\right)\approx E\left(\frac{1}{n}\left\|\boldsymbol{\eta}\right\|^2\right)=\sigma^2 \tag{6.44}$$

在地震波反演实践中，我们可以用初至波到达前的数据统计环境噪音水平，得到其方差。然后，在迭代反演过程中利用（6.44）式估计一个好的正则化参数。具体计算公式需要根据情况进行构造。

6.6 罚函数和数据偏差函数的定义

按 Bayes 估计的思想，需引入对模型参数分布的假设，因为由随机信号反演得到的模型参数本身也是随机的。为此引入各种所谓的罚函数去惩罚不符合概率分布假设的模型参数。

①能量最小准则

使反演结果的模最小，即

$$J(\mathbf{m})=\frac{1}{2}\left\|\mathbf{m}\right\|_2 \tag{6.45}$$

②总变差能量最小准则

使反演结果的总变差最小，即

$$J(\mathbf{m})=\frac{1}{2}\int_\Omega\sum_{i=1}^{d}\left(\frac{\partial\mathbf{m}}{\partial x_i}\right)^2 \tag{6.46}$$

还有各种范数可定义罚函数，如 Huber 范数，Cauchy 泛函，总变差能量最小准则也有许多变种。另外，数据偏差函数也不一定用观测数据与正演数据的差的 L2 范数来定义，具体见参考文献（Vogel, C. R., 2002）。

7 高斯性与线性性、非高斯性与非线性性

目前，各种地震波反演方法，包括层析反演及 FWI 等，均是基于预测误差是高斯分布的。在此假设下才有对应的 L_2 范数下定义的误差泛函，及相应的一系列线性化的最小二乘方法和梯度导引类的优化方法。

在线性化假设下，即预测算子（或正演算子）是线性的，L_2 范数下定义的误差泛函对应一个二次型函

数。非线性情况下，在某一个局部点，我们也把非线性问题转化为线性问题来解决。

但是，我们知道预测误差是高斯分布的这个假设与实际数据情况并不相符。预测误差的非高斯性，一方面是因为观测数据具有随机分布的特征，另一方面是因为我们构造的预测算子不能很好地预测所有的波现象。另外，模型参数的纵横向变化越剧烈，数据与模型之间关系的非线性性也越强。总之，只有观测数据或预测误差噪音满足高斯分布假设，并且系统是线性的，一套最小二乘方法得到的模型估计结果才是无偏的、方差最小的，或者说这是在满足上面所有假设条件下的最好的估计结果。

如果观测数据或是噪音的概率密度函数是非高斯的，上述解法方程的（线性）方法或是梯度导引类的（非线性）方法所得到的估计结果不一定是最佳的。另外，非高斯性与非线性性之间的关系是复杂的。首先，高斯分布的预测误差说明误差泛函的凸性很好，因为这样的概率分布函数理论上只有一个极小值点，而且函数的几何形状是或接近是超椭球体，求极值是比较容易的；其次，高斯性与线性性、非高斯性与非线性性并非是相同的概念。实质上它们尽管有联系，它们描述的完全不是同一类概念。高斯性和非高斯性描述的是观测数据或预测误差的概率分布，它决定了误差泛函的形式；线性性与非线性性指的是系统的性质（指系统参数的纵横向变化非常剧烈），或者指观测数据与模型参数之间的映射关系。线性映射是一（个数据体）对一（个模型）的，非线性映射是一（个数据）对多（个模型）的。最后，非线性性越强，误差范函的极值点越多，误差范函的凸性越差，求解全局极值点也就越困难。

根据上述分析和地球物理反演实践，我们知道，地球物理问题是高度非线性的，尤其是在那些构造复杂的探区，更不用讲地表条件复杂的探区。我们要做的工作是尽可能建立凸性好的泛函。选择合适的、符合数据噪音特点的概率分布函数是必须的，但这也是很困难的。因为目前我们对噪音的统计特征的研究做工作不多。然后，我们尽量把地球物理问题的非线性性减弱，我们认为多尺度反演和/或特征波场反演是正确的路线。另外，我们也可以发展出一系列的技术形成流程来最终解决全频带速度估计问题。

8 FWI 实用化途径分析

8.1 FWI 实用化方案

原始意义下的 FWI，就是得到全频带速度场的、解决储层级别的问题的、按照数据逼近方式的 FWI 在可以预计的时间内是不可能实用化的。

为了把原始意义的 FWI 应用于实际数据，需要进行变通。我们提出的 FWI 实用化的方案是用一组合适的技术组合基本实现原始意义下的 FWI。该方案由图 2 中的流程图描述。

8.2 FWI 实用化方案可行性分析

原始意义下的 FWI 之所以不能直接应用于实际数据，原因很多，很多文献已经明确地分析过。最根本地是 FWI 的非线性性。FWI 试图从炮集数据直接给出全频带的参数模型。图 2 给出的流程把它拆分成四个关键部分：首先，选合适的方法确定模型参数的背景变化，即低频或低波数成分。然后，通过线性化的反演（主要是最小二乘叠前深度偏移），确定模型参数的高波数变化的量（角度反射系数道集）和空间位置（成像剖面）。接着对成像剖面进行结构成像，就是用高精度的相关方法把反射结构的细节刻画出来，用于后续的精细的层析成像。最后，精细的层析成像确定模型参数的中波数成分。把低、中、高三个波数成分的参数融合成一个模型，完成一步更新。速度精度不满足要求，继续迭代更新，见图 2。流程图 2 实际上是通过一种多尺度反演流程实现 FWI。比较准确的速度场的背景部分是速度场中的最重要的部分。分析可知，这部分信息不宜用线性化方法估计，因为线性化方法一般地是描述速度场的扰动与场的扰动之间的关系。另一方面，这部分信息与波传播的旅行时和波场的低频成分关系密切。还有一点就是，估计速度场的低波数成分变化，不能依赖其中的高波数成分。分析可知，NMO/DMO 叠加、MVA 和透射波层析最具备确定速度场低波数分量的能力。利用极低频（<2Hz）地震数据进行 FWI 反演也可以确定这部分速度成分，但这部分地震数据极难有效地获取。因此，我们认为基于 Monta Carlo 随机采样机制的和 MVA 相结合的方法是估计这部分速度成分的重要方法。NMO/DMO 叠加不宜用于横向变速强的情况，透射波层析受限于观测方式。当然，目前常用的基于 PSDM 的 MVA 方法一般情况下也可以估计低波数速度变化。

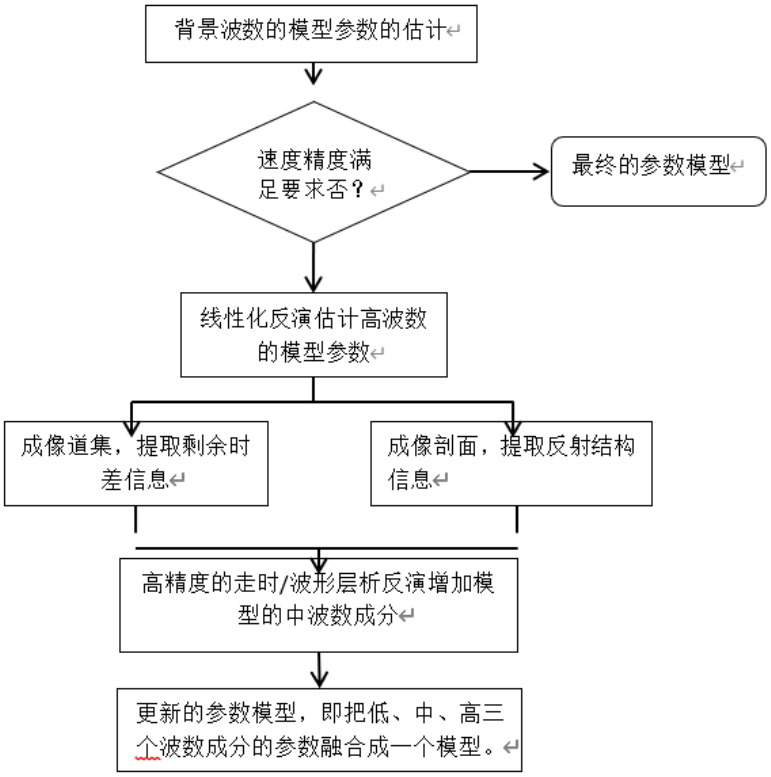


图 2 FWI 走向实用化的流程图

最小二乘叠前深度偏移实际上是一种波形层析反演方法，它利用地震波的反射振幅，估计的是速度的高波数的扰动量。迭代的最小二乘叠前深度偏移是通过反传播反偏移后的数据误差更新速度扰动量，得到扰动量的最佳估计。它产生的是 AVA 道集和叠加成像剖面。

在进行高精度的层析反演估计中波数的速度成分时，反射界面结构信息的控制是很重要的。反射界面结构信息的控制可以显著地降低反演模型的规模，提高反演的精度。反射界面结构信息可以从叠加成像剖面上通过高精度的相关方法获得。

高精度的层析速度反演需要迭代执行。它的初始模型可以是低波数的速度模型+最小二乘叠前深度偏移估计的高波数扰动+反射界面结构信息。此处，高精度的层析速度反演我们认为主要还是基于旅行时的反演，但可以是基于 Beam 波传播的。目前我们的方案是利用局部平面波形成的特征波构成旅行时层析成像和波形层析成像方法。

需要说明的是，这样的速度估计及建模流程的设计目标还是满足保精度的、保真的叠前深度偏移成像的需要，与 RTM 一起构成精确的地震成像的流程。当然，该流程也可以与各向异性参数估计和 Q 值参数估计统一在一起，在一个软件流程中实现三类参数的估计（速度参数、各向异性参数和 Q 值参数）。

9 总结与讨论

地震波反演理论在勘探地震学中是框架性的理论，我们从应用的角度比较详细地分析了它的核心环节。

正问题比较核心的是 Born 近似散射波场描述，它是最小二乘叠前偏移和层析成像的基础。

反演框架方面我们叙述了一种非常简洁的表述方式，把采样算子和震源函数也融合在正演过程及残差计算中。更清楚地看出数据采样、震源函数和模型参数在反演框架中的具体作用。

关于梯度这个核心问题，我们从变分原理的角度清楚地说明了线性情况下和非线性情况下如何计算梯度，以及梯度计算的具体含义。为地震波反演的实际计算奠定了基础。

关于线性反演及非线性反演的具体实现，我们对原始意义下的 FWI 讨论很少，因为它的框架已经很定型，也因为它在我国陆上数据中的应用目前还看不到明确的前景。我们主要讨论的是线性反演及非线性反

演的迭代实现，就是把迭代的最小二乘偏移和基于角度道集的层析反演结合在一起形成大的迭代过程，解决或部分解决全频带速度场的估计问题。这是我们坚持的实现“FWI”的方法之一。

正则化参数的选择在反演计算中具有特殊重要的地位，我们讨论了四种代表性的正则化参数选择的思想和方法。在此基础上，可以根据我们的反演实践构造出合适的方法。

关于预测误差的高斯性与非高斯性、模型空间和数据空间投影的线性性与非线性性，我们给出了比较明确的讨论，认为非线性性强是地震波反演困难的主要原因，高斯类和非高斯类的泛函建立是第二位的原因。据此得出进行特征波场反演和建立尽可能凸的泛函的结论。这是我们坚持的实现“FWI”的方法之二。

总之，通过本文的讨论，比较清楚地说明了地震波反演的核心环节的基本含义。基于此，指导我们的地震波反演的进一步实践。

致谢

感谢中国石油集团公司勘探开发研究院西北分院对同济大学波现象与反演成像研究组进行的全波形反演研究工作的资助。

参考文献

[1] Aster, R. C., Borchers, B., and Thurber, C. H., 2005, Parameter estimation and inverse problems; Elsevier Academic Press

[2] Baysal, E.,D.Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, 48, 1514–1524.

[3] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, 26, 99–108.

[4] Biondi, B., andW. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, 69, 1283–1298.

[5] Bouchon, M., M. Campillo, and S. Gaffet, 1989, A boundary integral equation—Discrete wavenumber representation method to study wave propagation in multilayered media having irregular interfaces: Geophysics, 54, 1134–1140.

[6] Brossier, R., J.Virieux, and S. Operto, 2008, Parsimonious finite-volume frequency-domain method for 2D P-SV wave modelling: Geophysical Journal International, 175, 541–559.

[7] Bunks, C., F.M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, 60, 1457–1473.

[8] Chapman, C., 1985, Ray theory and its extension: WKBJ and Maslov seismograms: Journal of Geophysics, 58, 27–43.

[9] Choi, Y., D.-J. Min, and C. Shin, 2008, Two-dimensional waveform inversion ofmulti-component data in acoustic-elastic coupled media: Geophysical Prospecting, 56, 863–881.

[10] Claerbout, J., 1971, Toward a unified theory of reflector mapping: Geophysics, 36, 467–481.

[11] Claerbout, J., 1976, Fundamentals of geophysical data processing: McGraw-Hill Book Co., Inc.

[12] Claerbout, J., and S. Doherty, 1972, Downward continuation of moveout corrected seismograms: Geophysics, 37, 741–768.

[13] Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic non-linear waveform inversion: Application to real data: Geophysics, 55, 527–538.

[14] Danecek, P., and G. Seriani, 2008, An efficient parallel Chebyshevpseudospectral method for large-scale 3D seismic forward modelling: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, P046.

[15] Dennis, J. E. Jr. and Schnabel, R. B., 1996, Numerical methods for unconstrained optimization and nonlinear equations; SIAM

[16] Djikpéssé, H., and A. Tarantola, 1999, Multi parameter 1-norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, 64, 1023–1035.

[17] Docherty, P., R. Silva, S. Singh, Z.-M. Song, and M.Wood, 2003, Migration velocity analysis using a genetic algorithm: Geophysical Prospecting, 45, 865–878.

[18] Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geophysics, 43, 1342–1351.

[19] Grote M. J. and I. Sim, 2010, Efficient PML for the wave equation: from internet: arXiv:1001.0319v1.

[20] Hansen, C., ed., 1998, Rank-defficient and discrete ill-posed problems —Numerical aspects of linear inversion: Society for Industrial and Applied Mathematics.

[21] Herrmann, F. J., C. R. Brown, Y. A. Erlangga, and P. P. Moghaddam, 2009, Curvelet-based migration preconditioning and scaling. Geophysics, 74, A41-A46.

[22] Koren, Z., K.Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, 96, 20289–20299.

[23] Kormendi, F., and M. Dietrich, 1991, Non-linear waveform inversion of plane-wave seismograms in stratified elastic media: Geophysics, 56, 664–674.

[24] Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

[25] Lambaré, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, 57, 1138–1154.

[26] Lecomte, I., 2008, Resolution and illumination analyses in PSDM: A ray based approach: The Leading Edge, 27, 650–663.

[27] Lee, K. H., and H. J. Kim, 2003, Source-independent full-waveform inversion of seismic data: Geophysics, 68, 2010–2015.

[28] Llimes, L., 2007, Sensitivity of seismic waves to the structure; Report 17, SW3D, Department of Geophysics, ChalesUniversity, Prague

[29] Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, 49, 533–549.

[30] Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

[31] Mora, P. R., 1987, Nonlinear two-dimensional elastic inversion of multi-offset seismic data: Geophysics, 52, 1211–1228.

[32] Mora, P. 1988, Elastic wavefield inversion of reflection and transmission data: Geophysics, 53, 750–759.

[33] Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, 100, 12431–12447.

[34] Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics, D. Reidel publishing company

[35] Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full waveform tomography: Application to the eastern Nankai trough: Journal of Geophysical Research, doi: 10.1029/2005JB003835.

[36] Operto, S., J. Virieux, P. Amestoy, J. Y. L'Excellent, L. Giraud, and H. Ben-Hadj-Ali, 2007, 3D finite-difference frequency-domain modeling of viscoacoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, 72, no. 5, SM195–SM211.

[37] Plessix, R. E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysics Journal International, 157, 975–987.

[38] Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method: Geophysical Prospecting, 38, 311–330.

[39] ———, 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, 64, 888–901.

[40] Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multisource cross-hole tomography, Part

1: Acoustic wave-equation method: Geophysical Prospecting, 38, 287–310.

[41] Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Multi-scale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, 159, 1032–1056.

[42] Ren, H., R. S. Wu, and H. Wang, 2011, Wave equation least square imaging using the local angular Hessian for amplitude correction: Geophysical Prospecting, Accepted.

[43] Saad, Y., 2003, Iterative methods for sparse linear systems, 2nd ed.: Society for Industrial andApplied Mathematics.

[44] Scales, J. A., P. Docherty, and A. Gersztenkorn, 1990, Regularization of nonlinear inverse problems: Imaging the near-surface weathering layer: Inverse Problems, 6, 115–131.

[45] Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, 173, 922–931.

[46] Shin, C., and Y. H. Cha, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, 177, 1067–1079.

[47] Sirgue, L., J. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F022.

[48] Snieder, R.,M. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem: Geophysics, 54, 991–1000.

[49] Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, 43, 23-48.

[50] Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: Geophysics, 74, WCA95-WCA107.

[51] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, 49, 1259–1266.

[52] Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation; SIAM

[53] Virieux, J., 1986, P-SV wave propagation in heterogeneous media, velocity stress finite difference method: Geophysics, 51, 889–901.

[54] Vogel, C. R., 2002, Computational Methods for Inverse Problems; SIAM.

[55] Wu, R. S., 2003, Wave propagation, scattering and imaging using dual-domain one-way and one-return propagation: Pure and Applied Geophysics, 160, 509–539.

[56] Yilmaz, O., 2001, Seismic data analysis: processing, inversion and interpretation of seismic data: SEG.

[57] Yu, J., J. Hu, G. T. Schuster, and R. Estill, 2006, Prestack migration deconvolution: Geophysics, 71, S53-S62.

[58] Zhdanov, M. S., 2002, Geophysical Inverse theory and regularization problems; Elsevier Science B. V.

[59] Zhou, B., and S. A. Greenhalgh, 2003, Crosshole seismic inversionwith normalized full-waveform amplitude data: Geophysics, 68, 1320–1330.

附录

附录 1：Fredholm 第一类积分方程及其共轭

用 $L^2(\Omega)$ 表示的 Ω 上的实值的平方可积函数在内积 $\langle f,g\rangle_{L^2}=\int_\Omega f(x)g(x)dx$ 和诱导范数 $\|f\|_{L^2}=\sqrt{\int_\Omega f(x)^2dx}$ 下形成一个 Hilbert 空间。Fredholm 第一类积分算子

$$(Kf)(x)=\int_\Omega k(x,y)f(y)dy,\quad x\in\Omega \tag{A1}$$

是有界的，假如 $B\triangleq\int_\Omega\int_\Omega k(x,y)^2dxdy<\infty$，此时 $\|K\|\le\sqrt{B}$。K 的共轭由

$$\left(K^{*}g\right)(y)=\int_{\Omega}k(x,y)g(y)dx,\quad y\in\Omega \tag{A2}$$

给出。当$K\left(x,y\right)=K\left(y,x\right)$时，$K$是自共轭的。

附录 2：Hilbert 空间中的最佳近似

假如$f\in\mathcal{H}$及S是$\mathcal{H}$的子空间。$s_{*}\in S$称为f的最佳近似，如果

$$s_{*}=\arg\min_{s\in S}\left\|s-f\right\| \tag{A3}$$

成立。这意味着对所有的$s\in S$存在$s^{*}\in S$，有$\left\|s_{*}-f\right\|\le\left\|s-f\right\|$。假如$s_{*}$存在，则$s_{*}$是唯一的。

正交投影定理：假如$s_{*}\in S$是f的最佳近似，则对于所有的$s\in S$成立

$$\left\langle s_{*}-f,s\right\rangle=0 \tag{A4}$$

定理说明了最小误差向量垂直于子空间S中的所有向量。对于有限维子空间S，存在基函数$\left\{\phi_{1},\phi_{2},\cdots,\phi_{N}\right\}$，正交投影定理给出了计算最佳逼近的公式

$$s_{*}=\sum_{j=1}^{N}\hat{\alpha}_{j}\phi_{j} \tag{A5}$$

其中$\hat{\alpha}=\left(\hat{\alpha}_{1},\hat{\alpha}_{2},\cdots,\hat{\alpha}_{N}\right)$通过解下面的法方程得到

$$G\alpha=\mathbf{b} \tag{A6}$$

矩阵$\left[G\right]_{ij}=\left\langle\phi_{j},\phi_{i}\right\rangle$，向量$\left[\mathbf{b}\right]_{i}=\left\langle f,\phi_{i}\right\rangle$。$N\times N$的矩阵$G$称为 Gram 矩阵。基向量$\left\langle\phi_{j},\phi_{i}\right\rangle$称为正交的，如果

$$\left\langle\phi_{j},\phi_{i}\right\rangle=\delta_{ij} \tag{A7}$$

此时，Gram 矩阵变为单位阵。最佳近似表示为，

$$s_{*}=\sum_{j=1}^{N}\left\langle f,\phi_{j}\right\rangle\phi_{j} \tag{A8}$$

附录 3 Gauss-Markov 定理

Gauss-Markov 定理：令$\mathbf{b}$是一个可表示为$\mathbf{b}=A\mathbf{x}+\mathbf{a}_{N}$的随机向量，其中$A$是$M\times P$矩阵（$M>P$）。其秩等于$P$；$\mathbf{x}$为一未知向量；$\mathbf{a}_{N}$为误差向量。若误差的均值等于零，方差为$\mathrm{Var}\left\{\mathbf{a}_{N}\right\}=\sigma^{2}I$（$\sigma^{2}$为常数），则存在一个方差最小的，无偏的估计子$\hat{\mathbf{x}}$。

全波形反演与波形层析成像的数值计算问题

王华忠 王雄文

波现象与反演成像研究组(WPI)，同济大学海洋与地球科学学院，上海，200092

【摘要】地震参数估计问题目前已经推进到高维全波形反演。至少从理论和实践上，高维全波形反演得到了充分的重视。它从理论上代表了地震波成像发展到了很高的层次。一般地，高维全波形反演从理论上归结为一个误差泛函的数值优化问题。它涉及到目标泛函的建立、迭代格式的导出、梯度的计算、Hessian 的讨论和具体的实现方案，其中还要考虑到正则化问题、迭代格式的收敛性和收敛效率问题。本文试图对上述问题做一个简明的总结。从数学理论和地震勘探的应用上建立一个全波形反演的实现框架。

【关键词】高维全波形反演，层析成像，最小二乘偏移，线性反演，非线性反演，误差泛函梯度，Hessian 矩阵

1 引言

全波形反演是勘探地震学中的引领性理论、方法和技术，这是毋庸置疑的。但是，全波形反演的完美理论并没有保证它的实用性。

旅行时层析成像和波形层析成像是全波形反演走向实用化的桥梁。它们本质上都是 Bayesian 框架下的非线性反演问题。但是，它们的作用是不同的。旅行时层析成像目标是估计速度场的背景成分；波形层析成像主要是估计速度场的扰动变化部分。全波形反演的目标是估计全频带的速度场。

我们知道，只有全频带的速度场（也包括其他参数场，譬如密度），才真正对储层预测起到直接的作用。但是，利用观测数据即使用全波形反演理论上也只能对速度场做到比较精确的估计，能否达到储层预测的要求还是很难预计的。

尽管全频带的速度场的估计非常困难，油气勘探实践的需求是无法回避的。很多的地球物理学家、数学家还是在不断地提出新的思想和方法来促进速度场的估计精度的提高。

层析成像与全波形反演的正过程描述和反演过程描述有很多的文献讨论过。我们在本文中对这两类问题做了详细的分析和对比。对其中的本质问题进行了归纳。指出了全频带的速度场估计的发展方向。在此基础上，我们较为详细地讨论了各种反演方法的数值计算问题。指出三种反演方法在数值计算上的共同之处就是利用数据残差的反传播估计参数修正量或梯度方向。数值计算上的另一个问题就是求解大规模的法方程或大规模的稀疏矩阵方程。解这类方程组的基本思想依然是迭代+误差反投影。直接解法在大规模的地球物理反演问题中基本无法使用，计算效率过低、存储量过大。

通过本文的分析，我们不仅仅想指出三类反演问题的数值计算方法，还想指明地震波反演到底如何做，才能实现计算效率高、计算精度高的双重目的。逐步接近估计全频带的速度场的目标。

2 全波形反演与波形层析成像的基本理论框架

2.1 线性化走时层析反演问题的数学描述

2.1.1 层析反演问题的一般性描述

在地震波反演成像中，旅行时反演是比较基础的工作。它主要刻画速度场的光滑部分，即背景速度场，为后续的更高波数的速度场的反演提供初始模型。

实质上，旅行时反演也完全可以在一般的框架下，即 Bayes 框架下展开讨论。此处，为了集中讨论反演问题的数值计算方法，我们还是沿用一般的方法来描述层析反演问题。旅行时反演的出发点是 Radon 变换，并且忽略预测误差的随机分布特征。假定预测误差是确定性的，预测误差与慢度扰动之间的关系是线性的，用误差反投影的方式直接估计慢度扰动。

旅行时与速度模型的关系可以用下面的积分方程进行表示

$$\begin{aligned} d^i &= g^i(\mathbf{m}) = \int_{R^i(\mathbf{m})} ds^i\left(\mathbf{x}^j,\mathbf{m}\right) m(\mathbf{x}^j) \\ &= \mathbf{ds}^i\left(\mathbf{x}^j,\mathbf{m}\right)\cdot\mathbf{m} \end{aligned} \tag{1}$$

其中 $\mathbf{m}$ 表示慢度场模型，$m\left(\mathbf{x}^j\right)$ 表示在点 $\mathbf{x}^j$ 处的慢度值，与速度场的关系为 $m\left(\mathbf{x}^j\right)=v^{-1}\left(\mathbf{x}^j\right)$，$ds^i\left(\mathbf{x}^j,\mathbf{m}\right)$ 表示在慢度模型 $\mathbf{m}$ 中第 i 根射线在网格 $\mathbf{x}^j$ 处的射线长度，$R^i(\mathbf{m})$ 代表在模型 $\mathbf{m}$ 中第 i 根射线的射线路径，d^i 是第 i 根射线旅行时。为了保证方程具有线性形式，通常会将模型的慢度值设为未知量，而不是将速度值本身高为未知量。因此，引入介质慢度扰动量后有，

$$\begin{aligned} g^i(\mathbf{m}_n+\delta\mathbf{m}) &= \mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}+\delta\mathbf{m}\right)\cdot\left(\mathbf{m}+\delta\mathbf{m}\right) \\ &= \begin{pmatrix} \mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)+\dfrac{\partial\mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)}{\partial\mathbf{m}}\delta\mathbf{m} \\ +O\left(\left\|\delta\mathbf{m}\right\|_2^2\right) \end{pmatrix}\cdot\left(\mathbf{m}+\delta\mathbf{m}\right) \\ &\approx \mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)\cdot\mathbf{m}+\left(J^T\mathbf{m}+\mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)\right)\cdot\delta\mathbf{m} \\ J &= \frac{\partial\mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)}{\partial\mathbf{m}} \end{aligned} \tag{2}$$

上式是对旅行时进行了 Taylor 展开并只保留一阶项的结果，其中 J 表示 Jacobi 矩阵。如果假设介质的小扰动不会引起射线路径太大的变化，即假设射线在光滑介质中走稳相路径，那么介质的小扰动带来的射线路径的扰动可以忽略不计，即 $J=0$。因此，上式可以化简为：

$$g^i(\mathbf{m}_n+\delta\mathbf{m}) = \mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)\cdot\mathbf{m}+\mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)\cdot\delta\mathbf{m} \tag{3}$$

因此，可以进一步得到慢度扰动量和旅行时扰动量之间的近似线性关系：

$$\begin{aligned} \delta d^i &= g^i\left(\mathbf{m}_n+\delta\mathbf{m}\right)-g^i\left(\mathbf{m}_n\right)=\mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)\cdot\delta\mathbf{m} \\ &= \int_{R^i(\mathbf{m})} ds^i\left(\mathbf{x}^j,\mathbf{m}\right)\delta m\left(\mathbf{x}^j\right) = \int_{R^i(\mathbf{m})} K\delta m\left(\mathbf{x}^j\right) \end{aligned} \tag{4}$$

因此，可以进一步得到慢度扰动量和旅行时扰动量之间的近似线性关系

$$\begin{aligned} \delta d^i &= g^i\left(\mathbf{m}_n+\delta\mathbf{m}\right)-g^i\left(\mathbf{m}_n\right)=\mathbf{s}^i\left(\mathbf{x}^j,\mathbf{m}\right)\cdot\delta\mathbf{m} \\ &= \int_{R^i(\mathbf{m})} ds^i\left(\mathbf{x}^j,\mathbf{m}\right)\delta m\left(\mathbf{x}^j\right) = \int_{R^i(\mathbf{m})} K\delta m\left(\mathbf{x}^j\right) \end{aligned} \tag{5}$$

上式是 Fredholm 第一类积分方程，其中矩阵 K 为该积分方程的核函数。上面的积分方程可以写成下面算子的形式

$$\delta d^i = \left(G_n\delta\mathbf{m}\right)^i = \int_{R^i(\mathbf{m}_n)} ds^i\delta m(\mathbf{x}^i) \tag{6}$$

当然也可以写成矩阵方程的形式。矩阵中的列向量可以视为展开慢度扰动函数的基函数族，在此坐标系中，存在如下的类 Parsaval 关系，

$$\begin{aligned} &\int_v dV(\mathbf{x})(G_n^T\delta\mathbf{d})\delta\mathbf{m}(\mathbf{x}) = \sum_i \delta\mathbf{d}^i(G_n\delta\mathbf{m})^2 \\ &= \sum_i \delta d^i \int_{R^i(\mathbf{m})} ds^i\delta\mathbf{m}\left(\mathbf{x}\right) \end{aligned} \tag{7}$$

当然，如果数据空间和模型空间是伴随的，存在伴随算子$(G_n)^*$把旅行时扰动投影成慢度扰动。

因此，综合（4）、（5）、（6）可得到旅行时反演算法如下：

$$\begin{aligned}\delta d_n^i &= \int\limits_{R(\mathbf{m}(\mathbf{x}^i))} ds^i \delta\mathbf{m}(\mathbf{x}^i) \\ \delta d_n^i &= \sum_j (C_D^{-1})^{i,j}\delta d_n^j \\ \delta\mathbf{m}_n(\vec{x}) &= \sum_j \delta\tilde{d}_n^j \psi^i(\mathbf{x})\end{aligned} \tag{8}$$

$$\gamma_n(\mathbf{x}) = \delta\mathbf{m}_n(\mathbf{x}) + (\mathbf{m}_n(\mathbf{x}) - \mathbf{m}_{prior}(\mathbf{x})) \tag{9}$$

$$\mathbf{m}_{n+1}(\mathbf{x}) = \mathbf{m}_n(\mathbf{x}) - \mu_n\gamma_n(\mathbf{x}) \tag{10}$$

$$\psi^i(\mathbf{x}) = \int\limits_{R_i(\vec{m}_n)} ds^i C_M(\mathbf{x}, \mathbf{x}(s_i)) \tag{11}$$

（7）第一式是一个基本的线性关系，它建立了旅行时扰动与慢度扰动之间的线性关系。可以用此关系形成稀疏矩阵，并用 ART/SIRT 方法解此矩阵方程，得到慢度扰动估计。这是基本层析成像的做法。

从更严谨的理论出发，层析反演过程应该按如下方式进行：首先，把慢度扰动量用加权算子（即数据空间的协方差矩阵）映射到其对偶空间；其次，用共轭转置算子把旅行时的对偶空间中的扰动量投影到慢度场的对偶空间中的慢度扰动量；最后，用模型加权算子（即模型空间的协方差矩阵）把对偶空间中的模型投影到模型空间。上述三次投影也可以用伴随算子一次完成。如果不考虑加权算子，即不考虑数据与模型的随机性，伴随算子与共轭转置算子等价，但此时得到的结果的精度和分辨率都会比较低。

很显然，（7）第三式就实现了加权反投影。数据加权函数和模型加权函数的定义是很有技巧的，它们的主要作用是提高反演结果的分辨率。实质上，它们分别与数据协方差矩阵和模型协方差矩阵的逆相对应。由于这两个协方差矩阵的不可知性或是不可逆性，我们可以从另外的角度理解和处理它。

层析速度反演可以用 beam-ray 思想修改（7）第三式的计算，多尺度反演思想也可以在这个过程中引入。

2.1.2 层析反演问题的矩阵方程

方程（4）是线性的，可以表达成如下的矩阵方程形式，

$$K_{M\times N}\delta\mathbf{m}_{N\times 1} = \delta\mathbf{d}_{M\times 1} \tag{12}$$

其中，M 代表射线条数；N 代表模型参数的总数（$NX\times NY\times NZ$）。很显然，矩阵方程（11）是不满秩的。当 $M > N$ 时方程是超定的，$M < N$ 时方程是欠定的。另外，该方程的系数矩阵是高度稀疏的，因为其中的非零元素代表射线穿过（$NX\times NY\times NZ$）个网格点中某些网格点的长度，单根射线形成一个方程。显然，单根射线仅仅穿过总网格点中的极少部分，矩阵每一个行向量中的非零元素分别与这些网格相对应。

对于勘探地震中的问题，一般教科书中讲的解矩阵方程的方法基本都不可行。解（11）式的基本思想是

$$\delta d^i = (G_n\delta\mathbf{m})^i = \int\limits_{R^i(\mathbf{m}_n)} ds^i \delta m(\mathbf{x}^i) \tag{13}$$

就是用系数矩阵 $K_{M\times N}$ 的转置作用于数据向量上，实现数据扰动向模型扰动的投影。具体计算时，我们可以对该转置矩阵做各种修改。ART 和 SIRT 方法就是实现的上述思想。

2.1.3 走时层析反演的迭代实现

（9）定义了走时层析反演的迭代实现公式。该公式实际上实现的是一个非线性的反演过程，该反演过程的每一个迭代步都是通过解一个线性化的矩阵方程来实现的，用一个简单的线性问题来逼近原来复杂的非线性问题。每次迭代中，整个模型得到了修改，在修改后的模型上计算新的射线和走时，形成新的系数矩阵。

事实上，每一个迭代步所解的线性化问题给出的是参数的扰动部分，或参数场的高波数成分。这部分高波数成分的求解也需要迭代完成。原因是数据的不完备以及参数选择不合适等。

因此，走时层析反演原则上应该用非线性迭代和线性迭代两重迭代来完成。

1.2 线性化波形层析反演问题的数学描述

1.2.1 Born 近似下散射波场的表达和对偶映射算子

符号规定

研究区域用 X 表示，区域内的点的坐标用 $\mathbf{x}$ 表示；其边界用 ∂X 表示；边界上的接收器坐标用 $\boldsymbol{\xi}$ 表示，炮点坐标用 $\boldsymbol{\eta}$ 表示。

因果关系

散射场是因果的，即 $u(t,\boldsymbol{\xi},\boldsymbol{\eta})=0$ 当 $t<0$ 时。因此有如下的 Fourier 变换，

$$\hat{u}(\omega,\boldsymbol{\xi},\boldsymbol{\eta})=\int_{-\infty}^{+\infty}u(\omega,\boldsymbol{\xi},\boldsymbol{\eta})e^{i\omega t}dt \tag{14}$$

Born 近似下的散射场表达，频率空间域的波动方程有如下的形式

$$\nabla^2\hat{u}(\omega,\mathbf{x})+S^2(\mathbf{x})\omega^2\hat{u}(\omega,\mathbf{x})=0 \tag{15}$$

在研究区域 X 内，介质的扰动关系可以表述为，

$$S^2(\mathbf{x})=S_0^2(\mathbf{x})+f(\mathbf{x}) \tag{16}$$

其中，$\mathrm{S}(\mathbf{x})$ 为介质的慢度，$\mathrm{S}_0(\mathbf{x})$ 为介质的背景慢度，$f(\mathbf{x})$ 为介质慢度的扰动，即散射源，或散射势。因此，散射波场可以表述为

$$\hat{u}(\omega,\boldsymbol{\xi},\boldsymbol{\eta})=-\omega^2\int_X\begin{pmatrix}G(\omega,\boldsymbol{\xi},\mathbf{x})f(\mathbf{x})\\u_{total}(\omega,\boldsymbol{\eta},\mathbf{x})\end{pmatrix}d\mathbf{x} \tag{17}$$

其中 $G(\omega,\boldsymbol{\xi},\mathbf{x})$ 表示背景慢度场的 Green 函数，该 Green 函数满足方程

$$\left(\nabla_{\mathbf{x}}^2+s_0^2(\mathbf{x})\omega^2\right)G(\omega,\mathbf{x})=\delta(\mathbf{x}-\boldsymbol{\xi}) \tag{18}$$

在 Born 近似假设条件下，有

$$\hat{u}(\omega,\boldsymbol{\xi},\boldsymbol{\eta})=-\omega^2\int_X\begin{pmatrix}G(\omega,\boldsymbol{\xi},\mathbf{x})f(\mathbf{x})\\u_{\mathrm{in}}(\omega,\boldsymbol{\eta},\mathbf{x})\end{pmatrix}d\mathbf{x} \tag{19}$$

其中 $u_{\mathrm{in}}(\omega,\boldsymbol{\eta},\mathbf{x})$ 表示背景慢度场下的入射波场。上式可以表示成 Fredholm 第一类积分方程的形式。

对偶空间、对偶算子与散射势估计的对偶表达

为了估计散射势 $f(\mathbf{x})$，把（1.18）式重写为，

$$\hat{u}(\omega,\boldsymbol{\xi},\boldsymbol{\eta})=-\omega^2\int_X \mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})f(\mathbf{x})d\mathbf{x} \tag{20}$$

其中，$\mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$ 是积分核函数。它代表脉冲源激发的波从震源到散射点，再到接收点的波传播过程后的响应。

Fredholm 第一类积分方程的求解是很困难的，主要是因为这类积分方程所对应的矩阵方程在大部分情况下条件数都很大。对于这类积分方程也可以用如下的对偶空间理论来近似求解。在选定的基函数下面，数据空间和模型空间可以认为是一个对偶空间。Born 近似下的 Green 函数就定义了这样一个基函数，它们也形成了一个映射 $\mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$，把模型空间函数 $f(\mathbf{x})$ 映射到数据空间函数 $\hat{u}(\omega,\boldsymbol{\xi},\boldsymbol{\eta})$。在对偶空间框架下存在映射 $\mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$ 的对偶映射 $\hat{\mathrm{M}}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$，定义下式，

$$\begin{aligned}&\left\langle u(\omega,\boldsymbol{\xi},\boldsymbol{\eta}),M(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})f(\mathbf{x})\right\rangle_{S_D}\\&\triangleq\left\langle f(\mathbf{x}),\hat{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})u(\omega,\boldsymbol{\xi},\boldsymbol{\eta})\right\rangle_{S_M}\end{aligned} \tag{21}$$

此处，S_D 代表数据空间；S_M 代表模型空间。上式是广义的 Parsaval 定理。

基于上述对偶映射，可以直接地形式地写出扰动的估计式

$$f_{est}(\mathbf{x})=2\,\mathrm{Re}\int_{\partial X_{rec}}\int_0^\infty \hat{M}(\omega,\xi,\eta,x)\tilde{u}(\omega,\xi,\eta)d\xi d\omega \tag{22}$$

上式就是利用单炮数据估计参数扰动的公式。对于某些积分核函数，有

$$\bar{\hat{\mathrm{M}}}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})=\hat{\mathrm{M}}(-\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) \tag{23}$$

则（1.21）式就变为，

$$f_{est}(\mathbf{x})=\int_{\partial X_{rec}}\int_0^\infty \hat{M}(\omega,\xi,\eta,x)\tilde{u}(\omega,\xi,\eta)d\xi d\omega \tag{24}$$

1.2.2 Born 近似下散射势的估计方法

用下面任意两个实函数以及它们的 Fourier 变换之间的关系，

$$\int_{-\infty}^{+\infty} g(t)h(t)dt=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\bar{\tilde{g}}(\omega)\tilde{h}(\omega)d\omega \tag{25}$$

因此，有

$$\begin{aligned}M(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})&=2\,\mathrm{Re}\int_0^{+\infty}\tilde{\bar{M}}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})e^{-i\omega t}dt\\\tilde{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}M(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})e^{-i\omega t}dt\end{aligned} \tag{26}$$

由（21）、（23）式可知，扰动参数估计问题变为了确定投影算子或共轭投影算子问题。可以给出如下的投影算子，

$$M\left(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{8\pi^2}\delta\left(t-\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)-\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)\frac{h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}$$
$$\tilde{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{16\pi^3}\frac{e^{-i\omega\left(\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)+\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)}}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right) \tag{27}$$

上式是时间域形式，将其转为频率域形式。

其中，相位函数和振幅函数首先满足如下的 Green 函数的射线近似解，即 WKBJ 解，其中接收点对应的 Green 函数的解为

$$\mathrm{G}^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)=-A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)e^{i\omega\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)} \tag{28}$$

炮点对应的 Green 函数的解为 $\mathrm{G}^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)=-A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)e^{i\omega\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}$

相位函数 $\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$ 和 $\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$ 满足如下的程函方程，

$$\left(\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)^2=S_0^2\left(\mathbf{x}\right)$$
$$\left(\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\right)^2=S_0^2\left(\mathbf{x}\right) \tag{29}$$

$\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$ 代表从炮点位置 $\boldsymbol{\eta}$ 到散射点 $\mathbf{x}$ 的在背景介质中的旅行时。$\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$ 代表从散射点 $\mathbf{x}$ 到接收点 $\boldsymbol{\xi}$ 的在背景介质中的旅行时。振幅函数 $A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$ 和 $A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$ 满足如下一阶输运方程，

$$A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\nabla_{\mathbf{x}}^2\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)+2\nabla_{\mathbf{x}}A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\cdot\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)=0$$
$$A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\nabla_{\mathbf{x}}^2\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)+2\nabla_{\mathbf{x}}A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\cdot\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)=0 \tag{30}$$

旅行时的计算可以直接解方程（28）或把上式变成各种形式的一阶常微分方程组。关于振幅项对（29）式的求解直接在笛卡尔坐标系中进行会遇到焦散问题。Maslov 方法（在相空间中解方程（29））和 Gaussian_Beam 方法（在射线中心坐标系中用傍轴射线理论解方程（29））会回避掉此问题。其他各种变种的方法也在不断地发展中。

（26）中的函数 $h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 是一个 Jacobian 变换，它把对 $\mathbf{x}$ 的积分变换为对 $\left(\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 的积分。Beylkin（1985）给出

$$h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)d\xi=S_0^3\left(1+\cos\psi\right)d\Omega \tag{31}$$

其中，Ω 是立体角。且有

$$\cos\psi\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)=\frac{\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\cdot\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\xi}\right)}{S_0^2\left(\mathbf{x}\right)} \tag{32}$$

$\psi\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 是由坐标 $\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 和介质分布决定的波传播的反射张角，即入射射线和出射射线的夹角。

（26）式定义了变背景和任意炮检点分布时的偏移成像中的积分核函数。事实上，（26）式中的 Green 函数还可以不用 WKBJ 近似来表示，此时，推出的积分核意义更广泛。则有下式

$$\tilde{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{16\pi^3}\frac{\hat{G}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)}{\left|\hat{G}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)\right|}\frac{\hat{G}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)}{\left|\hat{G}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)\right|}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right) \tag{33}$$

在时间域中，上式变为，

$$M\left(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\left(\begin{matrix}\breve{G}^{out}\left(t-\tau,\boldsymbol{\xi},\mathbf{x}\right)\breve{G}^{in}\left(t-\tau,\boldsymbol{\eta},\mathbf{x}\right)\\ h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)\end{matrix}\right)d\tau \tag{34}$$

其中，

$$\begin{aligned}\breve{G}^{out}\left(t,\boldsymbol{\xi},\mathbf{x}\right)&=2\,\mathrm{Re}\int_{-\infty}^{+\infty}\frac{G^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)}{\left|G^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)\right|^2}e^{-i\omega t}d\omega\\ \breve{G}^{in}\left(t,\boldsymbol{\eta},\mathbf{x}\right)&=2\,\mathrm{Re}\int_{-\infty}^{+\infty}\frac{G^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)}{\left|G^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)\right|^2}e^{-i\omega t}d\omega\end{aligned} \tag{35}$$

可以写出如下两种不同的积分形式，对应两种不同的偏移成像的说明。形式 I 为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{8\pi^2}\int_{\partial X_{rec}}\left(\begin{matrix}u\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)\big|_{t=\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)+\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}\\ \dfrac{h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}\end{matrix}\right)d\boldsymbol{\xi}d\omega \tag{36}$$

很显然，上式就是 Kirchhoff 积分的绕射叠加。就是说，从 Born 近似理论也可以推出 Kirchhoff 积分绕射叠加偏移公式。

把（26）或（32）式代入到（33）式可以得到另一种形式的偏移成像公式。形式 II 为

$$\begin{aligned}f_{est}\left(\mathbf{x}\right)&=-\frac{1}{16\pi^3}\mathrm{Re}\int_0^{\infty}\frac{e^{-i\omega\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}}{A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}\int_{\partial X_{rec}}\frac{e^{-i\omega\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)\tilde{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\omega\\ f_{est}\left(\mathbf{x}\right)&=-\frac{1}{16\pi^3}\mathrm{Re}\int_0^{\infty}\frac{\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\eta}\right)}{\left|\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\eta}\right)\right|^2}\int_{\partial X_{rec}}\frac{\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\xi}\right)}{\left|\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\xi}\right)\right|^2}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)\tilde{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\omega\end{aligned} \tag{37}$$

（36）对应的时域形式为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{8\pi^2}\int_0^{\infty}G^{in}\left(t,\mathbf{x},\boldsymbol{\eta}\right)w\left(t,\mathbf{x},\boldsymbol{\eta}\right)dt \tag{38}$$

其中，

$$w\left(t,\mathbf{x},\boldsymbol{\eta}\right)=\int_0^{\infty}\int_{\partial X_{rec}}G^{out}\left(t-\tau,\mathbf{x},\boldsymbol{\xi}\right)h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)u\left(\tau,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\tau \tag{39}$$

（38）式的含义就是（39）式进行检波点波场反传播，然后与下行波场进行相关成像。此为 Clearbout（1985）的成像条件。

（35）、（36）和（38）分别是不同形式的叠前偏移成像表达式。成像结果是地下介质的扰动（散射势）。上述三式用更深刻的数学背景阐述了地震波叠前偏移成像的本质。它就是 Born 近似下线性化反演的近似解。

1.2.3 线性化波形层析反演的矩阵表达

Born 近似线性化的 Fredholm 第一类积分方程（19）的求解还可以用矩阵方程求解的方式。方程（19）对应的矩阵方程可以写为

$$K_{M\times N}\delta\mathbf{m}_{N\times 1}\approx\delta\mathbf{u}_{M\times 1} \tag{40}$$

其中，M 代表炮检对的个数，即地震道的道数；N 代表模型参数的总数（$NX\times NY\times NZ$）。在频率空间域，$\delta\mathbf{u}_{M\times 1}$ 代表所有炮检对观测单频散射波场。由方程（18）可以看出，$K_{N\times M}$ 中元素代表每个炮检对和

每个地下网格点（默认它是散射点）形成的入射 Green 函数和出射 Green 函数的乘积再乘以 $-\omega^2$。因此，它总计有 M 行 N 列。该矩阵也是可非常稀疏的矩阵。一般而言，（39）是一个欠定的方程。如果模型剖分合适，减少参数的个数，它也可能是超定的。

同样地，（39）式的求解也可以用转置

矩阵 $K_{N\times M}^T$ 作用于散射场数据矢量上估计模型参数扰动，即

$$\delta\mathbf{m}_{N\times 1} \approx K_{N\times M}^T \delta\mathbf{u}_{M\times 1} \tag{41}$$

更多情况下，我们最小二乘方法求解方程（39）。首先建立如下二次型误差泛函，

$$\left\|e\right\|_2^2 = \left\|K_{M\times N}\delta\mathbf{m}_{N\times 1} - \delta\mathbf{u}_{M\times 1}^{obs}\right\|_2^2 \tag{42}$$

我们知道，误差泛函达到最小时得到的参数估计结果由如下法方程得到，

$$\left(K_{N\times M}^T K_{M\times N}\right)\delta\mathbf{m}_{N\times 1} = K_{N\times M}^T \delta\mathbf{u}_{M\times 1}^{obs} \tag{43}$$

其中，$H = K_{N\times M}^T K_{M\times N}$ 称为 Hessian 矩阵。它是一个非常庞大的、稀疏的矩阵。一般地，可以用共轭梯度法求解。

2.2.4 线性化波形层析反演的迭代实现

同样地，线性化的波形层析所估计的是速度场的扰动变化部分或高波数变换部分。最小二乘叠前偏移实际上就是线性化的波形层析，只不过我们常常把最小二乘叠前深度偏移解释为估计反射系数，实际上它估计的并不是反射系数，而是（15）定义的慢度平方差 $f(\mathbf{x})$。

因此，根据（43）定义的模型参数扰动量估计式，可以写出非线性的模型参数迭代公式，

$$\begin{aligned}\mathbf{m}_{N\times 1}^k &= \mathbf{m}_{N\times 1}^k + \mu^k \delta\mathbf{m}_{N\times 1} \\ &= \mathbf{m}_{N\times 1}^k + \mu^k \left(K_{N\times M}^T K_{M\times N}\right)^{-1} K_{N\times M}^T \delta\mathbf{u}_{M\times 1}^{obs}\end{aligned} \tag{44}$$

此时，矩阵 K 的计算利用当前迭代步的模型。（44）式也可以改写成梯度导引的形式，

$$\delta d^i = \left(G_n \delta\mathbf{m}\right)^i = \int_{R^i(\mathbf{m}_n)} ds^i \delta m(\mathbf{x}^i) \tag{45}$$

这样，可以更清楚地看出非线性迭代中每一步是如何用线性方法实现的。也可以看出最小二乘偏移成像与全波形反演的密切关系。这是用波形层析实现宽频带速度估计的理论解释。

2.3 非线性全波形反演问题的数学描述

2.3.1 非线性全波形反演中的正问题描述方法

非线性正问题的数学描述可以抽象为

$$A(\mathbf{m})\mathbf{u} = \mathbf{f} \tag{46}$$

其中，A 为正演算子。在地震波反演中，A 的形式为

$$A = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \Delta \tag{47}$$

A 的形式受所选择的波动方程的形式的影响。可以选常密度声波，变密度声波方程，各种拟声波方程，弹性波方程等。$\mathbf{u}$ 为方程（46）要求解的未知量，表示正演合成的波场，$\mathbf{f}$ 为震源函数。

野外观测到的地震数据可以表述为

$$\mathbf{d} = C\mathbf{u} + \boldsymbol{\eta} \tag{48}$$

此处，C 表示地震数据的采样算子，由观测系统决定。它使得连续的波场 $u(x,y,z,t)$ 变得规则离散化（或非规则离散化）及孔径有限化。采样算子的状态受实际条件的制约。η 表示噪音函数，在大部分情况下假设它是高斯白噪的。

连续波场受采样算子制约后，采样后的数据可以记为

$$[C\mathbf{u}] \approx \mathbf{u}(\mathbf{x}_i), \quad i = 1,2,\cdots,n \tag{49}$$

其中 $\mathbf{x}_i$ 表示第 i 个采样点的坐标。根据方程（46），可以将连续波场的正演计算形式地表达为

$$\mathbf{u} = \left(A(m)\right)^{-1}\mathbf{f} \tag{50}$$

2.3.2 非线性全波形反演的数学描述

基于前面对正问题的定义，根据数据残差最小的原则可以建立如下的优化问题：

$$\begin{aligned} &\min: \quad J(\mathbf{m}) \\ &S.T.: \quad C\mathbf{u} = \mathbf{d} \end{aligned}$$

可以用如下的方式定义上面优化问题所对应的误差泛函，

$$\min_{\mathbf{u}\in\mathcal{H}_2,\mathbf{m}\in\mathcal{H}_1} \frac{1}{2}\left\|C\mathbf{u}-\mathbf{d}\right\|_Y^2 + \alpha J(\mathbf{m}) \tag{51}$$

上式中的第一项代表正演波场与观测波场的波场残差，后一项代表对模型的先验约束，α 是正则化系数。一般地，我们可以通过假设模型中的元素满足某种概率分布条件来施加对模型的先验限制。把（50）式代入上式，得到

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2} \mathfrak{I}(\mathbf{m}) = \frac{1}{2}\left\|F(\mathbf{m})-\mathbf{d}\right\|_Y^2 + \alpha J(\mathbf{m}) \tag{52}$$

其中，与观测数据逼近的正演计算结果可以表达为

$$F(\mathbf{m}) = C\left(A(\mathbf{m})\right)^{-1}\mathbf{f} \tag{53}$$

将上述结果代入（52）式后，得到最后的泛函定义式，

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2} \mathfrak{I}(\mathbf{m}) = \frac{1}{2}\left\|C\left(A(\mathbf{m})\right)^{-1}\mathbf{f}-\mathbf{d}\right\|_Y^2 + \alpha J(\mathbf{m}) \tag{54}$$

通过此式，可以通过各种优化方法求解参数估计问题。假设上述定义的泛函至少是 Fréchet 二阶可微的，那么，可以对该泛函做如下的 Taylor 展开，

$$\begin{aligned} \mathfrak{I}(\mathbf{m}+\delta\mathbf{m}) &= \mathfrak{I}(\mathbf{m}) + \left\langle \mathrm{Grad}\mathfrak{I}(\mathbf{m}), \delta\mathbf{m}\right\rangle + \\ &\frac{1}{2}\left\langle \mathrm{Hess}\mathfrak{I}(\mathbf{m})\delta\mathbf{m}, \delta\mathbf{m}\right\rangle + o\left(\left\|\delta\mathbf{m}\right\|^3\right) \end{aligned} \tag{55}$$

如果泛函 $\mathfrak{I}(\mathbf{m}):\mathbb{R}^n \to \mathbb{R}$ 有连续的二阶偏导数，那么 $\mathrm{Grad}\mathfrak{I}(\mathbf{m})$ 是一个 n 维向量，表示泛函的梯度方向，$\mathrm{Hess}\mathfrak{I}(\mathbf{m})$ 是一个 $n\times n$ 的方阵，称为 Hessian 矩阵。其中，n 为模型参数的个数，m 为观测数据的个数。

定理：如果泛函 $\mathfrak{I}(\mathbf{m}):\mathbb{R}^n \to \mathbb{R}$ 在 $\mathbf{m}_0$ 点的某个邻域 $\Omega(\mathbf{m}_0,\delta)$ 内同时满足以下几个条件：

1. $\Im(\mathbf{m})$ 在该邻域内是 Fréchet 二阶可微的；

2. 梯度 $\mathrm{Grad}\Im(\mathbf{m}_0)=0$；

3. Hessian 矩阵 $\mathrm{Hess}\Im(\mathbf{m}_0)$ 是强正定的，那么 $\mathbf{m}_0$ 是泛函 $\Im(\mathbf{m}):\mathbb{R}^n\to\mathbb{R}$ 的严格极小点。

从此可以看出，首先根据（54）和（55）式求极小来估计模型参数值，得到的是局部极值点。另外，只有 Hessian 矩阵 $\mathrm{Hess}\Im(\mathbf{m})$ 是强正定的，才可以得到严格极小点。否则，估计结果不是真正的局部极小点，更不用讲全局极小点了。这进一步说明了目前的全波形反演方法对初始模型的依赖性。

2.3.3 全波形反演的地球物理学含义

为了进行全波形反演估计模型参数，首先定义如下的二次型误差泛函

$$\Im(\mathbf{m})=\frac{1}{2}f(\mathbf{m})^T f(\mathbf{m}) \tag{56}$$

其中，$f(\mathbf{m})$ 为观测数据与正演合成数据的拟合差或残差。误差泛函是一个映射 $\Im:\mathbb{R}^n\to\mathbb{R}$。我们主要考虑全波形反演的数值计算问题，关于正则化方面的考虑在此忽略。

参数估计问题就是找到一个参数 $\mathbf{m}$ 使误差泛函达到极小

$$\min_{\mathbf{m}\in\mathbb{R}^n}\varphi(\mathbf{m})=\min_{\mathbf{m}\in\mathbb{R}^n}\frac{1}{2}f(\mathbf{m})^T f(\mathbf{m}) \tag{57}$$

上述求极值问题，一般地归结为无约束最优化问题。典型的解法是 Gauss-Newton 方法和全 Newton 方法。

Gauss-Newton 迭代公式的推导

为了导出 Gauss-Newton 迭代公式，定义误差泛函的梯度并令其等于零

$$\mathrm{Grad}\Im(\mathbf{m})=Df(\mathbf{m})^T f(\mathbf{m})=0 \tag{58}$$

对拟合差函数 $f(\mathbf{m})$ 进行 Taylor 展开，仅仅保留线性项，有

$$f(\mathbf{m})\approx f(\mathbf{m}^*)+Df(\mathbf{m}^*)(\mathbf{m}-\mathbf{m}^*) \tag{59}$$

把上式代入梯度公式（58）

$$\begin{aligned}&Df(\mathbf{m}^*)^T\left(f(\mathbf{m}^*)+Df(\mathbf{m}^*)(\mathbf{m}-\mathbf{m}^*)\right)\\&=Df(\mathbf{m}^*)^T f(\mathbf{m}^*)+Df(\mathbf{m}^*)^T Df(\mathbf{m}^*)(\mathbf{m}-\mathbf{m}^*)\\&=0\end{aligned} \tag{60}$$

整理上式，得到

$$\mathbf{m}-\mathbf{m}^*=-\left(Df(\mathbf{m}^*)^T Df(\mathbf{m}^*)\right)^{-1}Df(\mathbf{m}^*)^T f(\mathbf{m}^*) \tag{61}$$

最后，得到 Gauss-Newton 迭代格式

$$\mathbf{m}^{k+1}=\mathbf{m}^k-\left(Df(\mathbf{m}^k)^T Df(\mathbf{m}^k)\right)^{-1}Df(\mathbf{m}^k)^T f(\mathbf{m}^k) \tag{62}$$

实际上，任何一个迭代格式应该有两个主要成分，一是梯度方向；另一个是步长。（62 式可以理解为，用线性 Hessian 的逆对梯度进行预条件处理，目的是微调梯度方向使收敛过程加速。明显地，我们可以引

入一个步长控制因子，进一步加快收敛过程。但步长的选择方法很关键。一维搜索是常用的步长计算方法。（62 以重写为

$$\mathbf{m}^{k+1}=\mathbf{m}^{k}-\mu^{k}\left(Df\left(\mathbf{m}^{k}\right)^{T}Df\left(\mathbf{m}^{k}\right)\right)^{-1}Df\left(\mathbf{m}^{k}\right)^{T}f\left(\mathbf{m}^{k}\right) \tag{63}$$

其中，μ 为引入的步长控制因子。

全 Newton 迭代公式的推导

为了导出全Newton迭代公式，假设 $\hat{\mathbf{m}}$ 是泛函 $\mathfrak{I}(\mathbf{m}):\mathbb{R}^{n}\to\mathbb{R}$ 的最小化估计值，则泛函 $\mathfrak{I}(\mathbf{m}):\mathbb{R}^{n}\to\mathbb{R}$ 的在 $\hat{\mathbf{m}}$ 附近的二次 Taylor 展开为

$$Q(\delta\mathbf{m})=\mathfrak{I}(\hat{\mathbf{m}})+\left\langle \mathrm{Grad}\mathfrak{I}(\hat{\mathbf{m}}),\delta\mathbf{m}\right\rangle+\frac{1}{2}\left\langle \mathrm{Hess}\mathfrak{I}(\hat{\mathbf{m}})\delta\mathbf{m},\delta\mathbf{m}\right\rangle \tag{64}$$

若 Hessian 矩阵 $\mathbf{Hess}\mathfrak{I}(\hat{\mathbf{m}})$ 是正定的，则

$Q(\mathbf{s})$ 存在唯一的极小值点，满足

$$\mathrm{Grad}\mathfrak{I}(\hat{\mathbf{m}})+\mathrm{Hess}\mathfrak{I}(\hat{\mathbf{m}})\delta\mathbf{m}=0 \tag{65}$$

事实上，令.$\partial Q/\partial\delta\mathbf{m}=0$.就可以得到上式。取

$$\hat{\mathbf{m}}^{k+1}-\hat{\mathbf{m}}^{k}=\delta\mathbf{m} \tag{66}$$

得到全 Newton 迭代格式

$$\hat{\mathbf{m}}^{k+1}=\hat{\mathbf{m}}^{k}-\left[\mathrm{Hess}\mathfrak{I}\left(\hat{\mathbf{m}}^{k}\right)\right]^{-1}\mathrm{Grad}\mathfrak{I}\left(\hat{\mathbf{m}}^{k}\right) \tag{67}$$

全 Newton 迭代格式中 Hessian 的计算是没有近似的，它的逆对梯度的预条件应该更精确，使得迭代的收敛性好于 Gauss-Newton 迭代格式。关于步长因子的引入，与 Gauss-Newton 迭代格式的考虑是一样的。（67）写为，

$$\hat{\mathbf{m}}^{k+1}=\hat{\mathbf{m}}^{k}-\mu^{k}\left[\mathrm{Hess}\mathfrak{I}\left(\hat{\mathbf{m}}^{k}\right)\right]^{-1}\mathrm{Grad}\mathfrak{I}\left(\hat{\mathbf{m}}^{k}\right) \tag{68}$$

其中，μ 为引入的步长控制因子。

2.4 模型空间搜索机制评价

总结上述三种典型的地球物理反演方法：走时层析、波形层析和全波形反演，它们的数值计算都可以归结一个迭代公式。因此，我们可以认为，目前的地球物理反演所做的数值运算可以抽象为按照一种搜索机制在所有可能的参数模型构成的模型空间中进行采样。采样到“最佳”的模型参数，就使得上述定义的范数达到极小。

目前的搜索机制可以抽象为：从一个初始值出发，在梯度方向的导引和步长因子约束下，向误差泛函下降的方向在模型空间进行采样。这样的采样是符合 Markov 链准则的。很显然，这样的采样机制不是最好的采样机制，既不是最高效的，也不是最有效的。不是有效的，指的是它的初值依赖性和梯度导引性，这样的机制只能导引到局部极值点。不是高效的，指的是在一个巨大的模型空间中搜索，应该依靠群体效应（类似蚂蚁寻迹方法），从不同的点出发来寻找最优点，而且不是或不完全是靠梯度方向来决定寻优方向，也许应该是吉布斯采样或 Metropolis 采样与梯度导引的某种结合。

总之，随着计算机能力的极大提高，目前梯度导引类的、依赖于初始值的反演方法很快将被随机采样类的方法所取代。

2.5 透射波走时层析/绕射波成像与反射波走时层析/反射波波形层析及全波形反演与波现象和介质描述方法的关系分析

在勘探地震学中，透射波走时层析成像用到的波路径从源点直接到检波点，中间发生的波现象仅仅是透射，穿过的介质是光滑背景介质，不考虑高波数介质扰动的存在。井间地震波层析，某些 3DVSP 数据层析，地表折射波层析、Diving-Wave 层析都是透射波层析。透射波层析可以直接估计速度场的低波数变化，它利用的是不同角度的射线路径的旅行时。反射波走时层析成像用到的波路径是从源点经反射后到达接收点，这首先需要高波数模型参数扰动的存在，否则不可能出现反射现象。两者波路径的差异很明显，路径的差异带来的是照明角度及射线覆盖范围的差异。射线的覆盖角度决定了走时层析的反演精度。反射波层析的一个重要特征是其需要高波数模型成分描述反射现象，在实际的反射层析中，必须要额外加入模型的反射界面信息，这是在实际层析反演中必须注意的。

绕射波波形层析与反射波波形层析理论上完全相同，名称上的差异首先体现在对介质参数描述方式的不同。绕射波波形层析认为地下介质是由光滑的背景加上孤立的绕射点形成；反射波波形层析认为地下介质是由反射界面、界面间的速度背景变化和/或速度背景中的孤立绕射点形成。理论上，二者都是基于 Born 近似，认为波场由一次散射波的叠加形成，所有高阶散射场都被忽略。目前的叠前深度偏移成像（尤其是最小二乘 PSDM）都是基于此理论，也说明了 PSDM 本身就是波形层析的一种实现方式。反射波波形层析成像时，要求有反射界面的存在，这样才能产生反射波。做反射波波形层析成像要求建立带反射界面的初始模型，这是地下速度模型受地质沉积与构造模式约束的特点决定的。另一方面，反射界面的存在可以对绕射体产生更宽角度的照明，提高反射波波形层析对小尺度异常体的反演精度。目前所谓的高分辨率层析成像（High-definition Tomography）讲的就是这个意思。

很显然，理论上全波形反演对波现象的描述是最完整的，它不依赖 Born 线性化近似，可以得到全频带的速度场。但是，描述全部的波现象需要全频带的速度场。基于光滑的初始模型的双向波正演不能描述全波场，这就难以发挥出全波形反演的优势。RTM 也存在这样的现象，理论上 RTM 可以对多次波成像，但多次波成像需要全频带的速度场。这又陷入到地球物理学中常见的“由未知求未知”的怪圈。

根据上述分析，可以清楚地看到，不同反演方法对模型描述方式的要求不同。透射波场、反射波场、一次波场、多次波场、全波场的正演描述与模型的描述方法紧紧联系在一起。对应地，各种反演方法就有了不同的特点。全波形反演是要把上述各种反演方法有机地串联在一起来实现，而不是仅利用 FWI 算法一步估计出全频带的速度场。

2.6 走时层析、波形层析和全波形反演方法对比分析

（11）和（9）分别定义了走时层析的法方程和迭代格式；（1.42）（1.44）分别定义了波形层析的法方程和迭代格式。（1.62）和（1.66）分别定义了全波形反演的迭代格式。

对比分析勘探地震学中三种反演方法的基本求解公式可以发现，它们有极大程度的相似性。三者都可以实现非线性的全频带速度场反演（外部迭代过程），非线性反演一般通过多步的线性化反演实现（线性化反演可以构造成内部迭代过程）。全波形反演似乎没有线性化的内部迭代过程，实际上求梯度这一步可以转化为内部迭代来实现。线性化反演主要确定介质参数的扰动部分，非线性反演试图确定“全”频带的速度场。介质参数的扰动发生在反射界面处，因此线性化反演都类似于一个“偏移”过程。迭代的线性化反演就是在做某种剩余偏移，从而不断修改介质参数的扰动量。扰动量修改的收敛性与介质参数的背景变化有密切关系，没有合适的介质参数背景变化，介质参数的扰动不可能估计准确。因此，两步迭代格式套用由波在介质中的传播实际决定。

另一个相似的地方是三种反演方法的数值求解方法非常类似，一是解大规模系数矩阵方程---法方程；二是利用梯度方向修正模型参数，梯度的估算通过残差的反传播实现。

它们的差异之处在于：层析反演的正问题描述的是一个线性化问题，它仅仅描述一次波的传播过程。这是层析反演方法不能反演速度场精细变化的根本原因。全波形反演的正问题描述了一个非线性正过程，

它描述了全波场的传播，既包括一次波也包括多次波。但是，这仅仅是理论。实际上，由于初始速度场中缺乏高波数成分，多次波是模拟不出来的，即使模拟出也可能不正确。因此，把波形层析推向考虑全波场，既包括一次波也包括多次波，是精确速度估计的方向。FWI 尽管理论上很完美，实际上不是一个可行的速度估计方法。

3 全波形反演与波形层析成像数值计算

根据前面的讨论，地震勘探中三种关键的反演方法：走时层析、波形层析和全波形反演的数值计算问题就比较明确了。

抽象地讲，用转置矩阵进行数据空间向模型空间的投影来估计模型参数的扰动量，实际上就是梯度方向，是最基本的方法。用 Hessian 的逆预条件梯度方向也是计算过程的关键步骤。计算步长因子是另一个关键步，这是全波形反演和层析反演数值计算的主流方法。它们是核心内容是数值最优化问题的求解。

另外，用最小二乘方法求解线性反演问题可以归结为求解法方程。这往往是一个巨大的、稀疏的矩阵方程。大规模稀疏矩阵的求解在全波形反演和层析反演中占据着重要地位，大规模稀疏矩阵的求解是数值代数中比较主要的内容。

3.1 走时层析成像问题的数值计算

走时层析主要是求解如下方程

$$\delta \mathbf{m}_{M\times 1} \approx K^{T}_{M\times N}\delta \mathbf{d}_{N\times 1} \tag{69}$$

它的地球物理含义是把旅行时扰动 $\delta \mathbf{d}_{N\times 1}$ 通过转置矩阵 $K^{T}_{M\times N}$ 反投影到模型空间得到模型参数扰动量 $\delta \mathbf{m}_{M\times 1}$ 的估计。M 是描述整个慢度场的网格点数，N 是射线条数。它的具体计算方法就是 ART 和 SIRT 方法。（2.1）式实现的是直接反投影，它的精度不高。ART 和 SIRT 方法实现的是加权反投影，提高了估计的精度。

3.1.1 ART 方法

上式表示的矩阵方程可以具体地写成如下方程组，

$$\begin{aligned}
\delta t^{1} &= l^{1}_{1}\delta s_{1} + l^{1}_{2}\delta s_{2} + \cdots + l^{1}_{M}\delta s_{M} \\
\delta t^{2} &= l^{2}_{1}\delta s_{1} + l^{2}_{2}\delta s_{2} + \cdots + l^{2}_{M}\delta s_{M} \\
\delta t^{N} &= l^{N}_{1}\delta s_{1} + l^{N}_{2}\delta s_{2} + \cdots + l^{N}_{M}\delta s_{M}
\end{aligned} \tag{70}$$

其中，l^{i}_{j} 是第 i 根射线在网格 j 中的射线长度；δs_{j} 是网格 j 中的慢度扰动量；M 是整个模型的网格点的总数。δt^{i} 是沿第 i 根射线的旅行时扰动量。射线的总条数为 N. 因此，上式的矩阵形式为，

$$A_{N\times M}\delta \mathbf{s}_{M\times \mathbf{1}} = \delta \mathbf{t}_{N\times 1} \tag{71}$$

其中 $\delta \mathbf{s}_{M\times 1}$ 是一个 M 维列向量，$\delta \mathbf{t}_{N\times 1}$ 是一个 N 维列向量。

由于每根射线仅仅穿过 M 个总网格中的很少几个网格点，因此，每个投影方程中的绝大部分系数（即射线长度）是零。系数矩阵 $A_{N\times M}$ 是一个高度稀疏的矩阵。

$\delta \mathbf{s}_{M\times 1}$ 的估计可以通过用系数矩阵 $A_{N\times M}$ 的转置矩阵反传播旅行时残差得到。即

$$\delta \mathbf{s}_{M\times \mathbf{1}} = A^{T}_{M\times N}\delta \mathbf{t}_{N\times 1} \tag{72}$$

就是说，慢度扰动可以通过叠加所有通过一个特定点的射线的投影（即旅行时残差）得到。其中，$A^{T}_{M\times N}$ 是系数矩阵的转置。

定义穿过第 j 个单元的所有射线的长度总和为

$$L_j = \sum_{i=1}^{N} l_j^i, j = 1,2,\cdots,M, \tag{73}$$

第 i 根射线的总长度为 $\Gamma^i = \sum_{j=1}^{M} l_j^i$。利用第 i 根射线的总长度对相应射线的投影值（旅行时残差）进行归一化或加权处理，得到

$$\delta\hat{t}^i = \frac{\delta t^i}{\sum_{j=1}^{M} l_j^i} = \frac{\delta t^{\ i}}{\Gamma^i} \tag{74}$$

利用穿过第 j 个单元的所有射线的长度总和对响应网格点的慢度值进行加权处理，就可以得到方程组

$$\begin{aligned} L_1\delta s_1 &= l_1^1\delta\hat{t}^1 + l_1^2\delta\hat{t}^2 + \cdots + l_1^N\delta\hat{t}^N \\ L_2\delta s_2 &= l_2^1\delta\hat{t}^1 + l_2^2\delta\hat{t}^2 + \cdots + l_2^N\delta\hat{t}^N \\ L_N\delta s_M &= l_M^1\delta\hat{t}^1 + l_M^2\delta\hat{t}^2 + \cdots + l_M^N\delta\hat{t}^N \end{aligned} \tag{75}$$

其中，L_j 是一个行加权因子，Γ^i 是一个列加权因子。方程（2.6）的矩阵形式为

$$L\delta\mathbf{s}_{M\times 1} = A_{M\times N}^T\delta\hat{\mathbf{t}}_{N\times 1} \tag{76}$$

可以清楚地看到此处的反投影与用波场外推算子进行残差波场的反传播有十分相似的意义。离散反投影方法是一个高效的方法，但是反演精度要差。该方法可以用在对反演精度要求不高的地方。

为了讨论一些可以实现的近似，我们稍微改变一下上面方程的形式

$$\delta s_j^p = \delta s_j^{p-1} + \frac{\delta t_i - q_i}{\sum_{k=1}^{N} l_{ik}^2} l_{ij} \tag{77}$$

其中，$q_i = \sum_{k=1}^{N} \delta s_k^{(p-1)} l_{ik}$。这就是 Kaczmarz 方法。定义，

$$\Delta\delta s_j^p = \delta s_j^p - \delta s_j^{p-1} = \frac{\delta t_i - q_i}{\sum_{k=1}^{N} l_{ik}^2} l_{ij} \tag{78}$$

注意，δt_i 被认为是沿第 i 根射线的旅行时扰动量，q_i 被认为是沿同样射线计算出的投影值

对 j 个网格的校正量 $\Delta\delta s_j$ 可以通过如下步骤得到：首先计算观测投影值和计算投影值的差；然后用 $\sum_{k=1}^{N} l_{ik}^2$ 归一化这个差值；接着把该值分配到第 i 穿过的所有成象单元上，每一个分配的值受 l_{ij} 的加权。

所有射线均如此处理，就实现了 ART 方法。

3.1.2 SIRT 方法

同时迭代重构方法（SIRT 方法）是 ART 方法改进。首先定义

$$\delta s_j^{(k+1)} = \delta s_j^{(k)} + \frac{\sum_i \left[l_{ij} \frac{\tau_i - \vec{l}_i^{\,T} s^{(k)}}{\sum_{j=1}^{N} l_{ij}} \right]}{\sum_i l_{ij}} \tag{79}$$

则第 j 个网格的校正量 $\Delta\delta s_j$ 定义为，

$$\Delta\delta s_j = \delta s_j^{(k+1)} - \delta s_j^{(k)} = \frac{\sum_i \left[l_{ij} \frac{\tau_i - \vec{l}_i^T s^{(k)}}{\sum_{j=1}^{N} l_{ij}} \right]}{\sum_i l_{ij}} \tag{80}$$

同样地，对 j 个网格的校正量 $\Delta\delta s_j$ 可以通过如下步骤得到：首先计算观测投影值和计算投影值的差；然后用 $\sum_{k=1}^{N} l_{ik}^2$ 归一化这个差值；接着把该值分配到第 i 穿过的所有成象单元上，每一个分配的值受 l_{ij} 的加权。所有射线均如此处理，就实现了 SIRT 方法

3.1.3 LS_QR 方法

另一种比较精确的方法是 LS_QR 方法，它比较精确地估算参数扰动量 $\delta\mathbf{m}_{N\times 1}$。与传统的最小二乘方法不同，LS_QR 方法并不是直接去求线性方程，而是在最小二乘思想的指导下在目标函数限定的解空间内寻找最佳的解。LSQR 技术主要的关键点有 3 个：1. Lanczos 迭代；2. 正交变换；3. 模型迭代。在第一个过程中，LSQR 方法运用 Lanczos 迭代方法产生模型空间的一组正交基 $\mathbf{v}_k, 1 \le k \le n$ 和数据空间的一组正交基 $\mathbf{u}_k, 1 \le k \le n$，并形成了 $(n+1)\times n$ 阶的下双对角矩阵 B 和 $n \times n$ 阶的上双对角矩阵 R。第二个过程的正交变换主要是通过一系列 Givens 平面旋转正交变换来实现的，在这个过程需要建立两个双对角矩阵 B 和 R 之间的联系，完成最小二乘与 QR 分解的结合。最后一步通过迭代运算求取方程的解，通过迭代的方式逐步逼近最小二乘解。LSQR 的主要优点包括要求的内存量小、易并行、计算效率高、精度高等，缺点是无法直接给出矩阵的广义逆，这样的话就无法得到模型的分辨率矩阵和协方差矩阵，这是所有迭代法的共同缺点。

LSQR 算法可以用如下的构架来表示：

初始化：

$$\beta_1 \mathbf{u}_1 = \mathbf{d}, \beta_1 = \|\mathbf{d}\|, \alpha_1 \mathbf{v}_1 = A^T \mathbf{u}_1, \alpha_1 = \|A^T \mathbf{u}_1\|,$$
$$\mathbf{w}_1 = \mathbf{v}_1, \mathbf{x} = 0, \varphi_1 = \beta_1, \rho_1 = \alpha_1$$

$for(k = 1,2,3,...)$，进入迭代求解过程

Lanczos 迭代算法，得到两个双对角矩阵：

$$\beta_{k+1}\mathbf{u}_{k+1} = A\mathbf{v}_k - \alpha_k \mathbf{u}_k$$
$$\alpha_{k+1}\mathbf{v}_{k+1} = A^T \mathbf{u}_{k+1} - \beta_{k+1}\mathbf{v}_k$$

Givens 正交变换：

$$\tilde{\rho}_k = \sqrt{\rho_k^2 + \beta_{k+1}^2}$$
$$c_k = \rho_k / \tilde{\rho}_k$$
$$s_k = \beta_{k+1} / \tilde{\rho}_k$$
$$\theta_{k+1} = s_k \alpha_{k+1}$$
$$\tilde{\varphi}_k = c_k \varphi_k$$
$$\rho_{k+1} = -c_k \alpha_{k+1}$$
$$\varphi_{k+1} = s_k \varphi_k$$

其中 ρ_{k+1}、φ_{k+1} 是下次迭代的初始值；

更新方程的解 **x** 和梯度方程 **g**：

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \left(\tilde{\varphi}_k / \tilde{\rho}_k\right)\mathbf{w}_k$$
$$\mathbf{w}_{k+1} = \mathbf{v}_{k+1} - \left(\theta_{k+1} / \rho_k\right)\mathbf{w}_k$$

如果不收敛，回第 2 步，否则迭代结束。

3.2 波形层析及最小二乘偏移成像问题的数值计算

由于波形层析反演时，模型参数维度非常大，导致形成的法方程规模巨大。用任何矩阵方程求解方法都是效率很低的。因此，首选的方法还是用梯度导引类的方法，或者说用转置矩阵反传播数据残差的方法。对于模型参数维度较小的问题才考虑用解大规模矩阵方程的方法。

3.2.1 梯度导引类的方法或误差反传播方法

为了推导梯度导引类的方法或误差反传播方法，还要利用如下的泛函变分求解极值问题。假设 $K:\mathrm{H}_1 \to \mathrm{H}_2$ 是一个有界的线性算子，定义如下的 Tikhonov 泛函，

$$\mathfrak{I}\left(\delta\mathbf{m}\right) = \frac{1}{2}\left\|K\delta\mathbf{m} - \delta\mathbf{u}\right\|_{\mathrm{H}_2}^2 + \frac{\alpha}{2}\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2 \tag{81}$$

对任意的介质扰动 $\delta\mathbf{m} \in \mathrm{H}_1$ 和扰动步长 $\tau \in \mathbb{R}$，总存在如下的泛函变分

$$\begin{aligned}
&\mathfrak{I}\left(\delta\mathbf{m} + \tau\delta\mathbf{h}\right)\\
&= \frac{1}{2}\left\|K\left(\delta\mathbf{m} + \tau\delta\mathbf{h}\right) - \delta\mathbf{u}\right\|_{\mathrm{H}_2}^2 + \frac{\alpha}{2}\left\|\delta\mathbf{m} + \tau\delta\mathbf{h}\right\|_{\mathrm{H}_1}^2\\
&= \frac{1}{2}\left\langle K\delta\mathbf{m} - \delta\mathbf{u}, K\delta\mathbf{m} - \delta\mathbf{u}\right\rangle_{\mathrm{H}_2} + \frac{\tau}{2}\begin{pmatrix}\left\langle K\delta\mathbf{m} - \delta\mathbf{u}, K\delta\mathbf{h}\right\rangle_{\mathrm{H}_2}\\ +\left\langle K\delta\mathbf{h}, K\delta\mathbf{m} - \delta\mathbf{u}\right\rangle_{\mathrm{H}_2}\end{pmatrix} + \frac{\tau^2}{2}\left\langle K\delta\mathbf{h}, K\delta\mathbf{h}\right\rangle_{\mathrm{H}_2}\\
&+ \frac{\alpha}{2}\left\langle\delta\mathbf{m}, \delta\mathbf{m}\right\rangle_{\mathrm{H}_1} + \frac{\alpha}{2}\tau\left(\left\langle\delta\mathbf{m}, \delta\mathbf{h}\right\rangle_{\mathrm{H}_1} + \left\langle\delta\mathbf{h}, \delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right) + \frac{\alpha\tau^2}{2}\left\langle\delta\mathbf{h}, \delta\mathbf{h}\right\rangle_{\mathrm{H}_1}\\
&= \frac{1}{2}\left(\left\|K\delta\mathbf{m} - \delta\mathbf{u}\right\|_{\mathrm{H}_2}^2 + \alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right) + \frac{\tau}{2}\begin{pmatrix}\left\langle K\delta\mathbf{m} - \delta\mathbf{u}, K\delta\mathbf{h}\right\rangle_{\mathrm{H}_2} +\\ \left\langle K\delta\mathbf{h}, K\delta\mathbf{m} - \delta\mathbf{u}\right\rangle_{\mathrm{H}_2}\end{pmatrix} +\\
&\frac{\alpha}{2}\tau\left(\left\langle\delta\mathbf{m}, \delta\mathbf{h}\right\rangle_{\mathrm{H}_1} + \left\langle\delta\mathbf{h}, \delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right) + \frac{\tau^2}{2}\left(\left\|K\delta\mathbf{h}\right\|_{\mathrm{H}_2}^2 + \alpha\left\|\delta\mathbf{h}\right\|_{\mathrm{H}_1}^2\right)
\end{aligned}$$

对上式定义的变分求如下的偏导数，可以定义误差泛函的梯度如下

$$\frac{\partial}{\partial\tau}\mathfrak{I}\left(\delta\mathbf{m}+\tau\delta\mathbf{h}\right)\Big|_{\tau=0}$$
$$=\left\{\frac{1}{2}\left(\left\langle K\delta\mathbf{m}-\delta\mathbf{u},K\delta\mathbf{h}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\delta\mathbf{m},\delta\mathbf{h}\right\rangle_{\mathrm{H}_1}\right)+\frac{\tau}{2}\left(\left\|K\delta\mathbf{h}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\delta\mathbf{h}\right\|_{\mathrm{H}_1}^2\right)\right\}\Big|_{\tau=0}$$
$$=\frac{1}{2}\left(\left\langle K\delta\mathbf{m}-\delta\mathbf{u},K\delta\mathbf{h}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\delta\mathbf{m},\delta\mathbf{h}\right\rangle_{\mathrm{H}_1}\right)$$
$$=\frac{1}{2}\left(\left\langle K^*\left(K\delta\mathbf{m}-\delta\mathbf{u}\right),\delta\mathbf{h}\right\rangle_{\mathrm{H}_1}+\alpha\left\langle\delta\mathbf{m},\delta\mathbf{h}\right\rangle_{\mathrm{H}_1}\right)$$
$$=\frac{1}{2}\left\langle K^*\left(K\delta\mathbf{m}-\delta\mathbf{u}\right)+\alpha\delta\mathbf{m},\delta\mathbf{h}\right\rangle_{\mathrm{H}_1}$$
$$=\frac{1}{2}\left\langle\left(K^*K+\alpha I\right)\delta\mathbf{m}-K^*\delta\mathbf{u},\delta\mathbf{h}\right\rangle_{\mathrm{H}_1}$$
$$=\frac{1}{2}\left\langle\mathrm{Grad}\mathfrak{I}\left(\delta\mathbf{m}\right),\delta\mathbf{h}\right\rangle_{\mathrm{H}_1}$$

由此可得梯度算子为

$$\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)=\left(K^*K+\alpha\mathbf{I}\right)\delta\mathbf{m}-K^*\delta\mathbf{u} \tag{82}$$

当梯度等于零时，可以得到线性系统的最小二乘解

$$\delta\hat{\mathbf{m}}=\left(K^*K+\alpha\mathbf{I}\right)^{-1}K^*\delta\mathbf{u} \tag{83}$$

一般地，用梯度导引类的方法或误差反传播方法来估计参数扰动，即用迭代的方式代替直接求逆来求解上面方程的解。为此，写出如下的迭代格式，

$$\delta\mathbf{m}^{k+1}=\delta\mathbf{m}^k-\mu^kK^*\left(K\delta\mathbf{m}^k-\delta\mathbf{u}^{obs}\right) \tag{84}$$

上式的地球物理含义是根据叠前成像结果，利用反偏移预测叠前数据， $K\delta\mathbf{m}^k$ 即为反偏移的结果，预测数据与观测数据的差为 $K\delta\mathbf{m}^k-\delta\mathbf{u}^{obs}$ 。 $K^*\left(K\delta\mathbf{m}^k-\delta\mathbf{u}^{obs}\right)$ 代表转置矩阵 K^* 作用于残差数据上实现残差数据的反传播，利用该算子可以计算出的范函的梯度方向。这样的计算在勘探地震中相当于做两次正演计算，效率很高。选择步长因子 μ^k 后就可以得到第 k 步的参数扰动修正量，然后进入下一步的迭代计算。

3.2.2 大规模矩阵方程的迭代解法

我们当然也可以用大规模矩阵方程的求解方法直接接法方程（2.23）。实际计算中，这仅仅在模型参数维度较小时是可行的。关于大规模矩阵方程的求解，我们推荐参考 Saad （2003）的专著《Iterative Methods for Sparse Linear Systems》。

此处，我们仅仅简述大规模矩阵的基本解法。众所周知，矩阵方程的求解分为直接法和迭代法。直接法的代表是 LU 分解方法，它用于小规模的非稀疏矩阵方程的求解是合适的。当系数矩阵非奇异时，方程的解可以是精确的。由于 LU 分解的误差和计算代价会随着矩阵规模的增大而增大，因此，对于大规模稀疏矩阵方程的求解一般地会使用迭代法。

迭代法分为基本的迭代方法、投影方法和 Krylov 子空间方法。基本的迭代方法包括 Jacobi 迭代方法、Gauss-Seidel迭代方法、逐次超松弛迭代方法。它们都是把某种数据残差作为迭代的修正方向，超松弛迭代加上了方向修正因子。我们知道线性方程组 $K\delta\mathbf{m}=\delta\mathbf{u}$ 的求解问题可以转化为二次型泛函求极值的问题，据此可以导出投影方法，即最速下降方法。它用二次型误差泛函的梯度方向作为迭代的修正方向。我们知道该梯度方向还是有数据残差决定的。

梯度方向并非二次型泛函的最好的下降方向，为此提出 Krylov 子空间方法，典型的方法就是共轭梯度方法。假设目标函数是凸的二次函数，可以写成如下形式：

$$\begin{aligned}\Im(\delta\mathbf{m}) &= \frac{1}{2}\left\|K\delta\mathbf{m}-\delta\mathbf{u}\right\|_2^2 \\ &= \frac{1}{2}\delta\mathbf{m}^H K^H K\delta\mathbf{m} - \left(K^H\delta\mathbf{u},\delta\mathbf{m}\right)_M + C \\ &= \frac{1}{2}\delta\mathbf{m}^H H\delta\mathbf{m} - \left(K^H\delta\mathbf{u},\delta\mathbf{m}\right)_M + C\end{aligned} \tag{85}$$

其中 $K^H\delta\mathbf{u}$ 表示地震数据的偏移结果，为误差范函的梯度方向，H 是 Hessian 矩阵，由于二次函数要求是凸的，所以这里假设 Hessian 矩阵 H 必须是正定的。由于 H 所形成的基函数经常是非正交的，所以直接求解 H 的逆是比较困难的。这里将 H 的基函数记成 $\mathbf{h}_1,\mathbf{h}_2,...,\mathbf{h}_n$，由它们构成的空间为 $S=span\{\mathbf{h}_1,\mathbf{h}_2,...,\mathbf{h}_n\}$。共轭梯度法就是通过寻找空间 S 的对偶空间 $\tilde{S}=span\{\tilde{\mathbf{h}}_1,\tilde{\mathbf{h}}_2,...,\tilde{\mathbf{h}}_n\}$，然后用对偶空间的基函数来逼近矩阵 H 的逆。其中对偶空间的基函数与原空间的基函数必须满足下面的条件：

$$\tilde{\mathbf{h}}_i H\mathbf{h}_j = 0,\ i\neq j \tag{86}$$

满足上述条件的基函数 $\tilde{\mathbf{h}}_i$ 与 $\mathbf{h}_j$ 称为关于矩阵 H 互共轭。共轭梯度法的梯度方法是通过迭代的方法在每步迭代过程中根据前面已有的梯度方向计算得到的，具体的迭代公式如下：

$$\tilde{\mathbf{h}}_i = -\mathbf{g}_i + \alpha_{i-1}\tilde{\mathbf{h}}_{i-1} \tag{87}$$

其中 $\mathbf{g}_i$ 表示目标函数在第 i 次迭代中的梯度，$\tilde{\mathbf{h}}_{i-1}$、α 分别表示在第 $i-1$ 次迭代中的共轭梯度方向和相应的步长。共轭梯度法根据步长 α 的选取方法不同可以分为不同的优化方法，但其总的思想框架都是一样的，不同方法之间只是在实现过程中存在差异而已。

下面简述最速下降方法和共轭梯度方法。

最速下降方法

对于上述问题，其梯度方向为

$$\mathbf{r}^k = -K^*\left(K\delta\mathbf{m}^k - \delta\mathbf{u}^{obs}\right) \tag{88}$$

则最速下降算法可以描述为：

$k := 0;$

$\delta\mathbf{m}^0 :=$ 初始值；

开始最速下降算法迭代过程

$\mathbf{r}^k = -K^*\left(K\delta\mathbf{m}^k - \delta\mathbf{u}^{obs}\right)$；负梯度方向的计算

$\mu := \arg\min_{\tau>0}\Im\left(\delta\mathbf{m}^k + \mu\mathbf{r}^k\right)$；步长的线性搜索

$\delta\mathbf{m}^{k+1}=\delta\mathbf{m}^k+\mu\cdot\mathbf{r}^k$；模型参数的更新

$k := k+1$；进入下一迭代步。

满足迭代收敛条件，结束迭代。

最速下降算法的收敛性讨论可以参考 Nocedal and Wright （1999）的专著《Numerical Optimization》。

共轭梯度方法

假如系数矩阵 K 是正定对称的，极小化 $\mathfrak{I}(\delta\mathbf{m})=c+\langle\delta\mathbf{u},\delta\mathbf{m}\rangle+\frac{1}{2}\langle K\delta\mathbf{m},\delta\mathbf{m}\rangle$ 等价于求解方程 $K\delta\mathbf{m}=\delta\mathbf{u}$ 。实际上，即使系数矩阵 K 不是正定对称的，也可以用共轭梯度法求解，而且收敛速度也要快于最速下降方法。

下面简述共轭梯度算法。

$k:=0;$

$\delta\mathbf{m}^0:=$ 初始值；

$\mathbf{r}^0=-K^*\left(K\delta\mathbf{m}^0-\delta\mathbf{u}^{obs}\right)$；初始负梯度方向的计算

$\mathbf{p}^0=\mathbf{r}^0$；初始共轭梯度方向的计算

开始共轭梯度算法迭代过程

$\mu:=\left\langle\mathbf{r}^k,\mathbf{r}^k\right\rangle\Big/\left\langle K\mathbf{p}^k,\mathbf{p}^k\right\rangle$;梯度步长因子计算

$\delta\mathbf{m}^{k+1}=\delta\mathbf{m}^k+\mu\cdot\mathbf{p}^k$;模型参数更新

$\mathbf{r}^{k+1}=\mathbf{r}^k-\mu\cdot K\mathbf{p}^k$;梯度方向的更新

$\beta:=\left\langle\mathbf{r}^{k+1},\mathbf{r}^{k+1}\right\rangle\Big/\left\langle\mathbf{r}^k,\mathbf{r}^k\right\rangle$;共轭梯度步长因子的计算

$\mathbf{p}^{k+1}:=\mathbf{r}^{k+1}+\beta\cdot\mathbf{p}^k$；共轭梯度方向计算

$k:=k+1$；进入下一迭代步。

满足迭代收敛条件，结束迭代。

存在一类预条件共轭梯度方法。所谓预条件是通过引入预条件矩阵使得系数矩阵 K 更接近正定对称的。具体可以参考 Nocedal and Wright （1999）的专著《Numerical Optimization》。

3.3 非线性梯度导引类全波形反演的数值计算问题

非线性波形反演当然可以转化为由每一迭代步都接一个线性化问题来解决。3.2 节中已经讨论过此问题。但是，主要的数值计算方法还是梯度导引类的方法。具体讲就是 Gauss-Newton 方法和全 Newton 方法。其中的核心问题当然是梯度的计算、Hessian 的计算核步长的计算。

3.3.1 梯度的计算问题

为了讨论梯度的计算，需要引入下面的数学概念。假设泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 是在 $\mathbf{m}$ 处的 Fréchet 可微，由 Reisz 表示定理，存在梯度 $\mathrm{Grad}\mathfrak{I}(\mathbf{m})\in\mathcal{H}$，泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 点的梯度满足

$$\mathfrak{I}'(\mathbf{m})\delta\mathbf{m}=\left\langle\mathrm{Grad}\mathfrak{I}(\mathbf{m}),\delta\mathbf{m}\right\rangle_{\mathcal{H}} \tag{89}$$

若泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 是在 $\mathbf{m}$ 处是 Fréchet 可微的，则对任意的 $\delta\mathbf{m}\in\mathcal{H}$，则映射 $\tau\mapsto\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})\ \mathbb{R}\to\mathbb{R}$ 在 $\tau=0$ 处是可微的，且有

$$\left.\frac{d}{d\tau}\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})\right|_{\tau=0}=\left\langle\mathrm{Grad}\mathfrak{I}(\mathbf{m}),\delta\mathbf{m}\right\rangle_{\mathcal{H}} \tag{90}$$

存在。因此，我们可以从上述 Gateaux 导数的定义计算泛函的微分。

勘探地震学中的大规模参数估计中，到底如何计算梯度是一个非常棘手的问题。Chavent（1991）在他的博士论文中提出了 Adjoint-State 方法（伴随状态方法），经其他学者完善后，很快在地球物理学领域得到推广应用。

下面我们从一个角度引入伴随状态方法来计算梯度。

已知算子 $A(m)$，它是一个投影，假定它是一个共轭算子。它建立了模型空间和数据空间的映射关系。$\frac{dA}{dm}$ 为算子微分，它定义了一个 Jacobi 矩阵。在一定的定义下，它是一个 Fréchet 微商。

定义

$$A(m)\left(A(m)\right)^{-1}=\mathbf{I} \tag{91}$$

对上式两边求导

$$\frac{dA(\mathbf{m})}{d\mathbf{m}}\left(A(\mathbf{m})\right)^{-1}+A(\mathbf{m})\frac{d\left(A(\mathbf{m})\right)^{-1}}{d\mathbf{m}}=0 \tag{92}$$

改写上式

$$\frac{d\left(A(\mathbf{m})\right)^{-1}}{d\mathbf{m}}=-\left(A(\mathbf{m})\right)^{-1}\frac{dA(\mathbf{m})}{d\mathbf{m}}\left(A(\mathbf{m})\right)^{-1} \tag{93}$$

定义数据拟合的残差矢量为

$$\mathbf{r}(\mathbf{m})=F(\mathbf{m})-\mathbf{d} \tag{94}$$

已知有如下的 Gateaux 导数定义式，

$$\left.\frac{dA(\mathbf{m}+\tau\mathbf{e}_i)}{d\mathbf{m}}\right|_{\tau=0}=\frac{dA}{d\mathbf{m}}\mathbf{e}_i \tag{95}$$

由此，可以定义如下的梯度计算式

$$\begin{aligned}
\left[G_{LS}\right]_i&=\frac{d}{d\tau}\Im_{LS}(\mathbf{m}+\tau\mathbf{e}_i)\big|_{\tau=0}=\left.\frac{dF}{d\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}\mathbf{r}(\mathbf{m})\\
&=\left\langle\left.\frac{dF}{d\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0},\mathbf{r}(\mathbf{m})\right\rangle=\left\langle C\frac{d}{d\mathbf{m}}\left(\left(A(\mathbf{m})\right)^{-1}\mathbf{e}_i\right)f,\mathbf{r}(\mathbf{m})\right\rangle_Y\\
&=-\left\langle C\left(A(\mathbf{m})\right)^{-1}\left(\frac{dA}{d\mathbf{m}}\mathbf{e}_i\right)\left(A(\mathbf{m})\right)^{-1}f,\mathbf{r}(\mathbf{m})\right\rangle_Y\\
&=-\left\langle\left(\frac{dA}{d\mathbf{m}}\mathbf{e}_i\right)\left(A(\mathbf{m})\right)^{-1}f,\left(\left(A(\mathbf{m})\right)^{-1}\right)^*C^*\mathbf{r}(\mathbf{m})\right\rangle_U=\left\langle\left(\frac{dA}{d\mathbf{m}}\mathbf{e}_i\right)\mathbf{u},\mathbf{z}\right\rangle_U
\end{aligned} \tag{96}$$

其中，$\mathbf{u}=\left(A(\mathbf{m})\right)^{-1}f$ 为正演波场；$\mathbf{z}=-\left(\left(A(\mathbf{m})\right)^{-1}\right)^*C^*\mathbf{r}(\mathbf{m})$ 为残差波场的反传播。内积可以视为零相位的互相关。

可以看出，波场 $\mathbf{z}$ 的求解需要解如下的微分方程

$$\left(A(\mathbf{m})\right)^*\mathbf{z}=-C^*\mathbf{r}(\mathbf{m}) \tag{97}$$

前已述及，$A(\mathbf{m})$是一个共轭算子，满足$\left(\left(A(m)\right)^{-1}\right)^{*}=\left(\left(A(m)\right)^{*}\right)^{-1}$。因此，解震源函数的正向传播和残差数据的反向传播两个正演模拟问题，然后把它们的结果进行点对点的零延迟相关运算可以得到空间上每一点的梯度值。可以看出，利用 Adjoint-State 方法可以方便高效地计算出数据残差泛函的梯度，用于非线性迭代公式的计算。

3.3.2 Hessian 的计算问题

Hessian 的定义：假如算子$\mathfrak{I}(\delta\mathbf{m})$：$\mathcal{H}_1\to\mathcal{H}_2$在$\delta\mathbf{m}\in\mathcal{H}_1$是 Fréchet 可微的，当且仅当存在$\mathfrak{I}'(\delta\mathbf{m})\in\mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)$，称为算子$A$在$\delta\mathbf{m}$点的一阶 Fréchet 导数，满足

$$\mathfrak{I}(\delta\mathbf{m}+\mathbf{h})=\mathfrak{I}(\delta\mathbf{m})+\mathfrak{I}'(\delta\mathbf{m})\mathbf{h}+o\left(\|\mathbf{h}\|_{\mathcal{H}_1}\right) \tag{98}$$

相应地，Hessian 被定义为，

$$\mathfrak{I}'(\delta\mathbf{m}+\mathbf{k})=\mathfrak{I}'(\delta\mathbf{m})+\mathfrak{I}''(\delta\mathbf{m})\mathbf{k}+o\left(\|\mathbf{k}\|_{\mathcal{H}_1}\right) \tag{99}$$

其中，$\mathfrak{I}''(\delta\mathbf{m})\in\mathcal{L}\left(\mathcal{H}_1,\mathcal{L}(\mathcal{H}_1,\mathcal{H}_2)\right)$为一阶 Fréchet 导数即 Hessian。

假如算子$\mathfrak{I}(\delta\mathbf{m})$：$\mathcal{R}^n\to\mathcal{R}$是连续可微的，则 Hessian 构成一个$n\times n$的矩阵，矩阵中的元素定义为

$$\left[\mathrm{Hess}\mathfrak{I}(\delta\mathbf{m})\right]_{ij}=\frac{\partial^2\mathfrak{I}}{\partial\delta m_i\partial\delta m_j},1\le i,j\le n \tag{100}$$

利用（1.63）式可以导出全 Newton 迭代格式（1.66），Hessian 在其中起着关键作用，它的逆对梯度方向进行有效的预条件处理。

但是，Hessian 的计算是非常不容易的。在勘探地震学中，线性的 Hessian（见（1.75）式）可以用 Green 函数来计算，尽管计算量很大。非线性的 Hessian 原则上应该用（2.41）式计算，但这几乎是不可能的。目前唯一可能的方法是根据（2.40）定义变分，再次利用共轭状态法计算 Hessian。这样的想法是可行的，目前已有文献报道。

2.3.3 关于梯度方向的预条件处理

从（1.62）和（1.66）两个迭代公式可以看出，迭代公式的收敛性主要取决于梯度方向和步长因子。

Hessian 矩阵的逆可以认为是对梯度方向的预条件处理。在勘探地震学中，我们知道 Hessian 矩阵中对角线上的元素代表地下每一成像点的照明总能量。由于 Hessian 矩阵非常庞大，求逆几乎是不可能的，一般地把（66）修改为

$$\hat{\mathbf{m}}^{k+1}=\hat{\mathbf{m}}^{k}-\mu^{k}\left[\mathbf{D}\left(\hat{\mathbf{m}}^{k}\right)\right]^{-1}\mathrm{Grad}\mathfrak{I}\left(\hat{\mathbf{m}}^{k}\right) \tag{101}$$

其中，$\mathbf{D}\left(\hat{\mathbf{m}}^{k}\right)$Hessian 矩阵的对角元素形成的对角矩阵，是用第k步迭代得到的模型计算出的地震波照明结果。

实际地震波反演过程中，鉴于梯度方向的极端重要性，应该考虑用勘探地震学的方法进行进一步的与条件处理。可行的方法包括：（1）对计算出的梯度进行多尺度分解，粗尺度的大梯度成分加权系数应该大一些，细尺度的梯度成分加权系数应该小一些；（2）把叠前深度偏移成像的构造信息约束梯度的计算。

应该还有很多思想方法可以预条件处理计算出的梯度。

3.4 典型的数值最优化方法

正如前面所分析的，最优化方法需要解决的最关键问题有两个：1. 梯度方向的求取；2. 步长的求取。要理解这两个问题如何决定最优化方法可以从一维优化方法开始讨论。典型的一维优化方法主要有牛顿法、割线法、多项式插值法、区间分割法等。这些不同的优化方法分别对目标函数做了不同的假设，例如牛顿法中，假设目标函数是一个二次函数，或是在信赖域内可以用一个二次函数比较好地逼近目标函数；割线法假设目标函数梯度的变化（即二阶项）在依赖域内是一个常数，或是变化比较小的值，这样的话就可以用差商法估计目标函数的二阶项，以计算优化步长；多项式插值法假设目标函数可以被某类多项式逼近，通过求取该类多项式的各项系数计算目标函数的极小值点；区间分割法假设目标函数是一个单峰函数，即在依赖域内只有一个极小值，然后根据目标函数的该特征通过分割区间的方法不断减小依赖区间，直接依赖区间收敛到极小值点上。从上面的分析可以看出，这些一维优化方法都对目标函数做了一定的假设，或是假设目标函数在某个依赖区间中满足一定的条件，然后再根据目标函数相应的特点寻找合适的方案和策略求取目标函数的极小值点。因此，在寻优过程中，对目标函数的分析是很重要的一项工作，只有充分把握了目标函数的几何、代数特征，才有可能提出最适合该目标函数的寻优方案。

高维的最优化方法主要是梯度导引类的方法，包括最速下降法、共轭梯度法、牛顿（拟牛顿）法等，还包括近期研究比较多的罚函数法、依赖域法、滤子方法、内点法等。其中梯度导引类的方法是比较传统的解决高维优化问题的方法，这种方法要求目标函数是一个凸函数，在整个解空间中存在一个极小值点，通过梯度方向的导引，逐步收敛到这个极小值点。在梯度导引类方法中，当梯度方法确定后，优化问题就退化为一维寻优问题，所有之前说的一维优化方法都可以用来确定步长的大小。梯度导引类方法要求目标函数在依赖区间内是一个凸函数，但这点在大部分情况下是不能满足的，因此需要用新的方法来寻找目标函数的全局极小值点。比较直接的想法就是用罚函数法，在目标函数的基础上加上一个罚函数做为约束条件，改变原问题解空间的分布，但得新的目标函数在新的解空间里为一个凸函数，然后再利用各种合适的梯度方法进行求解。由于罚函数方法会改变解空间的分布情况，因此对罚函数的选择要特别小心。好的罚函数可能会降低解空间的维数，改善目标函数的凸性，但不好的罚函数可能会让实际问题的真解不再包含于新的解空间，这样的话得到的解就不再具有任何意义了。还有一些近期研究比较多的依赖域法、滤子方法、内点法等都是为了解决目标函数不满足凸函数要求情况时的优化问题，这些方法也都是对目标函数做了各种假设，然后根据目标函数的几何、代数特征，提出寻优的方案。

综合上面的分析，不管是一维优化还是高维的优化问题，对目标函数的分析、把握目标函数本身的几何、代数特征是最关键的。只有充分把握了目标函数本身的特征，才可能根据目标函数的特征找到最合适的寻优方案，才可能利用最合理的优化方法解决所面对的优化问题。

数值优化问题是一个大问题，有很多专著讨论。譬如 Nocedal and Wright （1999）的专著《Numerical Optimization》。此处，我们罗列集中典型的数值优化算法，它们广泛地应用在非线性梯度导引类全波形反演问题的数值求解过程中。

3.4.1 非线性 CG（共轭梯度）方法

下面简述非线性共轭梯度算法。它可以用来求解非二次型的优化问题。

$k := 0;$

$\delta\mathbf{m}^0 :=$ 初始值；

$\mathbf{r}^0 = -\nabla\mathfrak{J}(\mathbf{m})$；初始负梯度方向的计算

$\mathbf{p}^0 = \mathbf{r}^0$；初始共轭梯度方向的计算

开始共轭梯度算法迭代过程

$\mu := \arg\min_{\tau>0} \mathfrak{J}(\mathbf{m}^k + \mu\cdot\mathbf{r}^k)$；梯度步长因子计算

$\mathbf{m}^{k+1}=\mathbf{m}^k + \mu\cdot\mathbf{p}^k$;模型参数更新

$\mathbf{r}^{k+1} = -\nabla\mathfrak{J}(\mathbf{m}^{k+1})$；梯度方向的计算

$\beta := \left\langle \mathbf{r}^{k+1}, \mathbf{r}^{k+1} \right\rangle \Big/ \left\langle \mathbf{r}^{k}, \mathbf{r}^{k} \right\rangle$;共轭梯度步长因子的计算

$\mathbf{p}^{k+1}:=\mathbf{r}^{k+1}+\beta\cdot\mathbf{p}^{k}$；共轭梯度方向计算

$k := k+1$；进入下一迭代步。

满足迭代收敛条件，结束迭代。这是一个比较常用的非线性优化的迭代格式。

3.4.2 Newton 方法

Newton 方法是最典型的非线性优化数值算法，就是（1.66）式的具体算法。则最速下降算法可以描述为：

$k := 0;$

$\delta\mathbf{m}^{0} :=$ 初始值；

开始最速下降算法迭代过程

$\mathbf{r}^{k} = -\nabla\Im\left(\mathbf{m}^{k}\right)$；负梯度方向的计算

$\left\|\mathbf{r}^{k}\right\|^{2} < \varepsilon$;结束迭代，退出

$\mathbf{p}^{k} = -\left[\nabla^{2}\Im\left(\mathbf{m}^{k}\right)\right]^{-1}\nabla\Im\left(\mathbf{m}^{k}\right)$　Hessisan 预条件梯度方向

$\mu := \arg\min_{\tau>0}\Im\left(\mathbf{m}^{k}+\mu\cdot\mathbf{p}^{k}\right)$；步长的线性搜索

$\mathbf{m}^{k+1}=\mathbf{m}^{k}-\mu\cdot\mathbf{p}^{k}$；模型参数的更新

$k := k+1$；进入下一迭代步。

满足迭代收敛条件，结束迭代。

3.4.3 Newton 方法的一种改进---置信区间方法

依赖域法是牛顿法的一种改进方法，该方法和牛顿法最大的区别在于引入了依赖域这个概念，并且要求每次迭代过程中的搜索步长必须在依赖域范围内。考虑如下的最优化问题：

$$\begin{aligned} &\min_{\mathbf{x}\in R^{n}} \quad f(\mathbf{x}), \\ &S.T. \quad A\mathbf{x}=\mathbf{b} \end{aligned}$$

假设在第 k 次迭代中的可行点为 $\mathbf{x}_k$，在引入依赖域后上面的优化问题在每次迭代过程中所需解决的子问题的形式如下：

$$\begin{aligned} &\min_{\mathbf{d}\in R^{n}} \quad \mathbf{g}_k^{T}\mathbf{d}+\frac{1}{2}\mathbf{d}^{T}B_k\mathbf{d} \;=\varphi_k(\mathbf{d}) \\ &S.T. \qquad\qquad A\mathbf{d}=0 \\ &\qquad\qquad \left\|\mathbf{d}\right\|_{\infty}\le\Delta_k \end{aligned}$$

其中向量 $\mathbf{g}_k$ 和矩阵 B_k 分别表示第 k 次迭代过程中的梯度向量和 Hessian 矩阵，新的目标函数 $\varphi_k(\mathbf{d})$ 相当于原来目标函数 $f(\mathbf{x})$ Taylor 展开后的一阶、二阶项。上面的优化问题是一个二次规划问题，第一个约束条件 $A\mathbf{d}=0$ 保证了新解依然是该优化问题的可行点，第二个约束条件 $\left\|\mathbf{d}\right\|_{\infty}\le\Delta_k$ 决定了解搜索的范围，即 Δ_k 为依赖区间的大小，它的大小可以由相邻的两个可行点通过下面的公式计算得到：

$$\Delta_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{d}_k)}{\varphi_k(0) - \varphi(\mathbf{d}_k)}$$

从 Δ_k 的表达式可以看出，在迭代过程中，如果目标函数 $f(\mathbf{x})$ 的收敛速度比较慢，则依赖区间比较小；如果目标函数 $f(\mathbf{x})$ 的一阶项、二阶项的值 $\varphi_k(0) - \varphi(\mathbf{d}_k)$ 越小，代表函数 $f(\mathbf{x})$ 越光滑，则依赖区间越大。所以依赖区间的大小和目标函数在可行点附近的正则性有关，可行点附近目标函数的正则性越好，即目标函数越光滑，依赖区间越大。

依赖域法可以通过下面的算法实现 $\varepsilon \geq 0, k = 1$

给出第一个可行点 $\mathbf{x}_1$，使得 $A\mathbf{x}_1 = \mathbf{b}$；

根据 $\mathbf{x}_1$ 计算 Hessian 矩阵 B_1；给出初始依赖域大小 Δ_1；；

求解优化问题：

$$\begin{aligned} \min_{\mathbf{d} \in R^n} \quad & \mathbf{g}_k^T \mathbf{d} + \frac{1}{2} \mathbf{d}^T B_k \mathbf{d} \ = \varphi_k(\mathbf{d}) \\ S.T. \quad & A\mathbf{d} = 0 \\ & \|\mathbf{d}\|_\infty \leq \Delta_k \end{aligned}$$

得到 $\mathbf{d}_k$，如果 $\|\mathbf{d}_k\|_\infty < \varepsilon$，迭代结束；

计算依赖域大小：

$$\Delta_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{d}_k)}{\varphi_k(0) - \varphi(\mathbf{d}_k)}$$

并更可行点：

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{x}_k + \mathbf{d}_k & if\ \Delta_k > 0 \\ x_k & else \end{cases}$$

如果 $\Delta_k \geq 0.25$，则跳到第 4 步；

否则，减小依赖域大小 $\Delta_k = \Delta_k / 2$，跳到第 5 步；

如果 $\Delta_k < 0.75$ 或者 $\|\mathbf{d}_k\|_\infty < \Delta_k$，则跳到第 5 步；

否则增大依赖域大小 $\Delta_k = 2\Delta_k$

更新依赖域大小 $\Delta_{k+1} = \Delta_k$，计算矩阵 B_{k+1}；令 $k = k + 1$，跳到第 2 步；

从上面的实现过程可以看出，与牛顿法相比，依赖域法只是多引入了依赖区间这个概念，利用依赖区间对解空间进行了进一步的约束。因此，在依赖域法中如果定义依赖区间的大小是整个方法的关键所在。

3.4.4 BFGS 和 L_BFGS 方法

下面简述 BFGS 算法。

$k := 0;$

$\delta\mathbf{m}^0 :=$ 初始值；

$\mathbf{H}_0 :=$ Hessian的初始值；

$\mathbf{r}^0 = -\nabla\mathfrak{I}\left(\mathbf{m}^0\right)$；初始梯度方向

开始最速下降算法迭代过程

$\mathbf{p}^k = -\mathbf{H}^{-1}\nabla\mathfrak{I}\left(\mathbf{m}^k\right)$；拟 Hessisan 预条件梯度方向

$\mu := \arg\min_{\tau>0}\mathfrak{I}\left(\mathbf{m}^k + \mu\cdot\mathbf{p}^k\right)$；步长的线性搜索

$\mathbf{m}^{k+1}=\mathbf{m}^k+\mu\cdot\mathbf{p}^k$；模型参数的更新

$\mathbf{r}^{k+1} = -\nabla\mathfrak{I}\left(\mathbf{m}^{k+1}\right)$；更新梯度方向；

$\mathbf{H}^{k+1} :=$ 更新拟Hessian；

$k := k+1$；进入下一迭代步。

满足迭代收敛条件，结束迭代。

其中拟 Hessian 的计算用

$$\mathrm{H}^{k+1} = \mathrm{H}^k - \frac{\mathbf{H}^k\mathbf{s}_k\mathbf{s}_k^T\mathbf{H}^k}{\left\langle\mathbf{H}^k\mathbf{s}_k,\mathbf{s}_k\right\rangle} + \frac{\mathbf{y}_k\mathbf{y}_k^T}{\left\langle\mathbf{y}_k,\mathbf{s}_k\right\rangle} \tag{102}$$

还存在一类有限内存的 BFGS 算法的变种，即 L_BFGS 方法。它直接计算拟 Hessian 的逆，用

$$\mathrm{H}_{k+1}^{-1} = \left(I - \frac{\mathbf{s}_k\mathbf{y}_k^T}{\left\langle\mathbf{y}_k,\mathbf{s}_k\right\rangle}\right)\mathrm{H}_k^{-1}\left(I - \frac{\mathbf{s}_k\mathbf{y}_k^T}{\left\langle\mathbf{y}_k,\mathbf{s}_k\right\rangle}\right) + \frac{\mathbf{s}_k\mathbf{s}_k^T}{\left\langle\mathbf{y}_k,\mathbf{s}_k\right\rangle} \tag{103}$$

对上述两式，有

$$\begin{aligned}\mathbf{s}_k &= \mathbf{m}_{k+1} - \mathbf{m}_k \\ y_k &= \nabla\mathfrak{I}\left(\mathbf{m}^{k+1}\right) - \nabla\mathfrak{I}\left(\mathbf{m}^k\right)\end{aligned} \tag{104}$$

3.4.5 线性搜索算法

步长因子在迭代算法中起着重要作用，但是合适的步长因子的计算是很困难的。基本思想是首先大致确定步长因子的大致范围，然后在此范围内用二次曲线逼近误差泛函曲线，找到误差下降最快的步长因子。这个过程是迭代执行的，首先定一个合适的较大的范围$[0,\mu_0]$，用下式定一个 μ_1

$$\mu_1 = \frac{-\mathfrak{I}(\delta\mathbf{m})\mu_0}{2\left[\mathfrak{I}(\delta\mathbf{m}+\mu_0\mathbf{h}) - \mathfrak{I}(\delta\mathbf{m}) - \mathfrak{I}'(\delta\mathbf{m})\mu_0\right]} \tag{105}$$

假如 μ_1 不可以接受，用 μ_1 取代 μ_0，进行下一轮计算，直到得到合适的步长因子为止。

4 结论与讨论

地震波层析成像（包括旅行时层析和波形层析）以及全波形反演是勘探地震学中的最核心问题。它们的目的是解决背景速度估计、叠前深度偏移（包括保真的 AVA 道集产生）、全频带速度估计（进入储层预测阶段）。可以看出，三个阶段的任务完成的好的话，勘探地震学中的问题就可以得到较为完满的解决。

当然，实际情况并不会如此理想。因此，勘探地震学需要发展一系列的方法技术来达到准确的油气预测的目的。但是，这并没有妨碍地震波层析成像与全波形反演的理论重要性。反而更印证了理论指导的不可或缺。

我们全面分析了旅行时层析反演、波形层析反演和全波形反演的正问题描述和反问题描述方法。指出了它们之间的联系与区别。特别指出它们的差异之处在于：层析反演的正问题描述的是一个线性化的问题，它仅仅描述一次波的传播过程。这是层析反演方法不能反演速度场的精细变化的根本原因。全波形反演的正问题描述的是一个非线性问题，就是说它描述的是全波场的传播，既包括一次波也包括多次波。但是，这仅仅是理论上的。实际上，由于初始速度场中缺乏高波数成分，多次波是模拟不出来的，也可能是不对的。因此，把波形层析推向考虑全波场，既包括一次波也包括多次波，是精确的速度估计的方向。FWI 尽管理论上很完美，实际上不是一个可行的速度估计的方法。这种分析实际上已经从本质上讲清楚了精确的速度估计的发展方向。

从计算的角度，我们指出了地震波层析成像（包括旅行时层析和波形层析）以及全波形反演的数值计算的核心问题是误差反传播，反传播的结果就是参数扰动量的估计或者是梯度计算结果。法方程的求解原则上是尽量少用的。从两种反演问题归结出的迭代格式，我们指出地震波反演数值计算的进一步方向必然是参数模型空间的、现代的、非梯度导引的采样机制的研究。吉布斯采样方法、Metropolis 采样机制、蚂蚁寻迹、模拟退火、遗传算法等方法必然会成为地震波层析成像和全波形反演的主要数值计算方法。

地震数据宽方位、高密度和全频带观测和 CPU/GPU 以及 CPU_MIC 高性能计算平台的发展，紧紧抓住地震波层析成像（包括旅行时层析和波形层析）技术和全波形反演技术是提高勘探地震技术水平的关键，可以起到带动全局的作用。

我们相信，用（前面已经指出的）变通的方式某种程度上实现 FWI 是可能的，也是必须的。

附录

附录 1 微商和 Gateaux 导数的数学概念

① Fréchet 意义下的强微分

强微分定义为映射增量的线性主部。

定义：设 X,Y 为实赋范线性空间，Ω 为 X 的开集，映射 $f:\Omega\to Y$，如果存在 $F\in B(X\to Y)$，使得

$$\lim_{\|\delta\mathbf{m}\|\to 0}\frac{\left\|f(\mathbf{m}_0+\delta\mathbf{m})-f(\mathbf{m}_0)-F\delta\mathbf{m}\right\|}{\|\delta\mathbf{m}\|}=0 \tag{106}$$

就称映射 f 在 $\mathbf{m}_0$ 处强可微；称 $F\delta\mathbf{m}$ 为映射 f 在 $\mathbf{m}_0$ 处的强微分或 Fréchet 微分，记为 $df(\mathbf{m}_0,\delta\mathbf{m})$；称 F 为映射 f 在 $\mathbf{m}_0$ 处的强导数或 Fréchet 导数，记为 $f'(\mathbf{m}_0)$ 或 $df(\mathbf{m}_0)$。如果映射 f 在 Ω 上每一点都强可微，就称映射 f 在 Ω 上强可微，称 $f':\Omega\to B(X\to Y)$ 为映射 f 的强导映射。

例 1 设映射 $f:R^n\to R^m$ 定义为

$$\begin{cases} y_1=f_1(x_1,x_2,\cdots,x_n)\\ y_2=f_2(x_1,x_2,\cdots,x_n)\\ \quad\vdots\\ y_m=f_m(x_1,x_2,\cdots,x_n)\end{cases} \tag{107}$$

其中，$\mathbf{x}=(x_1,x_2,\cdots,x_n)\in R^n$，$\mathbf{y}=(y_1,y_2,\cdots,y_m)\in R^m$

映射 f 在 $\mathbf{x}_0$ 处强可微等价于每个分量映射 $f_j:R^n\to R$ 在 $\mathbf{x}_0$ 处可微。对扰动 $\delta\mathbf{x}=(\delta x_1\quad\delta x_2\quad\cdots\quad\delta x_n)\in R^n$，存在

$$df(\mathbf{x}_0,\delta\mathbf{x})=\left(df_1(\mathbf{x}_0,\delta\mathbf{x})\quad df_2(\mathbf{x}_0,\delta\mathbf{x})\quad\cdots\quad df_m(\mathbf{x}_0,\delta\mathbf{x})\right) \tag{108}$$

其中，每一个导函数有如下形式，

$$df_j\left(\mathbf{x}_0,\delta\mathbf{x}\right)=\frac{\partial f_j\left(\mathbf{x}_0\right)}{\partial x_1}\delta x_1+\frac{\partial f_j\left(\mathbf{x}_0\right)}{\partial x_1}\delta x_2+\cdots+\frac{\partial f_j\left(\mathbf{x}_0\right)}{\partial x_n}\delta x_n\left(j=1,2,\cdots,m\right) \tag{109}$$

展开上式，得到如下的 Jacobi 矩阵

$$f'\left(\mathbf{x}_0\right)=\begin{pmatrix}\dfrac{\partial f_1\left(\mathbf{x}_0\right)}{\partial x_1} & \dfrac{\partial f_2\left(\mathbf{x}_0\right)}{\partial x_1} & \cdots & \dfrac{\partial f_m\left(\mathbf{x}_0\right)}{\partial x_1}\\ \dfrac{\partial f_1\left(\mathbf{x}_0\right)}{\partial x_2} & \dfrac{\partial f_2\left(\mathbf{x}_0\right)}{\partial x_2} & \cdots & \dfrac{\partial f_m\left(\mathbf{x}_0\right)}{\partial x_m}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{\partial f_1\left(\mathbf{x}_0\right)}{\partial x_n} & \dfrac{\partial f_2\left(\mathbf{x}_0\right)}{\partial x_n} & \cdots & \dfrac{\partial f_m\left(\mathbf{x}_0\right)}{\partial x_n}\end{pmatrix} \tag{110}$$

②Gateaux 意义下的弱微分

定义：设 X,Y 为实赋范线性空间，Ω 为 X 的开集，映射 $f:\Omega\to Y$，如果对 $\mathbf{m}\in\Omega,\delta\mathbf{m}\in X$，

$$\lim_{t\to 0}\frac{f\left(\mathbf{m}+t\delta\mathbf{m}\right)-f\left(\mathbf{m}\right)}{t} \tag{111}$$

存在，记为 $Df\left(\mathbf{m},\delta\mathbf{m}\right)$。则称 $Df\left(\mathbf{m},\delta\mathbf{m}\right)$ 是映射 f 在处沿方向 $\delta\mathbf{m}$ 的弱微分或 Gateaux 意义下的微分；如果对一切 $\delta\mathbf{m}\in X$，$Df\left(\mathbf{m},\delta\mathbf{m}\right)$ 存在，就称映射 f 在 $\delta\mathbf{m}$ 处弱可微或 Gateaux 意义下可微；如果映射 f 在 $\delta\mathbf{m}$ 处弱可微，且存在 $Df\left(\mathbf{m}\right)\in B\left(X\to Y\right)$ 使得

$$Df\left(\mathbf{m},\delta\mathbf{m}\right)=\left[Df\left(\mathbf{m}\right)\right]\delta\mathbf{m} \tag{112}$$

则称映射 f 在 $\delta\mathbf{m}$ 处弱可导或 Gateaux 意义下可导，称 $Df\left(\mathbf{m},\delta\mathbf{m}\right)$ 有界线性弱微分，称 $Df\left(\mathbf{m}\right)$ 为映射 f 在 $\delta\mathbf{m}$ 处的弱导数或 Gateaux 导数。

如果映射 f 在 Ω 上每点弱可微，就称映射 f 在 Ω 上弱可微；如果映射 f 在 Ω 上每点均具有界线性弱微分，就称映射 f 在 Ω 上具有有界线性弱微分或在 Ω 上弱可导。

定义：设 X 为实赋范线性空间，Ω 为 X 中的开集，如果泛函 $\varphi:\Omega\to R$ 在 Ω 上具有有界线性弱微分，且 $D\varphi\left(\mathbf{m}\right)=f\left(\mathbf{m}\right)$，就称 $f:\Omega\to X^*$ 为 φ 的梯度，记为 $f\left(\mathbf{m}\right)=grad\varphi\left(\mathbf{m}\right)$，并称 φ 为 f 的位势。

附录 2 有界线性算子及其共轭算子

①有界线性算子

一个线性算子 $A:\mathcal{H}_1\to\mathcal{H}_2$ 是有界的，当且仅当诱导算子范数

$\left\|A\right\|\triangleq\sup_{\left\|f\right\|_{\mathcal{H}_1}=1}\left\|Af\right\|_{\mathcal{H}_2}$ 是有限的。

有界线性算子是连续的。从 $\mathcal{H}_1$ 到 $\mathcal{H}_2$ 的有界线性算子空间由 $\mathcal{L}\left(\mathcal{H}_1,\mathcal{H}_2\right)$ 表示。

②有界线性算子的共轭. 一个有界线性算子 A 的共轭是算子 $A^*\in\mathcal{L}\left(\mathcal{H}_1,\mathcal{H}_2\right)$，它满足如下关系

$\left\langle Af,g\right\rangle_{\mathcal{H}_2}=\left\langle f,A^*g\right\rangle_{\mathcal{H}_1}$ 其中，$f\in\mathcal{H}_1,g\in\mathcal{H}_2$。

附录 3.3 Riesz 表示定理

假设 $f(\mathbf{x})$ 是 Hilbert 空间 $\mathcal{H}$ 中的线性泛函，它是有界的，即

$$\left|f(\mathbf{x})\right| \le const \left\|\mathbf{x}\right\|$$

则，泛函 $f(\mathbf{x})$ 具有如下形式

$f(\mathbf{x}) = (\mathbf{x}, \mathbf{y})$， $\mathbf{y} \in \mathcal{H}$

$\mathbf{y}$ 是唯一确定的。

参考文献

[1] 柯尔莫戈洛夫，A. H. 和 佛明，C. B. 著；段虞荣 郑洪深 郭思旭译， 2005，函数论与泛函分析初步（第七版）; 高等教育出版社。

[2] 夏道行，吴卓人，严绍宗 舒五昌编著，2009， 实变函数论与泛函分析; 高等教育出版社。

[3] Chavent, G., 2009, Nonlinear Least Squares for Inverse Problems: Theoretical Foundations and Step-by-Step Guide for Applications; Springer.

[4] Dennis Jr., J. E. and Schnabel, R. B., 1996, Numerical Methods for Unconstrained Optimization and Nonlinear Equations; SIAM.

[5] Lax, P. D., 2002, Functional Analysis,; John Wiley & Sons.

[6] Nocedal, J., and Wright, S. J., 1999,Numerical Optimization; Springer Science + Business Meida, Inc.

[7] Nolet, G., (Ed.), 1987, Seismic Tomography with Applications in Global Seismology and Explorations Geophysics; D. Reidel Publishing Company .

[8] Saad, Y., 2003, Iterative Methods for Sparse linear Systems; SIAM.

[9] Vogel, C. R., 2002, Computational Methods for Inverse Problems; SIAM.

[10] Zeidler, E., 1991, Applied Functional Analysis: Applications to Mathematical Physics; Springer

地震波传播的散射场表达与逆散射成像

王华忠，胡江涛

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】地震波成像目的首先是定位反射界面或散射点出现的空间位置，描述地下地质体的几何结构；然后是估计地下岩石介质的物性参数，主要是速度和密度参数。叠前地震数据偏移成像与反演成像是达到上述目标的主要方法。地震波的散射场表达是叠前偏移成像和反演成像的共同基础。本文以地震波传播的散射场表达与逆散射成像为切入点，首先说明逆散射成像的本质，然后再试图理清楚一般意义下的叠前偏移成像与逆散射成像之间的区别与联系，并说明叠前偏移成像把波场残差沿等时面投影，走时层析成像沿波路径投影，分别确定速度场的高波数成分和低波数成分。最后给出了迭代最小二乘偏移成像的具体实现方法和数值结果。

【关键词】散射场表达，逆散射，最小二乘偏移， 地震波传播

1 引言

油气勘探地震学的最终目标是储层。通向储层描述与刻画的途径有两条（目前看来，典型的路线有两条）。第一条是常规的，目前主要采用的。它的核心是保幅处理与成像，它的核心结果是（相对）保真/保幅的角度成像道集。保幅处理和成像是勘探地震学走向储层地震学的关键中间环节，它前面与野外地震数据采集紧密相关，后面为叠后波阻抗反演、叠前 AVA/AVO 反演等提供扎实的基础数据。在野外地震数据采集质量有保证的前提下，完成保真的角度道集的生成的最关键因素在于较准确的背景参数模型，主要是速度模型、各向异性参数模型和 Q 值模型。这是目前常用的从地震成像走向储层刻画的方法技术体系。该技术体系的核心包括三方面的关键因素：满足速度估计和保真成像要求的数据体、较准确的背景速度模型和保真成像技术系列、基于保真成像道集的储层参数估计及储层描述方法。

很显然，还存在一条更直接的储层描述和刻画的途径。基于满足速度估计和保真成像要求的数据体，和较准确的背景速度模型，直接估计精细的速度参数扰动，利用速度参数的扰动来解释储层的形态变换和参数变化，与井数据结合进入储层含油气性的描述。这一个途径是以 FWI 为代表的速度估计技术不断实用化后预示的储层描述与刻画的新途径。与目前常用的基于角度道集的储层刻画方法相结合，会使得储层描述与刻画技术得到新的提高，更有利于油田开发的精细化，挖掘产能潜力。

这两条路线的核心是在野外地震数据的质量能满足要求的前提下，尽可能估计精确的速度参数（也包括各向异性和 Q 值参数）。

一些数学家看来，石油地震勘探问题很简单，就是根据实测的叠前地震数据，用波动方程作为数据预测器，若基于一个假设的参数模型求解波动方程得到的合成数据与实测的叠前地震数据的总误差最小，该假设模型就是所求的。

事实上，这完全低估了勘探地震反演问题的复杂性。

从勘探地震学的初始，勘探地震学家走的是一条实际问题驱动的路线来解决成像问题。由 CMP 道集估计初始速度，进行叠后偏移成像；由成像道集进行偏移速度分析，有时结合层析成像反演速度模型，进行叠前偏移成像，是石油工业界发展出的实用化的地震波偏移成像的路线。经检验是行之有效的。更细致的、能用于储层描述的参数估计是基于 1D 波阻抗反演、AVA 角度道集反演。

事实上，一直有很多人在探讨是否能从更数学化的路线来解决成像问题。这分成两条路线，一是 Tarentola（1984）为代表的全波形拟合非线性反演速度（还有其他）参数的路线；一是 Bleistein（1985）为代表的基于散射波表达的线性反演速度（还有其它）参数的路线。

这些路线之间是紧密相联的。目的都是走向储层的精确刻画。在本论文集的其他四篇文章中，我们比较清楚地讲明了它们之间的关系。

本文以地震波传播的散射场表达与逆散射成像为切入点，首先说明逆散射成像的本质，然后再试图理清楚一般意义下的叠前偏移成像与逆散射成像之间的区别与联系。

2 地震波传播的散射场表达

2.1 基本的声波方程和声散射势的引入

一般地,我们假设波在地下介质中的传播可以由如下的声波方程控制，

$$\left[\frac{1}{v(\mathbf{x})^2}\frac{\partial^2}{\partial t^2}-\rho(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla\right)\right]u(\mathbf{x},t)=S(\mathbf{x},t) \tag{1}$$

其中，$u(\mathbf{x},t)$ 为声场势，它是时间 $t\in R$ 和空间坐标 $\mathbf{x}\in\Omega\subseteq R$ 的函数。$v(\mathbf{x})$ 代表速度场；$\rho(\mathbf{x})$ 代表密度场；$S(\mathbf{x},t)$ 代表体积场源。

在频率空间域，上式变为

$$\mathcal{L}_a(\mathbf{x},\omega)u(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{2}$$

其中，$\mathcal{L}_a(\mathbf{x},\omega)=\nabla\cdot\frac{1}{\rho}\nabla+\frac{\omega^2}{\rho v^2(\mathbf{x})}$ 为传播算子。

定义背景介质中的声波传播满足如下方程，

$$\mathcal{L}_{a0}(\mathbf{x},\omega)u_0(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{3}$$

其中，$\mathcal{L}_{a0}(\mathbf{x},\omega)=\nabla\cdot\frac{1}{\rho_0}\nabla+\frac{\omega^2}{\rho_0 v_0^2(\mathbf{x})}$ 为背景介质中的声波算子。

根据（2）和（3）定义声散射势，

$$\begin{aligned}\mathcal{V}_a(\mathbf{x},\omega)&\equiv\mathcal{L}_a(\mathbf{x},\omega)-\mathcal{L}_{a0}(\mathbf{x},\omega)\\&=\nabla\cdot\frac{a_\rho}{\rho}\nabla+\omega^2\frac{a_v}{\rho v^2(\mathbf{x})}\end{aligned} \tag{4}$$

其中，$a_\rho=1-\frac{\rho_0(\mathbf{x})}{\rho(\mathbf{x})}$，$a_v(\mathbf{x})=1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}$。

2.2 标量波方程和标量波散射势的引入

勘探地震学中，很多情况下关注标量波方程（或密度不变情况下的声波方程）。因此，类似地，写出如下的标量波散射势。

在频率空间域标量波方程形式为

$$\mathcal{L}_s(\mathbf{x},\omega)u(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{5}$$

其中 $\mathcal{L}_s(x,\omega)=\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}$ 为标量波传播算子。

定义背景介质中的声波传播满足如下方程，

$$\mathcal{L}_{s0}(\mathbf{x},\omega)u_0(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{6}$$

其中 $\mathcal{L}_{s0}(\mathbf{x},\omega)=\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}$ 为背景介质中的标量波算子。

根据（5）和（6）式，定义标量波散射势为，

$$\begin{aligned}\mathcal{V}_s(\mathbf{x},\omega)&\equiv\mathcal{L}_s(\mathbf{x},\omega)-\mathcal{L}_{s0}(\mathbf{x},\omega)\\&=a_s(\mathbf{x})\frac{\omega^2}{v_0^2(\mathbf{x})}\end{aligned}\tag{7}$$

其中，$a_s(\mathbf{x})=1-\dfrac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}$。

上式定义的散射势还可以写为如下的形式，

$$\begin{aligned}\mathcal{V}_s(\mathbf{x},\omega)&=\omega^2\frac{1}{v_0^2(\mathbf{x})}\left(1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}\right)\\&\approx\omega^2\frac{1}{v_0^2(\mathbf{x})}\frac{2\delta v\cdot v_0}{v_0^2(\mathbf{x})}=\omega^2\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\end{aligned}\tag{8}$$

上述形式引入 $\delta v(\mathbf{x})=v(\mathbf{x})-v_0(\mathbf{x})$ 更清楚地反映出速度的局部变化是散射势的本质。从散射势的定义也可以看出，估计散射势的绕射层析成像与估计或定位反射系数的叠前偏移成像是有严格区别的。

2.3 Lippmann-Schwinger 方程的导出

借助于算子恒等式 $\mathcal{L}=\mathcal{L}_0+\mathcal{L}_0\left(\mathcal{L}_0^{-1}-\mathcal{L}^{-1}\right)\mathcal{L}$ 和定义 $\mathcal{V}=\mathcal{L}_0^{-1}-\mathcal{L}^{-1}$ 可以得到

$$\mathcal{L}=\mathcal{L}_0+\mathcal{L}_0\mathcal{V}\mathcal{L}\tag{9}$$

算子 $\mathcal{L}$ 是对应声波和标量波的Green函数。它描述所有的波传播现象。$\mathcal{L}^{-1}$ 和 $\mathcal{L}_0^{-1}$ 是“逆”Green函数，它们经常被各自的共轭转置来代替。物理上，它们描述的是波的反传播现象。Green 函数可以描述从源到散射点，再到接收点的波传播；同样地，也可以描述从接收点到散射点，再到源点的波传播过程。假如波传播算子 $\mathcal{L}$ 用 Green 函数 G 表示，方程（9）可以重写为

$$G=G_0+G_0VG\tag{10}$$

方程（10）是非线性的，它能被进一步重写为

$$G=\left(I-G_0V\right)^{-1}G_0\tag{11}$$

对上式的右端项进行 Taylor 展开，

$$\begin{aligned}G=\left[\sum_{j=0}^{\infty}\left(G_0\mathcal{V}\right)^j\right]G_0=G_0+G_0\mathcal{V}G_0\\+G_0\mathcal{V}G_0\mathcal{V}G_0+G_0\mathcal{V}G_0\mathcal{V}G_0\mathcal{V}G_0+\cdots\end{aligned}\tag{12}$$

方程(10)和(12)被称为 Lippmann Schwinger 方程。

很显然，假如 $j<2$,方程（12）就仅描述了一个忽略了二阶以上散射波的波传播过程。因此，线性化的波传播算子仅仅描述波传播过程中的一阶散射。这就是

$$G=G_0+G_0\mathcal{V}G_0\tag{13}$$

上式就是 Born 近似下的波传播算子，其中方程右端的第一项描述了背景场中的传播现象，第二项描述了一阶散射波的传播现象。方程（10）-（13）清楚地说明了 Born 近似的物理含义。只有当散射势比较弱时，（13）式中的高于二阶的项才可以被忽略不计。注意，从（12）式可以清楚地看出，高阶散射场能否被忽略，除了散射势比较弱的条件外，还有散射体的体积也很重要，因为散射场是散射势的体积积分结果。小散射体和弱散射势是 Born 近似成立的条件。

2.4 散射场的表达

从（13）式可以看出，Born 近似线性化模拟的总波场包括两部分：第一部分是由背景 Green 函数所描述的背景波场；第二部分是由散射势引起的方程（12）中的第二项所描述的散射场。很显然，这存在一个背景速度场如何取才能得到符合 Born 近似要求的背景场的问题。原则性的要求是入射场中不能存在散射场成分。根据这样的原则，如何决定散射势中的背景速度 $v_0(\mathbf{x})$?这个问题是很难回答的！

事实上，在特别复杂，变化特别剧烈的速度场中，即使是 Lippmann-Schwinger 方程的有效性都是很值得质疑的。

假定 Born 近似假设成立，根据方程（13），线性化的总波场可以写成

$$\begin{aligned} u(\boldsymbol{\xi},\boldsymbol{\eta},\omega) &= u_0(\boldsymbol{\xi},\boldsymbol{\eta},\omega) + \\ &\omega^2 \int d\mathbf{x} G_0(\boldsymbol{\xi},\mathbf{x},\omega) \frac{a_s(\mathbf{x})}{v_0^2(\mathbf{x})} u_0(\mathbf{x},\boldsymbol{\eta},\omega) \end{aligned} \tag{14}$$

其中，Born 近似后的散射波场记为

$$\begin{aligned} & u_s(\boldsymbol{\xi},\boldsymbol{\eta},\omega) \\ &= \omega^2 \int d\mathbf{x} G_0(\boldsymbol{\xi},\mathbf{x},\omega) \frac{a(\mathbf{x})}{v_0^2(\mathbf{x})} u_0(\mathbf{x},\boldsymbol{\eta},\omega) \\ &= \omega^2 \int d\mathbf{x} G_0(\boldsymbol{\xi},\mathbf{x},\omega) u_0(\mathbf{x},\boldsymbol{\eta},\omega) f(\mathbf{x}) \end{aligned} \tag{15}$$

它代表从震源 $\boldsymbol{\eta}$ 到散射点 $\mathbf{x}$，再到接收点 $\boldsymbol{\xi}$ 的散射波场。

Born 近似的物理实质是，在背景场 $u_0(\boldsymbol{\xi},\boldsymbol{\eta},\omega)$ 中没有散射场，在扰动场中，只存在一次散射场。尽管用这种方式描述波传播与实际的波传播不同，但当二阶及二阶以上的散射场能量比较弱时，用 Born 近似进行 PSDM 和 Tomography 还是比较有效的。

关于子波问题，可以认为已经携带在背景场 $u_0(\mathbf{x},\boldsymbol{\eta},\omega)$ 中。也可以用 $G_0(\mathbf{x},\boldsymbol{\eta},\omega)S(\omega)$ 来表示，其中 $S(\omega)$ 代表子波的谱。

2.5 散射势的进一步解释

标量波方程定义的散射势形式为，

$$\mathcal{V}_s(\mathbf{x},\omega) \equiv a_s(\mathbf{x}) \frac{\omega^2}{v_0^2(\mathbf{x})} = a_s(\mathbf{x}) \left|\mathbf{k}_0(\mathbf{x})\right|^2 \tag{16}$$

可见散射势与场的波矢量之间是有密切关系的。在某种情况下，已知任意空间点上的全部的场的波矢量成分，可以完全地刻画散射势。

由于波矢量代表波的传播方向，且 $\mathbf{k}_0 = \mathbf{k}_0^S + \mathbf{k}_0^R$，通过这样的关系可以联系起散射势与角度反射系数之间的关系。

改变（16）式，得到

$$\begin{aligned} \mathcal{V}_s(\mathbf{x},\omega) &\equiv a_s(\mathbf{x}) \frac{\omega^2}{v_0^2(\mathbf{x})} \approx \left|\mathbf{k}_0(\mathbf{x})\right|^2 \frac{(v-v_0)(v+v_0)}{\left(\frac{v+v_0}{2}\right)\left(\frac{v+v_0}{2}\right)} \\ &= \left|\mathbf{k}_0(\mathbf{x})\right|^2 \frac{4(v-v_0)}{(v+v_0)} = \left|\mathbf{k}_0(\mathbf{x})\right|^2 4R(\theta=0^o) \end{aligned} \tag{17}$$

可见散射势可以写成与垂直反射系数有关的式子。原则上，非垂直入射的反射系数与散射势的关系可以用波场波矢量来规定。

上述一段说明试图把散射势的反演与反射系数的估计统一在一起。

3 逆散射问题分析

3.1 常速背景介质与平面波入射情形下散射场与散射势间的关系

首先给出常速介质中自由空间 Green 函数为

$$G(\mathbf{x}-\mathbf{x}',\omega)=\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}=\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|} \tag{18}$$

其中，$\mathbf{x}$ 代表场点位置，$\mathbf{x}'$ 代表源点位置，$|\mathbf{x}-\mathbf{x}'|$ 代表源点与场点之间的距离。$|\mathbf{k}|=\left|\omega/c\right|$ 代表场的波数。把（18）式代入（15）式定义的散射场中，得到

$$u_s(\mathbf{x},\omega)=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)u_0(\mathbf{x}',\omega)\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|} \tag{19}$$

一般地，在远场近似假设下面，入射场在常速介质中传播，可以用平面波近似，

$$u(\mathbf{x},\omega)=e^{i\mathbf{k}\cdot\mathbf{x}}=e^{i|k|\mathbf{s}_0\cdot\mathbf{x}} \tag{20}$$

其中，$\mathbf{s}_0$ 是代表平面波的入射方向的单位矢量。将（20）式代入（19）式得到，

$$u_s(\mathbf{x},\omega)=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\mathbf{x}'}\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|} \tag{21}$$

式(21)描述的波的传播关系如图 1 所示。

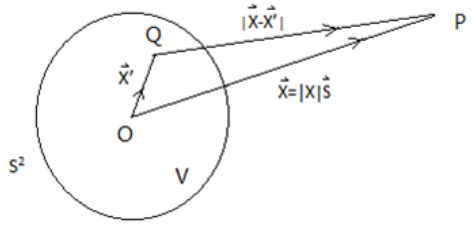


图 1 空间局部的散射体引起的散射波传播关系

总场表示为

$$\begin{aligned}u(\mathbf{x},\omega)&=u_0(\mathbf{x},\omega)+u_s(\mathbf{x},\omega)\\&=e^{ik\mathbf{s}_0\cdot\mathbf{x}}+\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\cdot\mathbf{x}'}\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}\end{aligned} \tag{22}$$

由图 1 可知，当场点 P 距离散射体比较远，散射体积源分布空间有限时，用 Q 代表散射体积源内的点，O 代表坐标原点，则可以做如下的近似推导

$$|\mathbf{x}-\mathbf{x}'|=|\mathbf{x}|-\mathbf{s}\cdot\mathbf{x}' \tag{23}$$

$$G(\mathbf{x}-\mathbf{x}',\omega)=$$

$$\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}=\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}\approx\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}e^{-ik\mathbf{s}\cdot\mathbf{x}'} \tag{24}$$

其中，$\mathbf{s}$ 是代表散射波传播的局部平面波方向的单位矢量。把（24）式代入（22）式，得到

$$
\begin{aligned}
u(\mathbf{x},\omega) &\approx e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\cdot\mathbf{x}'}\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}e^{-ik\mathbf{s}\cdot\mathbf{x}'} \\
&= e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'}\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \\
&= e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'} \\
&= e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}f(\mathbf{s},\mathbf{s}_0,\omega)
\end{aligned} \tag{25}
$$

$$
f(\mathbf{s},\mathbf{s}_0,\omega) = \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'} \tag{26}
$$

称为辐射势。注意到，辐射势是一个角度和频率决定的二维谱，它的完整性表示了散射场的完整性，也决定了散射势波数谱的完整性。

分析（26）式知，定义散射势的 Fourier 变换为，

$$
\tilde{\mathcal{V}}(\mathbf{k}') = \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-i\mathbf{k}'\cdot\mathbf{x}'} \tag{27}
$$

其中，$\mathbf{k}'$ 代表散射势的波数。已知所有的波数成分等价于确定了精确的散射势。对比（26）和（27）知，

$$
f(\mathbf{s},\mathbf{s}_0,\omega) = \tilde{\mathcal{V}}\left(k(\mathbf{s}-\mathbf{s}_0)\right) \tag{28}
$$

据此，可以总结出如下定理。

定理：在 Born 近似下，相对于散射体的远场区域，在单位向量 $\mathbf{s}$ 所规定的方向上，散射波的振幅谱完全取决于散射势的某一个 Fourier 分量，即下式定义的分量 $\mathbf{k}' = \frac{\omega}{c}(\mathbf{s}-\mathbf{s}_0)$。

由此可知，在常速介质情况下所谓逆散射问题就是根据观测到的散射波的频谱，确定散射势的 Fourier 的分量。散射势的所有的 Fourier 分量都确定了，通过 Fourier 逆变换，就可以完全精确地恢复散射势。

上述定理尽管是在平面波入射和常速背景情况下得到的，其理论指导意义是十分明显的。从 $\mathbf{k}'$ 的表达式中可以看出，确定 $\mathbf{k}'$ 的低波数部分，需要大角度散射波和低频数据；确定 $\mathbf{k}'$ 的高波数部分，需要小角度散射波和高频数据。总之，观测系统对反演成像的分辨率起着重要的作用。总之，震源子波的频带范围、观测系统和背景速度场的分布决定了反演成像的分辨率。

3.2 高频近似下的 Born 近似的物理含义解释

Born 近似下的散射场也可以表示为，

$$
u_s(\boldsymbol{\xi},\boldsymbol{\eta},\omega) = \omega^2\int d\mathbf{x}G_0(\boldsymbol{\xi},\mathbf{x},\omega)a(\mathbf{x})G_0(\mathbf{x},\boldsymbol{\eta},\omega) \tag{29}
$$

其中，速度扰动为 $a(\mathbf{x}) = 2c_0^{-3}(\mathbf{x})\delta c(\mathbf{x})$。用渐进 Green 函数（即 Green 函数的 WKBJ 近似解）代入上式，

$$
u_s(\boldsymbol{\xi},\boldsymbol{\eta},\omega) = \omega^2\int d\mathbf{x}a(\mathbf{x})A(\mathbf{x},\boldsymbol{\eta})A(\boldsymbol{\xi},\mathbf{x})\exp\left[-i\omega\left(T(\mathbf{x},\boldsymbol{\eta})+T(\boldsymbol{\xi},\mathbf{x})\right)\right] \tag{30}
$$

在散射点周围，对旅行时进行线性化，即围绕中心射线进行 Taylor 展开，

$$
T(\mathbf{x},\mathbf{x}') \approx T(\mathbf{x}_0,\mathbf{x}') + (\mathbf{x}-\mathbf{x}_0)\cdot\mathbf{p}(\mathbf{x}_0,\mathbf{x}') \tag{31}
$$

其中 $\mathbf{p}(\mathbf{x}_0,\mathbf{x}')$ 表示旅行时场 $T(\mathbf{x},\mathbf{x}')$ 在点 $\mathbf{x}_0$ 处的梯度，其方向为中心射线的传播方向。因此，（30）式中的指数部分可以化为，

$$\exp\left[-i\omega\left(T\left(\mathbf{x},\boldsymbol{\eta}\right)+T\left(\boldsymbol{\xi},\mathbf{x}\right)\right)\right]$$
$$=\exp\left[-i\omega\left(T\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+T\left(\boldsymbol{\xi},\mathbf{x}_0\right)\right)\right]\exp\left[-i\omega\left(\mathbf{x}-\mathbf{x}_0\right)\left(\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+\mathbf{p}\left(\boldsymbol{\xi},\mathbf{x}_0\right)\right)\right] \tag{32}$$
$$=\exp\left[-i\omega\left(T\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+T\left(\boldsymbol{\xi},\mathbf{x}_0\right)\right)\right]\exp\left[-i\mathbf{k}\cdot\left(\mathbf{x}-\mathbf{x}_0\right)\right]$$
$$\mathbf{k}=\omega\left(\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\xi}\right)\right) \tag{33}$$

式（33）的几何含义如图 2 所示。

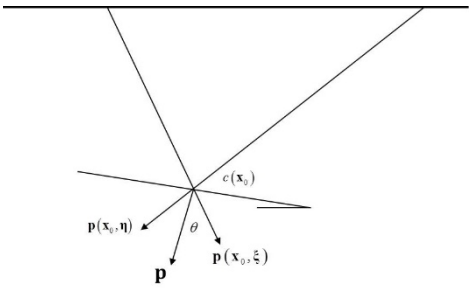


图 2 地震波传播过程中波矢量在反射点处的几何关系

如图 2 所示，存在关系 $\mathbf{p}=\omega\left(\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\xi}\right)\right)=2\mathbf{n}\cos\theta/c\left(\mathbf{x}_0\right)$。

其中 $\mathbf{n}$ 为界面的外法线单位方向向量。θ 为射线的入射角。$c\left(\mathbf{x}_0\right)$ 为 $\mathbf{x}_0$ 处的速度。根据上面的几何关系则有，

$$\mathbf{k}=\omega\mathbf{p}=\frac{2\omega\cos\theta}{c\left(\mathbf{x}_0\right)}\mathbf{n} \tag{34}$$

从 $\mathbf{k}$ 的表达式中可以看出，确定 $\mathbf{k}$ 的低波数部分，需要大角度散射波和低频数据；确定 $\mathbf{k}$ 的高波数部分，需要小角度散射波和高频数据。值得说明的是，此处的波数 $\mathbf{k}$ 与前面讨论的辐射势是关系密切的。因此，我们希望 $\mathbf{k}$ 规定的谱的完整性。

此处结论与全波形速度反演得到认识是一致的。

4 最小二乘散射势估计

4.1 散射势估计的几种方法

研究散射场表达的主要目的是建立起散射场与散射势之间的映射关系。上述映射关系实质上是一种广义 Radon 变换。散射场是投影结果，通过反投影来估计散射势就是逆散射问题。

逆散射问题的求解在不同应用领域存在较大的差异，主要看背景介质的复杂程度及相应的 Green 函数计算的复杂程度。相应地分为三大类。

（1）常速背景介质下，用平面波作为入射场，导出一类 Fourier 反变换方法估计 $a\left(\mathbf{x}\right)$。勘探地震中，相应地存在 $v\left(z\right)$ 介质（即水平层状介质）情况下的一类 Fourier 反变换方法。这是勘探地震领域上世纪 80 年代初关注的问题。

（2）缓变光滑背景介质下，利用拟微分算子理论导出一类直接反演方法估计 $a\left(\mathbf{x}\right)$。Bleistein（1985）在上世纪 80 年代中后期完整地讨论了此问题。Beylkin（1985）的研究工作更出色地解决了该问题。但这类方法基于拟微分算子理论，要求的数学基础非常高。

（3）任意变化的光滑背景介质情况下，基于 Bayes 估计理论，建立误差泛函，求解一类线性优化问题，估计 $a\left(\mathbf{x}\right)$。这是目前勘探地震学中逆散射问题的主流解法。相对而言，这种解法与勘探地震学面对的介质变化情况最接近，解法也比较容易掌握和理解。

4.2 拟微分算子理论下散射势的估计方法

4.2.1 对偶空间、对偶算子与散射势估计的对偶表达

为了估计散射势 $f(\mathbf{x})$，把（15）式重写为，

$$\hat{u}(\omega,\boldsymbol{\xi},\boldsymbol{\eta})=-\omega^2\int_X \mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})f(\mathbf{x})d\mathbf{x} \tag{35}$$

其中，$\mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$ 是积分核函数。它代表脉冲源激发的波从震源 $\boldsymbol{\eta}$ 到散射点 $\mathbf{x}$，再到接收点 $\boldsymbol{\xi}$ 的波传播过程后的响应。

Fredholm 第一类积分方程的求解是很困难的，主要是因为这类积分方程所对应的矩阵方程在大部分情况下条件数都很大。对于这类积分方程也可以用如下的对偶空间理论来近似求解。在选定的基函数下面，数据空间和模型空间可以认为是一个对偶空间。Born 近似下的 Green 函数就定义了这样一个基函数，它们也形成了一个映射 $\mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$，把模型空间函数 $f(\mathbf{x})$ 映射到数据空间函数 $\hat{u}(\omega,\boldsymbol{\xi},\boldsymbol{\eta})$。在对偶空间框架下存在映射 $\mathrm{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$ 的对偶映射 $\hat{\mathrm{M}}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$，定义下式，

$$\left\langle u(\omega,\boldsymbol{\xi},\boldsymbol{\eta}),M(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})f(\mathbf{x})\right\rangle_{S_D}\triangleq\left\langle f(\mathbf{x}),\hat{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})u(\omega,\boldsymbol{\xi},\boldsymbol{\eta})\right\rangle_{S_M} \tag{36}$$

此处，S_D 代表数据空间；S_M 代表模型空间。上式是广义的 Parsaval 定理。

基于上述对偶映射，可以直接地形式地写出扰动的估计式

$$f_{est}(\mathbf{x})=2\,\mathrm{Re}\int_{\partial X_{rec}}\int_0^\infty \hat{M}(\omega,\xi,\eta,x)\tilde{u}(\omega,\xi,\eta)d\xi d\omega \tag{37}$$

上式就是利用单炮数据估计参数扰动的公式。对于某些积分核函数，有

$$\bar{\hat{\mathrm{M}}}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})=\hat{\mathrm{M}}(-\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) \tag{38}$$

则（37）式就变为，

$$f_{est}(\mathbf{x})=\int_{\partial X_{rec}}\int_0^\infty \hat{M}(\omega,\xi,\eta,x)\tilde{u}(\omega,\xi,\eta)d\xi d\omega \tag{39}$$

4.2.2 Born 近似下散射势的估计方法

用下面任意两个实函数以及它们的 Fourier 变换之间的关系，

$$\int_{-\infty}^{+\infty} g(t)h(t)dt=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\bar{\tilde{g}}(\omega)\tilde{h}(\omega)d\omega \tag{40}$$

因此，

$$M(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})=2\,\mathrm{Re}\int_0^{+\infty}\bar{\tilde{M}}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})e^{-i\omega t}dt \tag{41a}$$

$$\tilde{M}(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})=\frac{1}{2\pi}\int_{-\infty}^{+\infty}M(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})e^{-i\omega t}dt \tag{41b}$$

由（37）、（39）式可知，扰动参数估计问题变为了确定投影算子或共轭投影算子问题。可以给出如下的投影算子，

$$M(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})=-\frac{1}{8\pi^2}\delta\left(t-\phi^{out}(\mathbf{x},\boldsymbol{\xi})-\phi^{in}(\mathbf{x},\boldsymbol{\eta})\right)\frac{h(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta})}{A^{out}(\mathbf{x},\boldsymbol{\xi})A^{in}(\mathbf{x},\boldsymbol{\eta})} \tag{42a}$$

上式是时间域形式，将其转为频率域形式可得

$$\tilde{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{16\pi^3}\frac{e^{-i\omega\left(\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)+\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)}}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right) \tag{42b}$$

其中，相位函数和振幅函数首先满足如下的 Green 函数的射线近似解，即 WKBJ 解，其中接收点对应的 Green 函数的解为

$$\mathrm{G}^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)=-A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)e^{i\omega\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)} \tag{43a}$$

炮点对应的 Green 函数解为

$$\mathrm{G}^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)=-A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)e^{i\omega\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)} \tag{43b}$$

相位函数 $\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$ 和 $\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$ 满足如下的程函方程，

$$\left(\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)^2=S_0^2\left(\mathbf{x}\right) \tag{44a}$$

$$\left(\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\right)^2=S_0^2\left(\mathbf{x}\right) \tag{44b}$$

$\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$ 代表从炮点位置 $\boldsymbol{\eta}$ 到散射点 $\mathbf{x}$ 的在背景介质中的旅行时。$\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$ 代表从散射点 $\mathbf{x}$ 到接收点 $\boldsymbol{\xi}$ 的在背景介质中的旅行时。振幅函数 $A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$ 和 $A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$ 满足如下一阶输运方程，

$$\begin{aligned}&A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\nabla_{\mathbf{x}}^2\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)+\\&2\nabla_{\mathbf{x}}A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\cdot\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)=0\end{aligned} \tag{45a}$$

$$\begin{aligned}&A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\nabla_{\mathbf{x}}^2\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)+\\&2\nabla_{\mathbf{x}}A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\cdot\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)=0\end{aligned} \tag{45b}$$

旅行时的计算可以直接解方程（44）或把上式变成各种形式的一阶常微分方程组。关于振幅项对（45）式的求解直接在笛卡尔坐标系中进行会遇到焦散问题。Maslov 方法（在相空间中解方程（45））和 Gaussian_Beam 方法（在射线中心坐标系中用傍轴射线理论解方程（45））会回避掉此问题。其他各种变种的方法也在不断地发展中。

（42）中的函数 $h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 是一个 Jacobian 变换，它把对 $\mathbf{x}$ 的积分变换为对 $\left(\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 的积分。Beylkin（1985）给出

$$h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)d\xi=S_0^3\left(1+\cos\psi\right)d\Omega \tag{46}$$

其中，Ω 是立体角。且有

$$\cos\psi\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)=\frac{\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\cdot\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\xi}\right)}{S_0^2\left(\mathbf{x}\right)} \tag{47}$$

$\psi\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 是由坐标 $\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$ 和介质分布决定的波传播的反射张角，即入射射线和出射射线的夹角。

（42）式定义了变背景和任意炮检点分布时的偏移成像中的积分核函数。事实上，（42）式中的 Green 函数还可以不用 WKBJ 近似来表示，此时，推出的积分核意义更广泛。则有下式

$$\tilde{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{16\pi^3}\frac{\hat{G}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)}{\left|\hat{G}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)\right|}\frac{\hat{G}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)}{\left|\hat{G}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)\right|}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right) \tag{48}$$

在时间域中，上式变为，

$$M\left(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\breve{G}^{out}\left(t-\tau,\boldsymbol{\xi},\mathbf{x}\right)\breve{G}^{in}\left(t-\tau,\boldsymbol{\eta},\mathbf{x}\right)h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)d\tau \tag{49}$$

其中，

$$\breve{G}^{out}\left(t,\boldsymbol{\xi},\mathbf{x}\right)=2\,\mathrm{Re}\int_{-\infty}^{+\infty}\frac{G^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)}{\left|G^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)\right|^{2}}e^{-i\omega t}d\omega \tag{50a}$$

$$\breve{G}^{in}\left(t,\boldsymbol{\eta},\mathbf{x}\right)=2\,\mathrm{Re}\int_{-\infty}^{+\infty}\frac{G^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)}{\left|G^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)\right|^{2}}e^{-i\omega t}d\omega \tag{50b}$$

可以写出如下两种不同的积分形式，对应两种不同的偏移成像的说明。形式 I 为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{8\pi^{2}}\int_{\partial X_{rec}}u\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)\Big|_{t=\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)+\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}\frac{h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}d\boldsymbol{\xi}d\omega \tag{51}$$

很显然，上式就是 Kirchhoff 积分的绕射叠加。就是说，从 Born 近似理论也可以推出 Kirchhoff 积分绕射叠加偏移公式。

把（42b）或（48）式代入到（49）式可以得到另一种形式的偏移成像公式。形式 II 为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{16\pi^{3}}\mathrm{Re}\int_{0}^{\infty}\frac{e^{-i\omega\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}}{A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}\int_{\partial X_{rec}}\frac{e^{-i\omega\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)\tilde{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\omega \tag{52a}$$

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{16\pi^{3}}\mathrm{Re}\int_{0}^{\infty}\frac{\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\eta}\right)}{\left|\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\eta}\right)\right|^{2}}\int_{\partial X_{rec}}\frac{\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\xi}\right)}{\left|\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\xi}\right)\right|^{2}}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)\tilde{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\omega \tag{52b}$$

（51）对应的时域形式为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{8\pi^{2}}\int_{0}^{\infty}G^{in}\left(t,\mathbf{x},\boldsymbol{\eta}\right)w\left(t,\mathbf{x},\boldsymbol{\eta}\right)dt \tag{53}$$

其中

$$w\left(t,\mathbf{x},\boldsymbol{\eta}\right)=\int_{0}^{\infty}\int_{\partial X_{rec}}G^{out}\left(t-\tau,\mathbf{x},\boldsymbol{\xi}\right)h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)u\left(\tau,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\tau \tag{54}$$

（53）式的含义就是（54）式进行检波点波场反传播，然后与下行波场进行相关成像。此为 Clearbout（1985）的成像条件。

（51）、（52）和（53）分别是不同形式的叠前偏移成像表达式。成像结果是地下介质的扰动（散射势）。上述三式用更深刻的数学背景阐述了地震波叠前偏移成像的本质。它就是 Born 近似下线性化反演的近似解。

值得注意的是，上述理论意义下的得到叠前深度偏移结果，就是对速度扰动量的估计。不是我们常认为的带限的反射系数的估计。

4.3 Bayes 框架下散射势估计方法

（15）式定义了一个 Fredholm 第一类积分方程，它描述的是一个线性问题。此处，我们把该 Fredholm 第一类积分方程的求解转化为一个 Bayes 估计问题。就是把该线性问题的求解，转化为利用泛函变分求极值问题。

基本的表述如下。假设（15）式定义的积分方程写出算子表达形式，

$$K\mathbf{m}=\mathbf{d} \tag{55}$$

其中，K 是 Fredholm 第一类积分方程的积分核，由（15）式知，积分核的定义为

$$K\left(v_{0}\left(\mathbf{x}\right),\mathbf{x},\omega\right)=\omega^{2}G_{0}\left(\boldsymbol{\xi},\mathbf{x},\omega\right)u_{0}\left(\mathbf{x},\boldsymbol{\eta},\omega\right) \tag{56}$$

$\mathbf{m}$ 代表由 $a\left(\mathbf{x}\right)\Big/v_{0}^{2}\left(\mathbf{x}\right)$ 作为元素形成的矢量；$\mathbf{d}$ 代表由每个炮检点对应的地震道的频谱值。

认为$K:\mathrm{H}_1\to\mathrm{H}_2$是一个有界的线性算子，定义如下的 Tikhonov 泛函，

$$\mathfrak{I}(\mathbf{m})=\frac{1}{2}\|K\mathbf{m}-\mathbf{d}\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\|\mathbf{m}\|_{\mathrm{H}_1}^2 \tag{57}$$

对上式求解的首要问题是计算泛函关于参数变化的导数，即泛函的梯度。为此，利用如下的变分方法。对任意的介质扰动$\delta\mathbf{m}\in\mathrm{H}_1$和扰动步长$\tau\in\mathbb{R}$，总存在如下的泛函变分

$$\begin{aligned}
&\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})\\
&=\frac{1}{2}\|K(\mathbf{m}+\tau\delta\mathbf{m})-\mathbf{d}\|_{\mathrm{H}_2}^2\\
&+\frac{\alpha}{2}\|\mathbf{m}+\tau\delta\mathbf{m}\|_{\mathrm{H}_1}^2\\
&=\frac{1}{2}\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}+\\
&\frac{\tau}{2}\begin{pmatrix}\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\\ \langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}\end{pmatrix}+\\
&\frac{\tau^2}{2}\langle K\delta\mathbf{m},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\frac{\alpha}{2}\langle\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}+\\
&\frac{\alpha}{2}\tau\begin{pmatrix}\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}+\\ \langle\delta\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}\end{pmatrix}+\frac{\alpha\tau^2}{2}\langle\delta\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}
\end{aligned}
=\begin{aligned}
&\frac{1}{2}\left(\|K\mathbf{m}-\mathbf{d}\|_{\mathrm{H}_2}^2+\alpha\|\mathbf{m}\|_{\mathrm{H}_1}^2\right)+\\
&\frac{\tau}{2}\begin{pmatrix}\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\\ \langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\rangle_{\mathrm{H}_2}\end{pmatrix}+\\
&\frac{\alpha}{2}\tau\left(\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}+\langle\delta\mathbf{m},\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\|K\delta\mathbf{m}\|_{\mathrm{H}_2}^2+\alpha\|\delta\mathbf{m}\|_{\mathrm{H}_1}^2\right)
\end{aligned} \tag{58}$$

对上式定义的变分求偏导数，

$$\begin{aligned}
&\frac{\partial}{\partial\tau}\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})\Big|_{\tau=0}\\
&=\left\{\begin{matrix}\left(\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\alpha\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\ +\tau\left(\|K\delta\mathbf{m}\|_{\mathrm{H}_2}^2+\alpha\|\delta\mathbf{m}\|_{\mathrm{H}_1}^2\right)\end{matrix}\right\}\Bigg|_{\tau=0}\\
&=\left(\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{\mathrm{H}_2}+\alpha\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\
&=\left(\langle K^*(K\mathbf{m}-\mathbf{d}),\delta\mathbf{m}\rangle_{\mathrm{H}_1}+\alpha\langle\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}\right)\\
&=\langle K^*(K\mathbf{m}-\mathbf{d})+\alpha\mathbf{m},\delta\mathbf{m}\rangle_{\mathrm{H}_1}\\
&=\langle(K^*K+\alpha I)\mathbf{m}-K^*\mathbf{d},\delta\mathbf{m}\rangle_{\mathrm{H}_1}\\
&=\langle\mathrm{Grad}\mathfrak{I}(\mathbf{m}),\delta\mathbf{m}\rangle_{\mathrm{H}_1}
\end{aligned} \tag{59}$$

其中，

$$\mathrm{Grad}\mathfrak{I}(\mathbf{m})=(K^*K+\alpha\mathbf{I})\mathbf{m}-K^*\mathbf{d} \tag{60}$$

是误差泛函相对于参数模型的梯度。经典地，我们利用该梯度构造最速下降迭代方法求解 Born 近似后的线性问题的解。迭代格式为，

$$
\begin{aligned}
&\mathbf{m}^{k+1}\\
=&\mathbf{m}^{k}-\mu\mathrm{Grad}\mathfrak{J}\left(\mathbf{m}\right)\\
=&\mathbf{m}^{k}-\mu\left[K^{*}\left(K\mathbf{m}-\mathbf{d}^{obs}\right)\right]\\
=&\mathbf{m}^{k}-\mu\left[K^{*}\left(K\mathbf{m}^{k}-\mathbf{d}^{obs}\right)\right]\\
=&\mathbf{m}^{k}-\mu\left[K^{*}\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)\right]
\end{aligned}
\tag{61}
$$

当梯度等于零时，可以得到线性系统的法方程

$$\left(K^{*}K+\alpha\mathbf{I}\right)\mathbf{m}=K^{*}\mathbf{d} \tag{62}$$

法方程的求解一般用共轭梯度方法。但是，由于勘探地震中模型描述用网格化方式，模型参数个数非常多，（62）式右端的 Hessian 矩阵异常庞大。若用 $NX\times NY\times NZ$ 表示要估计的模型参数的个数，那么 Hessian 矩阵的规模为 $\left(NX\times NY\times NZ\right)\times\left(NX\times NY\times NZ\right)$。当前的计算机技术还不能满足这样巨大规模的线性方程组的求解。形成 Hessian 矩阵本身就是一个巨大的挑战。所以，当前解法方程进行最小二乘偏移成像的基本做法是先进行叠前深度偏移成像，然后用对角矩阵对成像叠加后的结果进行照明补偿，就是用 Hessian 矩阵的对角元素去除成像结果。也可以用角度 Hessian 的对角元素去除每个共角度成像剖面，然后进行叠加产生最后的剖面，当然也可以不叠加就此产生照明补偿后的角度成像道集。最一般的做法应该对每个单炮道集进行照明补偿的最小二乘偏移，但这样做会由于弱照明产生很强的成像噪音（实质上是反问题的计算不稳定性）。另外，很多人试图把成像限制在局部范围内，在小的空间范围内实现全 Hessain 求逆进行最小二乘反演。

这也是迭代类的方法比较流行的原因。

迭代的最小二乘方法中，每次模型更新后，原则上都要重新计算核函数 K，并计算波场 $\mathbf{d}^{cal}$。重新计算核函数 K，需要计算出更新后的速度场，可以用下式进行

$$v\left(\mathbf{x}\right)=v_{0}\left(\mathbf{x}\right)\Big/\sqrt{1-\mathbf{m}\left(\mathbf{x}\right)v_{0}^{2}\left(\mathbf{x}\right)} \tag{63}$$

很多时候，往往并不更新核函数，因为我们假定核函数 $K\left(v_{0}\left(\mathbf{x}\right),\mathbf{x},\omega\right)$ 仅仅与背景速度 $v_{0}\left(\mathbf{x}\right)$ 有关。但这仅仅是从计算效率方面考虑的，本质上还是希望更新核函数。可能的计算策略当然包括间隔相邻的几个迭代步更新一次核函数。射线层析的迭代计算原则也是这样的。

（61）式中计算出的量 $a\left(\mathbf{x}\right)\Big/v_{0}^{2}\left(\mathbf{x}\right)$ 或者 $2\delta v\left(\mathbf{x}\right)\Big/v_{0}^{3}\left(\mathbf{x}\right)$ 代表的是速度的一种相对变化。只有在速度跃变的界面或绕射点处才会出现，因此，它反映了速度场中的高频变化部分。用这部分高波数速度扰动可以代表界面位置和绕射位置图像。这与一般的叠前深度偏移是很类似的。也可以精确估计这部分高波数速度扰动，用它们来刻画储层的性质。但是，我们认为，只有全波数带的速度场或合理的宽带速度场才可以进行有效的储层解释。这也是 FWI 技术用于储层解释时存在的最大问题，即不能提供合理有效的宽带速度场。

另一方面，在做数据空间的波场逼近时，还要考

虑 $\mathbf{d}^{cal}$ 的规则化问题，或者统一两个数据的空间采样方式。

4.4 层析成像中的残差反传播

迭代最小二乘绕射层析成像的基本思想可以从（61）式中清楚地看出。就是通过共轭算子 K^{*}，把利用 $v_{0}\left(\mathbf{x}\right)$，$a\left(\mathbf{x}\right)\Big/v_{0}^{2}\left(\mathbf{x}\right)$ 和 Green 函数计算出的预测波场和观测散射场的残差进行反传播。

我们希望，与射线层析沿着射线路径进行旅行时残差反投影一样，沿着波路径进行波场残差的反投影。另一方面，反射波的层析成像和透射波的层析成像是存在差异的，（15）式表达的散射场是背景介质中的散射势引起的，由背景介质中的 Green 计算的。很显然，（15）式自然地适应透射波的层析成像，但不合适

反射波的层析成像。主要是没有反射界面的存在，根本就不可能计算出炮点、绕射（反射）点到接收点的 Green 函数。可以说，（15）式根本就没有描述反射波情形下的波传播关系。这是对（15）式的解读必须认识到的。否则，基于此式的叠前深度偏移和层析成像就无法恰当地进行。利用（62）式直接进行反射波叠前深度偏移和层析成像根本就是不可能的，原因就在于此。

地震勘探主要是利用地震反射波进行反演成像，这是它非常不同于一般的无损探伤成像、医学层析成像的根本物理原因。

反射波的残差反投影有两种形式，一是把残差反投影到等时面上，修改反射界面处的高波数参数扰动；二是把残差沿波传播路径进行反投影，修改波传播路径上的背景参数扰动。前者对应一般的叠前深度偏移，包括 RTM；后者对应一般的旅行时层析，包括沿路径投影的波形层析。前者估计速度场的高波数成分；后者估计速度场的低波数成分。后者是前者的基础，必须很好地解决后者，前者才能得到较好的解决，根本原因是低波数速度成分与低频波场和旅行时之间存在更为线性化的关系，反演低波数成分相对容易进行。

反射波的残差反投影的这两种形式解释了叠前深度偏移和波形层析成像的本质，也说明了它们的差异与联系。我们认为当残差反投影沿着等时面进行时，就是在做叠前深度偏移；当残差反投影沿着波路径进行时就是在做波形层析。

为了沿路径反传播残差，首先需要在数据中引入波传播的方向，首先把地震数据修改成 $u_s\left(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s,\omega\right)$，该数据体称为立体地震数据，它是局部平面波域表达的叠前地震数据体。我们也称该数据提为特征波场数据。此时，（15）式被描述为

$$\begin{aligned}
&u_s\left(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s;\omega\right)\\
&=\omega^2\int d\mathbf{x}G_0\left(\boldsymbol{\xi},\mathbf{p}_r;\mathbf{x};\omega\right)\\
&\frac{a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}u_0\left(\mathbf{x};\boldsymbol{\eta},\mathbf{p}_s;\omega\right)\\
&=\int d\mathbf{x}K\left(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s;\mathbf{x};\omega;v_0\left(\mathbf{x}\right)\right)\frac{a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}\\
&=\int d\mathbf{x}K\left(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s;\mathbf{x};\omega;v_0\left(\mathbf{x}\right)\right)f\left(\mathbf{x}\right)
\end{aligned}\tag{64}$$

其中，$f\left(\mathbf{x}\right)=a\left(\mathbf{x}\right)\Big/v_0^2\left(\mathbf{x}\right)$ 为要反演的参数。

相应地，还要构建沿波路径进行投影和反投影的波传播算子。这是空间局部波传播算子被提出来研究的主要原因。

4.5 最小二乘反射波叠前偏移成像的实现

前已述及，反射波波动方程叠前偏移成像与波形层析成像从理论本质上是相同的，都是要估计散射势中的速度扰动量，用速度的高波数扰动来代表地下反射界面或绕射体的图像。

但是，二者的差异也是非常明显的。常规意义下，我们认为偏移成像是用反传算子外推到成像点（或绕射点/反射点）上的波场来定位绕射点/反射点出现的位置，并不估计反射系数或 $a\left(\mathbf{x}\right)$。Kirchhoff 积分偏移成像代表一类绕射叠加偏移方法，更清楚地反映了很多时候我们根本没有引入散射势的概念，同样进行叠前深度偏移。因此，像 Kirchhoff 积分这一类的绕射叠加叠前深度偏移和基于 Born 近似的叠前深度偏移甚至逆时深度偏移需要在理论上有一个清楚的界定。但是，到目前为止，叠前深度偏移本身及它和反射波波形层析成像的差异都没有清楚地分析说明。在后续的研究报告中，我们会深入地探讨这方面的问题。

下面我们具体讨论基于（15）式的最小二乘反射波叠前深度偏移成像。基于（15）式的反射波层析成像，在本论文集中冯波博士研究生会专文讨论。

为估计散射势，定义如下误差泛函，

$$
\begin{aligned}
J &= \frac{1}{2}\left\| u_s^{cal}(\xi,\eta,\omega) - u_s^{obs}(\xi,\eta,\omega) \right\|_2^2 \\
&= \frac{1}{2}\sum_{\omega}\sum_{\eta}\sum_{\xi}\left(u_s^{cal}(\xi,\eta,\omega) - u_s^{obs}(\xi,\eta,\omega) \right)^2
\end{aligned}
\tag{65}
$$

误差泛函能量最小化时，认为估计的散射势是所求的参数估计结果。其中 $u_s^{cal}(\xi,\eta,\omega)$ 有（15）式定义。

目标泛函的梯度是构造最速下降法的核心。该梯度定义为，

$$
\frac{\partial J}{\partial \mathbf{f}(\mathbf{x})} = \sum_{\omega}\omega^2 \sum_{\eta}\sum_{\xi} \frac{S(\omega)G_S(\mathbf{x},\eta,\omega)G_R(\xi,\mathbf{x},\omega)}{\left[u_s^{cal}(\xi,\eta,\omega) - u_s^{obs}(\xi,\eta,\omega) \right]^H}
\tag{66}
$$

其中 $S(\omega)G(\mathbf{x},\boldsymbol{\eta},\omega)$ 代表震源点 $\boldsymbol{\eta}$ 传到地下任一点 $\mathbf{x}$ 的场，$G_R(\xi,\mathbf{x},\omega)\left[u_s^{cal}(\xi,\eta,\omega) - u_s^{obs}(\xi,\eta,\omega) \right]^H$ 代表误差波场的反传波。

利用（66）定义的梯度可以进行最速下降法的最小二乘偏移成像的具体计算，

$$
\begin{aligned}
\mathbf{f}^{(k+1)}(\mathbf{x}) &= \mathbf{f}^{(k)}(\mathbf{x}) + \mu^{(k)}\delta\mathbf{f}^{(k)}(\mathbf{x}) \\
&= \mathbf{f}^{(k)}(\mathbf{x}) + \mu^{(k)}\frac{\partial J}{\partial \mathbf{f}(\mathbf{x})}
\end{aligned}
\tag{67}
$$

为了加速收敛，可以用 Hessian 矩阵的逆对梯度方向进行预条件处理，Hessian 矩阵的定义为，

$$
\begin{aligned}
&\mathbf{H}(\mathbf{x},\mathbf{y}) \\
&= \sum_{\omega}\omega^4 \sum_{\eta} S(\omega)G_S(\mathbf{x},\eta,\omega)G_S^H(\mathbf{y},\eta,\omega) \\
&\sum_{\xi} G_R(\xi,\mathbf{x},\omega)G_R^H(\xi,\mathbf{y},\omega)
\end{aligned}
\tag{68}
$$

该 Hessian 矩阵是由背景介质中的 Green 函数规定的，其中并不包括高阶反射/绕射效应，即不包括非线性部分。双向波计算 Green 函数可以包含这部分的作用，前提是速度场中包含高波数成分。

由 Hessian 矩阵的逆对梯度方向进行预条件处理后，最速下降法的迭代格式变为，

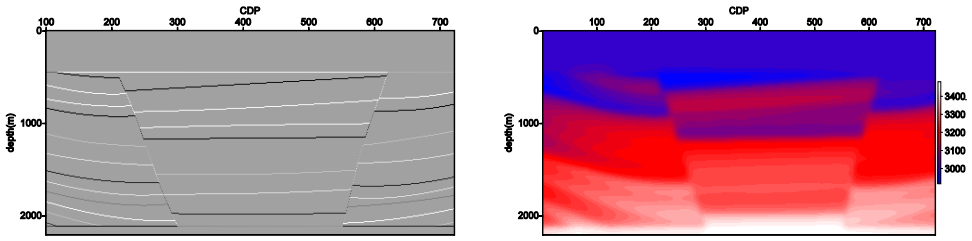
$$
\begin{aligned}
&\mathbf{f}^{(k+1)}(\mathbf{x}) \\
&= \mathbf{f}^{(k)}(\mathbf{x}) + \mu^{(k)}\delta\mathbf{f}^{(k)}(\mathbf{x}) \\
&= \mathbf{f}^{(k)}(\mathbf{x}) + \mu^{(k)}\mathbf{H}^{-1}\frac{\partial J}{\partial \mathbf{f}(\mathbf{x})}
\end{aligned}
\tag{69}
$$

由于 Hessian 规模较大，通常仅采用对角 Hessian 进行预条件处理。

迭代步长 a 一般可以采用试探的方式进行确定，一般采用试探加抛物拟合来选取最优下降步长来保证误差的下降。

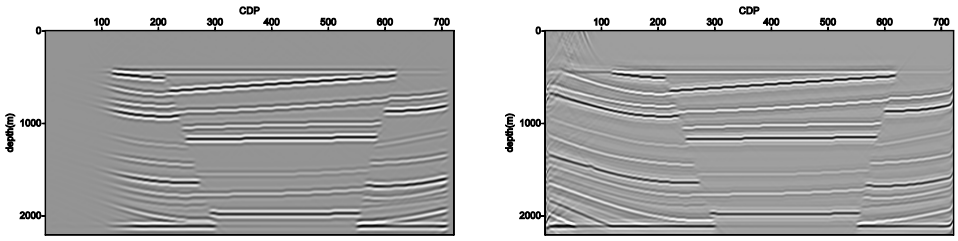
迭代的每一步需要在更新后对模型进行一次反偏移，然后将反偏移得到的正演数据从原始数据中减去，这样来判断误差是否达到要求，然后判断迭代是否终止。

图 3 为背景速度模型和反射系数模型。图 4 为常规偏移结果和最小二乘偏移结果。由图可以看出，最小二乘偏移结果在观测系统不充分的情况下也能较好地恢复反射系数模型，成像分辨率明显提升。由此可以看出，最小二乘偏移具有弥补观测系统不足对成像结果造成的畸变和提高成像分辨率的能力。



(a) (b)

图 3 反射系数模型（a）和背景速度模型（b）



(a) (b)

图 4 常规偏移结果（a）和最小二乘偏移迭代 10 次结果（b）

5 总结与讨论

地震波在复杂介质中传播的描述方法决定了后续的叠前深度偏移成像方法和层析反演方法。著名的描述方法是 Lippmann-Schwinger 方程以及基于此的散射场表达方法。Born 线性近似及对应的逆散射成像在勘探地震学及相关的学科中，譬如无损探伤成像、医学成像等，都有广泛的应用。

但是，勘探地震学中的逆散射成像不同于其他学科的逆散射成像，重大区别在于油气地震勘探中遇到的波现象以反射波为主，此时的逆散射成像方法必须用入射波与散射波相关的方式把波场残差放在反射界面上或绕射点上。如果用散射理论进行层析成像，需要沿波路径进行残差的反投影。

本文重点讨论了散射场的表达，以及基于此表达的最小二乘叠前深度偏移方法。指出这就是反射波的逆散射成像的一种实现方式。

本文还说明了一般意义下的叠前深度偏移和反射波的逆散射成像是有明显区别的。但目前二者都在使用，这对地震解释是有重要影响的，需要进一步的专文分析。

参考文献

[1] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, 26, 99-108
[2] Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, 52, 931-942
[3] Bleistein, N., Cohen, J. K., and Stockwell, J. W., Jr., 2000, Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion
[4] Bleistein, N., Cohen, J. K., and Hagin, F., 1985, Two and one-half dimensional Born inversion with an arbitrary reference: Geophysics, 52, 26-36
[5] Clayton, R. and Stolt, R., 1981, A Born-WKBJ inversion method for acoustic reflection data: Geophysics, 46, 1559-1567
[6] Cohen, J. K., Hagin, F. and Bleistein, N., 1986, Three dimensional Born inversion with an arbitrary reference: Geophysics, 52, 26-36
[7] Devaney, A. J., 2012, Mathematical Foundations of Imaging, Tomography and Wavefield Inversion: Cambridge
[8] Lambare, G., Operto, S., Podvin, P. and Thierry, P., 2003, 3D ray+Born migration/inversion, Part I: Theory: Geophysics, 68, 1348-1356

[9] Raz, S., 1981, Three-Dimensional velocity profile inversion from finite-offset scattering data: Geophysics, 50, 110-126
[10] Shuey, R., 1985, A simplification of the Zoeppritz equation: Geophysics, 50, 609-614
[11] Stolt, R., H., and Weglein, A. B., 1985, Imaging and Inversion of seismic data; Geophysics, 50, 2458-2472
[12] Stolt, R., H., and Weglein, A. B., 2012, Seismic Imaging and Inversion: Cambridge
[13] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, 49, 1259-1266
[14] Wapenaar, C.P.A., 1996, Inversion versus migration: A new perspective to an old discussion: Geophysics, 61, 804-814
[15] Weglein, A. B., Violette, P. B., and Keho, T. H., 1986, Using multiparameter Born theory to obtain certain exact multiparameter inversion goals: Geophysics, 51, 1069-1074

附录

A1. 标量散射理论的另一种表达

在介质速度场 $v(\mathbf{x})$ 中传播的标量场对应的齐次 Helmholtz 方程形式为，

$$\mathcal{L}_s(\mathbf{x},\omega)u(\mathbf{x},\omega)=\left(\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}\right)u(\mathbf{x},\omega)=0 \tag{A1.1}$$

引入背景介质速度场 $v_0(\mathbf{x})$，上式可以重写为，

$$\begin{aligned}&\mathcal{L}_s(\mathbf{x},\omega)u(\mathbf{x},\omega)\\&=\left(\nabla^2+\omega^2\left(\begin{array}{l}\dfrac{1}{v_0^2(\mathbf{x})}+\\ \dfrac{1}{v_0^2(\mathbf{x})}\left(\dfrac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}-1\right)\end{array}\right)\right)u(\mathbf{x},\omega)\\&=\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)u(\mathbf{x},\omega)+\frac{\omega^2 a(\mathbf{x})}{v_0^2(\mathbf{x})}u(\mathbf{x},\omega)=0\end{aligned} \tag{A1.2}$$

上式变为，

$$\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)u(\mathbf{x},\omega)=-\frac{\omega^2 a(\mathbf{x})}{v_0^2(\mathbf{x})}u(\mathbf{x},\omega) \tag{A1.3}$$

一般地，把上式中的 $\dfrac{\omega^2 a(\mathbf{x})}{v_0^2(\mathbf{x})}$ 称为散射势。实质上，上式的右端项定义了新的非齐次 Helmholtz 方程的（二次）震源项，这是一个体积源。

物理上，可以把在 $v(\mathbf{x})$ 中传播的波 $u(\mathbf{x},\omega)$ 分解成入射波场 $u_0(\mathbf{x},\omega)$（它在背景介质中传播），和（A1.3）中定义的散射势引起的散射场 $u_s(\mathbf{x},\omega)$ 的叠加。如图 A1 所示。

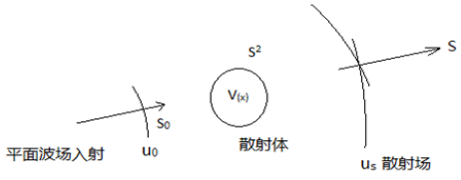


图 A1 散射波传播关系

背景场中传播的波场，假设满足齐次 Helmholtz 方程

$$\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)u_0(\mathbf{x},\omega)=0 \tag{A1.4}$$

（A1.4）代入（A1.3）可以得到散射势引起的散射场的表达式，

$$\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)u_s(\mathbf{x},\omega)=-\frac{\omega^2 a(\mathbf{x})}{v_0^2(\mathbf{x})}u(\mathbf{x},\omega) \tag{A1.5}$$

引入 Green 函数，满足方程

$$\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)G(\mathbf{x}-\mathbf{x}',\omega)=-\delta(\mathbf{x}-\mathbf{x}') \tag{A1.6}$$

用 $G(\mathbf{x}-\mathbf{x}',\omega)$ 乘（A1.5）式；$u_s(\mathbf{x},\omega)$ 乘（A1.6）式，结果相减，得到，

$$\begin{aligned}&u_s(\mathbf{x},\omega)\nabla^2G(\mathbf{x}-\mathbf{x}',\omega)-G(\mathbf{x}-\mathbf{x}',\omega)\nabla^2u_s(\mathbf{x},\omega)\\&=\frac{\omega^2 a(\mathbf{x})}{v_0^2(\mathbf{x})}u(\mathbf{x},\omega)G(\mathbf{x}-\mathbf{x}',\omega)-u_s(\mathbf{x},\omega)\delta(\mathbf{x}-\mathbf{x}')\end{aligned} \tag{A1.7}$$

如图所示，Green 函数和入射散射场的含义可以解释如下。利用 Green 定理，把左端项的体积分转化为面积分，得到，

$$u_s(\mathbf{x},\omega)=\int_V\frac{\omega^2 a(\mathbf{x}')}{v_0^2(\mathbf{x}')}u(\mathbf{x}',\omega)G(\mathbf{x}-\mathbf{x}',\omega)d\mathbf{x}+\oint_\Omega\left[u_s(\mathbf{x},\omega)\frac{\partial G(\mathbf{x}-\mathbf{x}',\omega)}{\partial n}-G(\mathbf{x}-\mathbf{x}',\omega)\frac{\partial u_s(\mathbf{x},\omega)}{\partial n}\right] \tag{A1.8}$$

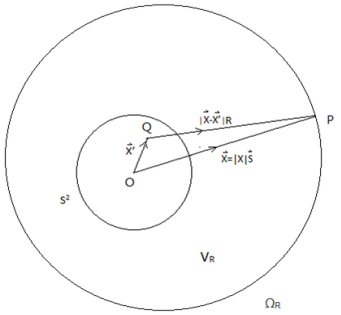


图 A2 空间局部范围内的散射体的散射场关系

物理上讲，散射势定义的体积源是空间有限的，当计算距离该体积源较远的场时，（A1.8）中的面积分项可以认为没有贡献。因为无论是 Green 函数还是散射场都是球面波场，远场情况下，忽略面积分符合索末菲辐射条件。

因此，散射势引起的散射场为

$$u_s(\mathbf{x},\omega)=\int_V\frac{\omega^2 a(\mathbf{x}')}{v_0^2(\mathbf{x}')}u(\mathbf{x}',\omega)G(\mathbf{x}-\mathbf{x}',\omega)d\mathbf{x} \tag{A1.9}$$

同样地，引入 Born 近似后，（A1.9）式变为，

$$u_s(\mathbf{x},\omega)=\int_V\frac{\omega^2 a(\mathbf{x}')}{v_0^2(\mathbf{x}')}u_0(\mathbf{x}',\omega)G(\mathbf{x}-\mathbf{x}',\omega)d\mathbf{x} \tag{A1.10}$$

仍有如下形式的总场表达式，

$$\begin{aligned} u(\mathbf{x},\omega) &= u_0(\mathbf{x},\omega) + u_s(\mathbf{x},\omega) \\ &= u_0(\mathbf{x},\omega) + \int_V \frac{\omega^2 a(\mathbf{x}')}{v_0^2(\mathbf{x}')} u_0(\mathbf{x}',\omega) G(\mathbf{x}-\mathbf{x}',\omega) d\mathbf{x} \end{aligned} \tag{A1.11}$$

这样的推导过程更清楚地体现了散射势所引起的散射场的物理本质。

地震波反演成像中的现代图像处理方法及应用

王华忠

①同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】勘探地震学中的地震波反演成像技术已经发展到以 FWI+RTM 为核心的阶段。精细的速度场（或参数场）和反射系数场是反演成像的目标。在对速度场（或参数场）和反射系数场的估计过程中及该过程完成后，压制参数场的图像中的噪音、提取其中的几何结构都是十分必要的步骤。尤其是几何结构信息对于参数反演和地质解释都是重要的有效信息。另外，在角度道集层析速度反演、立体层析速度反演，包括地震地质解释中，均需要从偏移成像结果中自动提取与反射界面和反射同相轴有关的信息。事实上，整个反射波地震勘探中的数据处理涉及到基于时空域地震波场的信号理论处理和波动理论处理和基于空间图像的图像处理。前者被广泛而深入地研究；后者正逐步得到重视。本文专门总结了用于地震波反演成像和处理过程中的图像处理方法。重点是抽象出图像处理的基本思想、模式、典型的方法技术，给出了一些在地震波反演成像和处理中的应用实例和对于地震反演图像处理中的建模方法的讨论。

【关键词】地震波反演成像，道集，剖面，反射同相轴，信号处理，图像处理，点处理，空间处理

1. 引言

油气勘探中地震波成像目前已经发展到以 FWI 和 RTM 为代表的反演成像阶段。即便如此，成像域和数据域的旅行时层析也远没有达到理想的实现效果（Sioni, et al. 2012；Hardy, 2013；Zdraveva, Hydal, and Woodward, 2013）。

地震数据处理和成像中对地震波场的认识主要是基于波动理论和信号理论的。基于波动理论，发展了一套速度估计和偏移成像的方法；基于信号理论，发展了一套信号分析、去噪音、反褶积的方法。实际上，基于波动理论和信号理论的对波场的处理应该是一体的，都应该统一在现代信号处理的理论框架下面，包括信号的表达和描述理论以及参数估计理论。譬如，最佳滤波、自适应滤波、谱估计与高阶谱估计、高维反褶积（FWI 可以认为是一种高维反褶积！）等都是信号表达和参数估计理论的具体应用。

但是，图像处理理论和方法在地震数据成像处理中似乎没有得到应有的尊重。

在速度分析/反演和速度建模过程中，无论是以 FWI 为代表的方法技术或是以成像道集剩余时差层析反演为代表的方法技术，把速度模型分解成反射界面所包围的块状层速度组合成的速度体都是对实际地下介质速度分布的一个较好的近似。这样的模型描述方式与地下介质组要由沉积层组成的客观现实（后期火山和构造运动会改变沉积层的几何形态）、地震数据由时距关系确定的反射子波形成的反射同相轴形成的实际是非常匹配的，也大大降低要反演参数的分布空间大小。因此，速度场中的几何结构，实际上也就是叠前深度偏移剖面上的反射同相轴的几何结构，对于地震波波形反演、旅行时层析反演和速度模型建立都是非常重要的信息。当然，这些几何结构信息也是地震地质解释的首要信息。如何从叠前深度偏移成像剖面（也包括道集）上凸显、自动识别这些几何结构是图像处理理论和方法要解决的问题。

叠前地震数据经过叠前深度偏移成像后就变成了一个空间图像，它从粗到细刻画了大的反射界面的几何结构和小的储层的几何结构。由于这个图像不一定能满足提取几何结构细节的要求，需要对它进行去模糊化处理（Debluring）和去噪音处理（Denoising）,更进一步地还要进行插值处理（Painting）。对于自动的模式识别，还要进行自动的图像分割处理（Segmentation）。上述四个图像处理的关键内容（Chan and Shen, 2005），在成像域的图像增强处理、反射界面自动识别、储层解释等方面有重要应用。通过图像处理得到的这些几何信息在地震参数反演中有非常重要的作用。

目前，地震数据采集数据量非常巨大，成像处理的效率必须大幅度提高才能满足处理周期的要求。自动的成像处理或尽可能自动的成像处理就成了必须要发展的技术。其中自动速度估计与建模是自动成像处理流程中的核心内容。要实现自动速度估计与建模，自动拾取方法，包括初至波到达时自动拾取、反射界面自动拾取、界面倾角的自动估计、成像道集中剩余深度差的拾取、自动的速度谱拾取（解释）是其中的关键技术内容。此时，先进的图像处理理论和方法必须受到特别的重视。统计估计和判别理论也会起到非常重要的作用，这些理论方法取代处理人员的判断作出决定。

图像处理是现代信号处理的一部分内容。但是，图像处理有其特殊性，它主要研究空间信号之间的相互关系。它与空间数据分析有天然的联系。

可以看出，地震波反演成像正逐步向更深、更广的方向发展。现代图像处理和空间数据分析理论和方法会得到充分的关注和深入的应用。

自 20 世纪 80 年代末，基于 Fourier 分析方法的数字图像处理有了深入发展。经典的数字信号处理往往基于信号的平稳性或者高斯随机过程假设，因此线性算法被认为是最优的。然而，实际的数字信号尤其是自然图像往往存在具有边缘、角点等具有重要视觉意义的非平稳几何结构，很难用简单的高斯过程进行有效刻画，所以数字图像处理逐渐倾向于非线性算法的研究。因此，数字图像处理的主流研究包括两个重要方面：一是数字图像的理解、表示与数学建模研究；二是基于图像模型的图像处理算法设计及性能研究。有效的图像建模对后续图像处理算法的设计和性能具有至关重要的作用。经过近 20 年的发展，研究者们已经从多个角度建立起图像的数学模型，主要包括：基于马尔科夫随机场(Markov random field; MRF)的统计图像建模方法，基于正则化函数空间的几何图像建模方法，基于调和分析的几何多尺度图像建模方法，基于多重分形的几何图像建模方法，以及基于人类视觉系统感知的图像建模方法。基于正则化函数空间的几何图像建模方法将导致基于变分泛函和几何偏微分方程(partial differential equation; PDE)的图像处理方法，统称为变分 PDE 方法。与其它方法相比，变分 PDE 方法在数学理论和数值计算上具有如下显著特点：可以直接处理图像中视觉重要的几何特征，如梯度、切线、曲率等；可以有效地模拟具有视觉意义的动态过程，如各向同性扩散、各向异性扩散等；可以运用成熟的数值计算方法或现代智能优化算法，如模拟退火、遗传算法等。

本文专门总结了用于地震波反演成像和处理过程中的图像处理方法。重点是抽象出图像处理的基本思想、图像处理的模式、典型的方法技术，给出了一些在地震波反演成像和处理中的应用实例和对于地震反演图像处理中的建模方法的讨论。

2. 现代图像处理基本思想与方法

2.1 图像的定义和图像处理

尽管图像处理是信号处理的一部分，但是图像处理的思想与方法本身还是有其特殊性。这主要表现在对信号和图像的认识各有侧重、图像处理所建立的模型也独具特点。当然，如果认为图像是一种比较特殊的信号，也是可以的。

对一个研究对象下一个严格的定义是首先要做的。但是对于图像下一个定义到目前为止还是不统一的，有很多图像处理的专著都不给图像下定义。下面罗列几种定义可以看出关于图像的定义的不统一。

图像是用各种观测系统以不同形式和手段观测客观世界而获得的，可以直接或间接作用于人眼并进而产生视知觉的实体（章毓晋编著，2007 年）。

Before we can define digital image processing, we must agree upon a definition for the word image. While most people have a notation of what an image is, a precise definition is elusive. Among the definitions of the word

in several of Webster's dictionaries are the following: " A representation, likeness, or imitation of an object or things, … a vivid or graphic description, … something introduced to represent something else. " Thus, in a general sense, an image is a representation of something else（Castleman, 1996）.

无论图像的定义多么含糊，二维矩阵表示的像素值构成了数字图像和三维数据体表示的像素值构成立体数字图像是一个比较严格的图像函数的说明。

数字图像处理的运算都是针对上述二维函数和三维函数的。基本的运算有单点运算和邻域运算两种。单点运算就是通过一个运算规则改变单点的像素值。邻域运算核心是设计一个滤波算子利用邻域数据对输出点数据进行滤波运算。

当然，图像处理的研究内容很多，模糊化处理（Debluring）、去噪音处理（Denoising）,插值处理（Painting）和图像分割处理（Segmentation）是四项基本内容。

尽管如此，单点运算和邻域运算是两项基础操作。

2.2 图像处理基础方法

2.2.1 单点处理基本思想

点处理的基本思想可以用下式表示

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{T}\left(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y})\right) \qquad (1.1)$$

很显然，$\boldsymbol{T}$ 为灰度变换函数。根据不同的目的可以给出各种形式的灰度变换函数。彩图的变换函数也可以用（1.1）式抽象。

2.2.2 单点处理方法

（1）Power-law 灰度变化

定义如下的变换函数

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{T}\left(\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y})\right)=\boldsymbol{c}(\boldsymbol{r})^{\gamma} \qquad (1.2)$$

其中，$\boldsymbol{c}$ 为常数。γ 是灰度调节参数。$\boldsymbol{r}$ 是输入灰度本身，也可以是灰度值的某种函数。 该函数是调整图像对比度的常用方法，连续变化 γ 值可以连续变化图像的对比度。

（2）直方图均衡方法

直方图均衡方法是自动调整图像对比关系的方法。基本思想是改变灰度的方差分布，使各种灰度成分的方差分布尽量均匀，不过于集中。

图 1 左上和左下是过于深和过于浅的两幅图，这是由于某种灰度成分过于集中导致的。左上图中深色灰度成分过多出现；左下浅色灰度成分过大出现。为了改变这种灰度分布过于集中的情况，利用直方图均衡方法对灰度成分分布进行调整。结果分别如右上和右下所示。

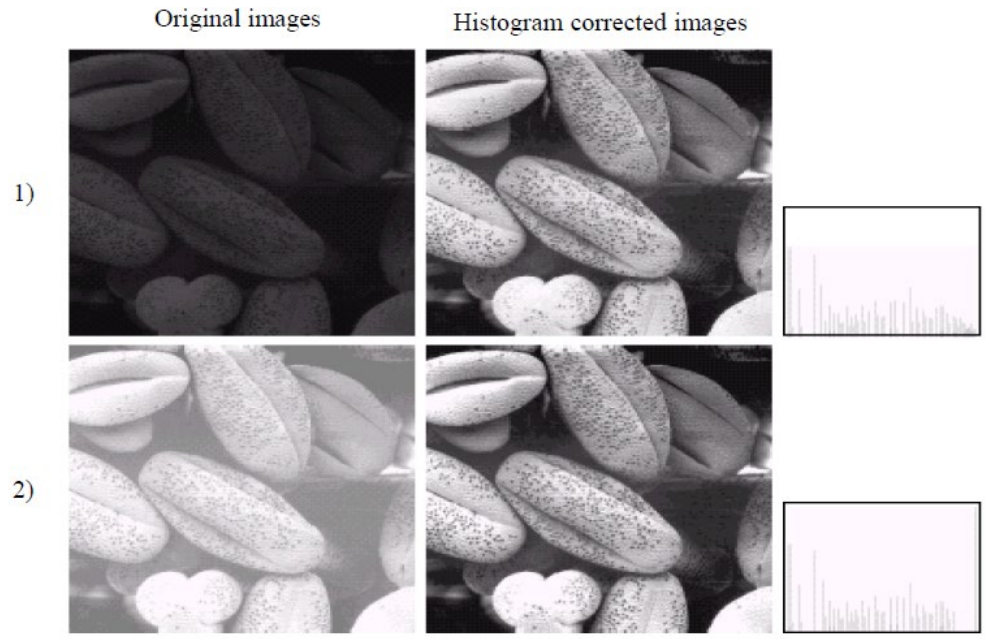


图 1 直方图均衡基本步骤得到的处理结果

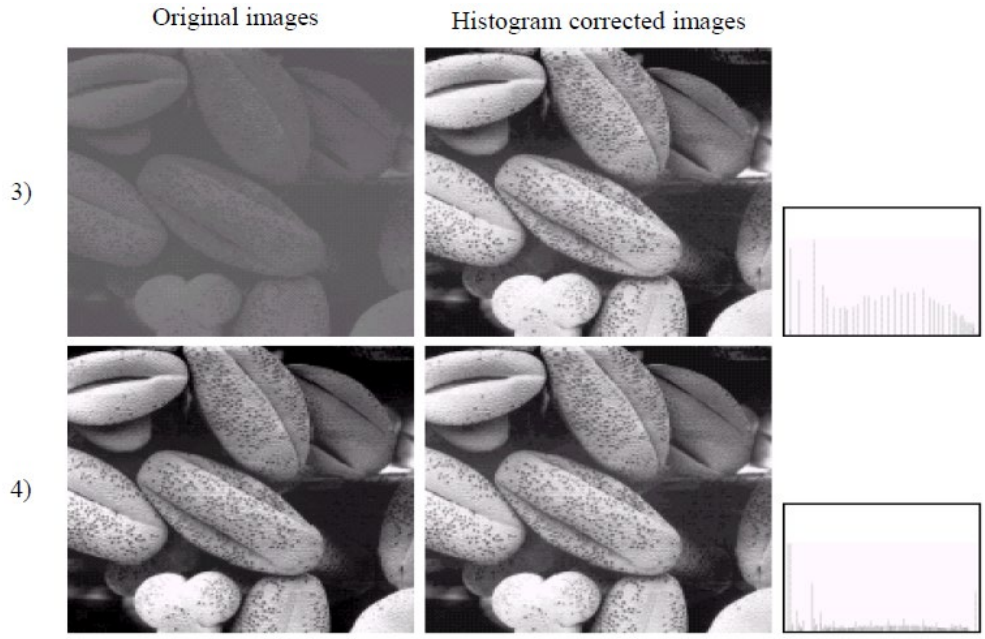


图 2 不同处理方法得到直方图均衡结果

直方图均衡方法的处理步骤：

①、统计一个图像中不同灰度成分的像素个数。

②、用总的像素对不同灰度成分的像素个数归一化，相当于计算不同灰度成分的像素出现的概率。

③、用最大灰度值乘以上述概率值，并取整。

④、把取整后的灰度值一对一地放到原像素点上，取代原像素值。

图 1 就是用上述四步处理后的结果。图 1 的右侧小图展示了处理后的像素的出现概率分布情况。

如果第三步不用最大灰度值乘概率值，而是用灰度均值、均方灰度值等去乘概率值，会得到不同的图像均衡结果。图 2 展示的就是用不用的方式处理得到的结果。这些处理目前的计算机可以实时处理，得到视觉上最好的结果。

2.2.3 邻域处理思想

对图像进行邻域处理是图像处理另一项基本操作。本质上，就是用如下的褶积型和相关型进行图像处理，

$$g(u,v)=\int_{s=-K/2}^{K/2}\int_{t=-K/2}^{K/2}h(s,t)f(u-s,v-t)dsdt \qquad (1.3a)$$

褶积型邻域处理公式。

$$g(u,v)=\int_{s=-K/2}^{K/2}\int_{t=-K/2}^{K/2}h(s,t)f(u+s,v+t)dsdt \qquad (1.3b)$$

相关型邻域处理公式。

很显然，上述公式是滑动滤波公式。它们的性质如何取决于积分核函数的设计。所有的滤波都表现为上述抽象的形式。它们可以用来 Denoising 处理，也可以进行 Debluring 处理，关键依然是积分核函数的设计。一般地，可以认为 Denoising 处理对应一个平滑褶积算子，Debluring 处理对应一个粗糙褶积算子。

2.2.4 邻域处理方法—各种滤波器类型

邻域处理方法的关键在于“滤波“算子的设计。滤波算子可以分为：低通的、高通的、和带通的；又可以分为线性的和非线性的。

从具体操作的角度，它们都是加权算子。比较典型的加权算子有：线性的均值滤波算子、高斯加权滤波算子（权系数都为正）、差分加权算子（权系数有正、负）；非线性的中值加权算子。

2.2.4.1 平滑和相关滤波器

加权系数都是正值和基本运算是相关运算是这类滤波器的特点。另一个特点是它们都起着平滑图像的作用，去除不相关的噪音。

2.2.4.1.1 均值滤波

$$g(m,n)=\frac{1}{Nu\cdot Nv}\sum_{i=-Nu/2}^{Nu/2}\sum_{j=-Nv/2}^{Nv/2}f(m-i,n-j) \qquad (1.4)$$

其中， $\boldsymbol{Nu}$ 和 $\boldsymbol{Nv}$ 为窗长度，以样点数表示。

2.2.4.1.2 高斯加权算子

$$\boldsymbol{h(u,v)}=\frac{1}{2\pi\sigma^2}\boldsymbol{e}^{-\frac{u^2+v^2}{\sigma^2}} \qquad (1.5)$$

其中，σ 为高斯窗长度参数。

2.2.4.1.3 中值滤波算子

$$med\{I,Z\}(p)=\underset{q\ \ sup\,p(Z+p)}{meidan}\{I(q)\} \tag{1.6}$$

按（1.6）式进行的中值滤波的基本步骤为：

①、I 是单调图像，即 1D 数组。

②、Z 定义为要处理的一个点 p 的邻域范围。形状可以根据处理的问题进行定义。

③、在该点邻域内，取 n 个像素点。

④、按值的大小排序成 I。或列表 $L(j), j=1,2,\cdots,n$

⑤、把列表中的中值 $L(m), m=\left[\frac{n}{2}\right]+1$ 输出作为该点 p 的中值滤波结果。其中，$\left[\frac{n}{2}\right]$ 代表取整。

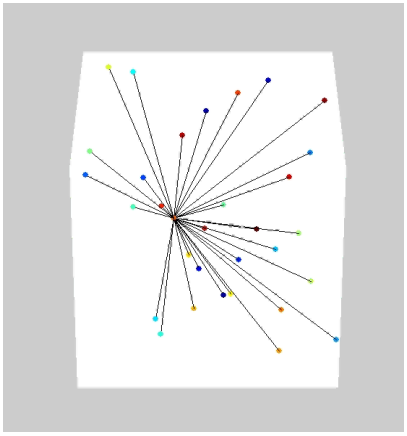


图 3 矢量中值滤波算子示意图

中值滤波还有矢量中值滤波、加权矢量中值滤波等各种变种。矢量中值滤波定义为，

$$v_m=\underset{k\neq j}{arg\,min}\left\{\left\|v_k-v_j\right\|\,v_k,v_j\in S\right\} \tag{1.7}$$

其中，S 是一个矢量集，v_m 是其中的中值矢量。如图 3 所示。考虑图像中几何结构特的中值滤波器的构造是必然的。

2.2.4.2 微分和褶积滤波器

加权系数有正有负，基本运算是褶积运算是这类滤波器的特点。另一个特点是它们都起着凸显图像不连续性的作用，刻画图像的几何特征，把光滑的背景成分压制掉。

2.2.4.2.1 一阶微分算子滤波器

定义为

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \approx \frac{\boldsymbol{f}\left(\boldsymbol{x}_{i+1}, \boldsymbol{y}\right)-\boldsymbol{f}\left(\boldsymbol{x}_{i-1}, \boldsymbol{y}\right)}{2 \Delta \boldsymbol{x}}=\nabla_{x} * \boldsymbol{f} \tag{1.8a}$$

其中，褶积算子定义为，

$$\nabla_{x}[\boldsymbol{u}, \boldsymbol{v}]=\frac{1}{2 \Delta \boldsymbol{x}}\begin{vmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{vmatrix} \tag{1.8b}$$

该滤波器对噪音很敏感，它凸显图像的纹理结构。改进的滤波器有 Sobol 差分算子，

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \approx \frac{1}{8 \Delta \boldsymbol{x}}\begin{bmatrix} \left(\boldsymbol{f}\left(\boldsymbol{x}_{i+1}, \boldsymbol{y}_{j-1}\right)-\boldsymbol{f}\left(\boldsymbol{x}_{i-1}, \boldsymbol{y}_{j-1}\right)\right) \\ +2\left(\boldsymbol{f}\left(\boldsymbol{x}_{i+1}, \boldsymbol{y}_{j}\right)-\boldsymbol{f}\left(\boldsymbol{x}_{i-1}, \boldsymbol{y}_{j}\right)\right) \\ +\left(\boldsymbol{f}\left(\boldsymbol{x}_{i+1}, \boldsymbol{y}_{j+1}\right)-\boldsymbol{f}\left(\boldsymbol{x}_{i-1}, \boldsymbol{y}_{j+1}\right)\right) \end{bmatrix} \tag{1.9a}$$

和

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \approx \frac{1}{8 \Delta \boldsymbol{y}}\begin{bmatrix} \left(\boldsymbol{f}\left(\boldsymbol{x}_{i+1}, \boldsymbol{y}_{j+1}\right)-\boldsymbol{f}\left(\boldsymbol{x}_{i+1}, \boldsymbol{y}_{j-1}\right)\right) \\ +2\left(\boldsymbol{f}\left(\boldsymbol{x}_{i}, \boldsymbol{y}_{j+1}\right)-\boldsymbol{f}\left(\boldsymbol{x}_{i}, \boldsymbol{y}_{j-1}\right)\right) \\ +\left(\boldsymbol{f}\left(\boldsymbol{x}_{i-1}, \boldsymbol{y}_{j+1}\right)-\boldsymbol{f}\left(\boldsymbol{x}_{i-1}, \boldsymbol{y}_{j-1}\right)\right) \end{bmatrix} \tag{1.9b}$$

2.2.4.2.2 梯度滤波器

定义为

$$\nabla \boldsymbol{f}=\begin{bmatrix} \frac{\partial}{\partial \boldsymbol{x}} \\ \frac{\partial}{\partial \boldsymbol{y}} \end{bmatrix} \cdot \boldsymbol{f} \approx \begin{bmatrix} \nabla_{x} * \boldsymbol{f} \\ \nabla_{y} * \boldsymbol{f} \end{bmatrix} 。 \tag{1.10}$$

而且有

$$|\nabla \boldsymbol{f}|=\sqrt{\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)^{2}+\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\right)^{2}} \approx \sqrt{\left(\nabla_{x} * \boldsymbol{f}\right)^{2}+\left(\nabla_{y} * \boldsymbol{f}\right)^{2}} \tag{1.11}$$

显然，$\nabla \boldsymbol{f}=\boldsymbol{i} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}+\boldsymbol{j} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}$ 刻画图像变化的方向。它代表与纹理（texture）垂直的方向。其中，矢量 $\nabla \boldsymbol{f}$ 与坐标系无关。而且有，

$$\nabla \boldsymbol{f}\cdot\boldsymbol{n}=\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{n}}\approx\frac{\boldsymbol{f}\left(\boldsymbol{p}+\boldsymbol{\varepsilon}\cdot\boldsymbol{n}\right)-\boldsymbol{f}\left(\boldsymbol{p}\right)}{\boldsymbol{\varepsilon}} \tag{1.12}$$

方向的估计可以用

$$\boldsymbol{\alpha}=\boldsymbol{tan}^{-1}\left(\frac{\partial \boldsymbol{f}/\partial \boldsymbol{y}}{\partial \boldsymbol{f}/\partial \boldsymbol{x}}\right)\approx\boldsymbol{tan}^{-1}\left(\frac{\nabla_y * \boldsymbol{f}}{\nabla_x * \boldsymbol{f}}\right) \tag{1.13}$$

2.2.4.2.3 二阶微分算子（Laplacian 算子）滤波器

定义为

$$\Delta \boldsymbol{f}=\frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{x}^2}+\frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{y}^2}=\begin{bmatrix}\frac{\partial}{\partial \boldsymbol{x}} & \frac{\partial}{\partial \boldsymbol{y}}\end{bmatrix}\begin{bmatrix}\frac{\partial}{\partial \boldsymbol{x}}\\ \frac{\partial}{\partial \boldsymbol{y}}\end{bmatrix}\cdot \boldsymbol{f}=\nabla\cdot\nabla \boldsymbol{f} \tag{1.14}$$

上式的含义是梯度的散度。具体计算式为，

$$\Delta \boldsymbol{f}=\frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{x}^2}+\frac{\partial^2 \boldsymbol{f}}{\partial \boldsymbol{y}^2}=\left(\Delta_x+\Delta_y\right)* \boldsymbol{f} \tag{1.15a}$$

其中，$\Delta_x=\frac{1}{2\Delta \boldsymbol{x}}\begin{bmatrix}0 & -1 & 0\\ 0 & 2 & 0\\ 0 & -1 & 0\end{bmatrix}$ 和 $\Delta_y=\frac{1}{2\Delta \boldsymbol{y}}\begin{bmatrix}0 & 0 & 0\\ -1 & 2 & -1\\ 0 & 0 & 0\end{bmatrix}$。一般地，$\Delta \boldsymbol{x}=\Delta \boldsymbol{y}=\boldsymbol{h}$。

显然，有 $\Delta_x+\Delta_y=\frac{1}{2\boldsymbol{h}}\begin{bmatrix}0 & -1 & 0\\ -1 & 4 & -1\\ 0 & -1 & 0\end{bmatrix}$。该褶积算子可以根据不同的情况取各种形式，譬如，$\begin{bmatrix}-1 & -1 & -1\\ -1 & 9 & -1\\ -1 & -1 & -1\end{bmatrix}$

或 $\begin{bmatrix}0 & -1 & 0\\ -1 & 5 & -1\\ 0 & -1 & 0\end{bmatrix}$ 或 $\begin{bmatrix}-2 & 1 & -2\\ 1 & 5 & 1\\ -2 & 1 & -2\end{bmatrix}$ 等等。

图 4 展示了灰度值、灰度一阶变化率和二阶变化率的曲线。可以看出二阶导数可以检出灰度值的变化最快的点，就是能进行边缘检测。

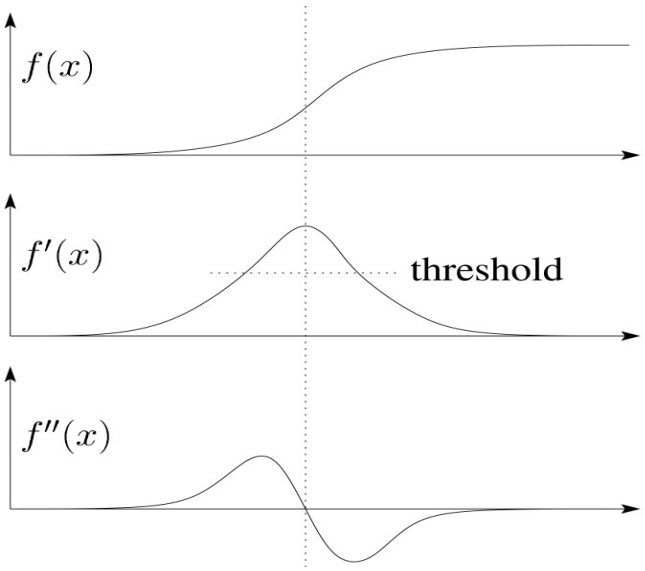


图 4 灰度值、灰度一阶变化率和二阶变化率的曲线

2.2.5 各种滤波器的数值结果

2.2.5.1 相关型滤波器---均值滤波、Gaussian 滤波和中值滤波方法

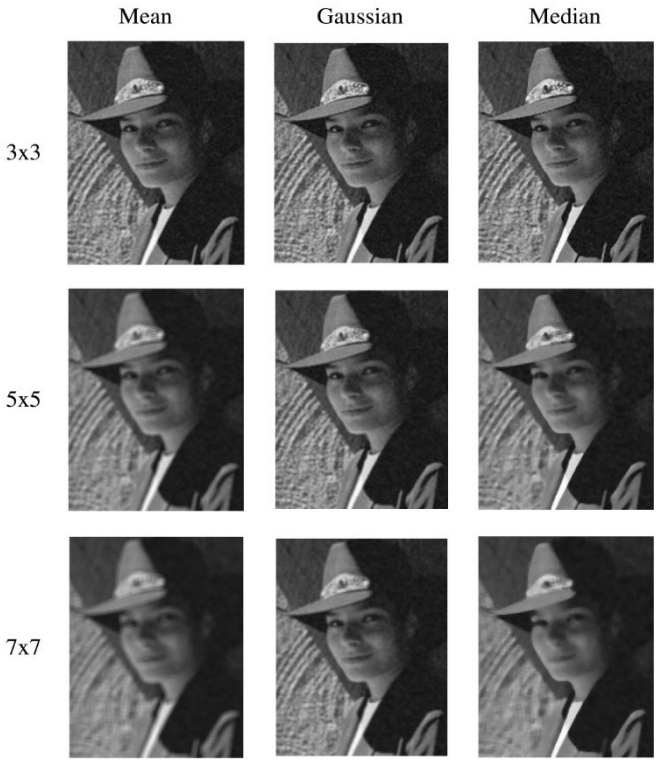


图 5 高斯噪音下三种滤波方法的对比

图 5 展示了 Gauss 噪音情况下，均值滤波、Gaussian 滤波和中值滤波方法对同一图像的滤波结果。其中 3×3、5×5 和 7×7 代表滤波窗的大小。对比可以看出，对于 Gauss 噪音，Gaussian 滤波器去噪效果最好。对于本图的噪音强度，小窗口的滤波效果要好于大窗口，大窗口滤波模糊效果太强。

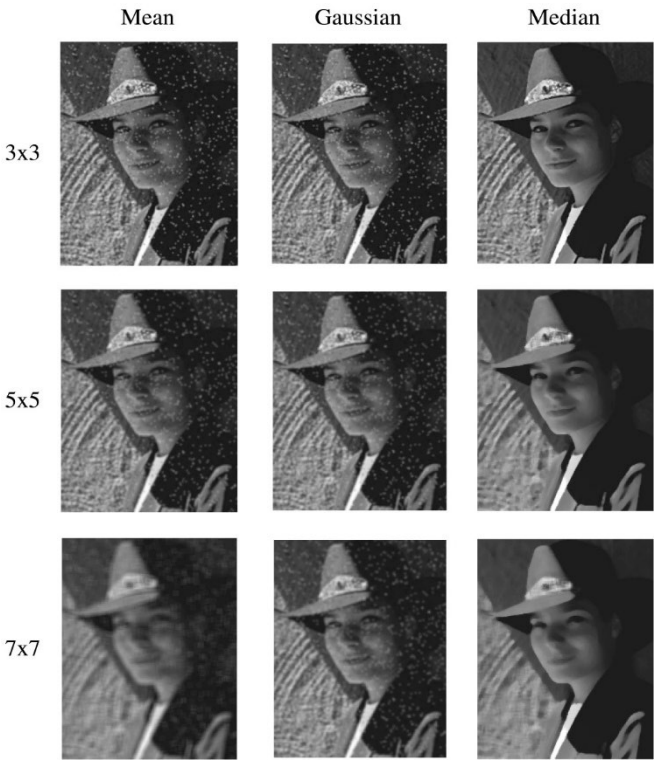


图 6 Salt and Papper 噪音下三种滤波方法的比较

图 6 展示了 Salt and Papper 噪音下，三种滤波方法的效果对比。此时，非线性的中值滤波方法的效果最好。关于窗口大小选择的结论是一样的，基本上宁小勿大的原则。

中值滤波和均值滤波在同时源采集的地震数据噪音压制中有特殊作用。因为同时源采集观测系统设计中，认为地提高不同源激发时间的随机性，目的就是加强伪解码、并重排序后噪音的随机性。这些强的不相干噪音，目前看来只能用非线性的中值滤波来消除，尽管中值滤波不是保幅的滤波方法。

因此，对于同时源数据分解，需要进一步发展振幅保真、效果更好的滤波方法消除伪解码、并重排序后强噪音。

图 7 展示了一阶差分算子滤波器作用于图像后的效果。很显然，沿不同方向的一阶差分算子凸显了图像中的像素值沿特定方向的变化。图 7a、图 7b 和图 7c 分别是原图、算子∇_x和∇_y作用后的结果。

图 8 中右图展示了梯度算子$\left[\nabla_x \quad \nabla_y\right]^T$作用于原图（左图）后的效果。很清楚地看出，梯度算子起到了图像边缘检测的作用。梯度本身就是反映图像灰度值的变化率，从此不难看出梯度算子应有的处理效果。

图 9 展示了 Laplace 滤波器作用于图像后的效果。很明显，它消除了背景中的光滑变化部分，凸显了其中的纹理。图 9b 和图 9c 是 Laplace 算子和 Sobel 算子作用图像后的对比，可以看出，Sobel 算子更能清楚地凸显纹理结构。

Laplace 算子滤波对 RTM 成像结果进行处理消除其中的背景变化，保留反射同相轴，是勘探地震中的典型应用。可以尝试把 Sobel 滤波算子作用于 RTM 成像结果上。



图 7 a 一阶差分滤波算子作用的原图

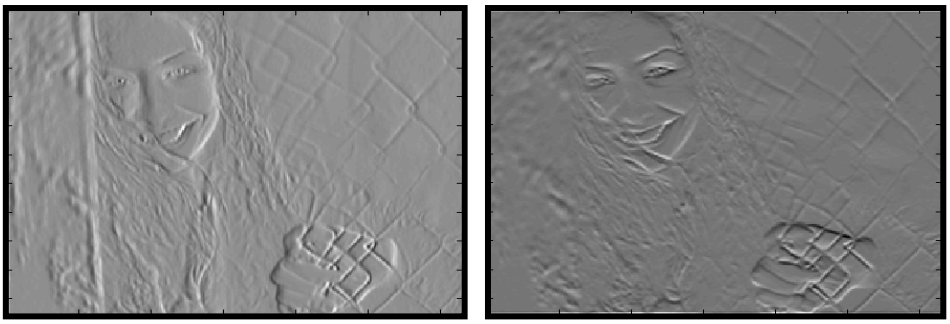


图 7 b 一阶差分滤波算子 $\nabla_x f$ 后的结果；图 7 c 一阶差分滤波算子 $\nabla_y f$ 后的结果



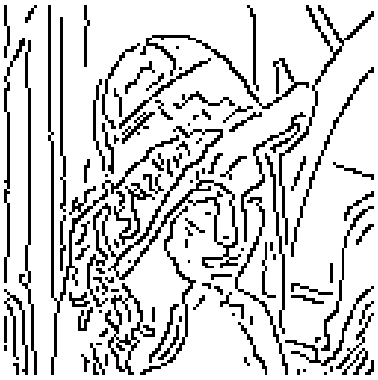
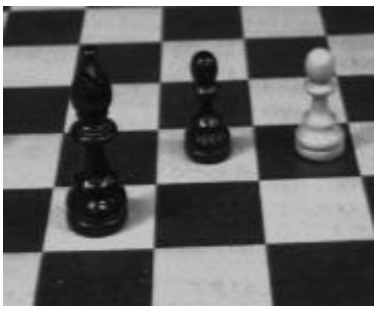
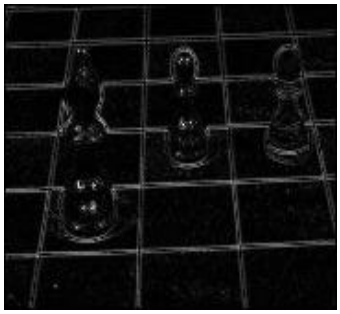


图 8 梯度算子的作用



（a）原图像



（b）Laplace 算子作用后的图像



（c） Sobel 滤波算子作用后的图像

图 9 二阶滤波算子的滤波效果

3. 基于结构张量的图像建模方法

3.1 基于结构张量的各向异性 PDE

基于 PDE 的图像分析与处理思想最早出现在 Witkin（1983）发表的文章“尺度空间滤波”，指出信号与具有不同尺度(方差)高斯函数卷积等价于以信号为初值的热传导方程。假设 $\boldsymbol{u}_0(\boldsymbol{x})$ 为初始信号，与 $\boldsymbol{u}_0(\boldsymbol{x})$ 相关的“尺度空间”分析归于求解下面的方程

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{t}} = \Delta \boldsymbol{u}(\boldsymbol{x},\boldsymbol{t}), \qquad \boldsymbol{u}(\boldsymbol{x},0) = \boldsymbol{u}_0(\boldsymbol{x}) \tag{2.1}$$

然而，高斯函数卷积作为低通滤波对应的是线性运算，或热传导方程本质上对应的是各向同性的扩散过程，图像中的边缘、角点等重要几何结构将被过分光滑，不利于后续的图像处理操作，如边缘检测、图像分割。事实上，数字图像的非高斯过程要求图像处理算法应该具有非线性和各向异性的特点，使得滤噪的同时能够保护图像中重要的几何信息。基于 PDE 的非线性图像处理算法始于 Perona-Malik（1987）提出各向异性扩散方程，通过在热传导方程中引入取值在 0 和 1 之间的单调下降“边缘停止”函数 $\boldsymbol{g}$，实现边缘保持的图像滤波

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{t}} = \nabla\left(\boldsymbol{g}\left(\left|\nabla \boldsymbol{u}\right|\right)\right)\nabla \boldsymbol{u}, \qquad \boldsymbol{u}(\boldsymbol{x},0) = \boldsymbol{u}_0(\boldsymbol{x}) \tag{2.2}$$

虽然 Perona-Malik 方程能够一定程度上实现边缘保持的图像滤波，但仍存在严重缺陷，例如：初始图像 $\boldsymbol{u}_0(\boldsymbol{x})$ 被噪声严重污染时，会导致大量伪边缘结构的出现；Perona-Malik 方程本身是不适定问题，函数 $\boldsymbol{g}$ 的选取往往导致逆扩散造成解的不稳定性，事实上，Perona-Malik 方程对应一种平滑-增强型扩散机制（Aubert and Kornprobst, 2000）。尽管如此，Perona-Malik 方程开创了基于 PDE 的非线性图像处理的先河。之后的 10 多年时间里，基于 PDE 的多数图像处理研究都是基于 Perona-Malik 方程展开的，包括 Weickert（1994）最先提出的基于结构张量的散度型各向异性 PDE 和 Tschumperlé（2002）最先提出的面向应用的迹型各向异性 PDE。

首先引入结构张量的概念（Weickert, 1994）。结构张量定义为：

$$\boldsymbol{J}_{\rho}\left(\nabla \boldsymbol{u}_{\sigma}\right) = \boldsymbol{G}_{\rho} * \left(\nabla \boldsymbol{u}_{\sigma} \otimes \nabla \boldsymbol{u}_{\sigma}\right) = \left(\boldsymbol{J}_{m,n}\right)_{m=1,2;n=1,2},$$

其中 $\boldsymbol{u}_{\sigma}$ 为图像 $\boldsymbol{u}$ 经过高斯滤波的图像(方差 $\sigma \geq 0$)，减少噪声对求导运算的影响，从而增强边缘检测的噪声鲁棒性；对张量 $\nabla \boldsymbol{u}_{\sigma} \otimes \nabla \boldsymbol{u}_{\sigma}$ 进行高斯滤波(方差 $\rho \geq 0$)，以增强几何结构方向的估计鲁棒性。

矩阵 $\boldsymbol{J}_{\rho}$ 对称且半正定，存在正交单位特征向量，分别记为 $\boldsymbol{w}$ 和 $\boldsymbol{v} = \boldsymbol{w}^{\perp}$ 。其中，$\boldsymbol{w} = \left(2\boldsymbol{J}_{12}, \boldsymbol{J}_{22} - \boldsymbol{J}_{11} + \sqrt{\left(\boldsymbol{J}_{22} - \boldsymbol{J}_{11}\right)^2 + 4\boldsymbol{J}_{12}^2}\right)^{\perp}$，$\boldsymbol{w} = \boldsymbol{w} \big/ |\boldsymbol{w}|$ 指向几何结构的最大对比度方向，称为主向量；相应地，$\boldsymbol{v}$ 指向几何结构的最小对比度方向，看作是几何结构的方向。它们相应的特征值分别记为 $\boldsymbol{\mu}$ 和 $\boldsymbol{\mu}^{\perp}$，定义为 $\boldsymbol{\mu} = \left(\boldsymbol{J}_{11} + \boldsymbol{J}_{22} + \sqrt{\left(\boldsymbol{J}_{22} - \boldsymbol{J}_{11}\right)^2 + 4\boldsymbol{J}_{12}^2}\right)^{\perp}$ 和 $\boldsymbol{\mu}^{\perp} = \left(\boldsymbol{J}_{11} + \boldsymbol{J}_{22} - \sqrt{\left(\boldsymbol{J}_{22} - \boldsymbol{J}_{11}\right)^2 + 4\boldsymbol{J}_{12}^2}\right)^{\perp}$。这两个

值可以作为局部几何结构的属性描述子：在平滑区域，$\mu \approx \mu^{\perp} \approx 0$；在边缘区域，$\mu \gg \mu^{\perp} \approx 0$；在角形区域，$\mu \gg \mu^{\perp} \gg 0$。

Weickert 基于结构张量提出如下散度型各向异性 PDE

$$\frac{\partial \boldsymbol{u}}{\partial t} = \nabla\left(\boldsymbol{D}\left(\boldsymbol{J}_{\rho}\left(\nabla \boldsymbol{u}_{\sigma}\right)\right)\right)\cdot\nabla \boldsymbol{u}, \qquad \boldsymbol{u}(\boldsymbol{x},0) = \boldsymbol{u}_0(\boldsymbol{x}) \tag{2.3}$$

其中，$\boldsymbol{D}\left(\boldsymbol{J}_{\rho}\left(\nabla \boldsymbol{u}_{\sigma}\right)\right)$ 称为扩散张量，对应 2×2 的(半)正定矩阵。显然，式（2.3）是 Perona-Malik 方程（2.2）的推广，当 $\boldsymbol{D}=\boldsymbol{g}\cdot\boldsymbol{I}$ 时式（2.3）即退化为式（2.2），其中 $\boldsymbol{I}$ 是单位矩阵。一方面，Weickert 从理论上讨论了式（2.3）解的存在性、唯一性、极值性等问题；另一方面，Weickert 等人将式（2.3）广泛应用于图像去噪、指纹增强、光流场计算等多种图像处理问题中。

在图像处理的实际应用中，Weickert 将扩散张量定义为 $\boldsymbol{D}\left(\boldsymbol{J}_{\rho}\left(\nabla \boldsymbol{u}_{\sigma}\right)\right)=\lambda_1 \boldsymbol{w}\boldsymbol{w}^{\boldsymbol{T}} + \lambda_2 \boldsymbol{v}\boldsymbol{v}^{\boldsymbol{T}}$，根据特征向量 $\boldsymbol{w}$，$\boldsymbol{v}$ 确定 PDE 的主要扩散方向，根据 λ_1，λ_2 确定两个扩散方向上的扩散强度。因此，PDE 式（2.3）的滤波性能主要由扩散张量 $\boldsymbol{D}\left(\boldsymbol{J}_{\rho}\left(\nabla \boldsymbol{u}_{\sigma}\right)\right)$ 具体刻画。

王正明等人（2006）关于 PDE 式（2.3）的滤波性能给出了如下定理。定理：设 $\boldsymbol{u}_0$ 为原始图像，$\boldsymbol{D}\left(\boldsymbol{J}_{\rho}\left(\nabla \boldsymbol{u}_{\sigma}\right)\right)=\lambda_1 \boldsymbol{w}\boldsymbol{w}^{\boldsymbol{T}} + \lambda_2 \boldsymbol{v}\boldsymbol{v}^{\boldsymbol{T}}$ 为连续的 2×2 扩散矩阵，$\boldsymbol{w}$，$\boldsymbol{v}$ 为其规范正交的特征向量分别对应图像的梯度方向和边缘方向，λ_1，λ_2 为其相应的特征根，利用扩散平滑模型（3）对 $\boldsymbol{u}_0$ 进行光滑近似于分别以 λ_1，λ_2 的速度沿 $\boldsymbol{w}$，$\boldsymbol{v}$ 方向光滑。

Albert 等人（2002）将主要 PDE 分成三类：平滑型（smoothing）PDE，平滑-增强型（smoothing-enhancing）PDE，增强型（enhancing） PDE，并将 PDE 式（2.3）划分为平滑型 PDE。Tschumperlé 等人（2002，2005）研究讨论了 PDE 式（3）的滤波性能，指出扩散张量驱动的散度型 PDE 不能够完全刻画扩散张量所要体现的滤波性能，并且提出一种扩散张量驱动的可计算迹型 PDE

$$\frac{\partial \boldsymbol{u}}{\partial t} = \boldsymbol{Trace}\left(\boldsymbol{D}\left(\boldsymbol{J}_{\rho}\left(\nabla \boldsymbol{u}_{\sigma}\right)\right)\cdot \boldsymbol{H}\right), \qquad \boldsymbol{u}(\boldsymbol{x},0) = \boldsymbol{u}_0(\boldsymbol{x}) \tag{2.4}$$

其中，$\boldsymbol{H}$ 为 Hessian 矩阵。Tschumperlé 等人将 PDE 式（2.4）广泛应用于图像去噪、图像修补、图像放大等多种图像处理问题中。事实上，本文研究分析将指出，PDE 式（3）实质上对应的是平滑-增强型（smoothingenhancing）滤波机制，而 PDE 式（2.4）对应的才是平滑型(smoothing)滤波机制。利用 PDE 式（2.3）对图像处理时，PDE 不仅分别以 λ_1，λ_2 的速度沿 $\boldsymbol{w}$，$\boldsymbol{v}$ 方向进行平滑，同时还具有增强角点和强边缘结构的作用。

3.2 基于结构张量的变分泛函

基于正则化函数空间的几何图像建模方法将导致基于变分泛函和几何 PDE 的图像处理方法。与上节中的边缘保持 PDE 类似，若干研究者相应地提出如下具有边缘保持性能的变分泛函

$$\boldsymbol{E}_r(\boldsymbol{u}) = \int_{\Omega} \varphi\left(\|\boldsymbol{u}\|\right) d\boldsymbol{x} \tag{2.5}$$

其中，$\varphi(\cdot)$称为势函数，满足如下基本条件：$\varphi(x)\geq 0,\varphi(x)=\varphi(-x),\varphi'(x)\geq 0$，并且$\varphi'(x)/x$在$(0,\infty)$上是单调下降函数。例如，当$\varphi(x)=|x|$时，式（3）即为 TV(total variation)模型（Aly and Dubois, 2005; Malgouyres and Guichard, 2001）；当$\varphi(x)$取为 Huber 范数时，式（2.3）即为 Huber-MRF 模型（Schultz and Stevenson, 1994）。对应于变分泛函（2.5）的 PDE 为

$$\frac{\partial \boldsymbol{u}}{\partial t}=\nabla\cdot\left(\frac{\varphi'(\|\nabla \boldsymbol{u}\|)}{\|\nabla \boldsymbol{u}\|}\cdot\nabla \boldsymbol{u}\right) \tag{2.6}$$

Weickert（2002）首次提出如下基于结构张量的变分泛函用于运动估计

$$\boldsymbol{E}_r(\boldsymbol{u})=\int_\Omega \boldsymbol{Trace}\varphi\left(\varphi(\mu)\boldsymbol{w}\boldsymbol{w}^T+\varphi(\mu^\perp)\boldsymbol{v}\boldsymbol{v}^T\right)d\Omega \tag{2.7}$$

其中，$\varphi(\cdot)$取为上述势函数。对应于式（7）的 PDE 为

$$\frac{\partial \boldsymbol{u}}{\partial t}=\nabla\left(\left(\varphi(\mu)\boldsymbol{w}\boldsymbol{w}^T+\varphi(\mu^\perp)\boldsymbol{v}\boldsymbol{v}^T\right)\cdot\nabla \boldsymbol{u}\right) \tag{2.9}$$

Tschumperlé 等人（2005）基于结构张量提出更为一般的变分泛函

$$\boldsymbol{E}_r(\boldsymbol{u})=\int_\Omega \psi(\mu,\ \mu^\perp)d\Omega \tag{2.10}$$

其中，$\psi(\cdot)$通常采用势函数$\varphi(\cdot)$定义。当结构张量$\boldsymbol{J}_\rho(\nabla \boldsymbol{u}_\sigma)$中的$\rho$ 和σ 均取 0 时，式（9）即退化为式（5）。对应于式（9）的 PDE 为

$$\frac{\partial \boldsymbol{u}}{\partial t}=2\cdot\nabla\left(\left(\frac{\partial \psi}{\partial \mu}\boldsymbol{w}\boldsymbol{w}^T+\frac{\partial \psi}{\partial \mu^\perp}\boldsymbol{v}\boldsymbol{v}^T\right)\cdot\nabla \boldsymbol{u}\right) \tag{2.11}$$

事实上，比较 PDE 式（8）和 PDE 式（9）知，式（7）也是式（9）的一种特殊情形。

注意到，Tschumperlé 等人在提出变分泛函式（9）时没有给出$\psi(\cdot)$的选取条件。事实上，不同的$\psi(\cdot)$选取方式对于式（9）滤波性能将有不同的影响。

3.3 冲击滤波器及图像特征增加

在图像处理中，增强型（enhancing）PDE 主要用于增强图像从而提高图像的清晰度，对应着图像平滑（smoothing）的逆过程。Osher 等人（1992）首次提出一种称为冲击滤波器（shock filter）的 PDE 图像增强方法。最常用的冲击滤波器定义如下

$$\frac{\partial \boldsymbol{u}}{\partial t}=-\boldsymbol{sign}\left(\boldsymbol{D}^2\boldsymbol{u}(\mu,\eta)\right)\|\nabla \boldsymbol{u}\| \tag{2.12}$$

其中，$\eta=\nabla u/\|\nabla u\|$ 为图像的梯度方向。可见，冲击滤波器根据边缘检测项 $D^2u(\mu,\eta)$ 的符号确定冲击流的方向，根据图像的梯度模确定冲击流的强度。然而，PDE（11）存在两点重要不足：其一，增强边缘的同时增强噪声点；其二，只能增强边缘不能增强角形结构，因此实质上是边缘冲击滤波。针对第一点不足，研究者们通过耦合边缘冲击滤波和平滑型 PDE（Alvarez and Mazorra, 1994），实现同时增强边缘和抑制噪声的目的。

4. Level Set 方法

4.1 函数 f 的 Level Set 的定义

定义：函数 f：$R^n \to R$ 的在 Level c 上的 Level Set 有点集

$$S=\{x: f(x)=c\} \tag{3.1}$$

组成。

对于 f：$R^2 \to R$，在 Level c 上的 Level Set 有下图的形象表示

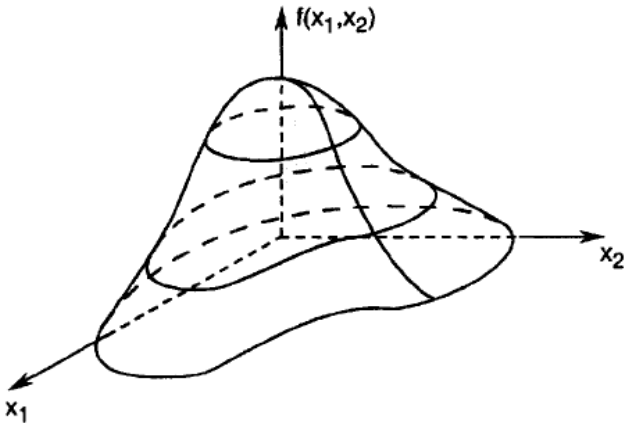


图 10 函数 f：$R^2 \to R$ 的 Level Set 的立体表示

3.2 Level Set 的几何特性

下面以函数 f：$R^2 \to R$ 为例来讨论 Level Set 的几何特性。我们称 x_0 在 Level Set 中的 level c 上是指 $f(x_0)=c$。

假设 Level Set 中有一曲线 γ，由一个可微的含参变量 t_0 的函数 $g(t_0)$ 表示：$g(t_0): R \to R^n$。假设 $g(t_0)=x_0$，$Dg(t_0)=v\neq 0$，则 v 在 x_0 处与曲线 γ 相切。

应用链式法则于 $h=f(g(t))$，在 t_0 处有，

$$\boldsymbol{h}'(t_0)=\boldsymbol{Df}\left(\boldsymbol{g}(t_0)\right)\boldsymbol{Dg}(t_0)=\boldsymbol{Df}\left(\boldsymbol{g}(t_0)\right)\boldsymbol{v} \qquad (3.2)$$

因曲线$\boldsymbol{\gamma}$是 Level Set 中的某一 Level 对应的曲线，则，$\boldsymbol{h}(t)=\boldsymbol{Df}\left(\boldsymbol{g}(t)\right)=\boldsymbol{k}$（常数）。因此，

$$\boldsymbol{h}'(t_0)=\boldsymbol{Df}\left(\boldsymbol{g}(t_0)\right)\boldsymbol{v}=\nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)^T\boldsymbol{v}=0 \qquad (3.3)$$

若函数$\boldsymbol{f}$：$\boldsymbol{R}^2\to\boldsymbol{R}$连续可微，存在如下定理。

定理：矢量$\nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)$在$\boldsymbol{x}_0$处正交于由$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{f}\left(\boldsymbol{x}_0\right)$定义的 Level Set 中的任意光滑曲线的切线。如图 11 所示

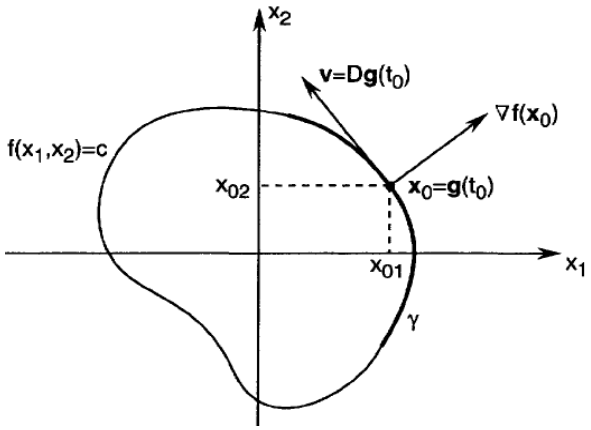


图 11 一个 Level Set 曲线的图示。

简言之，$\nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)$在$\boldsymbol{x}_0$处正交于 Level Set 函数$\boldsymbol{S}$。推理知，凡过$\boldsymbol{x}_0$点的切面（线）上的点$\boldsymbol{x}$与点$\boldsymbol{x}_0$形成的差矢量$\boldsymbol{\eta}=\boldsymbol{x}-\boldsymbol{x}_0$都与$\nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)$正交，即

$$\nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)\boldsymbol{\eta}=\nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)\left(\boldsymbol{x}-\boldsymbol{x}_0\right)=0,\quad \nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)\neq 0。 \qquad (3.4)$$

相对于 Level Set 函数而言，$\nabla\boldsymbol{f}\left(\boldsymbol{x}_0\right)$代表此处函数$\boldsymbol{f}$：$\boldsymbol{R}^2\to\boldsymbol{R}$变化最快的方向。

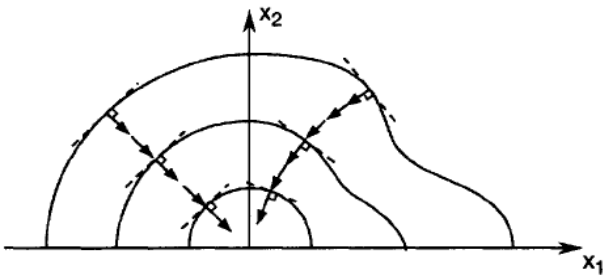


图 12 函数$\boldsymbol{f}$：$\boldsymbol{R}^2\to\boldsymbol{R}$的 Level Set 函数的平面展示。

关键是引入 Level Set 概念后，$\nabla \boldsymbol{f}\left(\boldsymbol{x}_0\right)$的几何意义更清楚。利用$\nabla \boldsymbol{f}\left(\boldsymbol{x}_0\right)$求极值的方法可以转化为利用 Level Set 求极值的问题。图 10 代表$\boldsymbol{f}\left(\boldsymbol{x}\right)$的立体图，图 11 代表$\boldsymbol{f}\left(\boldsymbol{x}\right)=\boldsymbol{c}$的某一个 Level Set。图 12 代表 Level Set 在平面上的投影。可见，Level Set 本身也是一个函数。图 12 中标示的最速下降方向与$\nabla \boldsymbol{f}\left(\boldsymbol{x}_0\right)\left(\boldsymbol{x}-\boldsymbol{x}_0\right)=0$是一致的。即点$\boldsymbol{x}_0$处的一个实值可微函数的最速上升方向正交于过此点的函数$\boldsymbol{f}\left(\boldsymbol{x}\right)$的 Level Set。

据此讨论，Level Set 概念可以用于泛函最优化求解。Lewis, Starr, and Vigh,（2012）,；Gazit, and Haber,（2012）； Zheglova, and Farquharson, （2012）以及 Guo, and de Hoop,（2013）讨论了 Level Set 方法在 FWI 中的应用。

4. 分析与讨论

图像处理是现代信号分析的一部分，它在勘探地震学中的应用越来越深入，这是一种不能不重视的现象。

地震波反演成像技术已经发展到以 FWI+RTM 为核心的阶段。但是，要得到符合地质意义的反演结果，在比较复杂的探区依然是十分困难的事情。其中，无法（自动地）得到反射界面的几何形态是关键因素。事实上，对任何客观实体的认识，它的几何外形都是十分重要的。仅仅是客观实体的几何外形本身就携带了极端重要的关于该客观实体的信息。勘探地震学中偏移成像（它就是提供反射界面的几何形态）十分重要，也比较成功地应用于油气勘探中几十年就是明证。

图像处理的目的很重要地有两点：提取图像中的纹理结构（几何信息）、研究图像中空间数据的（统计意义上的）关系。这两点在勘探地震学中的地震解释和反演成像正则化方面都是核心内容。

图像处理的最新成果，譬如 Level set 方法、TV 正则化方法、PDE 方法、结构张量分析方法、凸优化方法等等正逐步渗透到勘探地震学中。

本文仅仅是试图起到抛砖引玉的作用，希望引起 WPI 研究组研究生对现代信号处理和图像处理在地震波反演成像中的作用的关注。把这些他山之石用在地震波反演成像、自动成像的方法研究中。

参考文献

[1]. W.K. Pratt: Digital image processing. John Wiley and sons, New York, 1991
[2]. B. Jahne: Digital image processing. Springer Verlag, Berlin, 1995.
[3]. Castleman, K. R., 1996, Digital image processing, Prentice Hall, Inc. 1996
[4]. Tony, F. Chan and Jianhong Shen: Image Processing and Analysis---Variational, PDE, Wavelet and Stochastic Methods, SIAM, 2005
[5]. D. Comaniciu, P. Meer: Mean shift analysis and applications. 7th International Conference on Computer Vision, Kerkyra, Greece, Sept. 1999, 1197-1203.
[6]. D. Comaniciu, P. Meer: Mean shift: a robust approach toward feature space analysis. IEEE Trans. on PAMI Vol. 24, No. 5, May 2002, 1-18.
[7]. M. Elad: On the origin of bilateral filter and ways to improve it. IEEE trans. on Image Processing Vol. 11, No. 10, October 2002, 1141-1151.
[8]. M. Nagao, T. Matsuiama: A structural analysis of complex aerial photographs. Plenum Press, New York, 1980.
[9]. C. Tomasi, R. Manduchi: Bilateral filtering for gray and color images. Proc. Sixth Int'l. Conf. Computer Vision , Bombay, 839-846
[10]. 章毓晋编著，2007 年，图像工程（第二版），清华大学出版社
[11]. Chan, T. F., and Shen Jianghong, 2005, Imaging processing and analysis --- Variational, PDF, Wavelet and Stochastic Methods, SIAM

[12]. Aubert, G., and Kornprobst, P., 2006, Mathematical Problems in Image Processing --- Partial Differential Equations and the Calculus of Variations, Springer Science + Business Media, LLC
[13]. Scherzer, O., 2011, Handbook of Mathematical Methods in Imaging, Springer Science + Business Media, LLC
[14]. Kindermann, S., Osher, S., and Jones, P. W., 2005, Debluring and denoising of images by nonlocal functionals. Multiscale model simulations, 4(4), P1091-1115
[15]. Lou, Y., Zhang, X., Osher, S., and Bertozzi, A., 2010, Image recovery via nonlocal operators. J. sci. Comput. 42(2), P185-197
[16]. Rudin, L., Osher, S., and Fatemi, E., 1992, Nonlinear total variation based noise removal algo ithms. Phys. D., 60(1-4), P259-268
[17]. Mignotte, M., 2008, A nonlocal regularization strategy for image deconvolution. Pattern Recognit. Lett. 29(16), P2206-2212
[18]. Sioni, S., Guillaume, P. Lambaré, G., and Prescott, A., et al. 2012, High definition tomography brings velocities to light. SEG Las Vegas 2012 Annual Meeting
[19]. Hardy, P., 2013, Ongoing R&D in ray based tomography: well worth the effort. SEG Houston 2013 Annual Meeting
[20]. Zdraveva, O., Hydal, S., and Woodward, M., 2013,Tomography with geological constraints: an alternative solution for resolving of carbonates. EG Houston 2013 Annual Meeting
[21]. T.F Chan, J. Shen, L. Vese. Variational PDE models in image processing[J]. Notices of the American Mathematical Society, 2003, 50(1):14~26.
[22]. J. Weickert. Anisotropic diffusion in image processing[M]. Teubner, Stuttgart, 1998.
[23]. J. Weickert. Coherence-enhancing diffusion filtering[J]. International Journal of Computer Vision, 1999,31(2/3): 111~127.
[24]. J. Weickert. Scale-space properties of nonlinear diffusion filtering with a diffusion tensor[R]. Report No. 110,Laboratory of Technomathematics, University of Kaiserslautern, P. O. Box 3049, 67653 Kaiserslautern,Germany, 1994.
[25]. J. Weickert. Anisotropic diffusion in image processing[D]. University of Kaiserslautern, Germany, January 1996.
[26]. J. Weickert. A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance[J].Journal of Visual Communication and Image Representation, 2002, 13(1/2): 103~118.
[27]. D. Tschumperlé. PDE's based regularization of multivalued images and applications[D]. Université de Nice-Sophia Antipolis, 2002.
[28]. D. Tschumperlé, R. Deriche. Diffusion PDE's on vector-valued images[J]. IEEE Signal Processing Magazine,2002, 19(5): 16~25.
[29]. D. Tschumperlé, R. Deriche. Vector-valued image regularization with PDEs: a common framework for different applications[J]. IEEE Trans Pattern Anal Mach Intell, 27(4), (2005) 506~517.
[30]. A.P. Witkin. Scale-space filtering[A]. In Proceeding of IJCAI, Karlsruhe, 1983: 1019~1021.
[31]. P. Perona, J. Malik. Scale-space and edge detection using anisotropic diffusion[J]. IEEE Transactions on PAMI, 1990, 12(7):629~639.
[32]. G. Aubert, P. Kornprobst. Mathematical problems in image processing: partial differential equations and the calculus of variations[M]. Springer-Verlag New York, 2000, 111~118.
[33]. Bruhn, A., J. Weickert, C. Feddern, T. Kohlberger, C. Schnorr. Variational optical flow computation in real time[J]. IEEE Transactions on Image Processing, 2005, 14(5): 608~615.
[34]. Bruhn, A., J. Weickert. A Multigrid platform for real-time motion computation with discontinuity-preserving variational methods[J]. International Journal of Computer Vision, 2006, 70(3): 257~277.
[35]. P. Charbonnier, L.B. Féraud, G.. Aubert, M. Barlaud. Deterministic edge-preserving regularization in computed imaging[J]. IEEE Transactions on Image Processing, 1997, 6(2): 298~311.
[36]. H.A. Aly, E. Dubois. Image up-sampling using total-variation regularization with a new observation model[J]. IEEE Transactions on Image Processing, 2005, 14(10): 1647~1659.
[37]. F. Malgouyres, F. Guichard. Edge direction preserving image zooming: A mathematical and numerical analysis[A]. SIAM Journal of Numerical Analysis, 2001, 39(1): 1~37.
[38]. R.R. Schultz, R.L. Stevenson. A Bayesian approach to image expansion for improved definition. IEEE Transactions on Image Processing, 1994, 3(3): 233~242.
[39]. J. Weickert, C. Schnorr. A theoretical framework for convex regularizers in pde-based computation of image motion[J]. International Journal of Computer Vision, 2001, 45(3): 245~264.
[40]. S. Osher, L.I. Rudin. Feature-oriented image enhancement using shock filters[J]. SIAM Journal of Numerical Analysis, 1990, 27(4): 919~940.

[41]. L. Alvarez, L. Mazorra. Signal and image restoration using shock filters and anisotropic diffusion[J]. SIAM J. Numer. Anal., 1994, 31(2): 590~605.
[42]. G.Z. Yang, P. Burger, D.N. Firmin, S.R. Underwood. Structure adaptive anisotropic image filtering[J]. Image and Vision Computing, 14: 135~145.
[43]. G. Sapiro. Geometric Partial Differential Equations and Image Analysis[M], NewYork: Cambridge University Press, 2000, 241~243.
[44]. S. Farsiu, M.D. Robinson. Fast and robust multi-frame super resolution[J]. IEEE Transactions on Image Processing, 2004, 13(10): 1327~1344.
[45]. T.F. Chan, J.H. Shen. Mathematical models for local nontexture inpaintings[J]. SIAM J. Appl. Math., 2002, 62(3): 1019~1043.
[46]. Pemna, P., Malik, J., 1990, Scale-space and edge detection using anisotropic diffusion． IEEE Trans. Pattern. Anal. Machine Intell., 12(7):629.
[47]. You, Y. L., Kaveh, M., 2000, Fourth-order partial differential equations for noise removal. IEEE Trans. Imag. Proc., 9(10): 1723.
[48]. Whitaker, R. T., Pizer, S. M., 1993, A multi-scale approach to non-uniform diffusion. Comput. Vis. Graph. Imag. Proc. Image understand, 57(1):99-110.
[49]. Black, M. J. , Sapiro, G., Hegger, D., 1998, Robust anisotropic diffusion. IEEE Trans. on Imag. Proc. 7(3): 42l-432.
[50]. Lewis, W., Starr, B., and Vigh, D., 2012, A level set approach to salt geometry inversion in full-waveform inversion, SEG Annual Meeting
[51]. Gazit, M. H., and Haber, E., 2012, Joint Inversion Through A Level Set Formulation, ASEG 22nd International Geophysical Conference and Exhibition
[52]. Zheglova, P., and Farquharson, C., 2012, Level set method in seismic inversion: 2D reconstruction of boundaries, SEG Annual Meeting
[53]. Guo, Z. H., and de Hoop, M. V., 2013, Shape optimization and level set method in full waveform inversion with 3D body reconstruction, SEG Annual Meeting
[54]. Zhang, X., and Zhang, J., 2012, Edge preserving regularization for seismic traveltime tomography, SEG Annual Meeting
[55]. Martins, C. M., Lima, W. A., Barbosa, V. C. F., and João B. C. Silva, J. B. C., 2011, Total variation regularization for depth-to-basement estimate: Part 1 —Mathematical details and applications, GEOPHYSICS, 76(1); P. I1–I12.
[56]. Haukås, J., Ravndal, O. R., Fotland, B. H., Bounaim, A., and Sonneland, L., 2013, Automated salt body extraction from seismic data using the level set method, First Break, 31, April, P35-42
[57]. Arce, G. R., 2004, Nonlinear signal processing: A statistical Approach. Jone Wiley & Sons Inc.

地震波反演成像中的线性与非线性方程组的解法

王华忠

① 同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】地震波反演成像是当前地震波成像技术的核心内容,它把 1D 波阻抗反演、AVA 反演、各种旅行时层析反演、（最小二乘）偏移成像、FWI 反演统一在一个以 Bayes 估计为基本框架的估计理论下面。在 Bayes 估计框架下，层析成像、偏移成像和 FWI 实际上是有共同的理论背景的。层析成像是以 Radon 变换为正问题、以沿波路径的反投影为反问题基本操作的。偏移成像是利用观测波场的反传播来实现的。它们都是线性化的反演。FWI 基本上把层析成像和偏移成像的功能合在一起，但是每个迭代步的梯度计算依然是在和局部线性化的基础上进行的，就是说用线性化的迭代过程逼近非线性问题的求解，付出的代价是必须给出很好的初始值，否则非线性反演结果不收敛。本文试图在地震波反演理论框架下，用地球物理的语言，解释其中的线性和非线性方程组的解法。关于梯度类方法，在 2012 年度 WPI 研究报告《全波形反演与波形层析成像的数值计算疑问中》已经讨论过。本文重点讨论如何从线性和非线性方程组求解的角度理解地震波反演成像中的计算问题以及正则化和预条件的数学和勘探地震学含义。

【关键词】地震波反演成像，线性化反演，非线性化反演，方程组求解

引言

勘探地震学目前已经发展到以地震波反演成像为主要技术手段解决地下地质体的构造成像和弹性参数是估计的阶段。油气勘探地震学的基本目标是正确地定位含油气储集体、清楚地描述含油气储集体的几何形态、精确地估计含油气储集体的弹性参数（甚至其中的流体物性参数）。

地震波反演成像的主要特征是高维地震数据的反演，其中主要包括初至波层析成像、最小二乘偏移成像、反射波层析成像，FWI 反演成像。从计算的角度看，除了 FWI 反演成像每一步迭代是用共轭状态法计算梯度外（局部线性化后才可以计算梯度！），其他几种核心方法都需要求解大规模矩阵方程组，至少原则上是这样的。

大规模线性方程组的求解是上述反演成像方法在数值计算方面的核心问题。

从反演框架的角度，把线性方程组作为线性约束条件，和要反演参数的某种凸范数结合形成的凸优化问题的求解是最近学术界的研究热点，产生了丰富的研究成果。但是，如何把这些研究成果引入巨大规模的地震波弹性参数反演中，还是一个要研究的问题。

线性方程组直接求解方法，主要是 LU 分解方法无法用于大规模矩阵方程的求解。一是计算和存贮量大；而是勘探地震学中的线性方程组的系数矩阵不是满秩的。一般地，都是用迭代求解方法。Jacobi 迭代方法、Gauss-Seidel迭代方法、逐次超松弛迭代方法尽管很简单，但它们是迭代法发展的基础。基本思想是实现了一个残差反投影。根本问题是反投影的方法过于简单。但是，这毕竟指明了残差反投影是求解线性方程的基本思想，与直接法解线性方程组的基本思想是矩阵的初等变换有异曲同工之妙。把线性方程组的求解修改成一个二次型泛函的优化问题是线性方程组求解的另一经典思路。如此可以导出最速下降方法、共轭梯度方法、各种 Krylov 子空间方法。另一种值得一提的方法就是 ART 方法和 SIRT 方法，这是直接用反投影思想构造出的线性方程组的求解方法，它们以存贮量小，计算简单而成为层析成像（油气是旅行时层析成像）中的经典算法。

根据不动点原理，非线性方程（组）的求解，在构造出满足压缩映射定理的迭代算子的情况下，可以得到非线性方程（组）求解的迭代格式，其中 Newton 方法是最基本构造迭代格式的方法。很显然，Newton 方法是一种局部线性化的方法。另一种经典的求解非线性方程组的方法是梯度类方法，误差泛函的梯度来导引误差总能量的逐步减弱。本质上，梯度类方法也是一种逐步线性化方法。

总体看来，地震波反演成像中的大规模数值计算可以抽象出本质的东西。我们认为最本质的思想有两点：线性化的思想和反投影及反传播的思想。

预条件和正则化是反演成像计算过程中的两个非常重要的问题。预条件从线性方程组的求解方面引入；正则化从反演理论和反问题优化求解的角度引入。二者有十分紧密的联系。

本文在上述思想的指导下，用地球物理的语言来分析地震波反演成像中的线性方程组合非线性方程组

的解法。详细地对比分析正则化和预条件的数学和勘探地震学含义。希望这样的分析对地震波反演成像的求解有明显的指导性。

1 地震波反演成像中的线性和非线性方程组

随着野外宽方位、高密度地震数据采集技术和计算机技术的巨大进步，地震波反演成像已经成为油气勘探地震中的标志性技术。RTM 和 FWI 就是其中的代表。

地震波反演成像的分层次目标包括：1、用叠前深度偏移成像方法得到地下反射界面的几何形态信息；2、用保幅成像技术得到反射界面上的保持反射波波形特征的成像结果；3、用最小二乘偏移技术估计反射界面处的角度反射系数；4、用以 FWI 为代表的参数反演技术得到地下岩石介质的弹性参数（主要是速度、各向异性和 Q 值三类参数）。其中，基于线性近似的叠前深度偏移成像技术和层析成像技术是核心内容。它们最终都会形成大规模的线性方程组。线性方程的形成过程和具体形式见王华忠和王雄文（2012）。

勘探地震学家对于上述线性问题的求解，归结为层析成像中的残差反投影过程和地震波偏移成像中的残差反传播过程。残差反投影沿着波路径进行；残差反传播是把残差反传播到等时面上。

值得提及的是，波场的反传播是地球物理学家对数学利用的灵活应用，利用波场反传播实现每一个单炮道集的偏移成像和梯度计算。另外,很多地方不区分反传播与反投影,统称为反投影。我们认为分成残差沿波路径的反投影和残差反传播到等时面上有利于区别背景速度的反演和扰动速度的偏移成像。

2 线性方程组的解法

无论是从什么物理问题中归纳出的线性方程组，基本形式都可以写成

$$A\mathbf{x}=\mathbf{b} \tag{2.1}$$

应该认识到的是，系数矩阵 A 是近似的，有很多因素影响其结果，譬如观测系统、速度模型、正演方法、子波等。系数矩阵 A 往往是不满秩的。数据矢量 $\mathbf{b}$ 也是有误差的。这是实际的地震波反演成像都用最小二乘方法解（2.1）式的主要原因。

一般地，线性方程组的数值解法有两类：直接法和迭代法。前者是针对满秩、小规模和系数矩阵非稀疏情形的。后者是针对大规模矩阵方程进行求解的，其解是近似的。

1、直接法

直接法就是经过有限步算术运算，可求得线性方程组精确解的方法（若计算过程中没有舍入误差）。但实际计算中由于舍入误差的存在和影响。它的基本方法是 LU 分解；基本思想是对矩阵进行初等变换，化系数矩阵 为下三角阵或对角阵。在勘探地震学中，它主要用于信号处理等环节导出的效果矩阵的求解。

2、迭代法

迭代法就是用某种极限过程去逐步逼近线性方程组精确解的方法。迭代法具有需要计算机的存储单元较少、程序设计简单、原始系数矩阵在计算过程中始终不变等优点，但存在收敛性及收敛速度问题。迭代法是解大型稀疏矩阵方程组（尤其是由微分方程离散后得到的大型方程组）的重要方法。

基本的方法包括：

（1）经典的迭代方法，其中有 Jacobi 方法、Gauss-Seidel 方法、逐次超松方法。它们可以表示成一个统一的形式：

$$\mathbf{x}^{k+1}=\mathbf{x}^{k}+\omega\Sigma\mathbf{r} \tag{2.2}$$

其中，$\mathbf{r}$ 为残差，三种方法对应的 $\mathbf{r}$ 稍有不同；ω 为松弛系数，对于 Jacobi 方法、Gauss-Seidel 迭代方

法，$\omega=1$；Σ 为反投影矩阵，$\Sigma=\begin{bmatrix} \frac{1}{a_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{a_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{a_{nn}} \end{bmatrix}$。

这些方法可以求解满秩的、较大规模的矩阵方程。但是，这些方法的收敛很慢，遇到条件数大的稀疏矩阵收敛更慢。

（2）最速下降和共轭梯度方法

把（2.1）式的求解问题转化的二次型泛函的最优化问题是求解线性方程组的基本方法。问题可以提为，

$$J(\mathbf{x})=\frac{1}{2}(A\mathbf{x}-\mathbf{b})^{\mathrm{T}}\,\mathrm{W}(A\mathbf{x}-\mathbf{b}) \tag{2.3}$$

可以证明，当系数矩阵 A 是正定时，误差泛函 $J(\mathbf{x})$ 是凸的二次型，具有一个局部极小值。该极小值就是（2.1）式的解。在勘探地震学中，（2.3）式中的 W 是一个加权矩阵，它的目的是解开误差向量各分量之间的关联。本质上，它就是数据残差向量的协方差矩阵的逆。加入此矩阵可以很好地改善解的性质。

（2.3）式的梯度为

$$\nabla J(\mathbf{x})=\mathbf{b}-A\mathbf{x} \tag{2.4}$$

可见，上式定义了迭代过程中数据的残差。

最速下降法的迭代格式为

$$\mathbf{x}^{k+1}=\mathbf{x}^{k}-\mu_k\nabla J(\mathbf{x}^{k}) \tag{2.5}$$

对比（2.2）式，可以看出不同迭代法的本质是利用残差或者说残差的反投影来修改每次的迭代结果。

共轭梯度方法本质也是如此。

3、ART 方法和 SIRT 方法

这两种方法不是迭代方法，也不同于利用初等变换的直接方法。它们更清楚反映了线性方程组的求解可以用直接反投影的方法。此时，并不要求系数矩阵 A 是方阵。它是实际上可以认为是一种最小二乘法方程的近似求解方法。

假设（2.1）式矩阵方程不是方阵，更不可能是满秩的。此时，（2.1）式的求解必然要借助最小二乘方法，对应的法方程为，

$$A^{T}A\mathbf{x}=A^{T}\mathbf{b} \tag{2.6}$$

对于大规模矩阵方程，（2.6）式中的 Hessian 矩阵是很难处理的。为此，用 Hessian 矩阵代替它本身，即 $\Sigma\approx A^{T}A$，（2.6）可以写为：

$$\mathbf{x}\approx\Sigma^{-1}A^{T}\mathbf{b} \tag{2.7}$$

系数矩阵 A 的转置矩阵就是一个反投影算子。因此，上式就是利用对数据（或残差数据）的反投影来估计参数（或参数扰动量）。

上式可以用分量的形式重写一遍，

$$x_j = \frac{\sum_{i=1}^{N} A_{ji} b_i}{\sum_{i=1}^{N} A_{ij}^2} \tag{2.8}$$

用勘探地震学的语言讲，（2.8）式的分子部分进行反投影，分母部分进行照明补偿。用射线层析成像的语言讲,分子部分的含义是用穿透第 个网格的所有 根射线的长度对 个观测数据进行加权叠加,没有实际穿过的射线,加权系数为零。分母部分的含义是穿过第 个网格的所有 根射线的长度的平方和，某种意义下这就是照明强度。没有实际穿过的射线，穿过第 个网格的射线长度为零，平方当然也为零，没有照明贡献。

上述两种方法的本质就是如此。更具体的 SIRT 方法的解释，参考王华忠和王雄文（2012）。

此处，我们仅仅探讨线性方程组求解方法所蕴含的思想以及体现的勘探地震学涵义，具体的算法参考 Saad(2003)。

3 非线性方程组的解法

3.1 非线性方程组的基本形式

无论从任何物理问题提炼出的非线性问题，一般地可以用积分方程来表达，建立观测的物理量（数据）和引起该物理量变化的参数之间的关系，这个关系往往以高维褶积的形式出现，褶积算子或核函数具有重要的意义。离散化此积分方程会形成如下的非线性方程组，

$$\begin{cases} f_1(x_1, x_2, ..., x_m) = 0 \\ f_2(x_1, x_2, ..., x_m) = 0 \\ f_3(x_1, x_2, ..., x_m) = 0 \\ ..... \\ f_m(x_1, x_2, ..., x_m) = 0 \end{cases} \tag{3.1}$$

其中，$f_1, f_2, f_3, ..., f_m$ 为 $x_1, x_2, ..., x_m$ 的连续函数。

记 $\mathbf{x} = (x_1, x_2, ..., x_m), \mathbf{x} \in R^m$，则上式改为：

$$f_i(\mathbf{x}) = 0, i = 1, 2, ..., m \tag{3.2}$$

引入向量函数 $F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x}), ..., f_m(\mathbf{x}))^T$，则有更为简洁的非线性方程组的形式，

$$F(\mathbf{x}) = 0 \tag{3.3}$$

从勘探地震学的语言讲，（3.3）式就是一个残差计算的公式。它不可以写成类似（2.1）的矩阵形式，只有线性化后，才可以有线性化的矩阵方程。

3.2 非线性方程组的基本解法

对于非线性方程组（3.3）的求解有很多方法，例如：简单迭代法、Newton 法、Newton 下山法、牛顿松弛型迭代法、拟牛顿法与 Broyden 方法和非线性最小二乘问题数值方法等，下面我们逐个介绍这些方法。

3.2.1 解方程组 $\mathbf{x} = G(\mathbf{x})$ 的简单迭代法

提出的问题: $F(\mathbf{x})$ 实函数向量，求 $F(\mathbf{x}) = 0$ 的近似解。

3.2.1.1 基本思想方法

先将 $F(\mathbf{x}) = 0$ 化为等价方程

$$\mathbf{x} = G(\mathbf{x}) \tag{3.4}$$

从某个初始向量 $\mathbf{x}^{(0)}$ 出发，由如下的迭代格式，

$$\mathbf{x}^{(k+1)} = G(\mathbf{x}^{(k)}) \tag{3.5}$$

作向量序列 $\{\mathbf{x}^{(k)}\}$，其中 $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$，$G(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_3(\mathbf{x}))^T$。(3.5)式称为简单迭代或单点迭代或单步迭代法。映射 $G(\mathbf{x})$ 称为迭代映射。

假设每个 g_i 有连续的二阶导数：$\dfrac{\partial g_i}{\partial x_p \partial x_q}, 1 \le i, p, q \le n$，$G(\mathbf{x})$ 在点 $\mathbf{x}$ 的 F-导数为：

$$G'(x) = \begin{pmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial g_n}{\partial x_1} & \dfrac{\partial g_n}{\partial x_2} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{pmatrix}$$

设 $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_n^*,)$ 是（3.4）的解，则称 $\mathbf{x}^*$ 为 $G(\mathbf{x})$ 的不动点。当迭代(3.5)收敛时,极限点 $\tilde{\mathbf{x}}$ 又是 $G(\mathbf{x})$ 的连续点,则

$$\tilde{\mathbf{x}} = \lim_{k\to\infty} \mathbf{x}^{(k+1)} = \lim_{k\to\infty} G(\mathbf{x}^{(k)}) = G(\lim_{k\to\infty} \mathbf{x}^{(k)}) = G(\tilde{\mathbf{x}}) \tag{3.6}$$

即 $\tilde{\mathbf{x}}$ 是 $G(\mathbf{x})$ 的一个不动点。

3.2.1.2 简单迭代的收敛性

定义若对 $\forall \mathbf{x}, \mathbf{y} \in D$，存在常数 $L < 1$，成立

$\|G(\mathbf{x}) - G(\mathbf{y})\| \le L \| \mathbf{x} - \mathbf{y} \|$（3.6）则映射 $G(\mathbf{x})$ 在区域 D 上称为压缩的。常数 L 称为压缩因子。

定理(压缩不动点定理)：设映射 $G(\mathbf{x})$ 在区域 D 上满足：

（1）$G(\mathbf{x}) \in D, \forall \mathbf{x} \in D$;

（2）$G(\mathbf{x})$ 在区域 D 上是压缩映射，压缩因子为 L。

则对 $\forall \mathbf{x}^{(0)} \in D$，简单迭代格式(3.5)产生的序列 $\{\mathbf{x}^{(k)}\}$ 收敛于 $G(\mathbf{x})$ 在区域 D 上的唯一不动点 $\mathbf{x}^*$，且有误差估计：

$$\| \mathbf{x}^* - \mathbf{x}^{(k)} \| \le \frac{L^k}{1-L} \| \mathbf{x}^{(1)} - \mathbf{x}^{(0)} \|, k = 0, 1, \cdots. \tag{3.7}$$

定理（局部收敛定理） 若映射 $G(\mathbf{x})$ 在不动点 $\mathbf{x}*$ 处有 F 导数 $G'(\mathbf{x}^*)$，而且其谱半径小于 1：$\rho(G'(\mathbf{x}^*)) < 1$，则存在 $\delta > 0$，只要 $\mathbf{x}^{(0)} \in D_\delta$，由 $\mathbf{x}^{(k+1)} = G(\mathbf{x}^{(k)})$ 产生的迭代序列 $\{\mathbf{x}^{(k)}\}$ 收敛于 $\mathbf{x}^*$。

3.2.2 解方程组 $F(\mathbf{x}) = 0$ 的 Newton 法

基于不动点原理的简单迭代法对于构造迭代格式具有普适的指导性，但还需要具体的方法构造出可以计算的迭代格式。最基本的方法是 Newton 方法，它是基于局部线性化的思想。

3.2.2.1 Newton 法迭代公式

假设非线性方程组 $F(\mathbf{x}) = 0$ 的精确解或真解为 $\mathbf{x}^*$。$F(\mathbf{x})$ 在 $\mathbf{x}*$ 邻近有连续的 F 导数 $F'(\mathbf{x}*)$。且

$F'(\mathbf{x}^*)$非奇异，即$\det F'(\mathbf{x}^*)\neq 0$。

并设$\mathbf{x}^{(k)}$是（3.3）式的一个近似解。且$F(\mathbf{x})$在点$\mathbf{x}^{(k)}$有F导数$F'(\mathbf{x}^{(k)})$，则仿射映射$y=L(\mathbf{x})=F(\mathbf{x}^{(k)})+F'(\mathbf{x}^{(k)})(\mathbf{x}-\mathbf{x}^{(k)})$是映射(函数)$y=F(\mathbf{x})$的局部近似。

用$L(\mathbf{x})$近似$F(\mathbf{x})$，用$L(\mathbf{x})=0$的解作为（3.3）的改进解。很显然，$L(\mathbf{x})=0$的解为$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-[F'(\mathbf{x}^{(k)})]^{-1}F(\mathbf{x}^{(k)})$，其中，要求$\det F'(\mathbf{x}^{(k)})\neq 0$成立。

因此，有如下的Newton迭代公式

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-[F'(\mathbf{x}^{(k)})]^{-1}F(\mathbf{x}^{(k)})\ k=0,1,2,\cdots\cdots \tag{3.8}$$

3.2.2.2 牛顿法与简单迭代法的关系

定义映射$G(\mathbf{x})=\mathbf{x}-[F'(\mathbf{x})]^{-1}F(\mathbf{x})$，并要求$\det F'(\mathbf{x}^{(k)})\neq 0$，则Newton法迭代格式可以写为

$$\mathbf{x}^{(k+1)}=G(\mathbf{x}^{(k)}),\quad k=0,1,2,\cdots\cdots \tag{3.9}$$

与压缩映射定理类比，可以看出对映射算子$G(\mathbf{x})=\mathbf{x}-[F'(\mathbf{x})]^{-1}F(\mathbf{x})$的要求。

3.2.2.3 收敛定理

局部超线性收敛定理：如果$F(\mathbf{x})$在解$\mathbf{x}^*$邻近有连续的 F 导数, 且$\det F'(\mathbf{x}^*)\neq 0$。则存在$\delta>0$，只要$\|\mathbf{x}^{(0)}-\mathbf{x}^*\|\leq\delta$，Newton迭代生成的序列$\{\mathbf{x}^{(k)}\}$：超线性收敛于$\mathbf{x}^*$，即：

$$\lim_{k\to\infty}\frac{\|\mathbf{x}^{(k+1)}-\mathbf{x}^*\|}{\|\mathbf{x}^{(k)}-\mathbf{x}^*\|}=0\text{。} \tag{3.10}$$

牛顿法不仅收敛且在一般情况下收敛速度较快，是二阶收敛。

定理（Newton 法局部二阶收敛性）：如果$F(\mathbf{x})$的每个分量在解$F(\mathbf{x})$邻近有二阶连续偏导数，$\det F'(\mathbf{x}^*)\neq 0$。则存在$\delta>0$，只要$\|\mathbf{x}^{(0)}-\mathbf{x}^*\|\leq\delta$，Newton序列至少二阶收敛于$\mathbf{x}^*$，即

$$\|\mathbf{x}^{(k+1)}-\mathbf{x}^*\|\leq C\|\mathbf{x}^{(k)}-\mathbf{x}^*\|^2,\quad k=0,1,2,\cdots\cdots \tag{3.11}$$

3.2.3 牛顿法的变形与离散牛顿法

Newton法缺点是显而易见的，每步要计算Jacobi阵$F'(\mathbf{x}^{(k)})$的逆。因此，（3.8）式在勘探地震学中的大规模参数反演中是无法应用的。

存在很多改进的方法：m步Newton法、割线法、拟Newton法、Broyden方法、一般quasi- Newton法。

3.2.3.1 简化牛顿法和修正牛顿法

针对牛顿法计算量大的缺点，可将牛顿法（3.8）简化为

$$\mathbf{x}^{k+1}=\mathbf{x}^k-F'(\mathbf{x}^0)^{-1}F(\mathbf{x}^k),\quad k=0,1,... \tag{3.12}$$

此为上述简化牛顿法。除第 1 步外，它每步只计算一个$F(\mathbf{x})$值，但它只是线性收敛。若综合牛顿法收敛快和简化牛顿法计算量少的优点，将m步简化牛顿法组成一步牛顿迭代，则得

$$\begin{cases}\mathbf{x}^{k,i+1}=\mathbf{x}^{k,i}-F'(\mathbf{x}^k)^{-1}F(\mathbf{x}^{k,i}),\quad i=0,1,...,m-1\\ \mathbf{x}^k=\mathbf{x}^{k,0},\mathbf{x}^{k+1}=\mathbf{x}^{k,m},\quad k=0,1,...\end{cases} \tag{3.13}$$

称为修正牛顿法（也称萨马斯基（1967）技巧）。

3.2.3.2 牛顿下山法

牛顿法收敛速度快，但对初值 $\mathbf{x}_0$ 要求苛刻。在实际应用中不容易确定，有时往往由于初值选取不当而使迭代不收敛。Newton 下山法是一种降低对初值要求的修正的牛顿法。

引理：若 $F(\mathbf{x})\neq 0$， $F'(\mathbf{x})\neq 0$，则一定存在 $\Delta>0$，当 $0<t\leq\Delta$ 时，成立

$$\| F(\mathbf{x}-t[F'(\mathbf{x})]^{-1}F(\mathbf{x}))\| < \| F(\mathbf{x})\| \tag{3.14}$$

下山法：Newton 法的修正方向 $-[F'(\mathbf{x})]^{-1}F(\mathbf{x})$ 是 $F(\mathbf{x})$ 在 $\mathbf{x}$ 点的下山方向。在牛顿法中引进下山因子：$\omega_k\in(0,1)$，从而由

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-\omega_k[F'(\mathbf{x}^{(k)})]^{-1}F(\mathbf{x}^{(k)}),\quad k=0,1,2,\cdots\cdots \tag{3.15}$$

使 $\| F(\mathbf{x}^{(0)})\| > \| F(\mathbf{x}^{(1)})\| > \cdots$，呈下山状态。

通常取 $\omega_k\in\{1,\frac{1}{2},\frac{1}{4},\cdots\}$，使 $\| F(\mathbf{x}^{(k)}-2^{-i}[F'(\mathbf{x}^{(k)})]^{-1}F(\mathbf{x}^{(k)}))\| < \| F(\mathbf{x}^{(k)})\|$ 成立的最大值。

3.2.3.3 阻尼牛顿法

使用牛顿法（3.8）还要求 $F'(\mathbf{x}^k)$ 非奇异，故当 $F'(\mathbf{x}^k)$ 奇异或严重病态时都不能使迭代进行下去，为解决这一问题，可在（3.8）中引入参数 λ_k，使迭代序列改为

$$\mathbf{x}^{k+1}=\mathbf{x}^k-[F'(\mathbf{x}^k)+\lambda_k\mathbf{I}]F(\mathbf{x}^k),k=0,1,... \tag{3.16}$$

其中，λ_k 称为阻尼参数。方法（3.16）称为阻尼牛顿法，它可使 $F'(\mathbf{x}^k)+\lambda_k\mathbf{I}$ 非奇异，非病态，从而使迭代序列 $\{\mathbf{x}^k\}_0^\infty$ 有意义，这时，如果 $\{\mathbf{x}^k\}_0^\infty$ 收敛于 $\mathbf{x}^*$，只有线性收敛速度。

3.2.3.4 离散 Newton 法和 Newton-Steffensen 方法

有时 $F(\mathbf{x})$ 的导数 $F'(\mathbf{x})$ 不易计算，可采用均差近似导数，将（3.8）改为

$$\mathbf{x}^{k+1}=\mathbf{x}^k-J(\mathbf{x}^k,\mathbf{h}^k)^{-1}F(\mathbf{x}^k),k=0,1,... \tag{3.17}$$

其中，$J(\mathbf{x}^k,\mathbf{h}^k)=\left(\frac{1}{h_1^k}\left[F(\mathbf{x}^k+h_1^ke_1)-F(\mathbf{x}^k)\right],...,\frac{1}{h_n^k}\left[F(\mathbf{x}^k+h_n^ke_n)-F(\mathbf{x}^k)\right]\right)$ 为 $\mathbf{x}^k$ 处 $F(\mathbf{x})$ 的雅克比矩阵 $F'(\mathbf{x}^k)$ 的近似，$\mathbf{h}^k=(h_1^k,...,h_n^k)^T$ 是一个给定向量。（3.17）称为离散牛顿法。计算时步长向量 $\mathbf{h}^k=\mathbf{h}$（不依赖 k）。方法仍是超线性收敛，若取 $\mathbf{h}^k=F(\mathbf{x}^k)$，则（3.17）称为 Newton-Steffensen 方法，它与牛顿法同样是平方收敛的。

3.2.4 牛顿松弛型迭代法

用牛顿法求非线性方程组（3.3），除了上述（3.8）式定义的迭代格式外，还存在一类方法，即每一步都求解如下线性方程组

$$F'(\mathbf{x}^k)\Delta\mathbf{x}^k=-F(\mathbf{x}^k),\ k=0,1,... \tag{3.18}$$

得到当前步的修正量 $\Delta\mathbf{x}^k$，把修正量加到上一步的估计结果上，即 $\mathbf{x}^{k+1}=\mathbf{x}^k+\Delta\mathbf{x}^k$。

这样的方法是更明显地足逐步线性化方法求解非线性问题。这涉及到格式 $\mathbf{x}^{k+1}=\mathbf{x}^k+\Delta\mathbf{x}^k$ 的外迭代和（3.18）式迭代求解的内迭代。因此会产生牛顿－雅克比迭代、牛顿－高斯－赛德尔迭代及牛顿－SOR 迭

代。迭代法（3.18）的特点是不用求 Jacobi 阵 $F'(\mathbf{x}^{(k)})$ 的逆，计算简单，适用于解大型稀疏的非线性方程组。但它通常只有线性收敛速度。

可以想见，还存在一类线性-非线性解法相结合的双层迭代格式。如用牛顿法解这些一元非线性方程，而且牛顿法只是迭代一步，则有

$$x_i^{k+1}=x_i^k+\omega\frac{f_i(x_1^k,...,x_n^k)}{\frac{\partial}{\partial x_i}f_i(x_1^k,...,x_n^k)}\quad i=1,...,\mathrm{n},\mathrm{k}=0,1,... \tag{3.19}$$

称为一步雅克比-牛顿迭代。

类似地，可得一步 SOR-牛顿法

$$x_i^{k+1}=x_i^k+\omega\frac{f_i(x_1^{k+1},...,x_{i-1}^{k+1},x_i^k,x_n^k)}{\frac{\partial}{\partial x_i}f_i(x_1^{k+1},...,x_{i-1}^{k+1},x_i^k,x_n^k)}\quad i=1,...,\mathrm{n},\mathrm{k}=0,1,... \tag{3.20}$$

当 $\omega=1$ 时，（3.20）称为赛德尔-牛顿法。迭代（3.19）与（3.20）每步计算 n 个分量函数值和 n 个偏导数值，不用求逆，计算简单，存储量少，特别是（3.19）计算 $f_i(x)$ 及 $\frac{\partial f_i(x)}{\partial x_i}$ 用的都是前一步的信息，适合并行计算。

迭代法（3.19）及（3.20）在 $\omega\in(0,1]$ 及 $F(\mathbf{x})$ 满足一定条件下已证明迭代序列 $\{\mathbf{x}^k\}$ 是线性收敛的。

3.2.5 拟牛顿法与 Broyden 方法

3.2.5.1 拟牛顿法基本思想

求解方程组 $F(\mathbf{x})=0$ 的牛顿法收敛很快，但每步都要计算 $F'(\mathbf{x}^k)$，很不方便，有时计算很困难且计算量大。如何使每步不必计算 $F'(\mathbf{x}^k)$ 而又保持超线性收敛？拟牛顿法就是为解决这一问题而产生的，自 20 世纪 60 年代提出以来有很大发展。

首先考虑用较简单的矩阵 A_k 近似 $F'(\mathbf{x}^k)$，将迭代公式写成

$$\mathbf{x}^{k+1}=\mathbf{x}^k-A_k^{-1}F(\mathbf{x}^k),\qquad \mathrm{k}=0,1,... \tag{3.21}$$

此处，$A_k\in R^{n\times n}$ 依赖于 $F(\mathbf{x}^k)$ 及 $F'(\mathbf{x}^k)$。为了避免每步都重新计算 A_k，考虑下一步 A_{k+1} 只对 A_k 进行修正，它要求 A_{k+1} 满足方程

$$A_{k+1}(\mathbf{x}^{k+1}-\mathbf{x}^k)=F(\mathbf{x}^{k+1})-F(\mathbf{x}^k),\qquad \mathrm{k}=0,1... \tag{3.22}$$

称为拟牛顿方程。它表明矩阵 A_{k+1} 关于点 $\mathbf{x}^k$ 及 $\mathbf{x}^{k+1}$ 具有“差商”性质，即当 $n=1$ 时 A_{k+1} 即为 $F(\mathbf{x})$ 关于 $\mathbf{x}^k$ 及 $\mathbf{x}^{k+1}$ 的差商；但当 $n>1$ 时 A_{k+1} 并不确定，为此我们限制 A_{k+1} 是由 A_k 的的一个低秩修正矩阵得到的，即

$$A_{k+1}=A_k+\Delta A_k,\qquad rank(\Delta A_k)=\mathrm{m}>1 \tag{3.23}$$

其中 ΔA_k 是秩为 m 的修正矩阵，由（3.21），（3.22），（3.23）组成的迭代法就称为拟牛顿法。它通过给出的初始近似 $\mathbf{x}^0$ 及矩阵 A_0，由这三式可逐次计算得到 $\{\mathbf{x}^k\}_0^\infty$ 及 $\{A_k\}_0^\infty$，从而避免了每步计算 $F(\mathbf{x})$ 的雅

克比矩阵，使计算量减少，这种算法由于（3.23）中 ΔA_k 有不同的取法，因此可得到许多不同的拟牛顿算法，常用的拟牛顿算法是 ΔA_k 为秩 1 或秩 2 的方法。

下面列出拟牛顿法得到的序列 $\{\mathbf{x}^k\}$ 具有超线性收敛的理论。

定理：设 $F:D\subset R^n\to R^n,\mathbf{x}^*\in \mathrm{D}$ 使 $F(\mathbf{x}^*)=0$，且 $F'(\mathbf{x}^*)$ 非奇异，$F(\mathbf{x})$ 在 $S=S(\mathbf{x}^*,\delta)\subset \mathrm{D}$ 上有连续的 F- 导数，且 $\{A_k\}$ 为非奇异的矩阵序列，则对任何 $\mathbf{x}^0\in S$，由（3.21），（3.22），（3.23）组成序列 $\{\mathbf{x}^k\}$ 超线性收敛的充分必要条件是

$$\lim_{k\to\infty}\frac{\left\|[A_k-F'(\mathbf{x}^*)](\mathbf{x}^{k+1}-\mathbf{x}^k)\right\|}{\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\|}=0 \tag{3.24}$$

从此定理可以看到拟牛顿法方向 $\mathbf{s}^k=\mathbf{x}^{k+1}-\mathbf{x}^k=-A_k^{-1}F(\mathbf{x}^k)$ 与牛顿法方向 $\mathbf{s}_N^k=-F'(\mathbf{x}^k)^{-1}F(\mathbf{x}^k)$ 有以下关系：

$$\mathbf{s}^k-\mathbf{s}_N^k=\mathbf{s}^k+F'(\mathbf{x}^k)^{-1}F(\mathbf{x}^k)=F'(\mathbf{x}^k)^{-1}[F'(\mathbf{x}^k)-A_k]\mathbf{s}^k$$

于是等式（3.24）等价于 $\lim_{k\to\infty}\frac{\left\|\mathbf{s}^k-\mathbf{s}_N^k\right\|}{\left\|\mathbf{s}^k\right\|}=0$，它表明拟牛顿法的迭代序列 $\{\mathbf{x}^k\}$ 具有超线性收敛的条件是其修正方向 $\mathbf{s}^k$ 在大小和方向都与牛顿法修正方向 $\mathbf{s}_N^k$ 一致。

3.2.5.2 秩 1 拟牛顿法与 Broyden 方法

秩 1 算法是指（3.23）中修正矩阵 ΔA_k 的秩 $m=1$ 时的拟牛顿法，因为对任何 $n\times n$ 阶的秩 1 矩阵 ΔA 均可表示为 $\Delta A=\mathbf{u}\cdot\mathbf{v}^T$，其中 $\mathbf{u},\mathbf{v}\in R^n$，于是

$$\Delta A_k=\mathbf{u}_k\cdot\mathbf{v}_k^{\mathrm{T}},\qquad \mathbf{u_k},\mathbf{v}_k\in R^n \tag{3.25}$$

下面主要是选择合适的 $\mathbf{u}_k,\mathbf{v}_k$ 使它满足拟牛顿方程（3.22），若记 $\mathbf{s}^k=\mathbf{x}^{k+1}-\mathbf{x}^k,\mathbf{y}^k=F(\mathbf{x}^{k+1})-F(\mathbf{x}^k)$，则（3.22）为

$$A_{k+1}\mathbf{s}^k=\mathbf{y}^k \tag{3.26}$$

将（3.25）代入（3.23），则得

$$A_{k+1}=A_k+\mathbf{u}_k\cdot\mathbf{v}_k^T$$

代入（3.26），得

$$(A_k+\mathbf{u}_k\cdot\mathbf{v}_k^T)\mathbf{s}^k=\mathbf{y}^k$$
$$\mathbf{u}_k\cdot\mathbf{v}_k^T\cdot\mathbf{s}^k=\mathbf{y}^k-A_k\mathbf{s}^k$$

若 $\mathbf{v}_k^T\cdot\mathbf{s}^k\neq 0$，则有

$$\mathbf{u}_k=\frac{1}{\mathbf{v}_k^T\cdot\mathbf{s}^k}(\mathbf{y}^k-A_k\mathbf{s}^k),$$

将它代入（3.25），即得

$$\Delta A_k = \frac{1}{\mathbf{v}_k^T \cdot \mathbf{s}^k}(\mathbf{y}^k - A_k\mathbf{s}^k)\mathbf{v}_k^T, \qquad \mathbf{v}_k^T \cdot \mathbf{s}^k \neq 0 \tag{3.27}$$

若取 $\mathbf{v}_k = \mathbf{s}^k$，只要 $\mathbf{s}^k \neq 0$，就有 $(\mathbf{s}^k)^T \cdot \mathbf{s}^k > 0$，由（3.27）得

$$\Delta A_k = (\mathbf{y}^k - A_k\mathbf{s}^k)\frac{(\mathbf{s}^k)^T}{(\mathbf{s}^k)^T \cdot \mathbf{s}^k} \tag{3.28}$$

此时可得到秩 1 拟牛顿法

$$\begin{cases} \mathbf{x}^{k+1} = \mathbf{x}^k - A_k^{-1}F(\mathbf{x}^k), \\ A_{k+1} = A_k + (\mathbf{y}^k - A_k\mathbf{s}^k)\dfrac{(\mathbf{s}^k)^T}{(\mathbf{s}^k)^T \cdot \mathbf{s}^k}, \end{cases} \qquad \mathrm{k} = 0,1,... \tag{3.29}$$

此算法称为布罗伊登（Broyden）秩 1 方法。按（3.28）确定的 ΔA_k 满足拟牛顿方程（3.22）。实际上（3.28）两端右乘 $\mathbf{s}^k$，即得

$$(A_{k+1} - A_k)\mathbf{s}^k = (\mathbf{y}^k - A_k\mathbf{s}^k)$$

于是有 $A_{k+1}\mathbf{s}^k = \mathbf{y}^k$，故（3.29）是拟牛顿法的一种具体算法。

在（3.27）中若取 $\mathbf{v}_k = F(\mathbf{x}^{k+1})$，由（3.28）又有

$$\mathbf{y}^k - A_k\mathbf{s}^k = F(\mathbf{x}^{k+1}) - F(\mathbf{x}^k) - A_k\mathbf{s}^k = F(\mathbf{x}^{k+1})$$

于是由（3.27）可得秩 1 修正矩阵

$$\Delta A_k = (\mathbf{y}^k - A_k\mathbf{s}^k)\frac{(\mathbf{y}^k - A_k\mathbf{s}^k)^T}{(\mathbf{y}^k - A_k\mathbf{s}^k)^T\mathbf{s}^k}$$

显然 A_{k+1} 满足拟牛顿方程（3.22），并且 ΔA_k 对称，故当 A_0 对称，则所有 $A_{k+1}(\mathrm{k} = 0,1,...)$ 也对称，于是可得到对称秩 1 算法

$$\begin{cases} \mathbf{x}^{k+1} = \mathbf{x}^k - A_k^{-1}F(\mathbf{x}^k), \\ A_{k+1} = A_k + (\mathbf{y}^k - A_k\mathbf{s}^k)\dfrac{(\mathbf{y}^k - A_k\mathbf{s}^k)^T}{(\mathbf{y}^k - A_k\mathbf{s}^k)^T\mathbf{s}^k}, \end{cases} \mathrm{k} = 0,1,... \tag{3.30}$$

称为 Broyden 对称秩 1 算法，它适合于 $F'(\mathrm{x})$ 为对称的方程组求解。（3.29）与（3.30）是最常用的秩 1 拟牛顿法，可以证明这两个算法得到的迭代序列 $\{\mathbf{x}^k\}$ 均为超线性收敛。

另外，在所有秩 1 的公式中，若取 $\mathbf{v}_k = \mathbf{s}^k$ 的修正（3.28）在 $\|\bullet\|_F$ 泛数意义下具有极小性质。为了使下面定理书写简便，把（3.28）中的所有 k 标号取消，改写为

$$\overline{A} - A = \Delta A = \frac{(\mathbf{y} - A\mathbf{s})\mathbf{s}^T}{\mathbf{s}^T \cdot \mathbf{s}} \tag{3.31}$$

定理 设 $A \in R^{n\times n}$， $\mathbf{s},\mathbf{y} \in R^n$ 且 $\mathbf{s} \neq 0$，由（3.31）确定的 $\overline{A}$ 是极小问题

$\min\{\|\hat{A} - A\|_F \,|\, \hat{A}\mathbf{s} = \mathbf{y}\}$ 的解。即 $\|\overline{A} - A\|_F \leq \|\hat{A} - A\|_F$ 对 $\hat{A}\mathbf{s} = \mathbf{y}$ 的所有 $\hat{A}$ 成立。

定理表明 Broyden 方法（3.29）是所有秩 1 拟牛顿法中在 $\|\bullet\|_F$ 意义下 $\|\Delta A_k\|_F$ 是最小的。

在 Broyden 方法中每步都要求 A_k^{-1}，其计算量一般为 $O(\mathrm{n}^3)$。为了避免求逆可以利用以下引理。

引理（Sherman-Morrison 公式）：设 $A\in R^{n\times n}$，且 A 非奇异，$\mathbf{u},\mathbf{v}\in R^n$，当且仅当 $1+\mathbf{v}^T A^{-1}\mathbf{u}\neq 0$ 时，$A+\mathbf{u}\cdot\mathbf{v}^T$ 可逆，且有

$$(A+\mathbf{u}\cdot\mathbf{v}^T)^{-1}=A^{-1}-\frac{A^{-1}\mathbf{u}\cdot\mathbf{v}^T A^{-1}}{1+\mathbf{v}^T A^{-1}\mathbf{u}} \tag{3.32}$$

在（3.32）中，令 $H_k=A_k^{-1}$，并在引理中取 $A=A_k,\mathbf{v}=\mathbf{s}^k,\mathbf{u}=\dfrac{\mathbf{y}^k-A_k\mathbf{s}^k}{\|\mathbf{s}^k\|_2^2}$，可将（3.30）改写为逆 Broyden 秩 1 方法

$$\begin{cases}\mathbf{x}^{k+1}=\mathbf{x}^k-H_kF(\mathbf{x}^k),\\ H_{k+1}=H_k+(\mathbf{s}^k-\mathrm{H}_k\,\mathbf{y}^k)\dfrac{(\mathbf{s}^k)^T H_k}{(\mathbf{s}^k)^T H_k\mathbf{y}^k},\end{cases}\qquad k=0,1,... \tag{3.33}$$

这里 $(\mathbf{s}^k)^T H_k\mathbf{y}^k\neq 0$。

实际应用（3.29）或（3.33）求方程组（3.3）的数值解时，只要选择较好的初始近似 $\mathbf{x}^0$ 和初始矩阵 A_0 或 H_0，一般可得到足够精确的数值解。对于方法（3.30），与它互逆的秩 1 算法公式为

$$\begin{cases}\mathbf{x}^{k+1}=\mathbf{x}^k-H_kF(\mathbf{x}^k),\\ H_{k+1}=H_k+(\mathbf{s}^k-\mathrm{H}_k\,\mathbf{y}^k)\dfrac{(\mathbf{s}^k-\mathrm{H}_k\,\mathbf{y}^k)^T}{(\mathbf{s}^k-\mathrm{H}_k\,\mathbf{y}^k)^T\mathbf{y}^k},\end{cases}\qquad k=0,1,... \tag{3.34}$$

逆 Broyden 方法（3.33）及（3.34）不求逆，由 H_k 计算 H_{k+1} 的运算量为 $O(n^2)$ 比直接求 A_k^{-1} 计算量 $O(n^3)$ 节省，但这种方法计算 H_{k+1} 时稳定性较差，有时效果不理想。

4 非线性最小二乘问题数值解法

4.1 问题的基本提法

无论是勘探地震学中的问题，或是很多其他学科为问题，很多都归结为利用一个根据物理规律得到的正演函数（或称拟合函数）

$$y=g(x_1,...x_n) \tag{4.1}$$

其中，$x_1,x_2,...x_n$ 为正演或系统中的参数，$g(x_1,...x_n)$ 关于参数 $\mathbf{x}=(x_1,x_2,...,x_n)$ 为非线性的。若用 $g_i(\mathbf{x})$ 拟合实测数据 $y_i,i=1,2,...,m$，用最小二乘法确定参数时就得到关于

$$\varphi(\mathrm{x}_1,\mathrm{x}_2,...\mathrm{x}_n)=\sum_{i=1}^{m}[\mathrm{g}_i(\mathrm{x}_1,...\mathrm{x}_n)-y_i]^2,\qquad m>n \tag{4.2}$$

的极小问题。在勘探地震学中，上述问题的提法一般是在更广泛的 Bayes 估计的框架下给出的（Tarantola,2005）。

下面进行更为抽象的描述。设 $\mathbf{F}:D\subset R^n\to R^m,\mathbf{F}=(f_1,...,f_m)^T$，定义函数

$$\varphi(\mathbf{x})=\frac{1}{2}F(\mathbf{x})^T F(\mathbf{x}) \tag{4.3}$$

此时，$F(\mathbf{x})$ 代表数据残差向量，其中的分量 $f_i, i=1,2,\cdots,m$ 代表残差分量。显然 $\varphi: D\subset R^n \to R$，于是非线性最小二乘问题就是求解

$$\min_{\mathbf{x}\in D}\varphi(\mathbf{x})=\min_{\mathbf{x}\in D}\frac{1}{2}F(\mathbf{x})^T F(\mathbf{x}) \tag{4.4}$$

这是求多元函数 $\varphi(\mathbf{x})$ 极小的问题。

4.2 高斯-牛顿法

由极值存在的必要条件，若 $F(\mathbf{x})$ 在 D 上可微，$\varphi(\mathbf{x})$ 对 $\mathbf{x}$ 求导可得 $\varphi(\mathbf{x})$ 的极小点满足方程

$$g(\mathbf{x})=\nabla\varphi(\mathbf{x})=DF(\mathbf{x})^T F(\mathbf{x})=0 \tag{4.5}$$

称为法方程。其中 $DF(\mathbf{x})^T=\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$，$g(\mathbf{x})$ 是梯度向量。

为了构造求解方程（4.5）的迭代法，可将 $F(\mathbf{x})$ 在 $\mathbf{x}^k$ 点线性展开

$$F(\mathbf{x})\approx DF(\mathbf{x}^k)(\mathbf{x}-\mathbf{x}^k)+F(\mathbf{x}^k)=l_k(\mathbf{x})$$

在（4.5）中用 $l_k(\mathbf{x})$ 代替 $F(\mathbf{x})$，$DF(\mathbf{x}^k)$ 代替 $D\mathrm{F}(\mathbf{x})$，得

$$DF(\mathbf{x}^k)^T l_k(\mathbf{x})=0$$

并将它的解记作 $\mathbf{x}^{k+1}$，于是有

$$\mathbf{x}^{k+1}=\mathbf{x}^k-\left[DF(\mathbf{x}^k)^T DF(\mathbf{x}^k)\right]^{-1} DF(\mathbf{x}^k)^T F(\mathbf{x}^k) \tag{4.6}$$

称此公式为高斯-牛顿法，采用（4.5）的记号，并令

$$G(\mathbf{x})=DF(\mathbf{x})^T DF(\mathbf{x}) \tag{4.7}$$

则（4.6）可改写成

$$\mathbf{x}^{k+1}=\mathbf{x}^k-G(\mathbf{x}^k)^{-1} g(\mathbf{x}^k), \qquad k=0,1,... \tag{4.8}$$

它形式上与解方程（4.5）的牛顿法相似，但它不是牛顿法，因为解（4.5）的牛顿法要计算 $F(\mathbf{x})$ 的 2 阶导数，比较复杂。（4.5）式是全牛顿法。

而这里只计算 $DF(\mathbf{x}^k)$，较为简单，且仍具有较好的性质。（4.8）是 Gauss-newton 方法。事实上，若

$F(\mathbf{x})=A\mathbf{x}-\mathbf{b}$，其中 $A\in R^{m\times n},n<m,rank(A)=\mathrm{n}$，则（4.3）的 $\varphi(\mathbf{x})$ 是二次函数。这时有 $g(\mathbf{x})=\nabla\varphi(\mathbf{x})=A^T(A\mathbf{x}-\mathbf{b}),\quad G(\mathbf{x})=A^TA$。$A^TA$ 为对称正定矩阵。对于任意初始近似 $\mathbf{x}^0\in R^n$，用高斯-牛顿法（4.8）经一步迭代则可得问题（4.4）的解 $\mathbf{x}^1=\mathbf{x}^*=(\mathrm{A}^T\ A)^{-1}A^T\mathbf{b}$，称此性质为二次终止性。

对于一般非线性函数 $F(\mathbf{x})$，由（4.8）中记 $\mathbf{p}_k=-G(\mathbf{x}^k)^{-1}g(\mathbf{x}^k)=-G(\mathbf{x}^k)^{-1}\nabla\varphi(\mathbf{x}^k)$。于是 $\mathbf{p}_k^T\nabla\varphi(\mathbf{x}^k)=-\mathbf{p}_k^TG(\mathbf{x}^k)\mathbf{p}_k<0$。它表明向量 $\mathbf{p}_k$ 与一 $\nabla\varphi(\mathbf{x}^k)$ 的方向一致，是 $\varphi(\mathbf{x})$ 在点 $\mathbf{x}^k$ 处的下降方向。

为了防止 $G(\mathbf{x}^k)$ 奇异或病态，可加一个阻尼项，即令

$$\hat{G}(\mathbf{x}^k)=G(\mathbf{x}^k)+\mu_k\mathbf{I} \tag{4.9}$$

当阻尼因子 $\mu_k>0$ 时，显然 $\hat{G}(\mathbf{x}^k)$ 是对称正定矩阵，从而

$$\mathbf{p}_k(\mu_k)=-\hat{G}(\mathbf{x}^k)^{-1}\nabla\varphi(\mathbf{x}^k) \tag{4.10}$$

是 $\varphi(\mathbf{x})$ 在点 $\mathbf{x}^k$ 处的下降方向。于是可构造迭代法

$$\mathbf{x}^{k+1}=\mathbf{x}^k-\hat{G}(\mathbf{x}^k)^{-1}\nabla\varphi(\mathbf{x}^k),\qquad k=0,1,... \tag{4.11}$$

称为阻尼最小二乘法，或称为 Levenbrery-Marquardt 法，它是高斯-牛顿法的改进。当 $\mu_k>0$ 时总可以保证 $\varphi(\mathbf{x}^{k+1})<\varphi(\mathbf{x}^k)$，因而 $\{\mathbf{x}^k\}$ 总是收敛的，但当 $\mu_k>0$ 太大时序列 $\{\mathbf{x}^k\}$ 收敛速度下降，而若 μ_k 太小则收敛域过小，初始近似 $\mathbf{x}^0$ 受限制，当然要选得合适的 μ_k 较为困难，但原则上 μ_k 应取小的正数，例如 $\mu_k=10^{-4}\sim10^{-2}$。通常可取 $\mu_0=10^{-2}$，然后通过计算加以调节，具体计算方案可以有不同设计。

算法（阻尼最小二乘法）：

步骤 1 置初始近似 $\mathbf{x}^0\in R^n$，误差限 $\varepsilon>0$，阻尼因子 μ_0（可取 $\mu_0=10^{-2}$），缩放常数 $\nu>1$（取 $\nu=2,5$ 或 10），$0\to k$；

步骤 2 计算 $F(\mathbf{x}^k),DF(\mathbf{x}^k),\nabla\varphi(\mathbf{x}^k)$ 及 $G(\mathbf{x}^k),\varphi(\mathbf{x}^k),0\to j$；

步骤 3 解方程组 $[G(\mathbf{x}^k)+\mu_k\mathbf{I}]\mathbf{p}_k=-\nabla\varphi(\mathbf{x}^k)$，求得 $\mathbf{p}_k(\mu_k)=\mathbf{p}_k$；

步骤 4 计算 $\mathbf{x}^{k+1}=\mathbf{x}^k+\mathrm{p}_k(\mu_k)$ 及 $\varphi(\mathbf{x}^{k+1})$；

步骤 5 若 $\varphi(\mathbf{x}^{k+1})<\varphi(\mathbf{x}^k)$ 且 $j=0$，则取 $\mu_k=\dfrac{\mu_k}{\nu},1\to j$，转步 3，否则 $j\neq0$ 转步 7；

步骤 6 若 $\varphi(\mathbf{x}^{k+1})\geq\varphi(\mathbf{x}^k)$，则取 $\mu_k=\nu\mu_k,1\to j$ 转步 3；

步骤 7 若满足 $\|p_k\|\leq\varepsilon$ 或其他收敛准则，则 $\mathbf{x}^{k+1}$ 为极小点 $\mathbf{x}^*$ 的近似，停止；否则将 $\mathbf{x}^{k+1}\to\mathbf{x}^k,k+1\to k$，转步 2.

4.3 非线性方程组求解方法小结

4.3.1 非线性方程组的 Newton 迭代解法

对于方程组 $F(\mathbf{x})=0$，存在如下的 Newton 迭代格式：

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-[F'(\mathbf{x}^{(k)})]^{-1}F(\mathbf{x}^{(k)})$$

4.3.2 Gauss-Newton 法的基本思想

对于残差泛函 $\varphi(\mathbf{x})=\frac{1}{2}F(\mathbf{x})^T F(\mathbf{x})$ 得到梯度项，$\nabla\varphi(\mathbf{x})=DF(\mathbf{x})^T F(\mathbf{x})$，用 $l_k(\mathbf{x})=F(\mathbf{x}^k)+DF(\mathbf{x}^k)(\mathbf{x}-\mathbf{x}^k)$ 代替原函数 $F(\mathbf{x})$，令 $\nabla\varphi(\mathbf{x})=0$ 得到法方程，导出 Gauss-Newton 算法。很显然，本质上，该方法是在局部点 $\mathbf{x}^k$ 处对函数 $F(\mathbf{x})$ 做了线性近似，忽略了 $F(\mathbf{x})$ 的二阶导数。

4.3.3 全 Newton 法的基本思想

然而，对于残差泛函 $\varphi(\mathbf{x})=\frac{1}{2}F(\mathbf{x})^T F(\mathbf{x})$，进行 Taylor 展开式为 $Q(\mathbf{x})=\varphi(\mathbf{x}^*)+\left\langle\frac{\partial\varphi}{\partial\mathbf{x}},\mathbf{s}\right\rangle+\frac{1}{2}\left\langle\frac{\partial^2\varphi}{\partial\mathbf{x}^2}\mathbf{s},\mathbf{s}\right\rangle$。若在 $\mathbf{x}=\mathbf{x}^*$ 附近，$\frac{\partial^2\varphi}{\partial\mathbf{x}^2}>0$，则 $Q(\mathbf{x})$ 存在极小点，用二次型 $Q(\mathbf{x})$ 在 $\mathbf{x}^*$ 附近逼近 $\varphi(\mathbf{x})$，对二次型函数 $Q(\mathbf{x})$ 求极小值，代替对 $\varphi(\mathbf{x})$ 求极小值，这是全 Newton 法解非线性最小二乘问题的基本思想。

令 $\frac{\partial Q}{\partial\mathbf{x}}=0$，可以导出迭代公式：$\mathbf{x}^{k+1}=\mathbf{x}^k-\left[\frac{\partial^2\varphi}{\partial\mathbf{x}^2}\right]^{-1}\frac{\partial\varphi}{\partial\mathbf{x}}$

全 Newton 法比 Gauss-Newton 法收敛性更好，关键是二阶导数 $\frac{\partial^2\varphi}{\partial\mathbf{x}^2}$ 的计算。

4.3.4 方法对比

对比非线性方程组的 Newton 法、最小二乘非线性方程组的 Gauss-Newton 法和全 Newton 法的迭代公式，发现它们其实存在统一性：

非线性方程组的 Newton 法：

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-\left[F\left(\mathbf{x}^{(k)}\right)\right]^{-1}F\left(\mathbf{x}^{(k)}\right)$$

最小二乘非线性方程组的 Gauss-Newton 法：

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-\left[\mathrm{D}F\left(\mathbf{x}^{(k)}\right)^T\mathrm{D}F\left(\mathbf{x}^{(k)}\right)\right]^{-1}\mathrm{D}F\left(\mathbf{x}^{(k)}\right)^T F\left(\mathbf{x}^{(k)}\right)$$

最小二乘非线性方程组的 Gauss-Newton 法和全 Newton 法

$$\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}-\left[\frac{\partial^2\varphi}{\partial\mathbf{x}^2}\right]^{-1}\frac{\partial\varphi}{\partial\mathbf{x}}=\mathbf{x}^{(k)}-\left[\frac{\partial^2\varphi}{\partial\mathbf{x}^2}\right]^{-1}DF\left(\mathbf{x}^{(k)}\right)^T F\left(\mathbf{x}^{(k)}\right)$$

可以看出，三种方法本质上都是在反传播/反投影数据的残差 $F(\mathbf{x})$。用数据残差的反传来修改反演的参数值是迭代法的本质思想。在此思想的指导下，从勘探地震学的角度，可以更好地进行偏移成像和反演成像。

但是，非线性方程组的 Newton 法在勘探地震中并非常用。基本问题是它认为数据是没有误差的，不符合勘探地震学对数据的认识。

5 求解线性方程组的预条件思想和与求解优化问题的正则化思想

从前面的讨论可知，非线性方程组的求解实际上是都利用了局部线性化的概念。迭代过程中都可以通过解一个大规模的、线性的方程组来实现非线性方程（组）的迭代求解。

另一方面，解线性方程的预条件方法已经研究得很多。非线性优化过程中的正则化也有很深入的研究。很显然，二者应该是有深刻的联系的。下面我们先讨论线性方程组的预条件求解的思想和方法。线性方程组的预条件求解方法有下述三个方面。

关于数据的预条件方法：

$$M^{-1}A\mathbf{x}=M^{-1}\mathbf{b} \tag{5.1}$$

（5.1）式本质上是在对数据进行变换，变换矩阵是 M^{-1}。可以从不同的角度解释变换矩阵 M^{-1} 的含义，可以认为它是数据协方差的逆，也可以认为选一个基函数特征表达数据，本质上揭开矩阵向量各分量之间的相关性，得到稀疏的特征表达。数学解释（5.1）式的预条件是从改善系数矩阵 A 的性质的角度展开的。若取 $M^{-1}\approx A$，则可以大幅改善系数矩阵 A 的特征值展布。这与对数据进行系数表达是联系在一起的。

关于模型的预条件方法：

$$AM^{-1}\mathbf{u}=\mathbf{b} \tag{5.2a}$$

$$\mathbf{x}=M^{-1}\mathbf{u} \tag{5.2b}$$

（5.2）式本质上是对要估计的参数 $\mathbf{x}$ 进行稀疏的特征表达，可以认为用矩阵 M 的列作为坐标系来表达参数 $\mathbf{x}$ 得到稀疏的特征表达 $\mathbf{u}$。这可以大大降低模型的参数个数。矩阵 M 可以认为是参数向量 $\mathbf{x}$ 各分量之间的协方差矩阵的逆，作用在参数向量 $\mathbf{x}$ 实现各分量的解相关得到 $\mathbf{u}$。数学解释（5.2）式的预条件是从改善系数矩阵 A 的性质的角度展开的。若取 $M^{-1}\approx A$，也可以大幅改善系数矩阵 A 的特征值展布。

关于数据和模型的同时预条件方法：

$$M_L^{-1}AM_R^{-1}\mathbf{u}=\mathbf{b} \tag{5.3a}$$

$$\mathbf{x}=M_R^{-1}\mathbf{u} \tag{5.3b}$$

很显然，（5.3）式本质上是要对数据向量和模型向量同时做（5.1）式和（5.2）式分别定义的预条件处理。把（5.3a）写成 $AM_R^{-1}\mathbf{u}=M_L\mathbf{b}$ 应该更清楚地看到（5.3）式的含义，就是引入变换矩阵 M_L 和 M_R 分别对数据和模型进行特征的稀疏表达，数学计算上是改变了系数矩阵 A 的性质。

与矩阵方程 SVD 解法的对比：

对矩阵方程 $A\mathbf{x}=\mathbf{b}$ 进行 SVD 分解，也可以看出预条件的本质看含义。系数矩阵 A 进行 SVD 分解后有 $A\mathbf{x}=\mathbf{b}\Rightarrow U\Sigma V^T\mathbf{x}=\mathbf{b}$。其中右奇异矩阵 U 和左奇异矩阵 V 都是单位正交阵。重写上述 SVD 分解为，

$$\Sigma V\mathbf{u}=U^T\mathbf{b} \tag{5.4a}$$

$$\mathbf{x}=V\mathbf{u} \tag{5.4b}$$

应该注意到，$U^T=U^{-1}$，$V^T=V^{-1}$。SVD 分解解法可以有效地避开小奇异值和零奇异值的影响，预条件处理的本质目的也是如此。

与 Tikhonov 正则化思想的对比：

Tikhonov 正则化思想是反演理论中提出的改善解的性质的方法。尽管它不仅限于线性方程组的求解，但通过线性方程组求解的对比分析，可以把它的本质思想解释得更为清楚。Bayes 框架下，正则化思想的提法可以描述为：

$$\sigma_M(\mathbf{m})=\frac{\rho_M(\mathbf{m})\rho_D\left(A(\mathbf{m})\right)}{\int_M d\mathbf{m}'\rho_M(\mathbf{m}')\rho_D\left(A(\mathbf{m}')\right)}=\frac{\exp\{-S(\mathbf{m})\}}{Const} \tag{5.5a}$$

和

$$S(\mathbf{m})=\frac{1}{2}\left(\left(A(\mathbf{m})-\mathbf{d}_0\right)^H C_{\mathbf{D}}^{-1}\left(A(\mathbf{m})-\mathbf{d}_0\right)+\left(\mathbf{m}-\mathbf{m}_0\right)^H C_{\mathbf{M}}^{-1}\left(\mathbf{m}-\mathbf{m}_0\right)\right) \tag{5.5b}$$

（5.5a）是 Bayes 估计公式；（5.5b）是误差泛函的定义式。二则是等价的。（5.5b）的线性解为，

$$\left(A^H C_D^{-1}A+C_M^{-1}\right)\Delta\mathbf{m}_0=A^H C_D^{-1}\left(\mathbf{d}-A\mathbf{m}_0\right) \tag{5.6}$$

上式可以简记为，

$$\left(A^H C_D^{-1}A+C_M^{-1}\right)\Delta\mathbf{m}_0=A^H C_D^{-1}\Delta\mathbf{d} \tag{5.7}$$

可以看出，协方差矩阵 C_D^{-1} 的作用就是（5.1）中 M^{-1} 的作用。C_M^{-1} 的作用在上述预条件中是没有对应形式的。但是它的作用很清楚是为了改变法方程的系数矩阵的性质，使之是正定的。若引入变换矩阵 M^{-1}，修改上式为：

$$\left(A^H C_D^{-1}A+C_M^{-1}\right)M^{-1}\delta\mathbf{u}=A^H C_D^{-1}\Delta\mathbf{d} \tag{5.8a}$$

$$\Delta\mathbf{m}=M^{-1}\delta\mathbf{u} \tag{5.9b}$$

这将是更完善的正则化方法。

另外，还可以修改（5.5b）式为，

$$S(\mathbf{m})=\frac{1}{2}\left(\left(A(\mathbf{m})-\mathbf{d}_0\right)^H C_{\mathbf{D}}^{-1}\left(A(\mathbf{m})-\mathbf{d}_0\right)+\left(R\left(\mathbf{m}-\mathbf{m}_0\right)\right)^H C_{\mathbf{M}}^{-1}\left(R\left(\mathbf{m}-\mathbf{m}_0\right)\right)\right) \tag{5.10}$$

矩阵 R 的选择方式有很多。考虑问题的出发点不同，选择的矩阵 R 也不同。用 Laplace 算子构成矩阵 R 是常用的选择，因为 Laplace 算子处理后的剩余参数向量 $\mathbf{m}-\mathbf{m}_0$ 更符合高斯分布。

此时，（5.10）式的线性解为，

$$\left(A^H C_D^{-1}A+R^H C_M^{-1}R\right)\Delta\mathbf{m}_0=A^H C_D^{-1}\Delta\mathbf{d} \tag{5.11}$$

若引入变换矩阵 M^{-1}，修改上式为，

$$\left(A^H C_D^{-1}A+R^H C_M^{-1}R\right)M^{-1}\delta\mathbf{u}=A^H C_D^{-1}\Delta\mathbf{d} \tag{5.12a}$$

$$\Delta\mathbf{m}=M^{-1}\delta\mathbf{u} \tag{5.12b}$$

这基本上是最完善的正则化方法。

6 总结与讨论

地震波反演成像涉及的问题很多，譬如地震波正演计算、反演框架、反演成像问题的数值求解、反演结果的评价等。当然，还有反演理论的勘探地震学理解与解释，包括从勘探地震学的角度发展出更面向实用的方法技术。

仅仅考虑反演问题的数据计算，其中核心问题主要是大规模线性方程组的数值解法。因为非线性问题的求解方法目前主要还是用逐步线性化的方法。预条件和正则化是另一个重要问题。

线性方程组直接求解方法，主要是 LU 分解方法无法用于大规模矩阵方程的求解。一是计算和存贮量大；而是勘探地震学中的线性方程组的系数矩阵不是满秩的。一般地，都是用迭代求解方法。Jacobi 迭代

方法、Gauss-Seidel迭代方法、逐次超松弛迭代方法尽管很简单，但它们是迭代法发展的基础。基本思想是实现了一个残差反投影。根本问题是反投影的方法过于简单。但是，这毕竟指明了残差反投影是求解线性方程的基本思想，与直接法解线性方程组的基本思想是矩阵的初等变换有异曲同工之妙。

把线性方程组的求解修改成一个二次型泛函的优化问题是线性方程组求解的另一经典思路。如此可以导出最速下降方法、共轭梯度方法、各种 Krylov 子空间方法。

另一种值得一提的方法就是 ART 方法和 SIRT 方法，这是直接用反投影思想构造出的线性方程组的求解方法，它们以存贮量小，计算简单而成为层析成像（油气是旅行时层析成像）中的经典算法。

根据不动点原理，非线性方程（组）的求解，在构造出满足压缩映射定理的迭代算子的情况下，可以得到非线性方程（组）求解的迭代格式，其中 Newton 方法是最基本构造迭代格式的方法。很显然，Newton 方法是一种局部线性化的方法。由于勘探地震学中，总认为数据是有误差的，其求解方法要建立在非线性最小二乘基础上。也可以从 Bayes 反演框架来构建最小二乘反演方法。目前勘探地震学中求解这类问题基本上用梯度类方法，误差泛函的梯度来导引误差总能量的逐步减弱。梯度的计算用共轭状态法是普遍的。本质上，梯度类方法也是一种逐步线性化方法。

我们认为勘探地震反演成像的数值计算的最本质思想有两点：线性化的思想和反投影及反传播的思想。

预条件和正则化是改善反演解的基本思想。尽管不能讲它们是完全等价的，但它们之间有非常密切的关系。从勘探地震学解释预条件和正则化，看到的是对数据和参数的稀疏特征表达。从数学的角度看，它们是对系数矩阵的特征值分布的修改，尽可能让系数矩阵是正定的。

从数据误差的概率分布和参数矢量（差）的概率分布来选择泛函定义是改善反演解的另一个重要的方面。如何从勘探地震学的角度更好地理解与实现上述理念是需要进一步考虑的事情。这是一个重要议题，需要专文讨论。

参考文献

[1]. 王华忠、王雄文，2012，全波形反演与波形层析成像的数值计算问题，波现象与反演成像研究组年度研究报告（2012），同济大学

[2]. 王华忠、王雄文，2012，地震波反演问题理论分析，波现象与反演成像研究组年度研究报告（2012），同济大学

[3]. Saad, Y., 2003, Iterative Methods for Sparse linear Systems (Second Edition), SIAM

[4]. Dennis Jr., J. E. and Schnabel, R. B., 1996, Numerical Methods for Unconstrained Optimization and Nonlinear Equations; SIAM

[5]. Vogel, C. R., 2002, Computational Methods for Inverse Problems; SIAM.

[6]. Nocedal, J., and Wright, S. J., 1999, Numerical Optimization; Springer Science + Business Meida, Inc.

[7]. 李庆扬、关治、白峰杉编著，2000，数值计算原理，清华大学出版社

[8]. Bertsekas, D. P., 2009, Convex Optimization Theory, Athena Scientific Press

面向岩性油气藏的地震波反演成像理论

王华忠，冯波，王雄文

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】勘探地震学中，地震波反演成像估计速度参数是核心问题。但是，Tarantola（2005）经典意义下的全波形反演在实际地震数据处理中遇到极大的困难。原因也是众所周知的：目前的波动方程正问题不能完全预测实际数据；影响地震波振幅的因素太多，波阻抗仅仅是其中一种，近地表因素的影响要很好地消除；地震子波未知（也是要反演的参数）；实测数据和速度之间的非线性性很强，反演问题构造的泛函的凸性不够；数据中的噪音分布未知。根据勘探地震学的速度估计的实践，速度估计问题的严格提法应在统计决策的框架下进行，这是由地震数据与反演参数之间的严重的非线性性和地震数据的随机性决定的。统计决策问题的解决主要用 Metropolis-Hastings 算法、Gibbs 采样器和 Monte Carlo 优化算法来解决。假如地震数据与反演参数之间的非线性性不是很强，可以用局部线性化的方法来解决参数估计问题。数值优化问题提供了各种梯度导引类的局部优化算法。但是，数学表达的梯度算法离开了地球物理的引导，无法较好地解决地震波反演成像估计速度参数这个核心问题，其中反演过程中的合适的正则化或预条件是关键问题。把经典意义的 FWI 导引到实际应用，需要考虑的核心问题是：（1）局部线性近似的梯度计算在线性近似不满足时，不能导引出收敛的算法。改变反问题的提法，构造凸性更好的问题。反射波层析成像就是一种很好的尝试。（2）振幅信息不能可靠地用于速度场的估计，需要退化到用相位信息来反演速度。（3）本质上，如何构造出较凸的泛函是更深层次的问题。特征波场表达、特征参数表达和特征波传播算子是更基础的问题。（4）通过正则化和/或预条件方法局限解的范围。实际上，还是从地球物理角度构造较凸的泛函（此时，数据和模型之间的关系更为线性化！）是更基本的问题。但是，由于 FWI 反演是利用全部波现象的强非线性反演，上述问题很难解决。我们认为**基于多级 Born 近似的逐级线性化反演是高精度参数反演**和高精度高分辨率成像的很好的路线。逐次利用明显的波现象进行线性反演，最终可以达到 FWI 反演的目标。岩性油气藏的精细描述在现代勘探地震过程中，是利用高维地震反演估计精细的背景参数变化和高保真高分辨的高波数参数变化，结合岩石物理知识，进行油气预测。**多级 Born 近似的逐级线性化反演**一定是着力发展的技术。它是一项综合性的、解释处理性的技术体系，主要包括初至波层析反演（基于到达时、相位和波形的层析成像）、最小二乘偏移成像和反射波层析成像（基于到达时、相位和波形的层析成像）。现代图像处理理论和方法提取参数变化的几何形态在该项综合技术中起着重要作用。

【关键词】反演成像理论，多级 Born 近似，全波形反演，逐级线性化反演

引言

地震波反演成像包含两个层次，即反射系数估计（或称真振幅成像，其简化形式为一般的叠前偏移成像，它仅仅定位反射系数出现的位置）；全波形反演（Full Waveform Inversion，简写为 FWI）估计速度和密度（也可估计各向异性参数和吸收衰减参数等，但多参数反演的难度更大）。

地震波反演的基本思想是 Bayes 估计理论。Tarantola（1984,2005）详细而深入地分析了地震波全波形反演的理论问题。在整个地震数据处理与反演成像过程中，去噪音、反褶积、数据规则化、一维波阻抗反演、AVA（叠前）弹性参数反演、叠前偏移成像、速度分析、层析成像等都是在此框架下进行的。根据 Bayes 估计理论，首先假设存在一个正演算子，它可以预测观测数据，预测噪音为高斯白噪，此时的 Bayes 估计可以在最小二乘意义下实现；在一定的范数意义下建立一个预测误差，数据误差一般用 L2 范数，模型约束一般用 L1 或 Cauchy 范数等；根据正算子的线性与否，选择用线性最小二乘方法或非线性最小二乘方法求解反演问题。

预测观测数据的正算子的作用十分重要。褶积算子用于反褶积；一维波动方程用于预测自激自收的地震道（波阻抗反演）；AR 模型算子常用于线性信号预测（压制噪音）；旅行时计算算子用于预测实测数据中的旅行时（层析成像）；Zoeppritz 方程及各种简化形式预测角度反射系数（AVA 弹性参数反演）；Fourier 变换算子用于预测规则无假频数据的谱；动校正时距关系预测实测 CMP 道集中的双曲时距关系（估计均方根

速度）；各种形式的波动方程预测实际观测的炮记录（全波形反演）。

反演问题是典型的不适定问题，包括解的存在性、唯一性和稳定性问题。在地球物理反演中，解的存在性是有物理保证的，主要是解的不唯一性和计算过程中的不稳定性问题。从最小二乘解法来看，二者具有关联性，计算的不稳定性更加剧了解的不唯一性，但最根本的还是解的不唯一性问题。

反演问题解的不唯一性本质上是由数据空间向模型空间映射的非线性性引起的，更进一步地，是由模型参数变化的复杂性和描述波场和参数场关系的正算子决定的。不唯一性只可以减弱，不可以完全消除。减弱反演解的不唯一性的关键是减弱由数据空间向模型空间映射的非线性性，以及增加关于反演解的先验信息。

20 世纪 80 年代，地震波反演处于热潮时期，很多学者对全波形反演理论做了十分扎实的奠基工作。但受当时野外地震观测技术和高性能计算机技术的局限，无法将提出的理论方法转为实用。此后，一维波阻抗反演、AVA 叠前弹性参数反演得到不断发展，但从理论上看，这些反演方法都不具备高维叠前反演的潜力，施加给介质分布情况的假设过多。

近几年，地震波高维叠前反演又成为勘探地球物理方面的核心议题，这首先得益于地震数据采集技术和大规模高性能计算机技术的进步，很多研究者对地震波高维叠前反演研究的兴起也做了大量铺垫工作（Pratt，1990）。

首先从概率论的观点说明了地震波反演的本质，指出在假设观测噪音为高斯白噪的情况下，Bayes 估计可以在最小二乘意义下实现；在此基础上，我们提出**多级 Born 近似的逐级线性反演**的方法技术。为了对比，把非线性的 FWI 反演方法进行了详细的分析。最后，我们指出**多级 Born 近似的逐级线性反演是岩性油气藏描述的下一代核心技术。**

1. 地震波反演的概率论观点

Bayes 估计的思想可以简单概括为：估计的结果 $\hat{\mathbf{m}}$ 是由后验概率密度函数 $P\left(\mathbf{m}/\mathbf{y}\right)$ 所决定的均值 $E\left\{\mathbf{m}/\mathbf{y}\right\}$，其中 $\mathbf{y}$ 代表观测数据。

$$\hat{\mathbf{m}} = E\left\{\mathbf{m}/\mathbf{y}\right\} = \int \mathbf{m}P\left(\mathbf{m}/\mathbf{y}\right)d\mathbf{m} \tag{1.1}$$

式（1.1）表示参数估计的结果是由观测数据所对应的模型参数出现的概率所决定的。

由于后验概率密度 $P\left(\mathbf{m}/\mathbf{y}\right)$ 很难估计，所以将其转化为较容易估计的模型空间的先验概率密度函数 $P\left(\mathbf{m}\right)$、数据空间的概率密度函数 $P\left(\mathbf{y}\right)$ 和条件概率密度函数 $P\left(\mathbf{y}/\mathbf{m}\right)$ 来表示，其中 $P\left(\mathbf{y}/\mathbf{m}\right)$ 体现了正算子的不确定性，描述了从模型空间到数据空间的映射关系。该 Bayes 公式为

$$P\left(\mathbf{m}/\mathbf{y}\right) = \frac{P\left(\mathbf{m}/\mathbf{y}\right)P\left(\mathbf{m}\right)}{P\left(\mathbf{y}\right)} \tag{1.2}$$

关于模型空间的概率密度函数 $P\left(\mathbf{m}\right)$，其先验信息就是模型参数的均值和方差，$P\left(\mathbf{m}\right)$ 的取值反映了对模型的认识情况。$P\left(\mathbf{m}\right)$ 的均值和方差在反演求解过程中主要起到约束解的取值范围的作用。如果 $P\left(\mathbf{m}\right)$ 某个分量的方差较小，代表由先验信息所决定的该分量的值的可靠性就较高，因此该分量的估计结果主要由先验信息所决定；如果 $P\left(\mathbf{m}\right)$ 某个分量的方差较大，代表由先验信息所决定的该分量的值的可靠性就较低，

因此该分量的估计结果主要由观测数据所决定。$P(\mathbf{m})$初步定义了反演问题的解在整个模型空间的概率密度函数，而反演问题的求解过程就是利用所观测到的数据更新这一概率密度函数的过程。

关于数据空间的概率密度函数$P(\mathbf{y})$，可采用以下方式定义其均值和方差。地震数据可认为是随机过程的一次具体实现，由于该随机过程只实现一次，所以可将该次实现看作该次随机过程的均值，即$P(\mathbf{y})$的均值由采集到的地震数据所决定。而$P(\mathbf{y})$的方差主要由地震数据的观测方式所决定，地震数据在某个空间范围内的采样密度越高，该空间范围内的地震数据所对应的方差就越大。如果假设数据空间的概率密度函数为高斯函数，则数据空间的概率密度函数可以用上述估计到的均值和方差来决定。

关于条件概率密度函数$P(\mathbf{y}/\mathbf{m})$，它表示在给定模型$\mathbf{m}$的条件下，采集到数据$\mathbf{y}$的可能性的大小。实际上，$P(\mathbf{y}/\mathbf{m})$也体现了用来描述实际物理现象的正算子的精确程度，当所用的正算子可以准确无误地描述实际物理过程时，$P(\mathbf{y}/\mathbf{m})$应该是一个δ类函数，否则它也应该有自己的均值和方差。因此，正演问题描述方法的差异会显著地影响$P(\mathbf{y}/\mathbf{m})$的取值。一般情况下，假设正算子在某些假设条件下可以比较准确地描述地震波的物理传播过程，并且其概率密度函数也是高斯函数。但实际情况并非如此，由于使正算子成立所要求的那些假设条件在实际情况中不一定都能得到满足，因此用正算子预测的数据与观测数据会存在较大的差异，并且正算子的概率密度函数也不一定是高斯函数，再加上先验信息的不充分，反演问题实际上是一个非常复杂的问题，如果不加上合理的约束条件，则很难直接得到一个具有地质意义的结果。

根据 Bayes 公式可以看出，$P(\mathbf{m}/\mathbf{y})$的均值和方差主要由$P(\mathbf{m})$，$P(\mathbf{y})$和$P(\mathbf{y}/\mathbf{m})$来决定。从这个角度看，对反演问题的求解可以认为是在$P(\mathbf{y})$，$P(\mathbf{m})$和$P(\mathbf{y}/\mathbf{m})$的约束下，使$P(\mathbf{m}/\mathbf{y})$最大化，并且期望均值为$\hat{\mathbf{m}}=\int \mathbf{m}P(\mathbf{m}/\mathbf{y})d\mathbf{m}$时，模型空间的方差最小。目前，常规反演问题的基本做法是假设$P(\mathbf{m})$和$P(\mathbf{y}/\mathbf{m})$均为高斯函数，在忽略观测系统不规则性的前提下，假设$P(\mathbf{y})$是标准的正态分布（当地震数据是规则采样时，其各个分量的方差应均相等），把$P(\mathbf{m}/\mathbf{y})$的最大化问题转化为代价函数$S(\mathbf{m})$的最小化问题。一般地，如果假设$P(\mathbf{m})$和$P(\mathbf{y})$均为高斯分布，并且忽略正算子的不确定性，则代价函数$S(\mathbf{m})$可以简单地定义为

$$S(\mathbf{m})=\frac{1}{2}\left[(g(\mathbf{m})-\mathbf{d}_{obs})^H C_D^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})+\lambda(\mathbf{m}-\mathbf{m}_{prior})^H C_M^{-1}(\mathbf{m}-\mathbf{m}_{prior})\right] \tag{1.3}$$

式中：C_D为数据协方差矩阵；C_M为模型的协方差矩阵。

如果正演问题$g(\mathbf{m})$是线性函数，即$g(\mathbf{m})=G\mathbf{m}$，并且矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）是正定的，则$S(\mathbf{m})$是一个正定二次函数，有唯一的极小值点。求解$S(\mathbf{m})$极小值点的收敛速度和解的稳定性取决于矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）条件数的大小，因此可以用 Newton 类的方法快速得到其最优解。如果矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）是不定的，即其特征值有正有负，则要先对$S(\mathbf{m})$做适当的正则化处理（在$S(\mathbf{m})$的构造过程中加入适当的约束条件），再求取其极小值点。

如果正演问题$g(\mathbf{m})$是非线性问题，但代价函数$S(\mathbf{m})$是个严格凸的函数，那么$S(\mathbf{m})$的极小值点也是唯一的。求解$S(\mathbf{m})$极小值点的收敛速度和解的稳定性，取决于$S(\mathbf{m})$的二阶导数在模型空间各个方向上的

比值的大小，此时各种最优化方法均可以用来求解其极小值点。如果初始值在极小值点附近，可以用一个二次函数来近似$S(\mathbf{m})$在该局部范围内的值，这时也可以用 Newton 类方法加快收敛速度。如果$S(\mathbf{m})$不是严格凸的函数，则$S(\mathbf{m})$的极小值点不再唯一，解决此类问题可以用信赖域方法、Monte Carlo 方法等。这些方法都可以用来求解非线性最优化问题，但由于 Monte Carlo 方法计算量大，如果没有一个比较好的策略则很难得以实现，因此信赖域方法在这种情况下是一个比较好的选择。信赖域方法的具体实现方式有很多种，其主要思想就是通过相关的先验信息来约束解的范围，从而降低解空间的大小，使$S(\mathbf{m})$在约束后的解空间里为一个严格凸的函数，再用相应的最优化方法即可求取其极小值点。

通常也可以把反演过程视为在所有可能的参数模型形成的模型空间（也是方程的解空间）中进行采样。反演结果对应的采样点应该是能够使式（3）达到极小值，或是使后验概率密度$P(\mathbf{m}/\mathbf{y})$达到最大值的采样点。基于上述用采样的观点来理解反演问题，可以认为，在确定性的观点下利用梯度导引类的方法求解非线性反演问题仅仅是在初始值附近的局部范围内对可能的模型参数进行采样。但勘探地震学中模型空间规模巨大，初始值虽很重要，但其选择也很困难。如果初始值的选择不能接近真实值，梯度导引类的算法将不能得到有意义的反演结果。另一方面也说明所有可能的参数模型依赖于建模过程中对参数所选择的描述方式，其会直接影响到模型空间的采样方式。因此，在正演过程的建模中要尽可能选择最佳的描述方式以使得最终的模型空间足够小。

基于此观点，Monte Carlo 采样反演方法就显得非常直观。与梯度导引类方法不同，Monte Carlo 反演方法是基于一定的概率密度函数在所有可能的参数模型所形成的模型空间中进行采样，而不是基于梯度导引类方法在初始值附近对模型空间进行采样。Monte Carlo 方法对模型空间的采样所依据的概率密度函数体现了对模型的认识，这种认识主要来源于模型空间的先验信息。与梯度导引类方法相比，Monte Carlo 方法对初始值的依赖性要小很多，但无论是 Monte Carlo 随机采样方法还是梯度导引类采样方法，都应该是在 Bayes 估计的思想框架下进行的。

因此，对反演问题最后结果的评价原则上应该是统计决策问题，主要评价依据是反演结果的均值和方差，其中方差的意义尤其重要，它反映了反演结果的可靠程度。

目前，勘探地震学中的反演成像基本上以梯度导引类的线性和非线性优化反演方法为主，下面专门讨论这两方面的做法。

2. 多级线性反演成像

2.1 标量散射理论的表达

在介质速度场$v(\mathbf{x})$中传播的标量场对应的齐次 Helmholtz 方程形式为，

$$\mathcal{L}_s(\mathbf{x},\omega)u(\mathbf{x},\omega)=\left(\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}\right)u(\mathbf{x},\omega)=0 \qquad (2.1)$$

引入背景介质速度场$v_0(\mathbf{x})$，上式可以重写为，

$$\begin{aligned}\mathcal{L}_s\left(\mathbf{x},\omega\right)u\left(\mathbf{x},\omega\right)&=\left(\nabla^2+\omega^2\left(\frac{1}{v_0^2\left(\mathbf{x}\right)}+\frac{1}{v_0^2\left(\mathbf{x}\right)}\left(\frac{v_0^2\left(\mathbf{x}\right)}{v^2\left(\mathbf{x}\right)}\text{-}1\right)\right)\right)u\left(\mathbf{x},\omega\right)\\&=\left(\nabla^2+\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}\right)u\left(\mathbf{x},\omega\right)+\frac{\omega^2a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}u\left(\mathbf{x},\omega\right)=0\end{aligned}\tag{2.2}$$

上式变为，

$$\left(\nabla^2+\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}\right)u\left(\mathbf{x},\omega\right)=-\frac{\omega^2a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}u\left(\mathbf{x},\omega\right)\tag{2.3}$$

一般地，把上式中的$\frac{\omega^2a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}$称为散射势。实质上，上式的右端项定义了新的非齐次 Helmholtz 方程的（二次）震源项，这是一个体积源。

物理上，可以把在$v\left(\mathbf{x}\right)$中传播的波$u\left(\mathbf{x},\omega\right)$分解成入射波场$u_0\left(\mathbf{x},\omega\right)$（它在背景介质中传播），和（2.3）中定义的散射势引起的散射场$u_s\left(\mathbf{x},\omega\right)$的叠加。如图 1 所示。

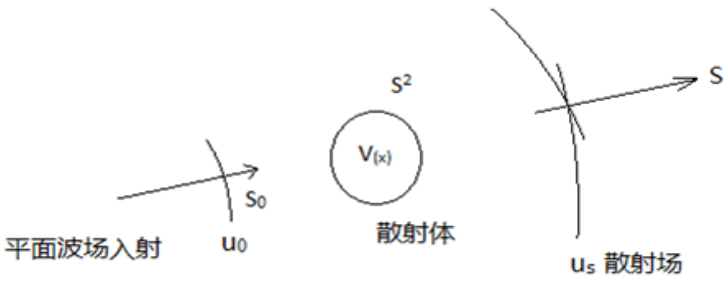


图 1 散射波传播关系

背景场中传播的波场，假设满足齐次 Helmholtz 方程

$$\left(\nabla^2+\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}\right)u_0\left(\mathbf{x},\omega\right)=0\tag{2.4}$$

（2.4）代入（2.3）可以得到散射势引起的散射场的表达式，

$$\left(\nabla^2+\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}\right)u_s\left(\mathbf{x},\omega\right)=-\frac{\omega^2a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}u\left(\mathbf{x},\omega\right)\tag{2.5}$$

引入 Green 函数，满足方程

$$\left(\nabla^2+\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)=-\delta\left(\mathbf{x}-\mathbf{x}'\right)\tag{2.6}$$

用$G\left(\mathbf{x}-\mathbf{x}',\omega\right)$乘（2.5）式；$u_s\left(\mathbf{x},\omega\right)$乘（2.6）式，结果相减，得到，

$$
\begin{aligned}
&u_s\left(\mathbf{x},\omega\right)\nabla^2 G\left(\mathbf{x}-\mathbf{x}',\omega\right)-G\left(\mathbf{x}-\mathbf{x}',\omega\right)\nabla^2 u_s\left(\mathbf{x},\omega\right)\\
&=\frac{\omega^2 a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}u\left(\mathbf{x},\omega\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)-u_s\left(\mathbf{x},\omega\right)\delta\left(\mathbf{x}-\mathbf{x}'\right)
\end{aligned}
\tag{2.7}
$$

如图所示，Green 函数和入射散射场的含义可以解释如下。利用 Green 定理，把左端项的体积分转化为面积分，得到，

$$
\begin{aligned}
&u_s\left(\mathbf{x},\omega\right)=\int_V \frac{\omega^2 a\left(\mathbf{x}'\right)}{v_0^2\left(\mathbf{x}'\right)}u\left(\mathbf{x}',\omega\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}\\
&+\oint_\Omega\left[u_s\left(\mathbf{x},\omega\right)\frac{\partial G\left(\mathbf{x}-\mathbf{x}',\omega\right)}{\partial n}-G\left(\mathbf{x}-\mathbf{x}',\omega\right)\frac{\partial u_s\left(\mathbf{x},\omega\right)}{\partial n}\right]
\end{aligned}
\tag{2.8}
$$

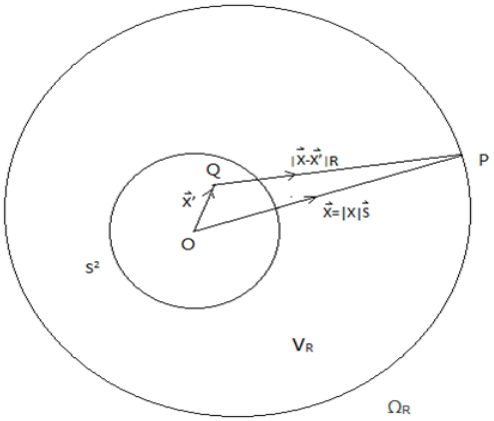


图 2 空间局部范围内的散射体的散射场关系

物理上讲，散射势定义的体积源是空间有限的，当计算距离该体积源较远的场时，（2.8）中的面积分项可以认为没有贡献。因为无论是 Green 函数还是散射场都是球面波场，远场情况下，忽略面积分符合索末菲辐射条件。

因此，散射势引起的散射场为

$$
\begin{aligned}
&u_s\left(\mathbf{x},\omega\right)=\int_V \frac{\omega^2 a\left(\mathbf{x}'\right)}{v_0^2\left(\mathbf{x}'\right)}u\left(\mathbf{x}',\omega\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'\\
&=\omega^2\int_V u\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'
\end{aligned}
\tag{2.9}
$$

其中，记 $\delta\mathbf{m}\left(\mathbf{x}\right)=a\left(\mathbf{x}\right)\Big/v_0^2\left(\mathbf{x}\right)$ 为速度扰动量。下面所讨论的利用 Born 近似的反演结果就是 $\delta\mathbf{m}\left(\mathbf{x}\right)$。

把（2.9）中的总场 $u\left(\mathbf{x},\omega\right)$ 用背景场 $u_0\left(\mathbf{x},\omega\right)$ 代替，则（2.9）式变为，

$$
u_s^1\left(\mathbf{x},\omega\right)\approx\omega^2\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'
\tag{2.10}
$$

此为一阶 Born 近似后的散射场。

接着也有如下形式的一阶 Born 近似后的总场表达式，

$$u^1(\mathbf{x},\omega)\approx u_0(\mathbf{x},\omega)+u_s^1(\mathbf{x},\omega)=u_0(\mathbf{x},\omega)+\omega^2\int_V u_0(\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}')G(\mathbf{x}-\mathbf{x}',\omega)d\mathbf{x}' \tag{2.11}$$

这样的推导过程更清楚地体现了散射势所引起的散射场的物理本质。也清楚地说明了一阶 Born 近似的物理含义。但是，上述一阶 Born 近似只有在背景速度 $v_0(\mathbf{x})$ 接近真速度 $v(\mathbf{x})$，也就是介质参数小扰动的情况下，是成立的；或计算出的散射场和总场是准确的。

2.2 多级 Born 近似

由于上述一阶 Born 近似仅在小扰动的情况下计算出的散射场和总场比较准，应该列出更高阶的 Born 近似，在小扰动条件不很成立时，也可以得到较高精度的散射场和总场的计算结果。实际上，二阶 Born 近似以及更高阶的 Born 近似对应了二次散射和更高次的散射现象。

二阶 Born 近似公式的导出方法为把（2.11）式代入（2.9）式，用一阶 Born 近似下的总场，取代其中的准确的总场，得到二阶 Born 近似下的散射场和总场，

$$\begin{aligned}
&u_s^2(\mathbf{x},\omega)\approx\omega^2\int_V u^1(\mathbf{x}'',\omega)\delta\mathbf{m}(\mathbf{x}'')G(\mathbf{x}-\mathbf{x}'',\omega)d\mathbf{x}''\\
&=\omega^2\int_V\left[u_0(\mathbf{x}',\omega)+\omega^2\int_V u_0(\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}')G(\mathbf{x}''-\mathbf{x}',\omega)d\mathbf{x}'\right]\delta\mathbf{m}(\mathbf{x}'')G(\mathbf{x}-\mathbf{x}'',\omega)d\mathbf{x}''\\
&=\omega^2\int_V u_0(\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}')G(\mathbf{x}-\mathbf{x}',\omega)d\mathbf{x}'\\
&+\omega^4\int_V\int_V u_0(\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}')G(\mathbf{x}''-\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}'')G(\mathbf{x}-\mathbf{x}'',\omega)d\mathbf{x}'d\mathbf{x}''
\end{aligned} \tag{2.12a}$$

$$\begin{aligned}
&u^2(\mathbf{x},\omega)\approx u_0(\mathbf{x},\omega)+u_s^2(\mathbf{x},\omega)\\
&=u_0(\mathbf{x},\omega)+\omega^2\int_V u_0(\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}')G(\mathbf{x}-\mathbf{x}',\omega)d\mathbf{x}'\\
&+\omega^4\int_V\int_V u_0(\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}')G(\mathbf{x}''-\mathbf{x}',\omega)\delta\mathbf{m}(\mathbf{x}'')G(\mathbf{x}-\mathbf{x}'',\omega)d\mathbf{x}'d\mathbf{x}''
\end{aligned} \tag{2.12b}$$

很显然，二阶 Born 近似下散射场和总场的精度要高于一阶 Born 近似下计算出的散射场和总场的精度。

分析知，（2.12b）式左端第一项代表背景、无散射介质中传播的波场，它从源点直接到接收点。它应该仅仅包括直达波和 Diving_Wave。频率越低、炮检距越大，观测到的 Diving_wave 影响深度越大。实际上，直达波也可以认为是 Diving_Wave 的一部分，低频时它的影响深度也很大。射线理论下，直达波仅仅在地表传播，地表水平时，这样的直达波不可以用来进行层析成像。波动理论看，情况并非如此。因此，波动理论的直达波层析成像与射线理论的直达波层析成像对观测系统的要求是不太相同的。折射波如何解释？事实上，Born 近似根本不打算解释折射波，它在 Born 近似公式中无明确的位置！也许把折射波隐含在多次散射波中。左端第二项代表一次散射波场，背景波场作用于散射体产生的散射场直接传到接收点得到的波场。明显地，**这是一般的偏移成像所处理的波场**。原则上，Claerbout 提出的入射波到达时等于反射波出发时的成像条件是一个含糊的说法，并没有指明是针对一次反射波（或绕射波）的。但是，实际情况下，我们仅仅能对一次反射波进行偏移成像。关键问题是如果对多次反射波或绕射波进行偏移成像必须要知道反射界面的准确位置和正确的背景速度，这基本上等于知道了精确的速度场，偏移成像已经没有必要了。这也解释了为什么 RTM 具有对多次波成像的能力，却不能对实测数据中的多次波进行准确成像的原因。就是因为我们不可能准确估计出严格满足 RTM 要求的速度场。左端第三项代表二次散射波场。相对于所关注的一个散射点而言，它影响的范围是有限的，就是 Fresnel 带，带内的二次源相互作用是加强的。这是波动理论对二次散射的基本解释。也是用二次散射（反射）波场进行反射波层析的物理基础。实际上，上述解释对复杂的散射体几何展布情况存在很多变化，但是左端第三项把所有可能的情况全都描述清楚了。只不

过我们很多时候仅仅关注但反射界面情形而已。

类似地，可以导出三阶 Born 近似的散射场和总场表达式，

$$
\begin{aligned}
&u_s^3\left(\mathbf{x},\omega\right)\approx\omega^2\int_V u^2\left(\mathbf{x}''',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'''\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}'''\\
&=\omega^2\int_V\left[\begin{array}{l}u_0\left(\mathbf{x}',\omega\right)+\omega^2\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'\\+\omega^4\int_V\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}''\end{array}\right]\delta\mathbf{m}\left(\mathbf{x}'''\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}'''\\
&=\omega^2\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'\\
&+\omega^4\int_V\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}''\\
&+\omega^6\int_V\int_V\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}'''-\mathbf{x}'',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'''\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}'d\mathbf{x}''d\mathbf{x}'''
\end{aligned}
\tag{2.13a}
$$

$$
\begin{aligned}
&u\left(\mathbf{x},\omega\right)\approx u_0\left(\mathbf{x},\omega\right)+u_s^3\left(\mathbf{x},\omega\right)\\
&=u_0\left(\mathbf{x},\omega\right)+\omega^2\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'\\
&+\omega^4\int_V\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}''\\
&+\omega^6\int_V\int_V\int_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}'''-\mathbf{x}'',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'''\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}'d\mathbf{x}''d\mathbf{x}'''
\end{aligned}
\tag{2.13b}
$$

更高阶的 Born 近似可以以此类推地导出。上述推导过程清楚地表达了各阶 Born 近似的物理含义。（2.13b）左端第 4 项的具体含义与第三项的含义是类似的，只不过更复杂而已。从数学上看，这是把（2.9）表达的计算变速介质下的波场的非线性问题转化成一系列线性问题的组合。数学化的表达是把上述过程描述为一个级数展开问题。我们从上述角度表达是为了清楚地说明 Born 级数的物理含义。这样可以更方便地理解利用各阶散射波进行速度反演的物理含义。

2.3 多级 Born 近似下的反演问题的梯度计算

上述各阶 Born 近似表达的散射场和总场的计算公式典型了地震波反演成像的基础。与 Bayes 框架下地震波反演理论结合，可以发展出一般叠前深度偏移成像理论、最小二乘叠前深度偏移（反演）成像理论、多次波成像理论、透射波层析成像理论、反射波层析成像理论、最终实现 FWI 反演成像。

下面我们以二阶 Born 近似下波场计算理论为基础，与 Bayes 框架下的反演理论结合，导出各种叠前偏移（反演）成像方法。此时我们忽略正则化思想，假定地震数据的预测误差满足高斯分布，可以定义如下的 L2 范数下的误差泛函

$$
\mathcal{J}\left(\delta\mathbf{m}\right)=\frac{1}{2}\left\|u^2\left(\mathbf{x},\omega\right)-u^{obs}\left(\mathbf{x},\omega\right)\right\|^2 \tag{2.14}
$$

把（2.12b）式定义的二阶 Born 近似下总场代入（2.14）式，得到，

$$
\begin{aligned}
\mathcal{J}\left(\delta\mathbf{m}\right) &= \frac{1}{2}\left\|u^2\left(\mathbf{x},\omega\right)-u^{obs}\left(\mathbf{x},\omega\right)\right\|^2 \\
&= \frac{1}{2}\left\|\begin{array}{l} u_0\left(\mathbf{x},\omega\right)+\omega^2\int\limits_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}' \\ +\omega^4\int\limits_V\int\limits_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}''-u^{obs}\left(\mathbf{x},\omega\right)\end{array}\right\|^2 \\
&= \frac{1}{2}\left\|\begin{array}{l} \omega^2\int\limits_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}' \\ +\omega^4\int\limits_V\int\limits_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}'' \\ -\left(u^{obs}\left(\mathbf{x},\omega\right)-u_0\left(\mathbf{x},\omega\right)\right)\end{array}\right\|^2
\end{aligned}
\tag{2.15}
$$

上式可以重写为，

$$
\mathcal{J}\left(\delta\mathbf{m}\right) = \frac{1}{2}\left\|\begin{array}{l} \omega^2\int\limits_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}' \\ +\omega^4\int\limits_V\int\limits_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}'' \\ -\delta u^{obs}\left(\mathbf{x},\omega\right)\end{array}\right\|^2 \tag{2.16}
$$

其中，$\delta u^{obs}\left(\mathbf{x},\omega\right)=\left\|u^{obs}\left(\mathbf{x},\omega\right)-u_0\left(\mathbf{x},\omega\right)\right\|$代表去除背景介质中传播的波的剩余波场。值得提及的是，利用$\delta u^{obs}\left(\mathbf{x},\omega\right)=\left\|u^{obs}\left(\mathbf{x},\omega\right)-u_0\left(\mathbf{x},\omega\right)\right\|$能量最小化建立泛函可以导出背景速度扰动量的估计。背景速度本身的估计是需要单独讨论的问题，譬如 CMP 叠加速度分析就是在估计背景速度本身，不过这样的方法过于粗糙而已。

该部分波场$\delta u^{obs}\left(\mathbf{x},\omega\right)$还要减去上式中的散射波场，若残差能量达到极小，就认为此时估计的参数$\delta\mathbf{m}$是我们期望的。在一般的叠前深度偏移成像中，要求$\delta u^{obs}\left(\mathbf{x},\omega\right)$中仅仅包含一次绕射波（反射波）。这是常规叠前深度偏移成像时特别要注意的。

在此，需要界定速度模型的表达方法。否者，可能引起混乱。但是这样的分解方法也仅仅是我们的观点而已。根据波场与速度反演的关系给出如下的分解：

①、背景场$u_0\left(\mathbf{x},\omega\right)\Rightarrow$背景速度估计$\mathbf{m}_0\left(\mathbf{x}'\right)$或背景速度扰动估计$\delta\mathbf{m}_0\left(\mathbf{x}'\right)$

②、一次散射场$\omega^2\int\limits_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'\Rightarrow$反射系数$R\left(\theta\right)$

③、二次散射波场$\omega^4\int\limits_V\int\limits_V u_0\left(\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}''$
$\Rightarrow$反射波层析速度扰动量$\delta\mathbf{m}\left(\mathbf{x}'\right)$

总体而言，估计参数$\delta\mathbf{m}$（就是反演成像）的基本的思想是用梯度导引的优化算法求解（2.16），核心问题是梯度的计算。下面对（2.16）推导相应的梯度$\partial\mathcal{J}\big/\partial\delta\mathbf{m}$。

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \delta \mathbf{m}} &= \frac{\partial}{\partial \delta \mathbf{m}}\left\{\frac{1}{2}\left\|\begin{array}{l}\omega^2\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}' \\ +\omega^4\int_V\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}'' \\ -\delta u^{obs}\left(\mathbf{x},\omega\right)\end{array}\right\|^2\right\} \\ &= \frac{\partial}{\partial \delta \mathbf{m}}\left\{\frac{1}{2}\left\|u_s^2\left(\mathbf{x},\omega\right)-\delta u^{obs}\left(\mathbf{x},\omega\right)\right\|^2\right\} \\ &= \frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial \delta \mathbf{m}}\delta\mathbf{d}\end{aligned} \tag{2.17}$$

其中，$\delta\mathbf{d}=u_s^2\left(\mathbf{x},\omega\right)-\delta u^{obs}\left(\mathbf{x},\omega\right)$。

$$\begin{aligned}\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial \delta \mathbf{m}} &= \omega^2 u_0\left(\mathbf{x},\omega\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right) \\ &+\omega^4\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}' \\ &+\omega^4\int_V u_0\left(\mathbf{x},\omega\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}''\end{aligned} \tag{2.18}$$

分析可知，上述梯度有两部分组成，一部分是

$$\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial \delta \mathbf{m}}\right|_T \delta\mathbf{d}=\left[\omega^2 u_0\left(x,\omega\right)G\left(x\text{ - }x',\omega\right)\right]\delta\mathbf{d}\,, \tag{2.19a}$$

第二部分是

$$\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial \delta \mathbf{m}}\right|_R \delta\mathbf{d}=\left[\begin{array}{l}\omega^4\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}' \\ +\omega^4\int_V u_0\left(\mathbf{x},\omega\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}''\end{array}\right]\delta\mathbf{d} \tag{2.19b}$$

梯度项第一部分的计算不需要已知模型中高波数成分；梯度项第二部分的计算需要已知模型中的高波数成分（包括它的位置和数量的正确！）。这两个梯度的量级是不同的，一次散射波梯度的量级显然比二次散射波梯度的量级大很多。因此，做好不要把两个梯度加在一起进行速度更新。而是串联迭代进行。已知，单独计算二次散射波的梯度需要付出很大的计算量。

很显然，用第一部分的梯度进行参数 $\delta\mathbf{m}$ 的估计，对反射波而言，得到的是最小二乘叠前深度偏移成像的结果。

$$I_{PSDM}\left(\mathbf{x}\right)=\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial \delta \mathbf{m}}\right|_T \delta\mathbf{d} \tag{2.20}$$

在各级 Born 近似的表达式中，透射波场 $u_0\left(\mathbf{x},\omega\right)$ 代表的含义是光滑的背景介质中传播的波场，这样的背景介质中，地震频带内的波传播不产生散射。它就是从炮点到达检波点的直达波和 Diving-Wave。折射波都不包含在内。

很显然，用 $u_0\left(\mathbf{x},\omega\right)$ 进行层析反演，目标是得到光滑的背景介质速度。对于地表观测的勘探地震数据而言，用 $u_0\left(\mathbf{x},\omega\right)$ 进行层析反演主要是估计地表附近的背景速度，这还与参与反演的波场的频率有关，低频的

Diving Wave，波路径影响的范围很深，利用低频 Diving-wave，也可以反演距地表较深的地层的背景速度。

用 $u_0(\mathbf{x},\omega)$ 进行层析反演借助如下的公式。首先，对 $u_0(\mathbf{x},\omega)$ 进行 Taylor 展开，并进行线性化，即

$$u_0(\mathbf{x},\omega;\mathbf{m}_0+\delta\mathbf{m}_0)\approx u_0(\mathbf{x},\omega;\mathbf{m}_0)+\frac{\partial u_0(\mathbf{x},\omega;\mathbf{m}_0)}{\partial\delta\mathbf{m}_0} \tag{2.21a}$$

$$\delta u_0(\mathbf{x},\omega;\mathbf{m}_0,\delta\mathbf{m}_0)=u_0(\mathbf{x},\omega;\mathbf{m}_0+\delta\mathbf{m}_0)-u_0(\mathbf{x},\omega;\mathbf{m}_0)\approx\frac{\partial u_0(\mathbf{x},\omega;\mathbf{m}_0)}{\partial\delta\mathbf{m}_0}\delta\mathbf{m}_0 \tag{2.21b}$$

其中，梯度项 $\frac{\partial u_0(\mathbf{x},\omega;\mathbf{m}_0)}{\partial\delta\mathbf{m}_0}$ 定义了透射波层析成像的核函数。

反射波梯度函数的响应如图 3a 所示，透射波梯度函数的响应如图 3b 所示。很显然，透射波的梯度函数的响应就是其波路径，反射波的梯度函数的响应就是偏移算子的脉冲相应。

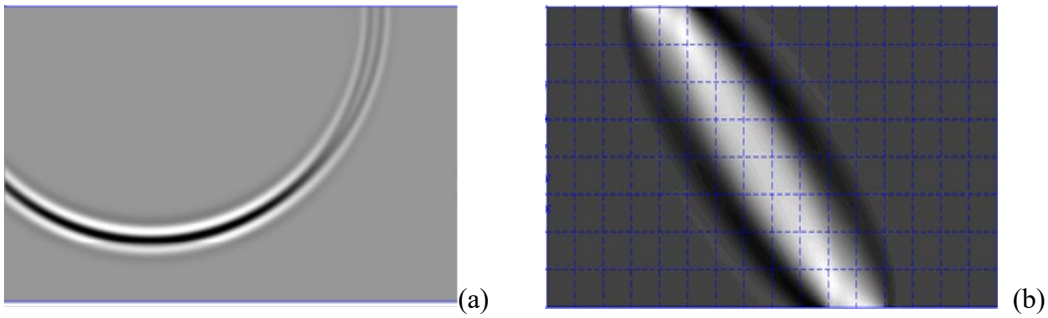

(a) (b)

图 3 （a）双层速度模型中，反射波对目标泛函梯度的贡献；（b）常速背景模型中，透射波对目标泛函梯度的贡献；

（2.20）式所做的是对中深层的反射波进行真振幅叠前深度偏移成像，得到角度反射系数；（2.21）利用透射波（一般在浅层传播，深层折射波出现在很大偏移距的地方！）进行反演成像，估计表层的速度，建立表层模型。

（2.19b）式中的第一部分代表背景入射场 $u_0(\mathbf{x},\omega)$ 入射到地下一个散射点，产生二次场，二次场对地下所有其它点激励后产生的场的变化；第二部分代表从源点出发的 Green 函数激励地下一个散射点，产生二次场，二次场对地下所有其它点激励后产生的场的变化。这两部分的叠加产生完整的梯度量。该梯度响应如图 4 所示。可以看出，该式进行的是反射波层析成像。利用二次散射波计算的积分核沿波路径进行残差波场的反传播。

$$\delta\mathbf{m}(\mathbf{x})=\left.\frac{\partial u_s^2(\mathbf{x},\omega)}{\partial\delta\mathbf{m}}\right|_R\delta\mathbf{d} \tag{2.22}$$

可以看出，（2.21）、（2.20）和（2.22）串联在一起可以进行初至波层析反演、最小二乘叠前深度偏移成像和反射波层析反演。

这三个公式是现代勘探地震学中地震波反演成像的核心公式。绝大部分的方法技术就是在此基础上的展开和应用研究。

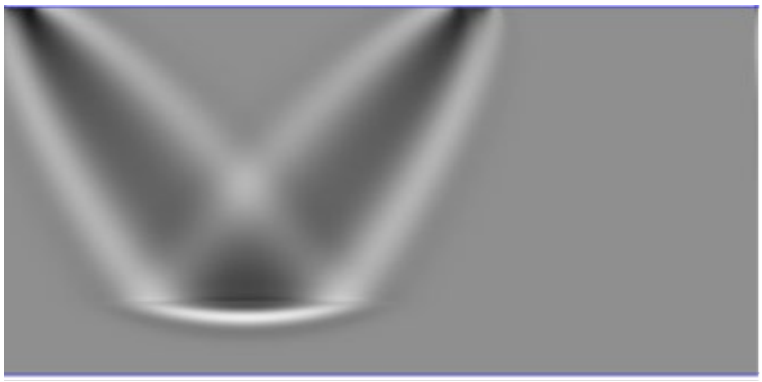


图 4 双层速度模型中，二次散射效应（即反射波）对目标泛函梯度函数的贡献

为了进一步发展起见，我们列出三阶 Born 近似的地震波反演成像公式体系。

$$\begin{aligned}\mathcal{J}\left(\delta\mathbf{m}\right)&=\frac{1}{2}\left\|u_s^3\left(\mathbf{x},\omega\right)-\delta u^{obs}\left(\mathbf{x},\omega\right)\right\|^2\\&=\frac{1}{2}\left\|\begin{array}{l}\omega^2\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)d\mathbf{x}'\\+\omega^4\int_V\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'d\mathbf{x}''\\+\omega^6\int_V\int_V\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}'''-\mathbf{x}'',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'''\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}'d\mathbf{x}''d\mathbf{x}'''\\-\delta u^{obs}\left(\mathbf{x},\omega\right)\end{array}\right\|^2\end{aligned}$$

（2.23a）

泛函的梯度为，

$$\frac{\partial\mathcal{J}}{\partial\delta\mathbf{m}}=\frac{\partial u_s^3\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}\delta\mathbf{d}\tag{2.23b}$$

$$\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}=\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}\right|_T+\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}\right|_{R^2}+\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}\right|_{R^3}\tag{2.23c}$$

其中，

$$\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}\right|_T=\omega^2u_0\left(\mathbf{x},\omega\right)G\left(\mathbf{x}-\mathbf{x}',\omega\right)\tag{2.24a}$$

$$\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}\right|_{R^2}=\omega^4\left\{\begin{array}{l}\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}'\\+\int_V u_0\left(\mathbf{x},\omega\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}-\mathbf{x}'',\omega\right)d\mathbf{x}''\end{array}\right\}\tag{2.24b}$$

$$\left.\frac{\partial u_s^2\left(\mathbf{x},\omega\right)}{\partial\delta\mathbf{m}}\right|_{R^3}=\omega^6\left\{\begin{array}{l}\int_V\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}'''-\mathbf{x}'',\omega\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}'d\mathbf{x}''\\+\int_V\int_V u_0\left(\mathbf{x},\omega\right)\delta\mathbf{m}\left(\mathbf{x}'\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)G\left(\mathbf{x}'''-\mathbf{x}'',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'''\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}'d\mathbf{x}'''\\+\int_V\int_V u_0\left(\mathbf{x},\omega\right)G\left(\mathbf{x}''-\mathbf{x}',\omega\right)\delta\mathbf{m}\left(\mathbf{x}''\right)G\left(\mathbf{x}'''-\mathbf{x}'',\omega\right)\delta\mathbf{m}\left(\mathbf{x}'''\right)G\left(\mathbf{x}-\mathbf{x}''',\omega\right)d\mathbf{x}''d\mathbf{x}'''\end{array}\right\}$$

（2.24b）

（2.23）和（2.24）中各量的地球物理含义及解释与二阶 Born 近似是类似的，不再赘述。

但是，值得指明的是，在某种意义上，用 Born 近似解释反射波层析成像是不合适的。现在，Born 近似理论在勘探地震中应用很广，但它的几何物理图景并不特别清楚，尤其用它来解释层间多次多次反射（散射）波时。它与波动理论的波传播之间的关系（譬如 Fresnel 体）是模糊的，需要进一步解释清楚。

3. 全波形（FWI）反演成像及其中的梯度计算

3.1 地震全波形反演中正问题的简洁表达

非线性正问题的数学描述可以抽象为

$$A(\mathbf{m})\mathbf{u} = -\delta(\mathbf{x}-\mathbf{x}_s)\mathbf{f} = -\mathbf{f}(\mathbf{x}_s,t) \tag{3.1}$$

其中，A 为正演算子。在地震波反演中，A 的形式为

$$A = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right) = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \Delta \tag{3.2}$$

A 的形式受所选择的波动方程的形式的影响。可以选常密度声波，变密度声波方程，各种拟声波方程，弹性波方程等。$\mathbf{u}$ 为要正演合成的波场，$\mathbf{f}$ 为震源函数。

野外观测到的地震数据可以表述为

$$\mathbf{d} = C\mathbf{u} + \boldsymbol{\eta} \tag{3.3}$$

此处，C 表示地震数据的采样算子，由观测系统决定。它使得连续的波场 $u(x,y,z,t)$ 变得规则离散化（或非规则离散化）及孔径有限化。采样算子的状态受实际条件的制约。η 表示噪音函数，在大部分情况下假设它是高斯白噪的。

连续波场受采样算子制约后，采样后的数据可以记为

$$[C\mathbf{u}] \approx \mathbf{u}(\mathbf{x}_i), \quad i = 1,2,\cdots,n \tag{3.4}$$

其中 $\mathbf{x}_i$ 表示第 i 个采样点的坐标。根据方程（3.1），可以将连续波场的正演计算形式地表达为

$$\mathbf{u} = \left(A(m)\right)^{-1}\mathbf{f} \tag{3.5}$$

采样后的离散地震数据表示为，

$$F(\mathbf{m}) = C\left(A(\mathbf{m})\right)^{-1}\mathbf{f} \tag{3.6}$$

3.2 全波形（FWI）反演及梯度计算

3.2.1 非线性全波形反演（FWI）基本问题概述

基于前面对正问题的定义，根据数据残差最小的原则可以建立如下的优化问题：

$$\begin{aligned} &\min: \quad J(\mathbf{m}) \\ &S.T.: \quad C\mathbf{u} = \mathbf{d} \end{aligned} \tag{3.7}$$

可以用如下的方式定义上面优化问题所对应的误差泛函，

$$\min_{\mathbf{u}\in\mathcal{H}_2,\mathbf{m}\in\mathcal{H}_1}\frac{1}{2}\left\|C\mathbf{u}-\mathbf{d}\right\|_Y^2+\alpha J\left(\mathbf{m}\right) \tag{3.8}$$

上式中的第一项代表正演波场与观测波场的波场残差，后一项代表对模型的先验约束，α 是正则化系数。一般地，我们可以通过假设模型中的元素满足某种概率分布条件来施加对模型的先验限制。
将（3.5）代入上式，得到泛函定义式，

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\mathfrak{I}\left(\mathbf{m}\right)=\frac{1}{2}\left\|C\left(A\left(\mathbf{m}\right)\right)^{-1}\mathbf{f}-\mathbf{d}\right\|_Y^2+\alpha J\left(\mathbf{m}\right) \tag{3.9}$$

通过此式，可以用各种优化方法求解参数估计问题。在无约束非线性优化理论中，（3.9）式的求解有很多不同的方法，譬如最速下降方法、全 Newton 方法、Gauss-Newton 方法、拟 Newton 方法等等。一般地，（3.9）式对应的的迭代求解格式可以写为

$$\mathbf{m}^{k+1}=\mathbf{m}^k-\mu_k\mathbf{g}_k \tag{3.10}$$

其中，μ_k 为第 k 个迭代步的步长，$\mathbf{g}_k$ 为第 k 个迭代步的梯度。在不同的迭代方法中，$\mathbf{g}_k$ 也可以认为是预条件修正后的梯度方向。步长的选择对迭代法的收敛性和收敛效率都影响甚大，具体地参考 Dennis 和 Schnabel（1996）以及 Nocedal 和 Wright（1999）。

（3.10）式中最核心的是梯度向量的计算。

3.2.2 共轭状态法梯度计算

3.2.2.1 非线性问题的梯度计算问题

假设泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 处是 Fréchet 可微的，由 Reisz 表示定理，存在梯度 $\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)\in\mathcal{H}$，称为泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 点的梯度，满足

$$\mathfrak{I}'\left(\mathbf{m}\right)\delta\mathbf{m}=\left\langle\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right),\delta\mathbf{m}\right\rangle_{\mathcal{H}} \tag{3.11}$$

若泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 处是 Fréchet 可微的，则对任意的 $\delta\mathbf{m}\in\mathcal{H}$，则关于变量 τ 的映射 $\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right):\mathbb{R}\mapsto\mathbb{R}$ 在 $\tau=0$ 处也是可微的，

$$\frac{d}{d\tau}\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right)\Big|_{\tau=0}=\left\langle\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right),\delta\mathbf{m}\right\rangle_{\mathcal{H}} \tag{3.12}$$

因此，我们可以从上述 Gateaux 导数的定义计算泛函的微分。

3.2.2.2 Jacobi 方法

计算非线性问题梯度的最简单方法是利用范函的差分（差商）值来代替微分值。该方法通过差分的方式计算范函在任意点 $\mathbf{m}$ 处的梯度值，其精度取决于差分的步长，表达式如下，

$$\left[\mathrm{Grad}\mathfrak{I}_n\left(\mathbf{m}\right)\right]_i=\frac{\mathfrak{I}_n\left(\mathbf{m}+\tau_i\mathbf{e}_i\right)-\mathfrak{I}_n\left(\mathbf{m}\right)}{\tau_i}+O\left(\tau_i^2\right),\quad i=1,2,\cdots,n \tag{3.13}$$

上面的式子表示范函沿任意一个模型空间方向 $\mathbf{e}_i$ 的导数值可以表示为一个差分的值加上一个二阶误差项，其中 $\mathbf{e}_i$ 表示模型空间中的第 i 个单位向量，n 表示模型空间的维数。所有导数值形成一个向量，构成了范函在点 $\mathbf{m}$ 处的梯度值。

3.2.2.3 Adjoint-State 方法（伴随状态方法）

已知算子 $A(\mathbf{m})=\dfrac{1}{v^2(\mathbf{x})}\dfrac{\partial^2}{\partial t^2}-\Delta$，它可以理解为一个投影算子，建立了模型空间和数据空间之间的映射关系。$\partial A/\partial\mathbf{m}$ 表示算子的微分，它定义了一个 Jacobi 矩阵。在一定的定义下，它是一个 Fréchet 微商。假设算子 $A(\mathbf{m})$ 存在逆 $A(\mathbf{m})^{-1}$，则有，

$$A(\mathbf{m})A(\mathbf{m})^{-1}=\mathbf{I} \tag{3.14}$$

上式两边对 $\mathbf{m}$ 求导可得，

$$\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}A(\mathbf{m})^{-1}+A(\mathbf{m})\frac{\partial A(\mathbf{m})^{-1}}{\partial\mathbf{m}}=0 \tag{3.15}$$

上式可改写为，

$$\frac{\partial A(\mathbf{m})^{-1}}{\partial\mathbf{m}}=-A(\mathbf{m})^{-1}\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}A(\mathbf{m})^{-1} \tag{3.16}$$

定义数据拟合的残差矢量为

$$\mathbf{r}(\mathbf{m})=\mathbf{d}_{cal}-\mathbf{d}_{obs}=F(\mathbf{m})-\mathbf{d}_{obs}=C\left(A(\mathbf{m})\right)^{-1}\mathbf{f}-\mathbf{d}_{obs} \tag{3.17}$$

其中 $\mathbf{r}(\mathbf{m})$ 表示波场残差，$F(\mathbf{m})$ 表示正算子根据模型 $\mathbf{m}$ 得到的数据，$\mathbf{d}_{obs}$ 表示观测到的数据。

重写数据逼近误差泛函为

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\mathfrak{I}_{LS}(\mathbf{m})=\frac{1}{2}\left\|C\left(A(\mathbf{m})\right)^{-1}\mathbf{f}-\mathbf{d}\right\|_Y^2=\frac{1}{2}\sum_t\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\mathbf{r}^2(\mathbf{m}) \tag{3.18}$$

首先，投影算子 $A(\mathbf{m})$ 的 Gateaux 导数可以用如下的方式进行定义，

$$\left.\frac{\partial A(\mathbf{m}+\tau\mathbf{e}_i)}{\partial\tau}\right|_{\tau=0}=\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\cdot\mathbf{e}_i \tag{3.19}$$

其中，$\dfrac{\partial A(\mathbf{m})}{\partial\mathbf{m}}$ 代表投影算子 $A(\mathbf{m})$ 的 Fréchet 微商。则误差泛函的梯度可以按如下形式导出

$$\begin{aligned}\left[G_{LS}\right]_i&=\frac{\partial}{\partial\tau}\mathfrak{I}_{LS}(\mathbf{m}+\tau\mathbf{e}_i)\big|_{\tau=0}=\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}\cdot\mathbf{r}(\mathbf{m})\\&=\left\langle\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0},\mathbf{r}(\mathbf{m})\right\rangle\end{aligned} \tag{3.20}$$

其中，$\left.\frac{\partial F}{\partial \tau}\left(\mathbf{m}+\tau \mathbf{e}_i\right)\right|_{\tau=0}$ 定义了正演波场随模型参数变化的方向导数---Gateaux 导数。定义误差泛函的梯度首先要给该导数。

由（3.6）式知，该导数可以写为，

$$\frac{\partial F}{\partial \tau}\left(\mathbf{m}+\tau \mathbf{e}_i\right)=C \frac{\partial}{\partial \mathbf{m}}\left(A\left(\mathbf{m}\right)^{-1} \mathbf{e}_i\right) f \tag{3.21}$$

利用（3.16）式，上式重写为

$$\begin{aligned} \frac{\partial F}{\partial \tau}\left(\mathbf{m}+\tau \mathbf{e}_i\right) &= -CA\left(\mathbf{m}\right)^{-1}\left(\frac{\partial A}{\partial \mathbf{m}} \mathbf{e}_i\right) A\left(\mathbf{m}\right)^{-1} f \\ &= -CA\left(\mathbf{m}\right)^{-1}\left(\frac{\partial A}{\partial \mathbf{m}} \mathbf{e}_i\right) \mathbf{u} \end{aligned} \tag{3.22}$$

把（3.22）定义的变分代入（3.20）式，

$$\begin{aligned} \left[G_{LS}\right]_i &= \left\langle \left.\frac{\partial F}{\partial \tau}\left(\mathbf{m}+\tau \mathbf{e}_i\right)\right|_{\tau=0}, \mathbf{r}\left(\mathbf{m}\right)\right\rangle_{\mathcal{H}_2} = -\left\langle CA\left(\mathbf{m}\right)^{-1}\left(\frac{\partial A}{\partial \mathbf{m}} \mathbf{e}_i\right) \mathbf{u}, \mathbf{r}\left(\mathbf{m}\right)\right\rangle_{\mathcal{H}_2} \\ &= -\left\langle \left(\frac{\partial A}{\partial \mathbf{m}} \mathbf{e}_i\right) \mathbf{u}, \left(A\left(\mathbf{m}\right)^{-1}\right)^* C^* \mathbf{r}\left(\mathbf{m}\right)\right\rangle_{\mathcal{H}_1} \\ &= \left\langle \left(\frac{\partial A}{\partial \mathbf{m}} \mathbf{e}_i\right) \mathbf{u}, \mathbf{z}\right\rangle_{\mathcal{H}_1} \end{aligned} \tag{3.23}$$

（3.23）中第二行利用了数据空间和模型空间存在的共轭关系（这种关系是假设的，理想弹性介质情况下应该是可以接受的!），用共轭算子 $\left(A\left(\mathbf{m}\right)^{-1}\right)^*$ 把数据残差投影到模型空间。定义 $\mathbf{u}=A\left(\mathbf{m}\right)^{-1} f$ 为正演波场；$\mathbf{z}=-\left(A\left(\mathbf{m}\right)^{-1}\right)^* C^* \mathbf{r}\left(\mathbf{m}\right)$ 为残差波场的反传播。

可以看出，波场 $\mathbf{z}$ 的求解需要解如下的微分方程

$$\left(A\left(\mathbf{m}\right)\right)^* \mathbf{z}=-C^* \mathbf{r}\left(\mathbf{m}\right) \tag{3.24}$$

前已述及，$A\left(\mathbf{m}\right)$ 是一个共轭算子，满足条件：$\left(A\left(m\right)^{-1}\right)^*=\left(A\left(m\right)^*\right)^{-1}$。

为了清楚起见，把正向波场外推方程和反向波场外推方程罗列如下：

正向波场外推方程：

$$\begin{cases} \left(\frac{1}{v^2\left(x\right)} \frac{\partial^2}{\partial t^2}-\Delta\right) u\left(\mathbf{x}, t\right)=-\delta\left(\mathbf{x}-\mathbf{x}_S\right) S\left(t\right) \\ u\left(\mathbf{x}, t\right)\big|_{\mathbf{x}=\Omega}=f_b\left(\Omega, t\right) \end{cases} \tag{3.25a}$$

其中，$f_b\left(\Omega, t\right)$ 代表边界条件。一般用 PML 边界条件。Ω 代表计算波场在空间上的边界。

反向波场外推方程：

$$\begin{cases}\left(\dfrac{1}{v^2(x)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)z(\mathbf{x},t)=0\\ z(\mathbf{x},t)\big|_{\mathbf{x}=\boldsymbol{\xi}}=u(\boldsymbol{\xi},t)\end{cases} \tag{3.25b}$$

其中，$\boldsymbol{\xi}$代表观测点坐标值。$u(\boldsymbol{\xi},t)$代表观测点处的波场。FWI 进行波场残差外推时，仅有观测点处有波场残差，因此，才有（3.25b）这样的边界条件的定义。实际上，波场逆时外推并不限于这样的边界条件。可见，（3.25a）是一个初值问题，沿时间正向外推计算；（3.25b）是一个边值问题，逆时间反向外推计算。（3.23）中计算梯度时，正演波场$\mathbf{u}$前还要作用一个算子$\dfrac{\partial A(\mathbf{m})}{\partial \mathbf{m}}\cdot\mathbf{e}_i$。它代表算子$A(\mathbf{m})$的微扰。

定义算子$A(\mathbf{m})$的微扰为，

$$\delta A(\mathbf{m})=\left(\frac{1}{(v+\delta v)^2}\frac{\partial^2}{\partial t^2}-\Delta\right)-\left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\Delta\right)=\frac{\delta v}{v^2}\frac{\partial^2}{\partial t^2} \tag{3.26a}$$

因此有，

$$\frac{\delta A(\mathbf{m})}{\delta \mathbf{m}}=\frac{1}{v^2}\frac{\partial^2}{\partial t^2} \tag{3.26b}$$

由此，最后的梯度公式为，

$$\mathbf{g}(\mathbf{m}(\mathbf{x}))=\sum_t\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\frac{1}{v^2(\mathbf{x})}\frac{\partial^2 u(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta})}{\partial t^2}z(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta}) \tag{3.27}$$

利用 Adjoint-State 方法求解梯度的地球物理意义可以解释为：误差函数对模型扰动的一阶导数值为震源函数激发的波场与残差波场的反传播得到的场的内积结果。这种计算梯度的方法需要求解方程$\mathbf{u}=A(\mathbf{m})^{-1}f$和$A(\mathbf{m})^{*}\mathbf{z}=-C^{*}\mathbf{r}(\mathbf{m})$，这是两次波场正演计算。可以看出，反问题的计算量主要集中在梯度的计算上，每次迭代原则上至少需要两次正演计算。

相对于 Jacobi 方法而言，Adjoint-State 方法可以方便、更高效地计算出泛函的梯度值，用于非线性迭代公式的计算。这是目前全波形反演中最常用的梯度计算方法。

3.3 FWI 具体实现方案

FWI 的核心工作是梯度的计算，用伴随状态法计算梯度，核心的计算量是（3.25）定义的正向和反向波场外推工作。首先定义单炮的梯度，然后累计各炮的梯度得到全模型空间的梯度。估计步长后，利用（4.4）式计算本次迭代修正量。基本计算步骤可罗列如下。

（1）单炮梯度的计算

①解方程（3.25a），计算并记录每一时刻的正传波场$\mathbf{u}(\boldsymbol{\xi},\boldsymbol{\eta},t)$，求其二阶导数，并用（3.26）定义的算子作用。

②计算检波点位置的波场残差$C^{*}\mathbf{r}(\mathbf{m})$。

③解方程（3.25b），将波场残差反传播到模型空间，得到反传播数据$\mathbf{z}(\boldsymbol{\xi},\boldsymbol{\eta},t)$。

④求正、反传波场的零时移互相关，得到单炮的梯度场。

（2）叠加所有炮求取的梯度场，得到整个模型空间的梯度场。

（3）利用（3.10）式得到本次迭代的修正结果。其中步长的估计参考 Dennis 和 Schnabel（1996）以及 Nocedal 和 Wright（1999）。

图 5 是 FWI 反演的数值实验。我们用两个初始模型测试 FWI，分别是一维线性递增的速度模型（从 1500m/s 到 4000m/s）和对真实速度模型平滑后的速度模型。图 5a 是初始模型采用线性递增速度模型的反演结果。图 5a 在浅层的反演结果与真实模型很接近，但深层速度的反演结果与真实模型相差较大。这是由于初始速度模型没有速度场的宏观背景部分，导致反演结果在深层没有收敛到真实模型。图 5b 是初始模型采用平滑过的真实速度模型的反演结果。对比图 5a 可以看出，由于初始模型包含了速度场的低波数部分，反演得到的结果好于图 5a 中采反演结果。图 5c 是初始模型采用线性递增速度模型的反演结果，但人为地去除了数据中小于 5hz 的频率分量。此时，初始速度模型与真实模型偏差较大，而地震数据中又缺少低频分量。图 5c 的结果证明了在此情况下，FWI 很难收敛。

通过理论分析以及数值实验验证，我们不能保证在实际数据上，FWI 方法技术是有效的。事实上，受初始模型精度，实际数据缺失低频，缺乏大偏移据的观测数据，地震子波波形未知，地下介质密度未知，噪音干扰等诸多因素限制，FWI 在实际应用上显得举步维艰，鲜有成功案例。

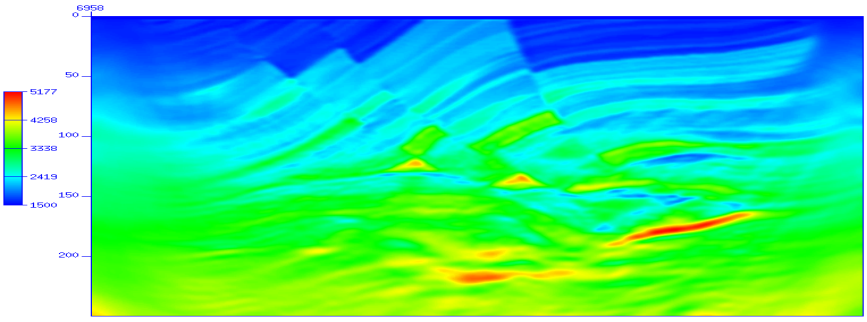


图 5a 初始模型采用线性递增速度模型得到的反演结果

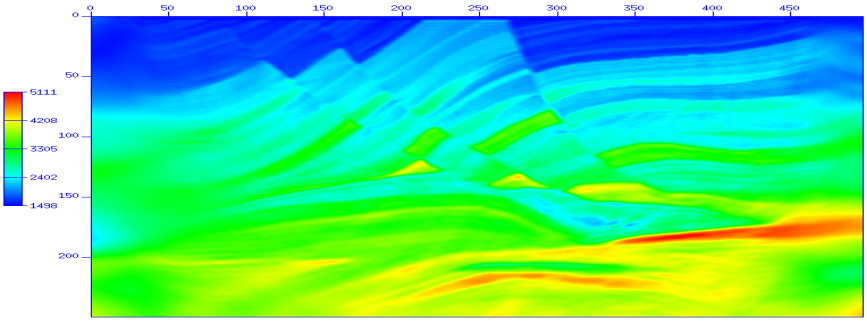


图 5b 初始模型采用平滑后的真实速度模型得到的反演结果

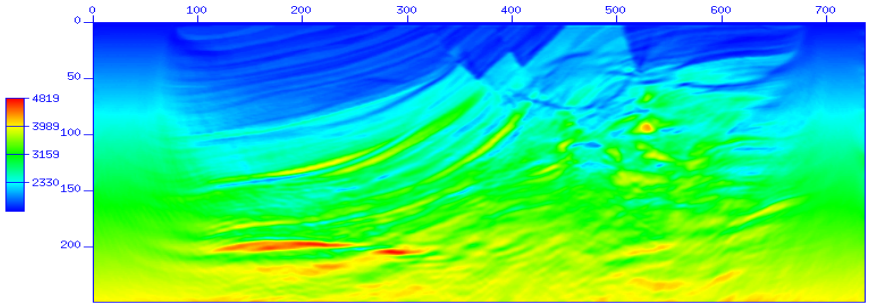


图 5c 初始模型采用线性递增速度模型得到的反演结果，但原始数据缺失低频（缺失 5hz 之下的频率分量）

4. 多级 Born 近似的逐步线性化反演与全波形反演（FWI）的区别与联系

根据前面两节对多级线性反演和非线性全波形反演（FWI）的讨论，可以看出，从计算的角度看，核心就是泛函梯度的表达。分析泛函梯度的计算可以了解两种反演方法的区别与联系。

多级线性反演的基础是多级线性 Born 近似。一级 Born 近似下的散射波场反映一次反射（绕射）波的激发与接收，也反映透射波（直达波、折射波和 Diving Wave）的激发与接收；二级 Born 近似下的散射波场反映一次反射（绕射）波和二次反射（绕射）波的激发与接收；三级 Born 近似下的散射波场反映一次反射（绕射）波、二次反射（绕射）波和三次反射（绕射）波的激发与接收。很显然，利用这样的波场表达进行地震波反演成像公式的推导，可以导出利用各次反射（绕射）波进行反演成像的公式。其中的基本步骤包括：透射波层析、各次反射波的最小二乘真振幅成像和各次反射波的层析成像。如果把所有的反射（绕射）波及透射波现象都用来进行反演成像，就是等价于 FWI 的反演过程。但是，值得注意的是，这样的过程会带来高次反射（散射）波的正演计算，这是极困难的事情。

从（3.27）和（3.9）式与多级 Born 近似反演相比，可以看出 FWI 反演是一个黑匣子。它利用了波在介质中传播的所有波现象。多级 Born 近似反演逐级对透射波、一次反射波、二次反射波和（原则上对）更高次的反射波进行反演成像。数据是有选择的、反演过程是可以控制的。

我们认为，基于多级 Born 近似的多级线性反演成像是非线性的 FWI 反演前必须要做的反演工作，也是把 FWI 技术推向精细的岩性油气藏勘探的正确的路线。

5. 正则化与预条件的关系

假设地震波传播的波场和速度之间的关系是近似线性的，及假如正问题可以描述为

$$\mathbf{b} \approx F\left(\mathbf{m}\right) - F\left(\mathbf{m}_0\right) = \left.\frac{\partial F}{\partial \mathbf{m}}\right|_{\mathbf{m}_0} \delta\mathbf{m} \tag{5.1}$$

波场扰动和参数扰动之间的局部线性化的关系可以用

$$\mathbf{A}\mathbf{x} \approx \mathbf{b} \tag{5.2}$$

这个矩阵方程来表示，并估计在 $\mathbf{m}_0$ 点附近的 $\delta\mathbf{m}$，即 $\mathbf{x}$。$\mathbf{b}$ 是实测数据。实际上，这个局部线性化的关系在勘探地震的速度估计中是很难得到的，尤其是在复杂的速度结构情况下。这也是需要从另外的角度得到线性化的（5.2）式的根本原因。

（5.2）式的求解经典地可以用 SVD 方法。实际上是假设 $\mathbf{A}\mathbf{x} = \mathbf{b}$，然后用 SVD 方法进行求解。当然，这中间隐含了对数据的假设，即 $\mathbf{b}$ 是高斯白噪的。实际上，也就是在用最小二乘的观点在处理上式，截断

SVD 是基本的想法。（5.2）式的最小二乘解与加权 SVD 的关系很多数学书上已经讨论很多（Vogel,2002）。

线性化的（5.2）式的求解，由于数据和系数矩阵中都存在误差，得到较好的解也是不容易的。正则化和预条件是两种基本的方法。根本的目的都是改变系数矩阵的条件数或特征值的分布。

一般地，我们把对（5.2）式的求解化为二次型泛函的求解问题

$$\mathcal{J}\left(\mathbf{x}\right)=\left\|\mathbf{Ax}-\mathbf{b}\right\|^2+\lambda\left\|R\mathbf{x}\right\|^2 \tag{5.3}$$

其中，R 是正则化算子。R 若取为二阶导数算子或 Laplace 算子，作用后的模型更类似随机量，因此，可以用 L_2 范数来构造模型的罚函数。

从反演问题的概率论认识中,可以看出，对于（5.3）式右端项中的数据误差性和模型误差项，都希望是满足高斯分布的。我们知道，数据项是比较符合高斯分布的，但模型项不容易符合高斯分布。

为此，我们需要对模型进行预处理，基本的思想是对它施加粗化算子，使它比较接近符合高斯分布。Laplace 算子就是其中一种选择。

从另一个角度看，反问题求解时，参数的光滑部分（低波数部分）是容易首先求解的。那么我们可以引入一个光滑算子 S，作用于原模型上，即

$$\mathbf{y}=\mathbf{Sx} \tag{5.4}$$

$\mathbf{y}$ 是参数的光滑部分。
把（5.4）代入（5.2）式得到，

$$AS^{-1}\mathbf{y}=\mathbf{b} \tag{5.5}$$

另一方面，若数据中有野值，需要引入一个算子去除它们。假设该算子为 L，代入（5.5）式得到

$$LAS^{-1}\mathbf{y}=L\mathbf{b} \tag{5.6}$$

假如对 A 矩阵做 SVD 分解（ $A=U\Sigma V^T$ ），并代入（5.6）式，

$$LU\Sigma V^TS^{-1}\mathbf{y}=L\mathbf{b} \tag{5.7}$$

其中，U 为左奇异矩阵，V 为右奇异矩阵。单纯地看（5.2）式中 A 矩阵的奇异值分解

$$U\Sigma V^T\mathbf{x}=\mathbf{b}\Rightarrow\Sigma V^T\mathbf{x}=U^T\mathbf{b}\Rightarrow\Sigma\mathbf{y}=\mathbf{d} \tag{5.8}$$

可以知道，各列正交的右奇异向量作用在数据上，对数据进行变换，到另一个域，表示为 $\mathbf{d}$；各列正交的左奇异向量作用在参数上，对参数进行变化，到另一个域，表示为 $\mathbf{y}$。在此情况下，数据 $\mathbf{d}$ 和参数 $\mathbf{y}$ 之间的关系与一对一的、满射的关系。参数 $\mathbf{y}$ 中各分量之间的关系是独立的。

从（5.7）式可以看出，假若数据与处理算子和模型与处理算子分别取为右奇异矩阵和左奇异矩阵的逆，反演问题会变得异常简单。

这就告诉我们，算子 L 和 S^{-1} 的选择更基本的含义是对数据和参数引入不同的变换矩阵，使得数据中的特征成分和参数中的特征成分建立起一对一的关系。这种关系是线性的。

从另一个角度，也可以看出，算子 L 和 S^{-1} 的选择是为了改善矩阵 A 的特征值的分布。相当于对 A 矩阵进行预条件处理，算子 L 是左预条件算子，算子 S^{-1} 是右预条件算子。一般数学教科书上我们取左预条件算子 $M\approx A^{-1}$，有

$$M\mathbf{Ax} \approx M\mathbf{b} \Rightarrow \mathrm{Q}\mathbf{x} \approx M\mathbf{b} \qquad (5.9)$$

此时，Q 接近一个对角阵。从勘探地震的角度，我们希望从（7）式来解释矩阵 A 的预条件的含义。

从上面的分析可见，正则化的直观目的是使得参数经正则化化算子的处理后更接近高斯分布。预条件的直观目的是改变矩阵 A 的特征值的分布。预条件又分施加于模型的左预条件和施加于数据的右预条件。

从反演的要求看，我们希望施加对数据的预条件和对参数的预条件，因此，预条件与数学上求解线性问题 $\mathbf{Ax} \approx \mathbf{b}$ 联系更紧密。正则化仅仅考虑参数的约束不是完善的处理方案。把正则化和预条件对比分析，可以更清楚地理解反演解的本质。

6. 地震波反演成像在岩性油气藏刻画中的作用

岩性油藏的刻画主要的方法还是依靠对岩性参数变化的预测。当前使用的方法技术，主要是在保持反射子波特征的保幅成像基础上利用反射子波波形的变化来进行油气藏描述和油气预测；在角度反射系数的基础，利用 Zeoprittz 方程，反演求解弹性参数，与岩石物理知识结合，估计岩性储层参数，进行油气藏描述和油气预测。当然，利用一般叠前深度偏移结果的储层的几何形态的描述也是很基本的。在高密度、宽方位、宽频带数据采集和前述的地震波反演成像的基础上，岩性油藏的刻画应该进入到一个新的阶段。高精度的背景速度估计+高保真和高分辨率的最小二乘参数反演成像，与岩石物理知识结合，直接进入（岩性）油气藏描述和油气预测。事实上，高保真和高分辨率的最小二乘参数反演成像也可以更精细地刻画储层的几何形态。基于此，地震地质解释也会进入到新的阶段。

7. 结论

岩性油气藏的描述，最根本的是依据地震弹性参数的变化。井资料和其他资料的应用应归于参数反演的正则化过程。地震参数的估计主要包括基于多级 Born 近似的多级线性反演和非线性的全波形反演。前者明确地知道在用哪些波线性进行反演，可以先利用明显的初至波反演浅层速度、已知偏移速度后利用一次反射波进行保真的最小二乘参数反演成像，接着进行二次散射波的层析反演成像。这样可控的过程更容易发展成实用化的技术。非线性的 FWI 是利用所有波现象的反演成像，受很多因素的限制。本文详细地分析了多级 Born 近似的多级线性反演和非线性的全波形反演的理论背景。指出了反演成像的发展方向。岩性油气藏描述的技术导向必定是以高维最小二乘参数反演成像提供基础数据，而不是空间中单点的 AVA 反演结果提供基础数据。**多级 Born 近似的多级线性反演是岩性油气藏描述的下一代核心技术。**

参考文献

[1] Aster, R. C., Borchers, B., and Thurber, C. H., 2005, Parameter estimation and inverse problems; Elsevier Academic Press

[2] Baysal, E.,D.Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, 48, 1514–1524.

[3] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, 26, 99–108.

[4] Biondi, B., andW. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, 69, 1283–1298.

[5] Bouchon, M., M. Campillo, and S. Gaffet, 1989, A boundary integral equation—Discrete wavenumber representation method to study wave propagation in multilayered media having irregular interfaces: Geophysics, 54, 1134–1140.

[6] Brossier, R., J.Virieux, and S. Operto, 2008, Parsimonious finite-volume frequency-domain method for 2D P-SV wave modelling: Geophysical Journal International, 175, 541–559.

[7] Bunks, C., F.M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, 60, 1457–1473.

[8] Chapman, C., 1985, Ray theory and its extension: WKBJ and Maslov seismograms: Journal of Geophysics, 58, 27–43.

[9] Choi, Y., D.-J. Min, and C. Shin, 2008, Two-dimensional waveform inversion ofmulti-component data in acoustic-elastic coupled

media: Geophysical Prospecting, 56, 863–881.

[10] Claerbout, J., 1971, Toward a unified theory of reflector mapping: Geophysics, 36, 467–481.

[11] Claerbout, J., 1976, Fundamentals of geophysical data processing: McGraw-Hill Book Co., Inc.

[12] Claerbout, J., and S. Doherty, 1972, Downward continuation of moveout corrected seismograms: Geophysics, 37, 741–768.

[13] Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic non-linear waveform inversion: Application to real data: Geophysics, 55, 527–538.

[14] Danecek, P., and G. Seriani, 2008, An efficient parallel Chebyshevpseudospectral method for large-scale 3D seismic forward modelling: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, P046.

[15] Dennis, J. E. Jr. and Schnabel, R. B., 1996, Numerical methods for unconstrained optimization and nonlinear equations; SIAM

[16] Djikpéssé, H., and A. Tarantola, 1999, Multi parameter 1-norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, 64, 1023–1035.

[17] Docherty, P., R. Silva, S. Singh, Z.-M. Song, and M.Wood, 2003, Migration velocity analysis using a genetic algorithm: Geophysical Prospecting, 45, 865–878.

[18] Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geophysics, 43, 1342–1351.

[19] Grote M. J. and I. Sim, 2010, Efficient PML for the wave equation: from internet: arXiv:1001.0319v1.

[20] Hansen, C., ed., 1998, Rank-defficient and discrete ill-posed problems —Numerical aspects of linear inversion: Society for Industrial and Applied Mathematics.

[21] Herrmann, F. J., C. R. Brown, Y. A. Erlangga, and P. P. Moghaddam, 2009, Curvelet-based migration preconditioning and scaling. Geophysics, 74, A41-A46.

[22] Koren, Z., K.Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, 96, 20289–20299.

[23] Kormendi, F., and M. Dietrich, 1991, Non-linear waveform inversion of plane-wave seismograms in stratified elastic media: Geophysics, 56, 664–674.

[24] Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

[25] Lambaré, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, 57, 1138–1154.

[26] Lecomte, I., 2008, Resolution and illumination analyses in PSDM: A ray based approach: The Leading Edge, 27, 650–663.

[27] Lee, K. H., and H. J. Kim, 2003, Source-independent full-waveform inversion of seismic data: Geophysics, 68, 2010–2015.

[28] Llimes, L., 2007, Sensitivity of seismic waves to the structure; Report 17, SW3D, Department of Geophysics, ChalesUniversity, Prague

[29] Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, 49, 533–549.

[30] Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

[31] Mora, P. R., 1987, Nonlinear two-dimensional elastic inversion of multi-offset seismic data: Geophysics, 52, 1211–1228.

[32] ———, 1988, Elastic wavefield inversion of reflection and transmission data: Geophysics, 53, 750–759.

[33] Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, 100, 12431–12447.

[34] Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics, D. Reidel publishing company

[35] Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full waveform tomography: Application to the eastern Nankai trough: Journal of Geophysical Research, doi: 10.1029/2005JB003835.

[36] Operto, S., J. Virieux, P. Amestoy, J. Y. L'Excellent, L. Giraud, and H. Ben-Hadj-Ali, 2007, 3D finite-difference

frequency-domain modeling of viscoacoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, 72, no. 5, SM195–SM211.

[37] Plessix, R. E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysics Journal International, 157, 975–987.

[38] Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method: Geophysical Prospecting, 38, 311–330.

[39] ———, 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, 64, 888–901.

[40] Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multisource cross-hole tomography, Part 1: Acoustic wave-equation method: Geophysical Prospecting, 38, 287–310.

[41] Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Multi-scale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, 159, 1032–1056.

[42] Ren, H., R. S. Wu, and H. Wang, 2011, Wave equation least square imaging using the local angular Hessian for amplitude correction: Geophysical Prospecting, Accepted.

[43] Saad, Y., 2003, Iterative methods for sparse linear systems, 2nd ed.: Society for Industrial andApplied Mathematics.

[44] Scales, J. A., P. Docherty, and A. Gersztenkorn, 1990, Regularization of nonlinear inverse problems: Imaging the near-surface weathering layer: Inverse Problems, 6, 115–131.

[45] Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, 173, 922–931.

[46] Shin, C., and Y. H. Cha, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, 177, 1067–1079.

[47] Sirgue, L., J. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F022.

[48] Snieder, R.,M. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem: Geophysics, 54, 991–1000.

[49] Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, 43, 23-48.

[50] Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: Geophysics, 74, WCA95-WCA107.

[51] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, 49, 1259–1266.

[52] Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation; SIAM

[53] Virieux, J., 1986, P-SV wave propagation in heterogeneous media, velocity stress finite difference method: Geophysics, 51, 889–901.

[54] Vogel, C. R., 2002, Computational Methods for Inverse Problems; SIAM.

[55] Wu, R. S., 2003, Wave propagation, scattering and imaging using dual-domain one-way and one-return propagation: Pure and Applied Geophysics, 160, 509–539.

[56] Yilmaz, O., 2001, Seismic data analysis: processing, inversion and interpretation of seismic data: SEG.

[57] Yu, J., J. Hu, G. T. Schuster, and R. Estill, 2006, Prestack migration deconvolution: Geophysics, 71, S53-S62.

[58] Zhdanov, M. S., 2002, Geophysical Inverse theory and regularization problems; Elsevier Science B. V.

[59] Zhou, B., and S. A. Greenhalgh, 2003, Crosshole seismic inversionwith normalized full-waveform amplitude data: Geophysics, 68, 1320–1330

面向岩性油气藏的最小二乘叠前深度偏移成像

王华忠，胡江涛

同济大学海洋与地球科学研究院波现象与反演成像研究组，上海，200092

【摘要】地震波成像的目的首先是定位反射界面或散射点出现的空间位置，描述地下地质体的几何结构；然后是估计地下岩石介质的物性参数，主要是速度和密度参数；最终与岩石物理结合描述含油气储层。事实上，地下地质体的几何结构携带了最重要的信息。高维叠前地震数据偏移成像与反演成像是达到上述目标的主要方法。地震波的散射场表达是叠前偏移成像和反演成像的共同基础（王华忠等，2013）。本文以地震波传播的散射场表达与逆散射成像为切入点，首先说明逆散射成像的本质，然后再试图理清楚一般意义下的叠前偏移成像与逆散射成像之间的区别与联系，并说明叠前偏移成像把波场残差沿等时面投影，波形层析成像沿波路径投影，分别确定速度场的高波数成分和低波数成分。最后给出了迭代最小二乘偏移成像的具体实现方法和数值结果。目的是发展高保真高分辨的最小二乘叠前深度偏移成像方法，用于岩性油气藏的描述与刻画。

【关键词】叠前偏移成像，逆散射成像，最小二乘偏移成像

引言

油气勘探地震学的最终目标是储层。通向储层描述与刻画的途径有两条（目前看来，典型的路线有两条）。第一条是常规的，目前主要采用的。它的核心是保幅处理与成像，它的核心结果是（相对）保真/保幅的角度成像道集。保幅处理和成像是勘探地震学走向储层地震学的关键中间环节，它前面与野外地震数据采集紧密相关，后面为叠后波阻抗反演、叠前 AVA/AVO 反演等提供扎实的基础数据。在野外地震数据采集质量有保证的前提下，完成保真的角度道集的生成的最关键因素在于较准确的背景参数模型的建立，主要是速度模型、各向异性参数模型和 Q 值模型。基于保真的角度成像道集，利用 1D 波阻抗反演和/或 AVA 叠前反演得到弹性参数，与岩石物理和测井结合描述储层参数，这是目前常用的从地震成像走向储层刻画的方法技术体系。该技术体系的核心包括三方面的关键因素：满足速度估计和保真成像要求的数据体、较准确的背景速度模型和保真成像技术系列、基于保真成像道集的储层参数估计及储层描述方法。

很显然，还存在一条更直接的储层描述和刻画的途径。基于满足速度估计和保真成像要求的数据体，和较准确的背景速度模型，直接估计精细的速度参数扰动，利用速度参数的扰动来解释储层的形态变换和参数变化，与井数据结合进入储层含油气性的描述。这一个途径是以 FWI 为代表的速度估计技术不断实用化后预示的储层描述与刻画的新途径。与目前常用的基于角度道集的储层刻画方法相结合，会使得储层描述与刻画技术得到新的提高，更有利于油田开发的精细化，挖掘产能潜力。

这两条路线的核心是在野外地震数据的质量能满足要求的前提下，尽可能估计精确的速度参数（也包括各向异性和 Q 值参数）。

一些数学家看来，石油地震勘探问题很简单，就是根据实测的叠前地震数据，用波动方程作为实测数据预测器，若基于一个假设的参数模型求解波动方程得到的合成数据与实测的叠前地震数据的总误差最小，该假设模型就是所求的。

事实上，这完全低估了勘探地震反演问题的复杂性。

从勘探地震学的初始，勘探地震学家走的是一条实际问题驱动的路线来解决成像问题。由 CMP 道集估计背景初始速度，进行叠后偏移成像；进入叠前偏移成像，由成像道集进行偏移速度分析，有时结合层析成像反演速度模型。迭代进行叠前偏移成像和速度模型的更新是石油工业界发展出的实用化的地震波偏移成像的路线。经检验是行之有效的。更细致的、能用于储层描述的参数估计是基于 1D 波阻抗反演、AVA 角度道集反演。

事实上，一直有很多人在探讨是否能从更数学化的路线来解决成像问题。这分成两条路线，一是 Tarentola（1984）为代表的全波形拟合非线性反演速度（还有其他）参数的路线；一是 Bleistein（1985）为代表的基于散射波表达的线性反演速度（还有其他）参数的路线。

这些路线之间是紧密相联的。目的都是走向储层的精确刻画。

本文以地震波传播的散射场表达与逆散射成像为切入点，首先说明逆散射成像的本质，然后再试图理清楚一般意义下的叠前偏移成像与逆散射成像之间的区别与联系。重点是发展出高保真高分辨的最小二乘叠前深度偏移成像方法，用于岩性油气藏的描述与刻画。指明在勘探地震学中，最小二乘叠前深度偏移成像就是相关学科中的逆散射成像，它们都归结为一个线性化的反演问题，其中的正问题是散射场的 Born 近似表达式，也可以称为是广义的 Radon 变换公式。勘探地震学中反射波占重要地位，反射波的逆散射成像尽管本质上与散射波的逆散射成像是一样的，但成像结果的表现形式是不同的，前者表现为角度反射系数，后者表现为速度扰动量。这又一次说明反射地震学中射线理论的广泛应用。小尺度散射体的逆散射成像只能用波动理论来解释。数值实验的结果验证了我们的分析。

1. 地震波传播的散射场表达

1.1 基本的声波方程和声散射势的引入

1.1.1 基本的声波方程

一般地,我们假设波在地下介质中的传播可以由如下的声波方程控制，

$$\left[\frac{1}{v(\mathbf{x})^2}\frac{\partial^2}{\partial t^2}-\rho(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla\right)\right]u(\mathbf{x},t)=S(\mathbf{x},t) \tag{1.1}$$

其中，$u(\mathbf{x},t)$为声场势，它是时间$t\in R$和空间坐标$\mathbf{x}\in\Omega\subseteq R$的函数。$v(\mathbf{x})$代表速度场；$\rho(\mathbf{x})$代表密度场；$S(\mathbf{x},t)$代表体积场源。

在频率空间域，上式变为

$$\mathcal{L}_a(\mathbf{x},\omega)u(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{1.2a}$$

其中

$$\mathcal{L}_a(\mathbf{x},\omega)=\nabla\cdot\frac{1}{\rho}\nabla+\frac{\omega^2}{\rho v^2(\mathbf{x})} \tag{1.2b}$$

为声波传播算子。

定义背景介质中的声波传播满足如下方程，

$$\mathcal{L}_{a0}(\mathbf{x},\omega)u_0(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{1.3a}$$

其中

$$\mathcal{L}_{a0}(\mathbf{x},\omega)=\nabla\cdot\frac{1}{\rho_0}\nabla+\frac{\omega^2}{\rho_0 v_0^2(\mathbf{x})} \tag{1.3b}$$

为背景介质中的声波算子。

1.1.2 声散射势的引入

根据（1.2）和（1.3）式，定义声散射势为，

$$\begin{aligned}\mathcal{V}_a(\mathbf{x},\omega)&\equiv\mathcal{L}_a(\mathbf{x},\omega)-\mathcal{L}_{a0}(\mathbf{x},\omega)\\&=\left(\nabla\cdot\frac{1}{\rho}\nabla+\frac{\omega^2}{\rho v^2(\mathbf{x})}\right)-\left(\nabla\cdot\frac{1}{\rho_0}\nabla+\frac{\omega^2}{\rho_0 v_0^2(\mathbf{x})}\right)\\&=\nabla\cdot\frac{a_\rho}{\rho}\nabla+\omega^2\frac{a_v}{\rho v^2(\mathbf{x})}\end{aligned}\tag{1.4}$$

其中

$$a_\rho=1-\frac{\rho_0(\mathbf{x})}{\rho(\mathbf{x})}\tag{1.5a}$$

$$a_v(\mathbf{x})=1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}\tag{1.5b}$$

1.2 标量波方程和标量波散射势的引入

勘探地震学中，很多情况下关注标量波方程（或密度不变情况下的声波方程）。因此，类似地，写出如下的标量波散射势。

在频率空间域标量波方程形式为

$$\mathcal{L}_s(\mathbf{x},\omega)u(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega)\tag{1.6a}$$

其中

$$\mathcal{L}_s(x,\omega)=\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}\tag{1.6b}$$

为标量波传播算子。$\mathcal{L}_s$ 代表 $\mathcal{L}_{\text{scalar}}$.

定义背景介质中的声波传播满足如下方程，

$$\mathcal{L}_{s0}(\mathbf{x},\omega)u_0(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega)\tag{1.7a}$$

其中

$$\mathcal{L}_{s0}(\mathbf{x},\omega)=\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\tag{1.7b}$$

为背景介质中的标量波算子。

根据（1.6）和（1.7）式，定义标量波散射势为，

$$\mathcal{V}_s(\mathbf{x},\omega)\equiv\mathcal{L}_s(\mathbf{x},\omega)-\mathcal{L}_{s0}(\mathbf{x},\omega)=\left(\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}\right)-\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)=a_s(\mathbf{x})\frac{\omega^2}{v_0^2(\mathbf{x})}\tag{1.8a}$$

其中

$$a_s(\mathbf{x}) = 1 - \frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})} \tag{1.8b}$$

上式定义的散射势，在速度扰动不大时，还可以写为如下的形式，

$$\begin{aligned} \mathcal{V}_s(\mathbf{x},\omega) &= \omega^2 \frac{1}{v_0^2(\mathbf{x})}\left(1 - \frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}\right) = \omega^2 \frac{1}{v_0^2(\mathbf{x})}\left(\frac{v^2 - v_0^2}{v^2}\right) \\ &= \omega^2 \frac{1}{v_0^2(\mathbf{x})} \frac{(v - v_0)(v + v_0)}{v^2(\mathbf{x})} \approx \omega^2 \frac{1}{v_0^2(\mathbf{x})} \frac{2\delta v \cdot v_0}{v_0^2(\mathbf{x})} = \omega^2 \frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})} \end{aligned} \tag{1.9}$$

上述形式引入 $\delta v(\mathbf{x}) = v(\mathbf{x}) - v_0(\mathbf{x})$ 更清楚地反映出速度的局部变化是散射势的本质。从散射势的定义也可以看出，估计散射势的绕射层析成像与估计或定位反射系数的叠前偏移成像是有严格区别的。

1.3 Lippmann-Schwinger 方程的导出

借助于算子恒等式 $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_0\left(\mathcal{L}_0^{-1} - \mathcal{L}^{-1}\right)\mathcal{L}$ 和定义 $\mathcal{V} = \mathcal{L}_0^{-1} - \mathcal{L}^{-1}$ 可以得到

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_0 \mathcal{V} \mathcal{L} \tag{1.10}$$

算子 $\mathcal{L}$ 是对应声波和标量波的 Green 函数。它描述所有的波传播现象。 $\mathcal{L}^{-1}$ 和 $\mathcal{L}_0^{-1}$ 是“逆”Green 函数，它们经常被各自的共轭转置来代替。物理上，它们描述的是波的反传播现象。Green 函数可以描述从源到散射点，再到接收点的波传播；同样地，也可以描述从接收点到散射点，再到源点的波传播过程。假如波传播算子 $\mathcal{L}$ 用 Green 函数 G 表示，方程（1.10）可以重写为

$$G = G_0 + G_0 V G \tag{1.11}$$

方程（1.11）是非线性的，它能被进一步重写为

$$G = \left(I - G_0 V\right)^{-1} G_0 \tag{1.12}$$

对上式的右端项进行 Taylor 展开，

$$G = \left[\sum_{j=0}^{\infty} \left(G_0 \mathcal{V}\right)^j\right] G_0 = G_0 + G_0 \mathcal{V} G_0 + G_0 \mathcal{V} G_0 \mathcal{V} G_0 + G_0 \mathcal{V} G_0 \mathcal{V} G_0 \mathcal{V} G_0 + \cdots \tag{1.13}$$

方程（1.11）、（1.13）被称为 Lippmann-Schwinger 方程。

很显然，假如 $j \le 2$,方程（1.13）就仅描述了一个忽略了二阶以上散射波的波传播过程。因此，线性化的波传播算子仅仅描述波传播过程中的一阶散射。这就是

$$G = G_0 + G_0 \mathcal{V} G_0 \tag{1.14}$$

上式就是 Born 近似下的波传播算子，其中方程右端的第一项描述了背景场中的传播现象，第二项描述了一阶散射波的传播现象。方程（1.11）-（1.14）清楚地说明了 Born 近似的物理含义。只有当散射势比较弱时，（1.14）式中的高于二阶的项才可以被忽略不计。注意，从（1.13）式可以清楚地看出，高阶散射场能否被忽略，除了散射势比较弱的条件外，还有散射体的体积也很重要，因为散射场是散射势的体积积分结果。小散射体和弱散射势是 Born 近似成立的条件。

1.4 散射场的表达

从（1.14）式可以看出，Born 近似线性化模拟的总波场包括两部分：第一部分是由背景 Green 函数所描述的背景波场；第二部分是由散射势引起的方程（1.13）中的第二项所描述的散射场。很显然，这存在一个背景速度场如何取才能得到符合 Born 近似要求的背景场的问题。原则性的要求是入射场中不能存在散射场成分。根据这样的原则，如何决定散射势中的背景速度$v_0(\mathbf{x})$？这个问题是很难回答的！

事实上，在特别复杂、变化特别剧烈的速度场中，一阶 Born 近似的应用效果是很值得考虑的。这也是人们逐步在考虑引入更高阶的 Born 近似，降低一阶 Born 近似弱散射要求的原因。但这样做也会引入高阶散射场计算上的复杂性。

假定 Born 近似假设成立，根据方程（1.14），线性化的总波场可以写成

$$u(\boldsymbol{\xi},\boldsymbol{\eta},\omega)=u_0(\boldsymbol{\xi},\boldsymbol{\eta},\omega)+\omega^2\int d\mathbf{x}G_0(\boldsymbol{\xi},\mathbf{x},\omega)\frac{a_s(\mathbf{x})}{v_0^2(\mathbf{x})}u_0(\mathbf{x},\boldsymbol{\eta},\omega) \tag{1.15}$$

其中，Born 近似后的散射波场记为

$$\begin{aligned}u_s(\boldsymbol{\xi},\boldsymbol{\eta},\omega)&=\omega^2\int d\mathbf{x}G_0(\boldsymbol{\xi},\mathbf{x},\omega)\frac{a(\mathbf{x})}{v_0^2(\mathbf{x})}u_0(\mathbf{x},\boldsymbol{\eta},\omega)\\&=\omega^2\int d\mathbf{x}G_0(\boldsymbol{\xi},\mathbf{x},\omega)u_0(\mathbf{x},\boldsymbol{\eta},\omega)f(\mathbf{x})\end{aligned} \tag{1.16}$$

它代表从震源$\boldsymbol{\eta}$到散射点$\mathbf{x}$，再到接收点$\boldsymbol{\xi}$的散射波场。

Born 近似的物理实质是：在背景场$u_0(\boldsymbol{\xi},\boldsymbol{\eta},\omega)$中没有散射场；在扰动场中，只存在一次散射场。尽管用这种方式描述波传播与实际的波传播不同，但当二阶及二阶以上的散射场能量比较弱时，用 Born 近似进行 PSDM 和 Tomography 还是比较有效的。这是当前最小二乘叠前深度偏移成像的基础。也预示了下一步的最小二乘叠前深度偏移成像需要考虑更高次的散射波才能得到更好的结果。

关于子波问题，可以认为已经携带在背景场$u_0(\mathbf{x},\boldsymbol{\eta},\omega)$中。也可以用$G_0(\mathbf{x},\boldsymbol{\eta},\omega)S(\omega)$来表示，其中$S(\omega)$代表子波的谱。但是，在最小二乘叠前深度偏移反演成像中，子波需要已知（或者也作为一个待反演的量），估计参数的高波数变化。因此，对子波的研究是最小二乘叠前深度偏移反演成像能否得到好结果的重要因素。

1.5 散射势的进一步解释

标量波方程定义的散射势形式为，

$$\mathcal{V}_s(\mathbf{x},\omega)\equiv a_s(\mathbf{x})\frac{\omega^2}{v_0^2(\mathbf{x})}=a_s(\mathbf{x})\left|\mathbf{k}_0(\mathbf{x})\right|^2 \tag{1.17}$$

可见散射势与场的波矢量之间是有密切关系的。在某种情况下，已知任意空间点上的全部的场的波矢量成分，可以完全地刻画散射势。

由于波矢量代表波的传播方向，且$\mathbf{k}_0=\mathbf{k}_0^S+\mathbf{k}_0^R$，通过这样的关系可以联系起散射势与角度反射系数之间的关系。

改变（1.17）式，得到

$$\begin{aligned}\mathcal{V}_s(\mathbf{x},\omega)&\equiv a_s(\mathbf{x})\frac{\omega^2}{v_0^2(\mathbf{x})}\approx|\mathbf{k}_0(\mathbf{x})|^2\frac{(v-v_0)(v+v_0)}{\left(\frac{v+v_0}{2}\right)\left(\frac{v+v_0}{2}\right)}\\&=|\mathbf{k}_0(\mathbf{x})|^2\frac{4(v-v_0)}{(v+v_0)}=|\mathbf{k}_0(\mathbf{x})|^2\,4R(\theta=0^o)\end{aligned}\tag{1.18}$$

可见散射势可以写成与垂直反射系数有关的式子。原则上，非垂直入射的反射系数与散射势的关系可以用波场波矢量来规定。

上述一段说明试图把散射势的反演与反射系数的估计统一在一起。

2. 逆散射问题分析

2.1 常速背景介质与平面波入射情形下散射场与散射势间的关系

首先给出常速介质中自由空间 Green 函数为

$$G(\mathbf{x}-\mathbf{x}',\omega)=\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}=\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}\tag{2.1}$$

其中，$\mathbf{x}$ 代表场点位置，$\mathbf{x}'$ 代表源点位置，$|\mathbf{x}-\mathbf{x}'|$ 代表源点与场点之间的距离。$|\mathbf{k}|=\left|\omega/c\right|$ 代表场的波数。把（2.1）式代入（1.16）式定义的散射场中，得到

$$u_s(\mathbf{x},\omega)=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)u_0(\mathbf{x}',\omega)\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}\tag{2.2}$$

一般地，在远场近似假设下面，入射场在常速介质中传播，可以用平面波近似，

$$u(\mathbf{x},\omega)=e^{i\mathbf{k}\cdot\mathbf{x}}=e^{i|k|\mathbf{s}_0\cdot\mathbf{x}}\tag{2.3}$$

其中，$\mathbf{s}_0$ 是代表平面波的入射方向的单位矢量。将（2.3）式代入（2.2）式得到，

$$u_s(\mathbf{x},\omega)=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\mathbf{x}'}\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}\tag{2.4}$$

（2.4）式描述的波的传播关系如图 1 所示。

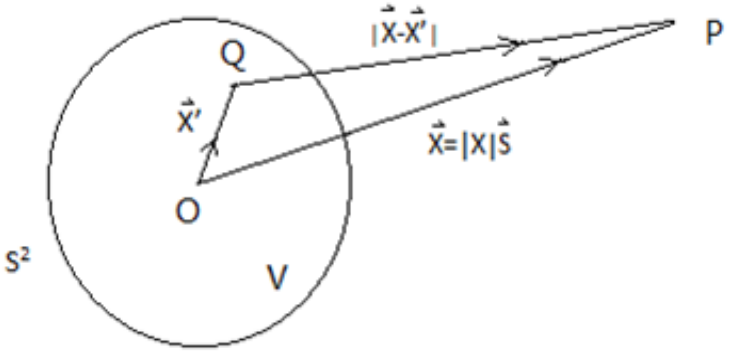


图 1 空间局部散射体引起的散射波传播关系

总场表示为

$$u(\mathbf{x},\omega)=u_0(\mathbf{x},\omega)+u_s(\mathbf{x},\omega)=e^{ik\mathbf{s}_0\cdot\mathbf{x}}+\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\cdot\mathbf{x}'}\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|} \tag{2.5}$$

由图 1 可知，当场点 P 距离散射体比较远，散射体积源分布空间有限时，用 Q 代表散射体积源内的点，O 代表坐标原点，则可以做如下的近似推导。

$$|\mathbf{x}-\mathbf{x}'|=|\mathbf{x}|-\mathbf{s}\cdot\mathbf{x}' \tag{2.6}$$

$$G(\mathbf{x}-\mathbf{x}',\omega)=\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}=\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}\approx\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}e^{-ik\mathbf{s}\cdot\mathbf{x}'} \tag{2.7}$$

其中，**s** 是代表散射波传播的局部平面波方向的单位矢量。把（2.7）式代入（2.5）式，得到

$$\begin{aligned}u(\mathbf{x},\omega)&\approx e^{ik\mathbf{s}_0\cdot\mathbf{x}}+\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\cdot\mathbf{x}'}\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}e^{-ik\mathbf{s}\cdot\mathbf{x}'}\\&=e^{ik\mathbf{s}_0\cdot\mathbf{x}}+\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'}\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}\\&=e^{ik\mathbf{s}_0\cdot\mathbf{x}}+\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'}\\&=e^{ik\mathbf{s}_0\cdot\mathbf{x}}+\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}f(\mathbf{s},\mathbf{s}_0,\omega)\end{aligned} \tag{2.8}$$

其中

$$f(\mathbf{s},\mathbf{s}_0,\omega)=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'} \tag{2.9}$$

称为辐射势。注意到，辐射势是一个角度和频率决定的二维谱，它的完整性表示了散射场的完整性，也决定了散射势波数谱的完整性。

分析（2.9）式知，定义散射势的 Fourier 变换为，

$$\tilde{\mathcal{V}}(\mathbf{k}')=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-i\mathbf{k}'\cdot\mathbf{x}'} \tag{2.10}$$

其中，$\mathbf{k}'$代表散射势的波数。已知所有的波数成分等价于确定了精确的散射势。对比（2.9）和（2.10）知，$f(\mathbf{s},\mathbf{s}_0,\omega)=\tilde{\mathcal{V}}(k(\mathbf{s}-\mathbf{s}_0))$成立。据此，可以总结出如下定理。

定理：在 Born 近似下，相对于散射体的远场区域，在单位向量 **s** 所规定的方向上，散射波的振幅谱完全取决于散射势的某一个 Fourier 分量，即下式定义的分量

$$\mathbf{k}'=\frac{\omega}{c}(\mathbf{s}-\mathbf{s}_0) \tag{2.11}$$

由此可知，在常速介质情况下所谓逆散射问题就是根据观测到的散射波的振幅谱，确定散射势的 Fourier 的分量。散射势的所有的 Fourier 分量都确定了，通过 Fourier 逆变换，就可以完全精确地恢复散射势。

上述定理尽管是在平面波入射和常速背景情况下得到的，其理论指导意义是十分明显的。从$\mathbf{k}'$的表达式中可以看出，确定$\mathbf{k}'$的低波数部分，需要大角度散射波和低频数据；确定$\mathbf{k}'$的高波数部分，需要小角度散射波和高频数据。总之，震源子波的频带范围、观测系统和背景速度场的分布决定了反演成像的分辨率。更细致的分析见刘培培等（2013）的讨论。

2.2 高频近似下的 Born 近似的物理含义解释

Born 近似下的散射场也可以表示为，

$$u_s\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)=\omega^2\int d\mathbf{x}G_0\left(\boldsymbol{\xi},\mathbf{x},\omega\right)a\left(\mathbf{x}\right)G_0\left(\mathbf{x},\boldsymbol{\eta},\omega\right) \tag{2.12}$$

其中，速度扰动为$a\left(\mathbf{x}\right)=2c_0^{-3}\left(\mathbf{x}\right)\delta c\left(\mathbf{x}\right)$。用渐进 Green 函数（即 Green 函数的 WKBJ 近似解）代入上式，

$$u_s\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)=\omega^2\int\ d\mathbf{x}a\left(\mathbf{x}\right)A\left(\mathbf{x},\boldsymbol{\eta}\right)A\left(\boldsymbol{\xi},\mathbf{x}\right)\exp\left[-i\omega\left(T\left(\mathbf{x},\boldsymbol{\eta}\right)+T\left(\boldsymbol{\xi},\mathbf{x}\right)\right)\right] \tag{2.13}$$

在散射点周围，对旅行时进行线性化，即围绕中心射线进行 Taylor 展开，

$$T\left(\mathbf{x},\mathbf{x}'\right)\approx T\left(\mathbf{x}_0,\mathbf{x}'\right)+\left(\mathbf{x}-\mathbf{x}_0\right)\cdot\mathbf{p}\left(\mathbf{x}_0,\mathbf{x}'\right) \tag{2.14}$$

其中$\mathbf{p}\left(\mathbf{x}_0,\mathbf{x}'\right)$表示旅行时场$T\left(\mathbf{x},\mathbf{x}'\right)$在点$\mathbf{x}_0$处的梯度，其方向为中心射线的传播方向。因此，（2.13）式中的指数部分可以化为，

$$\begin{aligned}&\exp\left[-i\omega\left(T\left(\mathbf{x},\boldsymbol{\eta}\right)+T\left(\boldsymbol{\xi},\mathbf{x}\right)\right)\right]\\&=\exp\left[-i\omega\left(T\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+T\left(\boldsymbol{\xi},\mathbf{x}_0\right)\right)\right]\exp\left[-i\omega\left(\mathbf{x}-\mathbf{x}_0\right)\left(\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+\mathbf{p}\left(\boldsymbol{\xi},\mathbf{x}_0\right)\right)\right]\\&=\exp\left[-i\omega\left(T\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+T\left(\boldsymbol{\xi},\mathbf{x}_0\right)\right)\right]\exp\left[-i\mathbf{k}\cdot\left(\mathbf{x}-\mathbf{x}_0\right)\right]\end{aligned} \tag{2.15}$$

其中

$$\mathbf{k}=\omega\left(\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\xi}\right)\right) \tag{2.16}$$

代表波场的一种波数。其几何意义由图 2 所示。

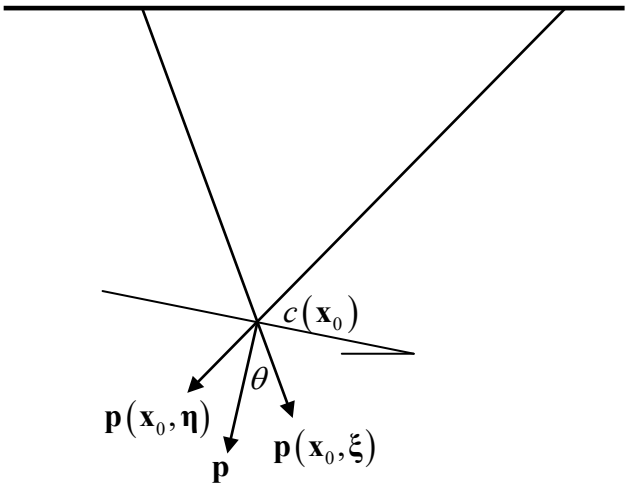


图 2 地震波传播过程中波矢量在反射点处的几何关系

如图 2 所示，$\mathbf{p}=\omega\left(\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\eta}\right)+\mathbf{p}\left(\mathbf{x}_0,\boldsymbol{\xi}\right)\right)=2\mathbf{n}\cos\theta/c\left(\mathbf{x}_0\right)$。其中$\mathbf{n}$为界面的外法线单位方向向量。$\theta$为射线的入射角。$c\left(\mathbf{x}_0\right)$为$\mathbf{x}_0$处的速度。根据上面的几何关系则有，

$$\mathbf{k}=\omega\mathbf{p}=\frac{2\omega\cos\theta}{c\left(\mathbf{x}_0\right)}\mathbf{n} \tag{2.17}$$

注意（2.17）式与（2.11）式的一致性。从$\mathbf{k}$的表达式中可以看出，确定$\mathbf{k}$的低波数部分，需要大角度散射波和低频数据；确定$\mathbf{k}$的高波数部分，需要小角度散射波和高频数据。值得说明的是，此处的波数$\mathbf{k}$与前面讨论的辐射势是关系密切的。因此，我们希望$\mathbf{k}$规定的谱的完整性。

此处结论与全波形速度反演得到认识是一致的。

3. 最小二乘散射势估计

3.1 散射势估计的几种基本方法

研究散射场表达的主要目的是建立起散射场与散射势之间的映射关系。上述映射关系实质上是一种广义 Radon 变换。散射场是投影结果，通过反投影来估计散射势就是逆散射问题。

逆散射问题的求解在不同应用领域存在较大的差异，主要看背景介质的复杂程度及相应的 Green 函数计算的复杂程度。相应地分为三大类。

（1）常速背景介质下，用平面波作为入射场，导出一类 Fourier 反变换方法估计$a\left(\mathbf{x}\right)$。勘探地震中，相应地存在$v\left(z\right)$介质（即水平层状介质）情况下的一类 Fourier 反变换方法。这是勘探地震领域上世纪 80 年代初关注的问题，仅仅为我们提供了对逆散射反演成像问题的理解的帮助。

（2）缓变光滑背景介质下，利用拟微分算子理论导出一类直接反演方法估计$a\left(\mathbf{x}\right)$。Bleistein（1985）在上世纪 80 年代中后期完整地讨论了此问题。Beylkin（1985）的研究工作更出色地解决了该问题。但这类方法基于拟微分算子理论，要求的数学基础非常高。

（3）任意变化的光滑背景介质情况下，基于 Bayes 估计理论，建立误差泛函，求解一类线性优化问题，估计$a\left(\mathbf{x}\right)$。这是目前勘探地震学中逆散射问题的主流解法。相对而言，这种解法与勘探地震学面对的介质变化情况最接近，解法也比较容易掌握和理解。

一般地，可以认为第二类是积分法的最小二乘叠前深度偏移成像方法，第三类是波动方程类的最小二乘叠前深度偏移成像方法。后者是目前研究和发展所侧重的。

3.2 拟微分算子理论下散射势的估计方法---积分法最小二乘叠前深度偏移成像

3.2.1 对偶空间、对偶算子与散射势估计的对偶表达

为了估计散射势$f\left(\mathbf{x}\right)$，把（1.16）式重写为，

$$\hat{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)=-\omega^2\int_X M\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)f\left(\mathbf{x}\right)d\mathbf{x} \tag{3.1}$$

其中，$M\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)$是积分核函数。它代表脉冲源激发的波从震源$\boldsymbol{\eta}$到散射点$\mathbf{x}$，再到接收点$\boldsymbol{\xi}$的波传播过程后的响应。

Fredholm 第一类积分方程的求解是很困难的，主要是因为这类积分方程所对应的矩阵方程在大部分情况下条件数都很大。对于这类积分方程也可以用如下的对偶空间理论来近似求解。在选定的基函数下面，数据空间和模型空间可以认为是一个对偶空间。Born 近似下的 Green 函数就定义了这样一个基函数，它们也形成了一个映射 $M\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)$，把模型空间函数 $f\left(\mathbf{x}\right)$ 映射到数据空间函数 $\hat{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)$。在对偶空间框架下存在映射 $M\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)$ 的对偶映射 $\hat{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)$，定义下式，

$$\left\langle u\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right),M\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)f\left(\mathbf{x}\right)\right\rangle_{S_D}\triangleq\left\langle f\left(\mathbf{x}\right),\hat{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)u\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)\right\rangle_{S_M} \tag{3.2}$$

此处，S_D 代表数据空间；S_M 代表模型空间。上式是广义的 Parsaval 定理。

基于上述对偶映射，可以直接地形式地写出扰动的估计式

$$f_{est}\left(\mathbf{x}\right)=2\,Re\int\limits_{\partial X_{rec}}\int\limits_{0}^{\infty}\hat{M}\left(\omega,\xi,\eta,x\right)\tilde{u}\left(\omega,\xi,\eta\right)d\xi d\omega \tag{3.3}$$

上式就是利用单炮数据估计参数扰动的公式。对于某些积分核函数，有

$$\bar{\hat{M}}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=\hat{M}\left(-\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right) \tag{3.4}$$

则（3.1.21）式就变为，

$$f_{est}\left(\mathbf{x}\right)=\int\limits_{\partial X_{rec}}\int\limits_{0}^{\infty}\hat{M}\left(\omega,\xi,\eta,x\right)\tilde{u}\left(\omega,\xi,\eta\right)d\xi d\omega \tag{3.5}$$

3.2.2 Born 近似下散射势的估计方法

用下面任意两个实函数以及它们的 Fourier 变换之间的关系，

$$\int\limits_{-\infty}^{+\infty}g\left(t\right)h\left(t\right)dt=\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\bar{\tilde{g}}\left(\omega\right)\tilde{h}\left(\omega\right)d\omega \tag{3.6}$$

因此，有

$$M\left(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=2\,Re\int\limits_{0}^{+\infty}\bar{\tilde{M}}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)e^{-i\omega t}dt \tag{3.7a}$$

以及

$$\tilde{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}M\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)e^{-i\omega t}dt \tag{3.7b}$$

由（3.3）、（3.5）式可知，扰动参数估计问题变为了确定投影算子或共轭投影算子问题。可以给出如下的投影算子，

$$M\left(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{8\pi^2}\delta\left(t-\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)-\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)\frac{h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)} \tag{3.8a}$$

上式是时间域形式，将其转为频率域形式可得

$$\tilde{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{16\pi^3}\frac{e^{-i\omega\left(\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)+\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)}}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right) \tag{3.8b}$$

其中，相位函数和振幅函数首先满足如下的 Green 函数的射线近似解，即 WKBJ 解，其中接收点对应的 Green 函数的解为

$$G^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)=-A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)e^{i\omega\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)} \tag{3.9a}$$

炮点对应的 Green 函数的解为

$$G^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)=-A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)e^{i\omega\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)} \tag{3.9b}$$

相位函数$\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$和$\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$满足如下的程函方程，

$$\left(\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\right)^2=S_0^2\left(\mathbf{x}\right) \tag{3.10a}$$

$$\left(\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\right)^2=S_0^2\left(\mathbf{x}\right) \tag{3.10b}$$

$\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$代表从炮点位置$\boldsymbol{\eta}$到散射点$\mathbf{x}$的在背景介质中的旅行时。$\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$代表从散射点$\mathbf{x}$到接收点$\boldsymbol{\xi}$的在背景介质中的旅行时。振幅函数$A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)$和$A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)$满足如下一阶输运方程，

$$A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\nabla_{\mathbf{x}}^2\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)+2\nabla_{\mathbf{x}}A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)\cdot\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)=0 \tag{3.11a}$$

以及

$$A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\nabla_{\mathbf{x}}^2\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)+2\nabla_{\mathbf{x}}A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\cdot\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)=0 \tag{3.11b}$$

旅行时的计算可以直接解方程（3.10）或把上式变成各种形式的一阶常微分方程组。关于振幅项对（3.11）式的求解直接在笛卡尔坐标系中进行会遇到焦散问题。Maslov 方法（在相空间中解方程（3.11））和 Gaussian_Beam 方法（在射线中心坐标系中用傍轴射线理论解方程（3.11））会回避掉此问题。其他各种变种的方法也在不断地发展中。

（3.8）中的函数$h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$是一个 Jacobian 变换，它把对$\mathbf{x}$的积分变换为对$\left(\boldsymbol{\xi},\boldsymbol{\eta}\right)$的积分。Beylkin（1985）给出

$$h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)d\xi=S_0^3\left(1+cos\,\psi\right)d\Omega \tag{3.12}$$

其中，Ω是立体角。且有

$$cos\,\psi\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)=\frac{\nabla_{\mathbf{x}}\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)\cdot\nabla_{\mathbf{x}}\phi^{in}\left(\mathbf{x},\boldsymbol{\xi}\right)}{S_0^2\left(\mathbf{x}\right)} \tag{3.13}$$

$\psi\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$是由坐标$\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)$和介质分布决定的波传播的反射张角，即入射射线和出射射线的夹角。

（3.8）式定义了变背景和任意炮检点分布时的偏移成像中的积分核函数。事实上，（3.8）式中的 Green 函数还可以不用 WKBJ 近似来表示，此时，推出的积分核意义更广泛。则有下式

$$\tilde{M}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{16\pi^3}\frac{\hat{G}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)}{\left|\hat{G}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)\right|}\frac{\hat{G}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)}{\left|\hat{G}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)\right|}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right) \tag{3.14}$$

在时间域中，上式变为，

$$M\left(t,\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}\right)=-\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\breve{G}^{out}\left(t-\tau,\boldsymbol{\xi},\mathbf{x}\right)\breve{G}^{in}\left(t-\tau,\boldsymbol{\eta},\mathbf{x}\right)h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)d\tau \tag{3.15}$$

其中

$$\breve{G}^{out}\left(t,\boldsymbol{\xi},\mathbf{x}\right)=2\,Re\int_{-\infty}^{+\infty}\frac{G^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)}{\left|G^{out}\left(\omega,\boldsymbol{\xi},\mathbf{x}\right)\right|^2}e^{-i\omega t}d\omega \tag{3.16a}$$

和

$$\breve{G}^{in}\left(t,\boldsymbol{\eta},\mathbf{x}\right)=2\,Re\int_{-\infty}^{+\infty}\frac{G^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)}{\left|G^{in}\left(\omega,\boldsymbol{\eta},\mathbf{x}\right)\right|^2}e^{-i\omega t}d\omega \tag{3.16b}$$

可以写出如下两种不同的积分形式，对应两种不同的偏移成像的说明。形式 I 为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{8\pi^2}\int_{\partial X_{rec}}u\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)\Big|_{t=\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)+\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}\frac{h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}d\boldsymbol{\xi}d\omega \tag{3.17}$$

很显然，上式就是 Kirchhoff 积分的绕射叠加。就是说，从 Born 近似理论也可以推出 Kirchhoff 积分绕射叠加偏移公式。

把（3.8b）或（3.14）式代入到（3.15）式可以得到另一种形式的偏移成像公式。形式 II 为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{16\pi^3}Re\int_0^{\infty}\frac{e^{-i\omega\phi^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}}{A^{in}\left(\mathbf{x},\boldsymbol{\eta}\right)}\int_{\partial X_{rec}}\frac{e^{-i\omega\phi^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}}{A^{out}\left(\mathbf{x},\boldsymbol{\xi}\right)}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)\tilde{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\omega \tag{3.18a}$$

以及

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{16\pi^3}Re\int_0^{\infty}\frac{\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\eta}\right)}{\left|\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\eta}\right)\right|^2}\int_{\partial X_{rec}}\frac{\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\xi}\right)}{\left|\tilde{G}\left(\omega,\mathbf{x},\boldsymbol{\xi}\right)\right|^2}h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)\tilde{u}\left(\omega,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\omega \tag{3.18b}$$

（3.1.36）对应的时域形式为

$$f_{est}\left(\mathbf{x}\right)=-\frac{1}{8\pi^2}\int_0^{\infty}G^{in}\left(t,\mathbf{x},\boldsymbol{\eta}\right)w\left(t,\mathbf{x},\boldsymbol{\eta}\right)dt \tag{3.19}$$

其中

$$w\left(t,\mathbf{x},\boldsymbol{\eta}\right)=\int_0^{\infty}\int_{\partial X_{rec}}G^{out}\left(t-\tau,\mathbf{x},\boldsymbol{\xi}\right)h\left(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\eta}\right)u\left(\tau,\boldsymbol{\xi},\boldsymbol{\eta}\right)d\boldsymbol{\xi}d\tau \tag{3.20}$$

（3.19）式的含义就是（3.20）式进行检波点波场反传播，然后与下行波场进行相关成像。此为 Clearbout（1985）的成像条件。

（3.17）、（3.18）和（3.19）分别是不同形式的叠前偏移成像表达式。成像结果是地下介质的扰动（散射势）。上述三式用更深刻的数学背景阐述了地震波叠前偏移成像的本质。它就是 Born 近似下线性化反演的近似解。

值得注意的是，上述理论意义下的得到叠前深度偏移结果，就是对速度扰动量的估计。不是我们常认为的带限的反射系数的估计。

3.3 Bayes 框架下散射势估计方法---波动方程最小二乘叠前深度反演成像

（1.16）式定义了一个 Fredholm 第一类积分方程，它描述的是一个线性问题。此处，我们把该 Fredholm 第一类积分方程的求解转化为一个 Bayes 估计问题。就是把该线性问题的求解，转化为利用泛函变分求极值问题。

基本的表述如下。假设（1.16）式定义的积分方程写出算子表达形式，

$$K\mathbf{m}=\mathbf{d} \tag{3.21}$$

其中，K 是 Fredholm 第一类积分方程的积分核，由（1.16）式知，积分核的定义为

$$K\left(v_0(\mathbf{x}),\mathbf{x},\omega\right)=\omega^2 G_0(\boldsymbol{\xi},\mathbf{x},\omega)u_0(\mathbf{x},\boldsymbol{\eta},\omega) \tag{3.22}$$

$\mathbf{m}$ 代表由 $a(\mathbf{x})\big/v_0^2(\mathbf{x})$ 作为元素形成的矢量；$\mathbf{d}$ 代表由每个炮检点对应的地震道的频谱值。

认为 $K:\mathrm{H}_1\rightarrow\mathrm{H}_2$ 是一个有界的线性算子，定义如下的 Tikhonov 泛函，

$$\mathfrak{I}(\mathbf{m})=\frac{1}{2}\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2 \tag{3.23}$$

对上式求解的首要问题是计算泛函关于参数变化的导数，即泛函的梯度。为此，利用如下的变分方法。

对任意的介质扰动 $\delta\mathbf{m}\in\mathrm{H}_1$ 和扰动步长 $\tau\in\mathbb{R}$，总存在如下的泛函变分

$$\begin{aligned}
&\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})=\frac{1}{2}\left\|K(\mathbf{m}+\tau\delta\mathbf{m})-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\left\|\mathbf{m}+\tau\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\\
&=\frac{1}{2}\left\langle K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\frac{\alpha}{2}\left\langle\mathbf{m}+\tau\delta\mathbf{m},\mathbf{m}+\tau\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\frac{1}{2}\left\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\tau^2}{2}\left\langle K\delta\mathbf{m},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}\\
&+\frac{\alpha}{2}\left\langle\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)+\frac{\alpha\tau^2}{2}\left\langle\delta\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}
\end{aligned}$$

$$
\begin{aligned}
&=\frac{1}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle \mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\left\langle \mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle \delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\left\langle K\delta\mathbf{m},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle \delta\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\frac{1}{2}\left(\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\\
&+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\left\langle \mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle \delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\left\|K\delta\mathbf{m}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)
\end{aligned}
\tag{3.24}
$$

对上式定义的变分求偏导数，

$$
\begin{aligned}
&\frac{\partial}{\partial\tau}\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right)\Big|_{\tau=0}\\
&=\left\{\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle \mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)+\tau\left(\left\|K\delta\mathbf{m}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\right\}\Big|_{\tau=0}\\
&=\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle \mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\left(\left\langle K^*\left(K\mathbf{m}-\mathbf{d}\right),\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\alpha\left\langle \mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\left\langle K^*\left(K\mathbf{m}-\mathbf{d}\right)+\alpha\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\left\langle \left(K^*K+\alpha I\right)\mathbf{m}-K^*\mathbf{d},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\left\langle \mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right),\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}
\end{aligned}
\tag{3.25}
$$

其中

$$
\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)=\left(K^*K+\alpha\mathbf{I}\right)\mathbf{m}-K^*\mathbf{d} \tag{3.26}
$$

式（3.26）是误差泛函相对于参数模型的梯度。K^* 是波场正传播算子的共轭算子，代表波场的反传播。经典地，我们利用该梯度构造最速下降迭代方法求解 Born 近似后的线性问题的解。迭代格式为，

$$
\begin{aligned}
&\mathbf{m}^{k+1}=\mathbf{m}^k-\mu\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}-\mathbf{d}^{obs}\right)\right]=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}^k-\mathbf{d}^{obs}\right)\right]\\
&=\mathbf{m}^k-\mu\left[K^*\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)\right]
\end{aligned}
\tag{3.27}
$$

利用上述迭代公式进行最小二乘叠前深度反演成像是现代勘探地震对成像精度提高的要求，并基于计算机能力的大幅度提高。

一般的最小二乘叠前深度偏移成像的基本做法是，令梯度等于零（实际上一般不会等于零!）时，可以得到线性系统的法方程

$$
\left(K^*K+\alpha\mathbf{I}\right)\mathbf{m}=K^*\mathbf{d} \tag{3.28}
$$

法方程的求解一般用共轭梯度方法。但是，由于勘探地震中模型描述用网格化方式，模型参数个数非常多，

（3.28）式右端的 Hessian 矩阵异常庞大。若用 $NX\times NY\times NZ$ 表示要估计的模型参数的个数，那么 Hessian 矩阵的规模为 $\left(NX\times NY\times NZ\right)\times\left(NX\times NY\times NZ\right)$。当前的计算机技术还不能满足这样巨大规模的线性方程组的求解。形成 Hessian 矩阵本身就是一个巨大的挑战。所以，**当前解法方程进行最小二乘偏移成像的基本做法是先进行常规的叠前深度偏移成像，然后用对角矩阵元素（就是地震波照明能量!）对成像叠加后的结果进行照明补偿，就是用 Hessian 矩阵的对角元素去除成像结果**。也可以用角度 Hessian 的对角元素去除每个共角度成像剖面，然后进行叠加产生最后的剖面，当然也可以不叠加就此产生照明补偿后的角度成像道集。最一般的做法应该对每个单炮道集进行照明补偿的最小二乘偏移，但这样做会由于弱照明产生很强的成像噪音（实质上是反问题的计算不稳定性）。另外，很多人试图把成像限制在局部范围内，在小的空间范围内实现全 Hessain 求逆进行最小二乘反演。

解法方程进行最小二乘叠前深度偏移有两个问题：法方程本身是误差泛函关于模型参数扰动的导数等于零和全 Hessian 的逆无法计算。而迭代法不存在这两个问题，这也解释了迭代类的最小二乘叠前深度偏移成像方法是今后叠前深度偏移发展方向的原因。

迭代法涉及残差计算 $\mathbf{r}^{k+1}=K\mathbf{m}^{k}-\mathbf{d}^{obs}$，需要一步反偏移；$K^{*}\mathbf{r}^{k+1}=K^{*}\left(K\mathbf{m}^{k}-\mathbf{d}^{obs}\right)$ 代表残差的反传播，需要一步波的反向传播。同样地，要考虑步长的计算、考虑用 Hessian 矩阵的逆预条件梯度，这四部分是一步迭代需要的计算量。一般地需要 n 步迭代计算。因此，迭代法的计算量远大于解法方程的计算量。

另外，值得讨论的是，（3.21）式中的积分核函数 K 对不同的波现象的处理形式及含义是不同的，基本地它代表波传播路径及沿路径的波场分布，但对于反射波和透射波（直达波、折射波和 Diving 波）而言，二者的计算方式是不同的，对于透射波而言，仅仅已知背景速度就可以计算 K；对于反射波而言，必须知道反射界面的位置才可以计算 K。反射波和透射波计算积分核函数 K 的差异导致利用反射波和透射波反演背景速度的方式的差异。（3.27）和（3.28）都是对透射波有效的反演算法。就是说最小二乘偏移成像的算法对于初至波层析成像是一样的。

迭代的最小二乘反演方法中，每次模型更新后，原则上都要重新计算核函数 K，并计算波场 $\mathbf{d}^{cal}$。重新计算核函数 K，需要计算出更新后的速度场，可以用下式进行

$$v\left(\mathbf{x}\right)=v_0\left(\mathbf{x}\right)\Big/\sqrt{1-\mathbf{m}\left(\mathbf{x}\right)v_0^2\left(\mathbf{x}\right)} \tag{3.29}$$

然而在很多时候，往往并不更新核函数，因为我们假定核函数 $K\left(v_0\left(\mathbf{x}\right),\mathbf{x},\omega\right)$ 仅仅与背景速度 $v_0\left(\mathbf{x}\right)$ 有关。但这仅仅是从计算效率方面考虑的，本质上还是希望更新核函数。可能的计算策略当然包括间隔相邻的几个迭代步更新一次核函数。射线层析的迭代计算原则也是这样的。

（3.27）式中计算出的量 $a\left(\mathbf{x}\right)\Big/v_0^2\left(\mathbf{x}\right)$ 或者 $2\delta v\left(\mathbf{x}\right)\Big/v_0^3\left(\mathbf{x}\right)$ 代表的是速度的一种相对变化。只有在速度跃变的界面或绕射点处才会出现，因此，它反映了速度场中的高频变化部分。用这部分高波数速度扰动可以代表界面位置和绕射位置图像。这与一般的叠前深度偏移是很类似的。也可以精确估计这部分高波数速度扰动，用它们来刻画储层的性质。但是，我们认为，只有全波数带的速度场或合理的宽带速度场才可以进行有效的储层解释。这也是 FWI 技术用于储层解释时存在的最大问题，既不能提供合理有效的宽带速度场。

另一方面，在做数据空间的波场逼近时，还要考虑 $\mathbf{d}^{cal}$ 的规则化问题，或者统一两个数据的空间采样方式。

3.4 层析成像中的残差反传播

迭代最小二乘波形层析成像的基本思想可以从（3.27）式中清楚地看出。就是通过共轭算子 K^{*}，把利用 $v_0\left(\mathbf{x}\right)$，$a\left(\mathbf{x}\right)\Big/v_0^2\left(\mathbf{x}\right)$ 和 Green 函数计算出的预测波场和观测散射场的残差进行反传播，得到速度扰动的

估计。这个问题就是逆散射成像问题，也是最小二乘叠前深度偏移成像问题。

实际上，速度场可被分为两部分：背景部分$v_0(\mathbf{x})$和扰动部分$a(\mathbf{x})$。我们希望，与射线层析沿着射线路径进行旅行时残差反投影一样，沿着波路径进行波场残差的反投影来估计速度场的背景部分$v_0(\mathbf{x})$。

估计速度场的背景部分需要利用波场中的不同偏移距的到达时信息或波场中的相位信息，它们是波传播的累计效应。一般基于 Radon 变换的反投影进行的层析成像利用的都是这一类的累计信息（积分结果），它们对要反演参数的高频变化是不敏感的。勘探地震学中，近地表速度场的估计就是利用的直达波、折射波和/或 Diving 波的到达时进行的。井中透射波的层析成像更是利用的非反射波的到达时。

反射波的层析成像和透射波的层析成像是存在差异的！

油气地震勘探中，中深层的地震波主要是反射波，反射波的旅行时累积的计算需要已知反射点位置，但它是未知的。这与透射波的旅行时计算是不同的。就是说，反射波旅行时层析时需要知道高波数速度异常体的位置，它需要同时或串联反演背景速度和反射界面位置参数。

前已述及，透射波层析反演使用的就是（3.27）式或（3.28）式以及与之相对应的一阶 Born 近似下的场的表达。

反射波的层析反演如何做？这是一个值得讨论的问题。我们已经提出要用二阶 Born 近似表达的波场以及相应的反演公式。

估计速度场的高波数扰动部分需要利用地震波的振幅信息，即波形变化信息。我们知道，正是反射界面处或散射点处的速度高波数变化（实际上是波阻抗变化）引起了波场中的同相轴。把波场中的波形残差反投影到反射界面上或散射点处需要正确的背景速度。前已述及，反射波层析估计背景速度需要正确的高波数速度异常体的位置信息，确定正确的位置信息需要正确的背景速度。这是耦合在一起的！

勘探地震学主要是利用地震反射波进行反演成像，这是它非常不同于一般的无损探伤成像、医学层析成像的根本物理原因。

利用地震反射波进行反演成像需要反射波层析成像（估计速度场的低波数变化）与反射波叠前偏移成像（估计速度场的高波数变化）的联合迭代。

反射波的残差反投影有两种形式，一是把残差反投影到等时面上，修改反射界面处的高波数参数扰动；二是把残差沿波传播路径进行反投影，修改波传播路径上的背景参数扰动。前者对应一般的叠前深度偏移，包括 RTM；后者类似一般的旅行时层析，包括沿路径投影的波形层析。前者估计速度场的高波数成分；后者估计速度场的低波数成分。后者是前者的基础，必须很好地解决后者，前者才能得到较好的解决，根本原因是低波数速度成分与低频波场和旅行时之间存在更为线性化的关系，反演低波数成分相对容易进行。

反射波的残差反投影的这两种形式解释了叠前深度偏移和波形层析成像的本质，也说明了它们的差异与联系。我们认为当残差反投影沿着等时面进行时，就是在做叠前深度偏移；当残差反投影沿着波路径进行时就是在做波形层析。

为了沿路径反传播残差，首先需要在数据中引入波传播的方向，把地震数据修改成$u_s(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s,\omega)$，该数据体称为立体地震数据，它是局部平面波域表达的叠前地震数据体。我们也称该数据体为特征波场数据。此时，（1.16）式被描述为

$$
\begin{aligned}
u_s\left(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s;\omega\right)&=\omega^2\int d\mathbf{x}G_0\left(\boldsymbol{\xi},\mathbf{p}_r;\mathbf{x};\omega\right)\frac{a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}u_0\left(\mathbf{x};\boldsymbol{\eta},\mathbf{p}_s;\omega\right)\\
&=\int d\mathbf{x}K\left(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s;\mathbf{x};\omega;v_0\left(\mathbf{x}\right)\right)\frac{a\left(\mathbf{x}\right)}{v_0^2\left(\mathbf{x}\right)}\\
&=\int d\mathbf{x}K\left(\boldsymbol{\xi},\mathbf{p}_r;\boldsymbol{\eta},\mathbf{p}_s;\mathbf{x};\omega;v_0\left(\mathbf{x}\right)\right)f\left(\mathbf{x}\right)
\end{aligned}
\tag{3.30}
$$

其中，$f\left(\mathbf{x}\right)=a\left(\mathbf{x}\right)\Big/v_0^2\left(\mathbf{x}\right)$为要反演的参数。

相应地，还要构建沿波路径进行投影和反投影的波传播算子。这是空间局部波传播算子被提出来研究的主要原因。

最后，我们对（1.16）式再做一个分析。首先它可以反映源点$\boldsymbol{\eta}$到地下任一点（散射点）$\mathbf{x}$到接收点$\boldsymbol{\xi}$的一次波传播关系，但是散射点$\mathbf{x}$的位置是未知的，利用（1.16）式进行背景速度的层析反演时，必须引入散射点$\mathbf{x}$的位置，才可以确定要反投影的波路径。利用（1.16）式进行高波数速度扰动量的定位和估计时，需要已知背景速度$v_0\left(\mathbf{x}\right)$和基于它计算处的入射场 Green 函数和散射场 Green 函数，两 Green 函数相遇时，$\mathbf{x}$点必然应该是高波数速度扰动量的出现位置。但仅仅已知相遇时间，导致高波数速度扰动量的出现位置位于等时面上；若利用特征数据体（或立体数据体）及空间局部投影算子，则可以把波场残差投影在反射或绕射点上。至于高波数速度扰动量的定量估计由下一节的最小二乘叠前深度偏移方法实现。

3.5 对非线性的最小二乘逆时深度偏移成像（LS_RTM）的分析

非线性的最小二乘逆时深度偏移成像（LS_RTM）是现代勘探地震高保真成像着力发展的方法技术。我们首先列出非线性的最小二乘逆时深度偏移成像（LS_RTM）所用的公式，然后对其进行分析。

具体实现时的基本公式罗列如下。

正向波场外推方程：

$$
\begin{cases}
\left(\dfrac{1}{v^2\left(x\right)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)u\left(\mathbf{x},t\right)=-\delta\left(\mathbf{x}-\mathbf{x}_S\right)S\left(t\right)\\
u\left(\mathbf{x},t\right)\big|_{\mathbf{x}=\Omega}=f_b\left(\Omega,t\right)
\end{cases}
\tag{3.31a}
$$

其中，$f_b\left(\Omega,t\right)$代表边界条件。一般用 PML 边界条件。Ω代表计算波场在空间上的边界。

反向波场外推方程：

$$
\begin{cases}
\left(\dfrac{1}{v^2\left(x\right)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)z\left(\mathbf{x},t\right)=0\\
z\left(\mathbf{x},t\right)\big|_{\mathbf{x}=\boldsymbol{\xi}}=u\left(\boldsymbol{\xi},t\right)
\end{cases}
\tag{3.31b}
$$

其中，$\boldsymbol{\xi}$代表观测点坐标值。$u\left(\boldsymbol{\xi},t\right)$代表观测点处的波场。FWI 进行波场残差外推时，仅有观测点处有波场残差，因此，才有（3.31b）这样的边界条件的定义。实际上，波场逆时外推并不限于这样的边界条件。

可见，（3.31a）是一个初值问题，沿时间正向外推计算；（3.31b）是一个边值问题，逆时间反向外推计算。

梯度公式为：

$$\mathbf{g}\left(\mathbf{m}\left(\mathbf{x}\right)\right)=\sum_{t}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\frac{1}{v^{2}\left(\mathbf{x}\right)}\frac{\partial^{2}u(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta})}{\partial t^{2}}z(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta}) \tag{3.32}$$

迭代公式为：

$$\mathbf{m}^{k+1}=\mathbf{m}^{k}-\mu\mathbf{g}\left(\mathbf{m}\left(\mathbf{x}\right)\right) \tag{3.33}$$

与前述基于一阶 Born 近似公式导出的 LS_PSDM 相比，LS_RTM 的梯度项中包含了对所有波现象的预测。当背景速度不准时，不能预测的波现象会干扰梯度的计算，导致收敛慢，甚至不收敛的情况。比较而言，一阶 Born 近似仅仅考虑一次散射波的预测，相对而言，梯度计算要更准确，使得 LS_PSDM 的收敛性更好。

更深入的分析有待进一步的理论公式的推导和数值实验的结合。

3.6 基于 Born 线性近似和逆散射成像理论的最小二乘反射波叠前偏移成像的具体实现方法

首先，必须有一个界定，一般地我们把利用波场传播的累计效应（即旅行时）的估计背景速度的方法称为层析成像；把估计速度场中的高波数成分的处理称波动方程叠前深度偏移，它一定是在背景速度已经比较准确已知的情况下，反演反射点处的速度场的高波数扰动。

即便做了如此界定，一般意义下的叠前深度偏移和估计估计速度场高波数成分的叠前深度偏移还是有差异的。常规意义下，我们认为偏移成像是用反传算子外推到成像点（或绕射点/反射点）上的波场来定位绕射点/反射点出现的位置，并不估计反射系数或$a\left(\mathbf{x}\right)$。Kirchhoff 积分偏移成像代表一类绕射叠加偏移方法，更清楚地反映了很多时候我们根本没有引入散射势的概念，同样进行叠前深度偏移。事实上，严格的叠前深度反演成像是把速度场的高波数成分的估计提为一个线性化的反演问题，用最小二乘方法实现。

下面我们具体讨论基于（1.16）式的最小二乘反射波叠前深度偏移成像方法。

为估计散射势，定义如下误差泛函，

$$\mathcal{J}=\frac{1}{2}\left\|u_{s}^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)-u_{s}^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right\|_{2}^{2}=\frac{1}{2}\sum_{\omega}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\left(u_{s}^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)-u_{s}^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right)^{2} \tag{3.34}$$

误差泛函能量最小化时，认为估计的散射势是所求的参数估计结果。其中$u_{s}^{cal}\left(\xi,\eta,\omega\right)$有（1.16）式定义。

目标泛函的梯度是构造最速下降法的核心。该梯度定义为

$$\begin{aligned}\frac{\partial\mathcal{J}}{\partial\mathbf{m}_{r}\left(\mathbf{x}\right)}&=\sum_{\omega}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\frac{\partial u_{s}^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)}{\partial\mathbf{m}_{r}}\left[u_{s}^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)-u_{s}^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right]^{H}\\&=\sum_{\omega}\omega^{2}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}S\left(\omega\right)G_{S}\left(\mathbf{x},\boldsymbol{\eta},\omega\right)G_{R}\left(\boldsymbol{\xi},\mathbf{x},\omega\right)\left[u_{s}^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)-u_{s}^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right]^{H}\end{aligned} \tag{3.35}$$

其中$S(\omega)G(\mathbf{x},\boldsymbol{\eta},\omega)$代表震源点$\boldsymbol{\eta}$传到地下任一点$\mathbf{x}$的场，$G_{R}\left(\boldsymbol{\xi},\mathbf{x},\omega\right)\left[u_{s}^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)-u_{s}^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right]^{H}$代表误残差波场的反传波。

利用（3.35）定义的梯度可以进行最速下降法的最小二乘偏移成像的具体计算，

$$\mathbf{m}_{r}^{(k+1)}\left(\mathbf{x}\right)=\mathbf{m}_{r}^{(k)}\left(\mathbf{x}\right)+\mu^{(k)}\delta\mathbf{m}_{r}^{(k)}\left(\mathbf{x}\right)=\mathbf{m}_{r}^{(k)}\left(\mathbf{x}\right)+\mu^{(k)}\frac{\partial\mathcal{J}}{\partial\mathbf{m}_{r}\left(\mathbf{x}\right)} \tag{3.36}$$

为了加速收敛，可以用 Hessian 矩阵的逆对梯度方向进行预条件处理，Hessian 矩阵的定义为，

$$\begin{aligned}\mathbf{H}(\mathbf{x},\mathbf{y})&=\frac{\partial^2\mathcal{J}}{\partial\mathbf{f}(\mathbf{x})\partial\mathbf{f}(\mathbf{y})}=\sum_{\omega}\omega^4\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}S(\omega)G_S(\mathbf{x},\boldsymbol{\eta},\omega)G_R(\boldsymbol{\xi},\mathbf{x},\omega)\left[\frac{\partial u_S^{cal}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)}{\partial\mathbf{m}_r(\mathbf{y})}\right]^H \\ &=\sum_{\omega}\omega^4\sum_{\boldsymbol{\eta}}S^2(\omega)G_S(\mathbf{x},\boldsymbol{\eta},\omega)G_S^H(\mathbf{y},\boldsymbol{\eta},\omega)\sum_{\boldsymbol{\xi}}G_R(\boldsymbol{\xi},\mathbf{x},\omega)G_R^H(\boldsymbol{\xi},\mathbf{y},\omega)\end{aligned} \tag{3.37}$$

该 Hessian 矩阵是由背景介质中的 Green 函数规定的，其中并不包括高阶反射/绕射效应，即不包括非线性部分。双向波计算 Green 函数可以包含这部分的作用。

由 Hessian 矩阵的逆对梯度方向进行预条件处理后，最速下降法的迭代格式变为，

$$\mathbf{m}_r^{(k+1)}(\mathbf{x})=\mathbf{m}_r^{(k)}(\mathbf{x})+\mu^{(k)}\delta\mathbf{m}_r^{(k)}(\mathbf{x})=\mathbf{m}_r^{(k)}(\mathbf{x})+\mu^{(k)}\mathbf{H}^{-1}\frac{\partial\mathcal{J}}{\partial\mathbf{m}_r(\mathbf{x})} \tag{3.38}$$

迭代步长 $\mu^{(k)}$ 一般可以采用试探的方式进行确定，一般采用试探加抛物拟合来选取最优下降步长来保证误差的下降。

迭代的每一步需要在更新后对模型进行一次反偏移，然后将反偏移得到的正演数据从原始数据中减去，这样来判断误差是否达到要求，然后判断迭代是否终止。

3.7 数值结果

图 3 展示了用公式（3.38）进行的最小二乘反射波叠前偏移成像的结果。可以看出，最小二乘偏移补偿了照明不足，提高了成像的分辨率，是成像结果的横向一致性更好。这样的结果更有利于后续的储层解释。

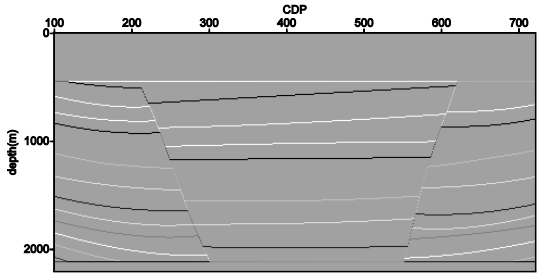


图 3a 反射系数模型

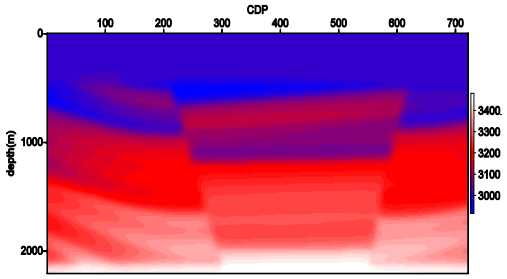


图 3b 背景速度模型

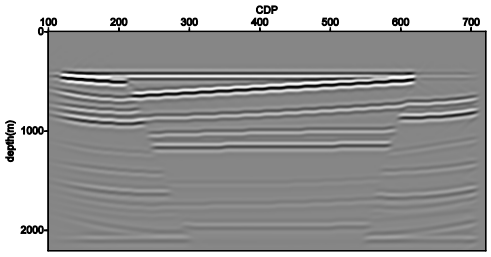


图 3c 一般叠前深度偏移结果

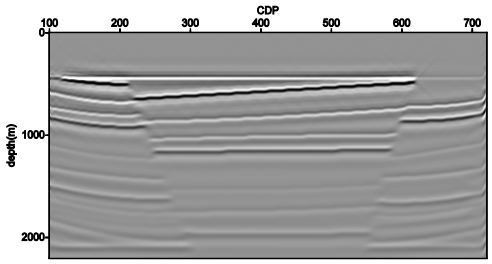


图 3d 迭代 20 次的叠前深度偏移结果

图 3 迭代法最小二乘叠前深度偏移成像结果

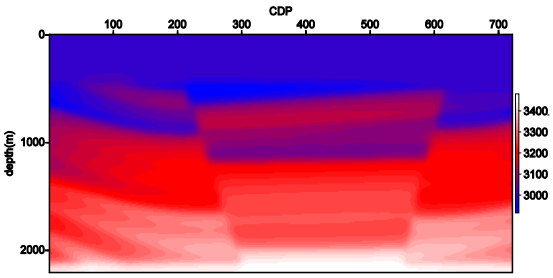


图 4a 初始速度模型

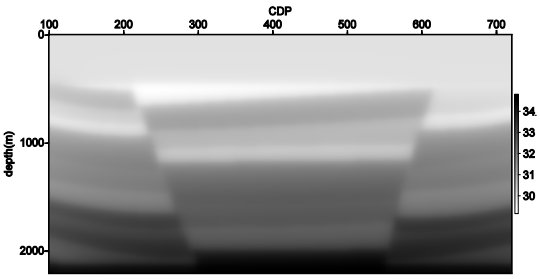


图 4b 初始 Q 值模型

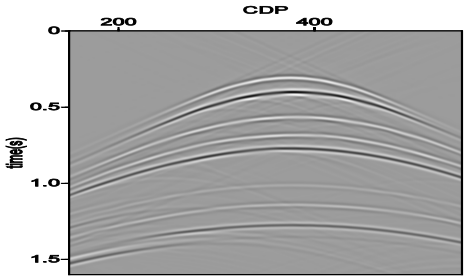


图 4c 声介质单炮道集

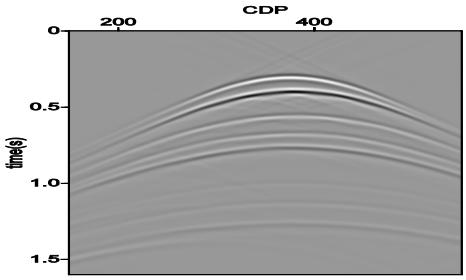


图 4d 粘声介质单炮道集

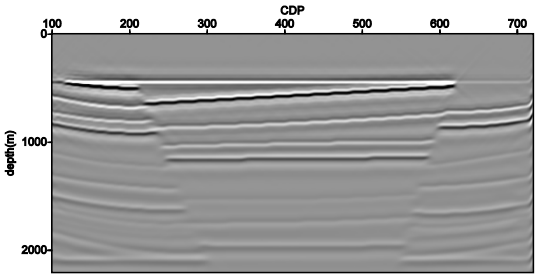


图 4e 声介质迭代 20 次的叠前深度偏移结果

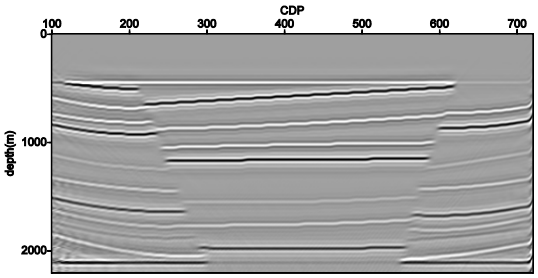


图 4f 粘声介质迭代 20 次的叠前深度偏移结果

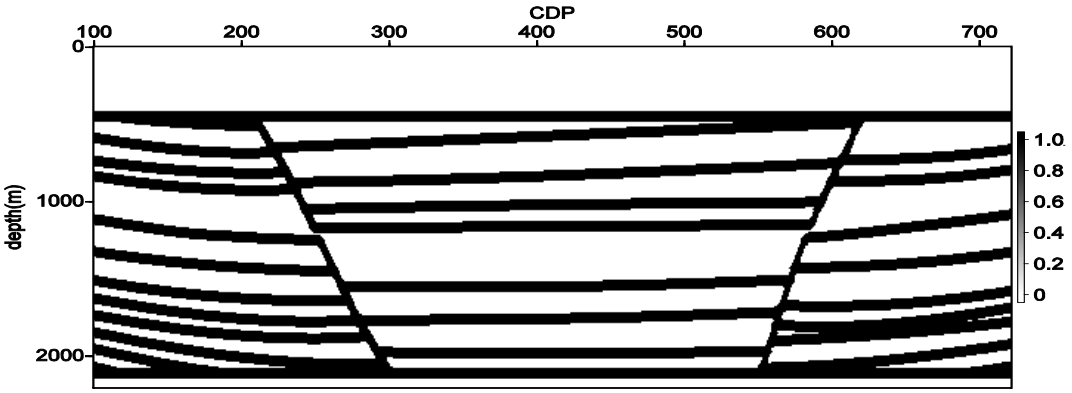


图 4g 正则化最小二乘偏移成像的加权系数

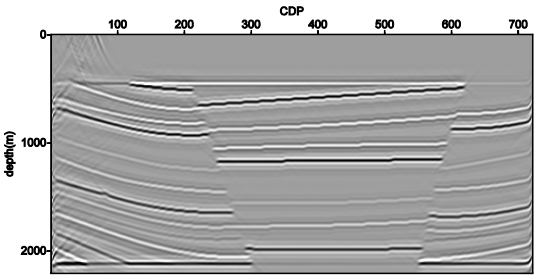


图 4h 粘声介质迭代 20 次的 Tkhonov 正则化叠前深度偏移结果

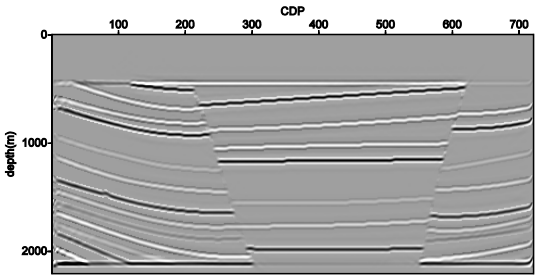


图 4i 粘声介质迭代 20 次的加权正则化叠前深度偏移结果

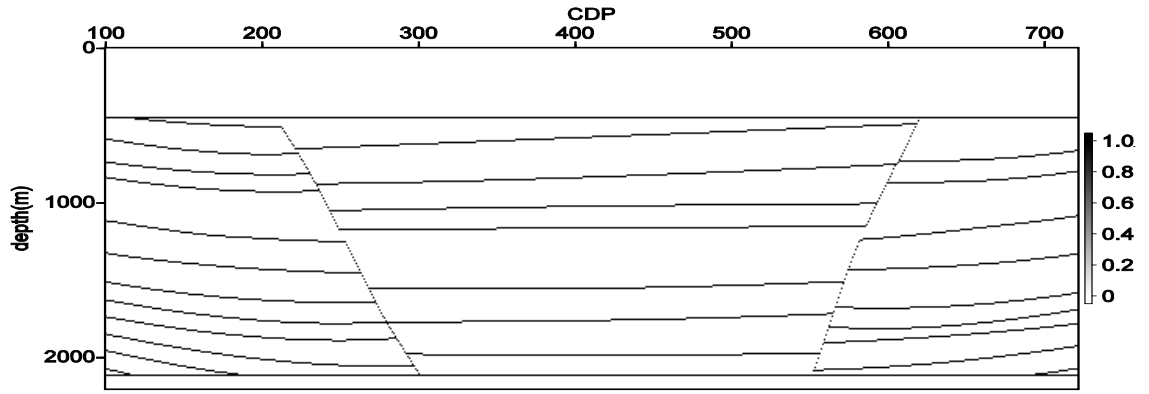


图 4j 正则化最小二乘偏移成像的加权系数

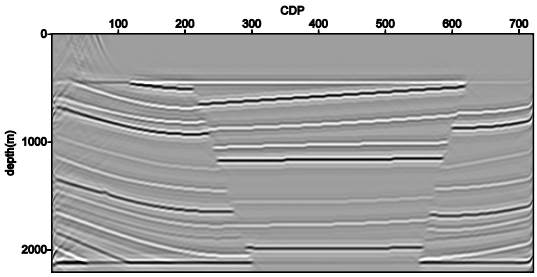


图 4k 粘声介质迭代 20 次的 Tkhonov 正则化叠前深度偏移结果

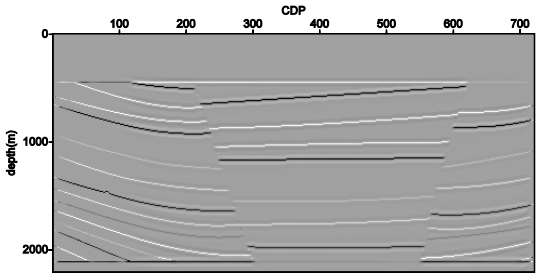


图 4l 粘声介质迭代 20 次的加权正则化叠前深度偏移结果

图 4 展示了声介质和粘声介质下最小二乘叠前深度偏移的对比，以及 Tkhonov 正则化和加权正则化情形下迭代最小二乘叠前深度偏移的对比，以及不同加权系数下迭代最小二乘叠前深度偏移的对比。可以看出，最后的粘声介质迭代 20 次的加权正则化叠前深度偏移结果与真实的模型的反射系数已经非常接近了。理论上讲，如此高分辨率和高保真的叠前深度成像结果完全胜任岩性油气藏的精细刻画的要求。当然，对实际数据而言，成像结果不会如此理想。但这并不改变上述理论预测的正确性。

总结与讨论

地震波在复杂介质中传播的描述方法决定了后续的叠前深度偏移成像方法和层析反演方法。著名的描述方法是 Lippmann-Schwinger 方程以及基于此的散射场表达方法。Born 线性近似及对应的逆散射成像在勘探地震学及相关的学科中，譬如无损探伤成像、医学成像等，都有广泛的应用。

但是，勘探地震学中的逆散射成像不同于其他学科的逆散射成像，重大区别在于油气地震勘探中遇到的波现象以反射波为主，此时的逆散射成像方法必须用入射波与散射波相关的方式把波场残差放在反射界

面上或绕射点上。如果用散射理论进行层析成像，需要沿波路径进行残差的反投影。

现代勘探地震学重点发展的技术就是高维弹性参数反演，用来进行岩性储层的精细描述。本文重点讨论了散射场的表达，以及基于此表达的最小二乘叠前深度偏移方法。深入地分析了利用透射波、反射波进行层析速度反演、最小二乘叠前深度偏移/反演成像的本质含义。发展了实用的最小二乘叠前深度偏移/反演成像，大量的数值实验验证了该技术具有高保真高分辨率成像的能力。

另外，一般意义下的叠前深度偏移和反射波的逆散射成像是有明显区别的。但目前二者都在使用，这对地震成像结果的解释是有重要影响的，需要进一步的专文分析。

参考文献

[1] 王华忠、冯波、王雄文，2013，面向岩性油气藏的地震波反演成像理论，波现象与反演成像研究组（WPI）2013 年度研究文集，同济大学海洋与地球科学学院

[2] 刘培培、胡江涛、王华忠，2013，基于 Hessian 算子的地震成像分辨率分析，波现象与反演成像研究组（WPI）2013 年度研究文集，同济大学海洋与地球科学学院

[3] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, 26, 99–108.

[4] Bleistein, N., 1987, On the imaging of reflectors in the earth. Geophysics, 52, 931-942

[5] Bleistein, N., Cohen, J. K., and Stockwell, J. W., Jr., 2000, Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion;

[6] Bleistein, N., Cohen, J. K., and Hagin, F., 1985, Two and one-half dimensional Born inversion with an arbitrary reference, Geophysics, 52, 26-36.

[7] Clayton, R. and Stolt, R., 1981, A Born-WKBJ inversion method for acoustic reflection data; Geophysics, 46, 1559-1567

[8] Cohen, J. K., Hagin, F. and Bleistein, N., 1986, Three dimensional Born inversion with an arbitrary reference, Geophysics, 52, 26-36.

[9] Devaney, A. J., 2012, Mathematical Foundations of Imaging, Tomography and Wavefield Inversion; Cambridge

[10] Lambar, G., Operto, S., Podvin, P. and Thierry, P., 2003, 3D ray+Born migration/inversion, Part I: Theory. Geophysics, 68, 1348-1356

[11] Raz, S., 1981, Three-Dimensional velocity profile inversion from finite-offset scattering data; Geophysics, 50, 110-126

[12] Shuey, R., 1985, A simplification of the Zoeppritz equation; Geophysics, 50, 609-614

[13] Stolt, R., H., and Weglein, A. B., 1985, Imaging and Inversion of seismic data; Geophysics, 50, 2458-2472

[14] Stolt, R., H., and Weglein, A. B., 2012, Seismic Imaging and Inversion; Cambridge

[15] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation; Geophysics, 49, 1259-1266

[16] Wapenaar, C.P.A., 1996, Inversion versus migration: A new perspective to an old discussion; Geophysics, 61, 804-814

[17] Weglein, A. B., Violette, P. B., and Keho, T. H., 1986, Using multiparameter Born theory to obtain certain exact multiparameter inversion goals; Geophysics, 51, 1069-1074

透射波与反射波层析、反射波最小二乘偏移、逆时偏移及全波形反演之间的关系与速度估计的合适方案选择

王华忠，冯波，胡江涛

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】速度估计是地震波成像的核心。速度估计的基本信息是数据体，数据体包含的基本信息是时距关系决定的波前面的到达时和波阻抗变化引起的波前面上的子波振幅变化。到达时信息是波传播的累积效应。波前面上的子波振幅变化是承载在到达时上的速度突变的反映。到达时对应速度场的背景变化，子波振幅变化对应速度场的跃变。透射波层析和反射波层析利用到达时来估计背景速度、反射波最小二乘偏移在已知背景速度的基础上估计速度的跃变。反射波层析的应用需要已知反射界面上的速度跃变。反射波层析和反射波最小二乘偏移是相互依赖的，需要迭代执行，这也是目前速度估计的经典思想。在数据缺乏低频成分时，FWI 根本不能用来估计速度场。如果提供了较好的初始速度场，FWI 的作用与 RTM 基本一样。因此，我们认为，在了解各种速度估计方法理论背景的基础上，合适的速度估计方案的选择取决于数据体的质量情况。

【关键词】透射波与反射波层析，最小二乘偏移，逆时偏移，全波形反演，速度估计

1．引言

油气地震勘探的目标是储层。储层描述的重要方式有两种，一是偏移速度估计、保真角度道集、AVA 反演、储层描述，这是目前常用的方法；另一种是背景速度估计、高波数速度场的估计、储层描述，这是目前 FWI+RTM 试图要走的储层刻画的路线。上述两条路线的核心都是地震波反演成像、目标是建立全频带（或全波数带）的速度场，通过精确的、精细的速度场的建立达到精确的、精细的储层描述的目的。

全频带或全波数带的速度场估计一直是地震学家、也是石油地震勘探学家所追求的目标。这也存在两条路线，一是石油地震勘探学家（Sheriff and Geldart, 1995）发展的一条实用化的路线；二是理论地震学家（Tarentola, 1984）发展的一条基于数学理论框架的路线。前者以 NMO 叠加速度分析、PSTM 偏移速度分析、PSDM 偏移速度分析和透射层析及反射波层析为估计背景速度的方法路线，结合 Migration 来估计速度场的高波数变化成分（很多时候是定位高波数变化成分出现的空间位置）。目前这依然是石油工业界的核心方法技术路线。后者是以 FWI 和 RTM 相结合，FWI 估计背景速度，RTM 估计速度场的高波数变化成分。但是，当实际地震数据中缺乏低频（3Hz 左右）成分时，FWI 的应用基本不成功，比常规路线改进不了什么（也可能没有任何改进！）。

从理论意义上讲，速度估计方法是基于一个完整的数学框架的，可以认为是贝叶斯估计理论。但实际上，速度估计因为数据质量的千差万别变得异常复杂。

我们认为，撇开贝叶斯估计理论不谈，速度估计的核心在于对实测的地震波中波现象的认识、在于对这些波现象的描述方法的理解、在于对利用这些传播算子进行的残差反投影的把握。

因此，本文从地震波传播过程的描述方法开始、讨论了反问题的建立方法、线性化的透射波层析、反射波层析和反射波最小二乘偏移的数学本质及对应的波现象；然后重点讨论了非线性的 FWI 和 RTM 反演成像，指出它们利用了多次散射/反射波，逆时波场外推算子是必须采用的波场传播算子。同时分析了 FWI 为什么一定要低频数据。最后，根据对这些不同的速度估计方法的特点分析，指出了速度估计与模型建立的合理的方法技术组合，特别指出合理的方法技术组合取决于实际数据体的质量情况，而不是方法技术的理论优势。

2．标量波方程及正问题表达

2.1 标量波方程、标量波散射势

勘探地震学中，很多情况下关注标量波方程（或密度不变情况下的声波方程）。因此，类似地，写出如下的标量波散射势。

$$\mathcal{L}_s\left(\mathbf{x},\omega\right)u\left(\mathbf{x},\omega\right)=-\delta\left(\mathbf{x}-\mathbf{x}_S\right)S\left(\omega\right) \tag{2.1a}$$

其中,

$$\mathcal{L}_s\left(x,\omega\right)=\nabla^2+\frac{\omega^2}{v^2\left(\mathbf{x}\right)} \tag{2.1b}$$

为标量波传播算子。$\mathcal{L}_s$ 代表 $\mathcal{L}_{\text{scalar}}$.

定义背景介质中的声波传播满足如下方程,

$$\mathcal{L}_{s0}\left(\mathbf{x},\omega\right)u_0\left(\mathbf{x},\omega\right)=-\delta\left(\mathbf{x}-\mathbf{x}_S\right)S\left(\omega\right) \tag{2.2a}$$

其中,

$$\mathcal{L}_{s0}\left(\mathbf{x},\omega\right)=\nabla^2+\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)} \tag{2.2b}$$

为背景介质中的标量波算子。

根据(2.1)和(2.2)式,定义标量波散射势为,

$$\begin{aligned}\mathcal{V}_s\left(\mathbf{x},\omega\right)&\equiv\mathcal{L}_s\left(\mathbf{x},\omega\right)-\mathcal{L}_{s0}\left(\mathbf{x},\omega\right)\\&=\left(\nabla^2+\frac{\omega^2}{v^2\left(\mathbf{x}\right)}\right)-\left(\nabla^2+\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}\right)\\&=a_s\left(\mathbf{x}\right)\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}\end{aligned} \tag{2.3a}$$

其中,

$$a_s\left(\mathbf{x}\right)=1-\frac{v_0^2\left(\mathbf{x}\right)}{v^2\left(\mathbf{x}\right)}。 \tag{2.3b}$$

上式定义的散射势还可以写为如下的形式,

$$\begin{aligned}\mathcal{V}_s\left(\mathbf{x},\omega\right)&=\omega^2\frac{1}{v_0^2\left(\mathbf{x}\right)}\left(1-\frac{v_0^2\left(\mathbf{x}\right)}{v^2\left(\mathbf{x}\right)}\right)\\&\approx\omega^2\frac{1}{v_0^2\left(\mathbf{x}\right)}\frac{2\delta v\cdot v_0}{v_0^2\left(\mathbf{x}\right)}=\omega^2\frac{2\delta v\left(\mathbf{x}\right)}{v_0^3\left(\mathbf{x}\right)}\end{aligned} \tag{2.4}$$

上述形式引入 $\delta v\left(\mathbf{x}\right)=v\left(\mathbf{x}\right)-v_0\left(\mathbf{x}\right)$ 更清楚地反映出速度的局部变化是散射势的本质。从散射势的定义也可以看出,估计散射势的绕射层析成像与估计或定位反射系数的叠前偏移成像是有严格区别的。

2.2 标量波方程的 Born 线性化

假如认为地表观测的地震波场是由于地下介质与背景介质存在差异,即有散射势的存在,而导致的散射波加入射波形成,则把总场记为

$$u\left(\xi,\eta,\omega\right)=u_0\left(\xi,\eta,\omega\right)+u_S\left(\xi,\eta,\omega\right) \tag{2.5}$$

依据(2.3)式定义的散射势得到散射场满足的方程为,

$$\mathcal{L}_{s_0}\left(\mathbf{x},\omega\right)u_S\left(\mathbf{x},\omega\right)=-\mathcal{V}_s\left(\mathbf{x},\omega\right)u\left(\mathbf{x},\eta,\omega\right) \tag{2.6}$$

Born 近似后的散射波场记为

$$u_s(\xi,\eta,\omega)= -\omega^2\int d\mathbf{x}G_0(\xi,\mathbf{x},\omega)\frac{a(\mathbf{x})}{v_0^2(\mathbf{x})}u_0(\mathbf{x},\eta,\omega) \tag{2.7}$$

它代表从震源 η 到散射点 $\mathbf{x}$，再到接收点 ξ 的散射波场。

Born 近似后，线性化的总波场可以写成

$$u(\xi,\eta,\omega)=u_0(\xi,\eta,\omega)+\omega^2\int d\mathbf{x}G_0(\xi,\mathbf{x},\omega)\frac{a_s(\mathbf{x})}{v_0^2(\mathbf{x})}u_0(\mathbf{x},\eta,\omega) \tag{2.8}$$

Born 近似的物理实质是，在背景场 $u_0(\xi,\eta,\omega)$ 中没有散射场，在扰动场中，只存在一次散射场。尽管用这种方式描述波传播与实际的波传播不同，但当二阶及二阶以上的散射场能量比较弱时，用 Born 近似进行 PSDM 和 Tomography 还是比较有效的。

关于子波问题，可以认为已经携带在背景场 $u_0(\mathbf{x},\eta,\omega)$ 中。也可以用 $G_0(\mathbf{x},\eta,\omega)S(\omega)$ 来表示，其中 $S(\omega)$ 代表子波的谱。

2.3 地震波反演中正问题的简洁表达

非线性正问题的数学描述可以抽象为

$$A(\mathbf{m})\mathbf{u}=\text{-}\delta(\mathbf{x}-\mathbf{x_s})\mathbf{f}=-\mathbf{f}(\mathbf{x_s},t) \tag{2.9}$$

其中，A 为正演算子。在地震波反演中，A 的形式为

$$A=\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)=\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\Delta \tag{2.10}$$

A 的形式受所选择的波动方程的形式的影响。可以选常密度声波，变密度声波方程，各种拟声波方程，弹性波方程等。$\mathbf{u}$ 为方程(2.9)要求解的未知量，表示正演合成的波场，$\mathbf{f}$ 为震源函数。

野外观测到的地震数据可以表述为

$$\mathbf{d}=C\mathbf{u}+\eta \tag{2.11}$$

此处，C 表示地震数据的采样算子，由观测系统决定。它使得连续的波场 $u(x,y,z,t)$ 变得规则离散化（或非规则离散化）及孔径有限化。采样算子的状态受实际条件的制约。η 表示噪音函数，在大部分情况下假设它是高斯白噪的。

连续波场受采样算子制约后，采样后的数据可以记为

$$[C\mathbf{u}]\approx\mathbf{u}(\mathbf{x}_i),\quad i=1,2,\cdots,n \tag{2.12}$$

其中 $\mathbf{x}_i$ 表示第 i 个采样点的坐标。根据方程（2.9），可以将连续波场的正演计算形式地表达为

$$\mathbf{u}=\left(A(m)\right)^{-1}\mathbf{f} \tag{2.13}$$

采样后的离散地震数据表示为，

$$F(\mathbf{m}) = C\left(A(\mathbf{m})\right)^{-1}\mathbf{f} \tag{2.14}$$

3．透射波层析、反射波层析和反射波最小二乘偏移

3.1 透射波层析

透射波层析利用的是从震源到接收点的波的旅行时信息，它是波传播的累积效应，与背景速度密切关联。

基本的表述如下。假设（2.7）式定义的积分方程写出算子表达形式，

$$K\mathbf{m} = \mathbf{d} \tag{3.1}$$

其中，K 是 Fredholm 第一类积分方程的积分核，由（2.7）式知，积分核的定义为

$$K\left(v_0(\mathbf{x}),\mathbf{x},\omega\right) = \omega^2 G_0\left(\xi,\mathbf{x},\omega\right)u_0\left(\mathbf{x},\eta,\omega\right) \tag{3.2}$$

$\mathbf{m}$ 代表由 $a(\mathbf{x})\Big/v_0^2(\mathbf{x})$ 作为元素形成的矢量；$\mathbf{d}$ 代表由每个炮检点对应的地震道的频谱值。值得说明的是，（3.2）式定义的积分核在透射波情形下是不关心散射势出现的位置的。这与反射层析是根本区别，反射层析必须已知反射界面才可以计算该积分核。

认为 $K:\mathrm{H}_1 \to \mathrm{H}_2$ 是一个有界的线性算子，定义如下的 Tikhonov 泛函，

$$\mathfrak{I}(\mathbf{m}) = \frac{1}{2}\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2 + \frac{\alpha}{2}\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2 \tag{3.3}$$

对上式求解的首要问题是计算泛函关于参数变化的导数，即泛函的梯度。为此，利用如下的变分方法。

对任意的介质扰动 $\delta\mathbf{m}\in\mathrm{H}_1$ 和扰动步长 $\tau\in\mathbb{R}$，总存在如下的泛函变分

$$\begin{aligned}\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m}) &= \frac{1}{2}\left(\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2 + \alpha\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\\ &+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2} + \left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)\\ &+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1} + \left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\ &+\frac{\tau^2}{2}\left(\left\|K\delta\mathbf{m}\right\|_{\mathrm{H}_2}^2 + \alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\end{aligned} \tag{3.4}$$

对上式定义的变分求偏导数，

$$\frac{\partial}{\partial\tau}\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})\bigg|_{\tau=0} = \left\langle \mathrm{Grad}\mathfrak{I}(\mathbf{m}),\delta\mathbf{m}\right\rangle_{\mathrm{H}_1} \tag{3.5}$$

其中

$$\mathrm{Grad}\mathfrak{I}(\mathbf{m}) = \left(K^*K+\alpha\mathbf{I}\right)\mathbf{m} - K^*\mathbf{d} \tag{3.6}$$

Born 近似后的线性问题的解。迭代格式为，

$$\begin{aligned}\mathbf{m}^{k+1}&=\mathbf{m}^{k}-\mu\mathrm{Grad}\mathfrak{J}\left(\mathbf{m}\right)\\&=\mathbf{m}^{k}-\mu\left[K^{*}\left(K\mathbf{m}-\mathbf{d}^{obs}\right)\right]\\&=\mathbf{m}^{k}-\mu\left[K^{*}\left(K\mathbf{m}^{k}-\mathbf{d}^{obs}\right)\right]\\&=\mathbf{m}^{k}-\mu\left[K^{*}\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)\right]\end{aligned}\tag{3.7}$$

当梯度等于零时，可以得到线性系统的法方程，

$$\left(K^{*}K+\alpha\mathbf{I}\right)\mathbf{m}=K^{*}\mathbf{d}\tag{3.8}$$

再次强调，透射层析和反射层析的根本差异在于计算积分核函数时是否需要反射界面处速度跃变的信息。

3.2 反射波层析

为实现反射波层析，要求初始速度模型可以产生反射波。在反射射线层析中，需要人为给出反射界面的位置以及切向导数，从而计算反射射线路径。在波动方程反射层析中，同样需要初始速度模型中包含反射界面的信息。假设背景速度 $v_{bg}(x)$ 可以用光滑的速度 $v_0(x)$ 以及反射系数 $R(x)$ 来表示：

$$v_{bg}(x)=v_0(x)\left(1+R(x)\right)\tag{3.9}$$

其中，反射系数 $R(x)$ 只是在反射界面处非零。将观测数据用光滑的速度 $v_0(x)$ 偏移，可以在成像剖面中提取反射系数 $R(x)$。若用模型用速度平方的倒数来表达，则 $\mathbf{m}_{bg}(x)=\left(v_{bg}^{2}(x)\right)^{-1}$。此时，模型扰动可以表示为：

$$\begin{aligned}\delta\mathbf{m}(x)&=\mathbf{m}_{bg}(x)-\mathbf{m}_0(x)\\&=\left(v_{bg}^{2}(x)\right)^{-1}-\left(v_0^{2}(x)\right)^{-1}\end{aligned}\tag{3.10}$$

若背景速度和光滑速度中的单频格林函数可以分别表示为 $G_{bg}\left(x;\omega\mid x_s\right)$ 和 $G_0\left(x;\omega\mid x_s\right)$，则模型扰动产生的扰动格林函数可以表示为：

$$\delta G_r\left(x;\omega\mid x_s\right)=G_{bg}\left(x;\omega\mid x_s\right)-G_0\left(x;\omega\mid x_s\right)\tag{3.11}$$

背景速度场中的 Born 波路径可以表示为：

$$\mathbf{L}_{bg}\left(x\mid x_r,x_s;\omega\right)=\omega^{2}G_{bg}\left(x\mid x_r;\omega\right)G_{bg}\left(x\mid x_s;\omega\right)\tag{3.12}$$

将式(3.11)代入(3.12)，有

$$\begin{aligned}\mathbf{L}_{bg}\left(x\mid x_r,x_s;\omega\right)=\omega^{2}\{&G_0\left(x;\omega\mid x_s\right)G_0\left(x;\omega\mid x_r\right)\\&+G_0\left(x;\omega\mid x_r\right)\delta G_r\left(x;\omega\mid x_s\right)\\&+\delta G_r\left(x;\omega\mid x_r\right)G_0\left(x;\omega\mid x_s\right)\\&+\delta G_r\left(x;\omega\mid x_s\right)\delta G_r\left(x;\omega\mid x_r\right)\}\end{aligned}\tag{3.13}$$

光滑速度场中的 Born 波路径可以表示为：

$$\mathbf{L}_0\left(x\mid x_r,x_s;\omega\right)=\omega^{2}G_0\left(x\mid x_r;\omega\right)G_0\left(x\mid x_s;\omega\right)\tag{3.14}$$

可以看出，波路径 $\mathbf{L}_{bg}$ 中的第一项代表光滑速度场中的波路径 $\mathbf{L}_0$，它是一个偏移算子，可以将观测数据投影到偏移响应上。$\mathbf{L}_{bg}$ 中的第二项与第三项则描述了反射波的波路径，分别代表了从炮点到反射点以及反

射点到检波点的波路径。$\mathbf{L}_{bg}$ 的最后一项为高阶项，可以忽略。为了描述反射波的波路径，我们希望将观测数据信息（走时残差，波形残差或者观测数据）投影到波路径上，从而更新宏观速度（即速度模型的低波数成分）。因此，反射波的波路径最终可以表示为

$$\begin{aligned}\mathbf{L}_r &= \mathbf{L}_{bg}\left(x \mid x_r, x_s;\omega\right) - \mathbf{L}_0\left(x \mid x_r, x_s;\omega\right)\\ &\approx \omega^2\{G_0\left(x;\omega \mid x_r\right)\delta G_r\left(x;\omega \mid x_s\right)\\ &\quad +\delta G_r\left(x;\omega \mid x_r\right)G_0\left(x;\omega \mid x_s\right)\}\end{aligned} \tag{3.15}$$

式(3.15)表达了背景速度场中的反射波波路径，它与地震反射全波形反演（Seismic reflection FWI=SRFWI，Xu 等，2012）构造的 FWI 目标泛函的 Frechet 导数的表达是一致的。事实上，若能够计算扰动格林函数 $\delta G_r\left(x;\omega \mid x_s\right)$ 以及 $\delta G_r\left(x;\omega \mid x_r\right)$，则反射波路径可以直接构造。式(3.15)的物理解释为：反射波波路径描述了从炮点到反射点以及反射点到检波点的波路径。因此，反射波波路径也可以也可以通过上下行波分解的方式构造。在频率域，格林函数的上下行波分解（Liu 等，2011）可以表示为

$$G\left(x,z;\omega\right) = \begin{cases} G^D\left(x,z;\omega\right), \omega k_z > 0 \\ G^U\left(x,z;\omega\right), \omega k_z < 0 \end{cases} \tag{3.16}$$

其中，上标 ^{U} 和 ^{D} 分别表示上行与下行。将上下行波分解后的格林函数表达式(3.16)代入背景速度场中的波路径 $\mathbf{L}_{bg}$ 中，有

$$\begin{aligned}&\mathbf{L}_{bg}\left(x \mid x_r, x_s;\omega\right)\\ &= \omega^2\{\left(G_{bg}^U\left(x \mid x_r\right) + G_{bg}^D\left(x \mid x_r\right)\right)\cdot\\ &\left(G_{bg}^U\left(x \mid x_s\right) + G_{bg}^D\left(x \mid x_s\right)\right)\}\\ &= \omega^2\{G_{bg}^U\left(x \mid x_r\right)G_{bg}^U\left(x \mid x_s\right)\\ &+G_{bg}^D\left(x \mid x_r\right)G_{bg}^D\left(x \mid x_s\right)\\ &+G_{bg}^U\left(x \mid x_r\right)G_{bg}^D\left(x \mid x_s\right)\\ &+G_{bg}^D\left(x \mid x_r\right)G_{bg}^U\left(x \mid x_s\right)\}\end{aligned} \tag{3.17}$$

为简化表达，上式中省略了格林函数中频率的显示表示。显然，背景介质中的单频格林函数经过上下行波分解后可以在光滑速度中传播的下行格林函数 G^D 以及在反射界面处产生的上行格林函数 G^U 。$\mathbf{L}_{bg}$ 中的前两项为炮检点同向传播的波路径，它与反射波路径的物理解释不符。而 $\mathbf{L}_{bg}$ 中的最后两项刚好描述了反射波路径。因此，经过上下行波分解之后，反射波路径可以表示为

$$\begin{aligned}&\mathbf{L}_r\left(x \mid x_r, x_s;\omega\right) = \omega^2\{G_{bg}^U\left(x \mid x_r\right)G_{bg}^D\left(x \mid x_s\right)\\ &+G_{bg}^D\left(x \mid x_r\right)G_{bg}^U\left(x \mid x_s\right)\}\end{aligned} \tag{3.18}$$

以上两种反射波路径虽然在形式上表达不同，但物理意义是相同的。它能够将观测数据信息投影到波路径上，从而更新宏观速度场。

可以看出，反射波层析实现的关键在于构造反射波对应的波路径，并沿波路径进行残差反投影。但是，构造反射波波路径需要背景速度场中有高波数反射界面的存在及高波数速度变化的信息，可见反射波层析和 PSDM/RTM 迭代执行才能很好地估计速度场，这既是 FWI 的理论预测，也是目前反射波层析和 PSDM 迭代进行来解决实际数据成像问题的实际证明。

3.3 反射波最小二乘偏移

利用透射层析和反射层析计算背景速度的重要目的是依据该背景速度估计速度场中的反射界面或散射点处的速度跃变，给出构造成像的结果，也包括保真成像的结果。

另一方面，反射层析计算背景速度需要速度的跃变来计算积分核。这说明速度的背景估计和速度的跃变估计二者是相关联的。

反射层析（包括 MVA）和 PSDM（及最小二乘偏移）迭代执行完成对速度场全波数成分的估计是速度估计的经典思想。

为了用最小二乘偏移估计散射势，定义如下误差泛函，

$$\begin{aligned}\mathcal{J} &= \frac{1}{2}\left\| u_s^{cal}\left(\xi,\eta,\omega\right) - u_s^{obs}\left(\xi,\eta,\omega\right)\right\|_2^2 \\ &= \frac{1}{2}\sum_{\omega}\sum_{\eta}\sum_{\xi}\left(u_s^{cal}\left(\xi,\eta,\omega\right) - u_s^{obs}\left(\xi,\eta,\omega\right)\right)^2\end{aligned} \tag{3.19}$$

误差泛函能量最小化时，认为估计的散射势是所求的参数估计结果。其中 $u_s^{cal}\left(\xi,\eta,\omega\right)$ 有（2.7）式定义。

目标泛函的梯度是构造最速下降法的核心。该梯度定义为，

$$\begin{aligned}\frac{\partial\mathcal{J}}{\partial\mathbf{f}\left(\mathbf{x}\right)} &= \sum_{\omega}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\frac{\partial u_s^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)}{\partial\mathbf{f}}\cdot \\ &\qquad \left[u_s^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right) - u_s^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right]^H \\ &= \sum_{\omega}\omega^2\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}S\left(\omega\right)G_S\left(\mathbf{x},\boldsymbol{\eta},\omega\right)G_R\left(\boldsymbol{\xi},\mathbf{x},\omega\right) \\ &\qquad \cdot\left[u_s^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right) - u_s^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right]^H\end{aligned} \tag{3.20}$$

其中 $S(\omega)G(\mathbf{x},\eta,\omega)$ 代表震源点 η 传到地下任一点 $\mathbf{x}$ 的场，$G_R\left(\xi,\mathbf{x},\omega\right)\left[u_s^{cal}\left(\xi,\eta,\omega\right) - u_s^{obs}\left(\xi,\eta,\omega\right)\right]^H$ 代表误残差波场的反传波。

利用（3.20）定义的梯度可以进行最速下降法的最小二乘偏移成像的具体计算，

$$\begin{aligned}\mathbf{f}^{(k+1)}\left(\mathbf{x}\right) &= \mathbf{f}^{(k)}\left(\mathbf{x}\right) + \mu^{(k)}\delta\mathbf{f}^{(k)}\left(\mathbf{x}\right) \\ &= \mathbf{f}^{(k)}\left(\mathbf{x}\right) + \mu^{(k)}\frac{\partial\mathcal{J}}{\partial\mathbf{f}\left(\mathbf{x}\right)}\end{aligned} \tag{3.21}$$

为了加速收敛，可以用 Hessian 矩阵的逆对梯度方向进行预条件处理，Hessian 矩阵的定义为，

$$\begin{aligned}\mathbf{H}(\mathbf{x},\mathbf{y}) &= \frac{\partial^2\mathcal{J}}{\partial\mathbf{f}\left(\mathbf{x}\right)\partial\mathbf{f}\left(\mathbf{y}\right)} \\ &\qquad S\left(\omega\right)G_S\left(\mathbf{x},\boldsymbol{\eta},\omega\right)G_R\left(\boldsymbol{\xi},\mathbf{x},\omega\right) \\ &= \sum_{\omega}\omega^4\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\cdot\left[\frac{\partial u_S^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)}{\partial\mathbf{f}\left(\mathbf{y}\right)}\right]^H \\ &= \sum_{\omega}\omega^4\sum_{\boldsymbol{\eta}}S^2\left(\omega\right)G_S\left(\mathbf{x},\boldsymbol{\eta},\omega\right)G_S^H\left(\mathbf{y},\boldsymbol{\eta},\omega\right) \\ &\cdot\sum_{\boldsymbol{\xi}}G_R\left(\boldsymbol{\xi},\mathbf{x},\omega\right)G_R^H\left(\boldsymbol{\xi},\mathbf{y},\omega\right)\end{aligned} \tag{3.23}$$

该 Hessian 矩阵是由背景介质中的 Green 函数规定的，其中并不包括高阶反射/绕射效应，即不包括非线性部分。双向波计算 Green 函数可以包含这部分的作用。

由 Hessian 矩阵的逆对梯度方向进行预条件处理后，最速下降法的迭代格式变为，

$$\begin{aligned}\mathbf{f}^{(k+1)}\left(\mathbf{x}\right) &= \mathbf{f}^{(k)}\left(\mathbf{x}\right) + \mu^{(k)}\delta\mathbf{f}^{(k)}\left(\mathbf{x}\right) \\ &= \mathbf{f}^{(k)}\left(\mathbf{x}\right) + \mu^{(k)}\mathbf{H}^{-1}\frac{\partial\mathcal{J}}{\partial\mathbf{f}\left(\mathbf{x}\right)}\end{aligned} \tag{3.24}$$

迭代步长 $\mu^{(k)}$ 一般可以采用试探的方式进行确定，一般采用试探加抛物拟合来选取最优下降步长来保证误差的下降。

迭代的每一步需要在更新后对模型进行一次反偏移，然后将反偏移得到的正演数据从原始数据中减去，这样来判断误差是否达到要求，然后判断迭代是否终止。

4．RTM 与 FWI 问题与特点分析

4.1 非线性反演问题概述

基于前面对正问题的定义，根据数据残差最小的原则可以建立如下的优化问题：

$$\begin{aligned} \min: &\quad J(\mathbf{m}) \\ S.T.: &\quad C\mathbf{u}=\mathbf{d} \end{aligned} \tag{4.1}$$

可以用如下的方式定义上面优化问题所对应的误差泛函，

$$\min_{\mathbf{u}\in\mathcal{H}_2,\mathbf{m}\in\mathcal{H}_1} \frac{1}{2}\left\|C\mathbf{u}-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{4.2}$$

上式中的第一项代表正演波场与观测波场的波场残差，后一项代表对模型的先验约束，α 是正则化系数。一般地，我们可以通过假设模型中的元素满足某种概率分布条件来施加对模型的先验限制。

将（2.13）代入上式，得到泛函定义式，

$$\begin{aligned} \min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2} \mathfrak{I}(\mathbf{m}) = &\frac{1}{2}\left\|C\left(A(\mathbf{m})\right)^{-1}\mathbf{f}-\mathbf{d}\right\|_Y^2 \\ &+\alpha J(\mathbf{m}) \end{aligned} \tag{4.3}$$

通过此式，可以用各种优化方法求解参数估计问题。在无约束非线性优化理论中，（4.3）式的求解有很多不同的方法，譬如最速下降方法、全 Newton 方法、Gauss-Newton 方法、拟 Newton 方法等等。一般地，（4.3）式对应的的迭代求解格式可以写为

$$\mathbf{m}^{k+1}=\mathbf{m}^k-\mu_k\mathbf{g}_k \tag{4.4}$$

其中，μ_k 为第 k 个迭代步的步长，$\mathbf{g}_k$ 为第 k 个迭代步的梯度。在不同的迭代方法中，$\mathbf{g}_k$ 也可以认为是预条件修正后的梯度方向。步长的选择对迭代法的收敛性和收敛效率都影响甚大，具体地参考 Dennis 和 Schnabel（1996）以及 Nocedal 和 Wright（1999）。（4.4）式中最核心的是梯度向量的计算。

4.2 共轭状态法梯度计算

非线性问题的梯度计算问题

假设泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 处是 Fréchet 可微的，由 Reisz 表示定理，存在梯度 $\mathrm{Grad}\mathfrak{I}(\mathbf{m})\in\mathcal{H}$，称为泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 点的梯度，满足

$$\mathfrak{I}'(\mathbf{m})\delta\mathbf{m}=\left\langle \mathrm{Grad}\mathfrak{I}(\mathbf{m}),\delta\mathbf{m}\right\rangle_{\mathcal{H}} \tag{4.5}$$

若泛函 $\mathfrak{I}:\mathcal{H}\to\mathbb{R}$ 在 $\mathbf{m}$ 处是 Fréchet 可微的，则对任意的 $\delta\mathbf{m}\in\mathcal{H}$，则关于变量 τ 的映射 $\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m}):\mathbb{R}\mapsto\mathbb{R}$ 在 $\tau=0$ 处也是可微的，

$$\left.\frac{d}{d\tau}\mathfrak{I}(\mathbf{m}+\tau\delta\mathbf{m})\right|_{\tau=0}=\left\langle \mathrm{Grad}\mathfrak{I}(\mathbf{m}),\delta\mathbf{m}\right\rangle_{\mathcal{H}}$$

因此，我们可以从上述 Gateaux 导数的定义计算泛函的微分。

4.2.1 Jacobi 方法

计算非线性问题梯度的最简单方法是利用范函的差分（差商）值来代替微分值。该方法通过差分的方式计算范函在任意点 $\mathbf{m}$ 处的梯度值，其精度取决于差分的步长，表达式如下，

$$\left[\mathrm{Grad}\mathfrak{I}_n\left(\mathbf{m}\right)\right]_i = \frac{\mathfrak{I}_n\left(\mathbf{m}+\tau_i\mathbf{e}_i\right)-\mathfrak{I}_n\left(\mathbf{m}\right)}{\tau_i} + O\left(\tau_i^2\right) \tag{4.6}$$

上面的式子表示范函沿任意一个模型空间方向 $\mathbf{e}_i$ 的导数值可以表示为一个差分的值加上一个二阶误差项，其中 $\mathbf{e}_i$ 表示模型空间中的第 i 个单位向量，n 表示模型空间的维数。所有导数值形成一个向量，构成了范函在点 $\mathbf{m}$ 处的梯度值。

4.2.2. Adjoint-State 方法（伴随状态方法）

已知算子 $A\left(\mathbf{m}\right)=\dfrac{1}{v^2\left(\mathbf{x}\right)}\dfrac{\partial^2}{\partial t^2}-\Delta$，它可以理解为一个投影算子，建立了模型空间和数据空间之间的映射关系。$\partial A/\partial\mathbf{m}$ 表示算子的微分，它定义了一个 Jacobi 矩阵。在一定的定义下，它是一个 Fréchet 微商。假设算子 $A\left(\mathbf{m}\right)$ 存在逆 $A\left(\mathbf{m}\right)^{-1}$，则有，

$$A\left(\mathbf{m}\right)A\left(\mathbf{m}\right)^{-1}=\mathbf{I} \tag{4.7}$$

上式两边对 $\mathbf{m}$ 求导可得，

$$\frac{\partial A\left(\mathbf{m}\right)}{\partial\mathbf{m}}A\left(\mathbf{m}\right)^{-1}+A\left(\mathbf{m}\right)\frac{\partial A\left(\mathbf{m}\right)^{-1}}{\partial\mathbf{m}}=0 \tag{4.8}$$

上式可改写为，

$$\frac{\partial A\left(\mathbf{m}\right)^{-1}}{\partial\mathbf{m}}=-A\left(\mathbf{m}\right)^{-1}\frac{\partial A\left(\mathbf{m}\right)}{\partial\mathbf{m}}A\left(\mathbf{m}\right)^{-1} \tag{4.9}$$

定义数据拟合的残差矢量为

$$\begin{aligned}\mathbf{r}\left(\mathbf{m}\right)&=\mathbf{d}_{cal}-\mathbf{d}_{obs}=F\left(\mathbf{m}\right)-\mathbf{d}_{obs}\\&=C\left(A\left(\mathbf{m}\right)\right)^{-1}\mathbf{f}-\mathbf{d}_{obs}\end{aligned} \tag{4.10}$$

其中 $\mathbf{r}\left(\mathbf{m}\right)$ 表示波场残差，$F\left(\mathbf{m}\right)$ 表示正算子根据模型 $\mathbf{m}$ 得到的数据，$\mathbf{d}_{obs}$ 表示观测到的数据。

重新数据逼近误差泛函为

$$\begin{aligned}\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\mathfrak{I}_{LS}\left(\mathbf{m}\right)&=\frac{1}{2}\left\|C\left(A\left(\mathbf{m}\right)\right)^{-1}\mathbf{f}-\mathbf{d}\right\|_Y^2\\&=\frac{1}{2}\sum_t\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\mathbf{r}^2\left(\mathbf{m}\right)\end{aligned} \tag{4.11}$$

首先，投影算子 $A\left(\mathbf{m}\right)$ 的 Gateaux 导数可以用如下的方式进行定义，

$$\left.\frac{\partial A(\mathbf{m}+\tau\mathbf{e}_i)}{\partial\tau}\right|_{\tau=0}=\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\cdot\mathbf{e}_i \tag{4.12}$$

其中，$\frac{\partial A(\mathbf{m})}{\partial\mathbf{m}}$ 代表投影算子 $A(\mathbf{m})$ 的Fréchet 微商。则误差泛函的梯度可以按如下形式导出

$$\begin{aligned}\left[G_{LS}\right]_i&=\frac{\partial}{\partial\tau}\Im_{LS}\left(\mathbf{m}+\tau\mathbf{e}_i\right)\big|_{\tau=0}\\&=\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}\cdot\mathbf{r}(\mathbf{m})\\&=\left\langle\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0},\mathbf{r}(\mathbf{m})\right\rangle\end{aligned} \tag{4.13}$$

其中，$\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}$ 定义了正演波场随模型参数变化的方向导数---Gateaux 导数。定义误差泛函的梯度首先要给该导数。

由（2.14）式知，该导数可以写为，

$$\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)=C\frac{\partial}{\partial\mathbf{m}}\left(A(\mathbf{m})^{-1}\mathbf{e}_i\right)f \tag{4.14}$$

利用（4.9）式，上式重写为

$$\begin{aligned}\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)&=-CA(\mathbf{m})^{-1}\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)A(\mathbf{m})^{-1}f\\&=-CA(\mathbf{m})^{-1}\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u}\end{aligned} \tag{4.15}$$

把（4.15）定义的变分代入（4.13）式，

$$\begin{aligned}\left[G_{LS}\right]_i&=\left\langle\left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0},\mathbf{r}(\mathbf{m})\right\rangle_{\mathcal{H}_2}\\&=-\left\langle CA(\mathbf{m})^{-1}\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u},\mathbf{r}(\mathbf{m})\right\rangle_{\mathcal{H}_2}\\&=-\left\langle\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u},\left(A(\mathbf{m})^{-1}\right)^*C^*\mathbf{r}(\mathbf{m})\right\rangle_{\mathcal{H}_1}\\&=\left\langle\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u},\mathbf{z}\right\rangle_{\mathcal{H}_1}\end{aligned} \tag{4.16}$$

（4.16）中第二行利用了数据空间和模型空间存在的共轭关系（这种关系是假设的，理想弹性介质情况下应该是可以接受的!），用共轭算子 $\left(A(\mathbf{m})^{-1}\right)^*$ 把数据残差投影到模型空间。定义 $\mathbf{u}=A(\mathbf{m})^{-1}f$ 为正演波场；$\mathbf{z}=-\left(A(\mathbf{m})^{-1}\right)^*C^*\mathbf{r}(\mathbf{m})$ 为残差波场的反传播。

可以看出，波场 $\mathbf{z}$ 的求解需要解如下的微分方程

$$\left(A(\mathbf{m})\right)^*\mathbf{z}=-C^*\mathbf{r}(\mathbf{m}) \tag{4.17}$$

前已述及，$A(\mathbf{m})$是一个共轭算子，满足条件：$\left(A(m)^{-1}\right)^{*}=\left(A(m)^{*}\right)^{-1}$。

为了清楚起见，把正向波场外推方程和反向波场外推方程罗列如下：

正向波场外推方程：

$$\begin{cases}\left(\dfrac{1}{v^2(x)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)u(\mathbf{x},t)=-\delta(\mathbf{x}-\mathbf{x}_S)S(t)\\ u(\mathbf{x},t)\big|_{\mathbf{x}=\Omega}=f_b(\Omega,t)\end{cases}\tag{4.17a}$$

其中，$f_b(\Omega,t)$代表边界条件。一般用 PML 边界条件。Ω代表计算波场在空间上的边界。

反向波场外推方程：

$$\begin{cases}\left(\dfrac{1}{v^2(x)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)z(\mathbf{x},t)=0\\ z(\mathbf{x},t)\big|_{\mathbf{x}=\xi}=u(\boldsymbol{\xi},t)\end{cases}\tag{4.17b}$$

其中，$\boldsymbol{\xi}$代表观测点坐标值。$u(\boldsymbol{\xi},t)$代表观测点处的波场。FWI 进行波场残差外推时，仅有观测点处有波场残差，因此，才有（4.17b）这样的边界条件的定义。实际上，波场逆时外推并不限于这样的边界条件。

可见，（4.17a）是一个初值问题，沿时间正向外推计算；（4.17b）是一个边值问题，逆时间反向外推计算。

（4.16）中计算梯度时，正演波场$\mathbf{u}$前还要作用一个算子$\dfrac{\partial A(\mathbf{m})}{\partial \mathbf{m}}\cdot\mathbf{e}_i$。它代表算子$A(\mathbf{m})$的微扰。

定义算子$A(\mathbf{m})$的微扰为，

$$\begin{aligned}\delta A(\mathbf{m})&=\left(\frac{1}{(v+\delta v)^2}\frac{\partial^2}{\partial t^2}-\Delta\right)-\left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\Delta\right)\\&=\frac{\delta v}{v^2}\frac{\partial^2}{\partial t^2}\end{aligned}\tag{4.18a}$$

因此有，

$$\frac{\delta A(\mathbf{m})}{\delta \mathbf{m}}=\frac{1}{v^2}\frac{\partial^2}{\partial t^2}\tag{4.18b}$$

由此，最后的梯度公式为，

$$\begin{aligned}\mathbf{g}(\mathbf{m}(\mathbf{x}))=\sum_t\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}&\frac{1}{v^2(\mathbf{x})}\\&\frac{\partial^2 u(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta})}{\partial t^2}z(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta})\end{aligned}\tag{4.19}$$

利用 Adjoint-State 方法求解梯度的地球物理意义可以解释为：误差函数对模型扰动的一阶导数值为震源函数激发的波场与残差波场的反传播得到的场的内积结果，其中震源函数激发的波场受（4.17）式定义的算子的作用。这种计算梯度的方法需要求解方程$\mathbf{u}=A(\mathbf{m})^{-1}f$和$A(\mathbf{m})^{*}\mathbf{z}=-C^{*}\mathbf{r}(\mathbf{m})$，这是两次波场正演计算。可以看出，反问题的计算量主要集中在梯度的计算上，每次迭代原则上至少需要两次正演计算。

相对于 Jacobi 方法而言，Adjoint-State 方法可以方便、更高效地计算出泛函的梯度值，用于非线性迭代公式的计算。这是目前全波形反演中最常用的梯度计算方法。

4.3 FWI 具体实现方案

FWI 的核心工作是梯度的计算，用伴随状态法计算梯度，核心的计算量是（4.17）定义的正向和反向波场外推工作。首先定义单炮的梯度，然后累计各炮的梯度得到全模型空间的梯度。估计步长后，利用（4.4）式计算本次迭代修正量。基本计算步骤可罗列如下：

（1）单炮梯度的计算

①解方程（4.17a），计算并记录每一时刻的正传波场 $\mathbf{u}\left(\xi,\eta,t\right)$，求其二阶导数，并用（4.18b）定义的算子作用。

②计算检波点位置的波场残差 $C^{*}\mathbf{r}\left(\mathbf{m}\right)$。

③解方程（4.17b），将波场残差反传播到模型空间，得到反传播数据 $\mathbf{z}\left(\xi,\eta,t\right)$。

④求正、反传波场的零时移互相关，得到单炮的梯度场。

（2）叠加所有炮求取的梯度场，得到整个模型空间的梯度场。

（3）利用（4.4）式得到本次迭代的修正结果。

步长的估计参考 Dennis 和 Schnabel（1996）以及 Nocedal 和 Wright（1999）。

4.4 RTM 与 FWI 问题与特点分析

RTM 和 FWI 是目前地震精确成像的理论上最佳的组合。实际应用上，它们二者都存在一定的问题。下面从地震波成像发展分析 RTM 与 FWI 的本质含义。

地震波叠前/叠后偏移成像的概念首先由 Hagedoorn（1954）提出。他直观地认为把地表观测的地震数据中蕴含的绕射波按一定的时距关系叠加到绕射点上可以形成地下绕射点/反射点的像。这种物理直观被 70 年代发展的 Kirchhoff 积分偏移成像方法所证实（Schneider,1978）。Claerbout(1971)提出激励时间成像条件及波场外推成像方法首先也是出自物理直觉。可以说，地震波偏移成像不是数学家推导出来的。而是地球物理学家首先提出概念性的设想和基本做法。Claerbout(1971)提出的入射波到达时等于反射波出发时的成像条件概念清楚地指出了地震波成像需要正向传播的入射波与逆向传播的外推后的反射波相关（偏移速度正确时在反射点处相关）是地震波偏移成像的本质。尽管他导出单向波（15o 方程）主要是处理波场沿深度外推，实质上已经预示了逆时偏移成像的可能性。

在 1982 年，SEG 年会上关于 Migration:Fundamental issues and future developments 的 Workshop 上，Whitemore 提出了逆时偏移的想法，1983 年，Baysal, et al.(1983) 和 Loewenthal and Mufti (1983)分别提出了时间空间与频率空间域的逆时偏移成像方法。Tarantola（1984）给出了 FWI 的详细的实现框架。这是从数学理论框架，不是从叠前地震成像的物理直觉出发进行地震波成像的尝试。但是，直到目前的研究表明基于地球物理直觉发展出的方法可以用数学理论框架完满地解释。

4.4.1 RTM 成像的特点及与其他 PSDM 的差异

前已述及，一般意义下的 PSDM 方法是从 Claerbout(1971)提出的地震波偏移成像的成像条件概念的基础上导出的。事实上，RTM 的初始提出（Whitemore，1982）也是基于该成像条件，即地震波激励时间成像条件：入射波到达时等于反射波出发时。该成像条件并没有指明入射波和反射波波场是如何计算的，它蕴含了成像条件可以用入射波和反射波在成像点处的相关来实现。这就是 RTM 偏移成像的逻辑。它与常规的 PSDM 的差异主要在于正向波场和逆时波场的全空间相关及逆时偏移所用的双向波方程正演计算可以模拟所有波现象，对标量波方程而言，包括直达波、折射波、Diving 波、一次反射波和多次反射波。正向波场和逆时波场的全空间相关意味着无论什么方向的波在成像点处都可以实现入射波和反射波的相关成像，因此

RTM 可以对大于 90o 的反射界面成像、可以对一次波和多次波成像。也可以对直达波、折射波、Diving 波这些透射波成像，尽管它们的成像结果是低频噪音（需要消除）。

正向波场和逆时波场的全空间相关意味着假如速度场中有反射界面的存在，从炮点到反射点、再到接收点的波路径上都会出现相关结果，产生低频噪音。没有反射界面的存在不能对多次波成像、有时也不能对大于 90o 的反射界面成像。

RTM 具有精确成像能力，但是需要精确的速度场，会产生低频噪音。如果知道了精确的速度场，偏移成像就没有意义了。因此，对 RTM 的本质有深入的认识、恰当地使用它是非常重要的。

另外，单炮道集的 RTM 的公式可以表示为，

$$R(\mathbf{x}) = \sum_t u(t,\mathbf{x}) z(t,\mathbf{x}) \tag{4.20}$$

其中正向传播的波场 $u(t,\mathbf{x})$ 和逆时间传播的波场 $z(t,\mathbf{x})$ 的计算，就是数值求解方程（4.17a）和（4.17b）式。（4.20）和（4.19）对比，可知 RTM 与 FWI 的重要差异。RTM 结果与 FWI 的第一次迭代结果也是不同的。

4.4.2 低频在 FWI 中不可或缺的原因

对前面讨论的 FWI 的理论及实现过程的分析可知，FWI 的本质操作是用正向外推波场和逆向外推波场的相关，即（4.19）式，计算速度场的更新量。正向外推波场和逆向外推波场的相关的基本要求是要相关的波场不能相差二分之一个波长。低频情况下，这个要求是很容易满足的，因为此时地震波波长很长。这是数据低频成分存在时，FWI 有效的根本原因。一般地，FWI 反演时有 10Hz 以下的低频数据就够了，足以提供满足偏移成像所需要的速度场。

但是，如果数据中缺乏低频成分，尤其是缺乏 3Hz 以下的低频成分，若没有正确的初始背景速度，FWI 迭代反演基本不会成功。

低频成分与速度场的背景变化部分存在密切的关系，这是波传播理论早已证实的。

4.4.3 FWI 到底处理什么波现象

FWI 可以处理正演方程中能模拟的波现象，就是说能用这样的波场进行速度场的反演。因此，FWI 用于实际地震数据反演时，地表附近的透射波现象的存在，譬如直达波、Diving 波，对于反演地表层的速度至关重要。这也是有些情况下 FWI 成功应用的关键。因为，地表速度的准确与否对于 FWI 能否发成功反演整个速度场起着核心作用，表层是跨不过去的。

利用反射波更新速度场，数据中必须有低频成分，此时 FWI 可以更新背景速度。反射界面处高波数速度场的更新，需要用（4.19）式定义的梯度来修正。因此，当数据中缺乏低频成分时，FWI 所起的作用与 RTM 似乎差不多，仅仅给出反射界面的存在。

因此，对实际数据而言，对于 FWI 的使用需要特别注意。盲目相信 FWI，可能得到事倍功半的效果。值得说明的是，对于透射波反演而言，其中是否缺乏低频成分，对 FWI 的成功应用并非关键的。

4.4.4 FWI 需要什么样的初始速度场

FWI 需要的初始速度场是使得正向外推波场和逆向外推波场中要相关的成分不能相差二分之一个波长。

高频情况下（譬如优势频率 30Hz）,达到此要求的初始场可以给出很好的 RTM 偏移结果。

这再一次说明，FWI 在数据缺低频时，不是速度反演的可行方案。

4.4.5 实际数据中 FWI 能起到什么作用

首先，实际数据中有 3Hz 及以上的有效信号成分时，FWI 可以给出能满足 RTM 成像要求的速度反演结果。

事实上，这存在两条面向储层解释的成像方案。一是常规的成像方案：用 FWI 的速度估计结果，进行 RTM 或其它 PSDM，给出角度成像道集，进行 AVA 叠前参数反演，进入储层解释；二是用 FWI 给出的速度成像结果，直接进行储层解释。第二条储层解释方案目前使用得不是很多，还需要很多的探讨。但是，不可轻易否认这样做的优点。

实际数据中缺乏低频时，需要用其它方法来估计低频成分。此时，如果还用第一条方案进入储层解释，使用 FWI 的意义就不是很大了。

5．速度估计的合适方案分析

5.1 透射波层析、反射波层析、反射波最小二乘偏移以及 RTM、FWI 之间的关系

透射波层析、反射波层析和反射波最小二乘偏移都是基于线性化的波传播关系，建立起炮点到接收点（透射波层析）或炮点到反射点再到接收点（反射波层析和反射波最小二乘偏移）之间的波路径。透射波层析反演低频背景速度，不需要反射界面的信息。反射波层析也是反演低频背景速度，但需要提供速度的高波数变换信息。两种层析方法利用的波线性不同，反演方法也很不相同，使用的环境也很不相同。透射波层析适用于地表层速度的反演，因为地表层存在直达波和 Diving 波。反射波层析反演适用于中深层速度反演，在这样的区域存在丰富的多角度的反射波。反射波最小二乘偏移适用于在已知背景速度的情况下，利用反射波波形信息估计中深层速度场的高波数变化。

波场线性化后，利用波场沿路径的 Green 函数，可以方便地实现透射波层析、反射波层析和反射波最小二乘偏移。

但是，对于非线性的波传播，由于存在丰富的多次反射/散射关系，很难利用沿波路径的 Green 函数来实现非线性的地震波成像。RTM 和 FWI 属于非线性地震波反演成像，它们都是通过波场的正时间外推和逆时间外推，然后对两个方向外推的波场进行相关实现的。沿时间进行双程波波场的计算，可以模拟各种波现象（指的是所用方程能模拟的所有波现象）。

线性反演方法与对应的波现象匹配；非线性反演方法处理所有波现象。线性反演类似一个灰匣子问题；非线性反演类似一个黑匣子问题。

由线性反演，通过考虑多次散射/反射，推进到非线性反演是目前地震反演的比较有代表性的思想方法（Berkhout, 2012）。

正问题的描述方法和所用的波现象是上述各种反演方法的核心差异。

5.2 速度估计的合适方案

根据前面我们对透射波层析，反射波层析和反射波最小二乘偏移，以及 RTM 和 FWI 的理论背景及方法特点分析，可以看出，速度估计方法主要是依据数据情况来选择的。

如果数据中包含 3Hz(更低更好)以上的有效波场成分，地震数据的规则性好、方位角/张角足够宽，地表速度能比较好地确定，FWI 是比较好的选择。

假如数据中缺乏有效的低频成分，利用透射波层析估计地表速度，反射波层析估计中深层背景速度，反射波最小二乘偏移确定中深层速度场的高波数变化量。这是比较好的方案。如果数据中缺乏有效的低频成分，信噪比很低，成像道集的聚焦速度分析是比较好的选择。这也是目前常用的方法。

总之，什么样的速度估计方案是合适的方案主要是取决于数据体的情况。

6．数值实验

6.1 透射波与反射波路径

Woodward（1992）给出了 Born 波路径的概念，指出反射能量可用波动方程反投影到 Born 波路径上。

图 6.1（a）展示了常速背景中的透射波路径，（b）为常梯度背景速度中的回转波波路径，（c）为双层速度模型中的 Born 波路径（背景速度为真实速度），（d）为双层速度模型中的 Born 波路径（背景速度为第一层的真实速度）。（e）为双层速度模型中的反射波波路径。可以看出，对于透射波或者回转波，观测数据信息（走时残差，波形残差或者观测数据）可以直接反投影到波路径上。而反射波的波路径的产生与背景速度中的高波数部分（反射界面）密切相关。若背景模型中有高波数成分，则 Born 波路径中包含反射波路径（图 6.1(c)所示）。若背景模型中无高波数成分，则 Born 波路径中无反射波路径（图 6.1(d)所示）。图 6.1(c)同样告诉我们，即使背景速度中有高波数成分，Born 波路径也不等价与反射波路径。Born 波路径包含反射波路径以及偏移相应算子！图 6.1(e)为从 Born 波路径中分离的反射波路径。此时，观测数据信息才可以被反投影到波路径上。对于反射地震数据，反射波路径的构造是至关重要的。

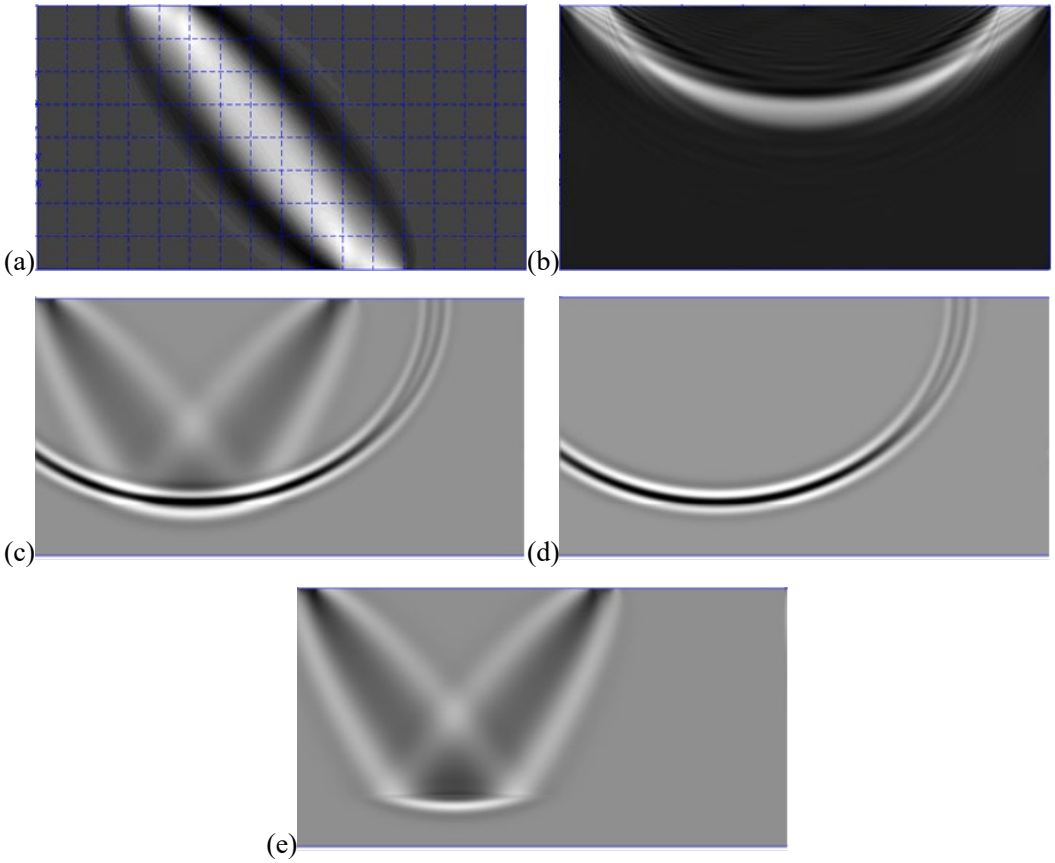
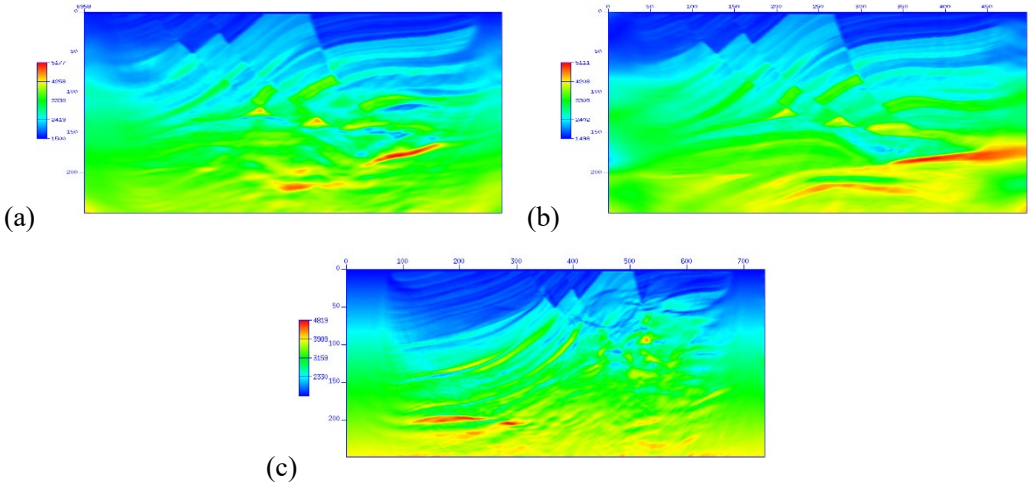


图 6.1 透射波（a）与回转波（b）波路径；（c）和（d）分别为真实模型（双层速度模型）与背景模型（速度为第一层的速度）中的 Born 波路径；（e）为双层速度模型中的反射波路径。

6.2 初始速度与低频数据对 FWI 的影响

为验证初始模型对 FWI 的影响，我们用两个初始模型测试 FWI，分别是一维线性递增的速度模型（从 1500m/s 到 4000m/s）和对真实速度模型平滑后的速度模型。图 6.2（a）和（b）分别是线性初始模型的反演结果以及真实速度模型光滑后作为初始模型的反演结果。可以看出，当初始模型较好时，FWI 反演的结果收敛性更好。图 6.2（c）为低频数据缺失（<5hz）情况下线性初始模型的反演结果。此时，FWI 陷入了局部极值。浅层的断层中的速度模型的构造出现较多假象，且深层的速度反演结果误差很大。



(a) (b)

(c)

图 6.2 线性初始模型的反演结果（a）；真实速度模型光滑后作为初始模型的反演结果（b）；低频数据缺失（<5Hz）情况下线性初始模型的反演结果(c)

7．结论与讨论

本文分析了透射波层析、反射波层析、反射波最小二乘偏移以及 RTM 和 FWI 的基本理论，从各自的理论可以知道它们的适用性。

速度反演的基础是多次覆盖的数据体。数据体包含的基本信息是时距关系决定的波前面的到达时和波阻抗变化引起的波前面上的子波振幅变化。到达时信息是波传播的累积效应。波前面上的子波振幅变化是承载在到达时上的速度突变的反映。到达时对应速度场的背景变化，子波振幅变化对应速度场的跃变。

透射波层析和反射波层析利用到达时来估计背景速度、反射波最小二乘偏移在已知背景速度的基础上估计速度的跃变。透射波层析的积分核不需要知道反射界面（透射波也不依赖于界面），这是它的主要特征，据此我们知道它适用于估计地表速度。

中深层的反射需要反射波来估计。反射波层析的应用需要已知反射界面上的速度跃变。反射波层析和反射波最小二乘偏移是相互依赖的，需要迭代执行，这也是目前速度估计的经典思想。

对于 FWI 的理论公式分析后知道，在数据缺乏低频成分时，FWI 根本不能用来估计速度场，此时需要提供给 FWI 一个较好的速度场使得正向外推波场和逆向外推波场中要相关的成分不能相差二分之一个波长。在数据缺乏低频成分的情况下，提供了较好的初始速度场时，由 FWI 的梯度公式知，FWI 的作用就与 RTM 基本一样。

我们认为，在了解各种速度估计方法理论背景的基础上，合适的速度估计方案的选择取决于数据体中的蕴含的波现象和数据质量情况。

参考文献

[1] Baysal, E., Kosloff, D.D., and Sherwood, J. W. C., 1983, Reverse time migration: Geophysics, 48, 1514-1524.

[2] Berkhout, A. J., 2012, Combining full wavefield migration and full waveform inversion, a glance into the future of seismic imaging: Geophysics, 77, S43–S50

[3] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, 26, 99-108

[4] Bleistein, N., Cohen, J. K., and Hagin, F., 1985, Two and one-half dimensional Born inversion with an arbitrary reference: Geophysics, 52, 26-36

[5] Claerbout, J. F., 1971, Towards a unified theory of reflector mapping: Geophysics, 36,467-481

[6] Clayton, R. and Stolt, R., 1981, A Born-WKBJ inversion method for acoustic reflection data: Geophysics, 46, 1559-1567

[7] Hagedoorn, J. G., 1954, A process of seismic reflection interpretation: Geophysical Prospection, 2, 85-127
[8] Loewenthal D. and Mufti I. R., 1983, Reverse time migration in spatial frequency domain, Geophysics, 48, 627-635.
[9] Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, 56, 645–653
[10] Plessix R. E. and Mulder W. A., 2004, Frequency-domain finite-difference amplitude-preserving Migration: Geophys. J. Int. 157, 975-987
[11] Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, 64, 888–901
[12] Schneider, W. A., 1978, Integral Formula for Migration in Two and Three Dimensions: Geophysics, 43, 49-76
[13] Sheriff and Geldart, 1995, Exploration Seismology, Second Edition, Cambridge
[14] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, 49, 1259–1266
[15] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation; Geophysics, 49, 1259-1266
[16] Weglein, A. B., Violette, P. B., and Keho, T. H., 1986, Using multiparameter Born theory to obtain certain exact multiparameter inversion goals: Geophysics, 51, 1069-1074
[17] Whitemore, N. D., 1982, Migration:Fundamental issues and future developments, SEG Expanded Abstract, 520
[18] Woodward, M. J., 1992, Wave-equation tomography: Geophysics, 57, 15–26

地震波反演成像方法与技术核心问题分析

王华忠 冯波 王雄文 胡江涛 李辉

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】现代勘探地震中,岩性油气藏的精细描述主要是利用高精度采样（宽方位、宽频带和高密度采样）的高维地震数据反演估计精细的背景参数变化（主要是速度、各向异性和 Q 值）和高保真及高分辨的高波数参数变化（主要是反射系数），结合岩石物理知识，进行油气预测。精细背景参数变化的估计主要是利用层析成像技术；高保真及高分辨的高波数参数变化估计方法主要是保真叠前深度偏移或最小二乘叠前深度偏移技术。现代地震勘探的主要特征是整个成像处理流程逐步进入以地震波反演成像为主的阶段，其中的每一个环节基本上都可以抽象为建立正演模型和 Bayes 参数估计两个关键步骤；另一个主要特征是现代的地震反演成像技术以岩性油气藏的描述为目标。

常规的地震反演成像分为偏移速度分析与层析成像、叠前深度偏移(角度道集产生)和 AVA 分析/反演三个重要环节，其中的关键技术是当前勘探地震学中的核心技术。

FWI 是最近几年重新受到业界特别关注的地震波反演成像核心技术。它是理论意义下十分完善的地震波反演成像理论框架。原则上，FWI 可以把上述三项常规的反演方法合为一体，给出比较理想的反演成像结果。但是，由于叠前数据的不完备、正演模型的不完善、初始模型不够精确、地震子波的未知和空变使得严格意义下的 FWI 方法并不能很好地解决实际问题。

本文分析经典 FWI 存在的问题，提出适合我国勘探地震数据情况的地震波反演成像技术发展路线。通过叠前地震数据的特征表达和地下介质模型的特征表达，提出把经典的 FWI 分成：透射波层析成像、最小二乘叠前深度偏移成像和反射波层析成像三个线性化反演方法的串联。指出把浅层速度模型建立、背景速度模型建立、成像道集层析速度模型建立、最小二乘叠前深度偏移成像、张角及界面倾角道集的产生以及与各项异性有关的裂缝和裂隙参数分析、小角度成像道集波阻抗反演作为当前油气地震勘探中近期可实现的地震波反演成像路线。

【关键词】反演成像理论，多级 Born 近似，全波形反演，逐级线性化反演，实用化地震反演成像技术路线

1 引言

一般意义下的地震波成像所指的就是地震波偏移成像。它是在假设偏移速度场已知的情况下，利用在该速度场中计算的波场反向外推算子，把地表记录的波场外推到绕射点上，用适当的成像条件提取成像值，目标是定量地定位反射系数出现的位置，并定性地保持反射系数的相对强度，尤其是角度反射系数的相对强度。

在此相对保真的叠前成像道集的基础上，进行地震子波属性的提取及基于地震子波属性的储层识别与描述。也可以进行 AVA 分析/反演进行半定量的储层识别与描述。

很显然，假定有比较合适的叠前地震数据，正确的背景速度场和保真的角度成像道集是进行后续的储层解释的基础。

地震波偏移成像与地震波反演成像之间的区别是明显的。原则上，前者定量地估计反射点的位置、定性地估计角度反射系数的大小；后者试图定量地估计全波数带的速度场（多参数反演情况下也包括其他参数的全波数带成分的估计）。

地震波反演成像的代表性技术是全波形反演（Full Waveform Inversion，简写为 FWI），它的目标是估计全波数带的速度场（也可以估计全波数带的其它参数场）。

地震波反演成像的基本思想是 Bayes 估计理论。Tarantola（1984,2005）详细而深入地分析了地震波全波形反演的理论问题。Bayes 估计理论更是现代信号分析中的核心内容(Kay，1993,1998)。在整个地震数据处理与反演成像过程中，去噪音、反褶积、高维数据规则化、一维波阻抗反演、AVA（叠前）弹性参数反演、最小二乘叠前偏移成像、层析成像等都是在此框架下进行的。根据 Bayes 估计理论，首先假设存在一个正演算子，它可以预测观测数据。若预测噪音为高斯白噪，此时的 Bayes 估计可以在最小二乘意义下实现，此为最大似然估计方法。假设已知关于模型的先验概率分布，可以建立一个联合的后验概率分布，数据误差一般用 L2 范数，模型约束一般用 L1 或 Cauchy 范数等；根据正算子的线性与否，选择用线性最小二乘方法或非线性最小二乘方法求解反演问题。此为经典的 Bayes 估计方法。

预测观测数据的正算子的作用十分重要。褶积算子用于反褶积；一维波动方程用于预测自激自收的地震道（波阻抗反演）；AR 模型算子常用于线性信号预测（压制噪音）；旅行时计算算子用于预测实测数据中的旅行时（旅行时层析成像）；Zoeppritz 方程及各种简化形式预测角度反射系数（AVA 弹性参数反演）；Fourier 变换算子用于预测规则无假频数据的谱；动校正时距关系预测实测 CMP 道集中的双曲时距关系（估计均方根速度）；各种形式的波动方程预测实际观测的炮记录（全波形反演）。可以看出每一类正算子对应一类反演问题。正算子建立起了要估计的参数与观测数据之间的线性或非线性关系。实际上，很多情况下，正演结果并不能很好地预测实测数据，导致反演结果精度降低，甚至反演过程根本不收敛。

反演问题是典型的不适定问题，包括解的存在性、唯一性和稳定性问题。在地球物理反演中，解的存在性是有物理保证的，问题主要是解的不唯一性和计算过程中的不稳定性。二者具有关联性，计算的不稳定性加剧了解的不唯一性，但最根本的还是解的不唯一性问题。反演问题解的不唯一性本质上是由数据空间向模型空间映射的非线性性引起的，更进一步地，是由模型参数变化的复杂性和描述波场和参数场关系的正算子决定的。不唯一性只可以减弱，不可以完全消除。减弱反演解的不唯一性的关键是减弱由数据空间向模型空间映射的非线性性、提高初始模型的精度、增加关于模型的先验信息，这些措施可以有效地提高反演解的精度。目前，FWI 技术发展过程中提出的所有方法都是基于上述三点的。

20 世纪 80 年代，地震波反演处于热潮时期，很多学者对全波形反演理论做了十分扎实的奠基工作（Beylkin, 1985; Bunks, et al, 1995; Lailly, 1983; Mora, 1987,1988; Nolet, 1987）。但受当时野外地震观测技术和高性能计算机技术的局限，无法将提出的理论方法转为实用。此后，一维波阻抗反演、AVA 叠前弹性参数反演这些计算需求小的反演方法得到了不断发展。但从理论上看，这些反演方法都不具备高维叠前反演的潜力，施加给介质分布情况的假设过多。从实用上看，它们的反演精度从来没有被认真地考究过。就是说这些反演结果并没有对地震解释产生过实质性的影响。

近几年，高维叠前地震数据的全波形反演理论与方法重新成为了勘探地球物理方面的核心议题，这首先得益于地震数据采集技术和大规模高性能计算机技术的进步，很多研究者对地震波高维叠前反演研究的兴起也做了大量铺垫工作（Pratt，1990, 1999; Biondi, and Symes, 2004; Lambaré, et al, 1992; Shin, and Cha, 2009）。宽方位、高密度和宽频带（尤其是低频）数据的采集技术和 GPU/CPU 异构计算机群技术为 FWI 技术的可能应用奠定了很好的基础。

但是，经典 FWI 技术的实际应用遇到了巨大的困难。真正成功的应用实例不多。叠前数据的不完备、正演模型的不完善、初始模型不够精确、地震子波的未知和空变是主要的四个因素。因此，近几年研究者们逐步认识到经典的 FWI 是很难得到合适的反演解的，提出了各种各样的变种，基本的思想是尽量克服上述四个方面的影响因素，使得反演结果更加稳定，反演过程的收敛性更强。

从理论发展的角度看，目前以地震波偏移成像为主的叠前地震数据成像处理走到以 FWI 技术为代表的地震波反演成像是无法回避的，但是 FWI 方法技术的实现还是要从系统的观点出发，依据实际数据的不同情况，构建出合适的技术流程，达到 FWI 的目的。它的目的是得到全（宽）波数带的速度估计结果（也可以包括其他参数的估计结果）。理论上讲，FWI 技术的实现等于把叠前数据采集、反演成像和储层解释有机地融合成在一起。

当前对 FWI 方法技术的定位一般地认为它可以提供满足 RTM 成像要求的更为精细的速度场。实质上，FWI 的真实目的是得到全（宽）波数带的参数估计，类似 RTM 的偏移成像方法是辅助 FWI 得到波数带更宽的参数估计结果，而不是 FWI 仅仅提供一个更好的偏移速度给 RTM（Samuel，2014）。据此可以看出，FWI 是期望从叠前数据中定量地估计弹性参数（甚至岩性参数和流体参数）的技术，但是能否达到这样宏大的目标，叠前数据的质量是关键因素。

因此，本文首先从叠前地震数据抽象认识入手，提出特征波的概念满足不同层次的 FWI 对数据的要求，接着提出对参数空间及地震波传播的抽象认识，建立反演参数和特征波场之间的关系，据此尽量强化二者之间的线性关系，同时说明正问题线性化的必要性及给出比较精确的初始模型的必要性。在此基础上，给出地震波反演成像的合理的技术流程，达到实现全（宽）波数带的参数估计。

我们认为，地震波反演成像的目的是通过弹性参数的估计刻画和描述储层。这是下一代地震成像技术的显著特征，标志着勘探地震学全面进入反演阶段。

2 油气勘探地震中关键问题分析

油气地震勘探的终极目标是尽可能精确地描述含油气储层。宏观地讲，这是一个更大的反问题。所依据的数据包括地震数据、测井数据、地质数据（包括基本地质知识）、岩石数据（包括基本岩石性质知识），反演过程归结为一个综合判断和决策过程。当前，世界上大的地球物理服务公司试图推出这样的综合性的软件系统。可以简称为采集、处理和解释一体化系统，但这样的系统在决策阶段存在软肋，解释方面也是以构造解释为主，含油气性的判断甚至远没有到达半定量的程度。在反演理论大框架下，向智能化的地震数据采集、处理、解释和决策一体化方向发展是没有异议的发展方向。目前存在如下一些共识：（1）地震数据采集向宽方位、高密度和宽频带数据采集发展；（2）反演成像向全（宽）波数带参数估计方向发展的趋势已十分明显；（3）地震地质解释的全三维化与智能化正逐步展开。这三点的目的都是指向更精确更智能化地描述储层。

描述地震储层时，地震波场中携带了哪些信息可以对储层的存在和储层的性质有所指示是油气地震勘探工作者必须思考和弄清楚的问题。这也是反演理论中的基本概念，即数据中一定包含要反演参数的变化信息，而且二者的关联性还要足够强。

目前，可以确认的是地震数据同相轴的到达时与地震波速度、各向异性参数是有密切关系的；地震波阻抗的变化与反射地震波振幅有密切关系；吸收衰减参数（Q 值）与反射地震波的振幅有明显的关系；地下介质的散射与地震波振幅、频率和相位存在明显的关系。充分利用上述关系来解决储层描述和刻画是关键问题。目前的地震波反演成像还远没有把上述关系利用得很彻底。

叠前地震数据是波在地下介质中传播后的累计效应，因此利用叠前地震数据反演得到的往往也是介质性质参数变化的光滑结果。对速度、各向异性参数和 Q 值的估计得到的都是这样的结果。从局部平面波传播的角度看，叠前地震数据与地下介质参数之间由局部平面波传播联系起来，方位角度成像道集或方位倾角成像道集是油气储层描述的核心的基础数据。它们是成像域的高维叠前数据体，直接携带了与储层及周边岩性变化有关的信息。到目前为止，在地震反演成像过程中，产生尽可能保真的方位角度成像道集或方位倾角成像道集是进一步描述裂缝储层、孔洞储层和其他类型岩性储层的基础。对波阻抗变化、裂隙方向和密度、散射体的估计需要在这样的道集上进行，利用成像道集上的振幅变化考察的是参数的高波数变化成分。

3 叠前地震数据的抽象认识

叠前地震反演成像依赖的是单炮道集表示的地震数据，一个探区的所有单炮道集可以表示为

$$\left\{D_S^i\left(\mathbf{X}_s,\mathbf{X}_r,t\right),i=1,2,\cdots,NS;\mathbf{X}_s\in R^2,\mathbf{X}_r\in R^2\right\} \tag{1}$$

其中，$\mathbf{X}_s=\left(x_s,y_s\right)$ 和 $\mathbf{X}_r=\left(x_r,y_r\right)$ 分别代表炮点和检波点坐标。

期望这样的数据是宽方位、高密度和宽频带（尤其包含低频成分）的。原因可以由下列定理清楚地表明。

定理(Born, M., and Wolf, E., 1999)：常速情况下，在一阶 Born 近似成立的条件下，在散射体的远场范围内，散射势与它的波数谱之间存在如下关系，

$$f\left(\mathbf{s},\mathbf{s}_0\right)=\int_V F\left(\mathbf{r}\right)e^{-i\mathbf{K}\cdot\mathbf{r}}d^3r \tag{2}$$

其中，$K=\dfrac{\omega}{v}\left(\mathbf{s}-\mathbf{s}_0\right)$，$\mathbf{s}_0$ 代表入射波方向，$\mathbf{s}$ 代表出射波方向。

由（2）式可见，如果得到宽波数带的散射势谱就可以通过 Fourier 反变换得到高分辨率的散射势的估计。散射势与速度扰动密切相关，很多文献上都有定义。散射势的波数分布范围由 $K=\dfrac{\omega}{v}\left(\mathbf{s}-\mathbf{s}_0\right)$ 给出。其中矢量 $\mathbf{s}-\mathbf{s}_0$ 定义了散射（或反射）张角，观测系统决定了张角的范围，张角范围越大，波数谱越宽，对应地，要求地震数据的观测角度越大，即要有充分长的偏移距和充分宽的方位角。其中频率范围也要求足够宽，尤其是低频成分对散射势的低波数成分贡献很大。至于叠前地震数据中的高密度采样，主要是从反假

频，尤其是对某些特殊的低视速度噪音需要密的空间采样，而提出来的。当然，增加覆盖次数，提高信噪比也是加密采样的基本考虑。

更进一步地，理论上来说，地震波反演成像要求叠前地震数据采集对地下任何一个绕射点（反射点）有广角度的、角度间隔均匀的、不产生采样假频的照明。同时期望每个角度的数据中仅仅有高斯白噪音。还期望各角度之间的子波特征是保持一致的。

目前，比较流行的经典意义下的 FWI 反演是利用变密度声波方程，甚至常密度声波方程进行地震波正演，当然，也有用弹性波方程进行地震波正演来预测实测炮道集的。

无论用何种波动方程来预测实测数据，都不可能很精确地预测实际数据。其中的影响因素实在太多。不能预测的部分变成数据残差，使得该残差不符合 Gauss 分布，问题的非线性性变强，解的非唯一性变强。

问题的解决办法是强调对数据的预处理，不强调使用复杂的正演模拟方程。为此我们提出了特征波的概念。所谓特征波是指单一震相的时空局部的波。可以是反射波、绕射波或透射波，也可以是某种波的波形、包络、相位、甚至是到达时。我们认为当前的反演成像还是要先针对单一震相的波现象进行。

另一方面，我们把单炮地震数据抽象地描述为：把地表（也可以是井中或海底（OBC）等处）观测的地震数据视为以旅行时、炮检点坐标（或中点半偏移距）为变量的函数，它的基本特征应该是时距关系规定的、带限反射子波表现出的同相轴，叠加上一定的随机噪音。速度（和各向异性参数）场的背景部分决定同相轴出现的位置，Q 值和波阻抗变化决定反射子波的相对振幅。观测数据中子波振幅的绝对值影响因素甚为复杂，海上地震勘探还比较容易分析清楚，陆上叠前地震数据中绝对振幅的具体含义及影响因素还要深入分析。这也说明了基于波形振幅差的一类 FWI 反演技术难于得到合理解的基本原因。所有基于反射振幅的反演方法都存在需要讲清楚实测数据的振幅是什么含义的问题。据此，我们更进一步地提出首先利用同相轴的到达时信息，反演背景速度、各向异性参数。引入吸收衰减旅行时也可以反演背景 Q 值参数。在此基础上，进行带吸收衰减的各向异性叠前深度偏移成像产生保真的方位角度成像道集或方位倾角成像道集，并进一步描述和刻画储层是目前比较现实的地震波反演成像技术路线。

更进一步地，把单炮道集中的波现象分为透射、反射波和绕射波等。当然也包括多次波和 AVA 现象。对应于近偏移距的折射波和直达波的这部分透射波，主要用来进行浅层近地表速度层析反演；对应于中长偏移距中的来自中深层的折射波和潜波（Diving Wave）的这部分透射波，主要用来进行中深层背景速度的估计。因为透射波是不受小尺度速度异常体的影响的，利用它也只能反演大尺度的背景速度场。目前叠前偏移成像主要用的是一次反射波和/或绕射波。RTM 原则上可以对多次波成像，但需要精确的、包含正确的反射界面的速度场和修正的成像条件。若能提供这样的速度场，RTM 也就没有必要了！二次反射波（绕射波）目前被用来进行反射波层析反演，估计背景速度的扰动量，提供更精确的偏移速度场。但是，仅仅考虑了 Fresnel 带内的二次散射波。

我们认为在这三个层次上认识地震数据，可以构造出比较合适的地震波反演成像流程。事实上，构建任何地震反演成像方法都要从叠前数据中所包含的信息能否实现反演目的出发。

4 参数空间及其中地震波传播的抽象认识

地震波在实际介质中的激发、传播和接收过程是地震波偏移成像和反演成像的基础。用一个人为构建的波传播算子进行地震波成像是其核心内容。地震波在实际介质中的传播过程和波场的检测过程是异常复杂的，而人为构建的波传播算子比较简单，只能描述复杂的物理波场的一部分，其他的波场成分要么处理掉，要么留下来或者增加偏移成像噪音或者增加反演成像的不唯一性。这是勘探地震工业中形成了一个复杂地震数据处理流程的基本原因。

我们把勘探地震所面对的地下介质抽象为广大的、横向缓慢变化的沉积层中分布着由于火山活动、后期地质构造运动和其他地球动力运动引起的小尺度速度异常体，譬如地震勘探中的绕射地质体。像断层、裂缝、地层尖灭、粗糙界面、孔洞等（大）小尺度的地质异常体是常见的油气运移通道和/或储集体，对它们的刻画和描述是油气资源勘探的重要目的。

针对这样的介质情况，我们提出如下的抽象：地下介质的速度场可以分为背景速度加反射界面处的速度扰动：

$$v(x,y,z)=v_B(x,y,z)+\delta v_B(x,y,z)+R(x,y,z;\gamma,\phi) \tag{3}$$

地震数据中不同偏移距接收到的反射同相轴的到达时间取决于背景速度 $v_B(x,y,z)$ 或 $v_B(x,y,z)+\delta v_B(x,y,z)$；同相轴上的反射波振幅取决于反射界面处波阻抗的变化。不同偏移距的数据包括透射波数据（直达波和 Diving Wave）和反射波数据（一次反射波和多次反射波）。上述观察决定了地震速度反演的基本方法。

地震波在这样的介质中的传播至少可以用两种模式来描述：WRW 模式和 Born 近似模式。WRW 模式借用了 Berkhout（2012）常用的描述波传播的方法，它认为地下介质是光滑的背景介质中叠加上具有一定反射系数的反射界面组成的。波在这样的介质中传播，地面观测的数据可以由 WRW 模式对应的正问题来描述，即下行波场遇到反射界面反传回地表形成地面记录。原则上，这样的模式还可以描述多次波传播。Born 近似模型视地下介质为背景速度加上小尺度的散射体。地表数据是所有地下散射体引起的散射波叠加形成的，可以仅仅包括一次散射，也可以同时包括多次散射。

当前，油气地震勘探的理论主要还是奠基在反射波的基础上。CMP 道集处理、角度成像道集处理、AVA 分析都隐含了反射波成像处理的假设。

基于上述参数分解方式及对应的波传播模式，可以构建出合理的地震波反演成像的实用化流程。

5 地震波反演成像的概率论观点及相应的解法

综合各种文献和我们的研究经验，认为地震波反演成像技术奠基在 Bayes 估计的理论框架上时，其理论基础最为扎实，对反演问题的认识和理解最为透彻。

首先，观测到的地震数据是随机信号，它是满足一定的概率分布的。实际上，假设地震波在实际介质中的传播的确可以由一个正问题比较精确地预测，预测结果与实测数据的残差也是随机信号。地震信号的随机性可以理解为确定性的传播波场叠加上不可预测的随机噪音。一般地，我们认为这部分不可预测的随机噪音满足高斯分布。因此，随机地震信号可以表示为

$$\mathbf{d}^{obs}=F(\mathbf{m})+\boldsymbol{\eta} \tag{4}$$

其中，$\mathbf{d}^{obs}$ 为实测波场；$F(\cdot)$ 为地震波正演算子；$\mathbf{m}$ 为要反演的介质参数；$\boldsymbol{\eta}$ 为随机噪音。上式是一个非线性方程组。

事实上，可以用 Newton 法直接解（4）式进行地震参数的反演。但是，这样的做法不利于认识地震参数反演问题的本质。地震波反演成像问题不是一个简单的、可以归结为非线性方程组求解的数学问题！

借助于 Bayes 估计的观点，对地震波反演成像的思想的理解会更为本质和深刻。Bayes 估计的观点为：参数 $\mathbf{m}$ 的反演结果或估计结果应该由后验概率密度函数 $P(\mathbf{m}|\mathbf{d}^{obs})$ 所定义的均值决定，即

$$\hat{\mathbf{m}}=\int \mathbf{m}P(\mathbf{m}|\mathbf{d}^{obs})d\mathbf{m} \tag{5}$$

条件概率密度 $P(\mathbf{m}|\mathbf{d}^{obs})$ 可以理解为：观测数据为实测数据 $\mathbf{d}^{obs}$ 时，对应不同的参数 $\mathbf{m}$ 时的概率。可见，用（5）式进行参数 $\mathbf{m}$ 的估计是非常合乎逻辑的。但是，高维参数反演情况下的后验概率密度函数 $P(\mathbf{m}|\mathbf{d}^{obs})$ 几乎是无法实际计算的。

首先，利用著名的 Bayes 公式把后验概率密度函数 $P(\mathbf{m}|\mathbf{d}^{obs})$ 的计算转化为先验概率密度函数的计算。即

$$P(\mathbf{m}|\mathbf{d}^{obs})=\frac{P(\mathbf{d}^{obs}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d}^{obs})} \tag{6}$$

（6）式中引入了实测数据 $\mathbf{d}^{obs}$ 对反演参数 $\mathbf{m}$ 的先验概率密度函数 $P(\mathbf{d}^{obs}|\mathbf{m})$、反演参数 $\mathbf{m}$ 本身的先验概

率密度函数 $P(\mathbf{m})$ 和实测数据 $\mathbf{d}^{obs}$ 的先验概率密度函数 $P(\mathbf{d}^{obs})$。实测数据 $\mathbf{d}^{obs}$ 对反演参数 $\mathbf{m}$ 的先验概率密度函数 $P(\mathbf{d}^{obs}|\mathbf{m})$ 反映了用任一个反演参数 $\mathbf{m}$ 借助正演模型预测实测数据时预测正确性的概率或预测误差满足的概率。反演参数 $\mathbf{m}$ 的先验概率密度函数 $P(\mathbf{m})$ 的引入相当于引入了对反演参数的某种正则化约束。关于实测数据 $\mathbf{d}^{obs}$ 的先验概率分布 $P(\mathbf{d}^{obs})$，一般地我们认为是没有任何先验信息，因此假定该先验概率分布是均匀分布，取常数值。

后验概率密度函数 $P(\mathbf{m}|\mathbf{d}^{obs})$ 取最大值时对应的参数 $\mathbf{m}$ 就是所要的反演结果从统计意义上看也是很有道理的。但是，由于（6）式中概率密度函数的复杂形式，直接对（6）式进行优化求解是非常不方便的。

因此，一般地，假设 $P(\mathbf{d}^{obs}|\mathbf{m})$ 和 $P(\mathbf{m})$ 这两个概率密度函数都是高斯函数，结合对 $P(\mathbf{d}^{obs})$ 的假设，有

$$P(\mathbf{m}|\mathbf{d}^{obs}) = C\cdot\exp\{-S(\mathbf{m})\} \tag{7}$$

其中，C 为常数；$S(\mathbf{m})$ 有如下的形式

$$S(\mathbf{m}) = \frac{1}{2}\left[\left(F(\mathbf{m})-\mathbf{d}^{obs}\right)^T C_D^{-1}\left(F(\mathbf{m})-\mathbf{d}^{obs}\right)+\left(\mathbf{m}-\mathbf{m}_{prior}\right)^T C_M^{-1}\left(\mathbf{m}-\mathbf{m}_{prior}\right)\right] \tag{8}$$

$S(\mathbf{m})$ 被称为代价函数。C_D 是数据自相关矩阵；C_M 是模型自相关矩阵。可以明显看出（8）式与 Tikhonov 正则化反演公式非常类似。说明了 Bayes 估计理论有更高的理论包容性，是反演理论的更为完整的框架。由于指数函数的单调性质，当 $S(\mathbf{m})$ 达到极小时，后验概率密度函数 $P(\mathbf{m}|\mathbf{d}^{obs})$ 取最大值。因此，反演问题转化为求如下优化问题

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\{S(\mathbf{m})\} \tag{9}$$

其中，$\mathcal{H}_1$ 代表参数函数取值空间；$\mathcal{H}_2$ 代表数据函数取值空间。

进入（9）式的具体求解方法之前，先分析实际地震数据反演与上述 Bayes 反演理论假设之间存在的距离。

Bayes 反演理论假设实测地震波场是可以被预测的，预测结果与实测数据的残差满足高斯分布。也可以近似满足 Gauss 分布，即满足一类广义 Gauss 分布。但是，实际情况下，地震波正演模拟不可能很精确地模式实际的地震波场。残差中既有观测随机噪音（$\boldsymbol{\eta}$），又有不能模拟的波场成分，后者有可能大于前者。这会导致残差满足或近似满足 Gauss 分布的假设失效，反演结果可能严重偏离实际情形。这是我们提出用特征波场或当前的地震波正演方法能较精确模拟的波场进行地震波反演的根本原因。地震子波的未知和空变也可以归结为地震波正演方法不能很好地模拟实测波场，但解决的方法是尽量不用振幅逼近，而是用波场相关或旅行时逼近来回避子波未知导致的正演波场振幅与实测波场振幅的不匹配。另一方面，实测波场的振幅影响因素非常复杂，涉及到震源与介质的耦合、实际震源的方向特征、检波器与介质的耦合、检波器本身的响应特征，以及波在复杂介质中的传播过程等等，并非解波动方程可以预测的。这也说明了理论模型数据上经典的全波形反演做得很好，但实际数据总做不好的原因。

地震波反演成像包含四个关键环节：地震波场数值模拟、模型表达（或模型参数化）、数据表达（或数据特征化）、反演方法。地震波数值模拟方法及线性化方法是其中最根本的。

（9）式描述的最优化的问题的求解，从根本上讲，是通过正问题的线性化把它变成一个真正的近似凸的或凸的优化问题。这应该是反问题最本质的事情（Boyd and Vandenberghe, 2004）。一旦变成了近似凸或凸的优化问题，其求解方法总是存在的。

化为近似凸或凸优化问题并没有现成的方法，我们认为从低尺度逐步过渡到高尺度的反演方法、利用特征波场的方法、把炮道集中的波场分为透射波、反射波串联反演的方法都是化非凸优化为近似凸优化的思想方法。

优化问题非凸，即存在多解性的根源，在于参数变化与波场变化之间的非线性性。该非线性性有两个

层次，一个是实际物理层面上的；一个是建立的数学物理模型层次上的。一般地，我们也不去区分这两个层次的区别。对非线性性的数学物理模型的线性化，就认为把一个非线性的反演问题化成了线性的，非凸优化问题就可以化为凸优化问题。实际上，我们构架的最小二乘反演问题、层析反演问题，看起来都是凸优化问题，但多解性依然很强的主要原因是数学模型层次上的线性化，并不等价于实际物理层次上的线性化。

一般地，我们只解近似凸优化问题。若 $S(\mathbf{m})$ 是凸的，最速下降方法、共轭梯度方法、Gauss-Newton 方法和全 Newton 方法都可以较好地实现反演方法的求解。如果 $S(\mathbf{m})$ 不是严格凸的函数，则 $S(\mathbf{m})$ 的极小值点不再唯一，解决此类问题可以用信赖域方法、Monte Carlo 方法(Mosegaard, and Tarantola, 1995)等。这些方法都可以用来求解弱非线性最优化问题，但由于 Monte Carlo 方法计算量大，如果没有一个比较好的策略很难实现，因此信赖域方法在这种情况下是一个比较好的选择。

对应非线性性很强的问题，难于找到一个合适的局部点（初始点）把正演问题 $F(\mathbf{m})$ 进行线性展开得到凸性较好的 $S(\mathbf{m})$。此类反演问题目前是无法求解的（Menke, 1984; Vogel, 2002; Aster, Borchers, and Thurber, 2005）。勘探地震中，小尺度发育的、复杂构造探区的反演成像连构造都无法确定就是此类问题。针对此类问题，能做的工作依然是不断地增加数据和参数两方面的、更丰富的信息。

对反演问题最后结果的评价原则上应该是统计决策问题，主要评价依据是反演结果的均值和方差，其中方差的意义尤其重要，它反映了反演结果的可靠程度。

事实上，Bayes 参数估计是一套概念，不是一种技术。在此概念下，地震波反演成像结果的好坏由很多因素决定。对 Bayes 框架下的地震波反演成像的本质问题深入理解并灵活运用才是关键的。

6 经典的 FWI 反演成像与多级线性反演成像

从上节的讨论可见，经典的 FWI 理论是精致而完善的，具体实现也不复杂。但是，FWI 反演结果总不能如理论预计。主要的问题前面已经述及，即：叠前数据的不完备、正演模型的不完善、初始模型不够精确、地震子波的未知和空变。因此，目前关于 FWI 的大量研究都是试图让 FWI 技术能在油气勘探的实践中发挥作用。

地震波反演成像的困难从根本上讲是观测数据与反演参数之间的线性性不强，以及参数的扰动不能引起观测数据的变化（Frechet 微商等于或接近于零）。前者说明反演问题解的非唯一性，后者说明地震反演成像并非对任何参数都可以进行。目前 FWI 技术发展侧重的几个方面是：（1）修改数据残差和模型差的先验概率分布，不再要求严格满足 Gauss 分布；（2）不用数据差而是用数据相关定义泛函，强化旅行时影响、弱化子波影响和振幅匹配；（3）不用原反射子波，改用反射子波的包络，来定义逼近泛函。弱化对低频的依赖性、弱化对初始模型精确性的依赖。（4）不用全波场逼近，分波型进行 FWI 反演，甚至退化为分波型的基于包络、基于相位、甚至基于旅行时的反演。

经典的 FWI 反演技术退化到分波型反演时，即认为 FWI 反演可以按一个流程来实现。原则上，（6）式定义的经典的 FWI 反演是完全自动化的反演成像技术，不需要对叠前数据做任何预处理。实际上，对于复杂的叠前数据情况和复杂的地表情况，FWI 技术要应用于实际数据上，必须构建合适的反演成像处理流程。

根据我国实际数据的情况，我们提出如下实用化的流程（称为 FWI 反演成像实用化流程）：

1、 利用透射波（折射波和 Diving Wave）估计背景速度 $v_B(x,y,z)$

利用中近偏移距数据的初至波旅行时层析可以得到近地表速度，利用中大偏移距的低频波场（早至波波场）可以估计中深层背景速度。

2、 利用反射波+$v_B(x,y,z)$ 确定反射系数 $R(x,y,z;\gamma,\phi)$

利用一次（或多次）反射和散射波进行最小二乘偏移估计反射系数，其中包括积分法的（LS_Kirchhoff_PSDM 和 LS_Beam_PSDM）和波动方程类的（LS_WE_PSDM 和 LS_RTM）。

3、利用反射波 + $v_B(x,y,z)$ + $R(x,y,z;\gamma,\phi)$ 确定 $\delta v_B(x,y,z)$，并更新背景速度 $v_B(x,y,z) \Leftarrow v_B(x,y,z)+\delta v_B(x,y,z)$。

利用背景速度和上一步估计的反射系数反偏移产生一次反射波。与实测的一次反射波逼近进行波动理论层析，也可称反射波 FWI，估计背景速度的扰动 $\delta v_B(x,y,z)$。同样地，此时可以利用全波形（相位或到达时）或包络波形（包络相位）等。

其中，第 2 步和第 3 步需要迭代执行，直到角度成像道集整体拉平为止。在此基础上，利用小角度反射系数可以反演波阻抗（或密度）。

上述流程中的三步核心技术都是一种线性化的反演方法，我们把这样的 FWI 实现方式称为逐步线性化反演或串联线性化反演。一般地，每项技术都可以进一步地用特征波的概念来实现，把复杂的波进一步分为单个震相，增加它们的实用化程度。

与上述流程相对应，还存在一个目前石油工业界正广泛使用的成像处理流程，产生保真的方位角度成像道集是核心。

该流程（称为常规叠前深度偏移成像处理流程）可以简单罗列如下：

1、（射线理论）初至波和/或折射波层析技术---估计近地表（背景）速度模型 $v_B^{Near_surface}(x,y,z)$

2、水平层状介质假设下动校正+速度分析技术---估计背景速度 $v_B(x,y,z)$

3、叠前偏移剩余速度分析---估计速度扰动量 $\delta v_B(x,y,z)$

4、角度成像道集层析速度反演---估计速度扰动量 $\delta v_B(x,y,z)$

5、波动方程叠前深度偏移---估计方位角度反射系数 $R(x,y,z;\gamma,\phi)$

对比上述两个流程，简单地看，二者的目标是一致的，都是希望得到全（宽）波数带的速度成分。二者的重要差异在于对背景速度的估计上。FWI 反演成像实用化流程用透射波（直达波、折射波和 Diving Wave）估计背景速度。存在低频长偏移距观测数据时，炮集中早至波成分主要是折射波和 Diving Wave，它们来自中深层。利用这些波现象进行 FWI 可以估计背景速度 $v_B(x,y,z)$。建立的背景速度的精度要高于常规叠前深度偏移成像处理流程。用低频长偏移距的透射波估计背景速度是当前 FWI 对实际数据处理的切实贡献。在很多海上数据数据的 FWI 处理中，都是因为背景速度精度的提高改善了成像质量。但改善并不显著。这就容易令人产生错觉：FWI 是为了提高偏移速度精度而存在的。实质上，FWI 的目标是得到全（宽）波数带的参数估计。它试图提供高精度的弹性参数估计结果，直接进入岩性储层的解释。当然，当前的常规叠前深度偏移成像处理流程本质目的也是得到全（宽）波数带的参数估计，但它往往定位在提供保真的方位角度成像道集上，基于此进入岩性储层的解释阶段。

7 针对陆上地震数据的 FWI 反演成像技术流程

针对我国陆上地震数据的特点，推进反演成像技术的进展，关键点在于：

（1） 地表因素空变引起的子波特征空变的减弱；

陆上地表条件横向变换剧烈，导致激发接收因素空间变化，反映到每一个地震道上，接收到的反射子波受炮检点地表因素的影响。

减弱空变子波特征的影响首先是在估计背景速度的变化时充分利用到达时信息和相位信息，少用或不用振幅信息。第二位的是做好子波空变特征的一致性处理。

（2） 初至波和早至波信息的充分利用；

陆上数据的反射波和绕射波信噪比低，初至波和早至波一般而言具有更高的信噪比。但是，目前往往仅仅有中小偏移距的初至波到达时被用来进行近地表速度层析。这一部分较高信噪比的数据没有

被充分地研究和利用。

（3） 背景速度估计与建模；

陆上数据一般地缺乏低频长偏移距数据，很难期望用透射波 FWI 估计背景速度$v_B(x,y,z)$。仅仅靠动校正和均方根速度估计在信噪比低和横向变速大的情况下背景速度估计精度是远远不够的，这会导致后续的背景速度扰动（$\delta v_B(x,y,z)$）估计方法不收敛，无法提高成像质量。

背景速度$v_B(x,y,z)$的估计是一个非线性反演过程。我们提出用基于 Kirchhoff PSDM 和成像道集拉平的 Monte Carlo 反演方法估计横向变速情况下的背景速度。2014 年 SEG 年会论文中有作者用 Monte Carlo 方法反演背景速度为后续的 FWI 提供初始速度模型（Sajeva, et al., 2014）。

（4） 做好角度成像道集的层析速度反演和建模；

角度成像道集中射线层析与建模是当前深度速度建模的主力工具。但是，它的估计精度与正则化思想与方法是密切相关的。最近几年因为现代图像处理技术的引入，反射界面有机地融入反演速度场中，极大地提高了估计精度。

对于陆上地震数据的角度成像道集的层析速度反演，在提高角度道集质量，自动拾取剩余时差，自动反射界面解释及估计倾角，Beam 层析，正则化和预条件等方面的改进会显著提高估计精度。另外，角度成像道集的层析速度反演与各向异性参数和 Q 值反演可以方便地融为一体，实现多参数联合建模。

（5） 做好最小二乘反演成像；

在上述几步工作的基础上，叠前深度偏移成像的目标就是至少做到保真的方位角度反射系数道集的生成。我们认为，从实用的角度看，Kirchhoff 积分角度域最小二乘偏移成像是比较具有广泛实用意义的。最小二乘 RTM 的计算量太大，难于大规模地应用于实际数据的成像处理中。

（6） 做好波阻抗反演工作。

波阻抗是与岩性变化关系更密切的参数。地震波反演成像的目标至少应提高到对波阻抗的反演成像上。简单地看，波阻抗与反射系数是积分和微分的关系。前者的微分是后者，反之，后者的积分是前者。从 FWI 反演成像角度看，小角度的反射系数与波阻抗的关系更为密切。Zhang 等（2013）和 Duan 等 (2014) 在 SGE 论文中提及用小角度反射系数估计波阻抗的方法。

上述 6 个关键点是针对陆上地震数据进行反演成像提出的，也可以把它们组成一个流程。与 FWI 反演成像实用化流程比，它更有可能促进反演成像方法技术的普及应用。

另外，特别值得关注的是，陆上地震数据噪音远大于海上地震，波场特征分解的思想更应体现在反演成像过程中，一是提高信噪比，二是降低正演模拟复杂波场的难度（特征波场模拟演比全波场模拟简单）。另一方面，背景部分的速度估计并不需要像叠前深度偏移那样要求数据很规则，利用波场的特征波分解可以把高信噪比的部分数据分出来，仅仅用这部分数据更新背景速度的扰动。

信号的特征提取对整个信号处理诸方面（譬如数据规则化、数据压缩、数据采样、去噪音等）都有重大影响，波场的特征分解在反演成像中同样会非常重要。2014 年 SEG 年会论文中已有作者关注波场特征分解与 FWI 反演数据预处理的关系（Hongbo Bi and Tao Lin 2014）。

陆上地震数据的 FWI 反演成像需要构建更现实的技术流程。若前述 6 个问题解决得比较好，而且每个阶段的反演方法中正则化思想运用得当，陆上地震数据的 FWI 反演成像会逐步向前推进。

8 地震波反演成像中的正则化与预条件

地震波反演成像的不唯一性和不稳定性是其典型的特征，得到稳定解的基本方法是正则化。正则化是地震波反演成像实用化过程中必须考虑的问题。

理论上，地震波反演成像只能进行线性反演或弱非线性反演（可以线性化的反演）。不失一般性，可以

对非线性的波场和参数之间的关系进行如下线性化处理，

$$\mathbf{d}=F\left(\mathbf{m}\right)\approx F\left(\mathbf{m}_0\right)+\left.\frac{\partial F}{\partial m}\right|_{\mathbf{m}_0}\delta\mathbf{m} \tag{10}$$

上式可以简写为矩阵方程

$$A\mathbf{x}=\mathbf{b} \tag{11}$$

该式用来估计在 $\mathbf{m}_0$ 点附近的 $\delta\mathbf{m}$。其中 $\mathbf{x}=\delta\mathbf{m}$，$\mathbf{b}=F\left(\mathbf{m}\right)-F\left(\mathbf{m}_0\right)$ 是数据残差。

由于数据和稀疏矩阵中都存在误差，（11）式本身的欠定或超定性质，矩阵 A 的不满秩， 一般地，我们把对（11）式的求解化为二次型泛函的求解问题，

$$\mathcal{J}\left(\mathbf{x}\right)=\left\|\mathbf{A}\mathbf{x}-\mathbf{b}\right\|^2+\lambda\left\|R\mathbf{x}\right\|^2 \tag{12}$$

其中，R 是正则化算子。从前面反演问题的概率论认识中可以看出，对于（8）式右端项中的数据误差性和模型误差项，都希望是满足高斯分布的。我们知道，数据项是比较符合高斯分布的，但模型项不容易符合高斯分布。为此，我们需要对模型进行预处理，基本的思想是对它施加粗化算子，使它比较接近符合高斯分布，Laplace 算子就是其中一种选择。

从另一个角度看，反问题求解时，参数的光滑部分（低波数部分）是容易首先求解的。那么我们可以引入一个光滑算子 S，作用于原参数上，即

$$\mathbf{y}=\mathbf{S}\mathbf{x} \tag{13}$$

$\mathbf{y}$ 是参数的光滑部分。把（13）代入（11）式得到，

$$A\mathbf{S}^{-1}\mathbf{y}=\mathbf{b} \tag{14}$$

可以看出估计 $\mathbf{y}$ 要容易得多，非唯一性也降低很多。

另一方面，不同数据分量的可靠性不同，需要引入加权系数。假设加权矩阵算子为 $\mathbf{L}$，代入（14）式得到

$$\mathbf{L}\mathbf{A}\mathbf{S}^{-1}\mathbf{y}=\mathbf{L}\mathbf{b} \tag{15}$$

（15）式实质上是引入了对于数据的正则化处理。

对比（8）式可以看出，关于模型的正则化和数据的正则化，实质上是试图弥补（12）式中丢弃了模型协方差阵和数据协方差阵的影响。对协方差阵的地球物理意义的理解可以帮助更好地选择（13）和（15）式中的正则化算子。

假如对矩阵 $\mathbf{A}$ 做 SVD 分解（$\mathbf{A}=\mathbf{U}\Sigma\mathbf{V}^T$），并代入（14）式，可以得到

$$\mathbf{L}\mathbf{U}\Sigma\mathbf{V}^T\mathbf{S}^{-1}\mathbf{y}=\mathbf{L}\mathbf{b} \tag{16}$$

其中，$\mathbf{U}$ 为左奇异矩阵，$\mathbf{V}$ 为右奇异矩阵。

为了对比，（11）式引入奇异值分解结果

$$\mathbf{U}\Sigma\mathbf{V}^T\mathbf{x}=\mathbf{b}\Rightarrow\Sigma\mathbf{V}^T\mathbf{x}=\mathbf{U}^T\mathbf{b}\Rightarrow\Sigma\tilde{\mathbf{y}}=\tilde{\mathbf{d}} \tag{17}$$

可以看出，各列正交的右奇异向量作用在数据上，把数据进行变换到另一个域，表示为 $\tilde{\mathbf{d}}$；各列正交的左奇异向量作用在参数上，把参数进行变化到另一个域，表示为 $\tilde{\mathbf{y}}$。在此情况下，数据 $\tilde{\mathbf{d}}$ 和参数 $\tilde{\mathbf{y}}$ 之间的关系与一对一的、满射的关系。参数 $\tilde{\mathbf{y}}$ 中各分量之间的关系是独立的。从（17）式可以看出，经过上述变换后，反演问题会变得异常简单。

对比（16）和（17）式，可知正则化算子$\mathbf{L}$和$\mathbf{S}^{-1}$的选择更基本的含义是对数据和参数引入不同的变换矩阵，使得数据中的特征成分和参数中的特征成分建立起一对一的、线性的关系。

这是最彻底的数据特征化表示和模型特征化表示。尽管上述思想不能用在大规模地震反演成像中，但它的启发意义是深远的。对数据、对参数进行多尺度特征分解，在此基础上进行反演是提高反演过程稳定性的重要思想和手段。

针对（12）式，梯度导引类的迭代反演方法的预条件主要是对梯度场进行的，其中包括施加（近似）Hessian 矩阵逆的方法，也有直接对梯度上进行各种预处理的方法，这里不再赘述。TV 正则化方法是十分值得关注的，很多文献（Vogel, 2002）上有具体的讨论，此处也不再赘述。

9 结论

地震波反演成像的基础是多次覆盖的数据体。目前叠前数据采集的最高技术是同时源激发进行宽频带、宽方位和高密度的叠前数据采集。利用这样的数据体对不特别复杂的地质构造进行偏移成像可以给出高精度的偏移成像结果，但是若地质情况特别复杂，即使较为可靠的偏移成像结果也难以得到。这主要不是偏移成像技术存在问题，而是估计精细偏移速度的反演成像技术存在问题。

数据体包含的基本信息是时距关系决定的波前面的到达时和波阻抗变化引起的波前面上的子波振幅变化。到达时信息是波传播的累积效应。波前面上的子波振幅变化是承载在到达时上的速度突变的反映。到达时对应速度场的背景变化，子波振幅变化对应速度场的跃变。据此，要反演的参数场一般地可以表达为背景部分和跃变部分。对应地，可以抽象出反射波在这样的参数场中的传播模式。

在数据表达模式、参数表达模式和波传播模式的基础上，可以很自信地构建出适合不同数据情况的反演成像处理流程。

反演成像的目标是得到全（宽）波数带的参数估计结果。以逆时偏移成像（RTM）技术为标志的、目的主要是刻画复杂构造界面形状的偏移成像技术已经发展到了接近成熟的阶段。地震成像技术的发展很快会转到以估计地下介质弹性参数为目标的反演成像技术上来。其中，各种层析成像技术（估计参数的背景变化）和最小二乘偏移成像技术（估计波阻抗的扰动）会成为两项核心技术。

另外，把要反演的参数场视为一个图像，它由纹理（高波数部分）和背景（低波数部分）组成，在反演过程中同时演化纹理的几何和背景的大小并达到收敛，是地震速度（也包括其他参数）反演的合理的思想。现代图像处理中的模式提取、模式识别和模式演化会对地震波反演成像产生很强的借鉴意义和促进作用。

参考文献

1. Aster, R. C., Borchers, B., and Thurber, C. H., 2005, Parameter estimation and inverse problems; Elsevier Academic Press
2. Berkhout, A. J., 2012, Combining full wavefield migration and full waveform inversion, a glance into the future of seismic imaging: Geophysics, 77, no. 2, S43–S50,
3. Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, 26, 99–108.
4. Biondi, B., andW. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, 69, 1283–1298.
5. Born, M., and Wolf, E., 1999, Principles of Optics, 7th edition, Cambridge University Press
6. Bunks, C., F.M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, 60, 1457–1473.
7. Boyd, S., and Vandenberghe, L., 2004, Convex Optimization, Cambridge University Press
8. Gaurav Dutta, Yunsong Huang, Wei Dai, Xin Wang, and G.T. Schuster, 2014, Making the most out of the least (squares migration), 82nd Annual International Meeting, SEG, Expanded Abstracts,,4405–4410.
9. Kay, S. M., 1993, Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory, Pearson Education, Inc
10. Kay, S. M., 1998, Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory, Pearson Education, Inc
11. Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.
12. Lambaré, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, 57, 1138–1154.
13. Lian Duan, Yu Zhang and Lingzi Peng,2014,Band-limited impedance perturbation inversion using cross-correlative least-squares RTM, 84th Annual International Meeting, SEG, Expanded Abstracts, 3720-3725
14. Hongbo Bi and Tao Lin 2014,Impact of adaptive data selection on full waveform inversion, 84th Annual International Meeting, SEG, Expanded Abstracts, 1094-1098

15. Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.
16. Mora, P. R., 1987, Nonlinear two-dimensional elastic inversion of multi-offset seismic data: Geophysics, 52, 1211–1228.
17. ———, 1988, Elastic wavefield inversion of reflection and transmission data: Geophysics, 53, 750–759.
18. Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, 100, 12431–12447.
19. Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics, D. Reidel publishing company
20. Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multisource cross-hole tomography, Part I: Acoustic wave-equation method: Geophysical Prospecting, 38, 287–310.
21. Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method: Geophysical Prospecting, 38, 311–330.
22. Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, 64, 888–901.
23. Sajeva, A., Bienati, N., Aleardi, M. and Stucchi, E., 2014, Estimation of velocity macro-models using stochastic full-waveform inversion, 84th Annual International Meeting, SEG, Expanded Abstracts,1227-1231
24. Samuel H. Gray，2014，Seismic imaging and inversion: What are we doing, how are we doing, and where are we going? 84th Annual International Meeting, SEG, Expanded Abstracts, 4416–4420.
25. Shin, C., and Y. H. Cha, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, 177, 1067–1079.
26. Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, 49, 1259–1266.
27. Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation; SIAM
28. Vogel, C. R., 2002, Computational Methods for Inverse Problems; SIAM.
29. Zhang, Y., A. Ratcliffe, L. Duan, and G. Roberts, 2013, Velocity and impedance inversion based on true amplitude reverse time migration: 83rd Annual International Meeting, SEG, Expanded Abstracts, 949–953.

地震反演及其在储层刻画中的应用

王华忠

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】目前地震地质解释最核心的工作方法是把成像剖面上的同相轴赋予地质层位的含义、并把同相轴的深度校正到和地质层位的深度一致；然后分析同相轴上的子波属性，根据子波属性的变化推测储层的变化。更深入的利用地震波振幅变化反演地下介质弹性参数的变化，根据弹性参数的变化，结合岩石物理来进行储层描述，也是地震地质解释中的重要方法。但是基于地震波振幅的反演，其结果的可靠性是严重缺乏评价的。本文比较全面地综述了地震波反演方法及其在储层描述中的作用。尤其简述了对基于振幅的反演的看法。希望引起业界对基于振幅的反演结果的可靠性的评价，真正把地震地质解释推进到定量或半定量的解释，提高油气预测的精度。

【关键词】地震波反演、地震波振幅、基于旅行时的反演、基于振幅的反演、反演不确定性

1 概述

地震勘探发展到现在，可以用系统的观点来看待整个油气地震勘探过程了。它包括了如下几个大的方面：地震波正演模拟；地震数据采集；近地表参数反演；数据处理（包括去噪音、去多次波和数据规则化）；偏移速度分析/Tomography/FWI/RTM；岩石物理建模；地震地质解释与储层描述和评价。若把油气地震勘探过程看成一个系统，它的输入是震源产生的地震子波；输出是接收到的叠前地震数据；系统本身是整个探区的岩性体（包括其中的储层）。

地震子波是未知的。由于海水介质简单，海上也许可以观测枪阵产生的激发响应（子波）。陆上地震子波的形态存在比较严重的空变。地震子波实际上是对震源激发响应的一种称谓，因为它是频率和时间展布都有限的。可以用 Ricker 子波之类的理论子波来近似它。

输出的地震数据除了地下介质的性质变化产生的复杂波现象外，地表引起的噪音是导致地震波成像处理复杂化的主因。总之，地下岩石介质变化引起的波现象是我们需要的输出，而地表条件变化引起的波场的变化及波现象是不期望的，因为油气地震勘探的目标在中深层。本质上，近地表也是油气地震勘探系统的一部分，而且这部分的影响不弄清或消除或弱化，中深层油气勘探的问题很难得到满意的解决。

油气勘探的终极目的是清楚地了解含油气储层的性质。事实上，含油气储层本身是一个复杂的集合体，往往尺度还比较小。另外，局部的、小尺度的含油气储层的定位、识别和评价必须建立在整个三维探区、甚至整个区域的构造、沉积关系清楚的基础上。无论是局部的、小尺度的储层或是区域的、大尺度的构造和沉积体的描述都需要某种抽象。一般地，在声波和弹性理论为基础的地震勘探中，用介质的弹性参数（譬如纵横波速度和密度等等）作为一个层次的模型来刻画波场与地下介质的声/弹性性质之间的关系、用储层参数（譬如孔隙度、渗透率、油气水饱和度等等）来描述储层性质、用岩石物理模型建立起弹性参数和储层参数之间的某种（半定量）关系。因此，从地震学的角度看，该系统应该是一套弹性参数来描述的。当然，弹性参数一定有时岩石性质参数（岩性、孔隙度、含流体性等）的函数。

认识地下岩石性质的途径，典型的方法就是利用地表（或井中）观测的地震数据进行地震波偏移成像和反演成像。实际上，简而言之的讲，就是把波场 $\boldsymbol{u}\left(\boldsymbol{x}_s,\boldsymbol{x}_r,\boldsymbol{t}\right)$ 映射成分辨率尽可能高的弹性参数场。最好是映射成分辨率尽可能高的岩石性质参数场。图 1 就是地震波反演成像把地震数据映射成地下岩石弹性参数（波阻抗）场的过程示意图。很遗憾，这样的映射不是一对一的，而是一对多的。叠前数据体并不能一对一的映射成高分辨率的地下波阻抗体。

由叠前（后）数据体映射成弹性参数场的过程广义地称为地震波反演。

高维全波形（FWI）反演理论是地震波反演的比较完整的理论，最近受到了广泛的关注。它已经逐渐在实际应用中见到效果，尽管只是在估计背景速度方面效果明显。但已经预示出由 1D FWI 反演上升到高维 FWI 并把其弹性参数反演结果在定量地震解释中进行应用的明显前景。当然，当前的定量地震解释还是依赖于 1D 反褶积模型的正反演。

地震数据映射成分辨率尽可能高的弹性参数场，其结果是高度不唯一的。原因很多：地震波正演方法过于简单，不能模拟实测的复杂的波现象；地震子波未知且空变；地震数据噪音不满足一定的先验概率分布；初始弹性参数场距真的弹性参数场过大；地震数据的频带过窄，不包含必要的低频和高频成分等等。

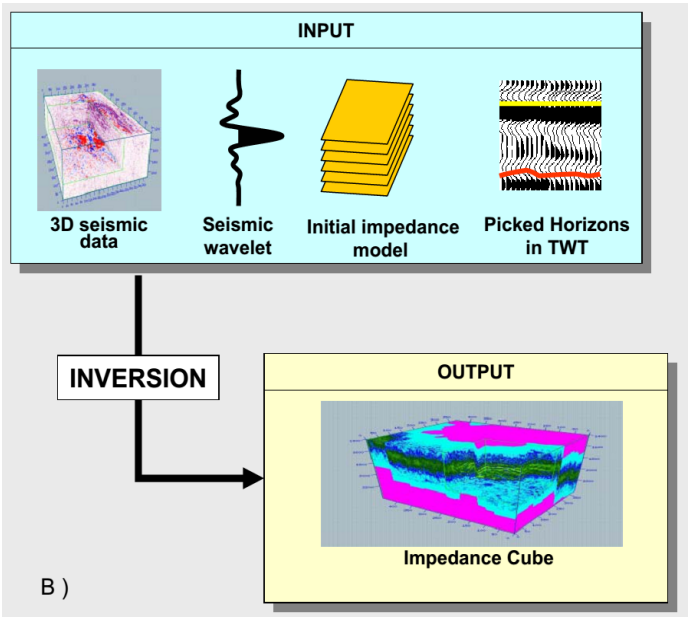


图 1 地震波反演成像把地震数据映射成地下岩石弹性参数（波阻抗）场的过程示意图

可以看出，除了反演方法要考虑更多的波现象外，要得到高精度的反演结果，先验的弹性参数（及测井数据）、合适的地震数据与处理、子波的估计是重要的因素。

综合的反演过程中的关键技术环节包括：

输入数据的质量控制和预（条件）处理；
井震匹配；数据的零相位化；子波的提取；
反演算法的执行，包括地震子波属性的提取、声/弹性阻抗的产生。
弹性参数转化成储层参数，进行储层识别与评价。

地震波反演最终落实到用反演理论、方法和技术统辖地震数据采集、数据预条件处理、反演成像（弹性参数估计）、储层建模（定量评价储层）全过程。这是现代油气勘探地震学的一个典型特征。真正抓住了问题的本质。

把当前高维 FWI 与已有的 1D 波阻抗反演和 AVA 反演有机地融合在一起是综合的地震波反演的核心任务，目标直指半定量和/或定量的储层建模。

2 地震数据的预条件

地震勘探的目的是找到含油气储层。含油气储层的识别和评价的主要依据是地震成像道集和剖面上反射子波的振幅变化，当然测井信息更为直接。但一定是以子波的振幅变化先进行预先判别才有后续的有效的钻井信息。基于振幅的半定量的地震解释已经成为地震解释的主要工作。

保振幅和保波形的地震数据成像处理是可靠的基于振幅的解释的基础。

保振幅和保波形的地震数据成像处理的关键技术环节包括：压制采集脚印的技术---地震数据规则化；可靠的振幅处理技术---球面扩散校正、陆上数据的地表一致性振幅校正、吸收衰减校正（反 Q 滤波）；保真的去噪音技术---多道反褶积、零相位全通滤波器；保真的多次波压制技术。

很明确地，一些处理技术不可以使用在保幅处理流程中。谱白化技术不要做，单道反褶积技术不要做，

小窗口的振幅均衡技术小心使用。整个处理过程中相对振幅变化的质量控制非常重要，每一步都要监控横向上振幅的变化特征有无明显的变化。

3 地震子波的提取

地震子波的作用在定量地震解释中是非常大的，因为解释员所看的主要是地震子波，而不是地下的反射系数。

子波的相位不同，形态就不同。因此为了更好地认识地下反射系数的分布，最好先把子波的形态进行统一。一般地是进行子波的零相位化，因为零相位子波是对称的，零相位子波有一个主瓣，把主瓣对应到反射系数出现的位置时，地下反射系数关系就可以通过地震剖面清楚地反映出来。

假如地震波在理想流体或弹性介质中传播，地下仅仅有一个反射层，反射子波 $\boldsymbol{w}(\boldsymbol{t})$ 的振幅谱一定是 $\boldsymbol{rw}(\boldsymbol{f})$。其中 $\boldsymbol{r}$ 为该反射层的反射系数，$\boldsymbol{w}(\boldsymbol{f})$ 为子波的振幅谱。因此，子波估计的过程首先是应该确定子波的振幅谱。利用 $\boldsymbol{logS}(\boldsymbol{f})=\boldsymbol{log}\,\boldsymbol{w}(\boldsymbol{f})+\boldsymbol{logr}(\boldsymbol{f})$，利用测井计算出 $\boldsymbol{logr}(\boldsymbol{f})$，也可以方便地计算出子波的振幅谱。其中 $\boldsymbol{S}(\boldsymbol{f})$ 是地震道的振幅谱，$\boldsymbol{r}(\boldsymbol{f})$ 是反射系数序列的振幅谱。然后，对具有该振幅谱的子波进行相位旋转，直到接近与零相位子波。

这样得到的子波还要通过井震匹配的标定，一是标定零相位子波的主瓣与已知反射系数的时移量；而是标定零相位子波的极性；三是标定零相位子波的振幅值。

很显然，井震标定存在的另一个关键问题是依据测井波阻抗计算的反射系数序列与子波褶积后得到的反射子波与实测井旁道对不起来，反射系数的双程旅行时存在问题。这需要用 Check-shot 或 VSP 观测进行调整。

从地震波反演的角度，并不需要子波的零相位化。子波的零相位化是方便反射层位解释和储层的几何形态解释以及 AVA 分析的。地震波反演理论仅仅需要子波的正确，不计较它的相位特点。

对于地震波反演而言，正确的子波本身就是要反演的目标。在进行弹性参数反演的同时，把正确的子波也反演出来。但是，这样做会大大地增加地震波反演的参数空间的规模，使得反演结果的不唯一性更强。

总之，在 FWI 反演已经正在实用化的今天，在 FWI 反演框架下如何看待子波，如何处理子波，是高维 FWI 反演的一个重要议题，也是它是否能走向储层描述的一个关键问题。

4 地震波反演成像基本方法

4.1 地震波反演基本思想

严格的地震波反演的定义是与地震波正演相对应的。已知地震子波、弹性参数场，解波动方程定解问题，模拟地震波的传播过程得到合成记录是地震波正演问题。地震波反演是已知实测的地震波场，已知波动方程及相应的定解问题，在一定的反演理论框架下，目前主要是 Bayesian 反演理论，估计地震子波和弹性参数场。也可以是假设子波已知（实际上并不知道!），估计弹性参数场。地震波正演问题是适定的、良态的；地震波反演问题是不适定的、病态的。所谓适定的问题是指问题的解是稳定的、唯一的、连续依赖于数据的。上述三条有一条不满足，该问题就是不适定的。反问题是十分典型的不适定的问题。这是反问题比正问题困难的根本原因。

广义地，可以把正反演问题放在一个系统下来看，由系统输入、系统本身和系统输出构成。勘探地震就可以视为这样一个系统，输入是地震子波；输出是地震数据体；系统是弹性参数（也可以是岩性性质参数）。正问题的研究可以认为是建立输入和输出之间的关系。在褶积系统中，我们把反射系数序列视为输入，1D 地震道视为输出，建立一个线性方程组 $\boldsymbol{Ar}=\boldsymbol{d}$，实际上它是一个线性预测器，预测算子是子波，预测结果是 1D 地震道。这说明一个问题，我们可以基于此建立一个很通用的正反演方法。输入可以是地震属性、可以是反射系数、可以是波阻抗；输出是地震数据。即认为输入和输出间存在一种线性关系。在井旁地震道的约束下，构建反问题，在输出与井旁道最佳的条件下，估计反射系数、或波阻抗或其他任何有意义的

反演结果。如图 2 所示。

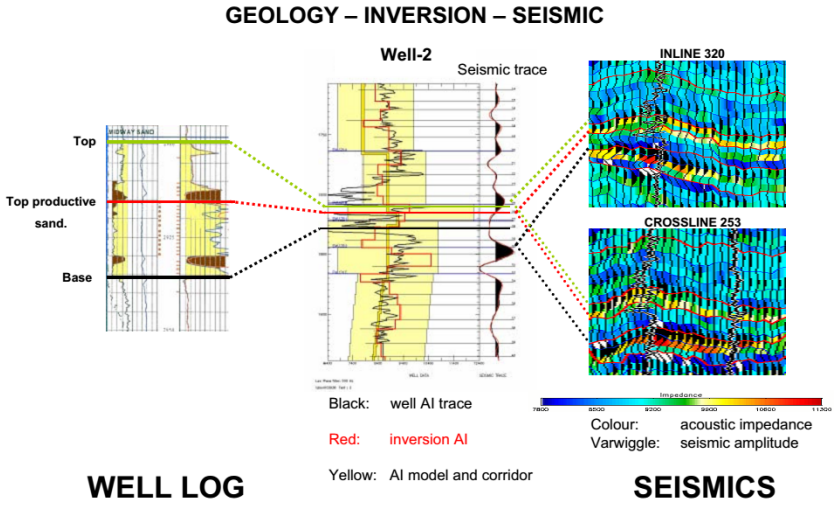


图 2 反演是把地震数据映射成类似测井记录

这实质上是 Wiener 滤波意义构建的正反演问题。

进一步的发展是不认为输入和输出间是线性关系，此时常用的反演思想是神经网络下映射，这涉及到神经网络的学习过程和应用过程。

据此可以看出，地震波反演问题可以退化成线性的 Weiner 滤波理论和反滤波理论。1D 滤波和反滤波是我们十分熟悉的。高维的滤波和反滤波是通过波动方程叠前正演和 FWI 来实现的。但是，这并没有改变问题的本质。

因此，在 FWI 技术逐步实用化的今天，定量的高维地震反演估计弹性参数，进而借助岩石物理模型进到储层参数或岩石性质参数估计是顺理成章的事。当然，以前在 1D 反演及实用化过程中积累的经验可以推广到高维的 FWI 反演中。

另外，地震波反演包括确定性的方法和随机反演方法两大类。更具体的可以分为经典的方法和高维 FWI 方法。按此分类是为了凸显高维 FWI 方法的理论框架的先进性。确定性的反演方法基本思想是估计弹性参数的一个值作为解的期望值。随机性的反演方法是估计弹性参数场的多个值进而计算解的期望和方差。

反演结果的精度主要取决于数据是否满足反演方法的要求；先验的弹性参数是否可以有效地降低反演解的非唯一性。当然，正问题是否可以预测实测数据也是很关键的影响因素。

下面罗列并简述常用的集中 1D 和高维反演方法。确定性的反演方法：叠后或单道反演方法：地震道积分、稀疏脉冲反演、有色反演、模型驱动的反演；叠前或多道反演方法：AVA 弹性阻抗反演、FWI 反演。随机反演方法：地震统计学反演。

①地震道积分

理论上，地震道积分的思想很简单，它不应归为反演方法。如果不论结果如何的话，做法也很简单。首先对地震数据做保幅处理，对子波零相位化，脉冲反褶积，提取反射系数序列。进行常规的速度分析，得到低频速度成分。根据对反射系数序列积分的思想，把速度的低频变化和高频变化融合成伪测井记录。

估计反射系数序列时，假设密度已知。反射层比较厚，容易提取反射系数序列。反射层比较平缓。

②有色反演

它不应该归结为一类反演方法。它实质上是一种滤波方法，不过滤波器的设计是利用测井数据（反射系数序列）的振幅谱与井旁道信号的振幅谱的比值来估计子波的振幅谱。如果反射系数的振幅谱假设是白

的，估计出的反射系数的振幅谱的是比较可靠的。

即 $w(f) \approx S(f)/r_{well}(f)$，其中 $r_{well(f)}$ 是来自井中的反射系数序列的振幅谱。有了子波的振幅谱，零相位假设下可以得到子波，它就是要估计的滤波器。用此滤波器对其他地震道进行反褶积可以把它们变成发射系数序列。

这就是有色反演的基本思想。图 3 展示了它的基本过程和结果。有色反演很简单、高效。但反演结果的精度很难保障。

③稀疏脉冲反演

这是理论上很完整的反演方法，它的主要问题应该是正问题的一维假设与实际地震波传播情形相差太远。

所谓的稀疏脉冲反演首先假设反射系数是稀疏的。在反演过程中，误差泛函建立时用 L1 范数或 Cauchy 范数来对应一类长尾概率分布函数，用来约束反射系数使之有稀疏特征。数据逼近还是用 L2 范数。求解该问题用的是 IRLS 方法（迭代再加权最小二乘方法）。

该方法还可以突破子波的最小相位假设，可以是任意的。还可以考虑 Q 引起的子波频散与振幅衰减。

稀疏脉冲反演的问题与高维 FWI 一样。差异仅仅是前者是单道反演。这样预示了 FWI 反演在储层建模中的发展趋势。不久的将来，FWI 不仅仅在地震波偏移成像，也要在地震波反演成像中有大的作用。

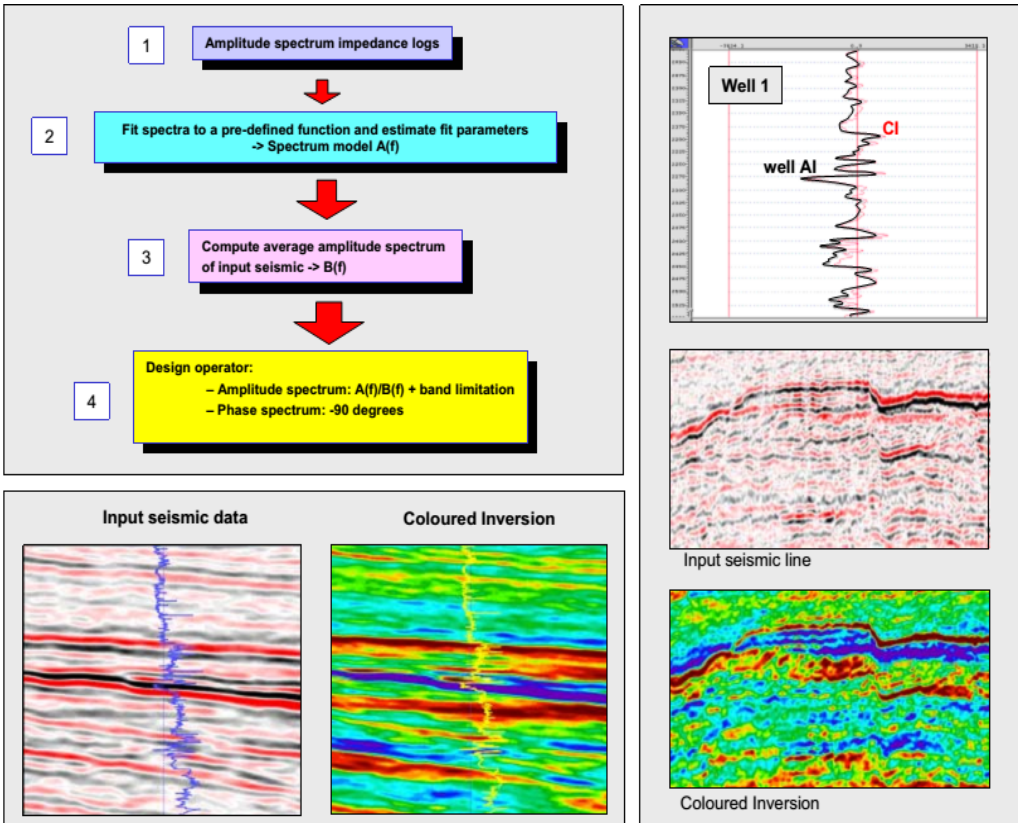


图 3 有色反演的基本过程

④基于模型的反演

基于模型的反演，基本思想与 Bayesian 框架下的反演没有本质差异，关键的问题是模型表达。基于模型的反演首先原则上是三维模型的迭代更新，数据是单道正演和单道逼近的。这两者是不那么协调的。但是，当地下界面和参数横向上变化不大时，可以接受。

基于模型的反演，不宜用梯度类的方法，最好用 Monte Carlo 类的反演。初始模型的建立用地震层位和

测井的综合，通过插值得到一个比较好的初始速度模型。反演结果就是在此模型附近的扰动，当合成数据与理论数据逼近较好时，反演结果就是所要的。

图 4 展示了模型驱动的反演的效果。模型驱动的反演有很强的约束，偏移成像的层位和测井的波阻抗，以及基于二者的插值已经建立了反演结果的基本框架，反演结果仅仅是在此基础上的变化。因此，该方法比较抗噪音。

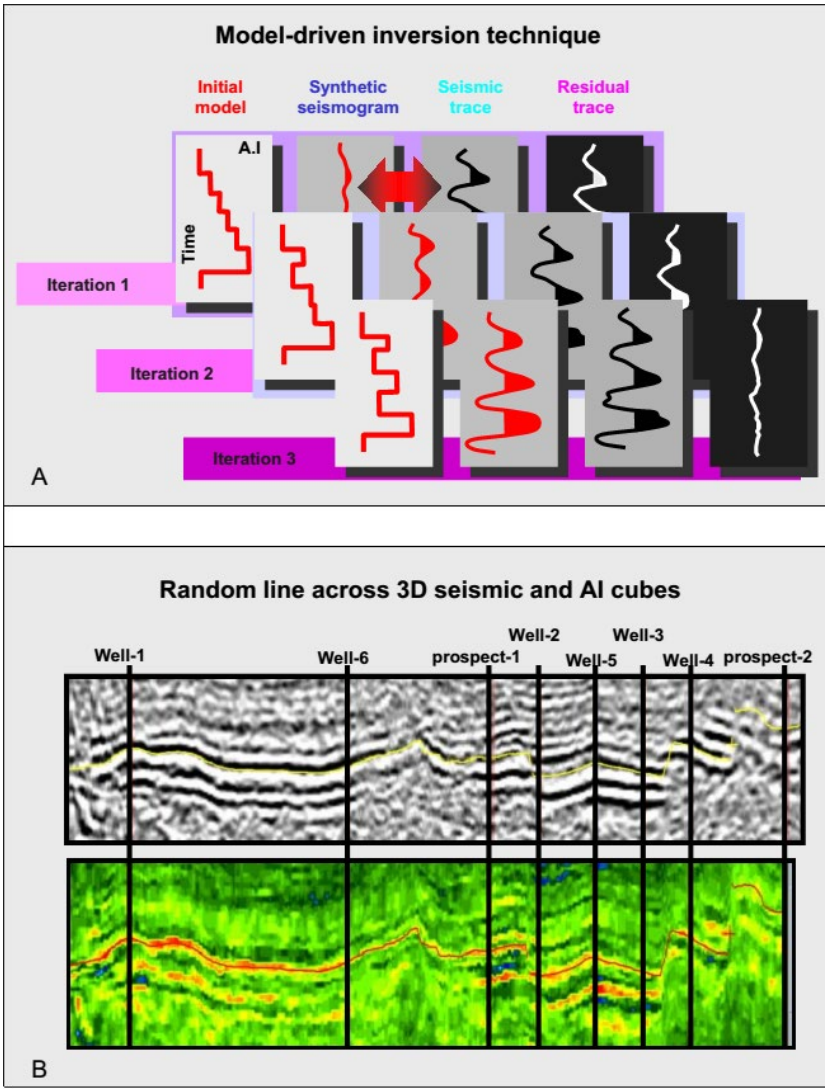


图 4 模型驱动的反演方法及反演结果

⑤叠前 AVA 弹性参数反演

前面四种方法都是基于叠后单道数据进行反演的。走向叠前反演有两种方法：一种是基于角度成像道集的 AVA 反演方法；一种是基于炮集数据的 FWI 方法。

AVA 反演是基于 Zeoppritz 方程的。该方程反映的是平面波入射到弹性参数分界面上发生的波的分裂---透射、转换和反射现象，建立了反射系数和透射系数与界面两边弹性参数的关系。

AVA 反演一般地用 Zeoppritz 方程的线性化的近似形式。它是纵波速度、横波速度和密度的函数。可以解三个角度的反射系数联立的方程组求解纵波速度、横波速度和密度扰动量。也可以用子波的反射系数褶积然后与实测 AVA 道集逼近来反演这些扰动量。但是，Zeoppritz 方程用来描述实际介质时，受子波分辨率限制，尤其是用解三个角度的反射系数联立的方程组求解纵波速度、横波速度和密度扰动量时更是如此。AVA 道集的不可靠更是很难评价。实际上，我们是在从一个很不可靠的数据出发进行反演估计弹性参数。

很难评价反演结果的可靠性。

实际中，反而是定性的 AVO 分析更有实用意义。

图 5 展示了基于 AVA 道集进行弹性波阻抗反演的基本过程。建立可靠的初始弹性参数（Vp,Vs,Rho）是非常重要的，否则既没有可靠的输入数据（AVA 道集），有没有可靠的先验约束，AVA 反演结果的可靠性无从谈起。

单个 AVA 道集的反演可能使得反演结果没有整体上的地质含义。因此，需要利用整体的 AVA 道集和整体的弹性参数模型建立起基于模型的或模型驱动的反演框架。基本思想类似于前述的基于模型的和单道数据的声阻抗反演流程。图 6 展示了这样的流程，其中对整体反演的思想体现不够。

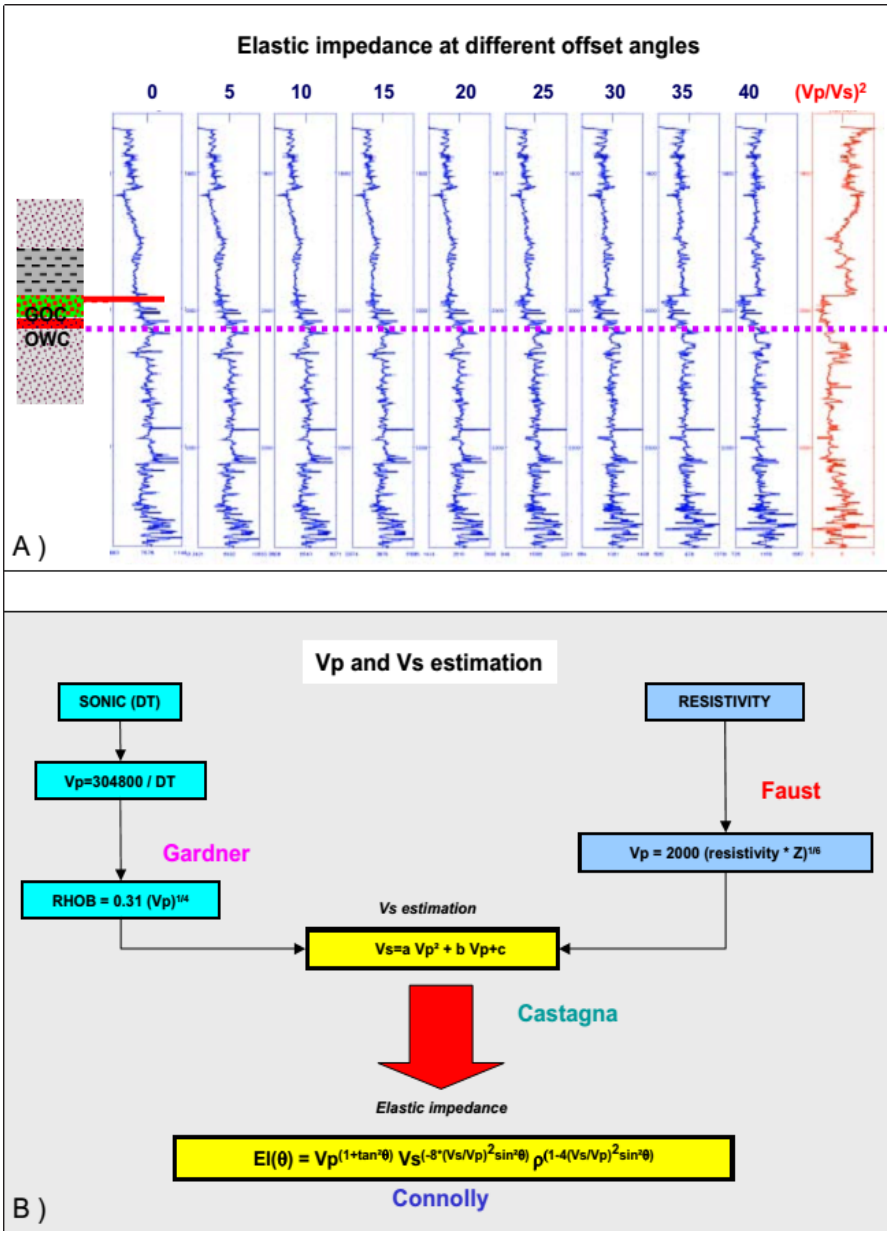


图 5 AVA 弹性波阻抗反演的基本过程

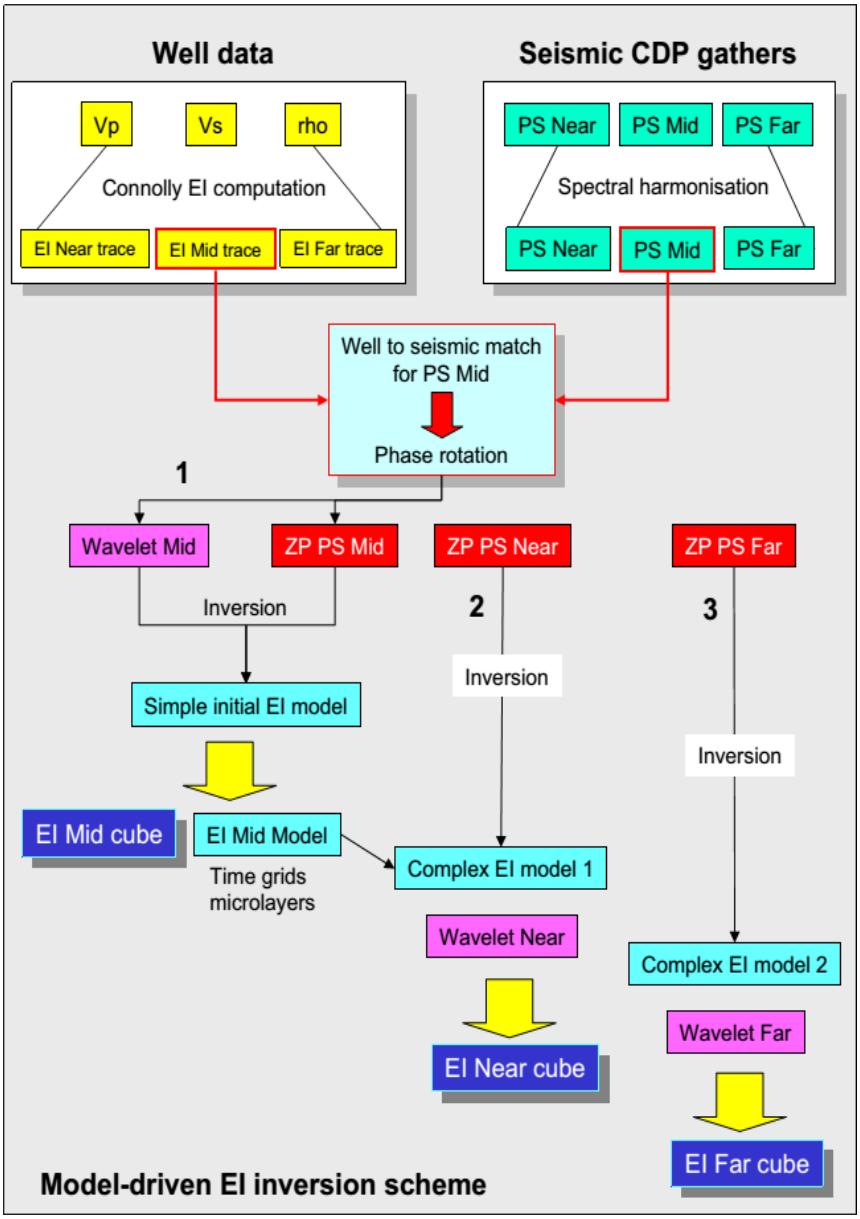


图 6 基于全体 AVA 道集和全局模型的模型驱动的弹性阻抗反演的流程

⑥高维全波形反演 FWI

全波形反演 FWI 是目前地震波反演的最完整的理论框架，可以说勘探地震学的所有反演问题都脱离不开此框架。1D 波阻抗反演、1D 反射系数估计、AVA 叠前反演等等都是高维 FWI 理论的某种简化。由于数据质量和计算机技术的进步，FWI 有逐步实用化的趋势。

4.2 全波性反演 FWI 基本理论

全波形反演的基本概念原则上要从 Bayesian 反演框架引申开来。但是假设已知数据残差的概率分布，可以根据数据残差最小的原则可以建立如下的优化问题：

$$\begin{aligned} \min: \quad & J(\mathbf{m}) \\ S.T.: \quad & C\mathbf{u} = \mathbf{d} \end{aligned} \tag{1}$$

可以用如下的方式定义上面优化问题所对应的误差泛函，

$$\min_{\mathbf{u}\in\mathcal{H}_2,\mathbf{m}\in\mathcal{H}_1}\frac{1}{2}\left\|C\mathbf{u}-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{2}$$

上式中的第一项代表正演波场与观测波场的波场残差，后一项代表对模型的先验约束，α 是正则化系数。一般地，我们可以通过假设模型中的元素满足某种概率分布条件来施加对模型的先验限制。（2）式还可以写为

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\mathfrak{I}(\mathbf{m})=\frac{1}{2}\left\|F(\mathbf{m})-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{3}$$

其中，与观测数据逼近的正演计算结果可以表达为

$$F(\mathbf{m})=C\left(A(\mathbf{m})\right)^{-1}\mathbf{f} \tag{4}$$

将上述结果代入（3）式后，得到最后的泛函定义式，

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\mathfrak{I}(\mathbf{m})=\frac{1}{2}\left\|C\left(A(\mathbf{m})\right)^{-1}\mathbf{f}-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{5}$$

通过此式，可以通过各种优化方法求解参数估计问题。假设上述定义的泛函至少是 Fréchet 二阶可微的，那么，可以对该泛函做如下的 Taylor 展开，

$$\mathfrak{I}(\mathbf{m}+\delta\mathbf{m})=\mathfrak{I}(\mathbf{m})+\left\langle \mathrm{Grad}\mathfrak{I}(\mathbf{m}),\delta\mathbf{m}\right\rangle+\frac{1}{2}\left\langle \mathrm{Hess}\mathfrak{I}(\mathbf{m})\delta\mathbf{m},\delta\mathbf{m}\right\rangle+o\left(\left\|\delta\mathbf{m}\right\|^3\right) \tag{6}$$

如果泛函 $\mathfrak{I}(\mathbf{m}):\mathbb{R}^n\to\mathbb{R}$ 有连续的二阶偏导数，那么 $\mathrm{Grad}\mathfrak{I}(\mathbf{m})$ 是一个 n 维向量，表示泛函的梯度方向，$\mathrm{Hess}\mathfrak{I}(\mathbf{m})$ 是一个 $n\times n$ 的方阵，称为 Hessian 矩阵。其中，n 为模型参数的个数，m 为观测数据的个数。

定理：如果泛函 $\mathfrak{I}(\mathbf{m}):\mathbb{R}^n\to\mathbb{R}$ 在 $\mathbf{m}_0$ 点的某个邻域 $\Omega(\mathbf{m}_0,\delta)$ 内同时满足以下几个条件：1. $\mathfrak{I}(\mathbf{m})$ 在该邻域内是 Fréchet 二阶可微的；2. 梯度 $\mathrm{Grad}\mathfrak{I}(\mathbf{m}_0)=0$；3. Hessian 矩阵 $\mathrm{Hess}\mathfrak{I}(\mathbf{m}_0)$ 是强正定的，那么 $\mathbf{m}_0$ 是泛函 $\mathfrak{I}(\mathbf{m}):\mathbb{R}^n\to\mathbb{R}$ 的严格极小点。

因此，（1）式所定义的优化问题的求解可以由如下迭代格式执行，

$$\mathbf{m}^{k+1}=\mathbf{m}^k+\delta\mathbf{m}^k=\mathbf{m}^k+\mu_k\mathrm{Grad}\mathfrak{I}(\mathbf{m}) \tag{7a}$$

该式为最速下降算法；

$$\mathbf{m}^{k+1}=\mathbf{m}^k+\delta\mathbf{m}^k=\mathbf{m}^k+\left(\mathrm{Hess}\mathfrak{I}(\mathbf{m})\right)^{-1}\mathrm{Grad}\mathfrak{I}(\mathbf{m}) \tag{7b}$$

该式为 Gauss-Newton 算法。

目前，FWI 计算的基本框架就是如此。

4.3 频率域全波形反演数值试验

分别用一维线性递增的速度模型（从 1500m/s 到 4000m/s）和对真实速度模型平滑后的速度模型作为初始模型。图 7a，b 是初始模型采用线性递增速度模型的反演结果。其中，图 7a 采用了 8 个离散的频率进行串联反演（f=3,4,6,8,10,12,15,18Hz），图 7b 采用了 30 频率进行串联反演（f=2~25Hz）。可以看出，随着采用的频率分量的增多，图 7b 在浅层的反演结果与真实模型很接近，但深层速度的反演结果与真实模型相差较大。这是由于初始速度模型没有速度场的宏观背景部分，导致反演结果在深层没有收敛到真实模型。图 8a，b 是初始模型采用平滑过的真实速度模型的反演结果。其中，图 8a 采用了 17 个离散的频率进行串联反演（f=1~30Hz），图 8b 采用了 35 频率进行串联反演（f=1~30Hz）。对比图 7b 与图 8a 可以看出，由于初始模型包含了速度场的低波数部分，即使采用了较少的频率分量用来反演（图 8a），得到的结果仍然好于图 7b 中采反演结果。图 8b 采用了更多的频率域数据串联反演，得到了比图 8a 中更为精细的结果。图 8b 的反演结果证明了若初始模型较好，FWI 可以给出一个精细的速度场。

从数值结果至少可以看出，在数据条件比较理想时，FWI 的反演结果可以对地震定量解释有很大的贡献。尤其是子波比较已知（譬如海上数据）、信噪比高、初始模型比较准、地震数据频带较宽并可以被利用起来时，FWI 应该成为重要的面向储层定量解释的反演工具。

实际情况中的问题比较复杂，FWI 距离用于储层描述的应用还有较大差距。但是，把在 1D 反演中积累

的经验用于 FWI 中，会逐渐地把 FWI 推向储层描述中，这是毋庸置疑的。

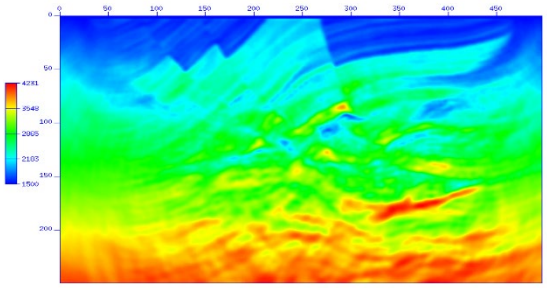


图 7a 用 8 个频率串联反演的速度场，f=3,4,6,8,10,12,15,18Hz

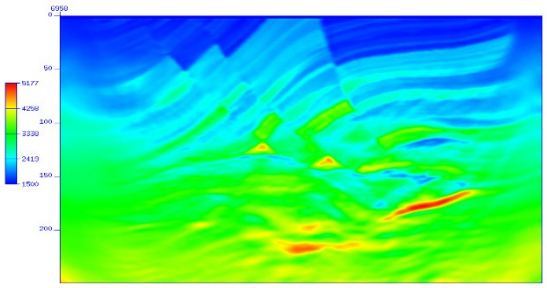


图 7b 用 30 个频率串联反演的速度场，f=2~25Hz

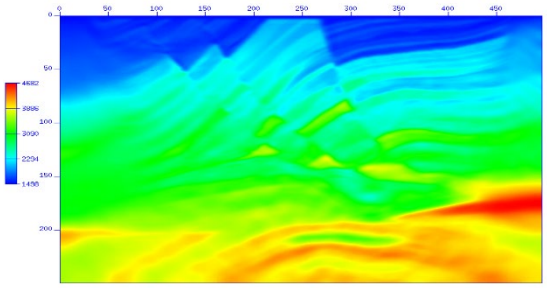


图 8a 用 17 个频率串联反演的速度场，f=1~30Hz

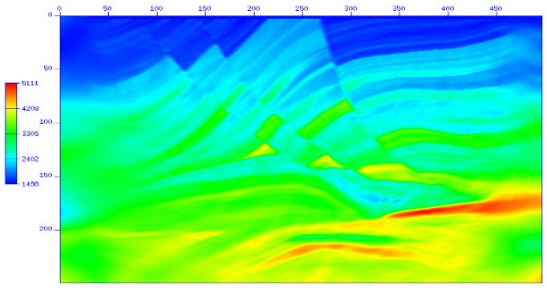


图 8b 用 35 个频率串联反演的速度场，f=1~30Hz

5 随机反演与地质统计学反演

地震波反演的目的也可以表述成地质建模，通过综合的地震反演建立弹性参数表示的地质模型或岩石性质参数表示的地质模型。

因此，对于反演而言，模型表达实质上已经蕴含了对地质模型的采样。反演也可以讲是在合成数据与实测数据逼近误差最小情况（也包括先验信息约束情况）下在所有可能的地质模型空间中的对最佳地质模型的搜索。

最佳的含义是基于此模型正演得到的合成数据与实测数据最逼近，也包括此模型与测井获取的地质模型最接近。

因此，除了上述确定性反演方法外，还有随机反演方法和地质统计学反演方法。一般地，确定性的反演按照误差泛函的梯度下降的方向收索下一个地质模型。这样的方法容易陷入局部极值点，得不到最佳的地质模型。

存在一大类随机反演方法，从一族初值地质模型出发，依据物理的（模拟退火）、生物进化的（遗传算法）、社会行为的（粒子群算法、蚂蚁算法等）模拟算法，进行下一族地质模型的搜索。它可以得到全局最佳的地质模型。还可以给出反演解的方差的估计。但是，计算效率要低于梯度类的搜索方法。

事实上，真正的地下地质模型是永远也不可能得到的。基于数据逼近和测井先验信息估计用弹性参数表示的地质模型或岩石性质参数表示的地质模型是我们解决认识储层问题的一个模式。应该还存在其他的模式。

由于采集的地震数据和测井数据的随机性，反演结果是不确定的。应该用随机反演的观点来进行地震数据反演及指导反演结果的应用。

地质建模显然是一个空间信息应用问题。如果有分布相对均匀的测井信息，对这些三维空间信息进行散乱数据的插值，可以得到一个地质建模的结果。这也是一种地质建模的方法，当然，它最好与确定性的或随机的反演方法结合在一起，前者给出的地质建模结果是后者的初始值和/或约束。地质统计学方法是实

现上述三维空间中散乱数据（这些数据是空间上有关联的!）插值的，（差不多）最有效的方法。因此，地质统计学方法用于反演和地质建模过程中是理所当然的。

地质统计学的基本方法是泛 Kirging 插值和协 Kirging 插值方法，它可以把空间不均匀分布的测井记录插值成一个有意义的地质模型。

6 反演在储层建模中的应用

储层建模是油气勘探的主要目的，油气开发就是在建立的储层模型上进行的。储层模型包括储层的几何结构和几何结构内的岩石性质参数（包括岩性、骨架组分、空隙度、流体成分、渗透率等等）。

最理想的储层建模的信息是假设有空间上非常密集的钻井和测井信息，经过简单的校正和插值可以得到分辨率极高的、可靠性极高的储层模型。但这样的代价是无法承受的。

根据地面地震数据进行地震波反演得到弹性参数，据此利用岩石物理模型建立储层模型是另一条比较可行的、经济上代价小的路线。但是，地震反演的多解性、低分辨率导致建立的储层模型无法满足含油气储层评价与开发的需求。

比较合理的路线应该是在宽方位、高密度、宽频带地震数据采集的基础上，利用叠前数据进行高精度的偏移速度反演和精确的、高分辨率的叠前深度成像解决大的反射层结构、大的断层的结构问题等等；对成像结果进行图像处理（相干分析）解决沉积相、小尺度储层的几何结构问题。描绘出细致的地质结构（地震相变）信息。

利用保真的叠前成像道集中的小偏移距数据体进行 1D 波阻抗反演；利用保真的叠前成像道集中的中、大偏移距数据体进行 AVA 叠前弹性波阻抗反演。已经建立的细致的地质结构信息作为反演的约束，建立三维的储层模型。这是地震尺度上的储层建模。

在细致的地质结构信息约束下，对井数据进行地质统计学建模，得到测井尺度下的储层模型。把地震尺度上的储层模型和测井尺度下的储层模型有机地融合在一起，得到更高分辨率的储层模型。这是反演在储层建模中的总体做法。

在上述过程中，其中的地震波反演方法包括，基于旅行时的层析反演、基于振幅的反射系数（最小二乘偏移）反演/1D 波阻抗反演/AVA 弹性参数反演，当然理论上也可以用 FWI 把两类反演方法统一在一起。但是，基于旅行时的层析反演和基于背景速度的叠前深度偏移成像是误差基本可控的、精度基本可以评价的反演方法。然而，基于地震波振幅的反演到目前为止还严重缺乏反演结果的精度评价。事实上，实际测量的地震波振幅的真实物理意义都值得仔细考虑，所用的波动方程在多大程度上能数值模拟实测地震波的振幅，震源子波的未知和空变引起多大的振幅计算误差，波场和反演参数之间的非线性性能否保证目前的基于凸或拟凸优化的反演方法收敛等等问题都使得目前的基于振幅的反演结果充其量不过是一种提取的子波属性，也不一定比直接在保真子波上进行属性提取和储层预测的精度高。这是作者的个人观点。今后基于 FWI 反演理论的弹性参数估计及基于此的储层预测能否做得更好，主要还要看地震波观测技术的进步，FWI 反演理论本身是没有问题的。

7 结论与讨论

油气地震勘探发展到现在，已经到了用 Bayesian 反演理论在系统思想的指导下，把“两宽一高”的数据采集、保真的成像处理、弹性参数反演、岩石物理建模、储层建模统一在一个思想框架下进行统一考虑的时间了。

常规的反演与储层描述特别关注井震匹配、关注子波估计，试图得到可靠的波阻抗等参数，建立储层模型。

但是，常规反演基于 1D 介质假设、基于声波假设、基于 Zeoppritz 方程，不能适应复杂的地质情况。

定量的地震地质解释需要在 Bayesian 估计框架下用高维全波形反演来估计弹性参数。弹性参数的高分辨率细节，要由精细的叠前偏移和偏移后的成像处理给出几何结构，并结合利用测井数据的地质统计学地质建模给出其中的参数，二者有机融合后给出。然后利用弹性波 FWI 调整给出的初始的高分辨率的弹性参

数场。在此过程中，要充分借鉴常规反演方法积累的井震匹配、子波估计的思想和方法。

这是定量化地震地质解释的思想，扬弃了目前比较定性的地震地质解释方法。逻辑上，这样的方法会大大地提高储层建模的精度，促进油气藏的有效开发。

参考文献

1. Ajlani, G., Al Kaabi, M. and Suwainna, O. [2003] Comparative analysis (CASP): a proposal for quantifying seismic data processing. The Leading Edge, 22, 1, 46-48.
2. Banchs, R. E. and Michelena, R.J. [2002] From 3D seismic attributes to pseudo-well-log volumes using neural networks: practical considerations. The Leading Edge, 21, 10, 996-1001.
3. Benabentos, M., Mallick, S., Sigismondi, M. and Soldo, J. [2002] Seismic reservoir description using hybrid seismic inversion: a 3D case study from the Maria Ines Oeste Field, Argentina. The Leading Edge, 21, 10, 1002-1008.
4. Buland, A. and Omre, H. [2003] Bayesian linearized AVO inversion. Geophysics, 68, 1, 185–198.
5. Cambois, G. [2000] AVO inversion and elastic impedance. Expanded abstracts, 70th SEG Annual Meeting, Calgary, 1-4.
6. Castagna, J.P. and Backus, M.M. [1993] Offset dependent reflectivity – Theory and practice of AVO analysis. SEG, Tulsa, Investigations in Geophysics, 8, 348.
7. Connolly, P. [1999] Elastic impedance, The Leading Edge, 18, 4, 438-452.
8. Cook, D.A. and Sneider, W.A. [1983] Generalized linear inversion of reflection seismic data. Geophysics,48, 665-676.
9. Coulon, J.P., Duboz, P. and Lafet, Y. [2000] Moving from seismic to layered impedance cube and porosity prediction in the Natih Emember. GeoArabia, 5, 1, 72-73.
10. Dix, C.H. [1955] Seismic velocities from surface measurements. Geophysics, 20, 68-86.
11. Duboz, P., Lafet, Y. and Mougenot, D. [1998] Moving to layered impedance cube: advantages of 3D stratigraphic inversion. First Break, 17, 9, 311-318.
12. Dubrule, O. [2003] Geostatistics for seismic data integration in earth models. EAGE Distinguished Instructor Series 6, 282.
13. Dutta, N.C. [2002] Geopressure prediction using seismic data: Current status and the road ahead. Geophysics, 67, 6, 2012-2041.
14. Fabre, N., Gluck, S., Guillaume, P. and Lafet, Y. [1989] Robust multichannel strati-graphic inversion of stacked seismic traces, 59th Annual Meeting SEG, 943.
15. Fatti, J.L., Smith, G.C., Vail, P.J., Strauss, P.J. and Levitt, P.R. [1994] Detection of gas in sandstone reservoirs using AVO analysis.: A 3-D seismic case history using the Geostack technique. Geophysics, 59, 1362-1376.
16. Faust, L.Y. [1951] Seismic velocity as a function of depth and geologic time. Geophysics, 16, 192-206.
17. Faust, L. Y. [1953] A velocity function including lithologic variations. Geophysics, 18, 271-287.
18. Gardner, G.H.F., Gardner, L.W. and Gregory, A.R. [1974] Formation velocity and density - The diagnostic basics for stratigraphic traps, Geophysics, 39, 770-780.
19. Gluck, S., Juve, E. and Lafet, Y. [1997] High resolution impedance layering through 3D stratigraphic inversion of post stack seismic data. The Leading Edge, 16, 1309- 1315.
20. Gluck, S., Deschizeaux, B., Mignot, A. Pinson and Huguet, F. [2000] Time-lapse impedance inversion of post-stack seismic data. Expanded abstracts, 70th Annual Meeting SEG, Calgary, 1509-1512.
21. Goffe, W.L., Ferrier, G.D. and Rodgers, J. [1994] Global optimisation of statistical functions with simulated annealing. Journal of Econometrics,60, 65-100.
22. Haas, A. and Dubrule, O. [1994] Geostatistical inversion – a sequential method of stochastic reservoir modelling constrained by seismic data. First Break, 12, 11, 561-569.
23. Helland-Hansen, D., Magnus, I., Edvardsen, A., and Hansen, E. [1997] Seismic inversion for reservoir characterisation and new well planning in the Snorre field. The Leading Edge, 16, 3, 269-273.
24. Lancaster, S. and Whitcombe, D. [2000] Fast track "coloured" inversion. Expanded abstracts, 70th SEG Annual Meeting, Calgary, 1572-1575.
25. Latimer, R.B., Davison, R. and Van Riel, P. [2000] Interpreter's guide to understanding and working with seismic derived acoustic impedance data. The Leading Edge, 19, 3, 242-256.
26. Lindseth, R. [1979] Synthetic sonic logs - a process for stratigraphic interpretation. Geophysics, 44, 3-26.
27. Ma, X. [2002] Simultaneous inversion of prestack seismic data for rock properties using simulated annealing. Geophysics, 67, 1877-1885.
28. Mallick, S. [1995] Model-based inversion of amplitude variations with offset data using a genetic algorithm. Geophysiscs, 60, 939-954.
29. Mallick, S., Lauve, J. and Ahmad, R. [2000] Hybrid seismic inversion: A reconnaissance tool for deep water exploration. The Leading Edge, 19, 38-43.
30. Pendrel, J., Stewart, R.R. and Van Riel, P. [1998] Interpreting sand channels from 3C-3D seismic inversion. Expanded Abstracts, 71th SEG Annual Meeting, 1588-1591.
31. Pica, A., Diet, J. and Tarantola, A. [1990] Non linear inversion of seismic reflection data in a laterally invariant medium. Geophysics, 55, 284-292.
32. Ronghe, S. and Surarat, K. [2002] Acoustic impedance interpretation for sand distribution adjacent to a rift boundary fault, Suphan basin, Thailand. AAPG Bulletin, 86, 10, 1753-1771.
33. Rowbotham, P., Marion, D., Eden, R., Williamson, P., Lamy, P. and Swaby, P. [2003] The implications of anisotropy for seismic impedance inversion. First Break, 21, 53-57.
34. Russell, B.H., Hedlin, K., Hilterman, F.J. and Lines, L.R. [2003] Fluid property discrimination with AVO: A Biot-Gassmann perspective. Geophysics, 68, 1, 29-39.
35. Sheriff, R.E. [2002] Encyclopedic dictionary of exploration geophysics.SEG, Tulsa, 429.

36. Shuey, R.T. [1985] A simplification of the Zoeppritz equations. Geophysics,50, 609-614.
37. Stewart, R.R., Gaiser, J.E., Brown, R. and Lawton, D.C. [2003] Converted-wave seismic exploration: applications. Geophysics, 68, 1, 40-57.
38. Sylta, O. and Krokstad, W. [2003] Estimation of oil and gas column heights in prospects using probabilistic basin modelling methods. Petroleum Geoscience, 9, No. 3, 243-254.
39. Tarantola, A. [1984] Inversion of seismic reflection data in the acoustic approximation. Geophysics, 49, 1259–1266.
40. Tarantola, A. [1986] A strategy for non-linear elastic inversion of seismic reflections data. Geophysics, 51, 1893-1903.
41. Torres-Verdin, C., Victoria, M., Merletti, G. and Pendrel, J. [1999] Trace-based and geostatistical inversion of 3-D seismic data for thin sand delineation: An application in San Jorge Basin, Argentina. The Leading Edge, 18, 9 , 1070-1077.
42. Trad, D., Ulrych, T. and Sacchi, M. [2003] Latest views of the sparse Radon Transform. Geophysics,68, 1, 386-399.
43. Van der Laan, J. and Pendrel, J. [2001] Geostatistical simulation of porosity and risk in a Swan Hills reef. Expanded abstracts, 71th SEG Annual Meeting, 1588-1591.
44. Van Riel, P., [2000] The past, present and future of reservoir characterization. The Leading Edge,19, 8, 878-881.
45. Vazquez, R., Mendoza, A., Lopez, A., Linares, M. and Bernal, H. [1997] 3-D seismic role in the integral study of the ArcabuzCulebra field, Mexico. The Leading Edge, 16, 12, 1763-1766.
46. Veeken, P.C.H., Rauch, M., Gallardo, R., Guzman, E. and Vila, R. Villasenor [2002] Seismic inversion of the Fortuna National 3D survey, Tabasco, Mexico. First Break, 20, 5, 287-294.
47. Velzeboer, C.J. [1981] The theoretical seismic reflection response of sedimentary sequences. Geophysics, 46, 843-853.
48. Walden, A.T. and Hosken, J.W.J. [1985] An investigation of the spectral properties of primary reflection coefficients. Geophysical Prospecting, 33, 400-435.
49. Weber, K.J. and Van Geuns, L.C. [1990] Framework for constructing clastic reservoir simulation models. Journal of Petroleum Technology, 42, 10, 1248-1297.
50. White, R. and Simm, R. [2003] Tutorial: Good practice in well ties. First Break, 21, 10, 75-83.
51. Yilmaz, O. [1987] Seismic data processing, Society of Exploration Geophysicists, Investigations in geophysics, 2,Tulsa, SEG, 526.
52. Yilmaz, O. [2001] Seismic data analysis, Volume 1 and 2. Society of Exploration Geophysicists, Investigations in geophysics, 10, Tulsa, SEG, 2027.

勘探地震学中可分解的速度反演

王华忠 冯波 王培尧

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】速度估计是勘探地震学的根本问题。到目前为止，地震波全波形反演（FWI）是速度估计的比较完整的理论框架。但是，基于振幅的速度估计方法在实际应用中由于反演过程的收敛性难于保证，并不能进入石油工业界的生产实践过程中。我们提出一套可分解的速度反演思想与方法，它包含两个层次的含义：第一个层次是基于多级 Born 近似，把 FWI 分解为透射波场的背景速度反演、基于一次反射（绕射）波场的最小二乘叠前深度偏移成像和基于二阶 Born 近似散射场的背景速度扰动量的反演；第二个层次的含义是基于一次反射（绕射）波的线性关系式构建可分的反演算法，同时反演速度在界面的扰动（包括其几何形态）和背景速度（扰动）。并以此为参照提出勘探地震学对计算数学的需求，指出勘探地震反演估计速度是一个系统性的工作，涉及到野外地震数据采集、数据的预处理、成像剖面上的图像处理、地震波传播的数值模拟、反演方法和反演解的评价。解决好此问题需要大量的现代信号分析、图像处理、统计判决和人工智能等方面知识的协助。

【关键词】全波形反演，非线性与线性的梯度计算，波场分解，可分解的速度反演

1、引言

21 世纪的勘探地震学正进入一个以现代信号分析为核心理论框架的发展时代，促使进入该时代的首要因素为野外广角度、高密度、宽频带的地震数据采集，然后是高性能的计算平台的出现（譬如每秒千万亿次、甚至万万亿次计算能力的计算机集群）。

把三维探区视为一个要研究的系统，在地表（或空中、或井中）接收到的信号和探区弹性参数构成了信号与系统的关系。研究它们之间的关系是现代信号分析学科的根本任务。现代信号分析本质上包含两部分研究内容：信号的表达与变换、系统参数估计。二者是相辅相成的。既期望得到信号和系统参数最稀疏的特征表达，也期望参数估计在稀疏表达的空间中进行。对应地，波现象预测方法也要表示在这样的空间中，还要有对应的波现象预测的数值计算方法，即波动方程最好在数据和参数的特征表达域中求解。

与波动方程预测地震波场一样，信号处理过程中的建模也是非常基础的工作，当前基追踪、匹配追踪、冗余字典等方法表达信号是比较有代表性的信号表达的思想，但对于地震信号处理而言，发展符合地震信号特点的表达方法对于信噪分离、剖面图像处理等是十分有意义的。

在信号和波场预测表达的基础上，估计理论框架下要解决的问题就是假设地震数据是符合某种分布的随机信号（就是说随机噪音的分布是可以用某种统计模型描述的），然后建立合适的误差泛函，并假设它是在局部上具有一定凸性的，接着求解一个局部最优问题，得到介质参数的某种估计。这其中，介质参数的几何结构是非常重要的信息。现代图像处理逐步地在勘探地震学中得到重要应用，从图像中自动地提取几何结构信息是关键点。这也是地震数据成像处理逐步走向（半）自动化的核心技术所在。图像的结构张量分析方法、Level-set 方法、TV 方法、高维小波变换方法就是代表性的自动图像信息提取技术。

一般情况下，介质参数估计归结为梯度导引类的迭代反演问题。用误差反传播或反投影来修正上一迭代步的参数模型，其中步长的选择是很关键的。误差反传播或反投影原则上可以认为是“RTM 过程”。因此，参数估计问题（FWI）最核心的运算就是 RTM。再退一步，RTM 的核心就是波场模拟或预测。因此，波现象以及它的数值预测方法的研究才是整个勘探地震学的理论核心。

但是，由于地震介质的复杂性导致的波场的不可预测性给勘探地震学的信号处理带了挑战，我们的观点是提出特征波场，增加波场的可预测性，把波形反演问题由一次特征波场，推到多次特征波场，逐步进展到全波场，把 FWI 逐步导引到实用化。其中，特征波场的预测方法依然是问题的核心之一。

另一方面，基于 Bayes 估计的 Monte Carlo 反演方法是在所有可能的参数模型空间中在一定的梯度导引下搜索寻找全局最优解的方法。在目前的超级计算机系统计算能力越来越强的情况下，Monte Carlo 反演方法在宏观背景速度估计方面会起到非常重要的作用。

无论何种反演方法，正则化或预条件都是非常重要的研究内容。现代图像处理在反演成像中的应用也体现于此。

最后，对反演解的评价，讨论它的无偏性和方差展布情况是反问题求解的不可分割的一部分。总之，勘探地震学中的地震数据采集、成像处理和参数反演逐步统一在了现代信号分析的理论框架下。在此总框架下，数学、物理方法相结合，最终解决参数估计问题（纵横波速度、密度；各向异性参数和吸收衰减参数）和储层描述问题是勘探地震学中最具代表性的技术路线。

鉴于勘探地震学中数据的不完整、包含强噪音、介质复杂性导致的波场的不可预测性等因素，我们提出可分解的速度反演方法，它包含两个层次的含义：第一个层次是基于多级 Born 近似，把 FWI 分解为透射波场的背景速度反演、基于一次反射（绕射）波场的最小二乘叠前深度偏移成像和基于二阶 Born 近似散射场的背景速度扰动量的反演；第二个层次的含义是基于一次反射（绕射）波的线性关系式构建可分的反演算法，同时反演速度在界面的扰动（包括其几何形态）和背景速度（扰动）。并以可分解的速度反演问题为参照，列举了五个方面的勘探地震学对计算数学的典型需求。

2、可分的反演问题

一般地,我们假设波在地下介质中的传播可以由如下的声波方程控制，

$$\left[\frac{1}{v(\mathbf{x})^2}\frac{\partial^2}{\partial t^2}-\rho(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla\right)\right]u(\mathbf{x},t)=S(\mathbf{x},t) \tag{2.1}$$

其中，$u(\mathbf{x},t)$为声场势，它是时间$t\in R$和空间坐标$\mathbf{x}\in\Omega\subseteq R$的函数。$v(\mathbf{x})$代表速度场；$\rho(\mathbf{x})$代表密度场；$S(\mathbf{x},t)$代表体积场源。忽略密度变化，（2.1）式在频率空间域标量波方程形式为

$$\mathcal{L}_s(\mathbf{m}(\mathbf{x});\mathbf{x},\omega)u(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{2.2a}$$

其中,

$$\mathcal{L}_s(\mathbf{m}(\mathbf{x});\mathbf{x},\omega)=\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})} \tag{2.2b}$$

为标量波传播算子。$\mathcal{L}_s(\mathbf{m}(\mathbf{x});\mathbf{x},\omega)$代表$\mathcal{L}_{\text{scalar}}(\mathbf{m}(\mathbf{x});\mathbf{x},\omega)$，$\mathbf{m}(\mathbf{x})$代表速度场。$\mathcal{L}_s(\mathbf{m}(\mathbf{x});\mathbf{x},\omega)$可以简记为$\mathcal{L}_s(\mathbf{m})$。

野外观测到的地震数据可以表述为

$$\mathbf{d}=C\mathbf{u}+\boldsymbol{\eta} \tag{2.3}$$

此处，C表示地震数据的采样算子，由观测系统决定。它使得连续的波场$u(x,y,z,t)$变得规则离散化（或非规则离散化）及孔径有限化。采样算子的状态受实际条件的制约。η表示噪音函数，在大部分情况下假设它是高斯白噪的。

连续波场受采样算子制约后，采样后的数据可以记为

$$[C\mathbf{u}]\approx\mathbf{u}(\mathbf{x}_i),\quad i=1,2,\cdots,n \tag{2.4}$$

其中$\mathbf{x}_i$表示第i个采样点的坐标。根据方程（2.1），可以将连续波场的正演计算形式地表达为

$$\mathbf{u}=(\mathcal{L}(m))^{-1}\mathrm{S} \tag{2.5}$$

采样后的离散地震数据表示为，

$$F(\mathbf{m})=C(\mathcal{L}(\mathbf{m}))^{-1}\mathrm{S} \tag{2.6}$$

基于前面对正问题的定义，根据数据残差最小的原则可以建立如下的优化问题：

$$\begin{aligned}&\min:\quad J(\mathbf{m})\\&S.T.:\quad C\mathbf{u}=\mathbf{d}\end{aligned} \tag{2.7}$$

可以用如下的方式定义上面优化问题所对应的误差泛函，

$$\min_{\mathbf{u}\in\mathcal{H}_2,\mathbf{m}\in\mathcal{H}_1}\frac{1}{2}\|C\mathbf{u}-\mathbf{d}\|_Y^2+\alpha J(\mathbf{m}) \tag{2.8}$$

上式中的第一项代表正演波场与观测波场的波场残差，后一项代表对模型的先验约束，α是正则化系数。一般地，我们可以通过假设模型中的元素满足某种概率分布条件来施加对模型的先验限制。

将（2.5）代入上式，得到泛函定义式，

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2}\Im(\mathbf{m})=\frac{1}{2}\left\|C\left(\mathcal{L}(\mathbf{m})\right)^{-1}\mathrm{S}-\mathbf{d}\right\|_Y^2+\alpha J(\mathbf{m}) \tag{2.9}$$

通过此式，可以用各种优化方法求解参数估计问题。在无约束非线性优化理论中，（2.9）式的求解有很多不同的方法，譬如最速下降方法、全 Newton 方法、Gauss-Newton 方法、拟 Newton 方法等等。一般地，（2.9）式对应的的迭代求解格式可以写为

$$\mathbf{m}^{k+1}=\mathbf{m}^k-\mu_k\mathbf{g}_k \tag{2.10}$$

其中，μ_k为第k个迭代步的步长，$\mathbf{g}_k$为第k个迭代步的梯度。在不同的迭代方法中，$\mathbf{g}_k$也可以认为是预条件修正后的梯度方向。步长的选择对迭代法的收敛性和收敛效率都影响甚大，具体地参考 Dennis 和 Schnabel（1996）以及 Nocedal 和 Wright（1999）。（2.10）式中最核心的是梯度向量的计算，一般地用共轭状态法计算其中的梯度$\mathbf{g}_k$。

但是，上述过程定义的 FWI 在勘探地震的实际过程中是很难保障收敛的。存在的问题包括：数据往往缺乏低频和长偏移距；地震子波未知，陆上地震子波空变严重；地震波正演不能模拟实际的波现象；影响地震波振幅的因素复杂，远非速度变化一项；速度横向变化剧烈，误差泛函的非线性性很强；难于提供满足 FWI 要求的初始速度场。

因此，需要探讨更具实用意义的速度估计方法和路线。

定义背景介质中的声波传播满足如下方程，

$$\mathcal{L}_{s0}(\mathbf{x},\omega)u_0(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{2.11a}$$

其中

$$\mathcal{L}_{s0}(\mathbf{x},\omega)=\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})} \tag{2.11b}$$

为背景介质中的标量波算子。

根据（2.2）和（2.11）式，定义标量波散射势为，

$$\mathcal{V}_s(\mathbf{x},\omega)\equiv\mathcal{L}_s(\mathbf{x},\omega)-\mathcal{L}_{s0}(\mathbf{x},\omega)=\left(\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}\right)-\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)=a_s(\mathbf{x})\frac{\omega^2}{v_0^2(\mathbf{x})} \tag{2.12a}$$

其中，

$$a_s(\mathbf{x})=1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})} \tag{2.12b}$$

上式定义的散射势还可以写为如下的形式，

$$\begin{aligned}\mathcal{V}_s(\mathbf{x},\omega)&=\omega^2\frac{1}{v_0^2(\mathbf{x})}\left(1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}\right)=\omega^2\frac{1}{v_0^2(\mathbf{x})}\left(\frac{v^2-v_0^2}{v^2}\right)=\omega^2\frac{1}{v_0^2(\mathbf{x})}\frac{(v-v_0)(v+v_0)}{v^2(\mathbf{x})}\\&\approx\omega^2\frac{1}{v_0^2(\mathbf{x})}\frac{2\delta v\cdot v_0}{v_0^2(\mathbf{x})}=\omega^2\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\end{aligned} \tag{2.13}$$

上述形式引入$\delta v(\mathbf{x})=v(\mathbf{x})-v_0(\mathbf{x})$更清楚地反映出速度的局部变化是散射势的本质。从散射势的定义也可以看出，估计散射势的绕射层析成像与估计或定位反射系数的叠前偏移成像是有严格区别的。

借助于算子恒等式$\mathcal{L}=\mathcal{L}_0+\mathcal{L}_0\left(\mathcal{L}_0^{-1}-\mathcal{L}^{-1}\right)\mathcal{L}$和定义$\mathcal{V}=\mathcal{L}_0^{-1}-\mathcal{L}^{-1}$可以得到

$$\mathcal{L}=\mathcal{L}_0+\mathcal{L}_0\mathcal{V}\mathcal{L} \tag{2.14}$$

假如波传播算子$\mathcal{L}$用 Green 函数G表示，方程（2.6）可以重写为

$$G=G_0+G_0VG \tag{2.15}$$

对应的 Born 级数序列为，

$$G = G_0 + G_0VG_0 + G_0VG_0VG_0 + G_0VG_0VG_0VG_0 + \cdots \tag{2.16}$$

假定 Born 近似假设成立，有 $G = G_0 + G_0VG_0$ （2.17）据此，一阶 Born 近似线性化的总波场可以写成

$$u(\boldsymbol{\xi},\boldsymbol{\eta},\omega) \approx u_0(\boldsymbol{\xi},\boldsymbol{\eta},\omega) + u_s^1(\boldsymbol{\xi},\boldsymbol{\eta},\omega) \tag{2.18}$$

其中，一阶 Born 近似后的散射波场记为

$$\begin{aligned} u_s^1(\boldsymbol{\xi},\boldsymbol{\eta},\omega) &= \omega^2 \int d\mathbf{x} G_0(\boldsymbol{\xi},\mathbf{x},\omega) \frac{a(\mathbf{x})}{v_0^2(\mathbf{x})} u_0(\mathbf{x},\boldsymbol{\eta},\omega) \\ &= \omega^2 \int d\mathbf{x} G_0(\boldsymbol{\xi},\mathbf{x},\omega) u_0(\mathbf{x},\boldsymbol{\eta},\omega) f(\mathbf{x}) \delta\mathbf{m}(\mathbf{x}) \end{aligned} \tag{2.19}$$

它代表从震源 $\boldsymbol{\eta}$ 到散射点 $\mathbf{x}$，再到接收点 $\boldsymbol{\xi}$ 的散射波场。其中，$\boldsymbol{\delta m}(\boldsymbol{x}) = \boldsymbol{a(x)} \big/ \boldsymbol{v_0^2(x)}$。事实上，还可以写出更高阶的 Born 近似后的波场。一般地，在勘探地震中，仅仅保留到二阶 Born 近似，得到二阶 Born 近似后的总波场

$$u(\boldsymbol{\xi},\boldsymbol{\eta},\omega) \approx u_0(\boldsymbol{\xi},\boldsymbol{\eta},\omega) + u_s^1(\boldsymbol{\xi},\boldsymbol{\eta},\omega) + u_s^2(\boldsymbol{\xi},\boldsymbol{\eta},\omega) \tag{2.20}$$

已知，（2. 10）式中定义的梯度为，

$$\mathbf{g}_k = \left[\frac{\partial u}{\partial m}\right]^T \left(u^{cal} - u^{obs}\right) \tag{2.21}$$

对应于二阶 Born 近似后的总波场，（2.21）式可以形式地可以重写为

$$\begin{aligned} \mathbf{g}_k &\approx \left[\frac{\partial u_0}{\partial m_B}\right]^T \left(u_0^{cal} - u_0^{0bs}\right) + \left[\frac{\partial u_s^1}{\partial \delta m}\right]^T \left({}^{cal}u_s^1 - {}^{obs}u_s^1\right) + \left[\frac{\partial u_s^2}{\partial \delta m_B}\right]^T \left({}^{cal}u_s^2 - {}^{obs}u_s^2\right) \\ &= \mathbf{g}_k^0 + \mathbf{g}_k^{s1} + \mathbf{g}_k^{s2} \end{aligned} \tag{2.22a}$$

此时，（2. 10）定义的迭代公式也可形式地写为

$$\mathbf{m}^{k+1} = \mathbf{m}^k - \left(\mu_k^0 \mathbf{g}_k^0 + \mu_k^{s1} \mathbf{g}_k^{s1} + \mu_k^{s2} \mathbf{g}_k^{s2}\right) \tag{2.22b}$$

就是说，FWI 可以用一种可以分解的方式来实现，即用地震记录中不同的波现象进行不用波数成分的速度场的反演。

针对（22）式引入的 Born 近似，速度场也被隐含地分解成

$$\mathbf{m} = \mathbf{m}_B + \delta\mathbf{m}_B + \mathbf{R}(\theta) \tag{2.23}$$

其中，$\mathbf{m}_B$ 与无散射的背景场对应，由透射波层析成像确定；$\delta\mathbf{m}_B$ 代表 $\mathbf{m}_B$ 的扰动，由反射波层析成像确定；$R(\theta)$ 代表角度反射系数，由最小二乘叠前深度偏移成像确定。这是当前勘探地震学中可分解的速度反演的典型的实现方式。基本过程可以被描述为

（1） 首先给定一个初始的 $\boldsymbol{m_B}$。它可以由透射波 $u_0(\boldsymbol{\xi},\boldsymbol{\eta},\omega)$ 的层析反演确定，更一般地是靠偏移速度分析确定。

（2） 基于此 $\boldsymbol{m_B}$ 进行（最小二乘）叠前深度偏移成像确定角度反射系数 $\mathbf{R}(\theta)$，即产生（角度）成像道集。假若道集拉平，认为给出的 $\boldsymbol{m_B}$ 可以产生正确的构造图像，据此产生的（最小二乘）叠前深度偏移结果可以解决勘探问题。成像工作到此结束。否则，转入步骤 3。

（3） 鉴于成像道集存在剩余深度差，$\boldsymbol{m_B}$ 不正确，进行基于（角度）成像道集的层析速度反演确定 $\delta\mathbf{m}_B$，并更新 $\boldsymbol{m_B}$。转入步骤 2。

很显然，目前的方法确定 $\boldsymbol{m_B}$、$\boldsymbol{\delta m_B}$ 和角度反射系数 $\mathbf{R}(\theta)$ 是分开的。但是该过程解了三个线性问题，而且确定 $\boldsymbol{m_B}$ 和 $\boldsymbol{\delta m_B}$ 涉及到大量的人工干预。

（2．18）式可以用线性方程组表示，

$$A(\mathbf{m}_B)\delta\mathbf{m} = \mathbf{d} \tag{2.19}$$

其中，右端项$\mathbf{d}$代表一阶 Born 近似散射场$u_s^1(\xi,\eta,\omega)$。$\boldsymbol{m}_B$代表背景速度，$\delta\boldsymbol{m}$代表背景速度的扰动，对反射波成像而言$\delta\boldsymbol{m}$对应角度反射系数$R(\theta)$。对$\boldsymbol{m}_B$而言，（2.19）式是一个非线性方程；若$\boldsymbol{m}_B$固定，对$\delta\boldsymbol{m}$而言，上式是一个线性方程。从勘探地震的角度看，模型参数可以视为一幅三维图像，其中蕴含一定的几何结构，对应反射界面处速度的越变。另一方面，模型参数的背景部分对应该三维图像的背景变化。$\delta\boldsymbol{m}$对应参数图像的奇变部分；$\boldsymbol{m}_B$对应它的光滑变化部分。事实上，我们希望用反射波场（也可以包括透射波场）同时确定$\boldsymbol{m}_B$和$\delta\boldsymbol{m}$。可以称此问题为可分的反演问题。认为至少存在两条途径：（1）盲反演方法；（2）Level-Set 方法。

3、盲反演方法

勘探地震中，把$\boldsymbol{m}_B$和$\delta\boldsymbol{m}$相结合的速度反演是比较典型的方法。盲反演试图同时反演$\boldsymbol{m}_B$和$\delta\boldsymbol{m}$。也可以用串联的方法对二者进行反演。下面我们讨论两种反演方法。首先给出如下误差泛函的定义

$$\begin{aligned}\min_{\boldsymbol{m}_B,\delta\boldsymbol{m}} \boldsymbol{E}(\boldsymbol{m}_B,\delta\boldsymbol{m}) &= \min_{\boldsymbol{m}_B,\delta\boldsymbol{m}}\left\{\left\|\boldsymbol{A}(\boldsymbol{m}_B)\delta\boldsymbol{m}-\boldsymbol{d}\right\|^2+\alpha\left\|\boldsymbol{R}\delta\boldsymbol{m}\right\|^2\right\}\\ &= \min_{\boldsymbol{m}_B,\delta\boldsymbol{m}_B}\left\{\left\|\begin{bmatrix}\boldsymbol{A}(\boldsymbol{m}_B)\\ \alpha\boldsymbol{R}\boldsymbol{I}\end{bmatrix}\delta\boldsymbol{m}-\begin{bmatrix}\boldsymbol{d}\\ 0\end{bmatrix}\right\|^2\right\}\end{aligned} \tag{3.1}$$

勘探地震学中，当背景速度$\boldsymbol{m}_B$正确时，还可以把像空间中在任何反射点上$\delta\boldsymbol{m}(\theta)$的排齐情况作为判断依据加入到（3.1）式的泛函中，对背景速度$\boldsymbol{m}_B$的反演进行约束。

3.1 盲反演问题的全耦合解法

首先，定义如下最小化问题

$$\min_{\boldsymbol{x}}\boldsymbol{\phi}(\boldsymbol{x})=\min_{\boldsymbol{x}}\frac{1}{2}\left\|\boldsymbol{\rho}(\boldsymbol{x})\right\|_2^2 \tag{3.2}$$

其中，

$$\boldsymbol{\rho}(\boldsymbol{x})=\boldsymbol{\rho}(\boldsymbol{m}_B,\delta\boldsymbol{m})=\begin{bmatrix}\boldsymbol{A}(\boldsymbol{m}_B)\\ \alpha\boldsymbol{R}\boldsymbol{I}\end{bmatrix}\delta\boldsymbol{m}-\begin{bmatrix}\boldsymbol{b}\\ 0\end{bmatrix}\text{和}\ \boldsymbol{x}=\begin{bmatrix}\boldsymbol{m}_B\\ \delta\boldsymbol{m}\end{bmatrix} \tag{3.3}$$

迭代反演所需要的几个关键公式为，

$$\boldsymbol{p}_k=\boldsymbol{\phi}'=\boldsymbol{J}_{\boldsymbol{\phi}}^T\boldsymbol{\rho} \tag{3.4a}$$

$$\boldsymbol{J}_{\boldsymbol{\phi}}^T=\begin{bmatrix}\partial\boldsymbol{\rho}(\boldsymbol{m}_B,\delta\boldsymbol{m})/\partial\boldsymbol{m}_B & \partial\boldsymbol{\rho}(\boldsymbol{m}_B,\delta\boldsymbol{m})/\partial\delta\boldsymbol{m}\end{bmatrix} \tag{3.4b}$$

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\tau_k\boldsymbol{p}_k \tag{3.4c}$$

利用（3.4）式可以同时更新$\boldsymbol{m}_B$和$\delta\boldsymbol{m}$。

3.2 盲反演问题的分离解法

上述问题还可以分成如下两步进行求解。第一步解一个线性问题，相当于求解一个最小二乘叠前深度偏移成像问题。

$$\delta\boldsymbol{m}^k=\arg\min_{\delta\boldsymbol{m}}\left\{\left\|\boldsymbol{A}(\boldsymbol{m}_B^k)\delta\boldsymbol{m}-\boldsymbol{b}\right\|^2+\alpha_k\left\|\boldsymbol{R}\delta\boldsymbol{m}\right\|^2\right\} \tag{3.5}$$

第二步解一个非线性问题

$$\boldsymbol{m}_B^k=\arg\min_{\boldsymbol{m}_B}\left\{\left\|\boldsymbol{A}(\boldsymbol{m}_B^k)\delta\boldsymbol{m}^k-\boldsymbol{b}\right\|^2+\beta_k\left\|\frac{\partial\delta\boldsymbol{m}^k(\theta)}{\partial\theta}\right\|^2\right\} \tag{3.6}$$

当数据残差最小和道集最佳拉平时，上述两步速度反演停止，认为$\boldsymbol{m}_B$和$\delta\boldsymbol{m}$即为所求的解。

4、Level-Set 方法

把要反演的参数值视为一个图像，其中包含纹理部分和背景部分。现代图像处理在图像反演重建过程

中引入 level-set 方法同时反演图像参数中的纹理部分和背景部分，即在迭代反演过程中，同时演化背景参数和纹理参数，当满足某种误差泛函定义的条件时，认为背景参数和纹理参数就是我们所要的反演结果。

上述思想在勘探地震中的速度反演中也有类似的做法。前已述及，把偏移成像和层析成像串联在一起，用偏移成像结果勾画反射界面的几何形态，用层析反演确定背景速度场，两步过程迭代进行，用图像处理中的几何界面检测方法勾画出反射界面，加到层析速度反演的建模过程中，建立 $\boldsymbol{m}_B$ 和 $\boldsymbol{\delta m}$ 两部分的速度模型。但是，上述方法还是经验性的，要求人工干预比较多。

Level-Set 方法可以把反演过程中 $\boldsymbol{m}_B$、$\boldsymbol{\delta m}$ 和 $\boldsymbol{\delta m}$ 所刻画的速度场的几何形态随迭代过程的演化有机地融合在一起。首先定义如下的泛函

$$\min_{\boldsymbol{m}_B,\boldsymbol{\delta m}} \boldsymbol{E}\left(\boldsymbol{m}_B,\boldsymbol{\delta m}\right)=\min_{\boldsymbol{m}_B,\boldsymbol{\delta m}}\left\{\left\|\boldsymbol{A}\left(\boldsymbol{m}_B\right)\boldsymbol{\delta m}-\boldsymbol{b}\right\|^2+\alpha_1\left\|\boldsymbol{D}\boldsymbol{m}_B\right\|^2+\alpha_2\int_{\Omega}\left|\boldsymbol{\phi}\left(\boldsymbol{\delta m}\left(\boldsymbol{x}\right)\right)\right|d\boldsymbol{x}\right\} \tag{4.1}$$

其中，$\boldsymbol{D}$ 为正则化化算子，一般地选为光滑算子。$\boldsymbol{\phi}\left(\boldsymbol{\delta m}\left(\boldsymbol{x}\right)\right)$ 为 level-set 函数，描述速度参数的界面形态。

上述泛函最优化的目标是在背景参数总能量最小、Level-set 函数（界面）总长度最小和正演模拟数据与实测数据最逼近的情况下确定 $\boldsymbol{m}_B$ 和 $\boldsymbol{\delta m}$ 的值。

Level-set 函数的演化满足如下的一阶常微分方程

$$\begin{cases}\dfrac{\partial\boldsymbol{\phi}}{\partial\boldsymbol{t}}+\boldsymbol{F}\left(\boldsymbol{x},\boldsymbol{t}\right)\left|\nabla\boldsymbol{\phi}\right|=0\\ \boldsymbol{F}\left(\boldsymbol{x},\boldsymbol{t}\right)=\boldsymbol{V}\left(\boldsymbol{x},\boldsymbol{t}\right)\cdot\boldsymbol{n}\left(\boldsymbol{x},\boldsymbol{t}\right)\end{cases} \tag{4.2}$$

其中，$\boldsymbol{\phi}$ 为 Level-set 函数，$\boldsymbol{V}\left(\boldsymbol{x},\boldsymbol{t}\right)$ 代表界面移动的“速度”，$\boldsymbol{n}\left(\boldsymbol{x},\boldsymbol{t}\right)$ 代表界面的外法线方向。显然，上式规定了反射界面随“时间 $\boldsymbol{t}$”演化的“速度”和方向。受上述泛函约束的界面演化“速度”和方向规定了反射界面的最终反演结果。

很显然，把 Levet-set 方法与 FWI 或 Tomography 结合时，演化方向（梯度方向）的计算得到修改。根据勘探地震反演的经验，$\boldsymbol{\delta m}$ 所刻画的速度场的几何形态在约束梯度方向方面是非常重要的因素。因此，利用（4.1）式定义的泛函进行速度模型的反演是非常符合勘探地震的实际情况的。

5、勘探地震学对计算数学的典型需求分析

勘探地震学中最核心的问题之一是利用观测的叠前地震数据估计尽可能精确的速度场。这是一个系统性的工作过程，涉及到野外地震数据采集、数据的预处理、成像剖面上的图像处理、地震波传播的数值模拟、反演方法和反演解的评价。

以可分解的速度反演问题为参照，可以列出如下五个方面的勘探地震学对计算数学的典型需求。

（1） 叠前地震数据采集与预处理
地震数据随机采样、地震数据规则化、地震数据稀疏特征表达与地震数据压缩、叠前去噪等。

（2） 现代图像处理
空间数据分析、图像分割、自动拾取与自动解释等。

（3） 地震波正演
3D Hemholtz 方程的差分求解、高频成分保真的波动方程有限差分解、高频近似下地震波场数值解、崎岖地表条件下弹性波数值模拟、小波变换域中的波动方程数值解等。

（4） 地震波反演
实际地震数据随机噪音分布特征确定、误差泛函建立方法、泛函正则化与预条件方法、大规模反演问题的数值解法、反演问题的收敛性等。

（5） 反演解的评价

后验概率及反演结果方差的计算、反演分辨率矩阵的计算、Bootstrap 反演方法、数据协方差矩阵和模型协方差矩阵的构建等。

勘探地震学中的速度估计方法目前以全波形反演为理论代表，试图利用尽可能好的观测数据估计宽波数带的速度场。实际上，由于数据的不完备、含较强的噪音、正演模拟不能描述复杂的波现象等使得 FWI 并不能很好地应用于勘探地震学中解决实际问题。提取波场中的一部分特征数据，在合适的正则化和/或预条件限制下，得到基本能满足工业需要的速度场，是更为合理的路线。这需要大量的现代信号分析、图像处理、统计判决和人工智能等方面知识的协助。

6、结论与讨论

勘探地震学中最核心的问题之一是利用观测的叠前地震数据估计尽可能精确的速度场。基于 Bayes 理论的全波形反演方法试图利用尽可能好的观测数据估计宽波数带的速度场似乎是最完整的理论方法。但是，由于数据的不完备、含较强的噪音、正演模拟不能描述复杂的波现象等使得 FWI 并不能很好地应用于勘探地震学中解决实际问题。

基于可分解的速度反演方法是更有实用意义的速度估计方法。它有两个层次的含义，第一个层次是基于多级 Born 近似，把 FWI 分解为透射波场的背景速度反演、基于一次反射（绕射）波场的最小二乘叠前深度偏移成像和基于二阶 Born 近似散射场的背景速度扰动量的反演；第二个层次的含义是基于一次反射（绕射）波的线性关系式构建可分的反演算法，同时反演速度在界面的扰动和背景速度（扰动）。Level-set 方法更是把速度值和速度几何结构在反演迭代步中同时演化，有望发展成实用化的方法技术。

勘探地震反演估计速度是一个系统性的工作，涉及到野外地震数据采集、数据的预处理、成像剖面上的图像处理、地震波传播的数值模拟、反演方法和反演解的评价。解决好此问题需要大量的现代信号分析、图像处理、统计判决和人工智能等方面知识的协助。

参考文献

[1] Matson, C. L., Borelli, K., S., Beckner, C. C., Jr, Hege, E. K., Lloyd-Hart, M., 2009, Fast and optimal multi-frame blind deconvolution algorithm for high-resolution ground-based imaging of space objects; Appl. Opt., 48(1): A75-A92

[2] Lödahl, M. G., 2002, multi-frame blind deconvolution with linear equality constraints; In: Bones, P. J., Fiddy, M. A., Millane, R. P., (Eds.) Image reconstruction from incomplete data II, Vol4792-21. SPIE, PP146-155

[3] Chung, J., Knepper, S., and Nagy, J. G., 2011, Large-scale inverse problems in imaging; In: Handbook of Mathematical Methods in Imaging; Scherzer, O., (Ed.), Springer

[4] Dorn, O., and Lesselier, D., 2011, Level-set methods for structural inversion and image reconstruction; In: Handbook of Mathematical Methods in Imaging; Scherzer, O., (Ed.), Springer

[5] Osher, S., Fedkiw, R., 2003, Level-set methods and dynamic implicit surfaces; Springer

[6] Bo-Feng, Huazhong-Wang, Xin Hu, 2014, Wave-equation first arrival travel-time tomography, EAGE Conference & Exhibition, EAGE, Expanded Abstracts, Th P02 04

[7] Kadlec, B. J., Dorn, G. A., and Tufo, H. M., 2008, Confidence and Curvature-Guided Level Sets for Channel Segmentation; SEG Annual Meeting, Expanded Abstracts.

地震波反演成像正则化思想与方法

王华忠

波现象与反演成像研究组（WPI），同济大学，上海，200092

【摘要】地震波反演成像的非线性性很强，正则化是获得稳定的反演解的必要措施。反问题的正则化思想与方法已有深入的论述。本文从地震波反演成像的角度进行新的解释。首先，对比分析了 Bayes 反演中的正则化思想与 Tikhonov 正则化思想的区别与联系，解释了协方差阵的逆作为正则化参数的意义。然后，通过数据协方差（逆）阵和模型协方差（逆）阵的正交分解，从波场和参数的特征表达的角度进一步阐释了反演成像中正则化的核心思想，并指出正则化的核心在于对梯度的预条件处理。最后，引出反射界面纹理约束的正则化算子，把正则化方法落脚到梯度预条件处理上。最小二乘叠前深度偏移数值试验证明了上述正则化思想的有效性。

【关键词】地震波反演成像，正则化，协方差阵，各向异性 Gauss 分布函数，图像纹理滤波算子，梯度

1 引言

地震波反演成像的理论框架已经相当成熟，但反演成像的结果远没有达到油气勘探专家的期望。勘探地震反问题的非线性性很强，解的不唯一性和不稳定性普遍存在。正则化是降低解的非唯一性，并获得稳定的反演解的必要措施。

正则化的数学定义是解一个与原反问题临近的良态的反问题。尽管概念上正确，但如何针对原反问题提一个临近的、良态的反问题是很困难的。

关于反问题的正则化，有大量的文献讨论此问题。这些文献大多是偏数学的，给出的正则化后的临近问题向原问题的解的收敛性证明对实际应用的指导十分有限。与勘探地震的具体应用有一定的差异。

因此，本文中我们试图从应用的角度讨论正则化的思想与方法。首先对比分析了 Bayes 反演中的正则化思想与 Tikhonov 正则化思想的区别与联系，解释了协方差阵的逆作为正则化参数的意义。重点是通过数据协方差（逆）阵和模型协方差（逆）阵的正交分解，从波场和参数的特征表达的角度进一步阐释了反演成像中正则化的核心思想，指出勘探地震反问题正则化的核心在于建立波场和模型参数之间的更为线性化的联系，最好是一对一的投影关系。我们认为数据的正则化是更为基础的正则化思想，首先应该从数据的正则化入手构建正则化方法。模型参数的正则化从参数的协方差（逆）阵引入会有更本质的认识，模型正则化一般地通过对梯度的预条件体现出来。最后通过去噪问题引出反射界面纹理约束的正则化算子，其可以起到对梯度预条件的作用，是一种比较实用的正则化方法。勘探地震实例证明了上述正则化思想与方法的有效性。

2 正则化的定义

地震波反演成像中，正则化起着至关重要的作用。正则化的定义存在着不同的表述。下面我们给出三种不同的定义。

定义 1（Engl，et al.，1996）：Regularization is the approximation of an ill-posed problem by a family of neighbouring well-posed problem。

定义 2（Kaipio and Somersalo，2005）：The basic idea of regularization methods is that, instead of trying to solve the original inverse problem exactly, one seeks to find a nearby problem that is uniquely solvable and that is robust in the sense that small errors in the data do not corrupt excessively this approximate solution。

上述两种定义的含义非常接近，从基本数学概念上说明了正则化的本质含义。但是，勘探地震实践中，上述正则化思想的具体实现方法多种多样。从数据和模型参数两个方面给出了不同的正则化方法，目的都是得到一个更稳定的反演解。因此，我们给出了一个更宽泛的、更具地球物理意义的正则化定义。

定义 3：正则化是使不适定（或非线性较强或凸性差）的地球物理反问题的解更稳定、更有地质意义的一类方法。

3 Bayes 框架下的反演与 Tikhonov 正则化反演的对比

地震波反演成像的基本框架应该首选 Bayes 估计框架。实测的地震数据被视为满足一定分布的随机信号，反演结果也是随机的。要反演的系统参数也被认为是随机的。用最大后验概率方法估计反演结果。其中，正则化思想主要体现在数据协方差逆阵和模型参数协方差逆阵上。Tikhonov 正则化反演可以认为是确定性的反演，且一般针对线性反演问题。关于数据和参数的正则化是从改善线性方程组或法方程组的系数矩阵的性质引入的。

3.1 确定性参数估计方法及正则化

地震波反演成像的正问题可以表示为：

$$\mathbf{d} = A(\mathbf{m}) + \boldsymbol{\eta} \tag{1}$$

其中，$A(m)$为传播算子。此处，隐含的假设是上述正问题是可以解释（或预测）信号（或波场）的。预测误差是可以用满足一定分布的加性噪音 $\boldsymbol{\eta}$ 来描述的。一般地，假设是高斯噪音。实际情况往往并非如此。本年度文集中“状态变量独立的地震波反演成像方法”一文较为详细地讨论了此问题。

定义上述正问题对实测数据的预测误差为

$$\Delta\left(\mathbf{d}^{obs}, \mathbf{d}^{cal}\left(=A(\mathbf{m})\right)\right) \tag{2}$$

对模型参数的估计可以定义为如下的优化问题

$$\hat{\mathbf{m}} = \arg\min_{\mathbf{m}} \left\{\Delta\left\{\mathbf{d}^{obs}, \mathbf{d}^{cal}\right\}\right\} \tag{3}$$

（2）式定义的预测误差度量可以有多种形式。譬如，L2 范数下的误差泛函定义为，

$$\Delta\left(\mathbf{d}^{obs}, A(\mathbf{m})\right) = \left\|\mathbf{d}^{obs} - A(\mathbf{m})\right\|^2 = \sum_i \left|d_i^{obs} - h_i(\mathbf{m})\right|^2 \tag{4a}$$

Lp（$1 < p < 2$ ）范数下的误差泛函定义为

$$\Delta\left(\mathbf{d}^{obs}, A(\mathbf{m})\right) = \left\|\mathbf{d}^{obs} - A(\mathbf{m})\right\|^p = \sum_i \left|d_i^{obs} - h_i(\mathbf{m})\right|^p,\ 1 < p < 2 \tag{4b}$$

K-L 范数下的误差泛函定义为

$$\Delta\left(\mathbf{d}^{obs}, A(\mathbf{m})\right) = \sum_i d_i^{obs} \ln \frac{d_i^{obs}}{h_i(\mathbf{m})} \tag{4c}$$

在勘探地震中，目前仅有（4a）式定义的偏差函数得到广泛应用。

确定性反演问题可以描述为

$$\hat{\mathbf{m}} = \arg\min_{\mathbf{m}} \left\{\Delta\left\{\mathbf{d}^{obs}, \mathbf{d}^{cal}\right\}\right\} \tag{5}$$

（5）式实质上是最大似然估计方法。它仅仅靠实测数据与正演计算数据的逼近来确定参数估计结果。可能存在很强的多解性，也可能存在解的不稳定性。需要引入正则化，减弱反问题的多解性，提高解的稳定性。直接写出（5）式应该蕴含有这样的假设：线性正问题可以很好地预测实际数据，线性正问题的系数矩阵（基本上）是满秩的。否者，直接写出（5）式进行反演，可能没有什么意义。这也从一个方面展现了引入正则化的必要性。

Tikhonov 正则化方法是经典的方法。在此框架下，模型参数估计方法被提为

$$\mathcal{J}(\mathbf{m})=\left\|\mathbf{d}^{obs}-A(\mathbf{m})\right\|_2^2+\lambda\left\|D\mathbf{m}\right\|_2^2 \tag{6}$$

其中引入了罚函数项，试图实现定义 1 和定义 2 给出的正则化定义。

（6）式对应的法方程为

$$\hat{\mathbf{m}}=\left(A^TA+\lambda D^*D\right)^{-1}A^T\mathbf{d}^{obs} \tag{7}$$

从中可以看出，引入罚函数项后，法方程是否满秩可以由罚函数项的系数决定。但是，如此引入罚函数项的物理含义是不清楚的，在勘探地震的反演实践中很难操作。

还可以有很多罚函数项的定义形式。把（6）式中罚函数项写成一般形式，

$$\mathcal{J}(\mathbf{m})=\left\|\mathbf{d}^{obs}-A(\mathbf{m})\right\|^2+\lambda\Omega(\mathbf{m}) \tag{8}$$

典型的罚函数项可以罗列如下：

经典平方正则化方法：

$$\Omega(\mathbf{m})=\left\|D\mathbf{m}\right\|^2=\sum_j\left|m_j-m_{j-1}\right|^2 \tag{9a}$$

稀疏正则化方法：

$$\Omega(\mathbf{m})=\left\|\mathbf{m}\right\|_1=\sum_j\left|m_j\right| \tag{9b}$$

总变差正则化：

$$\Omega(\mathbf{m})=\left\|D\mathbf{m}\right\|_1=\sum_j\left|m_j-m_{j-1}\right| \tag{9c}$$

组合正则化：

$$\Omega(\mathbf{m})=\sum_j\lambda_1\left|m_i\right|^2+\lambda_1\left|m_j-m_{j-1}\right| \tag{9d}$$

确定性的线性反演问题解的结构可以通过法方程系数矩阵的 SVD 分解得到清晰的说明（Vogel，2002）。

3.2 针对线性正问题的参数估计方法及正则化

线性正问题、对应的反演问题及正则化方法在地震波反演成像中处于核心地位。线性正问题表达为：

$$\mathbf{d}=G\mathbf{m} \tag{10}$$

其中，G 为线性化的传播算子。层析成像和最小二乘叠前深度偏移就是该问题的一个特例。本质上，勘探地震中的反演问题基本上都归结为线性正问题为基础的参数估计问题，非线性反演的每一个迭代步实质上都在解一个线性的反问题。Backus and Gilbert（1967）给出了非常抽象的线性反演问题的表达。

针对该正问题的确定性反演框架，建立的 Tikhonov 正则化反演泛函为

$$\Phi(\mathbf{m})=\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^TA^{-1}\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T+\mathbf{m}^TB^{-1}\mathbf{m} \tag{11}$$

上式对应的法方程为

$$\hat{\mathbf{m}}=\left(G^TA^{-1}G+B^{-1}\right)G^TA^{-1}\mathbf{d}^{obs} \tag{12}$$

对于上述反问题而言，正则化的核心是如何恰当地选择逆阵 A^{-1} 和 B^{-1}。一般地，逆阵 A^{-1} 选为数据的协方差矩阵；逆阵 B^{-1} 选为单位阵加上平方一阶导数算子形成的矩阵，或直接由 Laplace 算子构成的矩阵的逆。相比于引入超级参数（Super parameter），如此处理正则化问题，物理意义明晰，可操作性强。

Tikhonov 正则化反演泛函还可以按如下方式构建

$$\Phi(\mathbf{m})=\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T\Gamma_{\mathbf{d}}^{-1}\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T+\alpha\mathbf{m}^T\cdot\mathbf{I}\cdot\mathbf{m}+\beta\mathbf{m}^T\cdot D_1^2\cdot\mathbf{m} \tag{13}$$

其中，$D_1=\begin{bmatrix}\cdots & 0 & 0\\ -1 & 1 & 0\\ 0 & -1 & 1\\ 0 & 0 & \cdots\end{bmatrix}$

其中，上式中罚函数项包括两部分：前一部分惩罚反演的参数的总能量；后一部分惩罚反演的参数的变化率的总能量。两种考虑都是希望反演结果比较光滑。（13）式对应的法方程的解为

$$\hat{\mathbf{m}}=\left(G^T\Gamma_{\mathbf{d}}^{-1}G+\alpha\cdot\mathbf{I}+\beta\cdot\mathbf{D}_1^2\right)G^T\Gamma_{\mathbf{d}}^{-1}\mathbf{d}^{obs} \tag{14}$$

但是，（13）和（14）式中的正则化方法的物理意义依然不够明确，尽管已经比仅仅选择超级参数好了很多。

3.3 确定性反演问题的局限

确定性反演，无论从反演成像的理念上，或是从具体实现方法上，都存在一定的局限性。

1、认为正问题是可以预测实测波场中的信号的。

2、数据误差是确定性的信号，加上高斯白噪。

3、解是确定性的，不是随机的。缺乏对反演解的有效性的评价。

4、先验信息的表达是有局限的。

5、尽管有一些方法（比如 L 曲线方法，交叉验证方法等）可以确定引入的超级参数，但很难有效地确定。超级参数没有很明确的物理意义也是很难确定的原因。

3.4 随机性参数估计方法及正则化

3.4.1 随机性参数估计方法基本评价

随机反演方法的核心思想是：实测数据是随机信号，反演结果也是随机的，需要引入概率分布函数和统计量来评价反演解。不是从最小二乘反演问题对应的法方程的解的结构来评价解的质量。

基本的随机反演方法包括：贝叶斯估计方法和最大似然估计方法。最大熵估计方法也是常见诸文献的方法。

随机性参数估计方法优点：

1、显式地考虑误差和噪音的概率分布特征；

2、反演结果及要反演的参数是随机的，满足一定的概率分布；

3、可以用统计的方法来评价它们。

当然，随机性参数反演的缺点也是很明显的。譬如缺乏有效的计算方法计算后验概率密度和最大似然

函数，也很难从概率的角度直接计算反演解的期望和方差。

无论如何，随机性反演是地震波反演理论的一个更为抽象的理论框架。随机性反演问题的描述可以简述如下。

3.4.2 后验概率密度函数最大化（MAP）的参数估计方法

与控制方程相关的正问题可以描述为：

$$M:\mathbf{d}=A(\mathbf{m})+\boldsymbol{\eta} \tag{15}$$

其中，M 代表正问题是可以选择的，是反演中的一个变量。在本年度文集中，“状态变量独立的地震波反演成像方法”一文，通过在（15）式中引入状态变量，把正问题的选择对反演结果的影响具体体现出来。无论如何，必须充分意识到正问题的表达对反演结果有重要影响，并非仅仅考虑实测数据和模型参数的正则化就可以得到稳定的反演结果了。定义 1 和定义 2 实质上是不全面的正则化表述。定义 3 的表达之所以如此朦胧，也是有意避开正问题选择对反演解的影响。

基于上述抽象的正问题表达，可以写出贝叶斯估计的后验概率密度公式为

$$p(\mathbf{m}\,|\,\mathbf{d};M)=\frac{p(\mathbf{d}\,|\,\mathbf{m};M)\,p(\mathbf{m}\,|\,M)}{p(\mathbf{d}\,|\,M)} \tag{16}$$

上式中，假设选择的正问题 M 能模拟实测数据中的信号，实测数据与模拟数据的残差满足一定的概率分布 $p(\mathbf{d}^{obs}\,|\,\mathbf{m};M)=p_{\boldsymbol{\eta}}(\mathbf{d}-A(\mathbf{m}))$；假设先验信息满足一定的概率分布 $p(\mathbf{m}|\mathrm{M})$；假设实测数据满足一定的概率分布 $p(\mathbf{d}|\mathrm{M})$。

利用（16）式定义的后验概率密度进行参数估计的逻辑当然是用下式计算模型参数的期望值

$$\hat{\mathbf{m}}=E\{\mathbf{m}/\mathbf{d}\}=\int \mathbf{m}P(\mathbf{m}|\mathbf{d};M)\,d\mathbf{m} \tag{17}$$

但是，由于后验概率密度函数的计算是非常困难的，不具有可实现性，因此提出用后验概率密度函数的最大化进行参数估计的具体实现。但是，这样做是有条件的，基本的假设应该是后验概率密度函数的对称分布特征。否则，解释不清反演结果的含义到底是什么。目前，一般假设后验概率密度函数至少应该是广义 Gauss 分布的。

后验概率密度最大化（MAP）反演方法可以表示为

$$\begin{aligned}\hat{\mathbf{m}}&=\arg\max_{\mathbf{m}}\left\{p(\mathbf{m}\,|\,\mathbf{d}^{obs})\right\}=\arg\max_{\mathbf{m}}\left\{p(\mathbf{d}^{obs}\,|\,\mathbf{m})\,p(\mathbf{m})\right\}\\&=\arg\min_{\mathbf{m}}\left\{-\ln p(\mathbf{d}^{obs}\,|\,\mathbf{m})-\ln p(\mathbf{m})\right\}\end{aligned} \tag{18}$$

上式可重写为

$$\hat{\mathbf{m}}=\arg\min_{\mathbf{m}}\left\{-\left(Q(\mathbf{d}^{obs},A(\mathbf{m}))+\lambda\Omega(\mathbf{m})\right)\right\} \tag{19}$$

其中，

$$Q(\mathbf{d}^{obs},A(\mathbf{m}))=-\ln p(\mathbf{d}^{obs}\,|\,\mathbf{m}) \tag{20a}$$

$$\lambda\Omega(\mathbf{m})=-\ln p(\mathbf{m}) \tag{20b}$$

假若正问题对实测数据预测的数据残差的先验假设满足高斯分布 $\boldsymbol{\eta}\sim N(0,\sigma_{\boldsymbol{\eta}}^2\mathbf{I})$，则数据先验概率密度

可以写成

$$p\left(\mathbf{d}^{obs}\,|\,\mathbf{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\mathbf{d}}^{2}}\left\|\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right\|^{2}\right] \tag{21a}$$

或者

$$p\left(\mathbf{d}^{obs}\,|\,\mathbf{m}\right)\propto\frac{1}{\sqrt{\left(2\pi\right)^{N}\det\left(\Gamma_{\mathbf{d}}\right)}}\exp\left[-\frac{1}{2}\left(\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right)^{T}\Gamma_{\mathbf{d}}^{-1}\left(\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right)\right] \tag{21b}$$

假若先验的模型参数概率密度函数也符合高斯分布 $\mathbf{m}\sim N\left(0,\sigma_{\mathbf{m}}^{2}\mathbf{I}\right)$，则参数先验概率密度函数写为，

$$p\left(\mathbf{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\mathbf{m}}^{2}}\left\|\mathbf{m}\right\|^{2}\right] \tag{22a}$$

或

$$p\left(\mathbf{m}\right)\propto\frac{1}{\sqrt{\left(2\pi\right)^{K}\det\left(\Gamma_{\mathbf{m}}\right)}}\exp\left[-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{prior}\right)^{T}\Gamma_{\mathbf{m}}^{-1}\left(\mathbf{m}-\mathbf{m}_{prior}\right)\right] \tag{22b}$$

则 Bayesian 后验概率密度函数为

$$p\left(\mathbf{m}\,|\,\mathbf{d}^{obs}\right)\propto\exp\left[-\frac{1}{2\sigma_{\mathbf{d}}^{2}}\left[\left\|\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right\|^{2}+\frac{\sigma_{\mathbf{d}}^{2}}{\sigma_{\mathbf{m}}^{2}}\left\|\mathbf{m}\right\|^{2}\right]\right] \tag{23a}$$

或

$$p\left(\mathbf{m}\,|\,\mathbf{d}^{obs}\right)\propto\exp\left[-\frac{1}{2}\left(\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right)^{T}\Gamma_{\mathbf{d}}^{-1}\left(\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right)-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{prior}\right)^{T}\Gamma_{\mathbf{m}}^{-1}\left(\mathbf{m}-\mathbf{m}_{prior}\right)\right] \tag{23b}$$

据此，可以写出如下的后验概率密度函数最大（MAP）的参数估计公式

$$\hat{\mathbf{m}}=\arg\max_{\mathbf{m}}\left\{p\left(\mathbf{m}\,|\,\mathbf{d}^{obs}\right)\right\} \tag{24}$$

3.4.3 高斯分布假设下的 MAP 估计方法与 Tikhonov 正则化方法对比

高斯分布假设下的 MAP 估计方法为

$$\hat{\mathbf{m}}=\arg\max_{\mathbf{m}}\left\{p\left(\mathbf{m}\,|\,\mathbf{d}^{obs}\right)\right\}=\arg\min_{\mathbf{m}}\left\{\mathcal{J}\left(\mathbf{m}\right)\right\} \tag{25}$$

上式中，代价函数为，

$$\mathcal{J}\left(\mathbf{m}\right)=Q\left(\mathbf{d}^{obs},A\left(\mathbf{m}\right)\right)+\lambda\Omega\left(\mathbf{m}\right)=\left\|\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right\|_{2}^{2}+\lambda\left\|\mathbf{m}\right\|_{2}^{2},\lambda=\sigma_{\mathbf{d}}^{2}\Big/\sigma_{\mathbf{m}}^{2} \tag{26}$$

其中，数据偏差项为，

$$Q\left(\mathbf{d}^{obs},A\left(\mathbf{m}\right)\right)=-\ln p\left(\mathbf{d}^{obs}\,\middle|\,\mathbf{m}\right)=\frac{1}{2\sigma_{\mathbf{d}}^{2}}\left\|\mathbf{d}^{obs}-A\left(\mathbf{m}\right)\right\|^{2} \tag{27a}$$

罚函数项为

$$\lambda\Omega(\mathbf{m})=-\ln p(\mathbf{m})=\frac{1}{2\sigma_{\mathbf{m}}^2}\|\mathbf{m}\|_2^2 \tag{27b}$$

可见，高斯分布假设下的 MAP 估计方法形式上等价于 Tikhonov 正则化方法。

3.4.4 高斯分布假设下的 MAP 估计结果的解释

高斯分布假设下的 MAP 估计结果（即 Tikhonov 正则化方法估计结果）的解释为：

$$\mathbf{m}\,|\,\mathbf{d}^{obs}\sim N\left(\hat{\mathbf{m}},\hat{P}\right),with\ \hat{\mathbf{m}}=\left(A^TA+\lambda I\right)^{-1}A^T\mathbf{d}^{obs},\hat{P}=\sigma_{\boldsymbol{\eta}}^2\left(A^TA+\lambda\mathbf{I}\right)^{-1} \tag{28}$$

即：MAP 估计结果满足高斯分布 $\mathbf{m}\,|\,\mathbf{d}^{obs}\sim N\left(\hat{\mathbf{m}},\hat{P}\right)$，期望值是给出的反演结果（就是 Tikhonov 正则化方法估计结果） $\hat{\mathbf{m}}=\left(A^TA+\lambda I\right)^{-1}A^T\mathbf{d}^{obs}$；方差由 $\hat{P}\left(=\sigma_{\boldsymbol{\eta}}^2\left(A^TA+\lambda\mathbf{I}\right)^{-1}\right)$ 决定。

3.4.5 随机性反演方法和确定性反演方法中正则化方法的对比

若线性正问题表达为 $\mathbf{d}^{obs}=\mathbf{d}^{cal}+\eta=G\mathbf{m}+\eta$，认为地震数据（包括拾取的到达时）是随机的，$\eta$ 是随机噪音，则反演结果也是随机的。针对该正问题的随机性反演框架为：

1、建立 Bayesian 反演泛函

$$\Phi(\mathbf{m})=\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T\Gamma_{\mathbf{d}}^{-1}\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T+\mathbf{m}^T\Gamma_{\mathbf{m}}^{-1}\mathbf{m} \tag{29a}$$

2、给出上式的法方程，

$$\hat{\mathbf{m}}=\left(G^T\Gamma_{\mathbf{d}}^{-1}G+\Gamma_{\mathbf{m}}^{-1}\right)G^T\Gamma_{\mathbf{d}}^{-1}\mathbf{d}^{obs} \tag{29b}$$

确定性的 Tikhonov 正则化反演泛函及对应的解为

$$\Phi(\mathbf{m})=\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T\Gamma_{\mathbf{d}}^{-1}\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T+\alpha\mathbf{m}^T\cdot\mathbf{I}\cdot\mathbf{m}+\beta\mathbf{m}^T\cdot D_1^2\cdot\mathbf{m} \tag{30a}$$

$$\hat{\mathbf{m}}=\left(G^T\Gamma_{\mathbf{d}}^{-1}G+\alpha\cdot\mathbf{I}+\beta\cdot\mathbf{D}_1^2\right)G^T\Gamma_{\mathbf{d}}^{-1}\mathbf{d}^{obs} \tag{30b}$$

（29b）和（30b）对比可见，

$$\Gamma_{\mathbf{m}}^{-1}\Rightarrow\alpha\cdot\mathbf{I}+\beta\cdot\mathbf{D}_1^2 \tag{31}$$

由（31）知，模型参数协方差矩阵的逆的正则化作用可以由 Tikhonov 正则化方法来解释。那么如何确定模型参数协方差矩阵的逆 $\Gamma_{\mathbf{m}}^{-1}$ 是合适的？可以用如下（各向异性）的 Gauss 函数

$$\Gamma_{\mathbf{m}}(\mathbf{r}_1,\mathbf{r}_2)=\sigma^2\exp\left(-\frac{(\mathbf{r}_1-\mathbf{r}_2)^2}{2L^2}\right) \tag{32}$$

展开上式可得

$$\mathbf{m}\cdot\Gamma_{\mathbf{m}}^{-1}\cdot\mathbf{m}=\frac{1}{2\pi}\frac{1}{(\sigma L)^2}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{L^2}{2}\right)^n\nabla^n\mathbf{m}\cdot\nabla^n\mathbf{m} \tag{33}$$

进一步可以写成

$$\Gamma_{\mathbf{m}}^{-1}=\alpha\cdot\mathbf{I}+\beta\cdot\mathbf{D}_1^2+High-order-term \tag{34}$$

由上述讨论可知，模型参数协方差矩阵的逆$\Gamma_{\mathbf{m}}^{-1}$是更为完整的模型正则化方法，$\Gamma_{\mathbf{m}}$中的元素体现了模型参数各分量之间的相关性。$\Gamma_{\mathbf{m}}^{-1}$显然是起到解开各分量之间相关性的作用。可见，事先知道模型参数的二阶统计特征是模型正则化的关键。但是，这是非常难以做到的。因此，正则化方法的可操作性非常重要。

3.4.6 随机性反演方法解的评价

由于反演问题解的高度非唯一性，严重的不稳定性，对解的可靠性的评价显得异常重要。但是，评价解的可靠性很可能比估计反演解付出的计算代价更大。

无论如何，给出如下的简单评价，更详细的评价参考 Tarantola（2005）的专著。

随机性反演解和真解的理论关系可写为

$$\hat{\mathbf{m}}=G^{\dagger}\mathbf{d}^{obs}=G^{\dagger}G\mathbf{m}_{true} \tag{35}$$

其中，

$$G^{\dagger}=\left(G^T\Gamma_{\mathbf{d}}^{-1}G+\Gamma_{\mathbf{m}}^{-1}\right)G^T\Gamma_{\mathbf{d}}^{-1} \tag{36}$$

引入符号$\hat{\mathbf{m}}=G^{\dagger}G\mathbf{m}_{true}=\mathbf{R}\mathbf{m}_{true}$，其中$\mathbf{R}$称为分辨率矩阵。在 Bayes 框架下，$\mathbf{R}$定义为

$$\mathbf{R}=\mathbf{I}-\Gamma_{\mathbf{m}}^{posteriori}\Big/\Gamma_{\mathbf{m}}^{\Pr iori} \tag{37}$$

其中，$\Gamma_{\mathbf{m}}^{posteriori}$是模型参数的后验协方差阵。$\Gamma_{\mathbf{m}}^{\Pr iori}$模型参数的先验协方差阵。若$\Gamma_{\mathbf{m}}^{posteriori}$是零矩阵，$\mathbf{R}=\mathbf{I}$，分辨率最高。此情形说明仅仅依赖数据本身就可以精确反演，这对应线性正演算子G满秩的情况，不需要引入关于模型参数的先验信息。若后验协方差等于先验协方差，$\mathbf{R}=\mathbf{0}$是零矩阵，分辨率为零。此情形蕴含着先验信息的引入不能改善分辨率的意思，这对应另一个极端，正演算子G完全建立不起来数据和模型参数间的关系。

4 特征波场与特征参数场及与正则化的关系

行文至此，我们已经知道，地震波反演成像的基础首先是由地震波正演算子建立其模型参数向实测数据的投影关系，该投影算子应该把模型参数空间的可能元素投影到数据空间的可能元素上。否者，无论什么框架下的反演都无法进行，也许仅仅剩下依据大量的采样和统计方法了。这对高维勘探地震而言显然是不可行的。

正则化思想与方法的深层次的本质是建立其模型参数与数据之间的一对一的投影。理论上，若线性正演算子G满秩，且能通过$\mathbf{d}^{obs}=G\mathbf{m}$完全解释地震数据，就实现了所追求的这种一对一的投影关系。这样的理论分析可以通过如下的特征波场与特征参数场之间的投影来近似。

Bayesian 估计后验概率密度写为

$$p(\mathbf{m})\sim\exp\left[-\frac{1}{2}\left(L(\mathbf{m})-\mathbf{d}\right)^T\Gamma_{\mathbf{d}}^{-1}\left(L(\mathbf{m})-\mathbf{d}\right)-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{prior}\right)^T\Gamma_{\mathbf{m}}^{-1}\left(\mathbf{m}-\mathbf{m}_{prior}\right)\right] \tag{38}$$

对应的误差泛函定义为

$$S(\mathbf{m})=\frac{1}{2}\left(L(\mathbf{m})-\mathbf{d}\right)^{T}\Gamma_{\mathbf{d}}^{-1}\left(L(\mathbf{m})-\mathbf{d}\right)+\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_{prior}\right)^{T}\Gamma_{\mathbf{m}}^{-1}\left(\mathbf{m}-\mathbf{m}_{prior}\right) \tag{39}$$

对误差泛函求导

$$\partial S(\mathbf{m})\Big/\partial\mathbf{m}=\frac{1}{2}L^{T}\Gamma_{\mathbf{d}}^{-1}\left(L\mathbf{m}-\mathbf{d}\right)+\frac{1}{2}\Gamma_{\mathbf{m}}^{-1}\left(\mathbf{m}-\mathbf{m}_{prior}\right)=0 \tag{40}$$

得到法方程

$$L^{T}\Gamma_{\mathbf{d}}^{-1}L\mathbf{m}-L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}+\Gamma_{\mathbf{m}}^{-1}\left(\mathbf{m}-\mathbf{m}_{prior}\right)=0 \tag{41}$$

法方程的解为

$$\hat{\mathbf{m}}=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L+\Gamma_{\mathbf{m}}^{-1}\right)^{-1}\left(\Gamma_{\mathbf{m}}^{-1}\mathbf{m}_{prior}+L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\right) \tag{42}$$

上述法方程的解可以重写为

$$\begin{aligned}\hat{\mathbf{m}}&=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L+\Gamma_{\mathbf{m}}^{-1}\right)^{-1}\left(\Gamma_{\mathbf{m}}^{-1}\mathbf{m}_{prior}+L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\right)=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\Gamma_{\mathbf{m}}^{-1}\Gamma_{\mathbf{m}}\right)^{-1}\Gamma_{\mathbf{m}}\left(\Gamma_{\mathbf{m}}^{-1}\mathbf{m}_{prior}+L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\right)\\&=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\Gamma_{\mathbf{m}}\left(\Gamma_{\mathbf{m}}^{-1}\mathbf{m}_{prior}+L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\right)=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\left(\Gamma_{\mathbf{m}}\Gamma_{\mathbf{m}}^{-1}\mathbf{m}_{prior}+\Gamma_{\mathbf{m}}L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\right)\\&=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\left(\mathbf{I}\mathbf{m}_{prior}+\Gamma_{\mathbf{m}}L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\right)=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\left(\mathbf{m}_{prior}+\Gamma_{\mathbf{m}}L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\right)\\&=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\mathbf{m}_{prior}+\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\Gamma_{\mathbf{m}}L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d}\end{aligned} \tag{43}$$

在勘探地震中，很多情况下，我们仅仅反演参数的扰动量，此时往往令先验模型参数 $\mathbf{m}_{prior}$ 等于零，此时上式写为

$$\hat{\mathbf{m}}=\left(L^{T}\Gamma_{\mathbf{d}}^{-1}L\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\Gamma_{\mathbf{m}}L^{T}\Gamma_{\mathbf{d}}^{-1}\mathbf{d} \tag{44}$$

已知自相关函数是正定对称矩阵，可以做正交分解。模型的自相关函数 $\Gamma_{\mathbf{m}}$ 可以做正交分解

$$\Gamma_{\mathbf{m}}=U\Sigma_{\mathbf{m}}U^{T},\quad \Gamma_{\mathbf{m}}^{-1}=U^{T}\Sigma_{\mathbf{m}}^{-1}U \tag{45}$$

其中，U 是正交阵。U 的列向量被视为一组正交基。

数据的自相关函数 $\Gamma_{\mathbf{d}}$ 也可以做正交分解

$$\Gamma_{\mathbf{d}}=V\Sigma_{\mathbf{d}}V^{T},\quad \Gamma_{\mathbf{d}}^{-1}=V^{T}\Sigma_{\mathbf{d}}^{-1}V \tag{46}$$

其中，V 也是正交阵。V 的列向量被视为一组正交基。

据此，可以引入数据的特征分解为

$$\Gamma_{\mathbf{d}}^{-1}\mathbf{d}=V^{T}\Sigma_{\mathbf{d}}^{-1}\left(V\mathbf{d}\right) \tag{47}$$

所谓的特征分解，即把原数据投影到正交坐标系 V 中，得到高维数据中的方向谱分解，可以认为是提取出了数据的局部平面波。这样的特征分解是基于高维数据的二阶统计量（自相关函数）的，类似于选了一组高维 Fourier 变换进行局部平面波分解。当然，也可以用高维小波基或框架进行数据分解。

模型参数的特征分解与数据的特征分解是类似的，可以定义为

$$\Gamma_{\mathbf{m}}^{-1}\mathbf{m}_{prior}=U^T\Sigma_{\mathbf{m}}^{-1}\left(U\mathbf{m}_{prior}\right) \tag{48}$$

即把先验模型参数投影到正交坐标系U中，这相当于在 2D 或 3D 模型参数场中进行线性的结构提取。

把数据特征分解矩阵与传播算子结合，改善传播算子矩阵的性质，称为数据预条件过程。据此有

$$L^T\Gamma_d^{-1}\mathbf{d}=L^TV^T\Sigma_d^{-1}\left(V\mathbf{d}\right)=\left(VL\right)^T\Sigma_d^{-1}\left(V\mathbf{d}\right) \tag{49}$$

其中，VL是进行数据预条件处理，即对局部平面波数据残差进行反投影。

综合上述，特征表达后的解为

$$\hat{\mathbf{m}}=\left(\left(VL\right)^T\Sigma_{\mathbf{d}}^{-1}\left(VL\right)U\Sigma_{\mathbf{m}}U^T+\mathbf{I}\right)^{-1}\left(\mathbf{m}_{prior}+U\Sigma_{\mathbf{m}}\left(VLU\right)^T\Sigma_d^{-1}\left(V\mathbf{d}\right)\right) \tag{50a}$$

或

$$\hat{\mathbf{m}}=\left(\left(VL\right)^T\Sigma_{\mathbf{d}}^{-1}\left(VL\right)\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\left(\mathbf{m}_{prior}+\Gamma_{\mathbf{m}}\left(VL\right)^T\Sigma_d^{-1}\left(V\mathbf{d}\right)\right) \tag{50b}$$

特征表达后的解还可以重写为：

$$\hat{\mathbf{m}}=\left(\left(L'\right)^T\Sigma_{\mathbf{d}}^{-1}\left(L'\right)U\Sigma_{\mathbf{m}}U^T+\mathbf{I}\right)^{-1}\left(\mathbf{m}_{prior}+U\Sigma_{\mathbf{m}}U^T\left(L'\right)^T\Sigma_d^{-1}\left(V\mathbf{d}\right)\right) \tag{51a}$$

$$\hat{\mathbf{m}}=\left(\left(L'\right)^T\Sigma_{\mathbf{d}}^{-1}\left(L'\right)\Gamma_{\mathbf{m}}+\mathbf{I}\right)^{-1}\left(\mathbf{m}_{prior}+\Gamma_{\mathbf{m}}\left(L'\right)^T\Sigma_d^{-1}\left(V\mathbf{d}\right)\right) \tag{51b}$$

其中，$L'=VL$是引入的特征波传播算子。它的具体形式主要取决于特征波表达算子V。由数据自相关矩阵的正交分解引入特征波表达算子V，从计算的角度看是不方便的，因为传播算子L是解析表达的、V是数据驱动得到的。但是，这并非不可协调的。

综合结论：

1、通过对数据的预条件处理（即特征波分解），提出特征波的概念是反演成像中更为根本的正则化思想。可以把一个反演问题更好地提成凸问题，或把一个病态问题转化为一个临近良态的问题。此时，对应的波传播正问题相应改变（即引入L'），模型也要相应地进行特征表达。

2、建立特征表达后的波场和模型参数之间的关系是更为基础的正则化思想。当前很多正则化方法都基于这样的思想衍生出来。

3、模型参数的协方差阵$\Gamma_{\mathbf{m}}$是对梯度的预条件算子。协方差阵的特点是基本满足 Gauss 分布的。因此，取各向异性高斯分布函数作为$\Gamma_{\mathbf{m}}$的近似是有充分的统计和物理依据的。

5 正则化基本方法及勘探地震中的地质约束正则化方法

正则化有思想和具体实现两个层次的问题需要讨论。前两节主要是思想层次上的讨论。勘探地震中，地震波反演成像基本上都是梯度导引类的方法实现的。地质约束的梯度预条件处理成为了受广泛关注的方法。实质上，就是用模型参数的协方差阵$\Gamma_{\mathbf{m}}$对梯度进行预条件处理。

5.1 正则化基本方法

正则化的目的是提出原病态问题的临近的良态问题。在线性正问题对应的反演过程中，有如下三类基本的正则化方法：

1、截断 SVD 方法。其中截断奇异值的取舍是核心问题。

2、滤波 SVD 方法。超级参数的选择是关键。

3、迭代正则化方法，把迭代次数作为正则化参数。MOROZOV 偏差原则决定迭代次数。合理的噪音水平的获得是重点。

5.2 勘探地震中的地质约束正则化方法

勘探地震涉及的反演成像中，把地质信息合理地融入反演结果中是基本的追求。因此，提出了所谓的地质约束正则化方法。其中的基本思想是把反射界面的结构信息作为约束信息加入反演过程中。

我们知道模型参数的协方差阵 $\Gamma_{\mathbf{m}}$ 是模型参数的二阶统计量，包含了模型参数的几何结构信息。但是，直接用逆阵 $\Gamma_{\mathbf{m}}^{-1}$ 进行地质约束正则化显然是困难地。因此提出所谓的 Shaping Regularization 方法（Fomel，2007）。这算是一种有一定代表性的方法。

参数预条件或 Reparameterization 或在变换域表达参数的情况下，对应的超定问题 Tikhonov 正则化方法如下。首先引入对模型参数的变换

$$\mathbf{m}' = D\mathbf{m} \tag{52}$$

其中，D 是对模型参数的特征变换算子。

模型参数在原域和变换域的 Tikhonov 正则化反演泛函分别为

$$\mathcal{J}(\mathbf{m}) = \frac{1}{2}\left\|L\mathbf{m} - \mathbf{d}^{obs}\right\|_2^2 + \lambda^2 \left\|D\mathbf{m}\right\|_2^2 \tag{53a}$$

和

$$\mathcal{J}(\mathbf{m}') = \frac{1}{2}\left\|LD^{-1}\mathbf{m}' - \mathbf{d}^{obs}\right\|_2^2 + \lambda^2 \left\|\mathbf{m}'\right\|_2^2 \tag{53b}$$

它们分别对应的法方程为

$$\left(L^T L + \lambda^2 D^T D\right)\mathbf{m} = L^T \mathbf{d}^{obs} \tag{54a}$$

$$\left(\left(LD^{-1}\right)^T \left(LD^{-1}\right) + \lambda^2 \mathbf{I}\right)\hat{\mathbf{m}}' = \left(LD^{-1}\right)^T \mathbf{d}^{obs}$$

$$\left(\left(D^{-1}\right)^T L^T L\left(D^{-1}\right) + \lambda^2 \mathbf{I}\right)\hat{\mathbf{m}}' = \left(D^{-1}\right)^T L^T \mathbf{d}^{obs} \tag{54b}$$

原域和变换域两种解的含义：

原域的光滑解 $\mathbf{m}$ 对应（53a）和（54a）中的粗糙算子 D：一般选为 Laplace 算子。

变换域的稀疏非光滑解 $\mathbf{m}'$ 对应（53b）和（54b）中的光滑算子 D^{-1}，一般选为 Gauss 算子。

在原域和变换域中进行求解，从理论上看解的性质是等价的。如下的推导说明了这一点。对变换域求解的法方程进行变形可以导出原域中求解的法方程。

$$\left(\left(D^{-1}\right)^T L^T L\left(D^{-1}\right)+\lambda^2\mathbf{I}\right)\mathbf{m}'=\left(D^{-1}\right)^T L^T\mathbf{d}$$
$$\left(\left(D^{-1}\right)^T L^T L\left(D^{-1}\right)\mathbf{m}'+\lambda^2\mathbf{m}'\right)=\left(D^{-1}\right)^T L^T\mathbf{d}$$
$$\left(\left(D^{-1}\right)^T L^T L\mathbf{m}+\lambda^2 D\mathbf{m}\right)=\left(D^{-1}\right)^T L^T\mathbf{d}$$
$$\left(\left(D^T\right)^{-1} L^T L\mathbf{m}+\lambda^2 D\mathbf{m}\right)=\left(D^T\right)^{-1} L^T\mathbf{d}$$
$$\left(L^T L\mathbf{m}+\lambda^2 D^T D\mathbf{m}\right)=L^T\mathbf{d}$$
$$\left(L^T L+\lambda^2 D^T D\right)\mathbf{m}=L^T\mathbf{d}$$

应该指出，尽管理论上等价，实际上，两种处理方式的难易程度及处理效果都不相同。一般地，在变换域进行求解，正则化的效果更好，更能凸显地质约束正则化的思想。为了明确期间，我们再把 Shaping Regularization Operator 的导出罗列如下，更进一步地解释梯度预处理的方法。

由去噪问题引入 Shaping Regularization Operator。真解或实测信号 $\mathbf{m}_t$，光滑解或预测信号 $\mathbf{m}_s$，两者之间的关系为，

$$\mathbf{m}_t=\mathbf{m}_s+Noise \tag{55}$$

去噪反问题提为

$$\mathcal{J}\left(\mathbf{m}_s\right)=\frac{1}{2}\left\|\mathbf{m}_s-\mathbf{m}_t\right\|_2^2+\lambda^2\left\|D\mathbf{m}_s\right\|_2^2 \tag{56}$$

对误差泛函求导

$$\partial\mathcal{J}\left(\mathbf{m}_s\right)\Big/\partial\mathbf{m}_s=\mathbf{m}_s-\mathbf{m}_t+\lambda^2 D^T D\mathbf{m}_s=0 \tag{57}$$

得到法方程

$$(\mathbf{I}+\lambda^2 D^T D)\hat{\mathbf{m}}_s=\mathbf{m}_t \tag{58}$$

定义光滑算子

$$\mathbf{P}=(\mathbf{I}+\lambda^2 D^T D) \tag{59}$$

定义非光滑算子

$$\mathbf{S}=\mathbf{P}^{-1}=(\mathbf{I}+\lambda^2 D^T D)^{-1} \tag{60}$$

重写上式

$$\lambda^2 D^T D=\mathbf{S}^{-1}-\mathbf{I} \tag{61}$$

代入法方程 $\left(L^T L+\lambda^2 D^T D\right)\mathbf{m}=L^T\mathbf{d}^{obs}$ 得到，

$$\begin{aligned}&\left(L^T L+\mathbf{S}^{-1}-\mathbf{I}\right)\hat{\mathbf{m}}=L^T\mathbf{d}^{obs}\\&\left(L^T L-\mathbf{I}+\mathbf{S}^{-1}\right)\hat{\mathbf{m}}=L^T\mathbf{d}^{obs}\\&\left(\mathbf{S}\left(L^T L-\mathbf{I}\right)+\mathbf{I}\right)\hat{\mathbf{m}}=\mathbf{S}L^T\mathbf{d}^{obs}\end{aligned} \tag{62}$$

定义 Shaping Regularization Operator 为

$$\mathbf{S}=\mathbf{P}^{-1}=(\mathbf{I}+\lambda^2 D^T D)^{-1} \tag{63}$$

Shaping Regularization Operator 含义的解释：

$$\left(\mathbf{S}\left(L^TL-\mathbf{I}\right)+\mathbf{I}\right)\hat{\mathbf{m}}=\mathbf{S}L^T\mathbf{d}^{obs} \tag{64}$$

可以看出，$\mathbf{S}$ 算子作用在梯度（$L^T\mathbf{d}^{obs}$）上，目的是对梯度进行加强梯度纹理的滤波。可以看出算子 $\mathbf{S}$ 对应 Hessian 逆算子。Hessian 算子是一个典型的光滑算子，逆算子是非光滑算子。$\mathbf{S}=\mathbf{I}$ 时，退化为无正则化情形；$L^TL=\mathbf{I}$ 时，无需反演，仅仅是一个滤波过程。$\mathbf{S}=\lambda\mathbf{I}$ 时，其作用类似一个 Scaling 算子 $\hat{\mathbf{m}}\approx\lambda L^T\mathbf{d}^{obs}$。$\mathbf{S}$ 算子是 Hessian 的逆算子，实际上起到梯度的特征分解的作用。一般地，可以通过构造各向异性 Gauss 函数来构建。

综合评价：目前，地震波反演成像的核心都是用迭代法进行参数的逐步更新，梯度是最关键的因素。各种影响都反映在梯度中；各种正则化方法的最终体现也是在梯度中（L1 范数、TV 正则化、变换域表达模型）；梯度表现出符合地质意义的特点是地球物理的要求；模型参数的协方差阵 $\Gamma_{\mathbf{m}}$ 包含了模型参数的线性结构特征，梯度预条件算子的构造应该从 $\Gamma_{\mathbf{m}}$ 出发；$\mathbf{S}$ 算子是几何含义明确的梯度预条件算子。

6 正则化应用实例

地震波反演成像的解的非唯一性和不稳定性使得正则化方法不可避免。正则化的引入无非从数据约束和模型约束两个方面入手。正问题引入的解的不确定的正则化处理也可以通过数据和模型参数的特征化表达得到弱化。

据此，可以提出如下的正则化反演泛函

$$\mathcal{J}\left(\mathbf{m}\right)=\left\|\mathbf{w}_{\mathbf{d}}\left(\mathbf{d}^{obs}-L\mathbf{w}_{\mathbf{m}}\mathbf{m}\right)\right\|_2^2+\lambda\left\|D\mathbf{m}\right\|_p \tag{65}$$

其中，$\mathbf{w}_{\mathbf{d}}$ 是对数据的不确定性引入的加权系数。数据协方差阵很难获取的情况下，通过相对简单的统计和评价给出可靠性不同的加权系数是合理的选择。不同的文献给出了很多具体的处理方法。$\mathbf{w}_{\mathbf{m}}$ 一般地选为某种正交变换族或框架构成列向量的矩阵，目的还是在某个稀疏变换域反演特征化的模型参数。D 是某种算子，与 $p\left(0\le p\le 2\right)$ 的选择有关。一般地，还是要体现高维模型参数的某种稀疏特征。因为，在稀疏特征表达后，模型参数空间大幅度缩小。

在 $p=2$ 时，上式对应的反演解为

$$\hat{\mathbf{m}}=\left(\mathbf{w}_{\mathbf{m}}^TL^H\mathbf{w}_{\mathbf{d}}^T\mathbf{w}_{\mathbf{d}}L\mathbf{w}_{\mathbf{m}}+\lambda D^TD\right)^{-1}\mathbf{w}_{\mathbf{m}}^TL^H\mathbf{w}_{\mathbf{d}}^T\mathbf{d}^{obs} \tag{66}$$

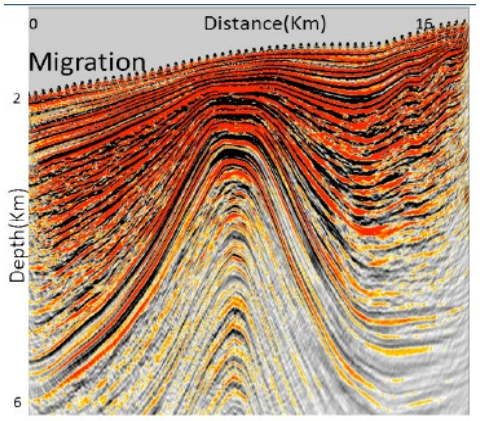


图 1a　实际数据一般的 Kirchhoff PSDM 成像结果

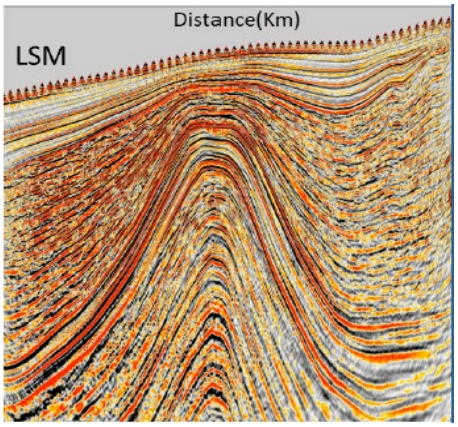


图 1b　实际数据基于（66）式的最小二乘 Kirchhoff PSDM 成像结果

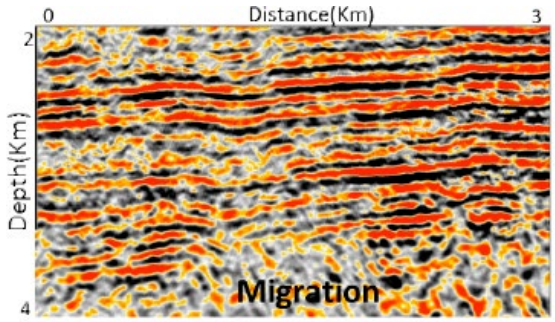


图 2a 实际数据一般的 Kirchhoff PSDM 成像结果

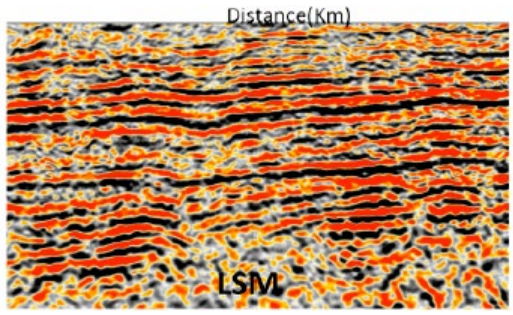


图 2b 实际数据基于（66）式的最小二乘 Kirchhoff PSDM 成像结果

图 1 和图 2 分别是两个具有不同地质特征的实际数据的 Kirchhoff PSDM 成像结果和基于（66）式的最小二乘 Kirchhoff PSDM 成像结果的对比（Trad，2015）。可以看出正则化的反演成像与一般成像结果之间的成像质量的差异。

特别值得注意的是，基于（66）式的反演成像，还是存在解的不唯一性和不稳定性。地下介质的几何形态和参数大小的定量估计的不确定性是永远无法彻底消除的。只能通过各种正则化方法把尽可能多的有助于得到可靠的反演解的先验信息恰当地加入到反演过程中。

7 总结与讨论

地震波反演成像的理论框架已经相当成熟，但反演成像的结果远没有达到油气勘探专家的期望。勘探地震反问题的非线性性很强，解的不唯一性和不稳定性普遍存在。正则化是降低解的非唯一性，并获得稳定的反演解的必要措施。

正则化的数学定义是解一个与原反问题临近的良态的反问题。尽管概念上正确，但如何针对原反问题提一个临近的、良态的反问题是很困难的。

我们首先对比分析了 Bayes 反演中的正则化思想与 Tikhonov 正则化思想的区别与联系，解释了协方差阵的逆作为正则化参数的意义。重点是通过数据协方差（逆）阵和模型协方差（逆）阵的正交分解，从波场和参数的特征表达的角度进一步阐释了反演成像中正则化的核心思想，指出勘探地震反问题正则化的核心在于建立波场和模型参数之间的更为线性化的联系，最好是一对一的投影关系。最后通过去噪问题引出反射界面纹理约束的正则化算子，可以起到对梯度预条件的作用，给出了一种正则化的实用的方法。

仅仅通过叠前地震数据要精确地反演地下介质的几何结构和参数大小是不可能的，永远会存在或强或弱的不确定性。正则化的目的本质上也是降低这种反演的不确定性。正则化的方法各种各样，本文可以说是对勘探地震反演成像中正则化思想与方法的一个较深入的分析，期望对后续的研究工作有明确的指导，基于此发展出更有效的正则化方法与技术。

参考文献

[1] Backus, G. E., and Gilbert, T. I., 1967, Numerical applications of a formalism for geophysical inverse problems, Geophys. J. R., Astr. Soc. 13, 247-276

[2] Engl, H. W., Hanke, M., and Neubauer, A., 1996, Regularization of inverse problem, Kluwer Academic Publishers

[3] Fomel, S., 2007, Shaping regularization in geophysical-estimation problems, Geophysics, 27(2), P. R29-R36.

[4] Kaipio, J., and Somersalo, E., 2005, Statistical and Computational Inverse Problems, Springer

[5] Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation; SIAM

[6] Trad, D., 2015, Least-square Kirchhoff depth migration: implementation, Challenges, and opportunities, 85th Annual International Meeting, SEG, Expanded Abstracts, 4238-4242

[7] Vogel, C.R., 2002, Computational Methods for Inverse Problems, SIAM

状态变量独立的地震波反演成像

王华忠 罗飞 冯波

波现象与反演成像研究组（WPI），同济大学，上海，200092

【摘要】地震波反演成像是勘探地震中的核心问题，在地下介质参数横向变化剧烈时，它是一个非线性很强的反演问题。不追求更好地解一个非线性（较）强的反问题，而是追求提一个更凸的反问题进行求解，是合理的反问题解决方案。本文首先将状态变量独立出来，把它作为一个反演参数，与模型参数地位等同，从状态方程约束反问题和数据约束反问题两方面出发，建立统一的理论框架，将常见的反演成像方法（传统全波形反演、微分相似度优化、层析成像、特征波反演等）纳入在此框架下，并分析它们的异同。期望在这样的理论框架下，发展出更有效的地震波反演成像理论、方法和技术。

【关键词】反演成像理论，状态方程，数据约束，成像条件

1 非线性反演或是更凸问题的反演？

地震波反演成像是勘探地震中的核心问题，它是一个标准的非线性反演问题，其非线性程度随着地下介质的复杂程度而提高。实测波场与地下介质之间的非线性程度决定了地震波反演问题的非线性程度。针对勘探地震中的参数反演问题，到底是专注于寻找解决非线性反演问题的思想与方法或是专注于提一个凸性更强的反演问题？

Rockafellar（1993）提出“In fact the great watershed in optimization isn’t between linearity and non-linearity, but convexity and non-convexity”。显然，提一个更凸的问题比追求更好地解一个非线性问题是更合理的解决反演问题的方案。

从 Tarantola（1984）建立以波动方程为基础的地震反演理论以后的近三十年来，尤其是最近几年来，随着以“两宽一高”为代表的地震数据采集技术以及以 GPU 为代表的高性能计算机技术的推进，全波形反演（FWI）方法和技术得到了长足的进步。在降低反演问题非线性方面做了很多有益的探索，提出了很多有效的方法。基本思想是从数据或波场入手来改进反演成像方法，典型的做法包括：从低频数据到高频数据的多尺度反演方法（Bunks et al. 1995；Sirgue & Pratt 2004）；从近偏移距数据到远偏移距数据的反演（Virieux & Operto 2009）；在 Laplace 域由小到大调节 Laplace 阻尼参数进行的由初至时刻不断放大时窗的反演（Shin & Cha 2009）；定义不同的失配函数的反演（Cara & Lévêque 1987；Luo & Schuster 1991；Gee & Jordan 1992；Fichtner et al. 2008；Van Leeuwen & Mulder 2010；Bozdag et al. 2011；Moghaddam & Mulder 2012；Warner , 2014）。还有另一类方法则是将传统的数据域反演扩展到成像域的反演（de Hoop et al. 2006；Shen & Symes 2008；Weibull et al. 2012）。

上述反演方法并没有从更为抽象的角度主动地把地震波反演问题提成一个更凸的问题。

本文从状态变量独立（即也作为一个反演参数，并与要反演的模型参数具有同等地位）的角度，重新表述地震波反演问题。提出状态变量和模型参数都进行特征化表达，把非线性较强的反演问题，（逐步地）转化为更凸的反演问题。最好的情形就是线性的二次凸问题。在此框架下，把经典 FWI、WRI（Herrmann，2013），AWI（Warner，2014）、DSO（Shen & Symes 2008）、数据域立体层析成像（Lambare, 2008）和成像道集层析成像重新表述，融合成一个整体。期望在这样的理论框架下，发展出更有效的地震波反演成像理论和方法。

2 Bayes 反演理论的缺陷

Bayes 反演理论是地震波反演成像的基础（Tarantola，2005）。Bayes 估计的思想可以简单概括为：估计的结果 $\hat{\mathbf{m}}$ 是由后验概率密度函数 $P\left(\mathbf{m}/\mathbf{y}\right)$ 所决定的均值 $E\left\{\mathbf{m}/\mathbf{y}\right\}$，其中 $\mathbf{y}$ 代表观测数据。

$$\hat{\mathbf{m}} = E\left\{\mathbf{m}/\mathbf{y}\right\} = \int \mathbf{m}P\left(\mathbf{m}/\mathbf{y}\right)d\mathbf{m} \qquad (1)$$

式（1）表示参数估计的结果是由观测数据所对应的模型参数出现的概率所决定的。

由于后验概率密度 $P\left(\mathbf{m}/\mathbf{y}\right)$ 很难估计，所以将其转化为较容易估计的模型空间的先验概率密度函数 $P\left(\mathbf{m}\right)$、数据空间的概率密度函数 $P\left(\mathbf{y}\right)$ 和条件概率密度函数 $P\left(\mathbf{y}/\mathbf{m}\right)$ 来表示，其中 $P\left(\mathbf{y}/\mathbf{m}\right)$ 体现了正算子的不确定性，描述了从模型空间到数据空间的映射关系。该 Bayes 公式为

$$P(\mathbf{m}/\mathbf{y}) = \frac{P(\mathbf{y}/\mathbf{m})P(\mathbf{m})}{P(\mathbf{y})} \tag{2}$$

关于模型空间的概率密度函数$P(\mathbf{m})$，其先验信息就是模型参数的均值和方差，$P(\mathbf{m})$的取值反映了对模型的认识情况。$P(\mathbf{m})$的均值和方差在反演求解过程中主要起到约束解的取值范围的作用。如果$P(\mathbf{m})$某个分量的方差较小，代表由先验信息所决定的该分量的值的可靠性就较高，因此该分量的估计结果主要由先验信息所决定；如果$P(\mathbf{m})$某个分量的方差较大，代表由先验信息所决定的该分量的值的可靠性就较低，因此该分量的估计结果主要由观测数据所决定。$P(\mathbf{m})$初步定义了反演问题的解在整个模型空间的概率密度函数，而反演问题的求解过程就是利用所观测到的数据更新这一概率密度函数的过程。

关于数据空间的概率密度函数$P(\mathbf{y})$，可采用以下方式定义其均值和方差。地震数据可认为是随机过程的一次具体实现，由于该随机过程只实现一次，所以可将该次实现看作该次随机过程的均值，即$P(\mathbf{y})$的均值由采集到的地震数据所决定。而$P(\mathbf{y})$的方差主要由地震数据的观测方式所决定，地震数据在某个空间范围内的采样密度越高，该空间范围内的地震数据所对应的方差就越小。如果假设数据空间的概率密度函数为高斯函数，则数据空间的概率密度函数可以用上述估计到的均值和方差来决定。

关于条件概率密度函数$P(\mathbf{y}/\mathbf{m})$，它表示在给定模型$\mathbf{m}$的条件下，采集到数据$\mathbf{y}$的可能性的大小。实际上，$P(\mathbf{y}/\mathbf{m})$也体现了用来描述实际物理现象的正算子的精确程度，当所用的正算子可以准确无误地描述实际物理过程时，$P(\mathbf{y}/\mathbf{m})$应该是一个δ类函数，否则它也应该有自己的均值和方差。因此，正演问题描述方法的差异会显著地影响$P(\mathbf{y}/\mathbf{m})$的取值。一般情况下，假设正算子在某些假设条件下可以比较准确地描述地震波的物理传播过程，并且其概率密度函数也是高斯函数。但实际情况并非如此，由于使正算子成立所要求的那些假设条件在实际情况中不一定都能得到满足，因此用正算子预测的数据与观测数据会存在较大的差异，并且正算子的概率密度函数也不一定是高斯函数，再加上先验信息的不充分，反演问题实际上是一个非常复杂的问题，如果不加上合理的约束条件，则很难直接得到一个具有地质意义的结果。

根据 Bayes 公式可以看出，$P(\mathbf{m}/\mathbf{y})$的均值和方差主要由$P(\mathbf{m})$，$P(\mathbf{y})$和$P(\mathbf{y}/\mathbf{m})$来决定。从这个角度看，对反演问题的求解可以认为是在$P(\mathbf{y})$，$P(\mathbf{m})$和$P(\mathbf{y}/\mathbf{m})$的约束下，使$P(\mathbf{m}/\mathbf{y})$最大化，并且期望均值为$\hat{\mathbf{m}} = \int \mathbf{m}P(\mathbf{m}/\mathbf{y})d\mathbf{m}$时，模型空间的方差最小。目前，常规反演问题的基本做法是假设$P(\mathbf{m})$和$P(\mathbf{y}/\mathbf{m})$均为高斯函数，在忽略观测系统不规则性的前提下，假设$P(\mathbf{y})$是标准的正态分布（当地震数据是规则采样时，其各个分量的方差应均相等），把$P(\mathbf{m}/\mathbf{y})$的最大化问题转化为代价函数$S(\mathbf{m})$的最小化问题。一般地，如果假设$P(\mathbf{m})$和$P(\mathbf{y})$均为高斯分布，并且忽略正算子的不确定性，则代价函数$S(\mathbf{m})$可以简单地定义为

$$S(\mathbf{m}) = \frac{1}{2}\left[(g(\mathbf{m}) - \mathbf{d}_{obs})^H C_D^{-1}(g(\mathbf{m}) - \mathbf{d}_{obs}) + \lambda(\mathbf{m} - \mathbf{m}_{prior})^H C_M^{-1}(\mathbf{m} - \mathbf{m}_{prior})\right] \tag{3}$$

式中：C_D为数据协方差矩阵；C_M为模型的协方差矩阵。

如果正演问题$g(\mathbf{m})$是线性函数，即$g(\mathbf{m}) = G\mathbf{m}$，并且矩阵$G^H C_D^{-1} G + \lambda C_M^{-1}$是正定的，则$S(\mathbf{m})$是一个正定二次函数，有唯一的极小值点。求解$S(\mathbf{m})$极小值点的收敛速度和解的稳定性取决于矩阵$G^H C_D^{-1} G + \lambda C_M^{-1}$条件数的大小，因此可以用 Newton 类的方法快速得到其最优解。如果矩阵$G^H C_D^{-1} G + \lambda C_M^{-1}$是不定的，即其特征值有正有负，则要先对$S(\mathbf{m})$做适当的正则化处理（在$S(\mathbf{m})$的构造过程中加入适当的约束条件），再求取其极小值点。

可见，Bayes 估计的思想具有很强的包容性。但是，利用（1）式或基于（2）式的 MAP 方法，都需要后验概率密度，计算上是不可行的。目前 Bayes 估计的主要方式是假设已知先验概率密度$P(\mathbf{y}/\mathbf{m})$和$P(\mathbf{m})$，一般地假设它们均满足高斯或广义高斯分布，基于 MAP 方法把估计问题转化为（3）式定义的代价函数求

解极小的问题。

（3）式存在的问题是隐含了一个很基本的假设：正问题 $g(\mathbf{m})$ 是可以解释实测数据 $\mathbf{d}^{obs}$ 的，不能解释的残差是符合高斯或某种广义高斯分布的。这种假设与勘探地震的实际是很不相符的，必须进行修改。

Tarantola（2005）指出："正问题不能很好地解释实测地震数据引入的数据残差与观测噪音（非信号部分）引入的数据残差可能是同一水平的"。我们认为，考虑到正演模型选择过于简单、仅考虑单参数反演、子波未知等等因素的影响，正问题的表达不合适引起的数据逼近误差很可能超过观测噪音引起的数据逼近误差。这就使经典的 FWI 完全失去了意义。

因此，把正问题的计算结果-状态变量-引入进来，并把它作为与模型参数同等地位的待反演参数看待，是深化对地震波反演问题的认识、把非线性的地震波反演成像问题提成一个更凸的反演问题的关键。

3 对状态变量独立的地震波反演问题的认识

状态变量是最优控制理论（赫孝良和葛照强，2015）描述系统状态的变量，与系统参数具有同等地位，共同满足控制方程。

勘探地震中，引入状态变量和系统参数的正问题的描述方法如下。把探区介质分布、激励和响应视为一个系统，对应该系统的非线性正问题的数学描述可以抽象为

$$A(\mathbf{m})\mathbf{u} = -\delta(\mathbf{x}-\mathbf{x}_s)\mathbf{f} = -\mathbf{f}(\mathbf{x}_s,t) \tag{4}$$

其中，A 为正演算子。在地震波反演中，A 的形式为

$$A = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right) = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \Delta \tag{5}$$

A 的形式受所选择的波动方程的形式的影响。可以选常密度声波，变密度声波方程，各种拟声波方程，弹性波方程等。$\mathbf{u}$ 为要正演合成的波场，即状态变量（系统响应）；$\mathbf{f}$ 为震源函数（激励函数）。

野外观测到的地震数据可以表述为

$$\mathbf{d} = P\mathbf{u} + \boldsymbol{\eta} \tag{6}$$

此处，P 表示状态变量的变换算子。η 表示噪音函数，在大部分情况下假设它是高斯白噪的。

实测数据和状态变量之间的关系由（6）式定义，其中蕴含了引入变换算子使得实测数据和状态变量进行匹配的思想，也可以说引入了让状态变量独立出来，作为一个可以处理的参量看待的思想。不再认为状态变量完全由系统参数、控制方程和激励函数决决定，尽管理论上的确由它们确定。这样的观点对后续的反演成像非常有意义。

我们认为，目前地震数据反演成像的基本逻辑主要包括相互有紧密联系的两条途径：一是 Bayes 估计理论；二是约束最优化理论。广义高斯分布假设下代价函数最小反问题优化求解方法基本上与约束最优化理论是等价的。下面我们从约束优化理论展开状态变量独立的地震波反演成像问题的讨论。

最简单的 FWI 反演成像目标函数定义为：

$$\min_{\mathbf{m}\in\mathbb{M}} \mathcal{J}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{d}^{obs} - \mathbf{d}^{cal}(\mathbf{m})\right\|^2 \tag{7}$$

其中，$\mathbf{d}^{obs}$ 为野外观测数据，$\mathbf{d}^{cal}(\mathbf{m})$ 为正演模拟数据。上式就是数据残差为高斯分布假设下的最大似然估计方法。另一个蕴含的假设是正问题的解能够解释实测地震数据，并使二者的残差满足高斯分布。这与地震波反演成像的实际相差甚远，反演结果有很强的多解性，完全无法满足应用的要求。因此，引出了各种各样的对对上式的修改，发展出了很多经典 FWI 的变种形式。一般地，是沿着 Bayes 估计的路线提出各种修改方法。概念上，各种典型的修改方法可以抽象为下式

$$\min_{\mathbf{m}\in\mathbb{M}} \mathcal{J}(\mathbf{m}) = \frac{1}{2}\left\| T\left(\mathbf{d}^{obs}\right) - T\left(\mathbf{u}(\mathbf{m})\right)\right\|^2 + \alpha \left\| D\mathbf{m}\right\|_p, \quad s.t. \mathbf{u}(\mathbf{m}) = \left(A(\mathbf{m})\right)^{-1}\mathbf{f} \tag{8}$$

其中，算子$T(\cdot)$ 可以选择各种各样的形式。譬如，可以对波场取对数，分别利用相位项的逼近和振幅项的逼近进行反演（$T(\mathbf{u}) = \ln \mathbf{u}$，Shin and Min，2006 ；$T(\mathbf{u}) = \mathrm{Re}(\ln \mathbf{u})$，$T(\mathbf{u}) = \mathrm{Im}(\ln \mathbf{u})$，Bedner，et al，2007；Sukjoon, et al，2007）；可以对波场进行 Laplace 变换，利用从初至到达时起始的不断扩大的窗函数截取不断增加的波场进行反演（$T(\mathbf{u}) = \int_W u(t)\mathrm{e}^{-st}\mathrm{dt}$，Shin and Cha，2008）；还可以对波场取包络，减轻 FWI 的 Cycle-Skipping 现象（$T(\mathbf{u}) = \sqrt{u^2(t) + \tilde{u}^2(t)}$，$u(t) = \frac{1}{\pi t} * u(t)$，Bozdag，et al，2011；Wu,et al，2014）。

但是，上述方法引入的对波场的变换没有涉及一个根本的问题，即实测波场不可以准确预测。我们认为，把正演计算波场作为系统的状态变量独立出来，是提高反演成像认识，发展出更有效反演成像方法的关键。

4 状态变量独立的地震波反演问题的提法

状态变量独立的地震波反演问题可以提成控制方程约束下的优化问题，即：

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \frac{1}{2}\left\| \mathbf{u} - P\left(\mathbf{d}^{obs}\right)\right\|_2^2 \quad s.t. A(\mathbf{m})P(\mathbf{u}) = \mathbf{q} \tag{9}$$

式中，$P(\cdot)$ 为施加在波场和观测数据上的变换算子。它可以是非线性算子，一般地定义为线性算子。$\mathbf{u}$ 为状态变量（波场），$\mathbf{d}^{obs}$ 代表实际观测数据，$\mathbf{m}$ 为系统（或模型）参数。$A(\mathbf{m})P(\mathbf{u}) = \mathbf{q}$ 为控制（或状态）方程，到底是变换后的波场或是未变换的波场满足控制方程依据反演成像的需要而定。$\mathbb{U}$ 和 $\mathbb{M}$ 分别是状态变量空间和系统（模型）参数空间。二者构成了满足控制方程 $A(\mathbf{m})P(\mathbf{u}) = \mathbf{q}$ 的联合空间。地震波反演成像就是在此联合空间中寻找最优解。事实上，激励函数 $\mathbf{q}\in\mathcal{Q}$ 也是需要反演的，但是在勘探地震中，一般假定激励函数是已知的。关于模型参数和状态变量的罚函数（约束函数），为了讨论简单，没有包括在上式中。这不会影响下述讨论的结论。

5 状态变量独立的地震波反演问题的解法

（9）式的典型解法有：拉格朗日乘子（Lagrange multiplier）法，即状态变量和系统参数同时反演的解法；消去状态变量的解法。后者就是经典的 FWI 方法。

5.1 状态变量和系统参数同时反演的解法

用拉格朗日乘子法（Lagrange multiplier）求解（9）式是典型的约束优化问题解法。通过引入一个新的参数 λ（即拉格朗日乘子），将等式（也可以是不等式）约束条件与代价函数结合在一起，形成新的代价函数，即

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \mathcal{J}(\mathbf{m},\mathbf{u},\lambda) = \min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \left(\frac{1}{2}\left\| \mathbf{u} - P\left(\mathbf{d}^{obs}\right)\right\|_2^2 + \left(A(\mathbf{m})P(\mathbf{u}) - \mathbf{q}, \lambda\right)\right) \tag{10}$$

上式中，$\mathbf{u}$为状态变量，$f(\mathbf{u},\mathbf{m}) = A(\mathbf{m})P(\mathbf{u}) - \mathbf{q}$ 是状态方程，λ为拉格朗日乘子或者称为伴随状态变量。式（10）的目标是在 $\mathbf{u}\in\mathbb{U}$ 和 $m\in\mathbb{M}$ 中找到一个满足 $f(\mathbf{u},\mathbf{m}) = 0$ 的 $\mathbf{u}^*$ 和 $\mathbf{m}^*$，使得 $\left\| \mathbf{u}^*\left(\mathbf{m}^*\right) - P\left(\mathbf{d}^{obs}\right)\right\|_2^2$ 达到最小。

为了求解（10）式，首先要求出如下的导函数，即：

$$\partial \mathcal{J}\left(\mathbf{m},\mathbf{u},\lambda\right)\Big/_{\partial \mathbf{m}}=\left[\partial f\left(\mathbf{u},\mathbf{m}\right)\Big/_{\partial \mathbf{m}}\right]^{*}\lambda \tag{11a}$$

$$\partial \mathcal{J}\left(\mathbf{m},\mathbf{u},\lambda\right)\Big/_{\partial \mathbf{u}}=\left[\partial f\left(\mathbf{u},\mathbf{m}\right)\Big/_{\partial \mathbf{u}}\right]^{*}\lambda \tag{11b}$$

$$\partial \mathcal{J}\left(\mathbf{m},\mathbf{u},\lambda\right)\Big/_{\partial \lambda}=f\left(\mathbf{u},\mathbf{m}\right) \tag{11c}$$

其中，（11a）式是代价函数对模型变量求偏导，得到模型参数迭代式的梯度项。（11b）式是代价函数对状态变量求偏导，得到了伴随状态方程；（11c）式是代价函数对拉格朗日乘子求偏导，得到状态方程。

对应的迭代公式为

$$\mathbf{m}^{k+1}=\mathbf{m}^{k}+\alpha_{\mathbf{m}}^{k}\left[\partial \mathcal{J}\left(\mathbf{m},\mathbf{u},\lambda\right)\Big/_{\partial \mathbf{m}}\right]^{k} \tag{12a}$$

$$\mathbf{u}^{k+1}=\mathbf{u}^{k}+\alpha_{\mathbf{u}}^{k}\left[\partial \mathcal{J}\left(\mathbf{m},\mathbf{u},\lambda\right)\Big/_{\partial \mathbf{u}}\right]^{k} \tag{12b}$$

$$\lambda^{k+1}=\lambda^{k}+\alpha_{\lambda}^{k}\left[\partial \mathcal{J}\left(\mathbf{m},\mathbf{u},\lambda\right)\Big/_{\partial \lambda}\right]^{k} \tag{12c}$$

依据牛顿法可以对上述三个变量同时进行迭代求解，当 $\left\|\nabla \mathcal{J}(m,\tilde{u},\tilde{a})\right\|_2 \le \varepsilon$ 时，迭代停止，得到最佳结果。

5.2 消去状态变量的解法

在勘探地震中，同时求解三个参数是不可能的。一般地，假定状态变量由控制方程确定下来，同时也假定了这样的状态变量可以预测实测波场中的信号，取消了状态变量的搜索空间 $\mathbb{U}$。

首先求解状态方程式（4），得状态变量

$$F\left(\mathbf{u},\mathbf{m}\right)=\mathbf{u}-A^{-1}(\mathbf{m})\mathbf{q}=0 \tag{13}$$

把上述状态变量代入，得到如下的误差泛函，

$$\min_{\mathbf{m}\in\mathbb{M}}\mathcal{J}(\mathbf{m})=\frac{1}{2}\left\|\mathbf{u}-P\left(\mathbf{d}^{obs}\right)\right\|_2^2=\frac{1}{2}\left\|A\left(\mathbf{m}\right)^{-1}\mathbf{q}-P\left(\mathbf{d}^{obs}\right)\right\|_2^2 \tag{14}$$

（14）式定义的最优化问题的求解，是经典的 FWI 问题解法。可以描述如下。

已知算子 $A\left(\mathbf{m}\right)=\dfrac{1}{v^2\left(\mathbf{x}\right)}\dfrac{\partial^2}{\partial t^2}-\Delta$，它可以理解为一个投影算子，建立了模型空间和数据空间之间的映射关系。$\partial A/\partial \mathbf{m}$ 表示算子的微分，它定义了一个 Jacobi 矩阵。在一定的定义下，它是一个 Fréchet 微商。假设算子 $A\left(\mathbf{m}\right)$ 存在逆 $A\left(\mathbf{m}\right)^{-1}$，则有，

$$A\left(\mathbf{m}\right)A\left(\mathbf{m}\right)^{-1}=\mathbf{I} \tag{15}$$

上式两边对 $\mathbf{m}$ 求导可得，

$$\frac{\partial A\left(\mathbf{m}\right)}{\partial \mathbf{m}}A\left(\mathbf{m}\right)^{-1}+A\left(\mathbf{m}\right)\frac{\partial A\left(\mathbf{m}\right)^{-1}}{\partial \mathbf{m}}=0 \tag{16}$$

上式可改写为，

$$\frac{\partial A(\mathbf{m})^{-1}}{\partial \mathbf{m}} = -A(\mathbf{m})^{-1}\frac{\partial A(\mathbf{m})}{\partial \mathbf{m}}A(\mathbf{m})^{-1} \tag{17}$$

定义数据拟合的残差矢量为

$$\mathbf{r}(\mathbf{m}) = \mathbf{d}^{cal} - \mathbf{d}^{obs} = F(\mathbf{m}) - \mathbf{d}^{obs} = C\left(A(\mathbf{m})\right)^{-1}\mathbf{f} - \mathbf{d}^{obs} \tag{18}$$

其中$\mathbf{r}(\mathbf{m})$表示波场残差，$F(\mathbf{m})$表示正算子根据模型$\mathbf{m}$得到的数据，$\mathbf{d}^{obs}$表示观测到的数据。

重写数据逼近误差泛函为

$$\min_{\mathbf{m}\in\mathcal{H}_1,\mathbf{d}\in\mathcal{H}_2} \Im_{LS}(\mathbf{m}) = \frac{1}{2}\left\| C\left(A(\mathbf{m})\right)^{-1}\mathbf{f} - \mathbf{d}\right\|_Y^2 = \frac{1}{2}\sum_t\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\mathbf{r}^2(\mathbf{m}) \tag{19}$$

首先，投影算子$A(\mathbf{m})$的 Gateaux 导数可以用如下的方式进行定义，

$$\left.\frac{\partial A(\mathbf{m}+\tau\mathbf{e}_i)}{\partial\tau}\right|_{\tau=0} = \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}}\cdot\mathbf{e}_i \tag{20}$$

其中，$\dfrac{\partial A(\mathbf{m})}{\partial \mathbf{m}}$代表投影算子$A(\mathbf{m})$的Fréchet微商。则误差泛函的梯度可以按如下形式导出

$$\begin{aligned}\left[G_{LS}\right]_i &= \frac{\partial}{\partial\tau}\Im_{LS}(\mathbf{m}+\tau\mathbf{e}_i)\Big|_{\tau=0} = \left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}\cdot\mathbf{r}(\mathbf{m}) \\ &= \left\langle \left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}, \mathbf{r}(\mathbf{m})\right\rangle\end{aligned} \tag{21}$$

其中，$\left.\dfrac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}$定义了正演波场随模型参数变化的方向导数---Gateaux 导数。定义误差泛函的梯度首先要给该导数。

由（13）式知，该导数可以写为，

$$\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i) = C\frac{\partial}{\partial\mathbf{m}}\left(A(\mathbf{m})^{-1}\mathbf{e}_i\right)f \tag{22}$$

利用（17）式，上式重写为

$$\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i) = -CA(\mathbf{m})^{-1}\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)A(\mathbf{m})^{-1}f = -CA(\mathbf{m})^{-1}\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u} \tag{23}$$

把（23）定义的变分代入（21）式，

$$\begin{aligned}\left[G_{LS}\right]_i &= \left\langle \left.\frac{\partial F}{\partial\tau}(\mathbf{m}+\tau\mathbf{e}_i)\right|_{\tau=0}, \mathbf{r}(\mathbf{m})\right\rangle_{\mathcal{H}_2} = -\left\langle CA(\mathbf{m})^{-1}\left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u}, \mathbf{r}(\mathbf{m})\right\rangle_{\mathcal{H}_2} \\ &= -\left\langle \left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u}, \left(A(\mathbf{m})^{-1}\right)^* C^*\mathbf{r}(\mathbf{m})\right\rangle_{\mathcal{H}_1} = \left\langle \left(\frac{\partial A}{\partial\mathbf{m}}\mathbf{e}_i\right)\mathbf{u}, \mathbf{z}\right\rangle_{\mathcal{H}_1}\end{aligned} \tag{24}$$

（24）中第二行利用了数据空间和模型空间存在的共轭关系（这种关系是假设的，理想弹性介质情况

下应该是可以接受的!），用共轭算子$\left(A(\mathbf{m})^{-1}\right)^{*}$把数据残差投影到模型空间。定义$\mathbf{u}=A(\mathbf{m})^{-1}f$为正演波场；$\mathbf{z}=-\left(A(\mathbf{m})^{-1}\right)^{*}C^{*}\mathbf{r}(\mathbf{m})$为残差波场的反传播。

可以看出，波场$\mathbf{z}$（共轭状态变量）的求解需要解如下的微分方程

$$\left(A(\mathbf{m})\right)^{*}\mathbf{z}=-C^{*}\mathbf{r}(\mathbf{m}) \tag{25}$$

前已述及，$A(\mathbf{m})$是一个共轭算子，满足条件：$\left(A(m)^{-1}\right)^{*}=\left(A(m)^{*}\right)^{-1}$。

为了清楚起见，把正向波场外推方程和反向波场外推方程罗列如下：

正向波场外推方程：

$$\begin{cases}\left(\dfrac{1}{v^{2}(x)}\dfrac{\partial^{2}}{\partial t^{2}}-\Delta\right)u(\mathbf{x},t)=-\delta(\mathbf{x}-\mathbf{x}_{S})S(t)\\ u(\mathbf{x},t)\big|_{\mathbf{x}=\Omega}=f_{b}(\Omega,t)\end{cases} \tag{26a}$$

其中，$f_{b}(\Omega,t)$代表边界条件。一般用 PML 边界条件。Ω代表计算波场在空间上的边界。

反向波场外推方程：

$$\begin{cases}\left(\dfrac{1}{v^{2}(x)}\dfrac{\partial^{2}}{\partial t^{2}}-\Delta\right)z(\mathbf{x},t)=0\\ z(\mathbf{x},t)\big|_{\mathbf{x}=\xi}=u(\boldsymbol{\xi},t)\end{cases} \tag{26b}$$

其中，$\boldsymbol{\xi}$代表观测点坐标值。$u(\boldsymbol{\xi},t)$代表观测点处的波场。FWI 进行波场残差外推时，仅有观测点处有波场残差，因此，才有**（26b）**这样的边界条件的定义。实际上，波场逆时外推并不限于这样的边界条件。

可见，（26a）是一个初值问题，沿时间正向外推计算，计算状态变量$\mathbf{u}$；（26b）是一个边值问题，逆时间反向外推计算，计算共轭状态变量$\mathbf{z}$。

（24）式中计算梯度时，正演波场$\mathbf{u}$前还要作用一个算子$\dfrac{\partial A(\mathbf{m})}{\partial\mathbf{m}}\cdot\mathbf{e}_{i}$。它代表算子$A(\mathbf{m})$的微扰。定义算子$A(\mathbf{m})$的微扰为，

$$\delta A(\mathbf{m})=\left(\frac{1}{(v+\delta v)^{2}}\frac{\partial^{2}}{\partial t^{2}}-\Delta\right)-\left(\frac{1}{v^{2}}\frac{\partial^{2}}{\partial t^{2}}-\Delta\right)=\frac{\delta v}{v^{2}}\frac{\partial^{2}}{\partial t^{2}} \tag{27}$$

因此有，

$$\frac{\delta A(\mathbf{m})}{\delta\mathbf{m}}=\frac{1}{v^{2}}\frac{\partial^{2}}{\partial t^{2}} \tag{28}$$

由此，最后的梯度公式为，

$$\mathbf{g}(\mathbf{m}(\mathbf{x}))=\sum_{t}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\frac{1}{v^{2}(\mathbf{x})}\frac{\partial^{2}u(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta})}{\partial t^{2}}z(t,\boldsymbol{\xi},\mathbf{x},\boldsymbol{\eta}) \tag{29}$$

利用 Adjoint-State 方法求解梯度的地球物理意义可以解释为：误差函数对模型扰动的一阶导数值为震源函数激发的波场与残差波场的反传播得到的场的内积结果。这种计算梯度的方法需要求解方程 $\mathbf{u}=A(\mathbf{m})^{-1}\mathbf{f}$ 和 $A(\mathbf{m})^{*}\mathbf{z}=-C^{*}\mathbf{r}(\mathbf{m})$，这是两次波场正演计算。可以看出，反问题的计算量主要集中在梯度的计算上，每次迭代原则上至少需要两次正演计算。下图解释了上述过程。如图 1 所示，（a）为纵向梯度变化速度场；（b）采用有限差分法求解状态方程（26a）得到波场 $\mathbf{u}$，可知状态变量为正传的波场值；求解式（26b）则得到图 2f 所示的反传波场值 $\mathbf{z}$，也就是伴随状态变量；将图 2b 和 2f 的结果代入式（29）中，得到如图 2g 所示的单炮梯度。上述过程就是经典的 FWI 方法的计算步骤。

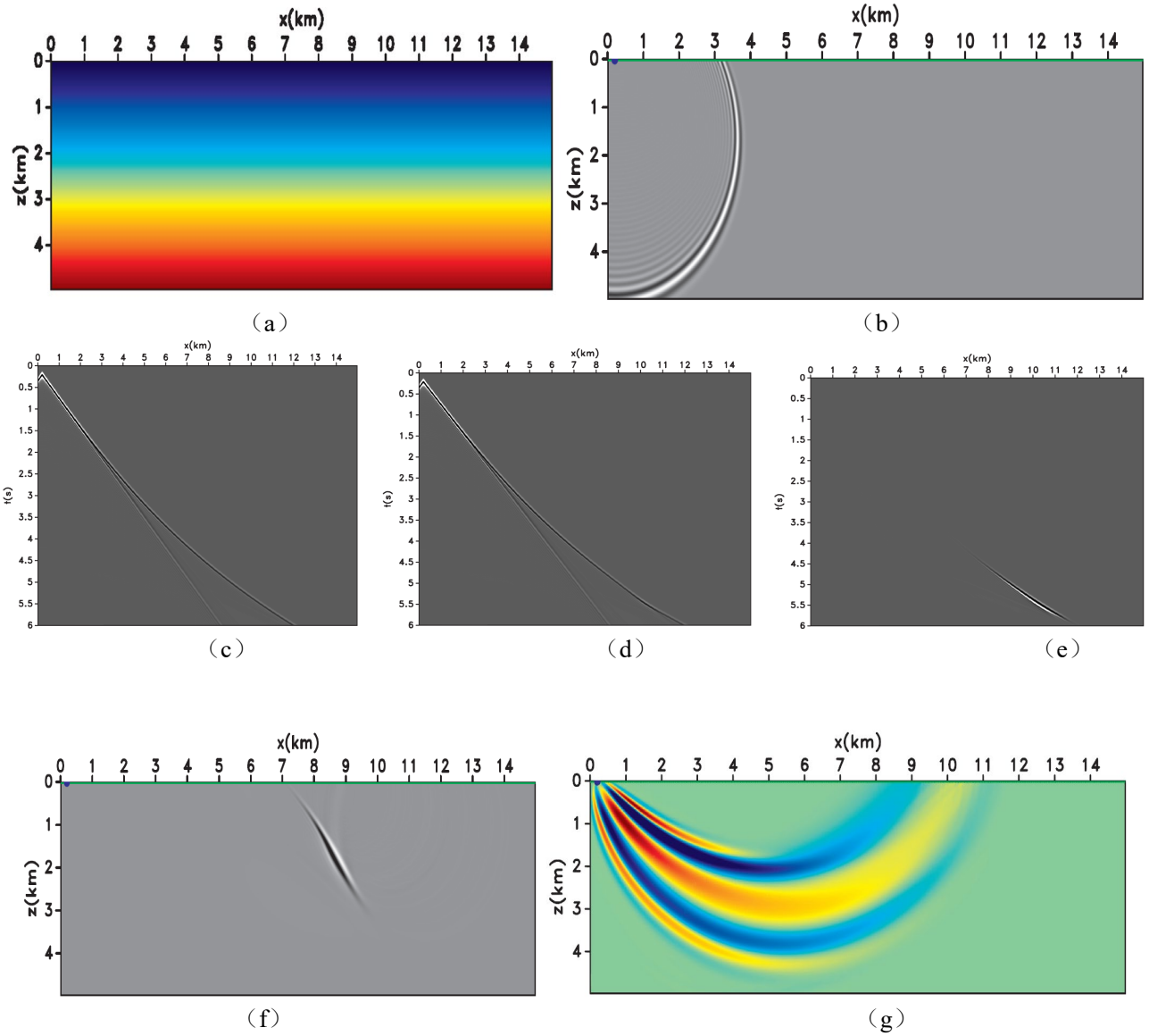


图 1 状态方程为声波方程时的示意图

（a）梯度速度场；（b）正传波场（状态变量 u）；（c）正演检波点处记录 $d^{cal}(m)$；（d）观测数据 d^{obs} ；（e）伴随源 $d^{cal}(m)-d^{obs}$ ；（f）反传波场（伴随状态变量 a）；（g）梯度

6 状态变量独立的地震波反演方法的意义

把状态变量独立出来看待地震波反演成像到底有什么意义？我们认为，把状态变量独立出来作为与模型参数同等的反演参数，同时满足控制方程，在一个联合空间 $\mathbb{U}\times\mathbb{M}$ 中寻找与实测数据最逼近的解，这样做本质上对正演计算的波场（状态变量）也可以提先验约束，不仅仅是对要反演模型参数提先验约束。从

两个角度同时提先验约束可以使非线性地震反演成像问题变得更凸。这是状态变量独立的地震波反演方法最本质的意义。

对状态变量提先验约束的基本出发点是让正演波场与模型参数之间的关系更加线性化，换句话讲是让二者之间的非线性关系弱化。更进一步地讲，是通过对实测波场的预处理，消除那些不能使正演波场与模型参数之间的关系更加线性化的波现象（譬如多次波）或波的表达（譬如振幅），使得实测波场、正演波场（状态变量）、模型参数、控制方程形成的体系在一个更凸的、更线性化的反演体系中统一起来。

上述思想内涵实质上已经在当前的地震波反演成像中有所体现，譬如特征波成像、立体层析成像、波场重构的成像等等。

7 状态变量独立的地震波反演方法实例

地震波反演从 Bayes 估计方法抽象的思想落实到具体实现上，无非要考虑如下方面的问题：状态变量的正演计算（其中包括了状态（控制）方程的选择，实质上也决定了对物理问题的认识，（单或多）参数的选择等）；状态变量与实测波场的逼近；状态变量的约束；模型参数的约束。

常规的 FWI 把状态变量由控制方程和模型参数决定下来显然是存在重大问题的，因为实际的物理过程不是可以由简单选择的控制方程及包含的模型参数决定的，这样的过程给出的状态变量不能很好地逼近实测数据。

波场重构反演（WRI）方法（Leeuwen，Herrmann and Peters，2014）和特征波场反演成像（CWI）方法（**王华忠等，2015**）是体现上述观点的两种反演成像方法。

7.1 波场重构反演（WRI=Wavefield Reconstruction Inversion）

Leeuwen, Herrmann and Peters（2014）提出如下的波动方程约束的优化反演优化问题，

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \lambda^2 \left\| A(\mathbf{m})\mathbf{u}-\mathbf{q}\right\|_2^2+\frac{1}{2}\left\|P\mathbf{u}-\mathbf{d}^{obs}\right\|_2^2 \tag{30a}$$

上式也可以写为，

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \left\| A(\mathbf{m})P^{-1}\mathbf{d}^{obs}-\mathbf{q}\right\|_2^2 \tag{30b}$$

可见波场重构反演方法的思想也是提出实测波场中满足控制方程的部分参与反演。为了对比起见，把常规 FWI 对应的反演优化问题罗列于此，

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \left\| PA(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}^{obs}\right\|_2^2 \tag{31}$$

（30b）式的反问题与（31）式反问题提法的基本思想正如前面的分析。

正如（10）式一样，勘探地震中，（30）式的直接求解几乎是不可能的。若固定模型参数 $\mathbf{m}$，求解（30a）式得到，

$$\left(\lambda^2 A(\mathbf{m})^* A(\mathbf{m})+P^*P\right)\mathbf{u}_\lambda=\lambda^2 A(\mathbf{m})^*\mathbf{q}+P^*\mathbf{d}^{obs} \tag{32}$$

（32）式被称为增广状态方程，$\mathbf{u}_\lambda$ 称为增广状态变量。

很显然，当 $\lambda\to\infty$ 时，（32）式退化为，

$$A(\mathbf{m})^* A(\mathbf{m})\mathbf{u}_\lambda=A(\mathbf{m})^*\mathbf{q}$$

即

$$A(\mathbf{m})\mathbf{u}_\lambda=\mathbf{q} \tag{33a}$$

增广状态方程退化为一般的状态方程。（30）式定义的反演问题退化为常规的 FWI 问题。当 $\lambda \to 0$ 时，（32）式退化为，

$$\left(P^{*}P\right)\mathbf{u}_{\lambda}=P^{*}\mathbf{d}^{obs} \tag{33b}$$

上式是一个滤波方程，从实测波场中抽取出想要的成分作为状态变量。上述分析清楚地解释了 WRI 方法的思想内涵。

但是，增广状态变量 $\mathbf{u}_{\lambda}$ 的物理含义很难解释。我们认为，来自实测波场的因素和来自控制方程的因素很难用一个 λ 算子来协调。协调的物理意义也不清楚。

Warner（2014）首次提出了自适应波形反演（AWI=Adaptive Waveform Inversion）方法，为了避免波场匹配时，遇到“周期跳跃”问题，该方法并不直接要求模拟数据与观测数据误差最小，而是借鉴了维纳滤波器的思想，建立如下目标泛函：

$$\mathcal{J}\left(\mathbf{m}\right)=\frac{1}{2}\left\|P\mathbf{w}-\mathbf{d}^{obs}\right\|+\frac{\alpha}{2}\frac{\left\|T\mathbf{w}\right\|}{\left\|\mathbf{w}\right\|} \tag{34}$$

上式中，第一项表示设计一个滤波器 $\mathbf{w}$ 作用在模拟数据 $\mathbf{d}^{obs}$ 上，目的是使其结果能够和观测数据相匹配。实质上，上述操作可以认为是提出了一个增广的状态变量与实测波场去逼近，而不是用既定的状态方程的输出与实测波场逼近。具体地，也不一定是振幅的逼近，完全可以仅仅设计一个相位逼近的 Weiner 滤波器。具体过程在 Warner（2014）的文章中并没有讲得很清楚，但是其蕴含的思想就是这样了。

7.2 特征波反演方法（CWI=Characteristic Wavefield Inversion）

王华忠等（2015）提出特征波反演成像的思想。我们认为无论如何选择控制方程都是不可能完全解释地震数据的。既然如此，我们从数据中筛选出能被控制方程较准确预测的波场。这部分波场称为特征波场。特征波场反演问题提为，

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}}\frac{1}{2}\left\|\mathbf{u}-P\left(\mathbf{d}^{obs}\right)\right\|_{2}^{2}\quad s.t.A^{P}(\mathbf{m})\mathbf{u}=\mathbf{q} \tag{35}$$

其中，P 为变换算子，可以取各种适当的形式，A^{P} 表述与变换算子有关的波场正传算子。上式还可以表述为，

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}}\frac{1}{2}\left\|C\left(A^{P}(\mathbf{m})^{-1}\mathbf{q}\right)-P\left(\mathbf{d}^{obs}\right)\right\|_{2}^{2} \tag{36}$$

特征波反演的核心问题在于解如下问题：

$$\mathcal{Q}\left(\mathbf{u}\right)=\frac{1}{2}\left\|P^{-1}\mathbf{u}-\mathbf{d}^{obs}\right\|_{2}^{2}+\alpha\left\|\mathbf{u}\right\|_{p} \tag{37}$$

得到特征波场 $\mathbf{u}$。把地震数据体变成特征数据体，推特征数据体进行反演成像。这基本等价于（33b）式所描述的情形。但我们没有显式地引入增广的状态变量，而是显式地提出了仅仅利用特征波场进行成像的思想。其中，我们认为对特征波场而言，控制方程是没有误差的，或误差可以忽略的（相当于（30a）式中的 $\lambda=0$）。相比于波场重构反演方法，特征波场反演的物理意义十分明确，可操作性很强，据此可以构建出有效的反演技术路线。尤其，（37）式的求解一般在压缩感知框架下进行，可以把先进的信号处理理论用于反演成像中。

更重要的是，特征波的提取不仅仅是现代信号框架（或压缩感知框架）下的操作，而是按照波动理论的物理意义进行操作，尽可能建立起特征波场与特征参数之间的联系，使得提出的反演成像问题更凸。

首先，给出特征波场的定义：时空局部的单震相的、方向带限的波场。此时我们就可以把一般的地震数据表达成特征数据，提取特征数据的过程可用下式来表示：

$$\mathbf{d}^C(\mathbf{x}_s, p_s, \mathbf{x}_r, p_r, t) = P(\mathbf{d}^{obs}(\mathbf{x}_s, \mathbf{x}_r, t)) \tag{38}$$

其中，$\mathbf{x}_s$ 代表了炮点的位置，$\mathbf{x}_r$ 代表了检波点的位置，p_s 代表了炮点端的射线参数，p_r 代表检波点端的射线参数，t 是记录到的旅行时场。$\mathbf{d}^C(\mathbf{x}_s, p_s, \mathbf{x}_r, p_r, t)$ 表示特征数据，$\mathbf{d}^{obs}(\mathbf{x}_s, \mathbf{x}_r, t)$ 代表常规地震数据，P 代表了变换算子。实质上，变换算子可以非常灵活，不同的变换算子代表我们希望用什么样的波场成分进行反演成像。

在特征波场的概念下，一般的正问题可写成积分形式：

$$u^C(\mathbf{x}_s, p_s, \mathbf{x}_r, p_r, t) = \int K(v_b(\mathbf{x}); \mathbf{x}_s, \mathbf{x}_r, \mathbf{x}, t) m^C(\mathbf{x}) d\mathbf{x} \tag{39}$$

其中，u^C 是特征波场数据，K 代表了灵敏度核函数，$\mathbf{m}^C(\mathbf{x})$ 是特征模型参数。当 K 的性质比较好（譬如其逆算子存在）的情况下，特征波场数据与特征模型参数之间建立起了一对一的映射关系。此时的反演成像问题凸性一定是好的。

在上述正问题的基础上，可以建立如下的地震波反演成像问题：

$$\mathcal{J}(v_b, \delta m^C) = \left\| \mathbf{u}^C(v_b, \delta m^C) - P(\mathbf{u}^{obs}) \right\|^2 \tag{40}$$

其中 $P(\mathbf{u}^{obs})$ 为特征波场，$\mathbf{u}^C$ 为计算的特征波场，$v_b(\mathbf{x})$ 为背景速度，$\delta\mathbf{m}^C(\mathbf{x})$ 则为速度扰动（反射系数）。$v_b(\mathbf{x})$ 和 $\delta\mathbf{m}^C(\mathbf{x})$ 是要反演的参数。

利用（40）式可以进行各种特征波现象的反演。譬如，若把 $P(\mathbf{u}^{obs})$ 和 $\mathbf{u}^C$ 界定为一次反射波场，可以进行利用一次反射波的反演。

可以利用（40）式进行一次反射波层析成像和最小二乘 PSDM，它们对应的梯度分别为 $\frac{\partial \mathcal{J}}{\partial v_b}$ 和 $\frac{\partial \mathcal{J}}{\partial \delta \mathbf{m}^C}$。

在波动方程反射波层析反演中，关键的问题是在于如何把来自于同一反射层的同震相的 $P(\mathbf{u}^{obs})$ 和 $\mathbf{u}^C$ 对应起来，保证不发生“串层”的情况，这是数据域反射波层析成像的关键所在。同样地，在 FWI（它不仅仅用了一次反射波和绕射波）中也存在相应地问题，这就是 Cycle-Skipping（周期跳跃）问题。在周期跳跃问题中，对数据的匹配要求更严格，不能超过半个子波周期。这也是 FWI 中初始模型参数要非常接近真实模型参数的原因，也就是在这样的条件下，FWI 反演问题才比较凸。

为了能够避免上述数据匹配存在的困难，可以把波动方程的反射波层析退化到反射波旅行时层析（RT_Tomo，或 Reflect Traveltime-Tomo）。此时的要求变成测出（自动测出）同一反射层的 $P(\mathbf{u}^{obs})$ 和 $\mathbf{u}^C$ 的旅行时差 Δt。这也是 RT_Tomo（反射波旅行时层析）的核心工作，剩下的就按照一般的反射层析框架工作即可。

从上述分析可以看出，定义合适的数据域中的波场残差或到达时残差是很困难的，必须提供相当准确的初始模型。这是地震波反演成像从数据域转移到成像域的主要原因。在像域进行地震波反演成像实质上是在利用不同成像点对应的不用照明角度的旅行时，这是像域中进行背景速度估计凸性更好的物理原因，因为到达时与背景速度的关系更为线性化。像域中信噪比高于叠前数据域的信噪比也是在像域进行背景速度估计的重要原因。

7.2 射线类立体层析成像

立体层析成像可以说是特征波成像的一个射线理论实现的实例。在立体层析成像过程中，数据空间可写成如下形式：

$$D^C(\mathbf{x}_s,\mathbf{x}_r,p_s,p_r,\mathrm{t}) \tag{41}$$

其中，$\mathbf{x}_s$ 代表了炮点的位置，$\mathbf{x}_r$ 代表了检波点的位置，p_s 代表了炮点端的射线参数，p_r 代表检波点端的射线参数，t 是记录到的旅行时场。射线理论立体层析成像的数据空间包含了上述三类变量。

对于模型空间，我们可以把模型参数分解为“背景+界面”的形式。模型空间写成：

$$M^C\left(v_b(\mathbf{x}),\mathbf{x},\alpha\right) \tag{42}$$

其中 $v_b\left(\mathbf{x}\right)$ 为背景速度，$\mathbf{x}=(x,y,z)$ 代表地下介质的界面位置，α 为地下界面的倾角。

从式（41）和式（42）可以看到，数据空间和模型空间都是多参数的，可以通过一个算子把数据空间和模型空间联系起来，这个算子也可以看作是模型空间向数据空间的一种映射关系，可表达为：

$$D^C\left(\mathbf{x}_s,\mathbf{x}_r,p_s,p_r,t\right)=F\left(M^C\left(v_b,\mathbf{x},\alpha\right)\right) \tag{43}$$

上式中，F 代表了模型空间向数据空间的映射关系，即代表了正演传播算子。在立体层析成像中，F 代表了射线类的正演传播算子（Farra,1987）。

通过（43）式，然后再加上不同的成像条件的约束，我们就可以建立起立体层析的成像过程，由于式（43）无论是数据空间还是模型空间都是多参数的，这给我们的反演过程带来了许多的麻烦，使得求解问题变得非常复杂。但是这也为我们提供了方便，可以比较灵活的选择参数进行反演。从而可以选择模型与数据之间关系比较敏感的参数进行反演。比如可以按照如下的方式进行反演。

1）首先在给定的初始速度下利用射线类算子进行偏移成像，得到地下介质的成像剖面，然后拾取成像剖面的位置和倾角（最好是自动化的拾取），然后利用拾取的位置和倾角向地表“打射线”，由于速度不准确，打到地表的射线和真实的射线不一致，其位置和射线参数都有可能不一致，利用 $t=t_s+t_r$ 的成像条件，建立射线参数扰动，炮检点位置扰动与地下速度扰动，成像位置和倾角的扰动之间的关系：

$$D(\delta p_s,\delta p_r,\delta x_s,\delta x_r)=\mathrm{F}(\delta \mathrm{v},\delta\vec{x},\delta\theta) \tag{44}$$

当然我们也可以只建立速度扰动和数据扰动之间的关系，把成像位置的更新和倾角的更新放在射线类偏移过程中，这个过程就和角度域层析成像一致。

2）假设数据空间的参数都已经得到，可以按照正确炮检点位置和射线参数向地下“打射线”，利用 $t=t_s+t_r$ 的成像条件进行成像，当速度模型正确时，由炮点端打下来的射线和检波点端打下的射线在正确位置相交，从而可以在正确的位置成像，但是当速度不正确时，炮检端打下来的射线不能相交，存在位置差，从而我们就可以建立速度扰动和位置扰动之间的关系，更新速度模型，建立的扰动关系可表示为：

$$\delta\vec{x}=F(\delta \mathrm{v}) \tag{45}$$

8 结论

Bayes 估计理论建立了地震波反演的完整的理论框架，原则上它已经考虑了正演问题的不确定性引入的反演解的不确定性。但是，从概率意义上，很难把正问题引入的不确定性融合到后验概率密度公式中。一般地，假设正问题输出是可以很好地预测实测数据的，即使有误差，也把它归结到观测误差中。这样的做法在勘探地震中是不可以容忍的。

把正问题的解（状态变量）独立出来，作为和模型参数同等地看待，可以从更高的程度上诠释地震波反演成像的实现方法。从反演成像的实现方法看，更本质的是把非线性的反演问题提成一个凸问题。提成一个凸问题的方法是综合考虑实测数据、状态方程、状态变量和模型参数，通过状态变量的独立，引出对实测数据的预条件（与独立的状态变量适配），再引出状态变量与模型参数的适配（强化它们之间关系的线性性），最后通过一个合适的状态方程把它们联系起来。通过这样的方式提出一个更凸的反演成像问题，发

展出实用化的地震波反演成像理论、方法与技术。

波场重构反演（WRI）方法和特征波反演（CWI）方法是上述思想的具体形式。自适应波形反演（AWI）可以看成 WRI 的一种具体实现，WRI 中的增广状态变量物理意义不清，很难直接发展出有效的算法，但 AWI 提出一种具体的实现方法。立体层析成像（Stero-Tomography）是特征波成像的一种具体实现。射线理论立体层析成像与射线理论角度道集层析又是等价的做法。特征反射波层析与角度成像域 DSO 是等价的做法。

我们认为，波场重构反演方法理论概念是正确的，但没有明确地指出具体实现方法。特征波反演成像方法物理意义十分清晰，与压缩感知信号特征表达框架、与凸问题的提出思想能非常好地融合在一起，是发展出实用化的地震波反演成像方法的正确途径。

参考文献

[1] Banerjee B, Walsh T F, Aquino W, et al. Large scale parameter estimation problems in frequency-domain elastodynamics using an error in constitutive equation functional[J]. Computer methods in applied mechanics and engineering, 2013, 253: 60-72.

[2] Bednar, J. B., C. Shin, and S. Pyun, 2007, Comparison of waveform inversion: Part 2: Phase approach: Geophysical Prospecting, 55, no. 4, 465–475

[3] Bozdağ E, Trampert J, Tromp J. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements[J]. Geophysical Journal International, 2011, 185（2）: 845-870.

[4] Bunks C, Saleck F M, Zaleski S, et al. Multiscale seismic waveform inversion[J]. Geophysics, 1995, 60（5）: 1457-1473.

[5] Cara M, Lévêque J J. Waveform inversion using secondary observables[J]. Geophysical Research Letters, 1987, 14（10）: 1046-1049.

[6] de Hoop, M.V., van der, H., Robert, D. & Shen, P., 2006. Wave-equation reflection tomography: annihilators and sensitivity kernels, Geophys. J. Int., 167(3), 1332–1352.

[7] Farra, 1987, Seismic Waveform Modeling In Heterogeneous Media by Ray Perturbation Theory, JOURNAL OF GEOPHYSICAL RESEARCH, 92, NO. B3, 2697-2712.

[8] Fichtner A, Kennett B L N, Igel H, et al. Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain[J]. Geophysical Journal International, 2008, 175（2）: 665-685.

[9] Gee L S, Jordan T H. Generalized seismological data functionals[J]. Geophysical Journal International, 1992, 111（2）: 363-390.

[10] Guasch L, Burgess T, Warner M. Optimised Adaptive Waveform Inversion-Improved Convergence Via Conjugate Gradients and Superior Step-length Calculation[C]//77th EAGE Conference and Exhibition 2015. 2015.

[11] Haber E, Ascher U M. Preconditioned all-at-once methods for large, sparse parameter estimation problems[J]. Inverse Problems, 2001, 17（6）: 1847.

[12] Lambare, G., Stereotomography. Geophysics, 2008. 73(5): VE25-VE34.

[13] Luo Y, Schuster G T. Wave-equation traveltime inversion[J]. Geophysics, 1991, 56（5）: 645-653.

[14] Moghaddam, P.P. & Mulder, W.A., 2012. The diagonalator: inverse data space full waveform inversion, SEG Technical Program Expanded Abstracts.

[15] Nocedal J, Wright S. Numerical optimization[M]. Springer Science & Business Media, 2006.

[16] Plessix R E. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications[J]. Geophysical Journal International, 2006, 167（2）: 495-503.

[17] Rockafellar, R. T., Lagrange multipliers and optimality, SIAM Review，35:183-283,1993

[18] Shin C, Cha Y H. Waveform inversion in the Laplace—Fourier domain[J]. Geophysical Journal International,

2009, 177（3）: 1067-1079.
[19] Shin, C. and D. Min, Waveform inversion using a logarithmic wavefield. Geophysics, 2006. 71(3): R31-R42
[20] Sirgue L, Pratt R G. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies[J]. Geophysics, 2004, 69（1）: 231-248.
[21] Sukjoon P, C., Shin, and J. B. Bednar, 2007, Comparison of waveform inversion: Part 2: amplitude approach: Geophysical Prospecting, 55, no. 4, 477–485
[22] Symes W W. Migration velocity analysis and waveform inversion[J]. Geophysical prospecting, 2008, 56（6）: 765-790.
[23] Symes W W. The seismic reflection inverse problem[J]. Inverse problems, 2009, 25（12）: 123008.
[24] Tarantola, A. & Valette, A, 1982. Generalized nonlinear inverse problems solved using the least squares criterion, Rev. Geophys. Space Phys., 20（2）, 129–232.
[25] Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation, Philadelphia: SIAM.
[26] Van Leeuwen T, Herrmann F J, Peters B. A New Take on FWI-Wavefield Reconstruction Inversion[C]//76th EAGE Conference and Exhibition 2014. 2014.
[27] Van Leeuwen T, Herrmann F J. A penalty method for PDE-constrained optimization in inverse problems[J]. arXiv preprint arXiv:1504.02249, 2015.
[28] Van Leeuwen T, Herrmann F J. Mitigating local minima in full-waveform inversion by expanding the search space[J]. Geophysical Journal International, 2013, 195（1）: 661-667.
[29] Van Leeuwen T, Mulder W A. A correlation-based misfit criterion for wave-equation traveltime tomography[J]. Geophysical Journal International, 2010, 182（3）: 1383-1394.
[30] Virieux J, Operto S. An overview of full-waveform inversion in exploration geophysics[J]. Geophysics, 2009, 74（6）: WCC1-WCC26.
[31] Warner M, Guasch L. Adaptive Waveform Inversion-FWI Without Cycle Skipping-Theory[C]//76th EAGE Conference and Exhibition 2014. 2014.
[32] Warner* M, Guasch L. Adaptive waveform inversion: Theory[M]//SEG Technical Program Expanded Abstracts 2014. Society of Exploration Geophysicists, 2014: 1089-1093.
[33] Weibull, W., Arntsen, B. & Nilsen, E., 2012. Initial velocity models for full waveform inversion, SEG Technical Program Expanded Abstracts 2012,pp. 1–4.
[34] Wu, R. S., B. Wu and J. Luo, Ultra-low-frequency information in seismic data and envelope inversion. 83rd Annual International Meeting, SEG, Expanded Abstracts, 2013: 3078-3082
[35] 赫孝良，葛照强，最优化与最优控制，2015，西安交通大学出版社
[36] 任浩然, 王华忠, 黄光辉. 地震波反演成像方法的理论分析与对比[J]. LITHOLOGIC RESERVOIRS, 2012, 24（5）.
[37] 王华忠, 冯波, 刘少勇等. 2015. 叠前地震数据特征波场分解、偏移成像与层析反演. 地球物理学报, 58(6): 2024-2034.
[38] 王华忠, 冯波, 王雄文等. 2015. 地震波反演成像方法与技术核心问题分析, 石油物探, 54(2): 115-125.

反射波成像与散射波成像的对比分析

王华忠，吴成梁，周阳，郭颂

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】地震波成像目前已经发展到以 FWI+LS_RTM 方法为核心的阶段，其中 FWI 主要利用透射波信息估计背景速度；LS_RTM 是一种基于逆散射思想的成像方法，试图估计高波数的波阻抗扰动或速度扰动。但是，完美的反演成像理论并没有带来能在实际中应用的可靠技术，FWI+LS_RTM 的技术组合远没有达到可以在任意数据体上都能产生明显的应用效果的程度。基本的观点是地震波正演模拟方法不能很好地模拟炮集中的复杂波场、噪音不满足高斯分布、震源子波空变且未知、初始模型的给出不能满足近似线性化反演的要求等等。实质上，还有一个重要的问题没有提及，就是仅仅有地面观测的背向传播的地震波场，与 FWI+LS_RTM 方法本身对数据的要求是不匹配的。本质上，FWI+LS_RTM 方法是一种逆散射成像方法，需要观测全向传播的地震波场，才有可能利用逆散射成像理论得到比较精确的介质参数成像结果。在仅仅有地面地震波场观测的情况下，我们认为应该引入勘探地震中面对的地下介质的特征和地震波场的特征，基于二者的特征构建以反射波为主的成像方法，以反射波成像结果作为逆散射成像的初始值，迭代的逆散射成像方法作为辅助更新反射波成像结果的手段。本文首先引入我们抽象出的勘探地震中介质和波场的特征表达；然后对反射波场和散射波场的特征进行分析；接着在地震波反演成像框架分析的基础上，提出合适的反射波成像和散射波成像的合理策略。

【关键词】散射波场，反射波场，层状介质，层状介质中波传播，散射介质，散射介质中波传播，反射波成像，散射波成像

1 引言

地震波成像目前已经发展到以 FWI+LS_RTM 方法为核心的阶段，其中 FWI 主要利用透射波信息估计背景速度；LS_RTM 是一种基于逆散射思想的成像方法，试图估计高波数的波阻抗扰动或速度扰动。FWI 方法本质上是包含了 LS_RTM 方法的。前者理论上是非线性的反演方法，可以估计从背景参数变化到高波数参数变化的所有成分。LS_RTM 是背景参数正确且已知的假设下的线性化反演，可以认为是 FWI 方法的一部分。FWI 方法完满的表达建立在 Bayes 估计理论框架技术上。

但是，完美的反演成像理论并没有带来能在实际中应用的可靠技术，FWI+LS_RTM 的技术组合远没有达到可以在任意数据体上都能产生明显的应用效果的程度。众多文献给出的基本的观点是地震波正演模拟方法不能很好地模拟炮集中的复杂波场、噪音不满足高斯分布、震源子波空变且未知、初始模型的给出不能满足近似线性化反演的要求等等，导致了 FWI+LS_RTM 不能给出满足实际生产要求的成像结果。

实质上，还有一个重要的问题没有提及，勘探地震情况下，我们仅仅在地面（或井中）布设检波器观测背向传播的地震波场，这与 FWI+LS_RTM 方法本身对数据的要求是不匹配的。本质上，FWI+LS_RTM 方法是一种逆散射成像方法，需要观测全向传播（即包围目标成像体的封闭面上）的地震波场，才有可能利用逆散射成像理论得到比较精确的介质参数成像结果。

幸运的是，勘探地震所面对的地下介质情况是比较特殊的，地下介质以横向缓变的层状沉积层为主，夹杂以由于地球动力运动导致的断层、尖灭、潜山等等异常地质体。波在这样的介质中传播以反射波为主，而不是以散射波为主。这正是勘探地震学有时被称为反射地震学的原因。我们认为勘探地震中的地下介质和地震波场是有特征的，基于二者的特征构建以反射波为主的成像方法，以反射波成像结果作为逆散射成像的初始值，迭代的逆散射成像方法作为辅助，更新反射波成像结果的手段，应该是合理的提高勘探地震中地震波成像结果质量的策略。

本文首先引入我们抽象出的勘探地震中介质和波场的特征表达；然后对反射波场和散射波场的特征进行分析；接着在地震波反演成像框架分析的基础上，提出合适的反射波成像和散射波成像的合理策略。

2 勘探地震中地下介质及地震波场的特征

图 1 展示了勘探地震中地下介质分布情况及一般的 3D 采集情况。可以看出勘探地震介质分布是有自身

特点的。

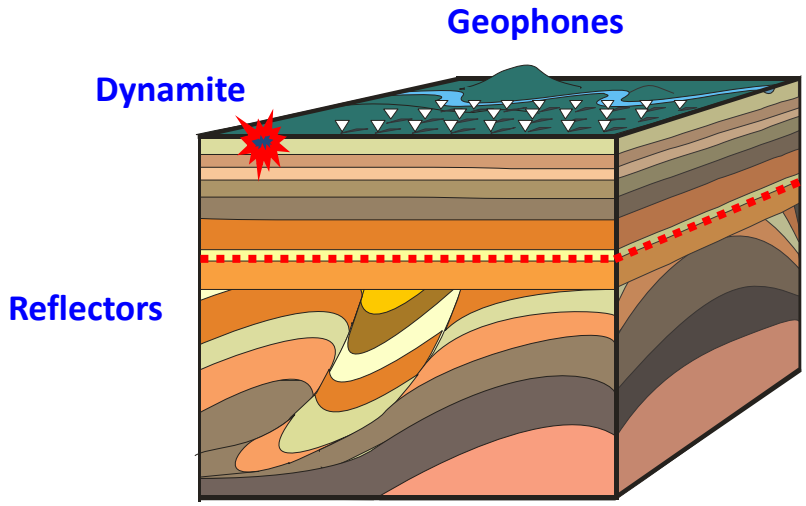


图 1 勘探地震中地下介质分布情况及采集系统示意图

我们把勘探地震的介质特征抽象为:**在空间上广泛分布的层状沉积层，加上火山活动，构造运动等引起的大、小尺度的速度异常体（或波阻抗变化的异常体）。**波在这样的介质中传播，必然以一次反射波、多次反射波和绕射波为主要的波现象。线性、非线性相干同相轴+不相干的噪音是地震信号的主要特点。人工地震震源（包括炸药、可控震源、气枪等）产生的是一个带限的、时间延续度有限的、震荡的、连续性较好（谱比较光滑）的子波（小波）。

我们把叠前地震数据的特征抽象为：**把地表（也可以是井中或海底（OBC）等处）观测的地震数据视为旅行时和炮点、检波点坐标为变量的函数，它的基本特征应该是时距关系规定的、带限反射子波表现出的同相轴。**

参数场中的反射界面和地震数据中的反射同相轴的广泛存在是勘探地震问题的主要特征。更进一步地，一次反射波同相轴是最主要的特征。反射波场可以视为仅仅是背向散射场，反射波场的成像仅仅需要一个面上的观测。参数场中的波阻抗扰动点和地震数据中的绕射同相轴（不是处处存在）是勘探地震问题的二级特征。散射波成像估计散射强度原则上需要包围散射点的封闭面上的场。

正是因为勘探地震中有着这样的介质分布特点和地震波特点，勘探地震中的地震波成像才发展成了独特的声成像方法技术体系。其中的关键技术环节包括：折射波层析与折射波静校正、初至波层析与静校正、NMO 叠加与均方根速度估计、NMO+DMO 叠加与均方根速度估计、PSTM 与对应的速度估计、PSDM 与对应的速度估计、成像道集及基于成像道集的层析速度反演，最后就是 AVA 分析和反演。这样的技术体系完全不同于医学声成像、材料声成像等类似学科中的技术。

因此，我们认为当前的常规地震波成像处理流程是按照地下介质的层状分布特点和一次反（散）射波的特点发展出来的有效的技术体系。基于该技术体系，石油工业取得了巨大的经济效益。

但是，该技术体系已经足够成熟了。FWI+LS_RTM 为代表的技术正备受关注，试图取代上述有效的技术流程。但实用化情况却步履维艰。

3 勘探地震中反射波和散射波传播的特征分析

勘探地震中面对介质分布的特殊性，决定了观测波场的特殊性和成像方法的特殊性。勘探地震中能用来成像的波场包括：透射波（初至波）；一次反射波；一次散（绕射）射波。其它复杂的波现象不能用于地震波偏移成像和层析成像中。

3.1 勘探地震中的介质特征与成像所用的波现象

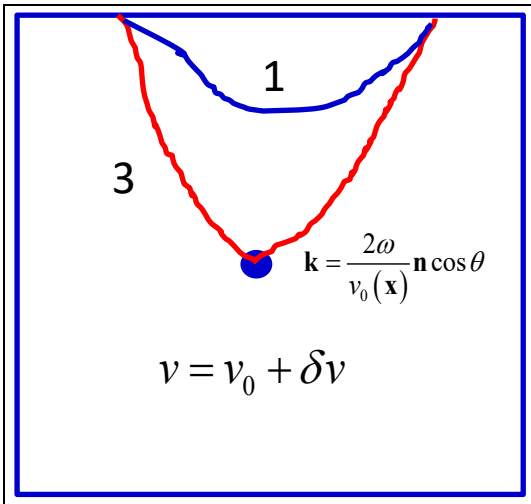


图 2a 透射波与散射波介质模型

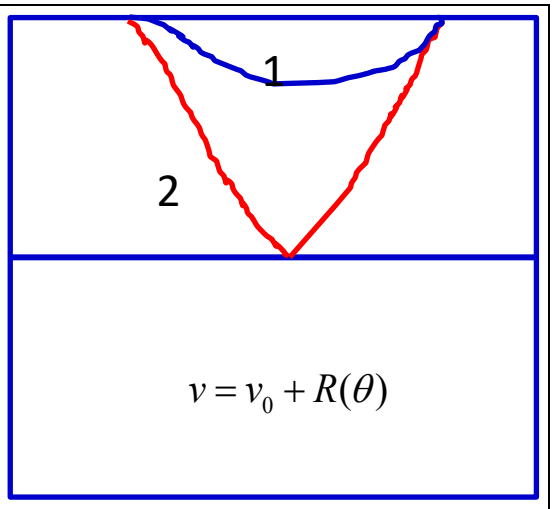


图 2b 透射与反射波介质模型

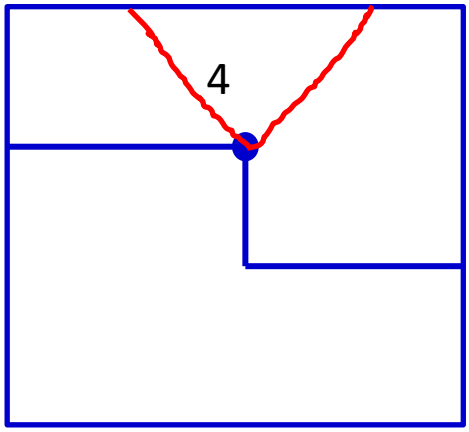


图 2c 绕射波介质模型

图 2 展示了三种勘探地震中常见的波传播和介质模型的示意图。到目前为止，所发展的各种成像方法都包含有针对地下介质分布特征的假设。

图 2a 展示背景+散射点的介质分布模式。这种介质分布假设在医学成像中是合适的，应该用 360 度的包围目标体的封闭面上的激发接收。用逆散射反演成像（即 FWI 反演成像）估计散射强度。在这样的介质展布和激发接收系统情况下，透射波和前向散射波是主要的波现象，用逆散射成像可以同时确定背景和跃变的速度成分。很遗憾，勘探地震中的介质分布特点和激发接收方式不是这样的，逆散射成像并非勘探地震中的首选成像方法。

图 2b 展示了背景+层状界面的介质模式，在勘探地震中广泛存在。仅有地表面上的激发接收，也可以对反（绕）射点进行叠前偏移成像。常规的以定位地下反（绕、散）射点位置为目标的叠前偏移成像方法仅仅靠地表面观测的背向传播的波场就可以达到其目的。但地表观测方式本质上不适合 FWI+LS_RTM 这类逆散射反演成像方法。勘探地震中的透射波（初至波）层析成像也是在不完整观测孔径的基础上进行的，导致初至波反演成像结果的精度很难评价。尽管初至波层析一直在勘探地震中广泛应用，但是我们仅仅用成像结果（或静校正结果）的改善来定性地判断初至波层析结果的精度。

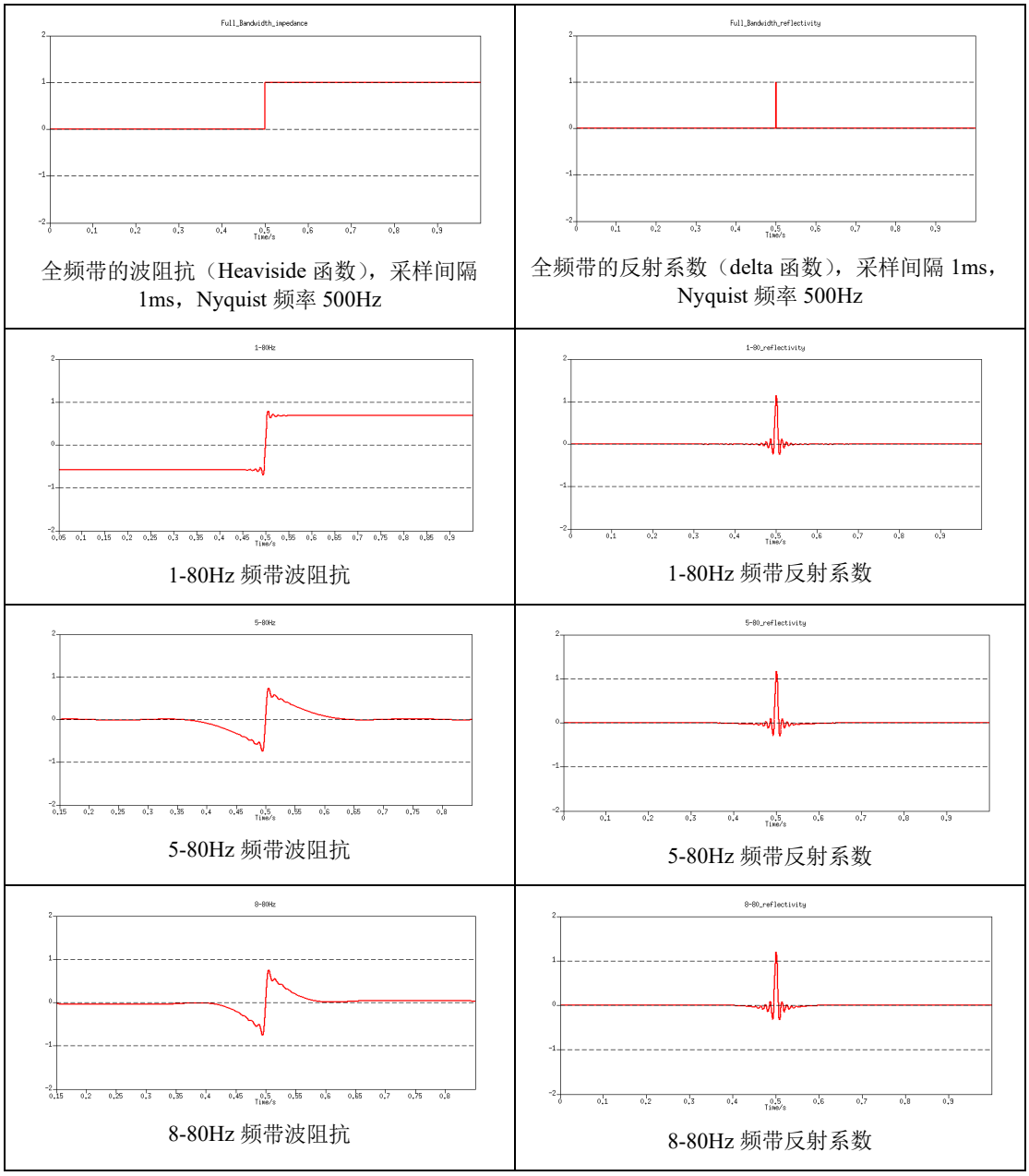
图 2c 展示背景+层状界面间断的介质模式，在勘探地震中也比较普遍。仅有地表激发接收，以对绕射点定位为目标的叠前偏移成像是可行的。但要估计间断点处的绕射强度，仅有背向波场是不够的，还必须要有

其它方向的绕射波场。这一点上，与逆散射成像是一致的。

总之，介质分布特征是决定观测方式和成像方法的根本。勘探地震中的地震波成像受限于仅仅只能在地表面激发接收数据，但受益于地下介质的近乎水平的分层分布特征，从而发展出了一套独特的地震波成像方法技术体系。

3.2 成像所用的波现象的低频特征

当前，宽频地震勘探已经成为勘探地震中的热点关注问题。由于大地介质的吸收作用，50Hz 以上的地震波很难在透过中深层的长偏移距地震记录中存在，所谓的宽频地震勘探主要是拓宽低频端（譬如 1-8Hz）的波场频率成分。这部分波场成分对于提高反演成像的分辨率和精度是至关重要的。



全频带的波阻抗（Heaviside 函数），采样间隔 1ms，Nyquist 频率 500Hz

全频带的反射系数（delta 函数），采样间隔 1ms，Nyquist 频率 500Hz

1-80Hz 频带波阻抗

1-80Hz 频带反射系数

5-80Hz 频带波阻抗

5-80Hz 频带反射系数

8-80Hz 频带波阻抗

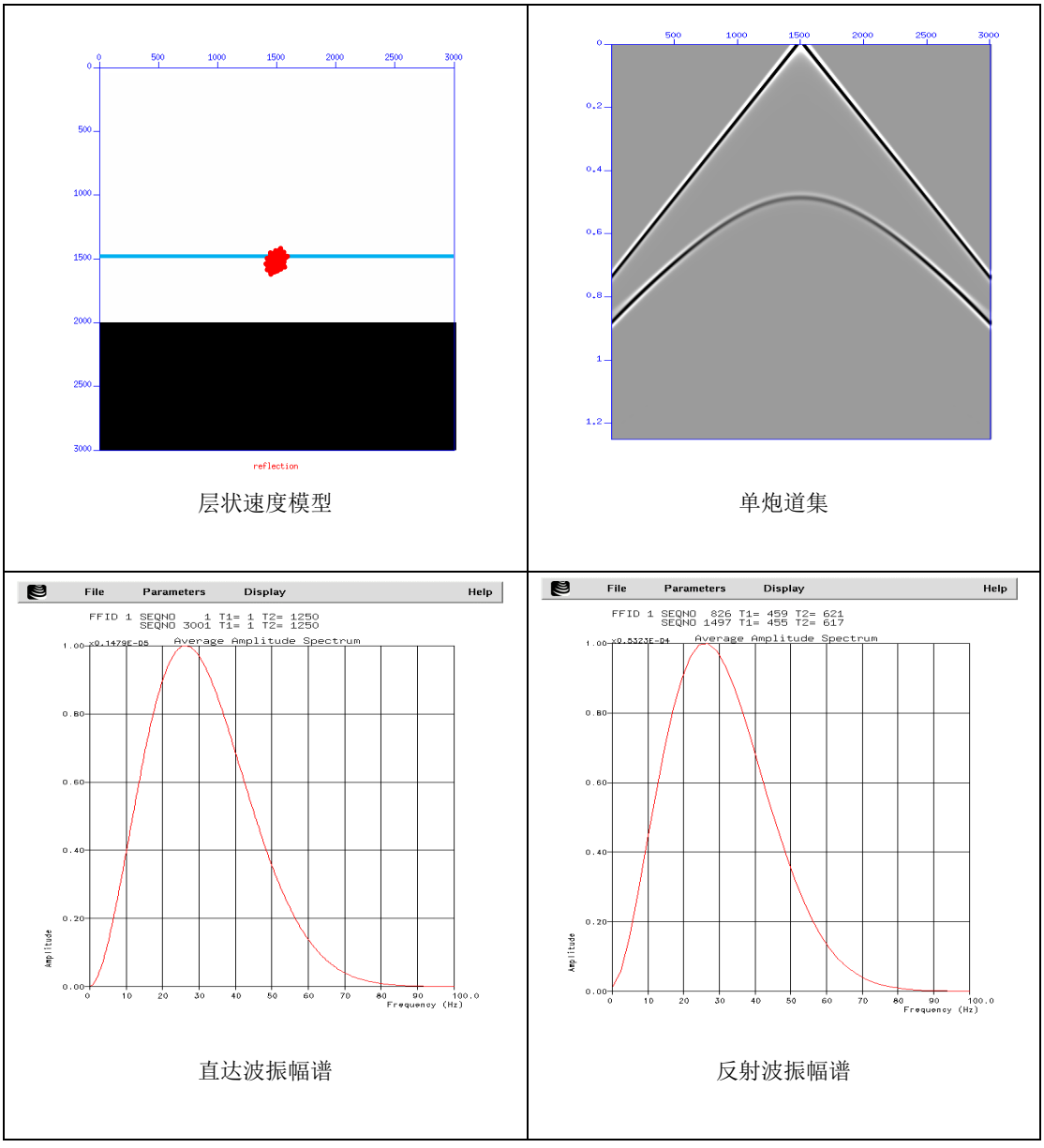
8-80Hz 频带反射系数

图 3 不同频带下的反射系数与波阻抗

图 3 显示了不同频带下反射系数与波阻抗的关系。首先提醒注意的是，此处的频带不是波场中地震子波的频带，而是滤波器的频带。波阻抗是由反射系数进行道积分得到的。但是，若从 1D 褶积系统看，该滤波器就是地震子波，滤波器的频带就是地震子波的频带。在高维叠前地震褶积系统中，成像剖面上的反射系

数与地震子波的关系是复杂的，成像剖面表现为地下介质反射系数与带限滤波器的褶积，该带限滤波器与观测系统孔径、采样、成像方法及地震震源子波都有关系，是它们的综合影响的结果。但仅仅根据 1D 褶积系统进行问题分析，并不影响如下结论的正确性。图 3 首先说明了一个问题，即反演系数估计对低频的依赖性不是很严重，缺乏低频，反射系数结果变化不大，这也是勘探地震中把反射系数成像作为主要目标的原因。波阻抗估计对低频的依赖性要严重得多，缺乏 1-7Hz（甚至 1-4Hz）的低频成分，波阻抗估计的分辨率（即精度）产生严重下降。但是，波阻抗反演结果与（岩性）油气藏识别与描述的关系更密切，得到更准确的波阻抗成像结果，提高油气藏识别和描述的精度才是当前宽带地震的本质含义。

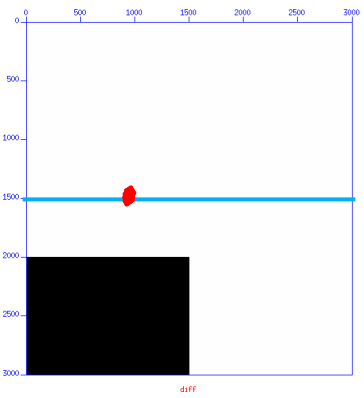
但是，不同的波现象对于低频地震波的响应是不同的。



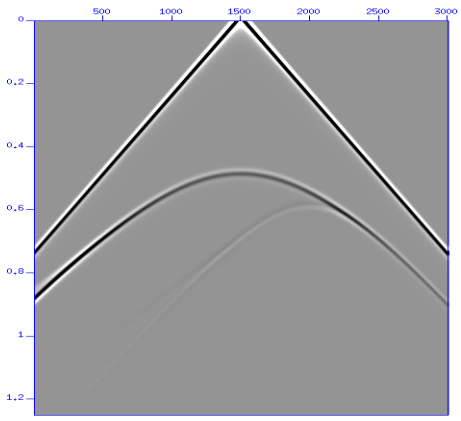
层状速度模型　　单炮道集

直达波振幅谱　　反射波振幅谱

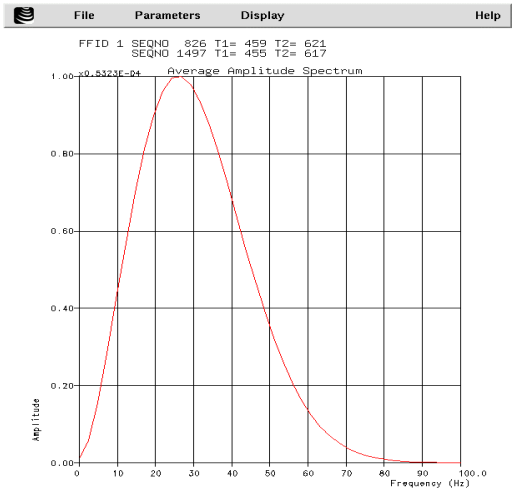
图 4　反射波的低频响应特征



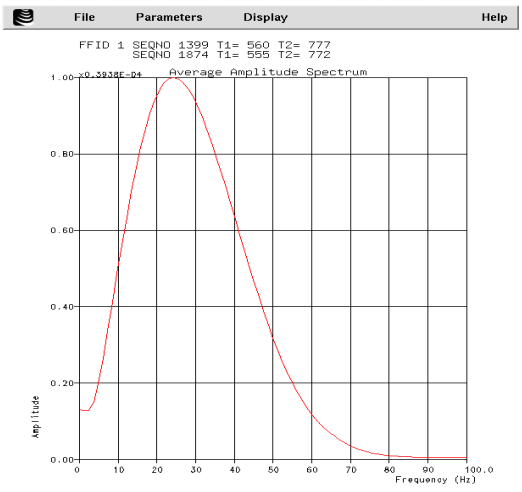
绕射波介质模型



单炮道集

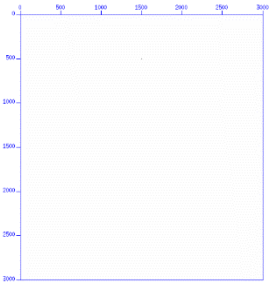


反射波振幅谱

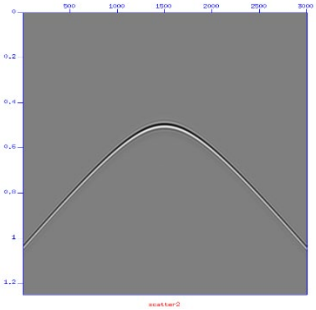


绕射波振幅谱

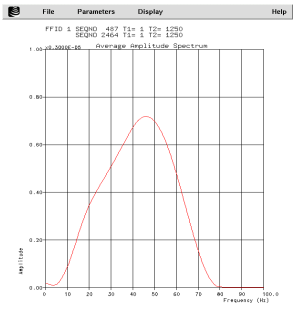
图 5 绕射波的低频响应特征



散射体半径 2m



炮集



振幅谱

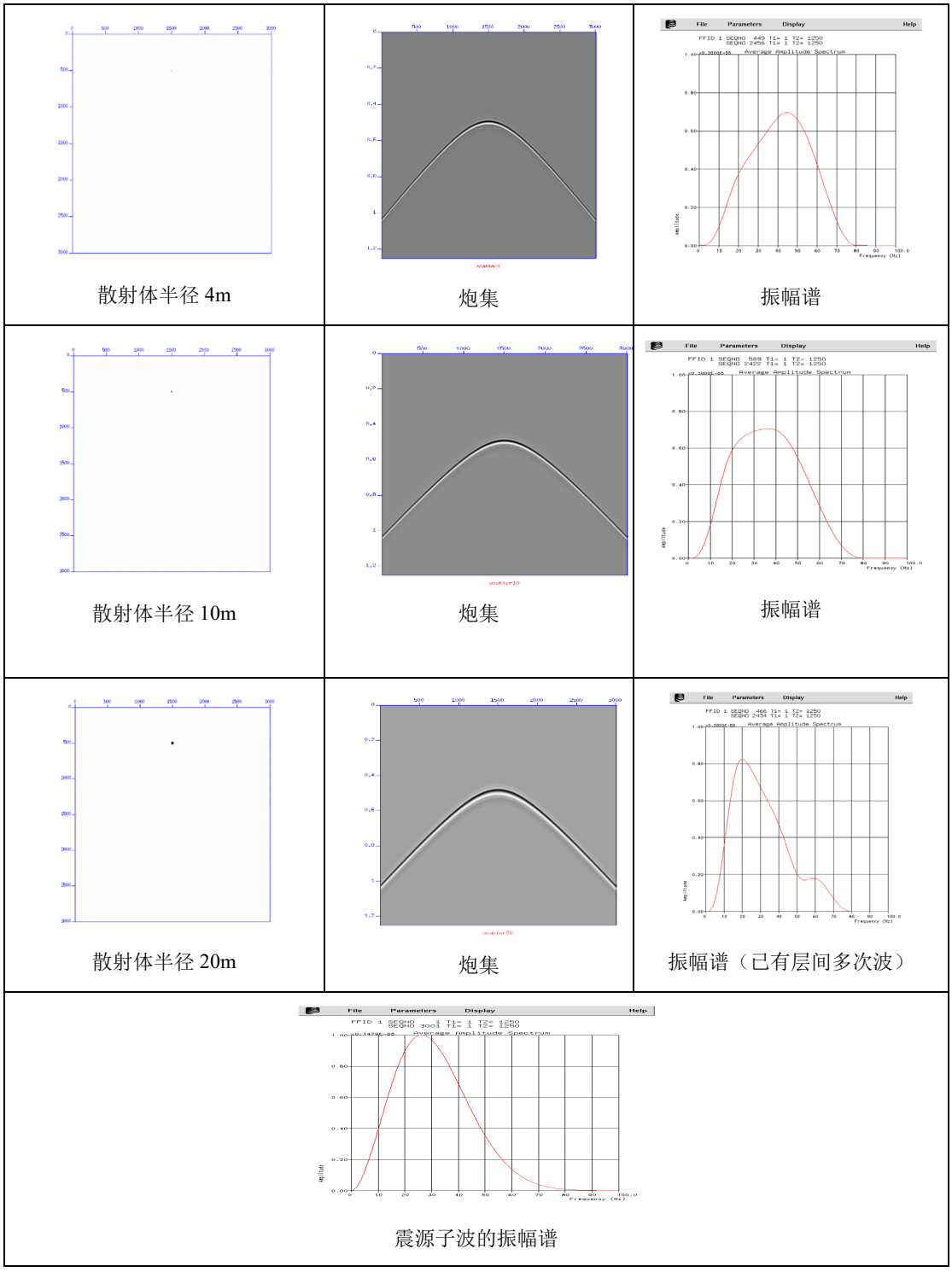


图 6 不同尺度散射体的散射波及振幅谱对比

图 4 到图 6 分别展示了反射波、绕射波和不同尺度散射体产生的散射波对于低频的响应情况。直达波和反射波的振幅谱成分基本一样，与震源激发子波的振幅谱成分相比基本相同，说明直达波和反射波中地震子波的低频成分没有缺失；绕射波低频成分与震源激发子波的振幅谱成分相比有缺失，说明低频波不容易产生绕射，地面检波器接收不到低频绕射波。图 6 中展示的不同尺度散射体的低频响应情况（数值试验参数：模型大小（3001m*3001m）；Dx=Dz=1m，扰动点中心位置（1501m，501m）背景速度 2000m/s，扰动速度 3000m/s。），小尺度散射体对于低频波没有响应，地表检波器接收不到低频的散射波。散射体尺度逐渐变大，地表接收到的散射场中低频波能量逐渐增加。

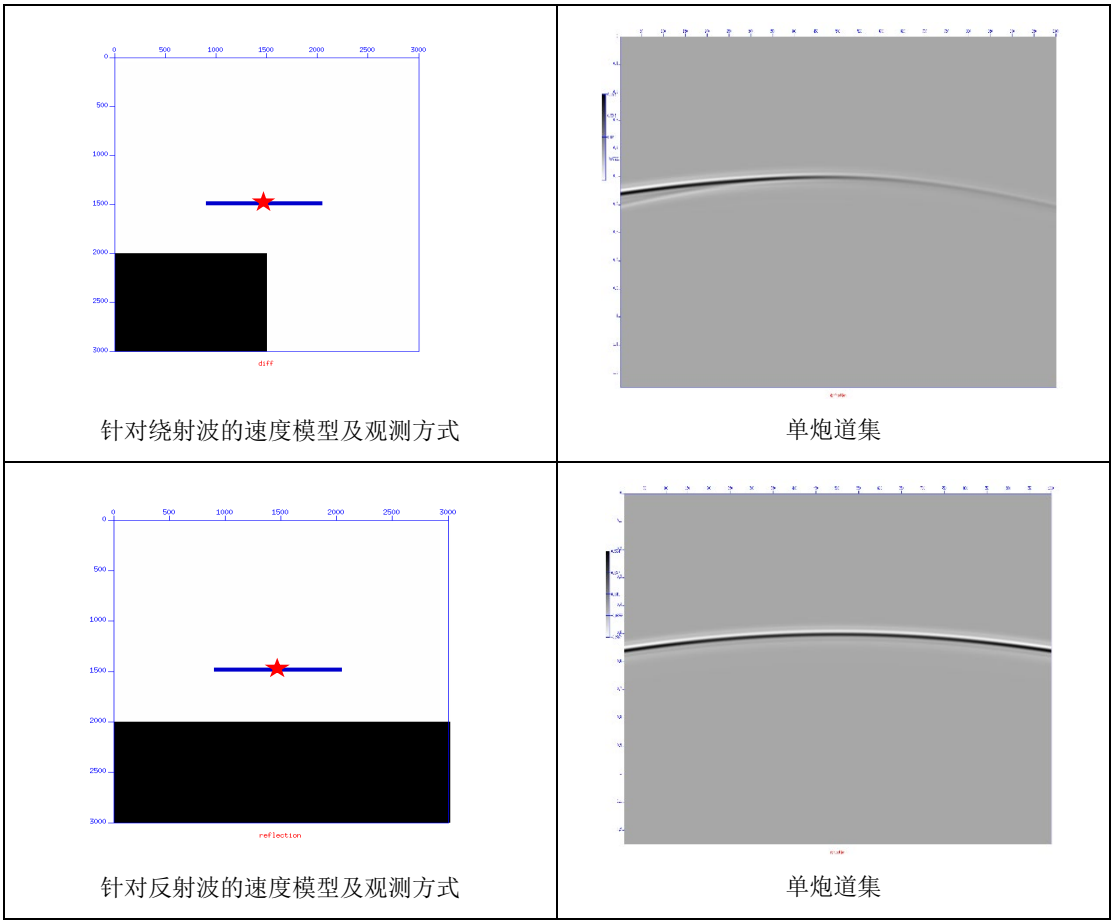
根据物理知识，我们已经知道上述数值试验的结论。但是，此处分析的目的是为了说明地表地震观测对

地震波成像方法技术的影响。宽带地震勘探目前主要是针对反射波进行的，提高反射波成像的分辨率，进而提高波阻抗估计的精度。“宽带”逆散射反演成像的含义还需要进一步探讨。

3.3 勘探地震观测系统下反射波、绕射波与散射波的能量特征分析

当前，深层油气藏地震勘探变得越来越重要，对于深层油气藏勘探应该选择什么样的成像方法，需要对波在介质中的传播特征有更深入的分析和认识。以下分析勘探地震观测系统下（即地表激发接收观测系统下）反射波、绕射波与散射波的能量特征。

为保证能量评价结果的合理性，对数值试验要求为：反射界面上下、绕射体内外、散射体内外速度差异相同；地表激发与接收观测；激发接收方式完全相同；观测空间相同；仅进行一炮试验，中间激发和双边接收。数值试验目的：分析这样的观测方式下，接收到的不同模式下的波场能量的差异。分析存在如此差异情况下，反射波成像和散射波成像方法应该采用的针对性措施。



针对绕射波的速度模型及观测方式　　单炮道集

针对反射波的速度模型及观测方式　　单炮道集

图 7 相同地面观测系统下反射与绕射波能量对比数值试验

图 7 中所采用的观测系统及速度模型：速度模型（震源位置(1501m，1501m)，检波器位置（1001m-2001m，1500m），nx=3001，dx=1m，nz=3001，dz=1m）。

图 8 中所采用的观测系统及速度模型：模型大小（3001m*3001m）；Dx=Dz=1m，扰动点中心位置（1501m，1501m）；背景速度 2000m/s，扰动速度 3000m/s；震源位置（1501m，1501m）检波器在深度 1500m。

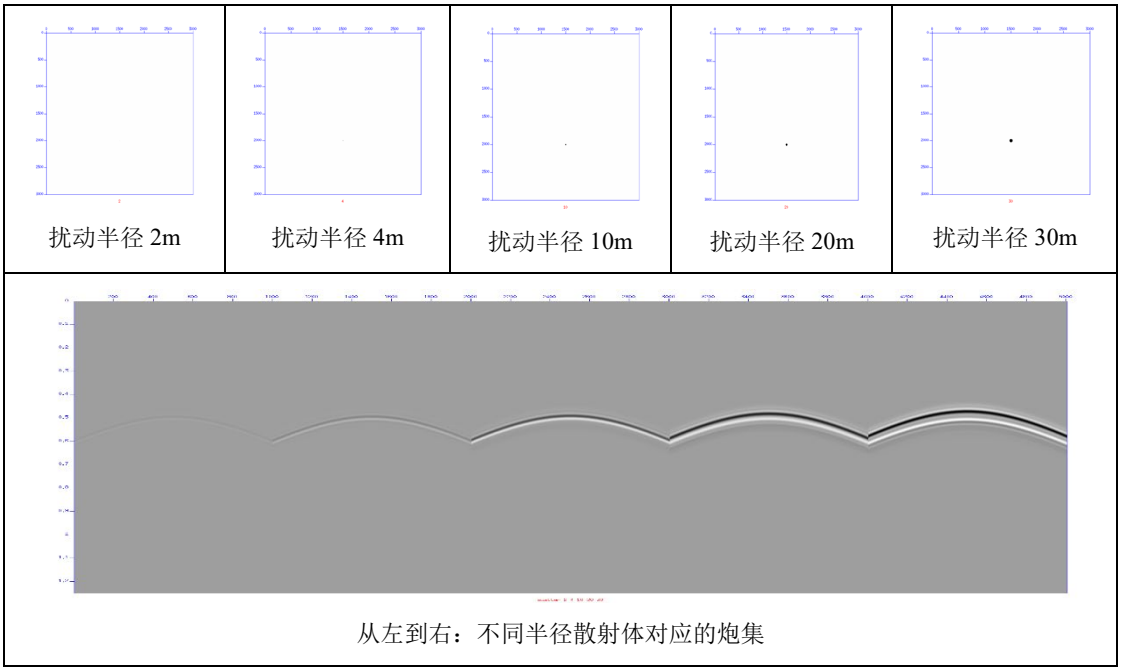


图 8 相同地面观测系统下不同尺度散射体背向散射波能量对比数值试验

从图 7-图 8 可以看出，不同波现象在相同地面观测系统下的能量差异很大，尤其是不同尺度散射体的背向散射能量差异更大。这定性地说明，对不同波现象进行地震波成像时，即使从信噪比上看，难度差异也是很大的。

为了更为定量地展示不同情况下波现象的能量差异，列出下面的表格和直方图。

表 1 不同情况下波现象的能量统计

介质分布模式	同孔径地表观测的能量
reflection	19214.320312
diffraction	1761.16687
scatter_30	1369.446777
scatter_20	874.631714
scatter_10	346.129608
scatter_4	26.222767
scatter_2	1.421453

从表 1 和图 9 展示的定量数据，可以更清楚看出在相同地表观测系统下，不同波现象的背向传播能量的差异。这更清楚地说明，在勘探地震中，为什么一直把反射波作为地震波成像的主要信息来使用。在当前 FWI+LS_RTM 这类逆散射反演成像为主要发展方向的情况下，如何充分地把反射波成像的核心地位凸显出来，融入到 FWI+LS_RTM 成像体系中，显得尤为重要。

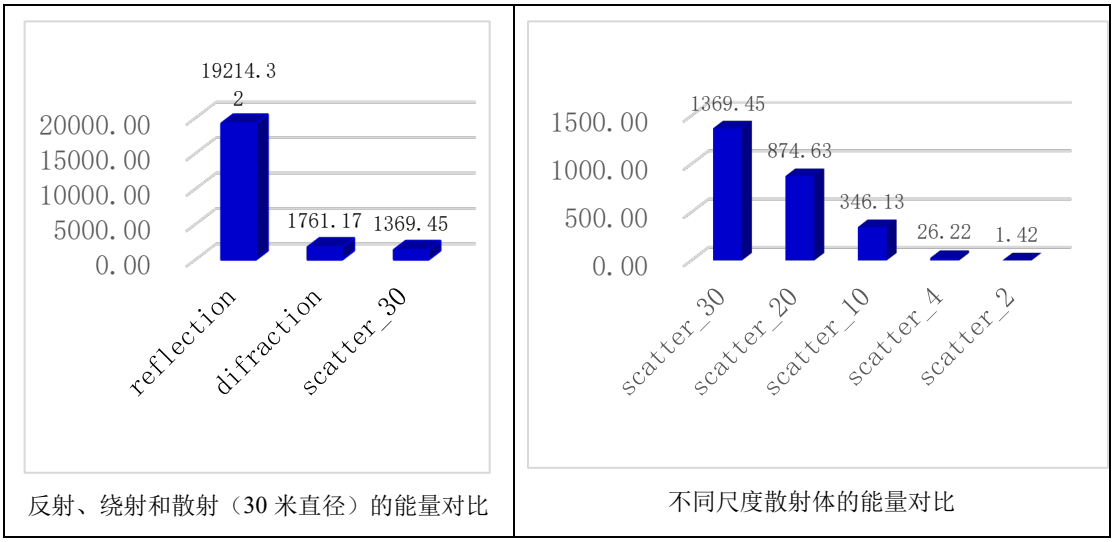


图 9 不同情况下波现象能量对比直方图

3.4 波现象、对应的观测方式与合适的成像方法

前已述及，到目前为止，勘探地震中地震波成像所用的波现象主要是透射波（初/早至波）、一次反射波和一次散（绕）射波。当然，从研究的角度，人们也在探索多次波成像。

从前面试验看出：透射波包含了源子波的全部谱成分；反射波包含了源子波的全部谱成分；散射波中不包含源子波的低频成分；绕射波中不包含源子波的低频成分。因此，宽频地震勘探只是针对透射波和反射波的。低频长偏移距透射波可以利用 FWI 估计更好的背景速度；反射波可以估计更高分辨率的反射系数。二者的结合可以更好地估计波阻抗。因此，到目前为止，地震波成像的核心依然是利用好透射波（初至波）和反射波。而且，从能量的角度看，地表观测系统得到的透射波（初/早至波）和反射波能量强，信噪比高，尤其是初至波具有最高的信噪比。

从地震波成像的角度看，不同的波现象应该有不同的观测系统。透射波（初/早至波）：低频长偏移距的观测+背景参数的层析成像；一次反射波：“两宽一高”观测+角度反射系数成像（可以用反射波层析反演进一步更新背景速度）；散射波：360 度全方位观测+FWI/逆散射成像；绕射波成像：原则上，类似散射波成像。

4 地震波成像的本质观点

4.1 地震波成像的本质问题

地震波成像起源于 Hargdoon（1945）对偏移成像的直观物理解释，即高频近似地震波场沿射线（或 Beam）在背景介质中传播到反射界面，Snell 定律反射到接收点。按绕射叠加的方式（输入道观点成像）和扩散到等时面的方式（输出道观点成像）进行地震波成像。Claerbout（1977）提出激励时间成像条件和波场外推的概念，使得偏移成像可以通过波场反传播的方式在计算机上编程来实现，促进了地震波成像在石油工业界的广泛应用。他提出的 15° 单向波方程外推偏移成像等一系列成像处理概念，为他及他建立的 SEP（斯坦福大学）研究组在石油工业界赢得了盛誉。

但是，这些（无疑是）正确的概念都是基于物理直觉提出的，缺乏严格的数学物理证明。本质上，地震波偏移成像和 Tomography 都可以归结为一个 Bayes 框架下的线性化的最小二乘反演问题（Tarantola，2000）。

4.2 梯度的处理

我们知道，在正问题线性化和残差波场高斯分布的假设下，最小二乘反演问题的核心在于梯度计算，尤其是带 Hessian 逆预条件的梯度的计算。

Bayes 框架下 L_2 范数最小二乘线性化反问题的梯度，本质上，是在提出严谨的正传播波场描述定解问

题和严谨的反传波场描述定解问题，并给出合适的定解问题解法的基础上，得到完整解决的。基于 Claerbout 成像条件的偏移成像之所以被称为定性的成像方法，就是它根本不关心严格的定解问题的求解，只要模拟的入射波场和反射波场在成像点能相关就达到了定位反射/绕射的目的，而且也不用区分反射和绕射情形。

事实上，基于 Claerbout 成像条件的偏移成像根本就不是一个线性化的最小二乘反演问题。一般的 RTM 之所以出现背景成像噪音，本质上也是没有把 RTM 成像问题上升到线性化的反演问题来看待，没有对梯度进行合适的处理。

4.3 基于高频近似正问题的保真成像（反射系数估计）的梯度

Bleistein，Cohen 和 Stockwell（2000）（以及 Beylkin，1985；Xu，Chauris，Lambaré，和 Noble，2001；Zhang，Xu，Bleistein 和 Zhang，2007）提出的保真反射系数的估计，理论上，就是在层状介质、高频近似描述波传播的情形下，解一个 Bayes 框架下的线性化的最小二乘反问题。只是把法方程在拟微分算子理论下进行了显式地推导，得到了估计反射系数的半解析表达式。仅此而已！问题的核心依然是梯度的正确处理。即：在提出严谨的正传播波场描述定解问题和严谨的反传波场描述定解问题，并给出合适的定解问题解法的基础上，得到正确的梯度处理。

仅仅基于 Claerbout 相关成像条件的粗糙处理，不可能得到保真的成像。梯度处理问题的核心在于波传播问题的正确处理，包括正传问题和反传问题。**不能被 Bleistein 等人弄的一堆基于微分算子的反演方法（譬如广义 Radon 逆变换，Beylkin，1985）迷住了我们对本质问题的判断！**

4.4 依据波动理论的线性化反演的梯度处理

前已述及，Bayes 框架下 L_2 范数最小二乘线性化反问题的梯度，本质上，是在提出严谨的正传播波场描述定解问题和严谨的反传波场描述定解问题，并给出合适的定解问题解法的基础上得到完整解决的。因此，波动理论的线性化反演的梯度处理，应该对正传播场的计算和反传播波场的计算提出严谨的定解问题，并给出合适的计算方法。只要这样处理梯度问题，得到的该线性反问题的解就是理论上应该给出的。

适当的分析是必要的，譬如在局部平面波传播框架下进行分析可以解释反射系数估计的各种影响因素，而不是像 LS_RTM 那样是一个黑匣子。当然，Bleistein 等人已经在地下介质层状假设，以及高频近似波传播框架下完整而清晰地说明了反射波成像（估计反射系数）的本质。

5 反射波成像

勘探地震面对的地下介质层状特性、地表激发观测系统、波场中反射波波现象占主导地位的特点，要求我们必须把反射波成像放在核心位置上，即便是 FWI+LS_RTM 已经成为技术研究核心的情况下依然如此。

由反射波成像理论知道，仅用地表观测的地震数据就可以确定地下层状介质反射系数的空间位置，也可以估计地下层状介质的（角度）反射系数的值。但其中包含了若干假设：①地下介质是理想声介质，可以用理想声介质中的波传播方程描述其中的波传播过程；②假设有无穷宽方位和长偏移距的连续的观测；③假设地震子波已知；④假设波传播过程可以仅用高频近似下的波传播理论描述（针对层状介质，高频近似一定是合理的）；⑤假设背景速度已知且正确；⑥假设地下层状介质的反射系数是弱的，满足波场的 Born 近似。

在上述 6 条假设下面，Bleistein（2000）给出了完整的角度反射系数估计的半解析公式。把保真成像的影响因素讲得非常清楚，主要是：几何扩散（包括球面扩散）因素；角度传播因素；采样因素。张宇等（2007）用单向波传播理论重新阐述了 Bleistein（2000）的理论框架，并没有增加新的意义。徐升的导师和徐升（2001）做的工作与前述两位的工作是一致的，仅是描述方式上有所区别。这三项研究工作标志着反射波成像理论的完善。

反射波成像理论可以总结为：在最小二乘成像理论框架下，假设地下介质是层状的（每层都终结于无穷远边界上），假设观测系统是无穷宽方位和长偏移距的、连续的观测（炮检点是对称的！），假设背景速度正确（满足 Born 近似）且子波已知，**方位角度反射系数可以准确地估计出来**，并有一个半解析的表达形式。

到目前为止，勘探地震学的最高理论阶段就是这样。石油工业界中具体的成像处理实践都没有越过此界

限。

反射波成像还有什么研究的空间？我认为可以用平面波传播理论重新描述角度反射系数成像，并建立 PSDM 角度反射系数与线性化的 AVA 反射系数之间的严格的对应关系。这是 WPI 关注的反射波成像的研究方向。

6 散射波成像

FWI（+LS_RTM）成像方法和技术，本质上，是针对散射介质分布和针对散射波场观测的。散射介质的特征：**背景介质中漂浮着局部不相连的波阻抗异常体。**散射波数据特征：散射同相轴为主，不仅仅包含背向散射，还必须有全方位的散射波场。即必须进行全方位的散射波场观测！在包围成像目标体的封闭面上进行激发和接收的观测。很显然，勘探地震的实践不满足 FWI(+LS_RTM)的介质假设和数据预期。

这是 FWI（+LS_RTM）不能很好地应用于勘探地震中的本质原因。当然，学者们还指出了 FWI（+LS_RTM）难于应用的其它非本质原因。譬如：正问题不能模拟所有观测波场中的波现象；震源子波空变且未知；FWI 反演的非线性性强；对初始模型的要求高。

勘探地震中的层状介质特征和地面激发接收的观测方式是勘探地震声成像区别于医学声成像（或与医学成像特点类似的问题）的根本之处。反射地震学的称谓不是没有道理的。

我们对当前 FWI 方法技术在实际中的应用情况，有如下的认识：①FWI 目前仅仅对低频长偏移距的透射波的层析成像是有效果的，所谓的效果也是在不完全数据下获得的，反演结果的精度是很难评价的。②本质上，基于透射波的比较精确的背景速度的估计**也需要全方位的观测数据，仅有地表观测进行透射波反演也存在精度不够高的问题**。③仅仅用背向散射数据的所谓的反射波 FWI 理论上是不成立的。此处的不成立主要是讲理论上不自洽！仅仅用背向散射波不可能完全确定散射强度，**理论上都无法给出估计结果的精度评价！**但是，层状介质中（角度）反射系数估计至少理论上是自洽的（利用地表观测的反射波数据可以得到（角度）反射系数的估计（Bleistein， 2000））！④仅仅用背向传播数据的、所谓的反射波 FWI（=LS_RTM）本质上是在完成反射系数估计的基础上，提高一下散射强度的估计精度而已。

这本质上指出了 FWI 在勘探地震观测系统下存在的根本问题。

因此，在 FWI+LS_RTM 的技术框架下，逻辑上应该重点发展的成像技术应该是：

①初至波（早至波）波动理论到达时/相位层析成像；

②以保真的角度反射系数成像提供初始值的 LS_RTM。

7 总结与讨论

主要利用地面观测的地震波场中的到达时信息进行地震波成像处理的阶段已经发展到以 FWI+LS_RTM 方法为核心的阶段，其中 FWI 主要利用透射波信息估计背景速度；LS_RTM 是一种基于逆散射思想的成像方法，试图估计高波数的波阻抗扰动或速度扰动。FWI 方法本质上是包含了 LS_RTM 方法的。前者理论上是非线性的反演方法，可以估计从背景参数变化到高波数参数变化的所有成分。LS_RTM 是背景参数正确且已知的假设下的线性化反演，可以认为是 FWI 方法的一部分。FWI 方法完满的表达建立在 Bayes 估计理论框架技术上。

但是，完美的反演成像理论并没有带来能在实际中应用的可靠技术，FWI+LS_RTM 的技术组合远没有达到可以在任意数据体上都能产生明显的应用效果的程度。众多文献给出的原因是地震波正演模拟方法不能很好地模拟炮集中的复杂波场、噪音不满足高斯分布、震源子波空变且未知、初始模型的给出不能满足近似线性化反演的要求等等，导致了 FWI+LS_RTM 不能给出满足实际生产要求的成像结果。另外一个重要原因是仅仅在地面（或井中）布设检波器观测背向传播的地震波场，这与 FWI+LS_RTM 方法本身对数据的要求是不匹配的，本质上，FWI+LS_RTM 方法是一种逆散射成像方法，需要观测全向传播（即包围目标成像体的封闭面上）的地震波场，才有可能利用逆散射成像理论得到比较精确的介质参数成像结果。

幸运的是，勘探地震所面对的地下介质情况是比较特殊的，地下介质以横向缓变的层状沉积层为主，夹

杂以由于地球动力运动导致的断层、尖灭、潜山等等异常地质体。波在这样的介质中传播以反射波为主，而不是以散射波为主。正是基于这样的地下介质特征和在这样的介质中传播的地震波场特征，以石油勘探为目标的勘探地震界几十年来发展的技术系列始终以反射波成像为核心。基于这样的技术系列，油气地震勘探获得了长足进步，极大地提升了石油工业的勘探和开发效益。

在以 FWI+LS_RTM 为核心的逆散射成像方法技术发展受阻的情况下，进一步的地震波成像技术发展的方向是什么？首先当然是不断地提升地震数据采集技术，尽可能得到“两宽一高”的地震数据体。第二是充分挖掘透射波层析成像的潜力，使背景速度估计结果尽可能准确。从勘探地震技术发展过程看，以前对透射波在勘探地震中的利用是不够的。第三，在对保真（或真振幅）成像充分重视的基础上，以保真（或真振幅）成像结果作为初始值，用逆散射成像（LS_RTM）进一步提升高波数参数扰动量的估计精度是符合逻辑的。不能一味地用基于振幅逼近的 Bayes 反演方法期望得到比较精确的高波数参数扰动量的估计结果。

总之，基于所要成像的介质的分布特征和在其中传播的波场的特征构建合适的地震数据采集方法技术和成像处理方法技术系列是勘探地震中成像技术发展的要义。

参考文献

[1] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, 26, 99–108.

[2] Bleistein, N., Cohen, J. K., and Stockwell, J. W., Jr, 2000, Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion, Springer

[3] Claerbout, J. F., 1972, Downward Continuations of Moveout Corrected Seismograms. Geophysics 37, 741-768

[4] Devaney, A. J., 2012, Mathematical Foundations of Imaging, Tomography and Wavefield Inversion, Cambridge University Press

[5] Hagedoorn, J. G., 1954, A Process of Seismic reflection interpretation. Geophysical Prospecting 2, 85-127

[6] Stolt, R. H., and Weglein, A., B., 2012, Seismic Imaging and Inversion – Application of Linear Inverse Theory, Cambridge University Press

[7] Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM.

[8] Troyan, V., and Kiselev, Y., 2010, Statistical Methods of Geophysical Data Processing, Word Scientific Publishing Co.

[9] Xu, S., H. Chauris, G. Lambaré, and M. Noble, 2001, Common-angle migration: a strategy for imaging complex media: Geophysics, 66, 1877‐1894.

[10] Zhang, Y., S. Xu, N. Bleistein, and G. Zhang, 2007, True amplitude angle domain common image gathers from one-way wave equation migrations: Geophysics, 72, no. 1, S49–S58.

勘探地震中应用泛函分析基础知识总结

王华忠

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】勘探地震中地震波反演成像已经全面进入到以 Bayes 参数反演为理论基础的发展阶段。Hagedoorn（1954）和 Claerbout（1977）基于地震波传播的几何物理直观提出的一套地震波偏移成像方法理论，在油气勘探地震领域中，甚至对整个石油工业，起到了巨大的作用。但是，常规意义下的、以定位反射（绕射）源位置为目标的地震波偏移成像方法研究已经走到尽头，尽管目前实际油气勘探地震的日常数据处理中还在大量应用这样的技术。以 FWI+LS_RTM 为标志的地震波反演成像已经成为当前地震波成像技术发展的核心，它以估计地下岩石介质性质的背景变化（低波数变化）和扰动变化（高波数变化）为目标。目前看来，Bayes 估计理论是地震波反演成像的最完美的理论框架。从概率统计意义下，把模型参数变化空间、数据采样空间、状态变量空间有机地联系在一起，给出了反演解的比较完整的解释。在系统线性和噪音高斯分布假设下，Bayes 估计理论很好地兼容了确定性反演理论，而确定性反演理论从解析的角度更好地分析了反演解的稳定性问题。地震波反演成像问题（事实上，所有的参数反演问题）都可以在泛函分析的理论框架下给出简洁而明晰地表达，尤其是用算子理论表达的情形下。本文试图给出地震波反演成像所需的泛函分析框架的简洁回顾，为 WPI 研究组更好地开展反演成像工作起到提纲挈领的引领作用。

【关键词】地震波偏移成像，反演成像，Beyes 反演理论，泛函分析，算子理论

1 引言

Hagedoorn（1954）和 Claerbout（1977）基于地震波传播的几何物理直观提出的一套地震波偏移成像方法理论。他们基本是在高频近似地震波传播理论及地震波在层状地下介质中传播的概念下提出了地震波偏移成像理论。核心是提出了反射（也包括散射）波成像的成像条件：入射波到达时等于反（出）射波的出发时。依据这个成像条件，与波传播的几何光学近似相结合，构建了整合几何学的理论和技术体系。其中的核心技术环节包括：NMO 叠加与速度分析，NMO+DMO 叠加+速度分析，CRS 叠加+速度分析，PSTM+速度分析，Kirchhoff 积分 PSDM+速度分析（成像道集层析）。这样的技术流程，到目前为止，还在统治着日常的地震波成像处理过程。

但是，这种以定位反射（绕射）源位置为目标的地震波偏移成像方法研究已经走到尽头。

以 FWI+LS_RTM 方法理论备受重视为标志，说明基于 Bayes 估计理论的地震波反演成像方法技术研究已经成为当前地震波成像技术发展的核心。地震波反演以估计地下岩石介质性质的背景变化（低波数变化）和扰动变化（高波数变化）为目标。

Bayes 估计理论是地震波反演成像的最完美的理论框架。它从概率统计意义下把模型参数变化空间、数据采样空间、状态变量空间有机地联系在一起，给出了反演解的比较完整的解释。在系统线性和噪音高斯分布假设下，Bayes 估计理论很好地兼容了确定性反演理论。以 Tikhonov 正则化反演理论为代表的一套确定性反演理论框架，很好地解释了线性系统和预测噪音为高斯白噪情形下的解的性质。

地震波反演成像问题（事实上，所有的参数反演问题）都可以在泛函分析的理论框架下给出简洁而明晰地表达，尤其是用算子理论表达的情形下。本文试图给出地震波反演成像所需的泛函分析框架的简洁回顾，为 WPI 研究组更好地开展反演成像工作起到提纲挈领的引领作用。

2 应用泛函分析简明基础

勘探地震中，最基本的正问题由如下算子表达的线性问题来表示：

$$L\mathbf{m}=\mathbf{d} \tag{2.1}$$

其中，L 代表一类线性算子，对应波传播算子；$\mathbf{m}$ 代表模型空间中的一类线性函数，代表要估计的模

型参数；$\mathbf{d}$ 代表数据空间中的一类线性函数，是某种观测数据（或预测的数据）。（2.1）式也代表由算子 L 把模型空间中的函数 $\mathbf{m}$ 映射到数据空间中的函数 $\mathbf{d}$。构造一类泛函

$$E(\mathbf{m})=\left\|\mathbf{d}^{obs}-L\mathbf{m}\right\|_2^2 \tag{2.2}$$

并最优化求解模型参数 $\hat{\mathbf{m}}=\lim\limits_{\mathbf{m}\in M}\left\|\mathbf{d}^{obs}-L\mathbf{m}\right\|$ 是勘探地震中的核心问题。

上述问题的求解，需要扎实的泛函分析基础知识。泛函分析最基础部分包括线性空间、线性赋范空间和内积空间的性质。其中，线性函数的基函数表达理论和范数概念是特别重要的。

2.1 可测集与可测函数

2.1.1 可测集与可测函数

定义 2.1.1 设 X 是基本空间，$\boldsymbol{R}$ 是 X 上的 $\sigma-$代数，且 $X=\bigcup\limits_{E\in\boldsymbol{R}}E$，则称 $(X,\boldsymbol{R})$ 是可测空间（measurable space），$\boldsymbol{R}$ 中的元素 E 是 $(X,\boldsymbol{R})$ 上的可测集。特别地，当 $X=\mathbf{R}^1$，$\boldsymbol{R}=\boldsymbol{L}$ 时，称 $(\mathbf{R}^1,\boldsymbol{L})$ 是 Lebsgue 可测空间；Lebsgue 可测空间上的可测集称为 Lebsgue 可测集；当 $X=\mathbf{R}^1$，$\boldsymbol{R}=\boldsymbol{S}(\boldsymbol{R}_0)=\boldsymbol{B}$ 时，称 $(\mathbf{R}^1,\boldsymbol{B})$ 是 Borel 可测空间；Borel 可测空间上的可测集（即：Borel 集）称为 Borel 可测集。

定义 2.1.2 设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，f 是定义在 E 上的有限实函数。若对一切实数 c，集 $E(c\le f)=\{x\,|\,c\le f(x),\ x\in E\}$ 都是 $(X,\boldsymbol{R})$ 上的可测集（即：$E(c\le f)\in\boldsymbol{R}$），则称 f 是 E 上关于 $\boldsymbol{R}$ 的可测的函数，简称 E 上的可测函数。特别地，当 $(X,\boldsymbol{R})=(\mathbf{R}^1,\boldsymbol{L})$ 时，称 f 是 E 上关于 $\boldsymbol{L}$ 的 Lebsgue 可测函数；当 $(X,\boldsymbol{R})=(\mathbf{R}^1,\boldsymbol{B})$ 时，称 f 是 E 上关于 $\boldsymbol{B}$ 的 Borel 可测函数。

定理 2.1.1 设 $(X,\boldsymbol{R})$ 是可测空间，f 是定义在 $E\subset X$ 上的有限实函数。则 f 是 E 上的可测函数的充分必要条件是：对任意实数 $c,\ d$，集 $E(c\le f<d)$ 是可测集。

证明：设 f 是可测函数，由于 $E(c\le f<d)=E(c\le f)-E(d\le f)$，而 $E(c\le f)$ 和 $E(d\le f)$ 都是可测集，所以 $E(c\le f<d)$ 是可测集。

反之，若已知对任意实数 $c,\ d$，集 $E(c\le f<d)$ 是可测集，则由 $E(c\le f)=\bigcup\limits_{n=1}^{\infty}E(c\le f<c+n)$ 立即得 $E(c\le f)$ 是可测集。

例 2.1.1 定义在闭区间 $E=[a,\ b]$ 上的任何一个连续函数 f 都是 E 上的 Lebsgue 可测函数。

证明：对任意实数 c，由 f 的连续性，集 $E(c\le f)=\{x\,|\,c\le f(x),\ x\in[a,\ b]\}$ 是 $[a,\ b]$ 中的闭集，因此 $E(c\le f)$ 是可测集；故 f 都是 $[a,\ b]$ 上的可测函数。

例 2.1.2 设函数 f 定义在 $E=(-\infty,\ \infty)$ 上，$\langle a_i,\ b_i\rangle\ (i=1,2,\cdots,n)$ 是一组互不相交的区间，函数

$$f(x)=\begin{cases}\alpha_i, & x\in\langle a_i,\ b_i\rangle\ (i=1,2,\cdots,n),\\ 0, & x\in(-\infty,\ \infty)-\bigcup\limits_{i=1}^{n}\langle a_i,\ b_i\rangle\end{cases}$$

称为阶梯函数，它是 E 上的 Lebsgue 可测函数。

证明：因为对任意实数 c，$E(c\le f)=\{x\,|\,c\le f(x),\ x\in(-\infty,\infty)\}$ 或是全直线，或是空集，或是有限个区间的并，而这些都是 Lebsgue 可测集，所以 f 是 $(-\infty,\infty)$ 上的可测函数。

例 2.1.3 设 $(X,\boldsymbol{R})$ 是可测空间，$E,\ E_i\in\boldsymbol{R}$，$i=1,2,\cdots,n$，$\bigcup_{i=1}^{\infty}E_i\subset E$，且 $E_i\cap E_j=\varnothing,\ i\neq j$，$f$ 是定义在 E 上的函数，且

$$f(x)=\begin{cases}\alpha_i, & x\in E_i\ (i=1,2,\cdots,n),\\ 0, & x\in X-\bigcup_{i=1}^{n}E_i,\end{cases}$$

则 f 是 E 上的可测函数。

例 2.1.4（不可测函数的例） $(\mathbf{R}^1,\boldsymbol{L})$ 是 Lebsgue 可测空间，Z 是 Lebsgue 不可测集，$f(x)$ 是 Z 的特征函数 $\chi_Z(x)$，$x\in\mathbf{R}^1$。因为 $\mathbf{R}^1\left(\frac{1}{2}\le\chi_Z\right)=\left\{x\,|\,\chi_Z(x)\ge\frac{1}{2},\ x\in\mathbf{R}^1\right\}=Z$ 是 Lebsgue 不可测集，所以函数 $\chi_Z(x)$ 不是 $\mathbf{R}^1$ 上的 Lebsgue 可测函数。

例 2.1.5 也有这样的可测空间 $(X,\boldsymbol{R})$，定义在 X 上的所有函数都是可测函数。例如，取 $\boldsymbol{R}=2^X$（此时 $\boldsymbol{R}$ 是一个 $\sigma-$ 代数），f 是定义在 X 上的任意一个有限实函数，对任意实数 c，显然 $X(c\le f)=\{x\,|\,c\le f(x),\ x\in X\}\in\boldsymbol{R}$，故 f 是 X 上的可测函数。

2.1.2 可测函数的性质

定理 2.1.2 设 $(X,\boldsymbol{R})$ 是可测空间，f 是定义在 $E\subset X$ 上的有限实函数，则

（1） 若 f 是 E 上的可测函数，则 E 必是可测集；反之不然。

（2） 若 f 是 E 上的可测函数，$E_1\subset E$ 可测，当 f 作为 E_1 上的函数时，f 是 E_1 上的可测函数；

（3） 设 $E_1\cap E_2=\varnothing,\ E_1\cup E_2=E$，若 E_1，E_2 是可测集，则 f 是 E 上的可测函数的充分必要条件是：f 是 E_1，E_2 上的可测函数。

（4） 集 E 是可测集的充分必要条件是：集 E 的特征函数 $\chi_E(x)$ 是 X 上可测函数。

证明：（1）因为 $E=\bigcup_{n=1}^{\infty}E(-n\le f)$，而根据可测函数的定义，集 $E(-n\le f)$ 是可测集，所以 E 是可测集。

反之不然。因为对 $\forall E\in\boldsymbol{L}$ 且 $m(E)\neq0$，都存在 $F\subset E,\ F\notin\boldsymbol{L}$. 若 $E\in\boldsymbol{L}$，其任意子集都 $\in\boldsymbol{L}$，则 $m(E)=0$.

（2）对任意实数 c，由于 $E_1(c\le f)=E(c\le f)\cap E_1$，而 $E(c\le f)$ 和 E_1 都是可测集，所以 $E_1(c\le f)$ 是可测集，即 f 作为 E_1 上的函数时，它是 E_1 上的可测函数。

（3）设 f 是 E 上的可测函数，由（2）知：f 是 E_1，E_2 上的可测函数。反之，若 f 是 E_1，E_2 上的可测函数，对任意实数 c，由于 $E(c\le f)=E_1(c\le f)\bigcup E_2(c\le f)$，所以 $E(c\le f)$ 是可测集，即 f 作为 E 上的

可测函数。

（4）必要性：设集 E 是可测集。因为 $X(c\le\chi_E(x))=\begin{cases}\varnothing, & c>1,\\ E, & 0<c\le 1,\\ X, & c\le 0,\end{cases}$，而 $\varnothing,E,X$ 都是可测集，所以 $\chi_E(x)$ 是 X 上的可测函数。

充分性：设 $\chi_E(x)$ 是 X 上的可测函数。由上面的式子知，当 $0<c\le 1$ 时，$E=X(c\le\chi_E(x))$。而 $\chi_E(x)$ 是 X 上的可测函数，故 $X(c\le\chi_E(x))$ 是可测集，即 E 是可测集。

注：性质（3）可以推广到有限个或可列个可测集 $E_1,E_2,\cdots,E_n,\cdots$，并且 $E_i\bigcap E_j\neq\varnothing$ 的情况。

定理 2.1.3 设 $(X,\boldsymbol{R})$ 是可测空间，f 是定义在 $E\subset X$ 上的有限实函数，则下面三个条件中的任何一个都是 f 是 E 上的可测函数的充分必要条件：

（1） 对任意实数 c，$E(c<f)$ 是可测集；

（2） 对任意实数 c，$E(f\le c)$ 是可测集；

（3） 对任意实数 c，$E(f<c)$ 是可测集。

定理 2.1.4 设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，f,g 都是 E 上的可测函数，则

（1） 对任意实数 α，αf 是 E 上的可测函数；

（2） $f+g$ 是 E 上的可测函数；

（3） $f\cdot g$ 及 f/g（对 $\forall x\in E,\ g(x)\neq 0$）是 E 上的可测函数；

（4） $\max(f,g),\ \min(f,g)$ 都是 E 上的可测函数。

推论 1 设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，$f_1,f_2,\cdots,f_n$ 都是 E 上的可测函数，则对任意实数 $\alpha_1,\alpha_2,\cdots,\alpha_n$，$\alpha_1f_1+\alpha_2f_2+\cdots+\alpha_nf_n$ 是 E 上的可测函数。

推论 2 设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，f 是 E 上的可测函数，则 $|f|$ 是 E 上的可测函数。

2.1.3 可测函数的极限

定理 2.1.5 设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，若 $\{f_n\}$ 是 E 上的一列可测函数，则当 $\{f_n\}$ 的上确界函数、下确界函数、上限函数、下限函数分别是有限函数时，它们都是 E 上的可测函数。

推论 设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，若 $\{f_n\}$ 是 E 上的一列有限的可测函数，若对一切 $x\in E$，$\lim\limits_{n\to\infty}f_n$ 存在，而且有有限值，则极限函数 $\lim\limits_{n\to\infty}f_n$ 是 E 上的可测函数。

定理 2.1.6 设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，若 f 是 E 上的有限可测函数，则必存在一列 $\{f_n\}$，每个 f_n 是可测集的特征函数的线性组合，使得 $\{f_n\}$ 在 E 上处处收敛于 f。

注：定理 2.1.6 说明：用可测集的特征函数的线性组合可以逼近可测函数。

推论：设 $(X,\boldsymbol{R})$ 是可测空间，$E\subset X$，若 f 是 E 上有界的可测函数，则必存在可测集的特征函数的线性组合的函数序列 $\{f_n\}$，使得 $\{f_n\}$ 在 E 上一致收敛于 f。

注： f 是 E 上的有界函数是指：$\exists M>0$，对 $\forall x\in E$，都有 $|f(x)|\le M$。

f 是 E 上的有限函数是指：$\forall x\in E$，都有 $|f(x)|<\infty$. 即：函数值都是有限实数的函数称为**有限函数**。显然有界函数是有限函数，反之则不然。例如：$f(x)=1/x$ 在 (0, 1) 内的任意函数值都是有限的，但它是 (0, 1) 内的无界函数。

2.1.4 Lebsgue 积分及其性质

定义 2.1.3 设 $(X, \boldsymbol{R}, \mu)$ 是测度空间，E 是一个可测集，$\mu(E)<\infty$，f 是定义在 E 上的可测函数，设 f 是有界的（即：存在实数 c, d，使得 $f(E)\subset(c,d)$），在 $[c, d]$ 中任取一分点组 $D: c=l_0<l_1<\cdots<l_{n-1}<l_n=d$。记 $\delta(D)=\max\limits_{1\le i\le n}(l_i-l_{i-1})$，$E_i=E(l_{i-1}\le f<l_i)$ 并任取 $\xi_i\in[l_{i-1}, l_{i-1}]$ $(i=1,2,\cdots,n)$，作和式 $S(D)=\sum\limits_{i=1}^{n}\xi_i\cdot\mu(E_i)$，称它为 f 在分点组 D 下的一个“和数”。若存在数 s，它满足如下条件：对 $\forall\varepsilon>0, \exists\delta>0$，使得对任意分点组 D，当 $\delta(D)<\delta$ 时，有 $|S(D)-s|<\varepsilon$（即：$s=\lim\limits_{\delta(D)\to 0}S(D)$），则称 f 在 E 上关于测度 μ 是可积（分）的，并称 s 是 f 在 E 上关于测度 μ 的积分，记作 $s=\int_E f\,\mathrm{d}\mu$。特别地，当测度空间 $(X, \boldsymbol{R}, \mu)$ 是 Lebsgue 测度空间 $(\mathbf{R}^1, \boldsymbol{L}, m)$，$f$ 关于测度 m 可积时，称 f 是 Lebsgue 可积函数；称 s 是 f 在 E 上关于测度 m 的 Lebsgue 积分，记作 $(L)\int_E f\,\mathrm{d}x$.通常就简记作 $\int_E f\,\mathrm{d}x$.当 $E=[a, b]$ 时，Lebsgue 积分又记作 $\int_a^b f\,\mathrm{d}x$。

定理 2.1.7 设 $(X, \boldsymbol{R}, \mu)$ 是测度空间，$E\in\boldsymbol{R}$，且 $\mu(E)<\infty$，则 E 上的一切有界可测函数 f（关于测度 μ）必是可积的。

定义 2.1.4 设 $E\subset\mathbf{R}^1$，f 是 E 上的函数。设 $x_0\in E$，若对 $\forall\varepsilon>0, \exists\delta>0$，当 $x\in E$，且 $|x-x_0|<\delta$ 时，有 $|f(x)-f(x_0)|<\varepsilon$，则称 x_0 是 f 的连续点；若 E 中每个点都是 f 的连续点，则称 f 是 E 上的连续函数。E 的孤立点都是 E 的连续点。

定理 2.1.8 设 $E\subset\mathbf{R}^1$ 是 Lebsgue 可测集，f 是 E 上的 Lebsgue 可测函数，则对 $\forall\varepsilon>0$，必有闭集 F_ε，使得 $m(E-F_\varepsilon)<\varepsilon$，且 f 是 F_ε 上的连续函数。

定义 2.1.5 设 $E\subset X$ 是测度空间 $(X,\boldsymbol{R},\mu)$ 的测度 $\sigma-$有限的集，$\{f_n\}$ 是 E 上一列函数，如存在 E 上的非负函数 g，使得在 E 上，$|f_n|\dot{\le}g$（即：在 E 上 $|f_n|$ 几乎处处不大于 g）对一切 n 成立，则称 g 是 $\{f_n\}$ 的控制函数。

定理 2.1.9（Lebsgue 的控制收敛定理） 设 $\{f_n\}$ 可测集 E 上的一列可测函数，g 是 $\{f_n\}$ 的一个可积的控制函数（即：在 E 上有 $|f_n|\dot{\le}g$，$n=1,2,3,\cdots$，而 g 在 E 上可积）。若 $\{f_n\}$ 依测度 μ 收敛于可测函数 f，则 f 在 E 上可积，且 $\lim\limits_{n\to\infty}\int_E f_n\,\mathrm{d}\mu=\int_E f\,\mathrm{d}\mu$。

定理 2.1.10（Lebsgue 的控制收敛定理） 设 $\{f_n\}$ 可测集 E 上的一列可测函数，g 是 $\{f_n\}$ 的一个可积的控制函数。若 $\{f_n\}$ 几乎处处收敛于可测函数 f，则 f 在 E 上可积，且 $\lim\limits_{n\to\infty}\int_E f_n\,\mathrm{d}\mu=\int_E f\,\mathrm{d}\mu$。

推论 1 设 E 是可测集，$\mu(E)<\infty$，$\{f_n\}$ 是 E 上的一列可测函数，且存在非负常数 K，使得 $|f_n|\dot{\le}K$，$n=1,2,3,\cdots$。若 $\{f_n\}$ 几乎处处收敛于（或：依测度 μ 收敛于）可测函数 f，则 $\lim\limits_{n\to\infty}\int_E f_n\,\mathrm{d}\mu=\int_E f\,\mathrm{d}\mu$。

推论 2 设$(X,\boldsymbol{R},\mu)$是完全测度空间，E是可测集，$\mu(E)<\infty$，$\{f_n\}$是E上的一列可测函数，g是$\{f_n\}$的一个可积的控制函数。若$\{f_n\}$几乎处处收敛于（或：依测度μ收敛于）函数f，则f在E上可积，且$\lim\limits_{n\to\infty}\int_E f_n d\mu=\int_E f\,d\mu$.

2.2 函数空间理论基础

2.2.1 线性空间

定义 2.2.1 设S是一非空集合，P是一数域。如果能规定两种对应法则，一种（加法）法则使任意的$\mathbf{x},\mathbf{y}\in S$有$S$中唯一元素---记为$\mathbf{x}+\mathbf{y}$---与之对应，而且满足下面的四个条件：

⑴ $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x},\quad \forall\mathbf{x},\mathbf{y}\in S$

⑵ $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z}),\quad \forall\mathbf{x},\mathbf{y},\mathbf{z}\in S$

⑶ S中存在唯一的一个元素记作$\mathbf{0}$，成为零元素，使得

$\mathbf{x}+\mathbf{0}=\mathbf{x},\quad \forall\mathbf{x}\in S$

⑷每个$\mathbf{x}\in S$于S中存在唯一的一个元素记作$-\mathbf{x}$，称为$\mathbf{x}$的负元素，使得

$\mathbf{x}+(-\mathbf{x})=\mathbf{0},\quad \forall\mathbf{x}\in S$

另一种（乘法）法则使任意的$\alpha\in P$与任意的$\mathbf{x}\in S$有S的唯一元素---记作$\alpha\cdot\mathbf{x}$---与之对应，而且满足下面的四个条件：

⑸ $1\cdot\mathbf{x}=\mathbf{x},\quad \forall\mathbf{x}\in S$

⑹ $(\alpha\beta)\cdot\mathbf{x}=\alpha(\beta\cdot\mathbf{x}),\quad \forall\mathbf{x}\in S,\alpha,\beta\in P$

⑺ $\alpha(\mathbf{x}+\mathbf{y})=\alpha\cdot\mathbf{x}+\beta\cdot\mathbf{y},\quad \forall\mathbf{x},\mathbf{y}\in S,\alpha\in P$

⑻ $(\alpha+\beta)\cdot\mathbf{x}=\alpha\cdot\mathbf{x}+\beta\cdot\mathbf{x},\quad \forall\mathbf{x}\in S,\alpha,\beta\in P$

前一规则称定义在集合S上的加法；后者称为定义在集合S上的乘法；统称为定义集合S和数域P上的线性运算或线性空间结构。集合S、数域P及“加”“乘”规则一起，称为集合S在数域P上的线性空间（或向量空间）。

集合S中的元素一般为向量，也可以是矩阵和函数。数域P可以是实数，也可以是复数。

线性空间的例子有很多。勘探地震中，所有可能的地震波场，就构成一个线性空间。

定义 2.2.2 设集合S是数域P上的线性空间。“+”与“·”是线性运算，V是S的一个非空子集。若集合V对“+”与“·”运算是封闭的，即

$\alpha\cdot\mathbf{x}+\beta\cdot\mathbf{y}\in V,\quad \forall\mathbf{x},\mathbf{y}\in V,\alpha,\beta\in P$

则集合V对“+”与“·”运算也构成数域P上的线性空间，称为集合S的一个（线性）子空间。

若$\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\in S$，则集合$\left\{\mathbf{x}=\sum_{i=1}^{n}\alpha_i\mathbf{x}_i\,\middle|\,\alpha_1,\alpha_2,\ldots,\alpha_n\in P\right\}$是集合$S$的子空间。称为集合$S$的由基（向量）$\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$张（生）成的子空间。记为$\mathrm{Span}(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n)$.

定义 2.2.3 设集合 S 是数域 P 上的线性空间，$\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n \in S$ 。若存在不全为零的数 $\alpha_1,\alpha_2,\ldots,\alpha_n \in P$，使得 $\alpha_1\mathbf{x}_1+\alpha_2\mathbf{x}_2+\cdots+\alpha_n\mathbf{x}_n=\mathbf{0}$，则称 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ 是线性相关的。否则，上述等式只有在 $\alpha_1=\alpha_2=\ldots=\alpha_n=0$ 时成立，称 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ 是线性无关的。

若集合 V 是集合 S 的子集，集合 V 是由线性无关的 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ 张成的，集合 V 是有限维的线性无关集。

定义 2.2.4 设集合 S 是数域 P 上的线性空间。若集合 S 是由 n 个线性无关的向量 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ 生成的，即 S=Span$\left(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\right)$，则称集合 S 是 n 维的。称 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ 是集合 S 的一组（有限维）基（向量）。

集合 S 中的任何一个元素都可以有这组基表示出来：

$$\mathbf{x}=\sum_{i=1}^{n}\alpha_i\mathbf{x}_i,\quad \alpha_1,\alpha_2,\ldots,\alpha_n\in P$$

其中，系数 $\alpha_1,\alpha_2,\ldots,\alpha_n\in P$ 称为 $\mathbf{x}$ 在基 $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ 上的投影（坐标）。

2.2.2 赋范线性空间和 Banach 空间

为了考察线性空间中两个元素之间的“距离”，需要引入一种度量方式。称这种度量方式为范数。范数是向量长度概念的推广。范数在反问题研究中具有非常重要的意义。

设 $\mathbf{x}=\left(x_1,x_2,\cdots,x_n\right)$ 是 n 维实数域 R^n 中的任一向量，其长度定义为：

$$\left|\mathbf{x}\right|=\left(x_1^2+x_2^2+\cdots+x_n^2\right)^{1/2}$$

其性质为：

（1）$\left|\mathbf{x}\right|\geq 0$，当且仅当 $\mathbf{x}=\mathbf{0}$ 时，$\left|\mathbf{x}\right|=0$；

（2）$\left|\alpha\mathbf{x}\right|=\left|\alpha\right|\left|\mathbf{x}\right|$；

（3）$\left|\mathbf{x}+\mathbf{y}\right|\leq\left|\mathbf{x}\right|+\left|\mathbf{y}\right|$。

2.2.2.1 赋范线性空间

定义 2.2.5（范数，赋范线性空间） 设 X 为是实（或：复）数域 $\boldsymbol{F}$ 的线性空间，若对 $\forall\mathbf{x}\in X$，存在一个实数 $\left\|\mathbf{x}\right\|$ 于之对应，且满足下列条件：

（1） $\left\|\mathbf{x}\right\|\geq 0$；且 $\left\|\mathbf{x}\right\|=0\Leftrightarrow\mathbf{x}=\mathbf{0}$；　　（非负性（non-negativity））

（2） $\left\|\alpha\mathbf{x}\right\|=\left|\alpha\right|\left\|\mathbf{x}\right\|$，$\alpha\in\boldsymbol{F}$；　　（正齐（次）性（positive homogeneity））

（3） $\left\|\mathbf{x}+\mathbf{y}\right\|\leq\left\|\mathbf{x}\right\|+\left\|\mathbf{y}\right\|$，$\mathbf{x},\mathbf{y}\in X$；　　（三角不等式（triangle inequality））

则称 $\left\|\mathbf{x}\right\|$ 为 $\mathbf{x}$ 的**范数（norm）**，称 $(X,\left\|\bullet\right\|)$（或：$X$）为**赋范线性空间（normed linear space）**，简称**赋范空间（normed space）**。

例 2.2.1 空间 $C[a,b]$ 是闭区间 $[a,b]$ 上的连续函数全体所成的线性空间。对 $\forall f\in C[a,b]$，规定

$$\|f\| = \max_{t\in[a,b]} |f(t)|, \tag{2.2.1}$$

易证$\|f\|$是f的范数，则$C[a,b]$按上述范数成为赋范线性空间。

例 2.2.2 设$\boldsymbol{L}[a,b]$是闭区间$[a,b]$上的 Lebesgue 可积函数全体所成的线性空间。对$\forall f \in \boldsymbol{L}[a,b]$，规定

$$\|f\| = \int_a^b |f(t)|\mathrm{d}t, \tag{2.2.2}$$

若将在$[a,b]$上满足$f(t)\doteq g(t)$的两个函数f，g视为同一个函数，即将在$[a,b]$上满足$f(t)\doteq 0$的函数f视为恒等于零的函数，即$f=0$，则在$\boldsymbol{L}[a,b]$上，$\|f\|$是f的范数，从而$\boldsymbol{L}[a,b]$按上述范数成为赋范线性空间。

例 2.2.3 在n维实向量空间$\mathbf{R}^n$或n维复向量空间（称为**酉空间**）$\mathbf{C}^n$中，对$\forall \mathbf{x}=(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$（或$\mathbf{C}^n$），令

$$\|\mathbf{x}\| = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}, \tag{2.2.3a}$$

$$\text{或}\ \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| \tag{2.2.3b}$$

$$\text{或}\ \|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i| \tag{2.2.3c}$$

它们都是x的范数，称（2.2.3）中的范数为 **Euclidean 范数**，$\mathbf{R}^n$按范数（2.2.3）所得到的赋范线性空间称为 **Euclidean 空间**。

例 2.2.4（空间$C^{(k)}[a,b]$） $C^{(k)}[a,b]$是闭区间$[a,b]$上连续，在$[a,b]$中处处k次连续可微的函数全体所成的线性空间。对$\forall f\in C^{(k)}[a,b]$，令

$$\|f\| = \max_{t\in[a,b]} \left\{|f(t)|, |f'(t)|, \cdots, \left|f^{(k)}(t)\right|\right\}, \tag{2.2.4}$$

则$\|f\|$是f的范数，$C^{(k)}[a,b]$按上述范数成为赋范线性空间。

定义 2.2.6 设$(X,\|\cdot\|)$是赋范线性空间，对$\forall \mathbf{x},\mathbf{y}\in X$，令

$$\rho(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|, \tag{2.2.5}$$

则称$\rho(\mathbf{x},\mathbf{y})$为**由范数$\|\cdot\|$决定的度量**。

注 1 易验证$\rho(\mathbf{x},\mathbf{y})=\|\mathbf{x}-\mathbf{y}\|$满足度量的 3 个条件。

注 2 今后对每个赋范线性空间总是按照（2.2.5）引入度量，使之成为度量空间。这样我们就可以在赋范线性空间中引入极限的概念。

定理 2.2.1 设(X,ρ)是线性的度量空间（即：(X,ρ)既是线性空间，又是度量空间），若度量ρ是由某个范数$\|\cdot\|$决定的，则ρ满足：对$\forall \mathbf{x},\mathbf{y}\in X$，$\alpha$是数，有

$$\rho(\mathbf{x},\mathbf{y}) = \rho(\mathbf{x}-\mathbf{y},\ 0),\quad \rho(\alpha\mathbf{x},0) = |\alpha|\rho(\mathbf{x},0). \tag{2.2.6}$$

反之，若 ρ 满足（2.2.6），则 $\|\mathbf{x}\|=\rho(\mathbf{x},0)$ 就是 $\mathbf{x}$ 的范数。

也就是说：(2.2.6)是线性的度量空间成为赋范线性空间(指范数与度量满足(2.2.5))的充分必要条件。

证明：设度量 ρ 是由某个范数 $\|\bullet\|$ 决定的，即

$\rho(x,y)=\|x-y\|$，

则对 $\forall x,y\in X$， α 是数，有

$$\rho(x,y)=\|x-y\|=\|(x-y)-0\|=\rho(x-y,0);$$

$$\rho(\alpha x,0)=\|\alpha x-0\|=\|\alpha x\|=|\alpha|\|x\|=|\alpha|\|x-0\|=|\alpha|\,\rho(x,0).$$

反之，若度量 ρ 满足（3.1.6），定义 $\|x\|=\rho(x,0)$。

（1）若 $\|x\|=\rho(x,0)=0 \Leftrightarrow x=0$；　　（由度量的定义即得）

（2）$\|\alpha x\|=\rho(\alpha x,\ 0)=|\alpha|\rho(x,\ 0)=|\alpha|\|x\|$；　　（由（3.1.6）得）

（3）$\|x+y\|=\rho(x+y,0)=\rho(x-(-y),0)=\rho(x,-y)$

$\le \rho(x,0)+\rho(-y,0)=\|x\|+\|-y\|=\|x\|+|-1|\|y\|=\|x\|+\|y\|$。

注：并不是所有的度量都是由某个范数所决定的。

例如,在数列空间 $s=\{\boldsymbol{x}=(x_1,x_2,x_3,\cdots)\ \big|\ x_i\in\mathbf{R}^1\ (\text{或 }\mathbf{C}^1)\}$ 中，度量为

$$\rho(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^{\infty}\frac{1}{2^i}\frac{|x_i-y_i|}{1+|x_i-y_i|},$$

若令

$$\|\boldsymbol{x}\|=\rho(\boldsymbol{x},\boldsymbol{0})=\sum_{i=1}^{\infty}\frac{1}{2^i}\frac{|x_i|}{1+|x_i|},$$

则对常数 $\alpha\neq 0$，显然它不满足正齐（次）性条件 $\|\alpha\boldsymbol{x}\|=|\alpha|\|\boldsymbol{x}\|$。

又如离散度量空间。设 X 为任一非空集， ρ: $X\times X\to R^1$ 定义如下：对 $\forall(x,y)\in X\times X$，

$$\rho(x,y)=\begin{cases}0, & x=y,\\ 1, & x\neq y.\end{cases}$$

若令

$$\|x\|=\rho(x,0)=\begin{cases}0, & x=0,\\ 1, & x\neq 0.\end{cases},$$

则对常数 $\alpha\neq 0,1$，显然它也不满足正齐（次）性条件 $\|\alpha x\|=\alpha\|x\|$。

（$\because\ \|\alpha x\| = \rho(\alpha x, 0) = \begin{cases} 0, & x = 0 \\ 1, & x \neq 0 \end{cases} = \rho(x,0) \neq \alpha\rho(x,0) = \alpha\|x\|$.）

由此可知：由范数所决定的度量主要是针对线性空间定义的度量，它除了满足一般度量的 3 个条件，还必须满足正齐（次）性，这是一般度量所不具备的。

另外，范数是向量长度的推广，因此向量 $\alpha\boldsymbol{x}$ 的长度 $\|\alpha\boldsymbol{x}\|$ 当然要满足：$\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$，而一般度量（距离）却没有这样的要求。

例 2.2.5（空间 $\boldsymbol{l}^p$） $\boldsymbol{l}^p = \{x = (x_1, x_2, x_3, \cdots) \mid x_i\ (i = 1,2,\cdots) \in \mathbf{R}^1\ (\text{或 } \mathbf{C}^1),\ \sum_{i=1}^{\infty} |x_i|^p < \infty,\ p \geq 1\}$（即 p 方绝对收敛的实（或：复）数列的全体），$\boldsymbol{l}^p$ 按照对每个坐标 $x_i\ (i = 1,2,\cdots)$ 的线性运算成为线性空间。对 $\forall \boldsymbol{x} = (x_1, x_2, x_3, \cdots) \in \boldsymbol{l}^p$，$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}$ 是 $\boldsymbol{x}$ 的范数，$\boldsymbol{l}^p$ 按照 $\|\boldsymbol{x}\|_p$ 成为赋范线性空间。

例 2.2.6（空间 $\boldsymbol{l}^\infty$） 设 $\boldsymbol{l}^\infty$ 是有界实（或：复）数列 $\boldsymbol{x} = (x_1, x_2, x_3, \cdots)$ 全体按通常的线性运算所成的线性空间（它是 $\boldsymbol{s}$ 的线性子空间）。

对于 $\forall \boldsymbol{x} = (x_1, x_2, x_3, \cdots) \in \boldsymbol{l}^\infty$，$\|\boldsymbol{x}\|_\infty = \sup_i |x_i|$ 是 $\boldsymbol{x}$ 的范数，则 $\boldsymbol{l}^\infty$ 按照 $\|\boldsymbol{x}\|_\infty$ 成为赋范线性空间。

例 2.2.7（空间 $V[a,b]$） 设 $V[a,b]$ 是区间 $[a,\ b]$ 上的实（或：复）有界变差函数的全体，按照通常的线性运算，它是一个线性空间。对 $f \in V[a,b]$，令

$$\|f\| = |f(a)| + \overset{b}{\underset{a}{\mathrm{V}}}(f), \tag{2.2.8}$$

则 $\|f\|$ 是 f 的范数，则 $V[a,b]$ 按范数 $\|f\|$ 成为赋范线性空间。

令 $V_0[a,b] = \{f \mid f \in V[a,b],\ f$ 在 (a,b) 中每点都是右连续的，且 $f(a) = 0\}$，它是 $V[a,b]$ 的线性子空间。在 $V_0[a,b]$ 上，范数 $\|f\|$ 等于全变差 $\overset{b}{\underset{a}{\mathrm{V}}}(f)$。

证明： $V[a,b]$ 按照通常的线性运算是线性空间，这一点是显然的。今证（2.2.8）定义的范数满足定义 2.2.5 的 3 个条件。

（1） 对 $f \in V[a,b]$，显然 $\|f\| \geq 0$；

$$\|f\| = |f(a)| + \overset{b}{\underset{a}{\mathrm{V}}}(f) = 0 \ \Rightarrow\ f(a) = 0 \ \text{且}\ \overset{b}{\underset{a}{\mathrm{V}}}(f) = 0 \Rightarrow\ f \equiv 0\ 。$$

（2） $\|\alpha f\| = |\alpha|\|f\|$ 显然；

（3） $\|f+g\| = |f(a)+g(a)| + \overset{b}{\underset{a}{\mathrm{V}}}(f+g) \leq |f(a)| + |g(a)| + \overset{b}{\underset{a}{\mathrm{V}}}(f) + \overset{b}{\underset{a}{\mathrm{V}}}(g) = \|f\| + \|g\|$；

故由定义 2.2.5 知：$\|f\|$ 是 f 的范数，且 $V[a,b]$ 按范数 $\|f\|$ 成为赋范线性空间。

2. 2. 2. 2 Banach 空间

定义 2.2.7 设$(X, \|\cdot\|)$是赋范线性空间，x_n $(n=1,2,\cdots)$、$x\in X$，若当$n\to\infty$时，

$$\|x_n - x\| \to 0,$$

则称点列$\{x_n\}$**依范数**$\|\cdot\|$**收敛于**x （$\{x_n\}$ **converges to** x **in norm**），记作

$$\lim_{n\to\infty} x_n = x \quad （或：\ x_n \to x \quad (n\to\infty)）.$$

并称$\{x_n\}$为**收敛点列**（**convergent sequence**），称x为点列$\{x_n\}$的**极限** （**limit**）。

定义 2.2.8 设(X,ρ)是度量空间，$\{x_n\}$是X中的点列。若对$\forall\varepsilon>0$，存在$N(\varepsilon)>0$，当$m,n\ge N(\varepsilon)$时，恒有$\rho(x_m,x_n)<\varepsilon$，则称$\{x_n\}$是$X$中的**基本点列**，或 **Cauchy 点列**。

定理 2.2.2 （1）度量空间X中的收敛点列$\{x_n\}$必是基本点列；（2）设$\{x_n\}$是度量空间X中的基本点列，若$\{x_n\}$有子点列$\{x_{n_k}\}$收敛于X中的点x，则$\{x_n\}$也收敛于x。

定义 2.2.9 设(X,ρ)是度量空间，若X中的每个基本点列都收敛，则称X是**完备（度量）空间**。若A是度量空间X的子空间，若A作为度量空间是完备的，则称A是X中的**完备子空间**。完备的赋范线性空间称为**巴拿赫（Banach）空间**。

注 1 一个不完备的度量空间可以有完备的子空间。注 2 完备度量空间X的闭子空间A必是完备的子空间；任何度量空间X的完备的子空间A必是X的闭子集。

例 3.1.8 n维 Euclidean 空间$\mathbf{R}^n$按范数$\|x\| = \left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2}$，$x=(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$，是 Banach 空间。

例 3.1.9 $C[a,b]$按范数$\|f\| = \max_{x\in[a,b]}|f(x)|$ $(f\in C[a,b])$是 Banach 空间。

证明:在空间$C[a,b]$中，因为对$\{f_n\in C[a,b] \mid n=1,2,\cdots\}$，$\lim_{n\to\infty}\|f_n - f\| = 0$ $\Leftrightarrow$ $\{f_n\}$在$[a,b]$上一致收敛于f，所以由数学分析知：$f\in C[a,b]$. 只要证明$C[a,b]$中的任意基本点列$\{f_n\}$是$[a,b]$上的一致收敛点列即可。事实上，设$\{f_n\}$是$C[a,b]$中的任意基本点列，即对$\forall\varepsilon>0$，存在自然数$N(\varepsilon)>0$，当$m,\ n\ge N(\varepsilon)$时，恒有$\rho(f_n,f_m) = \|f_n - f_m\| = \max_{a\le t\le b}|f_n(t)-f_m(t)| < \varepsilon$。从而只要$m,\ n\ge N(\varepsilon)$，对任何$t\in[a,b]$，必有$|f_n(t)-f_m(t)|<\varepsilon$。数列的 Cauchy 收敛条件知：$\{f_n\}$在$[a,b]$上收敛于某一个函数$f$。再在上式中令$m\to\infty$，得$|f_n(t)-f(t)|\le\varepsilon$因此由$t\in[a,b]$的任意性得：$\max_{a\le t\le b}|f_n(t)-f(t)|<\varepsilon$，即$\{f_n\}$在$[a,b]$上一致收敛于$f$。

例 2.2.10 $\boldsymbol{L}[a,b]$按范数$\|f\|_1 = \int_a^b|f(t)|\mathrm{d}t$ $\left(f\in\boldsymbol{L}[a,b]\right)$是 Banach 空间。

例 2.2.11 若在$C[a,b]$中定义范数$\|f\|_1 = \int_a^b|f(t)|\mathrm{d}t$ $(f\in C[a,b])$，即将$C[a,b]$看成$\boldsymbol{L}[a,b]$的子空间，则$\left(C[a,b],\ \|\cdot\|_1\right)$是不完备的空间。事实上，因为$C[a,b]$按范数$\|f\|_1 = \int_a^b|f(t)|\mathrm{d}t$ $(f\in C[a,b])$在$\boldsymbol{L}[a,b]$中稠密，又$C[a,b]\ne\boldsymbol{L}[a,b]$，所以$C[a,b]$不可能完备。

反例（Counterexample）：任取c $(a<c<b)$，作函数序列$\{f_n\}$：

$$f_n(t)=\begin{cases}-1, & a\le t\le c-\frac{1}{n},\\ n(t-c), & c-\frac{1}{n}\le t\le c+\frac{1}{n},\\ 1, & c+\frac{1}{n}\le t\le b.\end{cases}$$

证明：上述连续函数序列$\{f_n\}$按范数$\|\bullet\|_1$的极限函数不是连续函数。

2.2.3 内积、内积空间与完备的内积空间（Hilbert 空间）

2.2.3.1 内积与内积空间

为了研究线性空间中元素间的几何关系，需要定义内积。n 维实数域R^n中的任意两个向量$\mathbf{x}=(x_1,x_2,\cdots,x_n)$和$\mathbf{y}=(y_1,y_2,\cdots,y_n)$的内积定义为：

$$(\mathbf{x},\mathbf{y})=\sum_{i=1}^{n}x_iy_i=x_1y_1+x_2y_2+\cdots+x_ny_n$$

上述内积定义的性质：

（1）$(\mathbf{x},\mathbf{y})=(\mathbf{y},\mathbf{x})$；

（2）$(\alpha\mathbf{x},\mathbf{y})=\alpha(\mathbf{y},\mathbf{x})$；

（3）$(\mathbf{x}+\mathbf{y},\mathbf{z})=(\mathbf{x},\mathbf{z})+(\mathbf{y},\mathbf{z})$；

（4）$(\mathbf{x},\mathbf{x})\ge 0$，当且仅当$\mathbf{x}=\mathbf{0}$时，$(\mathbf{x},\mathbf{x})=0$。

很显然，由上述内积定义可以引申出向量长度和向量夹角的概念。

向量长度：$|\mathbf{x}|=(\mathbf{x},\mathbf{x})^{1/2}$；

向量夹角：$\theta=\langle\mathbf{x},\mathbf{y}\rangle=\cos^{-1}\dfrac{(\mathbf{x},\mathbf{y})}{|\mathbf{x}||\mathbf{y}|}$，$0\le\theta\le\pi$。

显然，$(\mathbf{x},\mathbf{y})=0$代表向量$\mathbf{x}=(x_1,x_2,\cdots,x_n)$和向量$\mathbf{y}=(y_1,y_2,\cdots,y_n)$正交。上述概念可以推广到复数域（$C^n$）线性空间。

定义 2.2.10：设集合S是数域P上的线性空间，$(\cdot,\cdot)$是一个对应规则，使得对任何$\mathbf{x},\mathbf{y}\in S$有唯一的用记号$(\cdot,\cdot)$表示的实数与之对应。若$\|\cdot\|$满足下列条件：

（1）$(\mathbf{x},\mathbf{y})=\overline{(\mathbf{x},\mathbf{y})}$，$\forall\mathbf{x},\mathbf{y}\in S$；

（2）$(\alpha\mathbf{x},\mathbf{y})=\alpha(\mathbf{y},\mathbf{x})$，$\forall\mathbf{x},\mathbf{y}\in S$；

（3）$(\mathbf{x}+\mathbf{y},\mathbf{z})=(\mathbf{x},\mathbf{z})+(\mathbf{y},\mathbf{z})$，$\mathbf{x},\mathbf{y},\mathbf{z}\in S$；

（4）$(\mathbf{x},\mathbf{x})\geq 0$，当且仅当$\mathbf{x}=\mathbf{0}$时，$(\mathbf{x},\mathbf{x})=0$

则称$(\cdot,\cdot)$为集合S上的内积。集合S与$(\cdot,\cdot)$一起，成为线性空间基础上定义的内积空间。

一般地，有如下的加权内积的定义。对应 n 维复数域（C^n）线性空间，有如下的加权内积：$(\mathbf{x},\mathbf{y})=\sum_{i=1}^{n}w_i x_i\overline{y}_i=w_1x_1\overline{y}_1+w_2x_2\overline{y}_2+\cdots+w_nx_n\overline{y}_n$，$\forall\mathbf{x},\mathbf{y}\in C^n$。其中，$w_i>0(i=1,2,\cdots,n)$。

更一般地，函数空间$C[a,b]$中定义的加权内积形式为：

$$(f,g)=\int_a^b w(t)f(t)\overline{g(t)},\quad \forall f,g\in C[a,b]$$

其中，加权函数$w(t)$非负。

定义 2.2.11 设$(\cdot,\cdot)$为内积空间X的内积，则

$$\|\mathbf{x}\|=(\mathbf{x},\mathbf{x})^{1/2}$$

为确定该内积空间的一个范数。称为由内积$(\cdot,\cdot)$诱导出的范数。

显然，该诱导范数满足范数的 3 条性质。

定义 2.2.12 设X是内积空间，$\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\in X$，矩阵

$$G=\begin{bmatrix}(\mathbf{x}_1,\mathbf{x}_1) & (\mathbf{x}_1,\mathbf{x}_2) & \cdots & (\mathbf{x}_1,\mathbf{x}_n)\\ (\mathbf{x}_2,\mathbf{x}_1) & (\mathbf{x}_2,\mathbf{x}_2) & \cdots & (\mathbf{x}_2,\mathbf{x}_n)\\ \vdots & \vdots & \ddots & \vdots\\ (\mathbf{x}_n,\mathbf{x}_1) & (\mathbf{x}_n,\mathbf{x}_2) & \cdots & (\mathbf{x}_n,\mathbf{x}_n)\end{bmatrix}$$

则，$\det G\neq 0$ 等价于$\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$线性无关。矩阵$G$称为 Gram 矩阵，$\det G$为矩阵的行列式。

在勘探地震的地震信号分析和反演成像中，一般地，由一族基函数形成 Gram 矩阵。完备的规范正交基对应的 Gram 矩阵是单位阵。Fourier 基函数就是这样的一组基函数。这也是 Fourier 基函数得到如此广泛应用的一个原因，而常用的 Radon 变换基函数就不具备这样的性质。利用 Gram-Schmidt 正交化方法可以把任意一组不相关的基函数转化为一组规范正交基。

基函数的选择及应用在函数（信号）逼近、插值及参数反演中有占有极为重要的地位。选定一组基函数就是选定了一个参考坐标系，在这个坐标系中进行信号的分析和处理。直观地，这代表了从一个既定的角度看待一个物理过程。很显然，不同的坐标系中看同一个物理过程，物理过程本身没有变化，但是得到认识深度和分析手段的难易程度大有不同。

2.2.3.2 基函数的正交性与正交坐标系

2.2.3.2.1 基函数的正交性

设H是内积空间，$\mathbf{x},\mathbf{y}\in H$，如果$(\mathbf{x},\mathbf{y})=0$，称$\mathbf{x}$与$\mathbf{y}$正交，记为$\mathbf{x}\perp\mathbf{y}$。设$M$是$H$的任意子集，如果$\mathbf{x}\in H$与$M$中每一元正交，称$\mathbf{x}$与$M$正交，记为$\mathbf{x}\perp M$；如果$M$，$N$是$H$中两个子集，对于任意

$\mathbf{x}\in M$，$\mathbf{y}\in N$，$\mathbf{x}\perp\mathbf{y}$，称$M$与$N$正交，记$M\perp N$。设$M$是$H$的子集，所有$H$中与$M$正交的元的全体称为$M$的正交补，记为$M^{\perp}$。

定理 2.2.3 设H是内积空间，则有：

（1）如果$\mathbf{x},\mathbf{y},\mathbf{z}\in H$，$\mathbf{x}=\mathbf{y}+\mathbf{z}$且$\mathbf{y}\perp\mathbf{z}$，则$\|\mathbf{x}\|^2=\|\mathbf{y}\|^2+\|\mathbf{z}\|^2$；

（2）如果L是H的一个稠密子集，即$\overline{L}=H$，并且$\mathbf{x}\perp L$，则$\mathbf{x}=0$；

（3）M是H的任意子集，则$M^{\perp}$是H的闭子空间。

定理 2.2.4 设M是内积空间H中的完备凸集，则对任意$\mathbf{x}\in H$，存在$\mathbf{x}_0\in M$，使得$\|\mathbf{x}-\mathbf{x}_0\|=d(\mathbf{x},M)=\inf\limits_{\mathbf{y}\in M}\|\mathbf{x}-\mathbf{y}\|$。

定理 2.2.5（正交分解） 设M是 **Hilbert** 空间H的闭子空间，则对任意$\mathbf{x}\in H$，存在唯一的$\mathbf{x}_0\in M$及$y\in M^{\perp}$，使得$\mathbf{x}=\mathbf{x}_0+\mathbf{y}$。

2.2.3.2.2 正交系

设$\{\mathbf{x}_\alpha\}$，$\alpha\in I$是内积空间H中的子集，如果$\alpha\neq\beta$时$(\mathbf{x}_\alpha,\mathbf{y}_\beta)=0$，称$\{\mathbf{x}_\alpha\},\alpha\in I$是中的一个正交坐标系。

设$\{\mathbf{x}_\alpha\}$，$\alpha\in I$是一个正交系，如果对每一上$\alpha\in I$，$\|\mathbf{x}_\alpha\|=1$,称$\{\mathbf{x}_\alpha\}$，$\alpha\in I$是一个标准正交系。

设$\{\mathbf{x}_\alpha\}$，$\alpha\in I$是H的一个正交系，如果包含它的最小闭子空间是全空间H，称$\{\mathbf{x}_\alpha\}$，$\alpha\in I$是的正交基。

定理 2.2.6 设$\{\mathbf{e}_n\}$是内积空间H中的标准正交系，$\mathbf{x}\in H$，$\alpha_1,\alpha_2,\cdots,\alpha_n$是$n$个数，则当且当仅$\alpha_k=(\mathbf{x},\mathbf{e}_k)(k=1,2,\cdots,n)$时，$\left\|\mathbf{x}-\sum\limits_{k=1}^{n}\alpha_k\mathbf{e}_k\right\|$取最小值。

定理 2.2.7（Bessel 不等式） 设$\{\mathbf{e}_n\}$是内积空间H中的标准正交系，则对任意$\mathbf{x}\in H$，有

$$\sum_{k=1}^{\infty}|(x,e_k)|^2\le\|x\|^2。$$

定理 2.2.8 设$\{\mathbf{e}_n\}$是内积空间中的一个标准正交系，则$\{\mathbf{e}_n\}$是完备的，当且仅当$\{\mathbf{e}_n\}$张成的子空间L在H中稠密。

定理 2.2.9 设H是 **Hilbert** 空间，$\{\mathbf{e}_n\}$是H中的标准正交系，则$\{\mathbf{e}_n\}$是完备的，当且仅当$\{\mathbf{e}_n\}$是完全的。

定理 2.2.10 设H是 **Hilbert** 空间，$\{\mathbf{e}_n\}$是H中的标准正交系，$\{\alpha_n\}\in l^2$，则存在$\mathbf{x}\in H$，使得$\alpha_k=(\mathbf{x},\mathbf{e}_k)(k=1,2,\cdots,n)$，并且$\sum\limits_{k=1}^{\infty}|\alpha_k|^2=\|\mathbf{x}\|^2$。

定理 2.2.11（正交化定理） 设$\{\mathbf{x}_n\}$是内积空间H中的可数子集，则在H中存在标准正交系$\{\mathbf{e}_n\}$，使得$\{\mathbf{x}_n\}$与$\{\mathbf{e}_n\}$张成的子空间相同。

2.2.3.2.3 可分空间的同构

定理 2.2.12 设 H 是任一可分的无穷维的 Hilbert 空间，则存在 H 上到 l^2 同构映射 ϕ，且 ϕ 保持内积。

这个定理表示任何一个无穷维中分空间可以表示为“坐标形式” l^2。

2.2.3.3 Riesz 表示定理，Hilbert 空间的共轭空间

2.2.3.3.1 Riesz 表示定理

定理 2.2.13（Riesz 表示定理） 设 H 是 **Hilbert** 空间，f 是 H 上任意有界线性泛函，则存在唯一的 $y_f \in H$，使得对于每一个 $\mathbf{x} \in H$，有 $f(\mathbf{x}) = (\mathbf{x}, \mathbf{y}_f)$，并且有 $\|f\| = \|\mathbf{y}_f\|$。

2.2.3.3.2 空间的共轭空间

设 H 是 **Hilbert** 空间，$A \in \beta(H)$，于是对任意 $\mathbf{y} \in H$，易见 $(A\mathbf{x}, \mathbf{y})(\mathbf{x} \in H)$ 是 H 上的一个有界线性泛函，因此由 **Riesz** 表示定理，存在唯一的 $\mathbf{z} \in H$，使得

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{z})(\mathbf{x} \in H)$$

其中，$B\mathbf{y} = \mathbf{z}$。

定义 2.2.13 设 H 是 **Hilbert** 空间，$A \in \beta(H)$，把上式确定的有界线性算子 B 称为 A 的共轭算子。即 A 的共轭算子 A^* 是由下式定义的算子：$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y})(\mathbf{x}, \mathbf{y} \in H)$。

定义 2.2.14 设 H 是 **Hilbert** 空间，A 是 H 上的有界线性算子，如果 $A^* = A$，即对任意 $(\mathbf{x}, \mathbf{y} \in H)$，有 $(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y})$，则称 A 是自共轭算子。

设 A 是 **Hilbert** 空间 H 的有界共轭算子，以下是算子 A 的一些简单性质：

（1）对任意 $\mathbf{x} \in H$，$(A\mathbf{x}, \mathbf{x})$ 是实的；

（2）$\|A\| = \sup_{\|\mathbf{x}\|=1} |(A\mathbf{x}, \mathbf{x})|$；

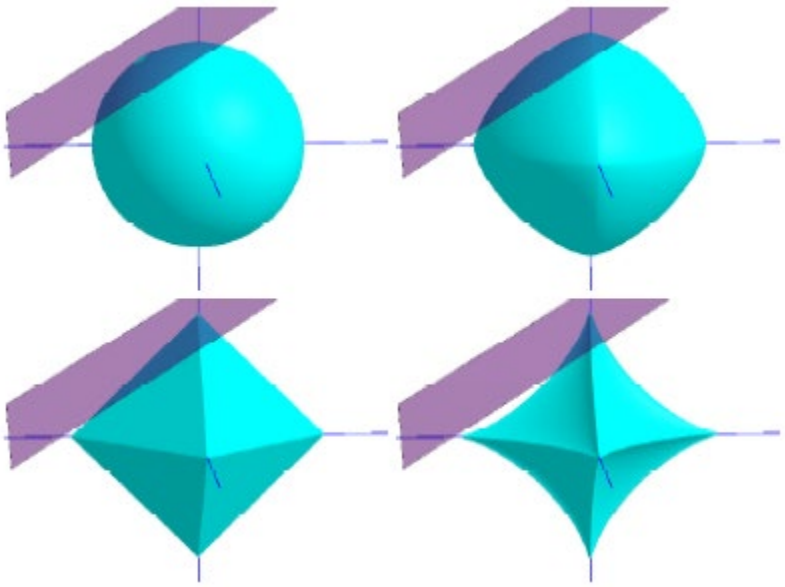
（3）算子 A 的特征值是实的；

（4）对应于算子 A 的不同特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 的特征向量 $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ 是正交的。

2.2.3.4 Hilbert 空间中的自共轭紧算子

引理 设 H 是 **Hilbert** 空间，A 是 H 上的有界共轭算子，如果存在 $\mathbf{x}_0 \in H$，$\|\mathbf{x}_0\| = 1$，使得泛函 $|\phi(\mathbf{x})| = |(A\mathbf{x}, \mathbf{x})|$ 在 $\mathbf{x}_0$ 点达到极大，则由 $(\mathbf{x}_0, \mathbf{y}) = 0$ 可推出 $(A\mathbf{x}_0, \mathbf{y}) = (\mathbf{x}_0, A\mathbf{y}) = 0$。

定理 2.2.14（Hilbert-Schmidt 定理） 设 A 是 **Hilbert** 空间 H 上的自共轭紧算子，则存在对应于特征值 $\{\lambda_n\}(\lambda_n \neq 0)$ 的特征向量构成的标准正交系 $\{\mathbf{e}_n\}$，使得每一元 $\mathbf{x} \in H$ 可唯一地表示为：$\mathbf{x} = \sum_k \alpha_k \mathbf{e}_k + \mathbf{x}_0'$，其中 $\mathbf{x}_0' \in N(A)$，即满足 $A\mathbf{x}_0' = 0$，同时 $A\mathbf{x} = \sum_k \lambda_k \alpha_k \mathbf{e}_k$。并且如果 $\{\mathbf{e}_n\}$ 是无穷的，则 $\lim_{n\to\infty} \lambda_n = 0$。



R^n 中定义的向量 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 的范数有三种形式：

L_∞ 范数 $\|x\|_\infty=\max\limits_{1\le i\le n}|x_i|$

L_1 范数 $\|x\|_1=\sum\limits_{i=1}^{n}|x_i|$

L_2 范数 $\|x\|_2=\left(\sum\limits_{i=1}^{n}|x_i|^2\right)^{1/2}$

矩阵 $A=(a_{ij}\in R^{n\times n})$ 的范数也有三种形式：

$$\|A\|_\infty=\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|$$

$$\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|$$

$$\|A\|_2=\left[\rho(A^TA)\right]^{\frac{1}{2}}$$

还有矩阵的 **Frobenius** 范数

$$\|A\|_F=(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2)^{\frac{1}{2}}$$

2.2.4 Fourier 级数与可分的 Hilbert 空间

2.2.4.1 Fourier 级数

在数学分析中，一个以 2π 为周期的函数可以展开为 Fourier 级数：

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n\cos nx+b_n\sin nx\right)$$

其中的傅里叶系数为：$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)dx$，$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos nxdx$，$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin nxdx$。

由收敛定理知：若函数 $f(x)$ 在一个周期内：①连续或只有有限个第一类间断点（可取或跳跃间断点）；②至多有有限个极值点，那么对于连续点 Fourier 级数收到 $f(x)$；③对于间断点级数收敛到左右极限的算术平均值。

可验证 $\left\{\mathbf{e}_n^*\right\}=\left\{\mathbf{e}_0,\mathbf{e}_n,\mathbf{e}_0'\middle|\mathbf{e}_0=\frac{1}{\sqrt{2}};\mathbf{e}_n=\cos nx;\mathbf{e}_n'=\sin nx;n=1,2,\cdots,n\right\}$ 是 $L^2[-\pi,\pi]$ 上的一个标准正交系。

内积定义为 $(f,g)=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)g(t)dt$，所以 $f(x)$ 的展开式中的系数可改写为：

$$f(x)=(f,\mathbf{e}_0)\mathbf{e}_0+\sum_{n=1}^{\infty}\left((f,\mathbf{e}_n)\mathbf{e}_n+(f,\mathbf{e}_n')\mathbf{e}_n'\right)$$

这种函数的展开式表示是研究很多实际问题的重要工具。下面将这种展开推广到一般的 Hilbert 空间。

2.2.4.2 Fourier 级数及其收敛性

定义 2.2.15 （广义）Fourier 级数

设 $\{\mathbf{e}_n\}$ 为内积空间 X 的标准正交系，$\mathbf{x}\in X$，称级数 $\sum_{k=1}^{\infty}(\mathbf{x},\mathbf{e}_k)\mathbf{e}_k=\sum_{k=1}^{\infty}\mathbf{c}_k\mathbf{e}_k$ 为 $\mathbf{x}$ 关于 $\{\mathbf{e}_n\}$ 的 Fourier 级数，其中 $\mathbf{c}_k=(\mathbf{x},\mathbf{e}_k),k=1,2,\cdots,n$ 为 $\mathbf{x}$ 关于 $\{\mathbf{e}_n\}$ 的（广义）Fourier 系数。

定理 2.2.15 设 $\{\mathbf{e}_n\}$ 为内积空间 X 的标准正交系，记子空间 $M=\mathrm{Span}\{\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_n\}$，则 $\forall\mathbf{x}\in X$，有 $\mathbf{x}_0=\sum_{k=1}^{n}(\mathbf{x},\mathbf{e}_k)\mathbf{e}_k$ 是 $\mathbf{x}$ 在 M 上的正交投影。$\mathbf{x}_0$ 也是 $\mathbf{x}$ 在 M 上的最佳逼近，且 $\mathbf{x}=\mathbf{x}_0+\mathbf{z}$，$z(=\mathbf{x}-\mathbf{x}_0)\perp M$。$\mathbf{z}$ 为最佳逼近误差向量。

定理 2.2.16（最佳逼近定理） 设 $\{\mathbf{e}_n\}$ 为内积空间 X 的标准正交系，$\mathbf{x}\in X$，$\mathbf{c}_k=(\mathbf{x},\mathbf{e}_k),k=1,2,\cdots,n$，则对任何坐标投影 $\{\alpha_1,\alpha_2,\cdots,\alpha_n\}\in R^n$ 有

$$\left\|\mathbf{x}-\sum_{k=1}^{n}c_k\mathbf{e}_k\right\|\le\left\|\mathbf{x}-\sum_{k=1}^{n}\alpha_k\mathbf{e}_k\right\|.$$

证明：设 $\mathbf{y}=\sum_{k=1}^{n}c_k\mathbf{e}_k$ 以及 $\mathbf{y}'=\sum_{k=1}^{n}\alpha_k\mathbf{e}_k$，则由定理 1.1 知 $\mathbf{y}$ 是 $\mathbf{x}$ 在 $M=\mathrm{Span}\{\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_n\}$ 上的正交投影，即 $\mathbf{y}\in M$，$\mathbf{x}-\mathbf{y}\perp M$．又因 $\mathbf{y}'\in M$，可见 $\mathbf{y}-\mathbf{y}'\in M$，于是 $\mathbf{x}-\mathbf{y}\perp\mathbf{y}-\mathbf{y}'$。根据勾股定理有：

$$\left\|\mathbf{x}-\sum_{k=1}^{n}\alpha_k\mathbf{e}_k\right\|_2^2=\|\mathbf{x}-\mathbf{y}'\|_2^2=\|\mathbf{x}-\mathbf{y}+\mathbf{y}-\mathbf{y}'\|_2^2=\|\mathbf{x}-\mathbf{y}\|_2^2+\|\mathbf{y}-\mathbf{y}'\|_2^2\geq\|\mathbf{x}-\mathbf{y}\|_2^2=\left\|\mathbf{x}-\sum_{k=1}^{n}c_k\mathbf{e}_k\right\|_2^2.$$

定理 2.2.17（Fourier 级数收敛的充要条件） 设 $\{\mathbf{e}_n\}$ 为内积空间 X 的标准正交系，$\mathbf{x}\in X$，则 $\mathbf{x}$ 关于 $\{\mathbf{e}_n\}$ 的 **Fourier 级数** $\sum_{k=1}^{n}(\mathbf{x},\mathbf{e}_k)\mathbf{e}_k$ 收敛于 $\mathbf{x}$（即 $\mathbf{x}=\sum_{k=1}^{n}(\mathbf{x},\mathbf{e}_k)\mathbf{e}_k$）的充要条件为：$\|\mathbf{x}\|_2^2=\sum_{k=1}^{n}|c_k|^2$，其中 $\mathbf{c}_k=(\mathbf{x},\mathbf{e}_k),k=1,2,\cdots,n$，称 $\|\mathbf{x}\|_2^2=\sum_{k=1}^{n}|c_k|^2$ 为 **Parseval** 公式。

证明：设 $\mathbf{y}=\sum_{k=1}^{n}c_k\mathbf{e}_k$，则 $\mathbf{x}-\mathbf{y}\perp\mathbf{y}$。则

$$\left\|\mathbf{x}-\sum_{k=1}^{n}c_k\mathbf{e}_k\right\|_2^2=\|\mathbf{x}-\mathbf{y}\|_2^2=\|\mathbf{x}\|_2^2-\|\mathbf{y}\|_2^2=\|\mathbf{x}\|_2^2-\sum_{k=1}^{n}|c_k|^2$$

因此级数 $\sum_{k=1}^{n}(\mathbf{x},\mathbf{e}_k)\mathbf{e}_k$ 收敛的等价条件为 $\mathbf{y}\to\mathbf{x}(k\to\infty)$（依范数），即 $\lim_{n\to\infty}\sum_{k=1}^{n}|c_k|^2=\|\mathbf{x}\|_2^2$。

注：由上述定理知，在内积空间 X 的标准正交系 $\{\mathbf{e}_n\}$ 下，当 Bessel 不等式中的等号成立（即 **Parseval** 公式成立）时，$\mathbf{x}$ 关于 $\{\mathbf{e}_n\}$ 的（广义）Fourier 级数 $\sum_{k=1}^{n}(\mathbf{x},\mathbf{e}_k)\mathbf{e}_k$ 就收敛于 $\mathbf{x}$。Bessel 不等式成为等式的条件是 X 中的标准正交系是完备的（即有没有选取足够多坐标）。

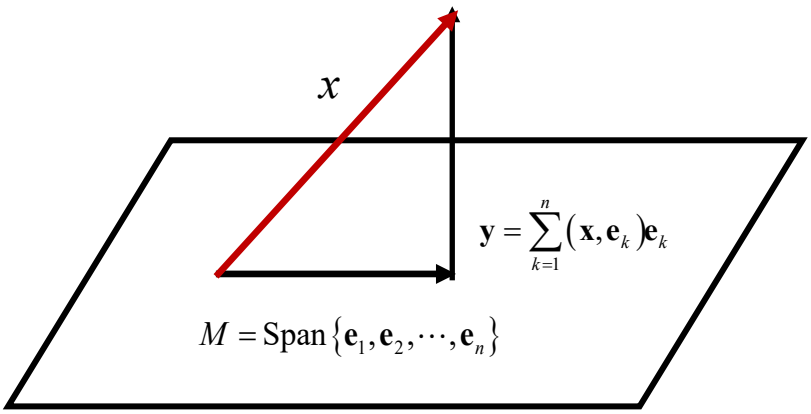


图 1 向量 $\mathbf{x}$ 在 $M=\mathrm{Span}\{\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_n\}$ 上的正交投影示意图

定义 2.2.16 完全标准正交系（基）

设 $\{\mathbf{e}_n\}$ 为内积空间 X 的标准正交系，如果 $\mathbf{x}\perp\mathbf{e}_k(k=1,2,\cdots,n)$，则 $\mathbf{x}=0$，那么称 $\{\mathbf{e}_n\}$ 为内积空间 X 的**完全标准正交系（或完全标准正交基）**。

定理 2.2.18 设 H 是 Hilbert 空间，则以下命题等价：

（1） $\{\mathbf{e}_n\}$ 是 H 的完全标准正交系；

（2） $\forall \mathbf{x}\in H$， $\mathbf{x}$ 关于 $\{\mathbf{e}_n\}$ 的 Fourier 级数 $\sum_{k=1}^{n}(\mathbf{x},\mathbf{e}_k)\mathbf{e}_k$ 收敛， $\mathbf{x}=\sum_{k=1}^{\infty}\mathbf{c}_k\mathbf{e}_k$；

（3） $\forall \mathbf{x}\in H$， $\|\mathbf{x}\|_2^2=\sum_{k=1}^{n}|c_k|^2$。

注：通过上述定理把数学分析中的 Fourier 展开式推广到抽象的 Hilbert 空间中，并揭示了完全标准正交系、Parseval 公式及 Fourier 展开式之间的本质联系。可以证明，在 $L^2[-\pi,\pi]$ 空间中，三角函数系 $\left\{\frac{1}{\sqrt{2}},\cos x,\sin x;\cos 2x,\sin 2x;\cdots,\cos nx,\sin nx;\cdots\right\}$ 是它的完全标准正交系。

2.2.4.3 **可分 Hilbert 空间的模型**

定义 2.2.17（线性等距同构） 设 X_1,X_2 为同一数域 K 上的内积空间，如果存在从 X_1 到 X_2 上的一一映射 $\varphi:X_1\to X_2$ 保持线性运算和内积，即 $\forall \mathbf{x},\mathbf{y}\in X_1,\alpha,\beta\in K$，有：

$$\varphi(\alpha\mathbf{x}+\beta\mathbf{y})=\alpha\varphi(\mathbf{x})+\beta\varphi(\mathbf{y});\ (\varphi(x),\varphi(y))=(\mathbf{x},\mathbf{y})$$

则称 X_1 与 X_2 **线性等距同构**。

事实上，可将线性等距同构的两个内积空间看作“同一”个空间。下面分别刻画有限维和无限维的 Hilbert 空间。

定理 2.2.19（有限维 Hilbert 空间模型） 设 H 是 n 维 Hilbert 空间，则 H 与复内积空间 C^n 线性等距同构。

证明：在 H 上取一组基 $\mathbf{g}_1,\mathbf{g}_2,\cdots,\mathbf{g}_n$，应用格拉姆-施密特（Gram-Schmidt）方法将其正交化，可得 H 的标准正交系 $\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_n$，作 H 到 C^n 上的映射 φ：

$$\varphi(\mathbf{x})=((\mathbf{x},\mathbf{e}_1),(\mathbf{x},\mathbf{e}_2),\cdots,(\mathbf{x},\mathbf{e}_n))\in C^n$$

容易验证 φ 是从 H 到 C^n 上的线性等距同构映射。

定理 2.2.20（Hilbert 空间可分充要条件） 无限维 Hilbert 空间 H 可分当且仅当 H 有完全标准正交系 $\{\mathbf{e}_n\}$。

定理 2.2.21 如果无限维 Hilbert 空间 H 可分，则 H 与 l^2 同构。

注：根据上述定理可知，复内积空间 C^n 是有限维 Hilbert 空间 H 的模型；复内积空间 l^2 是无限维 Hilbert 空间 H 的模型。

2.2.5 空间 $\boldsymbol{L}^p$

2.2.5.1 $\boldsymbol{L}^p$ $(\mathbf{0}\le p<+\infty)$上的范数

在分析中最常用的一类赋范线性空间是 $\boldsymbol{L}^p$。

定义 2.2.5.1 设$(X,\boldsymbol{R},\mu)$是一个测度空间，$E\in\boldsymbol{R}$，f是E上的实值（或复值）可测函数。若$|f|^p$在E上是可积的，其中p是一个固定的正数，则称函数f为**E 上关于测度 μ 的 p 方可积函数**，或简称为**p 方可积函数**。这类函数的全体称为**E 上关于测度 μ 的 p 方可积函数空间**，记为$\boldsymbol{L}^p(E,\boldsymbol{R},\mu)$，或简记为$\boldsymbol{L}^p(E,\mu)$。

特别地，用$\boldsymbol{L}^p(E)$表示E上关于 Lebsgue 测度m的p方可积函数空间。当$p=1$时，$\boldsymbol{L}^1(E,\mu)$就是E上的可积函数全体，也记作$\boldsymbol{L}(E,\mu)$. $E=[a,b]$时，改记$\boldsymbol{L}^p([a,b])$为$\boldsymbol{L}^p[a,b]$，又如$\boldsymbol{L}^p((-\infty,\infty))$改记为$\boldsymbol{L}^p(-\infty,\infty)$，等等。

注 1 $\boldsymbol{L}^p(E,\mu)$按通常的线性运算成一线性空间。

注 2 在$\boldsymbol{L}^p(E,\mu)$中，我们将几乎处处相等的两个可测函数f,g看成同一个向量，它可以用f表示，也可以用g表示。这时直接写$f=g$，而不必再写成$f\doteq g$. 经过这样“同一化”后，$\boldsymbol{L}^p(E,\mu)$仍然是线性空间。因为对几乎处处相等的函数，它们的积分相等，所以对$\boldsymbol{L}^p(E,\mu)$中的每个向量f可以做出一个确定的函数

$$\|f\|_p=\left(\int_E|f|^p\,\mathrm{d}\mu\right)^{1/p}\quad(p\ge1)$$

定理 2.2.5.1 对$\forall f\in\boldsymbol{L}^p(E,\mu)$，$p\ge1$，函数

$$\|f\|=\left(\int_E|f|^p\,\mathrm{d}\mu\right)^{1/p}\tag{2.2.5.1}$$

是$\boldsymbol{L}^p(E,\mu)$上的范数，$\boldsymbol{L}^p(E,\mu)$按照上述范数成为赋范线性空间。

定义 2.2.5.2 设$\{f_n\}$是$\boldsymbol{L}^p(E,\mu)$中一函数列，$f\in\boldsymbol{L}^p(E,\mu)$，$\{f_n\}$依范数$\|\cdot\|$收敛于$f$，即

$$\lim_{n\to\infty}\|f_n-f\|_p=\lim_{n\to\infty}\left(\int_E|f_n(t)-f(t)|^p\,\mathrm{d}\mu\right)^{1/p}=0\tag{2.2.5.2}$$

或

$$\int_E|f_n(t)-f(t)|^p\,\mathrm{d}\mu\to0\ \ (n\to\infty),\tag{2.2.5.3}$$

称之为$\{f_n\}$在E上**p 方平均收敛**于f（或简称$\{f_n\}$**平均收敛**于f）。

定理 2.2.5.2（平均收敛与依测度收敛的关系） 设$\{f_n\}$是$\boldsymbol{L}^p(E,\mu)$中一函数列，$f\in\boldsymbol{L}^p(E,\mu)$。若$\{f_n\}$在$E$上$p$方平均收敛于$f$，则$\{f_n\}$在$E$上依测度$\mu$收敛于$f$。

证明：因为$\{f_n\}$在E上p方平均收敛于f，所以

$$\int_E |f_n(t)-f(t)|^p \mathrm{d}\mu \to 0 \quad (n\to\infty).$$

对 $\forall \varepsilon>0$，有

$$\int_E |f_n(t)-f(t)|^p \mathrm{d}\mu \geq \int_{E(|f_n-f|\geq\varepsilon)} |f_n(t)-f(t)|^p \mathrm{d}\mu \geq \varepsilon^p \mu(E(|f_n-f|\geq\varepsilon)).$$

令 $n\to\infty$，则

$$\mu(E(|f_n-f|\geq\varepsilon))\to 0.$$

证毕！

推论 设 $\{f_n\}$ 是 $\boldsymbol{L}^p(E,\mu)$ 中一函数列，$f\in\boldsymbol{L}^p(E,\mu)$。若 $\{f_n\}$ 在 E 上 p 方平均收敛于 f，则必有 $\{f_n\}$ 的子函数列 $\{f_{n_k}\}$ 在 E 上几乎处处收敛于 f。事实上，由 F . Riesz 定理立得上述推论的结果。

注:定理 2.2.5.2 的逆命题不成立，即使函数列 $\{f_n\}$ 在有限可测集 E 上处处收敛于 f，也不能保证 $\{f_n\}$ 在 E 上 p 方平均收敛于 f。

例 2.2.5.1 作区间 $[0,1]$ 上的函数列 $\{f_n\}$ 如下：

$$f_n(t)=\begin{cases}0, & t=0 \text{ 或 } \frac{1}{n}\leq t\leq 1,\\ e^n, & 0<t<\frac{1}{n},\end{cases}$$

显然，$\{f_n\}$ 在 $[0,1]$ 上处处收敛于零函数 0；但是，当 $n\to\infty$ 时，对于 $\forall p>0$，有

$$\int_0^1 |f_n(t)|^p \mathrm{d}t = \int_0^{1/n} e^{np}\mathrm{d}t = \frac{1}{n}e^{np}\to\infty \quad (n\to\infty),$$

即 $\{f_n\}$ 并不 p 方平均收敛于零。

2. 2. 5. 2 $L^\infty(E,\mu)$ 上的范数

定义 2.2.5.3 设 $(X,\boldsymbol{R},\mu)$ 是一个测度空间，$E\in\boldsymbol{R}$，f 是 E 上的实值（或复值）可测函数，若 f 和 E 上的一个有界函数几乎处处相等（即：若存在 E 中的 $\mu-$零集 E_0，使得 f 在 $E-E_0$ 上有界），则称函数 f 为 **E 上（关于 μ）的本性有界可测函数**。E 上（关于 μ）的本性有界可测函数全体记作 $L^\infty(E,\mu)$。

注 $L^\infty(E,\mu)$ 按通常的线性运算成为一个线性空间。事实上，因为有限个零集的并仍是零集，所以任意有限个本性有界可测函数的线性组合仍是本性有界的。

定义 2.2.5.4 设 f 是 E 上的本性有界可测函数，称

$$\|f\|_\infty = \inf_{\substack{\mu(E_0)=0\\ E_0\subset E}}\left[\sup_{x\in E-E_0}|f(x)|\right] \qquad (2.2.5.4)$$

为 f 的**本性最大模**，也记作 $\operatorname{ess\,sup}_{x\in E}|f(x)|$.

命题 1 式（2.2.5.4）中的下确界是可以达到的，也就是说：存在零集 $\tilde{E}_0\subset E$，使得

$$\|f\|_\infty = \sup_{x\in E-\tilde{E}_0} |f(x)|.$$

证明：由下确界的意义，对每个 n $(n=1,2,3,\cdots)$，有 $E_n \subset E$，使得 $\mu(E_n)=0$，并且

$$\sup_{x\in E-E_n} |f(x)| < \|f\|_\infty + \frac{1}{n}.$$

作 $\tilde{E}_0 = \bigcup_{n=1}^{\infty} E_n$，则 $\mu(\tilde{E}_0)=0$，并且

$$\|f\|_\infty \le \sup_{x\in E-\tilde{E}_0} |f(x)| \le \sup_{x\in E-E_n} |f(x)| < \|f\|_\infty + \frac{1}{n},$$

令 $n\to\infty$，得

$$\|f\|_\infty = \sup_{x\in E-\tilde{E}_0} |f(x)|.$$

证毕！

命题 2 $\|f\|_\infty$ 是与 f 几乎处处相等的各个有界函数的绝对值的上确界的最小值。即：

$$\|f\|_\infty = \inf_{\substack{\mu(E_0)=0\\ E_0\subset E}} \left[\sup_{x\in E-E_0} \left\{ |g(x)| \;\middle|\; f(x)=g(x),\ g(x) \text{ 在 } E \text{ 上有界} \right\} \right]. \qquad (2.2.5.5)$$

定理 2.2.5.3 $\|f\|_\infty$ 是 $\boldsymbol{L}^\infty(E,\mu)$ 上的范数，$\boldsymbol{L}^\infty(E,\mu)$ 按此范数成为赋范线性空间。

定理 2.2.5.4 设 $\{f_n\}$ 是 $\boldsymbol{L}^\infty(E,\mu)$ 中一函数列，$f\in\boldsymbol{L}^\infty(E,\mu)$，$\{f_n\}$ 依范数 $\|\cdot\|_\infty$ 收敛于 f，即

$\|f_n - f\|_\infty \to 0 \quad (n\to\infty)$

$\Leftrightarrow \{f_n\}$ 在 E 上**几乎一致收敛**于 f，即：

$\{f_n\}$ 在 E 上除去一个零集后一致收敛于 f。

证明：“$\Rightarrow$”若 $\{f_n\}$ 依范数 $\|\cdot\|_\infty$ 收敛于 f（即 $\|f_n - f\|_\infty \to 0$ $(n\to\infty)$），则存在 $E_n\subset E$，$\mu(E_n)=0$ $(n=1,2,3,\cdots)$，使得

$$\|f_n - f\|_\infty = \sup_{x\in E-E_n} |f_n(x)-f(x)| \to 0 \quad (n\to\infty).$$

取 $E_0 = \bigcup_{n=1}^{\infty} E_n$，则 E_0 是一个零集，并且

$$\sup_{x\in E-E_0} |f_n(x)-f(x)| \to 0 \quad (n\to\infty). \qquad (*)$$

因为 E_0 是 E 的一个零集，所以 $(*)$ 说明：$\{f_n\}$ 在 E 上几乎一致收敛于 f。

“$\Leftarrow$” 显然(*)也是使$\|f_n - f\|_\infty \to 0 \ (n \to \infty)$的充分条件。 证毕!

定理 2.2.5.5 若$\mu(E) < \infty$，则对一切$p > 0$，有$\boldsymbol{L}^\infty(E,\mu) \subset \boldsymbol{L}^p(E,\mu)$，且

$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p .$$

2.2.6 紧集与有限维赋范线性空间

2.2.6.1 致密集的概念

实数直线上的 **Bolzano-Weierstrass 致密性定理（compactness theorem）**：任一有界数列必有收敛子列。

定义 2.2.6.1 设(X,ρ)是度量空间，$A \subset X$。若在A中的任何点列必有在X中收敛的子点列，则称A是（X中的）**致密集**。

若X自身是致密集，则称X是**致密空间**。

性质 1 有限点集是致密集。

注：点集和点列不一样，点列是取点集中的元素构成的，其各项可以重复，但点集中的元素却不能一样。因此，由于有限点集中的元素有限，所以要想构成点列，必然有同一个元素无数次重复，这样，这些重复的元素构成的子点列必然收敛。

性质 2 有限个致密集的并是致密集。

证明： 设$A_1, A_2, \cdots, A_m$是度量空间(X,ρ)的致密集，往证$A = \bigcup_{k=1}^{m} A_k$也是$(X,\rho)$的致密集。任取一点列$\{x_n\} \subset A$，则存在$A_\ell \ (1 \le \ell \le m)$，$\{x_n\}$有无限多项属于$A_\ell$，记其为$\{x_{n_k}\}$，即$\{x_{n_k}\} \subset A_\ell$。而$A_\ell$是致密的，所以必有在$X$中收敛的子点列$\{x_{n_{k_h}}\}$，使得

$$x_{n_{k_h}} \to x \in X \quad (h \to \infty),$$

即$\{x_n\}$在X中收敛的子点列$\{x_{n_{k_h}}\}$，故A也是(X,ρ)的致密集。证毕!

性质 3 致密集的任何子集是致密集。因此，任何一族致密集的交是致密集。

证明：只要证明“**致密集的任何子集是致密集**”即可，而“**任何一族致密集的交是致密集**”则是前者的直接推论。

设A是度量空间(X,ρ)的致密集，B是A的任一子集。任取一点列$\{x_n\} \subset B$，因为$B \subset A$，所以$\{x_n\} \subset A$。而A是致密的，因此点列$\{x_n\}$必有在X中收敛的子点列$\{x_{n_k}\}$，使得$x_{n_k} \to x \in X \quad (k \to \infty)$，故$B$也是致密的。 证毕!

性质 4 致密集的闭包是致密集。

证明：设(X,ρ)是度量空间，$A \subset X$是致密集，往证A的闭包$\overline{A} = A \cup A'$也是致密集。

任取一点列 $\{x_n\}\subset \overline{A}=A\bigcup A'$，则对每个 n，存在 $y_n\in A$（进而得一点列 $\{y_n\}\subset A$），使得 $\rho(x_n,y_n)<\dfrac{1}{n}$。因为 A 是致密的，所以点列 $\{y_n\}$ 必有在 X 中收敛的子点列 $\{y_{n_k}\}$，使得 $y_{n_k}\to y\in X\quad (k\to\infty)$，即 $\rho(y_{n_k},y)\to 0\quad (k\to\infty)$。

于是由三点不等式得

$$\rho(x_{n_k},y)\le \rho(x_{n_k},y_{n_k})+\rho(y_{n_k},y)<\frac{1}{n_k}+\rho(y_{n_k},y)\to 0\quad (k\to\infty)$$

即：点列 $\{x_n\}$ 必有在 X 中收敛的子点列 $\{x_{n_k}\}$，使得 $x_{n_k}\to y\in X\quad (k\to\infty)$.

由定义 **2.2.6.1** 知：A 的闭包 $\overline{A}$ 也是致密集。 证毕！

性质 5 致密集中的基本点列必然收敛。因此，致密的度量空间是完备的。

证明： 设 (X,ρ) 是度量空间，$A\subset X$ 是致密集。

若 $\{x_n\}$ 是 A 中的基本点列，则由 A 的致密性得：点列 $\{x_n\}$ 必有在 X 中收敛的子点列 $\{x_{n_k}\}$，使得 $x_{n_k}\to x\in X\quad (k\to\infty)$。于是 $x_n\to x\ (n\to\infty)$。证毕！

2.2.6.2 紧集

定义 2.2.6.2 （紧集） 称度量空间中的致密闭集为**紧集（Compact Set）**。

注: 显然，A 是紧集的充要条件是：**A 中任一点列必有收敛的子点列收敛于 A 中的一点**。

对于全空间来说，致密的概念和紧集的概念没有区别，致密的度量空间又称为**紧（度量）空间**。

性质 度量空间 (X,ρ) 中的紧集 A 看成 X 的子空间时是完备的。

定理 2.2.6.1（Gross） 设 A 是度量空间 (X,ρ) 中的紧集，$\boldsymbol{K}$ 是 X 中的一族开集。若 $\boldsymbol{K}$ 覆盖 A：

$$A\subset \bigcup_{B_\lambda\in \boldsymbol{K}} B_\lambda\ ,$$

则必有 $\boldsymbol{K}$ 中的有限个开集 $B_1,B_2,\cdots,B_n$ 覆盖 A：

$$A\subset \bigcup_{i=1}^{n} B_i\ 。$$

注: 这个定理也称为**有限覆盖定理**，它的逆命题也成立。

定理 2.2.6.2 设 A 是度量空间 (X,ρ) 中的点集，若 X 中每个覆盖 A 的开集族中必有有限个开集覆盖 A，则 A 是紧集。

根据定理 2.2.6.1 和定理 2.2.6.2，我们可以给出度量空间中紧集的另一个定义：

定义 2.2.6.3 （紧集） 设 A 是度量空间 (X,ρ) 中的子集，若 X 中每个覆盖 A 的开集族中必有有限个开集覆盖 A，则称 A 是**紧集**。

2.2.6.3 紧集上的连续映射

我们现在把闭区间上的连续函数的基本性质拓广到度量空间的紧集上。

定理 2.2.6.3 设 A 是度量空间 (X,ρ) 中的紧集，f 是 A 上的连续映射，则 A 的象 $B=f(A)$ 也是紧集。

推论 1 度量空间上的连续映射必然把致密集映射成致密集。

推论 2 度量空间中的紧集 A 上的连续函数 f 必然有界，而且上、下确界可以达到。

推论 3 集上的一对一的连续映射必是同胚映射。

证明: 设 f 是紧集 A 到 B 上的一对一的连续映射，往证逆映射 f^{-1} 也是连续的。事实上，只要证明 f^{-1} 的逆映射 f 将 A 的任意闭子集 E 映射成闭集即可。因为 A 是紧集，所以 A 的闭子集 E 也是紧集。因此 $f(E)$ 也是紧集。f 和 f^{-1} 一样都是连续映射，故 f 是拓扑映射。 证毕!

2.2.6.4 有限维赋范线性空间

有限维赋范线性空间又称为 **Minkowski 空间**。

定理 2.2.6.4 设 $(X_n,\|\bullet\|)$ 是 n 维赋范线性空间，$e_1,e_2,\cdots,e_n$ 是 X_n 的一个基，则必存在正数 C_1, C_2，使得对于 $\forall x=\sum_{i=1}^{n}x_ie_i\in X_n$，有

$$C_1\sqrt{\sum_{i=1}^{n}|x_i|^2}\le\|x\|\le C_2\sqrt{\sum_{i=1}^{n}|x_i|^2}\ , \tag{2.2.6.1}$$

并且映射

$$A:(x_1,x_2,\cdots,x_n)\mapsto\sum_{i=1}^{n}x_ie_i \tag{2.2.6.2}$$

是 n 维欧几里得空间 $\mathbf{R}^n$ 到 X_n 的同胚映射。

证明: 因为 $x=\sum_{i=1}^{n}x_ie_i\in X_n$，所以

$$\|x\|\le\sum_{i=1}^{n}|x_i|\|e_i\|\le\sqrt{\sum_{i=1}^{n}\|e_i\|^2}\times\sqrt{\sum_{i=1}^{n}|x_i|^2}\ . \tag{2.2.6.3}$$

取 $C_2=\sqrt{\sum_{i=1}^{n}\|e_i\|^2}$，则 $C_2>0$，且 $\|x\|\le C_2\sqrt{\sum_{i=1}^{n}|x_i|^2}$.

另一方面，作 $\mathbf{R}^n$ 中的单位球面：$S=\{(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n\ \big|\ \sum_{i=1}^{n}|x_i|^2=1\}$，考察 S 上的函数 $f(x_1,x_2,\cdots,x_n)=\left\|\sum_{i=1}^{n}x_ie_i\right\|$。显然 f 在 S 上处处大于零。现在证明 f 在 S 上的下确界 $C_1=\inf\limits_{(x_1,x_2,\cdots,x_n)\in S}\left\|\sum_{i=1}^{n}x_ie_i\right\|>0$。

因为 S 是 $\mathbf{R}^n$ 中的有界闭集，所以 S 是紧集。由**定理 2.2.6.3 的推论 2** 知：只要证明 f 是连续的，则 f 在 S 上的下确界 C_1 是 f 在 S 上某点的函数值，这样就能得到 $C_1>0$。

由式（2.2.6.3）得：$\left|f(x_1,x_2,\cdots,x_n)-f(y_1,y_2,\cdots,y_n)\right|=\left\|\sum_{i=1}^{n}x_ie_i\right\|-\left\|\sum_{i=1}^{n}y_ie_i\right\|$

$\le\left\|\sum_{i=1}^{n}(x_i-y_i)e_i\right\|\le C_2\sqrt{\sum_{i=1}^{n}|x_i-y_i|^2}$，因此 f 是 S 上的连续函数；由于 S 上没有零向量，根据范数的唯一性条件得 $C_1>0$。对于 $\mathbf{R}^n$ 中的非零向量 $x=\sum_{i=1}^{n}x_ie_i$，作 $\xi_j=\dfrac{x_j}{\sqrt{\sum_{i=1}^{n}|x_i|^2}}$，$j=1,2,\cdots,n$，则 $\sum_{j=1}^{n}|\xi_j|^2=1$，因此

$\dfrac{\left\|\sum_{i=1}^{n}x_ie_i\right\|}{\sqrt{\sum_{i=1}^{n}|x_i|^2}}=f(\xi_1,\xi_2,\cdots,\xi_n)\ge C_1$，即 $\left\|\sum_{i=1}^{n}x_ie_i\right\|\ge C_1\sqrt{\sum_{i=1}^{n}|x_i|^2}$，于是式（2.2.6.1）得证！

由式（2.2.6.1）得：$\|A\xi-A\eta\|\le C_2\|\xi-\eta\|$，且

$$\left\|A^{-1}x-A^{-1}y\right\|\le\frac{1}{C_2}\|x-y\|,\qquad(2.2.6.4)$$

故 A, A^{-1} 都是连续的映射。又显然 A 是 $\mathbf{R}^n$ 到 X_n 上的一一对应，因此 A 是 n 维欧几里得空间 $\mathbf{R}^n$ 到 X_n 的拓扑映射。证毕！

推论 1 设在有限维线性空间 X_n 上定义了两个范数 $\|\bullet\|_1$ 和 $\|\bullet\|_2$，则必存在常数 $k_1>0$ 和 $k_2>0$，使得对 $\forall x\in X_n$，都有 $k_1\|x\|\le\|x\|_1\le k_2\|x\|$。

为了说明**定理 2.2.6.4** 的结论，我们对一般的线性空间（不一定是有限维的）引入如下概念。

定义 2.2.6.4 设 X 是一个线性空间，$\|\bullet\|_1$，$\|\bullet\|_2$ 是定义在 X 上的两个范数。若存在正数 c_1,c_2，使得对一切 $x\in X$，都有

$$c_1\|x\|_2\le\|x\|_1\le c_2\|x\|_2,\qquad(2.2.6.5)$$

则称范数 $\|\bullet\|_1$ 和 $\|\bullet\|_2$ 是**等价**的。

注:设在线性空间 X 上有两个范数 $\|\bullet\|_1$，$\|\bullet\|_2$，X 按这两个范数成为赋范线性空间，记为 $(X,\|\bullet\|_1)$ 和 $(X,\|\bullet\|_2)$。则范数 $\|\bullet\|_1$ 和 $\|\bullet\|_2$ 等价的充要条件是在 $(X,\|\bullet\|_1)$ 和 $(X,\|\bullet\|_2)$ 中点列收敛的概念是一致的，即：

$\|x_n-x\|_1\to0\ (n\to\infty)$ 与 $\|x_n-x\|_2\to0\ (n\to\infty)$ 是等价的。

因此，当范数 $\|\bullet\|_1$ 和 $\|\bullet\|_2$ 等价时，赋范线性空间 $(X,\|\bullet\|_1)$ 和 $(X,\|\bullet\|_2)$ 是同胚的。

定理 2.2.6.4 的推论 1 说明：在有限维线性空间上，任何两个范数都是等价的，任何两个 n 维赋范线性空间都是同胚的。

推论 2 有限维赋范线性空间是完备的。

由于度量空间的完备子空间是闭的，所以得：

推论 3 任意赋范线性空间的有限维线性子空间是闭子空间。

从定理 2.2.6.4 又可得到：

定理 2.2.6.5 有限维赋范线性空间中任何有界集都是致密的。

证明：设 X_n 是 n 维赋范线性空间，则 X_n 和 n 维欧几里得空间 $\mathbf{R}^n$ 同胚。记 X_n 到 $\mathbf{R}^n$ 的拓扑映射为 f，则 f^{-1} 也是拓扑映射。对于 X_n 中的有界集 A，$f(A)$ 是 $\mathbf{R}^n$ 中的有界集。由**定理 2.2.6.1** 知：$f(A)$ 是 $\mathbf{R}^n$ 中的致密集，而致密集在连续映射 f^{-1} 下仍是致密的，故 $f(A)$ 的原象 A 是 X_n 中的致密集。证毕！

反之，我们可以证明：**若在一个赋范线性空间中每个有界集都是致密的，则这个空间一定是有限维的。**

为此,先介绍 F. Riesz 的一个引理。在欧几里得空间中，对任一线性真子空间，必有一单位向量与此子空间的“距离”等于 1。但对于一般的赋范线性空间，F. Riesz 证明了下面的 **Riesz 引理。**

Riesz 引理 设 E 是赋范线性空间 X 的闭子空间，且 $E \neq X$，则对 $\forall \varepsilon\ (0<\varepsilon<1)$，必存在 X 中的单位向量 x_0, $\|x_0\|=1$，使得 $\rho(x_0,E)>\varepsilon$.

证明：因为 E 是 X 的真子集，任取一点 $\overline{x} \in X-E$，又因为 E 是闭的，所以 $\rho(\overline{x},E)=d>0$.

若不然，即 $\rho(\overline{x},E)=0$，则 $\overline{x} \in \overline{E}=E$，矛盾!因为 $d/\varepsilon > d$，所以必有 $x' \in E$，满足 $\|\overline{x}-x'\|<d/\varepsilon$。作 $x_0=\dfrac{\overline{x}-x'}{\|\overline{x}-x'\|}$，则 $\|x_0\|=1$. 对 $\forall x \in E$，因为 $x'+\|\overline{x}-x'\|x \in E$，所以 $\left\|\overline{x}-\left(x'+\|\overline{x}-x'\|x\right)\right\| \geq d$，故 $\|x_0-x\|=\left\|\dfrac{\overline{x}-x'}{\|\overline{x}-x'\|}-x\right\|=\dfrac{1}{\|\overline{x}-x'\|}\left\|\overline{x}-\left(x'+\|\overline{x}-x'\|x\right)\right\| \geq \dfrac{d}{\|\overline{x}-x'\|}$.于是 $\rho(x_0,E) \geq \dfrac{d}{\|\overline{x}-x'\|}>\varepsilon$.证毕！

下面的定理指出了有限维空间和无限维空间的一个本质性的差别。

定理 2.2.6.6 若赋范线性空间 X 是无限维的，则 X 中必有不致密的有界集。

证明：证明单位球 $S=\{x \in X \mid \|x\| \leq 1\}$ 就不是致密集。事实上，在 X 中任取一个单位向量 x_1，$(\|x_1\|=1)$，用 $X_1=\{x \mid x=\alpha x_1,\ \alpha \text{ 是数}\}$ 表示由 x_1 张成的一维子空间，显然 $X_1 \neq X$。由**定理 2.2.6.4 的推论 3** 知：X_1 是 X 的闭子空间。由 Riesz 引理得：存在单位向量 $x_2 \in X$ $(\|x_2\|=1)$，使得 $\rho(x_2,X_1)>1/2$.用 X_2 表示由 x_1,x_2 张成的二维子空间，则 X_2 是 X 的闭子空间，且 $X_2 \neq X$。于是又可以对 X_2 运用 Riesz 引理。这样继续做下去，从 X 中选取了一列单位向量 $\{x_k \mid k=1,2,3,\cdots\}$ 和 X 的一列闭子空间 $\{X_k \mid k=1,2,3,\cdots\}$，$\{x_1,x_2,\cdots,x_k\} \subset X_k$，且 $\rho(x_{k+1},X_k)>\dfrac{1}{2}$，$k=1,2,\cdots$。因此，当 $m>n$ 时，$x_n \in X_{m-1}$，且 $\|x_m-x_n\| \geq \rho(x_m,X_{m-1})>\dfrac{1}{2}$.这样的点列 $\{x_k \mid k=1,2,3,\cdots\}$ 不可能含有收敛的子序列。于是有界集 $S=\{x \in X \mid \|x\| \leq 1\}$ 就不是致密集。证毕！

综上所述，得：

定理 2.2.6.7 一个赋范线性空间 X 是有限维的 $\Leftrightarrow$ X 的任何有界集都是致密的。

定理 2.2.6.8 一个赋范线性空间 X 是无限维的 $\Leftrightarrow$ X 中必有不致密的有界集。

注：上述定理指出了有限维空间和无限维空间的一个本质性的差别。

2.3 线性算子理论基础

2.3.1 线性算子与线性泛函的定义

算子概念起源于运算。例如，代数运算、求导运算、求不定积分和定积分、把平面上的向量绕坐标原点旋转一个角度等等。**在泛函分析中，通常把映射称为算子，而取值于实数域或复数域的算子也称为泛函数，简称为泛函**。本节着重考察赋范线性空间上的线性算子，它是出现在各个数学领域中具有线性性质的运算（例如线性代数中的线性变换；微分方程论、积分方程论中大量出现的微分、积分运算、积分变换等）的抽象概括。它是线性泛函分析研究的重要对象。

定义 2.3.1 设 F 是实数域或复数域，X,Y 是 F 上的两个线性空间，D 是 X 的线性子空间，$T:D\to Y$ 是一个映射。对 $\mathbf{x}\in D$，记 $\mathbf{x}$ 经 T 映射后的象为 $T\mathbf{x}$ 或 $T\left(\mathbf{x}\right)$。若对 $\forall \mathbf{x},\mathbf{y}\in D$ 及数 $\alpha,\beta\in F$，有

$$T(\alpha x+\beta y)=\alpha T(x)+\beta T(y)\ （或\ =\alpha Tx+\beta Ty）\tag{2.3.1}$$

则称 T 是**线性算子**。称 D 是 T 的**定义域**，记为 $\mathcal{D}(T)$；称集 $T(D)=\{Tx\mid x\in D\}$（或 TD）为 T 的**值域**（或**象域**），记为 $\mathcal{R}(T)$。

取值为实数或复数的线性算子 T（即：$\mathcal{R}(T)\subset\mathbf{F}$，$\mathbf{F}=\mathbf{R}^1$ 或 $\mathbf{C}^1$）分别称为**实的**或**复的线性泛函**，统称为**线性泛函**。**注：**后面所讨论的算子（泛函）都是线性算子（线性泛函）。

例 2.3.1 设 $X=C^1[0,1],\ Y=B[0,1]$（$[0,1]$ 上有界函数全体），定义 $(Tx)(t)=\dfrac{\mathrm{d}}{\mathrm{d}t}x(t)$，则 T 是 X 到 Y 的线性算子。

例 2.3.2 设 $X=C[a,b]$，$K(t,s)$ 是 $[a,b]\times[a,b]$ 上的二元连续函数，定义 $(Tx)(t)=\int_a^b K(t,s)x(s)\mathrm{d}s$。则 T 是 X 到 X 的线性算子。

例 2.3.3 设 $X=C[a,b]$，定义 $Tx=\int_a^b x(t)\mathrm{d}t$，则 T 是 X 上的线性泛函。

定义 2.3.2 设 $T_1,\ T_2$ 是 X 到 Y 的线性算子，它们的定义域分别是 $\mathcal{D}(T_1),\ \mathcal{D}(T_2)$。

（1） 对任一数 α,定义算子 αT_1：它以 $\mathcal{D}(T_1)$ 为定义域，而对任何 $x\in\mathcal{D}(T_1)$，

$$(\alpha T_1)x=\alpha(T_1x)\tag{2.3.2}$$

（2） 定义算子 T_1+T_2:它以 $\mathcal{D}(T_1)\bigcap\mathcal{D}(T_2)$ 为定义域，而对任何 $x\in\mathcal{D}(T_1)\bigcap\mathcal{D}(T_2)$，

$$(T_1+T_2)x=T_1x+T_2x\tag{2.3.3}$$

（3） 设 T_3 是以 $\mathcal{D}(T_1)$ 为定义域的 Y 到 Z 的线性算子，定义算子 $T_3\cdot T_1$（也记作 T_3T_1）。它以 $D=\left\{x\mid T_1x\in\mathcal{D}(T_3),\ x\in\mathcal{D}(T_1)\right\}$ 为定义域，而对任何 $x\in D$，

$$(T_3\cdot T_1)x=T_3(T_1x).\tag{2.3.4}$$

注：易知 αT_1（称 T_1 的 α 倍），T_1+T_2（称 T_1 与 T_2 的**和**），$T_3\cdot T_1$（称 T_3 与 T_1 的**积**）仍是线性算子。

定义 2.3.3 设T是以$\mathcal{D}(T)$为定义域的X到Y的线性算子。若从$x\neq 0$（$x\in\mathcal{D}(T)$）可推出$Tx\neq 0$，即T是单射，则T有逆映射T^{-1}：

$$T^{-1}(Tx)=x. \tag{2.3.5}$$

T^{-1}是一个以$\mathcal{R}(T)=\{Tx \mid x\in\mathcal{D}(T)\}$为定义域的线性算子，称$T^{-1}$为$T$的**逆算子**。设$T$是$X$到$X$的线性算子。如果对任何$x\in X$,有

$$Tx=x, \tag{2.3.6}$$

则称T是X的**单位算子**，或**恒等算子**,常用I_X表示,或简记为I。**注：**T^{-1}为T的逆算子当且仅当：

$$T^{-1}T=I_{\mathcal{D}(T)},\quad TT^{-1}=I_{\mathcal{R}(T)} \tag{2.3.7}$$

2.3.2 线性算子的有界性和连续性

定理 2.3.1 设T是赋范线性空间X到赋范线性空间Y的线性算子。若T在某一点$x_0\in\mathcal{D}(T)$连续，则T在$\mathcal{D}(T)$上处处连续。

定义 2.3.4 若算子T将其定义域$\mathcal{D}(T)$中的每个有界集都映射成一个有界集，则称T是**有界算子**。不是有界的算子就称为**无界算子**。

定理 2.3.2 设T是赋范线性空间X到赋范线性空间Y的线性算子，则T是有界算子的充要条件是：存在常数$M\geq 0$，使得对一切$x\in X$，有

$$\|Tx\|\leq M\|x\|。 \tag{2.3.8}$$

证明：“$\Rightarrow$”设T是有界线性算子，则T把单位球面$S=\{y \mid \|y\|=1,\ y\in X\}$映射成$Y$中的一个有界集。因此存在常数$M\geq 0$，对一切$y\in S$，都有$\|Ty\|\leq M$。当$\mathbf{x}=0$时，式（2.3.8）自然成立。当$\mathbf{x}\neq 0$，作$\frac{\mathbf{x}}{\|\mathbf{x}\|}$，由$\frac{\mathbf{x}}{\|\mathbf{x}\|}\in S$得：$\left\|T\frac{\mathbf{x}}{\|\mathbf{x}\|}\right\|=\left\|\frac{T\mathbf{x}}{\|\mathbf{x}\|}\right\|\leq M$，即$\|T\mathbf{x}\|\leq M\|\mathbf{x}\|$因而（2.3.8）对一切$x\in X$都成立。

“$\Leftarrow$”若（2.3.8）成立。设$A\subset X$是一个有界集，则存在常数$K\geq 0$，使得当$x\in A$时，$\|x\|\leq K$。因此，由（4.1.8）得：对一切$x\in A$，有$\|Tx\|\leq MK$，即TA是有界集。证毕！

注：如无特殊说明，有界线性算子T的定义域$\mathcal{D}(T)$总假定是全空间X，即$\mathcal{D}(T)=X$。因此，对赋范线性空间上的线性算子，以后可以用式（2.3.8）作为有界性的定义。

定义 2.3.5 设T是赋范线性空间X到赋范线性空间Y的线性算子，称

$$\|T\|=\sup_{\mathbf{x}\neq 0}\left\|\frac{T\mathbf{x}}{\|\mathbf{x}\|}\right\|\left(\mathbf{x}\in X\right) \tag{2.3.9}$$

为**算子 *T* 的范数**。

注：由**定理 2.3.2** 立得：有界线性算子的范数是有限的。

命题 2.3.1 对有界线性算子T，有

（1） $\|Tx\| \le \|T\|\|x\| \quad (x \in X)$; （2.3.10）

（2） $\|T\| = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|\le 1} \|Tx\| \quad (x \in X)$. （2.3.11）

证明：（1）因为 T 是有界线性算子，所以由定理 2.3.1 得：存在常数 $M \ge 0$，使得对一切 $x \in X$，有 $\|Tx\| \le M\|x\|$，于是 $\dfrac{\|Tx\|}{\|x\|} \le M$。

再考虑到 $\|T\| = \sup_{x\ne 0} \dfrac{\|Tx\|}{\|x\|}$，得 $\dfrac{\|Tx\|}{\|x\|} \le \|T\|$，即 $\|Tx\| \le \|T\|\|x\|$。

（2）一方面，显然 $\|T\| \ge \sup_{\|x\|\le 1} \|Tx\| \ge \sup_{\|x\|=1} \|Tx\|$. 另一方面，对 $\forall y\,(\ne 0) \in X$，因为 $\left\| \dfrac{y}{\|y\|} \right\| = 1$，所以 $\left\| T\dfrac{y}{\|y\|} \right\| = \dfrac{\|Ty\|}{\|y\|} \le \sup_{\|x\|=1} \|Tx\|$，故 $\sup_{y\ne 0} \dfrac{\|Ty\|}{\|y\|} \le \sup_{\|x\|=1} \|Tx\|$，即：$\|T\| \le \sup_{\|x\|=1} \|Tx\|$. 于是 $\|T\| = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|\le 1} \|Tx\|$。证毕！

注：当 $T = I$（单位算子）时，$\|I\| = 1$。

称 $\dfrac{\|Tx\|}{\|x\|}$ 为 T 在 x 方向的**伸张系数**，$\|T\|$ 的几何意义是一切方向伸张系数的上确界。一般说来，求出具体算子的范数的值并不容易。

例 2.3.4 设赋范线性空间 $L[a,b], C[a,b]$ 上的范数分别记为

$$\|f\|_L = \int_a^b |f(t)|\mathrm{d}t \quad (f \in L[a,b]), \qquad \|g\|_C = \max_{t\in[a,b]} |g(t)| \quad (g \in C[a,b]),$$

定义 $L[a,b]$ 到 $C[a,b]$ 的算子 T：对 $\forall f \in L[a,b]$，

$$(Tf)(x) = \int_a^x f(t)\mathrm{d}t, \tag{2.3.12}$$

则 $\|T\| = 1$.

证明： $\forall f \in L[a,b]$，使 $\|f\|_L = 1$. 因为

$$\begin{aligned} \|Tf\|_C &= \max_{x\in[a,b]} |(Tf)(x)| = \max_{x\in[a,b]} \left| \int_a^x f(t)\mathrm{d}t \right| \\ &\le \max_{x\in[a,b]} \int_a^x |f(t)|\mathrm{d}t \le \int_a^b |f(t)|\mathrm{d}t = \|f\|_L = 1, \end{aligned}$$

即 $\|T\| = \sup_{\|f\|_L=1} \|Tf\|_C \le 1$.

另一方面，取 $f_0(t) = \dfrac{1}{b-a} \in L[a,b]$，$\|f_0\|_L = \int_a^b |f_0(t)|\mathrm{d}t = \int_a^b \dfrac{1}{b-a}\mathrm{d}t = 1$，则又有

$$\|T\| = \sup_{\|f\|_L=1} \|Tf\|_C \geq \|Tf_0\|_C = \max_{x\in[a,b]} |(Tf_0)(x)| = \max_{x\in[a,b]} \left|\int_a^x f_0(t)\mathrm{d}t\right|$$
$$= \max_{x\in[a,b]} \left|\int_a^x \frac{1}{b-a}\mathrm{d}t\right| = \int_a^b \frac{1}{b-a}\mathrm{d}t = 1,$$

即$\|T\| \geq 1$. 综上所述，有$\|T\| = 1$.

定理 2.3.3 设X, Y是赋范线性空间，$T: X \to Y$是线性算子，则T是有界的充要条件是T是连续的。

2.3.3 有界线性算子

定义 2.3.6 设X和Y是两个线性空间，$(X \to Y)$表示由X到Y的线性算子的全体，即$(X \to Y) = \{T \mid T$是X到Y的线性算子$\}$。当$A, B \in (X \to Y)$，α是数时，对$\forall x \in X$，

$$(A+B)x = Ax + Bx, \quad (\alpha A)x = \alpha(Ax), \tag{2.3.13}$$

并称$A+B$为算子A与B的**和**；αA为数α与算子A的**积**。$(X \to Y)$按上述线性运算构成一个线性空间。

定理 2.3.4 设X和Y是两个赋范线性空间，$\mathscr{B}(X \to Y)$表示由X到Y的有界线性算子的全体，即$\mathscr{B}(X \to Y) = \{T \mid T$是$X$到$Y$的有界线性算子$\}$。若$A, B \in \mathscr{B}(X \to Y)$，则$A+B \in \mathscr{B}(X \to Y)$，$\alpha A \in \mathscr{B}(X \to Y)$，且$\mathscr{B}(X \to Y)$按线性运算式（2.3.13）构成一个线性空间。

若以式（2.3.9）定义的算子范数作为范数

$$\|T\| = \sup_{x\neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|\leq 1} \|Tx\| = \sup_{\|x\|=1} \|Tx\| \quad (x \in X), \tag{2.3.14}$$

则$\mathscr{B}(X \to Y)$是一个赋范线性空间。

定义 2.3.7 设X是赋范线性空间，X上的连续线性范函全体记做X^*，即

$X^* = \{f \mid f$是X上的连续线性泛函$\}$，

它按通常的线性运算：当$f, g \in X^*$，α是数时，对$\forall x \in X$，规定

$$(f+g)(x) = f(x) + g(x), \qquad (\alpha f)(x) = \alpha f(x); \tag{2.3.15}$$

及泛函的范数

$$\|f\| = \sup_{x\neq 0} \frac{|f(x)|}{\|x\|} = \sup_{\|x\|\leq 1} |f(x)| = \sup_{\|x\|=1} |f(x)| \quad (x \in X), \tag{2.3.16}$$

成为一个赋范线性空间，称为X的**共轭空间**。

注：由定理 2.3.4 知：若$f \in X^*$，则f一定有界；于是$X^* = \mathscr{B}(X \to Y)$，其中$Y = \mathbf{R}^1$或$\mathbf{C}^1$。

2.3.4 算子序列的收敛性

在经典分析中，一列函数的收敛性常常用到的是处处收敛和一致收敛的概念。由于所考察的问题的需要，不同场合采用不同的收敛概念。对于算子序列类似于函数列的一致收敛和处处收敛，也常常用到下面几种形式的收敛性。

定义 2.3.8 设 X 和 Y 都是赋范线性空间，若 $A_n, A \in \mathscr{B}(X \to Y)\ (n=1,2,\cdots)$，且

$$\|A_n - A\| \to 0 \quad (n \to \infty), \tag{2.3.17}$$

则称序列 $\{A_n\}$ **按算子范数收敛于** A，或**一致收敛于** A。若对每个 $x \in X$，都有

$$\|(A_n - A)x\| \to 0 \quad (n \to \infty) \tag{2.3.18}$$

则称序列 $\{A_n\}$ **强收敛于** A，记作

$$A_n \xrightarrow{\text{强}} A \quad \text{或} \quad A = (\text{强})\lim_{n\to\infty} A_n \quad \text{或} \quad A = (S)\lim_{n\to\infty} A_n .$$

若对每个 $x \in X$ 以及 $f \in Y^*$，都有

$$f(A_n x) \to f(Ax) \quad (n \to \infty) \tag{2.3.19}$$

则称序列 $\{A_n\}$ **弱收敛于** A，记作

$$A_n \xrightarrow{\text{弱}} A \quad \text{或} \quad A = (\text{弱})\lim_{n\to\infty} A_n \quad \text{或} \quad A = (W)\lim_{n\to\infty} A_n .$$

注：若序列 $\{A_n\}$ 一致收敛于 A，则 $\{A_n\}$ 必强收敛于 A；若序列 $\{A_n\}$ 强收敛于 A，则 $\{A_n\}$ 必弱收敛于 A。下面的例子说明：它们的逆命题一般不正确。

例 2.3.5（**强收敛而不一致收敛的算子序列**） 在 $\boldsymbol{l}^p\ (p \ge 1)$ 中，作“左移”算子 A 如下：当 $\boldsymbol{x} = (x_1, x_2, \cdots) \in \boldsymbol{l}^p$ 时，$A\boldsymbol{x} = (x_2, x_3, \cdots)$。则 A 是 $\boldsymbol{l}^p$ 到 $\boldsymbol{l}^p$ 上的有界线性算子。事实上，A 是 $\boldsymbol{l}^p$ 到 $\boldsymbol{l}^p$ 上的线性算子是显然的。因为 $\|A\boldsymbol{x}\|_p = \left(\sum_{i=2}^{\infty} |x_i|^p\right)^{1/p} \le \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p} = \|\boldsymbol{x}\|_p$，所以 A 是 $\boldsymbol{l}^p$ 到 $\boldsymbol{l}^p$ 上的有界算子。

而算子序列 $\{A^n\}$ 强收敛于零算子 0；但不一致收敛于零算子 0。事实上，对 $\forall \boldsymbol{x} = (x_1, x_2, \cdots) \in \boldsymbol{l}^p$，$A^n\boldsymbol{x} = (x_{n+1}, x_{n+2}, \cdots)$. 而

$$\|(A^n - 0)\boldsymbol{x}\|_p = \left(\sum_{i=n+1}^{\infty} |x_i - 0|^p\right)^{1/p} = \left(\sum_{i=n+1}^{\infty} |x_i|^p\right)^{1/p},$$

因为 $\left(\sum_{i=1}^{\infty} |x_i|^p\right)^{\frac{1}{p}} < \infty$，所以 $\|(A^n - 0)\boldsymbol{x}\|_p \to 0$，即序列 $\{A^n\}$ 强收敛于零算子 0。但是序列 $\{A^n\}$ 在 $\boldsymbol{l}^p$ 并不一致收敛于零算子 0。事实上，令 $\boldsymbol{e}_n = (0, \cdots, 0, 1, 0, \cdots)$（除第 n 个坐标为 1 外，其余坐标均为 0），显然 $A^n\boldsymbol{e}_{n+1} = \boldsymbol{e}_1$。因为

$$\|A^n - 0\| = \|A^n\| = \sup_{\|x\|=1} \|A^n\boldsymbol{x}\|_p \ge \|A^n\boldsymbol{e}_{n+1}\|_p = \|\boldsymbol{e}_1\|_p = 1,$$

所以序列 $\{A^n\}$ 在 $\boldsymbol{l}^p$ 并不一致收敛于零算子 0。

注： $\|A\|=\sup\limits_{\|x\|_p=1}\|Ax\|_p=\left(\sum\limits_{i=2}^{\infty}|x_i|^p\right)^{1/p}\le\left(\sum\limits_{i=1}^{\infty}|x_i|^p\right)^{1/p}=\|x\|_p=1$.取 $e_2=(0,1,0,\cdots)$（除第 2 个坐标为 1 外，其余坐标均为 0），则 $Ae_2=e_1$，且 $\|Ae_2\|_p=\|e_1\|_p=1$，于是 $\|A\|=1$。

例 2.3.6（弱收敛而不强收敛的算子序列） 在 $l^p\ (p\ge1)$ 中，作算子序列 $\{A_n\}$ 如下：当 $x=(x_1,x_2,\cdots)\in l^p$ 时，$A_nx=x_1e_n$，$n=1,2,\cdots$，其中 $e_n=(0,\cdots,0,1,0,\cdots)$（除第 n 个坐标为 1 外，其余坐标均为 0）。显然，A 是有界线性算子，且

$$\|(A_n-A_m)x\|_p=\|A_nx-A_mx\|_p=\|x_1e_n-x_1e_m\|_p=\left(|x_1|^p+|-x_1|^p\right)^{1/p}=|x_1|2^{1/p}$$

故当 $x_1\ne0$ 时，序列 $\{A_n\}$ 不是强收敛的。

然而，对 l^p 上任意连续线性泛函 y（必属于 l^q（见 4.2）），即 $y=(y_1,y_2,\cdots)\in l^q$，且对 $x=(x_1,x_2,\cdots)\in l^p$，有 $y(x)=\sum\limits_{i=1}^{\infty}x_iy_i$，$y(A_nx)=y(x_1e_n)=x_1y_n$．因为 $\sum\limits_{n=1}^{\infty}|y_n|^q<\infty$，所以 $y_n\to0\ \ (n\to\infty)$，于是 $y(A_nx)\to0\ \ (n\to\infty)$。这说明 $\{A_n\}$ 弱收敛于零。

3 泛函分析基础知识在地震波反演成像中的应用实例

地震波反演成像发展到以 FWI+LS_RTM 为主导的阶段，泛函分析知识对于成像方法的研究和理解已经变得十分必要。尽管 FWI 方法以非线性映射算子为标志，实质上到目前为止我们并没有有效的对应非线性映射的反演方法。凡是基于梯度导引类的优化反演算法本质上还是基于线性映射算子的。因此，前述的线性空间和线性算子基本理论基本上可以满足当前勘探地震中地震信号分析和地震波反演成像的需求。

勘探地震中地震信号分析基本上都是基于信号的线性预测算子（有界的平方可积算子），然后在估计理论的基础上实现信号去噪、插值（数据规则化）、压缩等等处理。地震波反演成像的核心工作：层析速度反演和 LS_PSDM 也是基于到达时和散射波场的线性预测算子，在估计理论的基础上实现背景速度和高波数速度的估计。可以认为，无论是信号分析或是反演成像，都是在 Hilbert 空间中进行的。把信号分析和反演成像奠定在泛函分析的基础上，可以得到对方法理论的更简洁的表达和更深入的理解。

3.1 散射波算子方程的导出

为了与前述泛函分析基础知识衔接，下面给出用于反演成像的线性算子的导出。在常密度声介质中，地震波传播受如下声波方程控制，

$$\frac{1}{v^2(\mathbf{x})}\frac{\partial^2u_{total}(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2}-\nabla^2u_{total}(\mathbf{x},t;\mathbf{s}_\mathbf{x})=f(t;\mathbf{s}_\mathbf{x}) \tag{3.1}$$

其中，$v(\mathbf{x})$ 为准确速度场，$f(t;\mathbf{s}_\mathbf{x})$ 为震源子波函数，t 为时间，$\mathbf{s}_\mathbf{x}=(s_x,s_y,s_z)$ 为炮点位置，$\mathbf{x}=(x,y,z)$ 为空间位置点。u_{total} 为地震全波场，它包含声介质中的所有波现象，比如直达波、回转波、折射波、反射波和多次波等。该波场在检波点位置处 $\mathbf{g}_\mathbf{x}=(g_x,g_y,g_z)$ 被接收到即形成了观测的地震波场，

$$d_{total}(\mathbf{g}_\mathbf{x},t;\mathbf{s}_\mathbf{x})=\delta(\mathbf{x}-\mathbf{g}_\mathbf{x})u_{total}(\mathbf{x},t;\mathbf{s}_\mathbf{x})$$

式（3.1）定义的全声波方程通常抽象地写成如下的算子方程的形式：

$$\mathbf{d}_{total}=\mathbf{L}_{total}(v)$$

其中 L_{total} 为全波场模拟算子。

地震勘探是由观测波场获取地下介质参数信息的一个过程，该过程可以表达为经典的反演问题，即已知观测数据，利用波动方程的投影关系，估计地下速度场（或参数场）信息。但是非线性的反投影目前缺乏有效的方法理论支撑。依据前述线性算子理论，需要导出线性算子投影方程。为此，采用 Born 近似来实现非线性问题的逐步线性化是比较普遍的方式。假设速度场可以分解为背景速度和速度扰动，

$$v(\mathbf{x})=v_0(\mathbf{x})+\delta v(\mathbf{x}) \tag{3.2}$$

同样的，将全波场也对应的分解为背景波场 u_0 和散射波场 u_s，

$$u_{total}=u_0+u_s \tag{3.3}$$

将公式（3.2）和（3.3）代入公式（3.1），整理可以得到如下方程，

$$\begin{aligned}&\frac{1}{(v_0(\mathbf{x})+\delta v(\mathbf{x}))^2}\frac{\partial^2(u_0(\mathbf{x},t;\mathbf{s}_\mathbf{x})+u_s(\mathbf{x},t;\mathbf{s}_\mathbf{x}))}{\partial t^2}-\\&\nabla^2(u_0(\mathbf{x},t;\mathbf{s}_\mathbf{x})+u_s(\mathbf{x},t;\mathbf{s}_\mathbf{x}))=f(t;\mathbf{s}_\mathbf{x})\end{aligned} \tag{3.4}$$

在速度扰动较小（即背景速度足够准确）的假设下，存在如下关系，

$$\frac{1}{(v_0(\mathbf{x})+\delta v(\mathbf{x}))^2}\approx\frac{1}{v_0^2(\mathbf{x})}-\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})} \tag{3.5}$$

将公式（3.5）代入公式（3.4）整理可以得到背景波场传播和散射波场传播的方程，

$$\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2u_0(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2}-\nabla^2u_0(\mathbf{x},t;\mathbf{s}_\mathbf{x})=f(t;\mathbf{s}_\mathbf{x}) \tag{3.6}$$

$$\begin{aligned}&\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2u_s(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2}-\nabla^2u_s(\mathbf{x},t;\mathbf{s}_\mathbf{x})\\&=\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2u_0(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2}+\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2u_s(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2}\end{aligned} \tag{3.7}$$

忽略高阶散射项，散射波场 u_s 退化为一阶散射波场 u_1，其对应的传播方程为，

$$\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2u_1(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2}-\nabla^2u_1(\mathbf{x},t;\mathbf{s}_\mathbf{x})=\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2u_0(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2} \tag{3.8}$$

在速度扰动量较小时，该一次散射波场可以近似代表地震记录中的一次反射波场及折射波。令

$$m(\mathbf{x})=\frac{2\delta v(\mathbf{x})}{v_0(\mathbf{x})} \tag{3.9}$$

其中 $\mathbf{m}(\mathbf{x})$ 为一个无量纲参数，通常将其叫做散射强度。由散射强度描述的一次散射波场的传播方程为，

$$\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2u_1(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2}-\nabla^2u_1(\mathbf{x},t;\mathbf{s}_\mathbf{x})=m(\mathbf{x})\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2u_0(\mathbf{x},t;\mathbf{s}_\mathbf{x})}{\partial t^2} \tag{3.10}$$

方程（3.10）所描述的波传播过程如图 2 所示，即由炮点出射的波场传播到散射点后，经过散射点作用产生一次散射波，该散射波进一步传播到检波点形成一次反射波记录。

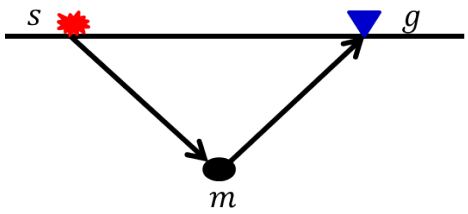


图 2 一阶散射波传播示意图

这样，由方程（3.10）得到的反射地震记录可以表达为，

$$d_1(\mathbf{g}_x,t;\mathbf{s}_x)=\delta(\mathbf{x}-\mathbf{g}_x)u_1(\mathbf{x},t;\mathbf{s}_x) \tag{3.11}$$

由方程（3.10）可以看出，散射波是震源函数、背景速度和散射强度的函数。因此可以利用散射波来实现背景速度和反射强度的反演。方程（3.10）也可以简洁地表示为如下的算子方程：

$$\mathbf{d}_1=\mathbf{L}_1(v_0)\mathbf{m} \tag{3.12}$$

其中，$\mathbf{L}_1$为一阶 Born 正演模拟算子，其对应于同时求解方程（3.6）和方程（3.8）。由方程可以看出，一次散射波与背景速度成非线性关系，而与散射强度成线性关系。可以定义如下泛函来同时估计背景速度和反射系数，

$$J_1\left(v_0(\mathbf{x}),\mathbf{m}(\mathbf{x})\right)=\left(\mathbf{d}_1^{obs}-L_1\left(v_0(\mathbf{x})\right)\mathbf{m}(\mathbf{x})\right)^T\left(\mathbf{d}_1^{obs}-L_1\left(v_0(\mathbf{x})\right)\mathbf{m}(\mathbf{x})\right) \tag{3.13}$$

这就是基于一次散射波的全波形反演方法。

相比于利用透射波反演（方程（(3.6)））,在地面勘探中，由于散（反）射波信息在地震记录的中小偏移距就非常丰富，所以基于散（反）射波的反演更为实用。但（3.13）定义的泛函对应非线性反问题，反演难度较大，因为背景速度和散射强度是相互联系的。一般地，先假设背景速度已经由层析反演方法确定，然后利用散（反）射波数据来进行散射强度（反射系数）估计（也可能仅仅是定位！）。此时，误差泛函可以写为，

$$J_1\left(\mathbf{m}(\mathbf{x})\right)=\left(\mathbf{d}_1^{obs}-L_1\left(v_0(\mathbf{x})\right)\mathbf{m}(\mathbf{x})\right)^T\left(\mathbf{d}_1^{obs}-L_1\left(v_0(\mathbf{x})\right)\mathbf{m}(\mathbf{x})\right) \tag{3.14}$$

这就是最小二乘偏移成像方法的基本理论。由于上述理论是基于双向波方程进行推导的，所以该方法又叫做最小二乘逆时偏移（LS_RTM）方法，其目的是估计散射强度（反射系数）。

利用背景介质中的 Green 函数可以半解析地表达投影算子$L_1\left(v_0(\mathbf{x})\right)$。在背景介质中，由炮点传播到散射点的频率域 Green 函数可以定义为，

$$\left(-\frac{\omega^2}{v_0^2(\mathbf{x})}-\nabla^2\right)G_0(x,\omega;s_x)=\delta(\mathbf{x}-\mathbf{s}_x) \tag{3.15}$$

类似地，可以定义由检波点到散射点的 Green 函数。这样，由公式（3.15）（3.6）定义的背景波场可以写为子波与 Green 函数的乘积，

$$u_0(\mathbf{x},\omega;\mathbf{s}_x)=f(\omega;\mathbf{s}_x)G_0(\mathbf{x},\omega;\mathbf{s}_x) \tag{3.16}$$

方程（3.10）的 Green 函数形式解为：

$$u_1\left(x,\omega;s_{\mathbf{x}}\right)=\sum_{\mathbf{x}}-\frac{\omega^2}{v_0^2\left(\mathbf{x}\right)}G_0\left(\mathbf{x},\omega;g_{\mathbf{x}}\right)\mathbf{m}\left(\mathbf{x}\right)f\left(\omega;s_{\mathbf{x}}\right)G_0\left(\mathbf{x},\omega;s_{\mathbf{x}}\right) \tag{3.17}$$

上式的算子形式为：

$$u_1\left(x,\omega;s_{\mathbf{x}}\right)=L_1\left(v_0\left(\mathbf{x}\right)\right)\mathbf{m}\left(\mathbf{x}\right) \tag{3.18}$$

（3.18）式就是用于 LS_RTM（或 LS_PSDM）的算子方程。

3.2 线性泛函的变分及梯度

（3.18）式定义了一个 Fredholm 第一类积分方程，它描述的是一个线性问题。在勘探地震中，无论是信号处理或是反演成像过程中，很多问题的都归结为这样一个 Fredholm 第一类积分方程。离散化后，该方程形成一个线性方程组。线性方程组的形态，即求解的唯一性、稳定性都取决于其中投影矩阵（算子）的列秩分布情形。

一般地，我们把该 Fredholm 第一类积分方程的求解转化为一个 Bayes 估计问题。就是把该线性问题的求解，转化为利用泛函变分求极值问题。（3.18）式算子方程重新记为：

$$K\mathbf{m}=\mathbf{d} \tag{3.19}$$

其中，K 是 Fredholm 第一类积分方程的积分核。

认为$K:\mathrm{H}_1\to\mathrm{H}_2$是一个有界的线性算子，定义如下的 Tikhonov 泛函，

$$\mathfrak{I}\left(\mathbf{m}\right)=\frac{1}{2}\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2 \tag{3.20}$$

对上式求解的首要问题是计算泛函关于参数变化的导数，即泛函的梯度。为此，利用如下的变分方法。

对任意的介质扰动$\delta\mathbf{m}\in\mathrm{H}_1$和扰动步长$\tau\in\mathbb{R}$，总存在如下的泛函变分

$$\begin{aligned}
&\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right)=\frac{1}{2}\left\|K\left(\mathbf{m}+\tau\delta\mathbf{m}\right)-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\left\|\mathbf{m}+\tau\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\\
&=\frac{1}{2}\left\langle K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\frac{\alpha}{2}\left\langle\mathbf{m}+\tau\delta\mathbf{m},\mathbf{m}+\tau\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\frac{1}{2}\left\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\tau^2}{2}\left\langle K\delta\mathbf{m},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}\\
&+\frac{\alpha}{2}\left\langle\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)+\frac{\alpha\tau^2}{2}\left\langle\delta\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\frac{1}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\left\langle K\delta\mathbf{m},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\delta\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\frac{1}{2}\left(\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\\
&+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\left\|K\delta\mathbf{m}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)
\end{aligned} \tag{3.21}$$

对上式定义的变分求偏导数，

$$\begin{aligned}
&\frac{\partial}{\partial\tau}\mathfrak{J}(\mathbf{m}+\tau\delta\mathbf{m})\Big|_{\tau=0}\\
&=\left\{\left(\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{H_2}+\alpha\langle\mathbf{m},\delta\mathbf{m}\rangle_{H_1}\right)+\tau\left(\|K\delta\mathbf{m}\|_{H_2}^2+\alpha\|\delta\mathbf{m}\|_{H_1}^2\right)\right\}\Big|_{\tau=0}\\
&=\left(\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\rangle_{H_2}+\alpha\langle\mathbf{m},\delta\mathbf{m}\rangle_{H_1}\right)\\
&=\left(\langle K^*(K\mathbf{m}-\mathbf{d}),\delta\mathbf{m}\rangle_{H_1}+\alpha\langle\mathbf{m},\delta\mathbf{m}\rangle_{H_1}\right)\\
&=\langle K^*(K\mathbf{m}-\mathbf{d})+\alpha\mathbf{m},\delta\mathbf{m}\rangle_{H_1}\\
&=\langle(K^*K+\alpha I)\mathbf{m}-K^*\mathbf{d},\delta\mathbf{m}\rangle_{H_1}\\
&=\langle\mathrm{Grad}\mathfrak{J}(\mathbf{m}),\delta\mathbf{m}\rangle_{H_1}
\end{aligned}\tag{3.22}$$

其中 $\mathrm{Grad}\mathfrak{J}(\mathbf{m})=(K^*K+\alpha\mathbf{I})\mathbf{m}-K^*\mathbf{d}$ （3.23）

是误差泛函相对于参数模型的梯度。K^*是波场正传播算子的共轭算子，代表波场的反传播。经典地，我们利用该梯度构造最速下降迭代方法求解 Born 近似后的线性问题的解。迭代格式为，

$$\begin{aligned}
&\mathbf{m}^{k+1}=\mathbf{m}^k-\mu\mathrm{Grad}\mathfrak{J}(\mathbf{m})=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}-\mathbf{d}^{obs}\right)\right]=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}^k-\mathbf{d}^{obs}\right)\right]\\
&=\mathbf{m}^k-\mu\left[K^*\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)\right]
\end{aligned}\tag{3.24}$$

利用上述迭代公式进行最小二乘叠前深度反演成像是现代勘探地震对成像精度提高的要求，并基于计算机能力的大幅度提高。

另外，基于图像反褶积的方法实现最小二乘叠前深度偏移成像的基本做法是，令梯度等于零（实际上一般不会等于零！）时，可以得到参数估计对应的法方程：

$$(K^*K+\alpha\mathbf{I})\mathbf{m}=K^*\mathbf{d}\tag{3.25}$$

之所以称上述做法为基于图像反褶积的方法，是因为算子$(K^*K+\alpha\mathbf{I})$本质上是一个模糊算子（图像处理中称为点扩散函数（PSF）），施加在要估计的图像上，起到平滑该图像的作用。对图像平滑后的结果就是（3.25）式的右端项（一般地，这就是常规偏移成像的结果）。把算子$(K^*K+\alpha\mathbf{I})$的逆算子作用在右端项上得到图像反褶积后的结果。

线性算子理论基础部分给出的例子 2.3.2，就是对应算子方程（3.18）式。在勘探地震中，这个例子对应有非常多的实际问题。据此可以看出，线性空间和线性算子理论在当今阶段的地震信号处理和反演成像方法研究中已经成为非常基础的知识需求。

4 总结与讨论

勘探地震中有两大类问题，一是信号分析问题；一是反演成像问题。信号分析包含的问题更为广泛。信号分析问题基本上是对平方可积函数空间（Hilbert 空间的例子！）中函数（信号）进行分析，主要的问题是函数的逼近问题。首先要解决的是选择一定的基函数族（最好完备的标准正交基函数族），在此基函数族中实现对函数（信号）的最佳逼近。在此过程中，解决信号的预测、去噪、插值、规则化、压缩等等方面的问题。地震波反演成像问题同样如此，首先也是建立模型空间向数据空间的（算子）映射方程（譬如方程（3.18）），这其中基函数的选择融入了积分核函数的构建中，不再显式地出现。然后在一定的范数意义下建立实测数据与预测数据之间的度量泛函。在此度量泛函下，求解一个优化问题，实现对参数的最佳估计。

可以看出，上述两个最核心的问题都是奠基在线性空间和线性算子的基础上的。因此，勘探地震技术发展到现在，有必要把技术发展的基础由微积分拓展到泛函分析。

本文仅仅起到一定的引领作用，为 WPI 研究组的研究生和研究人员指出方向。更全面深入的泛函分析理论学习，还要阅读其他参考文献（Zeidler，1995）。

参考文献

[1] 关治，陈景良，1990，数值计算方法，清华大学出版社。
[2] 夏道行，吴卓人，严绍宗，舒五昌编著，2010 年，实变函数论与泛函分析，高等教育出版社。
[3] 老大中，2015，变分法基础（第三版），国防工业出版社。
[4] Vogel, C. R., 2002, Computational Methods for Inverse Problem, SIAM
[5] Troyan, V., and Kiselev, Y., 2010, Statistical Methods of Geophysical Data Processing, World Scientific Publishing Co. Pte. Ltd.
[6] Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM
[7] Zeidler, E., 1995, Applied Functional Analysis---Applications to Mathematical Physics, Springer

扩散方程和椭圆方程在地震波反演成像中的应用

王华忠，李辉

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】勘探地震中地震波反演成像无疑是核心问题，其目的是期望弄清楚地下岩石介质的弹性参数分布，无论是低波数的光滑部分，或是高波数的跃变部分。但是，仅仅依赖地表观测的、观测孔径和频带有限的、采样不规则或采样密度不充分的、信噪比不高的、震源函数未知的叠前地震数据不可能获得完全满足油气藏勘探需求的反演成像结果。井中或其他来源的与地下岩石介质的弹性参数分布有关的信息必须引入来提高反演成像的稳定性和精度。另外，反演成像过程中，各种正则化思想和算法也是提高地震波反演成像精度的重要的、必不可少的措施。无论是充分利用井中或是其他来源的信息进行信息融合和建模，或是提取成像剖面上的反射界面结构信息用来进行反演过程中的正则化处理，都需要借用现代信号分析和现代图像分析方面新思想和新方法。扩散方程在现代图像结构提取中已经得到了广泛而有效的应用，各向异性扩散方程在图像结构提取中的作用与参数反演中的 TV 正则化方法是基本等价的；椭圆方程用在散乱数据插值过程中，最近也得到比较多的重视，这种插值方法体现出了更容易融合结构信息约束的特点，在地震参数建模过程中有明显的潜在优势。本文主要讨论地震波反演成像的特点、现代信号分析和现代图像分析思想在地震波反演成像中的可能应用，并具体给出椭圆方程和扩散方程在地震波反演成像中的应用实例。目的是引导 WPI 研究组成员关注现代信号分析和现代图像分析的进展，并借用这些思想和方法促进地震波反演成像精度的提高。

【关键词】地震波反演成像，图像背景，图像结构，现代信号分析，现代图像分析，正则化

1 引言

油气地震勘探的目标是完成对含油气目标体的定位、识别与描述，在此基础上进行油气藏的开发。其中，地震波反演成像无疑是核心问题，其目的是期望弄清楚地下岩石介质的弹性参数分布，无论是低波数的光滑部分，或是高波数的跃变部分。给出精确的岩石弹性介质低波数成分的估计，就可以完成线性化的高波数成分的估计。目前广泛用于油气藏识别的高波数成分是（角度）反射系数，正在向利用高维地震数据估计波阻抗的方向发展。但是，仅仅依赖地表观测的、观测孔径和频带有限的、采样不规则或采样密度不充分的、信噪比不高的、震源函数未知的叠前地震数据不可能获得完全满足油气藏勘探需求的反演成像结果。

稳定而可靠的地震波反演成像需要其他额外信息的约束，也需要各种正则化思想和方法的引入，尤其是反射地震界面的几何结构的约束。额外的信息主要来自钻井、测井和地质，这些与地下岩石介质的弹性参数分布有关的信息的引入可以有效地提高反演成像的稳定性和精度，一般地它们表现为空间离散数据的形式，需要有效的散乱数据插值方法把它们表示成初始模型的形式，约束反演成像的结果。另一方面，反演成像过程中，无论对数据残差和对模型本身都要提出有效的正则化思想和方法来约束反演解，使它具有明确的地质地球物理意义。

地震波反演成像发展到现在，Bayes 框架下的反演理论、方法等都已经十分清晰，增加解的稳定性的各种正则化方法都提出很多。但是，FWI+LS_RTM 一类的地震波反演成像核心方法依然不能很好地用于实际地震数据成像处理中。此时，需要借用现代信号分析和现代图像分析方面新思想和新方法于地震波反演成像、甚至地震成像数据的自动化地质解释中。首先是表达成散乱数据的各种与反演参数有关的先验信息需要融合成有地质和地球物理意义的数据体，融合后的数据体之所以体现出地质与地球物理意义主要是要在上述信息融合过程中引入层状沉积的反射界面及断层结构的约束。因此，需要改进现有的散乱数据插值方法，发展出受地质构造约束的散乱数据插值方法。这样的插值方法应该也可以推广到其他应用领域中。第二个重要的问题是自动抽取成像剖面上的反射结构特征，包括粗线条的和细节的几何特征。当然，也包括自动抽取特征前的各种图像预处理的方法；也包括自动抽取各种地震数据属性的方法。这些方法以前主要用在地震剖面的自动解释中。现在，在地震波反演成像中，从成像后的剖面上提取结构信息用来特征表达反演参数场，或约束反演成像结果变得必不可少。看起来，地震地质解释和地震波反演成像也逐渐再融合。当然，总的目的还是要完成含油气油藏的识别、描述与刻画。

我们已知，扩散方程在现代图像结构提取中已经得到了广泛而有效的应用，各向异性扩散方程在图像结构提取中的作用与参数反演中的 TV 正则化方法是基本等价的；椭圆方程用在散乱数据插值过程中，最近也得到比较多的重视，这种插值方法体现出了更容易融合结构信息约束的特点，在地震参数建模过程中有明显的潜在优势。

本文主要讨论地震波反演成像的特点、现代信号分析和现代图像分析思想在地震波反演成像中的可能应用，并具体给出椭圆方程和扩散方程在地震波反演成像中的应用实例。目的是引导 WPI 研究组成员关注现代信号分析和现代图像分析的进展，并借用这些思想和方法促进地震波反演成像精度的提高。

2 地震波成像处理流程中关键技术的理论背景

无论是地震数据的常规处理流程，或是地震波反演成像流程，本质上，都包含是三个或四个核心环节：地震数据的预条件；弹性参数场低波数光滑部分的估计；弹性参数场高波数跃变部分的估计；反演成像的后处理。

地震数据预处理的核心包括去噪、插值、数据规则化等；弹性参数场的估计包括背景部分和跃变部分的估计。二者的理论基础都可以归结为：正问题（预测）+反问题（参数估计）。信号处理理论和参数估计理论都可以统一在 Bayesian 估计的理论框架下，二者不同在于正问题描述方式的差异。二者都需要正则化约束，只是提法不同而已。

此处，我们以层析反演为例，说明上述统一的思想和方法。层析速度分析，一般是指基于非线性 Radon 变换的、基于旅行时的一类反投影方法。目的是估计背景速度（各向异性参数和 Q 值）。

2.1 旅行时速度层析的正问题描述

① Radon 变换/Fredhelm 第一类积分方程（N 代表射线条数）：

$$\mathbf{d}^i = g^i(\mathbf{m}) = \int_{R^i(\mathbf{m})} K\mathbf{m}(\mathbf{x}^i) = \int_{R^i(\mathbf{m})} ds^i \mathbf{m}(\mathbf{x}^i), \qquad i = 1,2,\cdots,N \tag{1}$$

②引入慢度场扰动（$\mathbf{m}_b$ 认为是初始背景速度）：

$$g^i(\mathbf{m}_b + \delta\mathbf{m}) = \int_{R^i(\mathbf{m}_b+\delta\mathbf{m})} ds^i(\mathbf{m}_b(\mathbf{x}^i) + \delta\mathbf{m}(\mathbf{x}^i) + \delta\mathbf{m}^2(\mathbf{x}^i) + \cdots) \qquad i = 1,2,\cdots,N \tag{2}$$

③Taylor 展开---正问题的线性化：

$$g^i(\mathbf{m}_b + \delta\mathbf{m}) = \int_{R^i(\mathbf{m}_b+\delta\mathbf{m})} ds^i(\mathbf{m}_b(\mathbf{x}^i) + \delta\mathbf{m}(\mathbf{x}^i) + \delta\mathbf{m}^2(\mathbf{x}^i) + \cdots) = \int_{R^i(\mathbf{m}_b)} ds^i\left(\mathbf{m}_b(\mathbf{x}^i) + \delta\mathbf{m}(\mathbf{x}^i)\right) + \mathrm{O}\left(\left\|\delta\mathbf{m}^2\right\|\right) \qquad i = 1,2,\cdots,N$$

重写上式：

$$g^i(\mathbf{m}_b + \delta\mathbf{m}) = \int_{R^i(\mathbf{m}_b)} ds^i \mathbf{m}_b(\mathbf{x}^i) + \int_{R^i(\mathbf{m}_b)} ds^i \delta\mathbf{m}(\mathbf{x}^i) = g^i(\mathbf{m}_b) + \int_{R^i(\mathbf{m}_b)} ds^i \delta\mathbf{m}(\mathbf{x}^i) + O(\|\delta\mathbf{m}\|^2)$$

得到慢度扰动和旅行时扰动之间的线性关系：

$$\delta\mathbf{d}^i = g^i\left(\mathbf{m}_b + \delta\mathbf{m}\right) - g^i\left(\mathbf{m}_b\right) = \mathbf{d}^i_{obs} - \mathbf{d}^i_{cal} = \int_{R^i(\mathbf{m}_b)} ds^i \delta\mathbf{m}(\mathbf{x}^i) \tag{3}$$

最后得到对应的矩阵形式的方程：

$$\mathbf{d}_{N\times 1} = G_{N\times M}\mathbf{m}_{M\times 1} \tag{4}$$

在背景介质 $\mathbf{m}_b$ 中进行射线追踪，得到第 i 根射线 $\mathbf{d}^i_{cal}$，形成一个线性方程，每个炮检对对应一个线性

方程，所有炮检对对应的线性方程形成一个方程组。解此线性方程组得到层析成像结果。

2.2 旅行时速度层析的反问题描述

①旅行时差的计算：

$$\delta \mathbf{d}^i = \int_{R^i(\mathbf{m}_b)} ds^i \delta \mathbf{m}(\mathbf{x}^i) \tag{5}$$

②旅行时差加权处理---数据正则化：

$$\delta \mathbf{d}^i = \sum_j (C_D^{-1})^{ij} \delta \mathbf{d}^j \tag{6}$$

③反投影路径加权处理—模型正则化：

$$\psi^i(\mathbf{x}) = \int_{R_i(\mathbf{m}_b)} ds^i C_M(\mathbf{x}, \mathbf{x}(s_i)) \tag{7}$$

④旅行时差反投影处理---得到梯度：

$$\delta \mathbf{m}(\vec{x}) = \sum_j \delta \tilde{\mathbf{d}}^j \psi^i(\mathbf{x}) \tag{8}$$

⑤慢度扰动修改方向---梯度预条件：

$$\gamma_k(\mathbf{x}) = \delta \mathbf{m}(\mathbf{x}) + (\mathbf{m}_b(\mathbf{x}) - \mathbf{m}_{prior}(\mathbf{x})) \tag{9}$$

⑥慢度迭代修改：

$$\mathbf{m}_{k+1}(\mathbf{x}) = \mathbf{m}_k(\mathbf{x}) - \mu_k \gamma_k(\mathbf{x}) \tag{10}$$

2.3 旅行时层析成像的核心步骤

①旅行时差拾取或计算；

②旅行时差质量评价---数据预条件；

③关于旅行时残差的数据正则化；

④观测系统与矩阵形成---射线追踪；

⑤关于模型的正则化；

⑥方程的求解---ART，SIRT，CG，LSQR。

2.4 正则化思想与方法

勘探地震中的地震波反演成像必须引入正则化方法，否则很难得到有意义的反演结果，对油气藏刻画产生有效的帮助。主要问题应该是地震勘探面对的地下介质的复杂多变、数据采集环境的恶劣（与医学成像相比）、数据采集方式的受限。

针对（4）式定义的正问题，根据确定性反演框架，建立 Tikhonov 正则化反演泛函：

$$\Phi(\mathbf{m}) = \left(\mathbf{d}^{obs} - G\mathbf{m}\right)^T A^{-1} \left(\mathbf{d}^{obs} - G\mathbf{m}\right) + \mathbf{m}^T B^{-1} \mathbf{m} \tag{11}$$

上式对应的法方程为：

$$\hat{\mathbf{m}}=\left(G^T A^{-1}G+B^{-1}\right)G^T A^{-1}\mathbf{d}^{obs} \tag{12}$$

关于正则化参数的选取，A^{-1} 一般地选为数据的协方差矩阵的逆；B^{-1} 一般选为单位阵加上平方一阶导数算子（相当于 Laplacian 算子）形成的矩阵。据此有，如下的 Tikhonov 正则化反演泛函：

$$\Phi(\mathbf{m})=\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T\Gamma_{\mathbf{d}}^{-1}\left(\mathbf{d}^{obs}-G\mathbf{m}\right)+\alpha\mathbf{m}^T\cdot\mathbf{I}\cdot\mathbf{m}+\beta\mathbf{m}^T\cdot D_1^2\cdot\mathbf{m} \tag{13}$$

其中，$D_1=\begin{bmatrix}\cdots & 0 & 0\\ -1 & 1 & 0\\ 0 & -1 & 1\\ 0 & 0 & \cdots\end{bmatrix}$。

上式对应的法方程解为：

$$\hat{\mathbf{m}}=\left(G^T\Gamma_{\mathbf{d}}^{-1}G+\alpha\cdot\mathbf{I}+\beta\cdot\mathbf{D}_1^2\right)G^T\Gamma_{\mathbf{d}}^{-1}\mathbf{d}^{obs} \tag{14}$$

若从随机信号的观点看，线性正问题应该表达为：

$$\mathbf{d}^{obs}=\mathbf{d}^{cal}+\eta=G\mathbf{m}+\eta \tag{15}$$

认为采集到的地震数据（包括拾取的到达时）是随机的。η 是随机噪音。当然，反演结果也是随机的。

针对该正问题的随机性反演，建立 Bayesian 反演泛函：

$$\Phi(\mathbf{m})=\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T\Gamma_{\mathbf{d}}^{-1}\left(\mathbf{d}^{obs}-G\mathbf{m}\right)^T+\mathbf{m}^T\Gamma_{\mathbf{m}}^{-1}\mathbf{m} \tag{16}$$

上式对应的法方程为：

$$\hat{\mathbf{m}}=\left(G^T\Gamma_{\mathbf{d}}^{-1}G+\Gamma_{\mathbf{m}}^{-1}\right)G^T\Gamma_{\mathbf{d}}^{-1}\mathbf{d}^{obs} \tag{17}$$

对比确定性反演和随机性反演框架的正则化反演结果，可以发现：

$$\Gamma_{\mathbf{m}}^{-1}\Rightarrow\alpha\cdot\mathbf{I}+\beta\cdot\mathbf{D}_1^2 \tag{18}$$

即模型参数协方差矩阵的逆的正则化作用可以由 Tikhonov 正则化方法来解释。众所周知，直接给出模型参数协方差矩阵的逆是不可能的。如何确定模型参数协方差矩阵的逆 $\Gamma_{\mathbf{m}}^{-1}$ 是合适的？协方差阵选成用如下（各向异性）的 Gaussian 函数是合理的，符合物理逻辑，

$$\Gamma_{\mathbf{m}}(\mathbf{r}_1,\mathbf{r}_2)=\sigma^2\exp\left(-\frac{(\mathbf{r}_1-\mathbf{r}_2)^2}{2L^2}\right) \tag{19}$$

只不过要给出上式中合理的参数是很难的。但这并不影响我们分析问题。可以给出如下表达式，

$$\mathbf{m}\cdot\Gamma_{\mathbf{m}}^{-1}\cdot\mathbf{m}=\frac{1}{2\pi}\frac{1}{(\sigma L)^2}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{L^2}{2}\right)^n\nabla^n\mathbf{m}\cdot\nabla^n\mathbf{m} \tag{20}$$

（20）式显然可以表达为，

$$\Gamma_{\mathbf{m}}^{-1}=\alpha\cdot\mathbf{I}+\beta\cdot\mathbf{D}_1^2+High-order-term \tag{21}$$

上述分析过程说明，尽管模型参数协方差矩阵的逆的引入是最合适的模型正则化方法，但是实际上，若真的知道了模型参数的协方差阵，也就知道了参数场的分布。这种要求显然是不合理的。但是，可以通过类似（21）式的方式加入模型的正则化约束。

2.5 随机性反演解的评价

反演解估计可以用如下伪逆算子来表达，

$$\hat{\mathbf{m}} = G^{\dagger}\mathbf{d}^{obs} = G^{\dagger}G\mathbf{m}_{true} \tag{22}$$

其中，

$$G^{\dagger} = \left(G^{T}\Gamma_{\mathbf{d}}^{-1}G + \Gamma_{\mathbf{m}}^{-1}\right)G^{T}\Gamma_{\mathbf{d}}^{-1}\text{。} \tag{23}$$

引入符号：

$$\hat{\mathbf{m}} = G^{\dagger}G\mathbf{m}_{true} = \mathbf{R}\mathbf{m}_{true} \tag{24}$$

其中 **R** 称为分辨率矩阵。

在 Bayesian 框架下，有

$$\mathbf{R} = \mathbf{I} - \Gamma_{\mathbf{m}}^{posteriori} \Big/ \Gamma_{\mathbf{m}}^{\mathrm{Pr}iori} \Rightarrow \Gamma_{\mathbf{m}}^{posteriori} = \left(\mathrm{I\text{-}R}\right)\Gamma_{\mathbf{m}}^{\mathrm{Pr}iori}\text{。} \tag{25}$$

可以看出，当分辨率矩阵是单位阵时，反演结果分辨率最高，等于获得了真实的参数分布。但（24）式告诉我们，通过反演算法我们永远也看不清真实世界的模样，只能看到被分辨率矩阵模糊后的真实世界的样子。另一方面，当分辨率矩阵是单位阵时，后验协方差阵等于零矩阵时，反演过程没有引入新的信息，先验信息已经足够精确。分辨率矩阵为零矩阵时，不能进行反演，此时先验信息的引入不能改善反演结果分辨率。这应该意味此时我们提供的数据和模型先验信息与要反演的结果之间根本没有关系。好在这样的事情应该仅仅在概念上会存在，实际反演过程中不会发生。但这并不意味着这种情形没有指导性，它告诫我们要切实地建立起观测数据（波场）和关于模型的先验信息与要反演参数之间的密切关系，而不是随意为之。

3 现代信号分析与现代图像处理的进展及在勘探地震中的可能应用

在所有自然科学和社会科学领域，都会遇到信号/图像的采样、处理和分析问题，勘探地震中的波场采样、成像处理和基于成像处理结果的地震地质解释仅仅是现代信号分析和现代图像处理理论方法的在具体问题中的应用。

这正是 WPI 研究组把地震波成像方法的研究置于现代信号分析框架的根本原因。

现代信号/图像分析的核心思想应该包括两部分：信号/图像的（特征）表达，相当于地震波成像中的正问题，即波场的预测问题；信号/图像（特征）参数的估计问题，相当于地震波成像中的反问题，即弹性参数的估计问题。正问题是反问题的基础，正是正问题的不同，才区分开了各个不同的专业领域。反问题基本上可以归结为 Bayesian 框架下的参数估计理论及具体实现，不同的专业领域差异不大。

现代信号/图像分析之所以成为各个相关专业的、通用的理论基础，正是在于抽象后的信号/图像总可以用 1D 或高维（大于等于 2D）离散的、平方可积函数在 Hilbert 空间中进行讨论。可以看出，这已经脱离了具体的专业领域。

1D 或高维（大于等于 2D）离散的、平方可积函数在 Hilbert 空间中的分析是现代信号/图像分析的核心研究内容。我们抽象出来的现代信号分析与现代图像处理的进展可以总结为下面的表述。

关于现代信号分析方面的进展，主要侧重于信号的特征表达。在正问题表达方面，Fourier 基函数（也包括类似 Chebyshev 正交基函数）的表达已经成为常规，各种 1D 和高维小波基函数（框架/标架）的构建是

研究的重点，Curvelet、BeamLet、RigLet、Seislet 等等以及（自适应）字典基（字典学习）等的构建有很多新的提法，目的是用与物理信号最匹配的基函数（字典）来提取或表达信号的特征。高维数据的奇异值分析（SSA=Singular Spectrum Analysis）与 Fourier 谱分析（FSA=Fourier Spectrum Analysis）对应方法，由于计算机技术的飞速进步，近几年也有长足的进步，广泛应用于不同领域的信号处理中。在信号特征参数反演方面，主要的进展体现在用凸优化和非凸优化算法进行参数的稀疏反演，提高特征参数反演的精度。总之，用正问题和反问题两个方面的进步来提高信号分析结果的精度。有非常多的参考文献说明上述方法理论进展。

关于图像分析方面的进展，首先需要说明的是图像是信号的特例，或者说图像也是信号，只不过图像处理有自己的特殊性，独立地进行图像分析的研究有一定的方便性而已。因此，现代信号分析方面的进展，某种程度上也是现代图像分析方面的进展。图像分析的核心思想是把它视为高维（至少 2D）的灰度图，然后研究灰度图的结构和背景（纹理）的特征，这是一种很高的抽象，完全脱离了具体的应用领域。图形研究的核心问题是图像结构特征的表达（正问题）、图像结构特征的提取（反问题）。这是图像分析的两个最基本的内容。更高级的研究内容是图像特征的分类（聚类）、（图像模式的自动）识别。再高级的研究内容就是图像认知和人工智能了。目前，深度学习方法是大家特别关注的，深度学习的思想方法到目前为止就是图像特征表达、提取、模式识别，还没有达到很好的图像认知的程度，基本上还是处在有监督的学习阶段。AlphaGo 计算机围棋软件系统基本是深度学习方法的一种具体实现。因此，图像结构特征的表达与提取才是图像分析的基石。这与信号特征的表达与提取是其核心问题是一致的。目前，图像特征表达和提取的进展主要体现在：图像的小波基函数展开、各向异性扩散方程在图像结构提取中的应用、变分问题在求解 ROF 模型（一种最核心的图像结构提取数学模型）、L2/L1/L0 范数下的 Bayesian 反演四个方面，图像特征的非线性表达（图像形态学）没有形成统一的方法。

从上述现代信号分析与现代图像处理的进展的描述可以看出，这些研究进展的思想和方法绝大部分都可以直接引入到勘探地震数据分析的各个阶段，包括数据采集、信号分析和处理、地震波成像和地震地质解释。事实上，我们不断地在吸收这些进展的思想与方法，譬如随机采样指导野外地震数据非规则采样与高效采集、压缩感知与稀疏反演用于反褶积和地震数据规则化及混叠数据分离、各向异性扩散方程用于图像去噪和结构提取、椭圆方程用于离散数据插值等等。图像处理方法进行辅助地震地质解释的研究也逐渐引起勘探地震工业界的关注。

4 椭圆方程在空间散乱数据插值及建模中的应用

散乱数据的空间插值与建模是勘探地震中一个很核心的问题。但是，它是一个没有引起足够关注的问题。

勘探地震中能用于提供地震波反演成像的信息主要有两类：在不同观测面上测量的叠前地震数据体；来自测井、钻井和地质等的数据体。前者一般是空间比较规则采样的，是地震波场或地震信号；后者是直接与地下弹性介质参数有关的信息，一般表现为空间散乱分布的形式，经常用作第一类数据反演成像时的约束信息。

空间散乱数据的插值，不同于地震数据的规则化，尽管两者有很多的联系。空间散乱数据有二维面上的散乱数据，勘探地震中更多是三维空间中的散乱数据，譬如钻井岩性数据、测井参数数据、来自地震剖面上沿界面的各种解释数据、地质知识提供的三维空间中的与地下岩性分布有关的信息。

对于散乱数据插值，首先要做的一件事是散乱数据的管理。到目前为止，石油地震勘探界并没有发布类似 SEGY 那样的数据格式标准。针对勘探地震而言，第二件要做的事是依据前述各种不同来源的空间数据信息插值成为有地震地质含义的弹性参数场，明确地讲就是得到由反射界面约束的各种弹性参数场。

离散数据插值的本质是：由实测空间/时间点处的函数值作为支撑点，选择一定的基函数族构建连续的逼近函数，在插值条件或逼近误差泛函最小的约束下，得到对应各基函数的投影值（或坐标值），产生一个连续的逼近函数，最后利用该连续函数计算任意点处的函数值。

对应的数学表述是：φ_j 是 $C\left(R^d\right)$ 上的线性泛函，给定数据 $f_j \in R, j=1,2,\cdots,n$，希望寻找 $C\left(R^d\right)$ 中的函数 $g\left(x\right)$，满足 $\varphi_j g\left(x\right)=f_j, j=1,2,\cdots,n$。

插值的基本思想包括：具有局部支撑性的基函数的选择；基函数的线性组合近似表达原函数；插值条件的应用确定插值函数。多项式基函数（Lagrange 插值、Newton 插值、Hermit 插值、样条插值）和 Sinc 函数插值最基础的。

1D 离散数据的插值方法，包括 Lagrange 插值方法、Newton 插值方法、Hermit 插值方法、分段插值方法、样条插值方法。

上述 1D 离散数据插值方法是用代数多项式函数作为基函数组构建插值多项式，这样的插值多项式存在唯一性。但它们不能直接推广到二维散乱数据插值，一般地空间散乱数据插值没有唯一解。在既定散乱数据基础上的插值都仅仅是对原函数（如果真能得到原函数的话！）的某一种近似。只要这种近似满足工程需求就可以了。

散乱数据插值方法一般地有局部方法和全局方法。主要的局部方法：分片插值、最近邻域点插值、自然邻域插值、距离反比插值；主要的全局方法：径向基函数插值和 Kriging 插值。

任何插值（逼近）方法都要考虑原函数本身的变化趋势与特征，才能得到有意义的插值结果。这就是有约束的插值。当前的高维信号数据规则化方法中已经在利用原数据中蕴含的结构信息进行插值。但是，散乱数据插值中如何引入约束信息还没有明确的技术路线。

利用椭圆方程进行散乱数据插值是一个可以引入约束的方法。椭圆方程是一个描述势场（静场）分布的方程，显然可以认为实测散乱数据是支撑整个势场函数的点，通过借椭圆方程得到其他非支撑点上的原函数值。这应当是一种全局插值方法。

（部分）空间散乱数据插值方法的特定可以总结如下：

表 1 空间散乱数据插值方法的优缺点对比

方法	优点	缺点	适用条件
最近邻点插值法 (Nearest Neighbor)	方法简单、计算效率高、不受前提条件的约束。	插值结果受样本点的影响大，缺乏对其他空间因素和变量自身所具有的固有规律的考虑。	样本点分布均匀、完整，空间变异性不很明显，只存在少量数据的缺失。
算术平均值插值法 (Arithmetic Mean)	算法简单、容易实现。	仅仅考虑了样本点的算术平均，忽略了样本点间存在的空间联系。	样本点分布均匀密集而且变化平缓，常用于局部降噪，适用范围窄。
距离反比插值法 (Inverse Distance Weighted，IDW)	算法简便易行是一种精确插值算法。	对权重函数中幂次选择十分敏感，易受采样集群影响，造成“牛眼”效应。	样本点的分布均匀、密集且样本点含有区域特征信息。
径向基函数插值法 (Radial Basis Function)	引入平滑因子，插值结果表面比较光滑。	当研究区域内采样点局部变异性大，采样数据具有较大的不确定时，难以保证估计数据的准确性。	采样数据局部变异不大时，像进行气象特征估计时可采用此法。
克立格插值法 (Kriging)	给出估计误差，充分考虑空间变量相关性，有效弥补数据集存在的聚类影响，插值精度高。	计算步骤繁琐，插值速度慢。	适用于采样区域的样本数据存在随机性和结构性两特征。

可以看出，基于原函数特征的空间散乱数据插值是很难做到的。目前没有现成的方法技术可以应用。

勘探地震中，空间散乱数据插值主要用于弹性参数场的建模过程中，而且希望得到的模型具有一定的地质含义，这就是地质构造约束下的建模。

勘探地震中散乱数据插值的主要方法是 Kriging 插值，它是基于径向基函数的。一般地，径向基函数族是由不同（与插值距离有关）方差的高斯函数组成的。很显然，Kriging 插值目前缺乏与地质构造的联系，不是一种地质构造约束下的插值方法。

下面讨论在调和插值、自然邻域插值基础上发展出的考虑地质构造约束的椭圆方程插值方法。

4.1 从调和插值到混合邻域插值

本节中回顾已知离散的点的坐标和值，通过偏微分方程（PDE）把数据插值到规则网格点上的方法。n 个已知数据的坐标记为

$$\mathrm{X}=\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\} \tag{26}$$

已知点的值记为

$$\Upsilon=\{v_1,v_2,\cdots,v_n\} \tag{27}$$

调和插值（Harmonic interpolation）与重调和插值（Biharmonic interpolation）

调和插值是一类广泛应用的插值方法。经典的调和插值满足如下的椭圆方程（拉普拉斯方程）

$$\begin{aligned}&\nabla^2 v(\mathbf{x})=0,\ \ \mathbf{x}\notin \mathrm{X}\\&v(\mathbf{x}_i)=v_k,\ \ \mathbf{x}\in \mathrm{X}\end{aligned} \tag{28}$$

重调和插值定义为如下方程的解

$$\begin{aligned}&(\nabla^2)^2 v(\mathbf{x})=0,\ \ \mathbf{x}\notin \mathrm{X}\\&v(\mathbf{x}_i)=v_k,\ \ \mathbf{x}\in \mathrm{X}\end{aligned} \tag{29}$$

从（28）和（29）中 PDE 的阶数可看出，调和插值效率较高，重调和插值效率很低。调和插值结果比较尖锐，重调和插值相对平滑，图 1 对比了两者的插值结果。

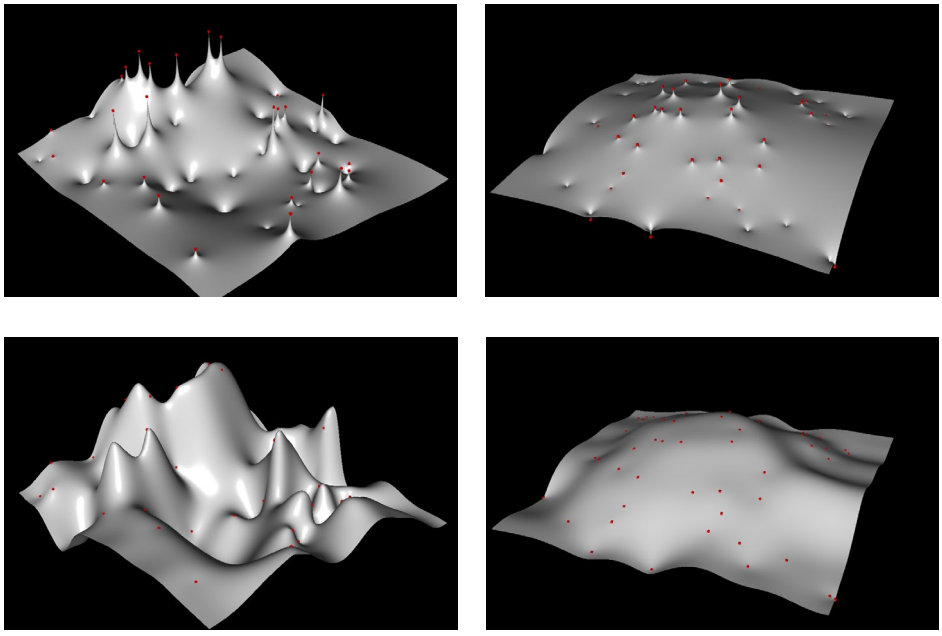


图 1 两组数据的调和插值（上）和重调和插值（下）结果对比。红色圆点已知，曲面是插值结果，下同

重调和插值可通过加入一个描述数据特征的张量修改成各向异性非均匀的插值方法：

$$\begin{aligned}&(\nabla\cdot\mathbf{D}(\mathbf{x})\nabla)^2 v(\mathbf{x})=0,\ \ \mathbf{x}\notin \mathrm{X}\\&v(\mathbf{x}_i)=v_k,\ \ \mathbf{x}\in \mathrm{X}\end{aligned} \tag{30}$$

直接求解方程（30）可实现构造约束的重调和插值。但加入张量之后的各向异性非均匀重调和插值计算效率进一步降低，导致实用性不强。

自然邻域插值（Natural neighbor interpolation）最近邻域插值

自然邻域插值由 Sibson（1981）定义，其形式为

$$v(\mathbf{x})=\frac{\sum_i a_i(\mathbf{x})v_i}{\sum_i a_i(\mathbf{x})} \tag{31}$$

其中系数 a_i 是 Voronoi 多边形的面积。图 2 给出了自然邻域插值的一个示例。

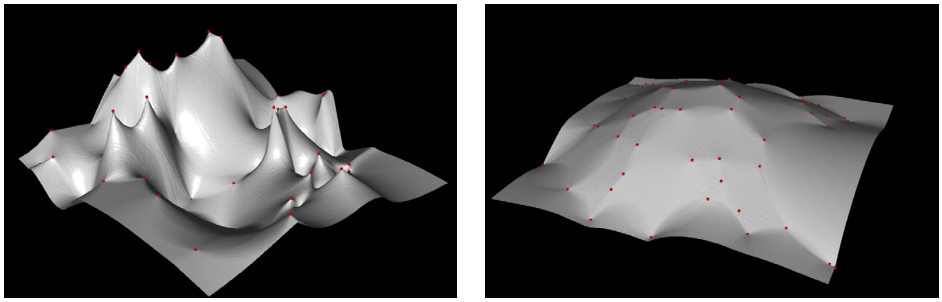


图 2 两组数据的自然邻域插值结果

定义两个变量：

1） d（$\mathbf{x}$）：最近距离，$\|\mathbf{x}-\mathbf{x}_i\|_2$ 最小的 $\mathbf{x}_i$ 对应的距离，$\mathbf{x}_i$ 是已知的点的坐标，即

$$d(\mathbf{x})=\min_i\|\mathbf{x}-\mathbf{x}_i\|_2 \tag{32}$$

2） p（$\mathbf{x}$）：最近距离的点 $\mathbf{x}_i$ 处的值，即

$$p(\mathbf{x})=v(\arg\min\|\mathbf{x}-\mathbf{x}_i\|_2) \tag{33}$$

利用上述两个变量修改自然邻域插值中的系数，并在插值点的一个邻域内取平均，即可得到所谓的“最近邻域插值”：

$$v(\mathbf{x})=\frac{\int h(\mathbf{x},\mathbf{y})p(\mathbf{y})d\mathbf{x}'}{\int h(\mathbf{x},\mathbf{y})d\mathbf{x}'} \tag{34}$$

其中 h 是阶跃函数：

$$h(\mathbf{x},\mathbf{y})=\begin{cases}1, & if\ \|\mathbf{x}-\mathbf{y}\|_2\le d(\mathbf{y})\\ 0, & otherwise\end{cases} \tag{35}$$

最近邻域插值的基本思想是：未知点的值等于距离最近的已知点的值。最近邻域插值的效果显然不能满足即便是初始模型建立的精度要求。

混合邻域插值（Blending nearest neighbors interpolation）

在最近邻域插值中，函数 h 可看作是一个对插值的硬约束。修改函数 h 使之变成一个范围内的软约束，可得到一般形式的混合邻域插值

$$v(\mathbf{x})-\frac{1}{e}\nabla\cdot d^2(\mathbf{x})\nabla v(\mathbf{x})=p(\mathbf{x}) \tag{36}$$

其中变量 $e\geq 2$。混合邻域插值的结果在已知点附近更加平滑。当 e=3 时，混合邻域插值等价于薄板径向基函数插值与双谐插值（Sandwell，1987）。图 3 给出了混合邻域插值结果，对应的变量 e 分别为 2、4 和 8。

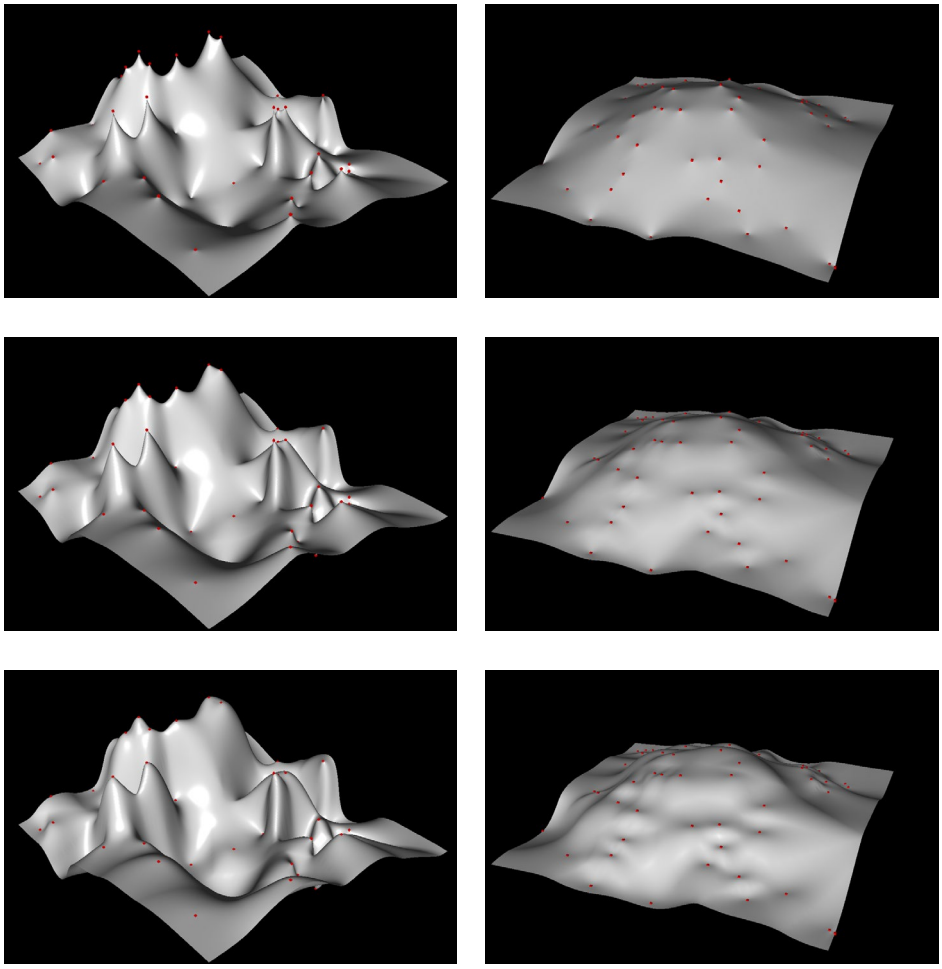


图 3 两组数据的修改的混合邻域插值（公式（14））结果。上：e=2；中：e=4；下：e=8

事实上，混合邻域插值的方程存在与径向基函数插值一一对应的形式：

$$v(\mathbf{x})-\frac{1}{e}d^{2-e}(\mathbf{x})\nabla\cdot d^e(\mathbf{x})\nabla v(\mathbf{x})=p(\mathbf{x}) \tag{37}$$

要求变量 $e\geq 2$。方程（37）中，随着 e 的增大，趋近于一类薄板线性基函数插值，同时方程的数值求解精度越低。混合邻域插值的目的是用 PDE 方法逼近最近自然插值的解，以提高插值的效率。同时，在方程中可以插入描述地质构造特征的张量。

混合邻域插值的实现可分成两部分：①求解最近距离函数 $d(\mathbf{x})$；②插值。第二步的插值可直接求解方程（36）实现。根据（32）式给出的最近距离的定义，第一步中最近距离函数满足如下的程函方程：

$$\begin{cases}\nabla d(\mathbf{x})\cdot\nabla d(\mathbf{x})=1, & \mathbf{x}\notin \mathrm{X}\\ d(\mathbf{x})=0, & \mathbf{x}\in \mathrm{X}\end{cases} \tag{38}$$

方程（38）不必用有限差分或其它算法求解，Park 等（2006）给出了一种 KD 树的快速求解算法。最近

距离对应的值 p（$\mathbf{x}$）可以在求解最近距离函数的同时获得。

张量约束的混合邻域插值

定义物理空间点之间的地质距离，地质距离不仅取决于两点之间的范数距离，也与空间中的地质张量场有关。空间点 $\mathbf{x}$ 处的地质张量记为 $\mathbf{D}$（$\mathbf{x}$），与方程（38）对应的空间中的“最小地质距离”g（$\mathbf{x}$）的方程为

$$\begin{cases}\nabla g(\mathbf{x})\cdot\mathbf{D}(\mathbf{x})\nabla g(\mathbf{x})=1, \ \mathbf{x}\notin \mathrm{X}\\ g(\mathbf{x})=0, \ \mathbf{x}\in \mathrm{X}\end{cases} \tag{39}$$

加入地质张量约束的混合邻域插值方程变为

$$v(\mathbf{x})-\frac{1}{e}\nabla\cdot g^2(\mathbf{x})\mathbf{D}(\mathbf{x})\nabla v(\mathbf{x})=p(\mathbf{x}) \tag{40}$$

这里令方程中的系数 e=4。

把离散数据插值转变为椭圆方程求解的优点：①可以把插值问题转化为解椭圆方程的问题，在离散点控制下的方程的解就是插值结果。②回答了散乱数据插值的解的含义。③可以比较方便地引入其它约束信息进行插值。

地质构造约束插值关键的问题是定义椭圆方程中的“距离”系数。在散乱数据插值方法中，用插值点与支撑点之间的“距离”定义影响权重，是散乱数据插值方法的核心思想。但是，此“距离”非彼“距离”。应该考虑它们之间的“影响距离”而不是空间距离进行权重的计算。

4.2 张量描述图像结构

勘探地震应用中，存在几种与构造有关的张量来表达反射界面的空间相关性信息。它们分别是：梯度平方张量；结构张量；描述图像纹理的扩散张量；地质张量。

1、基础的梯度平方张量：

$$\mathbf{G}=\nabla u(\nabla u)^T=\begin{pmatrix}u_{x_1}^2 & u_{x_1x_2}\\ u_{x_2x_1} & u_{x_2}^2\end{pmatrix} \tag{41}$$

2、对 G 的每个元素在物理空间上进行平滑，得到结构张量 C。

3、地质张量（Hale，2009）：

$$\mathbf{D}(\mathbf{x})=s\frac{\mathbf{C}^{-1}(\mathbf{x})}{1-c(\mathbf{x})} \tag{42}$$

其中，s 是伸缩系数，保证地质张量中最大特征值等于 1；c∈[0,1]是描述张量可信度的系数，越大可信度越高。

4、扩散张量（Weickert，1997）

$$\mathbf{C}=\begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{pmatrix}^{-1}\begin{pmatrix}\mu_1 & \\ & \mu_2\end{pmatrix}\begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{pmatrix} \tag{43}$$

其中，边缘增强型各向异性扩散系数定义为：

$$\begin{cases} \lambda_1 = \begin{cases} 1 & \mu_1 = 0 \\ 1-\exp\left(\dfrac{-3.315}{\mu_1^4}\right) & else \end{cases}, \\ \lambda_2 = 1 \end{cases} \tag{44a}$$

相干增强型各向异性扩散系数定义为：

$$\begin{cases} \lambda_1 = \alpha \\ \lambda_2 = \begin{cases} \alpha & \mu_1 = \mu_2 \\ \alpha + (1-\alpha)\exp\left(\dfrac{-1}{(\mu_1-\mu_2)^2}\right) & \mu_1 \neq \mu_2 \end{cases} \end{cases} \tag{44b}$$

已知偏移剖面和地下空间局部位置的速度值，可从偏移剖面中提取描述地质构造特征的张量，然后插值出规则的网格速度场。图 4 给出了已知的偏移剖面，图中示意了模拟的五口井的轨迹。从成像剖面中提取的相干增强各向异性扩散张量如图 5 所示。在此基础上利用混合邻域插值得到了规则的速度场，如图 6 所示。从中可看出，本实验利用 5 个一维的速度场即可插值出符合地质逻辑的二维速度场。

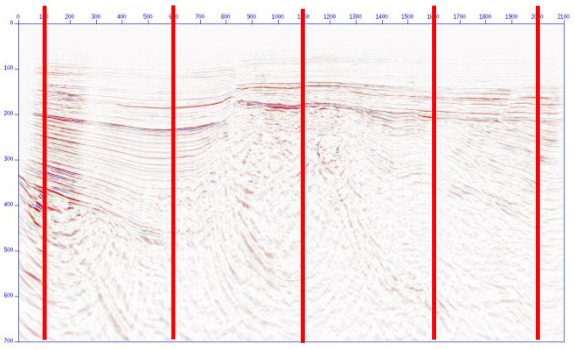


图 4 偏移剖面和井轨迹示意图

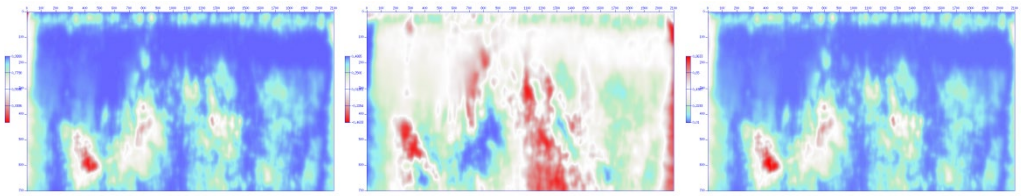


图 5 从偏移剖面提取出的扩散张量，左、中、右分别为 xx、xz、zz 分量

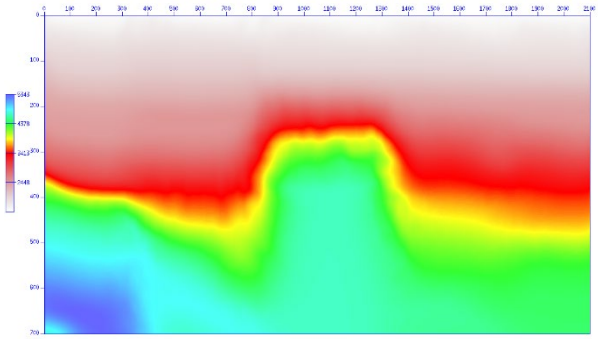


图 6 地质构造约束下的混合邻域插值结果

4.3 小结

对基于椭圆方程进行混合邻域插值的理解：①基于椭圆方程的散乱数据插值是一类值得关注的方法，可将一些先验信息加入反演中。②散乱数据插值中“距离”是非常重要的因素，不同学科有不同的理解和定义。③提取图像的特征，应用到“距离”的计算过程中，进而体现到插值中。④把地质构造约束插值方法应用到速度/各向异性参数/Q 值参数反演和建模的重要工作。

最后有如下总结性观点：⑴地震层析成像是参数（速度/各向异性参数/Q 值参数）建模的重要技术。⑵正则化是必须要考虑的因素。尽管正则化的最佳方法是加入模型和数据的协方差矩阵逆，但是这种方法基本是不现实的。反演过程中考虑构造约束的正则化有效方法：①基于图像结构张量的各向异性高斯滤波函数的构建。②基于各向异性扩散方程的 TV 正则化方法。⑶关于考虑地质构造约束的、基于散乱数据插值的、融合各种与反演参数有关的先验信息的初始模型建立方法，前述的考虑地质“距离”的椭圆方程插值方法是一种较好的选择。也可以用地质“距离”改进 Kriging 插值方法，以及局部邻域插值方法，发展更为有效的地质构造约束下的散乱数据插值方法。

5 扩散方程在勘探地震中的应用

从成像剖面中自动提取反射结构、甚至更为细节的地震相结构在地震波反演成像和地震地质解释中具有重要作用。这也是一般图像处理中结构提取要解决的问题。

图像结构提取的方法有很多，此处我们仅仅讨论扩散方程，尤其是各向异性扩散方程，在地震图像结构提取（也包括去噪）中的应用。下面的分析说明各向异性扩散方程用于图像结构提取，等价于 TV 正则化过程。这也从一个方面说明了 TV 正则化的几何意义，说明了 TV 正则化如何对反演成像中的梯度项进行处理。

在地震波反演成像中，广泛应用的模型是如下的 TV 正则化模型（也称 ROF 模型）：

$$T_\alpha = \frac{1}{2}\left\| K(\mathbf{m}) - \mathbf{d} \right\|^2 + \alpha \mathbf{TV}(\mathbf{m}) \tag{45}$$

图像处理中也广泛应用有如下的模型：

$$T_\alpha = \frac{1}{2}\left\| K\mathbf{u} - \mathbf{d} \right\|^2 + \alpha \mathbf{TV}(\mathbf{u}) \tag{46}$$

从两类问题具有如此类此的数学模型，可以看出图像处理中的很多思想方法移植到地震波成像中是顺理成章的。

为了避免误差泛函的不可导，勘探地震中的 TV 正则化修改为：

$$\mathrm{TV}(\mathbf{m}) = J_\beta(\mathbf{m}) = \int_0^1\int_0^1 \sqrt{\left(\frac{\partial \mathbf{m}}{\partial x}\right)^2 + \left(\frac{\partial \mathbf{m}}{\partial y}\right)^2 + \beta^2}\, dxdy \tag{47}$$

它来自于如下的误差泛函定义式

$$T_\alpha = \frac{1}{2}\left\| K(\mathbf{m}) - \mathbf{d} \right\|^2 + \alpha J_\beta(\mathbf{m}) \tag{48}$$

（47）式定义的 TV 正则化罚函数的计算可以表示为：

$$\begin{aligned}\mathrm{TV}(\mathbf{m}) &= J_\beta(\mathbf{m}) = \int_0^1\int_0^1 \sqrt{\left(\frac{\partial \mathbf{m}}{\partial x}\right)^2 + \left(\frac{\partial \mathbf{m}}{\partial y}\right)^2 + \beta^2}\, dxdy \\ &\approx \frac{1}{2}\psi \sum_{i=1}^{nx}\sum_{j=1}^{ny}\left(\left(D_{i,j}^x \mathbf{m}\right)^2 + \left(D_{i,j}^x \mathbf{m}\right)^2\right)\end{aligned} \tag{49}$$

其中，$D_{i,j}^x \mathbf{m} = \dfrac{\mathbf{m}_{i,j} - \mathbf{m}_{i-1,j}}{\Delta x}$，$D_{i,j}^y \mathbf{m} = \dfrac{\mathbf{m}_{i,j} - \mathbf{m}_{i,j-1}}{\Delta y}$。

罚函数的变分可以写成：

$$\begin{aligned}\frac{d}{d\tau} J_\beta(\mathbf{m}+\tau\mathbf{v})\Big|_{\tau=0} &= \frac{1}{2}\sum_{i=1}^{nx}\sum_{j=1}^{ny}\psi\left(\left(D_{i,j}^x \mathbf{m}\right)^2 + \left(D_{i,j}^x \mathbf{m}\right)^2\right) \\ &= \sum_{i=1}^{nx}\sum_{j=1}^{ny}\psi'_{i,j}\left[\left(D_{i,j}^x \mathbf{m}\right)\left(D_{i,j}^x \mathbf{v}\right) + \left(D_{i,j}^y \mathbf{m}\right)\left(D_{i,j}^y \mathbf{v}\right)\right]\end{aligned} \tag{50}$$

其中，$\psi'_{i,j} = \psi'_{i,j}\left[\left(D_{i,j}^x \mathbf{m}\right)^2 + \left(D_{i,j}^y \mathbf{m}\right)^2\right]$。

上式可以进一步写成，

$$\begin{aligned}&\frac{d}{d\tau} J_\beta(\mathbf{m}+\tau\mathbf{v})\Big|_{\tau=0} = \left\langle \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_x \mathbf{m}, D_x \mathbf{v}\right\rangle + \left\langle \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_y \mathbf{m}, D_y \mathbf{v}\right\rangle \\ &= \left\langle D_x^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_x \mathbf{m}, \mathbf{v}\right\rangle + \left\langle D_y^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_y \mathbf{m}, \mathbf{v}\right\rangle \\ &= \left\langle D_x^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_x \mathbf{m} + D_y^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_y \mathbf{m}, \mathbf{v}\right\rangle \\ &= \left\langle \mathrm{grad} J(\mathbf{m}), \mathbf{v}\right\rangle\end{aligned} \tag{51}$$

据此，可以写出罚函数对应的梯度：

$$\begin{aligned}&\mathrm{Grad} J(\mathbf{m}) = D_x^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_x \mathbf{m} + D_y^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_y \mathbf{m} \\ &= \begin{bmatrix} D_x^T & D_y^T \end{bmatrix} \begin{bmatrix} \mathrm{diag}\left(\psi'(\mathbf{m})\right) & 0 \\ 0 & \mathrm{diag}\left(\psi'(\mathbf{m})\right) \end{bmatrix} \begin{bmatrix} D_x \\ D_x \end{bmatrix} \mathbf{m}\end{aligned} \tag{52}$$

我们可以对罚函数的梯度做如下的解释。若把（52）式写成

$$\begin{aligned}&\mathrm{Grad} J(\mathbf{m}) = D_x^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_x \mathbf{m} + D_y^T \mathrm{diag}\left(\psi'(\mathbf{m})\right) D_y \mathbf{m} \\ &= \begin{bmatrix} D_x^T & D_y^T \end{bmatrix} \begin{bmatrix} \mathrm{diag}\left(\psi'(\mathbf{m})\right) & 0 \\ 0 & \mathrm{diag}\left(\psi'(\mathbf{m})\right) \end{bmatrix} \begin{bmatrix} D_x \\ D_x \end{bmatrix} \mathbf{m}\end{aligned} \tag{53}$$

梯度项可以解释为：

$$\frac{\partial \mathbf{m}}{\partial \tau} = -\nabla \cdot \left(\psi'(\mathbf{m}) \nabla \mathbf{m}\right) = -\frac{\partial}{\partial x}\left(\psi' \frac{\partial \mathbf{m}}{\partial x}\right) - \frac{\partial}{\partial y}\left(\psi' \frac{\partial \mathbf{m}}{\partial y}\right) \tag{54}$$

其中，$\psi'(\mathbf{m})=\psi'\left(\left(\frac{\partial \mathbf{m}}{\partial x}\right)^2+\left(\frac{\partial \mathbf{m}}{\partial y}\right)^2\right)$。可见（54）式是一个扩散方程的形式。进行上式的处理就等价于 TV 正则化的施加。具体地，在数据偏差泛函的梯度计算出来后，得到$\tilde{\mathbf{m}}'$。对$\tilde{\mathbf{m}}'$施加（54）式定义的处理，就在进行 TV 正则化。

扩散方程中的系数$\psi'(\mathbf{m})$决定了扩散方程的应用效果，即是否是保持结构的去噪。考虑图像灰度结构特征决定$\psi'(\mathbf{m})$，可以定义各向异性扩散方程，应用异性扩散方程可以得到保持图像结构特征的去噪。此处所谓的去噪也可以认为是结构特征保持，把非特征的灰度成分压制掉。

下面我们专门讨论一下图像去噪问题。首先是 L2 范数下的去噪问题定义为如下的反问题：

$$\inf_{u\in H^1(\Omega)}\left(\frac{1}{2}\|f-R\mathbf{m}\|^2+\alpha J_\beta(\mathbf{m})\right)=\inf_{u\in H^1(\Omega)}\left(\frac{1}{2}\|f-R\mathbf{m}\|^2+\alpha\int_\Omega|\nabla\mathbf{m}|^2d\mathbf{x}\right) \tag{55}$$

对应的变分方程为：

$$R^*(\mathbf{f}-R\mathbf{m})+\alpha\Delta\mathbf{m}=0 \tag{56}$$

解方程（56）等价于解一个如下的扩散问题：

$$\begin{cases}\dfrac{\partial\mathbf{m}}{\partial\tau}(\tau;\mathbf{x})=R^*(\mathbf{f}(\mathbf{x})-R\mathbf{m}(\tau;\mathbf{x}))+\alpha\Delta\mathbf{m}(\tau;\mathbf{x}),(\tau;\mathbf{x})\in(0,T]\times\Omega\\ \mathbf{m}(0;\mathbf{x})=\mathbf{f}(\mathbf{x}),\mathbf{x}\in\Omega\\ \left.\dfrac{\partial\mathbf{m}(\tau;\mathbf{x})}{\partial\mathbf{n}}\right|_{\partial\Omega}=0,(\tau;\mathbf{x})\in(0,T]\times\Omega\end{cases} \tag{57}$$

很多情况下，L2 范数下去噪不能得到最佳效果。为此，引入 TV 范数下的去噪问题：

$$\hat{\mathbf{m}}=\inf_{u\in H^1(\Omega)}\left(\frac{1}{2}\|\mathbf{f}-R\mathbf{m}\|^2+\alpha|\mathbf{m}|_{\mathrm{TV}}\right) \tag{58}$$

对应的变分方程为：

$$R^*(\mathbf{f}-R\mathbf{m})+\alpha\cdot\mathrm{div}\left(\frac{\nabla\mathbf{m}}{|\nabla\mathbf{m}|}\right)=0 \tag{59}$$

解方程（59）等价于解一个如下的扩散问题：

$$\begin{cases}\dfrac{\partial\mathbf{m}}{\partial\tau}(\tau;\mathbf{x})=R^*(\mathbf{f}(\mathbf{x})-R\mathbf{m}(\tau;\mathbf{x}))+\alpha\cdot div\left(\dfrac{\nabla\mathbf{m}}{|\nabla\mathbf{m}|}\right)(\tau;\mathbf{x}),(\tau;\mathbf{x})\in(0,T]\times\Omega\\ \mathbf{m}(0;\mathbf{x})=\mathbf{f}(\mathbf{x}),\mathbf{x}\in\Omega\\ \left.\dfrac{\partial\mathbf{m}(\tau;\mathbf{x})}{\partial\mathbf{n}}\right|_{\partial\Omega}=0,(\tau;\mathbf{x})\in(0,T]\times\Omega\end{cases} \tag{60}$$

（59）和（60）式对应一个各项同性扩散方程的去噪，不是在结构特征保持下的去噪，效果不佳。为此，修改（59）得到：

$$R^*\left(\mathbf{f}-R\mathbf{m}\right)+\alpha\cdot\mathrm{div}\left(\psi\left(\mathbf{m}\right)\frac{\nabla\mathbf{m}}{\left|\nabla\mathbf{m}\right|}\right)=0 \tag{61}$$

解方程（60）等价于解一个如下的扩散问题：

$$\begin{cases}\dfrac{\partial\mathbf{m}}{\partial\tau}\left(\tau;\mathbf{x}\right)=R^*\left(\mathbf{f}\left(\mathbf{x}\right)-R\mathbf{m}\left(\tau;\mathbf{x}\right)\right)+\alpha\cdot div\left(\psi\left(\mathbf{m}\right)\dfrac{\nabla\mathbf{m}}{\left|\nabla\mathbf{m}\right|}\right)\left(\tau;\mathbf{x}\right),\left(\tau;\mathbf{x}\right)\in\left(0,T\right]\times\Omega\\ \mathbf{m}\left(0;\mathbf{x}\right)=\mathbf{f}\left(\mathbf{x}\right),\mathbf{x}\in\Omega\\ \left.\dfrac{\partial\mathbf{m}\left(\tau;\mathbf{x}\right)}{\partial\mathbf{n}}\right|_{\partial\Omega}=0,\left(\tau;\mathbf{x}\right)\in\left(0,T\right]\times\Omega\end{cases} \tag{62}$$

从上述公式可以看出，带地质约束的 TV 正则化参数反演问题，可以统一地在 ROF 模型下得到解决。TV 正则化等价于解一个各向异性扩散方程进行去噪。

对比可见，可以把图像处理中的各向异性扩散方程去噪问题引入到勘探地震反演成像方法中。尽管看起来它们来自不同的途径，**但殊途而同归！从各向异性扩散方程的去噪角度更容易理解 TV 正则化的含义。**

对扩散方程应用的评价：①扩散方程本质上是一个耗散方程，因此随着迭代次数的不断增加，图像最终将达到均匀的状态。②各向异性扩散方程与各项同性的扩散方程相比，最大的突破在于引入了方向信息。结构张量可以对图像中的方向进行识别，但是由于高斯平滑的作用，导致边界上的尖锐部分会受到平滑作用的影响，变得光滑。需要考虑提出改进的保护尖灭、断点等图像特征的方法。③各向异性扩散方程用于结构信息的提取流程中，主要是起到提高资料品质的作用，是对地震图像进行预处理的一部分，对结构信息的提取仍然需要考虑其他的图像处理方法。譬如边缘检测、图像分割、结构特征提取等。

地震波层析成像或偏移成像中如何利用（各向异性）扩散方程：①LS_RTM/LS_RTM 中，主要是改善梯度的性质；②也可以在 FWI 中，实现层析成像的梯度项与偏移成像梯度项的分离。③在波动方程层析成像中，可以滤除隐含的偏移成像对应的梯度。④可以对成像叠加剖面进行结构提取，用于建模过程中。

6 总结与讨论

现代信号处理和现代图像处理研究的对象是平方可积的函数，这类函数构成一个线性空间，它是 Hilbert 空间中的典型实例。奠基在 Hilbert 空间基础上的线性运算和算子投影理论是信号/图像处理（分析）方法的思想基础。因此，所有应用领域的信号/图像分析都被置于现代信号处理和现代图像处理理论框架下。现代信号处理和现代图像处理领域是所有现代科学和技术分支中最为活跃、进展最显著的。

在人工智能思想和方法渗透所有相关领域的时候，勘探地震学这类与现代信号和现代图像分析学科本来就密不可分的相关学科更应该吸取两学科的先进思想。事实上，也的确在不断从现代信号和现代图像分析学科的进展中寻找借鉴的思想和方法。

油气地震勘探的目标是完成对含油气目标体的定位、识别与描述，在此基础上进行油气藏的开发。其中，地震波反演成像无疑是核心问题，其目的是期望弄清楚地下岩石介质的弹性参数分布，无论是低波数的光滑部分，或是高波数的跃变部分。给出精确的岩石弹性介质低波数成分的估计，就可以完成线性化的高波数成分的估计。目前广泛用于油气藏识别的高波数成分是（角度）反射系数，正在向利用高维地震数据估计波阻抗的方向发展。但是，仅仅依赖地表观测的、观测孔径和频带有限的、采样不规则或采样密度不充分的、信噪比不高的、震源函数未知的叠前地震数据不可能获得完全满足油气藏勘探需求的反演成像结果。

稳定而可靠的地震波反演成像需要其他额外信息的约束，也需要各种正则化思想和方法的引入，尤其是反射地震界面的几何结构的约束。额外的信息主要来自钻井、测井和地质，这些与地下岩石介质的弹性参数分布有关的信息的引入可以有效地提高反演成像的稳定性和精度，一般地它们表现为空间离散数据的形式，需要有效的散乱数据插值方法把它们表示成初始模型的形式，约束反演成像的结果。另一方面，反演成像过程中，无论对数据残差和对模型本身都要提出有效的正则化思想和方法来约束反演解，使它具有明确的地

质地球物理意义。

地震波反演成像发展到现在，Bayesian 框架下的反演理论、方法等都已经十分清晰，增加解的稳定性的各种正则化方法都提出很多。但是，FWI+LS_RTM 一类的地震波反演成像核心方法依然不能很好地用于实际地震数据成像处理中。此时，需要借用现代信号分析和现代图像分析方面新思想和新方法于地震波反演成像、甚至地震成像数据的自动化地质解释中。首先是表达成散乱数据的各种与反演参数有关的先验信息需要融合成有地质和地球物理意义的数据体，融合后的数据体之所以体现出地质与地球物理意义主要是要在上述信息融合过程中引入层状沉积的反射界面及断层结构的约束。因此，需要改进现有的散乱数据插值方法，发展出受地质构造约束的散乱数据插值方法。这样的插值方法应该也可以推广到其他应用领域中。第二个重要的问题是自动抽取成像剖面上的反射结构特征，包括粗线条的和细节的几何特征。当然，也包括自动抽取特征前的各种图像预处理的方法；也包括自动抽取各种地震数据属性的方法。这些方法以前主要用在地震剖面的自动解释中。现在，在地震波反演成像中，从成像后的剖面上提取结构信息用来特征表达反演参数场，或约束反演成像结果变得必不可少。看起来，地震地质解释和地震波反演成像也逐渐再融合。当然，总的目的还是要完成含油气油藏的识别、描述与刻画。

我们已知，扩散方程在现代图像结构提取中已经得到了广泛而有效的应用，各向异性扩散方程在图像结构提取中的作用与参数反演中的 TV 正则化方法是基本等价的；椭圆方程用在散乱数据插值过程中，最近也得到比较多的重视，这种插值方法体现出了更容易融合结构信息约束的特点，在地震参数建模过程中有明显的潜在优势。

到目前为止，我们认为物理信号和图像的特征表达和特征提取依然是最关键问题。当前信号和图像处理的效果不令人满意，关键原因就在于此。因此，字典学习成为核心问题逻辑上是很自然的。无监督学习是现代信号和图像智能处理的另一个瓶颈问题。这两个问题如果有本质突破，勘探地震的反演成像与地震地质解释精度会有一个跃变。

参考文献

[1] 冯象初 王卫卫编著，2009 年，图像处理的变分和偏微分方程方法，科学出版社

[2] 吴宗敏，2007 年，散乱数据拟合的模型、方法和理论，科学出版社

[3] Aster, R. C., Borchers, B., and Clifford Thurber, C., 2005, Parameter Estimation and Inverse Problems; Elsevier Academic Press

[4] Aubert, G., Kornprobst, P., 2006, Mathematical Problems in Image Processing – Partial Differential Equations and the Calculus of Variations （Second Edition）; Springer

[5] Chan, T. F., Shen, J_H,., 2005, Image Processing and Analysis: Variational, PDE, Wavelet and Stochastic Methods, SIAM Dave Hale, Image-guided blended neighbor interpolation, CWP-634, Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401, USA

[6] Iske, A., 2004, Multiresolution Methods in Scattered Data Modelling, Springer.

[7] Nolet, G. （Ed.）, 1987, Seismic Tomography with applications in Global Seismology and Exploration Geophysics, D. Reidel Publishing Company

[8] Park, S.W., L. Linsen, O. Kreylos, et al., 2006, Discrete Sibson interpolation: IEEE Transactions on Visualization and Computer Graphics, 12, 243-253

[9] Plata, 2008, A review on the piecewise polynomial harmonic interpolation, *Applied numerical mathematics*, 58:1168-1185

[10] Stein, M. L., 1999, Interpolation of Spatial Data - Some Theory for Kriging,

[11] Scherzer, O., 2011, Handbook of Mathematical Methods in Imaging, Springer

[12] Sandwell, D.T., 1987, Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data: Geophysical Research Letters, 14, 139-142

[13] Sibson, R., 1981, A brief description of natural neighbor interpolation, in V. Barnett, ed., Interpreting Multivariate Data: John Wiley & Sons, 21-36

[14] Weickert J., 1997, A review of nonlinear diffusion filtering, Scale-Space Theory in Computer Vision. Springer Berlin Heidelberg, 1-28

[15] Wendland, H., 2005, Scattered Data Approximation, Cambridge University Press

特征波反演成像理论框架

王华忠，冯波，王雄文，胡江涛，刘少勇，李辉，周阳

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】地震波反演成像的核心问题是把解一个非线性（较）强的反问题转化为提一个更凸的反问题进行求解。强非线性性的出现主要是由于实测数据与要反演的模型参数之间的关系远非线性的，也由于包含模型参数的控制方程不能很好地预测实际数据导致的。因此，我们提出了特征波反演成像的理论框架。基本思想是不追求对实测波场全部波现象的模拟，而是模拟其中的部分特征波场；也不一定追求波形逼近，而是尽可能利用到达时（相位）的逼近。相对于全波形反演 FWI（Full Waveform Inversion），特征波反演被称为 CWI（Characteristic Wave Inversion）。CWI 由四个基本步骤组成：特征波场（CWF = Characteristic Wave Field）的提取；波动理论的透射波层析成像；波动理论的一次反射波层析成像；最小二乘叠前深度偏移成像。特征波成像是基于对实测炮集中包含波现象的认识的一种反演框架。特征波场提取是其中重要的环节，包含三重含义：1 波现象的分解（譬如，矢量波分解成标量波、一次波和多次波分解）；2 时空局部的、单震相的、带方向的、带限波场的分解；3 同相轴上地震子波的分解（譬如，提出子波的达到时、相位等）。特征波提取基于压缩感知的框架进行。透射波波动理论层析基于初至和早至波到达时和相位。基于扰动波场的反演成像，首先考虑的是针对反射波、地下介质参数层状分布的情形，然后再考虑针对散射和绕射的情形。数据域的特征波反演估计低波数速度信息仅依赖同相轴上子波的到达时或/和相位信息，并尽量排除振幅对到达时和相位估计的影响。像域中的背景速度反演仅适宜进行基于到达时的反演，基于像的幅值的反演在理论上是不合理的。高波数参数估计时，首先进行方位角度反射系数的估计，在此基础上进行散射强度的估计。CWI 技术系列是推进经典 FWI 走向实用化的正确途径，初步的数值试验证明了上述判断。

【关键词】反演成像，特征波场，特征波反演，非线性问题，凸问题，数据域反演，像域反演

1 引言

岩性油气藏的精确描述是勘探地震学的重要目标。小尺度地质体的精确构造成像、振幅保真的成像、反射系数甚至其它弹性参数的精确估计是勘探地震学中的核心问题。随着高性能计算机技术的快速进步以及全方位、高密度、宽频带的地震数据采集技术的飞速发展，以估计“全或宽”波数带的弹性参数为目标的全波形反演方法（FWI）再次成为勘探地震学中的研究热点。而（各向异性）逆时偏移（RTM）方法的成功实用化进一步推动了对 FWI 技术的研究。然而，由于叠前地震数据的不完备（如有限观测孔径、不规则观测以及地表不一致的观测，也包括强噪音环境下的观测）、正演模型的不完善（正演算子无法很好地预测实际波场中的各种波现象）、初始模型距离反演的“真”参数模型太远、地震子波的未知且空变等多种因素制约，使得经典 FWI 难以达到其理论预计的目标，甚至在实际应用中根本不收敛。

本质上，经典 FWI 是一个非线性性很强的反演问题，根本原因在于实际观测的地震波场的变化与要反演参数的变化之间关系很弱，另外，包含要反演参数的控制方程不能很好地预测实测数据也是一个重要因素。而地震波反演成像的核心问题是把解一个非线性（较）强的反问题转化为提一个更凸的反问题进行求解。如何把一个非线性强的反问题转化为凸性更好的反问题进行求解，一直是地球物理反问题研究人员关注的焦点。典型地，包括引入波场的 Laplace 变换（Shin 和 Cha，2009）；引入波场的对数变换（Shin 和 Min，2006）；引入波场的 Hilbert 变换（Bozdag 等，2011；Wu 等，2014）。还有不进行数据逼近而是利用像域中成像道集的拉平或聚焦建立目标泛函进行反演（Shen 和 Symes，2008）。

更为全面地讨论如何构建较凸反问题的应该是波场重构的反演（WRI）方法（Van Leeuwen，Herrmann 和 Peters，2014；Van Leeuwen 和 Herrmann，2013），其基本思想是通过引入增广的状态变量解决正问题不能很好地预测实际地震数据中波现象的问题。但是，引入增广的状态变量并没有明确地指出仅仅特征波场和特征参数之间才存在更为密切的、更为线性的联系。勘探地震中地下介质参数的分布和叠前数据本身都是具有特征的，特征波场和特征参数之间才存在更为密切的关系（王华忠等，2015；王华忠等，2016）

本文明确提出仅仅用与特征参数关系密切的特征波场进行反演。特征波反演成像由四个基本步骤组成：特征波场的提取；波动理论的透射波层析成像；波动理论的一次反射波层析成像；最小二乘叠前深度偏移成像。特征波反演成像（CWI）的中心工作在于特征波场的提取。特征波场提取应包括三方面的含义：特征波型的提取（波型分解；一次波和多次波分解）；压缩感知框架下时空局部的方向波场的提取；反射子波的特征分解（譬如提取反射子波的到达时和相位）。此文重点分析了压缩感知框架下时空局部的方向波场的分解以及基于此波场的数据域和成像域的反演思想与方法。指出特征波场与反演参数之间的关系更紧密，基于特征波场的反问题凸性更好。另外，特征波反演成像中梯度的预条件要相对简单，因为此时特征波对应的梯

度干扰大大减弱。尽管此文没有重点讨论反射子波特征的提取，并非此问题对精确的反演成像不重要（冯波，2015）。

特征波成像是基于对实测炮集中所包含波现象的认识的一种反演框架。我们认为，只有经过相当长时间的特征波反演成像方法的研究，经典 FWI 中所蕴含的问题才能逐步得到认识和解决，进而发展出工业界实用化的反演成像技术。

2 求解强非线性反问题或是提出更凸的反问题？

地震波反演成像是勘探地震中的核心问题，它是一个标准的非线性反演问题，其非线性程度随着地下介质的复杂程度（譬如小尺度速度异常体存在、横向变速剧烈等）而提高。实测波场变化与地下介质变化之间的非线性程度决定了地震波反演问题的非线性程度。针对勘探地震中的参数反演问题，到底是专注于寻找解决较强非线性反演问题的方法或是专注于提一个凸性更强的反演问题来求解？

Rockafellar（1993）提出“In fact the great watershed in optimization isn't between linearity and non-linearity, but convexity and non-convexity”。显然，提一个更凸的反问题比奋力地解一个强非线性反问题更合理。

从 Tarantola（1984）建立以波动方程为基础的地震反演理论以后的近三十年（尤其是最近几年）来，随着以“两宽一高”为代表的地震数据采集技术以及以 GPU 为代表的高性能计算机技术的推进，全波形反演（FWI）方法和技术得到了长足的进步。在降低反演问题非线性性方面做了很多有益的探索，提出了很多有效的方法。基本思想是从数据或波场入手来改进反演成像方法，典型的做法包括：从低频数据到高频数据的多尺度反演方法（Bunks 等，1995；Sirgue 和 Pratt，2004），从近偏移距数据到远偏移距数据的反演（Virieux 和 Operto，2009）；由小到大调节 Laplace 阻尼参数进行的由初至时刻不断放大时窗的反演（Shin 和 Cha，2009）；定义不同的失配函数的反演（Cara 和 Lévêque，1987；Luo 和 Schuster，1991；Gee 和 Jordan，1992；Fichtner 等，2008；van Leeuwen 和 Mulder，2010；Bozdag 等，2011；Moghaddam 和 Mulder，2012；Warner，2014）。还有另一类方法则是将传统的数据域反演扩展到成像域的反演（de Hoop 等，2006；Shen 和 Symes，2008；Weibull 等，2012）。

上述反演方法并没有从更为抽象的角度主动地把地震波反演问题提成一个更凸的问题。本文希望在数据的特征表达和参数的特征表达概念下，把经典 FWI 转化为一个临近的、更凸的反问题进行求解。

3 Bayes 反演理论的缺陷

Bayes 反演理论是地震波反演成像的基础（Tarantola，2005）。Bayes 估计的思想可以简单概括为：参数估计的结果 $\hat{\mathbf{m}}$ 是由后验概率密度函数 $P(\mathbf{m}|\mathbf{y})$ 所决定的均值 $E\{\mathbf{m}|\mathbf{y}\}$，其中 $\mathbf{y}$ 代表观测数据。

$$\hat{\mathbf{m}} = E\{\mathbf{m}|\mathbf{y}\} = \int \mathbf{m} P(\mathbf{m}|\mathbf{y}) d\mathbf{m} \tag{1}$$

式（1）表示参数估计的结果是由观测数据所对应的模型参数出现的概率所决定的。

由于后验概率密度 $P(\mathbf{m}|\mathbf{d})$ 很难估计，所以将其转化为较容易估计的模型空间的先验概率密度函数 $P(\mathbf{m})$、数据空间的概率密度函数 $P(\mathbf{d})$ 和条件概率密度函数 $P(\mathbf{d}|\mathbf{m})$ 来表示，其中 $P(\mathbf{d}|\mathbf{m})$ 体现了正算子的不确定性。该 Bayes 公式为

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{m}|\mathbf{d})P(\mathbf{m})}{P(\mathbf{d})} \tag{2}$$

Bayes 估计的基本思想是后验概率密度 $P(\mathbf{m}|\mathbf{d})$ 最大时（即 MAP 方法）得到最好的反演解，反演解由期望值为 $\hat{\mathbf{m}} = \int \mathbf{m} P(\mathbf{m}|\mathbf{y}) d\mathbf{m}$ 给出，解的不确定性由对应的方差决定。利用（1）式或基于（2）式的 MAP 方法，都需要后验概率密度，计算上是不可行的。目前 Bayes 估计的主要方式是假设已知先验概率密度

$P(\mathbf{d}|\mathbf{m})$ 和 $P(\mathbf{m})$ ，一般地假设它们均满足高斯或广义高斯分布，基于 MAP 方法把参数估计问题转化为（3）式定义的代价函数求解极小的问题。若忽略正算子的不确定性，则代价函数 $S(\mathbf{m})$ 可以简单地定义为

$$S(\mathbf{m})=\frac{1}{2}\left[(g(\mathbf{m})-\mathbf{d}_{obs})^H C_D^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})+(\mathbf{m}-\mathbf{m}_{prior})^H C_M^{-1}(\mathbf{m}-\mathbf{m}_{prior})\right] \quad (3)$$

式中：C_D 为数据协方差矩阵；C_M 为模型协方差矩阵。$\mathbf{d}_{cal}=g(\mathbf{m})$ 代表正问题，$\mathbf{m}_{prior}$ 代表参数的先验值。

（3）式存在的问题是隐含了一个很基本的假设：正问题 $g(\mathbf{m})$ 是可以解释实测数据 $\mathbf{d}^{obs}$ 的，不能解释的残差是符合高斯或某种广义高斯分布的。这种假设与勘探地震的实际是很不相符的，必须进行修改。

Tarantola（2005）指出：“正问题不能很好地解释实测地震数据，其引入的数据残差与观测噪音（非信号部分）引入的数据残差可能是同一水平的”。

我们认为，考虑到正演模型选择过于简单（譬如仅考虑单参数）及震源函数未知等等因素的影响，正问题的表达不合适引起的数据逼近误差很可能超过观测噪音引起的数据逼近误差。这就使经典的 FWI 完全失去了意义。Bayes 估计理论似乎蕴含了正问题不准确引起的问题，但要么忽视它，要么没有具体的解决方案。

以下，我们把正问题的计算结果——**状态变量**——引入进来，并把**状态变量**作为与模型参数同等地位的待反演参数看待。这是深化对地震波反演问题的认识、把非线性的地震波反演成像问题提成一个更凸的反演问题的关键。

4 对状态变量独立的地震波反演问题的认识

状态变量是最优控制理论（赫孝良和葛照强，2015）描述系统状态的变量，与系统参数具有同等地位，共同满足控制方程。

勘探地震中，引入状态变量和系统参数的正问题的描述方法如下。把激励、探区介质分布、响应视为一个系统，对应该系统的非线性正问题的数学描述可以抽象为：

$$A(\mathbf{m})\mathbf{u}=-\delta(\mathbf{x}-\mathbf{x}_s)\boldsymbol{f}=-\boldsymbol{f}(\mathbf{x}_s,t) \quad (4)$$

其中，A 为正演算子。在地震波反演中，A 的一般形式为

$$A=\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)=\frac{1}{v^2}\frac{\partial^2}{\partial t^2}-\Delta \quad (5)$$

A 的形式受所选择的波动方程的影响，可以选常密度声波/变密度声波方程、各种拟声波方程、弹性波方程等。$\mathbf{u}$ 为要正演合成的波场，即状态变量（系统响应）；$\boldsymbol{f}$ 为震源函数（激励函数）。

野外观测到的地震数据可以表述为

$$P(\mathbf{d}^{obs})=P(\mathbf{u})+\boldsymbol{\eta} \quad (6)$$

此处，P 表示针对实测数据和状态变量的变换算子，也可以分别定义变化算子，但变换结果的物理意义应该是对应的（譬如都是透射波或一次反射波）。$\boldsymbol{\eta}$ 表示噪音函数，在大部分情况下假设它是高斯白噪的。

实测数据和状态变量之间的关系由（6）式定义，其中蕴含了引入变换算子使得实测数据和状态变量进行匹配的思想，也可以说引入了让状态变量独立出来，作为一个可以处理的参量看待的思想。不再认为状态变量完全由系统参数、控制方程和激励函数决定，尽管理论上的确由它们确定。这样的观点对后续的反演成像非常有意义。

地震数据反演成像方法主要包括相互有紧密联系的两条途径：一是 Bayes 估计理论；二是约束最优化理论。广义高斯分布假设下代价函数最小反问题优化求解方法基本上与约束最优化理论是等价的。下面我们从约束优化理论展开状态变量独立的地震波反演成像问题的讨论。

最简单的 FWI 反演成像目标函数定义为：

$$\min_{\mathbf{m}\in\mathbb{M}} \mathcal{J}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{d}^{obs} - \mathbf{d}^{cal}(\mathbf{m})\right\|^2 \tag{7}$$

其中，$\mathbf{d}^{obs}$ 为野外观测数据，$\mathbf{d}^{cal}(\mathbf{m})$ 为正演模拟数据。上式就是数据残差为高斯分布假设下的最大似然估计方法。另一个蕴含的假设是正问题的解能够解释实测地震数据，并使二者的残差满足高斯分布。这与地震波反演成像的实际相差甚远，反演结果有很强的多解性，完全无法满足应用的要求。因此，引出了各种各样的对上式的修改，发展出了很多经典 FWI 的变种形式。一般地，是沿着 Bayes 估计的路线提出各种修改方法。概念上，各种典型的修改方法可以抽象为下式

$$\min_{\mathbf{m}\in\mathbb{M}} \mathcal{J}(\mathbf{m}) = \frac{1}{2}\left\|T\left(\mathbf{d}^{obs}\right) - T\left(\mathbf{u}(\mathbf{m})\right)\right\|^2 + \alpha\left\|D\mathbf{m}\right\|_p, \quad s.t.\mathbf{u}(\mathbf{m}) = \left(A(\mathbf{m})\right)^{-1}\mathrm{f} \tag{8}$$

其中，算子 $T(\cdot)$ 可以选择各种各样的形式。譬如，可以对波场取对数，分别利用相位项的逼近和振幅项的逼近进行反演（$T(\mathbf{u}) = \ln\mathbf{u}$，Shin 和 Min，2006；$T(\mathbf{u}) = \mathrm{Re}(\ln\mathbf{u})$，$T(\mathbf{u}) = \mathrm{Im}(\ln\mathbf{u})$，Bedner 等，2007；Sukjoon 等，2007）；可以对波场进行 Laplace 变换，利用从初至到达时起始的不断扩大的窗函数截取不断增加的波场进行反演（$T(\mathbf{u}) = \int_W u(t)\mathrm{e}^{-st}\mathrm{d}t$，Shin 和 Cha，2009）；还可以对波场取包络，减轻 FWI 的 Cycle-Skipping 现象（$T(\mathbf{u}) = \sqrt{u^2(t) + \tilde{u}^2(t)}$，$\tilde{u}(t) = \frac{1}{\pi t} * u(t)$，Bozdag 等，2011；Wu 等，2014）。但是，上述方法引入的对波场的变换没有涉及一个根本的问题，即实测波场不可以准确预测。我们认为，把正演计算波场作为系统的状态变量独立出来，是提高反演成像认识，发展出更有效的反演成像方法的关键。

5 状态变量独立的地震波反演问题的提法

状态变量独立的地震波反演问题可以提成如下控制方程约束下的优化问题，即：

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \frac{1}{2}\left\|P^{cal}(\mathbf{u}) - P^{obs}\left(\mathbf{d}^{obs}\right)\right\|_2^2 \quad s.t. A(\mathbf{m})\mathbf{u} = \mathbf{q} \tag{9}$$

式中，$P^{cal}(\cdot)$ 为施加在正演波场上的变换函数，$P^{obs}(\cdot)$ 为施加在观测数据上的变换算子。它可以是非线性算子，一般地定义为线性算子。$\mathbf{u}$ 为状态变量（正演波场），$\mathbf{d}^{obs}$ 代表实际观测数据，$\mathbf{m}$ 为系统（或模型）参数。$A(\mathbf{m})\mathbf{u} = \mathbf{q}$ 为控制（或状态）方程。$\mathbb{U}$ 和 $\mathbb{M}$ 分别是状态变量空间和系统（模型）参数空间。二者构成了满足控制方程 $A(\mathbf{m})\mathbf{u} = \mathbf{q}$ 的联合空间。地震波反演成像就是在此联合空间中寻找最优解。事实上，激励函数 $\mathbf{q}\in\mathcal{Q}$ 也是需要反演的，但是在勘探地震中，一般假定激励函数是已知的。为了讨论简单计，关于模型参数和状态变量的罚函数（约束函数），没有包括在上式中。这不会影响下述讨论的结论。

6 特征波反演方法（CWI=Characteristic Wavefield Inversion）

我们认为无论如何选择控制方程及如何描述相应的定解问题都是不可能完全解释实测地震数据的。既然如此，我们从实测数据中筛选出能被控制方程较准确预测的波场。这部分波场称为特征波场（CWF=Characteristic Wave Field）。基于特征波场发展出一套特征波反演理论，特征地震波反演成像问题的凸性有明显的改善。为简洁计，下面的讨论没有考虑正则化问题。

基于特征波场（CWF）的特征波反演成像方法技术的核心在于特征波场数据的提取。实质上，特征波场也对应模型参数的特征成分。

压缩感知框架下的特征波场提取问题可以提为

$$\min\left\|\mathbf{d}_C^{obs}\right\|_p \quad s.t.\left\|\mathbf{d}_C^{obs}-\mathbf{d}^{obs}\right\|\le\delta \tag{10}$$

其中，$\mathbf{d}_C^{obs}$ 为提出的特征波场，$\mathbf{d}^{obs}$ 为实测波场。一般地，从高维地震数据体中提出的 $\mathbf{d}_C^{obs}$ 表现为线性相位变化的局部平面波场。因此，我们给出的特征波场的定义为：时空局部的、单震相的、带方向的带限波场。

在提出的特征波场的基础上，特征波场反演问题提为，

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}}\frac{1}{2}\left\|R\mathbf{u}^C-\mathbf{d}_C^{obs}\right\|_2^2 \quad s.t.A^C(\mathbf{m}^C)\mathbf{u}^C=\mathbf{q} \tag{11}$$

其中，R 是采样算子，A^C 描述特征波传播的波场正传算子，$\mathbf{u}^C$ 为正演计算的特征波场，$\mathbf{m}^C$ 为与特征波场对应的特征参数，$\mathbf{q}$ 代表震源函数。它们共同满足控制方程 $A^C(\mathbf{m}^C)\mathbf{u}^C=\mathbf{q}$。由于我们仅仅考虑特征波场，比较完整地进行特征波场的预测就是可能的。因此，（11）式可以重写为，

$$\min_{\mathbf{m}\in\mathbb{M}}\frac{1}{2}\left\|R\left(A^C(\mathbf{m}^C)^{-1}\mathbf{q}\right)-\mathbf{d}_C^{obs}\right\|_2^2 \tag{12}$$

Van Leeuwen，Herrmann 和 Peters（2014）提出如下的波动方程约束的优化反演问题（波场重构反演方法），

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}}\lambda^2\left\|A(\mathbf{m})\mathbf{u}-\mathbf{q}\right\|_2^2+\frac{1}{2}\left\|P\mathbf{u}-\mathbf{d}^{obs}\right\|_2^2 \tag{13}$$

上式也可以写为，

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}}\left\|A(\mathbf{m})P^{-1}\mathbf{d}^{obs}-\mathbf{q}\right\|_2^2 \tag{14}$$

可见波场重构反演方法的思想也是提出实测波场中满足控制方程的部分参与反演，即认为 $\mathbf{u}=P^{-1}\mathbf{d}^{obs}$。并利用（13）式定义了如下的增广状态变量，

$$\mathbf{u}_\lambda=\left(\lambda^2A(\mathbf{m})^*A(\mathbf{m})+P^*P\right)^{-1}\left(\lambda^2A(\mathbf{m})^*\mathbf{q}+P^*\mathbf{d}^{obs}\right) \tag{15}$$

但是，作者们没有提出计算可行的重构波场（或计算增广状态变量）的方法。

我们坚持用特征波场进行地震波反演的理念来自于如下线性问题求解方式的激发，

$$A\mathbf{m}=\mathbf{d}^{obs}\Rightarrow U\Sigma V^T\mathbf{m}=\mathbf{d}^{obs}\Rightarrow\mathbf{m}_C=\Sigma^{-1}\mathbf{d}_C^{obs} \tag{16}$$

其中，$\mathbf{m}_C=V^T\mathbf{m}$ 可以视为模型参数的特征表达；$\mathbf{d}_C^{obs}=U^T\mathbf{d}^{obs}$ 可以视为实测数据的特征表达。假若 Σ 是非零元素形成的对角阵，则 $\mathbf{m}_C$ 和 $\mathbf{d}_C^{obs}$ 之间就建立起了一对一的投影关系。据此可以认为，特征波场和特征模型参数之间的关系要比一般性的实测波场和模型参数之间的关系更密切，对应的正演问题的预测可靠性要提高很多，对应的反问题的凸性也会高很多。

特征波反演的核心问题之一在于解问题（10）。相比于波场重构反演方法（Van Leeuwen，Herrmann 和 Peters，2014），特征波场反演的物理意义十分明确，可操作性很强，据此可以构建出有效的反演技术路线。尤其，（10）式的求解一般在压缩感知框架下进行，可以把先进的信号处理理论用于特征波提取及反演成像中。更重要的是，特征波的提取不仅仅是现代信号框架（或压缩感知框架）下的操作，而是按照波动理论的物理意义进行操作（譬如仅仅提出透射波场和一次反射波场），尽可能建立起特征波场与特征参数之间的联系，使得提出的特征波反演成像问题更凸。

特征波反演成像的基本过程可以描述如下。首先，我们把一般的地震数据表达成特征数据，提取特征数据的过程可用下式来抽象地表示：

$$\mathbf{d}_C^{obs}(\mathbf{x}_s,p_s,\mathbf{x}_r,p_r,t)=P(\mathbf{d}^{obs}(\mathbf{x}_s,\mathbf{x}_r,t)) \tag{17}$$

其中，$\mathbf{x}_s$ 代表了炮点的位置，$\mathbf{x}_r$ 代表了检波点的位置，p_s 代表了炮点端的射线参数，p_r 代表检波点端的射线参数，t 是记录到的旅行时场。$\mathbf{d}_C^{obs}\left(\mathbf{x}_s,p_s,\mathbf{x}_r,p_r,t\right)$ 表示特征地震数据，$\mathbf{d}^{obs}(\mathbf{x}_s,\mathbf{x}_r,t)$ 代表常规地震数据。$P(\cdot)$ 代表某种变换算子，变换算子的选择可以非常灵活，不同的变换算子代表我们希望用什么样的波场成分进行反演成像。

在特征波场的概念下，一般的正问题可写成如下积分形式：

$$\mathbf{u}^C(\mathbf{x}_s,p_s,\mathbf{x}_r,p_r,t)=\int K(\mathbf{v}_b(\mathbf{x})+\delta\mathbf{v}_b(\mathbf{x});\mathbf{x}_s,\mathbf{x}_r,\mathbf{x},t)\mathbf{m}^C(\mathbf{x})d\mathbf{x} \tag{18}$$

其中，$\mathbf{u}^C(\mathbf{x}_s,p_s,\mathbf{x}_r,p_r,t)$ 是特征波场数据，K 代表了灵敏度核函数，$\mathbf{m}^C(\mathbf{x})$ 是特征模型参数（即反射界面的反射系数）。当 K 的性质比较好（譬如其逆算子存在）的情况下，特征波场数据与特征模型参数之间建立起了一对一的映射关系。此时的反演成像问题凸性一定是好的。

在上述正问题的基础上，可以建立如下的地震波反演成像问题：

$$\mathcal{J}\left(\delta\mathbf{v}_b(\mathbf{x}),\mathbf{m}^C(\mathbf{x})\right)=\left\|\mathbf{u}^C\left(\mathbf{v}_b(\mathbf{x})+\delta\mathbf{v}_b(\mathbf{x}),\mathbf{m}^C(\mathbf{x})\right)-\mathbf{d}_C^{obs}\right\|^2 \tag{19}$$

其中，$\mathbf{u}^C\left(\mathbf{v}_b(\mathbf{x})+\delta\mathbf{v}_b(\mathbf{x}),\mathbf{m}^C(\mathbf{x})\right)$ 为计算的特征波场，$v_b(\mathbf{x})$ 为背景速度，$\delta\mathbf{v}_b(\mathbf{x})$ 是背景速度的低波数扰动；$\mathbf{m}^C(\mathbf{x})$ 则为速度场的高波数扰动（反射系数），它代表了速度模型的几何特征（及反射界面的几何形态信息）。假定背景速度 $v_b(\mathbf{x})$ 是已知的，$\delta\mathbf{v}_b(\mathbf{x})$ 和 $\mathbf{m}^C(\mathbf{x})$ 是要反演的模型参数。事实上，背景速度 $v_b(\mathbf{x})$ 也要通过透射波层析反演来确定。

（18）和（19）式可以进行各种特征波现象的反演。关键是特征波场对应什么样的波现象，及对应波现象的正问题如何描述。譬如，若把 $\mathbf{d}_C^{obs}$ 和 $\mathbf{u}^C$ 界定为一次反射波场，可以进行利用一次反射波的反演。当然也可以定义为透射波的反演和多次波的反演。

一般地，常用（19）式进行一次反射波层析成像和最小二乘 PSDM，它们对应的梯度分别为 $\frac{\partial\mathcal{J}}{\partial\delta\mathbf{v}_b}$ 和 $\frac{\partial\mathcal{J}}{\partial\mathbf{m}^C}$。在数据域波动方程反射波层析反演中，关键的问题在于如何把来自于同一反射层的同震相的 $\mathbf{d}_C^{obs}$ 和 $\mathbf{u}^C$ 对应起来，保证不发生“串层”的情况，这是数据域反射波层析成像的关键所在。同样地，在 FWI（它不仅仅用了一次反射波和绕射波，还有多次反射和绕射波）中也存在相应地问题，这就是 Cycle-Skipping（周期跳跃）现象。在周期跳跃现象中，数据的匹配要求更严格，不能超过半个子波周期。这也是 FWI 中初始模型参数要非常接近真实模型参数的原因，也就是在这样的条件下，FWI 反演问题才比较凸，比较满足线性性。

为了能够避免上述数据匹配存在的困难，可以把波动方程的反射波层析退化到反射波旅行时层析（RT_Tomo=Reflection Travel-time Tomography）。此时的要求变成测出（自动测出）同一反射层的 $\mathbf{d}_C^{obs}$ 和 $\mathbf{u}^C$ 的旅行时差 Δt。这也是 RT_Tomo（反射波旅行时层析）的核心工作，剩下的就按照一般的反射层析框架工作即可。

成像域的反射波层析也可以基于特征波反演框架进行描述。事实上，在特征波场情形下，成像域的模型参数层析反演应该与数据域的模型参数层析反演等价。

如果进行旅行时层析，无论在数据域或是像域，基本的问题都是时差测量问题。数据域中，实测数据同相轴到达时是确定的，实测数据和合成数据对应同相轴的时差测量结果可以认为是绝对的；像域中，反射界面真深度是未知的，基于成像道集的剩余深度差（RMO）测量是相对的。这也是为什么需要在时间域进行层析速度反演的重要原因。实测数据和合成数据对应同相轴的时差测量是当前层析速度反演的一个核心问题，解决了此问题，FWI 中的周期跳问题就不存在了。AWI 方法（Warner 和 Guasch，2014a，2014b）的本质也是在解决这个问题。

7 特征波场提取的思想与方法

从实测的地震数据中，根据反演成像的要求，提取出特征波场，可以抽象为如下的变换，

$$\mathbf{d}_C^{obs} = P\left(\mathbf{d}^{obs}\right) \tag{20}$$

其中，变换算子 $P(\cdot)$ 可以根据反演成像方法的需要选各种形式，关键是构建特定的反演方法时想利用什么样的波现象、想利用该波现象的什么特征。可以提出各种各样的变换算子。譬如，对波场的 Laplace 变换（Shin 和 Cha，2009）、对波场的 Hilbert 变换（Bozdag 等，2011）、对波场的对数变换（Bedner 等，2007；Sukjoon 等，2007）等等。这些变换算子都是侧重于提取反射子波的特征。

但是，我们提出的特征波场，首先考虑的是区别利用不同的波现象，然后再进一步考虑利用特定波现象的更进一步的特征。我们认为，特定波现象的方向传播特征是很基本的波场特征。因此，特征波场变换算子至少包含三重含义：**区分波现象（区分波现象）；提出时空局部的方向波场；更进一步的子波波形特征的提取**。

区分波现象主要考虑分开纵横波、透射波和反射波、绕射波和反射波，还包括分开一次波和多次波；提出时空局部的方向波场主要有基于基函数族的和基于数据驱动的两种线性局部平面波提取方法；子波特征的提取主要是提取带限子波的振幅和到达时。

这三个方面目前都存在问题，尤其是在实测地震数据有较强噪音的情况下。带限子波的振幅和到达时的估计，因为与相位搅在一起，从概念上都还存在问题。因此，带限子波波形特征的表达和提取对反演成像而言是更为重要的研究内容，否则我们可能说不清楚振幅和到达时的含义，全波形反演的真正内涵也就讲不明白。

我们主要关注在压缩感知框架下如何提出时空局部的方向波场。这个问题被提为如下变换域的反问题，

$$\hat{\mathbf{d}}_C = \underset{\mathbf{d}_C}{\operatorname{argmin}} \left\|\mathbf{d}_C\right\|_1 \quad S.T. \left\|A\mathbf{d}_C - \mathbf{d}^{obs}\right\|_2 \le \varepsilon \tag{21a}$$

其中，A 是变换矩阵，列向量构成基函数族。数据的特征表达原则上也可以用如下数据驱动的方式，

$$\min_{\mathbf{X}} \left\|\mathbf{X}\right\|_* \quad s.t. \left\|\mathcal{A}\left(\mathbf{X}\right) - \mathbf{d}^{obs}\right\|_2 \le \varepsilon \tag{21b}$$

其中，$\mathcal{A} \coloneqq R\phi^H$ 是数据变换算子，代表把 SVD 分解后的奇异值恢复成与实测数据采样一致的数据。R 是采样算子，ϕ^H 是把 SVD 分解后的奇异值恢复到原数据域的算子。数据驱动的地震数据特征表达是目前地震数据预处理的一个研究热点（Kreimer，Stanton 和 Sacchi，2013；Oropeza 和 Sacchi，2011；Recht，2011）。相比于基于基函数的波场特征表达，数据驱动的方法物理意义要更难把握，毕竟特征向量的物理意义不是很明晰。

图 1 展示了抽取特征波场的基本思想和做法。发展这样的 CWF 产生方法的另一个目的是解决层析反演中实测数据同相轴和合成数据同相轴的对应的时差测量问题，试图通过这样的方式避开周期跳现象。

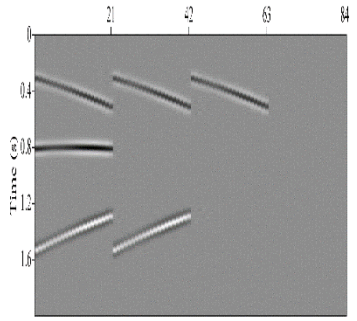


图 1a 匹配追踪方法抽取特征波场方法的示意图。从左到右，图 1 为包含三个同相轴的局部波场；图 2 为匹配追踪方法抽取平同相轴后的结果；图 3 为再抽取出下面一个斜同相轴后的结果；图 4 为三个同相轴全部抽取出后的结果。

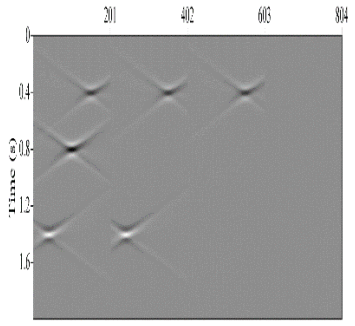


图 1b 与左图对应的 tau-p 谱。

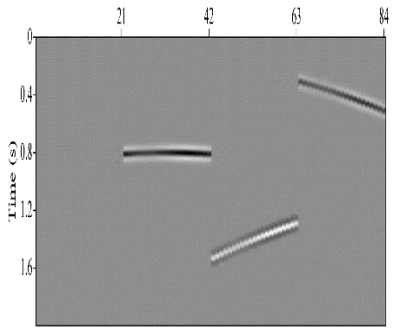


图 1c 数据恢复情况。从左到右，图 1 展示了原数据（图 1a 最左图展示）与反变换数据的差；图 2 展示了恢复出的水平轴；图 3 展示了恢复出的下面的斜轴；图 4 展示了恢复出的上面一个斜轴。比较完整地恢复出来原数据，实现了抽取和恢复的全过程。

8 数据域的特征波反演方法

数据域地震波反演的经典方法自然是 FWI 方法。经典 FWI 的目标是估计“全或宽”波数带的参数场。但是，到目前的实践表明，即使用经典 FWI 结合低频长偏移距数据估计低波数（背景）的速度也不是任何情形都能收敛的。这正是经典 FWI 衍生出很多变种的原因，但这也造成了 FWI 的滥觞。

此处我们要讨论的是基于特征波场（CWF）的特征波场反演（CWI）。CWI 包括四个关键步骤：用压缩感知方法提取所要的特征波场（目前主要是初至波、早至波、一次反射波；提出特征波的目的主要是用来进行层析成像估计背景参数的变化）；基于到达时（相位）的透射波波动理论层析反演；基于到达时（相位）的一次反射波波动理论层析反演；主要针对方位角度反射系数的最小二乘叠前深度偏移成像。

特征波场必须与要反演的参数之间有更紧密的关系才能使得构建的反演问题的凸性更好。我们提出如下的模型参数分解：

$$\mathbf{m}(\mathbf{x})=\mathbf{m}_0(\mathbf{x})+\Delta\mathbf{m}(\mathbf{x})+R(\mathbf{x},\gamma,\phi) \tag{22}$$

基于上述模型分解，波场也可以分为

$$\begin{aligned} G(x_r\,|\,x_s) &= G_0-\omega^2 G_0\delta\mathbf{m}G_0+\omega^4 G_0\delta\mathbf{m}G_0\delta\mathbf{m}G_0+\ldots \\ &= G_0-\omega^2\int G_0(x_r\,|\,\xi)G_0(\xi\,|\,x_s)\delta\mathbf{m}(\xi)d\xi \\ &+\omega^4\iint G_0(x_r\,|\,\eta)\delta\mathbf{m}(\eta)G_0(\eta\,|\,\xi)\delta\mathbf{m}(\xi)G_0(\xi\,|\,x_s)d\xi d\eta+\ldots \end{aligned} \tag{23}$$

其中，右端项中三项的具体含义分别为：第一项对应透射波；第二项对应一次反射波；第三项对应二次散射波。更高阶的散射波没有列出。地下存在散射点引起的场的扰动可以表示为

$$\begin{aligned}\delta G(x_r\,|\,x_s) &= G - G_0 \\ &= -\omega^2 \int G_0(x_r\,|\,\xi) G_0(\xi\,|\,x_s)\,\delta\mathbf{m}(\xi) d\xi \\ &\quad +\omega^4 \int G_0(x_r\,|\,\eta)\,\delta\mathbf{m}(\eta) G_0(\eta\,|\,\xi)\,\delta\mathbf{m}(\xi) G_0(\xi\,|\,x_s) d\xi d\eta + \ldots \\ &= \delta G_1(x_r\,|\,x_s) + \delta G_2(x_r\,|\,x_s) + \ldots\end{aligned} \tag{24}$$

其中， $\delta G_1(x_r|x_s)$ 代表一阶扰动场；$\delta G_2(x_r|x_s)$ 代表二阶扰动场。更高阶的扰动场被忽略。当若散射假设成立时，更高阶的扰动场是很弱的，忽略掉是合理的。

对应零阶扰动波场（透射波场）、一阶扰动波场和二阶扰动波场的波场扰动相对于介质扰动的梯度分别定义为：

零阶扰动波场梯度项：

$$\frac{\partial G}{\partial \mathbf{m}} = \lim_{\mathbf{m}\to\mathbf{m}_0} \frac{G(\mathbf{m}) - G(\mathbf{m}_0)}{\mathbf{m} - \mathbf{m}_0} \tag{25a}$$

一阶扰动波场梯度项：

$$\begin{aligned}\frac{\partial \delta G_1(x_r\,|\,x_s)}{\partial \delta\mathbf{m}(x)} &= \frac{\partial}{\partial \delta\mathbf{m}(x)}\left(-\omega^2 \int G_0(x_r\,|\,\xi) G_0(\xi\,|\,x_s)\,\delta\mathbf{m}(\xi) d\xi\right) \\ &= -\omega^2 G_0(x_r\,|\,x) G_0(x\,|\,x_s)\end{aligned} \tag{25b}$$

二阶扰动波场梯度项：

$$\begin{aligned}\frac{\partial \delta G_2(x_r\,|\,x_s)}{\partial \delta\mathbf{m}(x)} &= \frac{\partial}{\partial \delta\mathbf{m}(x)}\left(\omega^4 \int\int G_0(x_r\,|\,\eta)\,\delta\mathbf{m}(\eta) G_0(\eta\,|\,\xi)\,\delta\mathbf{m}(\xi) G_0(\xi\,|\,x_s) d\xi d\eta\right) \\ &= \frac{\partial}{\partial \delta\mathbf{m}(x)}\left(-\omega^2 \int G_0(\xi\,|\,x_s)\,\delta\mathbf{m}(\xi)\left(-\omega^2 \int G_0(x_r\,|\,\eta)\,\delta\mathbf{m}(\eta) G_0(\eta\,|\,\xi) d\eta\right) d\xi\right) \\ &= \frac{\partial}{\partial \delta\mathbf{m}(x)}\left(-\omega^2 \int G_0(\xi\,|\,x_s)\,\delta\mathbf{m}(\xi)\,\delta G_1(\xi\,|\,x_r) d\xi\right) \\ &= -\omega^2 G_0(x\,|\,x_s)\,\delta G_1(x\,|\,x_r) - \omega^2\left(\int G_0(\xi\,|\,x_s)\,\delta\mathbf{m}(\xi) \frac{\partial \delta G_1(\xi\,|\,x_r)}{\partial \delta\mathbf{m}(x)} d\xi\right) \\ &= -\omega^2 G_0(x\,|\,x_s)\,\delta G_1(x\,|\,x_r) + \omega^4 \int G_0(\xi\,|\,x_s)\,\delta\mathbf{m}(\xi) G_0(\xi\,|\,x) G_0(x\,|\,x_r) d\xi \\ &= -\omega^2 G_0(x\,|\,x_s)\,\delta G_1(x\,|\,x_r) + G_0(x\,|\,x_r)\,\omega^4 \int G_0(\xi\,|\,x_s)\,\delta\mathbf{m}(\xi) G_0(\xi\,|\,x) d\xi \\ &= -\omega^2 G_0(x\,|\,x_s)\,\delta G_1(x\,|\,x_r) - \omega^2 G_0(x\,|\,x_r)\,\delta G_1(x\,|\,x_s)\end{aligned} \tag{25c}$$

它们分别定义了联系波场扰动和介质参数扰动的积分核（Kernel）。波动理论下，三个梯度项的示意图分别为图 2a、图 2b 和图 2c。

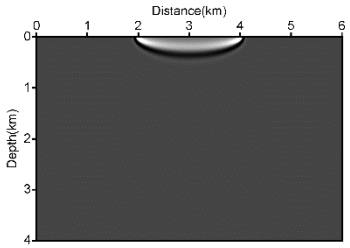


图 2a 透射波扰动响应

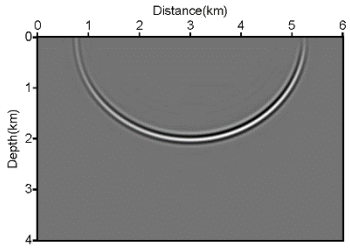


图 2b 一阶扰动场响应

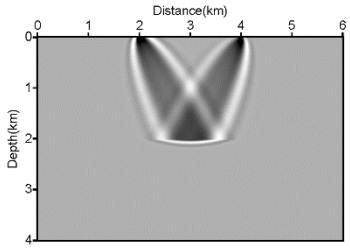


图 2c 二阶扰动场响应

更进一步地，对于一阶扰动和二阶扰动而言，必须充分利用地下介质的既有特征。我们把油气地震勘探面对的地下介质的特征表述为：在空间上广泛分布的层状沉积层，加上火山活动，构造运动等引起的大、小尺度的速度异常体（或波阻抗变化的异常体）。即地下介质的分层性是勘探地震所面对的介质的重要特征，产生绕射和散射的介质特征是次要的。

因此，利用二阶扰动波场的层析成像和利用一阶扰动波场的偏移成像都利用了这样的特征。对层析成像而言，可以利用反偏移计算图 2b 所示的二阶扰动响应；对偏移成像而言，可以针对方位角度反射系数进行成像。

实质上，FWI 是一种认为地下介质为背景+散射体（点）模式的反演成像方法。这样的成像方法应该对应封闭面上进行全方位激发接收的观测方式，而不是如勘探地震仅在地表面进行的观测方式。我们认为这也是 FWI 方法或逆散射成像方法不能很好地用于油气地震勘探领域的重要原因。

基于上述分析，我们认为把 FWI 分为如 CWI 所阐述的三个线性反演步骤在数学物理上是合理的。更进一步地，我们认为背景速度和各向异性参数与一阶扰动和二阶扰动场中的同相轴的到达时（相位）关系密切，这些参数的估计不仅要利用到达时（相位），还要尽可能剥离振幅对达到时（相位）测量的影响。事实上，层析反演的理论框架基本是定型的，实际数据到达时（相位）、合成数据到达时（相位）、以及对应同相轴上二者之间的差的测量成为了层析成像方法与技术的核心问题。我们提出压缩感知框架下的 CWF 提取，其中一个重要的目的就是发展出有效的对应同相轴上实际数据到达时（相位）和合成数据到达时（相位）之间的差的测量。

这是我们坚持的把 FWI 推进到实用化的具体路线。鉴于篇幅，具体数值结果不在此展示。具体地，参考冯波博士论文。

9 成像域的特征波反演

地震波反演也可以在成像域中进行。基本思想是利用不同角度（或偏移距）的数据产生的像理论上应出现在相同的空间位置上，即所谓的正确的成像速度产生正确的像，错误的成像速度产生错误的像。也可以说成像速度的扰动产生了像的扰动。成像速度的扰动与像的扰动之间存在线性关系。建立这样的线性关系就可以进行成像速度的反演。正确的像是指对同一地下像点而言，不同角度或偏移距的照明数据产生的像出现在该地下成像点上，表现为成像道集拉平或聚焦在一起。

但是，我们认为，这与 Bayesian 框架下的数据域的参数估计理论是有本质区别的。区别之处在于：（1）在成像域仅仅能估计光滑的成像速度，因为高波数的速度变化对产生拉平或聚焦的道集并无贡献。而道集的拉平或聚焦正是建立像域反演目标泛函的基础。（2）像域的成像速度估计缺乏概率统计意义下的解释。

特征数据体产生的像可以用下式表示，

$$\mathbf{I}\left(\mathbf{x}_I, z_I \middle| \gamma, \varphi\right) = \sum_S \sum_R \int_t u_S^C\left(\mathbf{x}, t; \mathbf{x}_s, p_s\right) u_R^C\left(\mathbf{x}, t; \mathbf{x}_r, p_r\right) dt \tag{26}$$

其中，$\mathbf{x}_I=\left(x_I,y_I\right)$是像点的平面坐标；$z_I$是像的深度。$\gamma$是张角，$\varphi$是方位角。$u_S^C\left(\mathbf{x},t;\mathbf{x}_s,p_s\right)$是外推到成像点处的震源端特征波场；$u_R^C\left(\mathbf{x},t;\mathbf{x}_r,p_r\right)$是外推到成像点处的检波端特征波场。

构建像域的反问题，有两条途径。一种是通过检测成像道集的拉平或聚焦情况，建立目标泛函。通过对目标泛函的求导，建立梯度导引类的优化反演方法，估计成像速度（Symes 和 Versteeg，1993；Shen 等，2005；Shen 和 Symes，2008；刘玉金，2014；Liu 等，2014），。另一种是直接建立像的扰动与成像速度扰动之间的线性关系（Baker，2015a，2015b），然后基于此线性关系进行成像速度的层析反演。

首先讨论第一类方法。Shen 等（2005）给出了两类所谓的差异相似性优化（DSO = Differential Semblance Optimization）目标函数，分别描述地下偏移距（h）道集在零偏移距处的聚焦特性和角度（θ）道集的拉平特性。目标函数分别写为：

$$\min_{\mathbf{v}} J_{\mathrm{h}}(\mathbf{m})=\frac{1}{2}\left\|P(h)\mathbf{I}\right\|_2^2=\frac{1}{2}\sum_x\sum_z\sum_h\left(h\mathbf{I}(x,z,h)\right)^2 \tag{27}$$

$$\min_{\mathbf{v}} J_{\theta}(\mathbf{m})=\frac{1}{2}\left\|P(\theta)\mathbf{I}\right\|_2^2=\frac{1}{2}\sum_x\sum_z\sum_\theta\left(\frac{\partial}{\partial\theta}\mathbf{I}(x,z,\theta)\right)^2 \tag{28}$$

其中，$P(h)=h$为零阶差分算子；$P(\theta)=\dfrac{\partial}{\partial\theta}$为一阶差分算子。

对于公式（27），当速度正确时，地下偏移距道集聚焦于$h=0$处，目标函数取得极小值。该目标函数的凸性较好，且不受周期跳跃问题制约，但对中、小尺度的速度扰动不敏感，速度估计的精度低。Shan 和 Wang（2013）指出，（27）式的收敛受到成像道集振幅的影响。当地下照明能量较弱时，需要引入照明补偿等（Yang 等，2012），否则强反射层的像会主导泛函能量，影响泛函收敛。此外，（27）式的梯度中含有一些近垂直的条带状假象，影响反演的收敛。Fei 和 Williamson 等（2010）通过修改成像差（用关于h的一阶差分算子作用于地下偏移距域的像）的定义，改善泛函梯度的性质并加速反演收敛。

根据成像原理，当速度正确时，角度道集的像为拉平的，相邻角度像的差异较小。由此可以得出公式（28）描述的角度道集 DSO 目标函数。由于（28）式只需要对相邻角度的像进行差分运算，所以弱化了周期跳跃问题的制约，降低了对初始模型的精度要求。然而，差分运算会放大角度道集中高频变化的影响，对角度道集的质量也有一定的要求。

Soubaras 和 Gratacos（2007）提出了角度道集叠加能量最大化方法，目标函数可以写为：

$$\max_{\mathbf{v}} J_{\mathrm{SPM}}(\mathbf{m})=\frac{1}{2}\sum_x\sum_z\left(\sum_\theta\mathbf{I}(x,z,\theta)\right)^2 \tag{29}$$

其中，$\mathbf{I}(x,z,\theta)$为角度域成像道集。该方法的精度高于上述 DSO 类的反演方法。然而，当初始速度与真速度的差异较大时，角度道集的 RMO 较大，甚至大于半个子波长度。此时，若按照公式（29）对所有角度叠加会遇到与 FWI 类似的问题（周期跳跃）。因此，该方法对初始模型的要求较高。

值得提及的是，地下偏移距成像道集是不区分反射和散射（绕射）的，都可以产生对应当前偏移速度的聚焦道集；角度成像道集是针对反射波的，只有反射界面才有对应的角度成像道集。因此，（27）式定义的泛函和（28）及（29）式定义的泛函应该适应于不同的介质参数分布情形。我们认为，反射界面的存在是勘探地震的特征，应该充分利用这样的特征。

在上述目标泛函特性分析的基础上，可以进入梯度导引的反演过程中。但是，我们认为利用地下偏移距聚焦成像道集，基于聚焦好坏建立目标反演进行成像速度的反演理论上是存在问题的。不能用聚焦好坏（能量聚焦度）作为目标函数进行优化反演更新成像速度，一是目标泛函的梯度不能指示速度修正量的正或负，这与数据逼近泛函是不同的；二是影响振幅的因素太多，聚焦能量最小，并不一定预示背景速度的修正是正

确的。角度道集叠加能量最大化目标泛函同样存在不合适的利用振幅的问题。利用角度道集之间的像差建立目标泛函进行速度反演是合适的逻辑，但是其中的 AVA 关系的影响，像的振幅畸变的影响会导致这样的方法不稳定。总之，我们不认为在像域进行上述基于像的振幅构建泛函，进行背景速度反演的方式是正确的选择。DSO 之类的像域背景速度估计概念二十多年来始终无法用于勘探地震实际应用中很能说明问题。

然后，讨论第二类方法。首先建立像的扰动与成像速度扰动之间的关系。由（26）式可见，像的扰动必然是其中波场的扰动引起，而场的扰动又是由于成像速度的扰动（即成像速度与成像真速度之间的差）引起。因此，引入像的扰动为，

$$\delta\mathbf{I}\left(\mathbf{x}_I,z_I\left|\gamma,\varphi\right.\right)=\sum_S\sum_R\int_t\left(\begin{array}{l}\delta u_S^C\left(\mathbf{x},t;\mathbf{x}_s,p_s\right)u_R^C\left(\mathbf{x},t;\mathbf{x}_r,p_r\right)\\+u_S^C\left(\mathbf{x},t;\mathbf{x}_s,p_s\right)\delta u_R^C\left(\mathbf{x},t;\mathbf{x}_r,p_r\right)\end{array}\right)dt \tag{30}$$

很显然，场的扰动与速度扰动之间的关系是由散射场的 Born 计算公式决定的，即

$$\delta u_S^C\left(\mathbf{x},t;\mathbf{x}_s,p_s\right)=\int G_0\left(\mathbf{y},\mathrm{t}-\tau;\mathbf{x}\right)\frac{2\delta\mathbf{v}_I\left(\mathbf{y}\right)}{\mathbf{v}_I^3\left(\mathbf{y}\right)}\frac{\partial^2}{\partial\tau^2}S\left(\mathbf{y},\tau;\mathbf{x}_s\right)d\tau d\mathbf{y} \tag{31a}$$

$$\delta u_R^C\left(\mathbf{x},t;\mathbf{x}_R,p_R\right)=\int G_0\left(\mathbf{y},\mathrm{t}-\tau;\mathbf{x}\right)\frac{2\delta\mathbf{v}_I\left(\mathbf{y}\right)}{\mathbf{v}_I^3\left(\mathbf{y}\right)}\frac{\partial^2}{\partial\tau^2}u\left(\mathbf{y},\tau;\mathbf{x}_s\right)d\tau d\mathbf{y} \tag{31b}$$

把（31）代入（30）可以导出像的扰动和成像速度扰动之间的线性关系

$$A\Delta\mathbf{m}=\delta\mathbf{I} \tag{32}$$

共轭梯度法求解上式可以得到更新的成像速度。公式（30）-（32）概念上是没有问题的，但是具体应用时，类似于上面的分析，还是要区分用于反射情形，或是用于散射（包含了反射的）情形。抽象地讲，这些公式可以用于不区分散射和反射的一般情形。但是，实际上，（32）式右端项表示的像的扰动是很难有效获取的，因为真像是未知的！

而在针对反射情形的角度成像道集中，像的扰动代表道集没有拉平，本次偏移速度得到的像与真像之间有差异，最主要的差异是成像深度差（RMO）！因此，还是要用到特征波场（局部平面波场）与特征像（角度成像道集）的概念，建立起它们之间的联系，进行基于 RMO 的、波动理论的角度成像道集层析成像是更为合理的选择。其中，有效的 RMO 的检测是最为核心的问题。在此基础上，有必要和可能的话，再基于公式（30）-（32）进行层析成像。这是我们一直坚持的像域中速度估计与建模的基本理念。鉴于篇幅，具体数值结果不在此展示。

10 结论与讨论

FWI 是勘探地震中地震波反演成像的核心理念。它把地下介质视为背景+散射（点）的模式，并试图对实测炮集中所有波现象进行基于振幅的波形逼近基础上的反演，其目标是估计“全或宽波数带”的速度场。但勘探地震中的实践表明，仅有地表观测的地震数据，不对勘探地震介质的特征和波场特征加以利用，FWI 方法即使仅用透射波进行基于低频长偏移距数据的背景参数的估计也是不很稳定的（视数据情况达到收敛或不收敛）。

我们认为，一定假设下的地震波正演方法可以预测实测地震波场是地震波反演最基本的要求。既然无论如何选择地震波正演模拟方法也无法完全预测实测数据中的波场，转而追求尽可能好地模拟其中的部分特征波场，进行基于特征波场的反演成像是合理的逻辑。特征表达参数模型和特征表达数据体，得到二者之间存在更密切联系的关系可以把强非线性的经典 FWI 问题转化为一个较凸的反问题进行求解。

基于上述理念提出特征波反演成像（CWI）方法，其中心工作在于特征波场的提取。特征波场提取包括三方面的含义：特征波型的提取（波型分解）；压缩感知框架下特征波场的提取（时空局部的方向波场的分解）；反射子波特征的提取（子波分解）。我们重点强调了压缩感知框架下时空局部的方向波场的分解以及基

于此波场的数据域和成像域的反演思想。指出特征波场与反演参数之间的关系更紧密，基于特征波场的反问题凸性更好。另外，特征波成像中梯度的预条件要相对简单，因为特征波对应的梯度干扰大大减弱。

CWI 包括四个关键步骤：用压缩感知方法提取所要的特征波场；基于到达时（相位）的透射波波动理论层析反演；基于到达时（相位）的一次反射波波动理论层析反演；主要针对方位角度反射系数的最小二乘叠前深度偏移成像。

到目前为止，参数的低波数成分估计方法基本是确定的，就是透射波层析和反射波层析，其中的反演方法框架已很成熟了（具体计算中还会有一些问题），但是对应的到达时（相位）差的测量是核心问题。实质上，FWI 中周期跳问题，就是没有解决对应同相轴的时差（相位）测量问题，否者的话，这个问题是不存在的。时差（相位）测量一定会受到振幅的影响，这也是物理上不可避免的，我们能做的只是减少这种影响。我们把时差测量问题作为 CWI 中的核心问题来对待。

关于参数的高波数成分估计，我们认为目前还没有什么公认的方法技术路线。FWI 方法假设地下介质是背景+散射（体），而且在一个统一的迭代过程中，由低波数成分到高波数成分全部估计出来。实践证明 FWI 方法中高波数分量估计的收敛性是十分没有保障的。即便是经典的最小二乘 RTM（LS_RTM）方法也是这样假设介质分布，在 Born 近似下进行数据逼近的反演，收敛性也没有什么保障。LS_RTM 经过这么多年的研究始终不能大规模处理实际数据也说明了这种技术的局限性。在射线+Born 近似下的角度反射系数的最小二乘反演估计可以算是利用了勘探地震地下介质特征的高波数分量估计方法，但是在小尺度散射（绕射）体分布比较多的探区，这样的成像方法显然是不合适的。兼顾反射波和散（绕）射波成像的高波数成分估计方法目前是有必要发展的，我们正试图沿这样的路线发展出比较实用的高波数分量估计方法。

参考文献

[1] Baker, P. M., Gerritsen, S., Cao, Q., Duveneck, E., 2015a, 3D RTM-based wave-path tomography: theory and applications to synthetic and field data, 85th Annual International Meeting, SEG, Expanded Abstract, 5259-5264

[2] Baker, P. M., Gerritsen, S., Cao, Q., Duveneck, E., 2015b, 3D RTM-based wave-path tomography tested at a realistic scale: 77th Conference & Exhibition – Workshops, EAGE, Extended Abstracts, WS02-C02.

[3] Bednar, J. B., C. Shin and S. Pyun, Comparison of waveform inversion, part 2: phase approach. Geophysical Prospecting, 2007. 55(4): 465-475.

[4] Bozdağ, E., J. Trampert and J. Tromp, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophys. J. Int., 2011. 185(2): 845-870.

[5] Bunks C, Saleck F M, Zaleski S, et al. Multiscale seismic waveform inversion[J]. Geophysics, 1995, 60(5): 1457-1473.

[6] Cara M, Lévêque J J. Waveform inversion using secondary observables[J]. Geophysical Research Letters, 1987, 14(10): 1046-1049.

[7] de Hoop, M.V., van der, H., Robert, D. & Shen, P., 2006. Wave-equation reflection tomography: annihilators and sensitivity kernels, Geophys. J. Int., 167(3), 1332–1352.

[8] Fei, W. and P. Williamson, On the gradient artifacts in migration velocity analysis based on differential semblance optimization. 80th Annual International Meeting, SEG, Expanded Abstracts, 2010: 4071-4076.

[9] Fichtner A, Kennett B L N, Igel H, et al. Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain[J]. Geophysical Journal International, 2008, 175(2): 665-685.

[10] Gee L S, Jordan T H. Generalized seismological data functionals[J]. Geophysical Journal International, 1992, 111(2): 363-390.

[11] Kreimer, N., Stanton, A., and Sacchi, M. D., 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, 78, No.6, V273-V284.

[12] Liu, Y., W. W. Symes and Z. Li, Extended Reflection Waveform Inversion via Differential Semblance Optimization. 84th Annual International Meeting, SEG, Expanded Abstracts, 2014: 1149.

[13] Luo Y, Schuster G T. Wave-equation traveltime inversion[J]. Geophysics, 1991, 56(5): 645-653.

[14] Moghaddam, P.P. & Mulder, W.A., 2012. The diagonalator: inverse data space full waveform inversion, SEG Technical Program Expanded Abstracts.

[15] Oropeza, V., and Sacchi, M. D., 2011, Simultaneous seismic data denoising and reconstruction via multichannel analysis: Geophysics, 76, No.3, V25-V32.

[16] Pyun S, Shin C, Bednar J B. Comparison of waveform inversion, part 3: amplitude approach[J]. Geophysical

Prospecting, 2007, 55(4): 477-485.
[17] Recht, B., 2011, A simpler approach to matrix completion: Journal of Machine Learning Research, 12, 3413-3430
[18] Rockafellar R T. Lagrange multipliers and optimality[J]. SIAM review, 1993, 35(2): 183-238.
[19] Sirgue L, Pratt R G. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies[J]. Geophysics, 2004, 69(1): 231-248.
[20] Shan, G. and Y. Wang, RTM BasedWave Equation Migration Velocity Analysis. 83rd Annual International Meeting, SEG, Expanded Abstracts, 2013: 4726-4731.
[21] Shin, C. and D. Min, Waveform inversion using a logarithmic wavefield. Geophysics, 2006. 71(3): R31-R42.
[22] Shin C, Pyun S, Bednar J. Comparison of waveform inversion, part 1: conventional wavefield vs logarithmic wavefield[J]. Geophysical Prospecting, 2007, 55(4): 449-464.
[23] Shin, C, Cha Y H. Waveform inversion in the Laplace—Fourier domain[J]. Geophysical Journal International, 2009, 177(3): 1067-1079.
[24] Shen, P., Differential semblance velocity analysis via shot profile migration. 75th Annual International Meeting,SEG, Expanded Abstracts, 2005. 24(1): 2249-2253.
[25] Shen, and W. W. Symes, Automatic velocity analysis via shot profile migration. Geophysics, 2008. 73(5):49-59.
[26] Soubaras, R. and B. Gratacos, Velocity model building by semblance maximization of modulated-shot gathers. Geophysics, 2007. 72(5): U67-U73.
[27] Sukjoon P, C., Shin, and J. B. Bednar, 2007, Comparison of waveform inversion: Part 2: amplitude approach: Geophysical Prospecting, 55, no. 4, 477–485
[28] Symes, W. W. and R. Versteeg, Velocity Model Determination Using Differential Semblance Optimization.63rd Annual International Meeting, SEG, Expanded Abstracts, 1993: 696-699.
[29] Tarantola, A., Linearized inversion of seismic reflection data. Geophysical Prospecting, 1984. 32: 998-1015.
[30] Tarantola A. Inverse problem theory and methods for model parameter estimation[M]. siam, 2005.
[31] Van Leeuwen, T., and Herrmann, F. J., 2013, Mitigating local minima in full-waveform inversion by expanding the search space[J]. Geophysical Journal International, 195(1): 661-667.
[32] Van Leeuwen, T., Herrmann, F. J., and Peters, B., 2014, A New Take on FWI-Wavefield Reconstruction Inversion. 76th EAGE Conference and Exhibition.
[33] Van Leeuwen T, Mulder W A. A correlation-based misfit criterion for wave-equation traveltime tomography[J]. Geophysical Journal International, 2010, 182(3): 1383-1394.
[34] Virieux J, Operto S. An overview of full-waveform inversion in exploration geophysics[J]. Geophysics, 2009, 74(6): WCC1-WCC26.
[35] Warner M, Guasch L. Adaptive Waveform Inversion-FWI Without Cycle Skipping-Theory[C]//76th EAGE Conference and Exhibition 2014a.
[36] Warner M, Guasch L. Adaptive waveform inversion: Theory[C]//2014 SEG Annual Meeting. Society of Exploration Geophysicists, 2014b.
[37] Weibull, W., Arntsen, B. & Nilsen, E., 2012. Initial velocity models for full waveform inversion, SEG Technical Program Expanded Abstracts 2012,pp. 1–4.
[38] Wu, R. S., C. Hu and B. Wang, Nonlinear sensitivity operator and inverse thin-slab propagator for tomographic waveform inversion. 84th Annual International Meeting, SEG, Expanded Abstracts, 2014: 928-933.
[39] Yang, T., J. Shragge and P. Sava, Illumination compensation for image-domain wavefield tomography. Geophysics, 2013. 78(5): U65-U76.
[40] 冯波，2015，数据域波动方程层析速度反演方法研究，同济大学博士学位论文
[41] 赫孝良，葛照强，最优化与最优控制，2015，西安交通大学出版社
[42] 刘玉金. 复杂介质条件下基于反演的地震成像方法研究:[博士学位论文]. 青岛: 中国石油大学(华东), 2014.
[43] 王华忠, 冯波, 刘少勇等. 2015. 叠前地震数据特征波场分解、偏移成像与层析反演. 地球物理学报, 58(6): 2024-2034.
[44] 王华忠，冯波，王雄文等. 2016. 压缩感知及其在勘探地震中的应用[J]. 石油物探. 55(4)：467-474

最小二乘叠前深度偏移成像理论与方法

王华忠[1]，胡江涛[1,2]，郭颂[1]

①同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

②油气藏地质及开发工程国家重点实验室，成都理工大学，四川成都，610059

【摘要】地震波成像的目的首先是定位反射界面或散射点出现的空间位置，描述地下地质体的几何结构；然后是估计地下岩石介质的物性参数，主要是速度和密度相关的参数；最终与岩石物理结合描述含油气储层。本文以地震波传播的散射场表达与逆散射成像为切入点，首先给出散射波的表达形式，讨论逆散射成像的本质。其中讨论了逆散射成像所需要的叠前地震观测数据及对应的观测方式。然后给出一般的线性化的最小二乘叠前深度偏移的基本理论框架，指出能给出其中 Green 函数计算的波传播算子就可以实现线性化的 LS_PSDM。接着讨论了当前广受关注的最小二乘逆时深度偏移（LS_RTM）方法，给出了迭代最小二乘偏移成像的具体实现方法和保反射界面结构的总变差正则化方法以及数值结果。最后分析了基于反射波的估计反射系数的成像和基于散射波的估计散射强度的成像方法的差异，指出针对勘探地震介质特征和波场特征的成像方法最具实用性的方法，逆散射成像应该在此基础上进行。基于散射波的成像方法能否满足储层描述的要求有待进一步研究工作的检验。

【关键词】散射场表达，逆散射成像，反射波成像，最小二乘逆时偏移成像，全波形反演

引言

油气勘探地震学的最终目标是储层。通向储层描述与刻画的途径有两条。第一条是常规的、基于反射波场的保幅或保真的角度反射系数的成像。基于保真的角度成像道集，利用 1D 波阻抗反演和/或 AVA 叠前反演得到弹性参数，与岩石物理和测井结合，进行储层识别和描述，这是目前常用的从地震成像走向储层刻画的方法技术体系（Wapenaar，1996）。该技术体系的核心包括三方面的关键因素：满足速度估计和保真成像要求的数据体、较准确的背景速度模型建立技术和保真角度反射系数成像技术、基于保真成像道集的储层参数估计及储层描述方法。当前，石油工业界十分关注的全波形反演（FWI）成像技术是基于散射和逆散射的反演成像技术，它的本质目标是估计宽（全）波数带的弹性参数场，基于此反演结果，直接进行储层描述。仅仅利用地表观测的有限带宽、有限孔径和无假频数据，FWI 的实用化并不成功，它仅仅在低频和长偏移距地表观测数据下，能得到背景速度的估计。弹性参数的跃变（高波数）成分的估计留给了最小二乘 RTM。但是，依赖散（反）射波振幅的 LS_RTM 也很难得到有效的收敛，给出满足储层描述要求的反演成像结果。

一些数学家看来，石油地震勘探问题很简单，就是根据实测的叠前地震数据，用波动方程作为实测数据预测器，若基于一个假设的参数模型求解波动方程得到的合成数据与实测的叠前地震数据的总误差最小，该假设模型就是所求的。事实上，这完全低估了勘探地震中弹性参数反演问题的复杂性（Clayton 和 Stolt，1981；Devaney，2012）。

从勘探地震学的初始，勘探地震学家走的是一条实际问题驱动的路线来解决成像问题。由 CMP 道集估计背景初始速度，进行叠后偏移成像；然后进入叠前偏移成像，由成像道集进行偏移速度分析，或进行层析成像建立速度模型。迭代进行叠前偏移成像和速度模型的更新是石油工业界发展出的实用化的地震波成像的路线。经检验是行之有效的。更细致的、能用于储层描述的参数估计是基于 1D 波阻抗反演和/或 AVA 角度道集反演（Shuey，1985）。

事实上，一直有很多人在探讨是否能从更数学化的路线来解决成像问题。这分成两条路线，一是 Tarentola 为代表的基于散射波的全波形拟合非线性反演路线（Tarantola，1984）；一是 Bleistein 为代表的基于反射波线性反演路线（Bleistein，1987；Bleistein 等，2000）。

本文以地震波传播的散射场表达与逆散射成像为切入点，首先给出散射波的表达形式，讨论逆散射成像的本质。其中讨论了逆散射成像所需要的叠前地震观测数据及对应的观测方式。然后给出一般的线性化的最小二乘叠前深度偏移的基本理论框架，指出能给出 Green 函数计算的算子都可以实现线性化的

LS_PSDM。接着讨论了当前广受关注的最小二乘逆时深度偏移（LS_RTM）方法。最后分析了基于反射波的估计反射系数的成像和基于散射波的估计散射强度的成像方法的差异。指出针对勘探地震介质特征和波场特征的成像方法最具实用性的方法，逆散射成像应该在此基础上进行。

1 地震波传播的散射场表达

FWI 和波动理论地震波层析成像本质上是针对地下介质为背景+散射体的情形的，因此，散射体引起的散射场的表达是其中最基本的问题。

1.1 基本的声波方程和声散射势的引入

1.1.1 基本的声波方程

一般地，我们假设波在地下介质中的传播可以由如下的声波方程控制，

$$\left[\frac{1}{v(\mathbf{x})^2}\frac{\partial^2}{\partial t^2}-\rho(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla\right)\right]u(\mathbf{x},t)=S(\mathbf{x},t) \tag{1.1}$$

其中，$u(\mathbf{x},t)$为声场势，它是时间$t\in R$和空间坐标$\mathbf{x}\in\Omega\subseteq R$的函数。$v(\mathbf{x})$代表速度场；$\rho(\mathbf{x})$代表密度场；$S(\mathbf{x},t)$代表体积场源。

在频率空间域，上式变为

$$\mathcal{L}_a(\mathbf{x},\omega)u(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{1.2a}$$

其中

$$\mathcal{L}_a(\mathbf{x},\omega)=\nabla\cdot\frac{1}{\rho}\nabla+\frac{\omega^2}{\rho v^2(\mathbf{x})} \tag{1.2b}$$

为声波传播算子。

定义背景介质中的声波传播满足如下方程，

$$\mathcal{L}_{a0}(\mathbf{x},\omega)u_0(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega) \tag{1.3a}$$

其中

$$\mathcal{L}_{a0}(\mathbf{x},\omega)=\nabla\cdot\frac{1}{\rho_0}\nabla+\frac{\omega^2}{\rho_0 v_0^2(\mathbf{x})} \tag{1.3b}$$

为背景介质中的声波传播算子。

1.1.2 声散射势的引入

根据（1.2）和（1.3）式，定义声散射势为，

$$\begin{aligned}\mathcal{V}_a(\mathbf{x},\omega)&\equiv\mathcal{L}_a(\mathbf{x},\omega)-\mathcal{L}_{a0}(\mathbf{x},\omega)\\&=\left(\nabla\cdot\frac{1}{\rho}\nabla+\frac{\omega^2}{\rho v^2(\mathbf{x})}\right)-\left(\nabla\cdot\frac{1}{\rho_0}\nabla+\frac{\omega^2}{\rho_0 v_0^2(\mathbf{x})}\right)\\&=\nabla\cdot\frac{a_\rho}{\rho}\nabla+\omega^2\frac{a_v}{\rho v^2(\mathbf{x})}\end{aligned}\tag{1.4}$$

其中

$$a_\rho=1-\frac{\rho_0(\mathbf{x})}{\rho(\mathbf{x})}\tag{1.5a}$$

$$a_v(\mathbf{x})=1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}\tag{1.5b}$$

1.2 标量波方程和标量波散射势的引入

勘探地震学中，很多情况下关注标量波方程（或密度不变情况下的声波方程）。类似地，写出如下的标量波散射势。

频率空间域标量波方程形式为

$$\mathcal{L}_s(\mathbf{x},\omega)u(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega)\tag{1.6a}$$

$$\mathcal{L}_s(x,\omega)=\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}\tag{1.6b}$$

式 1.6b 为标量波传播算子。$\mathcal{L}_s$ 代表 $\mathcal{L}_{\text{scalar}}$。

定义背景介质中的声波传播满足如下方程，

$$\mathcal{L}_{s0}(\mathbf{x},\omega)u_0(\mathbf{x},\omega)=-\delta(\mathbf{x}-\mathbf{x}_S)S(\omega)\tag{1.7a}$$

$$\mathcal{L}_{s0}(\mathbf{x},\omega)=\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\tag{1.7b}$$

1.7b 为背景介质中的标量波算子。

根据（1.6）和（1.7）式，定义标量波散射势为，

$$\mathcal{V}_s(\mathbf{x},\omega)\equiv\mathcal{L}_s(\mathbf{x},\omega)-\mathcal{L}_{s0}(\mathbf{x},\omega)=\left(\nabla^2+\frac{\omega^2}{v^2(\mathbf{x})}\right)-\left(\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right)=-a_s(\mathbf{x})\frac{\omega^2}{v_0^2(\mathbf{x})}\tag{1.8a}$$

$$a_s(\mathbf{x})=1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}\tag{1.8b}$$

1.8b 称为散射强度。（1.8a）式定义的散射势，在速度扰动不大时，还可以写为如下的形式，

$$\begin{aligned}\mathcal{V}_s(\mathbf{x},\omega)&=-\frac{\omega^2}{v_0^2(\mathbf{x})}\left(1-\frac{v_0^2(\mathbf{x})}{v^2(\mathbf{x})}\right)=-\frac{\omega^2}{v_0^2(\mathbf{x})}\left(\frac{v^2-v_0^2}{v^2}\right)\\&=-\frac{\omega^2}{v_0^2(\mathbf{x})}\frac{(v-v_0)(v+v_0)}{v^2(\mathbf{x})}\\&\approx-\frac{\omega^2}{v_0^2(\mathbf{x})}\frac{2\delta v\cdot v_0}{v_0^2(\mathbf{x})}=-\frac{\omega^2}{v_0^2(\mathbf{x})}\frac{2\delta v(\mathbf{x})}{v_0(\mathbf{x})}\end{aligned} \tag{1.9}$$

上述形式引入 $\delta v(\mathbf{x})=v(\mathbf{x})-v_0(\mathbf{x})$ 更清楚地反映出速度的局部变化是散射势的本质。从散射势的定义也可以看出，估计散射势的绕（散）射波动理论层析成像与估计（或定位）反射系数的叠前偏移成像是有严格区别的。根据（1.9）式定义的散射势，可以写出散射势引起的场的传播控制方程，

$$\left[\nabla^2+\frac{\omega^2}{v_0^2(\mathbf{x})}\right]u_s(\mathbf{x},\omega)=-\mathcal{V}_s(x,\omega)u(\mathbf{x},\omega) \tag{1.10}$$

求解该方程可以得到散射势引起的场 $u_s(\mathbf{x},\omega)$，进而进行波动理论的层析成像。这就是散射波 FWI 的线性化正问题基础。

1.3 Lippmann-Schwinger 方程的导出及散射场的具体表达

借助于算子恒等式 $\mathcal{L}=\mathcal{L}_0+\mathcal{L}_0\left(\mathcal{L}_0^{-1}-\mathcal{L}^{-1}\right)\mathcal{L}$ 并定义 $\mathcal{V}=\mathcal{L}_0^{-1}-\mathcal{L}^{-1}$ 可以得到

$$\mathcal{L}=\mathcal{L}_0+\mathcal{L}_0\mathcal{V}\mathcal{L} \tag{1.11}$$

算子 $\mathcal{L}$ 是对应声波和标量波的 Green 函数。$\mathcal{L}^{-1}$ 和 $\mathcal{L}_0^{-1}$ 是“逆”Green 函数，它们经常被各自的共轭算子来代替。共轭算子描述的是波的反传播现象。因此，Green 函数可以描述从源到散射点，再到接收点的波传播过程；同样地，也可以描述从接收点到散射点，再回到源点的波传播过程。

假如波传播算子 $\mathcal{L}$ 用 Green 函数 G 表示，方程（1.11）可以重写为

$$G=G_0+G_0VG \tag{1.12}$$

方程（1.12）是非线性的，它能被进一步重写为

$$G=\left(I-G_0V\right)^{-1}G_0 \tag{1.13}$$

对上式的右端项进行 Taylor 展开，

$$G=\left[\sum_{j=0}^{\infty}\left(G_0\mathcal{V}\right)^j\right]G_0=G_0+G_0\mathcal{V}G_0+G_0\mathcal{V}G_0\mathcal{V}G_0+G_0\mathcal{V}G_0\mathcal{V}G_0\mathcal{V}G_0+\cdots \tag{1.14}$$

方程（1.12）、（1.14）被称为 Lippmann-Schwinger 方程（Stolt 和 Weglein，2012）。

很显然，假如 $j\le 2$，方程（1.14）就仅描述一个忽略了二阶以上散射波的波传播过程。因此，线性化的波传播算子仅仅描述波传播过程中的一阶散射。这就是

$$G\approx G_0+G_0\mathcal{V}G_0 \tag{1.15}$$

上式就是 Born 近似下的波传播算子。其中方程右端的第一项描述了背景场中的传播现象，第二项描述了一阶散射波的传播现象。方程（1.12）-（1.15）清楚地说明了 Born 近似的物理含义。只有当散射势比较弱时，（1.15）式中的高于二阶的项才可以被忽略不计。注意，从（1.14）式可以清楚地看出，高阶散射场能否被忽略，除了散射势比较弱的条件外，还有散射体的体积也很重要，因为散射场是散射势的体积积分结果。小散射体和弱散射势是 Born 近似成立的条件。

从（1.14）式可以看出，Born 近似线性化算子模拟出的总波场包括两部分：第一部分是由背景 Green 函数所描述的背景波场；第二部分是由散射势引起的、方程（1.14）中的第二项所描述的散射场。很显然，这存在一个背景速度场如何取才能得到符合 Born 近似要求的背景场的问题。原则性的要求是入射场中不能存在散射场成分。事实上，在特别复杂、变化特别剧烈的速度场中，一阶 Born 近似的应用效果是很值得考虑的。这也是人们逐步在考虑引入更高阶的 Born 近似，降低一阶 Born 近似弱散射要求的原因。但这样做也会引入高阶散射场计算上的复杂性。

假定 Born 近似假设成立，根据方程（1.15），线性化的总波场可以写成

$$u\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)=u_0\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)+u_s\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right) \tag{1.16}$$

其中，Born 近似后的散射波场记为

$$u_s\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)=-\omega^2\int d\mathbf{x}G_0\left(\boldsymbol{\xi},\mathbf{x},\omega\right)f\left(\mathbf{x}\right)u_0\left(\mathbf{x},\boldsymbol{\eta},\omega\right) \tag{1.17}$$

它代表从震源 $\boldsymbol{\eta}$ 到散射点 $\mathbf{x}$，再到接收点 $\boldsymbol{\xi}$ 的散射波场。其中 $f\left(\mathbf{x}\right)=\dfrac{1}{v_0^2\left(\mathbf{x}\right)}\dfrac{2\delta v\left(\mathbf{x}\right)}{v_0\left(\mathbf{x}\right)}$ 被称为散射强度，是逆散射成像要估计的量。

Born 近似的物理实质是：在背景场 $u_0\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)$ 中没有散射场；在扰动场中，只存在一次散射场。尽管用这种方式描述波传播与实际的波传播不同，但当二阶及二阶以上的散射场能量比较弱时，用 Born 近似散射场进行最小二乘叠前深度偏移 PSDM 估计散射强度还是比较有效的。这是当前最小二乘叠前深度偏移成像的基础。也预示了下一步的最小二乘叠前深度偏移成像需要考虑更高阶次的散射波才能得到更好的结果（Weglein 等，1986）。

关于子波问题，可以认为已经携带在背景场 $u_0\left(\mathbf{x},\boldsymbol{\eta},\omega\right)$ 中。也可以用 $G_0\left(\mathbf{x},\boldsymbol{\eta},\omega\right)S\left(\omega\right)$ 来表示，其中 $S\left(\omega\right)$ 代表子波的谱。但是，在最小二乘叠前深度偏移成像中，子波假设已知，仅仅散射强度。事实上，子波的振幅谱和相位谱的不正确，都会影响散射强度估计的精度。因此，对子波的研究是最小二乘叠前深度偏移反演成像能否得到好结果的重要因素。

2 逆散射问题分析

2.1 常速背景介质与平面波入射情形下散射场与散射势间的关系

首先给出常速介质中自由空间 Green 函数为

$$G\left(\mathbf{x}-\mathbf{x}',\omega\right)=\frac{e^{i\mathbf{k}\cdot\left(\mathbf{x}-\mathbf{x}'\right)}}{\left|\mathbf{x}-\mathbf{x}'\right|} \tag{2.1}$$

其中，$\mathbf{x}$ 代表场点位置，$\mathbf{x}'$ 代表散射源点位置，$\left|\mathbf{x}-\mathbf{x}'\right|$ 代表源点与场点之间的距离。$\mathbf{k}$ 代表场的波数。把（2.1）式代入（1.17）式定义的散射场中，得到

$$u_s(\mathbf{x},\omega)=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)u_0(\mathbf{x}',\omega)\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|} \tag{2.2}$$

一般地，在远场近似假设下面，入射场在常速介质中传播，可以用平面波近似，

$$u(\mathbf{x},\omega)=e^{i\mathbf{k}\cdot\mathbf{x}}=e^{i|k|\mathbf{s}_0\cdot\mathbf{x}} \tag{2.3}$$

其中，$\mathbf{s}_0$ 是代表平面波的入射方向的单位矢量。将（2.3）式代入（2.2）式得到，

$$u_s(\mathbf{x},\omega)=\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\mathbf{x}'}\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|} \tag{2.4}$$

（2.4）式描述的波的传播关系如图 1 所示。

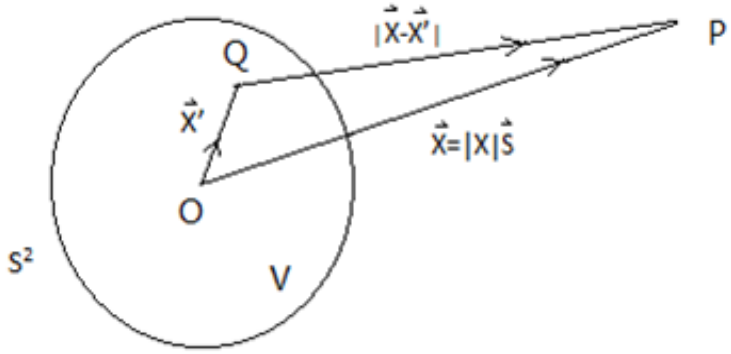


图 1 空间局部散射体引起的散射波传播关系

由（1.16）知，此时总场可表示为

$$u(\mathbf{x},\omega)=u_0(\mathbf{x},\omega)+u_s(\mathbf{x},\omega)=e^{ik\mathbf{s}_0\cdot\mathbf{x}}+\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\cdot\mathbf{x}'}\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|} \tag{2.5}$$

由图 1 可知，当场点 P 距离散射体比较远，散射体积源分布空间有限时，用 Q 代表散射体积源内的点，O 代表坐标原点，则可以做如下的近似推导。

$$|\mathbf{x}-\mathbf{x}'|=|\mathbf{x}|-\mathbf{s}\cdot\mathbf{x}' \tag{2.6}$$

$$G(\mathbf{x}-\mathbf{x}',\omega)=\frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{|\mathbf{x}-\mathbf{x}'|}=\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}\approx\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}e^{-ik\mathbf{s}\cdot\mathbf{x}'} \tag{2.7}$$

其中，$\mathbf{s}$ 是代表散射波传播的局部平面波方向的单位矢量。把（2.7）式代入（2.5）式，得到

$$
\begin{aligned}
u(\mathbf{x},\omega) &\approx e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{ik\mathbf{s}_0\cdot\mathbf{x}'}\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}e^{-ik\mathbf{s}\cdot\mathbf{x}'} \\
&= e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'}\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \\
&= e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}\int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'} \\
&= e^{ik\mathbf{s}_0\cdot\mathbf{x}} + \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|}f(\mathbf{s},\mathbf{s}_0,\omega)
\end{aligned}
\tag{2.8}
$$

其中

$$
f(\mathbf{s},\mathbf{s}_0,\omega) = \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-ik(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{x}'} \tag{2.9}
$$

称为散射波方向谱。定义散射势的 Fourier 变换为，

$$
\tilde{\mathcal{V}}(\mathbf{k}') = \int_V d\mathbf{x}'\mathcal{V}(\mathbf{x}',\omega)e^{-i\mathbf{k}'\cdot\mathbf{x}'} \tag{2.10}
$$

其中，$\mathbf{k}'$代表散射势的波数。对比（2.9）和（2.10）知，$f(\mathbf{s},\mathbf{s}_0,\omega) = \tilde{\mathcal{V}}\left(k(\mathbf{s}-\mathbf{s}_0)\right)$成立。即：散射势的 Fourier 变换等于散射波方向谱。

据此，可以总结出如下定理：在 Born 近似下，相对于散射体的远场区域，散射势的 Fourier 变换等于散射波方向谱。且有关系式

$$
\mathbf{k}' = \frac{\omega}{c}(\mathbf{s}-\mathbf{s}_0) \tag{2.11}
$$

成立（Born 和 Wolf，1999）。

注意到，散射波方向谱$f(\mathbf{s},\mathbf{s}_0,\omega)\left(=\tilde{\mathcal{V}}(\mathbf{k}')\right)$是由入射波和散射波夹角及散射波频率成分决定的，它的完整性取决于散射场记录的完整性，即散射波方向成分和散射波频率成分要是完整的。

由此可知，在常速介质情况下所谓逆散射问题就是根据散射波方向谱，确定散射势的 Fourier 的分量。散射势的所有的 Fourier 分量都确定了，通过 Fourier 逆变换，就可以完全精确地恢复散射势。

上述定理尽管是在平面波入射和常速背景情况下得到的，其理论指导意义是十分明显的。从$\mathbf{k}'$的表达式中可以看出，确定$\mathbf{k}'$的低波数部分，需要大角度散射波和低频数据；确定$\mathbf{k}'$的高波数部分，需要小角度散射波和高频数据。总之，震源子波的频带范围、观测系统和背景速度场的分布决定了散射强度估计结果的分辨率。因此，宽带、宽方位观测数据是高精度反演成像所需的。尽管（2.11）式中没有明确体现，高密度观测也是必需的。

2.2 高频近似下的 Born 近似的物理含义解释

Born 近似下的散射场（1.17）式也可以表示为，

$$
u_s(\boldsymbol{\xi},\boldsymbol{\eta},\omega) = -\omega^2\int d\mathbf{x}G_0(\boldsymbol{\xi},\mathbf{x},\omega)f(\mathbf{x})G_0(\mathbf{x},\boldsymbol{\eta},\omega) \tag{2.12}
$$

用渐进 Green 函数（即 Green 函数的 WKBJ 近似解）代入上式，

$$u_s(\boldsymbol{\xi},\boldsymbol{\eta},\omega)=\omega^2\int d\mathbf{x}f(\mathbf{x})A(\mathbf{x},\boldsymbol{\eta})A(\boldsymbol{\xi},\mathbf{x})\exp\left[-i\omega\left(T(\mathbf{x},\boldsymbol{\eta})+T(\boldsymbol{\xi},\mathbf{x})\right)\right] \tag{2.13}$$

在散射点周围，对旅行时进行线性化，即围绕中心射线进行 Taylor 展开，

$$T(\mathbf{x},\mathbf{x}')\approx T(\mathbf{x}_0,\mathbf{x}')+(\mathbf{x}-\mathbf{x}_0)\cdot\mathbf{p}(\mathbf{x}_0,\mathbf{x}') \tag{2.14}$$

其中$\mathbf{p}(\mathbf{x}_0,\mathbf{x}')$表示旅行时场$T(\mathbf{x},\mathbf{x}')$在点$\mathbf{x}_0$处的一阶导数，其方向为中心射线的切线方向。因此，(2.13)式中的指数部分可以化为，

$$\begin{aligned}&\exp\left[-i\omega\left(T(\mathbf{x},\boldsymbol{\eta})+T(\boldsymbol{\xi},\mathbf{x})\right)\right]\\&=\exp\left[-i\omega\left(T(\mathbf{x}_0,\boldsymbol{\eta})+T(\boldsymbol{\xi},\mathbf{x}_0)\right)\right]\exp\left[-i\omega(\mathbf{x}-\mathbf{x}_0)\left(\mathbf{p}(\mathbf{x}_0,\boldsymbol{\eta})+\mathbf{p}(\boldsymbol{\xi},\mathbf{x}_0)\right)\right]\\&=\exp\left[-i\omega\left(T(\mathbf{x}_0,\boldsymbol{\eta})+T(\boldsymbol{\xi},\mathbf{x}_0)\right)\right]\exp\left[-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}_0)\right]\end{aligned} \tag{2.15}$$

其中

$$\mathbf{k}=\omega\left(\mathbf{p}(\mathbf{x}_0,\boldsymbol{\eta})+\mathbf{p}(\mathbf{x}_0,\boldsymbol{\xi})\right) \tag{2.16}$$

代表波场的一种波数关系（Van Leeuwen 和 Mulder，2010）。其几何意义由图 2 所示。

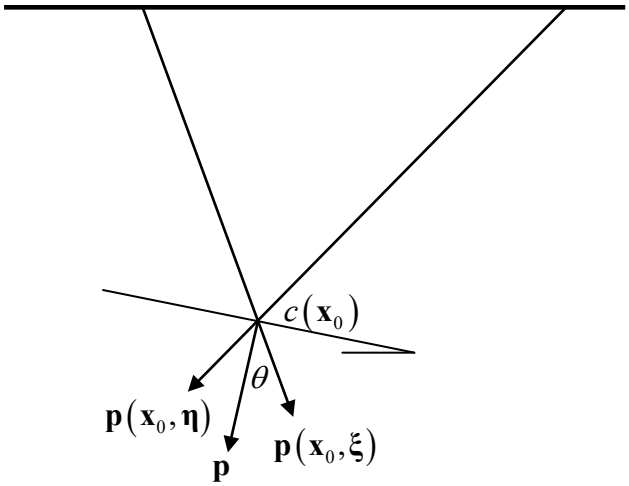


图 2 地震波传播过程中波矢量在反射点处的几何关系

如图 2 所示，$\mathbf{p}=\omega\left(\mathbf{p}(\mathbf{x}_0,\boldsymbol{\eta})+\mathbf{p}(\mathbf{x}_0,\boldsymbol{\xi})\right)=2\mathbf{n}\cos\theta/v_0(\mathbf{x}_0)$。其中$\mathbf{n}$为界面的外法线单位方向向量。$\theta$为射线的入射角。$v_0(\mathbf{x}_0)$为$\mathbf{x}_0$处的背景速度。根据上面的几何关系则有，

$$\mathbf{k}=\omega\mathbf{p}=\frac{2\omega\cos\theta}{v_0(\mathbf{x}_0)}\mathbf{n} \tag{2.17}$$

注意（2.17）式与（2.11）式的一致性。但这是从反射的角度来解释 从$\mathbf{k}$的表达式中可以看出，确定$\mathbf{k}$的低波数部分，需要大角度散射波和低频数据；确定$\mathbf{k}$的高波数部分，需要小角度散射波和高频数据。同样地，此处希望$\mathbf{k}$规定的谱的完整性。

这进一步证明了，高精度的逆散射反演成像需要宽频带和宽方位的地震数据观测。

3 Bayes 框架下基于 Born 近似散射场的散射势估计方法

（1.16）式定义了一个 Fredholm 第一类积分方程，它描述的是一个线性问题。此处，我们把该 Fredholm 第一类积分方程的求解转化为一个 Bayes 估计问题。就是把该线性问题的求解，转化为利用泛函变分求极值问题。基本的表述如下。假设（1.16）式定义的积分方程写出算子表达形式，

$$K\mathbf{m}=\mathbf{d} \tag{3.1}$$

其中，K 是 Fredholm 第一类积分方程的积分核，由（1.16）式知，积分核的定义为

$$K\left(v_0\left(\mathbf{x}\right),\mathbf{x},\omega\right)=\omega^2 G_0\left(\boldsymbol{\xi},\mathbf{x},\omega\right)u_0\left(\mathbf{x},\boldsymbol{\eta},\omega\right) \tag{3.2}$$

$\mathbf{m}$ 代表由 $f\left(\mathbf{x}\right)$ 作为元素形成的矢量；$\mathbf{d}$ 代表由每个炮检点对应的地震道的频谱值。

认为 $K:\mathrm{H}_1\rightarrow\mathrm{H}_2$ 是一个有界的线性算子，定义如下的 Tikhonov 泛函，

$$\mathfrak{I}\left(\mathbf{m}\right)=\frac{1}{2}\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2 \tag{3.3}$$

对上式求解的首要问题是计算泛函关于参数变化的导数，即泛函的梯度。为此，利用如下的变分方法。

对任意的介质扰动 $\delta\mathbf{m}\in\mathrm{H}_1$ 和扰动步长 $\tau\in\mathbb{R}$，总存在如下的泛函变分

$$\begin{aligned}
&\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right)=\frac{1}{2}\left\|K\left(\mathbf{m}+\tau\delta\mathbf{m}\right)-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\frac{\alpha}{2}\left\|\mathbf{m}+\tau\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\\
&=\frac{1}{2}\left\langle K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}+\tau K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\frac{\alpha}{2}\left\langle\mathbf{m}+\tau\delta\mathbf{m},\mathbf{m}+\tau\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\frac{1}{2}\left\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\tau^2}{2}\left\langle K\delta\mathbf{m},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}\\
&+\frac{\alpha}{2}\left\langle\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)+\frac{\alpha\tau^2}{2}\left\langle\delta\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\frac{1}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\left\langle K\delta\mathbf{m},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\delta\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\frac{1}{2}\left(\left\|K\mathbf{m}-\mathbf{d}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\\
&+\frac{\tau}{2}\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\left\langle K\delta\mathbf{m},K\mathbf{m}-\mathbf{d}\right\rangle_{\mathrm{H}_2}\right)+\frac{\alpha}{2}\tau\left(\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\left\langle\delta\mathbf{m},\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&+\frac{\tau^2}{2}\left(\left\|K\delta\mathbf{m}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)
\end{aligned} \tag{3.4}$$

对上式定义的变分求偏导数，

$$
\begin{aligned}
&\frac{\partial}{\partial\tau}\mathfrak{I}\left(\mathbf{m}+\tau\delta\mathbf{m}\right)\Big|_{\tau=0}\\
&=\left\{\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)+\tau\left(\left\|K\delta\mathbf{m}\right\|_{\mathrm{H}_2}^2+\alpha\left\|\delta\mathbf{m}\right\|_{\mathrm{H}_1}^2\right)\right\}\Big|_{\tau=0}\\
&=\left(\left\langle K\mathbf{m}-\mathbf{d},K\delta\mathbf{m}\right\rangle_{\mathrm{H}_2}+\alpha\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\left(\left\langle K^*\left(K\mathbf{m}-\mathbf{d}\right),\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}+\alpha\left\langle\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\right)\\
&=\left\langle K^*\left(K\mathbf{m}-\mathbf{d}\right)+\alpha\mathbf{m},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\left\langle\left(K^*K+\alpha I\right)\mathbf{m}-K^*\mathbf{d},\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}\\
&=\left\langle\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right),\delta\mathbf{m}\right\rangle_{\mathrm{H}_1}
\end{aligned}
\tag{3.5}
$$

其中

$$
\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)=\left(K^*K+\alpha\mathbf{I}\right)\mathbf{m}-K^*\mathbf{d} \tag{3.6}
$$

是误差泛函相对于参数模型的梯度。K^* 是波场正传播算子的共轭算子，代表波场的反传播。经典地，我们利用该梯度构造最速下降迭代方法求解 Born 近似后的线性问题的解。迭代格式为，

$$
\begin{aligned}
&\mathbf{m}^{k+1}=\mathbf{m}^k-\mu\mathrm{Grad}\mathfrak{I}\left(\mathbf{m}\right)=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}-\mathbf{d}^{obs}\right)\right]=\mathbf{m}^k-\mu\left[K^*\left(K\mathbf{m}^k-\mathbf{d}^{obs}\right)\right]\\
&=\mathbf{m}^k-\mu\left[K^*\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)\right]
\end{aligned}
\tag{3.7}
$$

当然，更一般性的解法是令（3.6）式定义的梯度等于零，得到如下的线性系统的法方程

$$
\left(K^*K+\alpha\mathbf{I}\right)\mathbf{m}=K^*\mathbf{d} \tag{3.8}
$$

法方程的求解一般用共轭梯度方法。但是，由于勘探地震中模型描述用网格化方式，模型参数个数非常多，（3.8）式右端的 Hessian 矩阵异常庞大。若用 $NX\times NY\times NZ$ 表示要估计的模型参数的个数，那么 Hessian 矩阵的规模为 $\left(NX\times NY\times NZ\right)\times\left(NX\times NY\times NZ\right)$。当前的计算机技术还不能满足这样巨大规模的线性方程组的求解。形成 Hessian 矩阵本身也是一个巨大的挑战。所以，**当前解法方程进行最小二乘偏移成像的基本做法是先进行常规的叠前深度偏移成像，然后用对角矩阵元素（就是地震波照明能量!）对成像叠加后的结果进行照明补偿，就是用 Hessian 矩阵的对角元素去除成像结果**。也可以用角度 Hessian 的对角元素去除每个共角度成像剖面，然后进行叠加产生最后的剖面，当然也可以不叠加产生照明补偿后的角度成像道集。（3.8）式代表的问题可以认为是图像反褶积问题。右端项是常规偏移成像得到的结果，它是模糊的图像。用共轭梯度法解（3.28）式相当于图像去模糊运算，可以显著地提高最终图像的质量。求解（3.8）式最关键的问题是难以形成 Hessian 矩阵，根据 Hessian 矩阵元素的物理含义（相邻点 Green 函数的互相关），提出了很多简化的 Hessian 矩阵生成方法。

实质上，（3.6）式计算的梯度不会等于零，因为正演数据和实测数据的残差永远不会等于零。这是最小二乘偏移成像中迭代解法占主要地位的根本原因。迭代法涉及残差计算 $\mathbf{r}^{k+1}=K\mathbf{m}^k-\mathbf{d}^{obs}$，这需要一步反偏移；$K^*\mathbf{r}^{k+1}=K^*\left(K\mathbf{m}^k-\mathbf{d}^{obs}\right)$ 代表残差的反传播，需要一步残差的反向传播；要考虑步长的计算；考虑用 Hessian 矩阵的逆对梯度项进行预条件处理。这四部分是一步迭代需要的计算量。一般地需要 n 步迭代计算。因此，迭代法的计算量也是很大的。

迭代的线性最小二乘反演方法中，并不更新核函数，因为我们假定核函数 $K\left(v_0\left(\mathbf{x}\right),\mathbf{x},\omega\right)$ 仅仅与背景

速度$v_0(\mathbf{x})$有关。但这仅仅是从计算效率方面考虑的，本质上还是希望更新核函数，因为背景速度与扰动速度是相互耦合的。存在较强多次散射波（反射波）时，要么事先消除多次波，要么构建非线性的最小二乘反演成像方法。

（3.7）或（3.8）式中计算出的量$\mathbf{m}(\mathbf{x})$代表的是速度的一种相对变化。只有在速度跃变的界面或绕射点处才会出现，因此，它反映了速度场中的高频变化部分。用这部分高波数速度扰动可以代表界面位置和绕射位置图像。这与一般的叠前深度偏移是很类似的。也可以精确估计这部分高波数速度扰动，用它们来刻画储层的性质。但是，我们认为，只有提供满足（2.11）或（2.17）要求的叠前数据体时，反演成像的结果才是比较精确的，才可以进行有效的储层解释。目前，仅仅利用地表观测的反射（散射）波场，加上噪音的存在，$\mathbf{m}(\mathbf{x})$的估计结果很难满足精确描述储层的要求。线性的最小二乘反演成像可以认为是背景速度正确时的反（散）射波 FWI。因此，目前的 FWI 要得到适用于储层解释的反演结果也是不现实的。速度（更应该是波阻抗）跃变的精确估计是一个比背景速度估计更难的问题。

值得指出的是，本节的讨论适用于利用各种波传播算子（包括 Kirchhoff 积分算子，Gauss-Beam 算子，单向波传播算子以及双向波传播算子）计算 Green 函数，然后进行线性化的最小二乘叠前深度偏移成像。

4 线性与非线性的最小二乘逆时深度偏移（LS_RTM）

FWI 的本质目的是估计宽或全波数带的速度（最好是波阻抗）。但是当前勘探地震数据采集并没有按（2.11）或（2.17）式的要求进行，而是仅仅在地表进行有限带宽和有限孔径的观测。这是当前的 FWI 分成估计低波数背景速度和高波数速度扰动两部分的主要原因。FWI 背景速度估计主要是利用低频长偏移距的透射波；FWI 速度扰动的估计是在假设背景速度正确的基础上利用反（散）射波进行。速度（或波阻抗）扰动的估计必须利用反（散）射波的振幅信息。基于目前的数据采集，它们的估计精度更为难以保障。众所周知，目前速度（或波阻抗）扰动的估计的技术热点是线性或非线性的最小二乘逆时深度偏移方法。

4.1 非线性的 LS_RTM

首先列出非线性的最小二乘逆时深度偏移成像（LS_RTM）所用的公式。正向波场外推方程为：

$$\begin{cases}\left(\dfrac{1}{v^2(x)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)u(\mathbf{x},t)=-\delta(\mathbf{x}-\mathbf{x}_S)S(t)\\ u(\mathbf{x},t)\big|_{\mathbf{x}=\Omega}=f_b(\Omega,t)\end{cases}\tag{4.1a}$$

其中，$f_b(\Omega,t)$代表边界条件。一般用 PML 边界条件。Ω代表计算波场在空间上的边界。

反向波场外推方程为：

$$\begin{cases}\left(\dfrac{1}{v^2(x)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)z(\mathbf{x},t)=0\\ z(\mathbf{x},t)\big|_{\mathbf{x}=\boldsymbol{\eta}}=r(\boldsymbol{\eta},t)=u^{cal}(\boldsymbol{\eta},t)-u^{obs}(\boldsymbol{\eta},t)\end{cases}\tag{4.1b}$$

其中，$\boldsymbol{\eta}$代表观测点坐标值。$u^{obs}(\boldsymbol{\eta},t)$代表观测点处的实测波场；$u^{cal}(\boldsymbol{\eta},t)$代表观测点处的正演合成波场。FWI 进行波场残差外推时，仅有观测点处有波场残差，因此，才有（4.1b）这样的边界条件的定义。实际上，波场逆时外推并不限于这样的边界条件。

可见，（4.1a）是一个初值问题，沿时间正向外推计算；（4.1b）是一个边值问题，逆时间反向外推计算。

误差泛函梯度公式为：

$$\mathbf{g}\left(\delta\mathbf{m}\left(\mathbf{x}\right),\delta\mathbf{m}_B\left(\mathbf{x}\right),\mathbf{m}_B\left(\mathbf{x}\right)\right)=\sum_t\sum_{\boldsymbol{\xi}}\frac{1}{v^2\left(\mathbf{x}\right)}\frac{\partial^2 u(\boldsymbol{\xi};\mathbf{x},t)}{\partial t^2}z(\boldsymbol{\xi};\mathbf{x},t) \tag{4.2}$$

其中，$\boldsymbol{\xi}$代表不同的炮点。对应的迭代公式为：

$$\delta\mathbf{m}^{k+1}=\delta\mathbf{m}^k-\alpha^k H_h^{-1}\mathbf{g}_h\left(\delta\mathbf{m}\left(\mathbf{x}\right),\delta\mathbf{m}_B\left(\mathbf{x}\right),\mathbf{m}_B\left(\mathbf{x}\right)\right) \tag{4.3a}$$

$$\delta\mathbf{m}_B^{k+1}=\delta\mathbf{m}_B^k-\beta^k H_l^{-1}\mathbf{g}_l\left(\delta\mathbf{m}\left(\mathbf{x}\right),\delta\mathbf{m}_B\left(\mathbf{x}\right),\mathbf{m}_B\left(\mathbf{x}\right)\right) \tag{4.3b}$$

其中，α^k和β^k分别为速度扰动和背景速度修正量的步长因子，H_h^{-1}和H_l^{-1}分别为对应速度扰动和背景速度的梯度向量的预条件矩阵。$\mathbf{g}_h$和$\mathbf{g}_l$分别为对应速度扰动和背景速度的误差泛函梯度向量。

与前述基于一阶 Born 近似公式导出的 LS_PSDM 相比，LS_RTM 的梯度项中包含了对所有波现象的预测。当背景速度不准时，不能预测的波现象会干扰梯度的计算，导致收敛慢，甚至不收敛的情况。比较而言，一阶 Born 近似仅仅考虑一次散射波的预测，相对而言，梯度计算要更准确，使得 LS_PSDM 的收敛性更好。但 LS_PSDM 要求$u^{cal}\left(\boldsymbol{\eta},t\right)$中仅仅包含一次散射（反射）波场，必须对实测数据$u^{obs}\left(\boldsymbol{\eta},t\right)$做很好的预处理，尽可能使之仅保留一次散射波场。

非线性 LS_RTM 的计算步骤：

（1）给定较准确的背景偏移速度$\mathbf{m}_B$；

（2）计算波场$u^{cal}\left(\boldsymbol{\eta},t\right)$和波场残差$r\left(\boldsymbol{\eta},t\right)$，并产生照明矩阵；

（3）计算梯度$\mathbf{g}\left(\delta\mathbf{m}\left(\mathbf{x}\right),\delta\mathbf{m}_B\left(\mathbf{x}\right),\mathbf{m}_B\left(\mathbf{x}\right)\right)$，并给出 RTM 成像结果。当成像道集拉平或数据拟合差达到要求，迭代停止。

（4）分离梯度$\mathbf{g}\left(\delta\mathbf{m}\left(\mathbf{x}\right),\delta\mathbf{m}_B\left(\mathbf{x}\right),\mathbf{m}_B\left(\mathbf{x}\right)\right)$为$\mathbf{g}_h\left(\delta\mathbf{m}\left(\mathbf{x}\right),\delta\mathbf{m}_B\left(\mathbf{x}\right),\mathbf{m}_B\left(\mathbf{x}\right)\right)$和$\mathbf{g}_l\left(\delta\mathbf{m}\left(\mathbf{x}\right),\delta\mathbf{m}_B\left(\mathbf{x}\right),\mathbf{m}_B\left(\mathbf{x}\right)\right)$；

（5）考虑梯度预条件和步长，利用（3.33）式分别修改速度扰动和背景速度。

（6）修改模型$\mathbf{m}\left(\mathbf{x}\right)=\mathbf{m}_B\left(\mathbf{x}\right)+\delta\mathbf{m}_B\left(\mathbf{x}\right)+\delta\mathbf{m}\left(\mathbf{x}\right)$；

（7）否则返回步骤（2）。

从上述非线性 LS_RTM 的计算步骤可以看出，非线性 LS_RTM 包含了速度扰动量的反演和背景扰动量的反演。两种成分的联合反演是目前地震波反演的导向性思想。

4.2 基于 Born 线性近似的 LS_RTM

4.2.1 基于 Born 线性近似的 LS_RTM 方法

目前对 LS_RTM 的讨论主要是基于（1.16）式的，我们称之为线性化的 LS_RTM。为估计散射势，定义如下误差泛函，

$$\mathcal{J}=\frac{1}{2}\left\|u_s^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)-u_s^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right\|_2^2=\frac{1}{2}\sum_{\omega}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\left(u_s^{cal}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)-u_s^{obs}\left(\boldsymbol{\xi},\boldsymbol{\eta},\omega\right)\right)^2 \tag{4.4}$$

误差泛函能量最小化时，认为估计的散射势是所求的参数估计结果。其中$u_s^{cal}\left(\xi,\eta,\omega\right)$由（1.16）式定义。

目标泛函的梯度是构造最速下降法的核心。该梯度定义为

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \mathbf{m}_r(\mathbf{x})}&=\sum_{\omega}\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}\frac{\partial u_s^{cal}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)}{\partial \mathbf{m}_r}\left[u_s^{cal}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)-u_s^{obs}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)\right]^H\\&=\sum_{\omega}\omega^2\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}S(\omega)G_S(\mathbf{x},\boldsymbol{\eta},\omega)G_R(\boldsymbol{\xi},\mathbf{x},\omega)\left[u_s^{cal}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)-u_s^{obs}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)\right]^H\end{aligned}\tag{4.5}$$

其中 $S(\omega)G(\mathbf{x},\boldsymbol{\eta},\omega)$ 代表震源点 $\boldsymbol{\eta}$ 传到地下任一点 $\mathbf{x}$ 的场，$G_R(\boldsymbol{\xi},\mathbf{x},\omega)\left[u_s^{cal}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)-u_s^{obs}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)\right]^H$ 代表残差波场的反传波。

利用（4.5）定义的梯度可以进行最速下降法的最小二乘偏移成像的具体计算，

$$\mathbf{m}_r^{(k+1)}(\mathbf{x})=\mathbf{m}_r^{(k)}(\mathbf{x})+\mu^{(k)}\delta\mathbf{m}_r^{(k)}(\mathbf{x})=\mathbf{m}_r^{(k)}(\mathbf{x})+\mu^{(k)}\frac{\partial \mathcal{J}}{\partial \mathbf{m}_r(\mathbf{x})}\tag{4.6}$$

为了加速收敛，可以用 Hessian 矩阵的逆对梯度方向进行预条件处理，Hessian 矩阵的定义为，

$$\begin{aligned}\mathbf{H}(\mathbf{x},\mathbf{y})&=\frac{\partial^2 \mathcal{J}}{\partial \mathbf{f}(\mathbf{x})\partial \mathbf{f}(\mathbf{y})}=\sum_{\omega}\omega^4\sum_{\boldsymbol{\eta}}\sum_{\boldsymbol{\xi}}S(\omega)G_S(\mathbf{x},\boldsymbol{\eta},\omega)G_R(\boldsymbol{\xi},\mathbf{x},\omega)\left[\frac{\partial u_S^{cal}(\boldsymbol{\xi},\boldsymbol{\eta},\omega)}{\partial \mathbf{m}_r(\mathbf{y})}\right]^H\\&=\sum_{\omega}\omega^4\sum_{\boldsymbol{\eta}}S^2(\omega)G_S(\mathbf{x},\boldsymbol{\eta},\omega)G_S^H(\mathbf{y},\boldsymbol{\eta},\omega)\sum_{\boldsymbol{\xi}}G_R(\boldsymbol{\xi},\mathbf{x},\omega)G_R^H(\boldsymbol{\xi},\mathbf{y},\omega)\end{aligned}\tag{4.7}$$

该 Hessian 矩阵是由背景介质中的 Green 函数规定的，其中并不包括高阶反射/绕射效应，即不包括非线性部分。双向波计算 Green 函数可以包含这部分的作用。由 Hessian 矩阵的逆对梯度方向进行预条件处理后，最速下降法的迭代格式变为，

$$\mathbf{m}_r^{(k+1)}(\mathbf{x})=\mathbf{m}_r^{(k)}(\mathbf{x})+\mu^{(k)}\delta\mathbf{m}_r^{(k)}(\mathbf{x})=\mathbf{m}_r^{(k)}(\mathbf{x})+\mu^{(k)}\mathbf{H}^{-1}\frac{\partial \mathcal{J}}{\partial \mathbf{m}_r(\mathbf{x})}\tag{4.8}$$

迭代步长 $\mu^{(k)}$ 一般可以采用试探的方式进行确定，一般采用试探加抛物拟合来选取最优下降步长来保证误差的下降。

迭代的每一步需要在更新后对模型进行一次反偏移，然后将反偏移得到的正演数据从原始数据中减去，这样来判断误差是否达到要求，然后判断迭代是否终止。

具体的计算步骤为：

（1）给定偏移速度场 $v_{mig}(\mathbf{x})$；

（2）用观测波场 $u_s^{obs}(\boldsymbol{\xi},\boldsymbol{\eta},t)$ 进行 RTM，作为初始的 LS_RTM 结果；

（3）解（1.10）式计算 $u_s^{cal}(\boldsymbol{\xi},\boldsymbol{\eta},t)$，计算残差波场；残差波场能量小于预设值，迭代停止。

（4）用残差波场进行 RTM，同时产生照明矩阵；

（5）用（4.8）式计算迭代成像结果；

（6）进行 RTM；

（7）转到步骤（3）。

步骤（3）也可以解法方程。但是，形成法方程本身就要巨大的计算量，解法方程也需要大量的内存空间和计算量。解法方程的优点是正则化和预条件的方法比较多。可以看出，此处的梯度计算和波场残差的计算在理论上是不够严谨的。

4.2.2 数值结果

图 3 展示了用公式（4.8）进行的最小二乘反射波叠前偏移成像的结果。可以看出，最小二乘偏移补偿了照明不足，提高了成像的分辨率，是成像结果的横向一致性更好。这样的结果更有利于后续的储层解释。

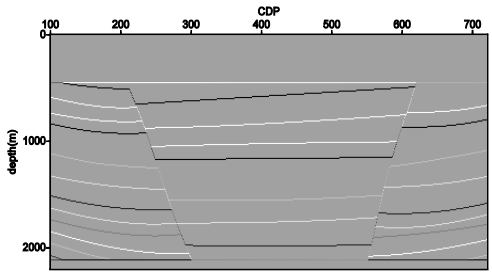


图 3-1 反射系数模型

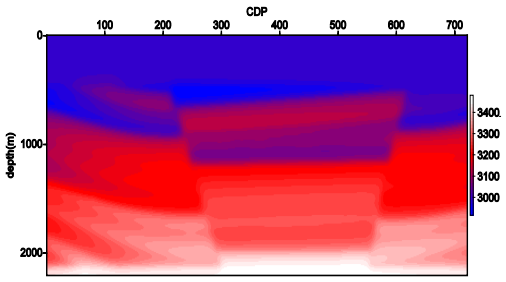


图 3-2 背景速度模型

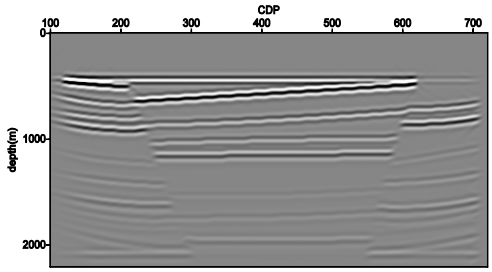


图 3-3 一般叠前深度偏移结果

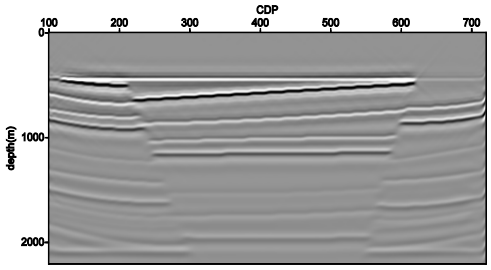


图 3-4 迭代 20 次的叠前深度偏移结果

图 3 迭代法最小二乘叠前深度偏移成像结果

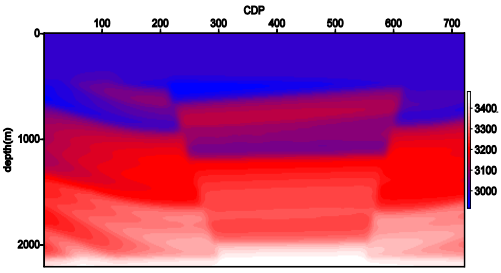


图 4-1 初始速度模型

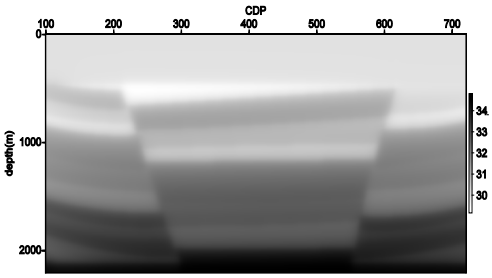


图 4-2 初始 Q 值模型

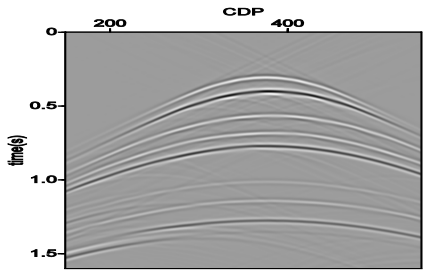


图 4-3 声介质单炮道集

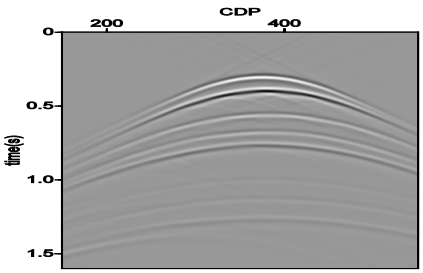


图 4-4 粘声介质单炮道集

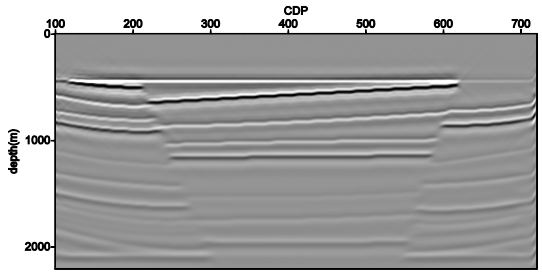


图 4-5 声介质迭代 20 次的叠前深度偏移结果

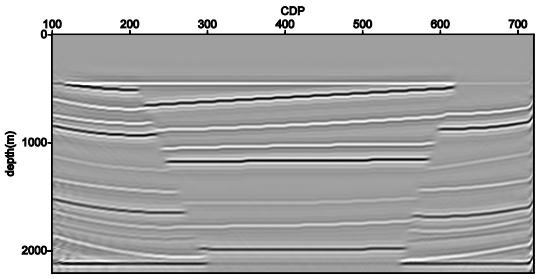


图 4-6 粘声介质迭代 20 次的叠前深度偏移结果

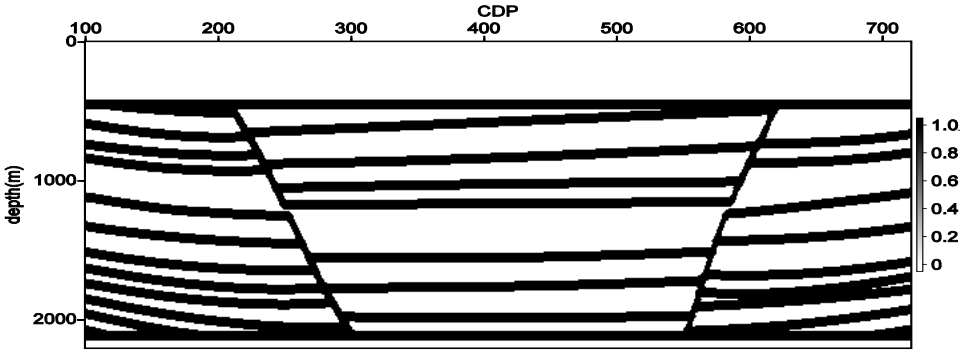


图 4-7 正则化最小二乘偏移成像的加权系数

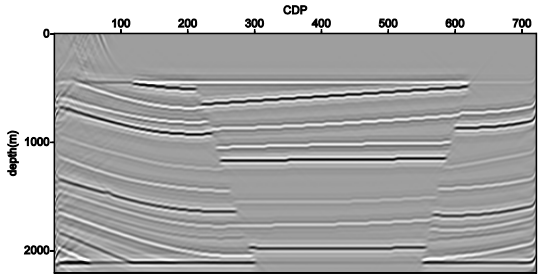


图 4-8 粘声介质迭代 20 次的 Tkhonov 正则化叠前深度偏移结果

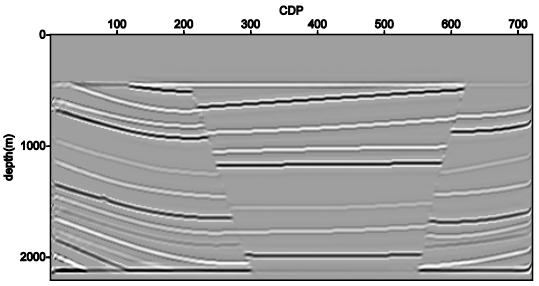


图 4-9 粘声介质迭代 20 次的加权正则化叠前深度偏移结果

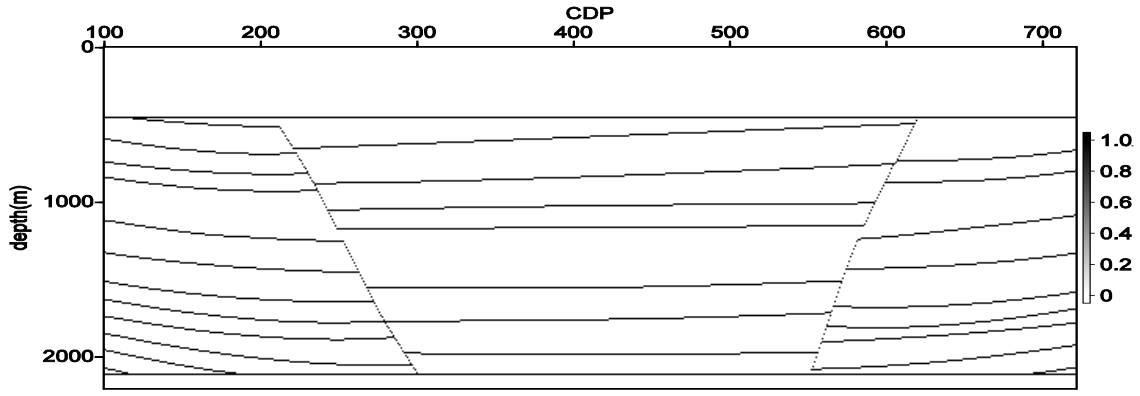


图 4-10 正则化最小二乘偏移成像的加权系数

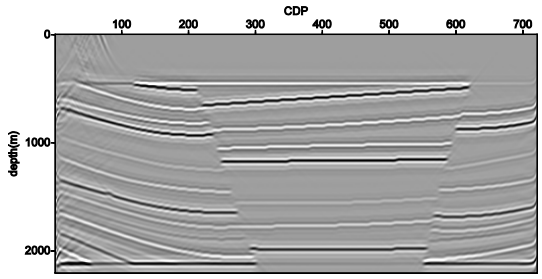


图 4-11 粘声介质迭代 20 次的 Tkhonov 正则化叠前深度偏移结果

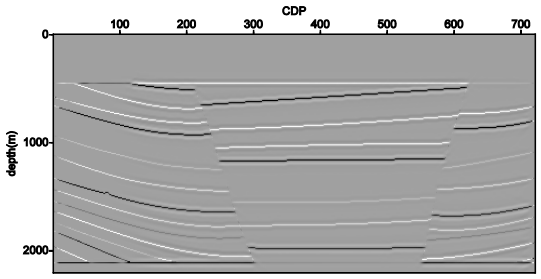


图 4-12 粘声介质迭代 20 次的加权正则化叠前深度偏移结果

图 4 展示了声介质和粘声介质下最小二乘叠前深度偏移的对比，以及 Tikhonov 正则化和加权正则化情形下迭代最小二乘叠前深度偏移的对比，以及不同加权系数下迭代最小二乘叠前深度偏移的对比。可以看出，最后的粘声介质迭代 20 次的加权正则化叠前深度偏移结果与真实的模型的反射系数已经非常接近了（胡江涛，2015）。理论上讲，如此高分辨率和高保真的叠前深度成像结果完全胜任岩性油气藏的精细刻画的要求。当然，对实际数据而言，成像结果不会如此理想。但这并不改变上述理论预测的正确性。

4.2.3 总变差（TV）正则化约束下的 LS_RTM

对于实际数据而言，多重因素的制约使得最小二乘反演成像成为一个不适定的反问题。譬如输入的叠前数据中的噪音（非一次反射波）、不准确的背景速度、正演（偏移）算子不能很好地模拟波现象、子波未知等一系列的因素会影响最小二乘反演成像迭代过程中的数据匹配，使不准确的数据残差反投影混叠在梯度中，进而影响最小二乘反演成像的收敛率。正则化是使不适定（或非线性较强或凸性差）的地球物理反问题的解更稳定、更有地质意义的一类方法。常用的 Tikhonov 正则化可以提升反问题解的稳定性，利用这种基于 L2 模的正则化方法得到的是一光滑解，而我们所期望的，是“棱角”更为分明的成像剖面，能突出地刻画地下的反射结构。所以，具备保留图像边界性质的 TV 正则化是更为合适的选择。

TV 正则化具有如下的形式：

$$R(\mathbf{m})=\|\mathbf{m}\|_{TV}=\sum_{1\le i,j\le n}\sqrt{\left|\left(\nabla_x\mathbf{m}\right)_{i,j}\right|^2+\left|\left(\nabla_z\mathbf{m}\right)_{i,j}\right|^2} \tag{4.9}$$

其中，$\left(\nabla_x\mathbf{m}\right)_{i,j}=m_{i+1,j}-m_{i,j}$，$\left(\nabla_z\mathbf{m}\right)_{i,j}=m_{i,j+1}-m_{i,j}$ 为 2D 情形下 $\mathbf{m}$ 在空间点 (i,j) 处的方向导数。TV 正则化在原点处不可微，使其难以求解，可利用如下策略将 TV 正则化加入 LSM 迭代步中 (Lin 和 Huang，2015)。

$$\begin{aligned}
&\text{初始化：}\ \mathbf{u}^0=\mathbf{m}^0 \quad k=0\\
&\mathbf{m}^{(k)}=\min_{\mathbf{m}}\left\{\left\|\mathbf{d}-L\mathbf{m}^{(k-1)}\right\|_2^2+\lambda_1\left\|\mathbf{m}-\mathbf{u}^{(k-1)}\right\|_2^2\right\}\\
&\mathbf{u}^{(k)}=\min_{\mathbf{u}}\left\{\left\|\mathbf{m}^{(k)}-\mathbf{u}\right\|_2^2+\lambda_2\left\|\mathbf{u}\right\|_{TV}\right\}\\
&k=k+1
\end{aligned}$$

上述迭代流程中，对每一迭代步中的像进行 TV 去噪，去噪结果作为下一迭代步的先验信息，可以有效的去除像中的偏移假象，随机噪音，提高成像质量。

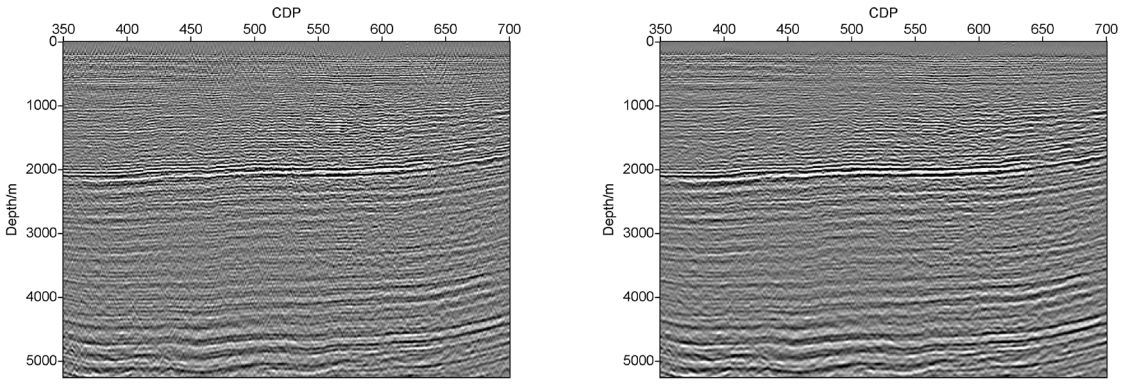


图 5 实际数据常规 LS-PSDM 结果（左）与 TV 正则化下 LS-PSDM 结果（右）

从图 5 实际数据常规 LS-PSDM 结果与 TV 正则化下 LS-PSDM 结果的对比中可以看出，TV 正则化去除了成像结果中的随机噪音干扰，提高了同相轴的连续性，从而提高成像质量。

5 针对反射系数的成像与针对散射强度成像的差异分析

到目前为止，本文的讨论都是建立在地下介质分布是背景+散射的基础上的。实质上，勘探地震所面对的介质可以描述为：在空间上广泛分布的层状沉积层，加上火山活动，构造运动等引起的大、小尺度的速度异常体（或波阻抗变化的异常体）。针对这样的介质情况，用背景+反射界面的方式来描述更为合适，体现了地下介质的特征及在其中传播的波场的特征。我们认为基于地下介质特征和波场特征的构建的成像方法和技术才是最实用的。

前已述及，针对散射的成像需要满足（2.11）和（2.17）式规定的野外采集方式得到的数据。当前，仅仅在地表进行有限带宽和有限孔径的观测显然无法满足逆散射成像的要求。

事实上，到目前为止的勘探地震成像技术都是基于反射波的，成像的目标是定位和估计反射界面的反射系数。勘探地震中，当前反射波成像理论可以总结为：在最小二乘成像理论框架下，假设地下介质是层状的（每层都终结于无穷远边界上），假设观测系统是无穷宽方位和长偏移距的、炮检（尽可能）对称的、无假频的，假设背景速度正确（满足 Born 近似）且子波已知，方位角度反射系数可以准确地估计出来，并有一个半解析的表达形式（Lambar 等，2003）。Bleistein 给出了上述成像方法的完整的理论描述（Cohen 等，1986；Bleistein 等，1985），这是勘探地震学家给出的反射波地震成像的最完整表达。

FWI（包括上述 LS_RTM）都是针对散射体的成像，试图估计散射强度 $f\left(\mathbf{x}\right)$。仅有地表观测数据，根据当前 FWI（LS_RTM）反演理论，很难给出反演解的精度评价。这是当前基于散射和逆散射成像理论的主要困境。

研究人员已经意识到了勘探地震的介质特点和波传播特点，提出了改进的 LS_RTM 方法。LS_RTM 的初始迭代进行反射系数的估计，即进行反射波最小二乘叠前深度偏移成像，然后再进入散射强度的估计。这可以有效地改善 LS_RTM 的收敛效率，提高最终成像质量。

6. 总结与讨论

油气勘探地震的目标是含油气储层的识别与描述。最核心的技术是地震波成像得到地下介质的几何特征图像和弹性参数（主要是反射系数和波阻抗）图像。叠前深度偏移给出介质的几何特征图像；FWI（LS_RTM）试图给出弹性参数的估计。

基于散射波理论的 FWI 根本目标是给出宽（全）波数带的弹性参数估计。但是，它需要的数据是目前仅仅地表观测给出的有限带宽和有限孔径的无假频观测所无法满足的。依据这样的不完整数据给出的反演

结果的精度也很难评价。目前的 FWI 仅仅可以在低频长偏移距数据存在的情况下给出背景速度的估计；LS_RTM 是在假设背景速度正确的情况下的散（反）射波 FWI。数据观测的不完整、叠前数据中的噪音、不准确的背景速度、正演（偏移）算子不能很好地模拟波现象、子波未知等因素导致 LS_RTM 很难收敛，解的精度难于评价。这是 LS_RTM 的困境，估计会一直伴随 LS_RTM 方法技术发展，无法得到根本解决。

方位角度反射系数的估计是针对勘探地震所面对的层状沉积地震特点以及波在层状介质中传播的特点发展的成像方法，是勘探地震学所独有的，勘探地震学发展到现在主要是基于这样的反射波成像理论。

基于散射及逆散射成像的方法理论，到目前为止，还没有得到实用性的有效证明。但是，首先进行基于反射系数的成像，然后进入散射强度的成像无疑是可行的方法。针对散射波成像的正则化方法也有待深入研究。

参考文献

[1] Bleistein, N., 1987, On the imaging of reflectors in the earth. Geophysics, 52, 931-942

[2] Bleistein, N., Cohen, J. K., and Stockwell, J. W., Jr., 2000, Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion;

[3] Bleistein, N., Cohen, J. K., and Hagin, F., 1985, Two and one-half dimensional Born inversion with an arbitrary reference, Geophysics, 52, 26-36.

[4] Clayton, R. and Stolt, R., 1981, A Born-WKBJ inversion method for acoustic reflection data; Geophysics, 46, 1559-1567

[5] Cohen, J. K., Hagin, F. and Bleistein, N., 1986, Three dimensional Born inversion with an arbitrary reference, Geophysics, 52, 26-36.

[6] Devaney, A. J., 2012, Mathematical Foundations of Imaging, Tomography and Wavefield Inversion; Cambridge

[7] Lambar, G., Operto, S., Podvin, P. and Thierry, P., 2003, 3D ray+Born migration/inversion, Part I: Theory. Geophysics, 68, 1348-1356

[8] Lin, Y., & Huang, L. (2015, December). Least-Squares Reverse-Time Migration with Modified Total-Variation Regularization. In 2015 SEG Annual Meeting. Society of Exploration Geophysicists.

[9] Shuey, R., 1985, A simplification of the Zoeppritz equation; Geophysics, 50, 609-614

[10] Stolt, R., H., and Weglein, A. B., 1985, Imaging and Inversion of seismic data; Geophysics, 50, 2458-2472

[11] Stolt, R., H., and Weglein, A. B., 2012, Seismic Imaging and Inversion; Cambridge

[12] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation; Geophysics, 49, 1259-1266

[13] Wapenaar, C.P.A., 1996, Inversion versus migration: A new perspective to an old discussion; Geophysics, 61, 804-814

[14] Weglein, A. B., Violette, P. B., and Keho, T. H., 1986, Using multiparameter Born theory to obtain certain exact multiparameter inversion goals; Geophysics, 51, 1069-1074

[15] Van Leeuwen, T. and W. A. Mulder, A correlation-based misfit criterion for wave-equation traveltime tomography. Geophys. J. Int., 2010. 182(3): 1383-1394.

[16] Born, M., and Wolf, E., 1999, Principles of optics, 7th expanded edition, Cambridge University Press

[17] 胡江涛，2015，最小二乘逆时偏移及角度道集提取方法研究，同济大学博士论文

地震数据成像处理中若干重要问题的思考

王华忠

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】油气地震勘探的核心任务是对各种数据进行处理和分析，获取有关油气藏状态的信息。数据分析理论和方法逐渐成为油气地震勘探的重要理论基础（本质上就应该如此，只不过大数据分析时代使得这种关系更明瞭而已！）。数据的随机性和系统的线性性（数据总是由系统产生出来的！）是数据分析的最基本假设。随机信号的高斯分布假设也是数据分析方法的基本假设。在这样的基本框架下，发展出了基于 Bayes 理论的数据分析方法与技术体系。勘探地震中的主要方法技术（不敢说所有的方法技术）莫不出于该理论体系。油气地震勘探包括数据（信息）采集、信息处理与分析以及信息解释与应用三个环节。本质上，与油气藏状态有关的信息提取和分析居于核心地位。勘探地震学已经发展了一套常规的、工业生产中实用的方法技术系列。但是，随着数据采集质量的提高（“两宽一高”的数据采集）、高性能计算机系统及并行计算技术的飞速发展，更高精度的地震数据成像处理方法（FWI+LS_RTM）以及更精细的信息解释方法（ML/SL/AI）正逐渐进入实质性的应用阶段，极大地提高了油气藏描述的精度。在线性或非线性系统、高斯或非高斯假设、以及 Bayes 反演框架下，信号分析、图像分析、弹性参数反演（偏移成像与层析成像）中，都存在一些问题需要思考。理论框架提供了强有力的理论工具，具体到发展出实用有效的方法与技术，还有很多问题需要研究。本文在（1）关于混采（同时源激发采集）及解混叠技术问题的看法；（2）关于 2D 中心 Wiener 滤波器设计的思想；（3）DTW 方法的基本思想分析；（4）关于信号的平稳性、高斯性以及非线性反演之间的关系；（5）子波相位及对反演成像问题影响的思考；（6）信号与图像的稀疏表达与地震波反演成像；（7）从系统辨识角度看待 FWI 反演再议；（8）线性系统反问题的观点；（9）关于勘探地震学中非线性反演问题的再议；（10）保真偏移成像与反演成像中梯度问题的认识；（11）评述当前反问题；（12）ML/AI 储层描述这 12 个方面的问题展开讨论。 体现我对上述 12 个方面问题的思考与认识。

【关键词】地震波成像，信号分析，图像分析，线性反演成像，非线性反演成像，机器学习与统计学习

1 引言

勘探地震学发展到今天，似乎到了变革的时刻。变革的触动因素主要有三个方面。首先是数据采集技术的发展。“两宽一高”是标志性的地震数据采集技术，到目前为止，并没有真正地实现这样的地震数据采集。但是，单点、高密度、宽带（尤其是低频）的震源、宽带的接收（数字检波）器、无线高速传输、长时间供电的电池、无人机、激光测振、光纤测振等等技术的快速发展似乎预示着自动化的“两宽一高”地震数据采集时代的到来。然后是高速发展的高性能计算技术。GPU 的广泛应用、FPGA 的探索应用，甚至量子计算机都有可能在不远的将来实用化，这些技术的进步为巨量的地震数据分析奠定了坚实的计算基础。最后，大数据分析方法与技术的进步深刻地改变着勘探地震中基于已知物理模型进行数据分析的基本做法。今后，基于模型的和基于数据的两种数据分析方法在勘探地震学的数据分析中会逐渐融合，在不同的阶段发挥作用，极大地提高地震数据分析的精度和效率。

尽管如此，数据分析的基本理论框架是不会有大的变化的。数据是由系统输出的，实测数据总是随机的。当前，系统是线性、实测数据的概率分布是高斯的基本假设在地震数据处理和分析的各个阶段依然占据核心地位。非线性系统、非高斯假设、非平稳信号的假设在勘探地震中逐步在推广应用。但是，基于这样假设的地震数据分析和处理方法在应用上并没有取得根本的成功。应该说，数据分析的理论框架在逐步成熟，但是理论框架的基本假设与实际数据不匹配，以及所用的模型与实际数据不匹配，使得由理论框架到实用技术之间的路线很难确定。即便是当前大热的大数据分析理念和方法引入勘探地震数据分析中，也很难改变当前油气地震数据分析结果不能满足勘探需求的局面。

因此，对勘探地震数据成像处理中的重要问题进行深入分析是非常必要的。

本文对如下 12 个问题展开思考与分析，它们分别是：（1）关于混采（同时源激发采集）及解混叠技术问题的看法；（2）关于 2D 中心 Wiener 滤波器设计的思想；（3）DTW 方法的基本思想分析；（4）关于信号的平稳性、高斯性以及非线性反演之间的关系；（5）子波相位及对反演成像问题影响的思考；（6）信号与图像的稀疏表达与地震波反演成像；（7）从系统辨识角度看待 FWI 反演再议；（8）线性系统反问题的观点；（9）关于勘探地震学中非线性反演问题的再议；（10）保真偏移成像与反演成像中梯度问题的认识；（11）评述当前反问题；（12）ML/AI 储层描述。分析的目的当然是希望弄清楚问题的本质，确定解决问题的正确方向，提出解决问题的有效方法。

2 关于混采（同时源激发采集）及解混叠技术问题的看法

（1）地震数据采集代表性技术是“两宽一高”采集，但是目前的硬件条件不能实现真正的“两宽一高”采集。

（2）我认为最好的地震数据采集方式是：压缩感知采集理论 + 时空随机采样 + 自动化布设相互独立的炮检点。

（3）当前，同时源激发采集是用震源同时激发（随机时间起振）来提高单位时间内采集效率，某种程度上达到高密度采样（尤其是炮密度增加）的要求。

（4）随机时间激发同时源采集带来的问题是多炮数据相干涉（串扰 Cross-Talk），当前的建模与成像方法不能直接用这样的数据，必须分成独立的单炮数据，即必须进行 Deblending 处理。

（5）Deblending 必然依赖于 Belending。解混叠必然要对混叠编码方式提出要求。当前主要体现在：随机激发时间编码方式；可控震源利用频率、振幅和相位进行编码。

我认为应该把成像最佳作为目标来提出数据采集的编码方式。但是，目前连以 Deblending 的结果作为 Blending 采集方式设计标准都做不到。这其中，信噪比不能把握是根本问题。

（6）Deblending 用伪解码数据的去噪作为基本的方法已成为共识。但强振幅的初至波、面波、面波散射波会对这样的方法产生较大影响，事先必须把初至波、面波等进行较好的消除。

（7）Deblending 严重依赖数据的信噪比，由于信噪比过低，在反射同相轴/散射同相轴不显著的数据上进行 Deblending 是不会成功的。

（8）过多的混叠炮数也会对 Deblending 带来挑战，尤其在混叠编码方式不是很随机的情况下更是如此。

（9）Deblending 归为高维信号预测下的去噪问题，需在基函数选择上下功夫，高维小波基、字典基等需下功夫磨合；非线性与线性信号预测器需相互结合。

（10）绕射/散射波的保护是 Deblending 中的关键问题。

（11）最好是发展混叠数据的直接成像。这样的直接成像一定是迭代的反演成像。

3 关于 2D 中心 Wiener 滤波器设计的思想

① 1D AR 模型 Wiener 滤波（等于 F-X 域 2D 滤波）的再审视

已知$\{x_1,x_2,\cdots,x_{N-1}\}$个数据点（可以是频率域的数据），它们应被视为不同点处的随机变量，由它们张成$N-1$维的子空间，它们被认为是基函数。基函数线性组合后的结果为：

$$\tilde{x}_N=\sum_{j=1}^{N-1}\alpha_j x_j+\eta$$

上式也可写成

$$\sum_{j=1}^{N}\beta_j x_j=\eta \qquad \beta=(-\alpha_1,-\alpha_2,\cdots-\alpha_{N-1},1)$$

利用已知数据$\{x_1,x_2,\cdots,x_{N-1}\}$，估计最佳$\boldsymbol{\alpha}$（滤波器、加权系数、回归系数）的方法：均方误差最小化。即

$$E\left\{\|\eta\|^2\right\}=E\left\{\left\|d-\sum_{j=1}^{N-1}\alpha_j x_j\right\|^2\right\}$$

据此导出的法方程为：

$$\mathbf{x}^H\cdot\mathbf{x}\boldsymbol{\alpha}=\mathbf{x}\cdot d \qquad \text{（法方程）}$$

$$\text{即 } R_{xx}\boldsymbol{\alpha}=R_{xd}$$

上述公式描述了 F-X 域的 2D 最佳滤波。可按前向进行，同时也可按后向进行，因为很多情况下，所有数据已观测到。对插值/去噪这类操作，情况就是这样。 前后向预测结果进行平均，应得到更好的预测结果。

② 关于 F-X 域预测去噪的空间不均匀采样的影响问题

F-X 域的数据可以表示为：

$$\begin{aligned}x_1 &= Se^{i\omega p_x x_1}\\ x_2 &= Se^{i\omega p_x x_2}\\ &\vdots\\ x_{N-1} &= Se^{i\omega p_x x_{N-1}}\end{aligned}$$

上式表示平面波源S沿射线参数p_x规定的方向传播。可见，相位（$\omega p_x x$）与位置是有关的，默认是均匀采样的。一旦出现如下图（图 1）所示的非均匀采样，若默认成均匀采样，相位就错，预测效果变差。去噪之前进行数据规则化是必须的，或插值与去噪同时进行。

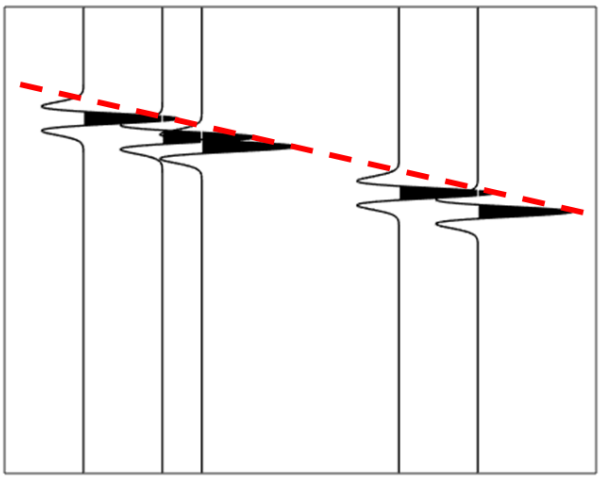


图 1 空间不规则采样的示意图

③ 如何把 1D F-X 域最佳预测滤波推广到 2D？

1D Wiener 滤波算子是一个 1D 平滑算子（加权算子），作用在数据$\{x_1, x_2, \cdots, x_{N-1}\}$上，是褶积算子（一维褶积运算）。这种褶积运算转化成矩阵方程，在最小二乘意义下得到法方程，估计最佳预测系数。

2D Wiener 中心滤波算子是一个 2D 平滑算子（加权算子），作用在 2D 数据上，是 2D 褶积运算。这种褶积运算转化成矩阵运算形不成矩阵方程(应是张量运算)，这正是 AR 模型在高维数据中应用的困难所在！

对于$\omega - x - y$域中单频数据中，中心 Wiener 加权滤波是对中心周围点进行加权处理，如下图所示：

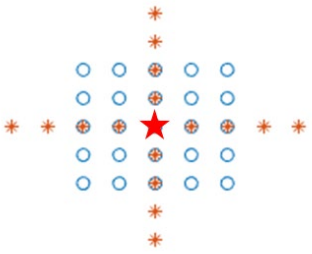


图 2 2D Wiener 中心滤波算子（褶积算子）的花样示意图

至于中心点周围哪些点参与加权，应可选不同的 stencil（花样）。应该有不同的形式，但最好是对称的褶积算子。

同样存在不规则采样对预测结果的影响问题，至少应先把数据投在规则点上，这是个合理的做法。可以用 Sinc 函数进行投影处理。

④ 2D Wiener 中心滤波算子的构造

2D 褶积算子与数据的作用方式为：

$$x_{i,j} = \sum_{k=i-M}^{i+M} \sum_{l=j-N}^{j+N} \alpha_{k,l} x_{k,l}$$

$x_{i,j}$ 是一个预测点，$\alpha_{k,l}$ 是一个向量呢？或是一个矩阵呢？$x_{k,l}$ 是排成一个向量呢？或是排成一个矩阵呢？

很显然，直观的看，就按 Hankel 矩阵形式排数据，$\alpha_{k,l}$ 对应地写成向量就行了！这是最令人放心的做法，这样形成的法方程一定保持数据中的相位线性特征。

4 DTW 方法的基本思想分析

目的是开展对 DTW（Dynamic Time Warping）基本思想的了解与批判，从中看能否提出适用于地震波场匹配的思想方法。

Y(t)						
1	0	0	1	2	1	-1
2	-1	-1	0	1	0	-2
3	-2	-2	-1	0	-1	-3
2	-1	-1	0	1	0	-2
1	0	0	1	2	1	-1
1	0	0	1	2	1	-1
0	1	1	2	3	2	0
	1	**1**	**2**	**3**	**2**	**0** **X(t)**

图 3 两条离散曲线形成的插值矩阵

对两个曲线 $x(t)$ ， $y(t)$ 进行离散表示。单个离散点组合相减，形成图 3 所示的差值矩阵。

不一定是方阵，因为两曲线点离散点数不一定相等！

形成这种差值矩阵的基本思想是什么？两段相似的波形会在矩阵中形成（在对角线附近或对角线平行线附近）会形成差值低谷。

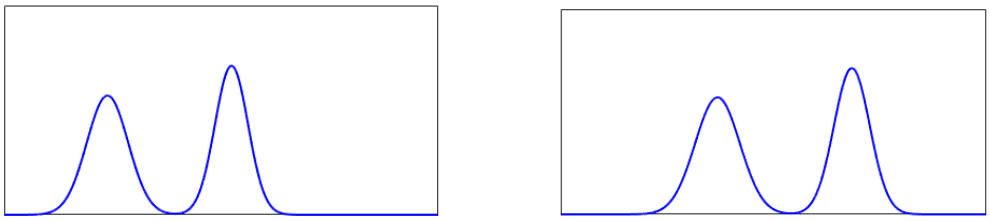


图 4 两个具有相同波峰的曲线的示意图

为什么会这样呢？因为对角线反映的是无时间移动的差值，而与对角线平行的是有时间移动的差值，譬如图 4 两曲线的离散点对应相减。

带状区域内的低谷反映了两个波形的差值。

构造这样一个矩阵的确很有意思！但自动地从这个矩阵中找出这个低谷带最低点的连线就相当于找到了两个要匹配的信号的时移量，从而完成两个时移信号的对齐。低谷带最低点的连线与对角线的差就是要匹配的信号的时移量。

但感觉这样的做法存在很多问题，譬如：

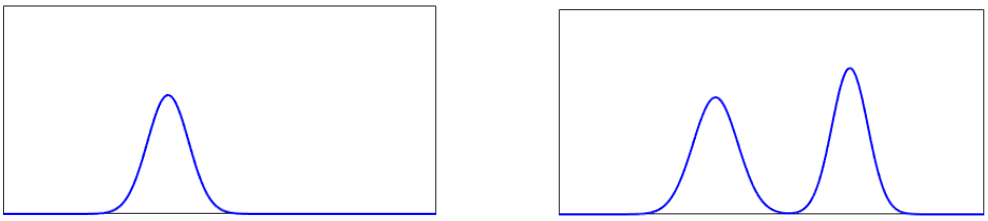


图 5 两个具有不同波峰的曲线的示意图

这两个信号相减形成矩阵，相对于对角线必然形成两个低谷带。测出的时移量有两个，到底向谁靠齐呢？

X(t)与 Y(t)匹配施加的要求：

1. 两个物理信号由类似的物理原因形成，其中包含的物理事件的个数相同。
2. 两个物理信号对应事件的起点和终点是明确的，二者的起点和终点可以对应起来，尽管二者采样产生的离散点数不同（至少可以不同）。
3. 两个物理信号对应或包含的物理事件的顺序不能人为地调换。
4. 两个物理信号的幅值应在相同的水平上，因为构成矩阵的方式是用二者离散值的差。这与相似系数和两信号幅值无关很不相同（相似系数计算中引入了归一化处理，若仅用互相关或协方差函数体现二者的相似性也与两信号的幅值有关！）。因此，生成上述差值矩阵应事先对两信号做归一化处理。

据此，把两信号排齐的问题，或寻找局部“波形”时差的问题转化为对差值矩阵的分析问题，即找到对角线或平行对角线附近的差值洼地最低点的连线！

应该有很多种做法，譬如转化为图像处理问题、变分问题、数值优化问题。把差值场转化成慢度场，进行射线追踪，找到山谷的或洼地的最低点连线是很自然的做法，相当于解一个变分问题，完全搬照射线追踪即可。

依然存在的问题是：

洼地最低点的连线是否代表了时移，或真正想要的时移？

这个问题与矩阵（差值矩阵）的构造方式是相关的，提出什么样的方式来构造矩阵？

目前的做法本质上是把对两曲线相似性评价问题转换成差值矩阵，通过对差值矩阵的“搜索”找到两曲线

点与点之间，本质上是“邻域曲线”之间的时移，然后进行排齐。

目前的做法存在很多问题，首先差值矩阵的构造方式存在问题，至少不应该仅是点对点求差，这样做不能敏锐地体现两“邻域曲线”的真正差异。

应提出描述“邻域曲线”属性的参数，定义抽象距离，再形成矩阵。

由差值矩阵找到对应的时移量，其中的基本思想需要考察！若把差值洼地最低点的连线作为寻优目标，首先应规定这根连线的起点位置。事实上，这个起点位置的确定应是不唯一的。必须引入先验信息。

若规定 A 曲线与 B 曲线的起点是已知的，且起点是相对应的，则差值矩阵的左下角应是连线的起点。

但这样的要求太严格。我个人认为应该由差值矩阵的上边或右边上的点出发，基本平行于对角线进行寻优处理。射线追踪依然是好的变分问题提法。至于由上边或右边上的点出发，于差值矩阵的生成方式有关。$x(t)-y(t)$ 或 $y(t)-x(t)$。差值不要分正负（取绝对值）、矩阵仅需要上三角或下三角。

由参考信号起点作为连线起点是合理的（否则也不叫参考信号了！）。但很多时候，我们进行任意两曲线的之间的排齐（匹配），起点该如何选择？

对地震信号而言，我觉得用任意包含子波的信号作为参考信号都可以，但应从差值矩阵的上边或右边上的点作为起点，向左下角方向进行射线追踪，可以进行信号的匹配或排齐。

总之，射线的起点是个很根本的问题，与时差测量的精度有重要关系。连起点都不不能确定下来，时差估计无从谈起。

5 关于信号的平稳性、高斯性以及非线性反演之间的关系

5.1 随机信号的平稳性

从信号的角度看，随机信号的所有阶次的统计量不随时间变化称严平稳。仅仅二阶以下统计量不随时间变化称广义平稳。

从系统的角度看，时（空）不变系统产生平稳信号。理想弹性介质和理想声介质构成时（空）不变系统。实际地下介质当然不符合这样的假设。

描述平稳信号可以使用 Fourier 基函数，它没有时（空）局部性。

相应地，有非平稳性的概念。严格地看，地震信号是非平稳的。但当地下介质粘（滞）弹性不很强时，用平稳信号的假设进行信号处理是可行的。事实上，常规地震数据处理中基本上都忽视了信号的非平稳性。精细的地震数据成像处理需要考虑介质的非理想弹性引起的信号的非平稳性。

5.2 线性性与高斯性

线性性一般指系统满足线性性，譬如褶积系统。所谓线性性是指系统满足加与和运算规则。

高斯性是指随机信号的概率分布函数是高斯的。

5.3 从最小二乘反问题理解线性性与高斯性之间的关系

L2 范数下定义的最小二乘反问题的误差泛函，从 Bayes 估计的观点看，满足高斯分布。L2 范数下的最小二乘反问题内含一个基本假设：数据中包含的信号是可以有一个模型进行预测的，预测误差满足高斯分布。另一个方面，在均方误差最小的意义下，反问题法方程的求解需要自相关函数，自相关函数是二阶统计量，自相关函数的 SVD 分解得到系统的线性表达。据此，高斯性要求与线性性要求联系在一起。

另外，系统的线性性与最小二乘反演问题的凸性是紧密相连的，此时 L2 范数误差泛函是一个二次型泛函，仅仅存在一个极小点。

另一方面，在系统线性性和误差高斯分布的假设下，后验概率密度函数也是高斯的。期望就是最小二乘反问题的估计结果，方差反映该估计结果的精度。

目前，我们遇到的绝大部分问题，都是在系统线性性和预测误差高斯分布假设下，用最小二乘方法进行求解的。譬如 Wiener 滤波/反滤波；FWI/LS_RTM/Tomography 等等。但是，这样的假设与实际情况的差异越来越不可忽视。系统线性性和误差高斯分布假设下的基本的反问题尽管其理论框架十分美妙，但在实际勘探地震反问题研究和应用方面，似乎已经走到了尽头。

5.4 系统非线性、误差非高斯、信号非平稳对反问题的影响

系统非线性首先导致 L2 误差泛函的非二次型，基于此误差泛函的反问题非凸。随着系统非线性程度的加大，误差泛函的极值点会非常多，梯度类反演方法无法得到有意义的解，只能用 Monte Carlo 之类的方法，即便如此，非线性性过强时，Monte Carlo 类方法也无法给出可行的反演解。针对这样的问题，目前基本没有办法。系统的非线性性强从另一个角度看就是系统参数（也包括输入）与输出之间的关系非常弱。

非高斯性是指随机信号的统计量不能仅仅由二阶统计量决定，必须引入高阶统计量。三阶统计量反映随机信号的概率密度函数偏离对称性的程度；四阶统计量反映随机信号的概率密度函数尖锐程度。随机信号的广义高斯分布函数一定程度上反映了非高斯性。在反问题构建中应该开始利用后验概率密度函数的广义高斯分布特点。

随机信号的非高斯分布一定与系统的非线性性有关，系统不再可用相互独立（正交）的特征向量进行线性叠加来描述，特征基函数之间是相互关联的。但是，目前还不知道非高斯性与非线性性之间如何进行定量描述（至少我还没有看到）。

信号的非平稳本质上也与系统的非线性性有关。频散介质（吸收衰减介质）系统就是一个非线性系统。根据频散介质中的波传播结果进行反演，涉及到了波场的非平稳性和系统的非线性性。另一方面，非平稳信号分析需要引入时频分析和小波变换。也许当要提出基于信号时频特征的反演方法时，需要借助于非平稳信号分析方法。但是，目前依然不清楚这是否是合适的方法技术路线。不知道即便这样做了，参数反演的精度是否会有本质上的提高。

在 L2 范数下的反演研究已经比较成熟的情况下，Lp(1<p<2)情况下的反演是必需要走的路线。但是，无论如何（无论到何时），反问题总是在信息不充分性情况下在寻找最优解。精度的提高永远没有止境！

6 子波相位及对反演成像问题影响的思考

6.1 地震子波（带限子波）的相位问题

单频谐波的相位$\phi(t)=\omega t+\phi_0$，ωt代表传播相位；ϕ_0代表初相位。地震子波相位的含义很清楚，它有两部分构成！一般地，同相叠加更关心传播相位，它由（各向异性）速度决定。信号处理更关心初相位，它决定子波的形态，影响去噪、精细叠加等信号处理效果。一般地，实际处理中要对地震子波进行零相位化处理。在偏移成像+AVA分析为技术核心的时代，对子波进行零相位化处理是没有问题的。但是，在地震波反演成像为主的时代，子波的相位情况未知，增加了反演结果的不准确性。

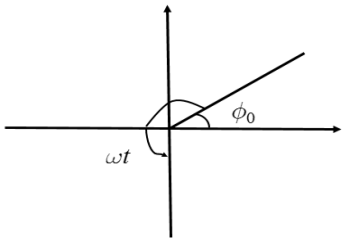


图6 子波相位变化示意图，$0\le\phi(t)\le 2\pi$主区间

6.2 地震子波的零相位化方法

1、提取子波的振幅谱，把子波的相位谱置成零

原子波：$W(\omega)=A_{(\omega)}e^{i\phi(\omega)}$

零相位化处理：

子波：$W(\omega)=A_{(\omega)}e^{i\phi(\omega)(=0)}$

2、取子波的振幅谱和相位谱$A(\omega)e^{i\phi(\omega)}$，对$\phi(\omega)$进行相位旋转，即：$\phi(\omega)\phi'(\omega)$（$\phi(\omega)$为常相位移）。然后检测子波的斜度/偏态，或峭度/峰态的检测。旋转到斜度/偏态（峭度/峰态）达到满意为止。

但最终结果并非真正的零相位子波，因为原子波不是零相位子波经常相位移变换得到的，上述逆常相位移处理并不能恢复出一个零相位子波。

地震子波可表达为如下谐波叠加：

$$W_{(\omega)}=\sum_{j=1}^{N\omega}A_{(\omega_j)}e^{i\phi(\omega_j)}$$
$$\phi(\omega_j)=\omega_j t+\phi_0(\omega_j)$$

当$\phi_0(\omega_j)$同相位时，上式定义一个零相位子波，否则，会给出一个混合相位子波。$W(\omega)e^{i\phi(\omega)(=c)}$显然不会让一个混合相位子波变成零相位子波。

3、Wiener 滤波整形成近似零相位子波。

可与子波整形一同完成！取参考零相位子波，可用 2.1 中方法生成，用 Wiener 滤波方法生成近似零相位子波。

6.3 子波的相位对后处理的影响问题

6.3.1 子波的最小相位化

为什么要改变原子波的相位特征？尤其是把原子波相位转成最小相位的？明显的目的是为了后续的反褶积方法，使得数据更满足反褶积方法的最小相位要求，否则，子波的最小相位化处理应找到更合适的理由才能施用。这在反演成像阶段是要特别注意的。

6.3.2 子波的零相位化处理

子波零相位化处理首先不会对 Migration 和速度估计产生重要影响，而且可改善成像叠加剖面的信噪比等。对 AVA 分析而言，子波零相位化处理应是十分必要的步骤。

6.3.3 修改子波相位对反演成像的影响

子波修改相位对参数估计方法，即对于 Bayes 反演方法（最小均方误差方法和最小二乘方法）有重要影响。原则上，不可以随意改变原子波的相位特征。

把原子波和正演子波都进行零相位化，尽管消除子波相位的未知，一定会改变反演结果。但到底对反演产生什么影响，好的或是坏的影响，还需要进一步分析。

6.4 1D 反射系数反褶积和高维反射系数反褶积的对比

关键出在正问题描述上，1D 反褶积建立在 1D 褶积模型基础上。高维反射系数反褶积并无对应的模型，尤其高维空间反射系数一定要考虑角度反射系数。

零角度高维反射系数反褶积（褶积）等价于 1D 反褶积（褶积）问题。

1D 褶积 Model/高维褶积 Model/波动理论层状介质反射波 Model/AVA Model，都在试图从一个方面描述层状介质反射波理论，前两个是线性化的，后两个是非线性的（当然也可以线性化！，实用中都是线性化后的模型）。

与$y_{(t)}=W_{(t)}*R_{(t)}$类似的高维褶积 Model 在勘探地震中并不存在，或并不严格存在，尽管可以抽象地写出一个来：

$$y(\mathbf{x}_s,\mathbf{x}_r;t)=h(\mathbf{x}_s,\mathbf{x}_r;\mathbf{x};t)*R(\mathbf{x},\theta,\phi)$$

实质上，$y(\mathbf{x}_s,\mathbf{x}_r;t)$是层状介质中线性化反射波场的表达式！也许我们不能执着于把 1D 褶积 Model 直接地推广到高维。但直接地推到高维应能帮我们更好地理解层状介质中线性反射波场所传达出的物理意义！

6.5 子波的相位特征对反演的影响

为什么一定要求子波的相位特征是最小相位的？

反褶积处理时仅仅已知$y(t)$，并不知道系统$h(t)$（即子波）的相位特征，因此，反褶积算子$\mathbf{a}$的估计过程中，R_{hh}的奇异性正是由于子波的相位未知而导致的，非奇异性是由于子波的最小相位性保持的。

$$\begin{aligned}\sum_{\mathbf{a}}\min &= \left|\mathbf{y}(t)-\mathbf{d}(t)\right|^2 \\ &= \left|h(t)*\mathbf{a}-\mathbf{d}(t)\right|^2 \\ R_{hh}\mathbf{a} &= r_{hd}\end{aligned}$$

对于地球物理参数反演问题也是一样的！不知道子波的相位特征，导致实测数据中反射子波与合成数据中反射子波不能完全相减，解的分辨率或精度降低。

6.6 高维反褶积公式与 1D 反褶积公式的对比

$$\begin{cases}1D & h(t)*R(t)=y(t) \\ \text{高维} & h(\mathbf{x}_s,\mathbf{x}_r;\mathbf{x};t)*R(\mathbf{x})=y(\mathbf{x}_s,\mathbf{x}_r;t)\end{cases}$$

1D 褶积公式中波传播因素的处理存在如下假设：

a、背景介质光滑，不存在透射损失；

b、弱扰动，即$R(t)$比较小，不引起多次反射；

c、水平层状介质，且形成的$R(t)$是白噪序列，即功率谱为常数；

d、理想声/弹介质，不存在介质频散现象，传播效应为纯相位因子$e^{i\omega\int\frac{zdz}{v(z)}}$所包含；又体现在$R_{(z)}\Rightarrow R(t)$中，因此，似乎不存在传播效应，实质上是有的，是层状介质中的一维波传播过程的线性化，正是这样传播结果等价于$y(t)=h(t)*R(t)$所描述的。

高维褶积公式中存在的假设：

a、背景介质光滑，不存在透射损失；

b、介质层状展布，不存在间断点；

c、介质弱扰动，即 $R(\mathbf{x})$ 比较小，不引起多起反射波；

d、理想声/弹介质，不存在介质频散或不平稳情形。

此时，传播效应不能由纯相位因子 $e^{i\omega\int\frac{2dz}{v(z)}}$ 来描述，这个纯相位因子本质上对应平面波的相位移。而是由球面波传播，及采样因素决定。主要是必须考虑球面波振幅衰减因素。振幅因素主要是球面扩散（几何扩散）（也可认为球面扩散也是几何扩散的一种）和方向因素（不同传播方向，振幅不同）决定。理论上讲，波场采样应是连续的，离散采样会影响波场振幅值。上述因素都可以体现在 $h(\mathbf{x}_s,\mathbf{x}_r;\mathbf{x};t)$（系统响应算子或点响应函数）中，用 Green 函数来表示，子波也可融入 Green 函数中。本质上，$h(\mathbf{x}_s,\mathbf{x}_r;\mathbf{x};t)$ 也表现为空间子波的情形，即每个角度/每个空间点都存在一个所谓的空间子波，近似地从逻辑上讲，高维反褶积应是该共角度空间褶积算子的逆。

目前，所谓的 LS_PSDM 估计 $R_{(\mathbf{x},\theta)}$，或真振幅成像估计 $R_{(\mathbf{x},\theta)}$，逻辑上应该是消除几何扩散（球面扩散）、方向因素及采样因素后，入射波与反射波（局部平面波）的比值，而且是振幅的比值。求比值这一步应该是子波相位对 $R_{(\mathbf{x},\theta)}$ 估计结果产生影响的时候，否则，所谓振幅比也没明确的含义。因此，目前的真振幅成像忽视了地震子波相位的影响，这是值得关注的地方！应该可以提出一种更精确的也更实用的 $R_{(\mathbf{x},\theta)}$ 估计方法。

1D 反褶积/褶积利用了系统的线性性（褶积 Model 是线性的！）。平稳性是指系统时不变性，反射子波不随时间变，即褶积系统是平稳的。因果系统的含义是无输入前不产生输出（最小相位系统与之对应）。高斯性是指对期望反射信号的预测误差是高斯的，也指存在反褶积算子使得反褶积结果是高斯白噪。

目前，这些假设无论在信号分析或是反演成像中都需要进一步修改。非线性系统应逐步开始考虑(压缩感知是典型的应用)；非平稳性也变得很平凡了；非高斯性也在逐渐普及。

专注于三个“非”，会产生新的反演成像思想和方法。我们应在这方面提出有意义的反演成像方法与技术。

7 信号与图像的稀疏表达与地震波反演成像

信号（高维信号=图像（包括视频））存在稀疏特征（尤其在变换域），引入实测信号 $\mathbf{y}$ 与原信号 $\mathbf{x}$ 之间的如下关系

$$\mathbf{y}=H\mathbf{x}+\boldsymbol{\varepsilon}=H\phi\boldsymbol{\alpha}+\boldsymbol{\varepsilon}$$

H 是系统模糊算子（点响应函数，其中也包括采样算子），$\boldsymbol{\varepsilon}$ 是加性高斯噪音 $N(0,\sigma^2)$（一般期望是高斯白噪音）。

针对 $\mathbf{y}$ 的典型的处理问题包括：

$$\begin{cases} a.Denoising \\ b.\text{高维数据规则化（插值）} \\ c.Deconvolution(Inversion) \\ d.Deblending(\mathbf{y}=H(\mathbf{x}_1+\mathbf{x}_2+...+\mathbf{x}_s)) \\ e.\text{数据压缩} \end{cases}$$

根本性的问题是 Inversion 问题，即估计 $\mathbf{x}$ 。需要关于 $\boldsymbol{\varepsilon}$ 或 $\mathbf{y}$ 的统计特征（假设已知，譬如高斯加性噪音）的先验信息。$\boldsymbol{\varepsilon}$ 或 $\mathbf{y}$ 的高斯统计假设本质上也蕴含着正问题可以模拟 $\mathbf{y}$ 中包含的信号，也蕴含着在该正问题对应的初始点处可能把正问题进行很好的线性化。当然也需要关于 $\mathbf{x}$ 的先验假设。

另一方面，希望 $\mathbf{x}$ 在变换域具有稀疏性是一种很根本的假设，某种程度上，稀疏性是信号（或图像）具有的本质特征。这种 $\mathbf{x}$ 具有的本质特征可以，也应该作为先验信息加以利用。

Bayes 框架下，不利用稀疏特征假设的 $\mathbf{x}$ 的估计

$$\begin{cases} \mathbf{x}=(H^TH)^{-1}H^T\mathbf{y} \\ \mathbf{x}=(H^TH+\lambda I)^{-1}H^T\mathbf{y} \end{cases}$$

这种基于 Tikhonov 正则化思想的反演方法已研究得很成熟，具有一套优美的算法。

若利用信号的稀疏特征，首先应选择一簇基函数（或若干簇基函数即字典基），把 $\mathbf{x}$ 投影到选定的基上。若投影系数具有快速衰减的性质，称信号在变换域中具有稀疏性。应该注意选基函数本身也代表引入了对 $\mathbf{x}$ 已有的先验认识。这些不同的基函数能刻画不同特征的信号，一种基函数对 $\mathbf{x}$ 的刻画比较稀疏，换一种基函数表达 $\mathbf{x}$ 就未必那么稀疏。

本质上，当然应自适应地依据数据来选择基函数。即便如此，自适应地选基函数的算法也隐含地引入了对 $\mathbf{x}$ 的先验认识，譬如用自相关函数的 SVD 分解就选择了线性基函数簇。

《Sparse Image & Signal Processing》一书花好几章讨论小波基以及不同小波基的适用性（对不同图像，特征抽取能力不同）。在应用环节，最关键的就是抽象出了几种不同的正则化约束反演模型。譬如

$$\begin{cases} MAP\text{问题} \\ (P_\lambda):\min\limits_{\boldsymbol{\alpha}\in R^T}\dfrac{1}{2}\left\|\mathbf{y}\text{-}H\boldsymbol{\phi}\boldsymbol{\alpha}\right\|^2+\lambda\psi(\boldsymbol{\alpha}) & \text{(A)} \\ \text{不等式约束问题} \\ (P_\sigma^q):\min\limits_{\boldsymbol{\alpha}\in R^T}\psi(\boldsymbol{\alpha})\ \text{s.t.}\ \left\|\mathbf{y}\text{-}H\boldsymbol{\phi}\boldsymbol{\alpha}\right\|_q\le\sigma & \text{(B)} \\ Dantzig\text{选择器问题} \\ (P_{DS}):\min\limits_{\boldsymbol{\alpha}\in R^T}\psi(\bar{\alpha})\ \text{s.t.}\ \left\|F^*(\mathbf{y}\text{-}F\boldsymbol{\alpha})\right\|_\infty\le\tau \end{cases}$$

(A) 式和(B)式还可提成：

$$(\alpha_A):\min_{\mathbf{x}\in R^N}\frac{1}{2}\|\mathbf{y}-H\mathbf{x}\|^2+\lambda\psi(\phi^*\mathbf{x})$$

$$(\alpha_\sigma^q):\min_{\mathbf{x}\in R^N}\frac{1}{2}\psi(\phi^*\mathbf{x})\ s.t.\ \ \|\mathbf{y}-H\mathbf{x}\|_q\le\sigma$$

还有非线性问题：

$$\mathbf{y}=\aleph(\mathbf{x})+\varepsilon$$

$$(P_\lambda):\min_{\boldsymbol{\alpha}\in R^T}\frac{1}{2}\|\mathbf{y}-\aleph\boldsymbol{\phi}\boldsymbol{\alpha}\|^2+\lambda\psi(\boldsymbol{\alpha})$$

接下来的事情是上述几个反演模型的求解。

$$\begin{cases} i.\text{LASSO算法是比较经典的算法；}\\ ii.\text{用}\mathrm{L}_1\text{代替}\mathrm{L}_0\text{的近似算法是常用的。可视为：一般LS梯度项+ 梯度滤波项；}\\ iii.\text{阈值方法（软阈值+硬阈值）是常用的。}\end{cases}$$

算法方面还需再总结。

至此，稀疏信号与图像处理基本如此了！关键是在不同的应用场景下，用好合适的基函数（发展出独特的基函数。当然很有创新性!!!），找到快速、有效的解法。

对比地震波成像中的参数估计问题

$$\mathbf{d}=F(\mathbf{m})+\boldsymbol{\varepsilon}\ \text{或}\ \delta\mathbf{d}=F\delta\mathbf{m}+\varepsilon$$

二者似乎形式上相同，但差异也是明显的。$\mathbf{d}$ 或 $\delta\mathbf{d}$ 为观测波场（扰动），$\mathbf{m}$ 或 $\delta\mathbf{m}$ 为参数（扰动）。数据和参数间的联系由波传播算子建立。

信号（图像）处理中，$\mathbf{y}$ 与 $\mathbf{x}$ 的物理意义相同，二者的联系由点扩散函数建立。地震波成像中的参数估计问题比信号和图像处理更复杂。图像的稀疏表达及应用更多地用来预处理 $\delta\mathbf{d}(\text{或者}\mathbf{d})$。当然也可以稀疏表达 $\delta\mathbf{m}$。

仅稀疏表达 $\delta\mathbf{m}$，本质上是对 $\delta\mathbf{m}$ 的正则化处理。已在地震波反演中广泛应用。但具体计算的有效性仍有待考察。基函数的选择也还值得审视。

$\delta\mathbf{d}(\text{或者}\mathbf{d})$ 的稀疏特征表达，然后基于此表达的成像也在不断受审视。单频波、平面波、局部平面波、编码波场、包络波场、到达时都是在基于特征波场进行成像。但到目前为止应用效果并不十分显著。

总之在对 $\delta\mathbf{m}(\mathbf{m})$ 和 $\delta\mathbf{d}(\mathbf{d})$ 的特征表达的观点下构建地震波反演方法一定是一个代表性的方向，然后进入多次波成像更是已成为当务之急的事。

FWI 之类黑匣子成像方法必须放在 $\delta\mathbf{m}(\mathbf{m})$ 和 $\delta\mathbf{d}(\mathbf{d})$ 特征表达的角度下审视才能导引出更有效的方法。

可以看出，信号（图像）处理与波场的反演成像无论从何种程度上、从任何角度上看，显然已融为一体。

8 从系统辨识角度看待 FWI 反演

Bayes 框架下的 FWI 反演，在关于数据预测误差和先验模型参数的分布符合高斯分布时，由 MAP 准则过渡到代价泛函最小化的优化问题。这种反演逻辑当然是不会错的。但仍然没有突显出正问题的选择对反演过程的影响。

从系统辨识的角度看反问题能清楚地认识到人为构造系统（正问题）是如何影响反问题的。

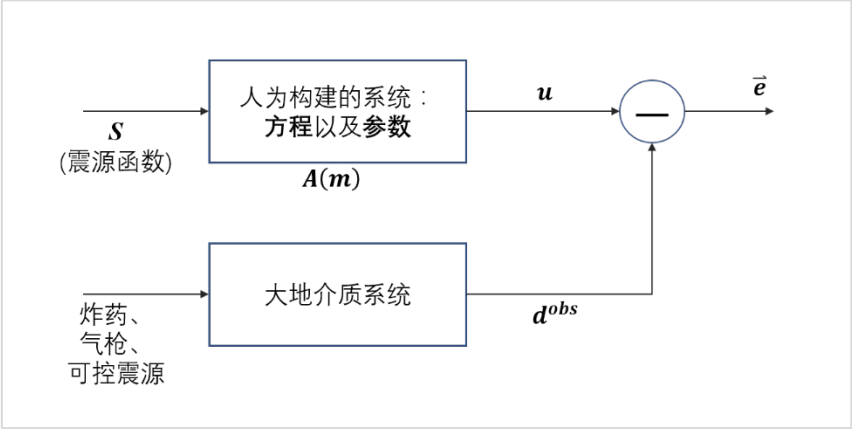


图 7 系统辨识理解 FWI 的示意图

在地震勘探中，引入状态变量和系统参数来描述正问题，如下式：

$$A(\mathbf{m})\mathbf{u} = -\delta(\mathbf{x}-\mathbf{x}_s)\mathbf{f} = -\mathbf{f}(\mathbf{x}_s,t) \tag{1}$$

其中 $A(\mathbf{m})$ 为正演算子，$\mathbf{u}$ 是正演的波场，$\mathbf{f}(\mathbf{x}_s,t)$ 是震源函数。上式可以简写成：

$$A(\mathbf{m})\mathbf{u} = \mathbf{f}$$

实际野外观测的数据可以写成：

$$\mathbf{d} = \mathbf{u}_{pred} + \mathbf{u}_{unored} + \boldsymbol{\eta} \tag{2}$$

其中，$\boldsymbol{\eta}$ 表示噪音函数，在大部分情况下假设它是高斯白噪的。$\mathbf{u}_{pred}$ 为可被系统状态变量预测的部分，$\mathbf{u}_{unpred}$ 为不可以被系统状态变量预测的部分。Tarantola（2005）曾经说正问题不能很好地解释实测地震数据引入的数据残差 $\mathbf{u}_{unpred}$ 与观测噪音引入的数据残差 $\boldsymbol{\eta}$ 可能是同一水平的。$\mathbf{u}_{unpred}$ 必须通过数据预处理方式消除掉。

把状态变量独立出来，我们认为在反问题的框架下不仅要逼近实测的数据，而且要正演结果（状态变量）满足控制方程（波动方程）。从这一角度可以看到，其本质是在要使得我们选择的正演过程更加符合实际的波传播过程。新的反问题可以写成：

$$\min_{\mathbf{m}\in\mathbb{M},\mathbf{u}\in\mathbb{U}} \frac{1}{2}\left\| P\mathbf{u}-d^{\text{obs}}\right\|_2^2 \quad s.t. A(\mathbf{m})\mathbf{u}=\mathbf{f} \tag{3}$$

这是 Hermann（2015）引入的反问题观点。要求逼近实测数据的正演波场必须满足一定的波动方程。我感到这样的逻辑很别扭。（2）式已经清楚地显示出，实测波场中一定存在不可预测的波现象，无论选择什么样的控制方程都是如此。另外 $\mathbf{u}$ 和参数 $\mathbf{m}$ 之间不是独立的，有非常强的耦合性。控制方程把它们耦合一起。同时反演这样两个参数，逻辑上站不住脚。

（3）式提法真正的意义在于提示我们要同时反演参数 A，$\mathbf{m}$ 和 $\mathbf{f}$，反问题才是更合理的。但是，我认为在勘探地震中，同时反演这三个量是不可能的。尤其是反演 A 是不可能的。也许在信号和图像处理领域中，反演 A 是可以的。我个人很难相信在勘探地震中同时反演 Green 函数与 $\mathbf{m}$ 和 $\mathbf{f}$。

我认为 Hermann（2015）的观点不可取！

我提出 CWI 的逻辑是，仅仅用包含某种可预测信号的数据进行反演。正问题就限定为描述这部分可预测信号的正问题。物理逻辑是十分清晰的。

Symes 教授领导的研究组的基本思想是清楚的。就是假设数据中仅包含可预测的信号及高斯噪音，在像域中进行背景速度的估计。这就是所谓的 DSO 的基本出发点。同时，他们把 DSO 与 FWI 用 Lagrange 乘子法结合在一起。

$$J_\lambda\left[\mathbf{g},\mathbf{m}\right]=\frac{1}{2}\left\| PA(\mathbf{m})^{-1}\mathbf{g}-d^{\text{obs}}\right\|_2^2+\frac{\lambda^2}{2}\left\|\frac{\partial\mathbf{m}}{\partial\theta}\right\|^2 \tag{4}$$

通过调节 Lagrange 乘子法可以先反演背景，再逐渐过渡到 FWI。等于把角度道集的层析与 FWI 结合在一起。

但是，这中间还有一个因素没有用上，就是震源。针对主动震源的情况，反问题可以提为，

$$J_\lambda\left[\mathbf{g},\mathbf{m}\right]=\frac{1}{2}\left\| PA(\mathbf{m})^{-1}\mathbf{g}-d^{\text{obs}}\right\|_2^2+\frac{\lambda^2}{2}\left\|\mathbf{g}-\mathbf{f}\right\| \tag{5}$$

动力学意义上的震源函数的反演，这样的提法是不合适的。因为真震源 $\mathbf{f}$ 是未知的。但是真震源 $\mathbf{f}$ 的空间位置是已知的，可以用（5）式的提法进行反演。如果仅仅进行震源函数的空间位置的约束反演，当 λ 比较大时，这是利用震源位置的层析反演。可以通过调节 λ，逐步过渡到 FWI 反演。逻辑上是可行的。

但是，这样的反问题多大程度上变成了凸的问题，还是很难说，尤其对高波数成分的反演而言。应该说，背景速度反演时，问题的凸性得到了改善。（5）式用透射波进行背景速度估计时，到达时（相位）的正确，再加上震源空间位置的正确，的确可以使反问题的凸性变好。（5）式用反射波进行反演时，同样地有上述结论。

可以看出，Symes 教授提出（4）和（5）式进行地震波反演，物理直觉更好，符合逻辑，更有可实现性。

事实上，Symes 教授的观点可扩展的地方在于，可以不区分一次震源和二次震源（像），提出融合一次震源和二次震源的反演公式，如果把二次震源体现的地质结构信息再融于反演过程中，将是更为完整的实现方案。所有源点的空间关系正确时，背景速度就是正确的。在背景速度正确时，逐步转入 FWI。

如果还是达不到 LSRTM 对背景速度的精度要求，我建议只能用 DTW 等方法来硬性消除残余的时差（速度误差），进行 LSRTM 的残差计算。

9 线性系统反问题的观点

地球物理反问题（勘探地震反问题，也包括信号/图像分析中涉及到的反问题）框架，存在两种不同的观点，一种是 Bayes 估计理论框架，一种是线性系统求逆（伪逆）的理论框架。Bayes 估计理论框架抽象程度很高，我非常喜欢在此框架下说明反问题，原则上也可以包容线性系统求逆这个反问题。但 Bayes 估计理论似乎没有讲明白反问题求解的内在条件，认为给出了后验概率密度函数就算基本完成了反问题的求解。当然，这中间距离实用存在非常大的距离。Bayes 估计理论包容性极强，易给人造成误解，以为直接用此方法就可以解决应用问题。

后验概率密度函数由关于噪音（即物理正问题对数据中包含的信号的逼近决定的噪音）和关于模型参数的先验概率决定，这两个先验概率密度函数也是未知的。

当我们分别假设他们分别满足（广义）Gauss 分布时，我们已经把某种约束施加在此反演框架上。首先，先验概率密度函数是单峰值的，那么后验概率密度函数也应是单峰值的，这是后验概率最大化方法求反问题的根本要求，否则 MAP 方法不能产生有意义的解。事实上，多峰值的后验概率密度函数情况下，如何决定反问题的解，目前看不到有效解法，至少我是这么认为的。如果要解的话，也是用强约束条件，在某个参数先验值周围，强制后验概率密度函数是单峰的，然后再套用 MAP 思想进行求解。

这是 Tarantola 书中始终强调他的讨论是弱非线性（moderate nonlinear）问题求解思想的根本原因。

这样反问题就可以退化为（准）二次型泛函的优化问题，因为（广义）Gauss 分布下的后验概率密度函数已退化为单峰的（也是凸的）函数。线性系统（ $\boldsymbol{Ax}=\boldsymbol{y}$ ）表示的二次型泛函可以很好地与 Gauss 分布函数假设一致起来。

这是目前大部分反问题总假设线性系统噪音高斯分布的根本原因。

可见线性（或拟线性，准线性）问题噪音高斯分布（也可是广义 Gauss 分布）下进行反问题求解，不是有了 Bayes 框架就可以置之不理的！这是反问题具体计算不可回避的根本性的问题。

即便当前非线性神经网络号称可以进行非线性优化，但根本上也是逐步求解上述拟线性问题。我认为当前抽取特征的方法都是线性方法，逐步（分层进行）特征表达的深度学习也不可能脱离上述反演框架，这是当前反问题求解的客观现实！

线性系统反问题的求解是大部分研究反问题的数学家所关注的，因为这里面有理论问题需讨论，而且可借助于数学分析这个强大的工具。

首先需定义线性空间、线性赋范空间、内积空间、Banach 空间、Hilbert 空间（完备的内积空间）。在 Hilbert 空间中，存在函数和算子的正交分解，即选择基函数族来表达函数和算子。

$\boldsymbol{x}\in H_1, \boldsymbol{y}\in H_2; A: H_1\to H_2$， H_1 和 H_2 维度可以不同， $\boldsymbol{Ax}=\boldsymbol{y}$ 。

存在伴随算子：$\left(A\mathbf{x},\mathbf{y}\right)_{H_2}=\left(\mathbf{x},A^*\mathbf{y}\right)_{H_1}$ 。

Fredholm 第一类积分方程是 $A\boldsymbol{x}=\boldsymbol{y}$ 的典型例子，很多应用问题可以表达成 Fredholm 第一类积分方程，线性化的 Fredholm 第一类积分方程是常见的。

线性算子 A 可写成矩阵形式，A^* 是 A 的共轭转置。（注：conjugate 共轭，adjoint 伴随）

$$A=A^*,\qquad A \text{ 是自共轭的/自伴随的。}$$

$$\left(A\boldsymbol{x},\boldsymbol{y}\right)\geq 0,\qquad A \text{ 是半正定的。}$$

$$A=A^T,\qquad A \text{ 是对称的。}$$

$$\left(A\boldsymbol{x},\boldsymbol{x}\right)>0,\qquad A \text{ 是正定的。}$$

A 可以是正定对称的。正定对称算子是可逆的：$A\rightarrow A^{-1}$。很显然，若 $A\boldsymbol{x}=\boldsymbol{y}$，$A$ 是正定对称算子，则 $\boldsymbol{x}=A^{-1}\boldsymbol{y}$。

事实上，$A\boldsymbol{x}=\boldsymbol{y}$ 这个问题不是物理意义上的，物理上的 $\mathbf{y}$ 永远包含噪音，而且噪音是随机的（这是 Bayes 估计理论的根本起因）。物理问题中，A 也是人为构建的，它甚至不能很好地预测 $\mathbf{y}$（观测数据）中的有效信号。

数学理论是一种抽象（可能来自物理，也可能来自数学逻辑），不能用数学公式代替客观的物理现实，要处处小心二者间的不一致。牛顿运动学/动力学观点/基本微积分的观点对人的思维有约束作用（指悟性差的人把这套概念等同于物理现实！）。

因此，物理现实中的反问题总需要考虑解的存在性、唯一性和稳定性，而且我们说到此三性是针对线性系统的反问题讲的，非线性反问题当然也存在此三性，但分析起来要复杂得多！线性问题解存在且唯一的基本条件包括如下两条：

1) $\mathrm{N}ull\left(A\right)=\left\{\mathbf{0}\right\}$ 线性算子 A 不存在零空间。

2) $\mathrm{Range}\left(A\right)\in H_2$ 线性算子 A 的值域在 H_2 内。这代表 A 要能预测 $\mathbf{y}$ 中的信号，否则，反问题完全失去基础。应该注意，Bayes 估计也要求这种基本假设成立。这应该是所有反问题的基础假设，保证解的存在性。

但是，$\mathrm{Range}\left(A\right)\in H_2$ 这个数学要求总被想当然地认为成立。我个人认为这是实际反问题中的一个核心影响因素。

当然在此假设成立的条件下，我们可以进入线性系统（$A\boldsymbol{x}=\boldsymbol{y}$）解的结构的分析中，此时线性算子 A 是“正确的”，$\mathbf{y}$ 中可包含小扰动误差！

线性问题解的结构的分析工具就是 SVD 分解。SVD 分解可以针对任意的线性系统，即算子矩阵 A 任意

形式都可以进行 SVD 分解。

算子矩阵 A 的 SVD 分解后得到的 Singular System 为 $\{\boldsymbol{u}_j,\boldsymbol{s}_j,\boldsymbol{v}_j\}$，$\mathbf{u}_j$ 是左奇异向量；$\mathbf{v}_j$ 是右奇异向量。

$A_{m\times n}\mathbf{x}_{n\times 1}=\mathbf{y}_{m\times 1}$ 这样一个线性系统是常见的，对应的奇异系统为：$\{\boldsymbol{u}_j,\boldsymbol{s}_j,\boldsymbol{v}_j\}_{j=1}^{r}$。

r 是矩阵的秩，是 $\mathrm{Range}(A)$ 的维度！ r 可能小于 n。此时 $\mathbf{v}_j,j=1,2,\cdots,r$ 形成 $V_{n\times r}^{r}$，即由 $\mathbf{v}_j,j=1,2,\cdots,r$ 构成 $\mathrm{Null}(A)^{\perp}$ 空间，可另外选个 $n-r$ 个正交基构成 $\mathrm{Null}(A)$ 空间 $V_{n\times(n-r)}^{0}$。$V_{m\times n}=\left[V_{n\times r}^{r},\ V_{n\times(\mathrm{n}-r)}^{0}\right]$。$r$ 也可能小于 m。此时 $\mathbf{u}_j,j=1,2,\cdots,r$ 成 $U_{m\times r}^{r}$，即由 $\mathbf{u}_j,j=1,2,\cdots,r$ 形成 $\mathrm{Range}(A)$ 空间，可另外选 $m-r$ 个正交基构成 $\mathrm{Range}(A)^{\perp}$ 空间 $U_{m\times(m-r)}^{\perp}$。$U_{m\times n}=\left[U_{m\times r}^{r},\ U_{m\times(m-r)}^{\perp}\right]$。

奇异值矩阵 D 写成：$D=\begin{pmatrix}\mathrm{diag}(s_1,\mathrm{s}_2,\cdots,s_r) & D_{r\times(n-r)}\\ D_{(m-r)\times r} & D_{(m-r)\times(n-r)}\end{pmatrix}$。

对于算子 A 有 $A\mathbf{v}_j=\mathbf{s}_j\mathbf{u}_j$。对于算子 A 的伴随算子 A^{*}，有 $A^{*}\mathbf{u}_j=\mathbf{s}_j\mathbf{v}_j$。

线性系统是一个整体，从物理上思考才能有这种整体观。单纯地从数学上讲这种整体观不是那么直接清楚的，以为 $A\boldsymbol{x}=\boldsymbol{y}$ 中各部分是独立的，尤其是算子 A 似乎是独立于 $\mathbf{x}$ 和 $\mathbf{y}$ 而存在的。事实上，算子 A 应是由 $\mathbf{x}$ 和 $\mathbf{y}$ 规定的，当然是由二者之间的因果关系决定的。

从上述观点出发再看 SVD 分解，就知道，左奇异向量（矩阵）和右奇异向量（矩阵）及零空间及其正交补及值域空间及正交补都有明确的物理意义。

另外，依据单峰后验概率密度函数可以由（拟）线性系统对应的二次型来近似的观点，可以看出梯度类方法解优化问题的关键所在。首先误差泛函是（近）二次型的，或者是比较凸的，然后梯度是其中的核心，Hessian 是加速收敛的重要推动力。

上述观点也明确地表现出：目前我们仅能解弱非线性问题，或由已知先验信息把强非线性问题转化成的弱非线性问题。即后验概率密度函数是拟高斯的（等价于误差泛函是可以（拟）线性系统对应的二次型来表示的）问题。

$A\boldsymbol{x}=\boldsymbol{y}$ 线性系统的解有三种解释方式：①TSVD，②滤波 SVD，③LS 求解。

首先，对算子矩阵 A 进行 SVD 分解：

$$A=U^{T}\sum V\rightarrow A\mathbf{x}=\mathbf{y}\rightarrow U^{T}\sum V\mathbf{x}=\mathbf{y}\rightarrow V\mathbf{x}=\frac{1}{\sum}U\mathbf{y}$$

若讨论线性问题的 LS 求解，则有法方程：

$$A^{T}A\mathbf{x}=A^{T}\mathbf{y}$$

$$\left(U^T\Sigma V\right)^T U^T\Sigma V\mathbf{x}=\left(U^T\Sigma V\right)^T\mathbf{y}$$

$$V^T\Sigma UU^T\Sigma V\mathbf{x}=V^T\Sigma U\mathbf{y}$$

$$V^T\Sigma I\Sigma V\mathbf{x}=V^T\Sigma U\mathbf{y}$$

$$V^T\Sigma^2 V\mathbf{x}=V^T\Sigma U\mathbf{y}$$

$$V\mathbf{x}=\frac{1}{\Sigma}U\mathbf{y}$$

可见，没有正则化参数的情况下，法方程的结果与 SVD 解线性方程的结果是一致的。

对于滤波 SVD，有$V\boldsymbol{x}=\frac{1}{\Sigma+\alpha}U\boldsymbol{y}$。它对应于 Tikhonov 正则化反问题下法方程的解，

$$\left(A^TA+\alpha I\right)\boldsymbol{x}=A^T\boldsymbol{y}$$

另外，用正交投影算子P对$\mathbf{y}$进行分解，即：

$$A\mathbf{x}=\mathbf{y}\rightarrow A\mathbf{x}=P\mathbf{y}+I\mathbf{y}-P\mathbf{y}\rightarrow A\mathbf{x}-P\mathbf{y}=I\mathbf{y}-P\mathbf{y}$$

$$\left\|A\mathbf{x}-P\mathbf{y}\right\|=\left\|\left(I-P\right)\mathbf{y}\right\|$$

$$\underset{P}{\mathrm{Lim}}\left\|A\mathbf{x}-P\mathbf{y}\right\|=\underset{P}{\mathrm{Lim}}\left\|\left(I-P\right)\mathbf{y}\right\|\rightarrow 0$$

$A\mathbf{x}-P\mathbf{y}$代表数据中可预测成分的逼近，属于$\mathrm{Range}\left(A\right)$；$(I-P)\boldsymbol{y}$为数据中不可预测部分，属于$\mathrm{Range}\left(A\right)^{\perp}$。

当n$\rightarrow\infty$时，$\left\|\boldsymbol{y}-AA^{\dagger}\boldsymbol{y}\right\|\rightarrow 0$,其中，$A^{\dagger}=(A^TA+\alpha_nI)^{-1}$

可见：$P=AA^{\dagger}$，P也称为数据分辨率矩阵。

上面的讨论都存在一个根本性假设，即算子矩阵A是正确的，至少存在 SVD 后建立起的$\mathbf{x}$与$\mathbf{y}$之间的关系。

可见，所有反问题的基础是正问题，无论基于物理规律的反演或是数据驱动的反演，因与果之间关系的确立都是核心，否则，由果求因根本不可能。

这是对反问题求解的最根本的认识。

10 关于勘探地震学中非线性反演问题的再议

反问题真正的难点在于非线性性，它是由于介质变化和场的变化之间的关系的复杂性引起的。因与果之间的非线性关系，是反问题非线性性的根源，只有深刻地理解这一点才能抓住非线性反演的本质。

介质变化的复杂性是原因，场的变化的复杂性是表象，依据场的复杂性，可预计介质变化的复杂性，也才能发展或设计合理的非线性问题反演方法。勘探地震中面对的地下介质变化，有几种典型的复杂情形：（a）层状介质，但介质层中存在强反射界面，即存在弹性参数（岩性）差异大的介质层。（b）弱反射沉积层中存在大块的盐体，与周围沉积层存在巨大的弹性参数（岩性）差异。（c）弱反射沉积层中，存在孤立的强异常体，尺度与地震波波长相当或小于地震波波长。当然，还有其它复杂的地下介质变化情形，譬如挤压、拉张引起的推覆构造和复杂断块等等。

复杂介质变化引起的问题：①多次反射或多次散射，它们强到不能被忽略；②很难找到合适的描述波场的线性化方法，譬如 Born 近似在很多情形下，很难精确成立，波场计算误差太大。连合适的正问题都没有，反问题很难有好的结果。这正是当前勘探地震反问题的根本症结所在。

解决问题的合适方式是什么？

我个人认为如下几个原则是合适的：①、针对介质变化特点，提出合适的介质与波场之间的线性化方法，譬如针对层状介质，应该用 $V_B(\mathbf{x})+R(\mathbf{x},\theta)$ 的介质描述方式，来推导线性化的波场正演公式。而不是用 Born 近似方式导出线性化的波场正演公式。②、针对大块盐体充填于沉积层中的情形，用 $R(\mathbf{x},\theta)$ 的估计来描述盐体顶部与沉积层的参数变化关系；波阻抗估计来确定盐体的内部参数，现代图像处理来决定盐体的边界，三者结合应能把大块盐体存在时的成像问题解决得比较好。③、充分关注数据的特点提出合适的数据预处理方法是十分必要的，譬如针对层状介质应由浅到深逐步增加数据进行反演；多次波很强的情形，应首先进行去除多次波的反演，然后逐步把多次波放进来进行反演；充分利用到达时进行反演，到达时实际上代表了波包能量的到达，到达时的估计方式应按波包的质心来进行，由统计的观点提取而不是由初至起振点或波峰/波谷点等来提取，最好用 Bootstrap 方式进行不同的到达时测量结果的统计反演。④、最好提出基于介质变化特点的变换域中多尺度参数表达与波场的表达来构建反演方法，并逐步把这种表达推广到能包括多次反射（散射）的情形。

勘探地震中反问题的非线性性是由于介质的中小尺度的强变化引起的。若 $FWI = T\text{omography} + LS_RTM$，$Tomography$ 不应处理非线性问题，非线性主要由 LS_RTM 来对付。而 LS_RTM 被提成线性反演问题逻辑上不太成立，也许进行非线性的 LS_RTM 是合理的！本质上，FWI 的后期迭代就是在进行所谓的 “非线性的 LS_RTM ”，但是，一般地，FWI 迭代不了这么多的次数就已经发散了。多次波是非线性的外在表现，与小尺度散射体（强反射面）的存在是有紧密关系的，但是，目前尚缺乏多次波成像进行高波数参数估计的方法。

应充分注意到低波数弹性参数的估计对小尺度高波数弹性参数估计的影响，二者实际上存在比较强的耦合性。从频带、孔径角度分析二者对非线性反演过程的影响是必要的。

关于 Tomography，主要是估计低波数速度成分。估计这一部分，问题的线性性是比较好的，不需要引入高波数成分（比如异常体外形）。但是，随着要估计的参数场细节的增加，波动理论 Tomography 和 LS_RTM 的功能不是可以截然分开的。我觉得中波数这一部分是如何出来的，理论上没有讲清楚。利用现代图像处理方法加入界面信息是一种必要的、也很有效的措施。但是，FWI 这种完全自动化的方式很难把现代图像处理方法有机地融合到自身的迭代过程中。本质上，利用现代图像处理方法加入界面信息是一种正则化方

法，正说明数据中缺乏估计中高波数参数的信息。

11 保真偏移成像与反演成像中梯度问题的认识

11.1 地震波成像的本质问题

地震波成像起源于 Hargdoon 对偏移成像的物理解释，即绕射叠加和扩散到等时面。Claerbout 又提出激励时间成像条件和波场外推的概念，使得偏移成像可以通过波场反传播来实现。

但是，这些正确的概念都是基于物理直觉提出的，缺乏严格的数学物理证明。本质上，偏移成像和 Tomography 都可以归结为一个 Bayes 框架下的线性化的最小二乘反演问题。我们知道，在正问题线性化和残差波场高斯分布的假设下，最小二乘反演问题的核心在于梯度计算，尤其是带逆 Hessian 预条件的梯度的计算。

11.2 梯度的处理

Bayes 框架下 L2 范数最小二乘线性化反问题的梯度，本质上，是在提出严谨的正传播波场定解问题和严谨的反传波场定解问题，并给出合适的定解问题解法的基础上得到完整解决的。基于 Claerbout 成像条件的偏移成像之所以称为定性的成像方法，就是它根本不关心严格的定解问题的求解，只要模拟的入射波场和反射波场在成像点能相关就达到了定位反射/绕射的目的，而且也不用区分反射和绕射情形。事实上，基于 Claerbout 成像条件的偏移成像根本就不是一个线性化的最小二乘反演问题。一般的 RTM 之所以出现背景成像噪音，本质上也是没有把 RTM 成像问题上升到线性化的反演问题来看待，没有对梯度项进行合适的处理。

11.3 基于高频近似正问题的保真成像（反射系数估计）的梯度

Bleistein 等人提出的保真反射系数的估计，理论上，就是在层状介质、高频近似描述波传播的情形下，解一个 Bayes 框架下的线性化的最小二乘反问题。只是把法方程进行了显式地推导，得到了估计反射系数的半解析表达式。仅此而已！问题的核心依然是梯度的正确处理。即：在提出严谨的正传播波场定解问题和严谨的反传波场定解问题，并给出合适的定解问题解法的基础上，得到正确的梯度处理。

仅仅基于 Claerbout 相关成像条件的粗糙处理，不可能得到保真的成像。梯度处理问题的核心在于波传播问题的正确处理，包括正传问题和反传问题。

不能被 Bleistein 等人弄的一堆基于微分算子的反演方法（譬如广义 Radon 逆变换）迷住了我们对本质问题的判断！

11.4 依据波动理论的线性化反演的梯度处理

前已述及：Bayes 框架下 L2 范数最小二乘线性化反问题的梯度，本质上，是在提出严谨的正传播波场定解问题和严谨的反传波场定解问题，并给出合适的定解问题解法的基础上得到完整解决的。因此，波动理论的线性化反演的梯度处理，应该对正传播场的计算和反传播波场的计算提出严谨的定解问题，并给出合适的计算方法。只要这样处理梯度问题，得到的该线性反问题的解就是理论上应该给出的。适当的分析是必要的，譬如在局部平面波传播框架下进行分析，但不能迷失在别人的话语体系中，还是要基于最基础的数学物理方程理论讨论问题。

12 评述当前反问题

①当前地震波成像的主要问题不在用更复杂的模型来描述或预测数据上。尽管模型是极重要的，但模型好坏的标准是适用，而不是复杂与简单。

② 地震波成像或地震数据分析，不同于大数据分析。勘探地震中叠前数据体（指对参数估计而言）不具备大数据特点。

但是，譬如智能去噪、初至拾取、同相轴或断层自动追踪应该是可以用 ML 实现的。

③ 地震数据分析，看 Yilmaz 的书，以前（第一版）名字叫做 Seismic Data Processing。仅讲到 Prestack Depth Migration 结束。但后来改成了 Seismic Data Analysis，原因是把储层解释放进去了。就是通过对井、震数据的综合分析得到对含油气储层的认识与判断。这才叫 Seismic Data Analysis。类比 Speech and Language Processing，从语音信号得到讲话人的真实语义，就相当于从 Seismic Data 得到含油气储层的认识。

④ 我认为地震数据参数估计这件事或 FWI 要做的事，与 ML 要解决的问题本质上是不同的。地震数据参数估计充其量不过是解决了地震信号的最佳表达，而且基本是在 Born 近似所建立的模型的基础上实现的最佳表达。

当然，这样的最佳表达一定是所谓的（并非真正的）最佳表达，这一点从系统辨识的角度可以清楚看出来。

⑤ 参数估计是勘探地震工作的核心。尽管大数据分析炒得很热，我们必须保持冷静。客观的评价它在勘探地震中的地位和作用。

参数估计的核心是如何解一个非线性问题。尽管大数据分析领域，很多问题的非线性性更强。但与勘探地震的非线性性问题的解决路线（至少在解决路线上）非常不同。

13 ML/AI 储层描述

储层描述的两件事：（1）储层几何形态的描述（2）储层几何形态内的弹性参数（可转化为岩性参数）的获取。探区地震地质解释也是这两件事，不过是从探区甚至更宏观的地质观点来解释探区的构造、沉积、岩性、含油气性。

储层解释是针对储层（目标地质体）的更细的认识。区域构造/沉积/岩性→探区构造/沉积/岩性→储层结构/岩性。大的控制小的，宏观的控制微观的。地质逻辑的合理性（先验信息的合理性）是基础。区域构造→局部断裂体系（断层体系）制约沉积格局。大的构造与沉积制约地震相/地震微相。岩相是地层岩性的具体表现。岩相由波阻抗来刻画，也可由测井信息的综合来刻画。因此到达地震储层描述阶段的解释一定是：针对储层岩性的描述与刻画。具体地讲，就是储层的几何形态的描述，几何形态中的物性参数的描述（估计）以及更根本的储层岩性参数（孔隙度、渗透率、饱和度）的描述。

地震地质解释是个复杂的认知过程，综合的地质知识（尽管非常重要）对解释的影响很难量化。ML/AL 很显然是适用于解释中的，但 ML/AI 所擅长的是从大量/高维数据中找到规律性的东西，并归类展现在解释员面前；ML/AI 也善于做统计分析工作，把统计结果展示出来或 ML/AI 自己依据统计结果进行判决。ML/AI 的所谓“智能”应体现在依据统计结果的判决上。

ML/AI 在地震解释中到底如何用？用来做什么？是基于我们对解释工作的本质认识来回答的。另外，地质（或石油地质）学家对解释的理解与地震学家（勘探地震学家）对解释的理解是不同的。前者更侧重

宏观地质逻辑（也很注意微观的岩心!）；后者更看重反射结构的图像识别、物性参数的估计及波场属性与物性/岩性之间关系的理解。

勘探地震学家进行的储层识别主要应用定量的方法。给出成像道集和剖面；提出地震波场属性；结合测井信息；用 ML/AI 中提供的各种算法并采用统计的思想，给出关于储层结构、物性参数及岩性参数的定量评估，达到储层描述的目的。

我认为基于保真的叠前成像道集，提取其中的属性，在井约束下面，结合深度学习算法，可能会对油气识别有一定的助益。

14 结论与讨论

勘探地震学的核心包括地震数据的采集与地震数据分析两大部分。地震数据采集技术的导向是自动化地实现“两宽一高”数据采集；地震数据分析的导向是（似乎是）在随机信号分析与 Bayes 估计理论框架下把基于模型的和基于数据的信息提取方式结合在一起获取更高精度的与储层有关的结果。

从系统的线性与非线性以及信号（数据）的概率分布是否高斯的基本假设，发展出了地震信号处理/分析方法（譬如去噪、插值、规则化、混叠书分离等等）和地震波成像方法（譬如层析成像、最小二乘偏移成像）。信号分析和参数估计本质上是基于相同的理论基础，不同的是信号分析用的是信号模型和弹性参数估计用的是波动方程模型。尽管理论结构很清晰，但从理论框架到实用化的技术之间有非常大的距离。我们知道这样的距离主要由上述理论框架的基本假设与实际数据不匹配，以及所用的模型与实际数据不匹配所导致的。

因此，需要对理论到应用中整个链条上的各个环节进行深入地分析。只有在深入的分析和清晰的认识基础之上，才有可能提出行之有效的方法技术。我们知道上述分析不是系统性的，只是在关键点上进行分析，体现了我们对问题的认识与判断。

参考文献

[1] Gelman, A., Carlin, J. B., and Stern, H. S., 2014, Bayesian Data Analysis, Third Edition, CRC Press.

[2] Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM.

[3] Vogel, C. R., 2002, Computational Methods for inverse problems, SIAM(Society for Industrial and Applied Mathematics.

[4] Jari Kaipio & Erkki Somersalo, 2005, Statistical and Computational Inverse Problem, Springer.

[5] Troyan, V., and Kiselev, Y., 2011, Statistical Methods of Geophysical Data Processing, World Scientific Press.

[6] Manolakis, A. G., Ingle, V. K., and Kogon, S. M., 2005, Statistical and Adaptive Signal Processing-Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing, Artech House Press.

[7] Leon Cohen, 1995, Time-Frequency Analysis, Prentice-Hall PTR

[8] Mallat , S., 2010, A wavelet tour of signal processing, The Third Edition., Elsevier.

[9] Michael Elad, 2010, Sparse and Redundant Representations---from theory to applications in signal and Image processing, Springer.

[10] Starck, J-L., Myrtagh, F., and Fadili, J, M., 2010, Sparse image and signal precessing – Wavelets, Curvelets, Morphological Diversity, Cambridge University Press

[11] Kay, S. M., 1993, Fundamentals of Statistical Signal Processing: Volume I, Estimations Theory, Prentice Hall PTR

[12] Kay, S. M., 1998, Fundamentals of Statistical Signal Processing: Volume II, Detection Theory, Prentice Hall PTR

[13] Theodoridis, S., 2015, Machine Lerning – A Bayesian and Optimization Perspective, Elsevier

[14] Hastie, T., Tibshirani, R., and Wsinwright, M., 2016, Statistical Learning with Sparsity – The Lasso and Generalizations, CRC Press

勘探地震学技术发展方向分析

王华忠

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】勘探地震学原则上可以归结为一种数据分析学科。数据来源主要是野外采集的叠前地震数据。数据分析的方法当前主要是基于模型的方法，而且主要分为两大类，一类基于信号模型的信号分析；一类是基于波动方程模型的弹性参数估计。经典的勘探地震学中的数据处理方法：各种静校正、去噪音、反褶积、NMO、叠前时间偏移、速度分析和叠前深度偏移，都可以视为一种简化的数据分析方法。当前的地震数据分析逐渐统一在 Bayes 理论框架下面，无论信号分析、图像分析和弹性参数估计都在同一个框架下进行。仔细分析当前勘探地震学中的关键技术环节，包括地震数据采集理论及方法、去噪理论及方法（包括数据规则化、数据压缩、数据解混叠等）、地震波层析成像和偏移成像、图像处理、机器学习等等，主要的思想方法都是奠基在 Bayes 理论框架上。本文中，我试图对当前勘探地震学技术发展方向进行一些点评，主要是依据最近几年 SEG 和 EAGE 年度论文集，当然也是依据波现象与反演成像研究组近几年来的研究工作做出的判断。

【关键词】勘探地震学，数据分析，Bayes 估计，信号分析，图像分析，弹性参数估计

一、引言

受 2000 年后一段时间高油价的刺激，当然计算机技术、大数据分析及人工智能技术也在这段时间有快速进步，勘探地震学技术发展非常快。常规的地震数据成像处理方法至少从研究的角度正逐渐淡出大家的视野。当前，从理论上看，勘探地震学似乎全面进入了以 Bayes 估计为理论基础的阶段，无论是信号处理与分析、图像处理与分析、弹性参数估计莫不如此。地震波正演当然是独立的研究范畴，但也可以认为它是为了进行 Bayes 估计而进行的建模方面的准备。在信号分析领域，小波变换本质上也可以视为是为了进行 Bayes 估计而进行的建模方面的准备。当然，也可以认为正演和反演是一体的，这只是认识问题的角度不同而已。尽管如此，我认为 Bayes 理论框架是包容性更强的一套优美的理论，可以更好地指导我们进行依据实测数据来估计未知系统、了解未知系统、认识未知系统，当然也包括估计系统输入。

在上述观念的导引下，仔细分析最近几年 SEG 和 EAGE 年度技术论文集中数千篇文章，对地震数据采集理论方法与技术、地震数据去噪方法、数据规则化方法、数据解混叠方法、鬼波压制方法、多次波压制方法、地震波层析成像方法、最小二乘成像方法、全波形反演方法等进行方法理论总结，可以明显地发现上述所有技术方向的方法理论都有一个明显的特点：正演预测+反演估计。Berkhout 教授在 2015 年度总结性文章中也提及了这样的观点，只不过他没有推广到如此广泛的程度。

最近几年大热的大数据分析，其核心是机器学习与统计学习，似乎把基于数据进行信息提取，进而实现判决的做法又提升一步，即不需要花费巨大的心思研究正问题，只要有充分的大数据，就可以得到未知系统（或未知对象）的正确的判决结果。尽管大数据分析在图像处理、语言分析等方面有重大成功，但在油气地震勘探中，尤其是由叠前数据直接到含油气储层识别方面，我认为不知何时或根本不知道能否会有很成功的应用。某些局部点上，譬如智能去噪、自动层位识别、初至拾取等，会有不错的应用效果。

从上述分析可以看出，当前的勘探地震学重点研究问题已经与十几年前大不相同了，勘探地震学教科书上的内容似乎都已经是老黄历了。

本文试图根据文献及个人研究的体验，简单地评述勘探地震学中的一些重点技术方向的技术要点，为波现象与反演成像研究组（WPI）的进一步研究工作指出方向。

二、地震数据采集代表性思想

a）地震波采样正从 Shannon 采样理论向 CS（Compressive Sensing）采样理论演化过程中，但本质上 CS 采样应是依据参数场（或成像结果）特点设计观测系统（所谓随机观测系统），达到最佳采样（最省炮检点数（即最省钱！）、最佳 Deblending、最佳成像）。但目前勘探地震中主要是随机激发时间的同时源数据采集，以提高采集效率为主。即便以 Deblending 效果最佳进行最佳效率的采样目前也没有成熟的理论支撑。

b）宽频、可控震源是采集技术中的一个核心。激发地震子波的特征：振幅谱光滑且对应高分辨率子波、低频成分有效（能用来估计波阻抗）、相位谱线性。目前气枪、炸药震源很难控制振幅谱，但整个振幅谱的形态决定子波的高分辨，控制振幅谱形态的震源才能真正实现高分辨子波的生成！可控震源似乎可以做到，但实际地震勘探结果看可控震源并没有给出更高分辨率的观测和成像结果，扫描信号和实际震源出力以及介质实际震动之间并不是协调一致的。

c）“两宽一高”地震数据采集是目前最能满足高精度地震成像要求的合理的观测方式。硬件技术是当前的瓶颈。震源、检波器相互独立，并且各自能自主移动、定位等，是今后“两宽一高”地震数据采集硬件技术的发展方向。总之，此类无人系统是数据采集技术方向。陆上和海上地震数据采集侧重点不同，海上更侧重自动化实现，陆上当前更侧重“节点”检波采集技术的实用化。

d）DAS 观测技术，因其廉价、轻便等诸多优点很受工业界重视。安全监控领域早已广泛应用。但用来测量精度要求很高的地震波场，看来技术还不成熟。

e）廉价地震观测系统是很有吸引力的方向。在城市地震勘探中，这可能是正确的技术方向。我相信针对地表层的高精度反射波勘探最终会打败面波地表勘探，目前的困扰就在于缺少合适的硬件系统进行针对近地表的“两宽一高”地震数据采集。

三、地震数据预处理代表性思想

a）最近几年来数据预处理的思想、观点逐渐成型，基本上没有什么大的变化。

（ i ）$\boldsymbol{\omega - x}$域线性预测器预测信号或高维小波基函数的线性组合预测信号；噪音能被认为是符合广义高斯分布的；建立$\mathbf{L}_2$范数、$\mathbf{L}_1$范数或$\mathbf{L}_1 - \mathbf{L}_2$混合范数反问题；实现插值、去噪、规则化、Deblending 等方面的处理。

（ ii ）上述思想转到数据驱动，即由数据自相关分解（正交分解），得到信号空间与噪音空间的分离，找到的正交基函数与人为选定的正交基函数本质上是类似的，这种称为 SSA（奇异值谱分析）的一类信号分析方法也是最近几年来的核心议题。最近几年主要是研究高维数据驱动下的奇异值谱分析。可以参考 Sacchi 教授的文章。

（iii）由数据驱动或事先人为选定基函数进行字典学习，然后进行去噪、规则化等是更进一步的方法技术，试图对信号表达达到更好的精度，实现更好的去噪效果。

b）Deblending 方法目前也逐渐地形成统一的思想，尽管做法上各有千秋。总体上可以表达为：伪解码；去噪音这两个核心步骤。去噪方法还是利用 a）中提及的思想与方法。由于去噪是基于 Bayes 估计理论，总是对噪音的统计分布有假设，因此去噪之前可以（最好是）加上对直达波、面波这类强干扰信号的预处理。另外，对 Deblending 方法结果再进行信号找回也是必要的，毕竟目前的任何去噪方法都不能保证不伤害有效信号。

Deblending 问题是一个典型的“鸡尾酒会”问题，但是勘探地震中的 Deblending 方法却不是用经典的独立分量分析方法，而是归结为伪解码后地震数据的去噪问题，这应该是与地震数据的随机采样和地震数

据具有典型的稀疏表达特征有关系的。Deblending 问题与压缩感知采样、地震数据的稀疏表达和噪音特点三个问题紧密相关。

c）非线性去噪方法的技术进展不大。我个人认为是当前的局部邻域滤波器很少体现信号的特征，所以非线性去噪不会成为去噪音的核心方法，只能在去除一些异常噪音方面发挥作用。去噪的根本在于建模预测信号，非线性的局部邻域滤波器是一种趋势加权滤波器，对信号的预测能力有限。可以把非线性信号的去噪统一成各向异性高斯滤波函数的设计问题，包括矢量中值滤波也可以统一在此框架下面。

d）陆上极低信噪比数据的去噪

针对陆上极低信噪比数据，2017 年度 SEG 论文中有几篇文章用 Beam-forming 的思想压噪。差不多十年前，我们就提出基于 Fresnel 半波带能同相叠加的物理观点对付极低信噪比数据的方法技术。总之，构建一个较好地消除来自 Fresnel 半波带的反射（散射）波时差的同相叠加滤波器，应能很好地压制陆上数据的强噪音。

e）鬼波及其它类型的多次波

我认为用编码的概念建立起一次波及各阶多次波之间的预测关系十分清晰明了，然后用解码进行反问题求解是很好的学术框架。WPI 在这方面的实践也证实了这种思想方式的有效性！上述思想可以方便地用于鬼波和水体相关多次波的消除。

四、关于地震波成像

当前，地震波成像无论从学术研究或应用角度看，都在由定位反射系数的 Migration 向估计反射系数，甚至波阻抗的 Inversion 方向转化。由叠前数据直接估计波阻抗的条件也逐渐变得成熟起来，这都是“两宽一高”的数据所带来的成像方面的变化。

我认为，WPI 提出的特征波成像（CWI）的路线是物理意义清晰的、正确的路线。CWI 包含四个关键技术步骤：

（ⅰ）CWF（特征波场）的抽取；

（ⅱ）透射波（直达、折射、Diving Wave）层析；

（ⅲ）反射波层析；

（ⅳ）高波数参数的估计。

其中 CWF 抽取归结为地震信号稀疏表达问题，属于信号预处理部分的研究。地震波成像主要关注其余三个方面的问题。

a）透射波层析

地震勘探技术发展越深入，对小尺度地质体的成像、岩性（弹性参数）成像越重视，近地表对这类成像（指小尺度地质体的成像和弹性参数反演成像）的影响越凸显出来。

我认为近地表问题的解决依然要依赖体波，而不是面波，面波只能起辅助作用。当前，估计近地表模型用透射波（直达、折射和 Diving Wave），属无奈之举，因当前针对深层目标的观测系统采集不到、也采集不全近地表模型中传播的反射和散射波，当然也无法利用它们进行更精确的成像。事实上，即便用透射波进行近地表模型建立，也是最近一段时间才真正地重视起来。

关于近地表建模，若基于当前针对中深层的观测系统，能利用的信息仅有直达波、折射波、Diving Wave。

直达波可以确定近 Offset 的表层速度，折射波能否用还要看探区是否存在折射层，Diving Wave 一般而言在中远 Offset 才明显，穿透深度比较大，三者相结合给出的也是表层、浅层、近地表模型的光滑部分，其中高波数的（或小尺度的）速度异常，地表高程突变引起的到达时突变不能放在透射波层析中。

选一个贴近地表的基准面进行高波数时差的消除是十分合理的，但我认为不存在一个公式定量地说明存在一个唯一的、最佳的基准面。

另外，与当前的炮初至相对应的在近地表介质中传播的体波的物理过程不够清楚，譬如 Diving Wave 在什么介质中、在什么频率范围、在什么 Offset 范围中比较发育，需进一步证实。只有对波场和介质的相互关系理解比较到位了，才能构建出合适的反演方法技术。

还有，非水平地表散射源产生的体波如何用在地表模型层析中，值得探讨。

尺度（速度异常体的尺度）在建模过程中起十分重要的作用，基于地震波到达时的反演不能估计小尺度的速度异常体的速度值，近地表建模中需把这些小尺度速度异常体的边界人为地刻画出来并填入合理的速度值。事实上，近地表建模也应引入（或最好引入）近地表模型的结构信息。但是这些先验信息从何而来？目前看来，露头、微测井、近地表地震等是有效的来源。

最后，近地表速度建模正确性或合理性的判断依据应该是标志性反射层对应同相轴的道间时差在任一 Offset 上都不应出现跳跃变化。这也是剩余静校正的判别依据，当然与中深层速度模型融合后，浅中深层反射波成像道集都拉平是更好判别标志。

b）反射波层析

首先必须指出，与透射波不同，反射波层析与反射波成像（指高波数部分的成像）是耦合在一起的，那种把利用反射波进行背景速度建模与速度扰动量的反演完全割裂开来看待的观点是不对的。二者可以同时反演，也可交替进行迭代反演。一般反射波 FWI（RFWI）是同时反演的。但我们还是坚持二者交替迭代反演的做法，这主要是从更实用化的角度来考虑的，当前石油工业界也是把背景建模与高波数扰动量估计分开进行作为主要做法。

背景速度场反射波层析成像分为像域和数据域中两种实现方式。

当前工业界常用的是像域（即基于成像道集中剩余深度差（RMO））的层析成像方法，又分为射线理论和波动理论层析成像方法，应用广泛的是射线理论的层析成像方法，当前工业界侧重发展的是 ORT 介质、宽方位数据的成像域射线理论层析成像方法，要点是：成像道集质量优化；构造层位、倾角、RMO 自动拾取及质控；全方位层析方程建立与求解方法；基于现代图像处理方法的构造约束建模。

DSO 方法（Symes 教授几十年来推崇的做法！）是像域波动理论的层析成像的代表性方法，目前还看不到广泛应用的前景。我不太认可这种层析建模的实现逻辑，试图利用像的振幅构建反演方法，本质却是要求道集拉平，消除 RMO，内在逻辑上不合理。因此，到目前为止，DSO 方法一直不温不火，也许会无疾而终。

数据域中反射波层析，即 RFWI，我不赞同用全波形逼近的方式进行。WPI 发展的是波动理论反射波旅行时层析。当前数据域反射波层析成像依然在不断发展中，实用化问题存在不少障碍。首要的问题是如何实现实测数据中同相轴与合成数据中同相轴的一对一匹配，完全克服 Cycle-skipping 问题；然后是子波的振幅与相位匹配或把二者的影响从测试数据和理论合成数据中剥离掉；正则化方法也是重点要考虑的问题。把这三点问题能解决好，数据域的反射波层析成像会成为有效的建模方法。

c）高波数参数估计

高波数参数估计是偏移成像方法所关注的。正在由对反射（绕射、散射）点的定位向反射系数（速度扰动量）估计方向发展中。从方法研究角度看，似乎由于宽带数据的获得，正在由角度反射系数估计向宽带的

波阻抗估计方向转移。

至于为什么不用 FWI 直接估计宽波数带的参数场，这个问题需花费大篇幅说明，简而言之是当前的方法不能做到从低波数参数估计到高波数参数估计的收敛，发散的反演方法和反演结果是没有什么价值的。

高波数参数估计的核心在于提一个真正的（符合勘探地震现实的）线性问题。基于反射波及层状介质模型的反射系数反演是比较符合这种要求的，最小二乘方法是基本的反演框架。

由于规模巨大，Hessian 矩阵逆的近似处理问题成为了最小二乘方法中的核心关注。存在较多实用化的高波数参数估计文章关注此问题。

目前存在的根本问题是强反差界面、非规则外形的反射体（盐体）、小尺度强异常体等引起的非线性问题。这是当前地震波反演成像的难点所在。

噪音或由于正问题不准确或正问题数值算法引起的“噪音”，对反问题的影响更是反问题的大瓶颈！

反问题的解决是看不到尽头的！理论意义上就该是这样。我们只能不断地随数据质量、正问题的精确化、模型先验认识的深入而不断进步。反问题的本质就是在不完全数据下做出的所谓的最佳估计。任何情况下，这句话都是反问题本质的合适表述。

五、关于图像处理、机器学习及大数据学习

地震成像的后处理基本上是面向图像或 3D 数据体的。针对这类数据体的处理主要用图像处理方法。这是当前层析建模、地震地质解释、储层描述与刻画中共同的议题。包括反射同相轴的识别、地震相识别、裂隙识别、盐体及各种尺度异常地质体的识别等。

现代图像分析理论是这类问题的共同基础。图像稀疏表达、偏微分方程图像处理、ROF 模型反演、图像分割以及模式识别等是其中的核心议题。

从数据中学习是人类学习知识的一个核心途径，数千年来一直如此。到目前为止，大家逐渐认识到数据是有关联的、有结构的，它们是与产生数据的物理原因直接相关的。就是说数据分析是有规律可循的。既然这样，计算机可以在一定的规则知道下对数据进行分析。当前的机器学习实际上就是在人类制定的学习算法或学习原则的基础上自主地从数据中获取信息、甚至知识。机器学习理论的逐步进展路线是：最佳 Weiner 滤波、Kalman 滤波、自适应（非平稳、缓平稳信号）滤波、有监督三层（浅层）神经网络发展到当前的无监督深层神经网络。无监督深层神经网络是当前大数据分析的重要工具。

从上述分析可知，大数据分析的算法原则上是简单的，与经典和现代信号分析的发展紧密相关，可以说就是现代信号分析的最新发展结果。大部分业界人员都可以弄清楚算法（不一定能创造出新算法），但是在大数据分析中，真正的问题还不是算法，而是对数据背后所蕴含的规律性的领悟。能基于数据与原因之间关系的领悟建立起合适的模型，然后进行大数据分析才是真正的大数据分析高手。

很遗憾，勘探地震中的核心问题地震波成像或弹性参数反演不是大数据分析问题。

机器学习思想方法影响勘探地震数据成像处理是不可避免的，选择像自动 CMP 道集速度谱拾取、自动噪音压制、自动初至拾取、自动同相轴拾取等具体问题作为切入点，把机器学习引入勘探地震数据成像处理中当然是合适的。但是，不要忘记，机器学习的真正强大在于试图代替人工进行含油气性判决。在我看来，这是一个长期的探索性研究工作。

应用很成功的语音识别，目前“听音达意”的成功率还是很高的。勘探地震中的“听音达意”就是“听波场知油藏”，这能做到吗？“听波场知油藏”的研究工作，从哪个点上切入呢？

六、关于 AVA 反演问题的评价

由 $R(\theta)$ 到 AVA 反演再到储层描述的路线，我认为目前是鸡肋性的技术路线，食之无味，弃之可惜。

AVA 反演的价值在于能解释 $R(\theta)$ 与弹性参数变化间的关系，指导我们做好 FWI 意义下的参数反演；另一个价值在于探索一下基于岩石物理关系进行储层参数估计的路线。总之，进行理论分析，产生一些理论指导性是 AVA 反演的价值所在。至于实际的 AVA 弹性参数反演由于估计结果本身方差太大，实用上没什么意义！这基本上被当前的勘探实践所证实。

今后储层描述应在 FWI 理论下由宽带、宽方位叠前数据直接进行波阻抗反演。结合岩石物理，进入储层描述。我认为这是“两宽一高”地震数据下进行储层描述的正确路线。这又一次体现出来地震数据采集对成像处理方法技术的促进作用。

七、结论

勘探地震学的技术发展进入了以反演成像为核心的阶段。以前简单的射线理论和偏移成像支撑起的常规地震数据成像处理方法技术尽管还在实际生产中起着重要的作用，但是已经是既往的技术了。

现在的观点是把叠前地震数据视为随机信号，在 Bayes 估计理论框架下，把信号处理、图像处理、弹性参数估计这些核心问题都置于同一个反演理论下，不同的只是信号预测模型、图像预测模型和波场预测模型这些正问题存在的差异。更进一步地，把不显式地基于模型的大数据分析也纳入在 Bayes 估计理论这个有高度包容性的理论平台下面。从这样的高度看待地震数据分析，可以把地震数据采集、地震数据预处理、弹性参数反演、地震数据解释的本质问题更准确地把握到。这样看待勘探地震中关键问题，并非忽视勘探地震的物理本质。震源激发的物理、地下介质的物理、波传播的物理、接收器的物理始终是至关重要的。但是，勘探地震本身总体上就是一个反问题，目标是弄清楚地下介质中含油气储层的几何位置、空间形态、岩石性质等。本质上，脱开 Bayes 估计理论的思想框架，正问题和反问题是一体的，可以说是同一个问题的两个方面而已。分成正反问题只是为了讨论、理解和说明问题的方便。

从这样的观点看待勘探地震学的技术发展，可以十分明晰地知道数据采集方法理论、数据预处理方法、弹性参数估计方法、包括地震数据解释方法（也可以姑且认为它是地震数据挖掘方法）至少从理论上讲都在向“非高斯、非线性、非平稳”的理论方法和技术方向发展。目的当然是突破当今现有的方法基于信号是“高斯、线性、平稳”的假设，我们知道这样的假设是不符合实际数据情况的，导致目前的信号处理和参数估计的精度不高。当然，上述评价过于宽泛，实在是如果过于关注细节，一是需要长篇大论，二是也会导致忽略当前技术发展的真正导向。

参考文献

[1] 《SEG2017 年度论文集》
[2] 《SEG2016 年度论文集》
[3] 《EAGE2017 年度论文集》
[4] 《EAGE2016 年度论文集》
[5] 《波现象与反演成像研究组研究文集》

从线性高斯反演成像到非线性非高斯反演成像的演变

王华忠，郭颂，吴成梁

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】地震波反演成像已经取代偏移成像成为了勘探地震学中地震波成像方法技术的发展方向。毫无疑问，“两宽一高”的地震数据采集及高性能计算机技术的发展促进了这种转变。但是，更本质地，Bayes 估计框架早已建立起了利用实测地震波场估计地下介质弹性参数相当完美的理论。在对波场预测算子线性化、预测误差高斯分布、L2 范数定义预测误差的前提下，线性反演问题的所有内涵都是清楚的。勘探地震学中的地震波线性反演方法与技术都在解决复杂的地下介质情况、复杂的地震波场、不完全的地震数据观测、不准确的地震波数值模拟情况下，如何把线性反演理论用在实际数据的反演过程中，从而得到更好的成像结果。地震波反演成像在正则化方面提出了很多有启发性的思想和实用化的方法。但是，非线性非高斯情形下的地震波反演成像基本没有明确的技术路线。本文首先阐述 Bayes 反演理论的基本思想，指出当前地震波线性化反演理论的完美性及实用化过程中遇到的关键问题。进一步地，重点分析当前弹性参数反演偏离线性反演理论假设的情形（即非线性，非高斯情形）下，地震波反演成像的观点与可能的方法技术。

【关键词】（广义）Gauss 分布，线性化正算子，非线性正算子，Bayes 估计，弹性参数估计，反演算法

1 引言

Tarantola（2005）的代表作“模型参数估计的反问题理论与方法（Inverse Problem Theory and methods for Model Parameter Estimation）”非常明确地指出地震波反演要估计的是波动理论正问题模型中所包含的模型参数。模型参数和地下介质弹性参数之间的关系由波动理论正问题能预测实际观测数据而建立起来。正是基于此，估计模型参数也就能在某种程度上“等价于”估计地下介质弹性参数。可以说我们直接地估计模型参数，间接地估计地下介质的弹性参数。从这一点可充分地看出波动理论正问题模型在地震波反演中到底有多重要！正问题描述与刻画的合理性才是有效地解决反问题的核心！

数据，实测数据，是估计地下介质弹性参数真正的信息源，波在地下介质中的传播过程受到地下介质弹性参数的改造，从而使得在地表接收到的数据（或波场）中携带了地下介质弹性参数的信息。勘探地震中的实测数据总是不完备的，所蕴含的信息不足以高精度地估计地下介质弹性参数。

因此，Bayes 框架下的弹性参数估计是当然的选择。 Bayes 估计理论真是一个完美的理论架构！

很遗憾，勘探地震中高维的弹性参数估计过程中并不能计算出后验概率密度函数，也不能据此算出估计结果的期望值。因此，实测数据与模型参数之间的先验概率分布被假设成 Gauss 的或广义 Gauss 的。同样地，必须要引入的模型参数的先验概率分布也被假设成满足 Gauss 的或广义 Gauss 的。引入如此假设的目的是试图利用 Gauss 或广义 Gauss 情形下后验概率密度函数的最大化来对模型参数估计结果的优劣做出判决。众所周知，Gauss 或广义 Gauss 情形下后验概率密度函数的最大化等价于对应的代价函数的最小化，从而把高维弹性参数估计问题退化为一个对代价函数求极小的最优化问题。

这还没有完！Gauss 或广义 Gauss 情形下后验概率密度函数本质上要求波动理论正问题模型是线性的或弱非线性的。强非线性性的波动理论正问题模型与后验概率密度函数满足 Gauss 或广义 Gauss 分布假设是矛盾的！因此，引入线性化的正问题进行数据预测就成了必须！

为了发展出有效的模型参数估计方法，我们不得不在各种假设下建立各种“模型”。这些“模型”与要解决的实际问题符合吗？显然差距甚大。

Gauss 线性反演理论是又一个完美的理论架构！在对波场预测算子进行线性化、预测误差满足 Gauss 分布（尤其是 Gauss 白噪分布）、L2 范数定义预测误差的前提下，Gauss 线性反演问题所有内涵全部是清楚的。最理想情况下，模型参数估计结果是无偏的和方差最小的。理论是完美的！现实却是差距甚大的！当前，勘探地震学中的地震波线性反演方法与技术都在解决复杂的地下介质情况、复杂的地震波场、不完全的地震观测、不准确的地震波数值模拟情况下，如何把线性反演理论用在实际数据的反演过程中，以得到更好的成像结果，各种正则化思想与方法也备受重视。

地震波线性反演方法与技术的进步是有目共睹的！

下一步应该怎么办？从上述分析可知，对后验概率密度函数（间接地也是对先验概率密度函数）必须要满足 Gauss 或广义 Gauss 的假设是（看起来是）必须要突破的！一旦没有这个假设，正问题的线性化假设也可以拿掉了！可是，没有了 Gauss 线性化假设，现有的一套反问题算法也不存在了！笔者认为在非 Gauss 和非线性情况下发展出有效的模型参数估计算法是问题的关键！

什么是有效的模型参数估计算法？

2 Bayes 框架下的地震波反演成像

地震波反演成像是奠基在 Bayes 框架下估计理论基础上的。整个地震数据处理与反演成像过程（去噪、叠前偏移成像、速度分析、层析成像等）都可认为在此框架下实现。

实际地震数据是一个随机过程，实测到的地震数据则认为是该随机过程的一次具体实现。因此，可以把观测的地震数据视为一组随机信号，其具有均值、方差等统计特性，满足一定的概率分布。同样地，反演的地下模型参数也应被认为是随机的。假设有模型空间$\boldsymbol{M}$和数据空间$\boldsymbol{D}$，在其定义的联合空间$\boldsymbol{M}\times\boldsymbol{D}$ 上，有同时反映观测数据$\mathbf{d}$以及模型参数$\mathbf{m}$信息的联合先验概率密度$\rho(\mathbf{d},\mathbf{m})$，同样有反映联系观测数据与模型参数物理规律的概率密度$\Theta(\mathbf{d},\mathbf{m})$。利用 Bayes 公式，可得出联合空间$\boldsymbol{M}\times\boldsymbol{D}$上的后验概率密度$\sigma(\mathbf{d},\mathbf{m})$：

$$\sigma(\mathbf{d},\mathbf{m}) = k\frac{\rho(\mathbf{d},\mathbf{m})\Theta(\mathbf{d},\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})} \tag{1}$$

其中，k为归一化常数，$\mu(\mathbf{d},\mathbf{m})$为联合空间中关于观测数据与模型参数的均匀概率密度函数。均匀概率密度函数是与坐标系统有关的，表示为特定数据和模型的参数化下（选定的坐标系统），概率空间中单位区域内的概率密度函数，代表了对模型空间与数据空间的参数化认识。

进一步地，关于模型空间的后验概率密度$\sigma_{\mathbf{M}}(\mathbf{m})$以及关于数据空间的后验概率密度$\sigma_{\mathbf{D}}(\mathbf{d})$可由边缘概率密度计算公式得到:

$$\sigma_{\mathbf{M}}(\mathbf{m})=\int_{\mathbf{D}}\sigma(\mathbf{d},\mathbf{m})d\mathbf{d} \tag{2}$$

$$\sigma_{\mathbf{D}}(\mathbf{d})=\int_{\mathbf{M}}\sigma(\mathbf{d},\mathbf{m})d\mathbf{m} \tag{3}$$

假设数据空间与模型空间为两个独立空间，联合空间中的均匀概率密度$\mu(\mathbf{d},\mathbf{m})$为：

$$\mu(\mathbf{d},\mathbf{m}) = \mu_{\mathbf{D}}(\mathbf{d})\mu_{\mathbf{M}}(\mathbf{m}) \tag{4}$$

其中，$\mu_{\mathbf{D}}(\mathbf{d})$和$\mu_{\mathbf{M}}(\mathbf{m})$分别表示关于观测数据和模型参数的均匀概率密度函数。该式表示模型空间中的模型参数化与数据空间中的数据参数化是无关的。

假设关于观测数据与模型参数的先验信息统计无关，公式（1）中的联合先验概率密度$\rho(\mathbf{d},\mathbf{m})$变为：

$$\rho(\mathbf{d},\mathbf{m}) = \rho_{\mathbf{D}}(\mathbf{d})\rho_{\mathbf{M}}(\mathbf{m}) \tag{5}$$

其中，$\rho_{\mathbf{D}}(\mathbf{d})$和$\rho_{\mathbf{M}}(\mathbf{m})$分别表示关于观测数据和模型参数的先验概率密度函数。

地球物理反演（包括其它基于数据的反演）的本质都是从数据中学习到有用的信息并据此得到客观认识。实测数据和要估计参数之间存在一定的关系是反演的根本基础，二者没有关系或关系极弱，不可能从实测数据中得到要估计的参数的信息。据此，联合概率密度$\Theta(\mathbf{d},\mathbf{m})$可用关于模型参数的均匀概率密度函数与其对应的条件概率密度函数来表示：

$$\Theta(\mathbf{d},\mathbf{m}) = \theta(\mathbf{d}\,|\,\mathbf{m})\mu_{\mathbf{M}}(\mathbf{m}) \tag{6}$$

其中，$\theta(\mathbf{d}\,|\,\mathbf{m})$ 为在给定的模型参数 $\mathbf{m}$ 的条件下能够预测观测数据 $\mathbf{d}$ 的条件概率密度函数，表达要反演的参数与实测数据之间存在的关系。均匀概率密度函数 $\mu_{\mathbf{M}}(\mathbf{m})$ 的引入表示模型参数的选择独立于地震数据的理论预测。

因此，关于模型空间的后验概率密度 $\sigma_{\mathbf{M}}(\mathbf{m})$ 变为：

$$\sigma_{\mathbf{M}}(\mathbf{m})=\int_{\mathbf{D}} k\frac{\rho_{\mathbf{D}}(\mathbf{d})\rho_{\mathbf{M}}(\mathbf{m})\theta(\mathbf{d}\,|\,\mathbf{m})}{\mu_{\mathbf{D}}(\mathbf{d})}d\mathbf{d} \tag{7}$$

Bayes 估计反演的模型参数结果 $\hat{\mathbf{m}}$ 可通过后验概率密度函数 $\sigma_{\mathbf{M}}(\mathbf{m})$ 的期望值 $E\{\mathbf{m}\,|\,\mathbf{d}\}$ 获得：

$$\hat{\mathbf{m}}=E\{\mathbf{m}\,|\,\mathbf{d}\}=\int_{\mathbf{M}}\int_{\mathbf{D}} k\mathbf{m}\frac{\rho_{\mathbf{D}}(\mathbf{d})\rho_{\mathbf{M}}(\mathbf{m})\theta(\mathbf{d}\,|\,\mathbf{m})}{\mu_{\mathbf{D}}(\mathbf{d})}d\mathbf{d}d\mathbf{m} \tag{8}$$

此时的估计结果是用后验概率密度函数得到的加权平均结果。只有在后验概率密度函数是单峰的，譬如高斯的或广义高斯的，此期望值才是有意义的。

其解可由方差 $\hat{\mathbf{C}}_{\mathbf{M}}$ 来评价：

$$\hat{\mathbf{C}}_{\mathbf{M}}=\int(\mathbf{m}-\hat{\mathbf{m}})(\mathbf{m}-\hat{\mathbf{m}})^{T}\sigma_{\mathbf{M}}(\mathbf{m})d\mathbf{m} \tag{9}$$

方差越大，说明参数估计的精度在概率意义上是越低的。

显然，从概率论的角度，利用公式（9）进行参数估计是合理的。但是目前缺乏有效的计算方法计算高维模型参数的联合后验概率密度函数。（8）式定义的积分也需要巨大的计算量。为此，不求后验概率密度函数的期望值，转为求取使得后验概率密度最大化（简称：MAP）的模型参数估计结果作为反演解：

$$\hat{\mathbf{m}}=\underset{\mathbf{m}}{arg\,max}\,\sigma_{\mathbf{M}}(\mathbf{m}) \tag{10}$$

由于公式（10）中的后验概率密度函数可能较为复杂，无法直接求解该式对应的最优化问题。需要进一步地引入先验信息。

首先假设模型空间$\boldsymbol{M}$和数据空间$\boldsymbol{D}$均为线性空间，而且对模型参数和数据在上述空间中的采样没有先验知识，则均匀概率密度函数为：

$$\mu_{\mathbf{M}}(\mathbf{m})=const,\ \mu_{\mathbf{D}}(\mathbf{d})=const \tag{11}$$

假设联系观测数据与模型参数物理规律的不确定性可以忽略：

$$\theta(\mathbf{d}\,|\,\mathbf{m})=\delta(\mathbf{d}-\mathbf{g}(\mathbf{m})) \tag{12}$$

假设关于模型参数$\boldsymbol{m}$的先验信息具有高斯分布的概率密度函数：

$$\rho_{\mathbf{M}}(\mathbf{m})=\mathrm{const}.\exp(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{prior})^{T}\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})) \tag{13}$$

其中，$\mathbf{m}_{prior}$ 为关于模型参数初始值的先验信息，$\mathbf{C}_{\mathbf{M}}^{-1}$ 为模型空间中模型参数的协方差逆算子。

假设数据空间中数据 $\mathbf{d}$ 与实测数据 $\mathbf{d}_{obs}$ 之差满足高斯分布：

$$\rho_{\mathbf{D}}(\mathbf{d})=\mathrm{const}.\exp(-\frac{1}{2}(\mathbf{d}-\mathbf{d}_{obs})^{T}\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{d}-\mathbf{d}_{obs})) \tag{14}$$

其中，$\mathbf{C}_{\mathbf{D}}^{-1}$ 为数据空间中残差数据的协方差逆算子。与（12）结合知道，数据残差实质上由正问题的不确定和观测的不确定性两种因素决定。勘探地震中正问题的不确定性实质上是不可以忽略的，一般地是通过数据的预处理来满足正问题被简单化的假设。

融合上述假设后，模型空间的后验概率密度 $\sigma_{\mathbf{M}}$ 变为：

$$\begin{aligned}&\sigma_{\mathbf{M}}(\mathbf{m})=const.\exp(-S(\mathbf{m}))\\&2S(\mathbf{m})=(\mathbf{m}-\mathbf{m}_{prior})^T\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})+(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})^T\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})\end{aligned} \tag{15}$$

其中，$S(\mathbf{m})$ 称为误差泛函。由单调性可知，后验概率密度函数最大化等价于误差泛函最小化：

$$\hat{\mathbf{m}}=\underset{\mathbf{m}}{arg\ min}\ S(\mathbf{m}) \tag{16}$$

进一步地，误差泛函有以下形式：

$$\begin{aligned}S(\mathbf{m})&=\frac{1}{2}\left\{(\mathbf{m}-\mathbf{m}_{prior})^T\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})+(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})^T\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})\right\}\\&=\frac{1}{2\mathbf{C}_{\mathbf{M}}^{-1}}\left\{\left\|\mathbf{m}-\mathbf{m}_{prior}\right\|^2+\lambda\left\|\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs}\right\|^2\right\}\end{aligned} \tag{17}$$

其中，$\lambda=\mathbf{C}_{\mathbf{D}}^{-1}/\mathbf{C}_{\mathbf{M}}^{-1}$，高斯白噪音情形下 $\lambda=\sigma_{\mathbf{M}}^2/\sigma_{\mathbf{D}}^2$（信噪比）。此时（17）式退化为 Tikhonov 正则化反演。

基于 Bayes 估计理论，可以从概率统计的角度理解反演成像，对反演结果进行精度评价。数据协方差（逆）阵和模型协方差（逆）阵携带了数据和模型参数的不确定性信息，据此可以更方便地进行正则化处理。另外，只有当观测数据与模型参数之间存在弱非线性关系，后验概率密度函数为 Gauss 或广义 Gauss 公式时，后验概率最大化以及据此发展出的代价函数最小化的方法才有可能进行实际计算。假若观测数据与模型参数之间是较强的非线性关系，（15）式给出的后验概率密度函数及对应的误差泛函会是多峰的、甚至极多峰的，此时用线性的或非线性的最小二乘方法给出的解严重依赖于初始值。当前，勘探地震中的旅行时速度层析成像和一次波叠前偏移成像是比较典型的线性反演问题。LS_RTM 在初始值提供较好的情况下应该是一个比较凸的反问题，但非“两宽一高”且信噪比低的数据情形下它也不一定是凸性较好的反问题。因此，从地震数据的实际情况出发，构建至少是凸的反问题进行地震波反演成像是基本的逻辑。

3 数据空间向模型参数空间的投影算子

从泛函空间和线性投影算子的角度讨论反问题是分析学家更喜欢的思想路线。这与上述统计意义下反问题的理论框架高度是一致的。

首先从算子投影的角度，引入数据和模型参数的原空间以及它们各自的对偶空间。假设 $\mathbf{M}$ 和 $\mathbf{D}$ 是两个线性空间，$\mathbf{M}^*$ 和 $\mathbf{D}^*$ 分别为其对应的对偶空间。原线性空间中的元素 $\mathbf{m}\in\mathbf{M}$ 和 $\mathbf{d}\in\mathbf{D}$，各自对偶空间中的元素 $\mathbf{m}^*\in\mathbf{M}^*$ 和 $\mathbf{d}^*\in\mathbf{D}^*$。令 $\mathbf{L}$ 为从原模型空间映射到原数据空间的线性算子，即：

$$\mathbf{d}=\mathbf{L}\mathbf{m} \tag{18}$$

$\mathbf{L}$ 的转置算子 $\mathbf{L}^T$ 定义为从对偶数据空间 $\mathbf{D}^*$ 映射到对偶模型参数空间 $\mathbf{M}^*$ 的线性算子，存在如下的对偶内积关系：

$$\left\langle\mathbf{m},\mathbf{L}^T\mathbf{d}^*\right\rangle_{\mathbf{M}}=\left\langle\mathbf{L}\mathbf{m},\mathbf{d}^*\right\rangle_{\mathbf{D}} \tag{19}$$

假设在线性空间 $\mathbf{V}$ 中存在一个加权算子 $\mathbf{W}$，该算子是一个线性、对称、正定的映射算子，它把线性空间 $\mathbf{V}$ 中的元素 $\mathbf{v}$ 映射到其对偶空间 $\mathbf{V}^*$ 中。线性空间 $\mathbf{V}$ 中任意两个向量 $\mathbf{u},\mathbf{v}$ 定义的标量积为：

$$\left(\mathbf{u},\mathbf{v}\right)=\left\langle \mathbf{Wu},\mathbf{v}\right\rangle \tag{20}$$

通过加权算子 $\mathbf{W}$ 建立了原空间与对偶空间的映射，其中 $\mathbf{W}$ 算子的逆算子 $\mathbf{C}=\mathbf{W}^{-1}$ 称为协方差算子。令 $\mathbf{v}$ 及 $\mathbf{v}^*$ 分别为原空间 $\mathbf{V}$ 与对偶空间 $\mathbf{V}^*$ 的两个元素，则有：

$$\begin{aligned}\mathbf{v}^*&=\mathbf{Wv}\\ \mathbf{v}&=\mathbf{Cv}^*\end{aligned} \tag{21}$$

因此可以说协方差矩阵算子将一个线性空间投影到其对偶空间中。而转置算子定义了从数据对偶空间向模型参数对偶空间的映射关系。标量积的定义引入了协方差算子，建立起原空间的元素到对偶空间元素的映射关系，反之亦然。

从映射的角度看线性反演，核心是建立起原数据空间向原模型参数空间的映射关系。这需要引入伴随算子的概念。$\mathbf{L}$ 的伴随算子 $\mathbf{L}^*$ 定义为从原数据空间 $\mathbf{D}$ 映射到原模型参数空间 $\mathbf{M}$ 的线性算子，且对所有的元素 $\mathbf{m}\in\mathbf{M}$ 和 $\mathbf{d}\in\mathbf{D}$ 满足：

$$\left(\mathbf{L}^*\mathbf{d},\mathbf{m}\right)_{\mathbf{M}}=\left(\mathbf{d},\mathbf{Lm}\right)_{\mathbf{D}} \tag{22}$$

经过标量积和转置算子的运算可得到：

$$\left(\mathbf{d},\mathbf{Lm}\right)_{\mathbf{D}}=\left\langle \mathbf{C}_{\mathrm{D}}^{-1}\mathbf{d},\mathbf{Lm}\right\rangle_{\mathbf{D}}=\left\langle \mathbf{L}^{T}\mathbf{C}_{\mathrm{D}}^{-1}\mathbf{d},\mathbf{m}\right\rangle_{\mathbf{M}}=\left(\mathbf{C}_{\mathrm{M}}\mathbf{L}^{T}\mathbf{C}_{\mathrm{D}}^{-1}\mathbf{d},\mathbf{m}\right)_{\mathbf{M}}=\left(\mathbf{L}^{*}\mathbf{d},\mathbf{m}\right)_{\mathbf{M}} \tag{23}$$

因此伴随算子为：

$$\mathbf{L}^{*}=\mathbf{C}_{\mathbf{M}}\mathbf{L}^{\mathbf{T}}\mathbf{C}_{\mathbf{D}}^{-1} \tag{24}$$

最终从原数据空间 $\mathbf{D}$ 映射到原模型参数空间 $\mathbf{M}$ 的映射过程为：

$$\mathbf{m}=\mathbf{L}^{*}\mathbf{d}=\mathbf{C}_{\mathbf{M}}\mathbf{L}^{\mathbf{T}}\mathbf{C}_{\mathbf{D}}^{-1}\mathbf{d} \tag{25}$$

据此可以看到，数据和模型的协方差（逆）算子连接起了原空间与对偶空间的桥梁。

从线性投影算子基本理论知道，若伴随算子是原数据空间 $\mathbf{D}$ 到原模型参数空间 $\mathbf{M}$ 的一对一的映射算子，反问题的解就是唯一的。此时，伴随算子是可逆算子。实际上，若伴随算子具备良好的局部支撑性（或紧性）已经非常好了。弱非线性的、凸的反问题应该对应这样的伴随算子。

即便对弱非线性反问题，对反演解引入正则化约束也是非常重要的。对数据引入预条件处理很多情况下也必不可少。从上面分析可以看出，数据协方差 $\mathbf{C}_{\mathbf{D}}$ 与模型协方差 $\mathbf{C}_{\mathbf{M}}$ 在模型参数的求解中具有特别重要的意义，关于数据和模型参数的先验信息很多都蕴含在这两个协方差算子中。所谓的关于数据的正则化和关于模型参数的正则化处理主要就是在考虑如何利用好这两个协方差矩阵。

一般地，人们并不区分转置算子与伴随算子。但是从上述讨论中的区分可以看出协方差阵的本质含义。已知协方差阵是对称正定阵，可以对其进行正交分解。数据协方差阵对数据向量进行变换可以认为是对数据进行了特征表达。模型参数对应的协方差阵有同样的作用。因此，可以认为对偶空间是特征数据和特征模型参数各自形成的线性空间。特征数据和特征模型参数之间的关系是更紧密的、或者更线性的、或更弱非线性的。转置算子建立起它们之间的更为一对一的投影关系。这对提高反演结果的稳定性是非常有益的。这应该是反演问题中正则化处理最本质的内涵了。

4 勘探地震中地震波反演成像具体思想方法

称地震波反演成像是勘探地震中最关键的技术环节一点也不为过，它给出了地震数据采集质量要求的理论基础，提供了储层描述所要求的基础数据。它永远是在信息不足情况下给出弹性参数“最佳”估计的过程。下面具体分析一下当前地震波反演成像的具体思想与方法，揭示它们存在的根本问题。

4.1 非线性正问题的线性化

任何反问题都奠基在正问题上，正问题建立起要反演的参数与实测数据之间的内在联系。要反演的参数和实测数据之间的关系最好是紧密的、线性的，至少是弱非线性的。正是要反演的参数与实测数据之间的关系很弱导致了反问题的极端困难。

在地震波反演中，深入研究要反演的参数对实测数据（或其中某些成分）的依赖关系，进而确定要反演的参数是解决反问题的重要一环，实质上这就是在选择合适的正问题参与反演。选择什么参数（或参数组合）、如何表达参数，决定了选择什么样的正问题，最终决定了反问题的凸性，或者说决定了反问题的难度。

对非线性的正问题 $A(\boldsymbol{m})=\boldsymbol{b}$ 进行 Taylor 展开的线性化，或 Born 近似的线性化，在地震波反演成像中是非常常见的。

$A(\boldsymbol{m})=\boldsymbol{b}$ 的 Taylor 展开：

$$A(\boldsymbol{m}_0+\Delta\boldsymbol{m})=A(\boldsymbol{m}_0)+\left.\frac{\partial A}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}_0}\Delta\boldsymbol{m}+\frac{1}{2}\left.\frac{\partial^2 A}{\partial \boldsymbol{m}^2}\right|_{\boldsymbol{m}_0}\Delta\boldsymbol{m}^2+\cdots\cdots \tag{26}$$

上式对应的线性化形式为：$A(\boldsymbol{m}_0+\Delta\boldsymbol{m})-A(\boldsymbol{m}_0)\approx\left.\frac{\partial A}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}_0}\Delta\boldsymbol{m}$，可以重写为：

$$\delta\boldsymbol{b}\approx\left.\frac{\partial A}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}_0}\Delta\boldsymbol{m} \tag{27}$$

其中，$\left.\frac{\partial A}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}_0}$ 是一个 Jacobi 矩阵，也被称为 Frechet 微商或灵敏度矩阵或点扩散函数（PSF）。

上述展开式也可以用如下的 Born 近似来解释：

$$\begin{aligned}G(\boldsymbol{m}_0+\Delta\boldsymbol{m})=&G_0(\boldsymbol{m}_0)+G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G^R(\boldsymbol{m}_0)\\&+G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G'(\boldsymbol{m}_0)\Delta\boldsymbol{m}_2G^R(\boldsymbol{m}_0)\\&+\cdots\cdots\end{aligned} \tag{28}$$

其中，$G_0(\boldsymbol{m}_0)$ 对应透射波，$G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G^R(\boldsymbol{m}_0)$ 对应一次反射或散射波，$G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G'(\boldsymbol{m}_0)\Delta\boldsymbol{m}_2G^R(\boldsymbol{m}_0)$ 对应一阶多次散射波。如图所示。

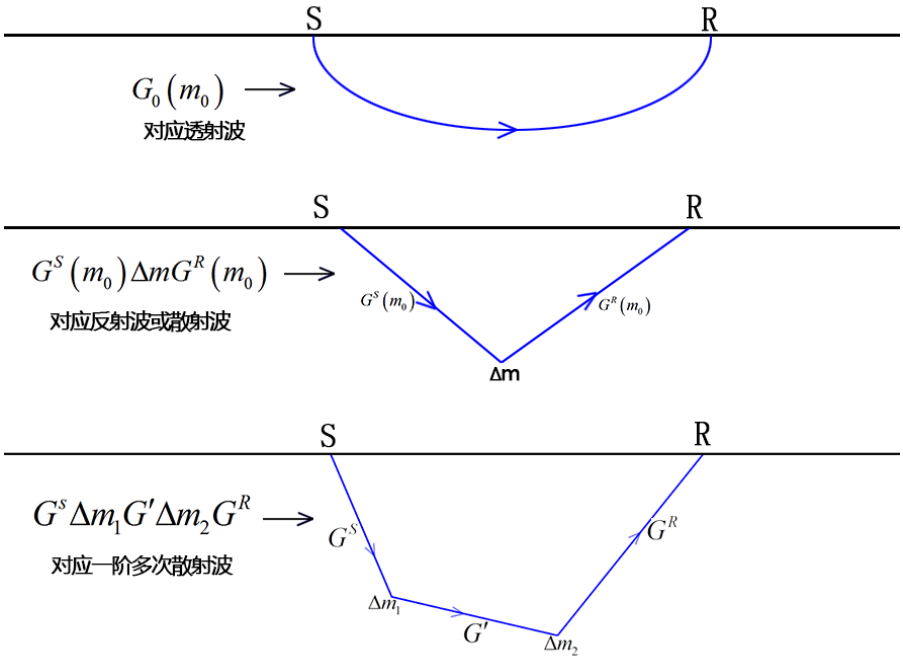


图 1 Born 近似后透射波、一次散射波和一阶多次散射波示意图

上式对应的线性化形式为：

$$G(\boldsymbol{m}_0+\Delta\boldsymbol{m})-G_0(\boldsymbol{m}_0)=G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G^R(\boldsymbol{m}_0) \tag{29a}$$

上式可以重写为：

$$\delta G=G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G^R(\boldsymbol{m}_0) \tag{29b}$$

线性化的正问题可以统一地用矩阵方程 $A\boldsymbol{m}=\boldsymbol{b}$ 来表示。Born 近似的物理意义十分清晰，因此得到了更广泛的应用。

实质上，（27）式和（29b）式代表的线性关系是非常广泛的。任何物理系统的线性表达都可归结为矩阵方程 $A\boldsymbol{m}=\boldsymbol{b}$ 的形式。L2 范数下把矩阵方程的求解问题转化为变分问题进行求解是非常经典的做法。在线性系统和高斯白噪声假设下，大量的文献从不同的角度把此类问题已经完全探讨清楚了。

勘探地震中的层析成像（包括反射波 FWI（RFWI））和 LS_RTM 之所以还存在问题，根本原因是实测数据（看起来是巨量数据）对弹性参数反演而言是非常不完备的，勘探地震系统是非常复杂的非线性系统，人为的线性化过程并不能很好地建立起对实测数据的预测。当然，预测误差不满足高斯白噪假设是必然的，震源函数未知更提升了弹性参数反演的非线性性。

4.2 线性化正问题情形下误差泛函的梯度

对于线性化正问题和非线性正问题情形下定义的 L2 范数误差泛函而言，其梯度的含义是十分不同。分析梯度的差异，对于深入理解梯度迭代类的反问题的本质很有帮助。

勘探地震中，对于线性化正问题 $A\boldsymbol{m}=\boldsymbol{b}$，线性化后的正问题仅表达从源到散射点和从散射点到检波点的传播过程（$A=GG^T$）。因此，线性正问题情形下 L2 范数误差泛函的梯度，有如下具体形式：

$$\mathcal{J}(\boldsymbol{m})=\frac{1}{2}(A\boldsymbol{m}-\boldsymbol{b})^2 \tag{30a}$$

$$\begin{aligned}\nabla\mathcal{J}(\boldsymbol{m}) &= A^T(A\boldsymbol{m}-\boldsymbol{b}) \\ &= A^T\delta\boldsymbol{u}\end{aligned} \tag{30b}$$

把 A^T 拆分成两部分，

$$A^T = \left(GG^T\right)^T = GG^T \tag{31}$$

把（31）代入（30）式得到：

$$\nabla\mathcal{J}(\boldsymbol{m}) = GG^T\delta\boldsymbol{u} \tag{32}$$

其中，G 对应源端下行波的 Green 函数，G^T 对应检端的地表残差波场（或扰动波场）反向外推。如图 1 中间图所示。二者相关，得到梯度项$\nabla\mathcal{J}(\boldsymbol{m})$。它就是常规的偏移响应。对于（29b）式定义的线性正问题，只能产生这样的偏移响应。如下图所示。

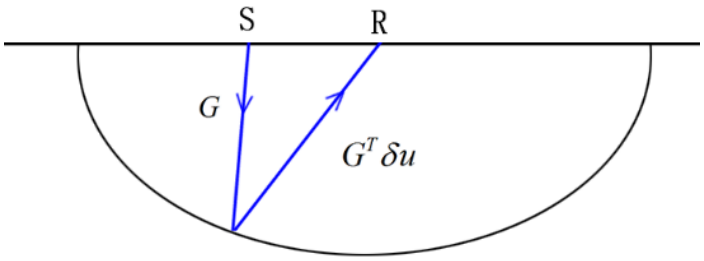


图 2 一次散（反）波对应的梯度响应

4.3 非线性正问题情形下误差泛函的梯度

利用非线性正问题 $A(\boldsymbol{m})=\boldsymbol{b}$ 描述的波传播过程进行梯度计算，其中包含了上述级数展开项（（28）式）中二阶（及二阶以上）项对应的波现象。这是用 $A(\boldsymbol{m})$ 与用 $A\boldsymbol{m}$ 处理正问题时出现梯度计算结果很不同的根本原因。勘探地震中的非线性反演也可以从这个角度进行解释。

用 $A\boldsymbol{m}$ 计算正问题本质上在用 $u_s' = \int G_S\Delta\boldsymbol{m}G_R d\boldsymbol{x}$ 来计算一阶散（反）射波。

用 $A(\boldsymbol{m})$ 计算正问题是在用 $u^{\text{total}} = u(\boldsymbol{m}_0) + u_s^1 + u_s^2 + \ldots\ldots$ 计算各阶散（反）射波，其中

$$\begin{aligned}u_s^1 &= G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G^R(\boldsymbol{m}_0) \\ u_s^2 &= G^S(\boldsymbol{m}_0)\Delta\boldsymbol{m}G'(\boldsymbol{m}_0)\Delta\boldsymbol{m}_2G^R(\boldsymbol{m}_0) \\ &\cdots\cdots\end{aligned} \tag{33}$$

对于非线性正问题情形下的误差泛函 $\mathcal{J}(\boldsymbol{m}) = \frac{1}{2}\left(A(\boldsymbol{m})-\boldsymbol{b}^{obs}\right)^2$，一般地用变分法导出其梯度公式：

$$g(m(\boldsymbol{x})) = \sum_{\boldsymbol{x}_s}\sum_t \frac{1}{v^2(\boldsymbol{x})}\frac{\partial^2 u(\boldsymbol{x}_s;\boldsymbol{x},t)}{\partial t^2} z(\boldsymbol{x}_s;\boldsymbol{x},t) \tag{34}$$

其中，$u(\boldsymbol{x}_s;\boldsymbol{x},t)$ 是解 Cauchy 问题得到的正时间方向传播的波场，$z(\boldsymbol{x}_s;\boldsymbol{x},t)$ 是沿逆时间方向，解边值问题得到的波场。把正向波场外推方程和反向波场外推方程罗列如下。正向波场外推方程：

$$\begin{cases}\left(\dfrac{1}{v^2(\boldsymbol{x})}\dfrac{\partial^2}{\partial t^2}-\Delta\right)u(\boldsymbol{x},t)=-\delta(\boldsymbol{x}-\boldsymbol{x}_S)S(t)\\ u(\boldsymbol{x},t)\big|_{\boldsymbol{x}=\Omega}=f_b(\Omega,t)\end{cases} \tag{35a}$$

其中，$f_b(\Omega,t)$代表边界条件。一般用 PML 边界条件。Ω代表计算波场在空间上的边界。

反向波场外推方程：

$$\begin{cases}\left(\dfrac{1}{v^2(x)}\dfrac{\partial^2}{\partial t^2}-\Delta\right)z(\boldsymbol{x},t)=0\\ z(\boldsymbol{x},t)\big|_{\boldsymbol{x}=\boldsymbol{\eta}}=r(\boldsymbol{\eta},t)=u^{cal}(\boldsymbol{\eta},t)-u^{obs}(\boldsymbol{\eta},t)\end{cases} \tag{35b}$$

其中，$\boldsymbol{\eta}$代表观测点坐标值。$u^{obs}(\boldsymbol{\eta},t)$代表观测点处的实测波场；$u^{cal}(\boldsymbol{\eta},t)$代表观测点处的正演合成波场。进行波场残差外推时，仅有观测点处有波场残差，因此，才有（35b）这样的边界条件的定义。实际上，波场逆时外推并不限于这样的边界条件。

梯度是两波场$u(\boldsymbol{x}_s;\boldsymbol{x},t)$和$z(\boldsymbol{x}_s;\boldsymbol{x},t)$在空间任一点的相关并对时间积分得到的，它是参数扰动场，因此有如下迭代公式：

$$\boldsymbol{m}^{k+1}=\boldsymbol{m}^k+\boldsymbol{\mu}_k\left.\frac{\partial J}{\partial \boldsymbol{m}}\right|_k \tag{36}$$

但是，当$A(\boldsymbol{m})$是非线性的正问题时（注：此处非线性是指波场与地下参数之间的关系是非线性的！包括了二阶及二阶以上的扰动场成分，或称多次波成分），梯度中包含的成分是很复杂的。

当然，我们应该首先弄清透射波，一次反（散）射波和二阶反射波的梯度响应。这三种类型的波对地震波反演成像而言是最核心的波场（特征波场），它们产生参数反演中贡献最显著的梯度成分。从层状介质的角度解释清楚多次反射波对应的梯度成分对理解非线性反演成像会有很大帮助。因为多次散射波对应的梯度成分更难分析清楚，从多次反射波开展分析会容易一些。

勘探地震中的透射波包括直达波、折射波和 Diving Wave，如下图所示。

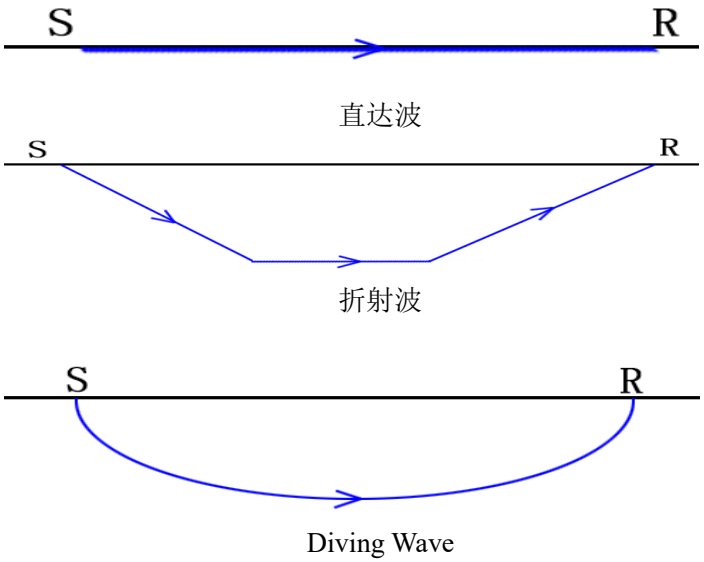


图 3 探勘地震地表观测数据对应的三种透射波示意图

为了有直观的印象，把三种重要波现象对应的梯度响应图示如下。透射波梯度：

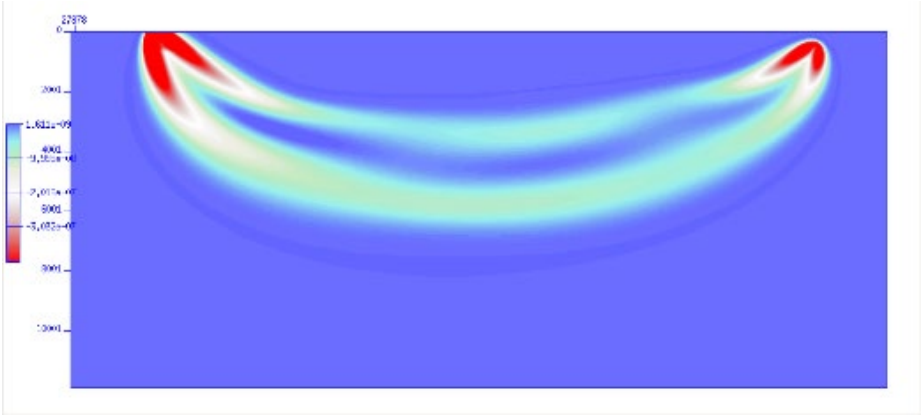


图 4a 透射波层析成像的梯度响应

一阶反（散）射波梯度：

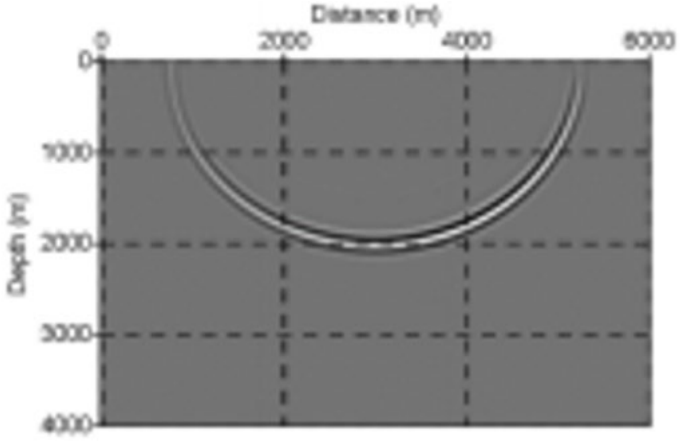


图 4b 一次散（反）射波偏移成像的梯度响应

二阶反射波梯度：

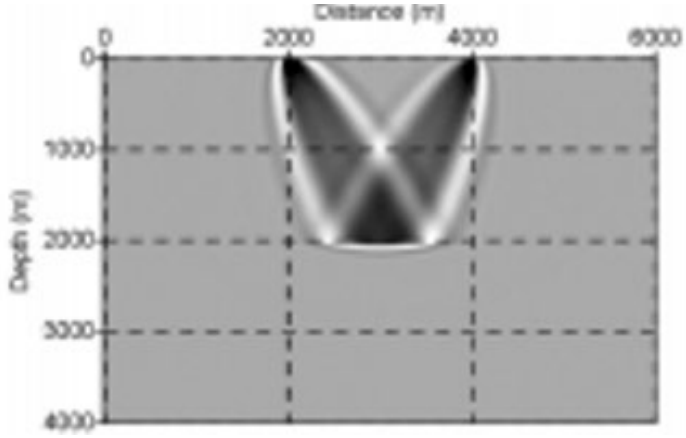


图 4c 二阶反射波层析成像的梯度响应

准确地讲，反射界面上相邻散射点之间二次散射波对应的梯度是所谓的“兔子耳朵”形状。如图 4c 所示。“相邻”指的是处在 Fresnel 半波带内的意思。我们常把这种梯度混淆成一次反射波对应的梯度，严格讲，这种解释是不严谨的。孤立的二次散射波的梯度响应不是所谓的“兔子耳朵”形状。

5 线性反演成像向非线性反演成像的转变

①从上述分析可知，所谓的线性化反演应该是针对单一波现象的。即便是弱非线性情况下的反演成像，不同波现象之间的梯度响应的串扰，也会导致迭代反演的收敛性降低。因此，在解开各种波现象之间的梯度串扰的基础上进行迭代反演成像是加速反演收敛的基本思路。WPI 提出的特征波成像方法就是该思想下的一种具体做法。矢量波成像也必须在波场分解的基础上构建单一波现象对应的梯度成分进行迭代反演，譬如 S 波梯度项和 P 波梯度项。总之，要求参数扰动引起的波场扰动是一对一的、比较敏感的，是反演成像收敛快且精度高的本质物理基础。

②声波方程正问题中，不考虑密度变化会引入错误的梯度；不考虑多次波现象会引入错误的梯度，不考虑（35）式中合适的定解问题会引入错误的梯度。另一方面，对噪声假设不合理会引入错误的梯度，对模型正则化不合理会引入错误的梯度。当然，子波不正确也会引入错误的梯度；背景速度的不正确也会引入错误的梯度。最后，对实际波传播过程做声波近似假设也会引入错误的梯度。

可以看出，梯度是弹性参数估计（FWI/LS-PSDM）中的核心问题，上述罗列出来的梯度项的错误若能得到修正，就可以改善当前反演成像的结果。这是当前反演成像最核心的问题。

③但很遗憾，能用梯度导引类的反演成像方法进行弹性参数估计的反问题要么是线性反演问题，要么是接近线性的反演问题。而且此时的后验概率密度函数是 Gauss 或接近高斯的问题。即便上述讨论中提到把多次散（反）射波对应的各自的梯度算出来进行迭代反演，也不等同于解非线性问题。

④事实上，我们对后验概率密度接近高斯的一类（准）线性反问题的求解的各方面都已经明白了（至少从概念上已经完全明白了此类反演的问题）!

⑤地震波反演成像的研究必然要转移到后验概率密度 σ_M 是非高斯、且正问题是非线性情况下的参数估计上。到目前为止，上述非线性非高斯情形下的反问题缺乏有效的解法。大数据分析领域、现代信号与图像分析领域等等都是如此。

⑥上述梯度计算中的 Cross-talk（各种波现象之间的干（串）扰产生的假梯度项）是 FWI 中的核心问题。它与反问题的非线性性以及反问题的迭代收敛性都存在密切关系。在层状介质假设下，也许是个切入点，通过利用各阶多次反射波对应的梯度，部分地解决勘探地震中非线性的弹性参数估计问题。实测数据中多次波的强弱从物理本质上展现了此时弹性参数估计问题非线性的强弱，多次波越发育，此时的弹性参数估计问题的非线性性越强。至少在层状介质情况下，依据多次波发育程度，在 $v_B(\boldsymbol{x})+R_\theta(\boldsymbol{x})$ 的模型描述下，用解耦 Cross-talk 的方式计算各波现象对应的梯度，结合合适的正则化方法，进行较强非线性情形下的弹性参数估计。

6 Monta Carlo 非线性反演与深度学习网络

①Monta Carlo 非线性反演是在巨大的模型空间中进行探测，寻找后验概率密度函数比较大的区域（或称重要区域）。探测机制主要包括：随机探测（不是沿梯度方向）加 Metropolis 准则。由于没有引入沿梯度方向的寻优方式，它具备很强的非线性问题寻优能力，至少理论上是这样的。存在的问题是：计算效率还是太低。Metropolis 准则中的接收（或拒绝）概率的计算存在问题。

②Monta Carlo 非线性反演在巨大的模型空间中的演化（探测、寻优）过程，结合 Metropolis 准则，其解决非线性问题的意义是非常清楚的。其每一步的演化都在既定的策略下进行。多点探测的 Monta Carlo 方法（譬如遗传算法）非常类似于多层网格系统。当前迭代步向下一迭代步的演化，类似两层网格的互联。全部迭代步的演化类似于一个深度（学习）网络系统。

③目前的深度学习神经网络，号称可以解决非线性寻优问题，但是这个过程似乎是个黑匣子。是否可以通过与 Monta Carlo 非线性反演中多点探测逐步演化寻优过程的对比，解释深度网络的非线性寻优能力？或者说用深度神经网络解决非线性寻优问题？

④关于非线性反演成像，目前没有公认有效的求解方法。FWI 是个非线性反演成像问题，但使用的 Newton 类逐步线性化的迭代算法，并不能保证收敛到全局最优解。FWI 远不能广泛实用的现状促使我们发

展出新的实现方案。用多点探测的 Monta Carlo 方法模拟深度学习网络也许是一条新路子。

7 “两宽一高”数据对非线性反演成像的贡献

“两宽一高”地震数据采集为高精度的地震波反演成像提供了更好的基础数据。从数据或波场的角度如何提炼非线性反演成像的有效方法？

①高密度采样的数据有利于噪音压制，展宽有效频带，尤其是低频端。有效的低频端数据（譬如 1Hz-8Hz）可以降低参数估计的非线性性。有效去噪后的数据体能使得反演中的数据误差更接近 Gauss 分布，从而提高反演成像的精度。

②宽方位长偏移距及高密度采样数据能提高数据对地下地质体的照明均匀度和照明密度，能有效地提高层析反演成像的精度。

③“两宽一高”地震数据中透射波反演成像能给出更好的初始弹性参数场，有效地降低反问题的非线性性。

④“两宽一高”地震数据提供更好的数据预条件基础，数据正则化的加入能提高反演成像的精度。

⑤基于“两宽一高”地震数据及更好的背景速度的高波数成像结果，在现代图像处理方法的辅助下，能提供更精确的弹性参数结构信息。精确的结构信息约束下的反演成像会有更高的精度。

“两宽一高”数据一定能促进反演成像质量的提高。尽管如此，这并不能解决强非线性反演成像问题。强非线性反演成像问题的解决主要还是依赖算法方面的创新。当前，量子计算技术也备受关注，预计不久就可以进入实用，笔者认为较强非线性反问题的解决主要还不是计算效率方面存在瓶颈而是缺乏保证收敛到有意义的解的算法。笔者还是认为，“两宽一高”能进行更好的数据或波场的特征化处理，基于这样的特征波场提出更凸的反演方法，是解决非线性非高斯问题的途径之一。

8 结论与讨论

地震波反演成像已经取代偏移成像成为了勘探地震学中地震波成像方法技术的发展方向。而且在波场预测算子线性化、预测误差高斯分布、L2 范数定义预测误差的前提下，线性反演问题所有内涵全部是清楚的。但是，FWI/LS_RTM 为代表的反演成像方法还是不能广泛地实用化。当然，即便高斯线性情形下的反演成像也存在很多实用化的问题。但是，这大多也是由于弹性参数估计反问题的非线性与非 Gauss 性引起的。

非线性反问题的物理实质到底是什么？笔者认为：非线性反演困难的根源是波场与要反演参数之间的关系不是线性的，在“两宽一高”数据情形下，这个问题更凸显。更进一步地，笔者认为：勘探地震中反问题的强非线性对应于地下介质中存在强反射系数反射层或强散射体，对应于小尺度、强变化异常体的存在，这是参数估计反问存在强非线性的物理根源。只有从这个物理根源出发，构建合适的反演方法才是解决非线性反问题的可行路线。不存在统一的、可行的数学框架解决千变万化的实际物理问题。只能是基于对物理问题的认识和既有的数学方法，构建合理的非线性反演技术路线，按一个技术路线来解决问题。

FWI 这类黑匣子的操作，每个迭代步要求线性化处理。当前迭代步的初始值如果不满足线性化假设，这个线性化处理不能保证收敛。

一般的非线性反演应该回避梯度类的方法，而是用 Monta Carlo 类的方法，但 Monta Carlo 类方法对付高维参数的反演是极为困难和不现实的。

当前梯度类方法（对应 L2 范数误差能量定义方法）蕴含的假设是正问题是线性的（解法方程），或弱非线性的（迭代法求解）。满足上述情形的线性 Gauss 反问题的解是可以满足实用要求的。当前，简单介质情形下的地震波成像处理的实际已经充分地证明了这一点。但复杂地表、复杂构造与复杂储层情况下如何提高成像精度？

针对非线性反问题的求解，需要循序渐进的思考如下问题：①强反射层状介质情形下的反演，就是包含层间多次波的反演，如何构建合适的非线性反演方法路线？②强散射体存在情形下，就是包含多次散射波

的情形下，如何构建合适的非线性反演方法路线？③二者同时存在情形下，如何构建合适的非线性反演方法路线？④非线性反演时如何进行参数场的描述？基于参数场描述方法的参数场与数据场之间关系的合适描述方法是解决非线性反演问题的核心。⑤显式地分开特征波场然后计算各自的梯度或是隐式地计算各特征波场的梯度是解决非线性反问题的一个关键问题。

总之，线性 Gauss 情形下的参数反演从方法研究看，已经差不多走到尽头。非线性反演还没有公认的方法路线。

参考文献

[1] Albert Tarantola, 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM
[2] Peter Bühlman and Sava van de Geer, 2011, Statistics for High-Dimensional Data---Methods, Theory and Applications, Springer
[3] Shahpoor Moradi, Daniel Trad, and Kristopher A. Innanen, 2018, Quantum computing in geophysics: algorithms, computational costs and future applications: 88th Annual International Meeting, SEG, Expanded Abstracts, 4649-4653.
[4] 王华忠，罗飞，冯波，2016，状态变量独立的地震波反演成像.波现象与反演成像研究组年度论文集，同济大学海洋与地球科学学院.

基于二阶统计量的信号分析方法存在问题及解决方案

王华忠

同济大学海洋与地球科学学院波现象与反演成像研究组，上海，200092

【摘要】“两宽一高”地震数据采集为高精度的参数估计或称为高精度的地震波反演成像提高了更好的数据基础。另一方面，反演成像需要更高信噪比的数据才能保证其收敛性。事实上，“两宽一高”的数据同样为去噪提供了更好的数据基础，而且把去噪方法推进到高维空间中能更好地把握数据中蕴含的信号特征，去噪更为彻底。地震信号分析理论中，实际观测的地震数据被视为确定性信号叠加上各种随机噪音形成所谓的随机信号。随机信号分析是基于随机信号的概率分布特征的。针对离散的实测数据的信号分析是基于它的统计特征的。一维随机信号的统计分析是常见的，并假设随机信号是平稳的和满足 Gauss 分布的。众所周知，地震信号是高维的,结构特征蕴含其中。基于随机信号统计特征的去噪方法或滤波器设计要充分地体现地震信号的结构特征。在 Bayes 框架下，用信号模型对实测信号进行预测，预测误差假定是 Gauss 的（尤其是 Gauss 白噪的），信号模型是线性的，优化求解得到最佳滤波器，从而实现观测数据的去噪音，是另一种经典的地震数据去噪思想。上述两种经典的去噪方法本质上都是仅仅利用了随机信号的二阶统计特征。本文分析了基于随机信号的统计特征和基于 Hilbert 空间中函数逼近进行去噪的基本思想，指出它们本质上的一致性；阐明了在高维数据空间中进行去噪的本质必然性；分析了勘探地震数据中振幅与相位的物理意义及在地震信号分析中对振幅和相位变化做施加的假设条件的逻辑含义；最后本文分析了基于二阶统计量的数据分析方法或称为信号分析方法存在的问题，并对信号分析方法的发展方向给出了预测。

【关键词】高维随机信号，实测离散数据，统计特征，信号预测器，信号分析，去噪音

1 引言

信号分析到底要分析什么？信号分析的思想方法是什么？

信号分析的目的是获取高维随机信号中包含的各个层次和各个方面的特征，也可以讲是对信号进行最佳的预测。信号分析的本质思想是实测离散数据是随机的、符合一定的概率分布函数的、可以用统计的方法来描述的。信号分析的基本方法包括对信号进行建模并在 Bayes 框架下估计信号模型中的参数从而实现对信号的最佳预测（或最佳的特征表达）。

信号处理主要包括两方面的核心内容：信号的建模和 Bayes 估计理论下对信号模型中所含参数的最佳估计。各种具体的信号分析或处理工作就是对上述两种内容的具体应用，譬如去噪、插值、信号恢复等等。

比较而言，信号的建模是更为基础的。在 Hilbert 空间中构建一组（或多组）基函数形成子空间，在子空间中对信号进行基函数线性组合的表达是信号建模的最基本方法或最常用方法。但是，物理世界中实测的数据都是随机的。针对随机数据（或称随机信号）用概率统计的方法进行统计建模，并基于此进行各种信号分析工作也是非常普遍的做法。信号建模是建立在对实测信号基本抽象的基础上的。我们认为实测地震数据是一个典型的五维数据体，可以抽象地认为该地震数据是线性和/或非线性（相位变化）的连续同相轴飘在满足不同的统计特征的随机噪声中，形成所谓的高维随机信号。地震信号分析的对象就是这个高维随机信号，更准确地讲是实测的、高维的、离散的、随机的数据体。这是笔者对实测地震数据建立的概念模型。

更具体地，可以认为实际测量的地震数据是确定性信号叠加上满足各种概率分布函数的随机噪音，即 $\mathbf{u}(x,y,t)=\mathbf{s}(x,y,t)+\boldsymbol{\eta}(x,y,t)$ 。其中，(x,y,t) 表示地震数据的空间时间坐标；$\mathbf{u}(x,y,t)$ 表示实际测量的地震数据；$\mathbf{s}(x,y,t)$ 表示对应的确定性信号（在 Bayes 反演框架下，要预测的信号也是随机的！）；$\boldsymbol{\eta}(x,y,t)$ 是满足某种概率分布的加性随机噪音，并假定各坐标点处的噪音分量为独立同分布的随机变量且与该点处的信号 $\mathbf{s}(x,y,t)$ 统计特征无关。应该注意的是：在实际地震信号的统计处理中，上述信号模型还可以表达在常见的五维空间（譬如 $\mathbf{u}(m_x,m_y,h_x,h_y,t)$，其中 m_x,m_y 为中心点坐标，h_x,h_y 为偏移距坐标）中，在更高维的空间中进行统计信号分析，统计特征会更符合既定的假设，去噪效果会更好。上述方式定义的信号模型的应用是非常普遍的，当前绝大部分的地震信号处理方法的构建都是从此模型出发的。地震信号分析的核心工作就是实现对信号 $\mathbf{s}(x,y,t)$ 的最佳预测或最佳（特征）表达，从而达到最佳去噪、插值、压缩等目的。

从统计意义下建立一个模型实现对信号的预测，就是确定上述概念模型中噪音满足的概率分布。众所周知，联合概率密度函数是刻画随机信号特征的最基本形式。在勘探地震中，最基本的随机噪音类型主要有两类：满足 Gauss 分布（包括广义 Gauss 分布）的和满足脉冲分布的。根据这样的认识设计出了各种各样的 Gauss 加权类的信号预测器（或称滤波器）和中值类滤波器。当然，随机信号的二阶统计特征，主要是自相关函数与功率谱密度；互相关函数和互功率谱密度，可以分析信号的功率（能量）谱、功率谱对应的基函数；信号的相关性及独立性（相似系数）、信号的相干性（相干函数）。当前，绝大部分的信号分析工作都是基于上述 4 个典型的二阶统计量完成的。即便是推广到非平稳信号情形下的时频分析，也主要是基于上述二阶统计量的拓展，譬如时间域波包的到达和延续；频率域波包的主频和延续就是用期望和方差定义，并且绝大部分是时频分析的算法都是基于二阶统计量推演出来的。

依据最大似然估计和 Hilbert 空间中的函数逼近理论，建立了信号（本质上是一个高维函数）的最佳预测或特征表达的系列方法，其中的核心问题是针对复杂多变的实测数据 如何在 Hilbert 空间中选择合适的基函数才能实现对数据中所蕴含信号的最佳预测或特征提取。如果基函数选择不合适，所谓的最佳预测或特征提取是做不到的。因此，发展了各种自适应数据变化的最佳预测或特征提取方法，尤其是基于字典（若干基函数构成的基函数字典）而不是基于单一基函数期望能最佳地预测信号。所谓 Learning From Data 的基本思想就是利用从数据中感知到的基函数（或基函数族）在某种范数最佳的意义下实现对信号的最佳预测。以深度学习为代表的机器学习是 Learning From Data 的核心思想方法。但是，从数据中感知其中所包含的信号的基本特征是 Learning From Data 的本质。据此可以看出，从实测数据中最佳地提取信号特征是现代信号分析的核心关注。

但是，我们 Learning From Data 的能力并没有本质的提高！

从 Gauss 线性情形下的信号分析发展到广义 Gauss 弱线性情形下的信号分析，从满足广义 Gauss 分布假设的数据中感知信号特征的能力有一定的提升是最近十几年信号（图像）分析的最大进步，凸优化问题的解法和计算机能力的提升促进了这种进步。仅此而已！

笔者认为：真正突破基于二阶统计量的信号分析思想与方法的局限，才能获得信号分析能力的大幅度提升。

但是，如何能做到？

2 统计信号分析

实测数据是确定性信号叠加上满足各种分布特征的随机噪音，高维的确定性信号具备相干性。实测数据被视为随机过程。不引起误解的情况下直接称要分析的信号是随机过程。这是对实测数据的基本的概念认识。正是基于这样的概念认识，才有了统计信号分析的理论与方法。

另一个重要的基础认识是任何实测信号都是观测到的物理系统的响应结果，这个响应结果必然是随机的，任意两次观测不可能是完全重复的。理论上，我们假设其中蕴含的信号是确定性的，可以由确定性的物理规律预测，该预测关系一般用高维的褶积关系来定义，该褶积关系是信号预测模型。勘探地震中，我们用地震波正演来进行波场（或数据中信号）的预测。

对于随机噪音（随机过程）只能用联合概率密度函数来描述。联合概率密度函数本质上是我们对随机噪音建立的数学模型，试图对随机噪音的分布特征进行模拟。不是预测它的数值大小，而是模拟随机噪音出现的概率。但是，高维随机过程的联合概率密度函数的确定只有理论意义，实际情况下不可能真的从实测数据中知道高维随机过程的联合概率密度函数。但是，高维随机过程的联合概率密度函数对于我们进行高维随机过程的分析具有重大的指导意义。

实际上，对于随机的实测数据我们总是用统计的思想与方法对它进行分析。最基本地，就是从理论上定义随机过程的各阶统计量。当然，最基本是就是一阶和二阶统计量。众所周知，一阶统计量是随机过程的期望值。二阶统计量主要包括时间域（或空间域）的自协方差（自相关）函数和互协方差（互相关）函数，期望值为零的时，相关函数和协方差函数是等价的；以及频率域（或波数域）的自功率谱函数和互功率谱函数。

为了简化对随机的实测数据的（概率）统计分析，我们总是对它施加一定的假设条件。随机过程的广义平稳（或二阶平稳）假设是很常用的。所谓的二阶平稳就是指随机过程的二阶统计量不随时间（或空间）变

化。这会给随机过程的统计分析带来很大的方便。另一个常用的假设是各态遍历性的假设，满足各态遍历性假设的随机过程可以用它的一次实现的统计性质来代替多次实现的统计性质。各态遍历性一定建立在满足平稳性假设的基础上的。满足各态遍历性的随机过程就可以用一次较长时间的观测代替多次观测。用多次观测数据进行统计分析一般情况下是做不到的，多次实际采样代价过于高昂了。对随机实测数据进行统计分析时，必要要知道已经施加了上述先验假设了。但是，往往会忽略掉这些先验假设。

对随机的实测数据进行统计分析时，实际上还要进一步施加先验假设。一般地，就是随机过程满足近似 Gauss 分布的假设。众说周知，随机过程满足的概率密度函数有很多种，譬如：Beta（β），二项分布，Cauchy 分布，卡方（Chi-Square）分布，指数分布，F（Fisher）分布，伽玛（γ）分布，几何分布，超（Hypergeometric）几何分布，正态（Gauss）分布，泊松（Poisson）分布，Student's (t)分布，威布尔（Weibull）分布。如下图所示。但是，在实际的地震数据分析和反演成像中，一般地假定实测数据满足正态分布。Cauchy 分布和 Student's (t)分布的假设也逐渐开始应用。

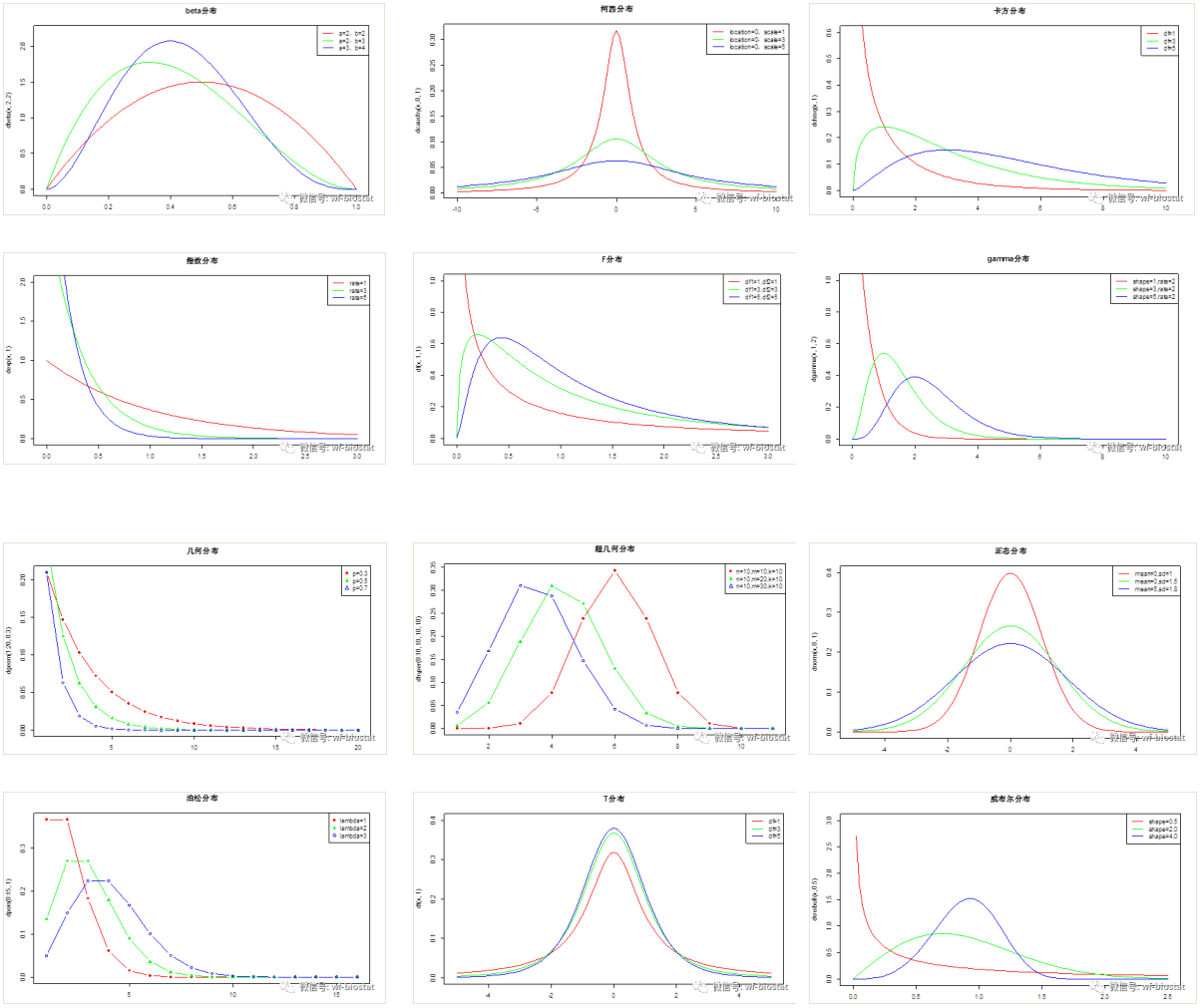


图 1 各种常用的概率分布函数

可以看出，正态分布、Cauchy 分布和 Student's (t)分布都有类似的单峰、中心对称的特征，它们的期望值和方差有十分明确的意义。更重要的，有如下不同分布的中心极限定理：设 $X_1, X_2, \cdots, X_n$ 是一列独立随机变量，它们的概率密度分别为 $f_{X_k}(x)$，并有 $E(X_k)=\mu_k$，$D(X_k)=\sigma_k^2$，$(k=1,2,\cdots)$，令：

$$B_n^2=\sum_{i=1}^{n}\sigma_k^2 \qquad Y_n=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n} \tag{1}$$

若对任意正数 τ，有 $\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\tau B_n}\left(x-\mu_k\right)^2 f_{X_k}\left(x\right)dx=0$，对任意 x，随机变量 Y_n 的分布函数 $F_n\left(x\right)$，满足 $\lim_{n\to\infty}F_n\left(x\right)=\lim_{n\to\infty}P\left\{\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}\le x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$ 。

该定理说明：如果所研究的随机变量是由大量独立且均匀的随机变量相加而成，那么它的分布将近似于正态分布。来自物理系统的实测随机数据用 Gauss 分布作为理论假设应该是最符合逻辑的选择了。在 Gauss 分布假设下面，随机过程的统计分析要简单很多，均值和方差的含义十分明了，不需要引入高阶统计量，均值和方差已经完整地描述了 Gauss 随机过程的统计信息。

对于一维的平稳 Gauss 随机过程，功率谱刻画了它的最基本特征。对于高维的平稳 Gauss 随机过程，方向谱刻画了它的最基本特征。功率谱（方向谱）的获取可以通过非参数化的 Fourier 变换类方法；更精确地是通过对信号建模并估计模型中的参数来精确地估计功率谱（方向谱）。在平稳 Gauss 随机过程假设下，一维情形下的谐波基函数叠加与高维情形下的平面波基函数叠加建立信号模型，并在 Bayes 或最大似然估计框架下得到最佳的谱估计结果是现代谱估计的典型做法。当然，还是在平稳 Gauss 随机过程假设下，对自相关矩阵进行特征分解（分解成信号子空间和噪音子空间），是一类基于数据驱动的谱估计方法。推广到高维情形，核心工作是构建 Hankel 矩阵，并对 Hankel 矩阵进行 SVD 分解，此时奇异值谱就是功率谱，奇异向量就是平面波基函数。这是所谓奇异值谱分析的典型做法。可以看出，对于平稳 Gauss 随机过程，基于 Fourier 变换的谱分析方法和基于自相关（或 Hankel）矩阵分解的谱分析方法是等价的。值得注意的是，一旦平稳 Gauss 随机过程的假设不成立，二者的等价性就失去了。

对于平稳 Gauss 随机过程的统计分析，另一个重要的方面应该是（定量地）测量信号间的差异性。互协方差函数和互功率谱密度从时间（或空间）域及频率（或波数）域考察两个信号间的差异性（反之，也是相似性）。归一化的互协方差函数被定义成相关系数；归一化的互功率谱密度被定义成相干函数。从相关系数的定义式可以清楚地看出，它是在减掉各自的背景（期望）成分后，定量地感知两个信号的同相成分乘积和异相成分乘积的差，完全没有异相成分，相似系数等于 1；完全没有同相成分，相似系数等于 0。相干函数测量两个信号的相干性，若是相干信号，相干函数等于 1；完全的非相干信号，相干函数等于 0。事实上，这两个函数只有对相干信号的考察，才有明确的意义。否则，所谓的相似系数和相干函数并不能真正揭示出两信号的差异性（或相似性）。

信号间差异性（或相似性）的感知与度量是信号分析中的一个极为重要的事情。但仅仅利用二阶统计量是无法做得很彻底的！二阶统计量无法感知两信号间的（初）相位差异，这也是为什么一定要引入相干信号的原因，相干信号是相互拷贝的信号，没有（初）相位差异性。

事实上，还有一个问题是如何度量信号间的差异性。笔者认为：相关系数和相干函数都不是很好的信号差异性测量方法，好的差异性测量方法应该度量的是信号特征之间的差异，而不是两个信号不同时刻的取值间的差异。

必须提及的是，平稳 Gauss 随机过程都是线性系统的输出结果。系统的线性性与随机过程的 Gauss 性是紧密联系的。至少在数学意义上，它们必须统一，否则二者就是相互矛盾的。线性系统下的 Gauss 随机过程情形下的各种信号分析问题涉及到的解法基本上可以认为就是（加权）最小二乘解法。这也说明，我们对线性系统下平稳 Gauss 随机过程的分析从数学上看意义得到了彻底的解决。

我们知道，针对非平稳信号的分析，首先假设信号的变化是缓平稳的或局部平稳的，然后把上述平稳 Gauss 随机过程的假设应用在局部平稳信号上，把相应的处理方法稍作修改应用在局部平稳信号的分析中。譬如 Kalman 滤波方法，自适应滤波方法等。更进一步地，发展了各种时频分析的方法，譬如短时（或加窗）Fourier 方法、小波变换方法等等。

信号分析的真正困难应该是随机过程非 Gauss 分布和系统的非线性性引起的。凸优化问题的引入解决 $L_p,1\le p<2$ 范数下的信号分析问题对于广义 Gauss 分布下随机信号的分析有了一定的适应性。通过 Monta Carlo 类反演方法也能解决部分非线性信号分析问题。但是，非高斯、非线性、非平稳情形下的信号分析问

题的真正解决还差距甚大。

3 基于 Hilbert 空间中函数逼近的信号分析

勘探地震中的地震波场，总可以表达成信号、相干噪音和随机噪音三部分叠加形成的。可以简记为 $D=S+N_c+N_r$ 。其中，D 为观测数据，S 为地震信号，N_c 为相干噪音，N_r 为随机噪音。噪音压制的思想框架可以认为是将含噪信号视为 Hilbert 空间（平方可积函数空间）中的一个平方可积函数，选取合适的基函数进行信号的稀疏的、近似的表达（信号预测的正问题）；然后在一定的估计理论基础上，基于上述信号预测正问题，求解最佳的信号预测器，从而实现信号的最佳估计。在这样的函数逼近思想下，地震数据的去噪音、插值、压缩和特征提取等是统一的。

信号是可预测的，相干噪音也是可预测的。此处指的是它们都不是随机信号，可进行确定性地预测。而随机噪音只能从概率或统计的角度进行预测，就是用某种联合概率密度函数来预测，或用联合概率密度函数计算出的某种统计量来预测。

对确定性的信号进行预测本质上就是寻找一个函数来逼近确定性的信号。函数逼近问题是在 Hilbert 空间中寻找一组基函数形成一个子空间，对一个函数的逼近就是在 L2 范数意义（也可在其它范数意义）下寻找逼近误差最小的该组基函数的线性组合来达到对函数的最佳逼近的目的。

对相干噪音的预测也用上述函数逼近的思想与方法。

一般地，对信号的预测和对相干噪音的预测是在两个不同基函数组构成的两个不同的子空间中进行的。这样的概念被认为是在基函数字典中实现对一个函数不同特征的逼近。实质上，高维空间中包含的确定性信号和相干噪音就可以认为是一个函数具有两种不同的特征。在基函数字典中对函数进行最佳逼近是函数逼近理论的当前进展。在更高维的空间（变换域）中进行函数的更为稀疏的特征表达是另一个层次的概念，高维小波变换就是在更高维的、也更为冗余的变换空间中实现的对函数的表达。很显然，高维小波变换与字典基概念的结合可以在更为高维的空间中实现对函数的更好的稀疏特征表达。但是，一味地在更高维的空间中对函数进行更为冗余的也更为稀疏的特征表达并不一定是最佳的选择。笔者认为，之所以需要如此做，还是因为我们不知道信号（或函数）中包含什么特征，所以才用所谓的高维小波变换与字典基结合的方法试图能获取信号的特征表达，实质上这样获得的稀疏特征表达不一定真的抓住了信号的本质特征。真正好的方式是用已知信号的真实特征（基函数）去匹配已知信号（实质上就是做内积），就可以提取出具有物理意义的信号的特征。笔者相信这种方法是非常有优势的，问题是需要信号分析人员对要分析的信号有足够深入的认识，能猜想到信号的重要特征是什么，用什么样的基函数可以表达它们。

预测误差（即随机噪音）的概率分布决定了选择什么样的范数来定义代价函数。一般地，选择 Gauss 分布。这与用基函数的线性组合进行函数逼近是一致的。线性 Gauss 的假设又一次施加在了基于 Hilbert 空间中函数逼近的信号分析思想与方法上面。当然，这样的假设同样存在问题，前一节已经阐述过其中的原因。在非 Gauss 概率分布假设和非线性信号预测的基础上进行更为深入细致的信号分析是必然的要求。

4 现代图像分析基本思想与方法

图像分析的思想与方法在地震波反演成像中的地位越来越重要。地震数据分析主要用信号分析的思想与方法，属于数据域中的处理；地震波成像结果的分析主要用图像处理的方法，属于图像域中的处理。地震波成像中的图像处理主要用于反演参数的正则化过程中，更明确地讲，就是提取图像中的结构信息用于约束模型参数的几何形态。当然，针对成像结果的图像处理还有很多其它的用处。

用成像结果表示的图像与灰度值表示的图像还是有明显的差异，因为成像结果基本上是零均值的震荡波形表示的图像。因此，一般的图像处理方法直接用在成像结果表示的图像上并不能产生很好的分析效果。

图像至少是二维的，图像是（正的）灰度值表示的。图像的概念模型为：有规律分布的、较强的灰度值表示的图像的结构与纹理飘在满足某种概率分布特征的随机噪音中。纹理可以认为是有规律的细致的结构。很显然，这样的概念模型可以转化为数学模型：$g(\boldsymbol{x})=u(\boldsymbol{x})+\eta(\boldsymbol{x})$ ，其中 $g(\boldsymbol{x})$ 代表实际获取的图像；$u(\boldsymbol{x})$ 代表无噪音干扰的图像；$\eta(\boldsymbol{x})$ 代表符合满足概率分布的噪音。

图像分析或处理分三个基本的层次：低层次；中等层次和高级层次。低层次的图像处理可以说是对图像进行预处理，包括灰度变换；去噪音；图像恢复；图像增强等。中等层次的图像处理包括图像分割、图像基本模式的识别。高级层次的图像处理主要是复杂模式的识别和图像认知等。图像去噪和图像恢复这些基本处理与信号处理方法是一致的。可以认为图像分析与处理是信号分析与处理的分支。但是，图像处理在中高级层次就显示出了它与信号处理的差异，尽管二者还是有非常紧密的联系。譬如信号的特征表达和图像的特征表达主要还是用 Hilbert 空间中的函数逼近的思想。

图像处理首先是对像素进行变换：$f(\boldsymbol{x})=T(g(\boldsymbol{x}))$ ，其中 $T(\cdot)$ 代表变换函数。基于图像灰度的直方图（某种概率分布）设计变换函数 $T(\cdot)$ 对图像 $g(\boldsymbol{x})$ 进行变换是常用的方法，主要目的是调节图像的对比度。

图像预处理中的去噪、图像反褶积（图像恢复）、随机噪音压制在信号分析中有全面的表述。理想滤波器（带通滤波器）、Butterworth 滤波器、Gauss 滤波器、Wiener 最佳滤波器、Wiener 反褶积；包括针对随机噪音的 Gauss 加权滤波器、各种中值类滤波器等，都是信号分析中的基本内容。频率域和时间域滤波的思想在信号分析中都有深入的描述。

从各向同性的 Gauss 滤波发展到各向异性的 Gauss 滤波；把各向异性扩散方程用于图像去噪；求解 ROF 模型（TV 正则化下的图像去噪模型），这三者方法是图像处理中比较新的去噪思想与方法，三者本质上存在极为密切的联系。目的都是在保持图像结构的前提下压制图像中的随机噪音。

基于函数微分思想的图像增强算子的构造与 Laplace 算子的有限差分格式的构造是类似的。目的是通过一阶或二阶差分运算提出图像中背景部分，锐化图像中的跃变部分，为进一步的图像分割和识别提高更好的基础数据。

进入图像分割才算进入了图像处理的核心阶段。这是图像处理特有的内容。可以认为，到目前为止并没有形成图像分割的统一的理论与方法，不存在一个适用于任何图像的图像分割理论框架。图像跃变点、分界线和分割面的检测应该是图像分割的基础。即便这个基础工作也没有形成统一的理论框架。在图像分割的基础上，如何进行图像（点、线、面、体）的定量描述也没有统一的理论。唯一比较成体系的图像表达方法应该是高维小波变换，它可以在变换域中表达图像的结构信息。利用 Hilbert 空间中对函数 $g(\boldsymbol{x})$ 的高维小波变换表达可以认为在一定程度上抓住了图像 $g(\boldsymbol{x})$ 的特征。当前，机器学习方法用于图像处理中的主要进步就是在 $L^p\left(1\le p<2\right)$ 范数下基于图像函数 $g(\boldsymbol{x})$ 的高维小波变换进行图像的特征表达。

图像的机器认知（或计算机认知）是图像分析与处理的终极目标。但是，要逼近人脑对图像的认知水平还有巨大的差异，本质上图像的认知已经超出了图像分析与处理的范围，进入了人工智能的范畴。

很明显，勘探地震中的各个环节都对图像处理提出了越来越高的要求。图像去噪音、图像特征表达当前还是主要处在用二阶统计量的阶段。非 Gauss 与非线性假设下的图像处理和分析方法的发展是无法回避的。

5 利用高阶累积量的信号分析

任何信号都是系统的响应。任何实测信号都包含随机噪音，都可以被认为是随机过程。实测信号包含确定性信号成分和加性随机噪音。系统建立起了输入和输出的关系，用信号预测的正问题进行表达。信号预测正问题一般是线性的。在 Bayes 反演框架下，实测的物理系统的响应中包含的信号也是随机的。至少概念上是这么认为的，利用实测的物理系统的响应（随机过程）在 Bayes 框架下估计的是其中所包含信号的期望值。这是现代信号分析最基本的出发点。

当前我们所做的信号分析工作绝大部分都是基于二阶统计量展开的。这些二阶统计量包括：即相关函数、协方差和功率谱。蕴含的基本假设是系统是可线性表达的。观测数据满足高斯分布（或噪音满足高斯白噪的）。去噪、插值等方面的信号分析工作都在此框架下进行。但实际数据情况并非如此。噪音不满足高斯分布是第一个问题，这也可能是由于信号模型不合适导致对实际数据中蕴含的信号预测不准引起的，不仅仅是实际物理系统的响应本质上会包含随机噪音。系统非线性是第二个问题。系统的线性表达也是人为施加给实际物理系统的，测试信号所对应的物理系统并非能用线性关系进行很好的逼近。注意此时物理系统主要是指对实测数据中所含信号的预测关系。信号非平稳是第三个问题，缓平稳情况下加窗数据分析效果一般情况下还是可以的。还有系统是否因果的及最小相位的也是当前实际问题中常遇到的，当然物理系统

本身一定是因果的，最小相位性假设是人为施加给物理的很难满足。这些实际问题是信号/图像处理中常见的，也是信号分析由二阶统计量转到高阶统计量的主要动力。

二阶统计量信号分析的核心工作是功率谱估计，它仅仅反映了信号的按频率展布的能量关系。由自相关函数（矩阵）的 SVD 分解可以估算功率谱，在 Gauss 线性假设下，与 Fourier 功率谱估计结果是一致的。但是相同功率谱的信号可以是非常不同的，这是由信号的相位特征决定的。功率谱完全不会反映信号的相位信息。功率谱估计的确是信号分析中的最核心问题，但绝不是唯一的核心问题。

对于非高斯信号的模型，如仅仅考虑利用自相关函数，就不可能充分获取隐含在数据中的信息。若信号不仅是非高斯的，而且是非最小相位的，采用基于自相关函数的估计方法所得到的模型参数，就不能反映原信号的非最小相位特点。当测量噪声较大，尤其当测量噪声有色（非白）时，基于自相关函数的估计方法所得到的模型参数有较大的估计误差。基于功率谱的理论把随机序列视为一系列统计不相关的谐波分量的叠加。把高维随机信号视为一系列统计不相关的平面波分量的叠加。实质上，认为系统是线性的。这样做忽略了各谐波之间的相位关系，实际上各谐波之间可能存在非线性的关系。存在非线性相位关系的信号的实例：$S_1(t)=\left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\frac{\alpha t^2}{2}} e^{j\left(\frac{\beta t^2}{2}+\omega_0 t\right)}$ ； $S_2(t)=\left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\frac{\alpha t^2}{2}} e^{j(m\sin\omega_m t+\omega_0 t)}$ 。

高阶累积量/高阶谱的确可以分析出观测信号是否满足高斯分布，某种程度上也可以定性地判断系统的非线性性。这是现代信号分析关注高阶累积量/高阶谱的基本原因。譬如，①基于累积量描述了信号与 Gauss 分布偏离的程度，从非 Gauss 信号中提取信息。②根据 Gauss 信号通过线性系统后仍为 Gauss 信号的规律，通过使用高阶谱检测未知系统在 Gauss 信号输入下，其输出信号相对于 Gauss 信号的偏离程度来分析系统的非线性特征。③高阶谱对 Gauss 噪声是零响应，或称为是盲 Gauss 的。可以较好地从 Gauss 噪声中分离出有色信号。④当谐波过程的各分量相互独立时，则此谐波信号的三阶累量恒为零。但当分量之间存在二次相位耦合关系时，则它的三阶累量不为零。因此，可以通过估计谐波信号的三阶累量及对应的双谱来检测二次相位耦合现象的存在。

但如何在非线性及非高斯情形下进行信号的分析，进行更合理的信号分析，高阶累积量/高阶谱分析并没有给出具体的做法。做出双谱、三谱，如何用？如何解决非高斯/非线性问题，这些都没有答案。利用高阶统计量如何开展更深入的信号与图像分析这样的问题没有回答。这都是高阶统计量信号分析理论几十年来（50-60 年代，高阶统计量分析理论已经建立了！）得不到长足发展的根本原因。

事实上，现代信号分析，80 年代开始的小波变换（更早出现的时频分析）对付信号的非平稳性；70 年代初开始的 L_1 范数下的反演对付非高斯情形，当然也进行稀疏反演及信号/图像的特征表达。DL（=Deep Learning）似乎在解决非线性信号预测问题（非线性反问题也与此有关！）。这是当前信号分析发展的基本路径。可见我们并未走利用高阶统计量进行信号分析的路线！

表面上的原因是高阶谱没有为非 Gauss、非线性和非平稳信号分析提供真正有效的分析工具！到目前为止，如何利用高价谱进行更深入的信号分析是不明朗的！

6 地震数据分析中的振幅与相位问题

地震数据被认为是带限的反射子波飘在满足各种概率分布假设的随机噪音中。带限的反射子波是地震频带内不同幅值的谐波叠加形成的，即 $u(t)=\sum_{j=\omega_S}^{\omega_E} a(\omega)e^{-i(\omega t+\varphi_0)}$ 。推广到高维空间中，地震波 $u(\boldsymbol{x},t)$ 被认为是平面谐波的叠加。选择这样的信号表达形式进行后续的各种各样的信号分析，实质上已经蕴含了对地震信号的假设。首先，振幅是不可以跳变（剧烈变化）的，否则就不是有限频带的信号，更不可能是（5-80Hz）的窄带信号。初相位 $\varphi_0(\omega)$ 是未知的，因此也不知道 $u(t)$ 是否是最小相位系统。相位 $\varphi(\omega)=\omega t+\varphi_0(\omega)$ 是一个线性相位关系，各谐波之间相位是没有关联的。典型的 AR 模型预测器更是蕴含了振幅线性变化、相位线性变化的假设。

上述线性化的信号模型当然与地震信号满足 Gauss 假设是一致的。地震信号分析基本上就做了 Gauss 线性假设下的处理工作。在 $L^p(1\le p<2)$ 范数下进行信号和图像分析工作在地震信号分析中并不广泛应用。

客观地讲，地震信号随频率变化的相位特征变化是缓慢的，否则整个地震数据分析（包括信号处理、反演成像和地震解释）将变得异常困难。地震波振幅的空间时间变化也是不很剧烈的。但是，尽管如此，目前的线性预测器还是不能很好地适应地震信号的实际情况。这应该是地震信号分析不够好的核心因素。但是，如何修改才能提出更好的、计算方便的预测器？改成弱非线性的预测器吗？在 $L^p\left(1 \le p < 2\right)$ 范数下进行信号和图像分析是否就比较彻底地解决了非 Gauss 随机噪音的问题？

6 结论与讨论

地震信号分析，实质上所有的信号分析，都是基于这样的概念信号模型：可以通过建模进行预测的信号飘在满足各种概率分布的随机噪音中构成了对实测信号（或数据）的概念性描述。与此概念信号模型对应的是其数学表达形式。但是，这是人为提炼出的模型，并不代表实测数据（任何实测数据都是物理系统的响应，都是随机的!）可以用上述模型很好地描述，基于这样的模型可以对信号进行完全深入的分析。

Gauss 线性假设是人为地施加在信号模型上的，并据此发展出了一系列到目前为止常用的信号分析方法。从统计意义上看，这些信号分析方法都是基于随机信号的二阶统计量而提出来的。

但是，线性预测器和预测误差 Gauss 分布的假设以及在此基础上发展的一整套完整的信号分析方法尽管还算有效，但已经过于成熟而缺乏发展的空间。实际信号的复杂多变性（指的是它是非平稳、非高斯的随机过程，对它的预测需要用非线性的预测器）以及对信号特征更准确描述的需求使得必须突破 Gauss 线性假设，提出更新更有效的信号分析方法。在 Hilbert 空间中进行信号的多分辨率分析（即小波变换和字典基下的变换）与在 $L^p\left(1 \le p < 2\right)$ 范数下进行信号和图像分析代表了信号分析的最新进展，试图突破 Gauss、平稳和线性假设。但是，即便有这样的信号（图像）分析的进展，也不代表信号分析方法与技术出现了革命性的变革，它们仅仅是对 Gauss 线性假设下基于二阶统计量的信号分析方法的改良。高阶统计量分析自上世纪 60 年代提出以来到现在为止基本没有实质性的进步，没有发展出有效的信号分析方法。目前我们对信号相位变化的估计能力基本是完全缺失的，很难捕捉住信号的细节变化特征。

笔者认为：对信号振幅的非线性变化，尤其是对其相位非线性变化，的预测能力的缺失是当前信号分析方法与技术停滞不前的主要原因。对信号的非线性建模并基于所建模型发展出有效的信号分析方法是真正的问题所在。Deep-Learning 是正确的方法技术方向吗？

参考文献

[1] Rafael C. Gonzalez, and Richard E. Woods, 2010, Digital Image Processing, Third Edition, Pearson Education.
[2] Otmar Scherzer, 2011, Handbook of Mathematical Methods in Imaging, Springer.
[3] Sergios Theodoridis and Konstantinos Koutroumbas, 2009, Pattern Recognition, Fourth Edition, Elsevier
[4] Boualem Boashash(Editor), 2016, Time-Frequency Signal Analysis and Processing – A Comprehensive Reference, Second Edition, Elsevier
[5] Leon Cohen, 1995, Time-Frequency Analysis, Prentice-Hall PTR
[6] Stéphane Mallat, 2009, A Wavelet Tour of Signal Processing, Third Edition, Elsevier
[7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, 2016, Deep Learning, Http://www.deeplearningbook.org/

PCA/RPCA/CCA/ICA 与地震信号建模

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】现代信号分析是地震数据分析的基础，地震数据分析的几乎每一个环节中都会用到现代信号分析的知识，譬如高维数据规则化、高维去噪音、包括 FWI 和 LS-RTM 都是建立在现代信号分析中信号预测和估计理论基础上的。实际观测的地震数据被视为确定性信号叠加上满足一定概率分布的随机噪音形成所谓的随机信号，这是地震信号分析的概念模型。定量地进行地震信号处理（其实也包括任何领域的信号处理）核心要务是对实际观测数据中包含的信号进行建模。最根本的模型是描述随机信号（随机过程）的联合概率密度函数。但是针对高维数据，用联合概率密度函数描述信号无论是理论上和实用上都是不可能或不实际的。信号建模的本质是用信号中包含的特征量（潜变量）的线性组合逼近信号。对地震信号而言，一维信号被认为是不同频率的谐波叠加而成的；高维信号被认为是不同方向的平面波叠加而成的，至少在局部窗内这样的观点是非常符合逻辑的。到目前为止，对信号进行建模的核心问题就是找到一组合适的特征量（潜变量）。一种方法是在 Hilbert 空间中根据对信号的认识预先设定一组（或多组）特征量；另一种方法是由数据驱动在一定的原则下估计一组（或多组）特征量。主分量分析（PCA）方法是最经典的数据驱动估计线性特征量的方法。它与 Gauss 假设+Fourier 基函数逼近相结合的 Fourier 信号分析理论一起构成了信号分析的基础。在信号分析理论中，前者被称为 SSA（Singular Spectrum Analysis）方法；后者被称为 FSA（Fourier Spectrum Analysis）方法。推广到高维信号处理中，对应地，有低秩矩阵分解方法和 Beamforming 方法。Robust PCA 针对的是线性信号上叠加了非高斯噪音，尤其是脉冲噪音（譬如椒盐噪音的情形），求解信号自相关矩阵核范数最小的约束凸优化问题，实现对线性信号的最佳建模（预测）。CCA（Canonical Correlation Analysis）方法的目的是用两个数据间包含的线性相关分量建立信号预测模型。PCA、RPCA 和 CCA 都仅仅利用了数据的二阶统计量，只能提取数据中的线性特征量，并用这些线性特征量的组合实现对信号的建模与逼近。二阶统计量仅反映数据中蕴含的信号的相关性，基于二阶统计量得到的特征量在表达信号时可能存在冗余，更进一步地是用独立特征量对信号进行建模表达，这是独立分解分析（ICA=Independent Component Analysis）要解决的问题。当然，最好是用导致信号变化的因变量来表达信号。但是由数据驱动的方法找到影响信号变化的原因变量是不现实的，这是人类用归纳综合的思想要解决的问题。不过，独立分量应该更接近于原因变量。本文重点阐述地震信号分析的核心理念，就是对信号进行建模。讨论 PCA、RPCA、CCA 和 ICA 这四种由数据驱动进行信号建模的思想与方法，简述 Hilbert 空间中用函数逼近的思想对信号建模的方法，最后对地震信号建模的思想与方法进行了综合评价。

【关键词】信号分析，随机信号，信号建模，PCA，RPCA，CCA，ICA，地震信号分析

1 引言

信号（也包括图像）分析的目的是获取高维随机信号中包含的各个层次和各个方面的特征，利用这些特征的（线性）组合对信号进行建模，用所建立的信号模型对信号进行最佳的预测，并基于此开展信号处理的各方面的应用，譬如插值、规则化、去噪等等。

任何信号分析都是基于一定的模型的。首先要有一个概念模型。实际观测的数据被视为确定性信号叠加上满足一定概率分布的随机噪音形成所谓的随机信号（随机过程），这是信号分析的最基本的概念模型。更具体地，实际测量的地震数据可以定量地由 $\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}(x,y,t)$ 表示。其中，(x,y,t) 表示地震数据的空间时间坐标；$\boldsymbol{u}(x,y,t)$ 表示实际测量的地震数据（也称地表观测到的地震波场）；$\boldsymbol{S}(x,y,t)$ 表示对应的确定性信号；$\boldsymbol{\eta}(x,y,t)$ 是满足某种概率分布的加性随机噪音，并假定各坐标点处的噪音分量为独立同分布的随机变量且与该点处的信号 $\boldsymbol{S}(x,y,t)$ 统计特征无关。应该注意的是：在实际地震信号的统计处理中，上述信号模型还可以表达在常见的五维空间（譬如 $\boldsymbol{u}(m_x,m_y,h_x,h_y,t)$，其中 (m_x,m_y) 为中心点坐标，

$\left(h_x,h_y\right)$为偏移距坐标）中，在更高维的空间中进行统计信号分析，统计特征会更符合既定的假设，去噪效果会更好。上述方式定义的信号模型的应用是非常普遍的，当前绝大部分的地震信号处理方法的构建都是从此模型出发的。地震信号分析的核心工作就是实现对信号$\boldsymbol{S}\left(x,y,t\right)$的最佳预测或最佳（特征）表达，从而达到最佳去噪、插值、压缩等目的。

信号分析更为本质的思想是假设实测离散数据是随机的、符合一定的概率分布函数的、可以用统计的方法来描述的。对随机信号（随机过程）的完整的数学表达当然是已知其联合概率密度函数。针对高维的离散数据，用联合概率密度函数进行建模与预测显然是做不到的。具体的信号分析（处理）主要包括两方面的核心内容：信号的建模和 Bayes 估计理论下对信号模型中所含参数的最佳估计。也可以认为就是信号的最佳建模，以及该最佳模型的具体应用（譬如去噪、插值、信号恢复等等）。

因此信号的建模是信号分析的核心。在 Hilbert 空间中构建一组（或多组）基函数形成子空间，在子空间中对信号进行基函数线性组合的表达是信号建模的最基本方法或最常用方法。信号建模是建立在对实测信号基本抽象和对信号深入认识的基础上的。对地震信号而言，我们认为实测地震数据是一个典型的五维数据体，可以抽象地认为该地震数据是线性和/或非线性（相位变化）的连续同相轴飘在满足不同的统计特征的随机噪声中，形成所谓的高维随机信号。地震信号分析的对象就是这个高维随机信号，更准确地讲是实测的、高维的、离散的、随机的数据体。这是笔者对实测地震数据建立的概念模型。从统计意义下建立一个高维随机信号的模型就是确定随机信号的联合概率密度函数，其中的二阶统计量（自相关函数、互相关函数）蕴含了高维信号中包含的线性特征。从这个意义上看，Hilbert 空间中选定基函数的信号建模和利用数据中的自相关函数分解进行信号建模二者在理论上是等价的，都是在一定的统计意义下建立信号的线性逼近模型。后者是 PCA 方法的理论基础。

PCA 方法、RPCA 方法、CCA 方法和 ICA 方法是当今数据驱动类信号建模的典型方法。主分量分析（PCA）方法是最经典的数据驱动估计线性特征量的方法。Robust PCA 是 PCA 方法基础上的改进，考虑的是线性信号上叠加了非高斯噪音，通过求解信号自相关矩阵核范数最小的约束凸优化问题，实现对线性信号的最佳建模（预测）。CCA（Canonical Correlation Analysis）方法优势通过用两个数据间包含的线性相关分量建立信号预测模型。PCA、RPCA 和 CCA 都仅仅利用了数据的二阶统计量。基于二阶统计量得到的特征量在表达信号时可能存在冗余。对 PCA 方法最进一步修改，就发展出来独立分量分析（ICA=Independent Component Analysis）方法。

本文重点阐述地震信号分析的核心理念，就是对信号进行建模。讨论 PCA、RPCA、CCA 和 ICA 这四种由数据驱动进行信号建模的思想与方法，简述 Hilbert 空间中用函数逼近的思想对信号建模的方法，最后对地震信号建模的思想与方法进行了综合评价。

2 基于 Hilbert 空间中函数逼近方法的信号建模

勘探地震中的地震波场，总可以表达成信号、相干噪音和随机噪音三部分叠加形成的。可以简记为$\boldsymbol{u}\left(x,y,t\right)=\boldsymbol{S}\left(x,y,t\right)+\boldsymbol{\eta}_C\left(x,y,t\right)+\boldsymbol{\eta}_R\left(x,y,t\right)$ 。其中，$\boldsymbol{u}\left(x,y,t\right)$为观测数据，$\boldsymbol{S}\left(x,y,t\right)$为后续要处理的地震信号，$\boldsymbol{\eta}_C\left(x,y,t\right)$为相干噪音，$\boldsymbol{\eta}_R\left(x,y,t\right)$为满足一定概率分布的随机噪音。（地震）信号分析的基本思想框架可以认为是将含噪信号视为 Hilbert 空间（平方可积函数空间）中的一个平方可积函数，选取合适的基函数进行信号的稀疏的、近似的表达（信号预测的正问题）；然后在一定的估计理论基础上，基于上述信号预测正问题，求解最佳的信号预测器，从而实现信号的最佳估计。很显然，这是一个典型的 Hilbert 空间中的函数逼近问题。在此思想框架下，地震数据的去噪、插值、压缩和特征提取等等处理过程可以统一成一件事。

信号是可预测的，相干噪音也是可预测的。此处指的是它们都不是随机信号，可进行确定性地预测。而随机噪音只能从概率或统计的角度进行预测，就是用某种联合概率密度函数来预测，或用联合概率密度函数计算出的某种统计量来预测。

对确定性的信号进行预测，本质上就是寻找一个函数来逼近确定性的信号（它就是一个连续函数）。函数逼近问题是在 Hilbert 空间中寻找一组基函数形成一个子空间，对一个函数的逼近就是在 L2 范数意义（也

可在其它范数意义）下寻找逼近误差最小的该组基函数的线性组合来达到对函数的最佳逼近的目的。很显然，对相干噪音的预测也用上述函数逼近的思想与方法。

一般地，对信号的预测和对相干噪音的预测是在两个不同基函数组构成的两个不同的子空间中进行的。这样的概念被认为是在基函数字典中实现对一个函数不同特征的逼近。实质上，高维空间中包含的确定性信号和相干噪音就可以认为是一个函数具有两种不同的特征。在基函数字典中对函数进行最佳逼近是函数逼近理论的当前进展。在更高维的空间（变换域）中进行函数的更为稀疏的特征表达是另一个层次的概念，高维小波变换就是在更高维的、也更为冗余的变换空间中实现的对函数的表达。很显然，高维小波变换与字典基概念的结合可以在更为高维的空间中实现对函数的更好的稀疏特征表达。但是，一味地在更高维的空间中对函数进行更为冗余的也更为稀疏的特征表达并不一定是最佳的选择。笔者认为，之所以需要如此做，还是因为我们不知道信号（或函数）中包含什么特征，所以才用所谓的高维小波变换与字典基结合的方法试图能获取信号的特征表达，实质上这样获得的稀疏特征表达不一定真的抓住了信号的本质特征。真正好的方式是用已知信号的真实特征（基函数）去匹配已知信号（实质上就是做内积），就可以提取出具有物理意义的信号的特征。笔者相信这种方法是非常有优势的，问题是需要信号分析人员对要分析的信号有足够深入的认识，能猜想到信号的重要特征是什么，用什么样的基函数可以表达它们。

预测误差（它是满足一定概率分布的随机噪音）的概率分布决定了选择什么样的范数来定义代价函数。一般地，选择 Gauss 分布。这与用基函数的线性组合进行函数逼近是一致的。线性 Gauss 的假设又一次施加在了基于 Hilbert 空间中函数逼近的信号分析思想与方法上面。线性 Gauss 假设下 Hilbert 空间中基函数（字典）线性组合的函数逼近建模与利用二阶统计量的 PCA/CCA 建模本质上是等价的。PCA/CCA 建模仅仅基于数据的一阶和二阶统计量，对实测数据的适应性更好一些，换句话讲，PCA/CCA 方法建立的模型可能更优。

3 主分量分析（PCA）方法

基于随机信号二阶统计量（自相关函数和互相关函数）的信号建模方法对实测数据的适应性更好，更能达到最佳逼近是 PCA 类方法得到广泛应用的理论基础，尤其是大数据分析和机器学习（ML）方法得到更广泛应用的时代，数据驱动的信号分析方法得到更充分的重视。

PCA 的出发点是一个 n 维随机向量 $\boldsymbol{x}$。注意，此处 n 维随机向量 $\boldsymbol{x}$ 是行向量。实际应用中，是该 n 维随机向量的一组离散采样值 $\{x_1, x_2, \cdots, x_n\}$。PCA 不对该随机向量的概率密度函数做明确的假设，只假设它的一阶和二阶统计量已知，实际上，就是假设可以由上述离散采样值统计得到它的一阶和二阶统计量。

本质上，n 维随机向量 $\boldsymbol{x}$ 的离散采样值是存在相互关联的，就是说用这些离散元素表达 $\boldsymbol{x}$ 是存在冗余的。这是 PCA 能广泛应用于信号处理和图像处理根本基础。若这些离散元素之间相互独立（它们之间的联合概率密度函数是 Gauss 的），PCA 什么也得不到。高维信号/图像处理中存在结构和纹理，就说明表达它们的离散元素之间存在相关性，PCA 方法可以提取出这些相关信息（即结构和纹理信息）。

在 PCA 中，首先要对 n 维随机向量 $\boldsymbol{x}$ 进行去中心化，即随机向量 $\boldsymbol{x}$ 减去其均值：

$$\overline{\boldsymbol{x}} \leftarrow \boldsymbol{x} - E\{\boldsymbol{x}\} \tag{1}$$

PCA 就是把 n 随机向量 $\overline{\boldsymbol{x}}$ 变化成另一个 m（$m<n$）维的随机向量 $\boldsymbol{y}$。实质上，就是引入一个线性变换 A，使得：

$$\boldsymbol{y}=A\overline{\boldsymbol{x}} \tag{2}$$

目的是通过线性变换 A 使得到的 m（$m<n$）维随机向量 $\boldsymbol{y}$ 的元素之间互不相关。很显然，K-L 变换是 PCA 的具体实现方法之一，而且是最广泛应用的方法。但是，方差极大化的 PCA 方法（Hotelling，1933）是与 PCA 更为名实相符的方法。下面详细描述方差极大化的 PCA 方法。

定义 y_1 是 n 维随机向量 $\overline{\boldsymbol{x}}$ 向向量 $\boldsymbol{w}_1$ 的投影，记为

$$y_1 = \sum_{k=1}^{n} w_{k1} x_k = \boldsymbol{w}_1^T \boldsymbol{x} \tag{3}$$

向量 $\boldsymbol{w}_1$ 就是 n 维随机向量 $\overline{\boldsymbol{x}}$ 中蕴含的特征向量，也可以称为特征成分或主成分。可以想见，n 维随机向量 $\overline{\boldsymbol{x}}$ 应该包含有若干个主成分，譬如有 m （$m < n$）个主成分，这些主成分的线性组合可以很好地表达 n 维随机向量 $\overline{\boldsymbol{x}}$ 。据此可以看出，PCA 分析与 Hilbert 空间中的信号（函数）逼近有本质上的一致性，可以说它们都是信号建模的最基本方法。

方差极大化 PCA 方法的数学提法：

$$\begin{aligned} \mathcal{J}_1^{\mathrm{PCA}}\left(\boldsymbol{w}_1\right) &= \arg\max E\left\{y_1^2\right\} = \arg\max E\left\{\left(\boldsymbol{w}_1^T \boldsymbol{x}\right)^2\right\} \\ &= \arg\max \boldsymbol{w}_1^T E\left\{\boldsymbol{x}\boldsymbol{x}^T\right\} \boldsymbol{w} = \arg\max \boldsymbol{w}_1^T C_{\boldsymbol{xx}} \boldsymbol{w} \end{aligned} \tag{4}$$

并满足约束条件：

$$\left\|\boldsymbol{w}_1\right\| = 1 \tag{5}$$

求解该优化问题，就可以实现对 n 维随机向量 $\overline{\boldsymbol{x}}$ 的 PCA 分析。

还存在最小均方误差 PCA 方法（Pearson，1901）。该方法的数学提法为：

$$\mathcal{J}_{\mathrm{MSE}}^{\mathrm{PCA}}\left(\boldsymbol{W}\right) = \arg\min E\left\{\left\|\boldsymbol{x} - \sum_{i=1}^{m}\left(\boldsymbol{w}_i^T \boldsymbol{x}\right)\boldsymbol{w}_i\right\|^2\right\} \tag{6}$$

矩阵 $\boldsymbol{W}$ 的列向量 $\boldsymbol{w}_i$ 是相互正交的。因此，（6）可以重写为：

$$\begin{aligned} \mathcal{J}_{\mathrm{MSE}}^{\mathrm{PCA}}\left(\boldsymbol{W}\right) &= \arg\min\left[E\left\{\left\|\boldsymbol{x}\right\|^2\right\} - E\left\{\sum_{i=1}^{m}\left(\boldsymbol{w}_i^T \boldsymbol{x}\right)^2\right\}\right] \\ &= \arg\min\left[tr\left(C_{\boldsymbol{xx}}\right) - \sum_{i=1}^{m} \boldsymbol{w}_i^T C_{\boldsymbol{xx}} \boldsymbol{w}_i^T\right] \end{aligned} \tag{7}$$

更为实用的 PCA 方法是 K-L（Karhunen-Loéve）变换。下面从解相关的角度导出如何确定特征矩阵 $\boldsymbol{W}$ 中各列向量 $\boldsymbol{w}_i\left(i = 1, 2, \cdots, m\right)$ 对应的主成分。首先定义随机向量 $\boldsymbol{x}$ 的自协方差函数，假定随机向量 $\boldsymbol{x}$ 已经做过去中心化处理，其自协方差函数(矩阵)与自相关函数(矩阵)相等，即：

$$\Psi_{\boldsymbol{xx}} = E\left\{\boldsymbol{x}\boldsymbol{x}^T\right\} = C_{\boldsymbol{xx}} = \begin{bmatrix} c(1,1) & c(1,2) & \cdots & c(1,n) \\ c(2,1) & c(2,2) & \cdots & c(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ c(n,1) & c(n,2) & & c(n,n) \end{bmatrix} \tag{8}$$

由于随机向量 $\boldsymbol{x}$ 各元素间存在相关性，自协方差/自相关矩阵不是对角阵（更不是单位阵）。为解开随机向量 $\boldsymbol{x}$ 各元素间存在的相关性，方法当然是引入变换阵 $\boldsymbol{W}$ ，并建立变换关系：

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} \tag{9}$$

期望变换后的随机向量 $\boldsymbol{y}$ 的各元素之间没有相关性。引入随机向量 $\boldsymbol{y}$ 的自相关矩阵：

$$\Psi_{yy}=E\left\{\boldsymbol{y}\boldsymbol{y}^T\right\}=E\left\{\boldsymbol{W}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{W}^T\right\}=\boldsymbol{W}E\left\{\boldsymbol{x}\boldsymbol{x}^T\right\}\boldsymbol{W}^T=\boldsymbol{W}C_{xx}\boldsymbol{W}^T=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_m\end{bmatrix} \tag{10}$$

众所周知，自相关矩阵 C_{xx} 具有（半）正定对称性，存在如下的正交分解，

$$C_{xx}=U\Sigma U^T \tag{11}$$

其中，U 是正交阵，U^T 是其转置阵，Σ 是对角阵。

把（11）代入（10 式），有

$$\Psi_{yy}=\boldsymbol{W}C_{xx}\boldsymbol{W}^T=\boldsymbol{W}U\Sigma U^T\boldsymbol{W}^T=P\Sigma P^T \tag{12}$$

其中，$P=WU$。很显然，若变换阵的 $\boldsymbol{W}$ 列向量选成正交阵的 U 行向量，则有 $P=\mathrm{I}$。对角阵 Σ 具有如下形式：

$$\Sigma=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0 & 0 & \cdots & 0\\ 0 & \lambda_1 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_m & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0 & \lambda_{m+1}=0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \lambda_n=0\end{bmatrix} \tag{13}$$

经过这样的变换后，随机向量 $\boldsymbol{y}$ 的各元素之间就没有了相关性。K-L 变换的本质就是这样。

实际应用中，也可以用如下提法：

$$\mathcal{J}_{MSE}^{\mathrm{PCA}}(\boldsymbol{W})=\arg\min\sum_{k=m+1}^{n}\lambda_k \tag{14}$$

进行 PCA 处理。

如果用下列提法：

$$\mathcal{J}_{Kerl}^{\mathrm{PCA}}(\boldsymbol{W})=\arg\min\sum_{k=1}^{m'}\lambda_k,\quad m'\le m \tag{15}$$

则（15）式奠定了用所谓的核范数进行 RPCA 的理论基础。

可以看出，无论是方差极大化 PCA 方法、最小均方误差 PCA 方法和 K-L 变换方法本质上是一样的。K-L 变换方法计算上更具操作性；方差极大化 PCA 方法几何含义更清晰；最小均方误差 PCA 方法与 Hilbert 空间中线性逼近的关系更密切。

另外，PCA 方法与信号的 Fourier 分析方法在高斯线性平稳假设成立的条件下，二者是完全等价的。局部平稳条件下，K-L 变换的结果更优。不论是 K-L 变换，还是 Fourier 变换，都是把信号变换到不同的正交基函数中去表达信号，解开信号各分量之前的相关性。一般来说，选定的基函数是正交的，这样会减小信号的冗余。K-L 变换是根据信号的统计特征来选择基函数的，基函数是变化的。Fourier 变换也是信号变换，

其基函数是 e^{iwt}，是固定的。K-L 变换是均方误差意义下的最佳变换，它在数据压缩技术中占有重要地位。

值得一提的是，PCA 方法也可以用在线方式实现，存在自适应 PCA 方法（迭代最小二乘 RLS PCA 方法），也可以用神经网络实现 PCA。

4 RPCA 方法

信号处理中噪音为 Gauss 分布的假设在对信号处理要求越来越高的时代严重限制了信号分析精度的提高。改变信号模型，考虑非高斯噪音情形，进行更高精度的信号处理成了必须要做的事情。

为此，把实测信号视为三部分：

$$\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{G}}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{S}}(x,y,t) \tag{16}$$

其中，$\boldsymbol{\eta}_G(x,y,t)$ 为 Gauss 噪声；$\boldsymbol{\eta}_S(x,y,t)$ 为 Sparse 噪声。$\boldsymbol{S}(x,y,t)$ 为数据中包含的信号。信号处理包括三种基本的模式。$\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{G}}(x,y,t)$ 是最经典的模式，常规线性信号模型+ℓ_2 范数下反问题可以很好地解决这类信号分析问题；$\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{S}}(x,y,t)$ 是稀释脉冲噪音（椒盐噪音，或非高斯噪音）存在的情形，RPCA 是解决此类问题的典型做法，提出的是线性模型和核范数+l_1 范数下反问题优化求解方法。$\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{G}}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{S}}(x,y,t)$ 是更复杂的模式，需要用线性模型和核范数+l_1 范数并结合 l_2 范数数据逼近下的反问题来对付此类问题。更为复杂的一类问题可以表达为 $\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{C}}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{G}}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{S}}(x,y,t)$，其中 $\boldsymbol{\eta}_C(x,y,t)$ 为相干噪音，此类问题的解决还要考虑基函数字典，用不同的基函数族来表示信号和相干噪音。

RPCA 应对的基本信号模型为：

$$\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{S}}(x,y,t) \tag{17}$$

一般地，假设要分析的是高维信号，具有空间相关性。当然一维信号分析同样可以在此框架下进行。不失一般性，总可以假设（排成矩阵形式后的）高维信号具有低秩性，则（17）式的处理借用如下反问题的求解：

$$\begin{aligned}&\mathcal{J}_{Kerl}^{\mathrm{RPCA}}(\boldsymbol{W})=\arg\min\left\{\sum_{k=1}^{m'}\lambda_k+\alpha\left\|\eta_S(x,y,t)\right\|_1\right\}\\&S.t.\left\|\sum_{k=1}^{m'}\left(\boldsymbol{w}_k^T\boldsymbol{u}\right)\boldsymbol{w}_k+\eta_S(x,y,t)\right\|^2\le\varepsilon\end{aligned} \tag{18}$$

对于稀疏脉冲噪音存在的情况下，为什么一定要解核范数下的反问题，即为什么用 RPCA 方法进行信号的建模及后续的去噪、插值、压缩等应用？主要原因是希望利用数据驱动的方式对局部窗数据在稀疏脉冲噪音存在的情形下，建立一个较优的模型。（18）式显然是个凸问题，与 SVD 方法结合，存在有效率较高的解法。

用基函数逼近的方法，必然要引入至少 L1 范数下的数据偏差项。这样的反问题提法，目前缺乏有效的算法。这是 RPCA 在高维数据（信号）分析中得到广泛应用的根本原因。　　关于 RPCA 中反问题（18）式的具体解法，参考 Generalized Principal Component Analysis（Vidal，Ma 和 Sastry，2016）一书。

5 CCA 方法

PCA 方法主要是解开 n 维随机向量 $\boldsymbol{x}$ 各元素间存在相关性，也可以说是利用主分量形成基函数空间从而最佳地表达维随机向量 $\boldsymbol{x}$。与 Fourier 分析方法相比，PCA 方法是数据驱动的，可以建立更优的模型逼近

n 维随机向量 $\boldsymbol{x}$。PCA 方法当然可以推广到高维情形，关于这方面的内容，参考“Nonnegative Matrix and Tensor Factorizations – Applications to Exploratory Multi-way Data Analysis and Blind Source Separation”一书。

但是，对于两个 n 维随机向量或多个 n 维随机向量如何抽取它们之间共同的特征量，从而实现各种应用目的？这正是 CCA 方法要应对的问题！

假设有两个 n 维随机向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，存在各自的主分量 $\boldsymbol{w}_{x,1}$ 和 $\boldsymbol{w}_{y,1}$，有

$$z_{x,1}=\left(\boldsymbol{w}_{x,1},\boldsymbol{x}\right)=\boldsymbol{w}_{x,1}^T\boldsymbol{x},\quad z_{y,1}=\left(\boldsymbol{w}_{y,1},\boldsymbol{y}\right)=\boldsymbol{w}_{y,1}^T\boldsymbol{y} \tag{19}$$

为了考察 $z_{x,1}$ 和 $z_{y,1}$ 之间的相关性，定义如下的相关系数：

$$\begin{aligned}\rho&=\frac{E\left\{z_{x,1}z_{y,1}\right\}}{\sqrt{E\left\{z_{x,1}^2\right\}E\left\{z_{y,1}^2\right\}}}=\frac{E\left\{\left(\boldsymbol{w}_{x,1}^T\boldsymbol{x}\right)\left(\boldsymbol{w}_{y,1}^T\boldsymbol{y}\right)\right\}}{\sqrt{E\left\{\left(\boldsymbol{w}_{x,1}^T\boldsymbol{x}\right)^2\right\}E\left\{\left(\boldsymbol{w}_{y,1}^T\boldsymbol{y}\right)^2\right\}}}=\frac{E\left\{\boldsymbol{w}_{x,1}^T\boldsymbol{x}\boldsymbol{y}^T\boldsymbol{w}_{y,1}\right\}}{\sqrt{E\left\{\boldsymbol{w}_{x,1}^T\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{w}_{x,1}\right\}E\left\{\boldsymbol{w}_{y,1}^T\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{w}_{y,1}\right\}}}\\&=\frac{\boldsymbol{w}_{x,1}^TE\left\{\boldsymbol{x}\boldsymbol{y}^T\right\}\boldsymbol{w}_{y,1}}{\sqrt{\boldsymbol{w}_{x,1}^TE\left\{\boldsymbol{x}\boldsymbol{x}^T\right\}\boldsymbol{w}_{x,1}}\sqrt{\boldsymbol{w}_{y,1}^TE\left\{\boldsymbol{y}\boldsymbol{y}^T\right\}\boldsymbol{w}_{y,1}}}=\frac{\boldsymbol{w}_{x,1}^T\Sigma_{xy}\boldsymbol{w}_{y,1}}{\sqrt{\boldsymbol{w}_{x,1}^TC_{xx}\boldsymbol{w}_{x,1}}\sqrt{\boldsymbol{w}_{y,1}^TC_{yy}\boldsymbol{w}_{y,1}}}\end{aligned} \tag{20}$$

其中，Σ_{xy} 代表 n 维随机向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的互相关或互协方差矩阵，零均值情形下，二者是等价的。

另外，我们定义如下等式：

$$E_{xy}=E\left\{\begin{bmatrix}\boldsymbol{x}\\\boldsymbol{y}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^T & \boldsymbol{y}^T\end{bmatrix}\right\}=\begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\\Sigma_{yx} & \Sigma_{yy}\end{bmatrix} \tag{21}$$

很显然，（20）式中的互相关和自相关矩阵都是未知的，需要由数据驱动来统计估计出来。在此基础上，通过最大化（20）式中的相关系数，可以找到 n 维随机向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 最相似的特征量，利用这样的特征量建模，可以得到十分相似的 $\overline{\boldsymbol{x}}=\boldsymbol{W}_x\boldsymbol{x}$ 和 $\overline{\boldsymbol{y}}=\boldsymbol{W}_y\boldsymbol{y}$。

根据上述分析，CCA 问题的提法应为：

$$\begin{aligned}&\mathcal{J}^{\mathrm{CCA}}\left(\boldsymbol{w}_{x,1},\boldsymbol{w}_{y,1}\right)=\arg\max_{\boldsymbol{w}_{x,1},\boldsymbol{w}_{y,1}}\boldsymbol{w}_{x,1}^T\tilde{\Sigma}_{xy}\boldsymbol{w}_{y,1}\\&s.t.\begin{cases}\boldsymbol{w}_{x,1}^T\tilde{C}_{xx}\boldsymbol{w}_{x,1}=1\\\boldsymbol{w}_{y,1}^T\tilde{C}_{yy}\boldsymbol{w}_{y,1}=1\end{cases}\end{aligned} \tag{22}$$

为了对比，把 PCA 问题的提法罗列一下：

$$\mathcal{J}^{\mathrm{PCA}}\left(\boldsymbol{w}_{x,1}\right)=\arg\max_{\boldsymbol{w}_{x,1}}\boldsymbol{w}_{x,1}^T\tilde{C}_{xx}\boldsymbol{w}_{x,1} \tag{23}$$

上述两式中的协方差函数加上标“~”代表它们都是用数据估计出来的。

对比可知，PCA 方法寻找 n 维随机向量 $\boldsymbol{x}$ 蕴含的特征量，对 n 维随机向量 $\boldsymbol{x}$ 建模；CCA 方法寻找 n 维随机向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间最一致的特征量，分别对 n 维随机向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 建模。很显然，CCA 方法可以找到 n 维随机向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间非常相关的部分。在很多应用场景中这无疑是非常有用的。

与 PCA 方法类似，我们同样不能直接求解（23）式定义的优化问题确定 $\overline{\boldsymbol{x}}=\boldsymbol{W}_x\boldsymbol{x}$ 和 $\overline{\boldsymbol{y}}=\boldsymbol{W}_y\boldsymbol{y}$。具体的计算方法参考“Canonical Correlation Analysis in Speech Enhancement”一书。

6 独立分量分析（ICA）方法

6.1 ICA 方法的概述

PCA/RPCA 方法在数据驱动的建模及相应的场景中得到了广泛的应用。可以说，深刻地领会了 PCA/RPCA 的基本思想基本上把握了当今数据分析的本质问题。但是，PCA/RPCA 只是在 L_2 范数（对应 Gauss 噪音假设）和 L1 范数（对应广义 Gauss）下对数据 $\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{G}}(x,y,t)$ 和数据 $\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_{\boldsymbol{S}}(x,y,t)$ 中的信号建立“最佳”的线性模型。它们仅仅利用了随机信号的二阶统计量。CCA 方法也仅仅利用了随机信号的二阶统计量。具体地，利用（半）正定对称的自相关函数的正交分解确定随机信号的特征（向）量，利用这些特征（向）量的线性组合对信号进行建模，进而解决各种场景下的应用问题。

从统计信号分析知，自相关函数提取出了一维和高维（大于等于二维）信号中的线性相关信息。在高维信号中，相关信息非常直观，就是（局部）数据体中蕴含的结构和/或纹理。

事实上，从 Learning From Data 的观点看，数据分析的核心问题是建立起实测数据与产生或影响实测数据的隐变量之间的关系，也就是建立起一个数学模型描述二者之间的关系。一般地，仅仅由数据本身驱动，我们只能建立起实测数据与表达该数据的隐变量之间的关系。PCA/RPCA 方法就是建立起这种关系的典型方法，它们属于无监督学习方法。

在有监督学习的情况下，我们很有可能建立起某种因果关系，当前这主要是利用标签样本，通过深度网络算法的训练，得到输入与输出间一定的因果关系。然后把训练好的这种因果关系用于决策判断。当前基于联结主义（深度神经网络）的所谓人工智能就是按上述方法解决一定的问题。应该说，一定场景下（譬如与图像识别和语音识别相关的场景）应用效果是显著的。

笔者认为，实测数据与其因变量的函数关系的建立，是人类智能擅长解决的问题，需要观察、实验、抽象、推理、分析、检验才可以建立比较可靠的函数关系。这类关系具有高度的抽象性和一般性。Newton 三定理对宏观物体的运动规律的广泛指导性就是这种因果关系的典型。人工智能远远达不到从实测数据中建立这种因果关系的程度。

建立实测数据与表达数据的潜变量之间的关系在数据分析中有更为现实的意义，当前的重点工作还在于此。但是，仅仅利用 PCA/RPCA 方法利用随机信号的二阶统计量，并用解相关后的特征（向）量（某种潜变量）的线性组合，建立一种模型来预测实测数据中的信号，是不彻底的。从统计理论指导，不相关的特征（向）量不一定是独立的。我们当然希望用独立的特征（向）量的线性组合对信号进行建模。不过，若随机信号是 Gauss 分布的，PCA 方法得到的特征（向）量也是统计独立的。

因此，若随机信号不是 Gauss 分布的，就需要引入独立分量分析 ICA（Independent Component Analysis）方法建立更好的模型来表达信号。很显然，RPCA 方法建立的模型与 ICA 方法建立的模型之间是存在一定关系的，这需要理论分析。能否修改 RPCA 来实现 ICA 也是值得考虑的。

下面简述一下 ICA 方法的基本思想和做法。

ICA 针对的典型问题是：假设信号向量 $\boldsymbol{x}$ 是由统计独立的潜变量 $\boldsymbol{s}$ 线性组合形成的，

$$\boldsymbol{x}=A\boldsymbol{s} \tag{24}$$

其中，矩阵 A 称为混合矩阵。已知实测离散数据 $\boldsymbol{x}$ 同时估计混合矩阵 A 和独立潜变量 $\boldsymbol{s}$。上述问题有很多对应的应用场景，最经典的就是盲源分离问题（鸡尾酒会问题）。混叠源分离是勘探地震中的典型应用。

假定存在一个变换阵 $\boldsymbol{W}$，作用于实测数据上，

$$\boldsymbol{z}=W\boldsymbol{x} \tag{25}$$

使得估计的独立潜变量 $\boldsymbol{z}$ 中的分量 $z_i, i=1,2,\cdots,n$ 是独立的。则变换阵 $\boldsymbol{W}$ 就称为解混矩阵。若期望估计的独立潜变量 $\boldsymbol{z}$ 等于独立潜变量 $\boldsymbol{s}$，就应该有

$$A = W^{-1} \tag{26}$$

因此，ICA 的核心问题就转换为估计解混矩阵。如何能做到对解混矩阵的估计？核心思想是要引入对独立潜变量统计特征的假设，即假设独立潜变量 $\boldsymbol{S}$ 的各分量满足非高斯分布！

6.2 独立潜变量 $\boldsymbol{S}$ 满足非高斯假设的必要性

假设独立潜变量满足如下均匀分布假设：

$$p(s_i) = \begin{cases} \dfrac{1}{2\sqrt{3}}, & 若 s_i \le \sqrt{3}, i = 1,2 \\ 0, & s_i 取其它值 \end{cases} \tag{27}$$

(s_1, s_2) 对应的联合概率分布如图 1。

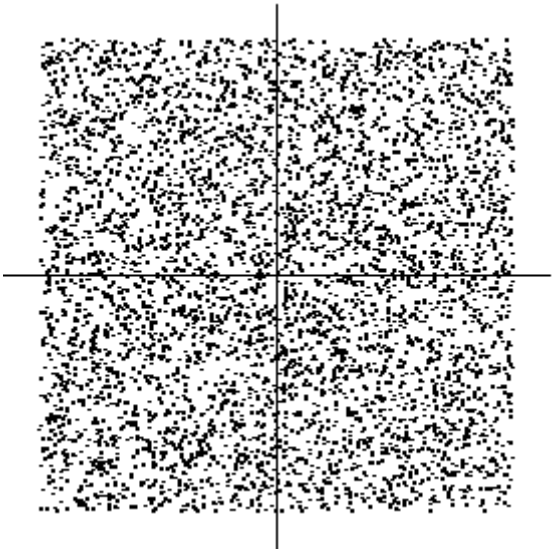


图 1 (s_1, s_2) 满足的联合概率密度函数

引入如下变换：

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \tag{28}$$

(x_1, x_2) 对应的联合概率分布如图 2。

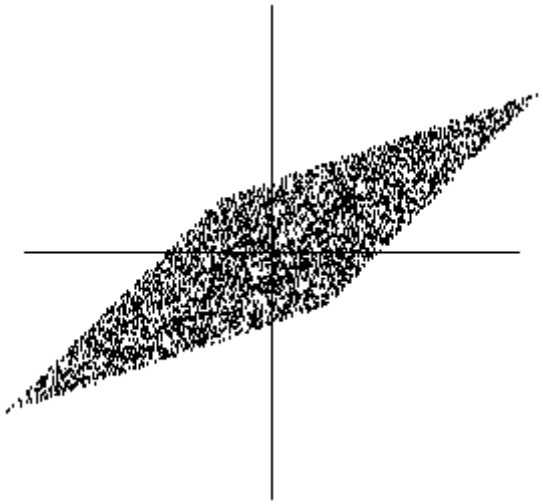


图 2 (x_1, x_2) 满足的联合概率密度函数

由图 2 可以看出，若能估计出 (x_1, x_2) 满足的联合概率密度函数，就是确定图 2 中的边，通过旋转就可以得到 (s_1, s_2) 满足的联合概率密度函数。就是说，这就可以得到对 (s_1, s_2) 的估计。但这样做的前提是 (s_1, s_2) 满足的联合概率密度函数是（27）式所定义的均匀概率分布。

若 (s_1, s_2) 满足如下的高斯概率分布，

$$p(s_1, s_2) = \frac{1}{2\pi} \exp(-\frac{s_1^2 + s_2^2}{2}) \tag{29}$$

经过 $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$ 的变换后，依然满足概率分布。如图 3 所示。这样就不能通过类似图 2 的方法由测量数据 (x_1, x_2) 估计 (s_1, s_2)。

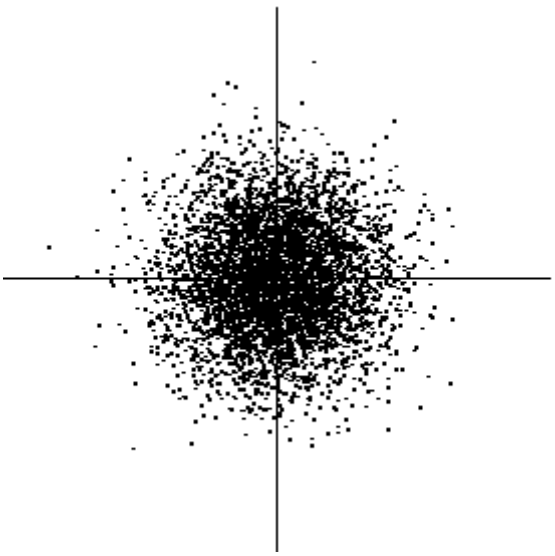


图 3 Gauss 联合概率密度函数

ICA 方法不能对满足 Gauss 分布的实测数据及独立变量进行估计。因此，假设独立潜变量 $\boldsymbol{s}$ 的各分量满足非高斯分布是开展 ICA 的先决条件。

另一方面，ICA 方法具有不唯一性。首先可以看到，

$$\boldsymbol{x} = As = \frac{1}{a} A \cdot a\boldsymbol{s} \tag{30}$$

因此，我们最好可以得到一个归一化的独立潜变量 $\boldsymbol{s}$。

$$\boldsymbol{x} = As = AP^{-1}P\boldsymbol{s} \tag{31}$$

一般地，取变换阵 P 为正交阵，有

$$\boldsymbol{x} = AP^{T}P\boldsymbol{s} \tag{32}$$

因此，我们不可能得到唯一的解混矩阵和唯一的独立潜变量 $\boldsymbol{s}$，只能得到对应某一正交阵 P 下的解混矩阵和独立潜变量 $\boldsymbol{s}$。

6.3 ICA 算法的基本思想

前已述及，ICA 方法不能对 Gauss 分布的独立潜变量 $\boldsymbol{s}$ 进行估计。它只能在下述情况下应用：① 所有非高斯分布的潜变量可以被估计（识别）；② 高斯分布潜变量的线性组合可以被估计（识别），但不可以估计每个的高斯分布潜变量；③ 假如只有一个高斯分布的潜变量，其他是非高斯分布的潜变量，所有潜变量可以被估计。

ICA 方法的核心问题是估计解混矩阵，而估计解混矩阵的基本思想是解混矩阵施加在实测数据上得到的独立潜变量 $\boldsymbol{s}$ 的各分量满足非高斯分布的假设。因此，独立潜变量 $\boldsymbol{s}$ 的各分量间非高斯分布的定量评价指标的定义成为了估计解混矩阵的关键！

独立分量的非高斯分布用什么定量指标来定义？

概率统计理论中的中心极限定理：多个独立随机变量的联合分布更趋向于高斯分布！两个独立随机信号的混合信号的分布比任一个单独的独立信号的分布更接近高斯分布。因此，

ICA 的关键是估计独立随机信号（独立潜变量 $\boldsymbol{s}$）的非高斯性！非高斯性就是独立性（Nongaussian is independent！）！

由概率统计理论知，只有 Gauss 随机变量的高阶累积量为零，则考察随机变量非高斯性就可以用高阶累积量，尤其是四阶累积量。

首先看一下矩和累积量的关系：

$$\begin{aligned}
&c_1(\tau) = m_1 = E[x] \\
&c_2(\tau_1,\tau_2) = m_2 - m_1^2 = E[(x-m_1)^2] \\
&c_3(\tau_1,\tau_2,\tau_3) = m_3 - 3m_1m_2 + 2m_1^3 = E[(x-m_1)^3] \\
&c_4(\tau_1,\tau_2,\tau_3,\tau_4) = m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4 \neq E[(x-m_1)^4] \\
&\cdots\cdots\cdots\cdots
\end{aligned} \tag{33}$$

可以看出，一阶矩和一阶累积量是相等的；若一阶矩（期望）等于零，二阶矩与二阶累积量相等，三阶矩也和三阶累积量相等。关于更多的矩和累积量的讨论，参考“Probability, Random Variables and Stochastic Processes”一书。

为了判断随机信号 $x(t)$ 的 Gauss 性与否，定义如下的偏度和峰度（峭度）

随机信号 $x(t)$ 的偏度：

$$s_x = \frac{c_3}{\sigma^3} \tag{34}$$

随机信号 $x(t)$ 的峰度（峭度）：

$$\varepsilon_x = \frac{c_4}{\sigma^4} = \frac{m_4 - 3m_2^2}{\sigma^4} = \frac{m_4 - 3\sigma^4}{\sigma^4} = \frac{m_4}{\sigma^4} - 3 \tag{35}$$

其中，σ 为方差。为方便，进一步定义如下归一化的峰度（峭度）：

$$k_1 = \frac{E\left\{x^4(t)\right\}}{E^2\left\{x^2(t)\right\}} \tag{36}$$

Gauss 随机信号，$k_1 = 3$；亚 Gauss 随机信号，$k_1 < 3$；超 Gauss 随机信号，$k_1 > 3$。峰度反映了随机信号 $x(t)$ 偏离 Gauss 函数对称性的程度；峰度（峭度）反映随机信号 $x(t)$ 偏离 Gauss 函数尖锐性的程度。图 4 是亚 Gauss（Sub-Gauss）和超 Gauss（Super-Gauss）随机变量概率分布展示，可见亚 Gauss（Sub-Gauss）随机变量概率密度函数有一个平顶；超 Gauss（Super-Gauss）随机变量概率密度函数比 Gauss（Sub-Gauss）随机变量概率密度函数更尖锐。

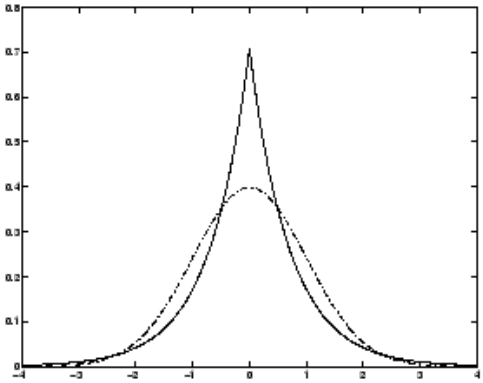


图 4 亚 Gauss（Sub-Gauss）和超 Gauss（Super-Gauss）随机变量概率分布图

由（35）式知，随机信号 $x(t)$ 的峰度不（一定）是非负的，作为考察随机信号 $x(t)$ 是否非线性性的定量度量是不合适的。但是，依然存在基于峰度最大化的梯度导引方法来估计解混矩阵（列向量）$\boldsymbol{w}_i$。遗憾的是，峰度（四阶累积量）对随机信号 $x(t)$ 中的野值异常敏感，基于此的 ICA 方法不稳健。因此，用（信息熵中）负熵代替峰度进行 ICA。

由信息论知道，在所有方差相同的随机变量中，高斯随机变量具有最大的熵！反过来看，熵可以用来测量随机变量的非高斯性。

定义负熵为：

$$\mathcal{J}(\boldsymbol{y}) = H(\boldsymbol{y}_{Gauss}) - H(\boldsymbol{y}) \tag{37}$$

其中，$\boldsymbol{y}_{Gauss}$ 为满足 Gauss 分布的随机变量。随机变量 $\boldsymbol{y}$ 的熵按常规被定义为

$$H(\boldsymbol{y}) = -\int p(\boldsymbol{y}) \log p(\boldsymbol{y}) d\boldsymbol{y} \tag{38}$$

为什么称（37）式为负熵，由高斯随机变量具有最大的熵和（38）式可知。

使用负熵考察随机信号 $x(t)$ 是否非线性性的定量度量，可以实现对非高斯性的最佳估计。但是由于计算上的困难，还要对负熵进行新一步简化。

一种方法是用高阶累积量近似：

$$\mathcal{J}(\boldsymbol{y}) \approx \frac{1}{12} E\left\{\boldsymbol{y}^3\right\}^2 + \frac{1}{48} Kurt(\boldsymbol{y})^2 \tag{39}$$

其中，峰度计算式为 $Kurt(\boldsymbol{y}) = E\{\boldsymbol{y}^4\} - 3(E\{\boldsymbol{y}^2\})^2$ （40）

另一种方法是基于最大熵方法提出的：

$$\mathcal{J}(\boldsymbol{y}) \approx c\ [\mathrm{E}\{G(\boldsymbol{y})\} - E\{G(v)\}]^2 \tag{41}$$

其中，v 是零均值单位方差 Gauss 随机变量。函数一般选为：

$$G_1(u) = \frac{1}{a_1}\log\cosh a_1 u,\ G_2(u) = -\exp(\frac{-u^2}{2}) \quad where\ 1 \le \mathrm{a}_1 \le 2. \tag{42}$$

求解上述定义下的优化问题，就可以实现 ICA 计算。

6.4 典型的 ICA 算法

此处仅仅罗列两种最基本的 ICA 算法，更多的算法讨论参考“Independent Component Analysis”一书。

6.4.1 采用峰度的梯度方法

峰度定义为：$Kurt(\boldsymbol{y}) = E\{\boldsymbol{y}^4\} - 3(E\{\boldsymbol{y}^2\})^2$，其中 $\boldsymbol{y} = \boldsymbol{w}^T\boldsymbol{z}$。首先要对实测数据进行预白化处理得到 $\boldsymbol{z}$，然后计算峰度的梯度，构造迭代公式，得到使峰度最大（非高斯性最大）的 $\boldsymbol{w}$。

峰度的梯度：$\dfrac{\partial\left|Kurt(\boldsymbol{w}^T\boldsymbol{z})\right|}{\partial\boldsymbol{w}} = 4\mathrm{sign}\left(Kurt(\boldsymbol{w}^T\boldsymbol{z})\left[E\left\{\boldsymbol{z}(\boldsymbol{w}^T\boldsymbol{z})^3\right\} - 3\boldsymbol{w}\|\boldsymbol{w}\|^2\right]\right)$ （43）

迭代公式为：$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + \lambda\left.\dfrac{\partial\left|Kurt(\boldsymbol{w}^T\boldsymbol{z})\right|}{\partial\boldsymbol{w}}\right|_k$

迭代收敛到稳定的解混矩阵的列向量。

6.4.2 基于负熵的梯度算法

具体算法包括如下步骤：

① 对实测数据进行去中心化处理，使其均值为 0；

② 然后进行白化，得到 $\boldsymbol{z}$；

③ 选择一个具有单位范数的初始化向量 $\boldsymbol{w}$，并赋予 γ 一个初始值；

④ 更新 $\boldsymbol{w}$，$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + \Delta\boldsymbol{w}$。其中，$\Delta\boldsymbol{w} \propto \gamma\boldsymbol{z}g\left(\boldsymbol{w}^T\boldsymbol{z}\right)$，$g\left(y\right) = \tanh\left(a_1 y\right), 1 \le a_1 \le 2$；

⑤ 标准（归一）化 $\boldsymbol{w}$，$\boldsymbol{w} \leftarrow \boldsymbol{w} \big/ \|\boldsymbol{w}\|$；

⑥ 如果 γ 的符号事先未知，更新 γ：$\Delta\gamma \propto \left(G\left(\boldsymbol{w}^T\boldsymbol{z}\right) - E\left\{G\left(\boldsymbol{v}\right)\right\}\right) - \gamma$；

⑦ 如果尚未收敛，返回步骤④。

6.5 ICA 方法的应用

我们已多次提及，现代信号分析的基础分析方法就是线性模型+Gauss 分布（L2 范数）下的 Fourier 分析方法和 PCA 方法。由 PCA 方法知道，这仅仅利用了数据间的相关信息（二阶统计量）进行建模来解决具体应用问题，ICA 分析中的建模进一步假设潜变量是独立的，比解相关更为本质。图 5 在图像去噪中的应用对比可以显著地体现出 ICA 的优势。

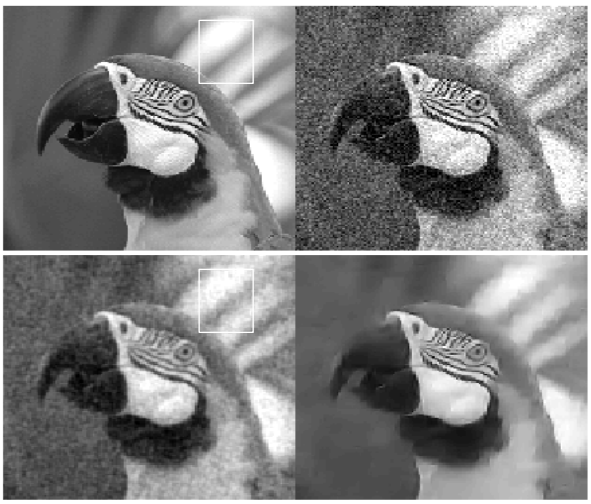


图 5 最佳 Wiener 滤波与 ICA 滤波结果对比图 5a 是无噪音图片；图 5b 是加噪音图片；图 5c 是 Weiner 最佳滤波结果；图 5d 是 ICA 方法的滤波结果。这与理论预计是一致的，ICA 方法的滤波效果远好于 Weiner 最佳滤波效果。

由于 ICA 方法建立的信号模型除了解相关（预白化处理）外，又考虑了特征（向）量间的独立性，因此能更好地刻画数据中蕴含的信号，当然地在去噪等应用中会取得更好的效果。

7 地震信号建模的综合评价

在复杂地表条件和复杂地下介质变化的情况下，地表观测的地震波场是异常复杂的。无论如何总可以用概念模型：$\boldsymbol{u}(x,y,t)=\boldsymbol{S}(x,y,t)+\boldsymbol{\eta}_C(x,y,t)+\boldsymbol{\eta}_G(x,y,t)+\boldsymbol{\eta}_S(x,y,t)$ 来表示。目前除了海上和东部地表比较简单的探区外，包括渤海湾浅水区域在内的广大西部、西北部和西南部探区，炮集的信噪比都比较差，山前带探区尤其差。近地表高程变化大和近地表不适合激发有效波是最大的噪音来源。可以说，单炮道集信噪比低的主要原因就是复杂地表条件引起的。

在这种的条件下，对地震信号（实际上大多是对成像有干扰的波场）的建模，最基本的方法当然是利用弹性波方程对复杂地表条件下的波场进行预测。但是这是很难做到的，一是复杂的近地表弹性参数场基本是不可能准确建立的；二是弹性波动方程也不一定能模拟复杂近地表介质条件下实际波场；三是非水平地表+复杂介质下的地震波正演模拟算法的精度很难保证。这也是到目前为止，复杂地表区勘探中并没有广泛应用波动方程预测并减去的方法进行去噪的根本原因。

低信噪比数据的去噪音主要还是靠对地震数据中信号（也包括对相干噪音）的更好建模实现对地震信号的各种预处理，主要是去噪音和数据规则化。高效采集中的数据解混叠也是重要的应用场景。

众所周知，多次覆盖的地震数据（信号）是高维的，单炮道集是三维的，更多的体现为五维的。现代信号分析理论清楚地说明：在高维空间中，可以更稀疏地表达信号，可以更好地感知到信号的相关性，可以对信号进行更好地建模。我们对地震数据的抽象为：线性和非线性的、由时距关系确定的、以及由地震子波体现出的同相轴飘在满足不同统计特征的随机噪音中。对地震数据进行加窗处理后，可以认为：线性的、由时距关系确定的，以及由地震子波体现出的同相轴飘在满足不同统计特征的随机噪音中的表述是地震数据的合理抽象。

在上述地震数据抽象下，PCA 和 RPCA 是非常合理的地震数据建模和处理方法。若进一步提高建模的精度，ICA 是合理的选择。

我认为，目前地震数据分析中真正困难点包括：数据采样率过低、数据采样过于不规则（存在较大的空白区域）、地表不一致性导致的振幅变化过快、直达波和面波（面波散射波）幅值远大于反射波幅值、静校正引起的相位非线性变化等。这些因素破坏了信号建模理论中的蕴含的基本假设。这些基本假设是：线性模型、统计平稳、噪音 Gauss 或广义 Gauss 分布。

我认为现代信号分析的理论不会突飞猛进地发展到可以考虑非线性、非平稳和非高斯分布的情形。我们只能在意识到上述问题存在的前提下，把当前最先进的方法恰当地用好。能推进到非线性信号预测最好，这是目前最急需改变的。

8 结论与讨论

地震波成像是油气地震勘探的核心，目前已经发展到反演成像为核心的阶段，对地震数据信噪比的要求远高于仅仅做偏移成像。复杂地表条件是导致地震数据信噪比低的根本原因，用波动理论对复杂的近地表波场进行预测并减去这些不能成像的干扰波场是众所周知的首要方法，但这是不可行的方法。实际油气地震勘探中的地震数据（信号）分析依然是依赖对实际数据中包含的信号进行高精度的建模，基于建立的模型进行各种应用。

对信号进行建模是数据分析的核心！

信号建模的基本理论包括如下两个在一定条件下可以统一在一起的方向：利用事先选定的基函数的线性组合进行信号表达，泛函分析中的函数逼近理论建立了数据中信号表示的数学分析基础；PCA/ICA 方法是用数据驱动的方式来探测数据中的信号是由什么潜变量生成的，其中 RPCA、字典学习、NMF（非负矩阵分解）是 PCA 的先进实现方式。PCA 是不完全的 ICA！深度学习（DL）也隐含地在做数据表达的工作。

当前的信号建模主要基于如下的基本假设：线性模型、统计平稳、噪音 Gauss 或广义 Gauss 分布。应对复杂的数据情形，在高维空间中，考虑非线性、非平稳和非高斯分布情形，提出新的信号建模方法，是信号（图像）分析的进一步发展方向。

参考文献

[1] René Vidal, Yi Ma and S. Shankar Sastry, 2016, Generalized Principal Component Analysis, Springer.

[2] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, 2001, Independent Component Analysis, John Wiley & Sons, Inc.

[3] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari, 2009, Nonnegative Matrix and Tensor Factorizations – Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, John Wiley & Sons, Inc.

[4] Jacob Benesty and Israel Cohen, 2018, Canonical Correlation Analysis in Speech Enhancement, Springer.

[5] Sergios Theodoridis, 2015, Machine Learning: A Bayesian and Optimization Perspective, Elsevier

[6] Rafael C. Gonzalez, and Richard E. Woods, 2010, Digital Image Processing, Third Edition, Pearson Education.

[7] Otmar Scherzer, 2011, Handbook of Mathematical Methods in Imaging, Springer.

[8] Sergios Theodoridis and Konstantinos Koutroumbas, 2009, Pattern Recognition, Fourth Edition, Elsevier

[9] Athanasios Papoulis and S. Unnikrishna Pillai, 2002, Probability, Random Variables and Stochastic Processes, The McGraw-Hill Companies, Inc

[10] Boualem Boashash (Editor), 2016, Time-Frequency Signal Analysis and Processing – A Comprehensive Reference, Second Edition, Elsevier

[11] Leon Cohen, 1995, Time-Frequency Analysis, Prentice-Hall PTR

[12] Stéphane Mallat, 2009, A Wavelet Tour of Signal Processing, Third Edition, Elsevier

[13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, 2016, Deep Learning, Http://www.deeplearningbook.org/

[14] Abma, R., and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, no. 6, E91–E97, doi: 10.1190/1.2356088.

[15] Asif, S., W. Mantzel, and J. Romberg, 2009, Random channel coding and blind deconvolution: Presented at the Allerton Conference on Communication, Control, and Computing.

[16] Beasley, C. J., 2008, A new look at marine simultaneous sources: The Leading Edge, **27**, 914–917, doi: 10.1190/1.2954033.

[17] Bobin, J., J. Starck, and R. Ottensamer, 2008, Compressed sensing in astronomy: Journal of IEEE, Selected Topics in Signal Process, **2**, no. 5, 718–726, doi: 10.1109/JSTSP.2008.2005337.

[18] Candès, E., J. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, no. 8, 1207–1223, doi: 10.1002/cpa.20124.

[19] Candès, E. J., and L. Demanet, 2005, The curvelet representation of wave propagators is optimally sparse: Communications on Pure and Applied Mathematics, **58**, no. 11, 1472–1528, doi: 10.1002/cpa.20078.

[20] Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: SIAM, Multiscale Modeling and Simulation, **5**, no. 3, 861–899, doi: 10.1137/05064182X.

[21] Candès, E. J., J. Romberg, and T. Tao, 2006b, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on Information Theory, **52**, no. 2, 489–509, doi:10.1109/TIT.2005.862083.

[22] Demanet, L., and L. Ying, 2007, Wave atoms and sparsity of oscillatory patterns: Applied and Computational HarmonicAnalysis, **23**, no. 3, 368–387, doi: 10.1016/j.acha.2007.03.003.

[23] Donoho, D., A. Maleki, and A. Montanari, 2009, Message passing algorithms for compressed sensing: Proceedings of the National Academy of Sciences.

[24] Donoho, D., and J. Tanner, 2009, Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing: Philosophical Transactions of the Royal Society A: Mathematical: Physical and Engineering Sciences, **367**, 4273–4293.

[25] Donoho, D. L., 2006, Compressed sensing: IEEE Transactions on Information Theory, **52**, no. 4, 1289–1306, doi: 10.1109/TIT.2006.871582.

[26] Donoho, D. L., and X. Huo, 2001, Uncertainty principles and ideal atomic decomposition: IEEE Transactions on Information Theory, **47**, no. 7, 2845–2862, doi: 10.1109/18.959265.

[27] Donoho, D. L., and B. F. Logan, 1992, Signal recovery and the large sieve: SIAM Journal on Applied Mathematics, **52**, no. 2, 577–591, doi: 10.1137/ 0152031.

[28] Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck, 2006, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit: Technical report, Stanford Statistics Department TR-2006-2. http://stat.stanford.edu/idrori/StOMP.pdf.

[29] Donoho, P., R. Ergas, and R. Polzer, 1999, Development of seismic data compression methods for reliable, low-noise performance: SEG International Exposition and 69thAnnual Meeting, 1903–1906.

[30] Hennenfent, G., and F. J. Herrmann, 2008, Simply denoise: wavefield reconstruction via jittered undersampling: Geophysics, **73**, no. 3, V19–V28.

[31] Herrmann, F., D. Wang, G. Hennenfent, and P. Moghaddam, 2008a, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, no. 1, A1–A5, doi: 10.1190/1.2799517.

[32] Herrmann, F. J., Y. A. Erlangga, and T. T. Lin, 2009b, Compressive simultaneous full-waveform simulation: Geophysics, **74**, no. 4, A35–A40.

[33] Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, no.1, 233–248, doi: 10.1111/j.1365-246X.2007.03698.x.

[34] Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2008b, Sparsity- and continuity-promoting seismic image recovery with curvelet frames: Journal of Applied and Computational Harmonic Analysis, **24**, no. 2, 150–173, doi: 10.1016/j.acha.2007.06.007.

[35] Krohn, C., and R. Neelamani, 2008, Simultaneous sourcing without compromise: 70th EAGE Conference & Exhibition, B008.

[36] Rauhut, H., K. Schnass, and P. Vandergheynst, 2008, Compressed sensing and redundant dictionaries: IEEE Transactions on Information Theory, **54**, no. 5, 2210–2219, doi: 10.1109/TIT.2008.920190.

[37] Romberg, J., 2009, Compressive sensing by random convolution: SIAM Journal on Imaging Sciences, **2**, no. 4, 1098–1128, doi: 10.1137/ 08072975X.

[38] Sacchi, M. D., T. J. Ulrych, and C.Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: IEEE Transactions on Signal Processing, **46**, no. 1, 31–38, doi: 10.1109/78.651165.

[39] Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: Geophysics, **59**, 938–945, doi: 10.1190/1.1443653.

[40] Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58, doi: 10.1190/1.1438154.

[41] Trad, D., T. Ulrych, and M. Sacchi, 2003, Latest views of the sparse radon transform: Geophysics, **68**, 386–399, doi: 10.1190/1.1543224.

[42] Trad, D. O., 2003, Interpolation and multiple attenuation with migration operators: Geophysics, **68**, 2043–2054, doi: 10.1190/1.1635058.

[43] Ulrych, T. J., and C. Walker, 1982, Analytic minimum entropy deconvolution: Geophysics, **47**, 1295–1302, doi: 10.1190/1.1441389.

[44] Unser, M., 2000, Sampling — 50 years after Shannon: Proceedings of the IEEE, **88**, no. 4, 569–587, doi: 10.1109/5.843002

[45] van Groenestijn, G. J.A., and D. J. Verschuur, 2009, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, no. 3, A23–A28, doi: 10.1190/1.3111115.

[46] Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, no. 4, V87–V95, doi: 10.1190/1.1993713.

[47] Zwartjes, P., and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: Geophysics, **71**, no. 6, V171–V186, doi:10.1190/1.2348633.

[48] Zwartjes, P. M., and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: Geophysics, **72**, no. 1, V21–V32, doi: 10.1190/1.2399442.

反问题提法与解法的本质逻辑探讨

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】油气地震勘探，无论狭义地或是广义地看，本质上都是一个反问题。广义地，凡涉及到由可观测量来估计或推演与可观测量有关的不可观测量的问题都可认为是一个反问题。这样定义的反问题在自然科学和社会科学的众多领域中都非常广泛地存在。狭义地，通过直接测量到的系统输出（数据），基于描述系统的状态方程，估计系统参数和系统激励的问题称为反问题，这是反问题的经典描述。在当今数据分析学科中，可观测量与潜变量之间关系的建立是核心问题，这也是一个典型的反问题。由于可观测量的不完整和受“噪音”污染，以及正问题的不确定性，反问题从提法到解法似乎都存在主观性。这就导致对反问题的认识很纠结，看起来主观构造的反问题却具有广泛的应用，成为各个领域中离不开的解决问题的利器。本文试图从反问题的提法和反问题的解法两个角度，探讨反问题的根本逻辑，回答反问题到底想做什么、能做什么、以及如何能做到什么。

【关键词】反问题，反问题的定义，反问题的提法，反问题的解法

1 引言

油气地震勘探目标的复杂化、地震数据采集技术的进步、计算机科学与技术的进步、大数据分析与人工智能时代的来临，推动着地震波成像理论方法与技术的不断深化。地震波偏移成像逐渐退出，地震波反演成像正成为地震波成像的核心。FWI+LS_RTM 的广受关注就是最有力的证明。事实上，油气藏的定量描述与刻画是一个更广泛意义下的反问题，而“制造”一个人工智能油气勘探系统（或一个油气勘探机器人）是在解一个更为广义的反问题。

大数据分析和人工智能在智能数据采集设备、互联网和 4G/5G 网络、高性能计算与云计算等等技术的促进下，最近几年有了突飞猛进的发展，迅速地改变了和改变着人们的生产和生活形态。大数据分析的最核心问题是完全由数据驱动的方式建立起可观测量和潜变量之间的依赖关系。这是一个新的老问题。经典的回归分析就是在做此事，但即便是非线性的回归分析也需要进一步的提升完善以适应当前大数据分析中更复杂的应用场景。

正是由于反问题的广泛适用性，人们对反问题的认识越来越深，造成的困扰反而越来越多。到底什么是反问题？

广义地，凡涉及到利用可观测量来估计或推演与可观测量有关的不可观测量的问题都可认为是一个反问题。这样的定义非常泛，包容性当然很强。

但是，众所周知，可观测量总是有限的、而且是受“噪音”污染的。利用这样的可观测量到底能在多大程度上或多高精度上估计出与之相关联的不可观测量？这个问题并没有严格的理论答案。正问题的建立，先提出一个假说、模式，然后不断地受到大量不同的实验、不同的实际观测（观察）的检验，出现偏差就修正正问题，直到正问题理论框架被确立。正问题应该排除了主观性！即便如此，同一个物理对象的尺度不同，考察的角度不同，正问题的描述方式也不同。某种程度上讲，正问题也存在不确定性。反问题看起来就是人为的、主观提出的问题。反问题往往基于一次观测结果就给出不可观测量的估计。尽管如此，更好地提出反问题，并努力求解反问题依然是人们坚持不懈做的事情！

为什么？因为这是解决实际问题的必由之路，绕不开！

可观测量的有限和受“噪音”污染是造成反问题的解存在不确定的基本原因。更重要的是反问题到底如何提法才能保证不可观测量的估计是有意义的，这才是反问题的根本逻辑。

反问题的根本逻辑是：**不可观测量与可观测量之间存在关系！**

否则，反问题就成了一个毫无意义的概念。

正是基于上述逻辑判断，本文试图从反问题的提法和反问题的解法两个角度，探讨反问题的根本逻辑，回答反问题到底想做什么、能做什么、以及如何能做到什么。

2 反问题的提法

前已述及，凡涉及到利用可观测量来估计或推演与可观测量有关的不可观测量的问题都可认为是一个广义的反问题；通过直接测量到的系统输出（数据），基于描述系统的状态方程，估计系统参数和系统激励的问题称为狭义的反问题。具体到勘探地震，三维探区形成一个三维地下介质体，震源激发能量下传，遇到反射界面或散射体或绕射体，地震波发生反射或散射或绕射，反传回地表，被检波器接收。震源、由弹性参数描述的三维地质体和接收到的波场构成一个系统。各种形式的波动方程作为描述该系统的状态方程。反问题是基于在地表（井中）观测的数据（波场）要么估计介质系统的弹性参数，要么估计震源函数，也可以二者同时估计。

推广一下，线性或非线性同相轴飘在满足各种概率分布的随机噪音中构成勘探地震信号或图像的基本模型。由观测的地震数据或已知的地震图像估计其中蕴含的感兴趣的信号或模式，尤其在低信噪比、或不规则采样、或有假频采样情况下，也是一个典型的、困难的反问题。此时，不可观测量（数据中的信号）与可观测量（实测数据）之间存在的关系也是要估计的，这是一个更为困难的盲反演问题。

即便是社会科学领域中，仅仅是基于（调查）采样数据，并假设数据满足一定的概率分布，用统计方法得到一些概率分布函数的某些统计量，这也是反问题。譬如自相关和功率谱就分别是随机信号的时域和频域二阶统计量。

反问题到底如何提法？

我还是认为 Bayes 估计的理论框架是包容性最好的，尽管很多时候看起来它过于抽象，不够实用。

反演的根本逻辑是：**不可观测量与可观测量之间存在关系。**定量地表达它们之间的关系应该有两条典型的途径：一是用二者之间的统计规律；二是用二者之间的函数关系。后者在科学和工程数据分析中更为常用。当然，函数关系有线性的和非线性的两种。

由于可观测量总是有限的（不完备的）、而且是受“噪音”污染的，不可能指望从可观测量中获取不可观测量的完备信息，因此，需要补充关于不可观测量的先验信息。这是反演的第二个根本逻辑。因此，我给反问题的基本定位是：**它总是一个不充分可观测信息下的不可观测量的估计问题。**

反问题之所以让人困扰，正是体现在各种不确定上。可观测量是不完备的、随机的；不可观测量与可观测量之间存在关系也是不确定的。但是，还一定要从这些不确定中得到一些有意义的认识。这也正是反问题的魅力所在！这才是解决实际问题应该有的逻辑思维！

放弃确定性，放弃对准确答案的追求，允许不确定性的存在，允许不精确答案的存在！这是提出反问题和解决反问题应该有的基本逻辑思维。

基于上述认识，假定不可观测量与可观测量之间存在线性或非线性的函数关系，在反问题的提法中，考虑两个方面的问题就是合理的。一是基于已知的函数关系对可观测量进行预测时的预测误差，预测误差小说明基于有限和不完美的可观测量对不可观测量的估计就是有意义的；二是需要增加（部分）先验信息降低仅仅考虑用不可观测量与可观测量之间的关系所带来的反演解的不确定性。

如何考虑这两个问题，与反问题的解法是密切相关的。我还是认为 Bayes 估计理论中用概率来表达对可观测量的预测结果的不确定和先验信息的不确定性是最抽象，也具备最高包容性的理论方法。对可观测量的预测结果的不确定性实质上应该包括：预测关系本身的不确定性；可观测量中的噪音带来的不确定性。我认为**还应该把不规则观测、不充分采样、有限孔径和有限频带观测统在对可观测量的预测结果的不确定性的考量中**，它们对最终反演解的不确定性（实质上就是反演解的精度）有很大的影响。我认为预测关系本身的不确定其中包括了系统参数的选择、系统状态方程的选择、系统参数的表达方法的选择，因为这三者对最终反演解的不确定也有重要的影响。当然，观测方式引起的不确定性和正问题引起的不确定性与反问题

的提法不是一个层次的问题。

在概率统计理论的指导下，用数学分析/泛函分析的方法提出反问题，从反问题求解的角度看更为合适。当然，从变分原理的角度提反问题也是非常符合逻辑的。

给出评价反演解不确定性的定量度量方式是给出反问题提法的根本。

假设 $\boldsymbol{d}^{obs}$ 是可观测量，$\boldsymbol{m}$ 是不可观测量。从 Bayes 估计理论看，评价反演解不确定性的定量度量应该用后验概率密度函数：

$$p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)\propto p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)p\left(\boldsymbol{m}\right) \tag{1}$$

上式也可以写成

$$p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)=k\cdot p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)p\left(\boldsymbol{m}\right) \tag{2}$$

其中，$p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)$ 代表已知不可观测量后对应出现可观测量的先验概率密度函数；$p\left(\boldsymbol{m}\right)$ 表示已知不可观测量的先验概率密度函数。事实上，不要把它们看成真实的概率密度函数，而是认为它们反映了某种不确定性。

从反问题求解的角度，希望后验概率密度函数是单峰的，最好是 Gauss 的，至少是广义 Gauss 的。只有这样，用后验概率密度函数的最大化得到的反问题的解才有明确的意义。

从数学分析/泛函分析（或变分原理）角度看，评价反演解不确定性的定量度量应该引入某种抽象的距离范数来定义某种泛函。物理上，这种泛函最小对应一个系统处于“能量”最低或熵最小的稳态。假定不可观测量和可观测量之间存在某种映射关系（预测关系），譬如：

$$\boldsymbol{d}^{obs}=g\left(\boldsymbol{m}\right)+\eta \tag{3}$$

而且存在有不可观测量的先验模型 $\boldsymbol{m}_{prior}$，则可以引入某种抽象的“距离”范数定义如下泛函：

$$\mathcal{J}\left(\boldsymbol{m}\right)=R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)+\lambda\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right) \tag{4}$$

其中，λ 是均衡系数，调整两个范数定义的度量结果对泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 的贡献。另外,

$$R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)=\left\|\boldsymbol{d}^{obs}-g\left(\boldsymbol{m}\right)\right\|_{W_d}^{q} \tag{5a}$$

W_d 是引入的关于可观测量的加权因子，q 代表某种范数的指标，上式一般地被称为数据偏差项。数据偏差项中，一般选为 $q=2$。实际上这是有问题的！

$$\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right)=\left\|\boldsymbol{m}-\boldsymbol{m}_{prior}\right\|_{W_m}^{p} \tag{5b}$$

W_m 是引入的关于不可观测量的加权因子，p 代表某种范数的指标，上式一般地被称为关于不可观测量的罚函数项。

同样地，希望泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 是单峰的或凸的；希望它是光滑可微分的；希望它是二次型函数。因为二次型泛函最容易获得的唯一最优解，光滑可微泛函可以用 Newton 类梯度迭代算法求解，单峰或凸的泛函存在唯一最优解。否则，泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 最优解的搜寻就要用到 Monte Carlo 方法。若泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 的峰值点异常多，目前没有什么有效的解法求解这样的反问题！

从数学分析/泛函分析（或变分原理）角度提出反问题的另一个典型形式为：

$$(\mathrm{P_J}):\quad \min \Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right),\quad Subject\quad to\, R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)\le\varepsilon \tag{6}$$

典型地，有如下 L_2 范数罚函数反演问题：

$$(\mathrm{P_2}):\quad \min\left\|\boldsymbol{m}-\boldsymbol{m}_{prior}\right\|_{W_m}^2,\quad Subject\quad to\, R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)\le\varepsilon \tag{7a}$$

L_1 范数罚函数反演问题：

$$(\mathrm{P_1}):\quad \min\left\|\boldsymbol{m}-\boldsymbol{m}_{prior}\right\|^1,\quad Subject\quad to\, R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)\le\varepsilon \tag{7b}$$

L_0 范数罚函数反演问题：

$$(\mathrm{P_0}):\quad \min\left\|\boldsymbol{m}-\boldsymbol{m}_{prior}\right\|^0,\quad Subject\quad to\, R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)\le\varepsilon \tag{7c}$$

和 L_∞ 范数罚函数反演问题：

$$(\mathrm{P_\infty}):\quad \min\left\|\boldsymbol{m}-\boldsymbol{m}_{prior}\right\|^\infty,\quad Subject\quad to\, R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)\le\varepsilon \tag{7d}$$

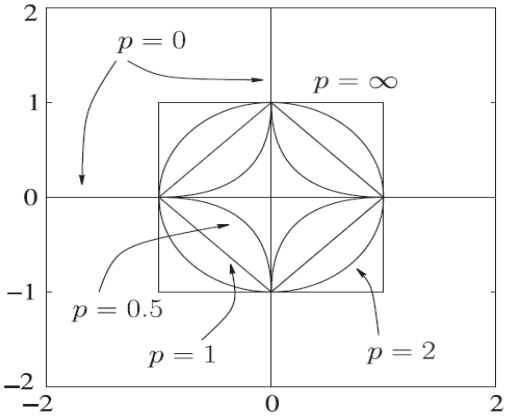


图 1 罚函数 $\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right)$ 的几何图样

罚函数 $\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right)=\left\|\boldsymbol{m}-\boldsymbol{m}_{prior}\right\|_{W_m}^q$ 定义成凸函数，此时 $p>1$。$p\le 1$ 对应的罚函数是非凸的，此时定义的罚函数是不能作为范数的。因此，L_0 范数的名称是被滥用的。p=1 是罚函数凸泛函的临界情况。图 1 展示了罚函数 $\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right)$ 的几何图样。值得提及的是图 1 展示的是没有施加加权函数时的情形。权函数 W_m 会对不同情形下 $\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right)$ 的几何图样会产生影响，当然这也就改变了反演问题的解。这正是施加权函数 W_m 要取得的效果。施加权函数 W_m 在反问题的提法中占有非常重要的地位，具体地提出反问题和求解反问题中必须予以充分的重视。

用类似一维情形展示罚函数 $\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right)$ 的几何形态对我们理解罚函数项的作用提供了另外的视角。图 2 是在类似一维情形下展示罚函数 $\Omega\left(\boldsymbol{m},\boldsymbol{m}_{prior}\right)$ 的几何形态。可以更直观地看出，随着 p 取值的不同，不同 $\boldsymbol{m}-\boldsymbol{m}_{prior}$ 对应的度量结果的变化。在泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 的优化求解中，由于不同的 $\boldsymbol{m}-\boldsymbol{m}_{prior}$ 的度量结果对

泛函 $\mathcal{J}(\boldsymbol{m})$ 的贡献不同，对它们的惩罚也是不同的。可以明显看出，L_2 范数罚下，大的分量被惩罚，小的分量得到保留，解就显得比较光滑，分辨率低。L_1 范数下对 $\boldsymbol{m}-\boldsymbol{m}_{prior}$ 中不同分量一视同仁，小的分量就被惩罚到零，仅有大的分量被保留，解的稀疏性就体现出来了，分辨率得到提高；L_0 范数下只有 $\boldsymbol{m}-\boldsymbol{m}_{prior}$ 中不等于零的分量对泛函 $\mathcal{J}(\boldsymbol{m})$ 有贡献，而且贡献相同，解的稀疏性是最好的。与正交匹配追踪这类稀疏提升算法结合，可以得到最稀疏的反演解。我认为不可以把稀疏的解与精确的解这两个概念混淆，二者并不等价！

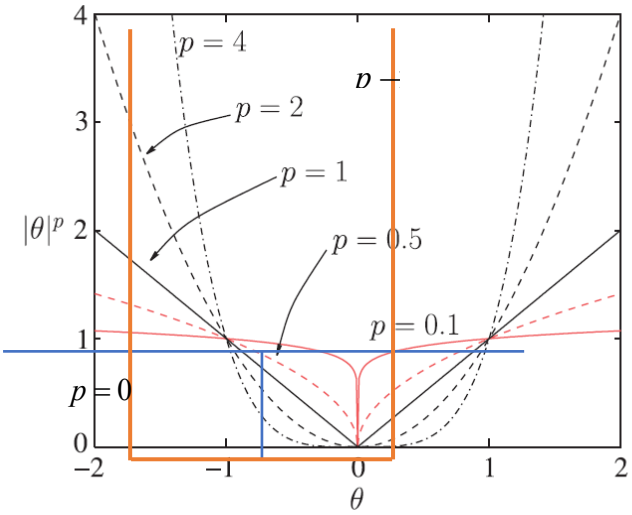


图 2 一维情形罚函数 $\Omega(\boldsymbol{m},\boldsymbol{m}_{prior})$ 的几何形态

关于 L_0 范数和 L_∞ 范数下罚函数的反问题求解，还需要做一些对比分析。向量 $\boldsymbol{m}$ 的 L_0 范数定义为：

$$\|\boldsymbol{m}\|^0=\lim_{p\to 0}\|\boldsymbol{m}\|_p^p=\lim_{p\to 0}\sum_{i=1}^{l}|m_i|^p=\sum_{i=1}^{l}\chi_{(0,\infty)}(|m_i|) \tag{8}$$

其中，函数 $\chi_{\mathcal{A}}(\tau):=\begin{cases}1 & \tau\in\mathcal{A}\\ 0 & \tau\notin\mathcal{A}\end{cases}$。可以清楚地看出，向量 $\boldsymbol{m}$ 的 L_0 范数就是向量 $\boldsymbol{m}$ 中分量不为零的个数。实际反演问题求解过程中，接近于零的分量当然也作为零分量进行处理。

向量 $\boldsymbol{m}$ 的 L_∞ 范数定义为：

$$\|\boldsymbol{m}\|^\infty=\lim_{p\to\infty}\left(\sum_{i=1}^{l}|m_i|^p\right)^{1/p}=\lim_{p\to\infty}\left(|m_i|^p\sum_{i=1}^{l}\left(\frac{|m_i|}{|m_i^{\max}|}\right)^p\right)^{1/p}=|m_i^{\max}| \tag{9}$$

向量 $\boldsymbol{m}$ 的 L_∞ 范数就等于绝对值最大的分量。

实际反演过程中，L_∞ 范数罚函数反演问题的求解本质是是选择若干大的分量形成反演解；L_0 范数罚函数反演问题的求解本质是选择不等于零的分量，而且也是尽可能选择大的 、不等于零的分量形成反演解。因此，L_0 范数罚函数反演问题的求解与 L_∞ 范数罚函数反演问题的求解本质上是一致的。因此，仅仅考虑 L_0 范数罚函数反演问题的求解就可以了。一类贪婪算法（譬如 OMP 算法）代表了目前 L_0 范数罚函数反演问题的求解思想与方法。具体地，参考“Sparse and Redundant Representations from Theory to Applications in Signal and Image Processing”一书。

仅仅考虑罚函数引入的约束，不能反映反问题的本质。反问题真正的困难还是由不可观测量和可观测量之间映射关系决定的。二者之间的强非线性关系引入了反问题的真正困难。从反问题的提法可以清楚地看出，数据偏差项定义了反演解的一个可行解集，罚函数项引入了正则化约束，从可行解集中选择一个符合约束条件的解作为反问题的最终结果。众所周知，非线性方程组 $F(\boldsymbol{m})=\boldsymbol{d}^{obs}-G\boldsymbol{m}=0$ 或 $\left|F(\boldsymbol{m})\right|=\left|\boldsymbol{d}^{obs}-G\boldsymbol{m}\right|\le\varepsilon$ 定义的可行解集可能是异常复杂的。因此，总假定，在某一点（一般是 $\boldsymbol{m}_{prior}$）附近，不可观测量与可观测量之间的关系是线性的，即 $\boldsymbol{d}^{obs}=G\boldsymbol{m}+\eta$。

此时，可行解集就由线性方程组中的不相关的超平面的“交点”界定。反问题的求解就是从可行解集中找到一个解向罚函数定义的凸集进行投影，投影结果就是反问题的解。关于可行解集，有必要进一步明确。对于欠定问题，可行解集落在 N 个 l 维超平面的“交点”上，它是一个 l-N 维的平面。它是一个仿射集（an affine set）。仿射集是一个线性子空间被一个常向量平移后得到的。

对于一个一般的欠定问题 $\boldsymbol{d}=G\boldsymbol{m},\boldsymbol{d}\in R^N,\boldsymbol{m}\in R^l,l>N$，总存在零空间 $\text{Null}(G)=\{\boldsymbol{z}\in R^l,G\boldsymbol{z}=\boldsymbol{0}\}$，若 $\boldsymbol{m}_0$ 是欠定问题 $\boldsymbol{d}=G\boldsymbol{m},\boldsymbol{d}\in R^N,\boldsymbol{m}\in R^l,l>N$ 的解，总存在：

$$G(\boldsymbol{m}-\boldsymbol{m}_0)=\boldsymbol{0} \tag{10}$$

则，$\boldsymbol{m}-\boldsymbol{m}_0\in\text{Null}(G)$。因此有，欠定问题 $\boldsymbol{d}=G\boldsymbol{m},\boldsymbol{d}\in R^N,\boldsymbol{m}\in R^l,l>N$ 的解空间为：

$$\mathbb{M}=\boldsymbol{m}_0+\text{Null}(G) \tag{11}$$

可见，欠定问题 $\boldsymbol{d}=G\boldsymbol{m},\boldsymbol{d}\in R^N,\boldsymbol{m}\in R^l,l>N$ 的解空间是一个仿射集。它不能给出欠定问题一个唯一解。为此，需要对反演解引入先验约束。上述提出的反问题系列 (P_j) 中就给出了不同的先验约束。

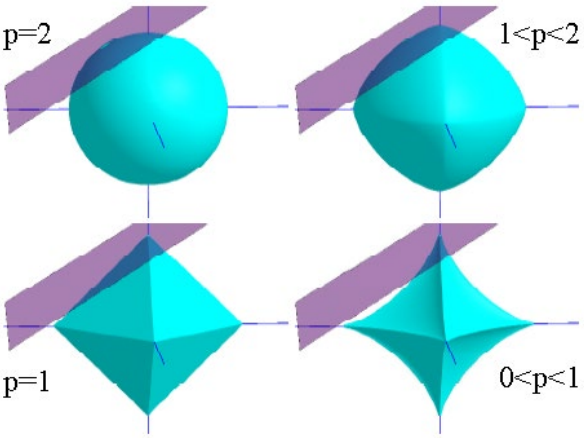


图 3 不可观测量与可观测量之间的关系是线性时，不同罚函数情形下反演解的几何说明

在不同罚函数情形下，基于线性正问题定义的反问题的解可用图 3 进行形象地展示。可以看出，L_2 范数罚函数下线性反问题是凸问题，可以归结为二次型泛函的求极值问题。此时反演解追求的是最小能量解，存在一套规范的解法，但反演解没有稀疏性。L_1 范数罚函数下线性反问题是凸问题与非凸问题的临界点。这是 L_1 范数下的反问题得到充分重视的理论基础。既有解的稀疏性，又有反问题的凸性，最近十几年在凸优化理论下发展出了有效的快速解法，**LASSO** 算法是典型的方法，存在若干 **LASSO** 算法的变种。L_0 范数下的反问题是非凸问题，稀疏性最好，但缺乏快速求解算法。一类贪婪求解算法是 NP-Hard 的，效率太低，不能用在大规模参数反演问题中。

关于数据偏差项的定义，首先存在的问题当然是不可观测量与可观测量之间的关系不能对可观测量做

出有效的预测或有效的解释。因此，不可观测量与可观测量之间的关系在反问题的建立和求解中占有至关重要的地位。在数据分析学科中，核心问题就是由数据驱动来建立这种关系。而在勘探地震学中，即便已经有了各种状态方程（波动方程）的存在，也不能说这个对实测数据的逼近问题已经得到了很好的解决。选择粘弹性波动方程、弹性波方程、变密度声波方程或是常密度声波方程进行实测叠前地震数据预测都是很难决定的，过于复杂的波动方程包含了更多要反演的弹性参数，使的反问题表现出更强的非线性性；过于简单的波动方程更难于预测复杂的实测波场，留下不可解释的残余误差，反演更难收敛。从这个角度，也可以部分地体现出反问题的主观性。

另外，数据偏差项定义成 $R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)=\left\|\boldsymbol{d}^{obs}-g\left(\boldsymbol{m}\right)\right\|_{W_d}^{q}$，即便考虑观测方式（$C$ 是引入的采样算子）后定义成 $R\left(\boldsymbol{d}^{obs},\boldsymbol{m}\right)=\left\|\boldsymbol{d}^{obs}-C\left(g\left(\boldsymbol{m}\right)\right)\right\|_{W_d}^{q}$，还是存在问题。众所周知，正问题是局部的和因果的，而反问题却没有局部性和因果性。这就会导致如此定义的数据偏差项不是同因之果的对减。用勘探地震学的语言讲就是，把不是来自同一个地下反射界面上的实测波场同相轴上的反射子波与理论合成波场同相轴上的反射子波进行对减，这是毫无意义的。但是这种物理上的合理要求很难在数据偏差项的定义中显示出来。在 FWI 反演中，笼统地归结为 Cycle-Skipping 现象。即便是相关泛函也回避不了这个问题。这是反问题中的一个很根本的问题，在当今的反演理论和实践中都被忽略了。我提出特征波反演，把来自同一反射界面的同相轴这种地震波场中的单震相提取出来，然后再定义抽象的距离度量。利用这样的方式部分地克服反问题没有因果性的问题。但是这样做会给数据预处理带来较大的压力。但无论如何，我坚信我提出的观点和做法是正确的。AWI 方法中引入的实测数据（波场）与合成数据（波场）的 Wiener 滤波最佳匹配和最佳输运泛函某种程度上都在尽可能实现同因之果之间的差异度量问题。初至波层析典型地实现了同因之果之间的差异度量，所以它的凸性要好很多。

最后，数据观测的不完整、不规则、欠采样问题没有体现在数据偏差项中，在反演过程弥补它们对反演精度的影响是必须要考虑的。初至波层析尽管凸性很好，但并未给出精度很高的反演结果，问题不出在反问题的提法和反问题的解法上，而是出在数据的不完备上。我认为在反问题的提法中尽可能考虑数据的不完备性对反演结果精度的影响是必须要做的事情。

3 反问题解法的基本逻辑

从前述反问题提法的基本逻辑中可以看出，尤其是从图 3 可以直观地看出，反问题的求解是从数据偏差项规定的可行解集中，寻找满足罚函数项给出的约束条件的、所谓的唯一解。

这是反问题求解的根本逻辑。但如何把上述根本逻辑转化为反问题的有效解法或体现在反问题解法的构造过程中？

事实上，除 Monte Carlo 算法外，目前所有的反问题解法都有一个基本假设，即泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 是单峰的或凸的，并且它的导数（梯度）是存在的，否则的话，至少次梯度是存在的。因此，梯度或次梯度导引类的迭代算法是求解反问题的根本方法。泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 是凸函数是目前梯度导引类优化算法的出发点。凸函数的定义：$\mathcal{J}\left(\boldsymbol{m}\right):C\subset H\to R$ 是凸泛函，假如

$$\mathcal{J}\left(\tau\boldsymbol{m}_1+\left(1-\tau\right)\boldsymbol{m}_2\right)\le\tau\mathcal{J}\left(\boldsymbol{m}_1\right)+\left(1-\tau\right)\mathcal{J}\left(\boldsymbol{m}_2\right);\boldsymbol{m}_1,\boldsymbol{m}_2\in C;0<\tau<1 \tag{12}$$

成立。

当泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 光滑可微，总存在如下的迭代公式：

$$\boldsymbol{m}^{\mathrm{k+1}}=\boldsymbol{m}^{\mathrm{k}}+\mu_k\left.\frac{\partial\mathcal{J}\left(\boldsymbol{m}\right)}{\partial\boldsymbol{m}}\right|_{\boldsymbol{m}=\boldsymbol{m}^{\mathrm{k}}} \tag{13}$$

是反问题求解的基本公式。其中，$\partial \mathcal{J}(\boldsymbol{m}) \big/ \partial \boldsymbol{m}$ 显然是一个 Jacobi 矩阵。

事实上，上式可以写成一个压缩映射形式：

$$\boldsymbol{m}^{\mathrm{k}+1}=G\left(\boldsymbol{m}^{\mathrm{k}}\right) \tag{14}$$

其中，G 是个压缩映射算子。

压缩映射算子的定义：若对 $\forall \boldsymbol{m}_i, \boldsymbol{m}_j \in D$ ，存在常数 $L<1$，成立 $\left\|G\left(\boldsymbol{m}_i\right)-G\left(\boldsymbol{m}_j\right)\right\| \leq L\left\|\boldsymbol{m}_i-\boldsymbol{m}_j\right\|$，则映射 $G(\boldsymbol{m})$ 在区域 D 上称为是压缩的。常数 $L(<1)$ 称为压缩因子。

压缩不动点定理：若映射 $G(\boldsymbol{m})$ 在区域 D 上满足：①$G(\boldsymbol{m}) \in D, \forall \boldsymbol{m} \in D$ ；②在区域 D 上 $G(\boldsymbol{m})$ 是压缩映射，压缩因子为 $L(<1)$，则对 $\forall \boldsymbol{m}_0 \in D$ ，迭代公式 $\boldsymbol{m}^{\mathrm{k}+1}=G\left(\boldsymbol{m}^{\mathrm{k}}\right)$ 产生的序列 $\left\{\boldsymbol{m}^k\right\}$ 收敛于映射 $G(\boldsymbol{m})$ 在区域 D 上的唯一不动点 $\boldsymbol{m}^*$，并且有误差估计式：

$$\left\|\boldsymbol{m}^k-\boldsymbol{m}^*\right\| \leq \frac{L^k}{1-L}\left\|\boldsymbol{m}^1-\boldsymbol{m}^0\right\|, k=0,1,2, \cdots \tag{15}$$

对泛函 $\mathcal{J}(\boldsymbol{m})$ 构造上述压缩映射迭代格式的基本方法就是 Newton 法。首先对泛函 $\mathcal{J}(\boldsymbol{m})$ 引入 Taylor 逼近式：

$$\mathcal{Q}(\boldsymbol{m})=\mathcal{J}(\boldsymbol{m}+\Delta \boldsymbol{m}) \approx \mathcal{J}(\boldsymbol{m})+\left(\operatorname{Grad} \mathcal{J}(\boldsymbol{m}), \Delta \boldsymbol{m}\right)+\frac{1}{2}\left(\operatorname{Hess} \mathcal{J}(\boldsymbol{m}) \Delta \boldsymbol{m}, \Delta \boldsymbol{m}\right) \tag{16}$$

（16）式表示用一个二次凸函数 $\mathcal{Q}(\boldsymbol{m})$ 在 $\boldsymbol{m}$ 点处逼近泛函 $\mathcal{J}(\boldsymbol{m})$。其中 $\operatorname{Grad} \mathcal{J}(\boldsymbol{m})$ 是泛函 $\mathcal{J}(\boldsymbol{m})$ 在点 $\boldsymbol{m}$ 处的一阶梯度；$\operatorname{Hess} \mathcal{J}(\boldsymbol{m})$ 是泛函 $\mathcal{J}(\boldsymbol{m})$ 在点 $\boldsymbol{m}$ 处的二阶梯度。对 $\mathcal{Q}(\boldsymbol{m})$ 求极值，即令 $\partial \mathcal{Q}(\boldsymbol{m}) \big/ \partial \Delta \boldsymbol{m}=0$ 得到：

$$\operatorname{Grad} \mathcal{J}(\boldsymbol{m})=-\operatorname{Hess} \mathcal{J}(\boldsymbol{m}) \Delta \boldsymbol{m} \tag{17}$$

把展开点处 $\boldsymbol{m}$ 记为 $\boldsymbol{m}^k$，则 $\Delta \boldsymbol{m}=\boldsymbol{m}^{k+1}-\boldsymbol{m}^k$，上式重写为标准的 Newton 迭代格式：

$$\boldsymbol{m}^{\mathrm{k}+1}=\boldsymbol{m}^{\mathrm{k}}-\left[\operatorname{Hess} \mathcal{J}(\boldsymbol{m})\right]^{-1} \operatorname{Grad} \mathcal{J}(\boldsymbol{m}) \tag{18}$$

对于线性正问题 $\boldsymbol{d}=G \boldsymbol{m}$ 对应的泛函 $\mathcal{J}(\boldsymbol{m})=\frac{1}{2}\left(\|G \boldsymbol{m}-\boldsymbol{d}\|^2+\lambda\|\boldsymbol{m}\|^2\right)$ 由变分法导出泛函 $\mathcal{J}(\boldsymbol{m})$ 的一阶梯度为（Vogel, 2002, P23）：

$$\operatorname{Grad} \mathcal{J}(\boldsymbol{m})=\left(G^T G+\lambda I\right) \boldsymbol{m}-G^T \boldsymbol{d} \tag{19}$$

不考虑正则化项（罚函数项）时，（19）式变为：

$$\operatorname{Grad} \mathcal{J}(\boldsymbol{m})=G^T(G \boldsymbol{m}-\boldsymbol{d}) \tag{20}$$

从（20）式可以看出，线性正问题情况下，梯度项的物理意义是残差数据的反传播。用勘探地震学的语言，也可以理解成入射波场（或下行波场）与出射波场（或上行波场）的互相关。

非线性正问题 $\boldsymbol{d}=G(\boldsymbol{m})$ 对应泛函 $\mathcal{J}(\boldsymbol{m})=\frac{1}{2}\left(\|G(\boldsymbol{m})-\boldsymbol{d}\|^2+\lambda\|\boldsymbol{m}\|^2\right)$ 的一阶梯度同样可以用变分法导出（Vogel, 2002, P88），

$$\mathrm{Grad}\mathcal{J}(\boldsymbol{m})=(W\boldsymbol{u},\boldsymbol{z}) \tag{21}$$

其中 $\boldsymbol{u}$ 为沿时间正向传播的波场，$\boldsymbol{z}$ 为逆时间正向传播的波场，W 是一种加权函数。

当泛函 $\mathcal{J}(\boldsymbol{m})$ 不存在一阶导数（一阶梯度）时，但只要它是凸泛函，就可以引入次梯度。（7b）式定义的反问题就可以按引入次梯度的方法进行求解。

次梯度的定义：矢量 $\boldsymbol{g}\in R^n$ 被称为凸函数 f 在点 $\boldsymbol{x}\in D$ 的次梯度，若

$$f(\boldsymbol{y})\geq f(\boldsymbol{x})+\boldsymbol{g}^T(\boldsymbol{y}-\boldsymbol{x}),\forall\boldsymbol{y}\in D \tag{22}$$

成立。

凸函数的次可微：凸函数 f 在 $\boldsymbol{x}\in D$ 上是次可微的，记成 $\partial f(\boldsymbol{x})$，并被记成如下集合：

$$\partial f(\boldsymbol{x})\triangleq\left\{\boldsymbol{g}\in R^n:f(\boldsymbol{y})\geq f(\boldsymbol{x})+\boldsymbol{g}^T(\boldsymbol{y}-\boldsymbol{x}),\forall\boldsymbol{y}\in D\right\} \tag{23}$$

如图 4 可见，不可微凸函数的次梯度不是唯一的，它构成一个集合！。在泛函 $\mathcal{J}(\boldsymbol{m})$ 光滑可微时，次微分就退化为一阶梯度：$\partial f(\boldsymbol{x})=\nabla f(\boldsymbol{x})$。

利用上述次梯度的概念，可以进行如下的 LASSO（The Least Absolute Shrinkage and Selection Operator）算法的示意性推导。

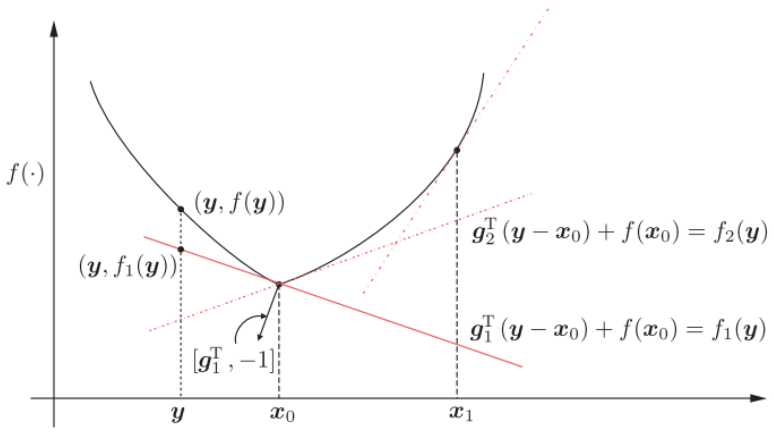


图 4 不可微凸函数的次梯度

若求解优化问题：$\hat{\boldsymbol{m}}=\underset{\boldsymbol{m}}{\mathrm{argmin}}\,\mathcal{J}(\boldsymbol{m})=\underset{\boldsymbol{m}}{\mathrm{argmin}}\left\{\frac{1}{2}\left(\|G\boldsymbol{m}-\boldsymbol{d}\|^2+\lambda\|\boldsymbol{m}\|^1\right)\right\}$，很显然泛函 $\mathcal{J}(\boldsymbol{m})$ 不是点点光滑可微的，但它是凸的，存在次梯度。次梯度等于零是极值存在的必要条件：

$$\boldsymbol{0}\in -G^T\boldsymbol{d}+G^TG\boldsymbol{m}+\frac{\lambda}{2}\partial\|\boldsymbol{m}\|^1 \tag{24}$$

假设矩阵 G 的列向量是正交的，$G^TG=\boldsymbol{I}$，上式可以写成分量形式：

$$\boldsymbol{0} \in -\hat{m}_{LS}^{i} + \hat{m}_{1}^{i} + \frac{\lambda}{2}\partial\left\|\hat{m}_{1}^{i}\right\|^{1}, \forall i \tag{25}$$

已知次梯度 $\partial\|\boldsymbol{m}\|^{1}$ 为：

$$\partial\|\boldsymbol{m}\|^{1} = \begin{cases} \{1\}, & \text{若} m>0 \quad \text{取正号} \\ \{-1\}, & \text{若} m<0 \quad \text{取负号} \\ [-1,1], & \text{若} m=0 \end{cases} \tag{26}$$

因此，分量形式的 LASSO 解为

$$\hat{m}_{1}^{i} = \begin{cases} \hat{m}_{LS}^{i} - \lambda/2 & \text{若} \hat{m}_{1}^{i} > s^{i} \\ 0 & -s^{i} \le \hat{m}_{1}^{i} \le s^{i} \\ \hat{m}_{LS}^{i} + \lambda/2 & \text{若} \hat{m}_{1}^{i} < -s^{i} \end{cases} \tag{27}$$

可见，**LASSO** 算法是在一个选择的阈值内，对最小二乘问题 $\mathcal{J}(\boldsymbol{m}) = \frac{1}{2}\|G\boldsymbol{m} - \boldsymbol{d}\|^{2}$ 的解进行软阈值处理。软阈值处理的几何如图 5 所示。

反问题求解的最基本的逻辑就是针对一个凸泛函，构建一个压缩映射算子以及对应的迭代格式。具体的构建方法就是基于 Newton 的思想，引入梯度下降的方向，同时兼顾对反演解施加的约束。

但是，大量的相关文献从不同的角度解释反演解的含义、构建求解反问题的方法，非常有必要建立一个基本统一的理论解释，把诸多方法的基本内涵统一在一个理论框架下。“Iterative Optimization in Inverse Problems”一书试图给出这样一个统一的框架。

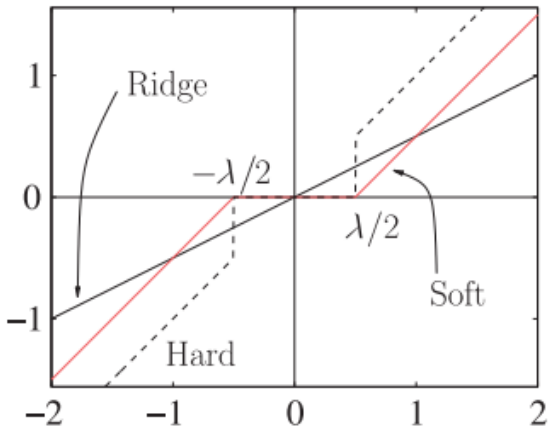


图 5 软阈值投影的几何

我们知道，线性正问题+Gauss 预测误差+L2 罚函数下的反问题求解，可以说存在一套统一的、完美的求解方案，从概率统计学角度（Gauss-Markov 定理）或是从分析学角度（SVD 分解）对反演解的意义做出了深入全面的分析。另外，基于二次型泛函表面形态的几何分析，深刻地剖析了最速下降算法、共轭梯度算法的本质。用于非二次型泛函的 Newton 方法本质上在局部点上还是利用了二次型泛函的求解思想和方法，通过逐步线性问题的二次型泛函的优化求解最终解决非二次型泛函的优化求解问题。

但是，对于凸的、不可微泛函的优化问题，甚至非凸的、不可微优化问题的求解以及解的含义、包括解的精度评价需要统一的、明晰的理论说明。我认为凸集投影思想和方法可以作为一个统一的框架来阐明凸的（甚至非凸的）、可微和不可微泛函的优化求解问题。

4 凸集投影方法的基本思想

投影思想和方法在约束反问题求解中占据举足轻重的地位。图 6 展示了投影的基本思想。图 6 的基本思想由下式表示：

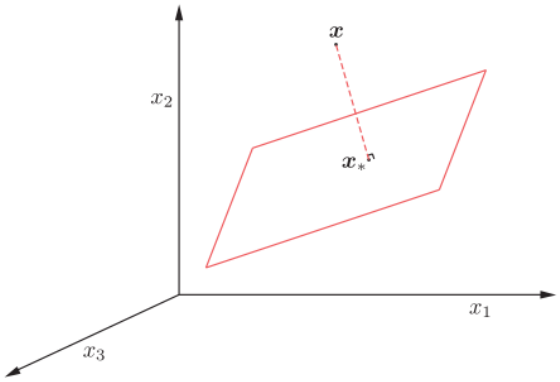


图 6 投影的基本思想

$$\boldsymbol{x}^{*}=\arg\min_{\boldsymbol{y}\in H}\left\{\left(x_1-y_1\right)^2+\left(x_2-y_2\right)^2+\left(x_3-y_3\right)^2\right\} \tag{28}$$

其中，$\boldsymbol{y}\in H$ 这个约束条件是很关键的。上述投影的思想在约束反问题求解中的应用由下述反问题：

$$\begin{cases}\boldsymbol{x}^{*}=\arg\min\limits_{\boldsymbol{y}}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|^{p}\\ s.t.\quad \boldsymbol{\theta}^{T}\boldsymbol{y}+\boldsymbol{\theta}_0=\boldsymbol{0}\end{cases} \tag{29}$$

的求解可以清楚地体现出来。所谓的约束优化就是从约束条件 $\boldsymbol{\theta}^{T}\boldsymbol{y}+\boldsymbol{\theta}_0=\boldsymbol{0}$ 规定的可行解中，寻找满足投影关系 $\boldsymbol{x}^{*}=\arg\min\limits_{\boldsymbol{y}}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|^{p}$ 的解。关键是如何基于上述投影的思想导出具有普遍意义的迭代公式，并用于各种提法的反问题的求解中。

首先给出投影的定义：假设 C 是 Hilbert 空间 H 中的一个非空的、闭的凸集，存在点 $\boldsymbol{x}\in H$，则存在一个用 $P_C\left(\boldsymbol{x}\right)$ 表示的唯一的点满足：

$$\left\|\boldsymbol{x}-P_C\left(\boldsymbol{x}\right)\right\|^{p}=\min_{\boldsymbol{y}\in C}\left\|\boldsymbol{x}-\boldsymbol{y}\right\| \tag{30}$$

上式表示点 $\boldsymbol{x}\in H$ 到凸集 C 上的投影。若 $\boldsymbol{x}\in C$，有 $P_C\left(\boldsymbol{x}\right)=\boldsymbol{x}$。

引理：若 $\boldsymbol{x}\notin C\subseteq H$，则投影 $P_C\left(\boldsymbol{x}\right)$ 落在凸集 C 的边界上。

下面简述投影算子 $P_C\left(\boldsymbol{x}\right)$ 的若干一般概念。投影算子 $P_C\left(\boldsymbol{x}\right)$ 是 R^n 到 R^n 的线性投影，且是等幂的，即 $P_C\cdot P_C=P_C$。若 P_C 是投影算子，$\boldsymbol{I}-P_C$ 也是投影算子，且有 $\mathrm{Ker}\left(P_C\right)=\mathrm{Ran}\left(\boldsymbol{I}-P_C\right)$。并有 $\boldsymbol{x}=P_C\left(\boldsymbol{x}\right)+\left(\boldsymbol{I}-P_C\right)\left(\boldsymbol{x}\right)$。若 $\boldsymbol{x}\in\mathrm{Ran}\left(P_C\right)$，则 $P_C\left(\boldsymbol{x}\right)=\boldsymbol{x}$；若 $\boldsymbol{x}\in\mathrm{Ker}\left(P_C\right)$，则 $P_C\left(\boldsymbol{x}\right)=\boldsymbol{0}$。若 $\boldsymbol{x}\in H$，且 H 是闭的 Hilbert 空间，则由 $\boldsymbol{x}=P_C\left(\boldsymbol{x}\right)+\left(\boldsymbol{I}-P_C\right)\left(\boldsymbol{x}\right)$ 式，该空间被分成下列直和的形式：$H=\mathrm{Ker}\left(P_C\right)\oplus\mathrm{Ran}\left(P_C\right)$。这种分解定义了一个唯一的投影。图 7 展示了投影算子 $P_C\left(\boldsymbol{x}\right)$ 对闭的 Hilbert 空间的直和分解。

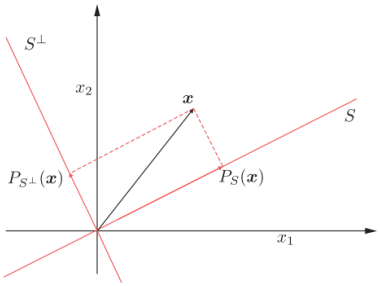


图 7 投影算子 $P_C(\boldsymbol{x})$ 对闭的 Hilbert 空间的直和分解

如果凸集投影算子 $P_C(\cdot)$ 是一个压缩的（至少是非膨胀的）算子，由前述压缩不动点原理，用它就可以构建反问题的迭代格式。把满足一定条件的可行解中的 $\boldsymbol{x}\in H$ 投影到凸集 C 上，得到最优化的结果。

据此可以看出，通过上述凸集投影算子，可以实现不动点理论下的迭代公式构造，与 Newton 法（或梯度导引类方法）平行地发展出一套反问题的数值求解算法。并且可以包容凸的（甚至非凸的）、可微和不可微泛函的优化求解。这显然是一套高于 Newton 梯度导引算法的思想体系。

凸集投影算子的性质，尤其是其非膨胀投影性质，在构造收敛的迭代格式中具有核心作用。

关于凸集投影算子的引理：假设 H 是 Hilbert 空间，$C\subseteq H$ 是闭凸集，且有 $\boldsymbol{x}\in H$。则投影 $P_C(\boldsymbol{x})$ 满足下列两条性质：

$$\mathrm{Real}\left\{\left(\boldsymbol{x}-P_C(\boldsymbol{x}),\boldsymbol{y}-P_C(\boldsymbol{x})\right)\right\}\le 0;\quad \forall \boldsymbol{y}\in C \tag{31}$$

和

$$\left\|P_C(\boldsymbol{x})-P_C(\boldsymbol{y})\right\|^2\le \mathrm{Real}\left\{\left(\boldsymbol{x}-\boldsymbol{y},P_C(\boldsymbol{x})-P_C(\boldsymbol{y})\right)\right\};\forall \boldsymbol{x},\boldsymbol{y}\in H \tag{32}$$

关于非膨胀映射算子的定义：映射 $T:H\mapsto H$ 是非膨胀的，若满足下式：

$$\forall \boldsymbol{x},\boldsymbol{y}\in H\quad \left\|T(\boldsymbol{x})-T(\boldsymbol{y})\right\|\le\left\|\boldsymbol{x}-\boldsymbol{y}\right\| \tag{33}$$

成立。

关于凸集投影算子是非膨胀的引理：若 C 是 Hilbert 空间 H 中的一个闭的凸集，则对应的投影算子 $P_C(\boldsymbol{x}):H\mapsto C$ 是非膨胀的，即

$$\forall \boldsymbol{x},\boldsymbol{y}\in H\quad \left\|P_C(\boldsymbol{x})-P_C(\boldsymbol{y})\right\|\le\left\|\boldsymbol{x}-\boldsymbol{y}\right\| \tag{34}$$

图 8 从几何上展示了投影算子 $P_C(\boldsymbol{x})$ 的非膨胀性质。正是由于投影算子 $P_C(\boldsymbol{x})$ 的非膨胀性质，类似前述的压缩映射不动点定理，可以通过投影算子 $P_C(\boldsymbol{x})$ 构造收敛的迭代公式。

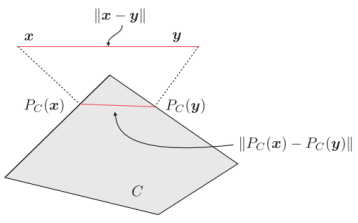


图 8 非膨胀的投影算子 $P_C(\boldsymbol{x})$

关于松弛的凸集投影算子的定义：若 C 是 Hilbert 空间 H 中的一个闭的凸集，如下定义的算子 $T_C : H \mapsto C$ 被称为松弛投影算子，

$$T_C \triangleq \boldsymbol{I} + \mu\left(P_C - \boldsymbol{I}\right), \quad \mu \in (0,2) \tag{35}$$

上式还可以表达成：

$$\forall \boldsymbol{x} \in H \quad T_C(\boldsymbol{x}) \triangleq \boldsymbol{x} + \mu\left(P_C(\boldsymbol{x}) - \boldsymbol{x}\right), \quad \mu \in (0,2) \tag{36}$$

很显然，μ=1时，$T_C(\boldsymbol{x}) = P_C(\boldsymbol{x})$。可以认为 $\mu > 1$ 是超松弛的；$\mu < 1$ 是欠（低）松弛的。

关于松弛的凸集投影算子是非膨胀的引理：松弛的投影算子也是非膨胀的，即有

$$\forall \boldsymbol{x}, \boldsymbol{y} \in H \quad \left\| T_C(\boldsymbol{x}) - T_C(\boldsymbol{y}) \right\| \le \left\| \boldsymbol{x} - \boldsymbol{y} \right\|, \quad \mu \in (0,2) \tag{37}$$

并有

$$\forall \boldsymbol{y} \in C, \quad \left\| T_C(\boldsymbol{x}) - \boldsymbol{y} \right\|^2 \le \left\| \boldsymbol{x} - \boldsymbol{y} \right\|^2 - \eta \left\| T_C(\boldsymbol{x}) - \boldsymbol{x} \right\|, \quad \eta = {(2-\mu)}/{\mu}, \mu \in (0,2) \tag{38}$$

图 9 给出了（38）式的几何解释。可以看出，松弛投影算子倾向于把点 $\boldsymbol{x} \in H$ 投影到更接近凸集 C 中（C 上）的点。正是基于这样的性质，松弛投影算子被广泛地用于构造约束反演问题中的迭代格式。

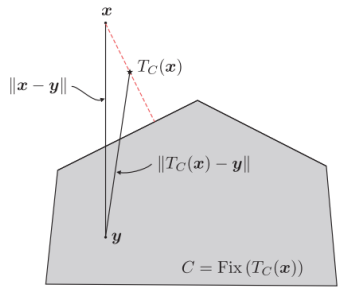


图 9 松弛投影算子的几何

临近算子方法是投影方法的进一步推广。临近算子定义被定义为：令 $f := R^m \mapsto R$ 是凸泛函，对应的 $\lambda-$临近算子为：$\mathrm{Prox}_{\lambda f}(\boldsymbol{x}) := R^m \mapsto R^m$，

$$\mathrm{Prox}_{\lambda f}(\boldsymbol{x}) := \arg\min_{\boldsymbol{v} \in R^m} \left\{ f(\boldsymbol{v}) + \frac{1}{2\lambda} \left\| \boldsymbol{x} - \boldsymbol{v} \right\|^2 \right\} \tag{39}$$

可以证明，临近算子也是非膨胀的，可以用来构造迭代算法。

对于一般的优化问题：

$$\boldsymbol{m}^* = \arg\min_{\boldsymbol{m}} \mathcal{J}(\boldsymbol{m}) = \arg\min_{\boldsymbol{m}} \left\{ f(\boldsymbol{m}) + g(\boldsymbol{m}) \right\} \tag{40}$$

其中，$f(\boldsymbol{m}) := R^m \mapsto R \cup \{+\infty\}$ 是连续不必点点可微的罚函数项，$g(\boldsymbol{m}) := R^m \mapsto R$ 为可微的数据偏差项。

由临近算子可以构造如下的迭代格式，

$$\boldsymbol{m}^{k+1} = \mathrm{Prox}_{\lambda_k f}\left(\boldsymbol{m}^k - \lambda_k \nabla g\left(\boldsymbol{m}^k \right) \right) \tag{41}$$

很显然，（39）式定义的临近算子可以由投影算子来构造。

5 凸集投影方法在参数估计反问题求解中的应用

正是由于（松弛的）投影算子具有非膨胀性质，使得可以用投影算子构建泛函 $\mathcal{J}(\boldsymbol{m})$ 的迭代求解公式。

若 $\mathcal{J}(\boldsymbol{m}):\mathrm{R}^m \mapsto (0,\infty)$ 是代价泛函，C 是 Hilbert 空间 H 中的一个闭的凸集，$C \subseteq R^m$。总存在如下的反问题：

$$\begin{cases} \boldsymbol{m}^* = \arg\min\limits_{\boldsymbol{m}} \mathcal{J}(\boldsymbol{m}) \\ s.t. \quad \boldsymbol{m} \in C \end{cases} \tag{42}$$

其中，代价泛函 $\mathcal{J}(\boldsymbol{m})$ 要求是凸的、连续的，但不必一定是点点可微的。

（42）式的经典求解方法是假设代价泛函 $\mathcal{J}(\boldsymbol{m})$ 是可微的，用 Newton 梯度导引类的算法求解。更一般的意义下，不对代价泛函 $\mathcal{J}(\boldsymbol{m})$ 施加点点可微性假设。在无约束情况下，即 $C \subseteq H$ 时，（42）式的求解有如下迭代公式：

$$\boldsymbol{m}^{\mathrm{k+1}} = \boldsymbol{m}^{\mathrm{k}} - \mu_k \partial \mathcal{J}(\boldsymbol{m}^k) \tag{43}$$

其中，$\partial \mathcal{J}(\boldsymbol{m})$ 是不可微代价泛函 $\mathcal{J}(\boldsymbol{m})$ 的次梯度。

当对 $\boldsymbol{m}$ 施加先验要求 $\boldsymbol{m} \in C$ 时，更一般的投影次梯度迭代格式为：

$$\boldsymbol{m}^{\mathrm{k+1}} = P_C\left(\boldsymbol{m}^{\mathrm{k}} - \mu_k \partial \mathcal{J}(\boldsymbol{m}^k)\right) \tag{44}$$

其中，$P_C(\cdot)$ 是向凸集 C 的投影算子。图 3 展示了引入的各种范数下的罚函数约束，不同范数下的罚函数约束构成了不同的凸集 C。

当代价泛函 $\mathcal{J}(\boldsymbol{m})$ 是可微的，就有如下的投影梯度方法：

$$\boldsymbol{m}^{\mathrm{k+1}} = P_C\left(\boldsymbol{m}^{\mathrm{k}} - \mu_k \nabla \mathcal{J}(\boldsymbol{m}^k)\right) \tag{45}$$

对应（42）式有典型的 Projected Landweber 算法，它对应的反问题为：

$$\begin{cases} \boldsymbol{m}^* = \arg\min\limits_{\boldsymbol{m}} \mathcal{J}(\boldsymbol{m}) = \arg\min\limits_{\boldsymbol{m}} \dfrac{1}{2}\|\boldsymbol{d} - G\boldsymbol{m}\|^2 \\ s.t. \quad \boldsymbol{m} \in C \end{cases} \tag{46}$$

其中，线性正算子 $G \in R^{n\times m}$，$\boldsymbol{d} \in R^n$。已知代价泛函 $\mathcal{J}(\boldsymbol{m})$ 的梯度为 $\nabla \mathcal{J}(\boldsymbol{m}) = G^T(G\boldsymbol{m} - \boldsymbol{d})$。

若代价泛函 $\mathcal{J}(\boldsymbol{m})$ 的梯度满足如下 $\gamma-$Lipschitz 连续：

$$\forall \boldsymbol{m}_1, \boldsymbol{m}_2 \in R^m, \quad \left\|\nabla \mathcal{J}(\boldsymbol{m}_1) - \nabla \mathcal{J}(\boldsymbol{m}_2)\right\| \le \gamma \|\boldsymbol{m}_1 - \boldsymbol{m}_2\|, \quad \gamma > 0 \tag{47}$$

若 $\mu \in \left(0, \dfrac{2}{\gamma}\right)$，从任意 $\boldsymbol{m}_0$ 点出发，投影梯度方法都收敛。更进一步地，若

$$\forall \boldsymbol{m}_1, \boldsymbol{m}_2 \in R^m, \quad \left\| G^T G\left(\boldsymbol{m}_1 - \boldsymbol{m}_2\right)\right\| \le \left\| G^T G\right\| \left\| \boldsymbol{m}_1 - \boldsymbol{m}_2\right\| \le \lambda_{\max} \left\| \boldsymbol{m}_1 - \boldsymbol{m}_2\right\| \tag{48}$$

其中，$\lambda_{\max}$ 是 Hessian 阵 $G^T G$ 的最大特征值。若 $\mu \in \left(0, \frac{2}{\lambda_{\max}}\right)$，从任意 $\boldsymbol{m}_0$ 点出发，投影梯度方法都收敛。（46）式的迭代求解格式为：$\boldsymbol{m}^{k+1} = P_C\left(\boldsymbol{m}^k - \mu_k \mathrm{G}^{\mathrm{T}}\left(\mathrm{G}\boldsymbol{m}^k - \boldsymbol{d}\right)\right)$。

当代价泛函 $\mathcal{J}\left(\boldsymbol{m}\right)$ 是不可微的，反问题为：

$$\begin{gathered} \boldsymbol{m}^* = \arg\min_{\boldsymbol{m}} \mathcal{J}\left(\boldsymbol{m}\right) = \arg\min_{\boldsymbol{m}} \left\{ \frac{1}{2} \left\| \boldsymbol{d} - G\boldsymbol{m}\right\|^2 + \lambda \left\| \boldsymbol{m}\right\|_{W_m}^1 \right\} \\ s.t. \quad \boldsymbol{m} \in C \end{gathered} \tag{49}$$

（49）式的迭代求解格式为：$\boldsymbol{m}^{k+1} = P_C\left(\boldsymbol{m}^k - \mu_k \partial \mathcal{J}\left(\boldsymbol{m}^k\right)\right)$。

投影算子 $P_C\left(\cdot\right)$ 被选为软门阈算子，此类迭代算法就被称为 **LASSO** 算法。**LASSO** 算法存在很多变种。关于 **LASSO** 算法更具体的内容，参考“Statistical Learning with Sparsity – The LASSO and Generalizations”一书和“Machine Learning - a Bayesian and Optimization Perspective”的第 8、9、10 章。

6 结论与讨论

反问题是各种涉及到由可观测量估计与之相关的不可观测量的统一称谓。由于可观测量总是有限的（不完备的）、而且是受“噪音”污染的，不可能指望从可观测量中获取不可观测量的完备信息，因此，需要补充关于不可观测量的先验信息。我给了反问题一个基本定位是：**它总是一个不充分可观测信息下的不可观测量的估计问题。**

反问题困扰人的根源正是体现在各种不确定上。可观测量是不完备的、随机的；不可观测量与可观测量之间存在的预测关系也是不确定的。但是，还一定要从这些不确定中得到一些有意义的结果和认识。事实上，这才是解决实际问题应该有的逻辑思维！放弃确定性，放弃对准确答案的追求，允许不确定性的存在，允许不精确答案的存在！这是提出反问题和解决反问题应该有的基本逻辑思维。

关于如何提出反问题，我认为 Bayes 估计理论中用概率来表达对可观测量的预测结果的不确定和先验信息的不确定性，进而确定待估计结果满足的后验概率密度，并且用该后验概率密度表达待估计结果（反演结果）的不确定性，是最抽象、也具备最高包容性的理论方法。对可观测量的预测结果的不确定性实质上应该包括：预测关系本身的不确定性；可观测量中的噪音带来的不确定性。我认为还应该把不规则观测、不充分采样、有限孔径和有限频带观测这些因素统在对可观测量的预测结果的不确定性的考量中，它们对最终反演解的不确定性（实质上就是反演解的精度）有很大的影响。我认为关于预测关系本身的不确定性，其中包括了系统参数的选择、系统状态方程的选择、系统参数的表达方法的选择，因为这三者对最终反演解的不确定也有重要的影响。上述观点和看法还要具体体现在反问题的实际求解过程中。

更为实用的、基于代价泛函提出反问题，我认为更多地需要考虑数据残差项的处理。同因之果的差异度量是我提出的最基本的概念，在数据残差的度量中必须只能对同源之果间的差异进行度量。差异度量的方式是重要的，但与同因之果的差异度量要求相比是第二位的。基本的原则是度量结果函数表现出更凸的性质。遗憾的是，度量结果函数的凸性更多地取决于实测数据与待反演参数之间的关系是否是线性的。差异度量方式是第二位的。差异度量方式对应的物理意义是重要的，希望它给出的结果对应系统的“能量”、或系统的信息熵等有唯一极小值的物理量。在数据偏差项中是否引入 L_∞ 或 $L_p, p \ne 2$ 之类的范数，我个人认为没有十分的必要，宁可通过去噪、特征数据（特征波场）及正问题的选择，使得数据残差满足 Gauss 分布。这样做的目的当然是希望能基于此给出高效的算法。对于数据偏差项，引入的差异度量函数要是可微的、物理意义明确的、凸性好的。我认为：不规则观测、不充分采样、有限孔径和有限频带观测这些因素应该反映在数据偏差项中。勘探地震中，罚函数的定义主要考虑把参数的结构信息能引入到反演解中，对参数的稀疏

性的要求最好与上述要求融为一体。最后，无论是数据偏差项或罚函数项的定义都应该是带加权函数的，加权函数的物理意义要非常明晰。

关于反问题的解法，我认为不动点理论是最高的理论框架，它简洁、抽象，但是它有足够的包容性。基于变分原理和 Newton 微积分思想，我们可以导出各种梯度导引类的迭代优化算法，但它仅能对凸的和点点可微的代价泛函给出适用的算法。Newton 算法只是不动点理论的一种具体实现。更有包容性的框架是投影算子方法。投影算子方法应该说是奠基在梯度导引类迭代算法基础上的。它的思想是把可行解集中的解用投影算子投影到约束凸集中，由于投影算子的非膨胀性，构造出的迭代格式可以逐步地把解集中的解收敛到一个不动点上。解集中的解的优化还是借用梯度导引类方法。临近算子方法，对罚函数进行临近分解（类似 Taylor 展开），然而临近算子的具体实现还是要利用投影算子。因此，基于投影算子方法，可以推进到对凸的、不必点点可微的代价泛函的优化求解。投影算子方法包容了梯度导引类算法，在更宽泛的函数类中给出了反问题迭代求解格式和各种不同的优化算法。

Monte Carlo 类算法是对付非线性性更强的一类反问题的求解方法，更多的讨论参考“Monte Carlo Statistical Methods”一书。

反问题的提法和解法目前大致就是这样。

参考文献

[1]. Bertaccini, D. and Durastante, F., 2018, Iterative Methods and Preconditioning for Large and Sparse Linear Systems with Applications: CRC Press

[2]. Boyd S. and Vandenberghe, L., 2009, Convex Optimization: Cambridge University Press

[3]. Byrne, C. L., 2014, Iterative Optimization in Inverse Problems: CRC Press

[4]. Cegielski, A., 2012, Iterative Methods foe Fixed Point Problems in Hilbert Spaces: Springer

[5]. Elad, M., 2010, Sparse and Redundant Representations from Theory to Applications in Signal and Image Processing: Springer

[6]. Hansen, P. C., 1998, Rank-Deficient and Discrete Ill-posed Problems: SIAM

[7]. Hastie, T., Tibshirani, R., and Wainwright, W., 2015, Statistical Learning with Sparsity – The LASSO and Generalizations: CRC press

[8]. Idier, J., 2008, Bayesian Approach to Inverse Problems (Edited): WILEY

[9]. Kaipio, J. and Somersalo, E., 2005, Statistical and Computational Inverse Problems: Springer

[10]. Kelley, C. T., 1995, Iterative Methods for Linear and Nonlinear Equations: SIAM

[11]. Kelley, C. T., 1999, Iterative Methods for Optimization: SIAM

[12]. Palomar D. P., and Eldar, Y. C., 2010, Convex Optimization in Signal Processing and Communications (Edited): Cambridge University Press

[13]. Robert, C. P., and Casella, G., 2004, Monte Carlo Statistical Methods: Springer

[14]. Saad, Y., 2003, Iterative Methods for Sparse Linear Systems (Second Edition): SIAM

[15]. Tarantola, A., 2005, Inverse problem theory and Methods for model parameter estimation: SIAM

[16]. Theodoridis, S., 2017, Machine Learning - a Bayesian and Optimization Perspective: Springer

[17]. Vogel, C. R., 2002, Computational Methods for inverse problems: SIAM

线性与非线性地震波反演成像理论问题分析

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】地震波反演成像归结为 Bayes 理论框架下的参数估计问题，把地表（包括井中）观测的波场（叠前数据）视为随机信号，已知一定的数学物理方程可以模拟（预测或解释）实测波场中的（部分）波现象，预测误差满足一定的概率分布（往往是 Gauss 分布），假设有来自井中或其他来源的关于要反演参数的先验信息，也把先验信息视为随机的（实质上这反映了先验信息的不确定性），基于 Bayes 公式获得要估计参数满足的后验概率密度函数，后验概率密度函数最大化（MAP 原则）对应的参数估计结果就认为是地震波反演问题的解。这是当前地震波反演成像的最基本的理论概念。非线性的全波形反演（FWI）以及对应的两种线性化形式---地震波层析成像和最小二乘逆时偏移成像（LS-RTM）都是上述理论概念在地震波反演成像中的具体化。理论框架是非常简洁、美妙和完善的。但是，勘探地震中的应用现实却与上述理论既定的假设条件相差很远，导致 FWI+LS-RTM 的实用化远不如预期。本文试图对线性与非线性地震波反演成像方法进行理论分析，说明当前 Bayes 估计理论的那些假设在地震波反演成像中是不满足的，重点分析线性反演理论问题，探讨非线性地震波反演成像的可行方案。

【关键词】Bayes 估计理论，线性反演，波动理论层析成像，最小二乘逆时偏移，非线性反演，全波形反演

1 引言

勘探地震学发展到现在，其中的核心问题逐渐聚焦，就是聚焦到把三维油气探区视为一个弹性介质系统，人工震源作为激励该系统的输入，地表观测的波场作为该系统的输出（理论上讲，应该在三维油气探区的所有边界上布设地震波接收器，才能采集到系统的完整输出。很显然，油气地震勘探中，仅仅在地表（或/和井中）观测了系统的部分输出），该实际物理系统被认为可以由人为建立的数学物理方程来描述，然后在上述正过程描述的基础上，基于 Bayes 估计理论对实际物理系统的弹性参数（纵波速度、横波速度和密度）进行估计。在弹性参数估计结果的基础上，结合岩石物理关系，进行岩石物性参数的估计，进而对储层含油气性进行评价，最后进入油气开发阶段。

这是当今油气地震勘探的基本逻辑。

据此，油气地震勘探的核心问题可以描述为：把地表（包括井中）观测的波场（叠前数据）视为随机信号，已知一定的数学物理方程可以模拟（预测或解释）实测波场中的（部分）波现象，预测误差满足一定的概率分布（往往是 Gauss 分布），假设有来自井中或其他来源的关于要反演参数的先验信息，也把该先验信息视为随机的（实质上这仅仅反映了先验信息的不确定性），基于 Bayes 公式获得要估计参数满足的后验概率密度函数，后验概率密度函数最大时对应的参数估计结果就被认为是地震波反演问题的解（MAP 原则）。

很显然，仅仅在地表（或/和井中）观测了实际物理系统的部分输出，而且观测是孔径受限的和频带受限的，采样是稀疏的（不满足 Shannon 采样定理），采样往往是不规则的，数据中包含很多非激励源引发的噪音，表明实际采集的地震数据非常不完备和不完美。地表高程（剧烈）变化和地表及地下介质（剧烈）空变使得要反演的弹性参数场包含了宽波数带成分，依据勘探尺度，它们从公分级别、米的级别、几十米的级别、几百米的级别到数千米和数十千米的级别之间变化，地震波在如此复杂变化的介质中传播，使得实测波场变化与不同尺度的介质弹性参数变化之间的关系是（高度）非线性的。另一方面，选择什么样的正问题参与地震波反演成像以及如何进行要反演参数的选择和描述，都严重影响所提出的反问题的非线性性。描述实际地震勘探系统的数学物理方程既不能过于简单，又不能过于复杂，简单的控制方程不能很好地预测实际观测波场；复杂的控制方程包含过多的未知参数。对要反演参数的选择和对它（它们）的描述方法很大程度上决定了介质系统参数反演问题求解的困难程度。从上述描述的系统的观点看地震波反演问题，有两个参数要反演：每炮的激励函数（震源子波）和介质系统的弹性参数。如果同时反演这两组参数，地震波反演成像的非线性性会显著增加，因此实际的地震波反演成像过程中，总假设每炮的激励函数（激发子波）是一致的和已知的。实际上，陆上地震勘探中每炮的震源子波显然是未知的，而且空变很明显。

从勘探地震学历史发展看，可以认为这就是一部描述实际介质系统正问题的方法不断演化的历史。最基础的正问题描述是用时距关系描述实际介质中的波传播过程：折射波时距关系和反射波时距关系（散射波时距关系和 CMP 道集中反射波时距关系）。勘探地震学发展过程中的大部分时间就是在基于这两种时距关系进行地震波成像。其间的单向波波动理论偏移、1D 波阻抗反演和基于 Zeoppritz 方程的 AVA 分析现在看就是当时条件下的权宜之计。上述这些技术才是勘探地震学独创的、有特色的技术。全面而深入地利用波动理论进行地震波成像才不过十几年，以 FWI 和 LS-RTM 为代表性技术，而且主要受大规模并行计算机技术的促进。

Bayes 框架下的、系统观点的弹性参数估计，由数学分析学家、统计学家、信号分析学家为主提出和完善，具有广泛的适用性。所有类似的基于观测数据推测其物理原因的专业学科都遵循同样的参数估计理论。有人统称这样的来自不同应用背景的问题为广义遥感（General Remote Sensing）问题。

前已述及，从勘探地震学历史发展看，地震波成像有其独特的观点和做法，它们来自于勘探地震所面对的特殊的地下介质情况，即以沉积形成的层状介质为主，中间夹杂着构造运动和火山活动等地质原因引起的介质局部畸变，波在这样的介质中传播具有自己的特点。勘探地震中的地表观测方式也深深地影响着地震波反演成像的方式与方法。因此，地震波反演成像不能完全照搬简洁而美妙的 Bayes 估计理论。总体而言，地震波成像从一开始就分成背景速度估计与叠前深度偏移成像佐证了上述观点。这不是在否认 Bayes 估计理论的普适性，而是强调具体问题（地震波反演成像）的特殊性。经典的 FWI 在实际数据处理中很难实用化的实践进一步说明了发展基于地震勘探介质特点、波场特征和实测数据采集方式的波动理论地震波反演成像的必要性。

基于 Bayes 估计理论进行实际介质系统的弹性参数估计才是 FWI 的真正内涵。

目前勘探地震中的 FWI 仅仅在利用透射波（长偏移距情况下的折射波和 Diving Wave）进行背景速度参数估计，得到较为精确的背景速度模型，提高后继的 RTM/LS-RTM 的成像精度。此时，还要求地震数据中包含低频（譬如 1Hz---6Hz）成分和长偏移距（譬如大于 8 千米，甚至达到 15 千米以上）而观测到的透射波。目前的 FWI 被称为透射波波动理论层析成像更名实相符。否者，既容易引起误解，也容易引起误用（有不少人认为 FWI 算法本身可以反演得到用于储层描述的波阻抗反演结果，在当前数据情况下，这是不可能的）。

本文试图在 Bayes 估计理论框架下，对 FWI 为代表的波动理论地震波反演成像的理论进行进一步分析，在线性反演理论分析的基础上，重点讨论非线性地震波反演的理论基础及可能的技术路线。然后指明根据目前的地震观测方式，FWI 在陆上数据中进行实际应用的技术方案。

2 油气地震勘探基本逻辑

油气勘探的目标是油气藏的识别、定位、描述与刻画。核心目标是获取油气藏的空间位置、几何形态、弹性参数（纵波速度、反射系数、波阻抗）、岩性参数（孔渗饱，但这需要与岩石物理结合）。高精度的地震波偏移成像和反演成像是关键的方法技术。FWI 是当今地震波反演成像指标性的理论方法与技术。图 1 展示了油气地震勘探中的关键问题、所用技术及基本逻辑关系。

从图 1 中可以清楚地看出，油气地震勘探问题可以抽象为利用各种不同来源、不同可信度的信息，对油气藏进行综合评价的问题。这应该是一个典型的人工智能和大数据分析问题。但是，这也是一个链条很长的、综合性的数据分析和决策问题，其中部分环节并不是大数据分析问题，譬如其中非常核心的地震波偏移成像与反演成像问题，也包括 1D 波阻抗反演、叠前 AVA 反演问题。无论如何，统一在 Bayes 框架下，用统计分析的观点，用信息融合和信息综合利用的观点，开展当今的油气勘探，尤其是油气藏的描述及含油气性判断是毋庸置疑的。因此，机器学习和人工智能理论与方法会在勘探地震整个链条中的各节点上首先得到广泛应用，但创建一个能有效应用的油气勘探机器（人）估计还需要很长的时间，但这应该成为不懈追求的长远目标。

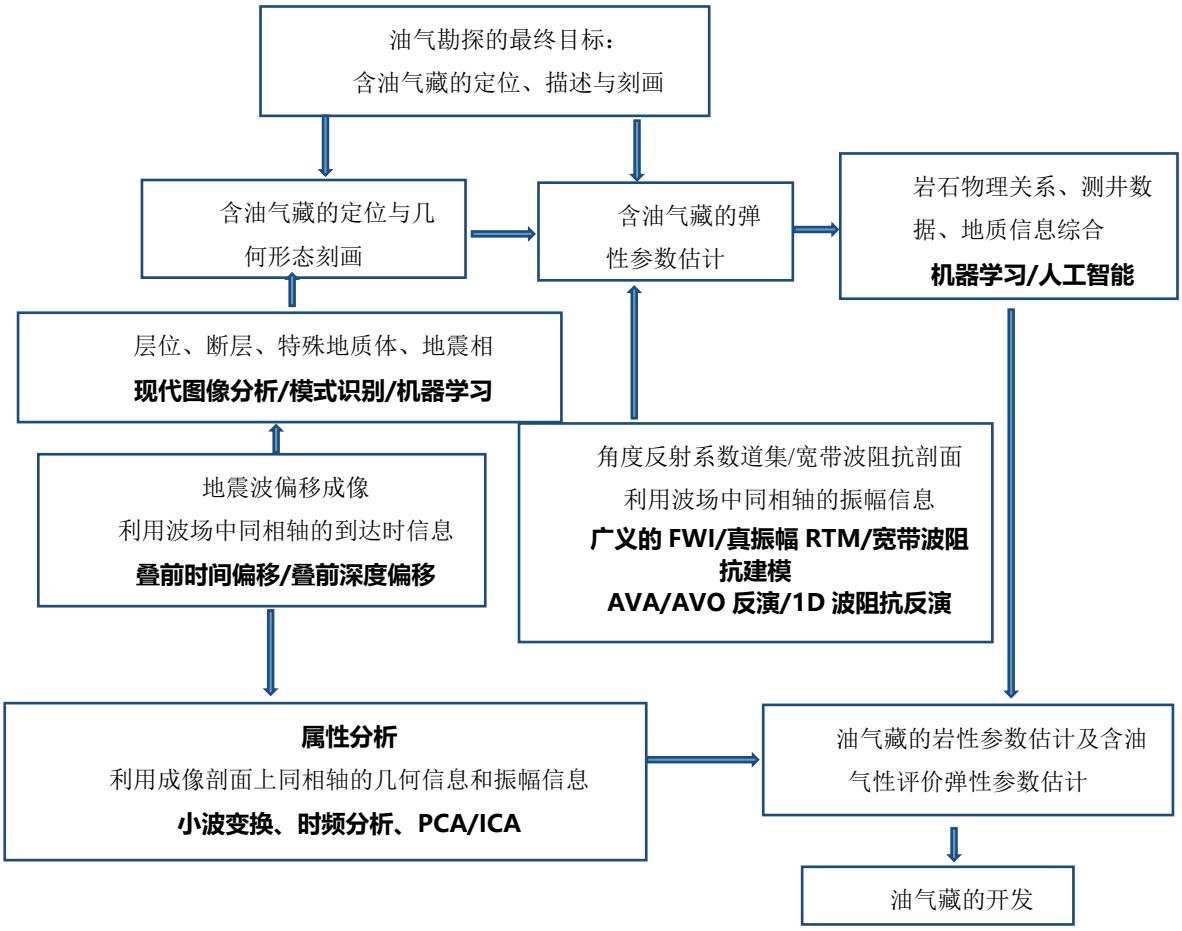


图 1 油气地震勘探中的关键问题、所用技术及基本逻辑关系

3 Bayes 反演理论框架

不夸张地讲，Bayes 估计理论应该是地震波成像的完美的理论框架。尽管勘探地震中地震波成像有独特的介质特点、波场特点和独特的观测方式，但是 Bayes 估计理论还是可以完美地包容地震波成像（也包括地震数据与处理）的理论与方法。事实上，当今的大数据分析、人工智能和机器学习的基本理论与方法也是建立在 Bayes 估计理论之上的。Bayes 估计理论是简洁优美的。它建立了参数估计的统一的思想框架。下面简述地震波成像中对应的 Bayes 估计理论。

Bayes 估计理论下的地震波参数反演涉及到如下五个线性空间：介质参数空间；估计参数空间；观测数据空间；状态变量空间；参数空间向状态变量的投影算子空间。观测数据是随机的，并认为依据随机的观测数据得到的参数估计结果也是随机的。假设由参数空间向状态变量的投影算子也具备不确定性。所有不确定性的描述都用相应的概率密度函数。

假设有模型空间M和数据空间D，在其定义的联合空间$M \times D$ 上，有同时反映观测数据$\boldsymbol{d}$以及模型参数$\boldsymbol{m}$信息的联合先验概率密度$\rho(\boldsymbol{d},\boldsymbol{m})$，同样有反映联系观测数据与模型参数物理规律的概率密度$\Theta(\boldsymbol{d},\boldsymbol{m})$。

由此可得出联合空间$M \times D$上的后验概率密度：

$$\sigma(\mathbf{d},\mathbf{m}) = k\frac{\rho(\mathbf{d},\mathbf{m})\Theta(\mathbf{d},\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})} \tag{1}$$

模型空间的后验概率密度 $\sigma_M(\boldsymbol{m})$可由边缘概率密度计算公式得到，并据此给出参数估计的期望与方差:

$$\sigma_{\mathbf{M}}(\mathbf{m})=\int_{\mathbf{D}}\sigma(\mathbf{d},\mathbf{m})d\mathbf{d} \tag{2}$$

$$\begin{cases}\hat{\mathbf{m}}=\int_{\mathbf{D}}\mathbf{m}\sigma_{\mathbf{M}}(\mathbf{m})d\mathbf{m}\\ \sigma^2=\int_{\mathbf{D}}(\mathbf{m}-\hat{\mathbf{m}})^2\sigma_{\mathbf{M}}(\mathbf{m})d\mathbf{m}\end{cases} \tag{3}$$

假设模型空间M和数据空间D均为线性空间，则它们各自的均匀概率密度为：

$$\mu_{\mathbf{M}}(\mathbf{m})=const.\ \ \mu_{\mathbf{D}}(\mathbf{d})=const. \tag{4}$$

假设测量数据与模型参数之间的物理规律（正问题）的不确定性可以忽略，则正问题的概率密度函数表示为：

$$\theta(\mathbf{d}\,|\,\mathbf{m})=\delta(\mathbf{d}-\mathbf{g}(\mathbf{m})) \tag{5}$$

假设待估计参数$\mathbf{m}$满足高斯分布的概率密度函数：

$$\rho_{\mathbf{M}}(\mathbf{m})=\text{const}.\exp(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{prior})^t\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})) \tag{6}$$

此处，引入关于估计参数的概率密度，只是用来表示估计结果的不确定性，并不是指介质参数是随机的。

再假设数据预测误差满足高斯分布的概率密度函数：

$$\rho_{\mathbf{D}}(\mathbf{d})=\text{const}.\exp(-\frac{1}{2}(\mathbf{d}-\mathbf{d}_{obs})^t\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{d}-\mathbf{d}_{obs})) \tag{7}$$

此处的数据预测误差满足高斯分布，既包括观测噪音引起的不确定性，也包括观测仪器引起的不确定性。实质上，在真实的参数反演过程中，也包括了正问题的不确定性引起的预测误差的不确定性。尤其是正问题引起的不确定性，可能在预测误差不确定性中占重要地位，这要在数据预处理中得到解决，通过数据预处理使得正问题与观测数据之间实现可预测性，预测误差满足既定的假设。

在上述各种假设下，待估计参数的后验概率密度 $\sigma_M(\mathbf{m})$为：

$$\sigma_{\mathbf{M}}(\mathbf{m})=K\rho_{\mathbf{M}}(\mathbf{m})\rho_{\mathbf{D}}(\mathbf{d}) \tag{8}$$

由于两种先验概率密度都满足高斯分布，后验概率密度函数形式为：

$$\sigma_{\mathbf{M}}(\mathbf{m})=const.\exp(-S(\mathbf{m})) \tag{9}$$

其中，

$$\begin{aligned}2S(\mathbf{m})&=(\mathbf{m}-\mathbf{m}_{prior})^t\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})+(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})^t\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})\\&=\|\mathbf{m}-\mathbf{m}_{prior}\|_{\mathbf{M}}^2+\|\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs}\|_{\mathbf{D}}^2\end{aligned} \tag{10}$$

后验概率密度函数最大化（MAP）对应的反演结果为：

$$\hat{\mathbf{m}} = arg \max_{\mathbf{m}} \sigma_{\mathbf{M}}(\mathbf{m}) = arg \min_{\mathbf{m}} S(\mathbf{m}) \tag{11}$$

MAP 反演准则在 Gauss 或广义 Gauss 假设下才是成立的，或至少是凸的反问题才是合理的。而事实上，地震波反演成像的真正困难在于其非线性，而且反问题的非线性取决于模型参数和数据之间的非线性关系，并不取决于估计方法。图 2 示意性地说明了地震波反演成像从线性到非线性性的逐渐演变过程。图 2a 示意性地展示了两个先验概率密度函数；图 2b 展示在线性正问题情况下，对应(11)式的地震波反演；图 2c 展示的是目前经典的全波形反演所对应的情形，此时初始条件给得非常好；图 2d 也是目前经典的全波形反演所对应的情形，此时初始条件给得比较差；图 2e 还是经典的全波形反演，但是先验概率密度函数不一定是满足广义 Gauss 分布的，初始值也不能给准，也可能是反问题提得非常差。此时，Monte Carlo 方法是应对这类反问题的求解方法。图 2f 对应强非线性反演情况，目前没有什么解法可以应对这种反问题的求解。

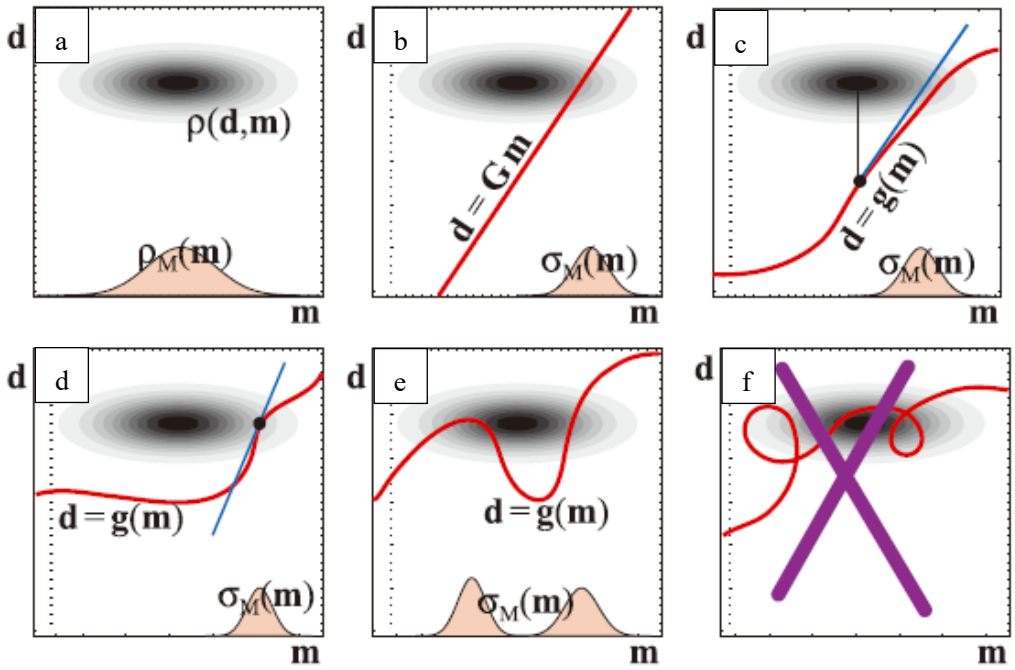


图 2 地震波反演成像从线性到非线性性的逐渐演变的示意图

事实上，关于反问题有如下一些描述是值得牢记的：正问题具有局部性和因果性；反问题没有局部性和因果性；正问题是良态的、适定的；反问题是病态的、不适定的；正问题是客观描述物理现实的；反问题是主观估计物理系统参数的。

关于反问题的主观性体现在：物理系统的描述方法（正问题）的选择是主观的；描述物理系统的参数的选择是主观的；描述物理系统参数的方式的选择是主观的；上述两种先验假设的选择是主观的；各种正则化思想和方法的选择也是主观的。

反问题的主观性可以任意吗？显然不行！最根本的限制是：参数估计结果要逼近真实物理系统的参数！否则反问题变得毫无意义了！另一个限制必然是：基于参数估计结果和正问题产生的对系统输出（实际观测数据）的预测误差要非常小！否则反问题就变得毫无意义了！

当前，勘探地震中存在两种极端的情形。正问题的线性性、后验概率密度函数的高斯性（单峰）与误差泛函是二次型的，三者是统一的，三位一体的，不可分离。决定因素是正问题（待估计参数与观测场之间的关系（或数据与潜变量之间的关系））的线性性！这是最简单的反问题情形。LS_RTM 和透射波层析成像是目前两种最典型的线性反演方法。后验概率密度函数多峰值，对应待估计参数与观测场之间的关系是强非线性的，理论上应该用 Monte Carlo 方法求解。这是非常复杂的反问题情形。当然，中间还有非线性的 FWI 情形。

4 全波形反演（FWI）存在的问题分析

4.1 非线性 FWI 存在的问题

勘探地震中，全波形反演（FWI）是一个典型的非线性反演问题。非线性主要体现在要反演的参数与实测数据之间的关系存在或强或弱的非线性性。我们把勘探地震中的数据抽象为：满足一定时距关系的同相轴飘在满足一定分布的随机噪音中。同相轴的到达时是背景速度决定的；同相轴上子波的振幅由反射界面两边波阻抗的扰动引起的。利用叠前地震数据反演背景速度和波阻抗扰动（尤其是反射系数）是最符合物理意义的参数选择。因此，基于同相轴到达时的旅行时层析成像和基于同相轴振幅的 LS_RTM 反射系数成像到目前为止还是勘探地震中地震波成像的两个最核心、也最有效的方法技术。

而全波形反演（FWI）是基于所选正问题，产生理论合成的炮集，预测实测炮集，并依据预测误差的某种范数最小，进行参数估计。选什么样的参数，选什么样的正问题会严重影响地震波反演的非线性性。这是经典的 FWI 很难用于实际勘探地震中的根本原因。

全波形反演（FWI）的理论假设与实际问题不适应性主要体现在：① 正问题，无论选得多么复杂，譬如双相介质地震波方程，也不可能完全模拟实测波场。目前，常规 FWI 中所常用的（常密度）声波方程理论上仅仅能模拟纯纵波。众所周知，实际观测波场是异常复杂的，尤其在复杂地表和复杂地下介质情形下。② 地震子波未知，无论从频带、相位和幅值上，正演模拟的地震反射（绕射）子波不可能与实测的地震反射（绕射）子波相同。理论上，应该同时反演地震子波和弹性参数。陆上勘探中，地震激发子波是逐炮空变的。地震子波反演的非线性比弹性参数反演的非线性性还要强！③ 利用波形差定义 L2 范数下的误差泛函是不合适的。数据逼近残差不（可能很好地）符合高斯分布！野外数据采集过程中，各种噪音不可控制！正问题（常密度声波方程）模拟不了实测数据中所有的波现象、子波未知等因素导致数据逼近残差不是由于要估计参数不准确引起的！有限差分数值频散以及各种其他的数值误差，甚至震源和检波器空间位置在正演模拟和实测数据中不匹配等等都会导致误差泛函的概率统计特征的变化。④ 高波数参数与实测波场之间存在较强的非线性关系。除背景速度和波阻抗扰动或反射系数之外的其他参数，譬如密度、各种弹性模量、甚至有人提出直接反演（岩石的孔隙度、渗透率和饱和度）等等参数，是在人为地提升地震波反演成像的非线性性，降低反演结果的精度或可靠性。也有人试图通过提升波场预测算子（正问题）的复杂度改善反演精度，我认为这都不是正确的途径。本质上，FWI 迭代反演只能对线性反问题或弱非线性的或凸的反问题能有收敛性保障。对于较强的非线性问题，初始解较差时，梯度类的优化反演方法不收敛！

事实上，连全波形反演（FWI）这个名称或提法都存在误导！名不正、言不顺！FWI 仅仅可以作为地震数据反演成像的名号，而不可视为一种单项技术！当前 FWI 成了一个大筐了，人人都在塞进自己的“私货”！FWI 成了一块橡皮泥，谁想怎么捏就怎么捏！

综合一下，即便仅仅估计背景速度，FWI 也是一个较强的非线性反问题。实际观测数据缺失低频或低频数据不可靠（频带缺失）；缺乏长偏移距的观测数据（有限观测孔径）；地震子波未知和空变；不满足 Gauss 分布的各种噪音（噪音分布）；正演算子无法预测所有观测到的波现象以及正问题数值求解的精度；仅仅在地表激发接收的地震观测方式，以及没有充分地利用地下介质的特征等等因素，加剧了 FWI 的非线性性、增强了不平稳性，提高了多解性。

地震波反演（或所有的系统参数估计）问题，都是用一个主观的理论框架，试图得到实际系统某种参数的估计。这样的一个主观的理论之所以还有效，核心在于可以用一个所建立的正问题，建立起系统参数和（部分）实测数据之间的（函数）关系，即用这样的正问题可以解释或预测（部分）实测数据。这永远是一个信息不足情况下的系统参数估计问题，数据的完整性和完美性是提高反演结果精度的最关键因素；合适的系统参数与（部分）实测数据之间的紧密联系（比较线性化的联系）是反演稳定性的根本；各种正则化方法是提高反演结果精度的必要措施。

4.2 非线性 FWI 具体公式

目前所谓的非线性 FWI，实际上是前述 Bayes 理论框架各种假设下，用后验概率最大化为原则进行参数估计的反演方法。实际上，就是使用如下的 Newton 梯度类算法:

迭代格式：

$$\tilde{\mathbf{m}}^{(k+1)}=\tilde{\mathbf{m}}^{(k)}-\left[\frac{\partial^2 S}{\partial\tilde{\mathbf{m}}^2}\right]^{-1}\nabla_{\tilde{\mathbf{m}}}S^{(k)} \tag{12}$$

求解数值优化问题：

$$\tilde{\mathbf{m}}=\underset{m}{arg\ min}\,S(\mathbf{m})=\underset{m}{arg\ min}\left\{\|\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs}\|_{\mathbf{D}}^2+\|\mathbf{m}\|_{\mathbf{W}}^p\right\} \tag{13}$$

其中，$\nabla_{\tilde{\mathbf{m}}}S^{(k)}$ 是负梯度，指示误差泛函下降最快的方向。$\left[\frac{\partial^2 S}{\partial\tilde{\mathbf{m}}^2}\right]^{-1}$ 是 Hessian 矩阵的逆，是梯度方向的预条件算子。

众所周知，非线性反问题(13)的 Newton 迭代方法是局部收敛算法，只有在初始点周围，误差泛函是二次型时，才会表现出二阶收敛性。但是，在复杂介质情形下，满足上述要求的初始值本身就是很难给出。

如果能给出满足上述要求的初始值，非线性的 FWI 完全可以转化成线性化问题（层析成像与 LS_RTM）来求解。首先对正问题进行线性化，建立参数扰动量与数据扰动量之间的关系：

$$\begin{aligned}&\mathbf{d}_{obs}=\mathbf{g}(\mathbf{m}_{\mathrm{B}}+\Delta\mathbf{m})=\mathbf{g}(\mathbf{m}_{\mathrm{B}})+\frac{\partial\mathbf{g}}{\partial\mathbf{m}}\Delta\mathbf{m}+\frac{1}{2}\frac{\partial^2\mathbf{g}}{\partial\mathbf{m}^2}\Delta\mathbf{m}^2+\cdots\\&\Delta\mathbf{d}_{obs}\approx\mathbf{g}(\mathbf{m}_{\mathrm{B}}+\Delta\mathbf{m})-\mathbf{g}(\mathbf{m}_{\mathrm{B}})=\frac{\partial\mathbf{g}}{\partial\mathbf{m}}\Delta\mathbf{m}\\&\Delta\mathbf{d}_{obs}=\mathrm{G}\Delta\mathbf{m}\end{aligned} \tag{14}$$

同样地，用如下迭代格式，

$$\mathbf{m}^{(k+1)}=\mathbf{m}^{(k)}-\mathbf{H}_a^{-1(k)}\nabla_{\mathbf{m}}S^{(k)}=\mathbf{m}^{(k)}-((\mathrm{G}^T\mathrm{G})^{(k)})^{-1}\mathrm{G}^T\Delta\mathbf{d}_{obs} \tag{15}$$

求解如下线性化的反问题：

$$\Delta\tilde{\mathbf{m}}=\underset{\mathbf{m}}{arg\ min}\,S(\Delta\mathbf{m})=\underset{\mathbf{m}}{arg\ min}\left\{\|\mathrm{G}\Delta\mathbf{m}-\Delta\mathbf{d}_{obs}\|_{\mathbf{D}}^2+\|\Delta\mathbf{m}\|_{\mathbf{W}}^p\right\} \tag{16}$$

当然，对线性化的反问题，也可以求解如下的法方程：

$$\mathrm{G}^T\mathrm{G}\Delta\mathbf{m}=\mathrm{G}^T\Delta\mathbf{d}_{obs} \tag{17}$$

事实上，非线性反演迭代格式(12)也可以认为每一个迭代步都在解一个线性反问题，即

$$\left[\frac{\partial^2 S}{\partial\tilde{\mathbf{m}}^2}\right]\Delta\tilde{\mathbf{m}}=\nabla_{\tilde{\mathbf{m}}}S^{(k)} \tag{18}$$

上述内容都是数值优化中的非常成熟的经典内容，但应用上述数值解法并不能很好地解决 FWI 方法在实际数据上的有效应用问题。FWI 不能很好地应用于实际数据中的根本原因，不在于优化算法，而在于经典的 FWI 中的各种假设不被满足，在于实际数据的不完整与不完美，在于经典反问题的提法。

4.3 解决较强非线性反问题的基本逻辑

首先应该明确，目前不存在求解强非线性问题有效的和广受认可的方案和具体算法。第二，假设有很好的初始值，在此基础上用梯度导引类算法得到不错的估计结果。这样假设下的做法，对复杂构造区的速度反演成像是不符合逻辑的，因为得到很好的初始值本身就是个很强的非线性问题。所以，经典的 FWI 也不能适用于地表复杂和构造复杂探区的成像。第三，地震数据的完整性的不断提高可以促进地震波成像质量的提升，但并不能彻底解决地震波反演成像问题。不能认为付出了很大的代价采集到了比较好的叠前地震数据，精确的地震波成像问题就一定得到较好的解决。第四，一味提升正问题的复杂程度（譬如：从声波方程提升到弹性波方程）并不是十分合理的选择。关键是正问题要与反演参数的选择及要利用的（特征）波场相匹配。第五，解决较强非线性的反问题，需要一套合理的方案，而不是一个看起来美妙的迭代算法。第六，用一组凸问题的组合来逼近较强非线性反问题的解是更合理的逻辑。第七，地震波反演成像问题永远是信息不足情况下的参数估计问题。信息的综合利用是较强非线性反问题求解的基本逻辑。第八，不存在（或永远也找不到）一个算法能彻底解决地震波成像问题。随着勘探复杂程度的逐渐提升，地震波成像方法技术一直要不断完善发展下去。

总之，我认为用一组凸的或良态的问题的组合来逼近原较强非线性问题的求解是根本的解决方案。另外，从信息综合利用的观点构建较强非线性反问题的求解路线是无法回避的理念。至于如何实现上述两点，就取决于各人对勘探地震反问题的理解了！

4.4 弱化 FWI 非线性性的具体措施

事实上，FWI 发展了这么多年，勘探地震学界已经探索了各种弱化 FWI 反演问题非线性性的各种方法与技术。具体地，包括：①多尺度反演策略。介质参数变化的尺度在反问题中是至关重要的物理因素，一般地，介质参数的大尺度变化引起对应数据的缓慢变化，二者之间的关系也更线性化，这就形成分频反演的做法。另外，迭代格式中的前几步，一般仅仅更新大尺度的参数变化，这就产生了把迭代次数作为正则化参数的反演方法。②利用波场的某种变换或波场中的部分波现象的策略。对波场取对数可以分离开振幅和相位，产生先利用相位，再利用振幅（或不利用振幅）的反演方法；对波场的 Laplace 变换可以通过调整变化系数，从初至波开始逐渐放入早至波、反射波的反演策略；对波场进行包络变换可以提出包络目标泛函，一定程度上克服缺乏低频时的周期跳（Cycle-skipping）现象。③对数据残差项定义不同差异测量泛函的策略。实际地震数据中包含的反（散）射子波的波形和振幅受到的影响因素太多，不是简单的正问题可以预测的。经典的波形对减的 L2 范数是最不可取的差异测量方法。首先要选择的理论合成数据与实测数据的差异测量泛函应该是互相关泛函；DSO 方法定义误差泛函一定是凸性很好的反问题；基于匹配滤波的误差泛函（AWI）定义方法一定程度上可以克服周期跳现象;最佳输运（OT）方法也是很好的差异测量的度量方法。④对代价函数中的罚函数项施加不同的正则化约束的策略。L2 范数定义罚函数得到的是低分辨率的反演结果，这是众所周知的。L1 和 L0 范数定义罚函数相对于假设参数分布具有稀疏性，在这样的先验假设下，进行参数估计。对于具有分层或分块特征的参数，施加 TV 正则化无疑是很合适的。

我（WPI 研究组）提出了特征波反演成像（CWI）方法，基本思想是：不是强解一个较强的非线性反演成像问题（FWI），而是解一组临近的凸问题，逼近较强反演成像问题的解。利用波现象选择+正问题选择+参数特征表达+误差泛函定义+正则化方法的基本理念，尽量把一个较强的非线性反问题提成若干个凸的反问题，分步进行求解，得到一个宽波数带的参数估计结果（具体而言就是宽带波阻抗估计结果）。特征波反演成像（CWI）包括的关键技术环节：①、基于信号稀疏特征表达的单震相特征提取，其中单震相特征：独立波形、包络、相位、到达时等，单震相波场的提取，可以彻底避开周期跳问题。②、地质信息与测井信息的空间表达+地质构造与沉积结构的智能提取。③、波动理论（波形或包络或相位或到达时）透射波层析成像。④、数据域波动理论（波形或包络或相位或到达时）反射波层析成像。⑤、最小二乘（保真）反演成像。⑥、非线性的多次波反演成像。

5 线性反问题

到目前为止，实质上，我们还只能有效地解决线性反问题，前已述及，非线性反问题在每一个迭代步中本质上也是按线性反问题求解的。因此，对线性反问题有深入认识是彻底理解非线性反演的基础。线性反问题的求解本质上可以归结为欠定、超定情形下线性方程组的求解问题。此处，我仅仅考察罚函数定义

为 L2 范数情形下的线性反问题。

5.1 欠定问题的最小二乘解

泛函分析理论框架下，假定存在一个物理过程的正问题描述，且该描述是线性的，即有 $\boldsymbol{b}=A\boldsymbol{x}$ 。反问题的误差泛函包含两部分：一是数据逼近的偏差函数 $\|\boldsymbol{b}-A\boldsymbol{x}\|$；一是关于解的罚函数 $\mathcal{J}(\boldsymbol{x})$。

一般情况下的反问题的提法为：

$$\underset{x}{\mathrm{Min}}\,\mathcal{J}(\boldsymbol{x}) \quad S.t. \quad \boldsymbol{b}=A\boldsymbol{x} \tag{19}$$

下面推导一下欠定的线性反问题的解。首先用 Larange 乘子法，上式可以写为

$$\mathfrak{L}(\boldsymbol{x})=\|\boldsymbol{x}\|_2^2+\lambda(A\boldsymbol{x}-\boldsymbol{b}) \tag{20}$$

上式的梯度为：

$$\frac{\partial\mathfrak{L}(\boldsymbol{x})}{\partial\boldsymbol{x}}=2\boldsymbol{x}+A^T\lambda \tag{21}$$

梯度等于零时，

$$\hat{\boldsymbol{x}}_{opt}=-\frac{1}{2}A^T\lambda \tag{22}$$

把最优解代入正问题线性方程中，$A\hat{\boldsymbol{x}}_{opt}=-\frac{1}{2}AA^T\lambda=\boldsymbol{b}$，据此可以得出

$$\lambda=-2\left(AA^T\right)^{-1}\boldsymbol{b} \tag{23}$$

最后，把 λ 代入(22)式定义的最佳解中，得到

$$\hat{\boldsymbol{x}}_{opt}=-\frac{1}{2}A^T\lambda=-\frac{1}{2}A^T\times-2\left(AA^T\right)^{-1}\boldsymbol{b}=A^T\left(AA^T\right)^{-1}\boldsymbol{b} \tag{24}$$

若引入不同的罚函数，(20)式可以写为：

$$\mathfrak{L}(\boldsymbol{x})=\|B\boldsymbol{x}\|_2^2+\lambda(A\boldsymbol{x}-\boldsymbol{b}) \tag{25}$$

(25)式的梯度项改为：

$$\frac{\partial\mathfrak{L}(\boldsymbol{x})}{\partial\boldsymbol{x}}=2B^TB\boldsymbol{x}+A^T\lambda \tag{26}$$

此时，最佳解为

$$\hat{\boldsymbol{x}}_{opt}=-\frac{1}{2}\left(B^TB\right)^{-1}A^T\lambda \tag{27}$$

代入正问题方程，

$$\boldsymbol{b}=A\hat{\boldsymbol{x}}_{opt}=-\frac{1}{2}A\left(B^TB\right)^{-1}A^T\lambda \tag{28}$$

据此，可得

$$\lambda=-2\left(A\left(B^TB\right)^{-1}A^T\right)^{-1}\boldsymbol{b} \tag{29}$$

代入(27)式，得到最佳解

$$\hat{\boldsymbol{x}}_{opt}=\left(B^TB\right)^{-1}A^T\left(A\left(B^TB\right)^{-1}A^T\right)^{-1}\boldsymbol{b} \tag{30}$$

若引入对数据的加权，(20)式又可以写为：

$$\mathfrak{L}\left(\boldsymbol{x}\right)=\left\|\boldsymbol{x}\right\|_2^2+\lambda^TW\left(A\boldsymbol{x}-\boldsymbol{b}\right) \tag{31}$$

对应的梯度为：

$$\frac{\partial\mathfrak{L}\left(\boldsymbol{x}\right)}{\partial\boldsymbol{x}}=2\boldsymbol{x}+\left(WA\right)^T\lambda \tag{32}$$

最优解为：

$$\hat{\boldsymbol{x}}_{opt}=-\frac{1}{2}\left(WA\right)^T\lambda \tag{33}$$

代入正问题方程，

$$\boldsymbol{b}=A\hat{\boldsymbol{x}}_{opt}=-\frac{1}{2}A\left(WA\right)^T\lambda \tag{34}$$

据此，可得

$$\lambda=-2\left(A\left(WA\right)^T\right)^{-1}\boldsymbol{b} \tag{35}$$

代入(33)式后的最优解为：

$$\hat{\boldsymbol{x}}_{opt}=\left(WA\right)^T\left(A\left(WA\right)^T\right)^{-1}\boldsymbol{b} \tag{36}$$

若同时考虑数据加权和罚函数项，(20)式还可以写为：

$$\mathfrak{L}\left(\boldsymbol{x}\right)=\left\|B\boldsymbol{x}\right\|_2^2+\lambda^TW\left(A\boldsymbol{x}-\boldsymbol{b}\right) \tag{37}$$

最优解为：

$$\hat{\boldsymbol{x}}_{opt}=\left(B^TB\right)^{-1}\left(WA\right)^T\left(A\left(B^TB\right)^{-1}\left(WA\right)^T\right)^{-1}\boldsymbol{b} \tag{38}$$

欠定问题的最小二乘问题解就是如此。但它们蕴含的勘探地震学意义还要进一步阐述，才能做好实际数据的反演。

5.2 超定问题与欠定问题伪逆的对比

超定情况下，线性问题的最小二乘问题提为：

$$\mathfrak{L}\left(\boldsymbol{x}\right)=\left(A\boldsymbol{x}-\boldsymbol{b}\right)^2+\lambda\left\|\boldsymbol{x}\right\|_2^2 \tag{39}$$

最优解为：

$$\hat{\boldsymbol{x}}_{opt}=\left(A^TA+\lambda I\right)^{-1}A^T\boldsymbol{b} \tag{40}$$

若同时考虑数据加权和罚函数项，(39)式还可以写为：

$$\mathfrak{L}\left(\boldsymbol{x}\right)=W\left(A\boldsymbol{x}-\boldsymbol{b}\right)^2+\lambda\left\|B\boldsymbol{x}\right\|_2^2 \tag{41}$$

对应的最优解为：

$$\hat{\boldsymbol{x}}_{opt}=\left(A^TW^TWA+\lambda B^TB\right)^{-1}\left(WA\right)^T\boldsymbol{b} \tag{42}$$

(40)式的勘探地震学意义是很清楚的。$A^T\boldsymbol{b}$ 实现数据空间向模型空间的对偶空间的投影；Hessian 矩阵把对偶空间的投影结果反投影到模型参数的原空间。Hessian 矩阵本身是一个 Gauss 型的光滑算子；Hessian 矩阵的逆算子是一个 Laplace 型的粗糙算子。Hessian 算子（矩阵）及其逆是原模型参数空间与其对偶空间之间的映射。Hessian 矩阵（及其逆）的勘探地震学意义是很清楚的。

对比一下欠定情形和超定情形下线性反问题的解：

$$\begin{cases}\hat{\boldsymbol{x}}_{opt}=\left(A^TA\right)^{-1}A^T\boldsymbol{b} & \text{超定情形}\\ \hat{\boldsymbol{x}}_{opt}=A^T\left(AA^T\right)^{-1}\boldsymbol{b} & \text{欠定情形}\end{cases} \tag{43}$$

显然，在欠定和超定情形下，AA^T 和 A^TA 是列满秩时，二种情形下的解是等价的。更具体地，上述两种情形的勘探地震学含义对比，参考本年度论文集中吴成梁在“数据域叠前偏移成像”一文中对此问题的分析。

5.3 Bayes 框架下线性反问题 MAP 解的统计含义

线性反问题的解，从统计意义上给出解释是非常有意义的。我认为线性反演问题的解的含义在统计意义下给出解释，更本质地说明了线性反演解的意义。

Bayes 框架下，正问题为线性情形时，MAP 对应的解为：

$$\begin{aligned}\mathbf{m} &= \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1}\left(G^t C_D^{-1}\mathbf{d}^{obs} + C_M^{-1}\mathbf{m}_{prior}\right)\\ &= \mathbf{m}_{prior} + \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1}\left(\mathbf{d}^{obs} - G\mathbf{m}_{prior}\right) \quad \text{超定情形}\\ &= \mathbf{m}_{prior} + C_M G^t\left(C_D^{-1} + G C_M G^t\right)^{-1}\left(\mathbf{d}^{obs} - G\mathbf{m}_{prior}\right) \quad \text{欠定情形}\end{aligned} \tag{44}$$

方差为：

$$\widetilde{C}_M = \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} = C_M - C_M G^t\left(C_D + G C_M G^t\right)^{-1} G C_M \tag{45}$$

下面给出(16)式定义的线性反问题 MAP 原则下给出的反演解的具体推导。已知如下的代价泛函：

$$S(\mathbf{m}) = \frac{1}{2}\left\{\left(G\mathbf{m} - \mathbf{d}^{obs}\right)^t C_D^{-1}\left(G\mathbf{m} - \mathbf{d}^{obs}\right) + \left(\mathbf{m} - \mathbf{m}_{prior}\right)^t C_M^{-1}\left(\mathbf{m} - \mathbf{m}_{prior}\right)\right\} \tag{46}$$

对应的梯度为：

$$\frac{\partial S}{\partial \mathbf{m}} = \frac{1}{2} G^t C_D^{-1}\left(G\mathbf{m} - \mathbf{d}^{obs}\right) + \frac{1}{2} C_M^{-1}\left(\mathbf{m} - \mathbf{m}_{prior}\right) \tag{47}$$

令梯度等于零：

$$\frac{\partial S}{\partial \mathbf{m}} = 0 \tag{48}$$

有，

$$G^t C_D^{-1} G\mathbf{m} - G^t C_D^{-1}\mathbf{d}^{obs} + C_M^{-1}\mathbf{m} - C_M^{-1}\mathbf{m}_{prior} = 0 \tag{49}$$

上式重写为：

$$\left(G^t C_D^{-1} G + C_M^{-1}\right)\mathbf{m} = G^t C_D^{-1}\mathbf{d}^{obs} + C_M^{-1}\mathbf{m}_{prior} \tag{50}$$

得到，

$$\tilde{\mathbf{m}} = \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1}\left(G^t C_D^{-1}\mathbf{d}^{obs} + C_M^{-1}\mathbf{m}_{prior}\right) \tag{51}$$

对上式进行如下的变换，得到

$$\begin{aligned}\tilde{\mathbf{m}} &= \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1}\left(G^t C_D^{-1}\mathbf{d}^{obs} + \left(G^t C_D^{-1} G + C_M^{-1} - G^t C_D^{-1} G\right)\mathbf{m}_{prior}\right)\\ &= \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1}\mathbf{d}^{obs} + \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1}\left(G^t C_D^{-1} G + C_M^{-1}\right)\mathbf{m}_{prior}\\ &\quad - \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1} G\mathbf{m}_{prior}\\ &= \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1}\mathbf{d}^{obs} + \mathbf{m}_{prior} - \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1} G\mathbf{m}_{prior}\\ &= \mathbf{m}_{prior} + \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1}\left(\mathbf{d}^{obs} - G\mathbf{m}_{prior}\right)\end{aligned} \tag{52}$$

这就证明了超定情形下，MAP 原则下 L2 范数线性反问题的解由(44)式给出。

把(52)式写成：

$$\begin{aligned}\tilde{\mathbf{m}} &= \mathbf{m}_{prior} + \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1} \left(\mathbf{d}^{obs} - G\mathbf{m}_{prior}\right) \\ &= \mathbf{m}_{prior} + \mathfrak{F} \cdot \left(\mathbf{d}^{obs} - G\mathbf{m}_{prior}\right)\end{aligned} \tag{53}$$

其中，$\mathfrak{F} = \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1}$ 是伴随算子。上式解释了 MAP 原则下 L2 范数线性反问题的解的统计意义。它是对先验参数值的修改，它是后验概率密度函数的均值（期望）。

下面证明超定和欠定情形下，MAP 反演解的统一。但是，应该注意二者的等价是有条件的。要利用如下的恒等式：

$$\begin{cases} G^t + G^t C_D^{-1} G C_M G^t = G^t C_D^{-1} \left(C_D + G C_M G^t\right) \\ G^t + G^t C_D^{-1} G C_M G^t = \left(C_M^{-1} + G^t C_D^{-1} G\right) C_M G^t \end{cases} \tag{54}$$

由(54)式显然有，

$$G^t C_D^{-1} \left(C_D + G C_M G^t\right) = \left(C_M^{-1} + G^t C_D^{-1} G\right) C_M G^t \tag{55}$$

进一步地，又有，

$$\left(C_M^{-1} + G^t C_D^{-1} G\right)^{-1} G^t C_D^{-1} = C_M G^t \left(C_D + G C_M G^t\right)^{-1} \tag{56}$$

因此，有欠定情形下 MAP 反演解：

$$\tilde{\mathbf{m}} = \mathbf{m}_{prior} + C_M G^t \left(C_D^{-1} + G C_M G^t\right)^{-1} \left(\mathbf{d}^{obs} - G\mathbf{m}_{prior}\right) \tag{57}$$

超定情形下 MAP 反演解：

$$\tilde{\mathbf{m}} = \mathbf{m}_{prior} + \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1} \left(\mathbf{d}^{obs} - G\mathbf{m}_{prior}\right) \tag{58}$$

可见在矩阵 $C_D + G C_M G^t$ 和矩阵 $C_M^{-1} + G^t C_D^{-1} G$ 非奇异的情形下，超定问题和欠定问题的 MAP 解是等价的。

下面再给出恒等式 $\left(BD^{-1}C + A\right)^{-1} = A^{-1} - A^{-1}B\left(D + CA^{-1}B\right)^{-1} CA^{-1}$ 的证明：

$$\begin{aligned} & C_M - C_M G^t \left(C_D + G C_M G^t\right)^{-1} G_M C_M \\ &= C_M - \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} G^t C_D^{-1} G C_M \\ &= \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} \left(\left(G^t C_D^{-1} G + C_M^{-1}\right) C_M - G^t C_D^{-1} G C_M\right) \\ &= \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} \left(G^t C_D^{-1} G C_M + C_M^{-1} C_M - G^t C_D^{-1} G C_M\right) \\ &= \left(G^t C_D^{-1} G + C_M^{-1}\right)^{-1} \end{aligned} \tag{59}$$

即有：

$$\widetilde{C}_M=\left(G^tC_D^{-1}G+C_M^{-1}\right)^{-1}=C_M-C_MG^t\left(C_D+GC_MG^t\right)^{-1}GC_M \tag{60}$$

这就证明了反演解的后验概率密度函数的方差由(45)式给出。后验概率密度函数的方差是对参数的先验概率密度函数方差的修正，相对于参数的先验概率密度函数方差，后验概率密度函数的方差应该是减小的。预示着估计结果的精度高于先验参数的精度。这正是进行参数估计反问题的目的。

值得注意的是，在矩阵 $C_D+GC_MG^t$ 和矩阵 $C_M^{-1}+G^tC_D^{-1}G$ 非奇异的情形下，存在如下两个恒等式：

恒等式 1：

$$\left(G^tC_DG+C_M^{-1}\right)G^tC_D^{-1}=C_MG^t\left(C_D+GC_MG^t\right)^{-1} \tag{61}$$

恒等式 2：

$$\left(G^tC_D^{-1}G+C_M^{-1}\right)^{-1}=C_M-C_MG^t\left(C_D+GC_MG^t\right)^{-1}GC_M \tag{62}$$

这是两个在信号和图像分析中也非常有用的恒等式。

6 地震波非线性反问题的提法及求解

6.1 关于勘探地震领域非线性反问题的基本认识

目前，线性高斯假设下的 L2 范数形式的误差泛函反问题，其内涵已经十分明晰。前述 Bayes 反演理论已经详述了这个问题，现代信号分析理论的参数估计部分更深入地分析过这类问题解的特征。可以说这类反问题的理论基础已经非常完善了，这是所有反问题中最常用、最基础、也最简单的反问题。很显然，线性、高斯、平稳假设下的反问题由于其过于严格的假设，导致与实际情况不符，因此反问题的研究必然向非线性、非高斯和非平稳情形下发展。

在勘探地震中，非线性反问题的本意是待估计参数与实测数据之间的关系是非线性的；高斯性指的是对参数的分布施加了 Gauss 分布的假设，同时也假设理论合成数据与实测数据的逼近误差满足高斯分布，因此后验概率密度也满足高斯分布，这也就有了上述 MAP 原则下反演解的统计意义；平稳性是指系统输出的信号是时间平稳的，事实上也希望是空间平稳的，强吸收衰减介质情形下的系统输出就破坏了系统输出的平稳性假设。

FWI 反演成像的求解基本都是使用基于梯度导引类算法，这是基于 Newton 微积分的思想而提出的。Newton 微积分思想的核心是对函数（包括泛函）进行局部分析。Newton 微积分是一种强大的、应用广泛的分析工具。但是，在代价泛函的优化求解中，过于依赖这种工具，就会限于局部化、线性化的方法体系中。而这样的方法无助于求解较强非线性反演问题。我认为较强非线性反问题的有效求解，关键根本不在优化算法上，而是反问题的提法上。我认为解决非线性问题的核心理念是：将问题描述为或转化为一系列的凸问题，而不是用所谓的“高级算法”硬解一个较强的非线性反问题。事实上，反问题非线性性强的根源在于正问题：即待估计参数与（部分）观测场之间的关系。因此，完全有必要把较强非线性反问题的求解从专注于数值方法上转到专注于反问题的提法上，转到建立更紧密的待估计参数与（部分）观测场之间的关系上。数据的完备性和完美性是提高反问题反演精度的关键因素，但一旦数据已经给出，反问题的解决还是依赖于所用的策略与方法。

6.2 勘探地震领域非线性反问题求解的基本理念

勘探地震领域中，关于反问题的研究，存在一些他山之石。譬如吴如山教授的研究从波传播的物理理

论切入，通过改善强绕射体存在时的正问题描述试图解决强扰动参数下的反演成像问题；Berkhout 教授带领的研究组基于 WRW 模型，实质上，是基于层状介质模型假设下的波传播，试图处理多次反射波存在下的高波数参数扰动量估计问题，具有鲜明的物理直觉，本质上是力图将反问题表述为一定的凸问题；另外从学术上看，Markov Monte Carlo 方法是比梯度导引类方法更好的解决途径，即通过在模型空间随机搜索，意图用模型空间中的随机采样刻画出完整的后验概率密度，得到更好的参数估计结果。

关于勘探地震领域非线性反问题求解，还有如下基本理念。首先是提出更好的“解的判决准则”。数据逼近残差 L2 范数作为衡量准则的确是不够合理的。它蕴含的假设太多，正问题$F(\boldsymbol{m})$必须能完全预测实测数据$\boldsymbol{d}^{obs}$中的各种物理波现象，仅仅余下满足高斯分布的随机噪声。众所周知，这是不可能的。另一方面，Gauss 分布是一个单峰的分布函数（凸函数），对应的误差泛函是二次型，这就预示着正问题$F(\boldsymbol{m})$是线性的（或接近线性的）。众所周知，正问题$F(\boldsymbol{m})$不是线性的，为了满足这个要求，必须提供非常接近真估计结果的初始值，正问题$F(\boldsymbol{m})$才能退化为线性的（或接近线性的）。这样的要求在复杂地表和复杂介质探区是不能满足的。关于反演问题解的判决依据，根本上看，还是从信息状态的角度才能得到深入的认识。我的观点：“反问题：永远是信息不足情况下的统计判决问题”。后验概率密度函数是最高判决依据。由于后验概率密度计算上很困难，退化是可以的，也是必须的。但退化成实测波形和预测波形的差的总能量，是否合适非常值得怀疑！FWI 的实践已经证明这是不行的！用包络代替波形、用加权旅行时代替波形、定义所谓的“最佳输运距离”、用互相关定义误差泛函、用道集拉平代替波形逼近定义误差泛函等，都可定义一个更凸的代价泛函，松弛了对初始值的严格要求，一定程度上，避免解陷入局部极值。但是，解的判决准则问题还是值得深入研究，尤其是从特征数据的角度提出新的误差泛函；从信息状态的角度，提出更好的、凸的误差泛函。

第二是如何构建更凸的反问题。我认为关键措施之一是从实测波场中筛选数据，即筛选出“部分可以很好预测的波场”。数学物理正问题模型相对于实际地下介质中的波传播情况都有所简化，导致正问题不能“解释”所有数据，这是反问题病态性的部分成因。因此，可以通过“修正”数据，使之能与正问题匹配，来缓解反问题的病态性。必须把特征数据筛选出来，匹配上能预测该特征数据的正问题，加上合适的模型参数先验信息，就可以构成一个很好的、比较凸的反问题。譬如，CMP 道集（数据）与基于层状介质的时距关系就是这样的特征数据与正问题匹配的典型实例。这看起来很简单，却支撑了地震波成像的实际应用。目前，这样的数据筛选工作做得很不到位，尤其在数据表达已经发展得很深入的情况下，并没有很好地借用来完善参数估计反问题。关键措施之二是特征化的模型参数表达。直接对模型参数进行网格离散，没有抓住要反演的模型参数的变化特征，徒然造成参数维度的指数上涨，反演难度加大。事实上，任何实际的物理系统对应的模型参数都是受自然规律控制的，都是有特征的。勘探地震中的介质特征就可以表示为：广大的沉积层中包含有构造运动、火山活动等引起的局部跃变的地质异常体。如能将模型参数进行特征表达、降维表达，那么反问题求解的难度可以下降很多。譬如，将模型拆分为宏观背景m_0+散射模型δm，若对散射模型做特征稀疏表达，可以进一步降维。实际上，勘探地震中介质以层状为主和波场以反射波为主的特征并没有在 FWI+LS_RTM 反演中得到真正的体现，导致这两种代表性的反演成像方法都难于真正实用化。

第三是单独考虑存在多次波时的非线性成像方法。勘探地震中，丰富的多次波的存在是地下介质存在强扰动、地震波反演出现较强非线性性的数据上的标志。强多次波存在时的地震波反演成像问题一定是强非线性反问题。首先要考虑的还是层状介质下的多次反射波存在下的偏移成像问题。这是最容易处理的问题。进一步地，要先估计高波数参数的变化，然后再考虑低中波数的参数反演。多次散射波反演成像要在对多次反射波成像积累经验后再进行。我不认为进行从低阶到高阶的、数据驱动的多次波分解，然后对各阶多次波进行线性成像是很合理的策略。我认为，**能够分阶次描述多次反射波传播的正问题**，是多次反射波偏移成像与反演成像的核心问题。基于此，可以展开对多次反射波的层析成像方法的研究。层状介质+Green 函数是构建**能够分阶次描述多次反射波传播的正问题**的基础。

第四是关于所谓的非线性 Hessian 在非线性反演成像中的作用的认识。所谓的非线性 Hessian 形式为：

$$H(\mathbf{m}) = F(\mathbf{m})^T F(\mathbf{m}) + \frac{\partial^2 F}{\partial \mathbf{m}^2}\Delta\mathbf{d} \tag{63}$$

我认为(63)式考虑了一阶多次反（散）波对 Hessian 的贡献。但这仅仅是梯度导引方法在当前迭代步的改进，仅仅适应弱非线性问题，不是从总体上试图解决非线性问题。

第五是考虑存在强介质参数扰动情形下正问题的改进。参数估计无论如何离不开用建立的数学物理关系来预测实测数据。若离开了这一点，任何反问题的逻辑基础就不存在了，变成了无中生有了！但是，在强介质参数扰动情况下，什么是合理的、恰当的数学物理关系？能与反问题很好地融合起来，非常方便后续的反演计算？我认为希望用一个考虑各阶多次波的正问题，尤其是（半）解析的正问题，譬如 Born 近似方法，是不现实的。我认为**能够分阶次描述多次反射波传播的正问题**是多次反射波偏移成像、层析成像的基础。

第六是考虑用 MCMC 方法+合理的解的评判准则来构建非线性反问题的求解策略。

理论上，MCMC 方法可能是更彻底的求解非线性反问题的途径。它不依赖与 Newton 的局部线性化假设，而是通过在模型空间随机搜索，意图用采样刻画出完整的后验概率密度。但是，本质上，MCMC 方法的有效性依然取决于后验概率密度的特征。对于具有非常多极值点的后验概率密度函数，再极端一些，对于存在很多高度一致的极值点的后验概率密度函数，MCMC 方法的有效性也是很存疑的！可以看出，根本问题在于反问题的提法，而不是反问题的数值求解方法。

我认为，可以用循序渐进的方法，首先用基于 CMP 道集层状介质下时距关系作为正问题，然后用 Kirchhoff PSDM 作为正问题，探索 MCMC 方法+后验概率密度函数生成下反演成像的合理逻辑。基于此，逐步开展 MCMC 方法在地震波反演成像中的研究。

6.3 勘探地震领域非线性反问题求解的具体策略

解决非线性问题，切实的出发点依然是从数据（Data）入手、从地震波传播的物理问题入手、从提出凸问题入手。反演理论的基本依据是正问题能预测或解释实测数据，这个基本依据永远是必要的！因为反问题的根本概念含义就是由可以直接观测的量估计与这些可以直接观测的量相关的潜（隐）变量。直接观测量及产生这些直接观测量的物理过程对反问题当然至关重要。正问题是联系（部分）直接观测量或特征观测量与待估计的潜（隐）变量之间的数学物理模型，当然也是至关重要的。直接观测量与由正问题数值解相减得到的残差满足 Gauss（广义 Gauss）分布是为了简化反问题的求解而施加的要求。这个人为施加的要求也不是可有可无的。要达到此要求：①对数据进行预处理，仅留下所选正问题能解释的数据；②改变正问题，更好地预测数据，但最好是用数据适应正问题而不是反之，因为改变正问题带来的是参数空间的变化，可能给反问题的求解带来困难；③施加正则化，引入更好的初始解。

总体而言，解决非线性问题的正确途径是从筛选数据入手。

地震勘探中炮集包含的特征数据子集可以罗列为：直达波；折射波；Diving Wave；一次反射波；一次散射波；多次反射波；多次散射波。上述特征数据（波现象）的分解是把 FWI 这个非线性反问题转换成一系列凸问题求解的关键。在人工智能（ML）系统下，可以把 FWI 包装成一系列凸问题组合，实现自动化与智能化的地震波反演成像。相应的关键技术包括：①特征波场的提出或提取，这需要高级的信号分析方法；②地质体几何结构的提取，这需要现代图像分析方法、模式识别和基于 ML 的智能图像识别；③包括多次反射波的层析成像与偏移成像技术；④包括多次散射波层析成像与偏移成像技术；⑤考虑复杂介质情况（弹性、各向异性、Q 值）下的上述一系列凸问题的求解；⑥上述一系列凸问题实现的一系列人工智能框架及相关的方法理论。

7 结论与讨论

线性、高斯、平稳假设下的地震波反演成像方法，事实上，也包括这种情形下的信号分析和图像分析方法，已经发展得十分完善，非线性、非高斯、非平稳假设下的地震波反演成像已经成为地震波成像方法研究的重点。

后验概率密度函数推广到广义 Gauss 分布情形，尤其是 L1 范数下进行稀疏反演，已经由稀疏信号表达和稀疏图像分析广泛地扩展到了勘探地震数据分析的相关环节中。非平稳信号分析更是由于短时 Fourier 变换、Gabor 变换和小波变换的引入成为了信号与图像分析的基本做法。非平稳、非高斯信号与图像分析也已

经成为常态。

但是，非线性参数估计问题，无论在信号和图像分析领域，还是在勘探地震领域中，都是迫切需要解决却没有得到很好解决的问题。尤其对于强非线性反问题，基本上没有什么有效的方法。FWI 就是这样一个典型的强非线性反问题。

对于 FWI 这样的强非线性反问题，我认为不可以期望找到一个有效的算法来解决，而是根据勘探地震的介质特征、波场特征，提炼出一系列凸问题的组合，来逼近该强非线性问题的解。鉴于地震数据仅仅在地表接收，而且叠前数据是频带带限的、孔径有限的，观测噪音是不可控的，仅仅靠叠前地震数据本身是无法满足储层描述所需要的高精度成像要求的，事实上，在复杂地表条件和复杂介质构造情况，即便获得准确的地下介质的构造像，都是难以做到的。因此，不要期望发展出一个完善的方法，能一劳永逸地解决地震波成像问题，不可能存在这样的方法。随着油气地震勘探目标逐渐复杂、尺度越来越小，埋藏越来越深，地震波成像问题永远是一个“信息不足情况下的弹性参数估计问题”。从综合信息利用的观点，看待地震波成像，逐步发展地震波成技术才是合理的思想逻辑。

求解 FWI 这样的强非线性反问题的正确路线，就是提出一系列凸问题的组合来逼近它的解。针对当前的具体情况，我提出如下的推进 FWI 实用化的基本做法：①、首先发展初至与早至波波动理论层析成像方法与技术。估计各向异性（VTI）速度和 Q 值。这中间初至波场与早至波场的提取与预处理是核心，不能像 Laplace 变换那样简单地处理。开始要发展具备自动的初至波到达时提取的波动理论层析成像方法。初至波的到达时要给出一个明确的定义，我提出用初至子波的质心来定义到达时。另外，要综合利用各种近地表模型的先验信息，譬如露头、高光谱、微测井等，减少近偏移距观测缺失带来的反问题解的不确定性。鉴于我国的油气探区大多位于陆上，首先要把该方法用在陆上地震数据成像处理中，使其能起到显著改善成像质量的作用，且能适用于山地、山前带、沙漠等不同探区的成像。然后再发展初至与早至波全波形波动理论层析成像方法，而且把前者的反演结果作为后者的初始值。② 然后发展以反射结构为中心的反射波波动理论层析成像方法与技术。其中的三个要点问题是：用偏移与反偏移结合筛选一次波数据；用机器学习算法自动提取反射界面、特殊地质体和断层，把上述信息用于速度模型更新的约束中，并随迭代过程不断演化构造样式；把梯度更新量累积起来，与地质结构约束一起，基于 Markov 演化的思想，分尺度修改速度模型，进一步提升梯度演化的稳定性。③发展最小二乘 RTM 及最小二乘 PSDM 成像方法。迭代的 LS-RTM 本质上就是估计高波数扰动量的 FWI，非线性性非常强，因此，必须要首先利用特征反射波场（一次反射波场），并在高频近似波传播框架下构建所谓的保真 RTM，本质上是用角度 Hessian 的逆对常规 RTM 结果进行预条件处理。另外，也可以在线性反问题成立的条件下，发展出像域最小二乘 RTM/PSDM 和数据域最小二乘 RTM/PSDM，本质上就是近似求解法方程。④发展宽带波阻抗成像方法。宽带波阻抗成像是当前带限反射系数成像的进一步扩展，宽带波阻抗成像分辨率更高，更有利于储层的描述与刻画。我提出了宽带波阻抗成像的具体策略解技术方案。⑤发展层状介质下的多次波偏移成像和层析反演成像。

总之，到目前为止，已经充分地认识到，由叠前地震数据估计地下介质的几何结构和弹性参数，进而结合岩石物理，估计地下介质的岩性参数，从而实现油气藏的描述与刻画是勘探地震的核心问题。抽象地看，也是个 Learning from Data 的问题。地震波成像问题不是一个大数据分析问题，油气藏的描述与刻画问题某种程度上是一个大数据分析问题。机器学习（ML）和人工智能（AI）可以在地震波成像的某些环节中（主要在正则化方面）起到辅助作用，但 FWI 参数反演问题只能视为 ML/AI 理论的一个具体应用实例，后者包含的理论和应有范围要宽泛很多。我相信总有一天可以发展出 ML/AI 理论下的弹性参数反演机器人，也可以发展出油气地震勘探机器人。当然，这并不代表人类的作为可以被这样的机器人所取代，而是人类在更高层次开展更为精确的弹性参数反演和油气预测。

参考文献

[1] 王华忠、冯波、王雄文、胡江涛、李辉，2015，地震波反演成像方法与技术核心问题分析[J]，石油物探，54（2），115-125

[2] 王华忠、胡江涛、郭颂，2017，最小二乘叠前深度偏移成像理论与方法[J]，石油物探，56（2），159-170

[3] 王华忠、郭颂、周阳，2019，宽带波阻抗建模[J]. 石油物探, 58(3): 313-324

[4] Billette, F. J., and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark: 67th Conference and

Exhibition, EAGE, Extended Abstracts, B035.

[5] Brenders, A. J., and R. G. Pratt, 2007, Waveform tomography of marine seismic data: What can limited offset offer?: 77th Annual International Meeting, SEG, Expanded Abstracts, 3024–3028, https://doi.org/10.1190/1.2793099.

[6] Chavent, G., F. Clément, and S. Gómez, 1994, Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example: 64th Annual International Meeting, SEG, Expanded Abstracts, 1179–1182, https://doi. org/10.1190/1.1822731.

[7] Datta, D., M. Sen, F. Liu, and S. Morton, 2016, Salt model building by shape-based parameterization and global FWI: 86th Annual International Meeting, SEG, Expanded Abstracts, 1069–1073, https://doi.org/10.1190/segam2016-13867592.1.

[8] Dellinger, J., A. Brenders, J. R. Sandschaper, C. Regone, J. Etgen, I. Ahmed, and K. J. Lee, 2017, The Garden Banks model experience: The Leading Edge, **36**, no. 2, 151–158, https://doi. org/10.1190/tle36020151.1.

[9] Dellinger, J., A. Ross, D. Meaux, A. Brenders, G. Gesoff, J. Etgen, J. Naranjo, G. Openshaw, and M. Harper, 2016, Wolfspar®, an "FWI-friendly" ultralow-frequency marine seismic source: 86th Annual International Meeting, SEG, Expanded Abstracts, 4891–4895, https://doi.org/10.1190/segam2016-13762702.1.

[10] Esser, E., L. Guasch, T. van Leeuwen, A. Aravkin, and F. J. Herrmann, 2015, Automatic salt delineation — Wavefield reconstruction inversion with convex constraints: 85th Annual International Meeting, SEG, Expanded Abstracts, 1337–1343, https://doi.org/10.1190/segam2015-5877995.1.

[11] Gomes, A., and N. Chazalnoel, 2017, Extending the reach of full waveform inversion with reflection data: Potential and challenges: 87th Annual International Meeting, SEG, Expanded Abstracts, 1454–1459, https://doi.org/10.1190/segam2017-17731403.1.

[12] Guasch, L., M. Warner, and F. J. Herrmann, 2016, Constrained waveform inversion — Automatic salt flooding with inclusions: 78th Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.201601547.

[13] Irabor, K., and M. Warner, 2016, Reflection FWI: 86th Annual International Meeting, SEG, Expanded Abstracts, 1136–1140, https://doi.org/10.1190/segam2016-13944219.1.

[14] Jonke, K., Z. Fu, B. Wray, and H. Shen, 2017, Improving minibasin and subsalt imaging with reflection full-waveform inversion: 87th Annual International Meeting, SEG, Expanded Abstracts, 1492–1496, https://doi.org/10.1190/segam2017-17661070.1.

[15] Kadu, A., T. van Leeuwen, and W. Mulder, 2016, A parametric level-set approach for seismic full-waveform inversion: 86th Annual International Meeting, SEG, Expanded Abstracts, 1146–1150, https://doi.org/10.1190/segam2016-13870276.1.

[16] Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: Theory and application, Society for Industrial and Applied Mathematics, 206–220.

[17] Lu, R., 2016, Revealing overburden and reservoir complexity with high-resolution FWI: 86th Annual International Meeting, SEG, Expanded Abstracts, 1242–1246, https://doi.org/10.1190/ segam2016-13872562.1.

[18] Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, no. 5, 645–653, https://doi.org/10.1190/1.1443081.

[19] Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and hybrid waveform inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 871–876, https://doi.org/10.1190/segam2013-0087.1.

[20] Mandroux, F., B. Ong, C. Ting, S. Mothi, T. Huang, and Y. Li, 2013, Broadband, long-offset, full-azimuth, staggered marine acquisition in the Gulf of Mexico: First Break, **31**, no. 6, 125–132.

[21] Michell, S., E. Shoshitaishvili, D. Chergotis, J. Sharp, and J. Etgen, 2006, Wide azimuth streamer imaging of Mad Dog: Have we solved the subsalt imaging problem?: 76th Annual International Meeting, SEG, Expanded Abstracts, 2905–2909, https://doi. org/10.1190/1.2370130.

[22] Michell, S., X. Shen, A. Brenders, J. Dellinger, I. Ahmed, and K. Fu, 2017, Automatic velocity model building with complex salt: Can computers finally do an interpreter's job?: 87th Annual International Meeting, SEG, Expanded Abstracts, 5250–5254, https://doi.org/10.1190/segam2017-17778443.1.

[23] Moldoveanu, N., and J. Kapoor, 2009, What is the next step after WAZ for exploration in the Gulf of Mexico?: 79th Annual International Meeting, SEG, Expanded Abstracts, 41–45, https://doi.org/10.1190/1.3255750.

[24] Mora, P., 1989, Inversion = migration + tomography: Geophysics, **54**, no. 12, 1575–1586, https://doi.org/10.1190/1.1442625.

[25] Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, no. 3, 888–901, https://doi.org/10.1190/1.1444597.

[26] Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2017, Salt model building at Atlantis with full-waveform inversion: 87th Annual International Meeting, SEG, Expanded Abstracts, 1507–1511, https://doi.org/10.1190/ segam2017-17738630.1.

[27] Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2018a, Full-waveform inversion: The next leap forward in subsalt imaging: The Leading Edge, **37**, no. 1, 67b1–67b6, https://doi.org/10.1190/tle37010067b1.1.

[28] Shen, X., L. Jiang, J. Dellinger, A. Brenders, C. Kumar, M. James, J. Etgen, D. Meaux, R. Walters, and N. Abdullayev, 2018b, High resolution full waveform inversion for structural imaging in exploration: 87th Annual International Meeting, SEG, Expanded Abstracts, 1098–1102, https://doi.org/10.1190/segam2018-2997202.1

[29] Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, no. 1, 231–248, https://doi.org/10.1190/1.1649391.

[30] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, no. 8, 1259–1266, https://doi.org/10.1190/1.1441754.

[31] Threadgold, I. M., K. Zembeck-England, P. G. Aas, P. M. Fontana, D. Hite, and W. E. Boone, 2006, Implementing a wide azimuth towed streamer field trial: The what, why and mostly how of WATS in Southern Green Canyon: 76th Annual International Meeting, SEG, Expanded Abstracts, 2901–2904, https://doi.org/10.1190/1.2370129.

[32] Wang, P., Z. Zhang, Z. Wei, and R. Huang, 2018, A demigration-based reflection full-waveform inversion workflow: 88th Annual International Meeting, SEG, Expanded Abstracts, 1138–1142, https://doi.org/10.1190/segam2018-2997404.1.

[33] Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambaré, 2012, Full waveform inversion for reflected seismic data: 74th Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.20148725.

[34] Yang, Z., S. Huang, and R. Yan, 2015, Improved subsalt tomography using RTM surface offset gathers: 85th Annual International Meeting, SEG, Expanded Abstracts, 5254–5258,

https://doi.org/10.1190/segam2015-5848366.1.

[35] Zhang, Z., J. Mei, F. Lin, R. Huang, and P. Wang, 2018, Correcting for salt misinterpretation with full-waveform inversion: 88th Annual International Meeting, SEG, Expanded Abstracts, 1143–1147, https://doi.org/10.1190/segam2018-2997711.1.

[36] Beaudoin, G., and A. A. Ross, 2007, Field design and operation of a novel deepwater, wide-azimuth node seismic survey: The Leading Edge, **26**, no. 4, 494–503, https://doi.org/10.1190/1.2723213.

[37] Billette, F. J., and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark: 67th Conference and Exhibition, EAGE, Extended Abstracts.

[38] Brenders, A. J., and R. G. Pratt, 2007, Waveform tomography of marine seismic data: What can limited offset offer?: 77th Annual International Meeting, SEG, Expanded Abstracts, 3024–3028, https://doi.org/10.1190/1.2793099.

[39] Dellinger, J., 2016, Challenges to extending the usable seismic bandwidth at the seafloor in the deepwater Gulf of Mexico: 86th Annual International Meeting, SEG, Expanded Abstracts, 66–70, https://doi.org/10.1190/segam2016-13762860.1.

[40] Dellinger, J., A. Ross, D. Meaux, A. Brenders, G. Gesoff, J. Etgen, J. Naranjo, G. Openshaw, and M. Harper, 2016, Wolfspar®, an "FWI-friendly" ultralow-frequency marine seismic source: 86th Annual International Meeting, SEG, Expanded Abstracts, 4891–4895, https://doi.org/10.1190/segam2016-13762702.1.

[41] Dellinger, J., A. J. Brenders, J. R. Sandschaper, C. Regone, J. Etgen, I. Ahmed, and K. J. Lee, 2017, The Garden Banks model experience: The Leading Edge, **36**, no. 2, 151–158, https://doi.org/10.1190/tle36020151.1.

[42] Gherasim, M., J. Etgen, B. Nolte, I. Ahmed, G. Xia, M. Reasnor, J. Howie, and D. Shepherd, 2010, Anisotropic velocity model building using OBS node tomography at Atlantis Field, Gulf of Mexico: 80th Annual International Meeting, SEG, Expanded Abstracts, 4359–4363, https://doi.org/10.1190/1.3513783.

[43] Plessix, R.-E., G. Baeten, J. W. de Maag, M. Klaassen, Z. Rujie, and T. Zhifei, 2010, Application of acoustic full waveform inversion to a low-frequency large-offset land data set: 80th Annual International Meeting, SEG, Expanded Abstracts, 930–934, https://doi.org/10.1190/1.3513930.

[44] Pratt, R. G., Z.-M. Song, P. Williamson, and M. Warner, 1996, Two-dimensional velocity models from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, no. 2, 323–340, https://doi.org/10.1111/j.1365-246X.1996.tb07023.x.

[45] Roberts, M., T. Dy, S. Ji, M. Reasnor, and D. Shepherd, 2011, Improving Atlantis TTI model building: OBN+NATS, prism waves & 3D RTM angle gathers: 81st Annual International Meeting, SEG, Expanded Abstracts, 3238–3242, https://doi. org/10.1190/1.3627868.

[46] Routh, P., J. Behura, and M. Tanis, 2016, Introduction to this special section: Full-waveform inversion Part I: The Leading Edge, **35**, no. 12, 1024–1024, https://doi.org/10.1190/tle35121024.1.

[47] Sirgue, L., 2006, The importance of low frequency and large offset in waveform inversion: 68th Conference and Exhibition, EAGE, Extended Abstracts.

[48] Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, no. 4, 65–70, https://doi.org/10.3997/1365-2397.2010012.

[49] Tanis, M., and J. Behura, 2017, Introduction to this special section: Full-waveform inversion Part II: The Leading Edge, **36**, no. 1, 58–58, https://doi.org/10.1190/tle36010058.1.

[50] Akcelik, V., 2002, Multiscale Newton-Krylov methods for inverse acoustic wave propagation: Ph.D. thesis,

Carnegie Mellon University.

[51] Aki, K., and P. Richards, 1980, Quantitative seismology: Theory and methods: W.H. Freeman & Co.

[52] Alerini, M., S. Le Bégat, G. Lambaré, and R. Baina, 2002, 2D PP- and PS-stereotomography for a multicomponent datset: 72nd Annual International Meeting, SEG, Expanded Abstracts, 838–841.

[53] Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

[54] Amundsen, L., 1991, Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion: Geophysics, **56**, 2027–2035.

[55] Amundsen, L., and B. Ursin, 1991, Frequency-wavenumber inversion of acoustic data: Geophysics, **56**, 1027–1039.

[56] Askan, A., 2006, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: Ph.D. thesis, Carnegie Mellon University.

[57] Askan, A., V. Akcelik, J. Bielak, and O. Ghattas, 2007, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: Bulletin of the Seismological Society of America, **97**, 1990–2008.

[58] Askan, A., and J. Bielak, 2008, Full anelastic waveform tomography including model uncertainty: Bulletin of the Seismological Society of America, **98**, 2975–2989.

[59] Barnes, C., and M. Charara, 2008, Full-waveform inversion results when using acoustic approximation instead of elastic medium: 78th Annual International Meeting, SEG, Expanded Abstracts, 1895–1899.

[60] Barnes, C., M. Charara, and T. Tsuchiya, 2008, Feasibility study for an anisotropic full waveform inversion of cross-well data: Geophysical Prospecting, **56**, 897–906.

[61] Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

[62] Bednar, J. B., C. Shin, and S. Pyun, 2007, Comparison of waveform inversion, Part 3:Amplitude approach: Geophysical Prospecting, **55**, 477–485.

[63] Ben Hadj Ali, H., S. Operto, and J. Virieux, 2008a, Velocity model building by 3D frequency-domain full-waveform inversion of wide-aperture seismic data: Geophysics, **73**, no. 5, VE101–VE117.

[64] Ben Hadj Ali, H., S. Operto, J. Virieux, and F. Sourbier, 2008b, 3D frequency-domain full-waveform tomography based on a domain decomposition forward problem: 78thAnnual International Meeting, SEG, Expanded Abstracts, 1945–1949

[65] Beydoun, W. B., and A. Tarantola, 1988, First Born and Rytov approximation: Modeling and inversion conditions in a canonical example: Journal of the Acoustical Society of America, **83**, 1045–1055.

[66] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, 99–108.

[67] Beylkin, G., and R. Burridge, 1990, Linearized inverse scattering problems in acoustics and elasticity: Wave Motion, **12**, 15–52.

[68] Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation from seismic reflection data by stereotomography: Geophysical Journal International, **135**, 671–680.

[69] Billette, F., S. Le Bégat, P. Podvin, and G. Lambaré, 2003, Practical aspects and applications of 2D stereotomography: Geophysics, **68**, 1008–1021.

[70] Biondi, B., and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

[71] Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis, 2007, Structure of the California Coast Ranges and San

Andreas fault at SAFOD from seismic waveform inversion and reflection imaging: Journal of Geophysical Research, doi: 10.1029/2006JB004611.

[72] Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52**, 931–942.

[73] Bouchon, M., M. Campillo, and S. Gaffet, 1989, A boundary integral equation — Discrete wavenumber representation method to study wave propagation in multilayered media having irregular interfaces: Geophysics, **54**, 1134–1140.

[74] Brenders, A. J., and R. G. Pratt, 2007a, Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data: Geophysical Journal International, **168**, 152–170.

[75] ———, 2007b, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.

[76] ———, 2007c, Waveform tomography of marine seismic data: What can limited offset offer?: 75thAnnual International Meeting, SEG, Expanded Abstracts, 3024–3029.

[77] Bube, K. P., and T. Nemeth, 2007, Fast line searches for the robust solution of linear systems in the hybrid 1/2 and Huber norms: Geophysics, **72**, no. 2, A13–A17.

[78] Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

[79] Capdeville, Y., Y. Gung, and B. Romanowicz, 2005, Towards global earth tomography using the spectral element method: A technique based on source stacking: Geophysical Journal International, **162**, 541–554.

[80] Carcione, J. M., G. C. Herman, and A. P. E. ten Kroode, 2002, Seismic modeling: Geophysics, **67**, 1304–1325.

[81] Cary, P., and C. Chapman, 1988, Automatic 1-D waveform inversion of marine seismic refraction data: Geophysical Journal of the Royal Astronomical Society, **93**, 527–546.

[82] Chapman, C., 1985, Ray theory and its extension: WKBJ and Maslov seismograms: Journal of Geophysics, **58**, 27–43.

[83] Chen, P., T. Jordan, and L. Zhao, 2007, Full three-dimensional tomography: A comparison between the scattering-integral and adjoint-wavefield methods: Geophysical Journal International, **170**, 175–181.

[84] Chironi, C., J. V. Morgan, and M. R.Warner, 2006, Imaging of intra-basalt and sub-basalt structure with full wavefield seismic tomography: Journal of Geophysical Research, **11**, B05313, doi: 10.1029/2004JB003595.

[85] Choi, Y., D.-J. Min, and C. Shin, 2008, Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupledmedia: Geophysical Prospecting, **56**, 863–881.

[86] Claerbout, J., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

[87] ———, 1976, Fundamentals of geophysical data processing: McGraw-Hill Book Co., Inc.

[88] Claerbout, J., and S. Doherty, 1972, Downward continuation of moveout corrected seismograms: Geophysics, **37**, 741–768.

[89] Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic non-linear waveform inversion:Application to real data: Geophysics, **55**, 527–538.

[90] Danecek, P., and G. Seriani, 2008, An efficient parallel Chebyshev pseudo-spectral method for large-scale 3D seismic forward modelling: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, P046.

[91] de Hoop, A., 1960, A modification of Cagniard's method for solving seismic pulse problems: Applied Scientific Research, **8**, 349–356.

[92] Dessa, J. X., and G. Pascal, 2003, Combined traveltime and frequency-domain seismic waveform inversion:

A case study on multi-offset ultrasonic data: Geophysical Journal International, **154**, 117–133.

[93] Devaney, A. J., 1982, A filtered back-projection algorithm for diffraction tomography: Ultrasonic Imaging, **4**, 336–350.

[94] Devaney, A. J., and D. Zhang, 1991, Geophysical diffraction tomography in a layered background: Wave Motion, **14**, 243–265.

[95] Djikpéssé, H., and A. Tarantola, 1999, Multiparameter 1-norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, **64**, 1023–1035.

[96] Docherty, P., R. Silva, S. Singh, Z.-M. Song, and M. Wood, 2003, Migration velocity analysis using a genetic algorithm: Geophysical Prospecting, **45**, 865–878.

[97] Dumbser, M., and M. Kaser, 2006, An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes — II. The three-dimensional isotropic case: Geophysical Journal International, **167**, 319–336.

[98] Dummong, S., K. Meier, D. Gajewski, and C. Hubscher, 2008, Comparison of prestack stereotomography and NIP wave tomography for velocity model building: Instances from the Messinian evaporites: Geophysics, **73**, no. 5, VE291–VE302.

[99] Ellefsen, K., 2009, A comparison of phase inversion and traveltime tomography for processing of near-surface refraction traveltimes: Geophysics, VOL. 74, NO. 6; P. WCB11–WCB24.

[100] Epanomeritakis, I., V. Akçelik, O. Ghattas, and J. Bielak, 2008, A Newton CG method for large-scale three-dimensional elastic full waveform seismic inversion: Inverse Problems, 24, 1–26.

[101] Erlangga, Y. A., and F. J. Herrmann, 2008, An iterative multilevel method for computing wavefields in frequency-domain seismic inversion: 78thAnnual International Meeting, SEG, ExpandedAbstracts, 1956–1960.

[102] Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, no. 5, J53–J64.

[103] Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2008, Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain: Geophysical Journal International, **175**, 665–685.

[104] Forgues, E., and G. Lambaré, 1997, Parameterization study for acoustic and elastic ray Born inversion: Journal of Seismic Exploration, **6**, 253–278.

[105] Futterman, W., 1962, Dispersive body waves: Journal of Geophysics Research, **67**, 5279–5291.

[106] Gao, F., A. R. Levander, R. G. Pratt, C. A. Zelt, and G. L. Fradelizio, 2006a, Waveform tomography at a groundwater contamination site: Surface reflection data: Geophysics, **72**, no. 5, G45–G55.

[107] ———, 2006b, Waveform tomography at a groundwater contamination site: VSP-surface data set: Geophysics, **71**, no. 1, H1–H11.

[108] Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.

[109] Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geophysics, **43**, 1342–1351.

[110] Gelis, C., J. Virieux, and G. Grandjean, 2007, 2D elastic waveform inversion using Born and Rytov approximations in the frequency domain: Geophysical Journal International, **168**, 605–633.

[111] Gelius, L. J., I. Johansen, N. Sponheim, and J. J. Stamnes, 1991, A generalized diffraction tomography algorithm: Journal of the Acoustical Society of America, **89**, 523–528.

[112] Gilbert, F., and A. Dziewonski, 1975, An application of normal mode theory to the retrieval of structural

parameters and source mechanisms from seismic spectra: Philosophical Transactions of the Royal Society of London, **278**, 187–269.

[113] Graves, R., 1996, Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences: Bulletin of the Seismological Society of America, **86**, 1091–1106.

[114] Grechka, V., L. Zhang, and J.W. Rector III, 2004, Shear waves in acoustic anisotropic media: Geophysics, **69**, 576–582.

[115] Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, 1310–1319.

[116] Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a Huber function norm: Geophysics, **74**, no. 3, R15–R24.

[117] Hak, B., and W. Mulder, 2008, Preconditioning for linearized inversion of attenuation and velocity perturbations: 78th Annual International Meeting, SEG, Expanded Abstracts, 2036–2040.

[118] Hansen, C., ed., 1998, Rank-defficient and discrete ill-posed problems —Numerical aspects of linear inversion: Society for Industrial and Applied Mathematics.

[119] Herrmann, F. J., Y. Erlangga, and T. T.Y. Lin, 2009, Compressive sensing applied to full-wave form inversion: 71st Conference & Technical Exhibition, EAGE, Extended Abstracts, S016.

[120] Hicks, G. J., and R. G. Pratt, 2001, Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit: Geophysics, **66**, 598–612.

[121] Hole, J. A., 1992, Nonlinear high-resolution three-dimensional seismic travel time tomography: Journal of Geophysical Research, **97**, 6553–6562.

[122] Hu, W., A. Abubakar, and T. M. Habashy, 2009, Simultaneous multifrequency inversion of full-waveform seismic data: Geophysics, **74**, no. 2, R1–R14.

[123] Hustedt, B., S. Operto, and J. Virieux, 2004, Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling: Geophysical Journal International, **157**, 1269–1296.

[124] Ikelle, L., J. P. Diet, and A. Tarantola, 1988, Linearized inversion of multioffset seismic reflection data in the -*k* domain: Depth-dependent reference medium: Geophysics, **53**, 50–64.

[125] Improta, L., A. Zollo, A. Herrero, R. Frattini, J. Virieux, and P. DellÁversana, 2002, Seismic imaging of complex structures by non-linear traveltimes inversion of dense wide-angle data:Application to a thrust belt: Geophysical Journal International, **151**, 264–278.

[126] Jaiswal, P., C. Zelt, A. W. Bally, and R. Dasgupta, 2008, 2-D traveltime and waveform inversion for improved seismic imaging: Naga thrust and fold belt, India: Geophysical Journal International, doi: 10.1111/j.1365-246X.2007.03691.x.

[127] Jaiswal, P., C. Zelt, R. Dasgupta, and K. Nath, 2009, Seismic imaging of the Naga thrust using multiscale waveform inversion: Geophysics, VOL. 74, NO. 6; P. WCC129–WCC140.

[128] Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, and D. Trézéguet, 1989,Wavelengths of earth structures that can be resolved from seismic reflection data: Geophysics, **54**, 906–910.

[129] Jin, S., and R. Madariaga, 1993, Background velocity inversion by a genetic algorithm: Geophysical Research Letters, **20**, 93–96.

[130] ———, 1994, Non-linear velocity inversion by a two-step Monte Carlo method: Geophysics, **59**, 577–590.

[131] Jin, S., R. Madariaga, J. Virieux, and G. Lambaré, 1992, Two-dimensional asymptotic iterative elastic inversion: Geophysical Journal International, **108**, 575–588.

[132] Jones, K., M. Warner, and J. Brittan, 1999, Anisotropy in multi-offset deepcrustal seismic experiments: Geophysical Journal International, **138**, 300–318.

[133] Kamei, R., and R. G. Pratt, 2008, Waveform tomography strategies for imaging attenuation structure for cross-hole data: 70th Conference & Technical Exhibition, EAGE, ExtendedAbstracts, F019.

[134] Kennett, B. L. N., 1983, Seismic wave propagation in stratified media: Cambridge University Press.

[135] Klem-Musatov, K. D., and A. M. Aizenberg, 1985, Seismic modelling by methods of the theory of edge waves: Journal of Geophysics, **57**, 90–105.

[136] Kolb P., F. Collino, and P. Lailly, 1986, Prestack inversion of 1-D medium: Proceedings of the IEEE, 498–508.

[137] Kolsky, H., 1956, The propagation of stress pulses in viscoelastic solids: Philosophical Magazine, **1**, 693–710.

[138] Komatitsch, D., and J. P. Vilotte, 1998, The spectral element method:An efficient tool to simulate the seismic response of 2D and 3D geological structures: Bulletin of the Seismological Society ofAmerica, **88**, 368–392.

[139] Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, **96**, 20289–20299.

[140] Kormendi, F., and M. Dietrich, 1991, Non-linear waveform inversion of plane-wave seismograms in stratified elastic media: Geophysics, **56**, 664–674.

[141] Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, A. Baumstein, M. D. Lacasse, and S. Lee, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, VOL. 74, NO. 6; P. WCC177–WCC188.

[142] Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

[143] Lambaré, G., 2008, Stereotomography: Geophysics, **73**, no. 5, VE25–VE34.

[144] Lambaré, G., and M. Alérini, 2005, Semi-automatic picking PP-PS stereotomography:Application to the synthetic Valhall data set: 75thAnnual International Meeting, SEG, ExpandedAbstracts, 943–946.

[145] Lambaré, G., S. Operto, P. Podvin, P. Thierry, and M. Noble, 2003, 3-D ray Born migration/inversion — Part 1: Theory: Geophysics, **68**, 1348–1356.

[146] Lambaré, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, **57**, 1138–1154.

[147] Lecomte, I., 2008, Resolution and illumination analyses in PSDM: A raybased approach: The Leading Edge, **27**, 650–663.

[148] Lecomte, I., S.-E. Hamran, and L.-J. Gelius, 2005, Improving Kirchhoff migration with repeated local plane-wave imaging? A SAR-inspired signalprocessing approach in prestack depth imaging: Geophysical Prospecting, **53**, 767–785.

[149] Lee, K. H., and H. J. Kim, 2003, Source-independent full-waveform inversion of seismic data: Geophysics, **68**, 2010–2015.

[150] Lehmann, I., 1936, P: Publications du Bureau Central Séismologique International, A14, 87–115.

[151] Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436.

[152] Liao, Q., and G. A. McMechan, 1995, 2.5D full-wavefield viscoacoustic inversion: Geophysical Prospecting, **43**, 1043–1059.

[153] ———, 1996, Multifrequency viscoacoustic modeling and inversion: Geophysics, **61**, 1371–1378.

[154] Lions, J., 1972, Nonhomogeneous boundary value problems and applications: Springer-Verlag, Berlin.

[155] Lucio, P. S., G. Lambaré, and A. Hanyga, 1996, 3D multivalued travel time and amplitude maps: Pure andApplied Geophysics, **148**, 113–136.

[156] Luo, Y., T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Seismic modeling and imaging based upon spectral-element and adjoint methods: The Leading Edge, **28**, 568–574.

[157] Malinowsky, M., and S. Operto, 2008, Quantitative imaging of the permomesozoic complex and its basement by frequency domain waveform tomography of wide-aperture seismic data from the polish basin: Geophysical Prospecting, **56**, 805–825.

[158] Malinowsky, M., A. Ribodetti, and S. Operto, 2007, Multiparameter fullwaveform inversion for velocity and attenuation — Refining the imaging of a sedimentary basin: 69th Conference & Technical Exhibition, EAGE, ExtendedAbstracts, P276.

[159] Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.

[160] McMechan, G. A., and G. S. Fuis, 1987, Ray equation migration of wide-angle reflections from southern Alaska: Journal of Geophysical Research, **92**, 407–420.

[161] Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

[162] Miller, D., M. Oristaglio, and G. Beylkin, 1987, Anew slant on seismic imaging: Migration and integral geometry: Geophysics, **52**, 943–964.

[163] Min, D.-J., and C. Shin, 2006, Refraction tomography using a waveform-inversion back-propagation technique: Geophysics, **71**, no. 3, R21–R30.

[164] Min, D.-J., C. Shin, R. G. Pratt, and H. S. Yoo, 2003, Weighted-averaging finite-element method for 2D elastic wave equations in the frequency domain: Bulletin of the Seismological Society ofAmerica, **93**, 904–921.

[165] Montelli, R., G. Nolet, F. Dahlen, G. Masters, E. Engdahl, and S. Hung, 2004, Finite-frequency tomography reveals a variety of plumes in the mantle: Science, **303**, 338–343.

[166] Mora, P. R., 1987, Nonlinear two-dimensional elastic inversion of multi-offset seismic data: Geophysics, **52**, 1211–1228.

[167] ———, 1988, Elastic wavefield inversion of reflection and transmission data: Geophysics, **53**, 750–759.

[168] Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447.

[169] Mulder, W. A., and B. Hak, 2009, Simultaneous imaging of velocity and attenuation perturbations from seismic data is nearly impossible: 71st Conference & Technical Exhibition, EAGE, ExtendedAbstracts, S043.

[170] Nihei, K. T., and X. Li, 2007, Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection TD-PSD: Geophysical Journal International, **169**, 1069–1078.

[171] Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.

[172] Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer.

[173] Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics: D. Reidel Publ. Co.

[174] Oldham, R., 1906, The constitution of the earth: Quarterly Journal of the Geological Society of London, **62**, 456–475.

[175] Operto, S., G. Lambaré, P. Podvin, and P. Thierry, 2003, 3-D ray-Born migration/inversion, Part 2:

Application to the SEG/EAGE overthrust experiment: Geophysics, **68**, 1357–1370.

[176] Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Quantitative imaging of complex structures from multifold wide aperture seismic data: A case study: Geophysical Prospecting, **52**, 625–651.

[177] Operto, S., J. Virieux, P. Amestoy, J. Y. L'Excellent, L. Giraud, and H. BenHadj-Ali, 2007, 3D finite-difference frequency-domain modeling of viscoacoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, no. 5, SM195–SM211.

[178] Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain fullwaveform tomography: Application to the eastern Nankai trough: Journal of Geophysical Research, doi: 10.1029/2005JB003835.

[179] Operto, S., J. Virieux, A. Ribodetti, and J. E. Anderson, 2009, Finite-difference frequency-domain modeling of viscoacoustic wave propagation in two-dimensional tilted transversely isotropic media: Geophysics, **74**, no.5, T75–T95, doi: 10.1190/1.3157243.

[180] Operto, S., S. Xu, and G. Lambaré, 2000, Can we image quantitatively complex models with rays?: Geophysics, **65**, 1223–1238.

[181] Paige, C. C., and M. A. Saunders, 1982a, Algorithm 583 LSQR: Sparse linear equations and least squares problems: Association for Computing Machinery Transactions on Mathematical Software, **8**, 195–209.

[182] ———, 1982b, LSQR: An algorithm for sparse linear equations and sparse least squares: Association for Computing Machinery Transactions on Mathematical Software, **8**, 43–71.

[183] Pica, A., J. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in laterally invariant medium: Geophysics, **55**, 284–292.

[184] Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

[185] ———, 2007, AHelmholtz iterative solver for 3D seismic-imaging problems: Geophysics, **72**, no. 5, SM185–SM194.

[186] ———, 2009, 3D frequency-domain full-waveform inversion with an iterative solver: Geophysics, VOL. 74, NO. 6, P. WCC149–WCC157.

[187] Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method: Geophysical Prospecting, **38**, 311–330.

[188] ———, 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

[189] Pratt, R. G., and N. R. Goulty, 1991, Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data: Geophysics, **56**, 204–224.

[190] Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

[191] Pratt, R. G., L. Sirgue, B. Hornby, and J. Wolfe, 2008, Cross-well waveform tomography in fine-layered sediments — Meeting the challenges of anisotropy: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F020.

[192] Pratt, R. G., Z. M. Song, P. R.Williamson, and M.Warner, 1996, Two-dimensional velocity model from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, 323–340.

[193] Pratt, R. G., andW. Symes, 2002, Semblance and differential semblance optimisation for waveform tomography:Afrequency domain implementation: Sub-basalt imaging conference, ExpandedAbstracts,

183–184.

[194] Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multisource cross-hole tomography, Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310.

[195] Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Veltering, 1986, Numerical recipes: The art of scientific computing: Cambridge University Press.

[196] Prieux, V., S. Operto, R. Brossier, and J. Virieux, 2009, Application of acoustic full waveform inversion to the synthetic Valhall model: 77thAnnual International Meeting, SEG, ExpandedAbstracts, 2268–2272.

[197] Pyun, S., C. Shin, and J. B. Bednar, 2007, Comparison of waveform inversion, Part 3: Phase approach: Geophysical Prospecting, **55**, 465–475.

[198] Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Multi-scale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, **159**, 1032–1056.

[199] Ribodetti, A., S. Operto, J. Virieux, G. Lambaré, H.-P. Valéro, and D. Gibert, 2000, Asymptotic viscoacoustic diffraction tomography of ultrasonic laboratory data: A tool for rock properties analysis: Geophysical Journal International, **140**, 324–340.

[200] Riyanti, C. D., Y. A. Erlangga, R. E. Plessix, W. A. Mulder, C. Vuik, and C. Oosterlee, 2006, A new iterative solver for the time-harmonic wave equation: Geophysics, **71**, no. 5, E57–E63.

[201] Riyanti, C. D., A. Kononov, Y. A. Erlangga, C. Vuik, C. W. Oosterlee, R. E. Plessix, and W. A. Mulder, 2007, A parallel multigrid-based preconditioner for the 3D heterogeneous high-frequency Helmholtz equation: Journal of Computational Physics, **224**, 431–448.

[202] Robertsson, J. O., B. Bednar, J. Blanch, C. Kostov, and D. J. van Manen, 2007, Introduction to the supplement on seismic modeling with applications to acquisition, processing and interpretation: Geophysics, **72**, no. 5, SM1–SM4.

[203] Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

[204] Saad, Y., 2003, Iterative methods for sparse linear systems, 2nd ed.: Society for Industrial andApplied Mathematics.

[205] Sambridge, M., and G. Drijkoningen, 1992, Genetic algorithms in seismic waveform inversion: Geophysical Journal International, **109**, 323–342.

[206] Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, 1–29.

[207] Sambridge, M. S., A. Tarantola, and B. L. N. Kennett, 1991, An alternative strategy for non linear inversion of seismic waveforms: Geophysical Prospecting, **39**, 457–491.

[208] Scales, J. A., P. Docherty, and A. Gersztenkorn, 1990, Regularization of nonlinear inverse problems: Imaging the near-surface weathering layer: Inverse Problems, **6**, 115–131.

[209] Scales, J. A., and M. L. Smith, 1994, Introductory geophysical inverse theory: Samizdat Press.

[210] Sears, T. J., S. C. Singh, and P. J. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56**, 843–862.

[211] Sen, M. K., and P. L. Stoffa, 1995, Global optimization methods in geophysical inversion: Elsevier Science Publ. Co., Inc. Sheng, J., A. Leeds, M. Buddensiek, and G. T. Schuster, 2006, Early arrival waveform tomography on near-surface refraction data: Geophysics, **71**,
no. 4, U47–U57.

[212] Shi, Y., W. Zhao, and H. Cao, 2007, Nonlinear process control of wave-equation inversion and its application in the detection of gas: Geophysics, **72**, no. 1, R9–R18.

[213] Shin, C., and W. Ha, 2008, A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains: Geophysics, **73**, no. 5, VE119–VE133.

[214] Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931.

[215] ———, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, **177**, 1067–1079.

[216] Shin, C., S. Jang, and D. J. Min, 2001a, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

[217] Shin, C., and D.-J. Min, 2006, Waveform inversion using a logarithmic wavefield: Geophysics, **71**, no. 3, R31–R42.

[218] Shin, C., S. Pyun, and J. B. Bednar, 2007, Comparison of waveform inversion, Part 1: Conventional wavefield vs. logarithmic wavefield: Geophysical Prospecting, **55**, 449–464.

[219] Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin, 2001b, Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion: Geophysics, **66**, 1856–1863.

[220] Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data: Geophysical Journal International, **151**, 325–344.

[221] Snieder, R., M. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem: Geophysics, **54**, 991–1000.

[222] Song, Z., P. Williamson, and G. Pratt, 1995, Frequency-domain acousticwave modeling and inversion of crosshole data, Part 2: Inversion method, synthetic experiments and real-data results: Geophysics, **60**, 786–809.

[223] Sourbier, F., A. Haiddar, L. Giraud, S. Operto, and J. Virieux, 2008, Frequency-domain full-waveform modeling using a hybrid direct-iterative solver based on a parallel domain decomposition method: A tool for 3D fullwaveform inversion?: 78th Annual International Meeting, SEG, ExpandedAbstracts, 2147–2151.

[224] Sourbier, F., S. Operto, J. Virieux, P. Amestoy, and J. Y. L'Excellent, 2009a, FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data — Part 1: Algorithm: Computers and Geosciences, **35**, 487–495.

[225] ———, 2009b, FWT2D:Amassively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data — Part 2: Numerical examples and scalability analysis: Computers & Geosciences, **35**, 496–514.

[226] Stekl, I., and R. G. Pratt, 1998, Accurate viscoelastic modeling by frequencydomain finite difference using rotated operators: Geophysics, **63**, 1779–1794.

[227] Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, **43**, 23–48.

[228] Taillandier, C., M. Noble, H. Chauris, and H. Calandra, 2009, First-arrival travel time tomography based on the adjoint state method: Geophysics, VOL. 74, NO. 6; P. WCB1–WCB10.

[229] Talagrand, O., and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation, I: Theory: Quarterly Journal of the Royal Meteorological Society, **113**, 1311–1328.

[230] Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

[231] ———, 1986, A strategy for nonlinear inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

[232] ———, 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Science Publ. Co., Inc.

[233] Thierry, P., G. Lambaré, P. Podvin, and M. Noble, 1999a, 3-D preserved amplitude prestack depth migration on a workstation: Geophysics, **64**, 222–229.

[234] Thierry, P., S. Operto, and G. Lambaré, 1999b, Fast 2D ray-Born inversion/migration in complex media: Geophysics, **64**, 162–181.

[235] Thomsen, L. A., 1986,Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

[236] van den Berg, P. M., and A. Abubakar, 2001, Contrast source inversion method: State of the art: Progress in Electromagnetics Research, **34**, 189–218.

[237] Vigh, D., and E. W. Starr, 2008a, 3D plane-wave full-waveform inversion: Geophysics, **73**, no. 5, VE135–VE144.

[238] ———, 2008b, Comparisons for waveform inversion, time domain or frequency domain?: 78thAnnual International Meeting, SEG, Expanded Abstracts, 1890–1894.

[239] Vigh, D., E. W. Starr, and J. Kapoor, 2009, Developing earth model with full waveform inversion: The Leading Edge, **28**, 432–435.

[240] Virieux, J., 1986, P-SV wave propagation in heterogeneous media, velocitystress finite difference method: Geophysics, **51**, 889–901.

[241] Virieux, J., S. Operto, H. B. H. Ali, R. Brossier, V. Etienne, F. Sourbier, L. Giraud, and A. Haidar, 2009, Seismic wave modeling for seismic imaging: The Leading Edge, **28**, 538–544.

Vogel, C., 2002, Computational methods for inverse problems: Society of Industrial and Applied Mathematics.

[242] Vogel, C. R., and M. E. Oman, 1996, Iterative methods for total variation denoising: SIAM Journal on Scientific Computing, **17**, 227–238.

[243] Wang, Y., and Y. Rao, 2009, Reflection seismic waveform tomography: Journal of Geophysical Research, doi: 10.1029/2008JB005916.

[244] Warner, M., I. Stekl, and A. Umpleby, 2007, Full wavefield seismic tomography — Iterative forward modeling in 3D: 69th Conference & Technical Exhibition, EAGE, ExtendedAbstracts, C025.

[245] ———, 2008, 3D wavefield tomography: synthetic and field data examples: 78th Annual International Meeting, SEG, Expanded Abstracts, 3330–3334.

[246] Williamson, P., 1991, A guide to the limits of resolution imposed by scattering in ray tomography: Geophysics, **56**, 202–207.

[247] Woodhouse, J., and A. Dziewonski, 1984, Mapping the upper mantle: Three dimensional modelling of earth structure by inversion of seismic waveforms: Journal of Geophysical Research, **89**, 5953–5986.

[248] Woodward, M. J., 1992, Wave-equation tomography: Geophysics, **57**, 15–26.

[249] Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: Geophysics, **73**, no. 5, VE5–VE11.

[250] Wu, R. S., 2003, Wave propagation, scattering and imaging using dual-domain one-way and one-return propagation: Pure and Applied Geophysics, **160**, 509–539.

[251] Wu, R. S., and K. Aki, 1985, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, 582–595.

[252] Wu, R.-S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic

imaging: Geophysics, **52**, 11–25.

[253] Yilmaz, O., 2001, Seismic data analysis: processing, inversion and interpretation of seismic data: SEG.

[254] Zelt, C., and P. J. Barton, 1998, Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe basin: Journal of Geophysical Research, **103**, 7187–7210.

[255] Zelt, C. A., R. G. Pratt, A. J. Brenders, S. Hanson-Hedgecock, and J. A. Hole, 2005, Advancements in long-offset seismic imaging:Ablind test of traveltime and waveform tomography: Eos, Transactions of the American Geophysical Union, **86**,Abstract S52A-04.

[256] Zhou, B., and S. A. Greenhalgh, 2003, Crosshole seismic inversion with normalized full-waveform amplitude data: Geophysics, **68**, 1320–1330.

勘探地震中地震信号的相位问题分析

王华忠，盛燊

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】勘探地震中的波场信号与其他广义遥感领域（譬如电磁波雷达、微波雷达、超声医学探测成像、超声无损探伤等）的信号相比有独到的特点，这些特点是地震信号分析和成像的物理基础。在以层状介质为主的、仅仅地表观测的勘探地震数据情形下，炮集中包含的波场可以认为是各种复杂的波现象飘在相干或随机的噪音中。这些复杂的波现象中，透射波（直达波、初至波和 Diving Wave）可被用来进行透射波层析成像（透射波 FWI）建立浅中层速度模型、一次反射波用来进行偏移成像和反射波层析成像（反射波 FWI）。自由地表相关多次波和层间多次波也被期望贡献于成像过程中。自由地表相关的波现象（海洋勘探中的自由地表相关多次波、陆上勘探中的面波和面波散射波等）在油气勘探中一般地是作为噪音被处理掉。无论是可用于成像的波现象或不能用于成像的干扰波现象，都可抽象为带限子波形成的同相轴飘在随机噪声中。带限子波携带了关于地下介质和震源的信息。因此，地震勘探中地震波成像就是利用不同炮检对观测的带限子波的到达时信息估计背景速度、利用带限子波的振幅信息估计（角度）反射系数。更精细的速度估计会用到同相轴的相位信息；吸收衰减参数 Q 值的估计会用到带限子波的振幅衰减信息。勘探地震中对地震数据分析和地震波成像的精度要求越来越高，因为是带限子波携带了地下介质和震源的信息，必然要对带限子波的表达进行更深入的研究。全波形反演代表了当前地震波成像的最先进方法，但是全波形反演是把带限子波视为一个整体，并没有对带限子波本身进行解剖性研究，这也间接地导致了 FWI 的实用化效果不能令人满意。勘探地震中带限子波被认为是各种幅值的谐波的叠加，这是符合物理实际的，但是各谐波的相位分布决定了带限子波的形态。在高精度的保真成像阶段和油藏描述阶段，对带限子波的相位问题进行深入地分析是很有必要的。本文从不同的角度分析带限子波的相位对地震波成像的影响。

【关键词】波现象，高精度成像，全波形反演，带限子波，振幅，相位

1 引言

利用在目标体外观测的波场数据，对目标体内介质的结构和物性参数进行估计或成像的学科领域都可称为广义的遥感领域。遥感领域的核心问题显然是波场数据的分析与成像。但是，由于不同的具体领域的数据特征相差很大，数据分析和成像的方法技术还是存在显著的差异。

勘探地震中的波场信号与其他广义遥感领域（譬如电磁波雷达、微波雷达、超声医学探测成像、超声无损探伤等）的信号相比有独到的特点，这些特点是地震信号分析和成像的物理基础。

在以层状介质为主的、仅仅地表观测的勘探地震数据情形下，炮集中包含的波场可以认为是各种复杂的波现象飘在相干或随机的噪音中。这些复杂的波现象中，透射波（直达波、初至波和 Diving Wave）可被用来进行透射波层析成像（透射波 FWI）建立浅中层速度模型、一次反射波用来进行偏移成像和反射波层析成像（反射波 FWI）。自由地表相关多次波和层间多次波也被期望贡献于成像过程中。自由地表相关的波现象（海洋勘探中的自由地表相关多次波、陆上勘探中的面波和面波散射波等）在油气勘探中一般地是作为噪音被处理掉。无论是可用于成像的波现象或不能用于成像的干扰波现象，都可抽象为带限子波形成的同相轴飘在随机噪声中。

带限子波携带了关于地下介质和震源的信息。地震勘探中地震波成像就是利用不同炮检对观测的带限子波的到达时信息估计背景速度、利用带限子波的振幅信息估计（角度）反射系数。更精细的速度估计会用到同相轴上带限子波的相位信息；吸收衰减参数 Q 值的估计会用到带限子波的振幅衰减信息。

勘探地震中对地震数据分析和地震波成像的精度要求越来越高。由于带限子波携带了地下介质和震源的信息，必然要对带限子波的表达进行更深入的研究。全波形反演代表了当前地震波成像的最先进方法，但是全波形反演是把带限子波视为一个整体，并没有对带限子波本身进行解剖性研究，这也间接地导致了 FWI 的实用化效果不能令人满意。

勘探地震中带限子波被认为是各种幅值的谐波的叠加，这是符合物理实际的。但是各谐波的相位分布决定了带限子波的形态。当前，信号分析与处理基本上都是基于信号的能量谱（功率谱）。但是功率谱相同的信号，形态上可能差异很大。功率谱刻画的是信号的重要部分，却是背景部分，更细节的部分是由更高阶的统计量决定的，某种程度上是相位变化决定的。

在高精度的保真成像阶段和油藏描述阶段，对带限子波的相位问题进行深入地分析是很有必要的。因此，本文从不同的角度分析带限子波的相位对地震波成像的影响。

2 勘探地震系统的线性性与相干信号及同相位叠加

根据信号与系统理论，带限子波（系统激励函数）通过一个线性系统后，系统输出与输入的带限子波之间构成相干信号。

相干信号的定义：假定系统输入的带限子波为 $f(\boldsymbol{x},t)$，系统输出的带限子波为 $g(\boldsymbol{x},t)=\mathrm{c}\cdot f(\boldsymbol{x},t)$，其中 $c=|c|e^{i\varphi}$ 是复数，则系数输出的带限子波为系统输入的带限子波的相干信号。

根据定义可以看出，相干信号之间只有振幅的常系数调整和常相位的时移，二者之间的形态是完全一样的。即激励带限子波在一个线性系统（介质）中进行传播，该系统并不改变带限子波中各谐波的初相位。因此，观测到的系统输出子波波形与输入子波波形是完全一致的。据此可知，相干信号的时差测量可以是精确的，相干信号可以实现完全的同相位叠加。

雷达系统、勘探地震系统很显然是近似的线性系统。因此，可以利用不同的数学算法实现相干信号的叠加，实现信号增强和成像，达到遥感探测的目的。

勘探地震数据的一个最重要的特点是针对地下同一反射点实行多次覆盖的观测，这当然是勘探地震独特的炮检观测系统实现的。为什么勘探地震中必须或必然会出现这样的观测系统？数据的低信噪比和层状介质下的速度估计与建模必须要有这样的观测系统。

勘探地震上百年的实践告诉我们，勘探地震系统基本上可以视为一个线性系统，来自地下同一反射点的、不同炮检对的反（散）射子波，相位特征基本一致，即便存在非弹性介质情形，它们之间的相位特征并没有显著的跃变。这就构成了地震波偏移成像（也包括反演成像）的坚实的物理基础。当前的地震地质解释和油藏识别很大程度上也依赖勘探地震系统是近似线性的物理现实。

对来自地下同一反射波点的、不同炮检对的、带限的反射子波的同相位叠加成为了勘探地震数据分析的核心问题。地震数据处理流程中的大部分处理步骤实质都是为了达成带限反射子波同相位叠加的目的。

勘探地震中同相轴上地震子波可以被表示为：$u(\boldsymbol{x},t)=A(\boldsymbol{x},t)e^{-i\phi(\boldsymbol{x},t)}$。其中，相位 $\phi(\boldsymbol{x},t)=\omega t(\boldsymbol{x})+\phi_0(\omega)$。影响同相位叠加的根本要素是炮点到反射点再到检波点的旅行时 $t(\boldsymbol{x})$，而旅行时主要取决于介质的背景速度 $v_b(\boldsymbol{x})$，背景速度分析与建模是勘探地震数据分析中最困难也是最重要的处理环节。各谐波成分的初相位 $\phi_0(\omega)$ 决定子波的形态。很显然背景速度是决定反射子波同相位叠加与否的主要因素；而各谐波成分的初相位 $\phi_0(\omega)$ 是次要因素。在当前高精度地震成像阶段，既要重视背景速度建模，也要关注子波的形态。尽管我们认为勘探地震系统基本是线性系统，但毕竟震源激发子波的空变、地下介质的粘弹性、细层叠合效应、散射效应都会导致反射子波间相位的不一致，最终影响同相位叠加，影响高精度成像。

勘探地震数据处理中存在两种典型的（近似）相干信号同相位叠加的方法。一种是 Beamforming 方法；另一种是各式各样的叠加成像和偏移成像方法。

Beamforming 是雷达数据处理的核心方法，勘探地震中称为线性 Radon 变换（更直接地称为倾斜叠加）。Beamforming 方法对应的物理过程可描述为：远场传来的波场被视为局部平面波，该局部平面波被面阵列检波器或线阵列检波器接收到，所有检波器的最优加权叠加形成 Beamforming 结果。勘探地震中，可以认为

是地下反射界面发出局部平面波，传播到地表面或线阵列检波器被接收到。通过 Beamforming 方法或线性 Radon 变换方法实现最佳的、同相位的叠加。

勘探地震中更为典型的同相叠加方法当然是各种成像方法，值得特别说明的是：静校正的目的也是为了实现同一反射点不同偏移距反射子波的同相位叠加。CMP 校正叠加、DMO 校正叠加、CRS 校正叠加、PSTM 及 CIG 道集叠加、PSDM 及 CIG 道集叠加、CIG 道集最优加权叠加。CMP 及 CRS 道集校正拉平视为水平层状介质的叠前偏移过程，拉平的 CMP 和 CRS 道集是 CIG 道集和 CIG 超道集。DMO 实质上是叠前部分偏移方法，与 NMO 结合（MZO=NMO+DMO）实现道集校正拉平，完成叠前全偏移给出 CIG 道集。基于成像的同相位叠加本质上都是把来自地下同一反射点的、不同炮检对的反（散）射子波拉平形成 CIG 道集，然后进行同相位叠加。很显然，背景速度的正确性起决定性作用。

决定子波形态的各谐波成分的初相位 $\phi_0(\omega)$ 并没有体现在偏移成像过程中。我们认为初相位信息是带限子波的更为细节的信息，很难从对地震波场的分析中获取。合理的做法应该是对基于背景速度产生的、“拉平”的 CIG 道集，进行基于子波形态的排齐与叠加。用这样的方式，间接地处理初相位不一致引起的子波形态不一致以及非同相位叠加问题。

3 勘探震源激发的带限子波

在广义遥感的电磁波探测、超声探测方法中，很难像地震勘探那样激发出明确的带限子波，带限子波作为系统的激励源，形成传播的波场，经历地下介质的反（散）射等过程后，被地表放置的检波器接收到。正是由于带限子波的存在，当然也对应于层状沉积为主的介质特征，勘探地震数据的特征被抽象为满足一定时距关系的、由带限子波表现出来的同相轴飘在不相干或随机噪音中。

勘探地震震源，典型地，包括海上的气枪震源和目前正在逐步商用化的海上可控震源、陆上的炸药震源和陆上可控震源。气枪震源和炸药震源都是脉冲源，激发出的带限子波的振幅谱（功率谱）特征还是比较明显的，即带限、存在一定的光滑性、存在优势频段、缺乏勘探地震所需求的低频和高频。可控震源也要特别设计扫描信号，通过与测量信号的互相关得到带限的相关子波。此时的相关子波已经没有了相位信息，假定是零相位的了！原则上，可控震源可以按客户定义产生所需求的振幅谱的带限子波，但是实际的可控震源难以产生具有客户期望的振幅的带限子波。从一个侧面，这也反映了设计一个机械系统或电子系统或机电系统产生期望的激发源子波是非常困难的。而基于波场的探测方法原则上都需要理想的带限子波作为激励源，接收器接收激励源产生的在系统中传播的到达系统外的波场，通过对这样的波场的分析达到广义遥感的目的。前已述及，正是检波器接收到的场中的带限子波携带了系统介质的信息和激励源函数的信息。这样的波场尤其适合在时间空间域中进行分析。

当前，勘探地震中，正在探索的连续的脉冲源激发以及直接对可控震源中单频谐波波场进行成像；其他遥感领域中众多单频谐波激发及谐波测量数据的直接处理，数学上当然是可行的。但对于测量数据的分析和后续的应用而言，并非是最佳的选择。譬如，在频率域噪音很难识别，频率域的数据非常不直观、不方便解释应用。勘探地震中，油气地质解释员直接对叠加剖面和成像剖面进行地质解释正是因为带限子波携带了地下介质的变化信息。

我们相信带限震源子波的概念也应在其他遥感领域中应用，无论激发出理想的震源子波是否困难。

由于激发介质和激发环境的不可控，勘探地震中，每次震源激发的带限子波都是不同的。谐波成分不同、每个对应谐波的振幅不同、相位也不同。最终的表现就是震源子波特征的空变导致接收到的波场中引入了这种不应有的变化。带限子波的这种空变之所以要消除，根本原因是带限子波携带了地下介质变化的信息，基于对这些信息的处理与分析达到对油气藏成像并识别油气藏的目的，这其中当然不能裹挟进激发震源与其周围介质相互作用的因素。

地震数据分析中，首先就是要消除这种由于激发介质和激发环境的不可控、不一致所导致的波场中带限子波的特征变化。我们能做的就是利用统计分析方法，感知到带限子波特征的异常变化，消除这种异常变化，保持带限子波的空间变化在趋势上的合理性。空间数据统计分析的主要目的就是把空间数据视为随机过程，对随机数据进行统计分析，得到其一阶统计量（均值、趋势的度量）、二阶统计量（偏离趋势的度量），据此就可以感知到空间随机数据的基本变化规律。地震数据中的各种一致性校正方法本质上就是依据上述

思想进行的。

4 带限子波的到达时及其测量

再次强调，正是测量波场（测量数据）中的带限子波携带了地下介质变化的特征。勘探地震数据分析中，最核心的两个处理步骤-偏移成像与背景速度建模-本质上都是利用了不同炮检对观测的波场中同相轴上带限子波的到达时信息。但是，带限子波的到达时该如何定义？这是勘探地震中的老问题，也是没有取得共识的问题。

带限子波在时空域应该具备的特征：时间或空间局部支撑性、正则性（光滑性、连续性）、震荡性（$\int u(\boldsymbol{x},t)dt=0$或$\int u(\boldsymbol{x},t)d\boldsymbol{x}=0$）；在频率域应该具备的特征：频带有限性（$\tilde{u}(\boldsymbol{x},\omega)=0$，当$|\omega|\geq\Omega$时）、谱正则性（光滑性、连续性）、存在优势频段。

物理上，一个波现象到达某一空间位置是由能量的到达为标识的。从波动的角度，就是某空间点上、某时刻开始有非零的震动能量。这也是把初至震动视为带限子波到达时的物理原因。但是，从地震数据分析和成像的角度看，这并非合理的定义。因为初至能量往往是很弱的，被淹没在噪声中，无法被有效地检测出来，这样的定义对初至拾取没有实质意义。

至于用带限子波的波峰或波谷作为带限子波的到达标识，由于带限子波间相位差异的存在，即带限子波形态差异的存在，也缺乏物理基础。在人工交互处理阶段，把带限子波的波峰或波谷作为带限子波的到达标识的确是方便的。这样的定义似乎缺乏数学意义。同时这样的定义受子波相位或形态的影响是显著的。

我们认为：对于带限子波$u(\boldsymbol{x},t)$，定义能量密度函数$P(t)=u_t^2(\boldsymbol{x})$，则带限子波到达时可以定义为$\hat{t}=\int tP(t)dt$。$\hat{t}$可以被称之为质心到达时。

图 1 展示了无噪、带限子波完全相同的 CMP 道集上的到达时拾取情况。从左往右，绿色圆点分别代表以初至能量、波峰、质心为到达时的拾取结果。这种情况下，基于带限子波波峰和质心的到达时估计结果基本一致，与基于初至能量的到达时拾取结果仅仅存在一个系统误差。图 2 展示了无噪、但带限子波存在相位变化（形态变化）的 CMP 道集上的到达时拾取情况。从左往右，绿色圆点代表图 1 中情形下的到达时拾取结果；红色圆点分别代表相位变化情况下以初至能量、波峰、质心为到达时的拾取结果。可以看出，带限子波的相位变化（形态变化）对质心到达时测量的影响是小的。因此，用质心到达时进行的测量结果是比较稳定的。当然如何孤立出带限子波，然后实施上述质心到达时的计算依然存在问题。这正说明了带限子波到达时估计的不确定性和困难程度。

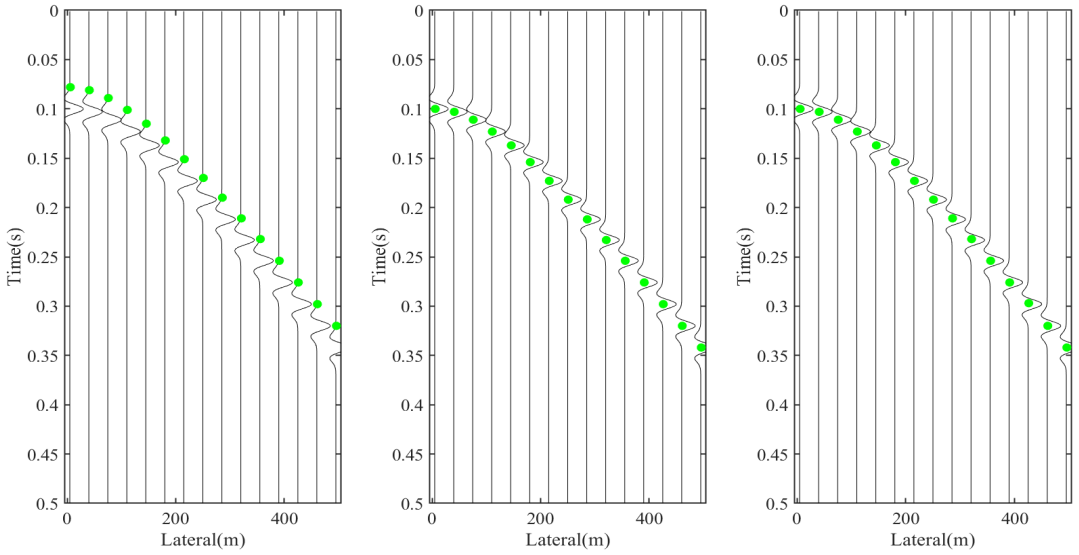


图 1 无噪、带限子波完全相同的 CMP 道集上的到达时拾取情况。从左往右，绿色圆点分别代表以初至能量、波峰、质心为到达时的拾取结果。

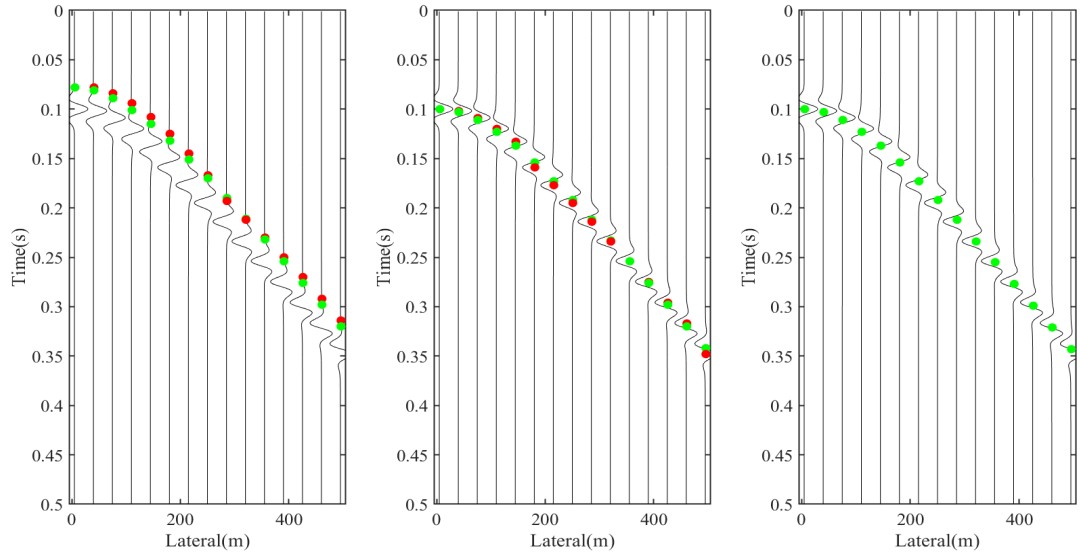


图 2 无噪、带限子波存在相位变化（形态变化）的 CMP 道集上的到达时拾取情况。从左往右，绿色圆点代表图 1 中情形下的到达时拾取结果；红色圆点分别代表相位变化情况下以初至能量、波峰、质心为到达时的拾取结果。

众所周知，由观测数据获取其内蕴的特征量是标准的反问题。我们认为，依据观测数据（波场）估计带限子波的到达时就是这样的一个反问题，因为实际数据中带限子波的到达时就是该子波的一个内蕴的特征量。当然，首先对要估计的特征有一个明确的数学物理定义是必须的。即便如此，基于某种原则定义的特征量未必适用于所有的情形。这种情况在复杂的数据分析中是很常见的。

5 带限子波的空间与时间变化

平稳随机信号被认为是谐波叠加形成的，可以通过 Fourier 分析、SSA 分析（或 PCA 分析）获取平稳随机信号的特征，对平稳随机信号进行有效的表达，进而对它进行各种既定的处理解决相关的应用问题。但是，一般的平稳随机信号中并没有带限子波的概念。

炸药和气枪脉冲源激发产生的带限子波是勘探地震中特有的概念。到目前为止，相关的遥感学科也没有充分地引入带限子波的概念。带限子波的引入使得勘探地震中有效信息的定义和辨识变得非常清晰，满足一定时距关系的、由带限子波标识的同相轴携带了地下岩性介质弹性参数变化的信息。我们可以称它们为特征波场，勘探地震成像本质上是特征波场的成像。

通信科学领域中，把编码后的有效信号调制到高频载波上，高频载波传播到目的地后，分离载波和有效信号，并对有效信号解码，实现通讯的目的。勘探地震中，有效信号就是满足一定时距关系的、由带限子波标识的同相轴。勘探地震中，没有明确的载波，地震波场既是载波也是信号，二者融为一体。勘探地震的目标与通讯科学是不同的。勘探震源激发的带限子波在地下介质中传播，与介质相互作用，形成各种时距关系决定的、带限子波表征的波现象。基于观测到的波现象进行地震波成像，获得地下介质结构分布和弹性参数分布的信息，从而解决油藏描述和开发的问题。

空间与时间变化的带限子波携带了地下介质变换的信息。

带限子波的到达时取决于地下介质的（各向异性）背景速度；带限子波的振幅取决于地下介质的波阻抗变化；带限子波的相位（形态）取决于震源及震源与围岩介质的相互作用、介质的非弹性、细层叠合效应、散射效应。勘探地震中的波以球面波传播为物理现实，振幅的球面扩散衰减也是带限子波振幅变化的重要形式。

勘探地震中，带限子波的到达时和介质波阻抗变化引起的振幅变化才是对地震波成像有贡献的信息。介质的非弹性效应引起的带限子波的振幅和相位变化；细层叠合效应引起的带限子波的振幅和相位变化；散射效应引起的带限子波的振幅和相位，理论上都可以被应用起来，探测地下介质的特征。但是，这是很难做到的。Q 值的估计很难剥离开细层叠合效应和散射效应。

震源及震源与围岩介质的相互作用引起不同炮检对之间带限子波的不一致。这种不一致也会严重影响成像的保真度。基于空间数据统计分析的方法把这种不一致统一到一个期望的趋势水平上。目前看来也只能如此！但是，这样做会潜在地引入对带限子波的人为改变。

球面波传播也是勘探地震的一大特点，很多相关的遥感学科都假设波是（局部）平面波传播的。众所周知，平面波传播时，波前面的振幅是不衰减的。勘探地震中，变速介质情况下球面波衰减的补偿也是要特别注意的事情。值得注意的是：波动理论地震波成像中球面波衰减是成像方法自动考虑的，不需要单独处理球面扩散引起的振幅衰减。但是，在整个复杂的处理流程中，球面扩散的合理处理要特别小心。

地质历史上沉积形成的层状介质是勘探地震介质的特征，带限子波在这样的介质中传播，透射波与反射波瓜分了入射波的能量，造成入射带限子波的振幅降低。但是，地震波成像被认为是线性化的成像，成像目标层之上的反射波很弱，认为可以忽略。

据上述分析可以看出，地震数据分析的主要对象就是带限子波。成像前要进行针对带限子波的保真的预处理；成像阶段目前核心工作依然是利用不同炮检对观测的带限子波到达时进行（各向异性）背景速度估计以及利用来自同一反射点、不同炮检对的反（散）射带限子波的同相叠加进行偏移成像。

6 带限子波的相位特征对保真偏移成像的影响

地震波偏移成像的数学本质是定位反射系数在地下介质的空间位置，所利用的信息是来自同一反射点、不同炮检对的反（散）射带限子波。对应的物理图景是：震源激发的带限子波沿着炮点到反射点的路径传播到反射界面，按 Snell 定理和角度反射系数在反射点处产生反射，检波器在地表观测到从反射点到检波点路径传来的带限子波。此物理图景清楚地展示了：如果仅仅考虑球面波传播、角度反射系数对带限子波的改造，假定其他因素不重要或可以被实现补偿掉，假定满足线性成像的假设，地震波偏移成像不仅仅可以定位反射系数出现的位置，成像结果也可以保持地下反射系数的相对关系（横向上振幅关系一致性和单个 CIG 道集上 AVA 关系的基本正确）。

但是，我们必须弄清楚是在什么假设下的保真成像才能在应用中不发生大的偏差。

一般地，Kirchhoff 积分叠前偏移成像是在（各向异性）背景速度场中计算炮点到反射点再到检波点的到达时，利用该旅行时把反射带限子波放在成像点上。很显然，构造正确的成像依赖于背景速度；保真成像

依赖于球面扩散校正、地表一致性振幅和相位校正、吸收衰减校正等。叠加成像的分辨率和保真度首先还是由（各向异性）背景速度决定的，也就是说来自地下同一反射点的、不同炮检距的带限反射子波的同相位叠加是首先要达到的。带限反射子波的相位（波形）是第二位的，但并非是不重要的。同一反射点不同炮检对带限反射子波波形差异过大，叠加成像结果的带限子波（成像子波也是带限子波！）的振幅和波形畸变会非常严重，根本无法表达地下介质的像，也无法进行后续的油藏描述。

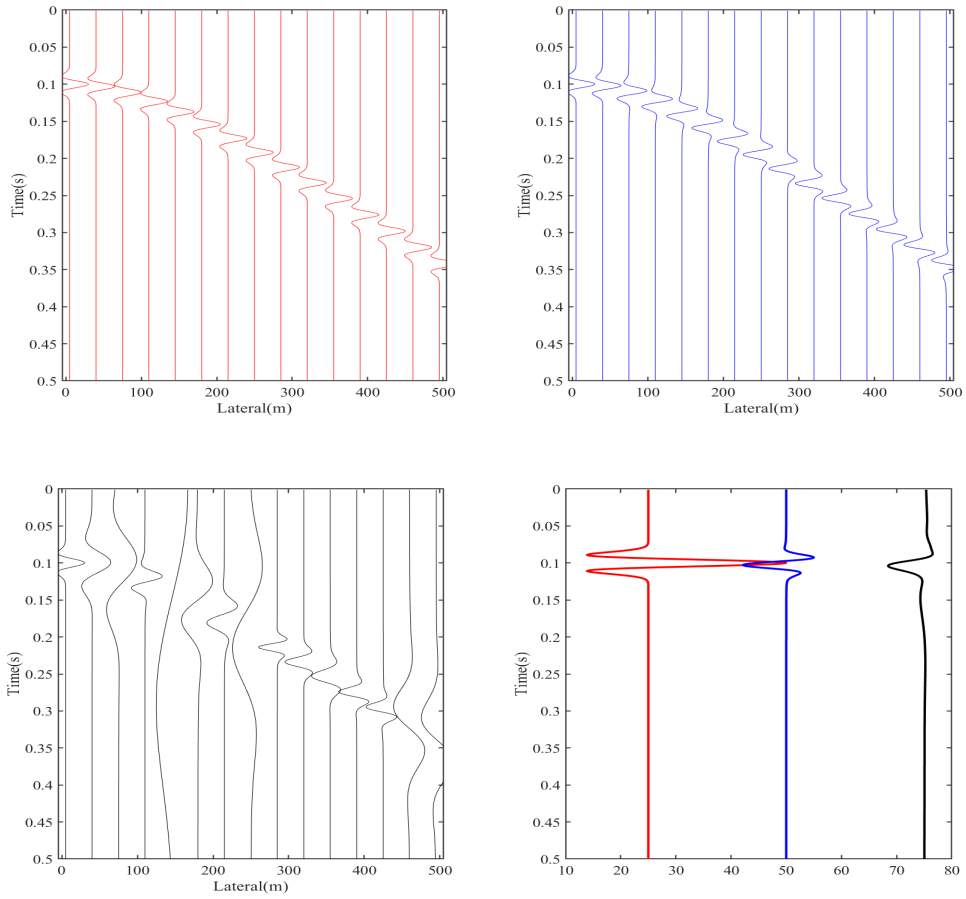


图 3 CMP 道集（或共成像点道集 CIG）中带限子波相位变化（形态变化）对叠加成像的影响

图 3 展示了 CMP 道集（也可以是共成像点道集 CIG）中带限子波相位变化（形态变化）导致的对保真成像的破坏作用。左上图中带限子波是相同的，右上图中带限子波存在常相位变化，左下图中带限子波相位变化非常大（设想的极端情况！），右下图是叠加结果，与红色曲线代表的同相位叠加相比，由于非同相叠加存在相削弱的情况，蓝色曲线所代表的常相位变化叠加的结果幅值较小，黑色曲线代表的随机相位剧烈扰动叠加结果更加失真。可以清楚地看出，带限子波相位变化（形态变化）对保真和高分辨成像的影响还是比较严重的，深入的数值分析能更系统地揭示出这种影响的细节。

幸运的是，勘探地震实践表明勘探地震系统基本上线性系统，震源激发的带限子波在地下介质中传播后，检波器接收的带限子波与震源激发的输入带限子波基本上是相干信号。如果这个物理现实不存在，勘探地震会变得异常困难！

而对波动理论的单向波逆深度外推偏移成像或逆时外推偏移成像，成像条件为：

$$I(\boldsymbol{x})=\int_0^T u_S(\boldsymbol{x},t)\cdot u_R(\boldsymbol{x},t)dt$$

其中，$u_S(\boldsymbol{x},t)\cdot u_R(\boldsymbol{x},t)$ 代表波场 $u_S(\boldsymbol{x},t)$ 与 $u_R(\boldsymbol{x},t)$ 在成像点 $\boldsymbol{x}$ 处的零延迟互相关。互相关是一个二阶统计量，互相关函数与互功率谱是一对 Fourier 变换，因此成像和梯度的计算可以在频率域中进行。频率域的

成像条件为：

$$I(\boldsymbol{x})=\sum_{\omega=\omega_{\min}}^{\omega_{\max}} u_S(\boldsymbol{x},\omega)\cdot u_R(\boldsymbol{x},\omega)$$

其中，$u_S(\boldsymbol{x},\omega)\cdot u_R(\boldsymbol{x},\omega)$ 代表波场 $u_S(\boldsymbol{x},t)$ 与 $u_R(\boldsymbol{x},t)$ 在成像点 $\boldsymbol{x}$ 处的互功率谱密度，求和结果就是偏移成像结果。由公式可以看出，仅仅取了正频率对应的互功率谱密度成分进行求和。

无论是时间域的成像条件或是频率域的成像条件，蕴含的物理意义都是：假设（各向异性）背景速度正确的条件下，传播到成像点处的正演带限子波与反传到成像点处的观测带限子波的零延迟互相关构成成像结果。

从高频近似的物理意义看，在成像点处反射带限子波 $w_R(\boldsymbol{x},t)$：

$$w_R(\boldsymbol{x},t)=r\left(\theta\right)w_S(\boldsymbol{x},t)$$

其中，$\boldsymbol{x}$ 代表成像空间中反射界面处的任一成像点，$w_S(\boldsymbol{x},t)$ 是入射带限子波，$r\left(\theta\right)$ 是张角为 θ 的 $\boldsymbol{x}$ 处的反射系数。

很显然，成像点 $\boldsymbol{x}$ 处的反射带限子波 $w_R(\boldsymbol{x},t)$ 与入射带限子波 $w_S(\boldsymbol{x},t)$ 的互相关结果为：

$$w_R(\boldsymbol{x},t)\cdot w_S(\boldsymbol{x},t)=r\left(\theta\right)w_S(\boldsymbol{x},t)\cdot w_S(\boldsymbol{x},t)$$

其中，$w_S(\boldsymbol{x},t)\cdot w_S(\boldsymbol{x},t)$ 是入射带限子波 $w_S(\boldsymbol{x},t)$ 的自相关。自相关函数的性质：它是一个正定对称的函数。看起来像一个零相位带限子波。但是，由于震源激发子波是空变的、未知的，即便假设地震勘探系统是线性系统，正演带限子波与入射带限子波也不是相干信号。互相关成像条件的成像结果就严重地依赖于正演带限子波的选择，当然也严重地依赖于（各向异性）背景速度的正确性。换句话说，互相关成像条件的保真成像结果依赖于相位函数 $\phi\left(\boldsymbol{x},t\right)=\omega t\left(\boldsymbol{x}\right)+\phi_0\left(\omega\right)$ 中的到达时 $t\left(\boldsymbol{x}\right)$ 和初相位 $\phi_0\left(\omega\right)$。

实际的保真成像方法技术本质上都是在处理到达时 $t\left(\boldsymbol{x}\right)$ 计算的准确性问题和对初相位 $\phi_0\left(\omega\right)$ 影响的消除问题（即带限子波的一致性问题）。到达时 $t\left(\boldsymbol{x}\right)$ 计算的准确性对保真成像影响是第一位的，因此（各向异性）速度建模在地震波成像中居于关键位置。成像条件的具体实现方法、地表一致性子波相位校正、吸收衰减补偿和成像道集后处理构成了保真成像过程中的其他几个关键技术环节。

7 带限子波的相位特征对层析成像的影响

地震波层析成像本质上是了利用炮点到检波点的传播路径多次沿不同方向穿过同一个要反演的速度单元体的带限子波的到达时信息。用 Radon 变换的语言讲，这个到达时是沿炮点到检波点的传播路径的路径积分（投影）结果。地震波层析成像的本质就是进行反投影，直接反投影就要把投影值按不同的权重放在整个路径上的速度单元体上。尽管这很简单，但已经把地震波成像与观测系统的关系、层析反演方法的基本思想都揭示清楚了。

检波器接收到的带限子波蕴含了勘探地震要估计的或要反演成像的介质信息。背景速度与带限子波的到达时密切相关；吸收衰减与带限子波的振幅衰减（带限子波振幅谱的质心频移）密切相关；保真成像与带限子波的振幅密切相关。

无论是带限子波的到达时定义、振幅衰减的定义、振幅值本身的定义都与其相位特征密切相关。众所周知，只有零相位的带限子波上述定义才可能是唯一的。否则就依据带限子波相位特征的变化而变化，没有唯一性。带限子波的到达时的定义、振幅衰减的定义和振幅值的定义的不唯一性也导致了反演成像结果的不唯一性，具体表现就是速度和 Q 值反演结果的精度（分辨率）降低，角度反射系数的估计结果和 AVA 分析结果的精度降低。

到目前为止，现代信号分析理论并没有提供关于相位分析的系统理论与方法。功率谱估计（包括时频谱估计）也是以提取信号的能量特征为主要目标。简化相位分析或降低相位影响的直观想法肯定是把带限子波中诸谐波成分的初相位都变成一样，即把带限子波调整成零相位子波。这样做使得带限子波的到达时的定义、振幅衰减的定义和振幅值的定义变唯一了，但并没有解决相位未知所导致的反演成像结果的不唯一。

8 结论与讨论

勘探地震中在以层状介质为主的、仅仅地表观测的勘探地震数据情形下，炮集中包含的波场可以认为是各种复杂的波现象飘在相干或随机的噪音中。更具体的，是满足各种时距关系的带限地震子波携带了关于地下介质和震源的信息。地震勘探中地震波成像就是利用不同炮检对观测的带限子波的到达时信息估计背景速度、利用带限子波的振幅信息估计（角度）反射系数。更精细的速度估计会用到同相轴上带限子波的相位信息；吸收衰减参数 Q 值的估计会用到带限子波的振幅衰减信息。

地震波成像的物理基础是地下介质系统基本是一个线性系统，震源激发的带限子波经过从炮点到检波点的传播、经过与地下介质相互作用的改造，被检波器接收到的输出带限子波近似是一个相干信号，因此来自地下同一点的、不同炮检对的反（散）射带限子波可以进行同相位叠加。静校正、NMO、NMO+DMO、CRS、PSTM 和 PSDM 的根本目的就是实现来自地下同一点的、不同炮检对的反（散）射带限子波的同相位叠加。这是地震波成像的核心问题。地震波成像中的同相位叠加是基于（各向异性）背景速度的，这是地震波成像的独特之处。正因为如此，（各向异性）背景速度估计与建模在地震波成像中具有特别重要的地位。同相叠加的另一个重要因素是带限子波的相位（波形）。尽管与（各向异性）背景速度相比，它是影响同相叠加的第二位因素，但不是不重要的因素。另一方面，同相位叠加也是高分辨率和高保真成像的基础。据此可以理解，高精度的（各向异性）速度建模、地表一致性相位处理、静校正、吸收衰减校正、成像道集后处理构成了高分辨率和高保真成像关键技术。

到目前为止，在高精度成像阶段，带限子波相位对成像结果的不唯一性或精度的影响并未引起足够的重视。对带限子波的研究应该再一次引起重视。科学和技术的进步就是这种螺旋上升的。

参考文献

[1]. Robinson, E. A., Geophysical Signal Analysis, SEG, 2000.

[2]. Robinson, E. A., and Treitel, S., Digital Imaging and Deconvolution: The ABCs of Seismic Exploration and Processing, SEG, 2008

[3]. 张贤达，现代信号处理（第三版），清华大学出版社，2015 年

勘探地震中强非线性反问题的解法

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】理论上讲，勘探地震领域中存在的具有引领性的研究议题应该是：强非线性反问题的求解。无论在强噪音和复杂信号特征下的信号建模问题、强噪音和复杂图像特征下的图像智能识别问题以及强噪音和复杂介质情况下的参数估计问题，都归结于强非线性反问题的求解。反问题的基本定位是：它总是一个不充分可观测信息下的不可观测量的估计问题。反问题的两个基本逻辑：不可观测量与可观测量之间存在线性或非线性的关系；可观测量总是含噪的，统计信息不完备的，不足以准确地估计不可观测量，需要补充关于不可观测量的先验信息。解决（强）非线性反问题，目前根本没有既定的理论框架。只有凸问题，我们可以确认有一个唯一的解，解的意义也是清楚的。目前，牛顿迭代优化算法（梯度导引类的算法）都是针对光滑凸问题的算法。与其无望地求解强非线性反问题，不如研究如何把强非线性问题提成合适的凸问题进行求解。我认为（至少基于目前的认识认为）：这是唯一正确的求解强非线性反问题的逻辑。对正问题进行深入研究，可以使我们更深入地理解参数变化与波场变化之间的关系，基于此可提出更合理的凸问题。但是，深入地研究正问题，并非意味着用更复杂的状态方程构建参数估计反问题。为了构建凸的反问题，提出更合适的参数与波场之间的关系式（算子），稳定地求解反问题才是更有意义的工作。所谓更合适的参数与波场之间的关系最好是一对一的因果关系。还要定义满足同因之果原则的差异度量方式（误差泛函）。我认为上述观点反映了求解强非线性反问题的本质。

【关键词】强非线性反问题，凸问题，可观测量，不可观测量，统计信息不充分，先验信息，状态方程，映射关系

1 引言

油气地震勘探发展到今天，地震数据分析和地震波成像过程中依然存在的真正没有得到解决的问题是：强非线性反问题的求解。

无论在强噪音和复杂信号特征下的信号建模问题、强噪音和复杂图像特征下的图像智能识别问题以及强噪音和复杂介质情况下的参数估计问题，都归结于强非线性反问题的求解。这三类问题恰恰是地震数据分析和地震波成像中最迫切希望解决的问题。

扩展一点看，大数据分析时代，复杂情况下的分类问题（包括聚类问题）也是个强非线性问题。

勘探地震学发展到现在，其中的核心问题逐渐在聚焦，就是聚焦到把三维油气探区视为一个弹性介质系统，人工震源作为激励该系统的输入，地表观测的波场作为该系统的输出（理论上讲，应该在三维油气探区的所有边界上布设地震波接收器，才能采集到系统的完整输出。很显然，油气地震勘探中，仅仅在地表（或/和井中）观测了系统的部分输出），该实际物理系统被认为可以由人为建立的数学物理方程来描述，然后在上述正过程描述的基础上，基于 Bayes 估计理论对实际物理系统的弹性参数（纵波速度、横波速度和密度）进行估计。在弹性参数估计结果的基础上，结合岩石物理关系，进行岩石物性参数的估计，进而对储层含油气性进行评价，最后进入油气开发阶段。这是当今油气地震勘探的基本逻辑。

据此，油气地震勘探的核心问题可以描述为：把地表（包括井中）观测的波场（叠前数据）视为随机信号，已知一定的数学物理方程可以模拟（预测或解释）实测波场中的（部分）波现象，预测误差满足一定的概率分布（往往是 Gauss 分布），假设有来自井中或其他来源的关于要反演参数的先验信息，也把该先验信息视为随机的（实质上这仅仅反映了先验信息的不确定性），基于 Bayes 公式获得要估计参数满足的后验概率密度函数，后验概率密度函数最大时对应的参数估计结果就被认为是地震波反演问题的解（MAP 原则）。

很显然，仅仅在地表（或/和井中）观测了实际物理系统的部分输出，而且观测是孔径受限的和频带受限的，采样是稀疏的（不满足 Shannon 采样定理），采样往往是不规则的，数据中包含很多非激励源引发的噪音，表明实际采集的地震数据非常不完备和不完美。地表高程（剧烈）变化和地表及地下介质（剧烈）空变使得要反演的弹性参数场包含了宽波数带成分。依据勘探尺度，它们从公分级别、米的级别、几十米的级

别、几百米的级别到数千米和数十千米的级别之间变化，地震波在如此复杂变化的介质中传播，使得实测波场变化与不同尺度的介质弹性参数变化之间的关系是（高度）非线性的。另一方面，选择什么样的正问题、如何进行要反演参数的选择和描述，都严重影响所提出的反问题的非线性性。描述实际地震勘探系统的数学物理方程既不能过于简单，又不能过于复杂，简单的控制方程不能很好地预测实际观测波场；复杂的控制方程包含过多的未知参数。对要反演参数的选择和对它（它们）的描述方法很大程度上决定了介质系统参数反演问题求解的困难程度。从上述描述的系统的观点看地震波反演问题，有两个参数要反演：每炮的激励函数（震源子波）和介质系统的弹性参数。如果同时反演这两组参数，地震波反演成像的非线性性会显著增加，因此实际的地震波反演成像过程中，总假设每炮的激励函数（激发子波）是一致的和已知的。实际上，陆上地震勘探中每炮的震源子波显然是未知的，而且空变很明显。正是复杂的数据情况以及复杂的介质情况导致了勘探地震中地震波反演成像的高度非线性性。

地震数据分析中，地震数据的预处理和成像结果的图像识别，前者是一个老问题，后者是大数据和 AI 时代的新问题，在复杂情形下，都具备高度的非线性性。

到目前为止，我们仅仅对线性高斯情形下的反问题有完整的求解理论和方法，对反问题的解有明确的理论评价。对于非凸的、强非线性的反问题没有统一的、有效的理论框架。只有凸问题，我们可以确认有一个唯一的解，解的意义也算是清楚的。目前，牛顿迭代优化算法（梯度导引类的算法）都是针对光滑的凸问题的。因此，与其无望地求解非凸的、强非线性的反问题，不如研究如何把强非线性问题提出合适的凸问题进行求解。

仔细分析勘探地震学中地震波成像的发展历史也可以领悟到：尽管地震波反演成像是一个强非线性问题，但是勘探地震学家并没有直接求解它，而是引入不同波现象的时距关系、高频近似下的波传播关系、常密度声波方程、Zeoppritz 方程等正问题构建一系列反问题的组合来实现（较）强非线性反问题的求解。几十年来的勘探实践表明这样的方法技术流程是行之有效的。

本文从反问题的提法、线性反问题的解法和当前非线性反问题解法的讨论入手，提出勘探地震中求解强非线性反问题的路线和方法。

2 非线性问题的根源与反问题的提法

前已述及，凡涉及到利用可观测量来估计或推演与可观测量有关的不可观测量的问题都可认为是一个广义的反问题；通过直接测量到的系统输出（数据），基于描述系统的状态方程，估计系统参数和系统激励的问题称为狭义的反问题。

具体到勘探地震，三维探区形成一个三维地下介质体，震源激发能量下传，遇到反射界面或散射体或绕射体，地震波发生反射或散射或绕射，反传回地表，被检波器接收。震源、由弹性参数描述的三维地质体和接收到的波场构成一个系统。各种形式的波动方程作为描述该系统的状态方程。反问题是基于在地表（井中）观测的数据（波场）要么估计介质系统的弹性参数，要么估计震源函数，也可以二者同时估计。

推广一下，线性或非线性同相轴飘在满足各种概率分布的随机噪音中构成勘探地震信号或图像的基本模型。由观测的地震数据或已知的地震图像估计其中蕴含的感兴趣的信号或模式，尤其在低信噪比、和/或不规则采样、和/或有假频采样情况下，也是一个典型的、困难的反问题。此时，不可观测量（数据中的信号或某种潜变量）与可观测量（实测数据）之间存在的关系也是要估计的。这是一个更为困难的盲反演问题。

无论是勘探地震系统的参数估计问题、实测数据中的信号估计问题或是成像结果中的图像模式智能识别问题，在复杂情形下都是强非线性问题。

反演的根本逻辑是：不可观测量与可观测量之间存在线性或非线性的关系。定量地表达它们之间的关系应该有两条典型的途径：一是用二者之间的统计规律；二是用二者之间的函数关系。前者多用于大数据分析中；后者在科学和工程数据分析中更为常用。当然，函数关系会有线性的和非线性的两种。

由于可观测量总是有限的（不完备的）、而且是受“噪音”污染的，不可能指望从可观测量中获取不可观测量的完备信息，因此，需要补充关于不可观测量的先验信息。这是反演的第二个根本逻辑。因此，我给反问题的基本定位是：它总是一个不充分可观测信息下的不可观测量的估计问题。

基于上述认识，假定不可观测量与可观测量之间存在线性或非线性的函数关系，在反问题的提法中，考虑两个方面的问题就是合理的。一是基于已知的函数关系对可观测量进行预测时的预测误差，预测误差小说明基于有限和不完美的可观测量对不可观测量的估计就是有意义的；二是需要增加（部分）先验信息降低仅仅考虑用不可观测量与可观测量之间的关系所带来的反演解的不确定性。

对可观测量的预测结果的不确定性实质上应该包括：预测关系本身的不确定性；可观测量中的噪音带来的不确定性。我认为还应该把不规则观测、不充分采样、有限孔径和有限频带观测统在对可观测量的预测结果的不确定性的考量中，它们对最终反演解的不确定性（实质上就是反演解的精度）有很大的影响。我认为预测关系本身的不确定应该包括系统参数的选择、系统状态方程的选择、系统参数的表达方法的选择，因为这三者对最终反演解的不确定也有重要的影响。

本质上讲，反问题就是一个决策问题。其核心当然是制定一个合理的决策依据。无论是 Bayes 估计、最大似然估计、无偏及方差最小估计、最小熵或最大熵估计等等都是某种决策依据。各种不同的反问题就是在上述决策依据下突出的。我们当然期望提出的反问题要是凸的、求解效率高的、能适应非高斯/非平稳/非线性情形的。事实上，这是很难兼顾的！

客观地讲，Bayes 估计有最高的抽象性和形式美。但是，如果不对先验概率分布函数施加一定的先验假设（譬如 Gauss 分布假设），Bayes 估计很难落地应用。

总之，给出评价反演解不确定性的定量度量方式是给出合理的反问题提法的根本！具体的各种反问题的提法，可以参考王华忠文集-Bayes 估计观点下的地震信号分析与地震波反演成像（2019），我们认为：系统响应→差异度量→反馈（误差反传播）→调整→达到最优是自然（进化）优化的基本逻辑，鉴于正问题具有局部因果性，反问题没有局部因果性，因此给出满足同因之果原则的差异度量方式是需要遵守的原则。

3 线性反问题的解法

勘探地震中，到目前为止，无论是参数估计或是信号分析，核心问题还是求解广义线性反问题。譬如参数估计中的层析成像与最小二乘偏移成像；信号分析中的信号建模与去噪和数据规则化等问题。

关于广义线性反问题的求解，本质上，要么是欠定方程组要么是超定（不满秩）方程组的求解。这是从线性方程组求解的角度看这个问题。因为任何线性系统的输入与输出之间的关系总可以归结为 Fredholm 第一类积分方程，它总可以退化为线性方程组（矩阵方程），所以线性方程组的求解具有十分广泛的应用场景。一般地，大规模矩阵方程的求解在变分原理的意义下进行。

3.1 线性预测误差泛函的几何性能及光滑凸问题的解的性质

假设线性系统对应的矩阵方程为 $\boldsymbol{Ax}=\boldsymbol{b}$，其中系数矩阵 $\boldsymbol{A}$ 是满秩的，最好是正定对称的。实质上，系数矩阵 A 也是一个线性投影算子。它把模型向量 $\boldsymbol{x}$ 投影成数据向量 $\boldsymbol{b}$。广义线性反演的另一个本质描述是寻找线性投影算子的逆算子或广义逆算子，对数据向量 $\boldsymbol{b}$ 进行反投影，估计模型向量 $\boldsymbol{x}$。实际情况下数据总是包含噪音的，此时 $\boldsymbol{Ax}\approx\boldsymbol{b}\left(\hat{\boldsymbol{b}}+\boldsymbol{\eta}\right)$。在变分原理意义下，定义如下的误差能量，

$$
\begin{aligned}
&\mathcal{J}(\boldsymbol{x})=(A\boldsymbol{x}-\boldsymbol{b})^T(A\boldsymbol{x}-\boldsymbol{b})\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A\boldsymbol{x})^T\boldsymbol{b}-\boldsymbol{b}^T(A\boldsymbol{x})+(A\boldsymbol{x})^T(A\boldsymbol{x})\\
&=\boldsymbol{b}^T\boldsymbol{b}-\boldsymbol{x}^TA^T\boldsymbol{b}-(A^T\boldsymbol{b})^T\boldsymbol{x}+\boldsymbol{x}^TA^TA\boldsymbol{x}\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})+(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})-\boldsymbol{x}^TA^T\boldsymbol{b}-(A^T\boldsymbol{b})^T\boldsymbol{x}+\boldsymbol{x}^TA^TA\boldsymbol{x}\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})+\left[(A^T\boldsymbol{b})^T(A^TA)^{-1}-\boldsymbol{x}^T\right](A^T\boldsymbol{b})-\left[(A^T\boldsymbol{b})^T-\boldsymbol{x}^TA^TA\right]\boldsymbol{x}\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})+\left[(A^T\boldsymbol{b})^T(A^TA)^{-1}-\boldsymbol{x}^T\right](A^T\boldsymbol{b})-\left[(A^T\boldsymbol{b})^T(A^TA)^{-1}-\boldsymbol{x}^T\right](A^TA)\boldsymbol{x}\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})+\left[(A^T\boldsymbol{b})^T(A^TA)^{-1}-\boldsymbol{x}^T\right]\left[(A^T\boldsymbol{b})-(A^TA)\boldsymbol{x}\right]\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})+\left[(A^T\boldsymbol{b})^T(A^TA)^{-1}-\boldsymbol{x}^T\right]A^TA\left[(A^TA)^{-1}(A^T\boldsymbol{b})-\boldsymbol{x}\right]\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})+\left[(A^TA)^{-1}(A^T\boldsymbol{b})-\boldsymbol{x}\right]^TA^TA\left[(A^TA)^{-1}(A^T\boldsymbol{b})-\boldsymbol{x}\right]\\
&=\boldsymbol{b}^T\boldsymbol{b}-(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})+(\boldsymbol{x}_o-\boldsymbol{x})^TA^TA(\boldsymbol{x}_o-\boldsymbol{x})\qquad \boldsymbol{x}_o=(A^TA)^{-1}(A^T\boldsymbol{b})\\
&=\boldsymbol{b}^T\boldsymbol{b}-\hat{\boldsymbol{b}}^T\hat{\boldsymbol{b}}+(\boldsymbol{x}_o-\boldsymbol{x})^TA^TA(\boldsymbol{x}_o-\boldsymbol{x})\\
&=\mathcal{J}_{\min}(\boldsymbol{x}_o)+(\boldsymbol{x}_o-\boldsymbol{x})^TA^TA(\boldsymbol{x}_o-\boldsymbol{x})
\end{aligned}
\tag{1}
$$

其中，$\mathcal{J}_{\min}(\boldsymbol{x}_o)$ 代表达到最佳解 $\boldsymbol{x}_o$ 时的误差能量，$\mathcal{J}_{\min}(\boldsymbol{x}_o)=\|\boldsymbol{\eta}\|^2$。为了公式推导的完整性，把如下公式列出，

$$
\begin{aligned}
&\hat{\boldsymbol{b}}=A\boldsymbol{x}_o\\
&\hat{\boldsymbol{b}}^T\hat{\boldsymbol{b}}=(A\boldsymbol{x}_o)^T(A\boldsymbol{x}_o)=\boldsymbol{x}_o^TA^TA\boldsymbol{x}_o\\
&=(A^TA\boldsymbol{x}_o)^T\boldsymbol{x}_o=(A^T\boldsymbol{b})^T\boldsymbol{x}_o\\
&=(A^T\boldsymbol{b})^T(A^TA)^{-1}(A^T\boldsymbol{b})
\end{aligned}
\tag{2}
$$

由矩阵论知，A^TA 具有对称正定性，存在 $A^TA=Q^T\Sigma Q$ 的正交分解，正交矩阵 Q 的列向量是特征向量，对角矩阵 Σ 中元素是与正交阵 Q 的列向量对应的特征值。因此，（1）式可以继续改写为，

$$
\begin{aligned}
&\mathcal{J}(\boldsymbol{x})=\mathcal{J}_{\min}(\boldsymbol{x}_o)+(\boldsymbol{x}_o-\boldsymbol{x})^TA^TA(\boldsymbol{x}_o-\boldsymbol{x})\\
&=\mathcal{J}_{\min}(\boldsymbol{x}_o)+(\boldsymbol{x}_o-\boldsymbol{x})^TQ^T\Sigma Q(\boldsymbol{x}_o-\boldsymbol{x})\\
&=\mathcal{J}_{\min}(\boldsymbol{x}_o)+\left(Q(\boldsymbol{x}_o-\boldsymbol{x})\right)^T\Sigma Q(\boldsymbol{x}_o-\boldsymbol{x})\\
&=\mathcal{J}_{\min}(\boldsymbol{x}_o)+\boldsymbol{q}^T\Gamma\boldsymbol{q}\qquad \boldsymbol{q}=Q(\boldsymbol{x}_o-\boldsymbol{x})\\
&=\mathcal{J}_{\min}(\boldsymbol{x}_o)+\sum_{k=1}^{M}\lambda_k|q_k|^2\\
&=\mathcal{J}_{\min}(\boldsymbol{x}_o)+\sum_{k=1}^{M}\frac{|q_k|^2}{1/\lambda_k}
\end{aligned}
\tag{3}
$$

其中，$\boldsymbol{q}=Q\left(\boldsymbol{x}_o-\boldsymbol{x}\right)$代表解的差向量$\boldsymbol{x}_o-\boldsymbol{x}$在正交坐标系$Q$中的投影。（3）式的几何形态可以抽象地用下图表示。

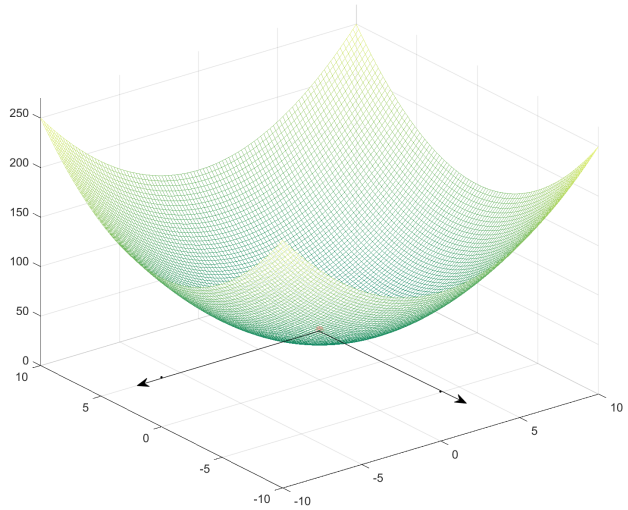


图 1 线性预测误差泛函的几何性能表面

图 1 清楚地展示出了广义线性反问题的本质。对于广义线性反演（线性正问题+Gauss 预测误差），线性预测误差泛函的几何形态是一个（超）椭球体。即关于线性反问题是一个凸问题。当系数矩阵 A 是满秩的，广义线性反问题存在唯一的解 $\boldsymbol{x}_o$。此时预测误差能量达到最低 $\mathcal{J}_{\min}\left(\boldsymbol{x}_o\right)=\left\|\boldsymbol{\eta}\right\|^2$。线性预测误差泛函的几何形态与系数矩阵 A 的条件数密切相关，系数矩阵 A 的条件数越大，（超）椭球体越扁长。迭代法求解广义线性反问题时收敛越慢，解的精度越低。

假若系数矩阵 A 不是满秩的，广义线性反问题不存在唯一的解。此时需要引入关于解 $\boldsymbol{x}$ 的先验信息。此时，最小二乘意义下的线性预测误差泛函被定义为，

$$\mathcal{J}\left(\boldsymbol{x}\right)=\left(A\boldsymbol{x}-\boldsymbol{b}\right)^T\left(A\boldsymbol{x}-\boldsymbol{b}\right)+\lambda\left\|\boldsymbol{x}\right\|^2 \tag{4}$$

（4）式定义的最小二乘意义下的广义线性反问题的解的含义为：

$$\hat{\boldsymbol{x}}=\min\left\|\overline{\boldsymbol{x}}\right\|^2 \quad S.T.\,\overline{\boldsymbol{x}}=\min\mathcal{J}\left(\boldsymbol{x}\right)=\min\left\{\left(A\boldsymbol{x}-\boldsymbol{b}\right)^T\left(A\boldsymbol{x}-\boldsymbol{b}\right)\right\} \tag{5}$$

（5）式的含义是：从无穷多的可行解 $\overline{\boldsymbol{x}}$ 中选一个能量最低的解，作为最小二乘意义下的广义线性反问题的解。要求解 $\boldsymbol{x}$ 是能量最低的就是引入的先验信息。具有能量最低属性的解就是光滑的解。据此，解决了解的唯一性问题。

统计学家给出的最小二乘意义下广义线性反问题解的解释是：它是后验概率密度函数为 Gauss 分布假设下的条件期望值。同时 Gauss-Markov 定理给出了解的无偏和方差最小的条件。

广泛应用的广义线性反问题的解的数学含义就是这样了。

3.2 线性问题的几何认知

关于线性方程组的几何认知，我们把矩阵方程 $A\boldsymbol{x}=\boldsymbol{b}$ 写成，

$$\boldsymbol{a}_1x_1+\boldsymbol{a}_2x_2+\cdots+\boldsymbol{a}_nx_n=\boldsymbol{b} \tag{6}$$

实际上，无论欠定、超定方程组都可以有上述表示。（6）式还可以写成，

$$A_{\mathrm{n\times m}}\boldsymbol{x}_{m\times 1}=\boldsymbol{b}_{n\times 1}\Rightarrow\sum_{j=1}^{m}\boldsymbol{a}_j x_j=\boldsymbol{b} \tag{7}$$

（6）和（7）式的含义可以解释为：数据向量 $\boldsymbol{b}$ 投影到由系数矩阵列向量形成的基函数空间中，各基函数上的投影值为解向量的对应分量。即线性方程组可表示为下图，

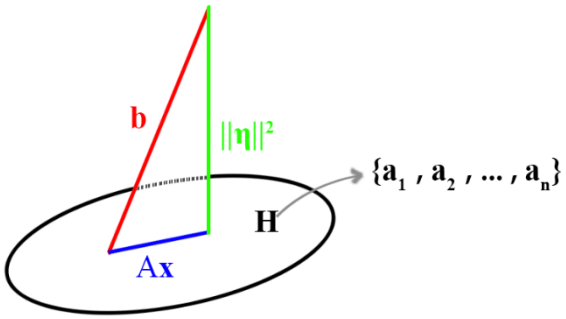


图 2 线性方程组的几何表达

进一步地，我们还可以对线性系统做更深入的解释。$\boldsymbol{b}=A\boldsymbol{x}=AS\boldsymbol{y}=C\boldsymbol{y}=\boldsymbol{c}_1y_1+\boldsymbol{c}_2y_2+\cdots+\boldsymbol{c}_ny_n$，此时 $\boldsymbol{c}_{\mathrm{j}},j=1,2,3,\cdots,n$ 为“特征向量”，用特征向量来表达数据向量 $\boldsymbol{b}$，期望投影系数 $\left(y_1,y_2,\cdots,y_n\right)$ 的分布有下图所示的特征，

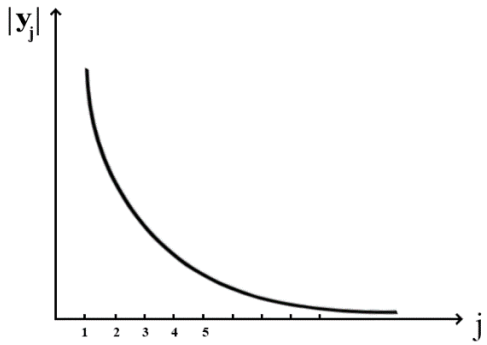


图 3“特征值”期望分布示意图

此时，仅用几个大的、或重要的“特征向量”的组合就可以很好地解释或预测数据向量 $\boldsymbol{b}$。求解新的线性问题 $\boldsymbol{b}=C\boldsymbol{y}$ 构成的广义线性反问题，收敛性和稳定性会有显著的改善。

3.3 广义线性反问题的迭代解法

一般地，数值数学仅重点讨论 $A\boldsymbol{x}=\boldsymbol{b}$、$A$ 适定且非奇异情况下的求解，存在唯一解，不需引入对解的正则化约束。除系数矩阵 A 为三对角阵或窄对角阵时可以用 LU 分解法求解 $A\boldsymbol{x}=\boldsymbol{b}$ 外，大规模的稀疏阵只能用迭代法求解。另外，LU 分解方法仅适用于系数非奇异的矩阵方程。

3.3.1 广义线性反问题直接解法的简述

关于广义线性反问题的直接解法，做一个简述。所谓直接解法就是对 $A\boldsymbol{x}=\boldsymbol{b}$ 进行不影响解的变换。A 非奇异时，总存在相似变换 P，P 也是非奇异的。为简单计算，一般取 P 为正交阵，正交阵是最简单的非奇异阵。也可以取高斯初等变换阵 G。这两种变换（实际上是坐标变换），作用于 $A\boldsymbol{x}=\boldsymbol{b}$，可以把 $A\boldsymbol{x}=\boldsymbol{b}$

变成 $\boldsymbol{Ix}=\boldsymbol{b}'$。当然也可以变成上三角阵 $U_U\boldsymbol{x}=\boldsymbol{b}'$ 或者下三角阵 $U_D\boldsymbol{x}=\boldsymbol{b}'$。据 SVD 分解当然可以解方程组，Householder 变换也可把 $A\boldsymbol{x}=\boldsymbol{b}$ 化为 $\sum\boldsymbol{x}=\boldsymbol{b}'$ 或 $\boldsymbol{Ix}=\boldsymbol{b}'$ 来解方程组。

线性系统 $A\boldsymbol{x}=\boldsymbol{b}$ 更常用的解法是迭代法。此时表面上不再假设系数矩阵 A 是非奇异的，也不再严格要求它是方阵。

3.3.2 广义线性反问题迭代解法的对比

从不同的角度可以引入线性系统 $A\boldsymbol{x}=\boldsymbol{b}$ 的迭代解法。

（1）设有 $\boldsymbol{x}_0, A\boldsymbol{x}-\boldsymbol{b}=\delta\boldsymbol{b}_1$，且系数矩阵 A 的条件数不是很大，有

$$A\left(\boldsymbol{x}_0+\delta\boldsymbol{x}_1\right)-\boldsymbol{b}=0$$

$$A\boldsymbol{x}_0+A\delta\boldsymbol{x}_1-\boldsymbol{b}=0$$

$$A\delta\boldsymbol{x}_1=\boldsymbol{b}-A\boldsymbol{x}_0=-\delta\boldsymbol{b}_1$$

$$\delta\boldsymbol{x}_1=-A\delta\boldsymbol{b}_1\Rightarrow\boldsymbol{x}_1=\boldsymbol{x}_0+\delta\boldsymbol{x}_1$$

$$A\boldsymbol{x}_1-\boldsymbol{b}=\delta\boldsymbol{b}_2$$

$$A\left(\boldsymbol{x}_1+\delta\boldsymbol{x}_2\right)-\boldsymbol{b}=0$$

$$A\boldsymbol{x}_1+A\delta\boldsymbol{x}_2-\boldsymbol{b}=0$$

$$A\delta\boldsymbol{x}_2=\boldsymbol{b}-A\boldsymbol{x}_1=-\delta\boldsymbol{b}_2$$

$$\delta\boldsymbol{x}_2=A^T\left(\boldsymbol{b}-A\boldsymbol{x}_1\right)=-A^T\delta\boldsymbol{b}_2\Rightarrow\boldsymbol{x}_3=\boldsymbol{x}_2+\delta\boldsymbol{x}_2$$

$$\cdots$$

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\delta\boldsymbol{x}^k=\boldsymbol{x}^k-A^T\left(\boldsymbol{b}-A\boldsymbol{x}^k\right)\tag{8}$$

（2）令 $A=M-N$，$A\boldsymbol{x}=\boldsymbol{b}$，则有

$$\left(M-N\right)\boldsymbol{x}=\boldsymbol{b}$$

希望矩阵的 M 逆存在，且形式简单易求解。上式变为

$$M\boldsymbol{x}=N\boldsymbol{x}+\boldsymbol{b}$$

$$\boldsymbol{x}=M^{-1}N\boldsymbol{x}+M^{-1}\boldsymbol{b}=M^{-1}\left(M-A\right)\boldsymbol{x}+M^{-1}\boldsymbol{b}=\left(I-M^{-1}A\right)\boldsymbol{x}+M^{-1}\boldsymbol{b}=\boldsymbol{Ix}+M^{-1}\left(\boldsymbol{b}-A\boldsymbol{x}\right)$$

进一步地，上式可转为如下的迭代形式：

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+M^{-1}\left(\boldsymbol{b}-A\boldsymbol{x}^k\right)\tag{9}$$

这是线性系统基本迭代法的基础公式。一般地，系数矩阵被分解为 $A=D-U-L$，其中，D：对角阵，U：上三角阵，L：下三角阵。当 $M=D$，$N=-U-L$ 时，形成 Jacobi 迭代格式；当 $M=D-L$，$N=-U$ 时，形成 Gauss–Seidel 迭代格式；当 $M=\dfrac{1}{w}\left(D-wL\right)$ 时，形成逐次超松弛迭代法。

对比（8）和（9），可以比较清楚地看出矩阵 M 的作用。

（3）定义反问题线性预测误差泛函 $J(\boldsymbol{x})=\frac{1}{2}(A\boldsymbol{x}-\boldsymbol{b})^T(A\boldsymbol{x}-\boldsymbol{b})$。假设 A 是满秩的、非奇异的，最好是对称正定的。这样的要求保证如下的迭代格式得到线性系统的唯一解。否者只能得到近似解。事实上，即便 A 是满秩的、非奇异的，条件数较大时，迭代法也很难得到精确解。这是所有迭代法都存在的问题。最速下降迭代格式：

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\mu^k\frac{\partial J}{\partial \boldsymbol{x}}$$

$$\frac{\partial J}{\partial \boldsymbol{x}}=-A^T(\boldsymbol{b}-A\boldsymbol{x})$$

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k-\mu^k A^T(\boldsymbol{b}-A\boldsymbol{x}) \tag{10}$$

μ^k 是最速下降要求引入的步长，令 $\varphi(\boldsymbol{x}^{(k)}+\mu\boldsymbol{r}^{(k)})$，有

$$\frac{\partial\varphi}{\partial\boldsymbol{x}}=0\ \rightarrow\ \mu_k=\frac{(\boldsymbol{r}^{(k)},\boldsymbol{r}^{(k)})}{(A\boldsymbol{r}^{(k)},\boldsymbol{r}^{(k)})} \tag{11}$$

共轭梯度法是最速下降法的改进，提高了收敛效率，并没改变梯度引导算法的本质。

（4）为对比起见，定义反问题非线性预测误差泛函 $J(\boldsymbol{x})=\frac{1}{2}(A(\boldsymbol{x})-\boldsymbol{b})^T(A(\boldsymbol{x})-\boldsymbol{b})$。对 $J(\boldsymbol{x})$ 做 Taylor 展开：

$$J(\boldsymbol{x}+\Delta\boldsymbol{x})\approx J(\boldsymbol{x})+(\mathrm{Grad}(J),\Delta\boldsymbol{x})+\frac{1}{2}(\mathrm{Hess}(J)\Delta\boldsymbol{x},\Delta\boldsymbol{x}) \tag{12}$$

记 $Q(\boldsymbol{x}+\Delta\boldsymbol{x})=J(\boldsymbol{x}+\Delta\boldsymbol{x})$。期望 $Q(\boldsymbol{x}+\Delta\boldsymbol{x})$ 为二次型泛函，本质上，期望在当前点 $\boldsymbol{x}$ 附近，$A(\boldsymbol{x})$=$\boldsymbol{b}$ 近似线性的，至少是弱非线性的。否则，上式的近似误差很大，后面导出的迭代算法很难收敛！

求极值 $\frac{\partial Q}{\partial\Delta\boldsymbol{x}}=0$ 得到：

$$\begin{aligned}&\mathrm{Grad}(J)-\mathrm{Hess}(J)\Delta\boldsymbol{x}\\&\quad=-\mathrm{Hess}(J)(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)\end{aligned}$$

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^k-\mathrm{Hess}(J)^{-1}\mathrm{Grad}(J) \tag{13}$$

对比上述四种迭代公式的推导，迭代公式的理论基础是有共通性的。本质上，上述迭代公式都希望 A 是满秩的，条件数好的，否则迭代解不收敛或很难收敛。遗憾的是，系数矩阵 A 是物理系统决定的，与观测方式关系很密切。除非改变正问题 $A\boldsymbol{x}=\boldsymbol{b}$ 的提法，否则系数矩阵 A 不是可以随意改变的。改变观测方式对系数矩阵 A 的性质的影响是第二位的。

3.3.3 广义线性反问题求解的若干要点

3.3.3.1 迭代格式中梯度项的认识

广义线性反问题的迭代解，原则上可统一成：

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^{k}-\mu_k \mathrm{Grad}\left(J(\boldsymbol{x}^{k})\right) \tag{14}$$

（a） $\mathrm{Grad}\left(J(\boldsymbol{x}^{k})\right)=-A^{T}\left(A\boldsymbol{x}^{k}-\boldsymbol{b}^{k}\right)$ 代表残差数据反投影，反投影算子为 A^{T} 。另外，$\mathrm{Grad}\left(J(\boldsymbol{x}^{k})\right)=\delta\boldsymbol{x}^{k}=-A^{T}\left(A\boldsymbol{x}^{k}-\boldsymbol{b}^{k}\right)$，则（14）式可写为，

$$\boldsymbol{x}^{k+1}=\boldsymbol{x}^{k}-\mu_k \delta\boldsymbol{x}^{k} \tag{15}$$

从任何一个迭代算法可见，对于 $J(\boldsymbol{x})$ 可微的代价泛函的求解，梯度是核心的问题。迭代算法的收敛快慢取决于梯度的正确性。步长可以统一地视为对梯度的预处理算子。Hessian 矩阵的逆是更好的梯度预处理算子。在实际数据处理中，对梯度进行各种依据问题的物理背景进行的预条件处理是保证收敛或得到有意义反演解的根本保障。由（15）式可以清楚地看出，梯度是某一迭代步要反演的系统参数的修正量，一般意义下，它应该与系统参数有相同的结构特征，或者它应该按照系统参数的几何结构来迭代地修改系统参数，这样能保证得到的反演解有物理意义，应该也能加快收敛。与现代图像分析方法结合、与空间数据统计方法结合是很符合逻辑的选择。

（b）若记 $\boldsymbol{x}^{k+1}=G\left(\boldsymbol{x}^{k}\right)$，则希望 $G\left(\boldsymbol{x}\right),\boldsymbol{x}\in D$ 是连续的，尤其是 Lipschtz 连续的，即

$$\left|G(\boldsymbol{x}_1)-G(\boldsymbol{x}_2)\right|\le L\left|\boldsymbol{x}_1-\boldsymbol{x}_2\right| \tag{16}$$

其中，$L<1$ 是压缩映射所要求的。对于收敛的迭代格式，算子 G 应满足上述要求。

3.3.3.2 迭代解的子空间投影表示

所有迭代法都会产生一个迭代解序列，迭代解有十分类似的表达形式。

（a）一般迭代法（Jacobi 方法、Gauss-Seidal 方法、逐次超松弛（SOR）方法）的迭代解：

$$\boldsymbol{x}_n=\boldsymbol{x}_0+\alpha_1\boldsymbol{r}_1+\alpha_2\boldsymbol{r}_2+\cdots+\alpha_n\boldsymbol{r}_n \tag{17}$$

其中，$\boldsymbol{r}_i=\boldsymbol{b}-A\boldsymbol{x}_i$ ， $i=1,2,...,n-1$，$\alpha_i=\dfrac{w_i}{a_{kk}}$, $i=1,2,...,n-1$，a_{kk} 为 $A\boldsymbol{x}=\boldsymbol{b}$ 中 A 矩阵的对角元素，松弛因子 w_i 可能取常值，$0\le w_i\le 2$ 。

（b）最速下降法的迭代解：

$$\boldsymbol{x}_n=\boldsymbol{x}_0+\alpha_1\mathrm{Grad}J(\boldsymbol{x}_1)+\alpha_2\mathrm{Grad}J(\boldsymbol{x}_2)+\cdots+\alpha_{n-1}\mathrm{Grad}J(\boldsymbol{x}_{n-1}) \tag{18}$$

其中，梯度向量 $\mathrm{Grad}J(\boldsymbol{x}_k)=-A^{T}\boldsymbol{r}_k$, $k=1,2,...,n-1$， α_k 为步长因子。

（c）全牛顿法的迭代解：

$$\boldsymbol{x}_n=\boldsymbol{x}_0+H^{-1}(\boldsymbol{x}_1)\mathrm{Grad}J(\boldsymbol{x}_1)+H^{-1}(\boldsymbol{x}_2)\mathrm{Grad}J(\boldsymbol{x}_2)+\cdots+H^{-1}(\boldsymbol{x}_{n-1})\mathrm{Grad}J(\boldsymbol{x}_{n-1}) \tag{19}$$

（d）共轭梯度法的迭代解：

$$\boldsymbol{x}_n=\boldsymbol{x}_0+\alpha_1\boldsymbol{p}_1(\boldsymbol{x}_1)+\alpha_2\boldsymbol{p}_2(\boldsymbol{x}_2)+\cdots+\alpha_{n-1}\boldsymbol{p}_{n-1}(\boldsymbol{x}_{n-1}) \tag{20}$$

其中， $\boldsymbol{p}_i=\boldsymbol{r}_i+\beta_i\boldsymbol{p}_{i-1}$ 为共轭梯度方向， β_i 为线性加权因子。

原则上，上述四种迭代格式中解的修正向量都可认为是数据残差的反投影形成，都是广义的梯度方向。可以看出，迭代解为精确解向一个子空间的投影，子空间是由广义梯度方向矢量（作为基函数）构成的！

3.3.3.3 关于线性方程组的近似投影解法

（a）把线性系统 $\boldsymbol{Ax}=\boldsymbol{b}$ 理解为一个投影过程

可以把 $\boldsymbol{Ax}=\boldsymbol{b}$ 理解成：$A_{n\times m}$ 是 m 个列向量（n 维列向量）形成，它们形成一个坐标系或基函数族，构成一个子空间。$\boldsymbol{x}_{m\times 1}$ 向量投影到此坐标系或基函数族上，形成向量 $\boldsymbol{b}_{n\times 1}$，即

$$\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\right]\begin{bmatrix}x_1\\x_2\\\vdots\\x_m\end{bmatrix}=\boldsymbol{b} \tag{21}$$

（21）式有更常用的表达形式，

$$\boldsymbol{b}=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_m\boldsymbol{a}_m=\sum_{j=1}^{m}x_j\boldsymbol{a}_j \tag{22}$$

可以认为这是一个广义 Fourier 变换，此时基函数族为$\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\right]$。据此可以看出，数据向量 $\boldsymbol{b}$ 是由$\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\right]$的线性组合形成的，不同的 $\boldsymbol{a}_j$, $j=1,2,...,m$ 对 $\boldsymbol{b}$ 的贡献是不同的。应该注意到，基函数族$\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\right]$中各元素是有物理含义的。基于对系数矩阵 $A_{n\times m}$ 的 SVD 分解可以找到对 $\boldsymbol{b}$ 贡献最大的、次大的、依次降序的有物理意义的分量。这是稀疏提升（sparse-promotion）方法解方程组（ $\boldsymbol{Ax}=\boldsymbol{b}$ ）的基本思想来源。

（b）线性方程组的直接反投影解法

由广义 Fourier 变换知，

$$\boldsymbol{x}=C_{m\times n}\boldsymbol{b}_{n\times 1}=\sum_{k=1}^{n}b_k\boldsymbol{c}_k \tag{23}$$

其中，矩阵 $C_{m\times n}$ 由基函数族$\left[\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_n\right]$构成列向量。矩阵 $C_{m\times n}$ 是如何构成的？它由如下方式给出，

$$C_{m\times n}A_{n\times m}\boldsymbol{x}_{m\times 1}=C_{m\times n}\boldsymbol{b}_{n\times 1} \tag{24}$$

若取 $C_{m\times n}$=$A^T_{m\times n}$，则上式为，

$$A^T_{m\times n}A_{n\times m}\boldsymbol{x}_{m\times 1}=A^T_{m\times n}\boldsymbol{b}_{n\times 1} \tag{25}$$

很显然，（25）式就是线性反问题对应的法方程。$A^T_{m\times n}A_{n\times m}$=$H$ 在实际的线性反问题中往往是一个庞大的矩阵。若 $A_{n\times m}$ 近似于正交矩阵，即$\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\right]$是互相正交的，则 $A^T_{m\times n}A_{n\times m}=\sum$，其中 $\sum$ 是对角阵。（25）式变为：

$$\boldsymbol{x}_{m\times 1}=\sum{}^{-1}A^T_{m\times n}\boldsymbol{b}_{n\times 1} \tag{26}$$

这样的假设情况在勘探地震中的参数估计中是很难成立的。但是，若$\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\right]$近似正交，可认为 $A^T_{m\times n}A_{n\times m}\approx\sum$，$\sum$ 是对角阵。（25）式也可以近似地写为，

$$\boldsymbol{x}_{m\times 1} \approx \sum^{-1} A^T_{m\times n}\boldsymbol{b}_{n\times 1} \tag{27}$$

为了计算的简单，这样的近似假设得到非常普遍的应用。譬如在所谓的保幅最小二乘偏移成像中，经常利用这样的假设照明补偿的成像。

假设 $A_{n\times m}$ 是列标准正交时， $A^T A = I$ ，其中 I 是单位阵。则（25）式变为，

$$\boldsymbol{x}_{m\times 1} \approx A^T_{m\times n}\boldsymbol{b}_{n\times 1} \tag{28}$$

（28）式被称为解线性方程组的直接反投影法。勘探地震中的 Migration，从数学上看，就是仅仅使用直接反投影法解线性方程组。直接反投影的物理含义是：穿过当前点的所有射线的投影结果的加权叠加。但是，直接反投影法得到的是线性系统 $A\boldsymbol{x}=\boldsymbol{b}$ 的近似解或模糊解。几何直观的解释是：系数矩阵 $A_{n\times m}$ 撑起的坐标系中，列向量不正交，导致参数 $\boldsymbol{x}_{m\times 1}$ 的照明不均匀。

对（28）式做分量层面上的分析，认为 $A^T_{m\times n}$ 是由 m 个行向量组成的，则（28）式可写成：

$$x_i = \sum_{j=1}^{n} a^{row}_{i,j} b_j \,, i = 1,2,...,m \tag{29}$$

可见，上式可以按行向量与数据向量的内积运算求解 $\boldsymbol{x}$ 的每个分量，仅仅涉及到两个向量的乘。

这样的方式特别适合并行计算，既可在 $x_i, i = 1,2,...,m$ 层次上并行，也可在 $\left(\boldsymbol{a}^{row}_i, \boldsymbol{b}\right)$ 内积层次上并行，尤其在 GPU 处理器上，可获得极高的加速。

显然， $\boldsymbol{x}_{m\times 1} = \sum^{-1} A^T_{m\times n}\boldsymbol{b}_{n\times 1}$ 就失去了如上的并行计算特征，从它的分量层面看：

$$x_i = \frac{\sum_{j=1}^{n} a^{row}_{i,j} b_j}{K_{i,i}} \,, i = 1,2,...,m \tag{30}$$

其中， $K_{m\times m} = A^T_{m\times n} A_{n\times m}$ 。其中对角元素的计算式为 $k_{i,i} = \sum_{j=1}^{n} a^T_{i,j} a_{j,i}$ ，很显然，对勘探地震中射线层析成像问题而言，对角元素的物理含义为：所有穿过当前点的射线的长度平方之和。（30）式中分子项的含义是：穿过当前点的所有射线的投影结果的加权叠加。由于 $K_{m\times m}$ 的引入，失去了并行计算特征，除非 $K_{m\times m}$ 事先已给出。到目前为止，直接投影法由于精度太低并不常用。

3.3.3.4 Kaczmarz 方法再议

对 Kaczmarz 方法的思想与做法可以总结如下。已知，线性方程组的经典迭代解法可以简述如下。首先给方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 一个初始解 $\boldsymbol{x}_0$，得到如下残差序列， $\boldsymbol{r}^{(0)} = \boldsymbol{b} - A\boldsymbol{x}^{(0)}$ ， …， $\boldsymbol{r}^{(k)} = \boldsymbol{b} - A\boldsymbol{x}^{(k)}$ ，然后，用 $A^T\boldsymbol{r}^{(k)} = \Delta\boldsymbol{x}^{(k)}$ 得到修正量，即形成迭代公式， $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \Delta\boldsymbol{x}^k$ 。但 Kaczmarz 方法的具体实现比较特殊，它是按分量形式进行迭代计算的。在最小二乘意义下，已知广义线性反问题的迭代公式为，

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + H^{-1}A^T\boldsymbol{r}^k \tag{31}$$

H^{-1} 仅用对角 Hessian 的逆来表示，分量形式表达的（31）式为，

$$x_i^{k+1}=x_i^k+\frac{\sum_{j=1}^{n}a_{ij}r_j^k}{\sum_{j=1}^{n}a_{ij}a_{ji}}\quad ,i=1,2,...,m \tag{32}$$

其中，利用了关系式 $A_{n\times m}=\left(a_{ji}\right)$, $A_{m\times n}^T=\left(\overline{a}_{ij}\right)$ 。式中分母代表对 Hessian 对角元素的简化。引入这样的简化，既保证了（32）式的并行计算特征，又提高了计算效率，这就是 ART 算法（ART=Algebraic Reconstruction Technique）。从射线层析看，分母项的物理含义是：所有穿过当前点的射线的长度平方之和；分子项的含义是：穿过当前点的所有射线的投影结果的加权叠加。

若对地下任一网格点而言，把穿过该网格的射线对应的梯度求和（加权求和），计算精度能进一步提升。这就引出了 SIRT 方法（SIRT=Simultaneous Iterative Reconstruction Technique），

$$x_i^{k+1}=x_i^k+\frac{1}{M_i}\sum_{l=1}^{M_i}\frac{\sum_{j=1}^{n}a_{ij}r_j^k}{\sum_{j=1}^{n}a_{ij}a_{ji}}\quad ,i=1,2,...,m \tag{33}$$

其中，$\sum_i$ 表示对穿过当前点的射线的梯度求和，M_j 代表穿过该点的射线的总条数。据此可以看出，ART/SIRT 并没什么特殊的地方，仅仅是针对法方程的、近似的、并行性好的算法而已！

4 非线性反问题的解法

我们多次断言：反问题的非线性性取决于可观测量与要反演的不可观测量之间的非线性性。对于线性预测系统 $\boldsymbol{Ax=b}$，如果预测误差满足高斯分布，即 $\boldsymbol{d}^{obs}=\boldsymbol{b}+\boldsymbol{\eta}$，$\boldsymbol{b}$ 是线性系统的状态变量，可被线性系统 $\boldsymbol{Ax=b}$ 准确描述，$\boldsymbol{\eta}$ 是预测误差，它满足高斯分布，尤其当它是 Gauss 白噪声时，基于误差泛函，

$$\Im\left(\boldsymbol{x}\right)=\frac{1}{2}\left(A\boldsymbol{x}-\boldsymbol{b}\right)^T\left(A\boldsymbol{x}-\boldsymbol{b}\right) \tag{34}$$

的最优估计结果是无偏的和方差最小的。这其中最关键的原因是 $\boldsymbol{Ax=b}$ 为线性的，误差泛函是二次型的、凸的、光滑可微的。由图 1 可见，凸反问题的解是唯一的。

假若可观测量 $\boldsymbol{d}^{obs}\left(=\boldsymbol{b}+\boldsymbol{\eta}\right)$ 与要反演的不可观测量 $\boldsymbol{x}$ 之间关系是非线性的，记为 $\boldsymbol{b}=A\left(\boldsymbol{x}\right)$，假若这个关系式具有 n 元 m 阶多项式的形式，则误差泛函，

$$\Im\left(\boldsymbol{x}\right)=\frac{1}{2}\left(A\left(\boldsymbol{x}\right)-\boldsymbol{b}\right)^T\left(A\left(\boldsymbol{x}\right)-\boldsymbol{b}\right) \tag{35}$$

至少是 m×m 阶 n 元多项式，这样的函数会有非常多的极小点，不再是凸的。

事实上，到目前为止，各种梯度导引类的优化算法，仅仅可以对凸的误差泛函进行优化求解。并期望凸的误差泛函是光滑可微的、最好二阶导数存在的。如果凸误差泛函存在不可微点，可以用引入次梯度的方法构建 LASSO 算法进行求解。

因此，可以断言，与其讨论（强）非线性反问题的求解，不如重点考虑凸问题的求解。对于凸误差泛函，若它又是光滑的、二阶可微的，Newton 迭代法及各种变种是典型的解法；若它是不可微的，LASSO 算法是典型的算法。具体的公式参考相应的文献（王华忠文集-Bayes 估计观点下的地震信号分析与地震波反演成像）。

一般地，认为强非线性反问题可以用 Monte Carlo 算法来求解。我们并不赞同这样的判断。

Monte Carlo 算法的基本目的是清楚的：对巨大的模型空间（要反演的参数的定义域）支撑的后验概率密度函数，进行全面探索，找到后验概率密度函数中贡献大的区域，最终找到后验概率密度函数最大值对应的模型参数（即参数估计结果）。

Monte Carlo 算法的精髓显然应该是：在巨大的参数空间中，能进行全面的高效探索，并能真正找到“最”重要区域的算法。

到目前为止，Monte Carlo 算法的设计借用了 Markov Chain 的思想：当前状态（模型参数）转移到下个状态（模型参数），按一定的规则进行。这个规则也可认为是确定步长和方向的规则。此时，模型参数在巨大定义域中转移方向的确定是基于 Metropolis 准则的。而不是把梯度方向作为转移方向。Metropolis 准则首先是概率性的，让误差泛函下降的方向并不一定是模型参数演变的方向；让误差泛函上升的方向也不一定不是模型参数演变的方向。这应该是 Monte Carlo 算法能进行全局搜索并能找到后验概率密度函数最大值（或误差泛函最小值点）的最本质原因。

但是，MCMC（Markov Chain Monte Carlo）算法的探索效率太低！因此，群体智能的引入就是必要的，这也是当前 Monte Carlo 算法与各种群体智能算法（遗传、粒子群、蚁群等算法）相结合的原因，主要目的是提高探测效率。

即使如此，也没有从理论上保证所设计的算法能全面地探索巨大的参数空间，也不能保证找到后验概率密度函数 σ_M 最大或误差泛函 $S(\mathbf{x})$ 最小所在的区域。

客观地评价，对于非线性问题不严重的情况，这样设计的 Monte Carlo 算法对于小维度的问题应该说性能还是不错的。另外，Monte Carlo 算法的设计，人的主观认识起到很大的作用，这与梯度类优化算法相比是很不相同的。

这进一步印证了，对于特别高维的、强非线性反问题的求解，目前还没有统一的、有效的数学理论框架。

5 勘探地震中非线性反问题的认识

5.1 关于勘探地震领域非线性反问题的基本认识

线性高斯假设下的 L2 范数形式的误差泛函反问题，其内涵已经十分明晰。现代信号分析理论的参数估计部分深入地分析过这类问题解的特征。这是所有反问题中最常用、最基础、也最简单的部分。很显然，线性、高斯、平稳假设下的反问题由于其过于严格的假设，导致与实际情况不符，因此反问题的研究必然向非线性、非高斯和非平稳情形下发展。

在勘探地震中，非线性反问题的本意是待估计参数与实测数据之间的关系是非线性的；高斯性指的是对参数的分布施加了 Gauss 分布的假设，同时也假设理论合成数据与实测数据的逼近误差满足高斯分布，因此后验概率密度也满足高斯分布，这也就有了上述 MAP 原则下反演解的统计意义；平稳性是指系统输出的信号是时间平稳的，事实上也希望是空间平稳的。

FWI 是 Bayes 估计理论在地震波反演成像中的一个应用实例。本质上，它要求解一个梯度导引类的凸优化问题。在代价泛函的优化求解中，FWI 局限于局部化、线性化的方法，而这样的解法要求反演参数与观测数据之间存在线性的或弱线性的关系，并不能有效地解决实际中遇到的强非线性的反演问题。

事实上，较强非线性反问题的有效求解，关键根本不在优化算法上，而是反问题的提法上。我认为解决非线性问题的核心观点应该是：将问题描述为或转化为一系列的凸问题，而不是用所谓的“高级算法”硬解一个较强的非线性反问题。事实上，反问题非线性性强的根源在于正问题：即待估计参数与（部分）观测场之间存在较强的非线性关系。因此，完全有必要把较强非线性反问题的求解从专注于数值方法上转到专注于反问题的提法上，转到建立更紧密的待估计参数与（部分）观测场之间的关系上。数据的完备性和完美性是提高反问题反演精度的关键因素，但一旦数据给定，反问题的解决还是依赖于所用的策略与方法。

5.2 勘探地震领域非线性反问题求解的基本理念

首先是提出更好的“解的判决准则”。数据逼近残差 L2 范数作为衡量准则的确是不够合理的。它蕴含的假设太多，正问题$F\left(\boldsymbol{m}\right)$必须能完全预测实测数据$\boldsymbol{d}^{obs}$中的各种物理波现象，仅仅余下满足高斯分布的随机噪声。众所周知，这是不可能的。另一方面，Gauss 分布是一个单峰的分布函数（凸函数），对应的误差泛函是二次型，这就预示着正问题$F\left(\boldsymbol{m}\right)$是线性的（或接近线性的）。众所周知，正问题$F\left(\boldsymbol{m}\right)$不是线性的，为了满足这个要求，必须提供非常接近真估计结果的初始值，正问题$F\left(\boldsymbol{m}\right)$才能退化为线性的（或接近线性的）。这样的要求在复杂地表和复杂介质探区是不能满足的。

关于反演问题解的判决依据，根本上看，还是从信息状态的角度才能得到深入的认识。我的观点：“反问题：永远是信息不足情况下的统计判决问题”。

后验概率密度函数是最高判决依据。由于后验概率密度计算上很困难，退化是可以的，也是必须的。但退化成实测波形和预测波形的差的总能量，是否合适非常值得怀疑！FWI 的实践已经证明这是不行的！用包络代替波形、用加权旅行时代替波形、定义所谓的“最佳输运距离”、用互相关定义误差泛函、用道集拉平代替波形逼近定义误差泛函等，都可定义一个更凸的代价泛函，松弛了对初始值的严格要求，一定程度上，避免解陷入局部极值。但是，解的判决准则问题还是值得深入研究，尤其是从特征数据的角度提出新的误差泛函；从信息状态的角度，提出更好的、凸的误差泛函。

第二是如何构建更凸的反问题。我认为关键措施之一是从实测波场中筛选数据，即筛选出“部分可以很好预测的波场”。数学物理正问题模型相对于实际地下介质中的波传播情况都有所简化，导致正问题不能“解释”所有数据，这是反问题病态性的部分成因。因此，可以通过“修正”数据，使之能与正问题匹配，来缓解反问题的病态性。必须把特征数据筛选出来，匹配上能预测该特征数据的正问题，加上合适的模型参数先验信息，就可以构成一个很好的、比较凸的反问题。譬如，CMP 道集（数据）与基于层状介质的时距关系就是这样的特征数据与正问题匹配的典型实例。这看起来很简单，却支撑了地震波成像的实际应用。目前，这样的数据筛选工作做得很不到位，尤其在数据表达已经发展得很深入的情况下，并没有很好地借用来完善参数估计反问题。关键措施之二是特征化的模型参数表达。直接对模型参数进行网格离散，没有抓住要反演的模型参数的变化特征，徒然造成参数维度的指数上涨，反演难度加大。事实上，任何实际的物理系统对应的模型参数都是受自然规律控制的，都是有特征的。勘探地震中的介质特征就可以表示为：广大的沉积层中包含有构造运动、火山活动等引起的局部跃变的地质异常体。如能将模型参数进行特征表达、降维表达，那么反问题求解的难度可以下降很多。譬如，将模型拆分为宏观背景m_0+散射模型δm，若对散射模型做特征稀疏表达，可以进一步降维。实际上，勘探地震中介质以层状为主和波场以反射波为主的特征并没有在 FWI+LS_RTM 反演中得到真正的体现，导致这两种代表性的反演成像方法都难于真正实用化。

第三是提出波场中存在多次波时的非线性成像理论与方法。勘探地震中，丰富的多次波的存在是地下介质存在强扰动、地震波反演出现较强非线性性的数据上的标志。强多次波存在时的地震波反演成像问题一定是强非线性反问题。多次反射波反演成像的非线性性源于包含多次波的全波场模拟需要已知背景速度和反射界面的反射系数，实际上它们都是要反演的目标。而背景速度和反射界面的反射系数对实测波场的影响是相互耦合的。降低反问题非线性性的措施体现在如何解开二者的耦合。线性 Born 反演成像和 FWI 反演成像的串联迭代应是合理的选择，具体做法为：假定道集基本拉平的情况下，Born 近似下的 LS_RTM 生成保真的零角度反射系数；基于空间数据融合方法，在背景速度场中融入反射系数形成包含速度跃变；基于此速度模型，非线性高波数 FWI 过程中可以模拟出层间多次波（也可以包括自由表面多次波），就可以进行包含层间多次波的反演成像。

第四是考虑存在强介质参数扰动情形下正问题的改进。参数估计无论如何离不开用建立的数学物理关系来预测实测数据。若离开了这一点，任何反问题的逻辑基础就不存在了，变成了无中生有了！但是，在强介质参数扰动情况下，什么是合理的、恰当的数学物理关系？能与反问题很好地融合起来，非常方便后续的反演计算？

第五是进一步考虑用群体智能算法改进的 MCMC 方法+合理的解的评判准则来构建新的非线性反问题求解策略。

梯度导引类的算法只能用于求解凸的反问题，这完全不适用求解强非线性反问题。前已述及，MCMC

方法求解高维、强非线性反问题也存在非常明显的限制。但是，在当前缺乏强非线性反问题有效算法的情况下，用群体智能算法改进 MCMC 方法，引入合理的正则化约束，再引入更好的解的评判准则，不失为一个符合逻辑的探索方向。

6 勘探地震领域非线性反问题求解的具体策略

解决非线性问题，切实的出发点依然是从数据（Data）入手、从地震波传播的物理问题入手、从提出凸问题入手。

参数反演理论的基本依据是（人为）建立的正问题能预测或解释实测数据中的信号（系统的状态变量），这个基本依据永远是必要的！

因为反问题的根本概念含义就是：由可以直接观测的量估计与这些可以直接观测的量相关的潜（隐）变量。直接观测量及产生这些直接观测量的物理过程描述方式对反问题当然至关重要。

直接观测量与由正问题数值解相减得到的残差满足 Gauss（广义 Gauss）分布是为了简化反问题的求解而施加的要求。这个人为施加的要求也不是可有可无的。要达到此要求，必须有：①对数据进行预处理，使得处理后的数据中仅包含所选正问题能解释的信号，其他的是 Gauss 随机噪声；②改变正问题，更好地预测数据中包含的信号。但最好是改变数据适应正问题而不是反之，因为改变正问题带来的是参数空间的变化，可能给反问题的求解带来额外的困难；③施加合理的正则化，引入更好的初始解。

总体而言，解决非线性反问题的正确途径是从筛选数据入手。

地震勘探中炮集包含的特征数据子集可以罗列为：直达波；折射波；Diving Wave；一次反射波；一次散射波；多次反射波；多次散射波。上述特征数据（波现象）的分解是把 FWI 这个非线性反问题转换成一系列凸问题求解的关键。在人工智能（ML）系统下，可以把 FWI 包装成一系列凸问题组合，实现自动化与智能化的地震波反演成像。相应的关键技术包括：①特征波场的提出或提取，这需要高级的信号分析方法；②地质体几何结构的提取，这需要现代图像分析方法、模式识别和基于 ML 的智能图像识别；③（包括）多次反射波的层析成像与偏移成像技术；④（包括）多次散射波层析成像与偏移成像技术；⑤考虑复杂介质情况（弹性、各向异性、Q 值）下的上述一系列凸问题的求解；⑥上述一系列凸问题实现的一系列人工智能框架及相关的方法理论。

7 结论与讨论

线性、高斯、平稳假设下的地震波反演成像方法，事实上，也包括这种情形下的信号分析和图像分析方法，已经发展得十分完善，非线性、非高斯、非平稳假设下的地震波反演成像已经成为地震波成像方法研究的重点。后验概率密度函数推广到广义 Gauss 分布情形，尤其是 L1 范数下进行稀疏反演，已经由稀疏信号表达和稀疏图像分析广泛地扩展到了勘探地震数据分析的相关环节中。非平稳信号分析更是由于短时 Fourier 变换、Gabor 变换和小波变换的引入成为了信号与图像分析的基本做法。非平稳、非高斯信号与图像分析也已经成为常态。

但是，非线性参数估计问题，无论在信号和图像分析领域，还是在勘探地震领域中，都是迫切需要解决却没有得到很好解决的问题。尤其对于强非线性反问题，基本上没有什么有效的方法。FWI 就是这样一个典型的强非线性反问题。

对于 FWI 这样的强非线性反问题，我认为不可以期望找到一个有效的算法来解决，而是根据勘探地震的介质特征、波场特征，提炼出一系列凸问题的组合，来逼近该强非线性问题的解。鉴于地震数据仅仅在地表接收，而且叠前数据是频带有限的、孔径有限的，观测噪音是不可控的，仅仅靠叠前地震数据本身是无法满足储层描述所需要的高精度成像要求的，事实上，在复杂地表条件和复杂介质构造情况下，即便获得准确的地下介质的构造像，都是难以做到的。因此，不要期望发展出一个完善的方法，能一劳永逸地解决地震波成像问题，不可能存在这样的方法。随着油气地震勘探目标逐渐复杂、尺度越来越小，埋藏越来越深，地震波成像问题永远是一个“信息不足情况下的弹性参数估计问题”。从综合信息利用的观点，看待地震波成像，逐步发展地震波成像技术才是合理的思想逻辑。

求解 FWI 这样的强非线性反问题的正确路线，就是提出一系列凸问题的组合来逼近它的解。针对当前的具体情况，我提出如下的推进 FWI 实用化的基本做法：

① 首先发展基于 CMP 道集和（类）CMP 道集的基于扫描的初始速度估计方法。用 ML 算法实现扫描初始速度建模的自动与智能实现。

② 接着发展初至与早至波波动理论层析成像方法与技术。估计各向异性（VTI）速度和 Q 值。这中间初至波场与早至波场的提取与预处理是核心，不能像 Laplace 变换那样简单地处理。开始要发展具备自动的初至波到达时提取的波动理论层析成像方法。初至波的到达时要给出一个明确的定义，我提出用初至子波的质心来定义到达时。另外，要综合利用各种近地表模型的先验信息，譬如露头、高光谱、微测井等，减少近偏移距观测缺失带来的反问题解的不确定性。鉴于我国的油气探区大多位于陆上，首先要把该方法用在陆上地震数据成像处理中，使其能起到显著改善成像质量的作用，且能适用于山地、山前带、沙漠等不同探区的成像。然后再发展初至与早至波全波形波动理论层析成像方法，而且把前者的反演结果作为后者的初始值。

③ 然后发展以反射结构为中心的反射波波动理论层析成像方法与技术。其中的三个要点问题是：用偏移与反偏移结合筛选一次波数据；用机器学习算法自动提取反射界面、特殊地质体和断层，把上述信息用于速度模型更新的约束中，并随迭代过程不断演化构造样式；把梯度更新量累积起来，与地质结构约束一起，基于 Markov 演化的思想，分尺度修改速度模型，进一步提升梯度演化的稳定性。

④ 发展最小二乘 RTM 及最小二乘 PSDM 成像方法。迭代的 LS-RTM 本质上就是估计高波数扰动量的 FWI，非线性性非常强，因此，必须要首先利用特征反射波场（一次反射波场），并在高频近似波传播框架下构建所谓的保真 RTM，本质上是用角度 Hessian 的逆对常规 RTM 结果进行预条件处理。另外，也可以在线性反问题成立的条件下，发展出像域最小二乘 RTM/PSDM 和数据域最小二乘 RTM/PSDM，本质上就是近似求解法方程。

⑤ 发展宽带波阻抗成像方法。宽带波阻抗成像是当前带限反射系数成像的进一步扩展，宽带波阻抗成像分辨率更高，更有利于储层的描述与刻画。我提出了宽带波阻抗成像的具体策略解技术方案。

⑥ 发展层状介质下的多次波偏移成像和层析反演成像。

总之，到目前为止，已经充分地认识到，由叠前地震数据估计地下介质的几何结构和弹性参数，进而结合岩石物理，估计地下介质的岩性参数，从而实现油气藏的描述与刻画是勘探地震的核心问题。抽象地看，也是个 Learning from Data 的问题。地震波成像问题不是一个大数据分析问题，油气藏的描述与刻画问题某种程度上是一个大数据分析问题。机器学习（ML）和人工智能（AI）可以在地震波成像的某些环节中（主要在正则化方面）起到辅助作用，但 FWI 参数反演问题只能视为 ML/AI 理论的一个具体应用实例，后者包含的理论和应用范围要宽泛很多。我相信总有一天可以发展出 ML/AI 理论下的弹性参数反演机器人，也可以发展出油气地震勘探机器人。当然，这并不代表人类的作为可以被这样的机器人所取代，而是人类在更高层次开展更为精确的弹性参数反演和油气预测。

参考文献

[1]. 王华忠，2020，王华忠文集-Bayes 估计观点下的地震信号分析与地震波反演成像，同济大学海洋与地球科学学院波现象与智能反演成像研究组
[2]. Guust Nolet (Edited), 1987, Seismic Tomography with Applications in Global Seismology and Exploration Geophysics, D. Reidel Publishing Company
[3]. Billette, F. J., and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark: 67th Conference and Exhibition, EAGE, Extended Abstracts, B035.
[4]. Brenders, A. J., and R. G. Pratt, 2007, Waveform tomography of marine seismic data: What can limited offset offer?: 77th Annual International Meeting, SEG, Expanded Abstracts, 3024–3028, https://doi.org/10.1190/1.2793099.
[5]. Chavent, G., F. Clément, and S. Gómez, 1994, Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example: 64th Annual International Meeting, SEG, Expanded Abstracts, 1179–1182, https://doi. org/10.1190/1.1822731.
[6]. Datta, D., M. Sen, F. Liu, and S. Morton, 2016, Salt model building by shape-based parameterization and global

FWI: 86th Annual International Meeting, SEG, Expanded Abstracts, 1069–1073, https://doi.org/10.1190/segam2016-13867592.1.
[7]. Dellinger, J., A. Brenders, J. R. Sandschaper, C. Regone, J. Etgen, I. Ahmed, and K. J. Lee, 2017, The Garden Banks model experience: The Leading Edge, **36**, no. 2, 151–158, https://doi. org/10.1190/tle36020151.1.
[8]. Dellinger, J., A. Ross, D. Meaux, A. Brenders, G. Gesoff, J. Etgen, J. Naranjo, G. Openshaw, and M. Harper, 2016, Wolfspar®, an "FWI-friendly" ultralow-frequency marine seismic source: 86th Annual International Meeting, SEG, Expanded Abstracts, 4891–4895, https://doi.org/10.1190/segam2016-13762702.1.
[9]. Esser, E., L. Guasch, T. van Leeuwen, A. Aravkin, and F. J. Herrmann, 2015, Automatic salt delineation — Wavefield reconstruction inversion with convex constraints: 85th Annual International Meeting, SEG, Expanded Abstracts, 1337–1343, https://doi.org/10.1190/segam2015-5877995.1.
[10]. Gomes, A., and N. Chazalnoel, 2017, Extending the reach of full waveform inversion with reflection data: Potential and challenges: 87th Annual International Meeting, SEG, Expanded Abstracts, 1454–1459, https://doi.org/10.1190/segam2017-17731403.1.
[11]. Guasch, L., M. Warner, and F. J. Herrmann, 2016, Constrained waveform inversion — Automatic salt flooding with inclusions: 78th Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.201601547.
[12]. Irabor, K., and M. Warner, 2016, Reflection FWI: 86th Annual International Meeting, SEG, Expanded Abstracts, 1136–1140, https://doi.org/10.1190/segam2016-13944219.1.
[13]. Jonke, K., Z. Fu, B. Wray, and H. Shen, 2017, Improving minibasin and subsalt imaging with reflection full-waveform inversion: 87th Annual International Meeting, SEG, Expanded Abstracts, 1492–1496, https://doi.org/10.1190/segam2017-17661070.1.
[14]. Kadu, A., T. van Leeuwen, and W. Mulder, 2016, A parametric level-set approach for seismic full-waveform inversion: 86th Annual International Meeting, SEG, Expanded Abstracts, 1146–1150, https://doi.org/10.1190/segam2016-13870276.1.
[15]. Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: Theory and application, Society for Industrial and Applied Mathematics, 206–220.
[16]. Lu, R., 2016, Revealing overburden and reservoir complexity with high-resolution FWI: 86th Annual International Meeting, SEG, Expanded Abstracts, 1242–1246, https://doi.org/10.1190/ segam2016-13872562.1.
[17]. Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, no. 5, 645–653, https://doi.org/10.1190/1.1443081.
[18]. Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and hybrid waveform inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 871–876, https://doi.org/10.1190/segam2013-0087.1.
[19]. Mandroux, F., B. Ong, C. Ting, S. Mothi, T. Huang, and Y. Li, 2013, Broadband, long-offset, full-azimuth, staggered marine acquisition in the Gulf of Mexico: First Break, **31**, no. 6, 125–132.
[20]. Michell, S., E. Shoshitaishvili, D. Chergotis, J. Sharp, and J. Etgen, 2006, Wide azimuth streamer imaging of Mad Dog: Have we solved the subsalt imaging problem?: 76th Annual International Meeting, SEG, Expanded Abstracts, 2905–2909, https://doi. org/10.1190/1.2370130.
[21]. Michell, S., X. Shen, A. Brenders, J. Dellinger, I. Ahmed, and K. Fu, 2017, Automatic velocity model building with complex salt: Can computers finally do an interpreter's job?: 87th Annual International Meeting, SEG, Expanded Abstracts, 5250–5254, https://doi.org/10.1190/segam2017-17778443.1.
[22]. Moldoveanu, N., and J. Kapoor, 2009, What is the next step after WAZ for exploration in the Gulf of Mexico?: 79th Annual International Meeting, SEG, Expanded Abstracts, 41–45, https://doi.org/10.1190/1.3255750.
[23]. Mora, P., 1989, Inversion = migration + tomography: Geophysics, **54**, no. 12, 1575–1586, https://doi.org/10.1190/1.1442625.
[24]. Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, no. 3, 888–901, https://doi.org/10.1190/1.1444597.
[25]. Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2017, Salt model building at Atlantis with full-waveform inversion: 87th Annual International Meeting, SEG, Expanded Abstracts, 1507–1511, https://doi.org/10.1190/ segam2017-17738630.1.
[26]. Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2018a, Full-waveform inversion: The next leap forward in subsalt imaging: The Leading Edge, **37**, no. 1, 67b1–67b6, https://doi.org/10.1190/tle37010067b1.1.
[27]. Shen, X., L. Jiang, J. Dellinger, A. Brenders, C. Kumar, M. James, J. Etgen, D. Meaux, R. Walters, and N. Abdullayev, 2018b, High resolution full waveform inversion for structural imaging in exploration: 87th Annual International Meeting, SEG, Expanded Abstracts, 1098–1102, https://doi.org/10.1190/segam2018-2997202.1

[28]. Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, no. 1, 231–248, https://doi.org/10.1190/1.1649391.
[29]. Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, no. 8, 1259–1266, https://doi.org/10.1190/1.1441754.
[30]. Threadgold, I. M., K. Zembeck-England, P. G. Aas, P. M. Fontana, D. Hite, and W. E. Boone, 2006, Implementing a wide azimuth towed streamer field trial: The what, why and mostly how of WATS in Southern Green Canyon: 76th Annual International Meeting, SEG, Expanded Abstracts, 2901–2904, https://doi.org/10.1190/1.2370129.
[31]. Wang, P., Z. Zhang, Z. Wei, and R. Huang, 2018, A demigration-based reflection full-waveform inversion workflow: 88th Annual International Meeting, SEG, Expanded Abstracts, 1138–1142, https://doi.org/10.1190/segam2018-2997404.1.
[32]. Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambaré, 2012, Full waveform inversion for reflected seismic data: 74th Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.20148725.
[33]. Yang, Z., S. Huang, and R. Yan, 2015, Improved subsalt tomography using RTM surface offset gathers: 85th Annual International Meeting, SEG, Expanded Abstracts, 5254–5258, https://doi.org/10.1190/segam2015-5848366.1.
[34]. Zhang, Z., J. Mei, F. Lin, R. Huang, and P. Wang, 2018, Correcting for salt misinterpretation with full-waveform inversion: 88th Annual International Meeting, SEG, Expanded Abstracts, 1143–1147, https://doi.org/10.1190/segam2018-2997711.1.
[35]. Beaudoin, G., and A. A. Ross, 2007, Field design and operation of a novel deepwater, wide-azimuth node seismic survey: The Leading Edge, **26**, no. 4, 494–503, https://doi.org/10.1190/1.2723213.
[36]. Billette, F. J., and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark: 67th Conference and Exhibition, EAGE, Extended Abstracts.
[37]. Brenders, A. J., and R. G. Pratt, 2007, Waveform tomography of marine seismic data: What can limited offset offer?: 77th Annual International Meeting, SEG, Expanded Abstracts, 3024–3028, https://doi.org/10.1190/1.2793099.
[38]. Dellinger, J., 2016, Challenges to extending the usable seismic bandwidth at the seafloor in the deepwater Gulf of Mexico: 86th Annual International Meeting, SEG, Expanded Abstracts, 66–70, https://doi.org/10.1190/segam2016-13762860.1.
[39]. Dellinger, J., A. Ross, D. Meaux, A. Brenders, G. Gesoff, J. Etgen, J. Naranjo, G. Openshaw, and M. Harper, 2016, Wolfspar®, an "FWI-friendly" ultralow-frequency marine seismic source: 86th Annual International Meeting, SEG, Expanded Abstracts, 4891–4895, https://doi.org/10.1190/segam2016-13762702.1.
[40]. Dellinger, J., A. J. Brenders, J. R. Sandschaper, C. Regone, J. Etgen, I. Ahmed, and K. J. Lee, 2017, The Garden Banks model experience: The Leading Edge, **36**, no. 2, 151–158, https://doi.org/10.1190/tle36020151.1.
[41]. Gherasim, M., J. Etgen, B. Nolte, I. Ahmed, G. Xia, M. Reasnor, J. Howie, and D. Shepherd, 2010, Anisotropic velocity model building using OBS node tomography at Atlantis Field, Gulf of Mexico: 80th Annual International Meeting, SEG, Expanded Abstracts, 4359–4363, https://doi.org/10.1190/1.3513783.
[42]. Plessix, R.-E., G. Baeten, J. W. de Maag, M. Klaassen, Z. Rujie, and T. Zhifei, 2010, Application of acoustic full waveform inversion to a low-frequency large-offset land data set: 80th Annual International Meeting, SEG, Expanded Abstracts, 930–934, https://doi.org/10.1190/1.3513930.
[43]. Pratt, R. G., Z.-M. Song, P. Williamson, and M. Warner, 1996, Two-dimensional velocity models from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, no. 2, 323–340, https://doi.org/10.1111/j.1365-246X.1996.tb07023.x.
[44]. Roberts, M., T. Dy, S. Ji, M. Reasnor, and D. Shepherd, 2011, Improving Atlantis TTI model building: OBN+NATS, prism waves & 3D RTM angle gathers: 81st Annual International Meeting, SEG, Expanded Abstracts, 3238–3242, https://doi. org/10.1190/1.3627868.
[45]. Routh, P., J. Behura, and M. Tanis, 2016, Introduction to this special section: Full-waveform inversion Part I: The Leading Edge, **35**, no. 12, 1024–1024, https://doi.org/10.1190/tle35121024.1.
[46]. Sirgue, L., 2006, The importance of low frequency and large offset in waveform inversion: 68th Conference and Exhibition, EAGE, Extended Abstracts.
[47]. Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, no. 4, 65–70, https://doi.org/10.3997/1365-2397.2010012.
[48]. Tanis, M., and J. Behura, 2017, Introduction to this special section: Full-waveform inversion Part II: The Leading Edge, **36**, no. 1, 58–58, https://doi.org/10.1190/tle36010058.1.
[49]. Akcelik, V., 2002, Multiscale Newton-Krylov methods for inverse acoustic wave propagation: Ph.D. thesis,

Carnegie Mellon University.
[50]. Aki, K., and P. Richards, 1980, Quantitative seismology: Theory and methods: W.H. Freeman & Co.
[51]. Alerini, M., S. Le Bégat, G. Lambaré, and R. Baina, 2002, 2D PP- and PS-stereotomography for a multicomponent datset: 72nd Annual International Meeting, SEG, Expanded Abstracts, 838–841.
[52]. Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.
[53]. Amundsen, L., 1991, Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion: Geophysics, **56**, 2027–2035.
[54]. Amundsen, L., and B. Ursin, 1991, Frequency-wavenumber inversion of acoustic data: Geophysics, **56**, 1027–1039.
[55]. Askan, A., 2006, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: Ph.D. thesis, Carnegie Mellon University.
[56]. Askan, A., V. Akcelik, J. Bielak, and O. Ghattas, 2007, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: Bulletin of the Seismological Society of America, **97**, 1990–2008.
[57]. Askan, A., and J. Bielak, 2008, Full anelastic waveform tomography including model uncertainty: Bulletin of the Seismological Society of America, **98**, 2975–2989.
[58]. Barnes, C., and M. Charara, 2008, Full-waveform inversion results when using acoustic approximation instead of elastic medium: 78th Annual International Meeting, SEG, Expanded Abstracts, 1895–1899.
[59]. Barnes, C., M. Charara, and T. Tsuchiya, 2008, Feasibility study for an anisotropic full waveform inversion of cross-well data: Geophysical Prospecting, **56**, 897–906.
[60]. Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.
[61]. Bednar, J. B., C. Shin, and S. Pyun, 2007, Comparison of waveform inversion, Part 3:Amplitude approach: Geophysical Prospecting, **55**, 477–485.
[62]. Ben Hadj Ali, H., S. Operto, and J. Virieux, 2008a, Velocity model building by 3D frequency-domain full-waveform inversion of wide-aperture seismic data: Geophysics, **73**, no. 5, VE101–VE117.
[63]. Ben Hadj Ali, H., S. Operto, J. Virieux, and F. Sourbier, 2008b, 3D frequency-domain full-waveform tomography based on a domain decomposition forward problem: 78thAnnual International Meeting, SEG, Expanded Abstracts, 1945–1949
[64]. Beydoun, W. B., and A. Tarantola, 1988, First Born and Rytov approximation: Modeling and inversion conditions in a canonical example: Journal of the Acoustical Society of America, **83**, 1045–1055.
[65]. Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, 99–108.
[66]. Beylkin, G., and R. Burridge, 1990, Linearized inverse scattering problems in acoustics and elasticity: Wave Motion, **12**, 15–52.
[67]. Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation from seismic reflection data by stereotomography: Geophysical Journal International, **135**, 671–680.
[68]. Billette, F., S. Le Bégat, P. Podvin, and G. Lambaré, 2003, Practical aspects and applications of 2D stereotomography: Geophysics, **68**, 1008–1021.
[69]. Biondi, B., and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.
[70]. Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis, 2007, Structure of the California Coast Ranges and San Andreas fault at SAFOD from seismic waveform inversion and reflection imaging: Journal of Geophysical Research, doi: 10.1029/2006JB004611.
[71]. Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52**, 931–942.
[72]. Bouchon, M., M. Campillo, and S. Gaffet, 1989, A boundary integral equation — Discrete wavenumber representation method to study wave propagation in multilayered media having irregular interfaces: Geophysics, **54**, 1134–1140.
[73]. Brenders, A. J., and R. G. Pratt, 2007a, Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data: Geophysical Journal International, **168**, 152–170.
[74]. ———, 2007b, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.
[75]. ———, 2007c, Waveform tomography of marine seismic data: What can limited offset offer?: 75thAnnual International Meeting, SEG, Expanded Abstracts, 3024–3029.
[76]. Bube, K. P., and T. Nemeth, 2007, Fast line searches for the robust solution of linear systems in the hybrid 1/2 and Huber norms: Geophysics, **72**, no. 2, A13–A17.
[77]. Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics,

60, 1457–1473.
[78]. Capdeville, Y., Y. Gung, and B. Romanowicz, 2005, Towards global earth tomography using the spectral element method: A technique based on source stacking: Geophysical Journal International, **162**, 541–554.
[79]. Carcione, J. M., G. C. Herman, and A. P. E. ten Kroode, 2002, Seismic modeling: Geophysics, **67**, 1304–1325.
[80]. Cary, P., and C. Chapman, 1988, Automatic 1-D waveform inversion of marine seismic refraction data: Geophysical Journal of the Royal Astronomical Society, **93**, 527–546.
[81]. Chapman, C., 1985, Ray theory and its extension: WKBJ and Maslov seismograms: Journal of Geophysics, **58**, 27–43.
[82]. Chen, P., T. Jordan, and L. Zhao, 2007, Full three-dimensional tomography: A comparison between the scattering-integral and adjoint-wavefield methods: Geophysical Journal International, **170**, 175–181.
[83]. Chironi, C., J. V. Morgan, and M. R.Warner, 2006, Imaging of intra-basalt and sub-basalt structure with full wavefield seismic tomography: Journal of Geophysical Research, **11**, B05313, doi: 10.1029/2004JB003595.
[84]. Choi, Y., D.-J. Min, and C. Shin, 2008, Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupledmedia: Geophysical Prospecting, **56**, 863–881.
[85]. Claerbout, J., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.
[86]. ———, 1976, Fundamentals of geophysical data processing: McGraw-Hill Book Co., Inc.
[87]. Claerbout, J., and S. Doherty, 1972, Downward continuation of moveout corrected seismograms: Geophysics, **37**, 741–768.
[88]. Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic non-linear waveform inversion:Application to real data: Geophysics, **55**, 527–538.
[89]. Danecek, P., and G. Seriani, 2008, An efficient parallel Chebyshev pseudo-spectral method for large-scale 3D seismic forward modelling: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, P046.
[90]. de Hoop, A., 1960, A modification of Cagniard's method for solving seismic pulse problems: Applied Scientific Research, **8**, 349–356.
[91]. Dessa, J. X., and G. Pascal, 2003, Combined traveltime and frequency-domain seismic waveform inversion: A case study on multi-offset ultrasonic data: Geophysical Journal International, **154**, 117–133.
[92]. Devaney, A. J., 1982, A filtered back-projection algorithm for diffraction tomography: Ultrasonic Imaging, **4**, 336–350.
[93]. Devaney, A. J., and D. Zhang, 1991, Geophysical diffraction tomography in a layered background: Wave Motion, **14**, 243–265.
[94]. Djikpéssé, H., and A. Tarantola, 1999, Multiparameter 1-norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, **64**, 1023–1035.
[95]. Docherty, P., R. Silva, S. Singh, Z.-M. Song, and M. Wood, 2003, Migration velocity analysis using a genetic algorithm: Geophysical Prospecting, **45**, 865–878.
[96]. Dumbser, M., and M. Kaser, 2006, An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes — II. The three-dimensional isotropic case: Geophysical Journal International, **167**, 319–336.
[97]. Dummong, S., K. Meier, D. Gajewski, and C. Hubscher, 2008, Comparison of prestack stereotomography and NIP wave tomography for velocity model building: Instances from the Messinian evaporites: Geophysics, **73**, no. 5, VE291–VE302.
[98]. Ellefsen, K., 2009, A comparison of phase inversion and traveltime tomography for processing of near-surface refraction traveltimes: Geophysics, VOL. 74, NO. 6; P. WCB11–WCB24.
[99]. Epanomeritakis, I., V. Akçelik, O. Ghattas, and J. Bielak, 2008, A Newton CG method for large-scale three-dimensional elastic full waveform seismic inversion: Inverse Problems, **24**, 1–26.
[100]. Erlangga, Y. A., and F. J. Herrmann, 2008, An iterative multilevel method for computing wavefields in frequency-domain seismic inversion: 78thAnnual International Meeting, SEG, ExpandedAbstracts, 1956–1960.
[101]. Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, no. 5, J53–J64.
[102]. Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2008, Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain: Geophysical Journal International, **175**, 665–685.
[103]. Forgues, E., and G. Lambaré, 1997, Parameterization study for acoustic and elastic ray Born inversion: Journal of Seismic Exploration, **6**, 253–278.
[104]. Futterman, W., 1962, Dispersive body waves: Journal of Geophysics Research, **67**, 5279–5291.
[105]. Gao, F., A. R. Levander, R. G. Pratt, C. A. Zelt, and G. L. Fradelizio, 2006a, Waveform tomography at a groundwater contamination site: Surface reflection data: Geophysics, **72**, no. 5, G45–G55.
[106]. ———, 2006b, Waveform tomography at a groundwater contamination site: VSP-surface data set:

Geophysics, **71**, no. 1, H1–H11.
[107]. Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.
[108]. Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geophysics, **43**, 1342–1351.
[109]. Gelis, C., J. Virieux, and G. Grandjean, 2007, 2D elastic waveform inversion using Born and Rytov approximations in the frequency domain: Geophysical Journal International, **168**, 605–633.
[110]. Gelius, L. J., I. Johansen, N. Sponheim, and J. J. Stamnes, 1991, A generalized diffraction tomography algorithm: Journal of the Acoustical Society of America, **89**, 523–528.
[111]. Gilbert, F., and A. Dziewonski, 1975, An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra: Philosophical Transactions of the Royal Society of London, **278**, 187–269.
[112]. Graves, R., 1996, Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences: Bulletin of the Seismological Society of America, **86**, 1091–1106.
[113]. Grechka, V., L. Zhang, and J.W. Rector III, 2004, Shear waves in acoustic anisotropic media: Geophysics, **69**, 576–582.
[114]. Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, 1310–1319.
[115]. Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a Huber function norm: Geophysics, **74**, no. 3, R15–R24.
[116]. Hak, B., and W. Mulder, 2008, Preconditioning for linearized inversion of attenuation and velocity perturbations: 78th Annual International Meeting, SEG, Expanded Abstracts, 2036–2040.
[117]. Hansen, C., ed., 1998, Rank-defficient and discrete ill-posed problems —Numerical aspects of linear inversion: Society for Industrial and Applied Mathematics.
[118]. Herrmann, F. J., Y. Erlangga, and T. T.Y. Lin, 2009, Compressive sensing applied to full-wave form inversion: 71st Conference & Technical Exhibition, EAGE, Extended Abstracts, S016.
[119]. Hicks, G. J., and R. G. Pratt, 2001, Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit: Geophysics, **66**, 598–612.
[120]. Hole, J. A., 1992, Nonlinear high-resolution three-dimensional seismic travel time tomography: Journal of Geophysical Research, **97**, 6553–6562.
[121]. Hu, W., A. Abubakar, and T. M. Habashy, 2009, Simultaneous multifrequency inversion of full-waveform seismic data: Geophysics, **74**, no. 2, R1–R14.
[122]. Hustedt, B., S. Operto, and J. Virieux, 2004, Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling: Geophysical Journal International, **157**, 1269–1296.
[123]. Ikelle, L., J. P. Diet, and A. Tarantola, 1988, Linearized inversion of multioffset seismic reflection data in the -*k* domain: Depth-dependent reference medium: Geophysics, **53**, 50–64.
[124]. Improta, L., A. Zollo, A. Herrero, R. Frattini, J. Virieux, and P. DellÁversana, 2002, Seismic imaging of complex structures by non-linear traveltimes inversion of dense wide-angle data:Application to a thrust belt: Geophysical Journal International, **151**, 264–278.
[125]. Jaiswal, P., C. Zelt, A. W. Bally, and R. Dasgupta, 2008, 2-D traveltime and waveform inversion for improved seismic imaging: Naga thrust and fold belt, India: Geophysical Journal International, doi: 10.1111/j.1365-246X.2007.03691.x.
[126]. Jaiswal, P., C. Zelt, R. Dasgupta, and K. Nath, 2009, Seismic imaging of the Naga thrust using multiscale waveform inversion: Geophysics, VOL. 74, NO. 6; P. WCC129–WCC140.
[127]. Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, and D. Trézéguet, 1989,Wavelengths of earth structures that can be resolved from seismic reflection data: Geophysics, **54**, 906–910.
[128]. Jin, S., and R. Madariaga, 1993, Background velocity inversion by a genetic algorithm: Geophysical Research Letters, **20**, 93–96.
[129]. ———, 1994, Non-linear velocity inversion by a two-step Monte Carlo method: Geophysics, **59**, 577–590.
[130]. Jin, S., R. Madariaga, J. Virieux, and G. Lambaré, 1992, Two-dimensional asymptotic iterative elastic inversion: Geophysical Journal International, **108**, 575–588.
[131]. Jones, K., M. Warner, and J. Brittan, 1999, Anisotropy in multi-offset deepcrustal seismic experiments: Geophysical Journal International, **138**, 300–318.
[132]. Kamei, R., and R. G. Pratt, 2008, Waveform tomography strategies for imaging attenuation structure for cross-hole data: 70th Conference & Technical Exhibition, EAGE, ExtendedAbstracts, F019.
[133]. Kennett, B. L. N., 1983, Seismic wave propagation in stratified media: Cambridge University Press.
[134]. Klem-Musatov, K. D., and A. M. Aizenberg, 1985, Seismic modelling by methods of the theory of edge

waves: Journal of Geophysics, **57**, 90–105.
[135]. Kolb P., F. Collino, and P. Lailly, 1986, Prestack inversion of 1-D medium: Proceedings of the IEEE, 498–508.
[136]. Kolsky, H., 1956, The propagation of stress pulses in viscoelastic solids: Philosophical Magazine, **1**, 693–710.
[137]. Komatitsch, D., and J. P. Vilotte, 1998, The spectral element method:An efficient tool to simulate the seismic response of 2D and 3D geological structures: Bulletin of the Seismological Society ofAmerica, **88**, 368–392.
[138]. Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, **96**, 20289–20299.
[139]. Kormendi, F., and M. Dietrich, 1991, Non-linear waveform inversion of plane-wave seismograms in stratified elastic media: Geophysics, **56**, 664–674.
[140]. Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, A. Baumstein, M. D. Lacasse, and S. Lee, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, VOL. 74, NO. 6; P. WCC177–WCC188.
[141]. Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.
[142]. Lambaré, G., 2008, Stereotomography: Geophysics, **73**, no. 5, VE25–VE34.
[143]. Lambaré, G., and M. Alérini, 2005, Semi-automatic picking PP-PS stereotomography:Application to the synthetic Valhall data set: 75thAnnual International Meeting, SEG, ExpandedAbstracts, 943–946.
[144]. Lambaré, G., S. Operto, P. Podvin, P. Thierry, and M. Noble, 2003, 3-D ray Born migration/inversion — Part 1: Theory: Geophysics, **68**, 1348–1356.
[145]. Lambaré, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, **57**, 1138–1154.
[146]. Lecomte, I., 2008, Resolution and illumination analyses in PSDM: A raybased approach: The Leading Edge, **27**, 650–663.
[147]. Lecomte, I., S.-E. Hamran, and L.-J. Gelius, 2005, Improving Kirchhoff migration with repeated local plane-wave imaging? A SAR-inspired signalprocessing approach in prestack depth imaging: Geophysical Prospecting, **53**, 767–785.
[148]. Lee, K. H., and H. J. Kim, 2003, Source-independent full-waveform inversion of seismic data: Geophysics, **68**, 2010–2015.
[149]. Lehmann, I., 1936, P: Publications du Bureau Central Séismologique International, A14, 87–115.
[150]. Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436.
[151]. Liao, Q., and G. A. McMechan, 1995, 2.5D full-wavefield viscoacoustic inversion: Geophysical Prospecting, **43**, 1043–1059.
[152]. ———, 1996, Multifrequency viscoacoustic modeling and inversion: Geophysics, **61**, 1371–1378.
[153]. Lions, J., 1972, Nonhomogeneous boundary value problems and applications: Springer-Verlag, Berlin.
[154]. Lucio, P. S., G. Lambaré, and A. Hanyga, 1996, 3D multivalued travel time and amplitude maps: Pure andApplied Geophysics, **148**, 113–136.
[155]. Luo, Y., T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Seismic modeling and imaging based upon spectral-element and adjoint methods: The Leading Edge, **28**, 568–574.
[156]. Malinowsky, M., and S. Operto, 2008, Quantitative imaging of the permomesozoic complex and its basement by frequency domain waveform tomography of wide-aperture seismic data from the polish basin: Geophysical Prospecting, **56**, 805–825.
[157]. Malinowsky, M., A. Ribodetti, and S. Operto, 2007, Multiparameter fullwaveform inversion for velocity and attenuation — Refining the imaging of a sedimentary basin: 69th Conference & Technical Exhibition, EAGE, ExtendedAbstracts, P276.
[158]. Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.
[159]. McMechan, G. A., and G. S. Fuis, 1987, Ray equation migration of wide-angle reflections from southern Alaska: Journal of Geophysical Research, **92**, 407–420.
[160]. Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.
[161]. Miller, D., M. Oristaglio, and G. Beylkin, 1987, Anew slant on seismic imaging: Migration and integral geometry: Geophysics, **52**, 943–964.
[162]. Min, D.-J., and C. Shin, 2006, Refraction tomography using a waveform-inversion back-propagation technique: Geophysics, **71**, no. 3, R21–R30.
[163]. Min, D.-J., C. Shin, R. G. Pratt, and H. S. Yoo, 2003, Weighted-averaging finite-element method for 2D

elastic wave equations in the frequency domain: Bulletin of the Seismological Society ofAmerica, **93**, 904–921.
[164]. Montelli, R., G. Nolet, F. Dahlen, G. Masters, E. Engdahl, and S. Hung, 2004, Finite-frequency tomography reveals a variety of plumes in the mantle: Science, **303**, 338–343.
[165]. Mora, P. R., 1987, Nonlinear two-dimensional elastic inversion of multi-offset seismic data: Geophysics, **52**, 1211–1228.
[166]. ———, 1988, Elastic wavefield inversion of reflection and transmission data: Geophysics, **53**, 750–759.
[167]. Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447.
[168]. Mulder, W. A., and B. Hak, 2009, Simultaneous imaging of velocity and attenuation perturbations from seismic data is nearly impossible: 71st Conference & Technical Exhibition, EAGE, ExtendedAbstracts, S043.
[169]. Nihei, K. T., and X. Li, 2007, Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection TD-PSD: Geophysical Journal International, **169**, 1069–1078.
[170]. Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.
[171]. Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer.
[172]. Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics: D. Reidel Publ. Co.
[173]. Oldham, R., 1906, The constitution of the earth: Quarterly Journal of the Geological Society of London, **62**, 456–475.
[174]. Operto, S., G. Lambaré, P. Podvin, and P. Thierry, 2003, 3-D ray-Born migration/inversion, Part 2: Application to the SEG/EAGE overthrust experiment: Geophysics, **68**, 1357–1370.
[175]. Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Quantitative imaging of complex structures from multifold wide aperture seismic data: A case study: Geophysical Prospecting, **52**, 625–651.
[176]. Operto, S., J. Virieux, P. Amestoy, J. Y. L'Excellent, L. Giraud, and H. BenHadj-Ali, 2007, 3D finite-difference frequency-domain modeling of viscoacoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, no. 5, SM195–SM211.
[177]. Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain fullwaveform tomography: Application to the eastern Nankai trough: Journal of Geophysical Research, doi: 10.1029/2005JB003835.
[178]. Operto, S., J. Virieux, A. Ribodetti, and J. E. Anderson, 2009, Finite-difference frequency-domain modeling of viscoacoustic wave propagation in two-dimensional tilted transversely isotropic media: Geophysics, **74**, no.5, T75–T95, doi: 10.1190/1.3157243.
[179]. Operto, S., S. Xu, and G. Lambaré, 2000, Can we image quantitatively complex models with rays?: Geophysics, **65**, 1223–1238.
[180]. Paige, C. C., and M. A. Saunders, 1982a, Algorithm 583 LSQR: Sparse linear equations and least squares problems: Association for Computing Machinery Transactions on Mathematical Software, **8**, 195–209.
[181]. ———, 1982b, LSQR: An algorithm for sparse linear equations and sparse least squares: Association for Computing Machinery Transactions on Mathematical Software, **8**, 43–71.
[182]. Pica, A., J. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in laterally invariant medium: Geophysics, **55**, 284–292.
[183]. Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.
[184]. ———, 2007, AHelmholtz iterative solver for 3D seismic-imaging problems: Geophysics, **72**, no. 5, SM185–SM194.
[185]. ———, 2009, 3D frequency-domain full-waveform inversion with an iterative solver: Geophysics, VOL. 74, NO. 6, P. WCC149–WCC157.
[186]. Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method: Geophysical Prospecting, **38**, 311–330.
[187]. ———, 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.
[188]. Pratt, R. G., and N. R. Goulty, 1991, Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data: Geophysics, **56**, 204–224.
[189]. Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.
[190]. Pratt, R. G., L. Sirgue, B. Hornby, and J. Wolfe, 2008, Cross-well waveform tomography in fine-layered sediments — Meeting the challenges of anisotropy: 70th Conference & Technical Exhibition, EAGE, Extended

Abstracts, F020.
[191]. Pratt, R. G., Z. M. Song, P. R.Williamson, and M.Warner, 1996, Two-dimensional velocity model from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, 323–340.
[192]. Pratt, R. G., andW. Symes, 2002, Semblance and differential semblance optimisation for waveform tomography:Afrequency domain implementation: Sub-basalt imaging conference, ExpandedAbstracts, 183–184.
[193]. Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multisource cross-hole tomography, Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310.
[194]. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Veltering, 1986, Numerical recipes: The art of scientific computing: Cambridge University Press.
[195]. Prieux, V., S. Operto, R. Brossier, and J. Virieux, 2009, Application of acoustic full waveform inversion to the synthetic Valhall model: 77thAnnual International Meeting, SEG, ExpandedAbstracts, 2268–2272.
[196]. Pyun, S., C. Shin, and J. B. Bednar, 2007, Comparison of waveform inversion, Part 3: Phase approach: Geophysical Prospecting, **55**, 465–475.
[197]. Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Multi-scale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, **159**, 1032–1056.
[198]. Ribodetti, A., S. Operto, J. Virieux, G. Lambaré, H.-P. Valéro, and D. Gibert, 2000, Asymptotic viscoacoustic diffraction tomography of ultrasonic laboratory data: A tool for rock properties analysis: Geophysical Journal International, **140**, 324–340.
[199]. Riyanti, C. D., Y. A. Erlangga, R. E. Plessix, W. A. Mulder, C. Vuik, and C. Oosterlee, 2006, A new iterative solver for the time-harmonic wave equation: Geophysics, **71**, no. 5, E57–E63.
[200]. Riyanti, C. D., A. Kononov, Y. A. Erlangga, C. Vuik, C. W. Oosterlee, R. E. Plessix, and W. A. Mulder, 2007, A parallel multigrid-based preconditioner for the 3D heterogeneous high-frequency Helmholtz equation: Journal of Computational Physics, **224**, 431–448.
[201]. Robertsson, J. O., B. Bednar, J. Blanch, C. Kostov, and D. J. van Manen, 2007, Introduction to the supplement on seismic modeling with applications to acquisition, processing and interpretation: Geophysics, **72**, no. 5, SM1–SM4.
[202]. Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.
[203]. Saad, Y., 2003, Iterative methods for sparse linear systems, 2nd ed.: Society for Industrial andApplied Mathematics.
[204]. Sambridge, M., and G. Drijkoningen, 1992, Genetic algorithms in seismic waveform inversion: Geophysical Journal International, **109**, 323–342.
[205]. Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, 1–29.
[206]. Sambridge, M. S., A. Tarantola, and B. L. N. Kennett, 1991, An alternative strategy for non linear inversion of seismic waveforms: Geophysical Prospecting, **39**, 457–491.
[207]. Scales, J. A., P. Docherty, and A. Gersztenkorn, 1990, Regularization of nonlinear inverse problems: Imaging the near-surface weathering layer: Inverse Problems, **6**, 115–131.
[208]. Scales, J. A., and M. L. Smith, 1994, Introductory geophysical inverse theory: Samizdat Press.
[209]. Sears, T. J., S. C. Singh, and P. J. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56**, 843–862.
[210]. Sen, M. K., and P. L. Stoffa, 1995, Global optimization methods in geophysical inversion: Elsevier Science Publ. Co., Inc. Sheng, J., A. Leeds, M. Buddensiek, and G. T. Schuster, 2006, Early arrival waveform tomography on near-surface refraction data: Geophysics, **71**,
no. 4, U47–U57.
[211]. Shi, Y., W. Zhao, and H. Cao, 2007, Nonlinear process control of wave-equation inversion and its application in the detection of gas: Geophysics, **72**, no. 1, R9–R18.
[212]. Shin, C., and W. Ha, 2008, A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains: Geophysics, **73**, no. 5, VE119–VE133.
[213]. Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931.
[214]. ———, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, **177**, 1067–1079.
[215]. Shin, C., S. Jang, and D. J. Min, 2001a, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.
[216]. Shin, C., and D.-J. Min, 2006, Waveform inversion using a logarithmic wavefield: Geophysics, **71**, no. 3,

R31–R42.
[217]. Shin, C., S. Pyun, and J. B. Bednar, 2007, Comparison of waveform inversion, Part 1: Conventional wavefield vs. logarithmic wavefield: Geophysical Prospecting, **55**, 449–464.
[218]. Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin, 2001b, Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion: Geophysics, **66**, 1856–1863.
[219]. Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data: Geophysical Journal International, **151**, 325–344.
[220]. Snieder, R., M. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem: Geophysics, **54**, 991–1000.
[221]. Song, Z., P. Williamson, and G. Pratt, 1995, Frequency-domain acousticwave modeling and inversion of crosshole data, Part 2: Inversion method, synthetic experiments and real-data results: Geophysics, **60**, 786–809.
[222]. Sourbier, F., A. Haiddar, L. Giraud, S. Operto, and J. Virieux, 2008, Frequency-domain full-waveform modeling using a hybrid direct-iterative solver based on a parallel domain decomposition method: A tool for 3D fullwaveform inversion?: 78th Annual International Meeting, SEG, ExpandedAbstracts, 2147–2151.
[223]. Sourbier, F., S. Operto, J. Virieux, P. Amestoy, and J. Y. L'Excellent, 2009a, FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data — Part 1: Algorithm: Computers and Geosciences, **35**, 487–495.
[224]. ———, 2009b, FWT2D:Amassively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data — Part 2: Numerical examples and scalability analysis: Computers & Geosciences, **35**, 496–514.
[225]. Stekl, I., and R. G. Pratt, 1998, Accurate viscoelastic modeling by frequencydomain finite difference using rotated operators: Geophysics, **63**, 1779–1794.
[226]. Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, **43**, 23–48.
[227]. Taillandier, C., M. Noble, H. Chauris, and H. Calandra, 2009, First-arrival travel time tomography based on the adjoint state method: Geophysics, VOL. 74, NO. 6; P. WCB1–WCB10.
[228]. Talagrand, O., and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation, I: Theory: Quarterly Journal of the Royal Meteorological Society, **113**, 1311–1328.
[229]. Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.
[230]. ———, 1986, A strategy for nonlinear inversion of seismic reflection data: Geophysics, **51**, 1893–1903.
[231]. ———, 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Science Publ. Co., Inc.
[232]. Thierry, P., G. Lambaré, P. Podvin, and M. Noble, 1999a, 3-D preserved amplitude prestack depth migration on a workstation: Geophysics, **64**, 222–229.
[233]. Thierry, P., S. Operto, and G. Lambaré, 1999b, Fast 2D ray-Born inversion/migration in complex media: Geophysics, **64**, 162–181.
[234]. Thomsen, L. A., 1986,Weak elastic anisotropy: Geophysics, **51**, 1954–1966.
[235]. van den Berg, P. M., and A. Abubakar, 2001, Contrast source inversion method: State of the art: Progress in Electromagnetics Research, **34**, 189–218.
[236]. Vigh, D., and E. W. Starr, 2008a, 3D plane-wave full-waveform inversion: Geophysics, **73**, no. 5, VE135–VE144.
[237]. ———, 2008b, Comparisons for waveform inversion, time domain or frequency domain?: 78thAnnual International Meeting, SEG, Expanded Abstracts, 1890–1894.
[238]. Vigh, D., E. W. Starr, and J. Kapoor, 2009, Developing earth model with full waveform inversion: The Leading Edge, **28**, 432–435.
[239]. Virieux, J., 1986, P-SV wave propagation in heterogeneous media, velocitystress finite difference method: Geophysics, **51**, 889–901.
[240]. Virieux, J., S. Operto, H. B. H. Ali, R. Brossier, V. Etienne, F. Sourbier, L. Giraud, and A. Haidar, 2009, Seismic wave modeling for seismic imaging: The Leading Edge, **28**, 538–544.
Vogel, C., 2002, Computational methods for inverse problems: Society of Industrial and Applied Mathematics.
[241]. Vogel, C. R., and M. E. Oman, 1996, Iterative methods for total variation denoising: SIAM Journal on Scientific Computing, **17**, 227–238.
[242]. Wang, Y., and Y. Rao, 2009, Reflection seismic waveform tomography: Journal of Geophysical Research, doi: 10.1029/2008JB005916.
[243]. Warner, M., I. Stekl, and A. Umpleby, 2007, Full wavefield seismic tomography — Iterative forward modeling in 3D: 69th Conference & Technical Exhibition, EAGE, ExtendedAbstracts, C025.

[244]. ———, 2008, 3D wavefield tomography: synthetic and field data examples: 78th Annual International Meeting, SEG, Expanded Abstracts, 3330–3334.
[245]. Williamson, P., 1991, A guide to the limits of resolution imposed by scattering in ray tomography: Geophysics, **56**, 202–207.
[246]. Woodhouse, J., and A. Dziewonski, 1984, Mapping the upper mantle: Three dimensional modelling of earth structure by inversion of seismic waveforms: Journal of Geophysical Research, **89**, 5953–5986.
[247]. Woodward, M. J., 1992, Wave-equation tomography: Geophysics, **57**, 15–26.
[248]. Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: Geophysics, **73**, no. 5, VE5–VE11.
[249]. Wu, R. S., 2003, Wave propagation, scattering and imaging using dual-domain one-way and one-return propagation: Pure and Applied Geophysics, **160**, 509–539.
[250]. Wu, R. S., and K. Aki, 1985, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, 582–595.
[251]. Wu, R.-S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11–25.
[252]. Yilmaz, O., 2001, Seismic data analysis: processing, inversion and interpretation of seismic data: SEG.
[253]. Zelt, C., and P. J. Barton, 1998, Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe basin: Journal of Geophysical Research, **103**, 7187–7210.
[254]. Zelt, C. A., R. G. Pratt, A. J. Brenders, S. Hanson-Hedgecock, and J. A. Hole, 2005, Advancements in long-offset seismic imaging:Ablind test of traveltime and waveform tomography: Eos, Transactions of the American Geophysical Union, **86**,Abstract S52A-04.
[255]. Zhou, B., and S. A. Greenhalgh, 2003, Crosshole seismic inversion with normalized full-waveform amplitude data: Geophysics, **68**, 1320–1330.

人体勘探与油气勘探方法技术的对比分析

王华忠[1]，周阳[2]

①同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

②成都理工大学，成都，610059

【摘要】受绿色能源发展、可持续发展和气候变暖等综合因素的影响，化石能源行业的未来不令人乐观。油气地震勘探的核心工作是地震数据处理与地震波成像，这与广义遥感领域中的不同技术方向有密切的联系。所谓广义遥感指的是：在目标物理系统外面观测该物理系统的状态变量，获得实际观测数据，通过对实测数据的处理与分析，获取关于目标物理系统内部介质性质的信息，进而解决实际应用问题。鉴于油气勘探行业可以预见的、不可避免的萎缩和式微，非常有必要对广义遥感领域中的相关技术方向进行对比分析，探索勘探地震中的数据分析方法推广应用到相关领域中的可能性。我们认为医学成像在今后相当长一段时间内有重要的发展和应用前景，因此本文试图把医学成像与地震成像进行对比。我们把利用各种物理场对人体内部介质状况的探测称为人体勘探，既方便与油气地震勘探进行对比，也清楚地显示出医学成像仅仅是人体勘探的一个重要环节。这正如地震波成像之于油气地震勘探的关系。无论是人体勘探或是油气勘探都是广义遥感的一个特定应用领域。核心问题都涉及到数据采集、数据预处理、介质参数估计或成像、成像结果的解释等。因此，本文从两种探测方式的介质特点、数据特点、观测方式、成像方法进行对比分析，试图找到把油气地震勘探中的方法技术推广应用到人体勘探中的切入点。

【关键词】可持续发展，油气勘探，人体勘探，介质特点，数据特点，成像方法对比

1 引言

受绿色能源发展、可持续发展和气候变暖等综合因素的影响，化石能源行业的未来不令人乐观。中国政府公布到 2060 年实现碳中和；北欧有些国家宣布到 2035 年实现碳中和。美国拜登政府也宣布重新加入巴黎气候协定。所有这些都预示着化石能源在当今社会的能源结构中的主要地位要逐渐让位于绿色能源，这是人类永久存续与发展必须要走的路线。

作为石油工业界的从业人员或与之相关联的行业人员，必须顺应这个社会发展趋势。正所谓社会大势浩浩荡荡，顺之者昌逆之者亡。

勘探地震学的学科属性应该是数据分析学科在油气勘探中的应用，属于广义遥感（Remote Sensing）领域。很显然，顺应当前变化趋势的、最快捷的转变方式是把勘探地震中积累的特色知识迁移应用到密切相关的领域中。这样的领域包括：超声无损探伤、工程/城市地球物理、医学成像、水声探测/通讯、雷达与遥感领域。更直接的做法是完全转向大数据分析和人工智能及它们的应用中。

一般地，认为实测数据是来源于某种物理系统输出。且都可以视为某种波场的记录。从甚低频到γ射线可以统一地用频率和波长来区分上述不同应用领域的广义遥感（Remote Sensing）。

勘探地震学中数据分析的特殊性体现在：它面对的地下岩石介质；勘探震源的激发；波在地下岩石介质中的传播；地震检波器接收属性；以及地表放置的炮检观测系统。勘探地震中，波穿透复杂的地下介质、野外数据观测带来复杂的信号预处理问题、仅仅地表观测方式带来的信息不充分情形下的参数估计问题、地下介质完全未知问题、成像结果的地质解释问题、地震数据分析的目标是油藏的岩性参数和流体参数等等，使得地震数据分析极为复杂。基本可以说，在所有广义遥感问题中（水声探测、地震波探测、超声波探测、电磁波探测、X-射线探测、γ射线探测等等），地震波数据分析与成像、地震数据解释是最困难的。

可以自信地讲：能把地震数据分析的精髓和方法技术理解和掌握很好，其他领域的数据分析都不在话下。但是，由于油气勘探工业界的逐渐式微，皮之不存毛将焉附！我们必须把所拥有的对勘探地震数据分析和成像的能力迁移到解决相关领域的数据分析问题上。

目前，超声成像和水声成像与地震波成像的关系最密切。电磁波雷达（与遥感）也与地震波成像有比较

密切的关系。这几个领域，波传播的理论框架基本一致，都满足双曲型波动方程。应该把水声成像、地震成像、超声成像统一在一起进行研究。（电磁）雷达波成像也应该纳入到统一的研究过程中。尤其是它们的参数估计问题应完全放在同一个理论框架下，即 Bayes 估计理论下，不同的仅仅是正问题和实测数据的特点。

不同遥感应用领域的信号处理和图像处理的理论、方法和技术的差异性更小。尤其是信号分析领域，可以说完全用相同的方法技术。

分析认为：比较而言，（超声）医学成像在今后相当长一段时间内有重要的发展和相对广阔的应用前景。因此本文试图把医学成像与地震成像进行对比。我们把利用各种物理场对人体内部介质状况的探测称为人体勘探，既方便与油气地震勘探进行对比，也清楚地显示出医学成像仅仅是人体勘探的一个重要环节。这正如地震波成像之于油气地震勘探的关系。无论是人体勘探或是油气勘探都是广义遥感的一个特定应用领域。核心问题都涉及到数据采集、数据预处理、介质参数估计或成像、成像结果的解释等。因此，本文从两种探测方式的介质特点、数据特点、观测方式、成像方法进行对比分析，试图找到把油气地震勘探中的方法技术推广应用到人体勘探中的切入点。

2 人体勘探的介质与数据观测

开展任何广义遥感领域的研究，首先应对目标介质系统有深入的理解。在此基础上，才能更好地理解波在此介质系统中的传播特征，也才能更好地理解观测数据的特点。对介质、波场和数据特点的把握是高精度成像的基础。

和油气勘探类似，人体勘探的目的是：寻找人体介质中的异常体的空间位置，最好估计出异常体与周围介质的性质差异。这些异常往往对应于人体的病变。

但是，人体介质显然不同于地下岩性介质。医学解剖学已经把正常的人体介质性质描述得很清楚了。不清楚的是人体介质发生的异常情况。人体介质主要包括骨骼、各种组织和流动的血液。甚高频电磁波（譬如 X-射线和 γ 射线等）穿透人体介质不同，衰减量不同。人体介质也是由存在波阻抗差异的骨骼、组织和流动的血液等组成的，超声波在人体介质中传播会产生反射波。下表展示了超声医学成像中关注的几种介质的声波速度和波阻抗。

表 1 超声医学成像中关注的几种介质的声波速度和波阻抗

材料	速度（m/s）	阻抗（$Z=\rho c$）10^6kg/m^{2}s
空气（25^{o}C）	346	0.000410
脂肪	1450	1.38
水（25^{o}C）	1493	1.48
软组织	1540	1.63
肝脏	1550	1.64
血液（37^{o}C）	1570	1.67
骨骼	4000	3.8 到 7.4
铝	6320	17.0

观测系统与波场特点对成像而言是至关重要的问题。尽管我们重点对比超声医学成像（超声医学包括超声疗法和超声诊断，超声成像是超声诊断技术），但是为了完整性，还是简述一下几种主要的医学成像方法的数据观测。

我们可以认为：人体勘探利用了从甚低频到 γ 射线这个广泛波段中的几个典型的波段。如图 1 所示。上图展示了从次声波、声波到超声波的频带范围；下图展示了从超声波波段过渡到无线电波波段，微波波段、红外波段、可见光波段、紫外波段、直到 X-射线波段和 γ 射线波段。医学成像主要用 X-射线波段和 γ 射线波段以及超声波波段。油气勘探地震中，地震波用的是大约 1Hz-120Hz 范围的波段，跨越次声波和声波波段。

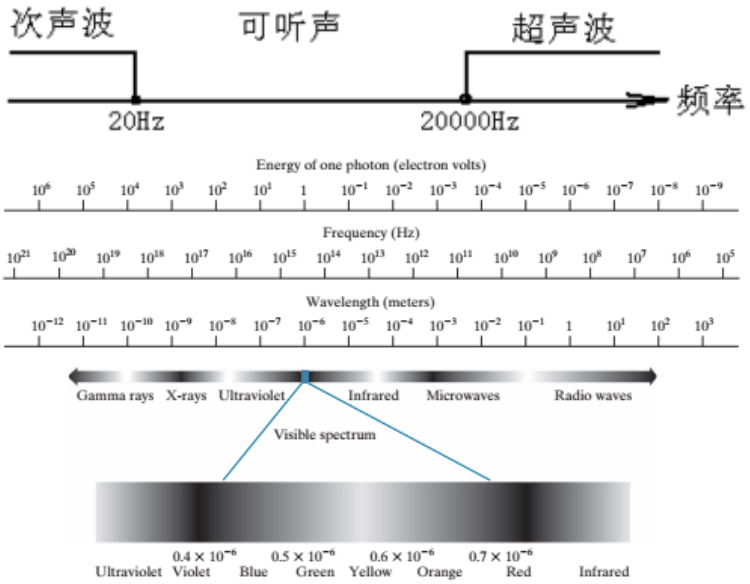


图 1 各种波的波段分布情况

和勘探地震数据观测方式相比，超声医学成像的源是基于压电效应的超声换能器，而且源和检是共体的，因此自激自收是主要的观测方式。如下图所示。

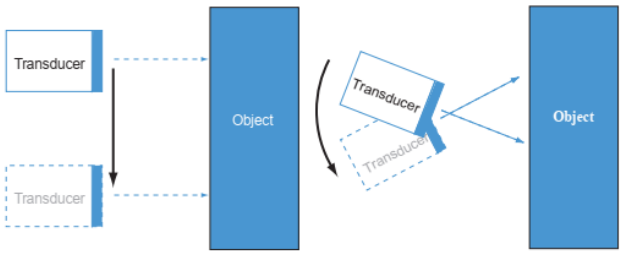


图 2a 超声医学成像 2D 数据观测系统。两种 B 超观测系统，左图是滑动式；右图是旋转式。

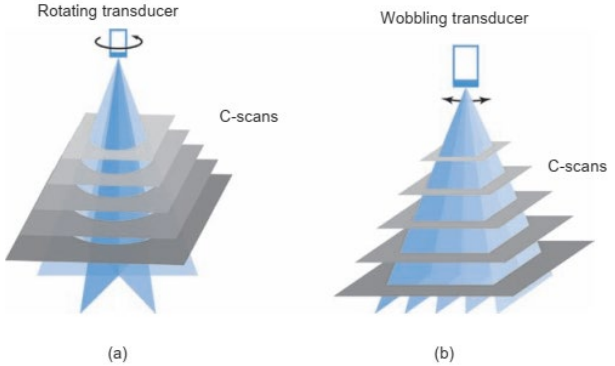


图 2b 超声医学成像 3D 数据观测系统。C 超观测系统。可以实现三维自激自收的三维数据观测。

从图 2 所示的观测系统可见，目前超声医学成像的核心不在于数据处理和成像，而是换能器的制造！与成像目标尺度相比，超声医学中超声波波长很短，观测的主要是反射波，可以用射线理论描述波传播。总结可知：超声医学成像中实测数据仅仅是零偏移距数据，没有激发和接收子波的概念，数据中仅仅有纵波分量，没有其他波现象的干扰。因此，超声医学成像根本没有涉及到线性反演成像理论下的成像问题。

X-Ray 成像技术是医学成像的另一种重要的技术形式。下图展示了 X-Ray 成像中观测系统的发展。从图中可见，第一代观测系统是单射线的激发与接收；第二代观测系统进化到扇状射线的激发与接收；第三代观测系统发展到扇状射线的弧形接收；第四代观测系统进展到接收器分布固定的圆环上，激发源围绕人体转动；最新一代观测系统发展到三维接收。但无论如何，X-Ray 成像技术仅仅观测沿直射线的投影，数据没有噪音（只有精度），没有波现象。

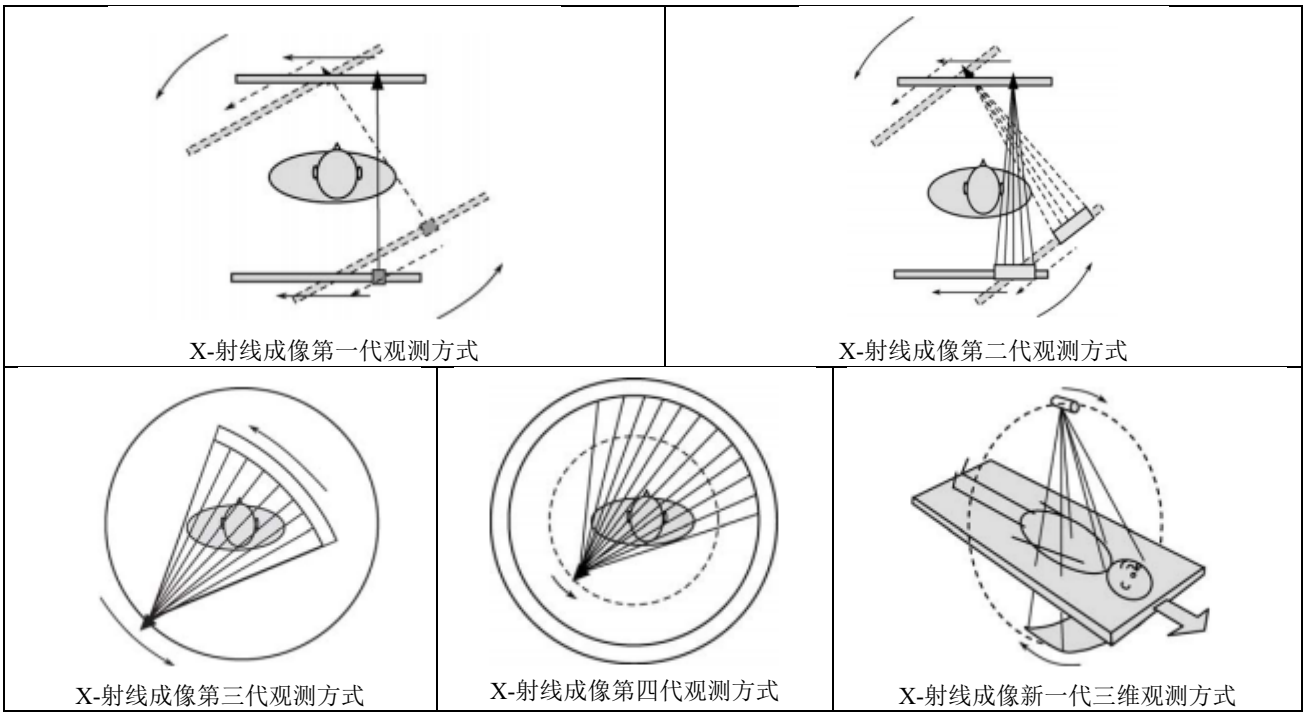


图 3 X-射线成像若干代观测系统及演化过程

其他几种典型的医学成像技术的观测系统与 X-射线成像技术是类似的，譬如核磁共振成像技术、正电子发射层析成像技术都是基于直射线投影值的观测结果进行成像。本质上，与 X-射线层析是完全一样的。

基本可以断定：医学成像的观测系统和数据复杂度远远低于地震勘探，医学成像方法的复杂程度绝不会超过地震波成像。

3 油气勘探的介质与数据观测

油气地震勘探面对的是复杂的地表和地下介质分布。任何一个新探区，介质分布都是未知的。油气地震勘探中，关于介质性质，我们知道的仅仅是地质和岩石物理的一般性知识。而人体勘探中，人体介质性质的正常状态是已知的，解剖学已经弄清楚了人体介质性质的正常状态。油气勘探中，钻井信息远没有达到人体勘探解剖学的地步，也不可能做到。

地震勘探中，源本身及与介质之间的关系、波传播及波与介质之间的关系、检波器本身及与介质的关系都非常复杂。勘探地震面对是矢量的弹性波，不是标量的声波。勘探地震中地下异常地质体的尺度与地震波长往往可以相比拟，绕射和散射波往往非常发育。勘探地震中，地震数据采集一般地只能在地表进行激发和接收，严重约束了数据采集孔径。勘探地震中震源激发的子波是重要的，它携带了与介质相关的信息。而超声医学成像中没有子波的概念，映衬出超声医学成像发展还比较低端。地震数据观测在野外进行，外源噪音非常强、随机噪音也比较强、不能用于成像的干扰波现象也非常强。

下图展示了地震数据特点。从图 4 可以看出，地震数据的极端复杂程度，有效的反射波很难从点炮道集上展示出来。

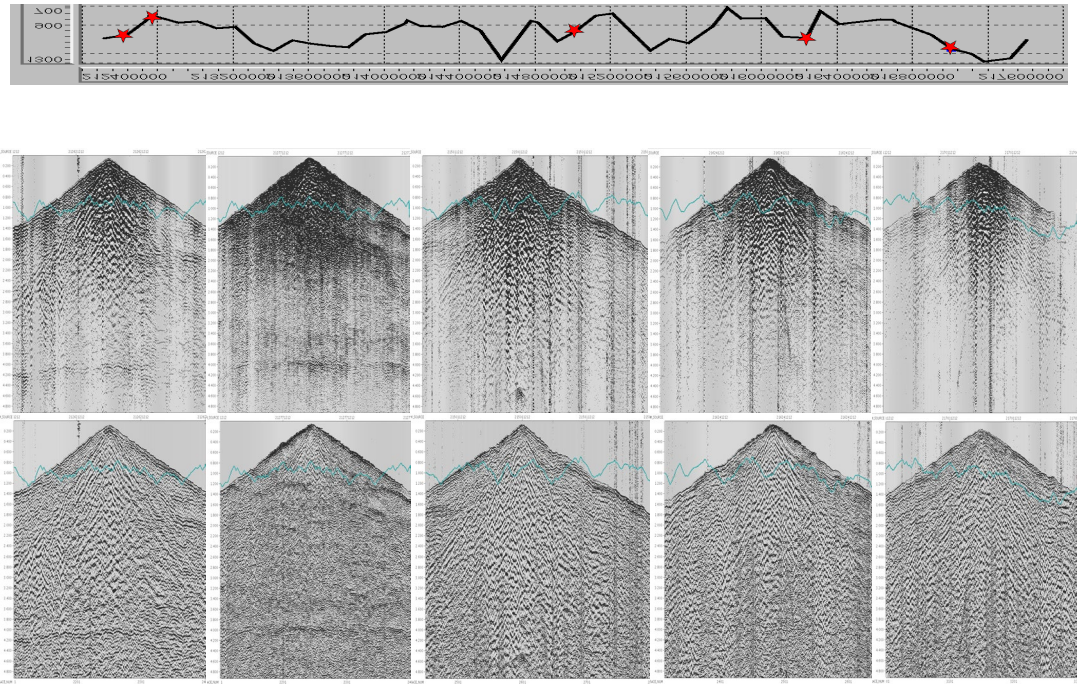


图 4 山前带地震数据-单炮道集。上图标识出各炮的采集位置。中图是未加 AGC 的显示。下图是 AGC 后的显示。

正是地震波数据采集方式的特殊性、地震数据和地震波场表现出的特殊性，才使得地震波成像技术独具特色。地震数据分析和地震波成像在油气勘探中占据了举足轻重的作用。

目前，超声医学成像还处在结构成像阶段。而油气勘探目前已经也必须进展到岩性成像阶段。估计储层岩石介质和其中的流体性质在油气勘探中是必须的，否则无法进行油气预测。地震波成像的非线性性远超超声医学成像。可以说，对超声医学成像而言，人体组织弹性参数的成像刚刚开始。

4 数据特点的对比分析

医学成像中，最常用的是 X 射线层析成像。X-射线医学成像观测的是直接穿透的、直线传播的透射衰减量。X-射线的频率非常高，波长非常短，完全没有波场的特征表现出来。因此，成像方法相对简单。X-射线观测时没有噪音，只有精度。只要观测系统对成像点的穿透覆盖足够高，角度均匀，角度足够大，成像精度是可以很高的。

由于安全性、廉价性、手持式可移动性，超声医学成像逐渐变得重要起来。超声医学成像观测的是自激自收的反射（超声）波。目前，超声医学成像中的观测方式只能获得自激自收的反射波，且仅仅观测到纵波，没有其他波现象，基本不会有外源干扰波（除非换能器自己产生干扰波），因为自然界很少会产生超声随机干扰。（至少目前）没有子波问题。（至少目前）还很难看到声阻抗（反射系数）的反演问题，仅仅是介质反射结构的成像问题。心肌搏动和血液流动速度的成像问题基于 Doppler 效应，不是波场传播效应。而且基于 Doppler 效应的医学成像非常简单，也许谈不上是成像，做了一个频率向速度的映射而已！

前已述及，正电子发射层析成像与核磁共振成像（MRI）观测也是沿直射线传播的物理量的投影值，它们的成像方法与 X-射线成像并无本质差异。

而油气勘探中地震波成像的数据观测，非常地不同寻常：本质上是矢量波场；其中包含的波现象很复杂；子波特征很复杂、不可控、特征空变；地震频带内，自然外源很多，噪音很大；非勘探目标区的近地表区域产生了非常复杂的干扰波场；地震频带内，自然随机噪音很强；仅仅在地表激发接收的多次覆盖观测。

正是油气勘探中地震数据的复杂性和丰富性使得地震波成像在 Remote Sensing 相关领域中，数据分析和成像最复杂、最有挑战、也最引人入胜。

只可惜行业式微了！

5 成像处理方法的对比分析

从前述的医学数据观测方式、数据特点分析可以看出，或可以想象得出，医学成像是比较简单的。成像后医学图像的分析，相对于油藏描述而言，也比较简单。毕竟医生关心的主要还是人体组织的几何图像问题。也许，医生已经知道了正常组织的弹性参数，异常组织的弹性参数不是非常重要。也许医生还不适应用人体介质的物性参数变化进行医疗诊断。但是，油气勘探中，必须要了解油藏的岩性参数和流体性质后才可能做出高效益的开发。

下面重点对比一下 X-射线层析成像和地震波层析成像方法。自激自收的超声医学成像和地震波偏移成像的对比仅做简述，因为自激自收反射波场的成像实在过于简单了！

X-射线频率极高，直射线传播产生投影值，对于医学介质成像而言波场特点可以忽略。因此，X 射线成像也就直射线透射层析成像。这是很成熟的方法技术！核心是 Fourier 切片定理：两维目标函数的投影的一维 Fourier 变换，等于过目标函数两维 Fourier 变换零点的、平行于检波器方向的一个切片。所有方向的切片都存在，目标函数两维 Fourier 变换就已知了。对目标函数两维 Fourier 变换进行反变换，就得到目标函数值。如下图所示。

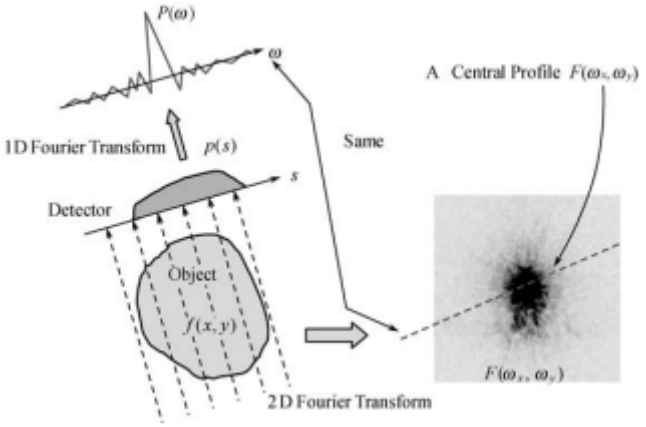


图 5a Fourier 切片定理。两维目标函数的投影的一维 Fourier 变换，等于过目标函数两维 Fourier 变换零点的、平行于检波器方向的一个切片。

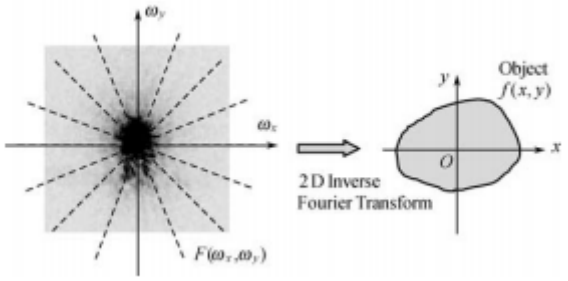


图 5b Fourier 切片定理。所有方向的切片都存在，目标函数两维 Fourier 变换就已知了。对目标函数两维 Fourier 变换进行反变换，就得到目标函数值。

其他的各种 X-射线观测系统的变种，如图 3 所示，也都是直射线的，无非是检波器不在一条直线上而已。但若检波器不在一条直线上，此时上述 Fourier 切片定理不再能直接应用。因此，其他变种的 X-射线观测系统下的 X-射线成像，无非是引入某种变换使得还能利用 Fourier 切片定理进行成像。当然可以用迭代法求解矩阵表示的线性方程组，但是计算效率和精度都会下降。

附件中讨论了 X-射线中的反投影与偏移成像中的反投影的理论对比分析。

但是，相对于油气勘探中地震波弯曲射线层析成像和绕射波波动理论层析成像，即便最新观测系统下的 3D X-Ray 射线成像，也没有很多要研究的东西!人体勘探中 X-射线层析成像要求实时产生成像结果是一个

典型的特点，这对数据观测方式和成像方法的选择有非常明显的限制。波动理论的层析成像在当前的计算机条件下很难用在人体勘探中。

X-Ray 成像，如果没有新的波现象的引入，就目前的观测方式和数据情况，层析成像方法已经很成熟了。但是，值得注意的是，X-Ray 成像后的图像处理和辅助医生进行智能识别（辅助诊断）还有较大的发展余地。尤其是所谓的 AI 技术自动识别图像在辅助医疗诊断中应该还有较大的发展空间。与油气勘探类似，人体勘探的核心工作也应该包括：数据采集、成像处理、图像解释，前两者已经比较成熟，后者还有较大的发展空间。

至于超声医学成像，即便是阵列换能器和对应的 3D 旋转激发接收方式，获得的依然是自激自收的反射波。从油气勘探地震成像的角度，这样的数据是不需要成像步骤的，仅仅需要一些数据预处理。勘探地震成像，由于勘探地震数据是带限子波激发的、球面传播的、反射波为主、绕射和散射波也比较发育的弹性地震波场，由于使用了“两宽一高”的观测，由于油藏描述的是对油藏进行岩性和流体性质的描述，成像方法非常复杂。Bayes 估计理论+波动方程地震波正演构成了地震成像的两大核心理论。由于波场的复杂性、信噪比低，Bayes 估计理论+信号预测模型构成了地震信号处理的核心。由于地下介质变化的复杂性，地震成像的不确定性，地震图像处理和解释也是非常重要的研究议题。总体而言，油气勘探地震数据分析是最有挑战的、最困难的。

相比而言，人体勘探中，超声医学观测系统接收的自激自收反射波成像的复杂度甚至赶不上上世纪 60 年代油气勘探中的手工成像技术。

超声医学成像中，按目前的观测方法，很难产生类似勘探地震的反射波为主的炮记录，人体结构应该也仅仅支持透射波观测。透射波 FWI 应该是发展方向。但鉴于人体介质与油气勘探介质的重大差异，LS_RTM 方法估计不太可能在超声医学中得到广泛应用。

6 结论与讨论

对比分析人体勘探和油气勘探的数据采集、数据特点和成像方法后，我们认为：并非人体勘探的数据分析方法多有吸引力，主要是因为油气行业的发展前景黯淡，使得我们不得不关注其它应用领域的问题。

与油气勘探关系密切的研究和应用领域，可以用广义遥感来称谓它们，其中包括甚低频天然地震、水声探测、超声探测、电磁波雷达探测（遥感）、X-射线及 γ 射线等。从波动的观点看，它们都是利用了某个波段的波源激励要探测的物理系统，观测物理系统的响应，通过对实测的系统响应数据的分析和成像，最后解决不同领域的应用问题。

很显然，广义遥感各相关领域中的数据分析方法有着共同的理论基础和共同的思想方法，与数据来自哪一个具体领域并非紧密关联。数据分析的根本理论基础来自于概率统计+数学分析（泛函分析）；数值数学主要是提供数据分析的工具；数据分析的基本思想与方法大多来自于现代信号分析以及现代图像分析，它们目前已经发展到机器学习阶段。数据分析的共同思想应该是：建立对数据中所含的系统状态变量的预测方程（各个广义遥感领域的分支应该都已经有了描述波在不同介质中传播的波动方程），根据观测的系统输出数据，在 Bayes 估计理论下，对介质系统参数进行成像（或反演）。信号和图像分析中的对数据中信号的建模奠基在 Hilbert 空间中的函数逼近理论基础上，当前的 ML（AI）方法是现代信号分析和现代图像分析发展的自然结果。

因此，作为大学的研究团队，研究的焦点还是要集中在数据分析的基本内容上。牢固建立概率统计、数学分析（泛函分析）、现代数值分析（最优化）、机器学习、现代信号分析、现代图像分析、人工智能的方面的理论基础。抽象出广义遥感的各相关领域（医学成像（尤其超声成像）、水声成像、电磁波雷达及遥感，包括地震勘探）中本质性的问题，譬如非高斯、非线性、非平稳情形下的信号预测方法；（较）强非线性情形下的系统参数估计方法等等。至于具体要关注哪一个广义遥感领域，这些只是我们的应用场景而已。

把应用场景仅仅局限在油气勘探地震上，实质上自我束缚手脚。当然，我们都知道每个领域的数据分析都有与其正问题紧密关联的特点和难度。但这也不应成为自我束缚手脚的理由。总之，数据分析的理论与技术方法研究的核心是要针对基础性的问题；应用场景的选择要考虑的是生存和发展问题。

参考文献

[1]. Gengsheng Lawrence Zeng, 2010, Medical Image Reconstruction – A Conceptual Tutorial, Higher Education Press
[2]. Paul Suetens, 2009, Fundamentals of Medical Imaging (Second Edition), Cambridge University Press
[3]. K. Kirk Shung, 2015, Diagnostic Ultrasound – Imaging and Blood Flow Measurements (Second Edition), CRC Press
[4]. Chris C. Shaw (Edited), 2014, Cone Beam Computed Tomography, CRC Press
[5]. Mostafa Analoui, Joseph D. Bronzino and Donald R. Peterson (Edited), 2013, Medical Imaging – Principles and Practices, CRC Press
[6]. R. Nick Bryan (Edited), 2010, Introduction to the science of medical imaging, Cambridge University Press
[7]. Robert E. Sheriff and Lloyd P. Geldart, 1995, Exploration Seismology, Cambridge University Press
[8]. Robert H. Stolt and Arthur B. Weglein, 2012, Seismic Iamging and Inversion – Application of Linear Inverse Theory, Cambridge University Press
[9]. Claerbout, J. F., Imaging the earth's interior, Blackwell Scientific Publications

附录

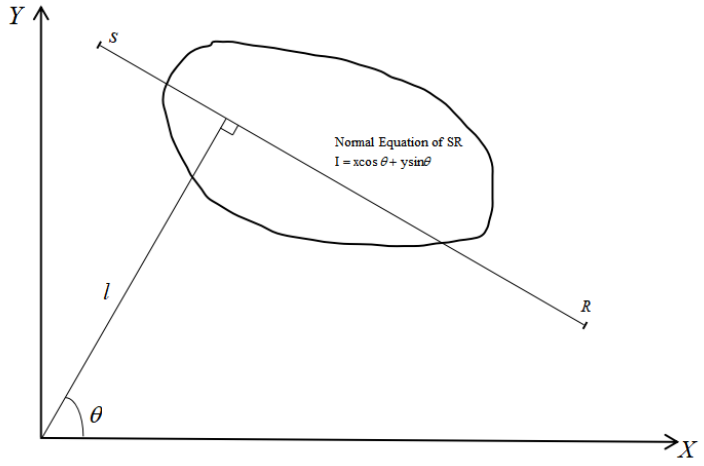
滤波反投影层析成像方法与Stolt偏移成像方法的对比

1 线性 Radon 变换

线性Radon变换有如下形式，

$$[Rf](l,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta(x\cos\theta+y\sin\theta-l)dxdy \tag{1}$$

其中$l=x\cos\theta+y\sin\theta$。其意义由附图1表示。据此，正投影的物理含义是：物理量沿图中直线穿过目标体从源点到检测点的投影结果。实际上，物理量是沿弯曲射线穿过目标体从源点到检测点产生投影结果。



附图1 Radon变换对应的直射线投影示意图

目标体内部的图像可以通过这些投影产生。这个过程称为反投影过程。Radon反变换对应于反投影过程。一般地，有四种产生目标图像的方法：Foruier投影方法、滤波反投影方法、ART/SIRT方法和矩阵求逆方法。前两种投影方法是基于常背景（连续）介质和全孔径观测导出的，它们常用于医学图像中；后两种方法常用于勘探地震学中，它们基于变背景（离散）介质假设和有限孔径观测。

2 Fourier （切片）投影定理

我们有如下的Radon变换，

$$[Rf](x,y)=P(l,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta(x\cos\theta+y\sin\theta-l)dxdy \tag{2}$$

和关于目标函数的二维Fourier变换

$$F(k_x,k_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)e^{-i2\pi\left(k_xx+k_yy\right)}dxdy. \tag{3}$$

令$k_x=k\cos\theta$和$k_y=k\sin\theta$，在地震学中$k=\dfrac{\omega}{v}$是垂直波数，那么方程（3）能被变为

$$F(k,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)e^{-i2\pi k(x\cos\theta+y\sin\theta)}dxdy \tag{4}$$

同时，Radon变换或投影的一维Fourier变换 $\tilde{P}(k,\theta)$ 为

$$\tilde{P}(k,\theta)=\int_{-\infty}^{\infty}P(l,\theta)e^{-i2\pi kl}dl \tag{5}$$

把方程（2）代入方程（4）产生

$$\begin{aligned}
&\int_{-\infty}^{\infty}P(l,\theta)e^{-i2\pi kl}dl\\
&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta(x\cos\theta+y\sin\theta-l)dxdy\right]e^{-i2\pi kl}dl\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\left[\int_{-\infty}^{\infty}e^{-i2\pi kl}\delta(x\cos\theta+y\sin\theta-l)dl\right]dxdy\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)e^{-i2\pi k(x\cos\theta+y\sin\theta)}dxdy\\
&=F(k,\theta)
\end{aligned} \tag{6}$$

因此，Fourier（切片）投影定理可以表述为：沿某方向的直线穿过目标函数的投影的一维Fourier变换等于二维目标函数的二维Fourier变换的一个切片。该定理的含义为：若已知二维目标函数的二维Fourier变换的所有切片，就相当于已知二维目标函数的二维Fourier变换，对它实施二维Fourier逆变换就可以给出目标的图象。

Fourier（切片）投影定理告诉我们：宽孔径的观测对于高精度成像是必须的，只有所有方向都有观测到的投影值，有对应的一维Fourier变换结果，目标函数的二维Fourier变换的波数谱才是完整的，其反变换结果才是高精度的目标函数反演结果。

3 直接反投影与滤波反投影

3.1 直接反投影

直接反投影是一个把来自不同直射线的、穿过二维目标函数同一物性参数剖分单元的投影值加在一起的运算，即

$$h(x,y)=\int_{0}^{\pi}P(l,\theta)d\theta \tag{7}$$

其中 $l=x\cos\theta+y\sin\theta$ 。因此，直接反投影的物理意义是：所有穿过当前介质单元的射线的投影值的（加权）和。我们注意到：反投影不是Radon变换的求逆，然而它是Radon变换求逆的一种粗略的近似，导致成像的模糊，因此需要引入滤波反投影方法。反投影不仅给出了一种简单的算法，而且它还指出了从观测数据得到目标图象的一般思路。

实际上，（1）式中的Radon变换可以用褶积形式来表达，

$$P(\boldsymbol{y})=\int f(\boldsymbol{x})K(\boldsymbol{x}-\boldsymbol{x}')d\boldsymbol{x}'$$

该式的矩阵形式为：$\boldsymbol{P}_{n\times 1}=K_{n\times m}\boldsymbol{f}_{m\times 1}$，其中 n 代表射线条数；m 代表介质单元的个数。矩阵形式的直

接反投影方法：$\boldsymbol{f}_{m\times1}\approx K^{T}_{m\times n}\boldsymbol{P}_{n\times1}$。物理含义显然是：穿过当前介质单元所有射线的加权和构成介质单元的参数的近似估计结果。

3.2 滤波反投影

从 Fourier（切片）投影定理可以知道，如果 Radon 变换的正投影可以保证得到完整的目标函数的二维 Fourier 变换结果，那么滤波反投影就是理论上精确的反问题解法。而且有足够快的计算效率，这对于要求实时成像的医学成像而言是非常关键的。

对方程（3）做从$\left(k_x,k_y\right)$到$\left(k,\theta\right)$的坐标变换得到

$$f(x,y)=\int_0^{2\pi}\int_0^{\infty}f(k,\theta)e^{-i2\pi k\left(x\cos\theta+y\sin\theta\right)}kdkd\theta. \tag{8}$$

在极坐标系中，有$\theta\in\left[0,2\pi\right]$和$k\in\left[0,\infty\right)$.通过在上式第二项中定义$\theta\to\theta+\pi$,我们可以重写方程（8）为

$$\begin{aligned}f(x,y)=&\int_0^{\pi}\int_0^{\infty}f(k,\theta)e^{-i2\pi k\left(x\cos\theta+y\sin\theta\right)}kdkd\theta\\&+\int_0^{\pi}\int_0^{\infty}f(k,\theta+\pi)e^{-i2\pi k\left(-x\cos\theta-y\sin\theta\right)}kdkd\theta\end{aligned} \tag{9}$$

通过在上式第二项中定义$k\to-k$，方程（8）可进一步写为

$$\begin{aligned}f(x,y)=&\int_0^{\pi}\int_0^{\infty}f(k,\theta)e^{-i2\pi k\left(x\cos\theta+y\sin\theta\right)}kdkd\theta\\&+\int_0^{\pi}\int_{-\infty}^{0}f(-k,\theta+\pi)e^{-i2\pi k\left(x\cos\theta+y\sin\theta\right)}kdkd\theta\end{aligned} \tag{10}$$

在极坐标系中，利用等式$f(-k,\theta+\pi)=f(k,\theta)$，可以得到

$$f(x,y)=\int_0^{\pi}\int_{-\infty}^{\infty}F(k,\theta)\left|k\right|e^{i2\pi\omega l}dkd\theta. \tag{11}$$

借助于Fourier投影定理，方程（11）可被重写为

$$f(x,y)=\int_0^{\pi}\left\{\int_{-\infty}^{\infty}\tilde{P}(k,\theta)\left|k\right|e^{i2\pi kl}dk\right\}d\theta. \tag{12}$$

因此，用这种方法，通过对一维Fourier变换过且经滤波过的投影值进行反投影，可以得到目标函数的图像。此处，频率加权因子$\left|k\right|$是一个反褶积因子，它用来消除反投影的模糊化。也可以说它是一个高通滤波器，这就是滤波反投影的由来。

对于弯曲射线Radon变换、不完整观测、含噪音的观测等等复杂情形，滤波反投影方法不合适或根本不能用。因此，计算机能力非常强大的情况下，各种正则化约束下的数值近似解法是更合理的选择。

4 Stolt 偏移方法

4.1 理论推导

根据爆炸反射面理论，把水平叠加剖面看作在反射界面上同时爆炸产生地震波，并以半速度向外传播，在地面上观测到的上行波剖面。对于这种观测结果，为了能够成像，要求向地下反向外推地震波场。

令$U(x,y,z=0,t)$表示未偏移的叠加剖面，$U(x,y,z,t=0)$表示偏移后剖面。在均匀各向同性完全弹性介质中，用半波速代替地震波速度，则标量波动方程变为：

$$\nabla^2 U-\frac{4}{v^2}\frac{\partial^2 U}{\partial t^2}=0 \tag{13}$$

对（13）式做关于x,y,z,t的傅立叶变换得到上式的象征方程，

$$k_x^2+k_y^2+k_z^2-\left(\frac{2\omega}{v}\right)^2=0 \tag{14}$$

其中，ω表示角频率；k_x,k_y,k_z分别表示水平方向和垂直方向的波数。重写上式为

$$\omega=\pm\frac{v}{2}\sqrt{k_x^2+k_y^2+k_z^2} \tag{15}$$

其中正号对应下行波、负号对应上行波。对方程(3.5)做如下的三维傅立叶变换

$$\frac{\partial^2\tilde{U}}{\partial t^2}+\frac{v^2}{4}\left(k_x^2+k_y^2+k_z^2\right)\tilde{U}=0 \tag{16}$$

将（15）式代入（16）中，得

$$\frac{\partial^2\tilde{U}}{\partial t^2}+\omega^2\tilde{U}=0\Rightarrow\left(\frac{\partial}{\partial t}+i\omega\right)\left(\frac{\partial}{\partial t}-i\omega\right)\tilde{U}=0 \tag{17}$$

（13）式的平面波解如下

$$\tilde{U}\left(k_x,k_y,k_z,t\right)=A\left(k_x,k_y,k_z\right)e^{-i\omega t} \tag{18}$$

对（18）式做三维傅立叶逆变换，得

$$U(x,y,z;t)=\left(\frac{1}{2\pi}\right)^3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk_xdk_y\int_{-\infty}^{\infty}A\left(k_x,k_y,k_z\right)e^{-i\omega t}e^{i\left(k_xx+k_yy+k_zz\right)}dk_z \tag{19}$$

该式表示空间中任意点的波场或者所有深度层上的波场。

令$z=0,$（19）式变为

$$U(x,y,0;t)=\left(\frac{1}{2\pi}\right)^3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk_xdk_y\int_{-\infty}^{\infty}A\left(k_x,k_y,k_z\right)e^{-i\left(\omega t-\left(k_xx+k_yy\right)\right)}dk_z \tag{20}$$

上式表示的就是地表观测记录。

我们还可以对实际观察的水平叠加剖面的三维傅立叶变换记为

$$B\left(k_x,k_y,0,\omega\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dxdy\int_{-\infty}^{\infty}U(x,y,0;t)e^{-i\left(\omega t-\left(k_xx+k_yy\right)\right)}dt \tag{21}$$

（21）式的逆变换为：

$$\begin{aligned}U(x,y,0;t)&=\left(\frac{1}{2\pi}\right)^3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk_xdk_y\int_{-\infty}^{\infty}B\left(k_x,k_y,0,\omega\right)e^{i\left(\omega t-\left(k_xx+k_yy\right)\right)}d\omega\\&=\left(\frac{1}{2\pi}\right)^3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk_xdk_y\int_{-\infty}^{\infty}B\left(k_x,k_y,0,\omega\right)e^{i\left(\omega t-\left(k_xx+k_yy\right)\right)}\frac{d\omega}{dk_z}dk_z\end{aligned} \tag{22}$$

当然，（20）和（21）式本质上是相同的，仅仅是推导的途径不同。对比（20）和（21）式，有

$$A\left(k_x,k_y,k_z\right)=B\left(k_x,k_y,0,\omega\right)\frac{d\omega}{dk_z} \tag{23}$$

由（15）式可以求出$\frac{d\omega}{dk_z}=\frac{v}{2}\frac{k_z}{\sqrt{k_x^2+k_y^2+k_z^2}}$，代入（23）式，得到

$$A\left(k_x,k_y,k_z\right)=B\left(k_x,k_y,0,\frac{v}{2}\sqrt{k_x^2+k_y^2+k_z^2}\right)\frac{v}{2}\frac{k_z}{\sqrt{k_x^2+k_y^2+k_z^2}} \tag{24}$$

因此，把（24）代入（19）得到频率-波数域三维叠后偏移公式为

$$\begin{aligned}U(x,y,z;0)=&\left(\frac{1}{2\pi}\right)^3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk_xdk_y\int_{-\infty}^{\infty}B\left(k_x,k_y,0,\frac{v}{2}\sqrt{k_x^2+k_y^2+k_z^2}\right)\\&\frac{v}{2}\frac{k_z}{\sqrt{k_x^2+k_y^2+k_z^2}}e^{i\omega t}e^{-i\left(k_xx+k_yy+k_zz\right)}dk_z\\=&\left(\frac{1}{2\pi}\right)^3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk_xdk_y\int_{-\infty}^{\infty}B\left(k_x,k_y,0,\frac{v}{2}\sqrt{k_x^2+k_y^2+k_z^2}\right)\\&\frac{v}{2}\frac{k_z}{\sqrt{k_x^2+k_y^2+k_z^2}}e^{-i\left(k_xx+k_yy+k_zz\right)}dk_z\end{aligned} \tag{25}$$

4.2 具体实现

由公式（25）知，F-K域偏移的具体实现步骤为：

（1）、对叠加剖面做由（21）式规定的三维Fourier变换。（2）、利用式（15）进行数据空间$D\left(k_x,k_y,\omega\right)$到$D\left(k_x,k_y,k_z\right)$的插值。（3）、对插值后的数据体乘上由$\frac{v}{2}\frac{k_z}{\sqrt{k_x^2+k_y^2+k_z^2}}$定义的振幅校正因子。（4）、对步骤3产生的数据体进行三维Fourier反变换，就可以得到最终的偏移剖面。

5 滤波反投影层析成像方法与 Stolt 偏移成像方法的对比

理论上讲，任何物理系统的输出都可以用投影来表示，都可以用一个褶积关系来表式，最终都可以用一个矩阵方程来表示，

$$\boldsymbol{P}_{n\times 1}=K_{n\times m}\boldsymbol{f}_{m\times 1} \tag{26}$$

对应的直接反投影是： $\boldsymbol{f}_{m\times 1}=K^{T}_{m\times n}\boldsymbol{P}_{n\times 1}$ （27）

这是泛函分析理论给出的统一的框架。数学本质就是如此！为什么不同领域的专家还专注此问题，提出很多不同的反投影算法。滤波反投影为核心的X-射线层析成像还获得了诺贝尔奖！这在泛函分析学家眼中是小儿科的东西！

事实上，一个问题，从不同的方面或角度来阐释是科学研究乃至科学进步的基本模式！对同一问题进行不同角度的对比分析是对问题加深理解的必有之路！

任何反投影方法都是基于正投影的。（26）和（27）尽管提供了高度抽象的形式，但对于解决具体的物理问题，还要明确投影的物理意义。

Radon变换定义的投影，

$$\left[Rf\right](l,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta(x\cos\theta+y\sin\theta-l)dxdy \tag{28}$$

其物理意义是十分明确的！

波动方程描述波在物理系统中的传播，算子形式的表达形式为，

$$Au\left(\boldsymbol{x},t\right)=f\left(\boldsymbol{x}_S,t\right) \tag{29}$$

其中，$A=\nabla^2-\dfrac{1}{v^2\left(\boldsymbol{x}\right)}$；$f\left(\boldsymbol{x}_S,t\right)$ 是震源函数；$\boldsymbol{x}_S$ 为源点坐标；$u\left(\boldsymbol{x},t\right)$ 为系统状态变量（即激励输出的波场）。但是，（29）式并没有建立起目标介质参数与实测波场之间的投影关系。对地震波成像而言，用反投影的语言统一解释地震波成像，必须给出正投影的定义。像FWI那样的笼统的定义也不能很好地说明问题的本质。必须给出具体的参数变化引起的波场（特征）变化这样的投影关系的定义才可以明确的反投影的定义，才可以做好反投影。

$u\left(\boldsymbol{x},t\right)$ 中包含的透射波层析成像的正投影可以用Radon变换来表述。但是，对速度层析而言，仅仅是背景速度产生了旅行时的累积投影。对于 $u\left(\boldsymbol{x},t\right)$ 中包含的反射波层析成像的正投影，概念上讲与透射波的正投影是一致的。但是不同的是：反射波测产生必须要有波阻抗界面的存在，即背景速度的层析成像与波阻抗界面是耦合的！当然，如果忽视这一点，反射波层析成像与透射波层析成像并没有理论上的差异。

如果要对背景速度中的参数跃变进行反投影意义下的成像，此时背景速度中的参数跃变引起的正投影的物理解释是非常不同的！

Born近似理论框架下，背景速度中的参数跃变引起的正投影的数学表达式为，

$$\Delta u\left(\Delta\boldsymbol{\xi},z=0,t\right)=\int_{S}^{R}G_S\left(\boldsymbol{x}_S,\boldsymbol{x}\right)\Delta m\left(\boldsymbol{x}\right)G_R\left(\boldsymbol{x},\boldsymbol{x}_R\right)d\boldsymbol{x} \tag{30}$$

其中，$G_S\left(\boldsymbol{x}_S,\boldsymbol{x}\right)$ 代表从炮点 $\boldsymbol{x}_S$ 到成像点 $\boldsymbol{x}$ 的Green函数；$G_R\left(\boldsymbol{x},\boldsymbol{x}_R\right)$ 代表从成像点 $\boldsymbol{x}$ 到检波点 $\boldsymbol{x}_R$ 的Green函数。对于固定的炮检点而言，沿着从炮点到检波点的传播路径“积累”的投影值是检波器测量的波场。奇怪的是，这个投影值并不像Radon变换投影值那样是一个累积结果，不是沿路径的积分结果！但是，

广义地看，检波器测量结果就应该是投影值！

（30）式对应的矩阵方程可以写为，

$$\Delta\tilde{\boldsymbol{u}}_{SR}\left(\boldsymbol{x}_S,\boldsymbol{x}_R;\omega\right)=\tilde{G}_{SR}\left(\boldsymbol{x}_S,\boldsymbol{x},\boldsymbol{x}_R;\omega\right)\Delta\boldsymbol{m}\left(\boldsymbol{x}\right) \tag{31}$$

（31）式的具体形式为，

$$\begin{bmatrix}\Delta\tilde{u}_{S_1R_1}\\ \Delta\tilde{u}_{S_1R_2}\\ \vdots\\ \Delta\tilde{u}_{S_1R_N}\\ \vdots\\ \Delta\tilde{u}_{S_NR_1}\\ \Delta\tilde{u}_{S_NR_2}\\ \vdots\\ \Delta\tilde{u}_{S_NR_N}\end{bmatrix}=\begin{bmatrix}
\overset{NZ=1}{\overbrace{\begin{matrix}G^{X_1Z_1}_{S_1R_1} & G^{X_2Z_1}_{S_1R_1}\\ G^{X_1Z_1}_{S_1R_2} & G^{X_2Z_1}_{S_1R_2}\\ \vdots & \vdots\\ G^{X_1Z_1}_{S_1R_N} & G^{X_2Z_2}_{S_1R_N}\\ \vdots & \vdots\\ G^{X_1Z_1}_{S_NR_1} & G^{X_2Z_2}_{S_NR_1}\\ G^{X_1Z_1}_{S_NR_2} & G^{X_2Z_2}_{S_NR_2}\\ \vdots & \vdots\\ G^{X_1Z_1}_{S_NR_N} & G^{X_2Z_2}_{S_NR_N}\end{matrix}}} & \cdots &
\overset{NZ=2}{\overbrace{\begin{matrix}G^{X_{NX}Z_1}_{S_1R_1} & G^{X_1Z_2}_{S_1R_1} & G^{X_2Z_2}_{S_1R_1}\\ G^{X_{NX}Z_1}_{S_1R_2} & G^{X_1Z_2}_{S_1R_2} & G^{X_2Z_2}_{S_1R_2}\\ \vdots & \vdots & \vdots\\ G^{X_{NX}Z_1}_{S_1R_N} & G^{X_1Z_2}_{S_1R_N} & G^{X_2Z_2}_{S_1R_N}\\ \vdots & \vdots & \vdots\\ G^{X_{NX}Z_1}_{S_NR_1} & G^{X_1Z_2}_{S_NR_1} & G^{X_2Z_2}_{S_NR_1}\\ G^{X_{NX}Z}_{S_NR_2} & G^{X_1Z_2}_{S_NR_2} & G^{X_2Z_2}_{S_NR_2}\\ \vdots & \vdots & \vdots\\ G^{X_{NX}Z_1}_{S_NR_N} & G^{X_1Z_2}_{S_NR_N} & G^{X_2Z_2}_{S_NR_N}\end{matrix}}} & \cdots & \cdots &
\overset{NZ=NZ}{\overbrace{\begin{matrix}G^{X_1Z_{NZ}}_{S_1R_1} & G^{X_2Z_{NZ}}_{S_1R_1} & \cdots & G^{X_{NX}Z_{NZ}}_{S_1R_1}\\ G^{X_1Z_{NZ}}_{S_1R_2} & G^{X_2Z_{NZ}}_{S_1R_2} & \cdots & G^{X_{NX}Z_{NZ}}_{S_1R_2}\\ \vdots & \vdots & \cdots & \vdots\\ G^{X_1Z_{NZ}}_{S_1R_N} & G^{X_2Z_{NZ}}_{S_1R_N} & \cdots & G^{X_{NX}Z_{NZ}}_{S_1R_N}\\ \vdots & \vdots & \cdots & \vdots\\ G^{X_1Z_{NZ}}_{S_NR_1} & G^{X_2Z_{NZ}}_{S_NR_1} & \cdots & G^{X_{NX}Z_{NZ}}_{S_NR_1}\\ G^{X_1Z_{NZ}}_{S_NR_2} & G^{X_2Z_{NZ}}_{S_NR_2} & \cdots & G^{X_{NX}Z_{NZ}}_{S_NR_2}\\ \vdots & \vdots & \cdots & \vdots\\ G^{X_1Z_{NZ}}_{S_NR_N} & G^{X_2Z_{NZ}}_{S_NR_N} & \cdots & G^{X_{NX}Z_{NZ}}_{S_NR_N}\end{matrix}}}
\end{bmatrix}\begin{bmatrix}\delta m_{1,1}\\ \delta m_{2,1}\\ \vdots\\ \delta m_{NX,1}\\ \vdots\\ \delta m_{1,NZ}\\ \delta m_{2,NZ}\\ \vdots\\ \delta m_{NX,NZ}\end{bmatrix}$$

（31）式的直接投影解，

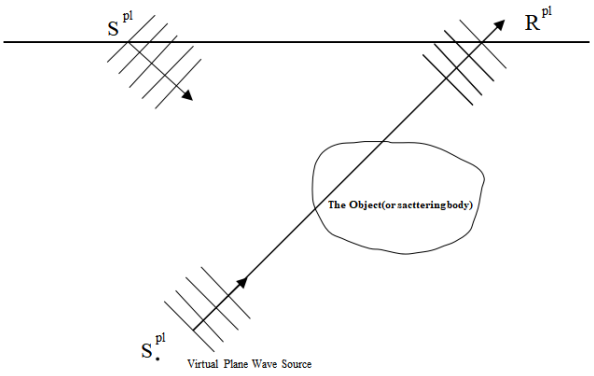
$$\Delta\boldsymbol{m}\left(\boldsymbol{x}\right)\approx\tilde{G}^{T}_{SR}\left(\boldsymbol{x}_S,\boldsymbol{x},\boldsymbol{x}_R;\omega\right)\Delta\tilde{\boldsymbol{u}}_{SR}\left(\boldsymbol{x}_S,\boldsymbol{x}_R;\omega\right) \tag{32}$$

（32）的含义是：地下任一点的速度扰动（或波阻抗扰动）量的估计等于所有炮检对观测的投影值的加权叠加。而每一个投影值的反投影不是沿着波传播路径的，而是沿着等时面的。常背景介质下零偏移观测对应的等时面是半圆；非零偏移距观测对应的等时面是半椭圆。

我认为，应该给这种正投影和反投影一个明确的称谓，以便与Radon变换正投影和对应的反投影明确区分开。

6 波动理论的 Fourier（切片）投影定理与波动理论透射波层析成像

考虑一个平面波入射到一个异常体或散射体上，同时由一个远离该异常体或散射体上的接收器接收散射波场。如附图2所示。在背景介质为常速时，可以假设虚拟平面波沿直线穿过目标体，产生观测平面波（投影）。目标体的像（或其波数谱）和散射场的关系可以被推导出来，它类似于Fourier投影定理。



附图2 平面波入射到散射波上产生散射波的示意图

此处仅考虑标量波方程，速度由$v(\boldsymbol{x})$来描述。对于无源区域中的压力场$\tilde{u}(\boldsymbol{x};\omega)$频域形式的标量波动方程（Helmholtz方程）为

$$\nabla^2\tilde{\boldsymbol{u}}(\boldsymbol{x};\omega)+\left(\frac{\omega}{v(\boldsymbol{x})}\right)^2\tilde{\boldsymbol{u}}(\boldsymbol{x};\omega)=0 \tag{33}$$

我们把目标函数 $\boldsymbol{m}(\boldsymbol{x})$ 定义成如下形式的速度摄动

$$\boldsymbol{m}(\boldsymbol{x})=1-\left(\frac{v_0}{v(\boldsymbol{x})}\right)^2 \tag{34}$$

把方程（34）代入（33）得到

$$\nabla^2\tilde{\boldsymbol{u}}\left(\boldsymbol{x};\omega\right)+\boldsymbol{k}^2\tilde{\boldsymbol{u}}\left(\boldsymbol{x};\omega\right)=k^2\boldsymbol{m}(\boldsymbol{x})\tilde{\boldsymbol{u}}\left(\boldsymbol{x};\omega\right) \tag{35}$$

其中 $\boldsymbol{k}=\dfrac{\omega}{v(\boldsymbol{x})}$。

把总波场分为入射波场 $\tilde{\boldsymbol{u}}_0(\boldsymbol{x};\omega)$ 和散射波场 $\tilde{\boldsymbol{u}}_S(\boldsymbol{x};\omega)$ 两部分，

$$\tilde{\boldsymbol{u}}(\boldsymbol{x};\omega)=\tilde{\boldsymbol{u}}_0(\boldsymbol{x};\omega)+\tilde{\boldsymbol{u}}_S(\boldsymbol{x};\omega) \tag{36}$$

并把方程（36）代入（35）得到

$$\nabla^2\tilde{\boldsymbol{u}}_S\left(\boldsymbol{x};\omega\right)+\boldsymbol{k}^2\tilde{\boldsymbol{u}}_S\left(\boldsymbol{x};\omega\right)=k^2\boldsymbol{m}(\boldsymbol{x})\tilde{\boldsymbol{u}}\left(\boldsymbol{x};\omega\right) \tag{37}$$

注意 $\nabla^2\tilde{\boldsymbol{u}}_0\left(\boldsymbol{x};\omega\right)+\boldsymbol{k}^2\tilde{\boldsymbol{u}}_0\left(\boldsymbol{x};\omega\right)=0$。方程（37）把散射场和目标函数及总场联系在一起。利用Green定理，可以得到（37）的解，

$$\tilde{\boldsymbol{u}}_S\left(\boldsymbol{x};\omega\right)=-\boldsymbol{k}^2\int_\Omega G\left(\boldsymbol{x}_S,\boldsymbol{x};\omega\right)\varepsilon\left(x\right)\tilde{\boldsymbol{u}}\left(\boldsymbol{x},\boldsymbol{x}_R;\omega\right)d\boldsymbol{x} \tag{38}$$

其中，积分是对目标体分布的范围 Ω 进行的。方程（38）把目标体外表记录数据和未知的散射势 $\boldsymbol{m}(\boldsymbol{x})$ 联系在一起。散射场是由于实际介质和背景介质间的波阻抗差异引起的。总场 $\tilde{\boldsymbol{u}}(\boldsymbol{x};\omega)$ 作用到散射势 $\boldsymbol{m}(\boldsymbol{x})$ 上产生相互作用 $\boldsymbol{m}(\boldsymbol{x})\tilde{\boldsymbol{u}}(\boldsymbol{x};\omega)$，形成二次源，散射场由此而产生。理论上，背景速度场可以任意选择。但背景场的选择直接影响散射势的大小。事实上，散射势不能变化过大。

对方程（38）引入Born近似得到，

$$\tilde{\boldsymbol{u}}_S\left(\boldsymbol{x};\omega\right)=-\boldsymbol{k}^2\int_\Omega G\left(\boldsymbol{x},\boldsymbol{x}_R;\omega\right)\varepsilon\left(x\right)\tilde{\boldsymbol{u}}_0\left(\boldsymbol{x}_S,\boldsymbol{x};\omega\right)d\boldsymbol{x} \tag{39}$$

方程（38）中的背景速度可以任意选，但近似式（39）只有在速度摄动 $\varepsilon(\vec{r})$ 较小时是比较准确的。Green函数 $G\left(\boldsymbol{x}_S,\boldsymbol{x};\omega\right)$ 只有在背景速度分布很简单时（譬如，常速或层状介质时）可以解析地写出。一般地，我们用数值近似或渐近近似（例如，WKBJ近似）来描述Green函数。

常背景速度情况下，入射平面波可以表示成，

$$\tilde{\boldsymbol{u}}_0\left(\boldsymbol{x};\omega\right)=e^{i\boldsymbol{k}\cdot(\boldsymbol{x}_S-\boldsymbol{x})} \tag{40}$$

其中 $\boldsymbol{k}$ 是入射波传播方向的波矢量。

同样地，常背景速度情况下，三维自由空间Green函数为，

$$G\left(\boldsymbol{x},\boldsymbol{x}_R;\omega\right)=\frac{1}{\left|\boldsymbol{x}_R-\boldsymbol{x}\right|}e^{i\boldsymbol{k}_R\cdot(\boldsymbol{x}-\boldsymbol{x}_R)} \tag{41}$$

把方程（40）和（41）代入方程（39）得到，

$$\begin{aligned}\tilde{\boldsymbol{u}}_S\left(\boldsymbol{x};\omega\right)&=-\boldsymbol{k}^2\int_\Omega\frac{e^{i\boldsymbol{k}_R\cdot\left(\boldsymbol{x}-\boldsymbol{x}_R\right)}}{\left|\boldsymbol{x}-\boldsymbol{x}_R\right|}\boldsymbol{m}\left(\boldsymbol{x}\right)e^{i\boldsymbol{k}_S\cdot\left(\boldsymbol{x}_S-\boldsymbol{x}\right)}d\boldsymbol{x}\\&=-\boldsymbol{k}^2\int_\Omega\frac{1}{\left|\boldsymbol{x}-\boldsymbol{x}_R\right|}\boldsymbol{m}\left(\boldsymbol{x}\right)e^{i\left(\boldsymbol{k}_S\cdot\left(\boldsymbol{x}_S-\boldsymbol{x}\right)+\boldsymbol{k}_R\cdot\left(\boldsymbol{x}-\boldsymbol{x}_R\right)\right)}d\boldsymbol{x}\\&=-\boldsymbol{k}^2\int_\Omega\frac{1}{\left|\boldsymbol{x}-\boldsymbol{x}_R\right|}\boldsymbol{m}\left(\boldsymbol{x}\right)e^{i\omega\left(T\left(\boldsymbol{x}_S,\boldsymbol{x}\right)+T\left(\boldsymbol{x},\boldsymbol{x}_R\right)\right)}d\boldsymbol{x}\end{aligned}\tag{42}$$

其中，$T\left(\boldsymbol{x}_S,\boldsymbol{x}\right)$代表源点到像点的旅行时；$T\left(\boldsymbol{x}_R,\boldsymbol{x}\right)$代表像点到检波点的旅行时。从源点到像点再到检波点的旅行时为，

$$T\left(\boldsymbol{x}_S,\boldsymbol{x},\boldsymbol{x}_R\right)=T\left(\boldsymbol{x}_S,\boldsymbol{x}\right)+T\left(\boldsymbol{x},\boldsymbol{x}_R\right)\tag{43}$$

总的旅行时有如下的Taylor展开形式，

$$\begin{aligned}&e^{i\omega\left(T\left(\boldsymbol{x}_S,\boldsymbol{x}\right)+T\left(\boldsymbol{x},\boldsymbol{x}_R\right)\right)}\\&\approx e^{i\omega\left(T\left(\boldsymbol{x}_S,\boldsymbol{x}_0\right)+T\left(\boldsymbol{x}_0,\boldsymbol{x}_R\right)\right)}e^{i\omega\left(\boldsymbol{x}-\boldsymbol{x}_0\right)\left(\boldsymbol{p}_S\left(\boldsymbol{x}_S,\boldsymbol{x}_0\right)+\boldsymbol{p}_R\left(\boldsymbol{x}_0,\boldsymbol{x}_R\right)\right)}\\&=e^{i\omega\left(T\left(\boldsymbol{x}_S,\boldsymbol{x}_0\right)+T\left(\boldsymbol{x}_0,\boldsymbol{x}_R\right)\right)}e^{i\bar{\boldsymbol{k}}\cdot\left(\boldsymbol{x}-\boldsymbol{x}_0\right)}\end{aligned}\tag{44}$$

把（44）式代入（42）式，

$$\begin{aligned}\tilde{\boldsymbol{u}}_S\left(\boldsymbol{x};\omega\right)&=-\boldsymbol{k}^2\int_\Omega\frac{1}{\left|\boldsymbol{x}-\boldsymbol{x}_R\right|}\boldsymbol{m}\left(\boldsymbol{x}\right)e^{i\omega\left(T\left(\boldsymbol{x}_S,\boldsymbol{x}\right)+T\left(\boldsymbol{x},\boldsymbol{x}_R\right)\right)}d\boldsymbol{x}\\&\approx-\boldsymbol{k}^2\int_\Omega\frac{1}{\left|\boldsymbol{x}-\boldsymbol{x}_R\right|}\boldsymbol{m}\left(\boldsymbol{x}\right)e^{i\omega\left(T\left(\boldsymbol{x}_S,\boldsymbol{x}_0\right)+T\left(\boldsymbol{x}_0,\boldsymbol{x}_R\right)\right)}e^{i\bar{\boldsymbol{k}}\left(\boldsymbol{x}-\boldsymbol{x}_0\right)}d\boldsymbol{x}\\&\approx-\boldsymbol{k}^2\frac{e^{i\omega\left(T\left(\boldsymbol{x}_S,\boldsymbol{x}_0\right)+T\left(\boldsymbol{x}_0,\boldsymbol{x}_R\right)\right)}}{\left|\boldsymbol{x}_0-\boldsymbol{x}_R\right|}\int_\Omega\boldsymbol{m}\left(\boldsymbol{x}\right)e^{i\bar{\boldsymbol{k}}\cdot\left(\boldsymbol{x}-\boldsymbol{x}_0\right)}d\boldsymbol{x}\end{aligned}\tag{45}$$

（44）式中波矢量$\bar{\boldsymbol{k}}=\omega\left(\boldsymbol{p}_S\left(\boldsymbol{x}_S,\boldsymbol{x}_0\right)+\boldsymbol{p}_R\left(\boldsymbol{x}_0,\boldsymbol{x}_R\right)\right)=\frac{2\omega\cos\theta}{v\left(\boldsymbol{x}_0\right)}\boldsymbol{n}$。针对（45）式，记

$$\tilde{\boldsymbol{m}}\left(\bar{\boldsymbol{k}}\right)=\int_\Omega\boldsymbol{m}\left(\boldsymbol{x}\right)e^{i\bar{\boldsymbol{k}}\cdot\left(\boldsymbol{x}-\boldsymbol{x}_0\right)}d\boldsymbol{x}\tag{46}$$

很显然，（46）式是一个关于目标函数的三维 Fourier 变换，可以认为是三维波动理论下的 Fourier 切片定理。只有波矢量$\bar{\boldsymbol{k}}=\frac{2\omega\cos\theta}{v\left(\boldsymbol{x}_0\right)}\boldsymbol{n}$的所有谱成分都存在时，或者说只有所有角度的炮检组合及所有频率成分都存在时，才能由上式的 Fourier 反变换得到散射体扰动量的精确解。很显然，这取决于入射波与散射波的张角展布和频率成分的展布。波动理论层析中的 Fourier 切片是由波矢量$\bar{\boldsymbol{k}}=\frac{2\omega\cos\theta}{v\left(\boldsymbol{x}_0\right)}\boldsymbol{n}$决定的，也是由观测系统决定的。反射波情形下稍微复杂一些。

AI 和 ML 在勘探地震中的应用前景分析

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】大数据和人工智能构成了当今社会的基本特色。从（大）数据中挖掘出用于决策的有效信息，执行最佳决策（或判断），是当前人工智能的根本内涵。油气地震勘探的目标是精确刻画与评价油气藏，做出是否钻井的决策。这个决策问题同样是基于从（大）数据中挖掘出的信息。油气地震勘探的基本数据是叠前地震数据体，还包括与地下介质及储层相关的各种先验信息（譬如地表露头、钻井、岩性、地质等），通过地震波反演成像得到地下介质弹性参数，与岩石物理结合转化为岩性参数（譬如孔隙度、渗透率和饱和度）。还要结合很多其它信息才能做出是否钻井的决策。这是一般领域的 AI 应用场景非常不同的问题。因此，非常有必要深入地分析自然智能、人类智能、人工智能的本质，分析勘探地震数据的本质、分析 ML 算法的本质，厘清 AI/ML 在勘探地震中所能发挥的作用。本文在分析智能本质的基础上，讨论了数据、信息与算法的内涵，阐述了地震数据分析的本质问题，给出了 AI 和 ML 在勘探地震中的应用前景的基本判断。期望在勘探地震中合理地应用 AI 和 ML 理论方法与技术能起到一定的促进作用。

【关键词】智能、人工智能、（大）数据、信息、决策、机器学习算法、地震数据分析

1 引言

达·芬奇曾经说过：大自然是大师中的大师，除非从它那里获得灵感，否则都是徒劳的。相对于其他自然科学领域，智能（Intelligence）是迄今为止自然界广泛存在、影响普遍而深远、却又没有得到深入认识的一个研究对象。自然智能（Natural Intelligence=NI）的表现多种多样，人类智能（Human Intelligence=HI）是自然智能中的最高存在形式。正因为如此，很容易把自然智能等同于人类智能。也正因为如此，研究和仿生人类智能（HI）并发展出人工智能（Artificial Intelligence=AI）成为智能科学研究的最重要目的。

尽管对智能并没有统一的定义，人类智能（HI）包括如下几个方面是基本都认可的。1、语言智能（Linguistic intelligence）：是指有效地运用口头语言或及文字表达自己的思想并理解他人，灵活掌握语音、语义、语法，具备用言语思维、用言语表达和欣赏语言深层内涵的能力，且能它它们结合在一起并运用自如的能力。2、数学逻辑智能（Logical-Mathematical intelligence）：是指有效地计算、测量、推理、归纳、分类，并进行复杂数学运算的能力。3、空间智能（Spatial intelligence）：是指准确感知视觉空间及周围一切事物，并且能把所感觉到的形象以图画的形式表现出来的能力。4、身体运动智能（Bodily-Kinesthetic intelligence）：是指善于运用整个身体来表达思想和情感、灵巧地运用双手制作或操作物体的能力。5、音乐智能（Musical intelligence）：是指人能够敏锐地感知音调、旋律、节奏、音色等能力。6、人际智能（Interpersonal intelligence）：是指能很好地理解别人和与人交往的能力。7、自我认知智能（Intrapersonal intelligence）：是指具备自我认识和拥有自知之明并据此做出适当行为的能力。8、自然认知智能（Naturalist intelligence）：是指善于观察自然界中的各种事物，对物体进行辩论和分类的能力。

人类智能（HI）过于复杂和微妙，人工智能（AI）不可能也不必要复制人类智能，而是充分认识和借鉴人类智能（严格地，应该是自然智能（NI））的基础上，发展出某些方面更强的人造智能，能更好地解决所面对的具体实际问题。

智能的本质到底是什么？我认为智能的本质是感知环境、理解环境、最佳决策、合理行动。我给出的智能（Intelligence）的定义：行为体依据“环境”或“请求”做出最佳决策（或响应）的能力。互联网给出的智能（Intelligence）定义为：个体对客观事物进行合理分析、判断及有目的地行动和有效地处理周围环境事宜的综合能力。

我们知道，智能（Intelligence）一定是群体行为才能体现出来的。智能（Intelligence）也是由群体在与环境的互动中发展出来的。大量单个个体及它们之间的相互作用（相互连接）在与环境的博弈中一定会产生智能，这是大自然的进化发展规律。可以借鉴群体相互作用产生智能的逻辑发展和优化人工智能，这也是人工智能（AI）发展的重要方向。

我认为智能（Intelligence）的根本要素应该包括：感知和理解环境；做出最佳决策；产生正确的行动。其中，感知和理解“环境“是最重要的环节。人类智能（HI）主要通过经验（表现为语言描述的规则或概念）和规则来理解”环境“。

我给出的人工（造）智能（AI）的定义：人造行为体依据“环境”或“请求”做出最佳决策（或响应）的一套算法。人工智能（系统）需要具备的要素：首先是具备感知能力，包括视觉感知、对语音和语言的理解等；然后是具备决策能力，能进行识别、预测，做出决策和判断；最后要有一套完整的系统，执行带反馈的行动。第三个要素对机器人才是必要的。因此，更具体地，可以说人工智能（AI）是人类在充分理解和抽象自然智能（NI）的基础上，由计算机+机器学习算法实现的、一种人造的智能形式。“人工智能 AI”是“自然智能”（特别是“人的智能”）的模拟、延伸或扩展。“人工智能 AI”的研究对象是“机器智能” 与“智能机器”。

计算机时代，所谓的环境就是广义的数据（环境=广义的数据），包括有结构的数据（产生自物理系统的数据）和无结构的数据。当今社会收集的大数据，85%以上是无结构数据。感知环境、理解环境、最佳决策就是机器学习算法要解决的问题。机器学习算法是 AI 实现的核心基础。

从数据中挖掘出用于决策的有效信息，执行最佳决策（或判断），是当前人工智能的根本内涵。这就解释了为什么**有大数据、有高性能的计算机、有有效的算法**构成了当前人工智能的三大基础。严格地讲，大数据的含义是包含做出最佳决策或响应所需要的充分信息，并非单纯地指数据量巨大或数据类型很多。用于决策的信息不足会导致决策结果质量降低，甚至做出错误决策。

我认为首先要正确地理解 AI 的本质，然后深入理解 ML 算法，结合地震数据的特点和勘探地震的任务，才能比较准确地预测 AI 和 ML 在勘探地震中到底能起什么作用。

2 （大）数据、信息（特征）、决策与算法

2.1 大数据时代和 AI 时代解决问题的基本模式

人类已经进入大数据时代、信息时代、智能时代。逐渐形成了图 1 所示的问题处理模式。

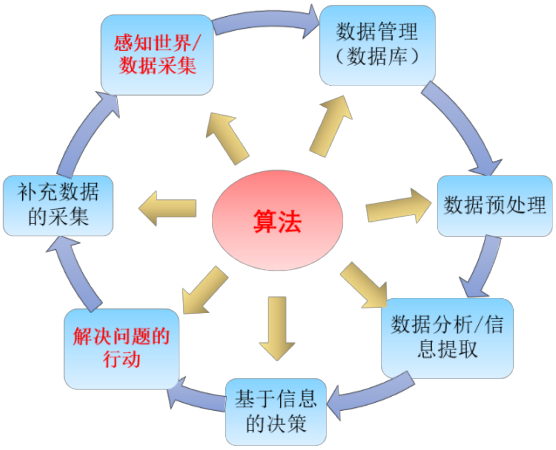


图 1 大数据时代解决问题的基本模式

在整个问题解决模式中，数据、信息、决策（推断）构成了整个环节中的核心要素。解决的基础或核心是算法。事实上，上述环节中的每一个节点都有相应的解决具体问题的算法。

2.2 大数据时代对数据的认识

人类已经进入大数据时代。数据甚至成了差不多最重要的生产要素和资源。谁拥有了数据，谁就拥有了

资本/资源。

什么是数据？数据是信息的载体。我比较认可这样的抽象定义。不包含信息的数据或不能从中提取出用来决策的有效信息的数据是不值得处理的。在计算机科学中，数据是指用于输入计算机进行处理，具有一定意义的数字、字母、符号和模拟量等的通称。

广义的数据包括：1、来自物理系统的、被各种传感器检测到的物理量。计算机上一般地以二进制数码表示这样的数据。2、非物理系统产生的“数据”，包括文字、符号、语音、文本、图片、图像、视频等等。计算机上依然以二进制数码表示为基础，但中间需要转换过程。一般地，来自物理系统的、被传感器检测到的数据是有明确的内在结构的。科学与工程领域所处理的数据主要是这类数据。它们可被称为狭义的数据。这类数据的管理不是困难问题，因此并没有引起大家的重视。而来自非物理系统的数据，一般地，初始状态往往是缺乏内在的逻辑关系的。针对这样的数据，首先应该在时空关系制约下建立内在的相互关系，形成有效的数据管理方式。数据管理本身就是数据处理的核心内容之一。数据库成为数据科学的重要内容之一是必然的。高维数据可以视为一个高维的函数。维度可以是原始采样的维度、也可以是变换后的维度。高维函数有内在的几何关系，或者说有内在的结构。所谓的信息就蕴含在这些结构中。

数据与信息的关系：数据中应包含信息，不包含信息的数据不值得花功夫处理；数据不等于信息；有些时候数据就是信息；同一数据包含的信息是多层次的，大多是潜藏的；同一个数据对不同的观察者而言包含的信息是不同的。

前已述及，从数据中挖掘出用于决策的有效信息，执行最佳决策（或判断），是当前人工智能的根本内涵。“数据”中有什么？数据能告诉我们什么？广义的“数据”中，蕴含着信息，也蕴含着知识。面向“知识”的 Learning From Data，目的是寻找因果关系，这更多地属于 AI 的范畴。狭义的“数据”中，蕴含着结构，蕴含着“特征”。面向“特征”的 Learning From Data，主要是寻找相关关系，这更多属于 ML 的范畴。

计算机可以进行数据处理、信息处理、甚至知识处理。但是，数据处理才是基础。最深层次的数据处理是奠基在数值数学提供的各种基础算法的基础上的。因此，数值数学才是数据分析学科的深层次基础。数据处理的根本目的：提供能用于决策的信息。

数据处理的基本逻辑：1、把数据预处理成能被算法处理的形式；2、一个或一组算法提取出数据中包含的信息。一般地，把数据视为随机变量（随机向量）的具体实现，挖掘其满足的统计模型（或统计模型中的参数），或建立起对数据中内蕴的几何结构的预测关系，认为提取出了数据中所包含的信息。譬如 PCA（Principal Component Analysis）、CCA（Canonical Correlation Analysis）、ICA（Independent Component Analysis）、NMF（Non-Negative Matrix Factorization，非负矩阵分解），字典学习、各种非线性降维方法等是经典的提取数据中几何结构的算法。

2.3 人工智能中的信息

信息是什么？信息是被消除的不确定性或被反映出的变化（Hartley, 1982；Ursul, 1971）。我特别认可这个来自信息科学或统计科学的信息定义。

信息的度量可被用来建立很基本的最优决策原则。这样的原则可以与 Bayes 后验概率密度函数最大化可以类比；与最大似然原则可以类比；与 Ockham 准则可以类比；与变分原理可以类比。我们知道，最佳推断、最佳决策或最佳估计本质上也是消除对推断/决策/估计结果的不确定性！

基于通讯科学产生的信息论，信息被定义为被消除的不确定性。不确定性显然与随机性有关。一个事件拥有不确定性，就认为这个事件拥有信息。信息量的度量自然应该与事件发生的概率联系起来。一个确定性的事件，概率为 1，认为是没有信息的。就是说，事件发生的概率越大，它包含的信息量越大。很显然，从这样的逻辑定义信息是非常合理的。

一个随机变量 x_i 对应的概率为 $P(x_i)$，它拥有的信息量定义为如下的形式，

$$I(x_i)=\log\left(\frac{1}{P(x_i)}\right) \tag{1}$$

(1)式称为随机变量 x_i 的自信息量。很显然，自信息是单个随机变量对应的某个事件发生的可能性的度量。按(1)式定义，自信息显然是一个随机变量。自信息熵是单个随机变量所有事件发生可能性的期望。自信息熵有如下定义，

$$H(X)=-\sum_{x\in\mathcal{X}}P(x)\log P(x) \tag{2}$$

上式可重写为：

$$H(X)=E_P I(X) \tag{3}$$

可以注意到，Kullback-Leibler distance、或 Kullback-Leibler Divergence 或相对熵在当前的机器学习算法或参数估计问题中有越来越广泛的应用，它被定义为，

$$D(p\|q)=\sum_{x\in\mathcal{X}}p(x)\log\frac{p(x)}{q(x)} \tag{4}$$

上式可以重写为，

$$D(p\|q)=E_p\log\frac{p(x)}{q(x)} \tag{5}$$

互信息被定义为：

$$I(X,Y)=\sum_{x\in\mathcal{X},\ y\in\mathcal{Y}}P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \tag{6}$$

同样地，(6)可以写为，

$$I(X,Y)=E_{P(x,y)}\log\frac{P(X,Y)}{P(X)P(Y)} \tag{7}$$

可以看出，相对熵 $D(p\|q)=\sum_{x\in\mathcal{X}}p(x)\log{p(x)}\big/{q(x)}$ 可以被解释为两种分布 $p(x)$ 和 $q(x)$ 之间“距离”或“差异”的度量。直观地，它是两种分布 $p(x)$ 和 $q(x)$ 规定的似然比对数的关于分布 $p(x)$ 的期望值。也可解释为：它是当真分布为 $p(x)$ 时，假设分布为 $q(x)$ 是无效性的度量。相对熵总是正的，只有在两种分布相等时，相对熵才等于零。可以证明：Kullback-Leibler distance、或 Kullback-Leibler Divergence 或相对熵是个凸函数。因此，相对熵总可被用来作为统计推断的原则。但是，相对熵不满足三角不等式。并不是一个真正的范数。

类比 Kullback-Leibler Divergence 或相对熵，(5b)式的互信息显然可以被重写为：

$$I(X,Y)=D\big(P(X,Y)\big\|P(X)P(Y)\big) \tag{8}$$

很显然，互信息是 Kullback-Leibler Divergence 或相对熵的一种特例。

除了相对熵和互信息被用来作为统计推断的原则外，一般意义下的地震数据分析中，往往用信息不足情形下的地震波成像、多信息联合的反演、信息综合利用等概念或提法来呼应大数据和信息时代对于高精度地、可靠地解决复杂勘探问题的需求。综合信息约束下地震波成像、油藏识别、地震解释，甚至钻井决策逐渐在油气勘探领域成为共识。相信综合信息利用会成为今后油气地震勘探领域的重要技术发展方向。

2.4 人工智能中的推断与决策

依据观测数据（随机变量\随机过程），提取其中的信息（统计推断\统计估计），进行最佳决策（统计决策），是大数据和智能时代，由 Learning Machine 来解决问题的基本逻辑。

Bayes 统计学派和经典统计学派（频率学派）发展了两种代表性的统计估计\统计推断\统计决策思想。Bayes 统计学派和经典统计学派思想的差异体现在：Bayes 统计学派把要估计的模型参数视为已知概率分布的随机变量（随机向量）；经典统计学派把要估计的模型参数视为确定的待估计变量。两者之间是有重大差异的。当前，Bayes 统计学派的思想完全占上风。

统计数据分析的本质就是在统计模型$\{P_\theta,\theta\in\Theta\}$的框架下估计随机数据中蕴含的信息，就是估计统计模型$\{P_\theta,\theta\in\Theta\}$ 中的参数θ或各阶统计量。本质上，这就是要建立随机数据（高维随机过程）的统计模型，也就是联合概率密度函数。我认为：这可以说是决策（推断）的最本质任务。实际数据分析中上，据此演化出了很多低级别的决策（推断）任务。

Bayes 统计思想的关键点：①把要估计的模型参数视为随机变量Θ，假设已知其分布p_Θ。②假设已知观测向量$\mathbf{X}$的先验条件分布$p_{X|\Theta}$。③由 Bayes 法则计算后验概率分布 $p_{\Theta|X}\left(\theta\middle|X=\mathbf{x}\right)$。④按照估计准则进行参数估计。

Bayes 统计推断思想中的估计准则包括：①后验条件期望作为参数估计结果，后验条件方差作为参数估计结果的精度评价；②MAP 准则；③均方误差最小准则；④线性均方误差最小准则，最经典的例子是 Wiener（反）滤波；⑤（加权）最小二乘准则，典型的实例是高维参数估计，具体地如 FWI、Tomography、LS_RTM、信号与图像建模等。

当前，推断、决策、估计或反演结果最佳的判决原则包括：①、Bayes 估计中后验概率密度函数的最大化。②、似然函数的最大化。④、估计结果的熵最大。4、交互熵（K-L 散度）和互信息最大。⑤如无必要勿增实体的原则（奥卡姆准则）。⑥变分原理。不同的原则背后的逻辑基础不同、理论假设不同、导出的算法不同、估计（或反演或推断）的结果也不尽相同。

关于统计估计\统计推断中的假设检验问题，其中假设检验的目的是判断我们依据统计模型做出的判断是否是可以接受的，或者说是否是合适的。假设检验问题是统计数据分析中的一个十分重要的问题，尤其在 ML/AI 算法中有重要地位。

（Bayes）统计决策理论： 依据样本数据（中蕴含的统计规律），在既定的决策代价函数下，基于最优决策的要求，结合先验信息，产生最佳决策行为和结果。样本空间；参数空间；决策行为空间是（ Bayes）统计决策的域。参数的先验信息和决策代价函数是（ Bayes）统计决策的另外两个问题。

（Bayes）统计决策是：依据数据样本产生的参数，并依据决策代价最小进行决策的。决策函数：设$X\sim P_\theta\left(\theta\in\Theta\right)$是随机变量（数据 X）满足的统计模型（概率密度函数，是参数θ的函数）。取值于决策（行动）空间$\mathcal{A}$的函数$\delta(\mathrm{X})$称为决策函数。决策风险函数：设$X\sim P_\theta\left(\theta\in\Theta\right)$是随机变量（数据 X）满足的统计模型，$L(\theta,a)$是损失函数。称决策函数$\delta(\mathrm{X})$平均损失$R(\theta,\delta)=E_\theta L\left(\theta,\delta(X)\right)$为决策风险函数。（Bayes）统计决策就是风险函数最小的决策。

综合评述：人类擅长的是（主要是）基于视觉模式进行抽象并形成概念认识。人类很不擅长基于大数据中蕴含的信息进行正确/合理的决策，这正是发展 AI、ML、DL 的根本动因。ML 算法借助高性能计算机可以从大数据中提炼出信息在 Bayes 决策理论指导下进行决策。这体现出的是人造的智能，它与人脑的智能是由本质区别的。人脑的智能是自然（进化）智能的最高代表。借鉴人脑智能，强化人工智能是必然的发展路线。

2.5 人工智能中算法

2.5.1 针对算法的基本观点

算法（Algorithm）是指解题方案的准确而完整的描述，是一系列解决问题的清晰指令，算法代表着用系统的方法描述解决问题的策略机制。算法能够对一定规范的输入，在有限时间内获得所要求的输出。如果一个算法有缺陷，或不适合于某个问题，执行这个算法将不会解决这个问题。不同的算法可能用不同的时间、空间或效率来完成同样的任务。一个算法的优劣可以用空间复杂度与时间复杂度来衡量。算法中的指令描述的是一个计算，当其运行时能从一个初始状态和（可能为空的）初始输入开始，经过一系列有限而清晰定义的状态，最终产生输出并停止于一个终态。

算法的特征：有穷性（ Finiteness ）：算法的有穷性是指算法必须能在执行有限个步骤之后终止；确切性（ Definiteness ）：算法的每一步骤必须有确切的定义；输入项（Input）：一个算法有 0 个或多个输入，以刻画运算对象的初始情况，所谓 0 个输入是指算法本身定出了初始条件；输出项（Output）：一个算法有一个或多个输出，以反映对输入数据加工后的结果。没有输出的算法是毫无意义的；可行性 (Effectiveness)：算法中执行的任何计算步骤都是可以被分解为基本的可执行的操作步骤，即每个计算步骤都可以在有限时间内完成（也称之为有效性）。

算法的要素包括：⑴、数据对象的运算和操作：计算机可以执行的基本操作是以指令的形式描述的。一个计算机系统能执行的所有指令的集合，成为该计算机系统的指令系统。一个计算机的基本运算和操作有如下四类：①算术运算：加减乘除等运算；②逻辑运算：或、且、非等运算；③关系运算：大于、小于、等于、不等于等运算；④数据传输：输入、输出、赋值等运算 。⑵、算法的控制结构：一个算法的功能结构不仅取决于所选用的操作，而且还与各操作之间的执行顺序有关。

对算法的基本观点：①算法指解题方案的准确而完整的描述。②算法代表着用系统的方法描述解决问题的策略机制。③不同的题目对应不同的算法。④有基本的算法和综（组）合的算法。⑤数值数学的目的就是提供在计算机上运行的、时间和空间复杂度低的、有效的算法。⑥数据处理的就是奠基在各种各样的算法的基础上的。⑦现代大数据和信息社会就是由各种各样的算法支撑起来的。⑧数学（数值数学）是现代大数据和信息社会的基石。

算法的主要处理对象：函数、泛函、算子、矩阵（标量、向量、张量）；算法针对的问题：包罗万象！算法针对的典型问题：①（离散数据表示的）函数的处理，插值、逼近、微分、积分；②常微分方程与偏微分方程的求解；③线性方程组与非线性方程组的求解；4 优化问题的求解。

2.5.2 地震数据分析中典型算法与 ANN 算法的对比

对于有结构的物理系统产生的数据，基于 Bayes 参数估计理论的参数化建模方法是根本性的数据分析方法。

非学习类算法的三种典型反问题模型：

图像处理中 ROF 模型：$J(\mathbf{u})=\frac{1}{2}(A\mathbf{u}-\mathbf{g})^2+\alpha TV(\mathbf{u})$ 。对于图像去噪问题， A=$\mathbf{I}$(单位矩阵)。

信号处理中去噪或反褶积模型：$J(\boldsymbol{\beta})=\frac{1}{2}\left(\Psi\boldsymbol{\beta}-d^{obs}\right)^2+\alpha|\boldsymbol{\beta}|_1$。

物理系统参数估计模型：$J(\mathbf{m})=\frac{1}{2}\left(A(\mathbf{m})-\mathbf{d}^{obs}\right)^2+\alpha\Omega(\mathbf{m})$。 $\Omega(\mathbf{m})$ 代表不同形式的正则化方式。

ANN/DNN 网络计算模式：ANN 是一种分布并行算法。ANN 模仿人脑计算的功能，是一种智能算法。ANN 的目的是建立起“广义数据”与决策目标之间的非线性映射关系

$$F(\mathbf{x},\mathbf{w})=\varphi\left(\sum_k w_{0k}\varphi\left(\sum_j w_{kj}\varphi\left(\cdots\varphi\left(\sum_i w_{li}\tilde{\mathbf{x}}_i\right)\right)\right)\right) \tag{9}$$

其中，$\tilde{\mathbf{x}}=\left(f_\theta^1(\mathbf{x}),\ f_\theta^2(\mathbf{x}),\cdots,\ f_\theta^b(\mathbf{x})\right)$， $\tilde{f}_\theta(\tilde{\mathbf{x}})=\sum_{j=1}^{b}\alpha_j\phi(\tilde{\mathbf{x}};\boldsymbol{\beta}_j)$。

人工神经网络是由高度抽象的、人工神经元构成的网络。本质上，这是一个计算模型。不过它仿生模拟了人脑神经元及人脑神经网络。事实上，是否真的仿生模拟了人脑神经元及人脑神经网络还很难说。经验地讲，很多仿生优化算法还是有比较好的结果的。既然人脑神经元构成的网络有如此强大的学习、联想、推断和决策能力，可以期望人工神经网络也应该有比较好的性能。

单个人工神经元的数学模型是描述由输入函数 **x** 到输出函数 **y** 之间关系的数学模型。我认为这个描述是没有问题的。输出函数 **y** 与输入函数 **x** 之间的关系可以用基函数族的线性组合来表示。泛函分析奠定了这方面的坚实基础。这个问题数学上绝对站得住脚。理论和实践也证明了这样的观点。但是输出函数 **y** 与输入函数 **x** 之间到底是否有关系，还是由人选择的合适的输出函数 **y** 与输入函数 **x** 决定的，即由标签数据决定的。风马牛不相及的输出函数与输入函数不可能找出什么有用的关系。因此，标签数据+DNN 方法是深度学习人工智能算法能否成果的关键。

做出正确/合理决策的根本基础：数据样本 X（或其特征）与决策目标 Y 之间存在较为线性的关系（或二者之间存在连续的关系）！因此，不能对（深度）ANN 算法（即 DL）进行神化！我认为：它仅仅是 Bayes 统计决策的智能化实现而已。

在采样数据 X 的统计信息不足（原因很复杂，譬如采样不充分、噪声太大等），且决策目标与采样数据 X 的关系过于非线性时，类比于密度 ρ 与到达时 t 的关系在 FWI 中估计密度 ρ ，无论构造什么样的网络构型和算法，也不可能训练它并做出正确的决策。

样本数据 X 与决策目标之间的强非线性性关系依然是做出正确/合理决策的最大障碍。数据与决策目标之间的映射关系往往是一个类似于非线性的“广义逆”。数学上，对非线性的“广义逆”研究很少。线性广义逆的存在也不能保证解的稳定性和收敛性。这应该是我们关注的真正的焦点问题！

（深度）ANN 算法并没有回答如何能在数据样本 X（或其特征）与决策目标之间存在（强）非线性关系（或不很连续的关系）时能给出最佳的判决。所有 AI 算法都没有回答这个问题！这是一个深层次的理论问题。目前的（深度）ANN 技术基本上是基于 30 年前的思想积累。未来，对人类大脑功能的理解、对数学的理解是深化 DL，甚至提出全新计算模式的根本途径。任重而道远！

2.5.3 对数据分析中重要算法的总体描述

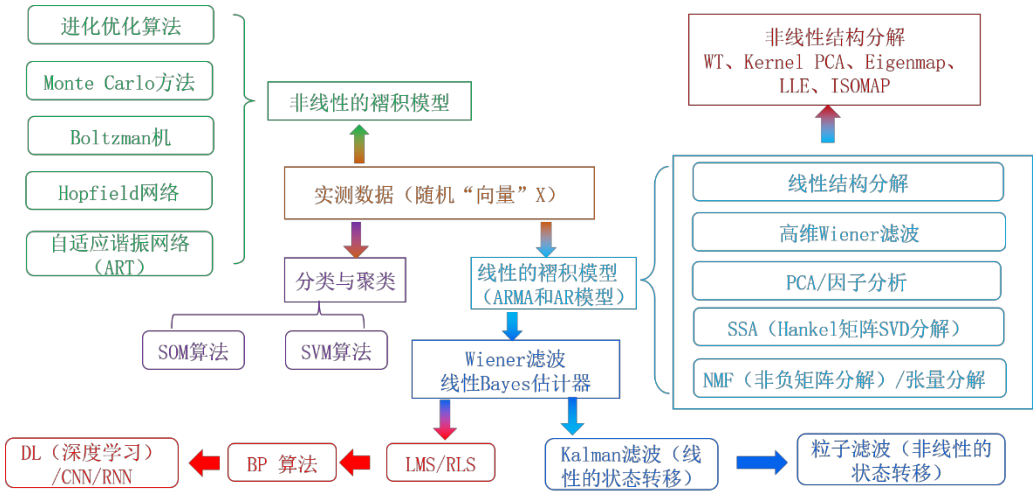


图 2 数据分析中主要算法的总体描述图

2.6 综合性结论

当今大数据、人工智能时代，数据+算法构成了这个社会正常运转的很根本的基础。由数据中提取信息，形成决策，构成了基本的运转模式。算法在其中起到了核心的作用。各种信息的融合，共同支持决策，提升

决策可靠性，是逐渐形成的基本逻辑。尽管数据处理依然是信息处理和智能处理的根基，但信息处理和智能处理正逐渐变成更普遍的工作内容。勘探地震数据分析也要逐渐从以处理狭义的物理系统数据为主（譬如信号处理、速度估计与偏移成像）转到处理狭义的物理系统数据和处理地质、测井、岩性这些非物理系统数据并重的阶段。以正确的决策为目标的数据分析阶段相信逐渐会很快到来。对人脑这个超强的信息处理和智能决策机器的仿真会不断地深化下去。

3 勘探地震数据分析的本质问题

油气地震勘探的终极目的是定位、刻画、描述和评价油气藏，提供井位并对油气藏开发提供支撑。油气地震勘探的核心问题之一是地震数据分析。

地震数据分析的核心是：由采集到的叠前地震数据以及其它关于要反演参数的先验信息，进行地震波反演成像，得到地下介质的定位、几何特征的估计和岩石物性参数（速度、波阻抗等）的估计。关键步骤包括：正问题（描述波的传播过程）+反问题（进行介质参数的估计）。

勘探地震数据分析的本质是：存在一个勘探地震系统，包括三维探区介质系统；震源激发；检波器接收；系统输出（叠前地震数据体）三个核心要素。存在一个对勘探地震系统的物理过程描述，一般地，用波动方程或由波动方程退化出来的高维褶积系统描述该勘探地震系统。高维褶积系统建立起地下介质参数变化与波场变化之间的关系，这是地震波成像最核心的基础。由实测数据体，基于高维褶积系统，在某种参数估计理论下构建系统参数估计反问题，它是地震数据数据分析中的核心问题。

Bayes 参数估计理论是地震波反演成像的理论基础。我们提出的对勘探地震反问题的定位：所有的反问题本质上都是在信息不充分情况下的统计决策问题！

地震波在实际介质中的激发、传播和接收过程是地震波偏移成像和反演成像的基础。勘探地震数据分析本质上是模型（高维褶积系统）驱动的参数估计问题。很显然，这与 Learning From Data 的机器学习是有本质差异的。一般性的数据分析，正因为缺乏类似高维褶积系统这样的因果关系式的存在，不得已从数据中寻找数据与其中的潜变量之间的关系式，然后基于这样的关系，进行后续的分类等决策判断。但是，机器学习算法挖掘出的数据与其中的潜变量之间的关系式，反映二者的相关关系，不一定是因果关系。因此，基于此关系做出的决策或判断更不可靠。本质上，机器学习算法特别需要大数据，期望用统计特征完整的大数据，使得找到的数据与其中的潜变量之间的关系式更可靠。基于这样的关系能做出可靠的决策或判断。再次特别指出，叠前地震数据只是体量大，并非统计特征完整的大数据，这是判断 AI/ML 在勘探地震中能否发挥核心作用、甚至能用在哪些场景下的关键因素。

勘探地震数据分析本质上是模型（高维褶积系统）驱动的参数估计问题。因此，它只要小样本数据就可以得到可靠的参数估计结果。当然，”复杂地表、复杂介质和复杂储层”，即所谓三复杂情形下，参数估计问题（FWI、Tomography、LS_PSDM/LS_RTM）也是非线性很强的问题，并不容易解决。各种复杂探区连地震波偏移成像都不能给出满足勘探需求的结果就充分证明了这一点。基于数据驱动的机器学习算法能在三复杂情形下能更好地解决地震波成像问题吗？我对此很不乐观或者说更不乐观！

影响实测地震波场（振幅、相位、波形特征）的因素可以归为如下四个方面：①与地表有关的因素，体现为地表介质与震源和检波器的相互作用及波在表层（风化层）的传播引起的道与道之间的振幅与相位的差异。②与反射界面有关的因素，主要体现为薄层调谐作用和反射界面的曲率等。③与传播有关的因素，包括球面扩散、透射损失、介质的非弹性吸收衰减、散射吸收衰减等。④与采集系统有关的因素，包括观测系统非规则、检波器组合的响应特征、震源的响应特征。

油气地震勘探中地震波传播正问题的描述主要体现在：在牛顿力学体系下，建立弹性波方程应力应变分析，与岩石物理学和连续介质力学提出广义 Hook’s Law，在牛顿第二定律下产生各种形式的波动方程。之所以在基础的波动方程基础上，发展出非常多的变种，主要是适应勘探地震介质以层状为主的情况、适应地表激发接收观测系统、适应以反射波为主的波场特征、适应地震数据分析算法复杂性的需求。这些变种可知可以罗列如下：21 个参数的各向异性（粘）弹性波方程；9 个参数的正交各向异性（粘）弹性波方程，进一步退化为标量的拟声波方程，进行 ORT_RTM 和积分法 ORT_PSDM；5 个参数的 TI 介质各向异性（粘）弹性波方程，并退化为标量的拟声波方程（各项异性介质成像的主要形式），进行 TI_RTM 和积分法 TI_PSDM；

2 个参数的各向同性（粘）弹性波方程，可以发展出弹性波 RTM，若退化为标量声波方程，可以发展出 ISO_RTM 和 ISO_PSDM；最后是退化为标量声波方程，这是勘探地震中最重要的波传播模型。发展出来很多种标量声波方程的数值解法：射线追踪、Gauss Beam 方法、傍轴近似方法、Maslov 近似方法，基于这些数值解法发展出各种 Kirchhoff 积分 PSDM 和射线层析成像方法；标量声波方程的 FD 解法、FE 解法等，发展出波动方程 PSDM 和波动理论层析成像方法。另外，进一步退化出的各向同性介质和各向异性介质中的直达波、反射波、折射波、Diving Wave 的时距关系式在地震波成像中起到了至关重要的作用。NMO、DMO、MZO、CRS、PSTM 都是基于上述时距关系发展出的成像方法。基于局部平面波入射到弹性介质分界面上的物理过程导出的 Zoeppritz 方程在勘探地震数据分析中有广泛的应用，这当然与勘探地震介质以层状为主，仅仅在地表的激发接收观测由密切关系。发展出了各种线性近似的 AVA 关系，反映振幅随入射角的变化。同样地，在 Bayes 参数估计理论下，提出了各种各样的 AVA 反演方法。数据基础的不扎实（即角度成像道集误差太大）导致 AVA 反演结果的可信度严重下降。小尺度目标地质体（而不是层状介质）的存在也侵蚀了 Zoeppritz 方程的物理基础。

总而言之，地震波在实际介质中激发传播与接收和地震波反演成像的传播模型（1D 褶积模型、Zoeppritz 方程、各种时距关系、程函方程、波动方程）是我们理解地震数据采集、地震波反演成像的根本。不同的波传播数学模型对应不同的地震波成像方法，譬如，1D 褶积模型---1D 反褶积、波阻抗反演；Zoeppritz 方程---AVA 弹性参数反演；各种时距关系---动校正、静校正、噪音压制、DMO，PSTM；程函方程---旅行时计算、Kirchhoff PSDM、Beam PSDM；波动方程---One-Way WE PSDM，RTM

地震波成像的根本在于地下介质变化引起地表实测波场的变化。因此，地震波在勘探地震系统中传播产生的重要波现象是地震波成像的物理基础。对地震波成像有贡献或存在重大应的波现象包括：①地震波衰减（Attenuation）；②地震波折射（Refraction）；③潜水波（Diving Waves）；④地震波反射（Reflection）；⑤地震波绕射（Diffraction）；⑥与海水面和/或自由地表有关的多次波和面波；⑦反射界面处波的分解（AVA/AVAz）。这些波现象与地震波反演成像息息相关。

从 Bayes 估计理论看，收集、评价、认识、利用与待反演参数相关的信息对于提高地震波成像的精度至关重要的。譬如探区地表情况、探区地质（构造、沉积、岩性等）情况、探区钻井信息、探区岩石物理知识等等对于提高地震波成像精度都是非常重要的。在大数据时代，所谓的大数据的含义真正体现在充分利用各种来源的、与储层介质有关的数据（通常被称为先验信息），提升地震波反演的可靠性（或精度）。

基于上述认识，提出图 3 表达的勘探地震数据分析的总体框架。

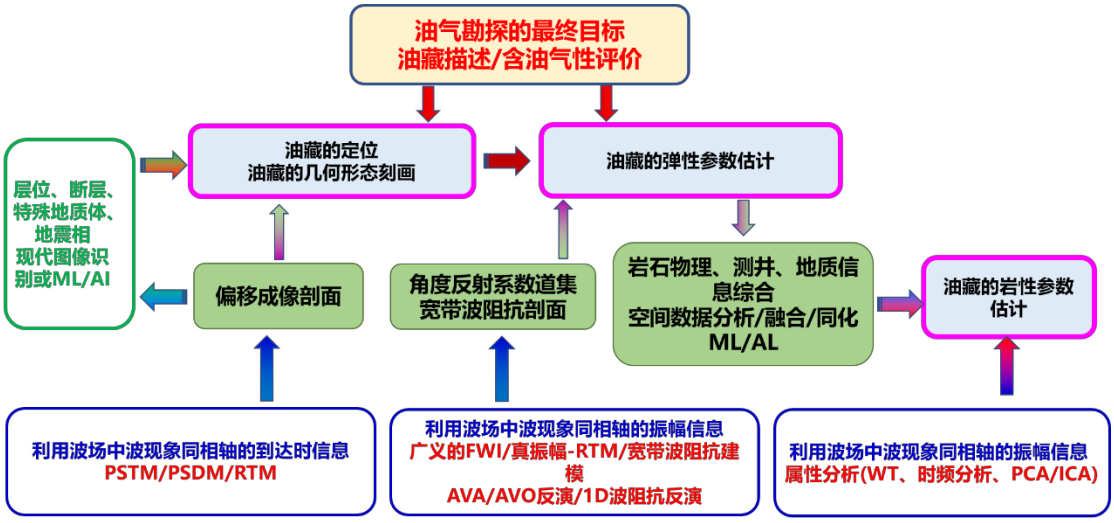


图 3 油气勘探地震数据分析的总体框架

依据图 3，可以列出地震数据分析的基本流程如下：①地震数据预处理；②射线与波动理论透射波层析成像建立近地表速度模型；③静校正+动校正+（类）CMP 道集反射波初始模型； ④（迭代的）角度成像道

集速度分析和/或射线理论层析速度反演与建模；⑤（迭代的）反射波波动理论层析速度反演与建模；⑥基于保真 PSDM 的角度反射系数成像道集产生及后处理；⑦AVA 分析与波阻抗反演；⑧油藏描述与评价。在上述流程结果的基础上，进入油气藏开发阶段。

关于地震数据预处理，我们把地震数据抽象为：线性（非线性）同相轴飘在不同统计特征的随机噪声中。对应的数学模型为：

$$d^{obs}(\mathbf{x},t)=S(\mathbf{x},t)+\eta(\mathbf{x},t) \tag{10}$$

其中，$S(\mathbf{x},t)=\Psi\boldsymbol{\beta}$。据此，提出一般性的反问题，

$$J(\boldsymbol{\beta})=\frac{1}{2}\left(S-d^{obs}\right)^2+\alpha|\boldsymbol{\beta}|_1=\frac{1}{2}\left(\Psi\boldsymbol{\beta}-d^{obs}\right)^2+\alpha|\boldsymbol{\beta}|_1 \tag{11}$$

当前，FWI 和 LS_RTM 形成了地震波成像的理论和技术发展的核心。实质上，FWI 和 LS_RTM 不过是下述 Bayes 参数估计理论在勘探地震中的一个应用实例。

针对非常一般性的正问题 $\mathbf{d}=\mathbf{g}(\mathbf{m})$，它可以来自各种领域中的物理问题，当然也包括油气勘探地震。可以写出如下的 Bayes 公式，

$$\rho_{\mathbf{M}}(\mathbf{m}\,|\,\mathbf{d})=\frac{\rho(\mathbf{d},\mathbf{m})}{\rho(\mathbf{d})}=\frac{\rho(\mathbf{d}\,|\,\mathbf{m})\rho(\mathbf{m})}{\rho(\mathbf{d})} \tag{12}$$

已知后验概率密度函数 $\rho_{\mathbf{M}}(\mathbf{m}\,|\,\mathbf{d})$ 后，参数 $\mathbf{m}$ 的估计结果应该是，

$$\hat{\mathbf{m}}=\int\mathbf{m}\rho_{\mathbf{M}}(\mathbf{m}\,|\,\mathbf{d})dm\quad C_M=\int(\mathbf{m}-\hat{\mathbf{m}})(\mathbf{m}-\hat{\mathbf{m}})^T\rho_{\mathbf{M}}(\mathbf{m}\,|\,\mathbf{d})d\mathbf{m} \tag{13}$$

由于后验概率密度函数 $\rho_{\mathbf{M}}(\mathbf{m}\,|\,\mathbf{d})$ 的期望和方差在高维参数估计时很难计算，因此有 MAP 反演解，

$$\hat{\mathbf{m}}_{\mathrm{MAP}}=\underset{\mathbf{m}}{arg\ max}\ \rho_{\mathbf{M}}(\mathbf{m}\,|\,\mathbf{d}) \tag{14}$$

由于后验概率密度函数 $\rho_{\mathbf{M}}(\mathbf{m}\,|\,\mathbf{d})$ 具体形式很难获知，假设先验概率密度函数 $\rho(\mathbf{d}\,|\,\mathbf{m})$ 和 $\rho(\mathbf{m})$ 都是 Guass 分布，因此有等价于 MAP 问题的代价泛函，

$$\begin{aligned}\hat{\mathbf{m}}&=\underset{\mathbf{m}}{arg\ min}\ \{S(\mathbf{m})\}\\&=\underset{\mathbf{m}}{arg\ min}\ \left\{(\mathbf{m}-\mathbf{m}_{prior})^T\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})+(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})^T\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})\right\}\end{aligned} \tag{15}$$

的极小化问题。

由于正问题 $\mathbf{d}=\mathbf{g}(\mathbf{m})$ 非线性导致代价泛函 $S(\mathbf{m})$ 非凸，因此引入对正问题的线性化，

$$\begin{aligned}\mathbf{d}_{obs}&=\mathbf{g}(\mathbf{m}_{\mathrm{B}}+\Delta\mathbf{m})=\mathbf{g}(\mathbf{m}_{\mathrm{B}})+\frac{\partial\mathbf{g}}{\partial\mathbf{m}}\Delta\mathbf{m}+\frac{1}{2}\frac{\partial^2\mathbf{g}}{\partial\mathbf{m}^2}\Delta\mathbf{m}^2+\cdots\\\Delta\mathbf{d}_{obs}&\approx\mathbf{g}(\mathbf{m}_{\mathrm{B}}+\Delta\mathbf{m})-\mathbf{g}(\mathbf{m}_{\mathrm{B}})=\frac{\partial\mathbf{g}}{\partial\mathbf{m}}\Delta\mathbf{m}\\\Delta\mathbf{d}_{obs}&=\mathrm{G}\Delta\mathbf{m}\end{aligned} \tag{16}$$

正问题线性化后的凸泛函反问题为，

$$\hat{\mathbf{m}}=\underset{\mathbf{m}}{arg\ min}\ \left\{(\mathbf{m}-\mathbf{m}_{prior})^T\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})+(\mathrm{G}\Delta\mathbf{m}-\Delta\mathbf{d}_{obs})^T\mathbf{C}_{\mathbf{D}}^{-1}(\mathrm{G}\Delta\mathbf{m}-\Delta\mathbf{d}_{obs})\right\} \tag{17}$$

事实上，这个线性化反问题既包括速度层析成像的基本理论，也包括 LS_RTM/LS_PSDM 的基本理论。还可以说，地震波成像、地震数据去噪、规则化、插值、去混叠等预处理都可以视为 Bayes 理论下的参数估计问题，因为它们都是线性正问题下的某种参数估计问题。

两宽一高”的叠前地震数据+关于地下介质参数的先验信息是地震波成像需要的基础数据。复杂介质情形下的勘探地震中的参数估计问题（地震波反演成像问题）是一个典型的强非线性反问题！FWI（LS_RTM）是地震波反演成像的代表性方法技术。它们是 Bayes 估计理论在勘探地震中的应用实例。复杂介质情形下，它们都是强非线性反问题。这是 FWI（LS_RTM）难以实用化的根本原因。

如何把 Bayes 参数估计理论下的强非线性的 FWI 推进到实际应用？

我们提出的具体方案如图 4 所示。本质思想是用一组凸的或较凸的反问题的组合来近似实现强非线性的 FWI 的成像目标。

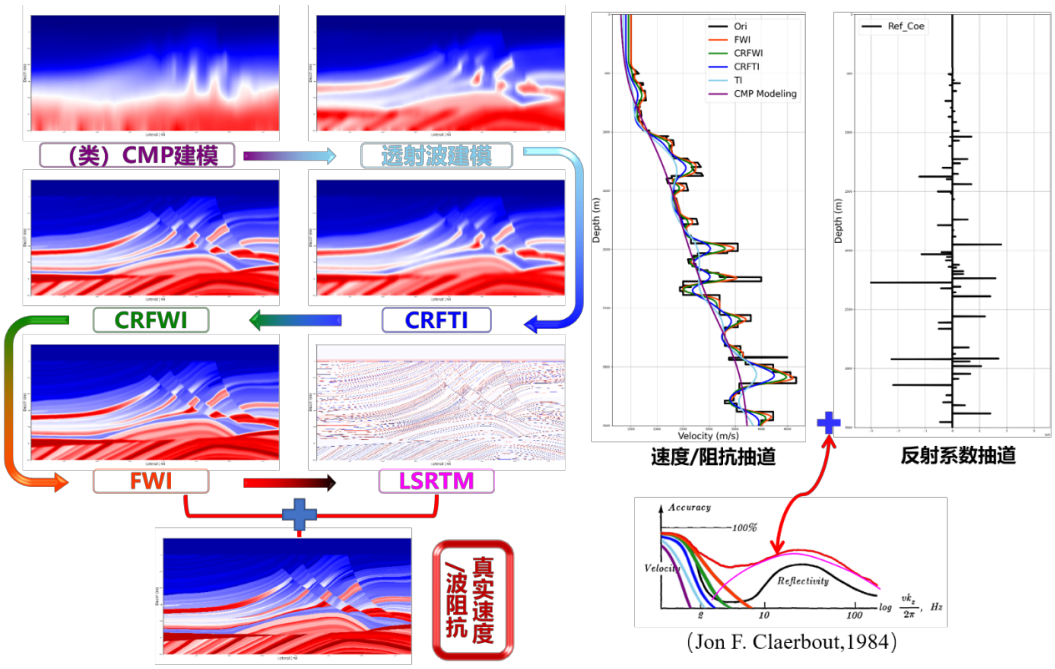


图 4 强非线性的 FWI 推进到实际应用的技术流程

为了构建凸的反问题，重点是提出更合适的介质参数与波场（波现象）之间的关系式（算子）。所谓更合适的介质参数与波场（波现象）之间的关系最好是一对一的因果关系。据此，WPI 提出特征波反演（CWI）成像理念和技术路线。并把 CWI 推进到宽带波阻抗成像。宽带波阻抗成像被表达为一个信息融合问题，而不是表达为一个强非线性的参数估计（或参数反演）问题。因此，在“两宽一高”数据支持下，具备了切实的可行性。

基于此，WPI 提出了“两宽一高”+Bayes 估计理论反演成像时代的新的地震数据分析技术流程。其中重要的步骤包括：①地震波+电磁波波场特征分析与地表岩性区域分划。②针对初至波的 CMP 道集速度扫描及初始速度建模。③高维多属性+进化优化算法+Markov 最佳演化的初至识别。④初至波到达时 Beam 层析；到达时波动理论层析。其中包括各向同性速度估计与建模；各向异性速度估计与建模；Q 值估计与建模。⑤基于 CMP 道集与类 CMP 道集（PSTM 及 PSDM CIG 转化来的 CMP 道集）自动与智能速度建模。⑥完善 CIG 道集射线层析速度建模。⑦特征反射波波动理论层析速度估计与建模。基于此，过渡到反射波 FWI 建模。⑧最小二乘保真成像方法，目标主要是保真的方位角度反射系数估计。基于此，过渡到所谓的全频带 FWI，或者称为 FWI-Imaging。⑨包括自由表面多次波的成像，最后发展到包含层间多次波的成像。⑩最终统一到宽带波阻抗成像。在上述流程结果的基础上，进入油气藏开发阶段。这样的地震数据分析流程显然是经典的地震数据分析流程的升级。它已经与大数据分析和 AI 有比较多的衔接。

4 AI 和 ML 在勘探地震中的应用前景的分析

千禧年后，油气地震勘探技术有了长足的进步。其中最关键的技术进步体现在三个方面：“两宽一高”地震数据采集技术；FWI+LS_RTM 地震波成像技术；大数据和人工智能在勘探地震中各环节的应用。

“两宽一高”地震数据采集是高精度地震成像的基本要求。事实上，“两宽一高”地震勘探技术是一项综合性的技术体系，是今后油气地震勘探领域居于中心地位的技术系列。它不仅仅是指“两宽一高”地震数据采集技术。更重要地，从海量的“两宽一高”地震数据中挖掘出所包含的与储层有关的信息，进而对油气储层进行更精确的描述，并渗透到储层开发阶段，提高整个勘探过程的经济效益，才是“两宽一高”地震勘探技术的真正目标。

遗憾的是，从“两宽一高”数据中挖掘出有助于油藏描述的信息的工作进展得相当地缓慢，FWI+LS_RTM 也并没有带来地震波成像和油藏描述精度的显著提升。大数据+高性能计算技术催生的 AI/ML 技术，给当今社会的很多领域带来了颠覆性的变化，也在油气勘探领域中的某些应用场景中得到了有效的应用。但是，AI 和 ML 能否为“两宽一高”地震数据中的信息深度挖掘并显著提升油藏描述精度带来变革性技术？

从数据中挖掘出用于决策的有效信息，执行最佳决策（或判断），是当前人工智能的根本内涵。大数据、信息挖掘、基于信息的最佳决策是人工智能的根本要点。在勘探地震领域，AI 和 ML 能否起到颠覆性的作用，根本上还在于是否有统计充分的大数据、是否能挖掘出能进行正确决策的信息、是否有有效的决策算法。总体而言，这些都不令人乐观，不能对此抱太大期望。最现实的，还是选择合适的具体场景，选择并优化合适的算法，把具体场景下的问题解决得好一些，超过常规方法技术。当前，AI 领域也逐渐从追求理论算法上的跨越转向解决好具体场景下的问题。

AI/ML 技术在勘探地震领域、在地震数据分析中的应用是无法回避的。它在信息挖掘的能力（或潜力）上是优于非学习类算法的。利用 AI/ML 技术，首先地震波成像处理流程的自动化和智能化程度一定会得到有效地改善，地震地质解释的自动化和智能化程度也一定会得到有效地改善，至少工作效率会有极大地提升。理论上讲，利用 AI/ML 在数据域/图像域的信息挖掘能力，还可以为参数估计方法（反演成像方法）提供更多的先验信息，从而提升反演成像的精度。事实上，AI/ML 技术还可以深度渗入到数据采集诸环节、数据预处理环节、地震地质解释环节、钻井与开发环节、整个油田生产的管理环节等方方面面。具体一些，我认为：在 Bayes 决策总体理论下，开展高维数据和不同类型数据中的相关信息挖掘，用在去噪、数据规则化、Deblending 等数据预处理环节；用在地表一致性处理环节；用在地表岩性空间变化识别环节；用在特征波识别与提取环节；用在成像剖面上的地质体几何结构识别环节；用在建模过程中的信息融合环节等，提高数据预处理、地震波成像、地震地质解释的自动化程度和精度，是机器学习在勘探地震中应用的、正确的技术发展方向。不能一开始就期望开发出一个油气勘探机器人，把当前油气勘探很多环节上的人员大都替换掉，期望极大地提高油气勘探的成功率和效益，这是不现实的。

下面，对勘探地震全过程可能涉及到的智能算法应用场景做一个梳理。数据采集阶段：智能的海上和陆上 Node 检波器系统；智能的可控震源系统，尤其是客户定制反射子波的可控震源系统；智能的高效采集方案实施控制系统；……。数据预处理阶段：智能的去噪、规则化、Deblending 等方法；智能的地表一致性处理方法；智能的保真高分辨预处理质量控制系统；……。自动与智能成像阶段：（高光谱）电磁波场与地震波场特征变化与地表岩性变化关系的智能辨识及应用；自动与智能的透射波层析反演方法技术，主要体现在智能算法+初至波辨识；智能的（类）CMP 道集速度谱解释与建模方法技术，主要体现在（T0,Vrms）对的合理选择；自动与智能的波动理论特征反射波旅行时层析反演方法技术，主要体现在智能算法+特征反射波辨识和智能算法+地质体形态与地震相识别；盲反演与超级分辨率反演的智能算法实现，主要体现在 PSF 的智能估算；……。地震地质解释阶段：层位与断层自动追踪与解释技术；五维地震数据中，随方位和张角变化的 AVAZ、VVAZ、FVAZ 信息的自动与智能提取方法技术；……。油藏描述阶段：井震数据的自动与智能匹配；成像剖面或成像道集属性信息及含油气性辨识；

……。

我认为：深刻理解（人工）智能的本质特点，深刻理解 AI/ML 算法的本质特点，深刻理解地震数据的本质特点，是把 AI 和 ML 在地震数据分析中充分发挥作用的基础。不要盲目地对 AI 产生迷信，在深刻地理解 AI/ML 的理论背景基础上，选择合适的应用场景+合理的算法才是 AI/ML 能发挥作用的根本。

5 结论性认识

从数据中挖掘出用于决策的有效信息，执行最佳决策（或判断），是当前人工智能的根本内涵。油气地震勘探的目标是精确刻画与评价油气藏，做出是否钻井的决策。这的确是适合 AI 解决的决策问题。

无论是否进入大数据和人工智能阶段，油气地震勘探的核心目标依然是精确地刻画和评价油气藏。“两宽一高”数据采集技术+广义的地震波反演成像技术（推进到宽带波阻抗成像）+多种信息融合的油藏描述与评价技术，始终是三个核心问题。

但是，发展替代人的智能的方法技术（AI 技术），替代人的部分智能劳动，提高工作效率、改善工作效果，是人类发展的必然阶段（即当前所谓的大数据和人工智能阶段！）。决策是智能劳动的核心任务，发展能自主决策的算法就是十分必要的。用智能学习算法在所有合适的应用场景改进当前地震勘探技术也是十分必要的。

我认为：发展出自动与智能的地震波成像处理系统应该是可以期待的。是否有必要和可能开发出能取代人进行含油气判决和钻井决策的机器人是需要慎重抉择的，尽管这是 AI 的真正目的。但我对此很不乐观！

参考文献

[1]. Abubakar A. Machine Learning for Geoscience Applications[C]//81st EAGE Conference and Exhibition 2019 Workshop Programme. European Association of Geoscientists & Engineers, 2019, 2019(1): 1-1.
[2]. Aharchaou M, Baumstein A. Deep learning-based artificial bandwidth extension: Training on ultrasparse OBN to enhance towed-streamer FWI[J]. The Leading Edge, 2020, 39(10): 718-726.
[3]. Alaudah Y, Michałowicz P, Alfarraj M, et al. A machine-learning benchmark for facies classification[J]. Interpretation, 2019, 7(3): SE175-SE187.
[4]. Alkhalifah, T., Almomin, A., & Naamani, A. (2021). Machine-driven earth exploration: Artificial intelligence in oil and gas. The Leading Edge, 40(4), 298-301.
[5]. Araya-Polo M, Farris S, Florez M. Deep learning-driven velocity model building workflow[J]. The Leading Edge, 2019, 38(11): 872a1-872a9.
[6]. Araya-Polo M, Jennings J, Adler A, et al. Deep-learning tomography[J]. The Leading Edge, 2018, 37(1): 58-66.
[7]. Baardman R H, Hegge R F. Machine learning approaches for use in deblending[J]. The Leading Edge, 2020, 39(3): 188-194.
[8]. Bergen K J, Johnson P A, de Hoop M V, et al. Machine learning for data-driven discovery in solid Earth geoscience[J]. Science, 2019, 363(6433): eaau0323.
[9]. Bianco M J, Gerstoft P, Traer J, et al. Machine learning in acoustics: Theory and applications[J]. The Journal of the Acoustical Society of America, 2019, 146(5): 3590-3628.
[10]. Caté A, Perozzi L, Gloaguen E, et al. Machine learning as a tool for geologists[J]. The Leading Edge, 2017, 36(3): 215-219.
[11]. Charu C. Aggarwal, Neural Networks and Deep Learning, Springer International Publishing, 2018
[12]. Chen Y, Schuster G T. Seismic inversion by Newtonian machine learning[J]. Geophysics, 2020, 85(4): WA185-WA200.
[13]. Chevitarese D, Szwarcman D, Silva R M D, et al. Seismic facies segmentation using deep learning[J]. AAPG Annual and Exhibition, 2018.
[14]. de Bragança R, Vamaraju J, Sen M K. Full-waveform inversion using machine learning optimization techniques[M]//SEG Technical Program Expanded Abstracts 2020. Society of Exploration Geophysicists, 2020: 865-869.
[15]. Fielding S R, Davies L. Drivers of imbalance in machine learning uptake across geology and geophysics[J]. First Break, 2019, 37(9): 59-63.
[16]. Hall B. Facies classification using machine learning[J]. The Leading Edge, 2016, 35(10): 906-909.
[17]. Hou S, Messud J. Machine learning for seismic processing: The path to fulfilling promises[C]//First International Meeting for Applied Geoscience & Energy. Society of Exploration Geophysicists, 2021: 3204-3208.
[18]. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learing, MIT Press, 2016
[19]. James O. Berger, Statistical Decision Theory and Bayesian Analysis, Springer, 1985

[20]. Jia Y, Ma J. What can machine learning do for seismic data processing? An interpolation application[J]. Geophysics, 2017, 82(3): V163-V177.
[21]. Jones I F. Tutorial: a geophysical perspective on machine learning[J]. First Break, 2021, 39(3): 35-42.
[22]. Kim Y, Nakata N. Geophysical inversion versus machine learning in inverse problems[J]. The Leading Edge, 2018, 37(12): 894-901.
[23]. Kong Q, Trugman D T, Ross Z E, et al. Machine learning in seismology: Turning data into insights[J]. Seismological Research Letters, 2019, 90(1): 3-14.
[24]. Li W, Hu W, Abubakar A. Machine learning and data analytics for geoscience applications—Introduction[J]. Geophysics, 2020, 85(4): WAi-WAii.
[25]. Maniar H, Ryali S, Kulkarni M S, et al. Machine-learning methods in geoscience[M]//SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists, 2018: 4638-4642.
[26]. Menon H. Machine Learning Applications in Oil & Gas Exploration & Production[C]//Second HGS and EAGE Conference on Latin America. European Association of Geoscientists & Engineers, 2020, 2020(1): 1-6.
[27]. Moghadas D, Badorreck A. Machine learning to estimate soil moisture from geophysical measurements of electrical conductivity[J]. Near Surface Geophysics, 2019, 17(2): 181-195.
[28]. Naeini E Z, Prindle K. Machine learning and learning from machines[J]. The Leading Edge, 2018, 37(12): 886-893.
[29]. Naeini E Z, Uwaifo J. Transfer learning and Auto-ML: A geoscience perspective[J]. First Break, 2019, 37(9): 65-71.
[30]. Qi J, Zhang B, Lyu B, et al. Seismic attribute selection for machine-learning-based facies analysisChoosing the best seismic attributes[J]. Geophysics, 2020, 85(2): O17-O35.
[31]. Rama Chellappa and Sergios Theodoridis (Edited), Signal Processing Theory and Machine Learning, Elsevier, 2014
[32]. Russell B. Machine learning and geophysical inversion—A numerical study[J]. The Leading Edge, 2019, 38(7): 512-519.
[33]. Sacrey D, Roden R. Solving exploration problems with machine learning[J]. First Break, 2018, 36(6): 67-72.
[34]. Sergios Theodoridis, Machine Learning – A Bayesian and Optimization Perspective, Elsevier, 2015
[35]. Shai Shalev-Shwartz and Shai Ben-David, Understanding Machine Learning – From Theory to Algorithms, Cambridge university Press
[36]. Sun B, Alkhalifah T. ML-descent: An optimization algorithm for full-waveform inversion using machine learning[J]. Geophysics, 2020, 85(6): R477-R492.
[37]. Trevor Hastie, Robert Tibshirani, and Martin Wainwright, Statistical Learning with Sparsity – The Lasso and Generalizations, CRC Press, 2015
[38]. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical learning - Data Mining, Inference, and Prediction, Springer, 2017
[39]. Wrona T, Pan I, Bell R E, et al. 3D seismic interpretation with deep learning: A brief introduction[J]. The Leading Edge, 2021, 40(7): 524-532.
[40]. Wrona T, Pan I, Gawthorpe R L, et al. Seismic facies analysis using machine learning[J]. Geophysics, 2018, 83(5): O83-O95.
[41]. Yang F, Ma J. Deep-learning inversion: A next-generation seismic velocity model building method[J]. Geophysics, 2019, 84(4): R583-R599.

L0 与 L1 正则化约束下反问题解的性质分析

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】自然和社会科学中非常多的问题都可以归结为输入、系统和输出，并由线性或非线性的褶积关系表达这个系统，表达系统输入和输出之间的关系。在应用层次上，表现为由实测的输出状态变量的观测值来估计系统本身（系统参数估计）、或估计系统输入。估计的基本原则是：由最佳估计得到的输入和/或系统参数能最佳地预测实测的系统输出。这是一个反问题。反问题的解往往是非唯一的，反问题的解应该具有什么性质？什么样的解才能解决实际问题？到目前为止，这个问题并没有一个完整的答案。原则上，反问题的解首先应该是由正问题决定的，或者说它的可行解集（可行解空间）是由正问题决定的。但是，这个可行解空间可能是巨大的。因此，必须要引入正则化条件，从巨大的可行解集中选择合理的解或满足先验要求的解。这种看似别扭的逻辑正是求解反问题的精髓。本文对比分析 Bayes 估计理论的解、最大似然估计理论下的解、L_2 范数下的解、L_1 范数下的解和 L_0 范数下的解，给出反演解性质的对比分析。并简单评价 L_∞ 范数下的解。目的是为在数据分析中具体进行反问题的求解做一些理论铺垫。

【关键词】正问题，褶积系统，反问题，可行解集，正则化，反演解性质，L_∞ / L_2 / L_1 / L_0 范数，Bayes 估计，最大似然估计

1 引言

自然和社会科学，主要解决两个问题。一是认识这个世界，即了解这个世界运行的内在规律，并建立描述其运行规律的某种数学物理模型；二是改变这个世界，即通过对世界的观测（数据），基于上述数学物理模型，构建一个反问题并求解，进而形成解决某类具体问题的方案、方法和技术。第二类问题是我们最关注、也是经常面对的。

这两类问题，在绝大部分情况下，总可以抽象为一个系统问题，包括系统的输入、系统本身和系统输出。建立起描述系统输出的数学物理关系，即褶积模型，属于正问题的范畴；由观测的系统输出，估计系统本身或系统参数以及估计系统输入属于反问题。

众所周知，正问题具有局部性、因果性；反问题没有局部性、因果性 。正问题是适定的、良态的；反问题是非适定的、病态的。正问题的解是唯一的；反问题的解是非唯一的。

观测数据属于正问题的值域是反问题解存在的最基本要求。至少部分观测数据属于临近正问题的值域更是反问题解存在的最基本要求。否者，反问题解完全不存在，努力求解这样的反问题只能是徒劳的。观测数据属于正问题的值域本质上还规定了反问题解的可行集。这个可行解集在很多实际问题中往往是巨大的。这说明了反演解具有不唯一性。如果落在可行解集中的可行解有很大的方差，必须引入其它先验的约束信息来缩小可行解集，降低方差。某种程度上，这也说明了为什么反问题的解要满足无偏和方差最小的原则。无偏和方差最小的解是精度最好的反问题解。

先验约束信息该如何引入？基本的方法可以分为三类：1、预置反演解的范围（也可称预置反演解的置信区间）；2、预置反演解的期望（实际上是反演解的初始值）和方差；3、规定反演解要满足的特性。反演解的特性包括能量范数最小（对应引入 L_2 范数）；反演解满足稀疏特性（对应引入 L_1 或 L_0 范数）；反演解的方向导数满足能量范数最小（对应 TV 正则化方法）。

引入正则化约束后，反演解的精度如何？反演解的精度该如何评价？为什么这样的反演解是反问题所期望的？这些问题在理论和应用层次上都是必须要回答的。

本文对比分析 Bayes 估计理论的解、最大似然估计理论下的解、L_2 范数下的解、L_1 范数下的解和 L_0 范数下的解，给出反演解性质的对比分析。并简单评价 L_∞ 范数下的解。目的就是厘清反问题解的性质，为反问题的实际应用奠定理论基础。

2 Bayes 参数估计理论与最大似然参数估计理论下反问题解的性质分析

对于已知系统输出的观测值，构建反问题，估计系统参数的问题，我们往往把系统输出的观测值视为随机变量（随机向量），并在 Bayes 估计理论下，讨论反问题、评价其解的性质。

Bayes 估计的思想可以简单概括为：估计的结果 $\tilde{\mathbf{m}}$ 是由后验概率密度函数 $P(\mathbf{m}/\mathbf{y})$ 所决定的均值 $E\{\mathbf{m}/\mathbf{y}\}$，其中 $\mathbf{y}$ 代表观测数据。已知后验概率密度函数 $P(\mathbf{m}/\mathbf{y})$ 后，估计结果写为，

$$\tilde{\mathbf{m}} = E\{\mathbf{m}/\mathbf{y}\} = \int \mathbf{m} P(\mathbf{m}/\mathbf{y}) d\mathbf{m} \tag{1}$$

式（1）表示参数估计的结果是由观测数据所对应的模型参数出现的概率所决定的。

由于后验概率密度 $P(\mathbf{m}/\mathbf{y})$ 很难估计，所以将其转化为较容易估计的模型空间的先验概率密度函数 $P(\mathbf{m})$、数据空间的概率密度函数 $P(\mathbf{y})$ 和条件概率密度函数 $P(\mathbf{y}/\mathbf{m})$ 来表示，其中 $P(\mathbf{y}/\mathbf{m})$ 体现了正算子的不确定性，正算子描述了从模型空间到数据空间的映射关系，它体现了待估计参数与实测实际之间的关系并非完全正确的，预测结果是存在不确定性的。后验概率密度由 Bayes 公式描述为

$$P(\mathbf{m}/\mathbf{y}) = \frac{P(\mathbf{y}/\mathbf{m}) P(\mathbf{m})}{P(\mathbf{y})} \tag{2}$$

据 Bayes 公式可以看出，$P(\mathbf{m}/\mathbf{y})$ 由 $P(\mathbf{m})$、$P(\mathbf{y})$ 和 $P(\mathbf{y}/\mathbf{m})$ 来决定。假设 $P(\mathbf{y})$ 均匀分布，没有引入先验信息的情况下，参数估计问题的求解可以认为是在 $P(\mathbf{m})$ 和 $P(\mathbf{y}/\mathbf{m})$ 的约束下，得到估计结果由期望均值为 $\tilde{\mathbf{m}} = \int \mathbf{m} P(\mathbf{m}/\mathbf{y}) d\mathbf{m}$ 决定，方差 $\sigma_{\tilde{\mathbf{m}}} = \int (\mathbf{m} - \tilde{\mathbf{m}})^2 P(\mathbf{m}/\mathbf{y}) d\mathbf{m}$ 决定。

更可行的做法是假设 $P(\mathbf{m})$ 和 $P(\mathbf{y}/\mathbf{m})$ 均为高斯函数，后验概率密度函数也是 Gauss 分布。此时，后验概率密度函数最大化（MAP）对应了代价函数

$$S(\mathbf{m}) = \frac{1}{2}\left[(g(\mathbf{m}) - \mathbf{d}_{obs})^H C_D^{-1} (g(\mathbf{m}) - \mathbf{d}_{obs}) + (\mathbf{m} - \mathbf{m}_{prior})^H C_M^{-1} (\mathbf{m} - \mathbf{m}_{prior}) \right] \tag{3}$$

的最小化。式中：C_D 为数据协方差矩阵；C_M 为模型的协方差矩阵。代价函数最小时（3）式的解认为是参数估计的反演解。（3）式的具体求解是数值优化方法要解决的问题。Newton 类方法是最常用的。实质上，Newton 类方法仅仅在（3）式是凸泛函时是有效的算法。否则，就要引入所谓的 Monte Carlo 全局寻忧方法求解（3）式。（3）式对应的反演解的意义是什么？为什么它能作为反问题的解？

事实上，（3）式对应的反问题的提法可以写为，

$$\begin{aligned} &min\left\{ \left(\mathbf{m} - \mathbf{m}_{prior}\right)^H C_M^{-1} \left(\mathbf{m} - \mathbf{m}_{prior}\right) \right\} \\ &Subject\ to \quad \left(g(\mathbf{m}) - \mathbf{d}_{obs}\right)^H C_D^{-1} \left(g(\mathbf{m}) - \mathbf{d}_{obs}\right) \le \varepsilon \end{aligned} \tag{4}$$

这样的反问题提法清楚地表明：要寻找能量最低的参数值，同时满足对观测数据的逼近要求。这样提出反问题看起来比 Bayes 估计理论更为直接。但是，如此提出反问题，对反演解评价缺乏严谨的理论支撑。

还是要回到 Bayes 估计理论。当（3）式中的非线性正问题$\mathbf{d}=g(\mathbf{m})$可以退化为线性正问题$\mathbf{d}=G\mathbf{m}$时，Gauss 分布的后验概率密度函数可以写为，

$$P\left(\mathbf{m}/\mathbf{y}\right)\sim e^{-(\mathbf{m}-\tilde{\mathbf{m}})^T\tilde{C}_{\mathbf{m}}^{-1}(\mathbf{m}-\tilde{\mathbf{m}})} \tag{5}$$

其中，后验概率密度函数的期望为

$$\begin{aligned}\tilde{\mathbf{m}}&=\left(G^TC_D^{-1}G+C_M^{-1}\right)^{-1}\left(G^TC_D^{-1}\mathbf{d}_{obs}+C_M^{-1}\mathbf{m}_{prior}\right)\\&=\mathbf{m}_{prior}+\left(G^TC_D^{-1}G+C_M^{-1}\right)^{-1}G^TC_D^{-1}\left(\mathbf{d}_{obs}-G\mathbf{m}_{prior}\right)\\&=\mathbf{m}_{prior}+C_MG^T\left(GC_MG^T+C_D\right)^{-1}\left(\mathbf{d}_{obs}-G\mathbf{m}_{prior}\right)\end{aligned} \tag{6a}$$

后验概率密度函数的方差为

$$\tilde{C}_{\mathbf{m}}=\left(G^TC_D^{-1}G+C_M^{-1}\right)^{-1}=C_M-C_MG^T\left(GC_MG^T+C_D^{-1}\right)^{-1}GC_M \tag{6b}$$

由（5）和（6）式清楚地看出，Bayes 估计理论下反问题的解是 Gauss 分布的后验概率密度函数的期望值。此时，反演解的精度自然应该由 Gauss 分布的后验概率密度函数的方差决定。Bayes 估计理论的强大理论价值体现在：它不仅说明了反问题的含义，也完整地刻画了反演解的意义。同时也用后验概率密度函数的方差界定了反演解的精度。

进一步地，假设数据协方差阵C_D退化为$\sigma_D^2\mathbf{I}$；模型参数协方差阵C_M退化为$\sigma_M^2\mathbf{I}$；$\mathbf{m}_{prior}=0$，则有

$$S(\mathbf{m})=\frac{\left(G\mathbf{m}-\mathbf{d}_{obs}\right)^2}{\sigma_D^2}+\frac{\mathbf{m}^2}{\sigma_M^2}=\left(G\mathbf{m}-\mathbf{d}_{obs}\right)^2+\frac{\sigma_D^2}{\sigma_M^2}\mathbf{m}^2=\left(G\mathbf{m}-\mathbf{d}_{obs}\right)^2+\lambda\mathbf{m}^2 \tag{7a}$$

$$\begin{aligned}\tilde{\mathbf{m}}&=\left(G^TC_D^{-1}G+C_M^{-1}\right)^{-1}\left(G^TC_D^{-1}\mathbf{d}_{obs}+C_M^{-1}\mathbf{m}_{prior}\right)=\left(\frac{G^TG}{\sigma_D^2}+\frac{1}{\sigma_M^2}\right)^{-1}\frac{G^T\mathbf{d}_{obs}}{\sigma_D^2}\\&=\left(G^TG+\frac{\sigma_D^2}{\sigma_M^2}\right)^{-1}G^T\mathbf{d}_{obs}=\left(G^TG+\lambda\mathbf{I}\right)^{-1}G^T\mathbf{d}_{obs}\end{aligned} \tag{7b}$$

$$\tilde{C}_{\mathbf{m}}=\left(G^TG+\lambda\mathbf{I}\right)^{-1} \tag{7c}$$

其中，$\lambda=\dfrac{\sigma_D^2}{\sigma_M^2}$。（7a）式显然定义了 Tikhonov 正则化代价泛函；（7b）给出了反演解的形式；（7c）定义了 Tikhonov 正则化反演问题的解的精度。Bayes 估计理论清楚地揭示了 Tikhonov 正则化反演的所有内涵。

最大似然推断可以认为是 Bayes 推断的特例，即假设没有引入关于待估计参数的先验信息$P(\mathbf{m})$，把先验概率密度函数$P(\mathbf{y}|\mathbf{m})$视为似然函数。这是频率统计学派坚持的观点。似然函数的最大化作为统计推断的原则。此时，待估计结果不被视为是随机的。这是最大似然估计的重大问题。如果不把待估计结果视为随机的，而是确定性的，很难全面地评价待估计解的精度。

3 线性预测误差泛函的几何性能及光滑凸问题解的性质分析

当前，基于 Newton 迭代法的反问题求解方法仅仅适用于凸的反问题，严格地讲在局部点上是凸的反问题。较强的非线性正问题对应的反问题，目前缺乏有效的解法。即便 Monte Carlo 算法也不能很好地解决此

类问题。

因此，下面的讨论都是针对凸的反问题的。而凸的反问题严格地讲，只能对应线性正问题。下面的讨论试图从本质上说明 L_2 范数下凸的反问题的解的性质。

假设线性系统对应的矩阵方程为 $G\mathbf{m}=\mathbf{d}$，其中系数矩阵 G 是满秩的，最好是正定对称的。实质上，系数矩阵 G 也是一个线性投影算子。它把模型向量 $\mathbf{m}$ 投影成数据向量 $\mathbf{d}$。广义线性反演的另一个本质描述是寻找线性投影算子的逆算子或广义逆算子，对数据向量 $\mathbf{d}$ 进行反投影，估计模型向量 $\mathbf{m}$。实际情况下数据总是包含噪音的，此时 $G\mathbf{m}\approx\hat{\mathbf{d}}\left(\hat{\mathbf{d}}=\mathbf{d}+\boldsymbol{\eta}\right)$，$\boldsymbol{\eta}$ 代表实测数据中的噪音。在变分原理意义下，定义如下的误差能量，

$$
\begin{aligned}
&\mathcal{J}(\mathbf{m})=(G\mathbf{m}-\mathbf{d})^T(G\mathbf{m}-\mathbf{d})\\
&=\mathbf{d}^T\mathbf{d}-(G\mathbf{m})^T\mathbf{d}-\mathbf{d}^T(G\mathbf{m})+(G\mathbf{m})^T(G\mathbf{m})\\
&=\mathbf{d}^T\mathbf{d}-\mathbf{m}^TG^T\mathbf{d}-\left(G^T\mathbf{d}\right)^T\mathbf{m}+\mathbf{m}^TG^TG\mathbf{m}\\
&=\mathbf{d}^T\mathbf{d}-\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)+\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)-\mathbf{m}^TG^T\mathbf{d}-\left(G^T\mathbf{d}\right)^T\mathbf{m}+\mathbf{m}^TG^TG\mathbf{m}\\
&=\mathbf{d}^T\mathbf{d}-\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)+\left[\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}-\mathbf{m}^T\right]\left(G^T\mathbf{d}\right)-\left[\left(G^T\mathbf{d}\right)^T-\mathbf{m}^TG^TG\right]\mathbf{m}\\
&=\mathbf{d}^T\mathbf{d}-\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)+\left[\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}-\mathbf{m}^T\right]\left(G^T\mathbf{d}\right)-\left[\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}-\mathbf{m}^T\right]\left(G^TG\right)\mathbf{m}\\
&=\mathbf{d}^T\mathbf{d}-\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)+\left[\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}-\mathbf{m}^T\right]\left[\left(G^T\mathbf{d}\right)-\left(G^TG\right)\mathbf{m}\right]\\
&=\mathbf{d}^T\mathbf{d}-\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)+\left[\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}-\mathbf{m}^T\right]G^TG\left[\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)-\mathbf{m}\right]\\
&=\mathbf{d}^T\mathbf{d}-\left(G^T\mathbf{b}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)+\left[\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)-\mathbf{m}\right]^TG^TG\left[\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)-\mathbf{m}\right]\\
&=\mathbf{d}^T\mathbf{d}-\left(G^T\mathbf{b}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)+(\mathbf{m}_o-\mathbf{m})^TG^TG(\mathbf{m}_o-\mathbf{m})\qquad \mathbf{m}_o=\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)\\
&=\mathbf{d}^T\mathbf{d}-\hat{\mathbf{d}}^T\hat{\mathbf{d}}+(\mathbf{m}_o-\mathbf{m})^TG^TG(\mathbf{m}_o-\mathbf{m})\\
&=\mathcal{J}_{min}(\mathbf{m}_o)+(\mathbf{m}_o-\mathbf{m})^TG^TG(\mathbf{m}_o-\mathbf{m})
\end{aligned}
\tag{8}
$$

其中，$\mathcal{J}_{min}(\mathbf{m}_o)$ 代表达到最佳解 $\mathbf{m}_o$ 时的误差能量，$\mathcal{J}_{min}(\mathbf{m}_o)=\|\boldsymbol{\eta}\|^2$。为了公式推导的完整性，把如下公式列出，

$$
\begin{aligned}
&\hat{\mathbf{d}}=G\mathbf{m}_o\\
&\hat{\mathbf{d}}^T\hat{\mathbf{d}}=\left(G\mathbf{m}_o\right)^T\left(G\mathbf{m}_o\right)=\mathbf{m}_o^TG^TG\mathbf{m}_o\\
&=\left(G^TG\mathbf{m}_o\right)^T\mathbf{m}_o=\left(G^T\mathbf{d}\right)^T\mathbf{m}_o\\
&=\left(G^T\mathbf{d}\right)^T\left(G^TG\right)^{-1}\left(G^T\mathbf{d}\right)
\end{aligned}
\tag{9}
$$

由矩阵论知，G^TG 具有对称正定性，存在 $G^TG=Q^T\Sigma Q$ 的正交分解，正交矩阵 Q 的列向量是特征向量，对角矩阵 Σ 中元素是与正交阵 Q 的列向量对应的特征值。因此，（8）式可以继续改写为，

$$
\begin{aligned}
&\mathcal{J}(\mathbf{m})=\mathcal{J}_{min}(\mathbf{m}_o)+(\mathbf{m}_o-\mathbf{m})^T G^T G(\mathbf{m}_o-\mathbf{m})\\
&=\mathcal{J}_{min}(\mathbf{m}_o)+(\mathbf{m}_o-\mathbf{m})^T Q^T \Sigma Q(\mathbf{m}_o-\mathbf{m})\\
&=\mathcal{J}_{min}(\mathbf{m}_o)+\left(Q(\mathbf{m}_o-\mathbf{m})\right)^{\mathrm{T}} \Sigma Q(\mathbf{m}_o-\mathbf{m})\\
&=\mathcal{J}_{min}(\mathbf{m}_o)+\mathbf{q}^{\mathrm{T}}\Gamma\mathbf{q} \qquad \mathbf{q}=Q(\mathbf{m}_o-\mathbf{m})\\
&=\mathcal{J}_{min}(\mathbf{m}_o)+\sum_{k=1}^{M}\lambda_k|q_k|^2\\
&=\mathcal{J}_{min}(\mathbf{m}_o)+\sum_{k=1}^{M}\frac{|q_k|^2}{1/\lambda_k}
\end{aligned}
\tag{10}
$$

其中，$\mathbf{q}=Q(\mathbf{m}_o-\mathbf{m})$代表解的差向量$\mathbf{m}_o-\mathbf{m}$在正交坐标系$Q$中的投影。（10）式的几何形态可以抽象地用下图表示。

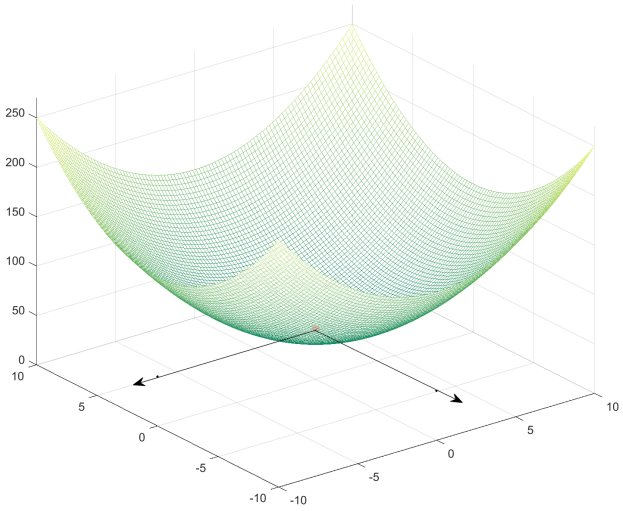


图 1 线性预测误差泛函的几何性能表面

图 1 清楚地展示出了广义线性反问题的本质。对于广义线性反演（线性正问题+Gauss 预测误差），线性预测误差泛函的几何形态是一个（超）椭球体。即关于线性反问题是一个凸问题。当系数矩阵G是满秩的，广义线性反问题存在唯一的解$\mathbf{m}_o$。此时预测误差能量达到最低$\mathcal{J}_{min}(\mathbf{m}_o)=\|\boldsymbol{\eta}\|^2$，达到实测数据的噪音能量水平。这也是偏差原则中用噪音能量水平作为正则化超参的理论基础。线性预测误差泛函的几何形态与系数矩阵G的条件数密切相关，系数矩阵G的条件数越大，（超）椭球体越扁长。迭代法求解广义线性反问题时收敛越慢，解的精度越低。

另外，从图 1 中还可以看出，实际数据分析中，我们并不追求误差泛函达到$\mathcal{J}_{min}(\mathbf{m}_o)=\|\boldsymbol{\eta}\|^2$时的反演解。这里面蕴含了$G\mathbf{m}\approx\mathbf{d}(\hat{\mathbf{d}}+\boldsymbol{\eta})$，$\boldsymbol{\eta}$是 Gauss 白噪这个假设往往是不成立的。$\hat{\mathbf{d}}$是实际数据中包含的信号，$G\mathbf{m}$并不能形成对$\hat{\mathbf{d}}$的完美的预测。实际上，我们常做的是：要么修改$G\mathbf{m}$提升对$\hat{\mathbf{d}}$的估计精度；要么对$\mathbf{d}$进行预处理使得$G\mathbf{m}$尽可能精确地预测其中的信号；还可以引入正则化约束，不追求$\mathcal{J}_{min}(\mathbf{m}_o)=\|\boldsymbol{\eta}\|^2$，而是在信号最佳逼近与正则化约束之间达到平衡。据此可以理解反演解是根据先验约束在可行解集中挑选一个能解决实际问题解的逻辑基础。

假若系数矩阵 G 不是满秩的，广义线性反问题不存在唯一的解。此时需要引入关于解 $\mathbf{m}$ 的先验信息。此时，最小二乘意义下的线性预测误差泛函被定义为，

$$\mathcal{J}\left(\mathbf{m}\right)=\left(G\mathbf{m}-\mathbf{d}\right)^T\left(G\mathbf{m}-\mathbf{d}\right)+\lambda\left\|\mathbf{m}\right\|^2 \tag{11}$$

（11）式定义的最小二乘意义下的广义线性反问题的解的含义为：

$$\tilde{\mathbf{m}}=min\left\|\bar{\mathbf{m}}\right\|^2 \quad S.T.\bar{\mathbf{m}}=min\,\mathcal{J}\left(\mathbf{m}\right)=min\left\{\left(G\mathbf{m}-\mathbf{d}\right)^T\left(G\mathbf{m}-\mathbf{d}\right)\right\} \tag{12}$$

（12）式的含义是：从无穷多的可行解 $\bar{\mathbf{m}}$ 中选一个能量最低的解，作为最小二乘意义下的广义线性反问题的解。要求解 $\hat{\mathbf{m}}$ 是能量最低的就是引入的先验信息。具有能量最低属性的解就是光滑的解。据此，解决了解的唯一性问题。

若正问题线性 $G\mathbf{m}=\mathbf{d}$ 且系数矩阵 G 是满秩的、预测误差 Gauss 白噪情形下，反问题 $\mathcal{J}\left(\mathbf{m}\right)=\left(G\mathbf{m}-\mathbf{d}\right)^T\left(G\mathbf{m}-\mathbf{d}\right)$ 的解是无偏和方差最小的。

数学分析学家从滤波 SVD 和截断 SVD 角度讨论 Tikhonov 正则化反问题的解，但没有从理论上严谨地说明 Tikhonov 正则化反问题的解的意义与精度评价。

这就从不同角度和不同层次说清楚了 L_2 范数下凸的反问题的解的性质。

4 L1 范数下 LASSO 问题解的性质分析

我一直认为：挑选合理的反演解是哲学层次上的议题，因为这不是可以按照逻辑推演出来的。反问题首先应该追求无偏与方差最小的反演解。按照 Bayes 估计和最大似然估计理论，这要求后验概率密度函数和似然函数应该是 Gauss 型的，而且该函数一定是凸的，有明确的无偏和方差最小的定义。此时，存在一套数学上看来十分优美的解法（共轭梯度法或 Newton 迭代法）。解的含义也是十分清楚的。但是，此时反问题解是光滑的、分辨率低的。

自然偏好简洁。反演解也应该体现出稀疏性，至少在变换域中体现出稀疏性，即变换域中变换系数存在快速衰减的性质。稀疏解往往是更合理的解。在已知反演解的可行解集的条件下，施加解的（原域或变换域中）稀疏性的约束，往往能得到精度更高的反演解。此时，反问题的提法显然应该是：

$$\begin{aligned} &min \quad \left\|\mathbf{m}\right\|_1 \\ &Subject\ to \quad \left(G\mathbf{m}-\mathbf{d}_{obs}\right)^H C_D^{-1}\left(G\mathbf{m}-\mathbf{d}_{obs}\right)\le\varepsilon \end{aligned} \tag{13}$$

与（13）式对应的反问题提法还有，

$$\tilde{\mathbf{m}}=argmin\left\{\left(G\mathbf{m}-\mathbf{d}_{obs}\right)^H C_D^{-1}\left(G\mathbf{m}-\mathbf{d}_{obs}\right)+\lambda\left\|\mathbf{m}\right\|_1\right\} \tag{14}$$

和

$$\begin{aligned} &min \quad \left(G\mathbf{m}-\mathbf{d}_{obs}\right)^H C_D^{-1}\left(G\mathbf{m}-\mathbf{d}_{obs}\right) \\ &Subject\ to \quad \left\|\mathbf{m}\right\|_1\le\delta \end{aligned} \tag{15}$$

（13）被称为 LASSO 问题；（15）被称为基追踪（Basis pursuit）问题。解（13）式的算法也称为 LASSO（Least Absolute Shrinkage and Selection Operator）算法。某种程度上，（13）、（14）和（15）定义的反问题是等价的。就是说，在最好的情况下，三个问题的解是相同的或非常接近的。（11）式定义的 Tikhonov 正则化问题也被称为 Ridge Regression（岭回归）问题。众所周知，岭回归问题的解为，

$$\tilde{\mathbf{m}}_{RR}=\left(G^TG+\lambda\mathbf{I}\right)^{-1}G^T\mathbf{d}_{obs} \tag{16}$$

LASSO 问题是个带线性不等式约束的凸优化问题，原则上可以用任何合适的方法进行求解。但是，目前遇到的参数估计问题的规模往往是巨大的，必须提出针对性的高效的计算方法。为此，首先重写（14）式为，

$$\begin{aligned} &\underset{(m_i,u_i)_{i=1}^{L}}{min} \quad \left(G\mathbf{m}-\mathbf{d}_{obs}\right)^{H} C_D^{-1}\left(G\mathbf{m}-\mathbf{d}_{obs}\right)+\lambda\sum_{i=1}^{L}u_i \\ &s.t. \quad \begin{cases} -u_i \le m_i \le u_i \\ u_i \ge 0 \end{cases} \quad i=1,2,\cdots,L \end{aligned} \tag{17}$$

然后，假定系数矩阵 G 为单位正交阵，即 $G^TG=\mathbf{I}$，则（11）式对应的最小二乘解为，

$$\tilde{\mathbf{m}}_{LS}=\left(G^TG\right)^{-1}G^T\mathbf{d}_{obs}=G^T\mathbf{d}_{obs} \tag{18}$$

则（16）是对应的岭回归问题的解修改为，

$$\tilde{\mathbf{m}}_{RR}=\frac{1}{1+\lambda}\tilde{\mathbf{m}}_{LS}=\frac{1}{1+\lambda}G^T\mathbf{d}_{obs} \tag{19}$$

可见，岭回归问题的解是由最小二乘解收缩来的。收缩系数为 $1/(1+\lambda)$。

对于 Lagrange 乘子法定义的 LASSO 问题对应的误差泛函（（14）式），它是不存在泛函梯度的，因为它在某些点上是不可导的。但是，通过定义次梯度可以求其极值点对应的反问题解。对于（14）式，次梯度等于零时的表达式为，

$$\mathbf{0}\in -2G^T\mathbf{d}+2G^TG\mathbf{m}+\lambda\partial\left\|\mathbf{m}\right\|_1 \tag{20}$$

其中，∂ 代表次微分。次微分是一个集合。

若 G 是单位正交阵，（20）式可写出分量形式，

$$0\in -2\tilde{m}_{LS,i}+2\tilde{m}_{1,i}+\frac{\lambda}{2}\partial\left|\tilde{m}_{1,i}\right| \quad i\in(1,2,\cdots,L) \tag{21}$$

上式中，次微分为，

$$\partial|m|=\begin{cases} \{1\}, & 若 m>0; \\ \{-1\}, & 若 m<0; \\ [-1,1], & 若 m=0。 \end{cases} \tag{22}$$

因此，LASSO 问题的解为，

$$\tilde{m}_{1,i}=\begin{cases} \tilde{m}_{1,i}-\dfrac{\lambda}{2}, & 若 \tilde{m}_{1,i}>0; \\ \tilde{m}_{1,i}+\dfrac{\lambda}{2}, & 若 \tilde{m}_{1,i}<0; \\ 0, & 若 -\dfrac{\lambda}{2}\le\tilde{m}_{1,i}\le\dfrac{\lambda}{2}。 \end{cases} \tag{23}$$

（23）式定义的 LASSO 问题的解可以图示为，

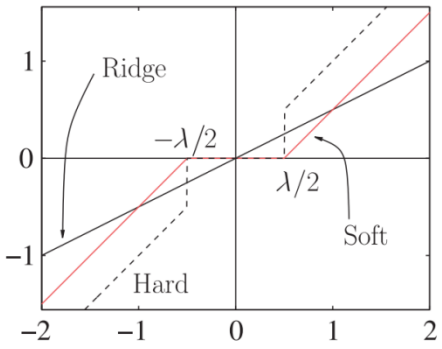


图 2 LASSO 问题的解的几何含义

上述简洁的推导过程清晰地说明了 L_1 范数约束下定义的 LASSO 问题的解的含义，它是由最小二乘问题 $\mathcal{J}(\mathbf{m})=(G\mathbf{m}-\mathbf{d})^T(G\mathbf{m}-\mathbf{d})$ 的解经过绝对值收缩后形成的。收缩幅度由超参 λ 决定。对比一下，Tikhonov 正则化解也是由最小二乘问题 $\mathcal{J}(\mathbf{m})=(G\mathbf{m}-\mathbf{d})^T(G\mathbf{m}-\mathbf{d})$ 的解经过收缩形成的，但它是由 $1/(1+\lambda)$ 一致收缩形成的。但 LASSO 问题的解是按分量的绝对值收缩形成的。直观地可以看出，凡是 $-\frac{\lambda}{2}\le\tilde{m}_{1,i}\le\frac{\lambda}{2}$ 的解分量一律收缩成零值。这正是 L_1 范数约束下定义的 LASSO 问题的解获得稀疏性的根源，即仅仅 $-\frac{\lambda}{2}\le\tilde{m}_{1,i}\le\frac{\lambda}{2}$ 之外大的分量被保留。

从上述分析过程可以看出，L_1 范数约束下定义的 LASSO 问题的解的确具备稀疏性质，在很多应用场景下确实具有比较高的精度。但是，为什么这样的解会更好，上述分析并没有能给出解释。事实上，很多应用场景下，稀疏解并非是更好的解。

任何反问题的解都应该是在正问题规定的可行解集中，通过引入某种正则化条件，挑选合理的解。这应该是最基本的认识。目前，并没有严谨的数学理论和方法给出公认的、好的反演解。实际数据分析中，只能是在充分理解反问题理论和方法的基础上，根据对要解决问题的深入认识，挑选合理的、有助于解决实际问题的反演解。

5 L0 范数下贪婪算法解的性质分析

我们知道，对待反演结果引入不同的范数，相当于对反演解施加了不同性质的约束。存在一般的 p 范数定义，

$$\|\mathbf{m}\|_p=\left(\sum_{i=1}^{L}|m_i|^p\right)^{1/p} \tag{24}$$

很显然，反演解的 L_1 范数为

$$\|\mathbf{m}\|_1=\sum_{i=1}^{L}|m_i| \tag{25a}$$

反演解的 L_∞ 范数为

$$\|\mathbf{m}\|_{\infty}=\lim_{p\to\infty}\left(|m_{i_{max}}|^{p}\sum_{i=1}^{L}\left(\frac{|m_i|}{|m_{i_{max}}|}\right)^{p}\right)^{1/p}=|m_{i_{max}}| \tag{25b}$$

反演解的 L_0 范数为

$$\|\mathbf{m}\|_{0}=\lim_{p\to 0}\sum_{i=1}^{L}|m_i|^{p}=\sum_{i=1}^{L}\chi_A(|m_i|),\quad \chi_A(\tau)=\begin{cases}1, & 若\tau\in A\\ 0, & 若\tau\notin A\end{cases} \tag{25c}$$

就是说，反演解的 L_∞ 范数是最大分量的绝对值；反演解的 L_0 范数是非零分量的计数结果。

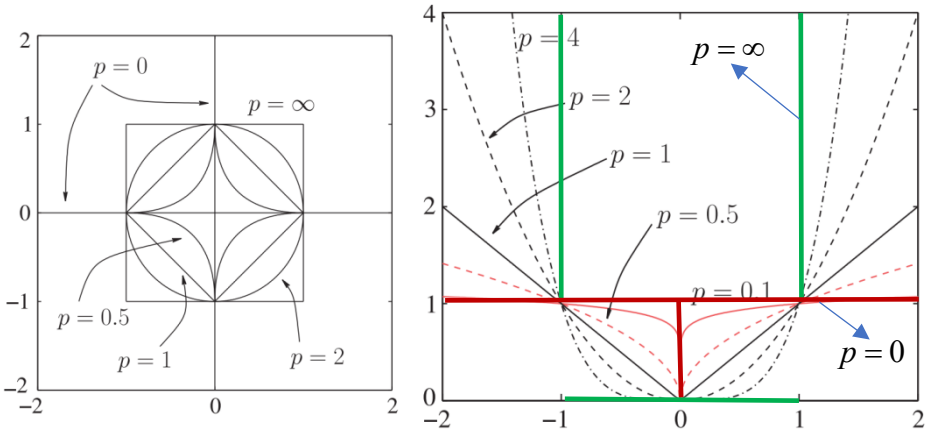


图 3 反演解不同范数的几何关系

图 3 展示了反演解不同范数的几何关系。从图中可以看出，对反演解引入不同的范数，在反演算法中相当于对反演解的不同分量施加了不同的惩罚。反演解不同分量的取值大小不同，在不同的范数下，受到的惩罚是不同的，从而使反演解表现出不同的特性。

对于 L_1 范数，与每个分量大小成等比例地受到惩罚，小的分量因此被削减（到零），从而使得反演解体现出稀疏性。对于 L_2 范数，大于 1 的分量受到的惩罚大，且分量值越大，惩罚越重，因此反演解中的大分量值被削弱，小于 1 的分量值相对地较少受惩罚，能得以保留，因此反演解中包含有大值和小值的分量，没有稀疏性，显得比较光滑。相比于 L_1 范数，L_0 范数对所有分量（零值除外）采用等权的惩罚，对于值小于 1 的分量，更为彻底地被削弱。因此，L_0 范数下的反演解，相比于 L_1 范数，表现出更强的稀疏性。对于 L_∞ 范数范数，凡是大于 1 的分量全部被惩罚掉，仅保留值小于 1 的分量。很显然，这样的范数不适于（勘探地震学中的）参数估计反问题中。

由图 3 可知，$0\le p<1$ 情形下的范数并不是凸的。$p=1$ 情形下的范数不是严格凸的。

我们还知道，对待反演结果引入不同的范数，相当于在 Bayes 估计理论中引入不同的关于待反演结果满足的、先验的概率密度函数。注意在 Bayes 估计理论下，我们总是把待估计结果视为满足一定概率密度函数的随机变量（随机向量）。$p=2$ 时的 L_2 范数对应于引入了待估计参数满足 Guass 分布的先验假设。$p=1$ 时的 L_1 范数对应于引入了待估计参数满足 Laplace 分布的先验假设。$p\to\infty$ 时的 L_∞ 范数对应于引入了待估计参数满足均匀分布的先验假设。但是，从先验概率密度函数角度解释反演解的稀疏性是比较困难的。

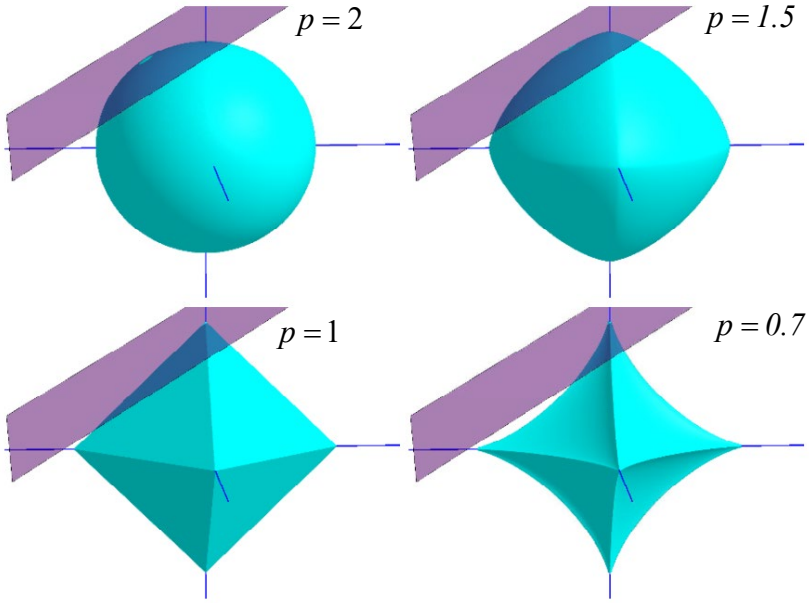


图 4 不同范数定义的“球”与超平面 $G\mathbf{m}=\mathbf{d}$ 的相切关系

图 4 从另一个角度展示了不同范数的约束条件为什么导致不同形式的反演解。理论上知道，超平面 $G\mathbf{m}=\mathbf{d}$ 与不同范数定义的“球”相切时切点的坐标定义了反演解。从图 4 可知，L_2 范数定义的球与超平面 $G\mathbf{m}=\mathbf{d}$ 相切时，切点的坐标分量可能都不为零。即解不存在稀疏性。L_1 范数定义的“球”与超平面 $G\mathbf{m}=\mathbf{d}$ 相切时，切点的坐标分量一定存在很多零。但是，可能存在一种极端情况，即超平面 $G\mathbf{m}=\mathbf{d}$ 与菱形的棱线或棱面平行，此时存在无穷多解。应该注意到，在高维情形下这是极难出现的情况。L_0 范数定义的“球”与超平面 $G\mathbf{m}=\mathbf{d}$ 相切时，切点的坐标分量一定存在很多零，且不存在 L_1 范数情形下的多解性问题。对着三种典型的情形进行综合评价，有如下的观点：L_1 范数约束下定义的 LASSO 问题的解的确具备稀疏性质，存在快速算法，是当前稀疏反演、稀疏表达、特征提取等等领域最常用的。

下面专门讨论 L_0 范数约束下定义的反问题（26）解的性质。

$$\begin{aligned} &min \quad \|\mathbf{m}\|_0 \\ &Subject\ to \quad \left(G\mathbf{m}-\mathbf{d}_{obs}\right)^H C_D^{-1}\left(G\mathbf{m}-\mathbf{d}_{obs}\right)\le\varepsilon \end{aligned} \tag{26}$$

前已述及，解任何一个（线性）反问题，首先要确定可行解集，然后在可行解集中选一个满足约束条件的解。可行解集显然是由正问题 $G\mathbf{m}=\mathbf{d}$ 决定的。

在 L_2 范数情形下，通过系数矩阵 G 的 SVD 分解决定了可行解集。把参数向量投影在左奇异向量构成的空间中；数据向量投影在右奇异向量构成的空间中得到的可行解集，在 L_2 范数下得到的反演解是不具备稀疏性的。主要原因就是左奇异向量和右奇异向量是正交坐标系形成的空间。好处是这样的解法可以有效地求解大规模参数估计问题。LASSO 问题的求解方法也是对上述最小二乘解进行按参数分量的绝对值大小的收缩产生的，当然也可以高效地求解大规模参数估计问题。

（26）式的求解根本思想是寻找最稀疏的一组反演解，投影在对应的、变换后的稀疏矩阵中，使得 $\left(G_S\mathbf{m}_S-\mathbf{d}_{obs}\right)^H C_D^{-1}\left(G\mathbf{m}-\mathbf{d}_{obs}\right)\le\varepsilon$，式中下标 S 是稀疏的意思。很显然，寻找到最好的 G_S 和 $\mathbf{m}_S$ 的过程对应一个枚举过程（这是 N-P Hard 问题）。对大规模参数估计问题，这完全是不可能的。

问题是如何能找到一种可行的解法！到目前为止，这样的典型算法被称为贪婪算法或正交匹配追踪算法。基本思想是把数据向量正交投影到系数矩阵的列向量空间中，寻找投影误差能量最小的分量。这体现了正

交匹配的含义。即，

$$\begin{aligned}\epsilon(j)&=\min_{m_j}\left\|\mathbf{d}-\mathbf{g}_j m_j\right\|^2=\left\|\mathbf{d}-\mathbf{g}_j\frac{\mathbf{g}_j^T\mathbf{d}}{\mathbf{g}_j^T\mathbf{g}_j}\right\|^2=\left(\mathbf{d}-\mathbf{g}_j\frac{\mathbf{g}_j^T\mathbf{d}}{\mathbf{g}_j^T\mathbf{g}_j}\right)^T\left(\mathbf{d}-\mathbf{g}_j\frac{\mathbf{g}_j^T\mathbf{d}}{\mathbf{g}_j^T\mathbf{g}_j}\right)\\&=\mathbf{d}^T\mathbf{d}-\mathbf{d}^T\mathbf{g}_j\left(\frac{\mathbf{g}_j^T\mathbf{d}}{\mathbf{g}_j^T\mathbf{g}_j}\right)-\mathbf{g}_j^T\mathbf{d}\left(\frac{\mathbf{g}_j^T\mathbf{d}}{\mathbf{g}_j^T\mathbf{g}_j}\right)+\mathbf{g}_j^T\mathbf{g}_j\left(\frac{\mathbf{g}_j^T\mathbf{d}}{\mathbf{g}_j^T\mathbf{g}_j}\right)^2\\&=\mathbf{d}^T\mathbf{d}-2\frac{\left(\mathbf{g}_j^T\mathbf{d}\right)^2}{\mathbf{g}_j^T\mathbf{g}_j}+\frac{\left(\mathbf{g}_j^T\mathbf{d}\right)^2}{\mathbf{g}_j^T\mathbf{g}_j}\\&=\mathbf{d}^T\mathbf{d}-2\frac{\left(\mathbf{g}_j^T\mathbf{d}\right)^2}{\mathbf{g}_j^T\mathbf{g}_j}\end{aligned}\tag{27}$$

基于（27）式，可以找到反演解的每一个分量。贪婪的含义是对（27）式进行修改，每一次都对剔除上次已找到的分量 m_j 及对应 $\mathbf{g}_j$ 在数据中贡献后的残差数据及余下的系数矩阵的列向量空间中施加（27）式寻找对应分量 $m_{j'}$ 和对应的 $\mathbf{g}_{j'}$，即

$$\begin{aligned}\epsilon(j')&=\min_{m_{j'}}\left\|\mathbf{r}^{k-1}-\mathbf{g}_{j'}m_{j'}\right\|^2=\left\|\mathbf{r}^{k-1}-\mathbf{g}_{j'}\frac{\mathbf{g}_{j'}^T\mathbf{r}^{k-1}}{\mathbf{g}_{j'}^T\mathbf{g}_{j'}}\right\|^2\\&=\left(\mathbf{r}^{k-1}-\mathbf{g}_{j'}\frac{\mathbf{g}_{j'}^T\mathbf{r}^{k-1}}{\mathbf{g}_{j'}^T\mathbf{g}_{j'}}\right)^T\left(\mathbf{r}^{k-1}-\mathbf{g}_{j'}\frac{\mathbf{g}_{j'}^T\mathbf{r}^{k-1}}{\mathbf{g}_{j'}^T\mathbf{g}_{j'}}\right)\\&=\left\|\mathbf{r}^{k-1}\right\|^2-\frac{\left(\mathbf{g}_{j'}^T\mathbf{r}^{k-1}\right)^2}{\left\|\mathbf{g}_{j'}\right\|^2}\end{aligned}\tag{28}$$

基于上述思想的贪婪算法或正交匹配追踪算法可以描述为，

迭代初次 k 的初始化 $k=0$；初始值赋值：$\mathbf{m}^0=\mathbf{0}$；$\mathbf{r}^0=\mathbf{d}-G\mathbf{m}^0=\mathbf{d}$。

初始化解的支撑集 $\mathcal{S}^0=\varnothing$。

主循环：$k=1,2,\cdots$

对基函数空间 $\{\mathbf{g}_1,\mathbf{g}_2,\cdots,\mathbf{g}_N\}$，扫描投影误差最小的解分量：$\epsilon(j)=\min_{m_j}\left\|\mathbf{r}^{k-1}-\mathbf{g}_j m_j\right\|^2$，得最佳解分量 $m_{j^o}=\dfrac{\mathbf{g}_{j^o}^T\mathbf{r}^{k-1}}{\mathbf{g}_{j^o}^T\mathbf{g}_{j^o}}$。

更新解的支撑集 $\mathcal{S}^k=\mathcal{S}^{k-1}\bigcup\{j^o\}$。

更新解 $\mathbf{m}^k$；更新残差 $\mathbf{r}^k=\mathbf{d}-G\mathbf{m}^k$。

若 $\left\|\mathbf{r}^k\right\|^2\le\varepsilon$，停止迭代；否则 $k+1$，进入下一次主循环。注意，总的循环次数必须小于参数向量 $\{m_1,m_2,\cdots,m_M\}$ 的个数 M。

贪婪算法（正交匹配追踪算法）提供了 L_0 范数约束下反问题的一个比较有效的解法。但是，解的精度很难评价。实际数据分析中反演解的意义不是很清楚。

6 L_∞ 范数在地震数据分析中可能应用的观点

前已述及，L_∞ 范数约束不适于用在大规模参数反演问题中，因为它只能给出要么是零要么是非零的反演解。也许在机器学习领域有其特殊的应用(Fuchs，2011；Studer 等，2012)。

但是，数据偏差项可以用 L_∞ 范数来定义。由于 L_∞ 范数由最大误差分量绝对值来定义，L_∞ 范数下的反问题又可成为最大（误差）最小化的反演问题。

到目前为止，还没有针对大规模参数估计问题的 L_∞ 范数反问题的有效解法。若把 L_∞ 范数反问题用于函数逼近情形，存在最佳一致逼近或 Chebysheve 逼近方法，因为这种方法利用了 Chebysheve 正交多项式基函数具有的如下性质，

$$max\left|f\left(\xi_i\right)-p\left(\xi_i\right)\right|,i=0,1,2,\cdots,n+1 \tag{29}$$

当且仅当 $f\left(\xi_{i+1}\right)-p^*\left(\xi_{i+1}\right)=-\left[f\left(\xi_i\right)-p^*\left(\xi_i\right)\right],i=0,1,2,\cdots,n$ 。

其中，$f\left(\mathbf{x}\right)$ 是被逼近任意函数；$p\left(\mathbf{x}\right)$ 是逼近函数。$p^*\left(\mathbf{x}\right)$ 代表最佳逼近函数。$\left(0,1,2,\cdots,n+1\right)$ 是交错点组。

Chebysheve 逼近中，$f\left(\mathbf{x}\right)$ 对应 Chebysheve 正交多项式基函数；$p\left(\mathbf{x}\right)=0$ 。此时，Chebysheve 正交多项式基函数的根决定一组交错点组。可以看出，Chebysheve 逼近方法的适用性很差，仅能用在对极为特殊的函数的逼近中（Soubaras，1996；Kosloff 等，2010）。

（29）式是普适性的最佳一致逼近的定义，但是如何找到满足定义要求的交错点组非常不易。只能针对特殊问题特殊对待。因此，最佳一致逼近只能用在某些函数逼近中。勘探地震中的应用也是如此。

7 结论性认识

自然和社会科学中非常多的问题都可以归结为对客观过程的数学模型描述（这对应于正问题）和基于实测数据及该数学模型而提出的反问题。可以说，正问题+反问题是解决非常多的实际问题的基本模式。但是，已知实际数据，提出反问题和求解反问题，包括对反问题解的评价都是有哲学意味的问题。反问题基于什么原则提出？为什么这些原则是有效的？反问题是多解的，本质上解反问题就是在正问题规定的可行解集中寻找合理的、或者说能在某种程度上解决实际问题的解。还有一个非常重要的环节是对反演解做出精度评价。

到目前为止，我们认为可行解集是靠正问题是否能解释实测数据来界定的，若与一个反演解对应的正问题能较好地预测或解释或逼近实测数据，就认为这个解处于可行解集中。这样的解会有非常多。然后，引入先验信息缩小可行解集，然后从缩小后的可行解集中作为我们寻找的解。更通常的做法是对反演解引入某种范数定义的约束，主要是 L_2 范数、L_1 范数和 L_0 范数。本文对三种范数下的反演解的性质给出了比较深入的分析。但是，到目前为止，还是仅有 Gauss-Markov 定理给出的反演解有明确的、或者说数学层次上的意义。就是说在线性正问题、对实测数据的逼近误差为高斯白噪、线性问题的系数矩阵是满秩的情形下，反问题的解是无偏的和方差最小的。L_1 范数下 LASSO 问题的解、L_0 范数下贪婪算法给出的解都没有对应的精度评价，不知道反演解的误差下限是什么。当前，很多机器学习算法也是这种状况。据此可以认识到反问题的解的性质的界定或者说反演解的误差限的界定是异常困难的事情。但是，这是必须要弄明白的，否者在实际应用中对反问题的求解总是缺乏足够的信心。同时，对反问题解的性质的研究可以不断地改进目前的算法。

参考文献

[1] M.J.D.Powell,1981,Approximation Theory and Methods;Cambridge University Press

[2] Michael Elad,2010,Sparse and Redundant Representations–From theory to Application in Signal and Image Processing;Springer

[3] Sergios Theodoridis,2015,Machine Learning–A Bayesian and Optimization Perspective;Elsevier

[4] Dan Kosloff,Reynam C.Pestana,and Hillel Tal-Ezer,2010,Acoustic and elastic numerical wave simulations by recursive spatial derivative operators；GEOPHYSICS 75:T167-T174.

[5] Soubaras，1996，Explicit 3-D migration using equiripple polynomial expansion and Laplacian synthesis,GEOPHYSICS 61:1386-1393.

[6] Studer,C.,Yin,W.,and Baraniuk,R.G.2012,Signal representations with minimum$\ell\infty$-norm.In 2012 50th Annual Allerton Conference on Communication,Control,and Computing(Allerton)(pp.1270-1277).IEEE.

[7] Fuchs,J.J.2011.Spread representations.Conference Record of the Forty Fifth Asilomar Conference on Signals,Systems and Computers(ASILOMAR)(pp.814-817).IEEE.

地震数据分析中的计算模式分析

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】自然和社会科学中，各种各样关系的数学抽象可以认为是高维（非）线性函数（算子）；各种各样的问题都可以用系统的观点来对待，系统的输出和输入的关系总可以用高维线性或非线性褶积算子来描述。在大数据和人工智能时代，数据分析构成了根本性的任务。数据分析的核心研究对象可以认为是高维（非）线性函数（算子）。研究方法主要不是用解析数学的方法，而是数值数学的方法。针对的核心问题可以描述为：已知高维实测数据体，提取其中蕴含的潜变量（也可称为其中蕴含的信息），基于这些提取的信息做出最佳决策。数值数学上看，这可以认为是求解高维（非）线性函数（算子）定义的变分问题，也可以说是通过能量泛函的最小化解决某种高维参数估计问题或统计推断问题。经典的数值数学、现代的机器学习存在种类繁多的数值算法，有必要对它们展开分析提炼出由一定统一性的计算模式。本文尝试把高维参数估计和机器学习中的输入与输出映射关系的建立这两类最普遍的计算模式（或计算问题）进行统一化，都归结为某种网格计算模型，为更深入的机器学习算法研究提供一定的基础。

【关键词】高维（非）线性函数（算子），数值数学，计算模式，网络结构，高维参数估计（FWI），深度神经网络，网格计算模型

1 引言

自然和社会科学中，各种关系的数学抽象可以表达：解析数学中的高维（非）线性函数；几何学中的高维流形。我们知道高维非线性函数（或高维空间中的流形）可以认为内含了复杂的几何结构，正是这些几何结构体现了或规定了自变量和因变量的关系。高维（非）线性算子，本质上，是两种空间的变换，是一种坐标变换关系。高维（非）线性泛函，可以认为是高维（非）线性函数的特例。

可以认为，高维（非）线性函数、高维（非）线性算子和高维非线性泛函是数学的基本研究对象。地震数据分析中的信号、图像、波场都是这样的高维（非）线性函数。所谓的高维，首先是自变量的高维，既包括空间时间坐标，也包括描述信号、图像和波场的属性参数。此时的高维（非）线性函数往往是物理现实决定的。然而，对信号、图像、波场的变换（即高维（非）线性算子），是科学家依据一定的目的人为构建的。对高维（非）线性函数引入变换的目的是在变换空间中，更能体现高维（非）线性函数的特征，并使得后续的处理更简单有效。

当今大数据和人工智能时代，很多问题的本质关系都可抽象为输入、系统和输出的关系。而描述这种关系的正是高维（非）线性函数（算子）。根据（随机的）实测数据体，提取出其中蕴含的信息，再依据这些信息，在一定的准则下做出最佳决策，是人工智能时代要解决的核心问题。

Bayes 估计、推断或决策理论奠定了数据分析的理论基础。但是，大数据和人工智能时代，算法是最基本的。（大）数据分析或机器学习算法可以认为核心目的就是实现最佳估计、最佳推断或最佳决策。这些算法是存在一些相通的模式的。

高维（非）线性算子和高维非线性泛函及相应问题的数值求解，本质上，都可以抽象为某种网格计算模式。按照当前人工神经网络的术语讲，这些数值求解方法可以抽象为某种网络结构。

以前，我们并没有专门地抽象很多重要的数值算法背后的逻辑相同性。本文对比线性方程组解法、非线性方程组解法、状态空间转移的 Monte Carlo 方法、深度神经网络（以及各种变种神经网络）、动力系统演化方法等等重要的算法，抽象出共同的网格计算模式。期望这样的统一性工作能带来机器学习算法上的研究深化和创新。

2 计算数学的基本内涵

大数据和人工智能时代，万事万物都在试图数字化、信息化、智能化。数据处理是信息化和智能化的基础。数据处理的算法基础是计算数学。

计算数学的研究对象是什么？我认为：计算数学的直接对象还是高维（非）线性函数（算子），只不过是离散后的高维（非）线性函数（算子）。计算数学的目的是：发展出在计算机上实际可行的、理论可靠的、计算复杂性好的各种常用算法。因此，发展出能在计算机上执行的、针对高维（非）线性函数和高维（非）线性算子的算法就是计算数学的最基本内涵。

达到此目的的最根本的思想：①引入对高维（非）线性函数（算子）的最佳逼近。即用多项式、有理分式或三角多项式基函数的线性组合对原高维（非）线性函数进行最佳逼近。原高维（非）线性函数往往是未知形式的、或者仅仅由一些离散观测数据表达的，必须引入上述函数逼近才能处理这样的具体问题！计算机仅仅能进行加、减、乘、除和逻辑判断的基本操作，必须引入上述函数逼近才能处理变化万千的各种具体问题！Hilbert 空间中的函数逼近是计算数学最核心的思想与方法。对逼近函数进行插值处理、数据规则化处理、积分和微分处理、求解常微分和偏微分方程，就可以产生计算机上的有效算法。②引入对高维（非）线性函数（算子）的变换算子。应该知道高维（非）线性算子本质上就是变换。目前讨论的变换主要是线性变换。直接对高维（非）线性算子或变换的研究很少。离散、线性算子就是矩阵。矩阵也可以认为是二阶张量。矢量是一阶张量。高阶张量的研究目前还处于基础的阶段。很少见到张量分析的书籍说明张量分析还没有形成理论体系。某种程度上这也佐证了高维（非）线性算子或变换的研究处于更浅层次的状况。变换的本质就是坐标系的转换。坐标系可以认为是基函数族。 $\boldsymbol{y}=H\boldsymbol{x}$ 本质上就是把向量 $\boldsymbol{x}$ 投影在矩阵 H 规定的列向量坐标系中得到 $\boldsymbol{y}$ 。数学上，希望 $\boldsymbol{y}$ 的各分量间是解耦合的。解耦合的本质是把 $\boldsymbol{y}$ 投影在特征向量构成的坐标系中。这样对 $\boldsymbol{y}$ 的研究就变得更为简单且更深入。

综合一下，可以认为计算数学的最基本内涵就是如下两点：①选择计算机上可以运行的基函数（多项式、有理分式、三角函数），在 Hilbert 空间中进行对原函数的逼近，用对逼近函数的处理代替对原函数的处理，解决数据处理中一大类问题。原函数往往只是表现为离散采样数据。②计算机上能处理的算子仅仅是矩阵，矩阵是离散、线性、有限性维算子。矩阵是一组坐标系。变换是从一个坐标系变换到另一个坐标系。变换的基本目的是把向量投影在特征基函数组成的坐标系中，实现对向量的特征表达。或者把向量表达在另一个坐标系中（特征基函数组成的坐标系），向量的各分量不再相关。

我认为：如果有了上述认识，计算数学的基本思想就理解了！我认为：基本上所有的计算机算法，都是在上述基本内涵的基础上构建出来的。

自然和社会科学中的实际问题是无穷无尽的。对付这些问题的数值算法思想是高度抽象和简洁的。多种多样的数值算法既然来自高度抽象和简洁的思想，有一个相对统一的计算模式吗？

3 计算模式的抽象

数值算法有基本模式吗？我认为计算数学中各种各样的数值算法是可以抽象出一定的计算模式的。

计算模式抽象的出发点依然要遵循系统的观点。无论是物理系统，或是非物理系统，数学形式上，我认为总可以抽象为高维褶积表达的输入与输出之间的关系。这样的关系具有广泛的普适性。

离散、线性、有限维情形下，高维褶积系统可以表达为：

$$\mathbf{d}(\mathbf{x})=\sum_{\Omega}K(\mathbf{x},\mathbf{y})\mathbf{m}(\mathbf{y})d\mathbf{y} \tag{1}$$

上式对应矩阵方程为 $\mathbf{d}=K\mathbf{m}$ 。由计算数学知，只有矩阵形式的 $\mathbf{d}=K\mathbf{m}$ 才能进行具体的数值运算。对于大于二维的情形， K 是四阶或四阶以上的张量。张量的运算还是通过某种形式退化为矩阵运算。

我们知道，（1）式定义的线性正问题及对应的反问题在自然和社会科学中具有十分广泛的应用。因此，可以认为 $\mathbf{d}=K\mathbf{m}$ 和 $\mathbf{m}\approx K^T\mathbf{d}$ 的数值算法代表了一种通用的计算模式。$\mathbf{m}\approx K^T\mathbf{d}$ 是对伪逆简化后的、称为

直接反投影的反问题解法。

可以认为$\mathbf{d}=K\mathbf{m}$的算法对应如下网格计算模型：

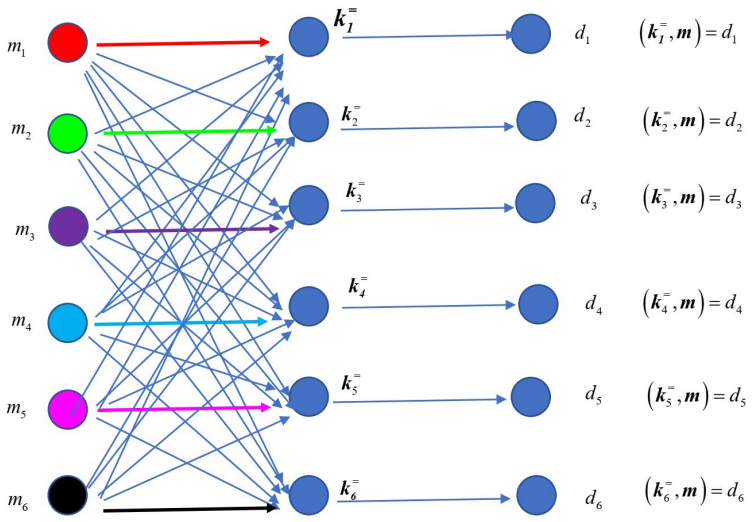


图 1 高维褶积问题对应的网格计算模型

方程中系数矩阵K表示为行向量基函数族$K=\left(\boldsymbol{k}_1^{=},\boldsymbol{k}_2^{=},\cdots,\boldsymbol{k}_N^{=}\right)$。

类似地，可以认为$\mathbf{m}\approx K^T\mathbf{d}$对应如下网格计算模型：

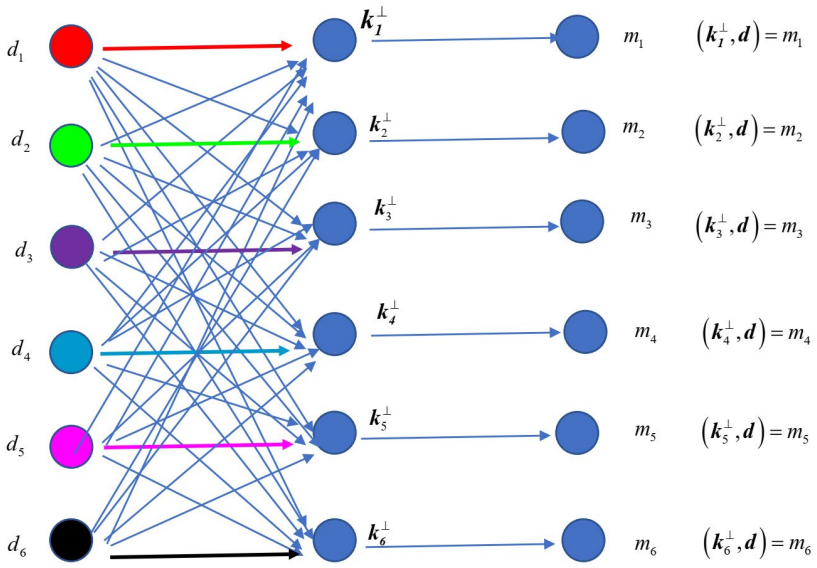


图 2 高维反褶积问题对应的网格计算模型

方程中的系数矩阵K^T表示为列向量基函数族$K=\left(\boldsymbol{k}_1^{\perp},\boldsymbol{k}_2^{\perp},\cdots,\boldsymbol{k}_N^{\perp}\right)$。

我们知道，线性问题的反问题求解往往是转化为变分问题，即误差能量泛函取极小的问题。只要系数矩阵K是正定对称阵，误差能量泛函取极小时的解就是线性问题$\mathbf{d}=K\mathbf{m}$的解。此时，存在基本通用的迭代格式$\mathbf{m}^{k+1}=\mathbf{m}^{k}+\eta K^T\left(\mathbf{d}^{obs}-K\mathbf{m}^{k}\right)=\mathbf{m}^{k}+\eta K^T\delta\mathbf{d}^{k}$。

针对上述迭代格式，可以认为对应如下网格计算模型：

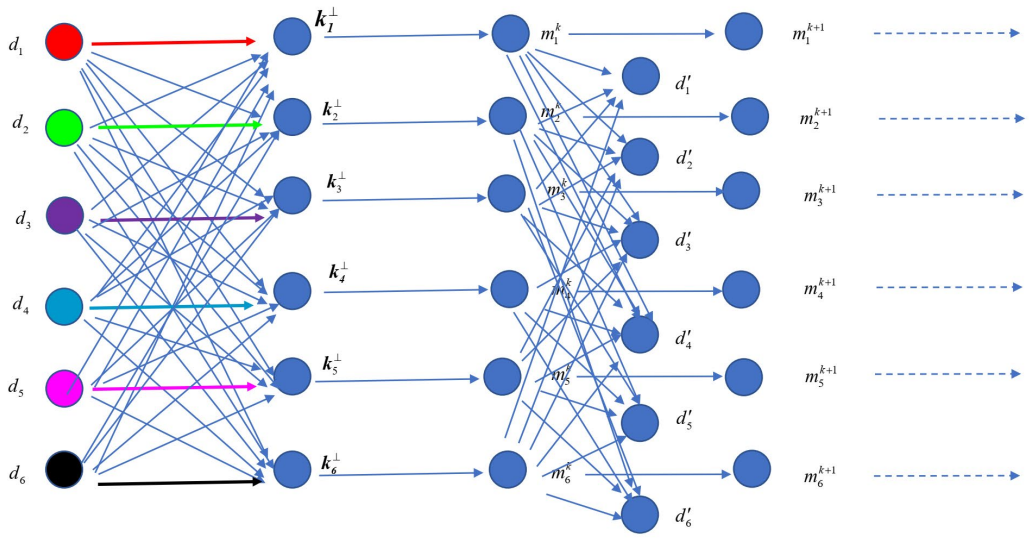


图 3 迭代格式对应的网格计算模型

其中系数矩阵表示为列向量基函数族 $K=\left(\boldsymbol{k}_1^{\perp},\boldsymbol{k}_2^{\perp},\cdots,\boldsymbol{k}_N^{\perp}\right)$。

事实上，非线性正问题对应的反问题的数值计算框架，在逐步线性化模式下，与离散、线性、有限性维情形是一致的。即每一个迭代步解一个方程（1）规定的线性系统对应的反问题。

非线性正问题对应的反问题的数值计算理论框架可以罗列如下。首先定义误差能量泛函，

$$arg\ \ min\ \ \ S(\mathbf{m})=(\mathbf{m}-\mathbf{m}_{prior})^t\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})+(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})^t\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs}) \tag{2}$$

为了导出 Newton 算法，对（2）引入 Taylor 展开近似，

$$S\left(\mathbf{m}_0+\Delta\mathbf{m}\right)=S\left(\mathbf{m}_0\right)+\frac{\partial S}{\partial\mathbf{m}}\Delta\mathbf{m}+\frac{1}{2}\frac{\partial^2 S}{\partial\mathbf{m}^2}\Delta\mathbf{m}^2+\mathrm{O}(\Delta\mathbf{m}^2) \tag{3}$$

对 Taylor 近似进行截断处理，仅仅保留二次项，得到局部近似的误差能量泛函，

$$\tilde{S}\left(\mathbf{m}\right)\triangleq S\left(\mathbf{m}_0+\Delta\mathbf{m}\right)=S\left(\mathbf{m}_0\right)+\frac{\partial S}{\partial\mathbf{m}}\Delta\mathbf{m}+\frac{1}{2}\frac{\partial^2 S}{\partial\mathbf{m}^2}\Delta\mathbf{m}^2 \tag{4}$$

对局部近似的误差能量泛函进行求极值运算。即求如下的一阶导数，并令其等于零，

$$\frac{\partial\tilde{S}}{\partial\Delta\mathbf{m}}=0 \tag{5}$$

由（5）式得到，

$$\frac{\partial\tilde{S}}{\partial\mathbf{m}}=-\frac{\partial^2\tilde{S}}{\partial\mathbf{m}^2}\Delta\mathbf{m} \tag{6}$$

显然，（6）式给出当前局部点的解的更新量，

$$\Delta\mathbf{m}=\left[\frac{\partial^2\tilde{S}}{\partial\mathbf{m}^2}\right]^{-1}\frac{\partial\tilde{S}}{\partial\mathbf{m}} \tag{7}$$

把更新量加入到上一步迭代解就给出如下的迭代格式，

$$\mathbf{m}^{k+1}=\mathbf{m}^{k}-\left[\frac{\partial^2\tilde{S}}{\partial\mathbf{m}^2}\right]^{-1}\frac{\partial\tilde{S}}{\partial\mathbf{m}} \tag{8}$$

我们知道，上述网格计算模式，是由既定的数理关系（因果关系）确定伪逆，给出解估计结果。这与仿生人脑神经元组成的深度神经网络是不同的。但也可以看到它们之间的类似性。

伪逆是个线性的概念，建立的是由数据空间到参数空间的线性投影关系。这个投影关系的优点是梯度项的物理意义很清楚，尤其在 FWI、层析成像和最小二乘 RTM 中。这种有明确几何或物理意义的梯度项，十分有利于引入特定的正则化处理，提升解估计的稳定性。

针对非线性正问题对应的反问题，存在另一个典型的计算模式。它就是 Markov 状态转移。寻找非线性正问题对应的反问题的解的过程可以认为是一个在可能的解空间中随机游走的过程。其中的核心思想是：对参数空间（或可能的解空间）进行重要性采样。

对参数空间（或可能的解空间）的重要性采样可以解决高维（非）线性泛函的全局优化问题。存在一大类 Monte Carlo 重要性采样思想与方法。Metropolis 准则下的 Markov 状态转移是实现重要性采样的基本思想。但是，低效率和难以证明真的达到了重要性采样是存在的根本问题。为了进一步提高重要性采样的效率，引入了模拟退火、遗传算法、粒子群算法、蚁群算法等等。这些大都来自于仿真物理过程或仿真生物进化过程。

很显然， Metropolis 准则下的 Markov 状态转移方法实现了由数据空间到参数空间的非线性映射。这个映射是在 Metropolis 准则，沿着复杂的游走路径实现的。但一大类 Monte Carlo 方法显然不是仿真人脑的学习类算法。Metropolis 准则下的 Markov 状态转移方法也可以抽象成某种网络计算模式。我认为：目前，我们显然没有能真正地理解高效率重要性采样的本质，也没有上升到建立由数据空间到参数空间的非线性映射的层次上理解 Monte Carlo 方法。也许在 DNN 的激励下，能引入 DNN 的某些想法，更好地实现重要性采样，发展出十分有效的新算法。

另外，必须要说明的是：任何计算模式，解决的是数据空间到参数空间（更合适地，参数应称为决策目标）的映射问题或是伪逆估计问题。但是，估计理论明确指出：好的估计应该是无偏和方差最小的。令人遗憾的是，好的估计不取决于计算模式，而是取决于数据空间到参数空间（决策目标）之间的关系。另外，计算模式很少考虑如何兼顾估计的无偏和方差最小的要求。这应该是机器学习算法存在的比较严重的问题。

4 多层神经网络与 FWI 计算模式对比

人脑是一个强大的信息处理系统，具备十分强大的信息能力。相对于计算机系统在某些方面有巨大的优势。当然，它也有巨大劣势的方面，譬如大数据处理能力、计算能力远远不如计算机。用计算机模拟人脑的信息处理能力是人类不懈追求的一个目标。

按照（认知）神经科学的研究，人脑可以认为是一个高度并行的、分布式的计算网络。人工神经网络就是把基本的神经元（感知机）按照一定的网络结构组合成深度神经网络，解决各种各样的应用问题。

本质上，人工神经网络就是一个计算模型。不过它仿生模拟了人脑神经元及人脑神经网络。事实上，是否真的仿生模拟了人脑神经元及人脑神经网络还很难说。经验地讲，很多仿生优化算法还是有比较好的结果的。既然人脑神经元构成的网络有如此强大的学习、联想、推断和决策能力，自然可以期望人工神经网络也应该有比较好的性能。关键是如何能做到！人类发展新的科学和技术往往是向自然界寻求灵感的。但人类对自然的仿生，从来不是生搬硬套的。也不可能做到生搬硬套。人工神经网络的基本模式已经确定了，当前的重点是根据应用问题发展特定的网络算法。

单个人工神经元（感知机）可以由图 4 来描述。

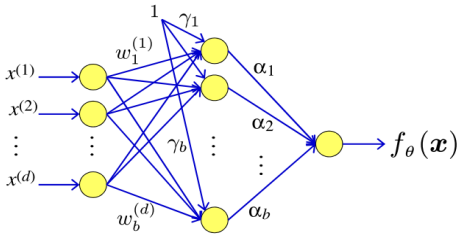


图 4 单个人工神经元（感知机）示意图

图 4 对应的计算模式可以数学抽象为

$$f_{\boldsymbol{\theta}}(\boldsymbol{x})=\sum_{j=1}^{b}\alpha_j\phi(\boldsymbol{x};\boldsymbol{\theta}_j) \tag{8}$$

其中，

$$\phi(\boldsymbol{x};\boldsymbol{\theta})=\frac{1}{1+\exp(-\boldsymbol{x}^T\boldsymbol{w}-\gamma)} \tag{9}$$

式中的 $\boldsymbol{\theta}$ 定义为 $\boldsymbol{\theta}=(\boldsymbol{w}^T,\gamma)^T$ 。

（8）和（9）式显然描述了一个非线性逼近问题。它建立起输入与输出之间的非线性关系。理论已经证明：只要隐藏层足够宽，该计算模式的输出可以无限逼近于给定的期望输出。

从估计理论知道：上述最佳逼近并不能保证推断结果或决策结果的最佳。原因何在：输入数据含噪、输入数据不全、输入数据与决策目标之间的存在高度非线性。上述逼近理论给出的最佳逼近的结论并没有回答估计或判决理论关心的问题。估计理论的期望是：估计结果无偏且方差最小；逼近理论的要求是：感知机给出的输出尽可能等于已知的期望输出（即图 4 中的 $f_{\boldsymbol{\theta}}(\boldsymbol{x})$）。

由单个神经元组成的人工神经网络和深度神经网络可以由图 5 来表示。

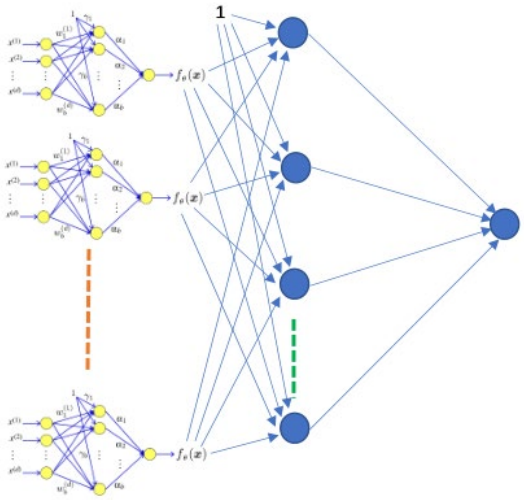


图 5 单个人工神经元（感知机）组成的人工神经网络的示意图

图 5 所示，由单个神经元组成人工神经网络（ANN/DNN），是单个神经元功能的扩展，并没有改变非线性映射的本质。ANN/DNN 仅仅是又一个网络结构！很显然，ANN 是一种分布并行算法。ANN 模仿人脑

计算的功能，是一种智能算法。ANN/DNN 的目的也是建立起“广义数据”与决策目标之间的非线性映射关系。这种映射关系可以抽象为如下的数学模型，

$$F\left(\boldsymbol{x},\boldsymbol{w}\right)=\varphi\left(\sum_{k}w_{0k}\varphi\left(\sum_{j}w_{kj}\varphi\left(\cdots\varphi\left(\sum_{i}w_{li}\tilde{\boldsymbol{x}}_{i}\right)\right)\right)\right) \tag{10}$$

其中，$\tilde{\boldsymbol{x}}=\left(f_{\theta}^{1}\left(\boldsymbol{x}\right),\ f_{\theta}^{2}\left(\boldsymbol{x}\right),\cdots,\ f_{\theta}^{b}\left(\boldsymbol{x}\right)\right)$，$\tilde{f}_{\theta}\left(\tilde{\boldsymbol{x}}\right)=\sum_{j=1}^{b}\alpha_{j}\phi\left(\tilde{\boldsymbol{x}};\boldsymbol{\beta}_{j}\right)$。

对于 ANN/DNN 网络，关键是训练过程，即用大量的训练数据建立起“广义数据”与决策目标之间的非线性映射关系。训练算法（BP 算法）借鉴了自适应滤波中的 LMS（Least Mean Square）方法。但是，本质上，BP 算法并非用了误差的均方，而是单点预测误差，这就会导致梯度随机。很难期望这样的算法能对付强非线性问题。

BP 训练算法可以简单罗列如下：

训练数据集：$\mathscr{J}=\left\{\boldsymbol{x}\left(n\right),\boldsymbol{d}\left(n\right)\right\}_{n=1}^{N}$ （11a）

数据误差：$e_{j}\left(n\right)=d_{j}\left(n\right)-y_{j}\left(n\right)$ （11b）

误差泛函：$\mathbb{E}\left(n\right)=\sum_{j\in C}\mathbb{E}_{j}\left(n\right)=\frac{1}{2}\sum_{j\in C}e_{j}^{2}\left(n\right)$ （11c）

总平均误差泛函：$\mathbb{E}_{avg}\left(n\right)=\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left(n\right)=\frac{1}{2N}\sum_{n=1}^{N}\sum_{j\in C}e_{j}^{2}\left(n\right)$ （11d）

本质上，这也是一个非线性正问题对应的反问题。同样地，存在 Newton 算法，

$$\begin{aligned}&\mathbb{E}\left(\boldsymbol{w}\left(n\right)+\Delta\boldsymbol{w}\left(n\right)\right)\\&=\mathbb{E}\left(\boldsymbol{w}\left(n\right)\right)+\boldsymbol{g}^{T}\left(n\right)\Delta\boldsymbol{w}\left(n\right)+\frac{1}{2}\Delta\boldsymbol{w}^{T}\left(n\right)H\left(n\right)\Delta\boldsymbol{w}\left(n\right)+\mathrm{O}\left(\left(\Delta\boldsymbol{w}\left(n\right)\right)^{2}\right)\end{aligned} \tag{12a}$$

$$\Delta\boldsymbol{w}\left(n\right)=H^{-1}\left(n\right)\boldsymbol{g}^{T}\left(n\right) \tag{12b}$$

$$\boldsymbol{w}^{k+1}\left(n\right)=\boldsymbol{w}^{k}\left(n\right)+\Delta\boldsymbol{w}^{k}\left(n\right)=\boldsymbol{w}^{k}\left(n\right)+H^{-1}\left(n\right)\boldsymbol{g}^{T}\left(n\right) \tag{12c}$$

其中，梯度项为 $\boldsymbol{g}\left(n\right)=\left.\frac{\partial\mathbb{E}\left(\boldsymbol{w}\right)}{\partial\boldsymbol{w}}\right|_{\boldsymbol{w}=\boldsymbol{w}\left(n\right)}$，Hessian 项为 $H\left(n\right)=\left.\frac{\partial^{2}\mathbb{E}\left(\boldsymbol{w}\right)}{\partial\boldsymbol{w}^{2}}\right|_{\boldsymbol{w}=\boldsymbol{w}\left(n\right)}$。

下面展开对 FWI 算法和 ANN/DNN 算法的对比分析。

FWI 本质上建立了一个伪逆，实现由数据到参数的映射。FWI 有大迭代/外迭代和小迭代/内迭代组成。每个外迭步对应一个局部二次型的优化问题；内迭代是求解每个局部二次型的优化问题（如果按 $\Delta\mathbf{m}^{k}=\left[\frac{\partial^{2}\tilde{S}\left(\mathbf{m}^{k}\right)}{\partial\mathbf{m}^{2}}\right]^{-1}\frac{\partial\tilde{S}\left(\mathbf{m}^{k}\right)}{\partial\mathbf{m}}$ 计算，内迭代步可以一次完成！）。可以看出，每个线性迭代步都有一个伪逆。每一个伪逆算子都是线性的。伪逆由当前状态的正算子决定！这反映了非线性映射的建立是由每个线性映射步组合实现的。这正是 Newton 法的思想本质。可以清楚地看出，Newton 法是仅仅针对局部光滑可微的、凸泛函有效的！只能得到局部最优的解估计！Newton 法应该不是求解非线性泛函优化问题的针对性方法；

但它的确是针对连续可微的、凸泛函的针对性方法。FWI 计算上的局限性是非常明显的！但是，这就是目前我们所拥有的算法！

ANN/DNN 本质上也是建立了一个伪逆，实现由数据到参数的映射。公式 $F\left(\boldsymbol{x},\boldsymbol{w}\right)=\varphi\left(\sum_k w_{0k}\varphi\left(\sum_j w_{kj}\varphi\left(\cdots\varphi\left(\sum_i w_{li}x_i\right)\right)\right)\right)$ 就反映了这种映射。这个映射应该包括网络拓扑结构、激活函数 $\varphi(\bullet)$、和权系数向量 $\boldsymbol{w}$。但是，目前 DNN 一般仅仅训练估计权系数向量 $\boldsymbol{w}$。公式 $F\left(\boldsymbol{x},\boldsymbol{w}\right)=\varphi\left(\sum_k w_{0k}\varphi\left(\sum_j w_{kj}\varphi\left(\cdots\varphi\left(\sum_i w_{li}x_i\right)\right)\right)\right)$ 反映出这个伪逆是非线性的。非线性的伪逆算子的几何结构是很难解释的。线性的伪逆也可以视为一个坐标系，有明确的几何解释。连续凸泛函下有明确的解的意义。

$F\left(\boldsymbol{x},\boldsymbol{w}\right)=\varphi\left(\sum_k w_{0k}\varphi\left(\sum_j w_{kj}\varphi\left(\cdots\varphi\left(\sum_i w_{li}x_i\right)\right)\right)\right)$ 理论上可以做到由数据（输入）到参数（决策目标）的高度逼近。数学上的逼近定理也已经证明了这一点。BP 算法训练它，也在很多应用上做到这一点。但是，在训练数据不全、训练数据有噪音、训练数据对（标签数据）关系错误的情况下，这种伪逆算子的泛化应用能力完全没有保障。再次说明：逼近问题与估计问题不是等价的！FWI 在高度非线性情况下的应用也完全没有保障，这已经在复杂的勘探地震数据中得到了证明！

在高度非线性的反问题中，强调构建一个最佳逼近的函数，并不能保证反演解（或 DNN 泛化应用）的有效性。很多人混淆了逼近问题与估计问题的本质差异！只要选复杂的逼近函数，一般地，总可以实行非常高精度的逼近。但无偏+方差最小的估计却是极难做到的！

在有数学物理预测关系控制或是完全由数据驱动情况下，根据实测数据来估计系统参数或逼近决策目标，抽象地，可以认为都在估计一个伪逆算子。伪逆算子的性态应该是有界的、紧的、连续的、……，最好是可逆的。我认为无论全 Newton 算法还是 DNN 算法，本质上，伪逆算子的上述性态都不可以回避的；估计的无偏和方差最小的特性也是不可以忽视的。

由实测数据映射成决策目标的问题，本质上就是一个 Bayes 最佳决策问题。ANN/DNN 算法仅仅是一种新的计算模式下实现而已，必须要在 Bayes 决策和 Bayes 估计的理念（强调无偏+方差最小的估计）下，进一步地借鉴人脑的生物学进化特征，优化 DNN 算法，让 DNN 算法的伪逆更有泛化能力。对训练数据的理解、对网络结构的理解、对 BP 算法的理解、对正则化方法的理解，综合在一起才能用好 DNN，发挥它的潜力。

我认为，当前的 ANN/DNN 算法距离真正好地解决非线性正问题对应的反问题尚有巨大的差异。

5 结论性认识

建立（高维）观测数据到潜变量（或待估计参数、或决策目标）之间的映射关系广泛地存在于数据科学和认知科学领域，相关的应用领域更是很难计数。这既是一个科学问题，又是一个技术问题。科学上，这是一个（强）非线性反问题，极为具有挑战性。可以说强非线性反问题的求解是目前科学上最为束手无策的问题之一！技术上这是必须要给出可行解的问题。

某种程度上，很多种机器学习算法是解决（强）非线性反问题的新探索。首要的是仿生人脑的功能。当然，也仿生各种自然进化功能。主分量分析、独立分量分析、稀疏与冗余表达、甚至模拟动力系统的演化都成了解决（强）非线性反问题的探索方向。发展出（强）非线性反问题的有效算法是基本追求。

建立（高维）观测数据到潜变量（或待估计参数、或决策目标）之间的映射关系存在各种不同的计算模式。各种典型的计算模型都可以抽象为某种网格计算模式。线性褶积系统及其对应的反问题；非线性褶积系统及其对应的反问题；有监督学习所用的深层网络系统；无监督学习对应的各种网络系统，本质上都可以抽

象成一个网络计算模型。更一般地，很多数值计算模式都可以抽象为状态转移问题，状态转移问题的解决都可以用某种网络计算模型来完成。因此，可以说模仿人脑功能的人工神经网络模型并非是很特殊的，它仅仅是网络计算模型的一种。

由于（强）非线性反问题不存在有效的解法，必须从技术环节上引入各种正则化方法，牺牲解的无偏性，提升解的稳定性（降低方差）。但是，当前的网络计算模式并不或很难考虑估计结果的无偏和方差最小的要求。尤其对 DNN 和 Monte Carlo 算法。估计结果的无偏和方差最小是指导反演问题的最基本原则要求，绝对不能仅仅盯着建立一个能对期望输出进行高度逼近的非线性函数！

总之，数据分析问题基本上是一个统计决策问题。信息不足、含噪、非线性导致各种决策算法效果不彰。基本情况就是如此，我们要做的只是深入理解数据、深入理解算法，给出合理的解决问题的方案，产生可用的、“最佳的”决策结果（估计结果）。

参考文献

[1]. Simon Haykin, 2009, Neural Networks and Learning Machines; Third Edition，Pearson Educcation, Inc.

[2]. Moon, T. K., and Stirling, W. C., 2000, Mathematical Methods and Algorithms for Signal Processing; Prentice Hall

[3]. Gene H. Golub, and Charles F. Van Loan, 1996, Matrix Computation, Third Edition, The Johns Hopkins University Press

[4]. Don Simon, 2013, Evolutionary Optimization Algorithms: Biologically Inspired and Population-Based Approaches to Computer Intelligence; John Wiley & Sons, Inc.

[5]. Alfio Quarteroni, Riccardo Sacco, Fausto Saleri, 2000, Numerical Mathematics; Springer Scoence + Business Media, Inc.

[6]. Michael S. Gazzaniga, Richard B. Ivery, George R. Mangun, 2009, Cognitive Neuroscience – The Biology of the Mind （3rd Edition）; W. W. Norton & Company

反褶积问题的基本理论与在勘探地震中应用的认识

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组（WPI），上海

【摘要】科学和工程中绝大部分的应用问题都可归结为依据对系统输出变量的观测结果进行估计系统参数，甚至估计系统的输入。在勘探地震中，与之对应的一个典型问题是褶积与反褶积问题。更为广泛的意义下，这个问题应该描述为 Bayes 估计理论下的系统参数估计问题，因为系统输出变量的观测结果往往被认为是一个满足一定概率分布的随机向量，待估计的系统参数也是满足一定概率分布的随机量。地震数据分析中的若干处理环节都会涉及到不同反褶积方法，但是这些方法技术往往并没有被统一在同一个抽象的理论框架下，导致理不清问题的本质，方法技术的发展达不到最优状态。本文试图简洁而深入地讨论反褶积问题的本质，并把它们体现在地震数据分析的相关具体环节中，为提升地震数据分析的方法技术层次做些理论铺垫。

【关键词】物理系统，输入与输出，褶积系统，反褶积，Bayes 估计理论，地震数据分析

概述

反褶积是一个涵盖范围很广的议题，因为褶积过程（广义的褶积）包含的内容实在太多了！科学和工程中绝大部分的应用问题，甚至非常多的社会科学问题也可以抽象成一个系统的输入输出模式。通过观测系统的输出，对系统本身进行辨识、参数估计，甚至对输入进行估计，是自然科学、工程和社会科学领域中广泛存在的议题。

一般地，物理系统由于冲击而产生输出，这个过程就对应一个褶积问题。输出是输入被系统响应函数褶积的结果，系统响应函数往往表现出带通（低通）滤波器的特点。褶积公式被表达为：

$$y(t)=h(t)*x(t) \tag{1}$$

其中，$h(t)$是系统响应函数，$x(t)$是系统冲激函数（激励函数）。$y(t)$表示系统的输出。一般地，认为$y(t)$是满足一定概率分布的随机向量。

反褶积就是由实测的系统输出$d(t)\left(=y(t)+\eta(t)\right)$估计系统参数$h(t)$，甚至同时估计系统输入$x(t)$。

对（1）式还存在一个十分流行的解释，即输入是一个真像，$h(t)$是一个模糊算子，$y(t)$是被模糊的真像。反褶积的目的是尽可能完整地恢复真像。实际上，高维的、时变的、非线性的褶积系统是更为普遍的模型。高维情形下，噪音$\eta(\mathbf{x},t)$很可能不满足高斯分布。

另外，褶积模型的建立并非一定要针对物理系统，只要能抽象成类似的物理系统，都可以建立类似的褶积关系。可以看出公式(1)具有十分广泛的包含性，据此产生了反褶积问题在不同领域中的广泛存在。

一般地，反褶积的目的是（尽可能地）消除系统响应函数的模糊（平滑、低通滤波）效应，恢复系统输入函数（冲击函数）的原样。系统参数估计问题不被称为反褶积问题，尽管从 Bayes 参数估计理论看，二者没有什么理论上的差异。当前，我们处理的绝大多数反褶积问题（包括参数估计问题）都是在（系统）线性（甚至要求系统最小相位）、观测噪音高斯、系统时不变假设下进行的。这样的假设显然是与实际情况存在偏差的。这就要求数据分析方法技术向高维、时变、非线性、非高斯情形下推进。

地震数据分析中的若干处理环节都会涉及到不同反褶积方法，但是这些方法技术往往并没有被统一在同一个抽象的理论框架下，很难弄清楚不同的反褶积方法之间的理论关系。本文试图简洁而深入地讨论反褶积问题的本质，把它们统一在 Bayes 估计的理论框架下、统一在 Fredholm 第一类积分方程的框架下、统

一在线性反演理论的框架下，为提升地震数据分析的方法技术层次做些理论铺垫。

勘探地震中的褶积模型物理

勘探地震中，褶积模型的来源有特殊的物理解释。

首先，我们假设平面波入射到水平层状介质中，假若垂直入射到水平层状介质，波在此情形下的传播可用一维波动方程来描述。

$$\frac{\partial^2 u}{\partial z^2}=\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}+f_{(z=0,t)}^{plane_w} \tag{2}$$

这个一维波动方程描述的波传播，包含了层间多次波、自由表面多次波，也包括一次反射波，公式(2)描述的平面波垂直入射到水平层状介质中的波传播过程可认为是一个非线性的问题。可以认为这是个非线性的褶积问题。

而一般地，总假定褶积问题是线性的，即 $h(t)$ 是线性褶积算子，公式(1)可以用矩阵方程来表示。勘探地震中常用的线性褶积算子可由 Zoeppritz 方程导出。平面波入射到水平层状介质在界面上发生的物理，可由图 1 展示：

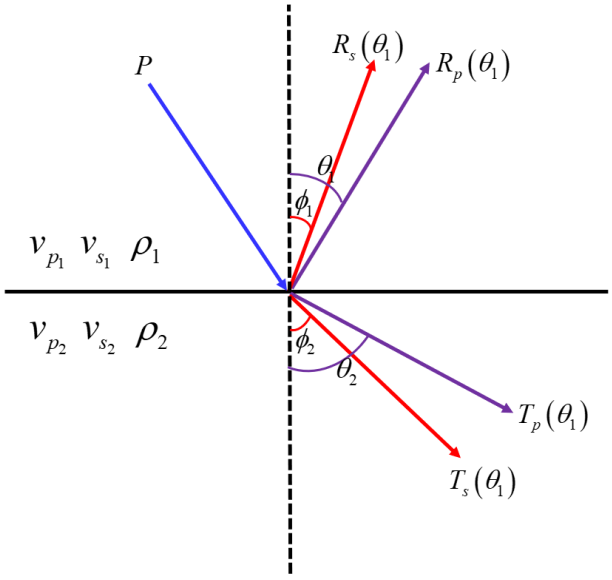


图 1 平面波入射到水平层状介质物理示意图

平面波入射到图 1 所示的水平层状界面上，波场连续条件结合合理的边界条件可以导出如下的 Zoeppritz 方程，

$$\begin{bmatrix} R_p(\theta_1) \\ R_s(\theta_1) \\ T_p(\theta_1) \\ T_s(\theta_1) \end{bmatrix}=\begin{bmatrix} -\sin\theta_1 & -\cos\phi_1 & \sin\theta_2 & \cos\phi_2 \\ \cos\theta_1 & -\sin\phi_1 & \cos\theta_2 & -\sin\phi_2 \\ \sin 2\theta_1 & \frac{v_{p_1}}{v_{s_1}}\cos 2\phi_1 & \frac{\rho_2 v_{s_2}^2 v_{p_1}}{\rho_1 v_{s_1}^2 v_{p_2}}\cos 2\phi_1 & \frac{\rho_2 v_{s_2} v_{p_1}}{\rho_1 v_{s_1}^2}\cos 2\phi_2 \\ -\cos 2\phi_1 & \frac{v_{s_1}}{v_{p_1}}\sin 2\phi_1 & \frac{\rho_2 v_{p_2}}{\rho_1 v_{p_1}}\cos 2\phi_2 & \frac{\rho_2 v_{s_2}}{\rho_1 v_{p_1}}\sin 2\phi_2 \end{bmatrix}^{-1}\begin{bmatrix} \sin\theta_1 \\ \cos\theta_1 \\ \sin 2\theta_1 \\ \cos 2\phi_1 \end{bmatrix} \tag{3}$$

其中横波是 sv 波，即横波转换波。当平面波垂直入射，Zoeppritz 方程退化为：

$$\begin{bmatrix} R_p(0°) \\ R_s(0°) \\ T_p(0°) \\ T_s(0°) \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & \dfrac{v_{p_1}}{v_{s_1}} & 0 & \dfrac{\rho_2 v_{s_2} v_{p_1}}{\rho_1 v_{s_1}^2} \\ -1 & 0 & \dfrac{\rho_2 v_{p_2}}{\rho_1 v_{p_1}} & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \tag{4}$$

公式(4)可以改写为：

$$\begin{bmatrix} R_p(0°) \\ R_s(0°) \\ T_p(0°) \\ T_s(0°) \end{bmatrix} = \begin{bmatrix} 0 & \dfrac{\rho_2 v_{p_2}}{\rho_2 v_{p_2} + \rho_1 v_{p_1}} & 0 & \dfrac{-\rho_1 v_{p_1}}{\rho_2 v_{p_2} + \rho_1 v_{p_1}} \\ \dfrac{-\rho_2 v_{s_2}}{\rho_2 v_{s_2} + \rho_1 v_{s_1}} & 0 & \dfrac{-\rho_1 v_{s_1}^2}{v_{p_1}\left(\rho_2 v_{s_2} + \rho_1 v_{s_1}\right)} & 0 \\ 0 & \dfrac{\rho_1 v_{p_1}}{\rho_2 v_{p_2} + \rho_1 v_{p_1}} & 0 & \dfrac{\rho_1 v_{p_1}}{\rho_2 v_{p_2} + \rho_1 v_{p_1}} \\ \dfrac{\rho_1 v_{s_1}}{\rho_2 v_{s_2} + \rho_1 v_{s_1}} & 0 & \dfrac{\rho_1 v_{s_1}^2}{v_{p_1}\left(\rho_2 v_{s_2} + \rho_1 v_{s_1}\right)} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \tag{5}$$

因此，有

$$\begin{aligned} & R_s(0°) = T_s(0°) = 0 \\ & R_p(0°) = \frac{\rho_2 v_{p_2} - \rho_1 v_{p_1}}{\rho_2 v_{p_2} + \rho_1 v_{p_1}} = \frac{z_2 - z_1}{z_2 + z_1} \\ & T_p(0°) = 1 - R_p(0°) \end{aligned} \tag{6}$$

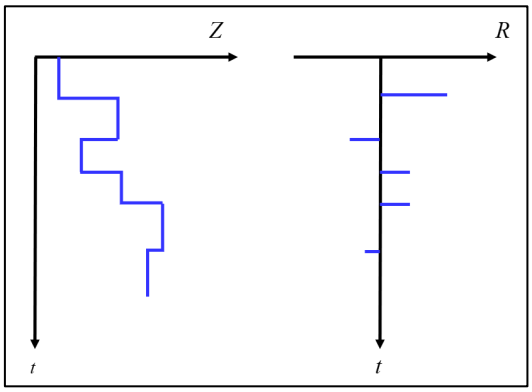


图 2 波阻抗与反射系数示意图

水平层状介质波阻抗变化引起反射系数的存在，见图 2。反射系数序列与地震子波的褶积，构成勘探地震中常用的褶积模型，

$$S(t) = W(t) * R(t) \tag{7}$$

事实上，（6）式仅仅给出了零角度反射系数与地层波阻抗变化之间的关系，即它是相对波阻抗（扰动）。

由（6）式转到（7）式并非按理论逻辑导出的。严格地，$S(t)=W(t)*R(t)+f\left(W(t),R(t)\right)$，其中$f\left(W(t),R(t)\right)$代表自由表面和层间多次波的贡献。这是非线性的褶积模型，$f\left(W(t),R(t)\right)$具有很复杂的形式。很显然，只有地下反射界面反射系数很小，且自由地表不存在或自由地表反射系数也很弱时，才能忽略$f\left(W(t),R(t)\right)$，得到（7）式的线性褶积模型。

对于实际地震记录，水平层状介质情况下的实际地震记录，有如下线性褶积模型来描述：

$$d(t)=S(t)+\eta(t)=W(t)*R(t)+\eta(t) \tag{8}$$

公式(7)和公式(8)是勘探地震中的褶积模型，前述讨论说明了它们的物理背景。

本质上，(7)和(8)式也可以从系统的角度来理解，此时，系统响应函数为子波$W(t)$，期望它是最小相位的；而系统的激励函数为反射系数序列$R(t)$，一般地，假设它是高斯白噪的。

更进一步地，图像 PSF 褶积模型，地表一致性褶积模型等，各种情形都可以抽象出特定的线性褶积模型，尤其是 Born 近似下一阶扰动波场与地下介质扰动关系：

$$\delta u(\mathbf{x},t)=\int G(\mathbf{x}',\mathbf{x},t)\delta m(\mathbf{x}')d\mathbf{x}' \tag{9}$$

是更为经典的勘探地震中的高维线性褶积模型。其中，$\delta m(\mathbf{x}')$代表地下像点的介质参数扰动；$\delta u(\mathbf{x},t)$代表地表观测到的、由于地表震源激发的波场激活地下介质扰动而产生的、传播到地表的散射场的线性叠加结果；$G(\mathbf{x}',\mathbf{x},t)$代表连接地表震源、地下散射点和地表接收点的 Green 函数。（9）式的具体推导在很多地震波成像的文献中可以查到。

通过上述不同的褶积模型，可以把信号分析、图像分析和参数估计统一在褶积与 Bayes 估计理论下的反褶积的理论框架下。广义线性反演问题是反褶积问题的统一称谓！

地表一致性褶积模型的地球物理解释

勘探地震中，还有一个广泛存在的褶积模型。它与（7）和（9）式的物理解释差异很大。这就是地表一致性褶积模型。地震波在地下介质中传播，同相轴上稳定的子波$W(t)$受到的影响因素包括：源端的复杂因素影响、源端到反射界面的传播因素的影响、反射界面的影响、反射界面到检端的传播因素影响、检端的复杂因素影响。

很显然，整个过程可以用褶积模型表示出来。褶积模型可以写为：

$$d=W_s*W_+*(W*R)*W_-*W_R \tag{10}$$

其中，W_s被认为是源端复杂因素影响；W_+被认为是源端到反射界面的传播因素的影响；W_-被认为是反射界面到检端的传播因素的影响；W_R：被认为是检端复杂因素影响。W 是传播到反射界面处的、期望的地震子波；R 是反射系数。d 当然是实测的地震道。

核心问题时如何在地表一致性假设下，消除源端和检端的复杂因素影响，即W_s和W_R的影响。这就是地表一致性反褶积要解决的问题，其中W_+和W_-主要通过各向异性/Q 介质中的波场反传来消除。

很显然，（10）式也是个褶积模型。但它并不是推导出来的，而是按照一般褶积系统的物理含义直接写出来的。因此，广义地理解褶积模型是必要的、也是必须的。不能困于褶积模型只能来自于物理系统的机械认识。

反褶积问题的基本概念

从根本上讲，反褶积问题就是一个线性反演问题。在 Bayes 估计理论下，可以对反褶积问题的各方面产生深入的认识。可以说反褶积是褶积模型与线性反演理论结合产出的方法技术，反褶积是线性、高斯和平稳假设情形下的 Bayes 估计问题。

反褶积有多种多样的实际应用情形。勘探地震中，最典型的反褶积应是 Spike Deconvolution（脉冲反褶积）。

假定已知地震子波$W(t)$，脉冲反褶积应提为最小二乘意义下线性反演问题：

$$\varepsilon(f)=\left\|f(t)*W(t)-\delta(t)\right\|^2 \tag{11}$$

用来估计反子波$\mathrm{f(t)}=\mathrm{W(t)}^{-1}$。反子波的具体计算过程为，

$$\frac{\partial\varepsilon}{\partial f}=0\Rightarrow\mathbf{W}^H\mathbf{W}\mathbf{f}=\mathbf{W}\boldsymbol{\delta} \tag{12}$$

$$R_{ww}\mathbf{f}=\mathbf{r}_{w\delta} \tag{13}$$

$$\mathbf{f}=R_{ww}^{-1}\mathbf{r}_{w\delta} \tag{14}$$

公式(13)是公式(11)对应的法方程，(14)是反子波的估计结果。$\hat{\delta}(t)=f(t)*d^{obs}(t)$给出脉冲反褶积结果，即反射系数序列。

(14)式存在唯一解的条件显然是子波自相关矩阵R_{ww}是正定对称的，或对角占优的。更具体的，R_{ww}是满秩的、可逆的！与此对应地，要$W(t)$是最小相位子波。

这是勘探地震中反褶积的最基础做法。

统一的最佳 Wiener 滤波（反滤波）理论框架

最佳 Wiener 滤波是 L2 范数下的整形滤波器（波形逼近滤波），是 Bayes 线性估计，即线性均方误差最小的估计。最佳 Wiener 滤波是反褶积问题的更普遍的实现方法。

假定实测数据为$d^{obs}(t)$，根据褶积模型(7)，对应的最佳 Wiener 滤波的逼近误差定义为：

$$\varepsilon(f)=\left\|W(t)*R(t)-d^{obs}(t)\right\|^2 \tag{15}$$

对应的法方程为：

$$C_{WW}R=\mathbf{r}_{Wd} \tag{16}$$

其中，$C_{WW}=W^HW$是子波行向量外积构成的系数矩阵，它是一个自相关矩阵。(16)式给出的估计结果$\hat{R}(t)$与$\hat{\delta}(t)=f(t)*d^{obs}(t)$给出的结果是等价的。

(15)和(16)式是在最佳 Wiener 滤波框架下描述的反褶积方法。信号分析中广泛应用的 AR Model 也可认为是类似(7)式的线性褶积模型，是滤波算子（滤波器）与已知数据的线性褶积。因此，所有基于线性褶积模型的问题都可以按照(15)和(16)式定义的方式来解决。

因此，我们说线性褶积问题和反褶积问题实质上代表了一大类类似的实际应用问题，它们都可以在线性反演理论下得到解决。

可以看出，最佳 Wiener 滤波是线性、高斯、平稳假设下的线性 Bayes 估计。Gauss-Markov 定理告知：线性褶积模型、平稳信号及噪音高斯白噪假设下，Bayes 估计结果是无偏和方差最小的估计。这就清楚地说明了线性反演理论的解的本质含义。

脉冲反褶积情形下的最小相位子波的要求是在上述假设下的更进一步要求，使得脉冲反褶积理论上存在唯一解。实际上，即便做了地震子波的最小相位化，也不能保证脉冲反褶积系数序列的估计结果是唯一的。噪音分布的未知或非高斯会带来反演解的理论模糊性。

因此，线性、高斯、平稳假设下最佳 Wiener 滤波理论是更具解释能力的线性反演理论。各种最佳滤波、最佳反滤波都在此框架下实现。

预测反褶积

勘探地震中，另一个经典的反褶积方法是预测反褶积。已知地震子波 $W\left(t\right)$ 和含有多次波的地震道 $\mathbf{d}^{obs}\left(t\right)$，估计反射系数序列 $r\left(t\right)$。定义如下反问题：

$$\varepsilon\left(\mathbf{r}\left(t\right)\right)=\left\|\mathbf{d}^{obs}\left(t\right)-W\left(t+\alpha\right)\mathbf{r}\left(t\right)\right\|^2 \tag{17}$$

对应的法方程为：

$$C_{ww}^{\alpha}\mathbf{r}=c_{Wd}^{\alpha} \tag{18}$$

其中，$W\left(t+\alpha\right)$ 是时延后的子波，α 代表时延量，$\hat{r}\left(t\right)$ 为估计出的反射系数序列，解法方程(18)可以得到 $\hat{r}\left(t\right)$。

时变反褶积

时变反褶积是勘探地震中的另一个典型问题。首先，时变褶积模型定义为：

$$d\left(t\right)=\int W\left(t-\tau\right)r\left(t\right)d\tau \tag{19}$$

其中，$W\left(t\right)$ 代表时变子波，与时间 t 有关。在吸收衰减介质中波传播时，地震子波显然是时变的。

对应的反问题提为：

$$\varepsilon\left(r\left(t\right)\right)=\left\|d^{obs}\left(t\right)-\int W\left(t-\tau\right)r\left(t\right)d\tau\right\|^2 \tag{20}$$

每个时刻 t 均要对应地求解上述反问题。

对应的，公式(20)的法方程为：

$$C_{ww}^{t}\mathbf{r}=c_{wd}^{t} \tag{21}$$

其中，自相关函数 C_{WW}^{t} 是时变的，与时间 t 有关的，这正是非平稳情形自相关函数应有的情况。

可以用迭代最小二乘近似地实现时变反褶积，迭代最小二乘的计算效率要远高于求解问题(19)和(20)式。

地表一致反褶积

地表一致性反褶积也是勘探地震中典型的反褶积问题。存在不同的地表一致性褶积模型。我认为：(10)式更符合物理实际。地表一致性反褶积的真正目的是消除炮端和检端由于W_s和W_R带来的对同相轴上期望子波的振幅和相位的影响。

从偏移成像和信号处理的角度，不希望来自地下同一稳定反射层的子波出现振幅和形态的变化（剧变），这肯定不是W_+、W_-和反射系数R引起的。

问题时如何达到来自同一地下稳定反射层的子波的振幅和形态的一致性？显然应该是统计方法。根据统计方法，把子波的振幅和形态统一到期望的水平上。

实际中，地表一致性反褶积的具体实现可以有不同的途径。

偏移成像结果的反褶积

当前，对偏移成像结果进行图像反褶积，提高反褶积结果的分辨率和保真度，形成了勘探地震中又一个典型的反褶积问题。我们知道，偏移成像基于(9)式定义的褶积模型，Born 近似下，它是一个线性的褶积模型。定义反褶积对应的线性反问题为，

$$J\left(\delta\mathbf{m}\right)=\left(F_{born}\delta\mathbf{m}-\delta\mathbf{u}^{obs}\right)^T\left(F_{born}\delta\mathbf{m}-\delta\mathbf{u}^{obs}\right) \tag{22}$$

(22)式对应的法方程为，

$$\left(F_{born}^TF_{born}\right)\delta\mathbf{m}=F_{born}^T\delta\mathbf{u}^{obs} \tag{23}$$

$$H\delta\mathbf{m}=\delta\mathbf{m}_{mig} \tag{24}$$

其中，$H=F_{born}^TF_{born}$，$\delta\mathbf{m}_{mig}=F_{born}^T\delta\mathbf{u}^{obs}$。很显然,(24)描述了一种高维褶积关系。即，常规偏移成像结果是地下真像的模糊结果，模糊算子是 Hessian 矩阵。Hessian 的每一行对应地下任一成像点的 PSF。PSF 是个光滑算子，正是 PSF 模糊了地下介质的真像。图像反褶积就是求解(24)式对应的反褶积问题，得到地下介质真像的估计结果。

一般地，用共轭梯度法求解如下反问题，

$$J\left(\delta\mathbf{m}\right)=\left(H\delta\mathbf{m}-\delta\mathbf{m}_{mig}\right)^T\left(H\delta\mathbf{m}-\delta\mathbf{m}_{mig}\right) \tag{25}$$

用 EM 算法同时求解 PSF 和真像$\delta\mathbf{m}$也是值得探索的。

自由地表多次波数据的褶积模型和反褶积

海上自由表面是一个很特殊的物理面，把海上探区地下介质视为一个系统，可以认为它是波场的反馈面。正是由于它的存在，导致了一次反射波的多次反馈。根据这样的认识，可以提出如下的褶积系统。首先定义$\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)=G\left(\mathbf{x}_r,\mathbf{x},\mathbf{x}_s;t\right)*R\left(\mathbf{x}\right)$为地下介质和 Green 函数构成的一次上行波场。用r代表自由表面反射系数。自由表面和地下介质系统构成了一个反馈体系。如图 3 所示。在该反馈体系中传播的地震波可以用如下褶积系统来描述，

$$\begin{aligned}&\mathbf{u}^{obs}\left(\mathbf{x}_r,\mathbf{x}_s;t\right)=\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)+\left\{\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\cdot r\right\}\cdot\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)+\left\{\left[\left\{\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\cdot r\right\}\cdot\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\right]\cdot r\right\}\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)+\cdots\\&=\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\left\{1+\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\cdot r+\left[\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\cdot r\right]^2+\left[\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\cdot r\right]^2+\cdots\right\}\\&=\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\frac{1}{1-\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)\cdot r}\end{aligned}\tag{26}$$

很显然，(26)式定义了一个非线性的褶积系统。

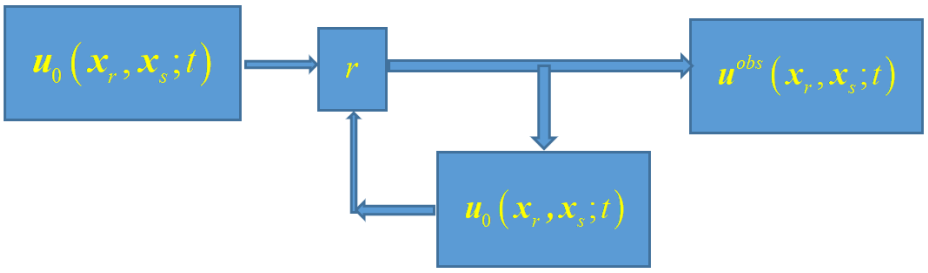


图 3 自由表面、炮检在海面情况项，多次波场生成的反馈机制

已知 $\mathbf{u}^{obs}\left(\mathbf{x}_r,\mathbf{x}_s;t\right)$ 和自由界面反射系数，求解上述非线性褶积系统，可以估计出一次反射波场 $\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)$。通过迭代法可以求解上述非线性问题，得到一次反射波场 $\mathbf{u}_0\left(\mathbf{x}_r,\mathbf{x}_s;t\right)$。

换一个角度，如果区分上下行波场，适应 OBS 观测数据，即利用上下行波场分别估计一次反射波，甚至估计一次波对应的 Green 函数场，同时剥离掉震源子波和自由表面多次波的影响，要引入图 4 表示的反馈褶积模型。

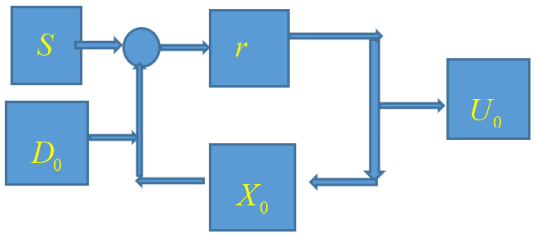


图 4 OBS 情形下，多次波生成的反馈机制

图 4 中，S 代表震源函数；r 依然是自由表面反射系数；X_0 代表一次波对应的 Green 函数场；D_0 是包含多次波的下行波场；U_0 是包含多次波的上行波场。根据图 4 展示的反馈机制，可以写出如下的非线性褶积模型。

下行波场对应的非线性褶积模型为：

$$\begin{aligned}&D_0=S+\left(X_0r\right)S+\left(X_0r\right)^2S+\cdots=S\left(1+X_0r+\left(X_0r\right)^2+\cdots\right)\\&=S\left(\frac{1}{1-X_0r}\right)\end{aligned}\tag{27}$$

上行波场对应的非线性褶积模型为：

$$U_0 = X_0 S + X_0\left(X_0 r\right)S + X_0\left(X_0 r\right)^2 S + \cdots = S\left(X_0 + X_0\left(X_0 r\right) + X_0\left(X_0 r\right)^2 + \cdots\right)$$
$$= S\left(\frac{X_0}{1 - X_0 r}\right) \tag{28}$$

从(27)和(28)可以看出，

$$U_0 = X_0\left(\frac{S}{1 - X_0 r}\right) = X_0 D_0 \tag{29}$$

应该指出，上述(27)~(29)都是在把数据表达在频率空间域导出来的。另外，同样可以看出，由上行波场或下行波场估计一次波对应的 Green 函数场的非线性反褶积的求解，原则上都需要迭代方法。这一点与线性反演理论下的反褶积是显著不同的。

反褶积的进一步发展方向

从前面的讨论可以看出，所谓的褶积与反褶积本质上是一个涵盖非常广泛的概念。从系统的观点看，非常多的问题都可以用褶积关系描述它们的正过程。对应地，可以用反褶积方法来解决相应的具体问题。

当前，线性的褶积和反褶积问题已经非常清楚了。高斯、线性、平稳假设下的无偏和方差最小的解就是反褶积的最佳结果。但是，实际情况下，首先是预测误差的高斯分布很难满足，严格的平稳性也很难满足，非线性的系统也经常出现。我们能做的是依据对目标问题的理解，尽量把问题修改成近似满足高斯、线性、平稳假设的要求，尽可能得到较好的应用效果。

从方法理论的发展角度看，必须尽量突破高斯、线性、平稳假设！稀疏约束反褶积是推广到反射系数非高斯白噪的情形！高维反褶积、噪音非高斯情形下的反褶积、时变反褶积、盲反褶积都是有必要研究的。

非线性情形下的反褶积更是要点问题。水平层状介质下，存在多次波情形下的反射系数估计，就是个典型的非线性反褶积问题。提出合理的非线性褶积模型，直接求解非线性反褶积问题是一个必要的发展发现。AR 模型是线性预测模型（褶积模型）的代表，是时候推进到应用反馈褶积模型（ARMA 模型）进行非线性反褶积的阶段了。

结论与讨论

科学和工程中绝大部分的应用问题，甚至很多社会科学中的很多问题都可归结为一个系统问题。系统的输入和输出之间的关系对应一个褶积模型。系统的输出对应实测的数据。由系统输出，进行反褶积处理，估计系统参数或系统输入，是一大类问题的目的。因此，褶积模型对应十分广泛的应用场景，看起来差异很大的问题都可以归结在褶积与反褶积的理论框架下得到很好地解决。

到目前为止，线性褶积模型及对应的反褶积算法依然是最广泛应用的。（时不变）线性（褶积）模型、高斯预测误差、平稳随机的实测数据是基于线性褶积模型进行反褶积的基本假设。线性褶积模型的反褶积算法都是非常成熟的。总体而言，线性反演理论是线性褶积模型进行反褶积的理论基础。

但是，实际数据处理中，高维褶积与反褶积的计算效率、正则化方法、时变反褶积、盲反褶积的具体实现还是有很多困难的。更为重要的是，实际数据并不满足线性、高斯、平稳假设，基于线性反演理论的反褶积并不能得到很好的效果。我们能做的是引入各种预处理方法，譬如数据加窗、数据去噪、引入广义线性模型等。

本质上讲，各种线性褶积模型都可以认为等价于“AR 模型”。我们对基于“AR 模型”的线性问题的反问题，以及它的推广，譬如 Kalman 滤波、自适应滤波、甚至粒子滤波都已经很清楚了。机器学习中多层感知机的 BP 算法是 LMS 算法的扩展。但是，如果没有更新对褶积模型的认识，仅仅在反褶积算法层次上下

功夫并非是解决具体问题的正确途径。我认为，推进到“ARMA 模型”进行反褶积应该是反褶积问题的重要发展方向。这应该与 RNN 算法由非常密切的关系，遗憾的是，RNN 算法的物理意义不是很明确。更重要的是提出新的、所谓的非线性褶积模型，譬如包含多次波的褶积模型，这就进入了非线性反褶积的阶段了。针对非线性褶积与反褶积的研究一定会大幅度促进地震数据处理的进展。

参考文献

[1]. Aki, K., and Richards, P. G., 1980, Quantitative seismology: W. H. Freeman and Co.
[2]. Amundsen, L., 1993, Wavenumber-based filtering of marine point source data: Geophysics, 58, 1335–1348.
[3]. Lasse Amundsen., 2001, Elimination of free-surface related multiples without need of the source wavelet: GEOPHYSICS, 66, 327–341
[4]. Amundsen, L., and Ikelle, L. T., 2000, Multidimensional signature deconvolution and free-surface multiple elimination of 4C data: 62nd Ann. Mtg., Eur. Assn. Geosci. Eng., L-54. Amundsen, L., Ikelle, L. T., and Martin, J., 2000a, Multiple 2attenuation and P=S splitting of multicomponent OBC data at a heterogeneous sea floor: Wave Motion, 32, 67–78.
[5]. Amundsen, L., and Reitan, A., 1995, Decomposition of multicomponent sea floor data into upgoing and downgoing P- and S-waves: Geophysics, 60, 563–572.
[6]. Amundsen, L., Secrest, B., and Arntsen, B., 1995, Extraction of the normal component of the particle velocity from marine pressure data: Geophysics, 60, 212–222.
[7]. Amundsen, L., Vestby, J., and Berg, L. E., 2000b, Improved multiple attenuation and imaging of OBC data using a wave theory based approach: 62nd Ann. Mtg., Eur. Assn. Geosci. Eng., L-53.
[8]. Antia, H. M., 1991, Numerical methods for scientists and engineers: Tata McGraw-Hill Publ. Co., Ltd.
[9]. Araujo, F. V., Weglein, A. B., Carvalho, P. M., and Stolt, R. H., 1994, Inverse scattering series for multiple attenuation: An example with surface and internal multiples: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1039–1041.
[10]. Barr, F. J., and Sanders, J. I., 1989, Attenuation of water-column reverberations using pressure and velocity detectors in a water-bottom cable: 59th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstract, 653–656.
[11]. Berkhout, A. J., 1999, Multiple removal based on the feedback model: The Leading Edge, 18, 127–131.
[12]. Carvalho, P. M., and Weglein, A. B., 1994, Wavelet estimation for surface-related multiple attenuation using a simulated annealing algorithm: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1481–1484.
[13]. Carvalho, P. M., Weglein, A. B., and Stolt, R. H., 1991, Examples of a nonlinear inversion method based on the T matrix of scattering theory: Application to multiple suppression: 61st Ann. Internat., Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1319–1322.
[14]. Carvalho P M, Weglein A B, Stolt R H. Nonlinear inverse scattering for multiple suppression: Application to real data. Part I[M]//SEG Technical Program Expanded Abstracts 1992. Society of Exploration Geophysicists, 1992: 1093-1095.
[15]. Dragoset, B., and MacKay, S., 1993, Surface multiple attenuation and subsalt imaging: 63rd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1099–1102.
[16]. Enders A. Robinson and Sven Treitel, 2004, Geophysical Signal Analysis : SEG（Reprinted）
[17]. Fokkema, J. T., and van den Berg, P. M., 1990, Removal of surfacerelated wave phenomena: The marine case: 60th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1689–1692.
[18]. Gary Hampson and Greg Szumski ,2020, Down/down deconvolution : 90th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 3099–3103
[19]. Holvik, E., Osen, A., Amundsen, L., and Reitan, A., 1999, On P- and S-wave separation at a liquid-solid interface. J. Seis. Expl., 8, 91–100.
[20]. Ikelle, L. T., 1998, Deghosting and free-surface multiple attenuation of multicomponent OBC data: 68th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1234–1237.
[21]. Ikelle L T. Using even terms of the scattering series for deghosting and multiple attenuation of ocean-bottom cable data[J]. Geophysics, 1999, 64(2): 579-592.
[22]. Ikelle, L. T., Amundsen, L., and Eiken, O., 1997a, Multiple attenuation at primary/multiple interferences: The Troll example: The Leading Edge, 16, 1751–1753.
[23]. Ikelle, L. T., and Jaeger, R., 1997, A sensitivity analysis of inverse scattering multiple attenuation to shallow water: J. Seis. Expl., 8, 331–350.

[24]. Ikelle, L. T., Roberts, G., and Weglein, A. B., 1997b, Source signature estimation based on the removal of first-order multiples: Geophysics, 62, 1904–1920.
[25]. Kennett, B. L. N., 1979, The suppression of surface multiples on seismic records: Geophys. Prosp., 27, 584–600.
[26]. Osen, A., Amundsen, L., and Reitan, A., 1998, Towards optimal spatial filters for multiple attenuation and P/S splitting of OBC data: 68th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 2036–2039.
[27]. Osen A, Amundsen L, Reitan A. Removal of water-layer multiples from multicomponent sea-bottom data[J]. Geophysics, 1999, 64(3): 838-851.
[28]. Osen A, Amundsen L, Reitan A. Approximate plane-wave decomposition and de-multiple processing of point-source line-profile data by the 1-D Fourier transform[J]. Journal of Seismic Exploration, 1999, 8(1): 15-25.
[29]. Osen, A., Amundsen, L., Reitan, A., and Helgesen, H. K., 1996, Removal of water-layer multiples from multicomponent sea-bottom data: 66th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1531–1534.
[30]. Osen, A., Amundsen, L., Secrest, B. G., and Reitan, R., 1995, Wavelet estimation from two marine field measurements by plane wave decomposition: J. Seis. Expl., 4, 345–358.
[31]. Schalkwijk, K. M., Verschuur, D. J., and Wapenaar, C. P. A., 1997, Decomposition of ocean bottom cable (OBC) data: 67th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 8–11.
[32]. Todd K.Moon and Wynn C. Stirling, 2000, Mathematical Methods and Algorithms for Signal Processing: Prentice Hall
[33]. van Borselen, R. G., Fokkema, J. T., and van den Berg, P. M., 1996, Surface-related multiple elimination: Geophysics, 61, 202–210.
[34]. van Borselen, R. G., van den Berg, P. M., and Fokkema, J. T., 1992, Removal of surface-related wave phenomena in the double radon transform domain: 62th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1085–1088.
[35]. Verschuur, D. J., 1992, Adaptive surface-related multiple elimination: Geophysics, 57, 1166–1177.
[36]. Weglein, A. B., 1999a, Multiple attenuation: an overview of recent advances and the road ahead: The Leading Edge, 18, 40–44.
[37]. Weglein A B. How can the inverse-scattering method really predict and subtract all multiples from a multidimensional earth with absolutely no subsurface information?[J]. The Leading Edge, 1999, 18(1): 132-136.
[38]. Weglein, A. B., Gasparotto, F. A., Carvalho, P. M., and Stolt, R. H., 1997, An inverse-scattering series method for attenuating multiples in seismic reflection data: Geophysics, 62, 1975–1989.
[39]. Weglein, A., and Secrest, B. G., 1990, Wavelet estimation for a multidimensional acoustic or elastic earth: Geophysics, 55, 902–913.
[40]. Ziolkowski, A., Parkes, G., and Haugland, T., 1982, The signature of an air gun array: Computation from near-field measurements including interactions: Geophysics, 47, 1413–1421.
[41]. Ziolkowski, A., Taylor, D. B., and Johnston, R. G. K., 1998, Multiple wavefields: separating incident from scattered, up from down, and primaries from multiples: 68th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1499–1502.
[42]. Ziolkowski A M, Taylor D B, Johnston R G K. Marine seismic wavefield measurement to remove sea-surface multiples[J]. Geophysical Prospecting, 1999, 47: 841-870.

复杂地表地震勘探的认识及复杂地表探区地震数据成像处理的观点

王华忠，许荣伟

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】油气地震勘探已经进入了复杂地表、复杂构造、复杂储层和深层目标阶段。“两宽一高”的地震数据采集已经成为地震数据采集的常规要求。高精度和高保真成像也已经是必须的。但是，勘探地震的复杂性主要由近地表因素导致，这个判断在“两宽一高”油气勘探阶段依然没有改变，依然是成立的。近地表因素依然是复杂地表探区的最大障碍。问题的关键是复杂地表引起了剧变的道间时差和极低的信噪比，哪怕初至波也极有可能被噪音严重干扰，这就导致了速度估计精度和地震波成像质量的严重下降，难以解决油气勘探的地质目标问题。本文从复杂山地地震数据的静校正、预处理、速度建模、偏移成像、复杂山地地震数据成像处理的基本原则及处理流程等方面对山地地震勘探问题展开讨论，试图给出解决山地地震勘探问题的合理技术方案。

【关键词】山地地震勘探，复杂地表，剧变道间时差，低信噪比，预处理，速度建模，叠前偏移成像，合理技术方案

1 引言

油气地震勘探已经进入了复杂地表、复杂构造、复杂储层和深层目标阶段，也已经进入了“两宽一高”和 FWI+LS_RTM 为技术代表的时代，追求的是小尺度和隐蔽油气藏的描述，高保真和高分辨的地震波成像是必然的要求，复杂近地表条件影响高保真和高分辨地震波成像。

我们已经清楚地知道：复杂地表区油气地震勘探的真正复杂性是由近地表因素引起的，其它影响因素排在近地表影响因素之后。

我们也已经清楚地知道：地震勘探介质系统可以近似地视为一个线性系统，线性系统的输入信号 $S(\mathbf{x}_s,t)$ 和输出信号 $u(\mathbf{x}_r,t)$ 近似是相干信号，即 $u(\mathbf{x}_r,t)=c\cdot S(\mathbf{x}_s,t)$，$c=|c|e^{i\phi(t)}$。其中，$\mathbf{x}_s$ 是激发震源坐标点；$\mathbf{x}_r$ 是接收波场的坐标点。近似相干的信号总是可以同相位叠加增强的。这是当前地震波偏移成像最本质的物理基础。“两宽一高”观测方式下，地下同一反射点甚至有上千次的覆盖。来自同一反射点的接收子波的同相位叠加是高保真和高分辨地震波成像最基本要求。从地表到中深层的（各向异性）速度模型的正确是对来自同一反射点的接收子波实现同相位叠加的根本保障。近地表速度模型的正确性对接收子波能否同相位叠加的影响最严重。近地表高程（剧烈）变化、低降速带厚度变化和近地表速度的横向变化，引起近地表波场复杂，这是最严重的噪声源，主要的噪声波场包括：面波、面波散射波、一次和多次折射波、折射波产生的散射波、近地表小尺度速度异常引起的散射波等等。然后上述近地表变化还会导致道间时差（也可以称为静校正量）的快速变化。道间时差的快速变化进一步降低观测数据的信噪比。高精度的速度估计严重地依赖高信噪比数据。高质量的地震波成像非常依赖正确的速度建模结果。据此，复杂地表区的地震勘探陷入了一个似乎无法解开的、纠结的闭环！尽管地震勘探已经进入了“两宽一高”和 FWI+LS_RTM 为技术代表的时代，并没有针对复杂地表探区的十分有效的方法技术。

可以看出：复杂地表探区的真正问题在于地震数据的低信噪比！

尽管我们知道了问题的症结所在，但是复杂山地地震数据的低信噪比是固有的。真正的解决问题的措施在于业界专业人员能深入地理解近地表介质情况、理解复杂近地表介质中的波传播过程、通晓噪音产生的机制。在此基础上，基于现有的方法技术设计出最佳的解决问题方案和技术流程。

因此，从复杂山地地震数据的静校正、预处理、速度建模、偏移成像、复杂山地地震数据成像处理的基本原则及处理流程等方面对山地地震勘探问题展开讨论，试图给出解决山地地震勘探问题的合理技术方案。

2 关于静校正的基本观点

静校正是相对于动校正而提出的。动校正是在水平地表、层状介质假设下成立的。动校正是与叠后时

间偏移和叠前时间偏移组合应用的时间域成像方法技术。静校正+NMO/DMO+叠后时间偏移/叠前时间偏移形成了常规时间域处理流程中的核心技术。

近地表高程（剧烈）变化、低降速带厚度变化和近地表速度的横向变化导致了道间时差，引起了静校正问题。静校正的目的很明确，把受道间时差改造的非双曲时距关系恢复成近似双曲的，它是 NMO 的预处理，也是去噪及子波一致性校正的预处理。常规时间域处理流程中，静校正是核心的问题。

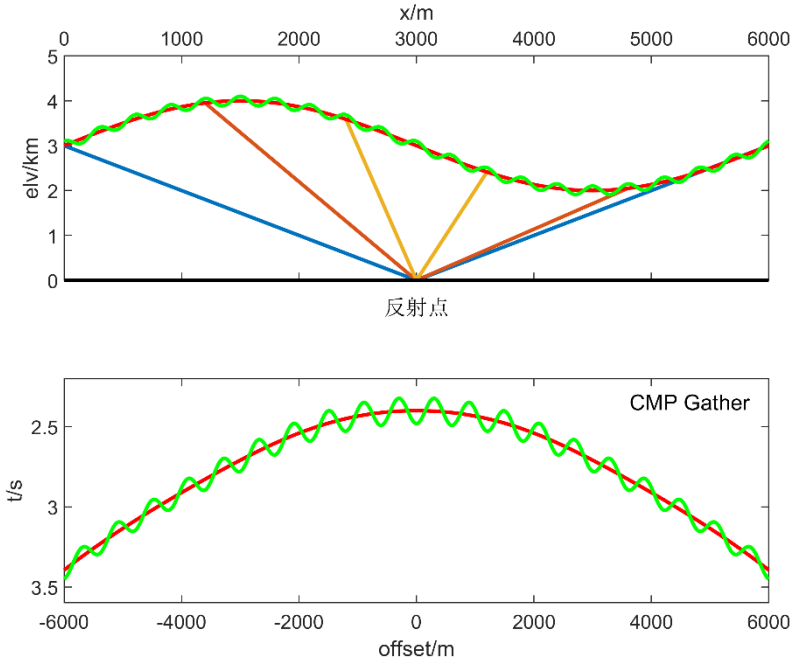


图 1 高波数（短波长）静校正示意图

由于近地表高程（剧烈）变化、低降速带厚度变化和近地表速度的横向变化快慢程度不同、异常体的尺度不同，静校正量有短波长、中波长和长波长之分。有必要给出一些半定量的指标来界定静校正的高、中和低频成分的分界。高、中和低波长静校正量的引入对于实际处理是很有帮助的。

为此，给出关于高程变化和速度变化引起的静校正量的表达。E 代表高程；V 代表速度。风化层厚度变化可以归结为速度的横向变化。S 代表静校正量。

静校正的高波数变化表示为：

$$S_h = E_h + V_h \tag{1}$$

其中，E_h 代表高程快速变化引起的短波长静校正量；V_h 代表近地表速度横向快速变化引起的短波长静校正量；S_h 代表短波长静校正量。

类似地，静校正的中波数变化表示为：

$$S_m = E_m + V_m \tag{2}$$

其中，E_m 代表高程中等程度变化引起的中波长静校正量；V_m 代表近地表速度横向中等变化引起的中波长静校正量；S_m 代表中波长静校正量。

同样地，静校正的长波数变化表示为：

$$S_l = E_l + V_l \tag{3}$$

其中，E_l 代表高程慢变化引起的长波长静校正量；V_l 代表近地表速度横向慢变化引起的长波长静校正量；S_l 代表长波长静校正量。

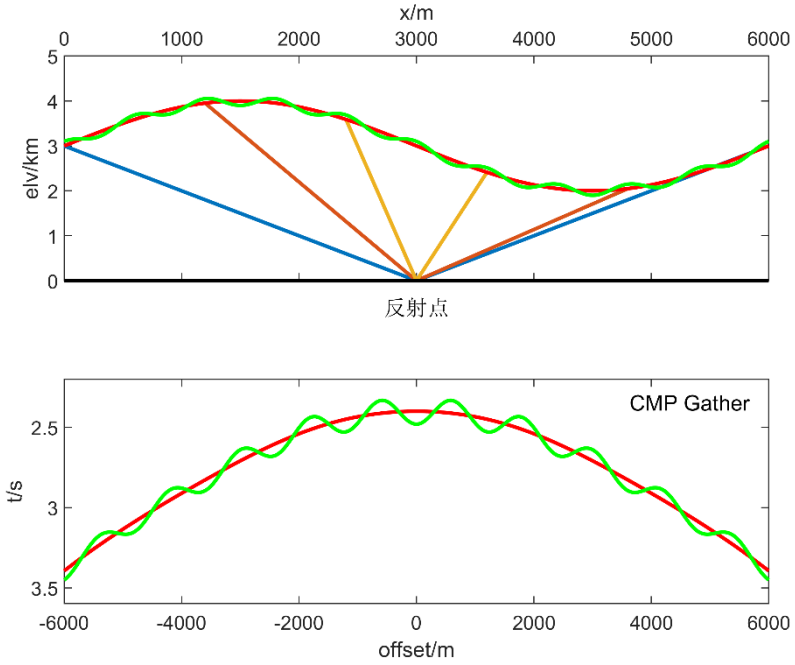


图 2 中波数（中波长）静校正示意图

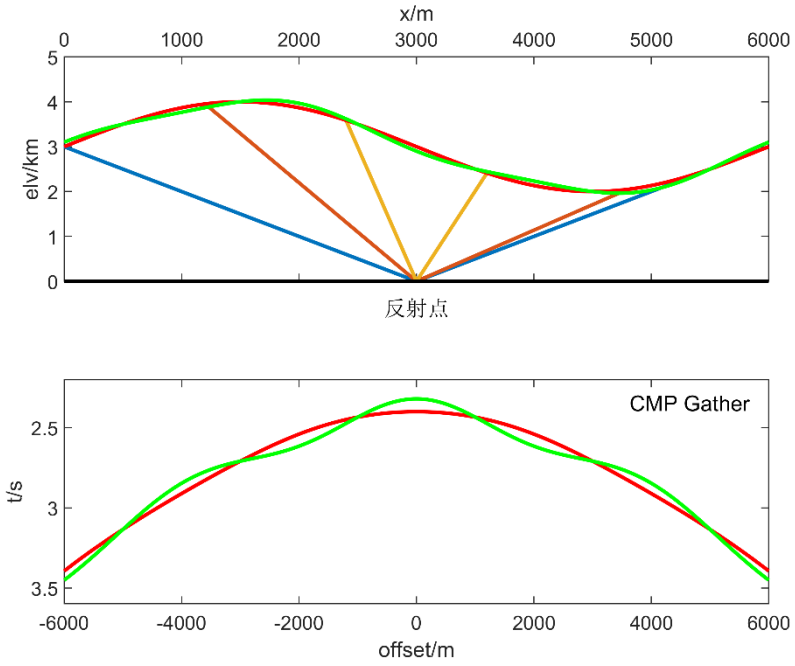


图 3 低波数（长波长）静校正示意图

但到底是地下介质的哪一部分引入了不同的静校正量？这依赖于近地表介质模型。根据地表一致性假设，静校正主要取决于低降速带的形态。风化层底面之上的介质情况定义了低降速带：地表面的几何形态、风化层底面之上的速度变化和低降速带底面几何形态。众所周知，静校正量就是由地表面的几何形态和风化层底面之上的速度变化构成的速度模型计算出来的。其中包括高波数（或高频）变化、中波数（或中频）变化和低波数（或低频）变化。静校正量的这些快慢不同的变化量，对应于近地表速度模型中的快慢不同的变化量。

实际勘探中，近地表速度模型中的快慢不同的变化量如何界定？实际上，这是比较定性的描述！

我认为，山地勘探或地表崎岖地区的勘探中，高程剧变是引起S_h的主要原因。也包括近地表模型中包含的小尺度速度异常体以及风化层厚度的快速变化。高程快速变化引起高频静校正量对成像的影响容易分析。小尺度速度异常体、风化层厚度的快速变化引起的高频静校正对成像的影响分析起来很不容易。事实上，静校正基准面的选择和填充速度也对静校正量的计算结果由比较大的影响。当然也会影响后续的 NMO 叠加和均方根速度估计结果。

本质上讲，时间域处理中引入的地表一致性假设、地表水平且地层水平层状的假设，蕴含了风化层底界也应该是水平的且静校正基准面应该与此一致的假设，在松辽盆地、渤海湾盆地还基本可行，但在西部山地探区已经偏离太多了。

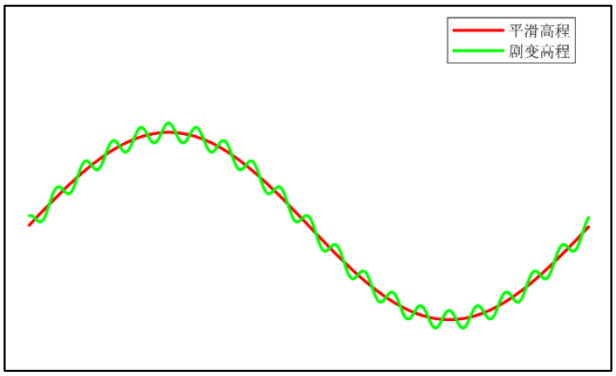


图 4 高程剧变是引起 S_h 的主要原因

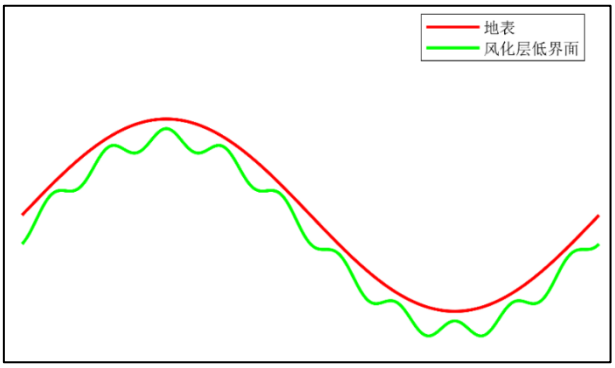


图 5 风化层厚度横向变化是引起 S_h 的另一个主要原因

很多探区，有风化层存在，其厚度在横向上变化很快，这是引起S_h的另一个主要原因。但很多山地探区基岩出露地表，基本上没风化层，或仅局部地区有风化层，这种探区是不存在所谓静校正量的。

关于静校正的具体含义，图 6 展示了它的几何意义。地表一致性假设为：$\vec{x}$ 点处的t_s与 offset 无关，无论什么炮检对的波，在$\vec{x}$ 点处都垂直传到地表。正是上述地表一致性假设，才有了所谓的“静校正”的做法。

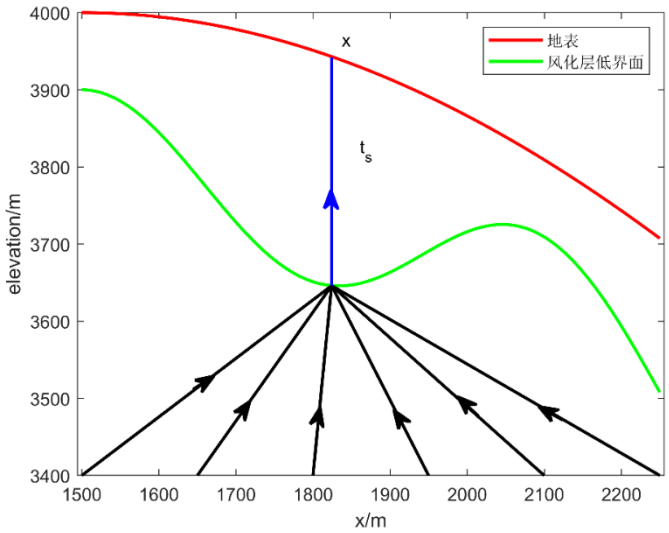


图 6 地表一致性静校正的几何解释

理论上，静校正就是要校正掉高程变化、风化层厚度横向变化和近地表横向变速带来的道间时差。因此，静校正依赖于风化层厚度、近地表速度、地表高程、最终浮动基准面和充填速度。可见，静校正的影响因素也很多!必须深刻理解探区的近地表地质情况才有可能做好静校正。

在复杂地表区做好静校正及以后的 NMO/DMO、叠后时间偏移与叠前时间偏移也不容易！事实上，在复杂地表和存在横向变速的介质情况下，时间域的处理无论如何都很别扭。这是时间域处理的理论假设不符合介质变化的实际所决定的，注定了时间域的处理在山前带探区只能起辅助作用。必须用深度域成像的思维，看待山前带地震数据处理，才能构建合理的处理流程。

3 与复杂地表相关问题的基本观点

陆上地震勘探 80%的问题为复杂近地表条件引起，由高程变化和地表岩性横向变化所引起。但对此说法还要做具体分析。因为复杂地表条件是空间域的（(x,y,z)域），而炮集是时间域的（(x,y,t)域），二者之间的相互关联非常复杂。只有把二者之间的分析清楚与透彻了，后继的数据处理、建模和偏移成像才能真正做到心中有数，有的放矢。

3.1 关于复杂地表条件所引起的噪音

凡不能贡献于当前的一次反射波、散射波、绕射波成像的其他波现象及随机噪音都应被消除掉。从地震波激发、传播与接收的物理过程，结合实测单炮道集看，噪声基本由近地表因素产生；从能量上看，95%以上的噪音由近地表因素产生，多次波噪音在总能量占比中不大（陆上数据！层间多次波!）。面波相关的噪音又占了大部分，压掉这部分噪音是最主要的工作。

地表高程变化（也包括横向变速）引起的道间时差变化进一步降低同相轴（有效信号）的信噪比，使去噪更困难。复杂近地表探区去噪与静校正二者是耦合在一起的，应迭代进行。

3.2 同相轴的光滑性问题

地震信号处理（去噪、规则化等）要求同相轴是光滑的；地下同一反射点、不同 Offset 上反射子波的拉平也要求同相轴是光滑的。这样的要求来自于当前的去噪方法和偏移成像方法都是线性理论下的方法。

同相轴上地震子波的时差跳跃是要预处理掉的。振幅跃变也是要消除掉的。

3.3 复杂地表条件引起的道间地震子波振幅与形态（相位）差异问题

基于统计原则和地表一致性假设进行地震子波的振幅和形态（相位）的一致性校正是必须做的一步工作，尤其是地表一致性振幅处理必须要做。

地表一致性振幅和相位校正又是基于同相轴存在连续性的、信噪比较高的假设条件的。因此，道间时差的消除对去噪和地表一致性振幅与相位校正而言又是优先的。

3.4 静校正量与道间时差的关系

这是两个不同的概念，道间时差包含了静校正量，但不同道集中表现出来的道间时差却不仅仅由地表高程变化和风化层厚度变化所引起，一个道集涉及到的其他区域的速度横向变化也是产生道间时差的重要根源。应把这两个概念（静校正量和道间时差）区分开！

用道间时差的概念可以包容更多的问题。

3.5 静校正的合理做法

静校正量由于地表高程变化和风化层厚度横向变化引起的，是服务于 NMO 叠加的！各种静校正方法及其合理性均奠基于此！静校正方法的目的是消除地表高程变化和近地表风化层厚度变化引起的道间时差，恢复 CMP 道集中同相轴的双曲时距关系。应该注意到风化层内的分层横向变速使得道与道之间的时差变化更复杂。

静校正的做法：

① 层析成像建立风化层底面以上的近地表速度模型，选择充填速度和最终浮动基准面，计算每一点处的静校正量。

这是最基本的静校正方法，去除主要的静校正量。速度建模的方式不同，产生不同名称的静校正方法。譬如折射层析静校正；初至层析静校正等等。都是在建立了风化层底面以上的速度模型后，再计算静校正量的。

② 剩余静校正。

由于高程测量，包括井深测量，以及风化层底面以上的速度建模精确不够等因素，中、高频静校正量消除不干净（长波长静校正量也消除不干净！），存在剩余道间时差。在炮集中，表现为初至到达时中存在跃变；在 Common Offset 道集中，表现为同相轴中存在跃变。这些跃变的存在表明存在剩余的中高频静校正量。这些剩余的中高频静校正量要用统计的方法来检测或估计，不能再依赖于风化层以上的速度模型，凡偏离统计背景的道间时差跃变均表明存在剩余静校正量。有地表一致性初至剩余静校正方法和地表一致性反射剩余静校正方法。

3.6 满足静校正需求的建模

静校正量的计算依赖于风化层底面之上的速度模型。这个模型有几个关键要素：① 风化层底界的几何形态；② 真地表面的几何形态；③ 两个面之间的速度值；④ 两个面之间可能的速度分层结构，它可以被融入速度值中，但能利用分层结构建模的话，建模精度会更高。

这个模型建构时的困难包括：① 大炮数据对建此模型的贡献不大，在此模型中走的射线不多。用折射波进行基于折射波时距关系的建模可能更好些，但很多探区，大炮记录中，不能识别出有效的折射波。② 大炮初至层析，近偏移距道数太少，道间距太大，层析成像建立风化层底面以上的速度模型，误差会很大。③ 风化层底面很难确定。风化层底面的几何形态会引入静校正量的计算误差。

我对构建适用于静校正的风化层底面之上速度模型的观点：① 地表高程面（真地表面）要准确的描述出来；② 风化层底面只能尽可能通过各种信息来确定。要清楚地了解或知道风化层底面几何形态不准时会给静校正量的计算带来的影响；③ 两个面之间速度值的填充，我认为还是要充分地利用近地表调查信息，微测井、露头、岩性变化、高光谱、大炮初至波特征变化等等。把空间信息融合作为重要工具，建立初始速度模型，然后再与初至波波动理论旅行时层析结合，估计速度值。不能过于依赖初至波旅行时信息，这个信息对于估计与建立风化层底面上的速度模型太不充分。

我对风化层底面以上速度模型对静校正影响的看法：① 这个模型不可能是很精确的，静校正量的计算是有误差的。事实上，什么是正确的静校正量都是不定的！② 风化层底面之下与之上速度变化不能太快！风化层面的形态变化不能太剧烈。这个要求本质上是时间域处理成像所需求的，否则时间域处理成像处理流程不能用于这样的探区。但这个假设或要求并没有人指出来（或我没有看到有人指出来），更不会受到重视！正是这个假设或要求区别开了时间域成像处理流程与深度域成像处理流程。如果这个要求得不到满足，即使做了静校正，CMP 道集中时距关系依然不能恢复成近似双曲的时距关系，而 CMP 道集中同相轴满足双曲时距关系是整个时间域处理的根本要求。

3.7 风化层之上速度建模不准对静校正的影响

静校正依赖于风化层厚度、速度、地表高程、最终浮动基准面和充填速度。风化层模型对静校正的影响是最重要的。

如图 7 所示，假设 CMP 点距为 20 米或道间距为 20 米，A、B、C、D、E、F 点处存在偏离背景的速度异常，这些地震道上子波到达时会存在跃变！

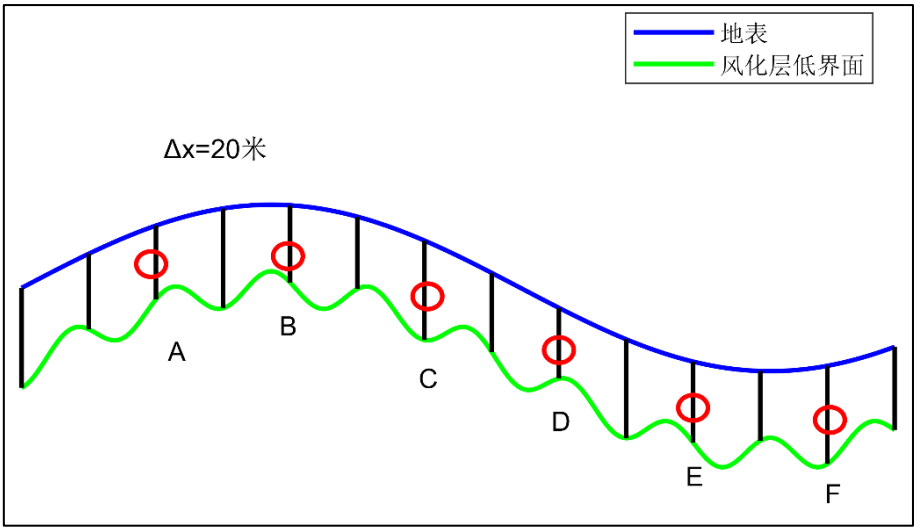


图 7 风化层模型对静校正的影响示意图

但这些尺度在 20 米左右的速度异常是无法由目前的近地表层析反演所把握的。在风化层底面的空变太快（或风化层底面的高程出现跃变）的情况下，若风化层上下速度跃变，此时也等于引入了小尺度速度异常，这也是近地表层析所无法控制的，只能寄希望于风化层上下速度跃变很小。但在大多数山前带探区，这种要求不符合实际。我国东部地区大多符合这种要求。但还要具体分析！东部地区有些探区近地表速度（风化层附近的速度）也横向变化很快！

3.8 地表高程的变化对复杂地表区勘探及对高波数静校正的影响

复杂地表区勘探应对高程剧变引起足够的重视，它对勘探过程、数据品质和静校正都会引起很大的影响。可以罗列如下：① 高程剧变导致施工困难；② 引起面波、直达波散射噪音很强；③ 高程剧变导致炮集质量（信噪比）显著下降。折射波，初至波难以准确识别；④ 高程测量的误差映射在速度模型中导致速度模型不准；⑤ 高程剧变引起剧变的道间时差，进一步降低信噪比；⑥ 高程剧变时，充填速度与真速度差异大，风化层底面之上速度模型建立误差大时，会引入显著的剩余静校正量。

3.9 与成像相关的近地表建模

风化层底面之上的近地表速度模型是应用于静校正的。但，前已分析过，风化层之上的近地表模型仅靠大炮数据更难建立，这个区域中，仅有很少的射线穿过（覆盖），无法用层析反演建立较好的速度模型，需结合其他的信息。这部分速度模型对到达时的影响不能忽视，因为近地表速度低，尽管波在其中走的路径较短，但时间还是比较可观的，对同相位叠加的影响还是比较大的。从图 8 可见，从 S_i 到 R_i 长传播路径上有小尺度的速度异常，对到达时的影响一般不太大，很难引起显著的道间距时差 $(t_i - t_{i+1})$。但近地表速度误差对道间时差的影响却是显著的。这是为什么在山地勘探中，在复杂地表勘探中，特别重视对风化层之上速度模型进行研究的原因。

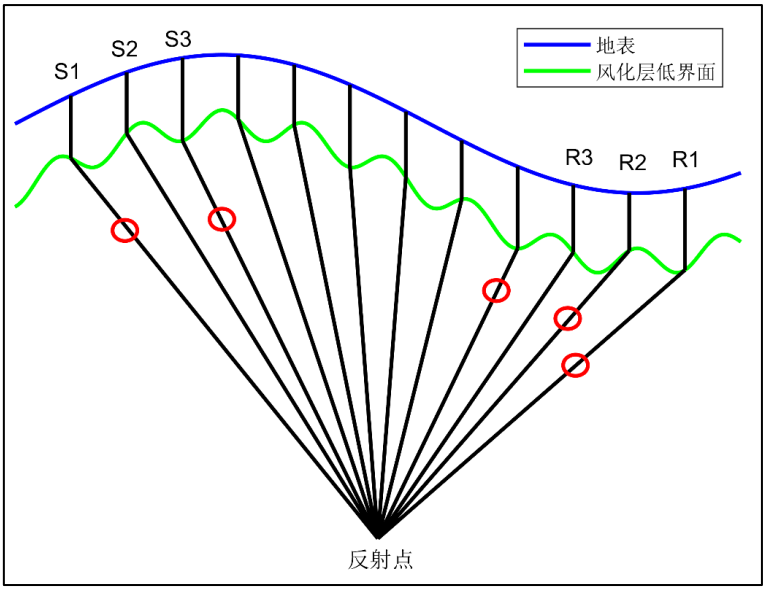


图 8 风化层速度异常与长路径上速度异常体对成像到达时的影响

本质上，若能准确建立真地表面之下的速度模型，加上叠前深度偏移，地震波成像工作可以结束了！关于成像所用的近地表模型的建立，目前主要用透射波波动理论层析成像方法来建立。这是受最近十几年 FWI 方法技术发展所促进的。成像所用的近地表模型的建立见图 9。

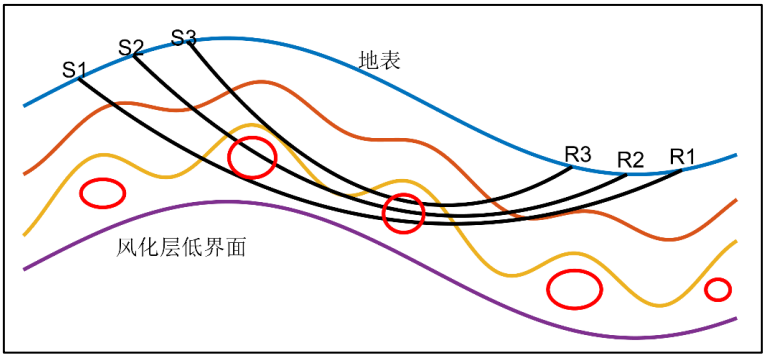


图 9 近地表速度对成像的影响

成像所用的近地表速度当然包括风化层底面之上那一部分，但成像所用的近地表速度模型有更多的透射波（折射+Diving Wave）和反射波穿过，用层析反演，在当前“两宽一高”数据下，应有更高的精度。这是最近十几年来 FWI 技术所促进的。目前的 FWI 就是利用透射波估计更精确的浅层速度模型（0 米-1000 米？）。而过浅那部分（0 米-100 米）的速度模型（譬如风化层底面之上的速度模型）严重缺乏射线穿过的信息，FWI 层析反演的精度很低。因此，我们无法完全抛开静校正。

在当前追求完全深度域成像的过程中，到底静校正还能发挥多少作用？如何发挥作用？通过小平滑面的选择+高（中）频静校正+波动理论透射波层析的技术组合能否把复杂地表地震勘探的地震波成像成功地推进到全深度域？尽管目前的实践支持上述观点，还是有很多问题需要分析！

总之，陆上地震勘探近地表建模是个令人头痛的问题，复杂地表情况下造成严重的成像困难，需综合的解决方案。小平滑面的选择+高（中）频静校正+波动理论透射波层析的技术组合是目前较好的技术方案。还需要不断完善！

3.10 时间域成像处理与深度域成像处理的分界点

我认为：**时间域成像处理与深度域成像处理的分界点**就在于基于风化层层底面之上的速度模型进行静校正后，CMP 道集中同相轴满足双曲时距关系的假设。

我认为：理论上讲，静校正问题就是近地表速度建模的一部分，近地表速度模型能较准确的建立时，不需要进行静校正处理，也不存在所谓的静校正。

但仅在地表观测的地震勘探方式导致了仅仅利用大炮数据无法准确地建立近地表速度模型。因此，只能进一步引入水平层状介质的假设，从而产生了静校正+NMO/DMO+叠后时间偏移/叠前时间偏移的常规流程。说好，这是充分利用地震层状介质分布的特点和地表数据观测的特点；说不好，时间域处理就是无奈之举！

深度域处理流程，同样困扰于如何建立比较准确的速度模型，尤其是近地表速度模型。尽管深度域成像处理也有自己的困难，但毕竟引入的假设较少，普适性更强。

我个人认为：深度域成像处理是大势所趋，无法改变。这是必须走的路线！核心是尽量建立比较准确的、包括风化层底面之上的近地表速度模型。

4 复杂山地地震数据的预处理

4.1 小平滑地表面

山地数据处理的正确做法等于“小平滑面+全深度域成像”是毋庸置疑的了。

小平滑面的基本选择逻辑是：基于小平滑面的短（中）波长静校正后，其它的波传播旅行时都可以由目前的速度分析与建模方法进行较为准确的预测得到很好的速度建模结果，进而得到很好的 PSDM 成像结果。

当前的建模方法不能预测的短波长旅行时跃变，应该通过选择小平滑面+静校正从数据中消除掉。PSDM 从小平滑面开始。

平滑面的选择与速度建模技术的精度是相关的，与数据观测质量和近地表调查数据的丰富程度品质是关联的。也与去噪音的方法技术联系在一起。

没有什么定量的方法指导小平滑面的选择。只有一些原则可以遵循：① 尽量贴近真地表；② 尽量满足去噪的要求以及地表一致性处理等的要求。③ 尽量满足速度分析方法的要求。

很多时候我们只能消除短波长量后，用同相轴的光滑程度来判断。

高程静校正+初至剩余静校正就可以了。不要过多地考虑反射剩余静校正，留给 CIG 道集的后处理来解决其余的剩余高频静校正量。

上述讨论能做到的话，“真”走时问题应该是有保证的。可以满足从小平滑面上开展 PSDM 的要求。

4.2 关于近地表速度模型与静校正在 PSDM 处理流程中的地位问题

理论上，此时必须有地表一致性假设的成立，才有理论上合理的静校正。必须重视这个理论假设的含义，以及实际勘探中这个假设的不成立程度及其带来的影响。

近地表速度模型+静校正基准面+充填速度决定静校正量。这其中包括：短波长、中波长、长波长静校正量。

对 PSDM 流程，能被速度模型描述的旅行时，都应在 PSDM 中进行描述。尽量不改变数据中的旅行时（因此仅仅消除数据中的高波数成分）、尽量不改变真地表（因此选贴近地表的小平滑面）。

近地表与浅层速度建模都是只能描述真实地层的背景速度。因为近地表与浅层的速度低，对 PSDM 成像的影响是不能忽视的。这就说明为什么高精度成像越来越重视近地表与浅层速度建模。

不要再过于关注长波长静校正的重要性。只要近地表速度模型正确，可以在 PSDM 中解决长波长静校正问题。

中短波长静校正才是对去噪、地表一致性处理、高分辨率叠加等的关键影响因素。必须在高程静校正+初至剩余静校正+反射剩余静校正+CIG 道集拉平中各阶段仔细处理。应该体会到：反射剩余静校正可以合

并到 CIG 道集拉平过程中。

5 复杂山地地震数据处理中的速度建模

5.1 表层速度建模

大炮数据对于（譬如 0-100 米的）表层速度模型建立是不能提供有效信息的。采样太粗了！频率太低了！波长太长了！近地表速度变化尺度太小了！

近地表建模必须靠独立的数据采集系统：微测井、小折射、微 VSP、针对地表的极高精度地震勘探、露头等等信息。近地表建模是个综合信息挖掘和融合的问题。

目前对近地表建模能把综合信息挖掘和融合问题解决好，建立其背景速度成分已经很好了。短波长部分和部分短中波长部分静校正（对应速度场的小尺度变化成分）只能通过时间域的“静”校正方法解决。

如果真能建立从地表到深层的准确的偏移速度模型，从真地表开始偏移就行了。绝大部分的预处理根本不必做了！但这种理想情况大多不符合实际！

表层速度建模并不是单纯的地表微测井加散乱数据插值问题，如果表层速度建模没有考虑地质因素（譬如露头、结构等），而是纯数学的插值操作，这样的速度建模是没有道理的。

5.2 浅层速度建模

利用 2500 米偏移距内的数据建立的近地表速度模型计算静校正量逻辑上是说不通的。特别要注意静校正的“地表一致性”假设。把风化层底界面以下的速度场也纳入静校正，会破坏炮集中的时距关系，降低后续的速度分析和成像精度，导致吃力不讨好！

把近地表速度模型定义为风化层底界以上的速度场，且满足“地表一致性”假设，是不能破坏的原则。

实际上，很多探区风化层并不明显存在。这就要仔细把握时间域处理与深度域处理的界限。并尽早进入深度域处理。

浅层速度严格地讲可以说是从风化层底界开始到折射波和 Diving Wave 能较为准确描述的区域。这显然与观测系统和数据品质是关联的。

当前，宽方位长偏移距数据逐渐增多的情况下，浅层速度部分也在逐渐扩张，有与中层速度重叠的趋势。我们当然期望重叠得越多越好，重叠区域的照明更充分，速度建模的精度会更高。

近地表速度模型的建立中，露头和地质信息的约束要远比数学插值方法重要。与速度有关的信息的收集、整理、数字化、表达、融合是近地表建模的重要环节。

浅层速度建模的最主要波现象就是直达波、折射波和 Diving Wave，最主要的波场就是初至波。还包括初至波后续的早至波。最好是波动理论的初至波层析反演。陆上数据中，初至波和早至波的信噪比太低是波动理论的初至波层析反演技术很好应用的阻力，尤其在可控震源激发情形下。但是，这是必须要克服的技术困难。应该是绕不过去的。

表层与浅层的一体化建模当然是有必要的。有一个好的表层初始模型，会很有利于浅层速度建模的精度提升。

浅层速度建模主要依赖的信息源就是初至波。但是，很多探区初至波识别与走时检测严重受制于信噪比。此时，其它信息源的补充就非常必要。对浅层而言，能有的信息源依然是测井和地质先验认识。

复杂地区速度建模是一个标准的反问题。由于叠前道集的信息不足，必须借用先验信息，降低速度建模的不准确性。测井和地质认识是最重要的信息。关键是如何用好这些信息。

近地表的各向异性在某些探区是不可忽视的。近地表的各向异性取决于特定的地质特征。中深层介质地震各向异性往往是尺度各向异性的，各向异性参数由各向异性介质下的时距关系进行估计。逻辑上讲，近地

表各向异性参数也是如此。但是，近地表照明比中深层更不均匀，更要依赖测井和岩石地球物理来给出合理的近地表各向异性参数。

5.3 中深层速度建模

对于山地或崎岖地表区的地震勘探，充分利用初至波（早至波）,进行波动理论的到达时反演，建立浅层速度模型是速度建模中不可逾越的环节。海上勘探成像质量的提升主要得益于长偏移距数据中初至波（Diving Wave 和浅层折射波）的应用。陆上勘探中，由于数据中初至波信噪比太低、可控震源炮集初至难以识别等等问题，初至波波动理论层析的应用不是很多。但必须对该项技术重视起来。

反射波速度建模还是中深层速度建模的主要技术。它的基本逻辑：① 反射界面和层速度值共同构成速度场，二者是耦合的。反射界面的合理性在速度建模中往往有更重要的作用。地质知识的约束主要在于提供合理的反射结构。② 速度值的估计，（类）CMP 道集的利用是第一位的，初始反射波速度模型还是要由基于（类）CMP 道集的拉平来判断。在（类）CMP 道集速度建模中，要充分考虑构造的合理性。低信噪比情形下要充分利用超道集的概念。这对陆上低信噪比数据是必须的。比较复杂的探区，由一般 CMP 道集，到 PSTM（类）CMP 道集，再过渡到 PSDM（类）CMP 道集，由 NMO 速度分析过渡到剩余速度分析。把剩余速度修改放在前述整个迭代过程的后面阶段。③ 关于界面和速度值的解耦，建议用 Map Migration 的思想，把时间域界面映射到深度域界面，构建起大致的深度域反射结构，在此基础上进行迭代细化。④ CIG 道集层析速度反演建模，是目前最常用的反射波层析建模技术。关键点：道集质量；层位拾取；RMO 测量，尤其是 RMO 测量，决定速度建模的精度。层位拾取也是引入反射结构约束的环节。

简单评价一下立体层析：好处是引入了波传播方向的逼近，比仅仅靠到达时逼近，反演精度会更高一些。缺点是：参数空间变大，收敛更慢。立体层析有自己的特点，但优势也不是很显著。仅仅 CGG 公司（法国的一批研究人员在坚持这种做法），其它公司和大学并没有跟进。

时间域处理仅仅适用于缓变地表和缓变地层情况下，否者 CMP 道集速度会误导。提供的初始速度不能保证后续的层析反演收敛。

地表和地层横向变化较快的情形下，只有（类）CMP 道集 NMO 提供的初始速度才是比较合理的。

5.4 针对表层的高速砾石问题

高速砾岩区是天山北缘，也许任何山口冲积扇存在的区域，出现的特殊地质分布。还有地表第四系沉积覆盖下的冲积河道，也表现为高速砾岩体的特征。高速砾岩区出现在浅层，层内没有明显分界面，速度高是典型特征。对付这样的孤立异常体的措施还是很多的。但是，对速度建模而言，关键是识别出高速砾岩体的边界形态。高速砾岩体内速度值的精确估计，没有单一的方法，只能靠各种技术组合：钻井+测井（浅井 VSP）提供对速度分布的基本认识；反射波速度扫描提供背景速度；反射层析提供更精确的速度。接近地表，大炮观测过于稀疏是根本问题。建议布置针对高速砾岩体的高密度小排列观测方式，目前浅层高分辨地震勘探已经成为非常重要的地表调查方式。不能指望大炮观测能很好地解决浅层问题。

引入对山口冲积扇的地质知识，非常有助于建立较好的冲积扇体速度模型。

5.5 速度建模 Gamma 质控

用 Gamma（偏移速度与真速度的比值）谱进行速度建模的质量控制，是道集拉平与否的一种量化展示。

但是，道集拉平并非速度模型正确的最合适标准。

速度模型是由界面+层速度值构成的。也可以用网格模型来表达速度模型。但是，网格模型进行速度估计会极大地增加多解性。加入界面的合理约束会极大提高速度建模的精度，这正是地质知识介入速度建模的理论依据。

速度估计是典型的反问题，多解性是反问题的典型特征。道集拉平不是反问题唯一解的判断标准，只是合理解的标准。

高低速、薄厚层组合形成的地层，具备一定的统计效应，使得某种等效速度可以计算出合适的走时，可以实现偏移聚焦，可以进行不错的拉平，但等效速度不等于真速度。

正确的速度如何得到的？充分的信息：叠前炮集中包含的与速度估计有关的信息+与速度有关的先验信息。反问题本质就是多解的，我们能做的就是充分利用信息，降低多解性。地震勘探的本质就是这样，能做的就是提高认识，工作扎实，把成像尽可能做好。并没有什么灵丹妙药!

5.6 速度建模应与地质认识相结合

层速度的变化是有趋势的。单个地层内的速度趋势应该是平稳变化的。

地质、沉积和岩性的概念始终要体现在建模过程中。

数据品质差、速度场复杂时，速度建模问题必须引入合理的约束，不能仅靠数学算法。

6 复杂山地地震数据处理中的偏移成像

6.1 时间偏移与深度偏移的认识

时间偏移和深度偏移的分界：成像射线（Imaging Ray）发生严重偏折。预示着时间偏移的成像点位置与真像点之间出现严重偏离。时间偏移的像被严重畸变。地下介质发生较大横向变化时，时间偏移是不能用的。

山地或崎岖地表勘探，尽量不要再用 PSTM。

地下构造复杂，横向变速严重，更不能用 PSTM。

偏移成像质量判别的一个基本逻辑：构造越简单，越符合地质逻辑，正确的可能性越大。

6.2 成像的合理性

成像合理有很多定性标准：地质逻辑合理（地层不交叉、接触关系合理而清晰），构造简单；不过偏移，不欠偏移；聚焦程度好，同相轴不发虚；断层面清晰等等。

建议做一些正演，看看炮集逼近程度。进一步判断速度的合理性。

引入井信息的判断也是必要的。

速度正确性的判决问题本来是成像工作的重要环节。目前，没有受到充分重视！

7 复杂山地地震数据处理的基本原则

在复杂探区构造合理的速度建模有一些基本原则：从探区构造简单的区域开始，用时间域的（类）CMP 速度分析构建初始速度模型。包括层位的解释和层速度值的确定。根据地质逻辑，把构造简单区域的地质分层和层速度信息向复杂区域延展（一般地，一个探区中复杂区域占比总是少的）。比较复杂的区域用 PSDM+（类）CMP 道集，结合延展过来的信息建立初始速度模型。据此建立比较合理和精确的初始速度模型。

好的初始速度模型才能保证后续的 CIG 道集的层析速度反演的迭代收敛性。

地质构造信息、沉积信息和岩性信息是速度建模所依赖的重要的先验信息。

由粗到细的建模原则，大的层位和层速度可以控制住下一级尺度上的速度变化范围。特别主要大的速度异常体的几何形态的描述，譬如火山岩分布区域、高速砾岩区域、岩盐存在的区域等。

由浅到深的原则。

由于叠前数据不足以使当前的速度建模技术获得满足成像要求的偏移速度模型，建模问题就是一个艺

术问题，是不断迭代逼近对真实地下地质情况的认识的过程。

仅仅靠数学算法是不行的，复杂构造区，必须引入先验信息，才有可能建立较好的深度速度场。

所有的反问题都是：信息不足情形下的最佳决策问题。包括速度场的建立，井位的确定都是如此。

事情的本质就是这样的，技术人员能做的就是提升自己的综合能力，提升对方法技术的理解，把成像问题做好。不存在灵丹妙药。

8 山前带地震数据处理基本观点

理论上看，时间域成像处理的核心技术：NMO/DMO+叠后时间偏移/叠前时间偏移 NMO/DMO 都是假设炮检在一个水平面上，不存在横向变速。

实际上，地表非水平已经破坏了上述理论假设。而且地表、浅层、中层和深层中的横向变速是普遍存在的。时间域成像，至少从理论上看是不适合用于山地地震数据处理的。

理论上，只要有准确的地表、浅层、中层，到深层的速度模型，结合叠前深度偏移方法，从真地表面开始，进行叠前深度偏移成像，构造成像问题应该可以得到完满的解决。

山地地震数据成像处理，深度域成像处理流程是必须的。至少从理论上看，各种关键技术的基本假设，与复杂地表和复杂构造情况，没有显著的矛盾。但是，地表附近的速度模型特别缺乏合适的叠前地震数据进行估计与建模。因此，引入小平滑面作为速度建模域成像的基准面是必须的。依据这个小平滑面，消除高频及部分中频道集时差。

如果可能的话，再做一些剩余的高、中频静校正处理，譬如初至剩余静校正和反射波剩余静校正。

若还有余下的高、中频剩余静校正，在叠前成像道集上进一步消除。

预处理方面，以小平滑面为基础进行静校正处理，去噪和地表一致性校正也要基于此面进行。小平滑面的选择也要考虑保证去噪能较好地进行。地表相关噪音的消除比较复杂，多域、高维去噪是基本原则。主要依赖噪音和信号的特征，选择不同的去噪技术。

山地勘探炮集中的噪音特征变化多端，即便相邻的单炮道集，噪音特征也变化很大。无论如何，基本的噪音特点为：局部线性；强振幅；频散；低频；假频；野值。基于上述特征设计或选择滤波器是没有问题的。但是，噪音特征与道间时差（的消除）有紧密关系，静校正和去噪要迭代进行。先保证初至（初至受噪音干扰最小!）剩余静校正做得比较好，然后再去噪。球面扩散校正后再去噪。

9 山前带地震数据处理流程

尽管从理论上看，时间域处理不适合用于山前带地震数据处理中，但横向变速比较小的情况下，时间域处理给出的成像，也包括速度模型，可以给深度域成像处理提供借鉴。

9.1 时间域处理流程

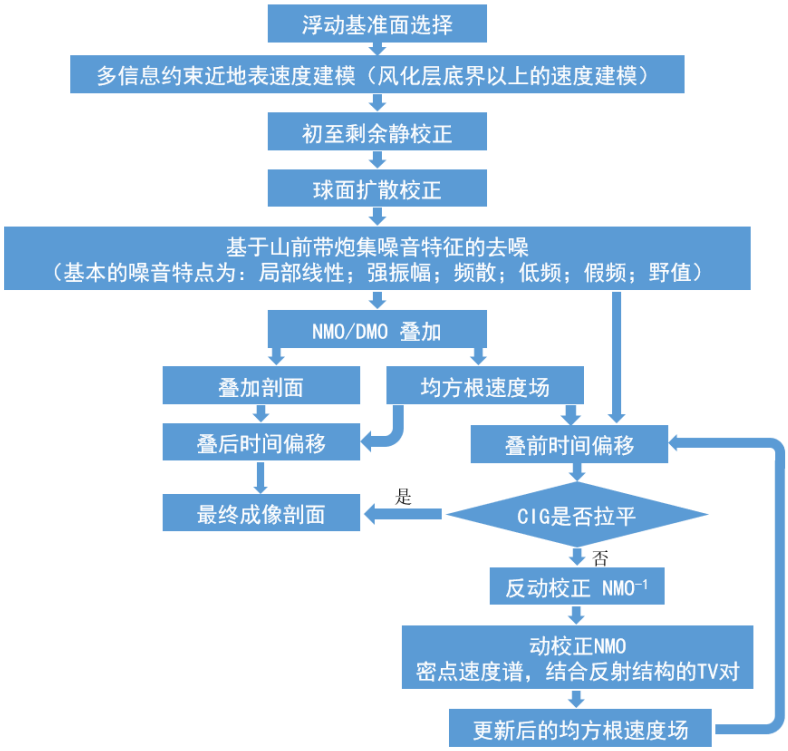


图 10 时间域处理流程示意图

9.2 深度域处理流程

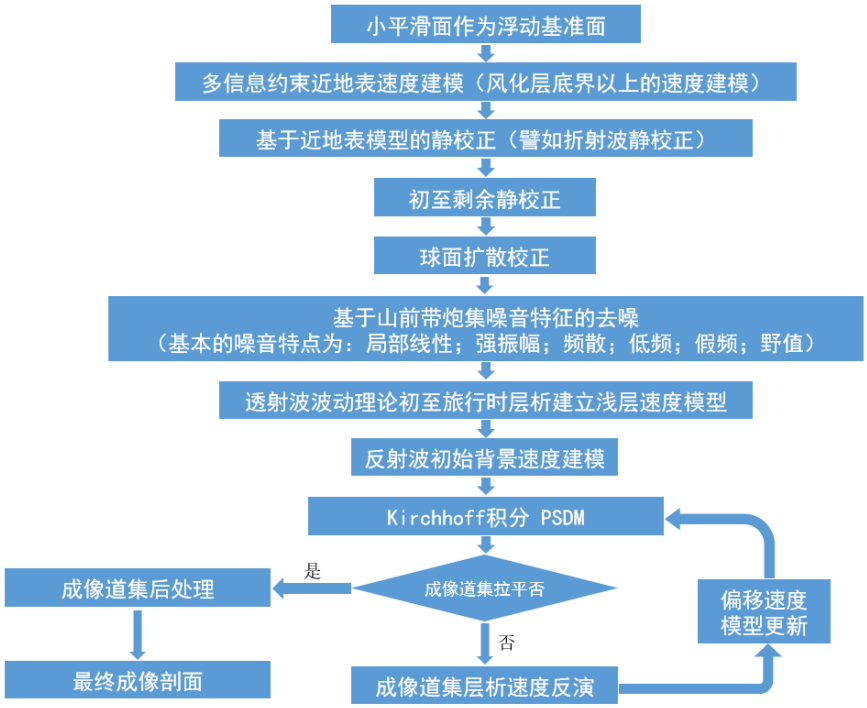


图 11 深度域处理流程示意图

10 结论与讨论

油气地震勘探已经进入了所谓的复杂地表、复杂构造、复杂储层和深层目标阶段。“两宽一高”的地震数据采集已经成为了地震数据采集的常规要求。高精度和高保真成像也已经是必须的。但是，近地表因素导致了勘探地震的复杂性，近地表因素依然是复杂地表油气探区的最大障碍。而且，复杂地表引起了剧变的道间时差和极低的信噪比，哪怕初至波也极有可能被噪音严重干扰。

事实上，复杂地表区的勘探是很特殊的，叠前地震数据中剧变的道间时差和极低的信噪比使得当前所谓的高精度成像技术根本不能用于复杂地表探区的成像处理中。在复杂地表区地震波成像的根本逻辑要靠来自地下同一反射点的、高覆盖的观测子波的同相位叠加。但是，这要依赖于宽方位、长偏移距的地震观测数据提供的高精度的（各向异性）速度建模结果。而高精度的速度建模依赖于高信噪比的叠前地震数据。遗憾的是，信噪比低（极低）正是复杂地表地震数据的典型特征。这也印证了为什么剧变的道间时差和极低的信噪比成为了复杂地表地震勘探的真正障碍。

我国的复杂地表探区的油气勘探是无国外经验可以借鉴的。西方先进国家很少花大功夫在复杂地表区展开油气勘探，没有积累出可以让我们借鉴的经验。

我们已经知道，打开复杂地表探区油气勘探的突破口在于提升对近地表地质问题的认识。地表高程变化、基岩出露地表的情况、近地表风化层情况、近地表岩性变化情况、地表岩性对激发和接收的影响等等的认识是必须要掌握的信息。野外观测以及相关的近地表调查，都要充分地关注上述信息，通过各种各样的观测手段尽可能多的获取关于近地表地质和岩性变化的认知，为后续的数据预处理和成像提供尽可能多的近地表结构和速度变化的信息。复杂山地数据处理，不能像海上高品质数据成像处理那样依赖 FWI 这样的高端成像技术，而是要在地震地质认识的指导下，充分地理解复杂地表探区的噪音、道间时差、近地表速度建模、中深层速度建模、同相位叠加偏移成像的本质，组合出合理、合适、有效的成像处理流程，追求得到一个最佳的成像结果。

把复杂地表探区油气勘探描述成一个信息不足情况下的最佳决策问题是合适的。充分利用各种信息，提升速度建模的精度，进而提高偏移成像的精度，最终提高油气勘探的成功率和效益。

参考文献

[1]. Cox, Michael J. G., 1999, Statistic Correction for Seismic Reflection Survey; SEG
[2]. Butler, D. K., 2005, Near - Surface Geophysics; SEG
[3]. Al-Husseini, M. I., Glover, J. B., and Barley, B. J., 1981, Dispersion patterns of the ground roll in eastern Saudi Arabia: Geophysics, 46, 121–137.
[4]. Anno, P. D., 1983, Statics solutions and event correlations: two critical aspects of shear-wave exploration: 53rd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 347–349.
[5]. Beavis, F. C., 1985, Engineering geology: Blackwell Scientific Publications, Inc.
[6]. Booker, A. H., Linville, A. F., and Wason, C. B., 1976, Long wavelength static estimation: Geophysics, 41, 939–959.
[7]. Clymer, R., Hirshorn, B., and McEvilly, T. V., 1986, Rainfall controlled near-surface traveltime changes: Presented at the 3rd Ann. Summer Res. Workshop, Soc. Expl. Geophys.
[8]. Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, 20, 68–86.
[9]. Domenico, S. N., 1982, Acoustic wave propagation in air-bubble curtains in water—part I: History and theory: Geophysics, 47, 345–353.
[10]. Ewing, M., Crary, A. P., and Thorne, A. M., Jr., 1934, Propagation of elastic waves in ice, part I: Physics, 5, 165–168.
[11]. Ewing, M., and Worzel, J. L., 1948, Long-range sound transmission, in Propagation of sound in the ocean: Geol. Soc. Am., Mem. 27.
[12]. Friedmann, V., 1988, Near-surface sand effects: First Break, 6, 83–93.
[13]. Gardener, R., 1984, Seismic refraction as a tool in evaluation of rock quality for dredging and trenching: The Leading Edge, 3, No. 12, 68–74.
[14]. Haeni, F. P., 1986, Application of seismic refraction methods in groundwater modeling studies in New England: Geophysics, 51, 236–249.
[15]. Herd, L., 1990, Seismic acquisition in the loess hills of China: 60th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 958–961.

[16]. Hildebrant, G., 1990, Acquisition in the transition zone of the Cook Inlet, in 1989 SEG workshop on data acquisition in problem areas: The Leading Edge, 9, No. 9, 32.
[17]. Hoke, J. H., 1967, Refraction prospecting in the Rub' Al Khali, in Musgrave, A. W., Ed., Seismic refraction prospecting: Soc. Expl. Geophys., 493–504.
[18]. Hunter, J. A., Pullan, S. E., Burns, R. A., Gagne, R. M., and Good, R. L., 1984, Shallow seismic reflection mapping of the overburden-bedrock interface with the engineering seismograph—some simple techniques: Geophysics, 49, 1381–1385.
[19]. Johnston, D. H., 1989, Recent advances in exploitation geophysics: The Leading Edge, 8, No. 9, 22–28.
[20]. Jones, J. L., Leslie, C. B., and Barton, L. E., 1958, Acoustical characteristics of a lake bottom: J. Acoust. Soc. Am., 30, 142–155.
[21]. Jones, J. L., Leslie, C. B., and Barton, L. E., 1964, Acoustic characteristics of underwater bottoms: J. Acoust. Soc. Am., 36, 154–157.
[22]. King, G. A., Dunlop, K. N. B., and Graebner, R. J., 1988, Surface seismic monitoring of an active water flood: 58th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 177–179.
[23]. Lester, O. C., Jr., 1932, Seismic weathered or aerated surface layer: AAPG Bull., 16, No. 12, 60–64.
[24]. Levin, F. K., 1980, Seismic velocities in transversely isotropic media, II: Geophysics, 45, 3–17.
[25]. Levin, F. K., 1984, Anisotropy due to bedding—a computer study: Geophys. Prosp., 32, 187–197.
[26]. Meekes, J. A. C., Scheffers, B. C., and Ridder, J., 1990, Optimization of high-resolution seismic reflection parameters for hydrogeological investigations in the Netherlands: First Break, 8, 263–270.
[27]. Meissner, R., and Theilen, F., 1984, Attenuation of seismic waves in sediments: Proc. 11th World Petrol. Congr.: John Wiley & Sons, Inc., 2, 363–379.
[28]. Miller, R. D., and Steeples, D. W., 1990, A shallow seismic reflection survey in basalts of the Snake River Plain, Idaho: Geophysics, 55, 761–768.
[29]. Monk, D. J., Flud, R., and Plumlee, M., 1994b, Evaluation of sea bottom features from seismic calibration data—a case history: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 865–867.
[30]. Rahn, P. H., 1986, Engineering geology: an environmental approach: Elsevier Science Publ. Co., Inc.
[31]. Ramsden, C. R. T., 1985, The application of high resolution seismic processing to low relief structures—Harriet oil accumulation: Expl. Geophys., 16, 335–348.
[32]. Rice, J. A., Krohn, C. E., and Houston, L. M., 1991, Shallow near-surface effects on seismic waves: 61st Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 747–749.
[33]. Sams, M. S., Worthington, M. H., King, M. S., and Khanshir, M. S., 1993, A comparison of laboratory and field measurements of P-wave anisotropy: Geophys. Prosp., 41, 189–206.
[34]. Schatz, P. M., and Lindblade, P. V., 1986, Refraction statics: key to removing statics problems in shallow water areas: 56th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 277–279.
[35]. Selem, A. M., 1955, Reflection survey in rough topography: Geophys. Prosp., 3, 246–257.
[36]. Sheriff, R. E., 1991, Encyclopedic dictionary of exploration geophysics: Soc. Expl. Geophys.
[37]. Steeples, D. W., Miller, R. D., and Black, R. A., 1990, Static corrections from shallow-reflection surveys: Geophysics, 55, 69–775.
[38]. Tatham, R. H., 1982, VP/VS (ratio of seismic compressional and shear-wave velocities) and lithology: Geophysics, 47, 336–344.
[39]. Thralls, H. M., and Mossman, R. W., 1952, Relation of seismic correction to surface geology: Geophysics, 17, 218–228.
[40]. Uhrig, L. F., and van Melle, F. A., 1955, Velocity anisotropy in stratified media: Geophysics, 20, 774–779.
[41]. Wiest, B., and Edelmann, H. A. K., 1984, Static corrections for shear wave sections: Geophys. Prosp., 32, 1091–1102.
[42]. Wilson, W. D., 1960a, Speed of sound in sea water as a function of temperature, pressure, and salinity: J. Acoust. Soc. Am., 32, 641–644.
[43]. Wilson, W. D., 1960b, Equation for the speed of sound in sea water: J. Acoust. Soc. Am., 32, 1357.

高维局部数据体中线性信号预测基本理论与方法

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】若干线性结构（可以视为局部平面波）飘在具有不同概率分布特征的局部高维数据体中是地震数据分析的核心概念模式。对局部高维数据体中的线性结构进行最佳预测，从而解决去噪、数据规则化、Deblending 等问题，是地震数据处理中非常基本的环节。这包括模型驱动的方法和数据驱动的方法。前者是由选定的局部平面波基函数的线性叠加来表示局部高维数据体中包含的信号；后者由数据矩阵（张量）分解的方法推断局部高维数据体中包含的线性结构。本文比较全面地分析了频率-空间域高维 Wiener 滤波方法、自相关矩阵及 Hankel 矩阵正交分解方法（SSA 方法）、高维线性 Radon 变换方法（高维 Beamforming 方法）和张量分解方法的基本理论，为进行局部高维数据体中线性信号预测及各种应用奠定理论基础。

【关键词】局部高维数据体，线性结构，最佳预测，高维 Wiener 滤波方法，高维 SSA 方法，高维线性 Radon 变换方法，张量分解方法，去噪与数据规则化

1 引言

高精度地震波成像对噪声压制的要求越来越高，譬如 FWI/LS_RTM 这些高精度成像的代表性方法都是在高斯噪音假设下导出的。

从实测地震数据中压制相干或非相干或随机噪音，本质上是把实测数据视为随机过程，并对随机过程进行建模。建立实测数据满足的概率分布模型是理论本质，但具体的方法技术并不追求建立实测数据的概率分布模型，而是假设已知这个概率分布模型（常常假设它是 Gauss 分布模型）的条件下建立测试数据中所含信号的表达模型。为简单计，进一步假设数据中所含的信号可以用线性模型（线性基函数的叠加）来表达。

因此，我们提出：勘探地震数据分析中，l 个局部平面波飘在满足不同统计分布特征的随机噪音中是线性信号预测的基本出发点。线性去噪的本质，即对 n 维局部数据块中包含的 l 个局部平面波进行最佳预测，实现线性去噪、数据规则化等目的。最佳的意义体现在 Bayes MAP 估计或最大似然估计下。

数学语言表达的线性去噪的本质：认为 n 维局部数据块中包含有 l 个线性基函数为特征的信号，并叠加上满足 Gauss 分布的随机噪音，在 Bayes 估计理论下对 l 个线性基函数进行最佳预测，即可实现线性去噪和数据规则化等目的。

实现上述线性信号预测和噪声压制目的的典型方法技术包括：（高维）Wiener 滤波方法、（高维）线性 Radon 变换方法（也可称为高维 Beamforming 方法）、针对 Hankel 矩阵的 SSA（Singular Spectrum Analysis）方法。张量分解方法也是越来越常用的数据驱动的线性信号预测方法。

本文的目的就是全面地分析频率-空间域高维 Wiener 滤波方法、自相关矩阵及 Hankel 矩阵正交分解方法（SSA 方法）、高维线性 Radon 变换方法（高维 Beamforming 方法）和张量分解方法的基本理论，为进行局部高维数据体中线性信号预测及各种应用奠定理论基础。

2 一维信号的傅里叶谱分析（FSA）与奇异谱分析（SSA）的等价性证明

一维实测数据可视为由不同频率的谐波叠加而成的信号，再加上满足不同概率分布的随机噪音所构成。数学上可记为 $\boldsymbol{d}(t)=S(t)+\boldsymbol{\eta}(t)$。实测数据 $\boldsymbol{d}(t)$ 被视为随机过程。

谐波叠加形式的一维信号可表示为：

$$\boldsymbol{S}(t)=\sum_{k=1}^{K}\boldsymbol{\varphi}_k(t)=\boldsymbol{\Psi}\boldsymbol{I} \tag{1}$$

其中，$\boldsymbol{\Psi}=[\boldsymbol{\varphi}_1(t)\quad \boldsymbol{\varphi}_2(t)\quad \cdots\quad \boldsymbol{\varphi}_K(t)]$ 代表一组谐波基函数向量形成的矩阵，$\boldsymbol{\varphi}_k(t)$ 代表不同频率和幅值的

谐波基函数，$\boldsymbol{I}$ 代表标准等权叠加系数形成的单位矩阵。

$S(t)$ 的自相关函数（矩阵）由如下的外积形成：

$$\begin{aligned}&\boldsymbol{S}^H(t)\cdot\boldsymbol{S}(t)=\boldsymbol{I}^H\boldsymbol{\Psi}^H\boldsymbol{\Psi}\boldsymbol{I}\\&=\boldsymbol{I}^H\begin{bmatrix}(\varphi_1,\varphi_1) & (\varphi_1,\varphi_2) & \cdots & (\varphi_1,\varphi_K)\\(\varphi_2,\varphi_1) & (\varphi_2,\varphi_2) & \cdots & (\varphi_2,\varphi_K)\\\vdots & \vdots & \ddots & \vdots\\(\varphi_K,\varphi_1) & (\varphi_K,\varphi_2) & \cdots & (\varphi_K,\varphi_K)\end{bmatrix}\boldsymbol{I}=\boldsymbol{I}^H R\boldsymbol{I}=\begin{bmatrix}F_1^2(\omega) & & & \\ & F_2^2(\omega) & & \\ & & \ddots & \\ & & & F_K^2(\omega)\end{bmatrix}\end{aligned}\tag{2}$$

上式是 Parseval 关系的另一种形式。也证明了一维信号自相关函数的分解相当于找到了数据中所含的谐波成分，即 $F(\omega_k)\neq 0$ 的成分。这就在一维情形上证明了信号的 Fourier 谱分析（FSA）与奇异谱 $\left(F^2(\omega_k)\right)$ 分析（SSA）的等价性。

对于含噪声的实测数据，有如下的情形：

$$\begin{aligned}&\boldsymbol{d}^H(t)\cdot\boldsymbol{d}(t)=\boldsymbol{I}^H\boldsymbol{\Psi}^H\boldsymbol{\Psi}\boldsymbol{I}+\boldsymbol{\eta}^H\boldsymbol{\eta}=\\&=\begin{bmatrix}F_1^2(\omega) & & & \\ & F_2^2(\omega) & & \\ & & \ddots & \\ & & & F_K^2(\omega)\end{bmatrix}+\begin{bmatrix}\sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2\end{bmatrix}\\&=\begin{bmatrix}F_1^2(\omega)+\sigma^2 & & & \\ & F_2^2(\omega)+\sigma^2 & & \\ & & \ddots & \\ & & & F_K^2(\omega)+\sigma^2\end{bmatrix}\end{aligned}\tag{3}$$

此时，假设噪音 $\boldsymbol{\eta}(t)$ 是 Gauss 白噪声。很显然，理论上讲，上述分析并不用限于一维情形，n 维情形同样如此！不过高维情形缺乏上述一维情形下简洁的矩阵关系而已。

3 局部高维数据体中线性信号的表达及反问题提法

对于二维及以上的局部高维数据体情形，显然已不可能用自相关函数（矩阵）分解进行信号分析。首先，我们认为二维及以上的局部高维数据体是由若干个局部平面波飘在满足不同统计特征的随机噪音中形成的。因此，二维及以上的局部高维数据体中的信号表达显然可用局部平面波的叠加来进行。以二维情形为例，把 (x,t) 域中的信号变换到 (x,ω) 域，即 $S(x,t)\xrightarrow{FT}S(x,\omega)$。$(x,\omega)$ 域中的信号可以用局部平面波基函数的叠加来逼近，

$$S(x,\omega)=\sum_{l=1}^{L}S(p_x^l)e^{ik_x\cdot x}=\sum_{l=1}^{L}S(p_x^l)e^{i\omega p_x^l\cdot x}=\begin{bmatrix}\boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_L\end{bmatrix}\cdot\begin{bmatrix}s_1\\s_2\\\vdots\\s_L\end{bmatrix}=\boldsymbol{\Psi}\cdot\boldsymbol{S}\tag{4}$$

其中 $s_l=S(p_x^l)$，$\boldsymbol{\varphi}_l=e^{i\omega p_x^l\cdot x}$。很显然这就是线性 Radon 变换（或 Beamforming）的正问题。

推广到三维数据 $\mathbf{d}(x,y,t)$ 情形，显然有下面讨论的关系。首先还是进行 Fourier 变换，把时空域三维数据 $\mathbf{d}(x,y,t)$ 变换到频率空间域 $\mathbf{d}(x,y,\omega)$，即 $\mathbf{d}(x,y,t)\xrightarrow{FT}\mathbf{d}(x,y,\omega)$。$(x,y,\omega)$ 域中的信号可以用局部平面波基函数的叠加来逼近，即：

$$\mathbf{d}(x,y,\omega)=\sum_{l=1}^{L}S(\boldsymbol{p})e^{i\mathbf{k}\cdot\boldsymbol{x}}=\sum_{l=1}^{L}S(\boldsymbol{p}^l)e^{i\omega\boldsymbol{p}^l\cdot\boldsymbol{x}}=\begin{bmatrix}\boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_L\end{bmatrix}\cdot\begin{bmatrix}s_1\\ s_2\\ \vdots\\ s_L\end{bmatrix}=\boldsymbol{\Psi}\cdot\boldsymbol{S} \tag{5}$$

其中 $s_l=S(\boldsymbol{p}^l)$，$\boldsymbol{\varphi}_l=e^{i\omega\mathbf{p}^l\cdot\mathbf{x}}$，$\mathbf{k}=\left(k_x,k_y\right)$，$\boldsymbol{p}=\dfrac{\mathbf{k}}{\omega}=\left(\dfrac{k_x}{\omega},\dfrac{k_y}{\omega}\right)=\left(p_x,p_y\right)$，$\boldsymbol{x}=\left(x,y\right)$。这是三维数据块 $\mathbf{d}(x,y,t)$ 情形下的线性 Radon 变换（或 Beamforming）的正问题。应该注意到，(5)式并非一个矩阵方程，仅仅是三维以上频率空间域局部数据体中所含信号的形式表达。

很显然，(4)和(5)式可以推广到 $\mathbf{d}(x_1,x_2\ldots x_N,t)$ 的 N 维情形，如果有必要的话！但 N 维情形下 $\boldsymbol{\Psi}$ 表示的这个基函数空间太大了，导致参数空间 $\boldsymbol{S}$ 变得巨大，Radon 变换（或 Beamforming）反问题计算效率会很低。因此高维 Radon 变换（或 Beamforming）反问题必须引入先验约束，预知 $s_l=S(p_x^l)$ 的大致范围。事实上，高维线性预测本质都存在同类问题（也可称为维度灾难）!

从实测数据 $\boldsymbol{d}(t)=S(t)+\boldsymbol{\eta}(t)$ 中最佳估计其中局部平面波成分可以提为各种不同正则化约束下的反问题。譬如，

$$\mathcal{J}(\mathbf{S})=\frac{1}{2}\left\|\boldsymbol{d}(t)-\Psi\cdot\mathbf{S}\right\|^2+\lambda\left\|\mathbf{S}\right\|^p \tag{6}$$

其中，$p=2,1,0$。当 $p=2$ 时，引入的是 L2 范数的约束；$p=1$ 或 $p=0$ 时引入的是稀疏约束。注意，$p=0$ 时定义的范数并不是真正的范数。

4 高维（>2）频率空间域数据的 SSA 分析的基本思想分析

首先简述一下 PCA（Principal Component Analysis）方法的基本思想。假设实测数据向量（可以是 1D 时间域或空间域也可以是 2D 频率空间域的随机向量），记为 $\mathbf{X}$。假设其均值等于零，即 $E[\mathbf{X}]=0$。存在一个向量 $\mathbf{q}$，$\mathbf{X}$ 向 $\mathbf{q}$ 的投影记为，

$$\boldsymbol{\sigma}=\mathbf{X}^T\mathbf{q} \tag{7}$$

注意，$\mathbf{X}^T$ 是行向量，$\mathbf{q}$ 是列向量。很显然，$\boldsymbol{\sigma}$ 应是随机变量。其方差可以写为，

$$E\left[\boldsymbol{\sigma}^2\right]=E\left[\left(\mathbf{X}^T\mathbf{q}\right)^T\left(\mathbf{X}^T\mathbf{q}\right)\right]=E\left[\mathbf{q}^T\mathbf{X}\mathbf{X}^T\mathbf{q}\right]=\mathbf{q}^TE\left[\mathbf{X}\mathbf{X}^T\right]\mathbf{q}=\mathbf{q}^T\mathbf{R}_{\mathbf{XX}}\mathbf{q} \tag{8}$$

其中，$\mathbf{R}_{\mathbf{XX}}$ 是随机向量 $\mathbf{X}$ 的自相关矩阵。基于(8)式可以定义一个关于 $\mathbf{q}$ 的方差函数，

$$\Phi(\mathbf{q})=E\left(\boldsymbol{\sigma}^2\right)=\mathbf{q}^TR\mathbf{q} \tag{9}$$

对(9)式求极大对应的 $\mathbf{q}$：

$$\mathbf{q}=\min_{\mathbf{q}\in Q}\Phi(\mathbf{q})=\min_{\mathbf{q}\in Q}\left(\mathbf{q}^T\mathbf{R}_{\mathbf{XX}}\mathbf{q}\right)\quad s.t.\quad \mathbf{q}^T\mathbf{q}=1 \tag{10}$$

就是 PCA 的基本思想。

用 Lagrange 乘子法求解(10)式定义的优化问题，

$$\mathbf{q}=\min_{\mathbf{q}\in Q}\Phi(\mathbf{q})=\min_{\mathbf{q}\in Q}\left(\mathbf{q}^T\mathbf{R}_{\mathbf{XX}}\mathbf{q}+\lambda\left(\mathbf{q}^T\mathbf{q}-1\right)^2\right) \tag{11}$$

求解(11)式可以导出：

$$\mathbf{R}_{\mathbf{XX}}\mathbf{q}=\lambda\mathbf{q} \tag{12}$$

很显然，向量$\mathbf{q}$应该是自相关矩阵的特征向量，λ是与之对应的特征值。PCA 方法的具体实现利用了自相关矩阵是正定对称矩阵，正定对称矩阵可以分解为：

$$\mathbf{R_{XX}}=\mathbf{Q\Lambda Q}^T \tag{13}$$

其中，$\mathbf{Q}$是标准正交阵，即$\mathbf{QQ}^T=\mathbf{I}$。正交矩阵Q的所有列向量包含了要找的所有的特征向量$\mathbf{q}$。这就实现了 PCA。

据此，可以提炼出 PCA 的基本思想：把随机向量X投影到正交矩阵$\mathbf{Q}$的列向量形成的标准正交基向量构成的坐标系中，每个基函数向量上的投影坐标构成了所谓的方差值σ。反之，通过(10)式确定的向量$\mathbf{q}$，即正交阵$\mathbf{Q}$的列向量中的某一个。显然，借助于自相关矩阵$\mathbf{R_{XX}}$的正交分解是确定向量$\mathbf{q}$的更快速的方法。SSA 就是通过自相关矩阵$\mathbf{R_{XX}}$的正交分解实现的 PCA。二者本质上是等价的。但 SSA 还可以借助 Hankel 矩阵的 SVD 分解推广到更高维的空间中。

SSA（Singular Spectrum Analysis）方法在勘探地震（以及高维图像处理）中的线性信号预测方面有广泛的应用。因为 SSA 是数据驱动的方式，用滑动窗选数据块的方式，可以适用缓平稳信号情形。FSA 严格地讲只能适用于广义平稳信号情况下。下面分析一下 SSA 背后的本质含义。

对于频率空间域数据块$\tilde{\mathbf{d}}(x,\omega)$，若其中仅含有一个同相轴（只有一个局部平面波），则有：

$$\begin{aligned}\tilde{\mathbf{d}}(x,\omega)&=(d_0,d_1,\cdots,d_N)\\&=(s,s\cdot\varphi(\Delta x),s\cdot\varphi(2\Delta x),\cdots,s\cdot\varphi(N\Delta x))\\&=(s,s\theta,s\theta^2,\cdots,s\theta^N)\end{aligned} \tag{14}$$

其中，$\varphi(k\Delta x)=e^{i\omega pk\Delta x}=\theta^k$，$\theta=e^{i\omega p\Delta x}$,单频数据$\left(s,s\cdot\varphi(\Delta x),s\cdot\varphi(2\Delta x),\cdots,s\cdot\varphi(N\Delta x)\right)$之间表现出类似 AR 模型的数据间的预测关系。

实质上，一阶微分方程

$$\frac{\partial u}{\partial x}=a\frac{\partial u}{\partial t} \tag{15}$$

也表现了相邻数据点之间存在的可线性预测关系，因为上式可以写为$dt/dx=a$。很显然，方程(15)中的系数就是$u(x,t)$中线性结构的斜率。

把频率空间域中局部数据块$\mathbf{d}(x,\omega)$排成 Hankel 矩阵，有

$$\mathbf{D}=\begin{bmatrix}d_0&d_1&\cdots&d_k\\d_1&d_2&\cdots&d_{k+1}\\\vdots&\vdots&\ddots&\vdots\\d_{n-k}&d_{n-k+1}&\cdots&d_n\end{bmatrix}=\begin{bmatrix}s&s\theta&\cdots&s\theta^k\\s\theta&s\theta^2&\cdots&s\theta^{k+1}\\\vdots&\vdots&\ddots&\vdots\\=s\theta^{n-k}&s\theta^{n-k+1}&\cdots&s\theta^n\end{bmatrix} \tag{16}$$

可见，该矩阵是秩为 1 的矩阵，对应于只有一个基函数。因此，如此把频率空间域中局部数据块$\tilde{\mathbf{d}}(x,\omega)$排成 Hankel 矩阵，对 Hankel 矩阵进行 SVD 分解，可以保证估计出的基函数向量是正确的。

若频率空间域中局部数据块$\tilde{\mathbf{d}}(x,\omega)$中存在$L$条同相轴，即：

$$\begin{aligned}\tilde{\mathbf{d}}(x,\omega)=(d_0,d_1,\cdots,d_n)=[&(s_1+s_2+\cdots+s_L)+(s_1\theta_1+s_2\theta_2+\cdots+s_L\theta_L)\\&+(s_1\theta_1^1+s_2\theta_2^2+\cdots+S_L\theta_L^2)+\cdots+(s_1\theta_1^n+s_2\theta_2^n+\cdots+s_L\theta_L^n)]\end{aligned} \tag{17}$$

则它对应的 Hankel 矩阵应是：

$$\mathbf{D}=\begin{bmatrix} d_0 & d_1 & \cdots & d_{\ k} \\ d_1 & d_2 & \cdots & d_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n-k} & d_{n-k+1} & \cdots & d_n \end{bmatrix}=\begin{bmatrix} s_1 & s_1\theta_1^1 & \cdots & s_1\theta_1^k \\ s_1\theta_1 & s_1\theta_1^2 & \cdots & s_1\theta_1^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_1\theta_1^{n-k} & s_1\theta_1^{n-k+1} & \cdots & s_1\theta_1^n \end{bmatrix} + \begin{bmatrix} s_2 & s_2\theta_2^1 & \cdots & s_2\theta_2^k \\ s_2\theta_2 & s_2\theta_2^2 & \cdots & s_2\theta_2^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_2\theta_2^{n-k} & s_2\theta_2^{n-k+1} & \cdots & s_2\theta_2^n \end{bmatrix}+\cdots\cdots+\begin{bmatrix} s_L & s_L\theta_L^1 & \cdots & s_L\theta_L^k \\ s_L\theta_L & s_L\theta_L^2 & \cdots & s_L\theta_L^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_L\theta_L^{n-k} & s_L\theta_L^{n-k+1} & \cdots & s_L\theta_L^n \end{bmatrix} \tag{18}$$

可见，若频率空间域数据块 $\tilde{\mathbf{d}}(x,\omega)$ 中存在 L 条同相轴，对 Hankel 矩阵进行 SVD 分解，同样可以得到对应基函数。即找到其中对应的线性信号（局部平面波）。但是，(18)式是对数据中包含的线性结构的矩阵表达，并非是 SVD 的分解结果（$\mathbf{D}=\mathbf{Q}\Lambda\mathbf{Q}^T$）。SVD 结果仅仅是把数据向量 $\tilde{\mathbf{d}}(x,\omega)$ 投影在单位正交矩阵 $\mathbf{Q}$ 的列构成的正交坐标系中。构建一个共轭的投影关系，可以重构数据本身。理论上，可以实现完全的重构。

但是，SSA 并没有把矩阵 D 进行(18)式的展开，产生对数据中线性结构的一对一的分解。因此，SSA 中重构数据时，线性结构的个数 L 是未知的，它严重影响重构精度的参数。

对于三维以上的数据，譬如局部三维数据 $\mathbf{d}(x,y,t)\xrightarrow{\text{FT}}\tilde{\mathbf{d}}(x,y,\omega)$，更需要必须排成 Hankel 矩阵的形式，进行 SVD 分解。但是，本质上，三维情况下已经不存在(17)和(18)式的定义矩阵形式。下面从地震数据分析的角度，探讨频率空间域中局部三维数据 $\tilde{\mathbf{d}}(x,y,\omega)$ 对应的 Hankel 矩阵的物理意义。假设 $\tilde{\mathbf{d}}(x,y,\omega)$ 中仅含有一个同相轴，二维数据中线性结构可以表示为：

$$\tilde{\mathbf{d}}(x,y,\omega)=\mathbf{D}=\begin{bmatrix} S & S\theta_y & S\theta_y^2 & \cdots & S\theta_y^{ny} \\ S\theta_x & S\theta_x\theta_y & S\theta_x\theta_y^2 & \cdots & S\theta_x\theta_y^{ny} \\ S\theta_x^2 & S\theta_x^2\theta_y & S\theta_x^2\theta_y^2 & \cdots & S\theta_x^2\theta_y^{ny} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S\theta_x^{nx} & S\theta_x^{nx}\theta_y & S\theta_x^{nx}\theta_y^2 & \cdots & S\theta_x^{nx}\theta_y^{ny} \end{bmatrix} \tag{19}$$

记上式为 $\mathbf{D}=(\mathbf{d}_x^0,\mathbf{d}_x^1,\cdots,\mathbf{d}_x^{ny})$，其中 $\mathbf{d}_x^i(x,y,\omega),i=0,1,2,\cdots,ny$ 代表矩阵 $\mathbf{D}$ 的列向量。此时，也可以认为 $\mathbf{D}$ 是一个向量。把该向量再次排成 Hankel 矩阵，有

$$\tilde{\mathbf{d}}(x,y,\omega)=(\mathbf{d}_x^0,\mathbf{d}_x^1,\cdots,\mathbf{d}_x^{ny})=\begin{pmatrix} \mathbf{d}_x^0 & \mathbf{d}_x^1 & \cdots & \mathbf{d}_x^k \\ \mathbf{d}_x^1 & \mathbf{d}_x^2 & \cdots & \mathbf{d}_x^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{d}_x^{ny-k} & \mathbf{d}_x^{ny-k+1} & \cdots & \mathbf{d}_x^{ny} \end{pmatrix} \tag{20}$$

其中，上述矩阵中，每一元素（向量）又可以排成一个 Hankel 子矩阵。

$$\mathbf{d}_x^0=\begin{bmatrix} d_{0,0} & d_{1,0} & d_{2,0} & \cdots & d_{k,0} \\ d_{1,0} & d_{2,0} & d_{3,0} & \cdots & d_{k+1,0} \\ d_{2,0} & d_{3,0} & d_{4,0} & \cdots & d_{k+2,0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{nx-k,0} & d_{nx-k+1,0} & d_{nx-k+2,0} & \cdots & d_{nx,0} \end{bmatrix}=\begin{bmatrix} s & s\theta_x & s\theta_x^2 & \cdots & s\theta_x^k \\ s\theta_x & s\theta_x^2 & s\theta_x^3 & \cdots & s\theta_x^{k+1} \\ s\theta_x^2 & s\theta_x^2 & s\theta_x^2 & \cdots & s\theta_x^{k+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s\theta_x^{nx-k} & s\theta_x^{nx-k+1} & s\theta_x^{nx-k+2} & \cdots & s\theta_x^{nx} \end{bmatrix}$$

$$\mathbf{d}_x^1=\begin{bmatrix} d_{0,1} & d_{1,1} & d_{2,1} & \cdots & d_{k,1} \\ d_{1,1} & d_{2,1} & d_{3,1} & \cdots & d_{k+1,1} \\ d_{2,1} & d_{3,1} & d_{4,1} & \cdots & d_{k+2,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{nx-k,1} & d_{nx-k+1,1} & d_{nx-k+2,1} & \cdots & d_{nx,1} \end{bmatrix}=\begin{bmatrix} s\theta_y & s\theta_x\theta_y & s\theta_x^2\theta_y & \cdots & s\theta_x^k\theta_y \\ s\theta_x\theta_y & s\theta_x^2\theta_y & s\theta_x^3\theta_y & \cdots & s\theta_x^{k+1}\theta_y \\ s\theta_x^2\theta_y & s\theta_x^2\theta_y & s\theta_x^2\theta_y & \cdots & s\theta_x^{k+2}\theta_y \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s\theta_x^{nx-k}\theta_y & s\theta_x^{nx-k+1}\theta_y & s\theta_x^{nx-k+2}\theta_y & \cdots & s\theta_x^{nx}\theta_y \end{bmatrix}$$

$$\mathbf{d}_x^2=\begin{bmatrix} d_{0,2} & d_{1,2} & d_{2,2} & \cdots & d_{k,2} \\ d_{1,2} & d_{2,2} & d_{3,2} & \cdots & d_{k+1,2} \\ d_{2,2} & d_{3,2} & d_{4,2} & \cdots & d_{k+2,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{nx-k,2} & d_{nx-k+1,2} & d_{nx-k+2,2} & \cdots & d_{nx,2} \end{bmatrix}=\begin{bmatrix} s\theta_y^2 & s\theta_x\theta_y^2 & s\theta_x^2\theta_y^2 & \cdots & s\theta_x^k\theta_y^2 \\ s\theta_x\theta_y^2 & s\theta_x^2\theta_y^2 & s\theta_x^3\theta_y^2 & \cdots & s\theta_x^{k+1}\theta_y^2 \\ s\theta_x^2\theta_y^2 & s\theta_x^2\theta_y^2 & s\theta_x^2\theta_y^2 & \cdots & s\theta_x^{k+2}\theta_y^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s\theta_x^{nx-k}\theta_y^2 & s\theta_x^{nx-k+1}\theta_y^2 & s\theta_x^{nx-k+2}\theta_y^2 & \cdots & s\theta_x^{nx}\theta_y^2 \end{bmatrix}$$

$$\mathbf{d}_x^{ny}=\begin{bmatrix} d_{0,ny} & d_{1,ny} & d_{2,ny} & \cdots & d_{k,ny} \\ d_{1,ny} & d_{2,ny} & d_{3,ny} & \cdots & d_{k+1,ny} \\ d_{2,ny} & d_{3,ny} & d_{4,ny} & \cdots & d_{k+2,ny} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{nx-k,ny} & d_{nx-k+1,ny} & d_{nx-k+2,ny} & \cdots & d_{nx,ny} \end{bmatrix}=\begin{bmatrix} s\theta_y^{ny} & s\theta_x\theta_y^{ny} & s\theta_x^2\theta_y^{ny} & \cdots & s\theta_x^k\theta_y^{ny} \\ s\theta_x\theta_y^{ny} & s\theta_x^2\theta_y^{ny} & s\theta_x^3\theta_y^{ny} & \cdots & s\theta_x^{k+1}\theta_y^{ny} \\ s\theta_x^2\theta_y^{ny} & s\theta_x^2\theta_y^{ny} & s\theta_x^2\theta_y^{ny} & \cdots & s\theta_x^{k+2}\theta_y^{ny} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s\theta_x^{nx-k}\theta_y^{ny} & s\theta_x^{nx-k+1}\theta_y^{ny} & s\theta_x^{nx-k+2}\theta_y^{ny} & \cdots & s\theta_x^{nx}\theta_y^{ny} \end{bmatrix}$$

由(20)式可以看出，当仅有一个同相轴时，矩阵 $\tilde{\mathbf{d}}(x,y,\omega)=(\mathbf{d}_x^0,\mathbf{d}_x^1,\cdots,\mathbf{d}_x^{ny})$，是秩为 1 的矩阵，仅一个特征向量，即一个同相轴对应的局部平面波基（Hankel 矩阵的每一列都规定了该基函数）。多个同相轴的情形类似于前面一个同相轴的情形。据此可以看出，对 $\omega-x-y$ 域数据排成 Hankel 矩阵的几何意义（或地震数据分析下局部平面波的意义）。

Hankel 矩阵的这个特点可以从线性信号的 AR 模型预测得出，见后面的具体讨论。

更高维数据也是上述 2D 和 3D 情形的直接扩展。但 Hankel 矩阵的规模会进一步扩张。SVD 分解的工作量会快速上升。仅仅对于 3D 数据而言，决定 Hankel 矩阵规模的参数为 $(k+1)\times(k+1)\times nx\times ny$。可见用 Hankel 矩阵的正交分解来进行线性信号的最佳预测，计算量依然是个巨大的障碍。同样地，局部高维数据体中蕴含的线性结构的个数是信号重构时的未知参数。

5 二维情形下的 $\omega-x$ 域线性信号预测（AR 模型 Wiener 滤波）

事实上，二维数据 $\omega-x$ 域线性信号预测还可以换为 AR 模型预测的形式。对于 $\mathbf{d}=\mathbf{s}+\boldsymbol{\eta}$ 而言，线性信号间显然存在如下关系：

$$S(\omega,x_2)=\alpha\cdot S(\omega,x_1) \tag{21}$$

其中 $\alpha=e^{i\omega p\Delta x}$。上式等价于信号的时空关系满足微分方程：$\dfrac{\partial S}{\partial x}=\alpha\dfrac{\partial S}{\partial t}$。即 $\omega-x$ 域线性信号情形下，两数据点 $d(\omega,x_1)$ 与 $d(\omega,x_2)$ 之间存在由 $e^{i\omega p\Delta x}$ 定义的线性（相位）预测关系。扩展上述关系，可得到 AR 模型来表示的 $\omega-x$ 域的线性信号预测关系：

$$\sum_{k=1}^{P}\alpha_k S_{n-k}=0 \quad\Rightarrow\quad S_n=\alpha_1 S_{n-1}+\alpha_2 S_{n-2}+\ldots+\alpha_p S_{n-P} \tag{22}$$

众所周知，AR 模型是由对系统响应函数的有理分式逼近导出的。但是，对于 $\omega-x$ 域中存在线性信号的二维数据块，类似 AR 模型的线性信号预测关系(22)可以直接写出来。

对于实测信号而言，假设噪音 $\boldsymbol{\eta}$ 满足高斯白噪假设，线性信号预测 AR 模型可写为：

$$d_n\approx\alpha_1 d_{n-1}+\alpha_2 d_{n-2}+\ldots+\alpha_P d_{n-P}=\mathbf{d}^H\cdot\boldsymbol{\alpha} \tag{23}$$

上式的基本含义即 d_{n-1}，d_{n-2} … d_{n-P} 相邻的两个数据之间存在 $e^{i\omega p\Delta x}$ 的相位差，那么 S_n 相当于沿斜率为 p 的同相轴线性叠加（与线性 Radon 变换本质是一致的！）。上式的 AR 模型解释：用斜率为 p 的同相轴上的 P 个数据对 $P+1$ 个点进行线性预测。

那么滤波器（本质是加权系数）应该是什么样的？理论上应是 $\dfrac{1}{P}e^{i\omega p\Delta x}$。这个从(21)式可以看出来。

下面从反问题的角度再分析 $\omega-x$ 域线性预测的本质含义。首先定义 L2 范数的误差泛函：

$$\min J(\boldsymbol{\alpha})=\frac{1}{2}\left(\boldsymbol{d}^H\cdot\boldsymbol{\alpha}-S_n\right)^T\left(\boldsymbol{d}^H\cdot\boldsymbol{\alpha}-S_n\right) \tag{24}$$

对应的法方程为：

$$\boldsymbol{d}(\boldsymbol{d}^H\cdot\boldsymbol{\alpha}-S_n)=0\Rightarrow\mathbf{R}_{dd}\boldsymbol{\alpha}=\boldsymbol{d}\cdot S_n \tag{25}$$

其中，$\boldsymbol{\alpha}=\left[\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_P\right]^T$，$\boldsymbol{d}^H=\left[d_1,\ d_2,\ \ldots,\ d_P\right]$。对应的自相关矩阵正交分解（$\mathbf{R}_{dd}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T$）后，得到法方程的解，

$$\mathbf{U}^T\boldsymbol{\alpha}=\mathbf{U}^T\boldsymbol{d}\cdot\left(\frac{S_n}{\boldsymbol{\Sigma}}\right) \tag{26}$$

可见 $\boldsymbol{\alpha}$ 与 $\boldsymbol{d}$ 中基函数向量平行方向的向量。这就从理论上证明了上面对滤波器系数的几何含义的推断。滤波器算子就是一个线性结构筛选器。雷达测向(Beamforming)中应用了类似的思想。可以想到，线性 Radon 变换也是沿线性结构方向进行投影。

6 FSA、SSA、高维 Wiener 滤波与线性 Radon 变换（Beamforming）等价性的说明

到目前为止，我们知道：对于 FSA 而言，一维、二维到 n 维并无本质差异；线性 Radon 变换（Beamforming）可以说是 FSA 的变种实现，认为信号由 n 维平面波编码或叠加而形成，即 $S(\mathbf{x})=\sum_{k=1}^{K}\alpha_k\boldsymbol{\varphi}_k(\mathbf{x})$，$\boldsymbol{\varphi}_k(\mathbf{x})$ 为 n 维平面波基。但二者的具体实现上，还是有区别的。线性 Radon 变换（Beamforming）可以自己决定基函数族的选择，也可以提不同反问题进行求解，可以考虑不同的正则化方法，具体实现显得更灵活。

FSA 是最常用的线性信号分析方法，但往往不是最合适的线性信号分析方法。

SSA 对于一维、二维、三维至 n 维，具体实现上很不相同。一维局部数据的 SSA 仅仅对一维局部数据的自相关函数进行 SVD 分解即可；二维 SSA 需要把局部数据变换到 $\omega-x$ 域，进行每个频率切片局部数据 Hankel 矩阵（或自相关矩阵）的构建，然后进行 SVD 分解，得到局部平面波的预测。若对时间空间域局部数据 $\mathbf{d}(x,t)$ 直接进行处理，只有对 $\mathbf{d}(x,t)$ 中有一个水平同相轴时间空间的图像（数据块）形成矩阵，然后 SVD 能得到仅一个特征基函数。若 $\mathbf{d}(x,t)$ 中包含若干同相轴，必须在 $\omega-x$ 域中处理，即 SSA 应在 $\omega-x$ 域中执行。三维以上的数据的 SSA，前面已经讲得很清楚了。

线性 Radon 变换（Beamforming）也等价于 SSA。首先，频率空间域数据块 $\tilde{\mathbf{d}}(x,\omega)$ 中包含若干局部平面波，则线性 Radon 变换（Beamforming）正问题表述为，

$$\tilde{\mathbf{d}}=\begin{bmatrix}\tilde{d}(x_1,\omega)\\ \tilde{d}(x_2,\omega)\\ \vdots\\ \tilde{d}(x_N,\omega)\end{bmatrix}=\begin{bmatrix}e^{i\omega p_1x_1} & e^{i\omega p_2x_1} & \ldots & e^{i\omega p_mx_1}\\ e^{i\omega p_1x_2} & e^{i\omega p_2x_2} & \ldots & e^{i\omega p_mx_2}\\ \vdots & \vdots & \ddots & \vdots\\ e^{i\omega p_1x_N} & e^{i\omega p_2x_N} & \ldots & e^{i\omega p_mx_N}\end{bmatrix}\cdot\begin{bmatrix}s(\omega,p_1)\\ s(\omega,p_2)\\ \vdots\\ s(\omega,p_m)\end{bmatrix}=\mathbf{s}\Psi^H \tag{27}$$

数据向量 $\tilde{\mathbf{d}}(x,\omega)$ 的外积定义的自相关矩阵应为：

$$\tilde{\mathbf{d}}\cdot\tilde{\mathbf{d}}^H=\left(\mathbf{s}\Psi^H\right)\left(\mathbf{s}\Psi^H\right)^H=\mathbf{s}\Psi^H\Psi\mathbf{s}^H=\Sigma=\begin{pmatrix}s_1^2 & 0 & \cdots & 0\\ 0 & s_1^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & s_m^2\end{pmatrix} \tag{28}$$

可见，线性 Radon 变换（Beamforming）正问题构成自相关矩阵，正交分解结果正是 SSA 的结果。因此，用事先选择的基向量作为秩 1 向量，外积生成数据张量逼近数据体，这也是一种可行的路径。事实上，

秩 1 向量并不一定对应线性结构。后面讨论的张量的 CP 分解和 Tucker 分解并不要求秩 1 向量是线性结构（即平面波基函数）。高维线性 Radon 变换（高维 Beamforming）本质上可以认为是用实现选定的平面波基函数（秩 1 向量）通过外积运算构建数据中内蕴信号的逼近信号。

F-X 预测本质上也是对法方程（系数矩阵（Hankel 矩阵）、自相关矩阵）进行正交分解来确定最佳预测系数的。可见平面波（线性基函数）飘在 Gauss 随机噪声中的线性信号预测的本质就是这样，绝大部分的信号处理新方法都是在上述线性信号预测基础上的升级。

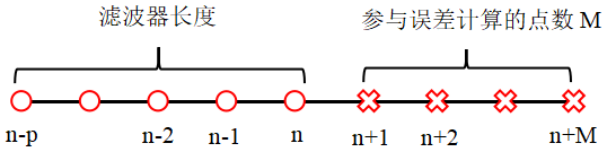
到此可以知道，无论是线性 Radon 变换（Beamforming）、高维数据块的 Hankel 矩阵（自相关函数矩阵）分解、或是二维及以上 AR 模型，这三个问题，从物理背景看，都是完全一样的，即高维局部平面波漂在高维数据体中，从不同角度用不同的方法对局部平面波进行“最佳”预测。殊途同归！

科学和技术中常出现这种不同视角审视同一物理问题得到不同理解和解决问题方法的情况。

7 高维局部数据体 AR 模型及 Wiener 中心滤波器

7.1 高维数据中线性结构的 AR 前向预测模型

二维情形下 F-X 域前向预测问题有如下的讨论。预测器定义为：$d_n \approx \alpha_1 d_{n-1} + \alpha_2 d_{n-2} + ... + \alpha_P d_{n-P}$。下图展示了 F-X 域前向预测中数值计算的具体情况。



*

图 1 二维情形下 F-X 域前向预测中数值计算的具体情况

由图 1 得到的前向滤波器矩阵方程：

$$\begin{pmatrix} d_n & d_{n-1} & \cdots & d_{n-p} \\ d_{n+1} & d_n & \cdots & d_{n+1-p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n+\mathrm{M}-1} & d_{n+M-2} & \cdots & d_{n+M-1-p} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n+M-1-p} \end{pmatrix} = \begin{pmatrix} \hat{d}_{n+1} \\ \hat{d}_{n+2} \\ \vdots \\ \hat{d}_{n+M} \end{pmatrix} \tag{29}$$

上式可以简记为矩阵方程形式：$\mathbf{A}\boldsymbol{\alpha} = \hat{\mathbf{d}}$。针对参与误差计算的数据点，可以定义如下的预测误差方程：

$$\begin{pmatrix} \hat{d}_{n+1} \\ \hat{d}_{n+2} \\ \vdots \\ \hat{d}_{n+M} \end{pmatrix} - \begin{pmatrix} d_{n+1} \\ d_{n+2} \\ \vdots \\ d_{n+M} \end{pmatrix} = \begin{pmatrix} d_n & d_{n-1} & \cdots & d_{n-p} \\ d_{n+1} & d_n & \cdots & d_{n+1-p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n+\mathrm{M}-1} & d_{n+M-2} & \cdots & d_{n+M-1-p} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n+M-1-p} \end{pmatrix} - \begin{pmatrix} d_{n+1} \\ d_{n+2} \\ \vdots \\ d_{n+M} \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix} \tag{30}$$

预测误差方程也可以简记为矩阵方程形式：$\mathbf{e} = \hat{\mathbf{d}} - \mathbf{d}^{obs} = \mathbf{A}\boldsymbol{\alpha} - \mathbf{d}^{obs}$。系数矩阵的转置矩阵为，

$$\mathbf{A}^T = \begin{pmatrix} d_n & d_{n+1} & \cdots & d_{n+M-1} \\ d_{n-1} & d_n & \cdots & d_{n+M-2} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n-p} & d_{n+1-p} & \cdots & d_{n+M-1-p} \end{pmatrix} \tag{31}$$

误差泛函定义为 $J(\boldsymbol{\alpha}) = \frac{1}{2}\mathbf{e}^T\mathbf{e} = \frac{1}{2}(\mathbf{A}\boldsymbol{\alpha} - \mathbf{d}^{obs})^T(\mathbf{A}\boldsymbol{\alpha} - \mathbf{d}^{obs})$。优化求解误差泛函有如下的法方程，

$$[\mathbf{A}^T\mathbf{A}]_{(p+1)\times(p+1)}\boldsymbol{\alpha}_{(p+1)\times1}=\mathbf{A}^T{}_{(p+1)\times M}\mathbf{d}^{obs}_{M\times1}=[\mathbf{A}^T\mathbf{d}^{obs}]_{(P+1)\times1} \tag{32}$$

共轭梯度法解上述法方程，就可以得到滤波器系数向量。把滤波算子施加于数据上，就可以实现实际数据的滤波。应该注意到滤波器系数向量是复的，其中包含了不同斜率的线性结构（同相轴）。

二维情形下 F-X 域后向预测的具体实现有类似的结论。见下图：

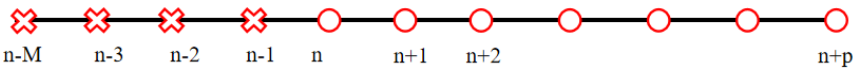


图 2 二维情形下 F-X 域后向预测中数值计算的具体情况

对应的预测误差方程为：

$$\begin{pmatrix} d_n & d_{n-1} & \cdots & d_{n-p} \\ d_{n+1} & d_n & \cdots & d_{n+1-p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n+\mathrm{M}-1} & d_{n+M-2} & \cdots & d_{n+M-1-p} \end{pmatrix}\begin{pmatrix} a_0^* \\ a_1^* \\ \vdots \\ a_p^* \end{pmatrix}-\begin{pmatrix} d_{n+1}^{obs} \\ d_{n+2}^{obs} \\ \vdots \\ d_{n+M}^{obs} \end{pmatrix}=\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix} \tag{33}$$

简记成如下矩阵方程形式：

$$\mathbf{A}_{M\times(p+1)}\boldsymbol{\alpha}^*{}_{(p+1)\times1}-\mathbf{d}_{M\times1}=\mathbf{e}_{M\times1} \tag{34}$$

同样地，由 L2 误差泛函的极小化可以得到对应的法方程的解：

$$[\mathbf{A}^T\mathbf{A}]_{(p+1)\times(p+1)}\boldsymbol{\alpha}^*_{(p+1)\times1}=\mathbf{A}^T{}_{(p+1)\times M}\mathbf{d}^{obs}_{M\times1}=[\mathbf{A}^T\mathbf{d}^{obs}]_{(P+1)\times1} \tag{35}$$

理论上看，仅求 $\boldsymbol{\alpha}$ 或者 $\boldsymbol{\alpha}^*$，然后滤波结果由 $\hat{S}=(\mathbf{d}_f^H\boldsymbol{\alpha}+\mathbf{d}_b^H\boldsymbol{\alpha}^*)/2$ 给出即可。一般地，双向滤波结果好于前向或后向滤波结果。

三维频率空间域数据块 $\tilde{\mathbf{d}}(x,y,\omega)$ 的 AR 模型滤波问题该如何具体求解？三维情形下，AR 前向预测模型可以写为：

$$\mathbf{d}_{m+1,n+1}=\sum_{j=0}^{Npx}\sum_{k=0}^{Npy}\alpha_{j,k}\mathbf{d}_{m-j,n-k} \tag{36}$$

可见，上式中数据和滤波器系数都变成了矩阵。图 3 展示了高维（三维及以上）情形下必须面对的问题，关键是在规定的误差泛函定义区域约束下，基于(36)式形成具体的矩阵方程。

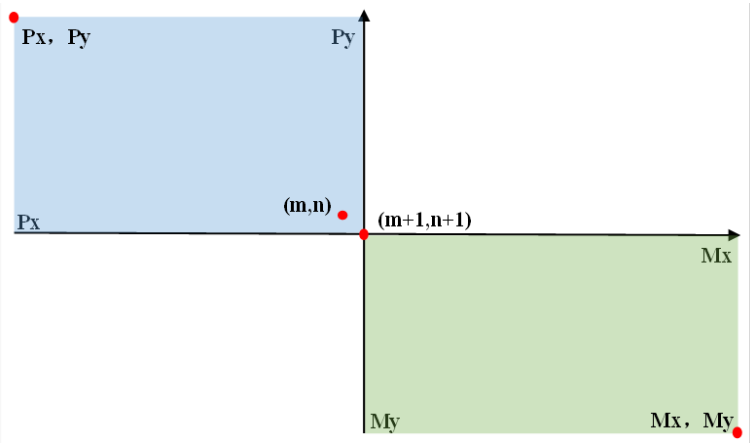


图 3 三维情形下 AR 模型前向预测器范围和预测误差范围

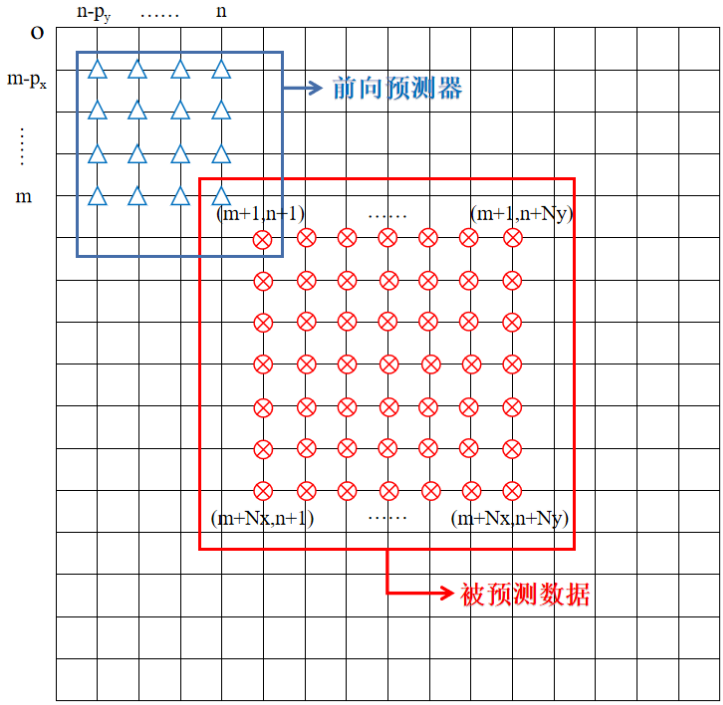


图 4 三维情形下 AR 模型及前向预测误差网格点，其中三角点代表 AR 模型预测器涉及的网格点；圆点代表预测误差计算涉及的网格点

基于选定的预测误差计算区域，有如下的误差泛函定义式，

$$\varepsilon=\sum_{J=0}^{Mx}\sum_{K=0}^{My}\left(d^{obs}(m+J,n+k)-\tilde{d}(m+J,n+K)\right)^2 \tag{37}$$

其中，$J=1,2,\cdots,Nx;K=1,2,\cdots,Ny$。

为了具体地排出 $\omega-x-y$ 情形下前向预测矩阵方程，具体的预测关系由图 4 规定。根据图 4，类似于二维情形，可以排出 $\omega-x-y$ 域前向预测的矩阵方程为：

(38)

构建 L2 范数下的预测误差泛函并优化求解，可得求解上述矩阵方程(34)中的滤波器系数 $\boldsymbol{a}_{m\times n}$，然后按上图可进行滤波（前向滤波）。

上述方程(34)也可记为如下矩阵方程：

$$\mathbf{A}'_{[Nx\cdot Ny]\times[(px+1)\cdot(py+1)]}\cdot\boldsymbol{a}'_{[(px+1)\cdot(py+1)]\times 1}=\boldsymbol{d}'_{[Nx\cdot Ny]\times 1} \tag{39}$$

上式对应的 L2 范数下的法方程为：

$$(\mathbf{A}'^{T}\mathbf{A}')_{\substack{[(px+1)\cdot(py+1)]\\ \times[(px+1)\cdot(py+1)]}}\boldsymbol{a}'_{[(px+1)\cdot(py+1)]\times 1}=(\mathbf{A}'^{T}\boldsymbol{d}')_{[(px+1)\cdot(py+1)]\times 1} \tag{40}$$

这是由 $\omega-x$ 域直接上升到 $\omega-x-y$ 域的矩阵方程。但是，没有对系数矩阵 $\mathbf{A}'$ 的性质进行分析，没建立起系数矩阵 $\mathbf{A}'$ 的 Hankel 矩阵表示。为了说明 AR 模型预测与其它方法的联系，需要对(39)式中系数矩阵 $\mathbf{A}'$ 进行深入的分析。

7.2 AR 模型滤波器系数的 Hankel 矩阵形式分析

AR 模型滤波器系数矩阵与 SSA 和线性 Radon 变换（Beamforming）中的数据矩阵是有密切关系的。为此，我们先分析 $\omega-x$ 域中，数据点之间的预测关系。该关系可以用下图表示。

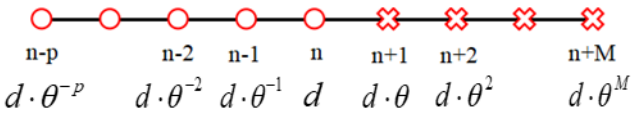


图 5 频率空间域 AR 模型预测数据点间的关系

若 $\omega-x$ 域中仅有一个线性同相轴，则以 d_n 为参考，记为 d，则其他数据与 d_n 之间仅有线性相位移关系，即：

$$e^{j\omega pk\cdot\Delta x}=(e^{j\omega p\Delta x})^k=\theta^k \tag{41}$$

AR 模型滤波方程的矩阵形式为：

$$\begin{pmatrix} d_n & d_{n-1} & \cdots & d_{n-p} \\ d_{n+1} & d_n & \cdots & d_{n+1-p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n+\mathrm{M}-1} & d_{n+M-2} & \cdots & d_{n+M-1-p} \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n+M-1-p} \end{pmatrix}=\begin{pmatrix} d_{n+1} \\ d_{n+2} \\ \vdots \\ d_{n+M} \end{pmatrix} \tag{42}$$

其中的系数矩阵可写为：

$$\begin{pmatrix} d & d\cdot\theta^{-1} & \cdots & d\cdot\theta^{-P} \\ d\cdot\theta & d & \cdots & d\cdot\theta^{1-P} \\ \vdots & \vdots & \ddots & \vdots \\ d\cdot\theta^{M-1} & d\cdot\theta^{M-2} & \cdots & d\cdot\theta^{M-1-P} \end{pmatrix} \tag{43}$$

可见这是个秩为 1 的矩阵，尽管这与一般 Hankel 矩阵的排序稍有不同，但本质上是一样的，可见 $\omega-x$ 域中 AR 模型系数矩阵的构成就是按 Hankel 矩阵构成的。同理，对于更高维的频率空间域数据的线性信号 AR 模型预测，系数矩阵都有 Hankel 矩阵的特征。基于 Hankel 矩阵的 SVD 分解的去噪方法（SSA）就与 AR 模型高维线性预测完全联系起来了。

至此，AR 模型推广到高维情形的问题通过系数矩阵的 Hankel 矩阵构造得到完美的解决！

7.3 高维 AR 中心预测模型及高维 Wiener 中心滤波器

假设一个数据块中仅包含平面波结构，可知两道之间有如下关系：

$$\begin{aligned}\tilde{d}(x+\Delta x,y+\Delta y,\omega)&=\tilde{d}(x,y,\omega)e^{i\omega p_x\Delta x}e^{i\omega p_y\Delta y}\\&=\alpha_x\cdot\alpha_y\cdot\tilde{d}(x,y,\omega)\end{aligned} \tag{44}$$

即 $\tilde{d}(x+\Delta x,y+\Delta y,\omega)$ 与 $\tilde{d}(x,y,\omega)$ 之间仅有线性相位预测关系。其中，$\alpha_x=e^{i\omega p_x\Delta x}$，$\alpha_y=e^{i\omega p_y\Delta y}$。

从统计意义上，可拓展出如下的二维 AR 模型：

$$S(n_x,n_y)=\sum_{j=0}^{P_x}\sum_{k=0}^{P_y}\alpha_{jk}S(n_x-j,n_y-k) \tag{45}$$

其中 $\alpha_{jk}\left(j=0,1,2,\cdots,P_x;k=0,1,2,\cdots,P_y\right)$ 代表线性预测算子。(44)式反映的是一种前向预测关系。内在的

预测关系是由相位移算子 $\alpha_x = e^{i\omega p_x \Delta x}$ 和 $\alpha_y = e^{i\omega p_y \Delta y}$ 规定的。我们知道，一般的 AR 模型的导出是利用有理分式逼近系统响应函数，物理意义很不明显。从上述线性预测的本质上，完全没有必要拘泥于前向或后向预测，利用一段或一片数据进行中心点的预测完全是可以的。这就是 AR 模型的 Wiener 中心滤波。

对于 $\omega-x-y$ 域的情形，AR 模型中心预测关系可以图示为：

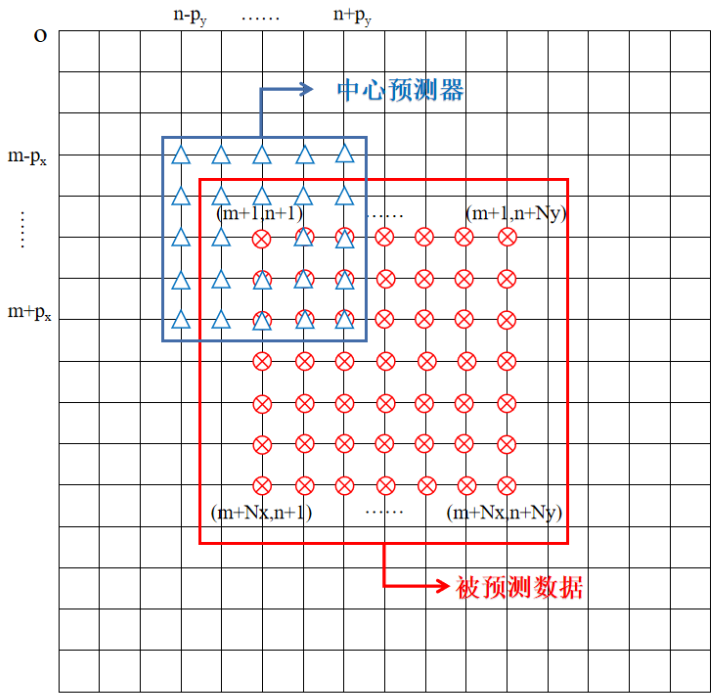


图 6 AR 模型的中心预测关系

对应地，$\omega-x-y$ 域 Wiener 中心滤波写为：

$$\hat{d}_{m+1,n+1} = \sum_{j=-p_y}^{p_y} \sum_{i=-p_x}^{p_x} a_{ij} d_{m-i,n-j} \tag{46}$$

按图 6 中的区域定义预测误差为：

$$\varepsilon(\mathbf{a}) = \sum_{L=1}^{Ny} \sum_{K=1}^{Nx} (\hat{\mathbf{d}}_{m+K,n+L} - \mathbf{d}^{obs}_{m+K,n+L})^2 \tag{47}$$

问题是(47)式对应的系数矩阵方程如何构建。分析可知，把图 6 中的二维中心滤波器在数据上滑动，可以构建出(47)式对应的矩阵方程。具体构建结果由(46)式给出。然后对(46)进行预测误差最小化的求解，可以得到三维 $\omega-x-y$ 情形下的法方程，进而得到需要的滤波器系数矩阵 $[a_{i,j}]$ 中的元素，然后就可以对数据进行滤波处理。

具体计算当然可以用共轭梯度方法。对于高维问题，法方程的系数矩阵规模可能过大了。直接解方程不一定是高效的方法。

关于 Wiener 中心滤波，既可以针对要滤波的数据点点滑动进行，也可以直接对误差逼近区域内的点用(48)式的最佳逼近结果作为滤波结果。但是，理论上，还是用前者更合适。

分析可知，(46)式的系数矩阵具有 Hankel 矩阵的结构。该矩阵与高维数据 SSA 中的 Hankel 矩阵是完全一样的。据此可知，高维 Wiener 中心滤波与高维 SSA 本质上是等价的。

事实上，可以说所有高维数据体中线性结构预测方法本质上都是相同的，只不过具体的实现方法不同而已。

8 张量分解概念下的高维线性信号预测方法

8.1 数据张量分解的概念及数据驱动的高维信号线性预测

数据驱动的高维信号分析借用了张量的概念。但物理学中的张量是物理量的表示，与此处高维数据驱动的张量分解进行数据分析的侧重点是十分不同的。

高维数据一般用张量进行表示。类似地，一组离散数据用向量表示；两维数据体用矩阵表示。

若干“局部平面波”飘在满足不同统计分布特征的随机噪声中代表对勘探地震中数据块的基本认识。应该注意到：三维空间中局部平面波是可见的，高于三维的局部平面波是数据中蕴含的抽象的线性结构，但这并不阻碍用“局部平面波”来表示更高维的数据结构。这样的高维数据块用 $\mathbf{d}(x_1,x_2,\cdots,x_N;\omega)$ 表示。勘探地震中，典型地用 $\mathbf{d}(m_x,m_y,h_x,h_y;\omega)$ 表示频率空间域的五维 CMP 道集数据体，也可统一表示中点-半偏移距坐标系下的所有类型五维数据体。高维线性信号（或称线性结构）飘在满足不同统计分布特征的随机噪声中就是张量分解进行信号分析的元问题。

数据张量的定义： N 阶张量 $\mathcal{X}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ 可以被写为 N 个向量的外积：

$$\mathcal{X}=\mathbf{a}^{(1)}\circ\mathbf{a}^{(2)}\circ\cdots\circ\mathbf{a}^{(N)} \tag{48}$$

其中，符号 “∘”代表向量的外积。这说明 N 阶张量 $\mathcal{X}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ 中的每个元素是对应的向量的乘积：

$$x_{i_1i_2\cdots i_N}=a_{i_1}^{(1)}a_{i_2}^{(2)}\cdots a_{i_N}^{(N)}\quad\text{for all }\ 1\le i_n\le I_n;I_n\in\left(I_1,I_2,\cdots,I_N\right) \tag{49}$$

根据前面的讨论， $(x_1,x_2,\cdots,x_N;\omega)$ 域中的（广义）线性信号预测关系形式上可以表示为，

$$d(x_1,x_2,\cdots,x_N;\omega)=\sum_{l=1}^{L}S(\boldsymbol{p}^l)e^{i\omega\boldsymbol{p}^l\cdot\boldsymbol{x}}=\begin{bmatrix}\boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_L\end{bmatrix}\cdot\begin{bmatrix}s_1\\ s_2\\ \vdots\\ s_L\end{bmatrix}=\boldsymbol{\Psi}\cdot\boldsymbol{S} \tag{50}$$

其中 $S_l=S(\boldsymbol{p}^l)$ ， $\boldsymbol{\varphi}_l=e^{i\omega\mathbf{p}^l\cdot\mathbf{x}}$ ， $\mathbf{k}=\left(k_{x_1},k_{x_2},\cdots,\ k_{x_N}\right)$ ， $\boldsymbol{x}=\left(x_1,x_2,\cdots,\ x_N\right)$ ，$\boldsymbol{p}=\dfrac{\mathbf{k}}{\omega}=\left(\dfrac{k_{x_1}}{\omega},\dfrac{k_{x_1}}{\omega},\cdots,\dfrac{k_{x_N}}{\omega}\right)=\left(p_{x_1},p_{x_2},\cdots,p_{x_N}\right)$。

根据上述关系，可以对 $(x_1,x_2,\cdots,x_N;\omega)$ 域中的数据 $\mathbf{d}(x_1,x_2,\cdots,x_N;\omega)$ 进行张量结构的表达。我们知道，类比于 $(x_1;\omega)$ 情形下， $\boldsymbol{\varphi}_l=e^{i\omega\mathbf{p}^l\cdot\mathbf{x}}$ 代表一个平面波基向量；对于 $(x_1,x_2,\cdots,x_N;\omega)$ 情形， $\boldsymbol{\varphi}_l=e^{i\omega\mathbf{p}^l\cdot\mathbf{x}}$ 代表一个 N 维空间中的线性张量结构。即：

$$\boldsymbol{\varphi}_l=e^{i\omega\mathbf{p}^l\cdot\mathbf{x}}=e^{i\omega\mathbf{p}_{x_1}^l\cdot\mathbf{x}_1}\circ e^{i\omega\mathbf{p}_{x_2}^l\cdot\mathbf{x}_2}\circ\cdots\circ e^{i\omega\mathbf{p}_{x_N}^l\cdot\mathbf{x}_N} \tag{51}$$

可见，(52)式是勘探地震数据情形下的、相当于(49)式表示的张量分解。(52)式中每一项都是沿某个维度的基向量。所有维度的基向量按(52)式支撑起了 $(x_1,x_2,\cdots,x_N)$ 中的一个“局部平面波”或一个局部线性结构。这就清楚地说明了数据驱动的高维信号线性预测的根本问题（元问题）。

8.2 基于张量分解进行高维线性信号预测的思想与方法

① 一维时间（或空间）域一阶张量（即向量）分解的基本思想

首先对时间或空间域一阶张量 $\boldsymbol{y}(t)$：

$$\boldsymbol{y}(t)=\{y_1,\ \ y_2,\ \ \cdots\ \ y_n\} \tag{52}$$

进行分析。一阶张量 $\mathbf{y}(t)$ 的向量外积定义为：

$$\boldsymbol{y}\circ\boldsymbol{y}=\boldsymbol{y}^T\cdot\boldsymbol{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\cdot\begin{bmatrix}y_1,\ \ y_2,\ \ \cdots\ \ y_n\end{bmatrix}=\begin{pmatrix}(y_1,y_1)&(y_1,y_2)&\cdots&(y_1,y_n)\\(y_2,y_1)&(y_2,y_2)&\cdots&(y_2,y_n)\\\vdots&\vdots&\ddots&\vdots\\(y_n,y_1)&(y_n,y_2)&\cdots&(y_n,y_n)\end{pmatrix} \tag{53}$$

这就是时间或空间域一维信号的自相关 $\mathbf{R}_{\mathbf{yy}}$，可进行正交分解或 SVD 分解：

$$\mathbf{R}_{yy}=\mathbf{V}\cdot\mathbf{V}^T=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \tag{54}$$

若记 $\mathbf{U}=\{\boldsymbol{u}_1,\ \ \boldsymbol{u}_2,\ \ \cdots\ \ \boldsymbol{u}_J\}$，$\boldsymbol{u}_j, j=1,2,\ldots,J$ 对应特征向量：

$$\begin{aligned}\mathbf{R}_{yy}&=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T=\left(\boldsymbol{u}_1^H,\ \ \boldsymbol{u}_2^H,\ \ \cdots\ \ \boldsymbol{u}_J^H\right)\begin{bmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_J\end{bmatrix}\cdot\begin{pmatrix}\boldsymbol{u}_1\\\boldsymbol{u}_2\\\vdots\\\boldsymbol{u}_J\end{pmatrix}\\&=\sum_{j=1}^{J}\lambda_j\boldsymbol{u}_j^H\circ\boldsymbol{u}_j,\ \ \boldsymbol{u}_j=\left(u_{j,1},\ \ u_{j,2},\ \ \cdots\ \ u_{j,n}\right)\end{aligned} \tag{55}$$

其中，$(\lambda_1,\lambda_2,\cdots,\lambda_J)$ 是与特征向量对应的特征值。

(56)式必须是这种形式，否则两边不可能全是矩阵，也不可能有等式成立，$\boldsymbol{u}_j^H\circ\boldsymbol{u}_j, j=1,2,\cdots,J$ 对任意的 j 都是一个秩 1 矩阵，即 $\mathbf{R}_{yy}$ 可分解成秩 1 矩阵和的形式，这是矩阵近似表达的基础，也是张量进行低秩逼近的基础。

很显然，对 $\mathbf{R}_{yy}$ 进行低秩逼近，首先选择的数值方法就是对 $\mathbf{R}_{yy}$ 进行正交分解（或 SVD 分解），选择其中的主要特征向量，实现对 $\mathbf{R}_{yy}$ 的最佳低秩逼近；第二种方法是用事先已知的（人为构建的）正交基 $\boldsymbol{u}_j, j=1,2,...,J$ 来计算 $\sum_{j=1}^{J}\boldsymbol{u}_j^H\circ\boldsymbol{u}_j$ 来逼近 $\mathbf{R}_{yy}$。可以看出，最佳的特征向量的个数 J 对于最佳低秩逼近有重要意义。

据此可知，数据张量的算子表达，譬如(49)和(50)式，仅仅是数学家为讨论问题方便而引入的，并非为了数值计算而引入的。张量分解或张量逼近的数值算法还是要基于各种分解形式，在逼近误差最小的目标下进行构建。应该注意到，所谓的张量分解（或矩阵（二阶张量）分解）本质上是对它们进行某种变换，达到用很少的特征成分表达原张量的目的，据此进行近似表达或逼近，进而解决各种应用问题。这和高维函数的逼近本质上并无区别。

② 频率空间（即二维时间空间）域一阶张量（矩阵）分解的基本思想

把时间转化为频率，由于频率成分间的独立性，频率空间域一阶张量分解，本质上就是在对二维时空域信号/图像进行分析。

$$y(x,t)\xrightarrow{FT}\tilde{y}(x,\omega) \tag{56}$$

对于单频数据 $\tilde{y}(x,\omega)$，数据向量记为 $\tilde{\boldsymbol{y}}=\{y_1,\ \ y_2,\ \ \cdots\ \ y_n\}$，对应的向量外积定义为：

$$\tilde{\boldsymbol{y}}^H\cdot\tilde{\boldsymbol{y}}=\mathbf{R}_{\tilde{y}\tilde{y}}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T=\sum_{j=1}^{J}\lambda_j\boldsymbol{u}_j^H\circ\boldsymbol{u}_j \tag{57}$$

可见时间（空间）域一阶张量分解与频率空间域一阶张量分解是完全一样的。此时，正交基

$\mathbf{u}_j, j=1,2,\cdots,J$ 对应的是二维平面波基。但是，(56)式的分解结果对应的是谐波基。

③一般性问题的张量分解的基本思想

首先要说明，一般的张量分解仅仅用于数据体中包含线性结构的情形。到目前为止，自相关矩阵的分解仅仅适用于一维时间（或空间）域数据；Hankel 矩阵的分解适用于高维数据块。它们构成了经典的 SSA。也可以认为它们是 PCA 的一种具体实现方法。当然，PCA 的具体实现也有其它各种各样的途径。

SSA 的本质是把包含线性结构的数据体（或高维线性结构信号）表达在一个正交坐标系中，实现了原数据体中线性信号的解相关。

严格地讲，PCA 并不限制数据中信号成分一定是线性结构的。但是，在离散数据情形下，用矩阵正交分解的方式进行主分量分析，就要限制主分量是线性结构的。这是算法层次上引入的限制，并非 PCA 理论本身引入的。

高维数据体中显然可能内蕴很复杂的结构体。如何感知和表达这些内蕴的复杂结构是数据分析的本质任务。前已述及，高维函数的表达和逼近要解决的也是同类问题。连续小波变换代表了高维函数分析的最高理论进展。离散小波变换也是感知和表达高维数据中复杂结构的有效方法。进一步地，用多种基函数组进行对高维数据中复杂结构的刻画是所谓的字典学习要解决的核心问题。

(49)式表达的是最基本的张量秩 1 分解形式，也称为 Canonical Decomposition 或 Parallel Factor Analysis，前者简称 CP 分解，后者简称 PARAFAC。与之对应的变种形式称为 Tucker 分解。 CP 分解中各秩 1 成分之间是不正交的；Tcuker 分解中各秩 1 成分之间是正交的。前已述及，张量的 CP 分解和 Tucker 分解并不约束秩 1 向量一定是线性结构（局部平面波基函数）。

我们认为，线性 Radon 变换（Beamforming）本质上是用预先选定的平面波基函数（秩 1 向量），通过外积运算实现对数据中线性结构的逼近的。这相当于 CP 分解是分析（Analysis）方法；线性 Radon 变换（Beamforming）是合成（Synsisis）方法。然而，线性 Radon 变换（Beamforming）具有鲜明的地球物理含义。

下面以三维线性 Radon 变换（Beamforming）为例分析 CP 分解和线性 Radon 变换（Beamforming）的对应关系。三维线性 Radon 变换（Beamforming）包含了由所有 p_x'、 p_y' 规定的平面波基函数，可以视为一个四阶张量：

$$\boldsymbol{R}=\begin{bmatrix} R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^1,p_y^1) & R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^2,p_y^1) & \cdots & R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^{npx},p_y^1) \\ R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^1,p_y^2) & R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^2,p_y^2) & \cdots & R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^{npx},p_y^2) \\ \vdots & \vdots & & \vdots \\ R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^1,p_y^{npy}) & R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^2,p_y^{npy}) & \cdots & R(\omega,\boldsymbol{x},\boldsymbol{y},p_x^{npx},p_y^{npy}) \end{bmatrix} \tag{58}$$

上式中每个矩阵元素又被表达为：

$$\begin{aligned}\mathbf{R}(\omega,\boldsymbol{x},\boldsymbol{y},p_x',p_y') &= \begin{bmatrix} e^{i\omega(p_x'\cdot x_1+p_y'\cdot y_1)} & e^{i\omega(p_x'\cdot x_2+p_y'\cdot y_1)} & \cdots & e^{i\omega(p_x'\cdot x_{nx}+p_y'\cdot y_1)} \\ e^{i\omega(p_x'\cdot x_1+p_y'\cdot y_2)} & e^{i\omega(p_x'\cdot x_2+p_y'\cdot y_2)} & \cdots & e^{i\omega(p_x'\cdot x_{nx}+p_y'\cdot y_2)} \\ \vdots & \vdots & & \vdots \\ e^{i\omega(p_x'\cdot x_1+p_y'\cdot y_{ny})} & e^{i\omega(p_x'\cdot x_2+p_y'\cdot y_{ny})} & \cdots & e^{i\omega(p_x'\cdot x_{nx}+p_y'\cdot y_{ny})} \end{bmatrix} \\ &= \begin{bmatrix} e^{i\omega p_y'\cdot y_1} \\ e^{i\omega p_y'\cdot y_2} \\ \vdots \\ e^{i\omega p_y'\cdot y_{ny}} \end{bmatrix}\cdot\begin{bmatrix} e^{i\omega p_x'\cdot x_1} & e^{i\omega p_x'\cdot x_2} & \cdots & e^{i\omega p_x'\cdot x_{nx}} \end{bmatrix}=\left(\boldsymbol{u},\boldsymbol{v}^H\right)\end{aligned} \tag{59}$$

因此，(60)式可以简记为，

$$\mathbf{R}=\begin{bmatrix}\left(\boldsymbol{u}_1,\boldsymbol{v}_1^H\right) & \left(\boldsymbol{u}_1,\boldsymbol{v}_2^H\right) & \cdots & \left(\boldsymbol{u}_1,\boldsymbol{v}_{npx}^H\right)\\ \left(\boldsymbol{u}_2,\boldsymbol{v}_1^H\right) & \left(\boldsymbol{u}_2,\boldsymbol{v}_2^H\right) & \cdots & \left(\boldsymbol{u}_2,\boldsymbol{v}_{npx}^H\right)\\ \vdots & \vdots & \ddots & \vdots\\ \left(\boldsymbol{u}_{npy},\boldsymbol{v}_1^H\right) & \left(\boldsymbol{u}_{npy},\boldsymbol{v}_2^H\right) & \cdots & \left(\boldsymbol{u}_{npy},\boldsymbol{v}_{npx}^H\right)\end{bmatrix}=\begin{bmatrix}\boldsymbol{u}_1\\ \boldsymbol{u}_2\\ \vdots\\ \boldsymbol{u}_{npy}\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{v}_1^H & \boldsymbol{v}_2^H & \cdots & \boldsymbol{v}_{npx}^H\end{bmatrix}=\sum_{j=1}^{J}\boldsymbol{u}_j\circ\boldsymbol{v}_j^H \tag{60}$$

可以清楚地看出，*npx* 和 *npy* 共同决定了平面波的个数。实际勘探地震局部数据块中，平面波的个数是很少的，譬如 1-3 个。可以看出(61)是和(49)式的一致性。即线性 Radon 变换（Beamforming）本质上与 CP 分解是等价的。

总而言之，在实际地震数据问题中，张量分解是针对局部数据体进行的，而局部数据体可以描述为平面波飘在不同统计特征的随机噪声中，对局部数据体的分解应该可以等价为对三维平面波基函数的分解。

9 有限离散线性算子表达与逼近的基本观点

1、矩阵 **A** 是一个离散有限线性算子。还可以上升为更高维的张量形式。张量/矩阵对应的连续形式可以认为是高维函数（或算子）。基于 Weierstrass 定理的函数逼近理论从概念上是可以用来指导张量/矩阵的逼近的。

2、矩阵 **A** 是一个离散有限线性算子。把 $\boldsymbol{x}$ 转换成 $\boldsymbol{y}$，即 $\boldsymbol{y}=\mathbf{A}\boldsymbol{x}$，可以把 **A** 视为列向量表示的坐标系，$\mathbf{A}\boldsymbol{x}$ 即把 $\boldsymbol{y}$ 表示在该坐标系中，$\boldsymbol{x}$ 是各坐标系的系数。

3、矩阵 **A** 的特征值/特征向量分解，即对列向量坐标系的正交化。线性算子 **A** 的谱分解即特征值/特征向量分解，或等价于特征值/特征向量分解。

4、矩阵 **A** 的奇异值分解(SVD 分解)之所以应用很广泛，比矩阵正交分解更广，是因为 SVD 分解可以应用于任意尺寸的 $\mathbf{A}_{m\times n}$ 矩阵中，也包括复矩阵。

$$\mathbf{A}_{m\times n}\cdot\mathbf{A}_{n\times m}^T\Rightarrow\text{正交分解}\Rightarrow\text{左奇异向量}$$
$$\mathbf{A}_{n\times m}^T\cdot\mathbf{A}_{m\times n}\Rightarrow\text{正交分解}\Rightarrow\text{右奇异向量}$$

可见 SVD 分解也是用 QR 正交分解实现的，QR 正交分解是特征值/特征向量计算的根本方法。

5、奇异向量与特征向量是等价的，但奇异值与特征值是不同的。高维离散有限线性算子的分解应该用 SVD 分解/QR 分解，即应该借用矩阵分解的思想方法。分解的目的当然是提取线性算子的主成分，或知道线性算子由哪些正交分量组成。

6、n 维数据体也可以视为一个高维离散线性算子，其中包含有线性结构，这些线性结构是不相关的，但不一定是正交的。

7、提取 n 维数据体中的线性结构的目标不一定要用正交分解的方法来实现，因为正交分解的结果并不对应线性结构，但二者是同构关系，用正交分解的反变换并不会伤害有效的线性结构，但条件是知道线性结构的维度，可惜这往往是不知道的。

8、若 n 维离散数据体包含若干线性结构，对它的逼近或用其中的线性结构表达它是问题的核心。Hankel 矩阵+SVD 分解+特征空间维度确定，的确可以解决很多问题，但前已讲过，特征向量并不等价于线性结构，应用起来不是很好。

9、Beamforming(线性的或广义的，对应于线性或抛物 Radon 变换)引入并不正交的线性结构，作为基函数线性组合来逼近包含线性（或广义线性）结构的 n 维离散数据体，施加 L1 范数/L0 范数的稀疏约束，可以在某些情况下更好地解决问题，毕竟对 n 维离散数据中线性（广义线性）结构的估计更准确，放弃了正交分解得到分辨率更高的结构估计结果。

10、更理论化的表达是对任何 Hankel 矩阵 **A** 进行分解，秩 1 张量分解是最基本的选择，本质上秩 1 张量分解的选择可以引入类似 beamforming 的平面波线性结构，相当于引入了先验信息。

11、再进一步，即可得到字典学习，用不同基函数族来拟合 n 维线性数据体中的结构。

10 结论性认识

勘探地震中基本的数据图景是若干局部平面波飘在满足不同统计分布特征的随机噪声中。实测的数据被视为随机过程，假设在一个窗函数规定的数据块中，满足高斯、平稳、时不变的假设。勘探地震中的线性信号分析理论是在线性信号预测关系的基础上，解一个满足线性、高斯、平稳假设下的 Bayes 估计问题，实现对局部平面波的最佳预测。这应该是勘探地震中数据分析最基本的元问题描述。

局部数据体中内蕴的线性结构的预测是最核心的问题。目前线性结构的预测方法主要包括：线性 Radon 变换（或 Beamforming）方法、SSA 方法和高维 Wiener 滤波方法。三种方法的本质思想是一致的，只是具体的实现途径不同。但是，它们的处理效果、计算效率、抗假频的方便程度都是不同的。张量分解适应于解决更高维数据中的线性结构预测问题，当然也可以进行非线性结构预测问题，只要这些结构能被分解成秩 1 向量的外积能被描述的形式即可。这显然比上述三种方法前进了一步。

勘探地震中的高维线性信号预测，涉及到窗函数选择、抗假频处理、非规则采样、计算效率和并行计算等问题。如果把上述四个方面的问题得到比较完善的解决，就可以发展出性能较好的线性滤波模块，解决实际数据中的去噪问题。数据规则化、混叠数据的解混叠往往也是基于高维线性信号预测的。

到目前为止，绝大部分的信号和图像分析问题还是建立在高斯、平稳、线性假设下的。为适应信号的缓平稳和非平稳，发展了自适应率波方法，譬如 Kalman 滤波方法、LMS 方法以及 RLS 方法、粒子滤波方法等，最著名的是小波域处理方法；克服高斯性假设，发展了 L1+L1 范数下的信号最佳逼近方法；为克服线性假设，发展了广义线性逼近模型，基于此模型的 L2+L1 误差泛函优化有一定的能力解决非线性问题。但是，复杂信号/图像模式+复杂噪音下的最优信号预测依然还是一个强非线性问题，没有发展出有效的方法技术。

参考文献

[1] Powell, M. J. D., 1981, Approximation theory and methods; Cambridge University Press

[2] Cichocki, A., Zdunek, R., Phan, A. H., and Amari, Shun-ichi, 2009, Nonnegative Matrix and Tensor Factorizations – Applications to Exploratory Multi-way Data Analysis and Blind Source Separation; John Wiley & Sons, Ltd

[3] Horn, R., and Johnson, C. R., 1986, Matrix Analysis；Cambridge University Press

[4] Moon, T. K., and Stirling, W. C., 2000, Mathematical Methods and Algorithms for Signal Processing; Prentice Hall

基于信息理论的推断原则和学习算法

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】人类的日常活动或生产科研活动，无时无刻不在做各种层次的决策。所谓决策，本质上就是权衡付出代价的大小，做出代价最小的决定和行动。定义决策代价看起来不是很复杂的，但能给出一个保罗万象问题的决策代价函数决非易事。这基本上是一个哲学层析的、形而上的问题。Bayes 估计理论（或称统计推断理论、统计决策理论）给出了一个很高层次的理论框架。首先它认为观测数据是随机的，决策变量与观测数据是存在一定关系的，它们之间的关系可以用一定的条件概率密度来表示；认为决策变量也是随机的，满足一定的概率分布；根据一般的 Bayes 公式得到统计推断所用的后验概率密度函数；后验概率密度函数的期望给出估计推断结果，方差界定估计推断结果的精度。最大似然估计是频率学派给出的估计理论准则。他们认为给定决策变量的先验信息逻辑上是不合理的，也不接收受估计结果可以认为是满足一定先验概率分布的假设。反映观测数据与决策变量之间关系的似然函数的最大化作为统计推断的准则。信息是广泛存在的。学术上认为不确定性减少是信息存在的体现，并据此定义出信息量，它是一个随机变量。信息量的期望值被定为为熵。相对熵和互信息的凸性使得它们成为了当前机器学习算法中非常重要的决策准则。决策准则定下来后，余下的就是相对具体的优化计算方法。上述逻辑已经成为了解决各种领域问题的基本逻辑框架。本文主要是对上述逻辑进行相对深入的分析，奠定解决具体问题的框架性思维。

【关键词】统计决策，Bayes 估计理论，最大似然估计，熵，相对熵，互信息，机器学习算法

1 引言

人类已经进入了所谓的大数据、人工智能时代。大数据时代的支撑力量，首先是大数据。当前，大数据的来源主要是无处不在的、各种各样的物理传感器采集到的数据；也包括人类的生产、生活活动产生的、并非由物理传感器采集的数据，这类数据样式更复杂、也许量更大。第二就是互联网，数据要有互联互通。第三是计算机技术，要有强大的存储、管理、计算、分析能力的计算机系统。更为重要的，要有算法。算法的主要目的是：管理数据、预处理数据、分析数据、信息提取、形成决策。最后最好是控制“机器人”做出行动。这是当今大数据和人工智能时代的基本模式。

算法的基础是什么？众所周知，算法是分层次的。存在各种各样或复杂或简单、或单一或综合的算法。总体而言，算法是要完成或解决一项或大或小的任务。这就涉及到最佳决策问题或最优化问题。算法是在决策目标上发展出的一套解决问题的、可在计算机上执行的策略或步骤。因此，算法的基础在于提出合理的决策原则。

解决最佳决策问题或最优化问题的最高优先级的事情是给出最佳决策的判决准则。所谓决策，本质上就是权衡付出代价的大小，做出代价最小的决定和行动。代价函数是要首先定义好的。然后才能进入代价函数最小化的算法设计阶段。这是一个总体的逻辑。

这基本上是一个哲学层析的、形而上的问题。当前，能称得上定义决策原则的思想方式可以包括：Bayes 估计理论；最大似然估计理论；信息熵最大或最小理论；变分原理。还可以包括奥卡姆剃刀原理。

Bayes 估计理论（或称统计推断理论、统计决策理论）给出了一个很高层次的理论框架。首先它认为观测数据是随机的，决策变量与观测数据是存在一定关系的，它们之间的关系可以用一定的条件概率密度来表示；认为决策变量也是随机的，满足一定的概率分布；根据一般的 Bayes 公式得到统计推断所用的后验概率密度函数；后验概率密度函数的期望给出估计推断结果，方差界定估计推断结果的精度。最大似然估计是频率学派给出的估计理论准则。他们认为给定决策变量的先验信息逻辑上是不合理的，也不接收受估计结果可以认为是满足一定先验概率分布的假设。反映观测数据与决策变量之间关系的似然函数的最大化作为统计推断的准则。

变分原理是数学分析学家给出的准则。它是建立在物理世界基本原理的基础上，即稳定的状态一定出现在能量最低的状态。事实上，Bayes 估计理论和最大似然估计理论是包容了变分原理的。

奥卡姆剃刀原理的基本逻辑是：如无必要勿增实体。这个基本逻辑是基于对世界的观察提出的，即问题的最合理或最优的解应该是最简洁的。数学提法是最优的解在某一个域中应该是最稀疏的。

从算法的角度看，最佳决策函数或最优化目标函数被期望具有凸性。但是，上述原则下并不能自动地给出凸的最佳决策函数或最优化目标函数。因此，随着机器学习算法的应用不断普及和研究不断深入，基于相对熵和互信息的学习算法变得越来越受关注，因为它们本身就具有凸性。参数估计领域也在借用相对熵和互信息作为最优化目标函数发展出更为稳定的具体算法。

本文主要是对上述最优决策逻辑进行相对深入的分析，奠定解决具体问题的框架性思维。据此可以发展出各种解决具体的算法。

2 推断问题的基本原则

基于具有随机特性的观测结果，提取出其中的信息，基于信息进行统计推断，做出“最佳”决策，这是各个领域中最核心的任务。推断的最核心任务是首先建立推断的基本原则。这些原则首先来自于思维逻辑，然后来自于自然原则。

来自思维逻辑的推断准则包括：Bayes 估计、推断、或决策原则；最大似然原则；信息熵最大或最小的原则。来自自然原则的推断准则包括：变分原理；奥卡姆剃刀原理。低一个层次，我们希望基于推断原则产生的代价函数应该是凸的。

估计、推断或决策的基础是数据或信息。数据可以被抽象为“系统”的输出。信息可以认为是来自于观测数据的。数据可以是抽象的、广义的“数据”。物理系统的输出被各种物理传感器接收是最直观、最有结构化的数据。但是，日常生活、生产等等类型的活动产生的“数据”是五花八门、千差万别的。基于这样的数据进行决策，首先要做的是“数据”的规范化。本质上，数据规范化的目的是按照某种逻辑关系建立起这些数据间空间时间、或其它维度上的关联性。这些关联性内蕴了研究对象的某种内在性质。规范化后从数据中提取信息的算法就相对比较简单和程式化了。数据分析的基本逻辑就应该是这样的。当今，互联网领域数据分析的最困难问题不是数据分析算法，而是数据规范化工作，正体现了上述观点。某种程度上，发展概率图模型也是为了对付这种复杂的“数据”。

“观测数据”（事实上，很多数据并非观测的，而是收集来的）被视为随机的，是当前数据分析领域广泛接收的观点，进行统计估计或统计决策理论框架下的数据分析已经成为共识。绝大部分的机器学习算法都是在统计估计/推断/决策理论下发展出来的。

3 Bayes 推断与最大似然推断

Bayes 推断（Bayesian Inference）可以认为是 ML（Machine Learning）的基本理论基础。某种程度上可以认为：机器学习就是统计学习。绝大部分、甚至所有的，ML 算法都是奠基在 Bayes 推断思想上的！统计数据分析的本质就是在统计模型 $\{P_\theta, \theta \in \Theta\}$ 的框架下估计随机数据中蕴含的信息，就是估计统计模型 $\{P_\theta, \theta \in \Theta\}$ 中的参数 θ 或各阶统计量。本质上，这也可以说是建立随机数据（高维随机过程）的统计模型，也就是联合概率密度函数。然后基于联合概率密度函数做出最佳决策。可以认为估计统计理论和最佳统计决策是一回事。

Bayes 估计的思想可以简单概括为：估计的结果 $\hat{\mathbf{m}}$ 是由后验概率密度函数 $P(\mathbf{m}/\mathbf{y})$ 所决定的均值 $E\{\mathbf{m}/\mathbf{y}\}$，其中 $\mathbf{y}$ 代表观测数据。已知后验概率密度函数 $P(\mathbf{m}/\mathbf{y})$ 后，估计结果写为，

$$\hat{\mathbf{m}} = E\{\mathbf{m}/\mathbf{y}\} = \int \mathbf{m} P(\mathbf{m}/\mathbf{y}) d\mathbf{m} \tag{1}$$

式(1)表示参数估计的结果是由观测数据所对应的模型参数出现的概率所决定的。

由于后验概率密度 $P(\mathbf{m}/\mathbf{y})$ 很难估计，所以将其转化为较容易估计的模型空间的先验概率密度函数 $P(\mathbf{m})$、数据空间的概率密度函数 $P(\mathbf{y})$ 和条件概率密度函数 $P(\mathbf{y}/\mathbf{m})$ 来表示，其中 $P(\mathbf{y}/\mathbf{m})$ 体现了正算

子的不确定性，正算子描述了从模型空间到数据空间的映射关系，它体现了待估计参数与实测实际之间的关系并非完全正确的，预测结果是存在不确定性的。后验概率密度由 Bayes 公式描述为

$$P(\mathbf{m}/\mathbf{y})=\frac{P(\mathbf{m}/\mathbf{y})P(\mathbf{m})}{P(\mathbf{y})} \tag{2}$$

Bayes 公式中的三个概率密度函数，在不同的应用中应该有不同的解释。在参数估计问题中，我们给出如下的含义说明。关于模型空间的概率密度函数 $P(\mathbf{m})$，首先体现了待估计结果被认为是随机的，先验信息是假定已知的概率分布函数。很多时候，我们假定 $P(\mathbf{m})$ 是满足 Gauss 分布的，其均值和方差在待估计参数的求解过程中主要起到约束解的取值范围的作用。如果 $P(\mathbf{m})$ 某个分量的方差较小，代表由先验信息所决定的该分量的值的可靠性就较高，因此该分量的估计结果主要由先验信息所决定；如果 $P(\mathbf{m})$ 某个分量的方差较大，代表由先验信息所决定的该分量的值的可靠性就较低，因此该分量的估计结果主要由观测数据所决定。 $P(\mathbf{m})$ 初步定义了反演问题的解在整个模型空间的概率密度函数，而反演问题的求解过程就是利用所观测到的数据更新这一概率密度函数，把它转化成后验概率密度函数的过程。关于数据空间的概率密度函数 $P(\mathbf{y})$，在 Bayes 框架下，我们认为没有关于实测数据的先验信息。这样的先验假设应该是合理的。因此，概率密度函数 $P(\mathbf{y})$ 被认为是均匀分布，就是说采样到什么数据的概率都是相等的。

关于条件概率密度函数 $P(\mathbf{y}/\mathbf{m})$，它表示在给定模型 $\mathbf{m}$ 的条件下，出现数据 $\mathbf{y}$ 的可能性的大小。实际上， $P(\mathbf{y}/\mathbf{m})$ 显然应该与用来描述实际物理现象的正算子的精确程度，以及实测数据本身的不确定决定。后者主要是观测仪器、噪音、非规则观测引起的不确定性。这两种因素的影响很难区分开，即很难进一步分解成譬如 $P^{Model}(\mathbf{y}|\mathbf{m})$ 和 $P^{Noise}(\mathbf{y}|\mathbf{m})$。但是，我们要清楚地知道是这两种因素决定了 $P(\mathbf{y}/\mathbf{m})$。

根据 Bayes 公式可以看出，$P(\mathbf{m}/\mathbf{y})$ 由 $P(\mathbf{m})$、$P(\mathbf{y})$ 和 $P(\mathbf{y}/\mathbf{m})$ 来决定。假设 $P(\mathbf{y})$ 均匀分布，没有引入先验信息的情况下，参数估计问题的求解可以认为是在 $P(\mathbf{m})$ 和 $P(\mathbf{y}/\mathbf{m})$ 的约束下，得到估计结果由期望均值为 $\hat{\mathbf{m}}=\int \mathbf{m}P(\mathbf{m}/\mathbf{y})d\mathbf{m}$ 决定，方差 $\hat{\mathbf{m}}=\int(\mathbf{m}-\hat{\mathbf{m}})^2P(\mathbf{m}/\mathbf{y})d\mathbf{m}$ 决定。

更可行的做法是假设 $P(\mathbf{m})$ 和 $P(\mathbf{y}/\mathbf{m})$ 均为高斯函数，后验概率密度函数也是 Gauss 分布。此时，后验概率密度函数最大化（MAP）对应了代价函数

$$S(\mathbf{m})=\frac{1}{2}\left[(g(\mathbf{m})-\mathbf{d}_{obs})^H C_D^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})+\lambda(\mathbf{m}-\mathbf{m}_{prior})^H C_M^{-1}(\mathbf{m}-\mathbf{m}_{prior})\right] \tag{3}$$

的最小化。式中：C_D 为数据协方差矩阵；C_M 为模型的协方差矩阵。

如果正演问题 $g(\mathbf{m})$ 是线性函数，即 $g(\mathbf{m})=G\mathbf{m}$，并且矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）是正定对称的，则 $S(\mathbf{m})$ 是一个正定二次函数，有唯一的极小值点。求解 $S(\mathbf{m})$ 极小值点的收敛速度和解的稳定性取决于矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）条件数的大小，因此可以用 Newton 类的方法快速得到其最优解。如果矩阵（$G^H C_D^{-1}G+\lambda C_M^{-1}$）是不定的，即其特征值有正有负，则要先对 $S(\mathbf{m})$ 做适当的正则化处理（在 $S(\mathbf{m})$ 的构造过程中加入适当的约束条件），再求取其极小值点。

如果正问题 $g(\mathbf{m})$ 是非线性问题，但代价函数 $S(\mathbf{m})$ 往往是个非凸的函数，那么 $S(\mathbf{m})$ 的极小值点可能是非常多的。为了让代价函数 $S(\mathbf{m})$ 成为一个凸的函数，可能的做法之一是给定一个非常接近真解的初始解 $\mathbf{m}_0$；或者修改数据和正问题把 $S(\mathbf{m})$ 变成凸问题，但这已经不是在解原问题了，而是在解原问题的一

个临近问题。

通常也可以把反演过程视为在所有可能的参数模型形成的模型空间（也是方程的解空间）中进行采样。反演结果对应的采样点应该是能够使式（3）达到极小值，或是使后验概率密度 $P(\mathbf{m}/\mathbf{y})$ 达到最大值的采样点。基于上述用采样的观点来理解反演问题，可以认为，在确定性的观点下利用梯度导引类的方法求解非线性反演问题仅仅是在初始值附近的局部范围内对可能的模型参数进行采样。但勘探地震学中模型空间规模巨大，初始值虽很重要，但其选择也很困难。如果初始值的选择不能接近真实值，梯度导引类的算法将不能得到有意义的反演结果。另一方面也说明所有可能的参数模型依赖于建模过程中对参数所选择的描述方式，其会直接影响到模型空间的采样方式。因此，在正演过程的建模中要尽可能选择最佳的描述方式以使得最终的模型空间足够小。

基于此观点，Monte Carlo 采样反演方法就显得非常直观。与梯度导引类方法不同，Monte Carlo 反演方法是基于一定的概率密度函数在所有可能的参数模型所形成的模型空间中进行采样，而不是基于梯度导引类方法在初始值附近对模型空间进行采样。Monte Carlo 方法对模型空间的采样所依据的概率密度函数体现了对模型的认识，这种认识主要来源于模型空间的先验信息。与梯度导引类方法相比，Monte Carlo 方法对初始值的依赖性要小很多，但无论是 Monte Carlo 随机采样方法还是梯度导引类采样方法，都应该是在 Bayes 估计的思想框架下进行的。

最大似然推断可以认为是 Bayes 推断的特例，即假设没有引入关于待估计参数的先验信息 $P(\mathbf{m})$，把先验概率密度函数 $P(\mathbf{y}|\mathbf{m})$ 视为似然函数。这是频率统计学派坚持的观点。似然函数的最大化作为统计推断的原则。此时，待估计结果不被视为是随机的。这是最大似然估计的重大问题。如果不把待估计结果视为随机的，而是确定性的，很难全面地评价待估计解的精度。

4 变分原理

变分法是研究依赖于某些函数的定积分型泛函极值的分析学分支。目的是求泛函的极值对应的函数。Euler 和 Lagrange 分别用 The Calculus of Variation 和 The methods of Variation 来命名变分法。

变分问题的直接方法指不通过求解 Euler 方程而直接从泛函本身出发，利用给定的边界条件，求出使泛函取得极值的极值函数的近似表达式（或精确解）。间接方法自然是求解满足边界条件的 Euler 方程得到极值函数的近似表达式（或精确解）。

数学家弗里德里希斯证明：对于一个正定算子方程，一定有一个与之等价的泛函极小值问题。即若能找到泛函极值问题的解，就可找到相应的算子方程的解。大部分褶积系统建立起了输入与输出之间的算子方程。而我们知道，褶积系统对应的算子方程是十分广泛的存在。作为求解算子方程的基本理论框架的变分原理具有广泛的适用性。非常多的问题都可以在变分原理的框架下得到求解。

二阶偏微分方程的边值问题的求解总可以提成二次型泛函的极值问题。二次型泛函可以认为是误差能量泛函。误差能量泛函的最小化解对应二阶偏微分方程边值问题的解。这就建立起了 Bayes 估计及最大似然估计之间的联系。

5 交互熵与互信息推断

信息是自然和社会科学领域中所关注的千差万别的对象所具有的基本属性。到目前为止，人们也没有给信息一个统一的定义。也许正是因为信息具有的这种广泛存在特点和反映事物本质属性的特点，使得可以由信息定义出一种推断原则。

基于通讯科学产生的信息论，信息被定义为被消除的不确定性。不确定性显然与随机性有关。一个事件拥有不确定性，就认为这个事件拥有信息。信息量的度量自然应该与事件发生的概率联系起来。一个确定性的事件，概率为 1，认为是没有信息的。就是说，事件发生的概率越大，它包含的信息量越大。很显然，从这样的逻辑定义信息是非常合理的。

一个随机变量 x_i 对应的概率为 $P(x_i)$，它拥有的信息量定义为如下的形式，

$$I(x_i) = log\left(\frac{1}{P(x_i)}\right) \tag{4}$$

(4)式称为随机变量 x_i 的自信息量。很显然，自信息是单个随机变量对应的某个事件发生的可能性的度量。按(4)式定义，自信息显然是一个随机变量。自信息熵是单个随机变量所有事件发生可能性的期望。自信息熵有如下定义，

$$H(X) = -\sum_{x\in\mathcal{X}} P(x) log\, P(x) \tag{5}$$

上式可重写为：

$$H(X) = E_P I(X) \tag{6}$$

我们更关心的是两个随机变量之间的关系，类似地，定义互信息熵为：

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y) log\, P(x,y) \tag{7}$$

上式可以重写为：

$$H(X,Y) = E_P I(X,Y) \tag{8}$$

两个随机变量间往往存在依赖关系，因此有条件信息熵，

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} P(x) H(Y|X=x) = -\sum_{x\in\mathcal{X}} P(x) \sum_{y\in\mathcal{Y}} P(y|x) log\, P(y|x) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x) P(y|x) log\, P(y|x) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y) log\, P(y|x) \end{aligned} \tag{9}$$

很显然，条件信息熵是条件概率密度函数 $P(y|x)$ 定义的信息量的期望值。因此，应该有 $H(Y|X) \neq H(X|Y)$。同时应该有，

$$H(X) \geq H(X|Y) \tag{10}$$

和

$$H(Y) \geq H(Y|X) \tag{11}$$

成立。

根据条件熵和联合概率密度函数关系 $P(x,y) = P(x)P(y|x)$，互信息熵可以表达为，

$$
\begin{aligned}
H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x,y) \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x)P(y|x) \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(y|x) \\
&= -\sum_{x\in\mathcal{X}} P_y(x)\log P(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(y|x) \\
&= H(X) + H(Y|X)
\end{aligned}
\tag{12}
$$

即：

$$
H(X,Y) = H(Y) + H(X|Y) \tag{13}
$$

根据条件熵和联合概率密度函数关系 $P(x,y) = P(y)P(x|y)$，互信息熵还可以表达为，

$$
\begin{aligned}
H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x,y) \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(y)P(x|y) \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(y) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x|y) \\
&= -\sum_{y\in\mathcal{Y}} P_x(y)\log P(y) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x|y) \\
&= H(Y) + H(X|Y)
\end{aligned}
\tag{14}
$$

即：

$$
H(X,Y) = H(X) + H(Y|X) \tag{15}
$$

由(13)和(15)可以容易得到，

$$
H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{16}
$$

和

$$
H(X,X) = H(X) \tag{17}
$$

可以注意到，Kullback-Leibler distance、或 Kullback-Leibler Divergence 或相对熵在当前的机器学习算法或参数估计问题中有越来越广泛的应用，它被定义为，

$$
D(p\|q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \tag{18}
$$

上式可以重写为，

$$
D(p\|q) = E_p \log\frac{p(x)}{q(x)} \tag{19}
$$

互信息被定义为：

$$I(X,Y)=\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)log\frac{P(x,y)}{P(x)P(y)} \tag{20}$$

同样地，(20)可以写为，

$$I(X,Y)=E_{P(x,y)}log\frac{P(X,Y)}{P(X)P(Y)} \tag{21}$$

可以看出，相对熵$D(p\|q)=\sum_{x\in\mathcal{X}}p(x)log\,p(x)/q(x)$可以被解释为两种分布$p(x)$和$q(x)$之间“距离”或“差异”的度量。直观地，它是两种分布$p(x)$和$q(x)$规定的似然比对数的关于分布$p(x)$的期望值。也可解释为：它是当真分布为$p(x)$时，假设分布为$q(x)$时无效性的度量。相对熵总是正的，只有在两种分布相等时，相对熵才等于零。可以证明：Kullback-Leibler distance、或 Kullback-Leibler Divergence 或相对熵是个凸函数。因此，相对熵总可被用来作为统计推断的原则。但是，相对熵不满足三角不等式。并不是一个真正的范数。

类比 Kullback-Leibler Divergence 或相对熵，(21)式的互信息显然可以被重写为：

$$I(X,Y)=D\left(P(X,Y)\middle\|P(X)P(Y)\right) \tag{22}$$

很显然，互信息是 Kullback-Leibler Divergence 或相对熵的一种特例。

互信息还可以表达为，

$$\begin{aligned}
I(X,Y)&=\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)log\frac{P(x,y)}{P(x)P(y)}=\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)log\frac{P(x,y)/P(y)}{P(x)}\\
&=\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)log\frac{P(x|y)}{P(x)}\\
&=\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)log\,P(x|y)-\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)log\,P(x)\\
&=\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)log\,P(x|y)-\sum_{x\in\mathcal{X}}P_y(x)log\,P(x)\\
&=H(X)-H(X|Y)
\end{aligned} \tag{23}$$

即

$$I(X,Y)=H(X)-H(X|Y) \tag{24}$$

由(10)知道，互信息$I(X,Y)$总是大于等于零的。

同时，互信息还可以表达为，

$$
\begin{aligned}
I(X,Y) &= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P(x,y)\log\frac{P(x,y)/P(x)}{P(y)} \\
&= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P(x,y)\log\frac{P(y|x)}{P(y)} \\
&= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P(x,y)\log P(y|x) - \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P(x,y)\log P(y) \\
&= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P(x,y)\log P(y|x) - \sum_{y\in\mathcal{Y}} P_x(y)\log P(y) \\
&= H(Y) - H(Y|X)
\end{aligned}
\tag{25}
$$

即

$$
I(X,Y) = H(Y) - H(Y|X) \tag{26}
$$

同样地，由(11)式知，可以知道，互信息 $I(X,Y)$ 总是大于等于零的。由(23)和(25)式的推导过程，还可以知道，

$$
I(X,Y) = I(Y,X) \tag{27}
$$

根据(23)和(25)式，还可以导出，

$$
\begin{aligned}
I(X,Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \left(H(X,Y) - H(X)\right) \\
&= H(X) + H(Y) - H(X,Y)
\end{aligned}
\tag{28}
$$

由(16)还可以得到：

$$
I(X,X) = H(X) \tag{29}
$$

从机器学习的角度，互信息的含义可以从如下系统的观点得到某种解释。系统输入为 x；系统输出为 y。



图 1 系统的观点解释互信息熵

从公式(23)可以看出，互信息显然是由于输出 y 的发生所导致的输入 x 的不确定性的降低或熵减。本质含义是输入 x 与输出 y 之间的相关性（最好不要提相互依赖性）越强，互信息量越大！若输入 x 与输出 y 完全相关，互信息量达到最大。这是互信息用于无监督学习的根基！可以证明，互信息具有凸性，前已述及互信息 $I(X,Y)$ 总是大于等于零的，这是互信息能作为推断或决策原则的理论基础。按照图 1 所示，把互信息作为 ML 算法的基础原则是合理的，因为很多机器学习算法重点关注输入和输出存在的关系。

下面用图示的形式，列出各种信息熵与互信息之间的关系。

$$H(X)+H(Y)-I(X,Y)=H(X|Y)+H(Y|X)+I(X,Y)=H(X,Y)$$

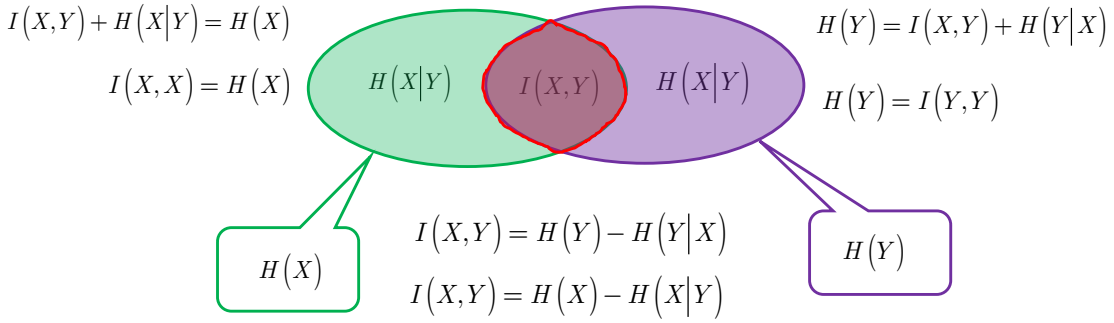


图 2 互信息与互信息熵的关系

6 基于信息理论的学习算法

机器学习的目的是构建一个 LM（Learning Machine）。其中的核心问题是：（随机）数据信息挖掘+基于所挖掘信息的决策。人工智能的核心就是机器学习算法。机器学习算法主要是基于概率统计理论的。机器学习算法总被抽象为建立起输入与输出之间关系的网络结构，完成各种具体的功能，譬如回归与分类及聚类。典型地，机器学习算法包括：有监督学习算法和无监督学习算法。

智能学习类算法的定义：从数据中提取信息，并能据此进行自主决策的一类算法。 “决策代价”最小化作为算法构建的依据。推断或决策原则在机器学习算法中占据核心地位。

可以证明：相对熵（K-L 散度）和互信息都是凸函数。它们都是大于等于零的。因此，它们都可以作为推断、决策、估计或反演的原则。相对熵（K-L 散度）作为两种分布差异的度量，更合适用在有监督学习算法中，或者更适合用于对已知的输出进行最佳逼近的场景中。而互信息作为两随机变量相关程度的度量，更适合用于各种无监督学习场景中。

图 3 展示了几种典型的互相息作为 ML 算法决策准则的场景。下面分几种应用场景说明互信息作为决策准则的机器学习算法。

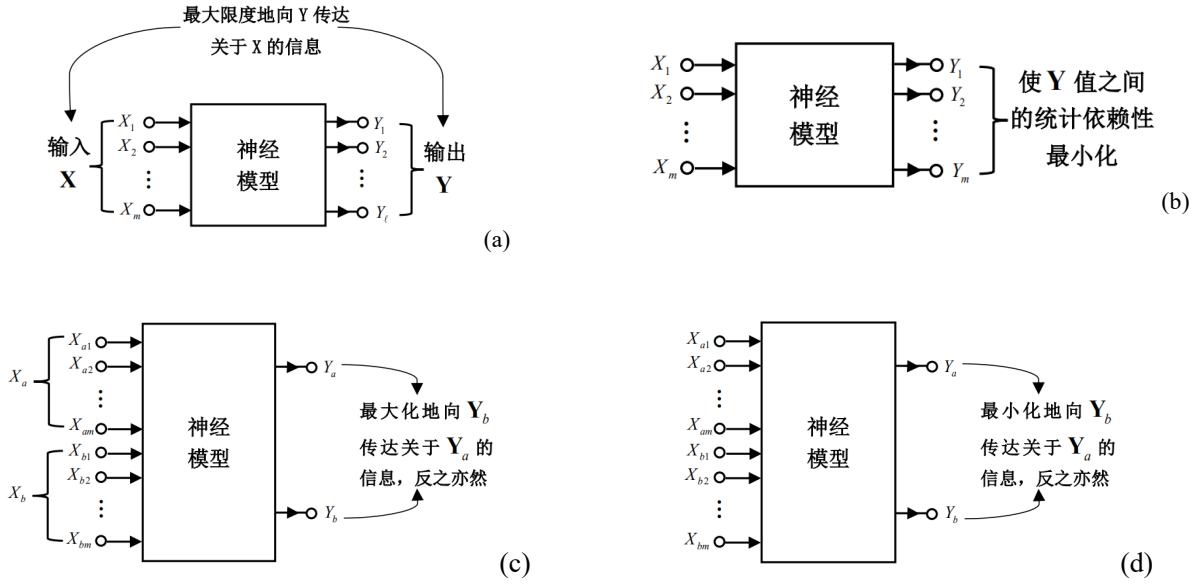


图 3 互相息作为 ML 算法决策准则的场景

情形 I：图 a 可以作为挖掘输入与输出之间相关性的场景。目标是输入与输出之间有尽可能多的相关性。这比较类似于无监督学习的自编码。

情形 II：图 d 是典型的独立分量分析的应用场景，通过强化输出端各分量间的统计独立性或最小化输出端各分量间的统计相关性，达到独立分量分析的目的。独立分量分析具有广泛的应用场景。

情形 III：图 b 通过最大化输出端分量的互信息实现分类。显然可以实现与自组织映射（SOM）类似的工作。也可以用于检测输入图像的相似性，解决与此相关的各种问题。

情形 IV：图 c 是图 b 的相反情形，用于检测输入图像的差异性，解决与此相关的各种问题。情形 III 和情形 IV 与 CCA（Canonical Correlation Analysis，典则相关分析）要解决的问题是类似的。

相对熵（K-L 散度）总是凸函数，而非线性的误差泛函不一定总是凸函数。因此，用相对熵（K-L 散度）作为度量函数进行推断、决策、估计或反演的准则在一定条件下应该可以得到更好的决策结果。互信息测量两个随机变量之间的相关性，这与人脑的学习机制（Hebb 学习机制）非常类似，应该特别适用于无监督学习中。类似地，互信息也总是凸函数。至少表面上看，相对熵和互信息都没有引入高斯假设，原则上可以考虑非高斯情形下的推断、决策、估计或反演问题。

用相对熵或互信息作为开发新的机器学习算法的决策原则，应该由很大的探索空间或余地，推断或决策效果可能更好一些。

7 结论性认识

当今自然科学和社会科学中许许多多问题都可以抽象为基于（随机）观测数据的决策问题，这也是人类社会进入大数据和智能社会的一个逻辑基础。Bayes 统计推断/ Bayes 参数估计理论已经被我们所熟知。据此发展出的均方误差最小估计子和（加权）最小二乘估计子已经广泛应用在信号处理、图像处理和地震波成像（参数反演）的各种场景中。最大似然估计可以认为是 Bayes 统计推断/ Bayes 参数估计理论的特例。事实上，基于能量最小化泛函的变分原理也可以认为是 Bayes 统计推断/ Bayes 参数估计理论在两个先验概率密度函数都 假定满足 Gauss 分布下的情形。但是，上述几种准则都不能保证给出的决策函数是凸的，使得后续的算法受解的不唯一的困扰。

基于 Shannon 信息论发展出的相对熵和互信息，给出了不同的、适用于推断、决策、估计或反演问题的准则。相对熵提供了两种分布的差异度量；互信息提供了一个随机变量包含另一个随机变量的信息量的度量。它们的最大特点是：它们均为非负的凸函数。据此，它们应该更适合用作决策准则，能给出更稳定的解。尤其在无监督学习即将成为或者说已经成为 AI/ML 领域的主要方向的情况下，互信息准则很显然会成为核心的决策准则，据此会发展出各种各样的无监督学习算法。

我认为这些无监督学习算法在特征波（初至波和特征反射波等）识别、地震相识别、层位和断层识别中起到重要作用，为提高地震波成像的精度提供更多的额外信息。也许可以把互信息用于多次波压制过程中、用于 OBN 四分量数据的处理中。

参考文献

[1]. 陈家鼎、郑忠国，2007，概率与统计，北京大学出版社

[2]. 老大中，2015，变分法基础，第三版，国防工业出版社

[3]. Moon, T. K., and Stirling, W. C., 2000, Mathematical Methods and Algorithms for Signal Processing; Prentice Hall.

[4]. Simon Haykin, 2009, Neural Networks and Learning Machines; Third Edition，Pearson Educcation, Inc.

[5]. Thomas M. Cover, and Joy A. Thomas, 2006, Elements of Information Theory; Second Edition, John Wiley & Sons, Inc.

[6]. Sergios Theodoridis, 2015, Machine Learning: A Bayesian and Optimization Perspective; Elsevier Ltd.

[7]. Dimitri P. Bertsekas, and John N. Tsitsiklis, 2008, Introduction to Probability, Srcond Edition, Posts and Telecom Press.

Bayes 参数估计与 Bayes 决策理论方法及勘探地震数据分析中的应用

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】依据从物理系统中输出的数据和非物理系统产生的数据，从数据中获得信息，然后进行决策解决相关问题是各行各业通用的解决问题的基本模式。总体而言，这都可以归结为统计数据分析要解决的问题。把广义的、无论是物理系统产生的或非物理系统产生的“数据”视为随机过程的实现或采样，依据概率基础上的统计理论，从随机的“数据”中提取出有用的信息，产生决策行为，是统计数据分析的核心内涵。Bayes 学派和频率学派各自发展了统计数据分析的思想方法，二者存在密切的关系。随机“数据”的总体分布模型包含了所有的信息，一般地用总体分布模型中的参数代表“数据”中包含的信息。利用随机“数据”的实现对总体分布模型中参数的估计；建立“数据”与潜变量之间的函数描述的关系；依据随机“数据”和其总体分布模型进行关于统计假设的判断问题（假设检验问题）构成了统计数据分析的三个主要内容。把待估计参数视为随机变量，引入相应的先验概率分布，据此产生的 Bayes 统计数据分析的思想方法，由于在解决实际问题中表现出广泛的适用性，得到了更好的应用效果而得到普遍应用。本文试图提取 Bayes 参数估计和 Bayes 决策理论方法的核心思想概念，开展较为深入的分析，并讨论它们在勘探地震中的具体应用。

【关键词】统计数据分析，统计模型，参数估计，假设检验，Bayes 估计，Bayes 决策，地震数据分析

1 引言

一般而言，当前各行各业采集到的有结构和无结构的数据，可以抽象地认为是来自物理系统和非物理系统的，它们都被认为是随机变量或随机向量，或是更高维的随机数据体。大数据时代，解决问题的基本模式是从这些随机“数据”中提取相关“信息”，基于“信息”进行决策，从而解决相关的问题。

勘探地震学的数据主要是人工震源激发的振动在地下介质中传播，经过波与介质的相互作用，在地表（井中）的检波器采集到的有结构的地震波场。当然也包括与地下介质弹性（岩性）参数相关的各种信息。广义的遥感学科，都有类似的数据特征。目标都是通过某种参数估计方法弄清楚系统内部的介质参数分布，并据此进行决策，从而解决各自学科中的具体问题。

统计数据分析是数据科学的基础理论。对“数据”建立总体分布模型，用总体分布模型中的参数代表“数据”中包含的信息。利用随机“数据”的实现对总体分布模型中参数进行估计（参数估计问题）；建立“数据”与潜变量之间的、函数描述的关系（回归分析问题）；依据随机“数据”和其总体分布模型进行关于统计假设的判断（假设检验问题）构成了统计数据分析的三个主要内容。

统计数据分析分为 Bayes 学派和频率学派。Bayes 学派认为总体分布模型中的参数是随机的，满足一定的先验分布的。通俗地讲，在基于观测数据和总体分布模型进行参数估计时对待估计参数是有先验认识的。频率学派认为待估计参数是未知但确定的，不是随机的，不需要假定先验认识，完全由充分的样本实现和合理的总体分布模型决定。Bayes 学派的典型参数估计方法是后验概率密度最大化方法；频率学派的典型参数估计方法是似然函数最大化方法。从实际应用情况看，Bayes 学派的统计数据分析显然占了上风。

把所考察对象的总和称为总体。当刻画总体的随机向量 $\mathbf{X}^t$ 的分布族 $\{F_\theta,\theta\in\Theta\}$ 确定后，$\mathbf{X}^t\sim\{F_\theta,\theta\in\Theta\}$ 称为总体模型。实测数据向量 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 是随机向量 $\mathbf{X}=(X_1,X_2,\cdots,X_n)$ 的一组观测值，其中 $X_1,X_2,\cdots,X_n\sim \text{iid}F_\theta$。随机向量 $\mathbf{X}=(X_1,X_2,\cdots,X_n)$ 是总体随机向量 $\mathbf{X}^t$ 的样本，$\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 是样本的实现，就是观测的数据。随机向量 $\mathbf{X}$ 的取值空间 $\mathcal{X}$ 称样本空间，随机向量 $\mathbf{X}$ 和它的分布族 $\{P_\theta,\theta\in\Theta\}$ 称为统计模型。一般地，实际数据分析中，我们与统计模型打交道。若 $\mathbf{X}\sim\{P_\theta,\theta\in\Theta\}$ 是一个统计模型，定义在样本空间 $\mathcal{X}$ 上的任何函数 $T(\mathbf{x})(\mathbf{x}\in\mathcal{X})$ 都称统计量。对统计模型中的参数 θ 进行估计是假设统计模型形式已知，由观测数据最终确定统计模型所必须要做的。参数估计是统计数据分析的基础。FWI 和 LS_RTM 就是 Bayes 参数估计理论在勘探地震中的应用实例。

统计数据分析另一个重要问题是假设检验。设 $\mathbf{X}^t\sim\{F_\theta,\theta\in\Theta\}$ 为总体模型，假设检验问题是两个关于总体真值的互相对立的判断 $(\theta\in\Theta_0;\theta\in\Theta_1)$ 的鉴别问题，其中 Θ_0 是 Θ 的一个真子集，$\Theta_1=\Theta\setminus\Theta_0$ 为 Θ_0 的余集。判断 $\theta\in\Theta_0$ 为零假设（Null Hypothesis）,记为 H_0;判断 $\theta\in\Theta_1$ 为对立假设（Alternative Hypothesis）,记为 H_1。更为一般地，做出判断需要依赖于样本观测值。为此，需要定义一个样本空间的子集 $\mathcal{W}$，当样本

观测值 $\mathbf{x}\in\mathcal{W}$ 时，否定零假设 $H_0\left(\theta\in\Theta_0\right)$。$\mathcal{W}$ 被称为否定域。设 $\left(\Theta_0,\Theta_1\right)$ 为总体模型 $\mathbf{X}^t\sim\left\{F_\theta,\theta\in\Theta\right\}$ 的一个假设检验问题，$\mathbf{X}=\left(X_1,X_2,\cdots,X_n\right)$ 是总体随机向量 $\mathbf{X}^t$ 的样本，$\mathcal{W}$ 为检验问题的一个否定域，$\alpha\in(0,1)$ 为一个常数，则①称定义在 Θ 上的函数 $\beta_{\mathcal{W}}\left(\theta\right)\triangleq P_\theta\left(\mathbf{X}\in\mathcal{W}\right)$ 为 $\mathcal{W}$ 的功效函数；②若 $\mathcal{W}$ 满足条件 $\sup\limits_{\theta\in\Theta_0}\beta_{\mathcal{W}}\left(\theta\right)\le\alpha$，则称 $\mathcal{W}$ 为检验问题 $\left(\Theta_0,\Theta_1\right)$ 的一个显著性水平为 α 的否定域。又称 $\mathcal{W}$ 为检验问题 $\left(\Theta_0,\Theta_1\right)$ 的一个显著性水平为 α 的一致最大功效否定域（简称 UMP（Uniformly most powerful）否定域）,若①$\mathcal{W}$ 是水平为 α 的否定域；②对任何其他的水平为 α 否定域 $\tilde{\mathcal{W}}$，均有 $\beta_{\mathcal{W}}\left(\theta\right)\ge\beta_{\tilde{\mathcal{W}}}\left(\theta\right),\forall\theta\in\Theta_1$。假设检验问题又可分为：决策性假设检验问题和显著性假设检验问题。Newman-Pearson 引理（简称 N-P 引理）尽管很简单，但奠定了假设检验问题的基本思想基础。简单假设检验问题：设有总体模型 $\mathbf{X}^t\sim\left\{F_\theta,\theta\in\Theta\right\}$，其中 $\theta\in\Theta\left(\Theta=\left\{\theta_0,\theta_1\right\}\right)$，对应的假设检验问题 $H_0:\theta=\theta_0\leftrightarrow H_1:\theta=\theta_1$。Newman-Pearson 引理：设 $\mathbf{X}^t$ 的一个样本为 $\mathbf{X}=\left(X_1,X_2,\cdots,X_n\right)$，假设 $\mathbf{X}^t$ 的概率密度函数为 $f\left(x,\theta\right),\theta=\theta_0/\theta=\theta_1$。记 $L\left(\mathbf{x},\theta\right)=\prod\limits_{i=1}^{n}f\left(x_i,\theta\right)$，其中 $\mathbf{x}=\left(x_1,x_2,\cdots,x_n\right)\in\mathcal{X}$。若对于给定的 $\alpha\in(0,1)$，n 维空间中的集合 $\mathcal{W}=\left\{\mathbf{x}:L\left(\mathbf{x},\theta_1\right)>\lambda_0L\left(\mathbf{x},\theta_0\right)\right\}$，其中 $\lambda_0\ge0$ 是常数，满足 $P_{\theta_0}\left(\mathbf{X}\in\mathcal{W}\right)\triangleq\int\cdots\int_{\mathcal{W}}\prod\limits_{i=1}^{n}f\left(x_i,\theta_0\right)dx_1\cdots dx_n=\alpha$，则 $\mathcal{W}$ 就是简单假设检验问题的水平为 α 的 UMP 否定域。事实上，$L\left(\mathbf{x},\theta\right)$ 是观测向量 $\mathbf{x}=\left(x_1,x_2,\cdots,x_n\right)$ 与参数 θ 共同决定的似然函数。广泛应用的似然比假设检验准则就是由 Newman-Pearson 引理衍生出来的。各行各业中存在大量的假设检验问题。机器学习领域中的分类（聚类）问题是假设检验理论与方法的典型应用。

统计数据分析的第三个核心问题是：建立观测量和与观测量相关的潜变量之间的依赖关系式。这就是典型的回归分析方法要解决的问题。回归分析可以说是上述参数估计理论和假设检验理论在统计数据分析中的应用。广义地讲，有监督学习神经网络就是实现回归分析的典型算法，是经典线性回归分析在理论和算法上的提升。

最后，统计决策也可以独立出来进行讨论，因为决策问题涉及到决策行动和决策代价。当然，决策性假设检验与统计决策是存在密切关联的，把统计决策归为决策性假设检验也是可以的。统计决策涉及到三方面的信息：①观测样本 $\mathbf{X}=\left(X_1,X_2,\cdots,X_n\right)$ 提供的信息；②行动引起的代价，用 $L\left(\theta,a\right)$ 表示行动 a 引起的损失。③关于参数的先验信息。设有统计模型 $\mathbf{X}\sim\left\{P_\theta,\theta\in\Theta\right\}$，取值于行动空间 A 的样本函数 $\delta\left(\mathbf{X}\right)$ 称为决策函数。又有损失函数 $L\left(\theta,a\right)$，则决策函数 $\delta\left(\mathbf{X}\right)$ 对应的平均损失 $R\left(\theta,\delta\left(\mathbf{X}\right)\right)=E_\theta\left[L\left(\theta,\delta\left(\mathbf{X}\right)\right)\right]$ 被定义为其风险函数。又称决策函数 $\delta\left(\mathbf{X}\right)$ 为可允许的，若不存在另一个决策函数 $\bar{\delta}\left(\mathbf{X}\right)$ 使得 $R\left(\theta,\bar{\delta}\left(\mathbf{X}\right)\right)\le R\left(\theta,\delta\left(\mathbf{X}\right)\right),\forall\theta\in\Theta$。又称决策函数 $\delta^*\left(\mathbf{X}\right)$ 为极大极小的，若对任意其他决策函数 $\delta\left(\mathbf{X}\right)$ 均有 $\underset{\theta\in\Theta}{\mathrm{Max}}R\left(\theta,\delta\left(\mathbf{X}\right)\right)\ge\underset{\theta\in\Theta}{\mathrm{Max}}R\left(\theta,\delta^*\left(\mathbf{X}\right)\right)$。假如 $\pi\left(\theta\right)$ 为 Θ 上的概率密度函数，又称决策函数 $\delta\left(\mathbf{X}\right)$ 为相对于先验概率密度 $\pi\left(\theta\right)$ 的 Bayes 决策函数，若有 $\int_{\theta\in\Theta}R\left(\theta,\delta\left(\mathbf{X}\right)\right)\pi\left(\theta\right)d\theta\le\int_{\theta\in\Theta}R\left(\theta,\delta'\left(\mathbf{X}\right)\right)\pi\left(\theta\right)d\theta$，$\delta'\left(\mathbf{X}\right)$ 是其他决策。很显然，强化学习算法是在统计决策理论下发展出来的。

据此，可以清楚地看出，勘探地震数据分析完全应该奠基在统计数据分析的理论基础上。油气地震勘探所有环节涉及到的数据分析问题都可以在统计数据分析的基础上找到相应的思想、方法，并发展出相应的具体技术。包括当前机器学习（深度学习）中的各种算法都是奠基在统计数据分析的理论基础上的，无论是有监督学习算法、无监督学习算法均是如此。

是时候把油气地震勘探领域的数据分析方法都置于统计数据分析的基础上，用 Bayes 参数估计理论和 Bayes 决策（假设检验）理论统一地震数据预处理（信号处理）、介质弹性参数估计（地震波成像、AVA 反演）、地震图像解释（图像处理）、测井信息提取与表达、地质信息提取与表达、甚至油气藏识别与判断的计算机辅助等各环节。在一个统一的理论框架下，开发一套全新的地震数据分析软件系统，全面提升油气地震勘探的自动化和智能化水平。

2 Bayes 参数估计理论与方法

从上述统计数据分析的思想，所谓的 Bayes 参数估计，就是在 Bayes 统计数据分析的思想下，基于总

体随机向量 $\mathbf{X}^t$ 的样本随机向量 $\mathbf{X}=(X_1,X_2,\cdots,X_n)$ 的具体实现 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$，估计统计模型 $\{P_\theta,\theta\in\Theta\}$ 中的参数。

勘探地震中，统计模型就是由观测的地表波场和待估计的介质参数模型共同决定的后验概率密度函数 $P(\mathbf{m}|\mathbf{d})$。Bayes 估计的思想可以简单概括为：估计的结果 $\hat{\mathbf{m}}$ 是由后验概率密度函数 $P(\mathbf{m}|\mathbf{d})$ 所决定的均值 $E(\mathbf{m}|\mathbf{d})$，其中 $\mathbf{d}$ 代表观测数据。已知后验概率密度函数 $P(\mathbf{m}|\mathbf{d})$ 后，估计结果写为，

$$\hat{\mathbf{m}}=E\{\mathbf{m}|\mathbf{d}\}=\int\mathbf{m}P(\mathbf{m}|\mathbf{d})d\mathbf{m} \tag{1}$$

式(1)表示参数估计的结果是由观测数据所对应的模型参数出现的概率（即后验概率密度函数）所决定的。同时，参数估计结果的方差（一定程度上代表估计精度）也可以由下式确定

$$\sigma_{\hat{\mathbf{m}}}=\int(\mathbf{m}-\hat{\mathbf{m}})^2P(\mathbf{m}|\mathbf{d})d\mathbf{m}$$

由于后验概率密度 $P(\mathbf{m}|\mathbf{d})$ 很难估计，所以将其转化为较容易估计的模型空间的先验概率密度函数 $P(\mathbf{m})$、数据空间的概率密度函数 $P(\mathbf{d})$ 和条件概率密度函数 $P(\mathbf{d}|\mathbf{m})$ 来表示，其中 $P(\mathbf{d}|\mathbf{m})$ 体现了正算子的不确定性，正算子描述了从模型空间到数据空间的映射关系，它体现了待估计参数与实测数据之间的关系并非完全正确的，预测结果是存在不确定性的。后验概率密度由 Bayes 公式描述为

$$P(\mathbf{m}|\mathbf{d})=\frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})} \tag{2}$$

Bayes 公式中的三个概率密度函数，在不同的应用中应该有不同的解释。在参数估计问题中，我们给出如下的含义说明。关于模型空间的概率密度函数 $P(\mathbf{m})$，首先体现了待估计结果被认为是随机的，先验信息是假定已知的概率分布函数。很多时候，我们假定 $P(\mathbf{m})$ 是满足 Gauss 分布的，其均值和方差在待估计参数的求解过程中主要起到约束解的取值范围的作用。如果 $P(\mathbf{m})$ 某个分量的方差较小，代表由先验信息所决定的该分量的值的可靠性就较高，因此该分量的估计结果主要由先验信息所决定；如果 $P(\mathbf{m})$ 某个分量的方差较大，代表由先验信息所决定的该分量的值的可靠性就较低，因此该分量的估计结果主要由观测数据所决定。$P(\mathbf{m})$ 初步定义了反演问题的解在整个模型空间的概率密度函数，而反演问题的求解过程就是利用所观测到的数据更新这一概率密度函数，把它转化成后验概率密度函数的过程。关于数据空间的概率密度函数 $P(\mathbf{d})$，在 Bayes 框架下，我们认为没有关于实测数据的先验信息。这样的先验假设应该是合理的。因此，概率密度函数 $P(\mathbf{d})$ 被认为是均匀分布，就是说采样到什么数据的概率都是相等的。事实上，$P(\mathbf{d})=\int P(\mathbf{d}|\mathbf{m})P(\mathbf{m})d\mathbf{m}$ 不影响后验概率密度 $P(\mathbf{m}|\mathbf{d})$ 的极值分布，即不影响 MAP 估计结果。

关于条件概率密度函数 $P(\mathbf{d}|\mathbf{m})$，它表示在给定模型 $\mathbf{m}$ 的条件下，出现数据 $\mathbf{d}$ 的可能性的大小。实际上，$P(\mathbf{d}|\mathbf{m})$ 显然应该与用来描述实际物理现象的正算子的精确程度，以及实测数据本身的不确定性决定。后者主要是观测仪器、噪音、非规则观测引起的不确定性。这两种因素的影响很难区分开，即很难进一步分解成譬如 $P^{Model}(\mathbf{d}|\mathbf{m})$ 和 $P^{Noise}(\mathbf{d}|\mathbf{m})$。但是，我们要清楚地知道是这两种因素决定了 $P(\mathbf{d}|\mathbf{m})$。

根据 Bayes 公式可以看出，$P(\mathbf{m}|\mathbf{d})$ 由 $P(\mathbf{m})$、$P(\mathbf{d})$ 和 $P(\mathbf{d}|\mathbf{m})$ 来决定。假设 $P(\mathbf{d})$ 均匀分布，没有引入先验信息的情况下，参数估计问题的求解可以认为是在 $P(\mathbf{m})$ 和 $P(\mathbf{d}|\mathbf{m})$ 的约束下，得到估计结果由期望均值为 $\hat{\mathbf{m}}=\int\mathbf{m}P(\mathbf{m}|\mathbf{d})d\mathbf{m}$ 决定，方差 $C_{\hat{\mathbf{m}}}=\int(\mathbf{m}-\hat{\mathbf{m}})^2P(\mathbf{m}|\mathbf{d})d\mathbf{m}$ 决定。

更可行的做法是假设 $P(\mathbf{m})$ 和 $P(\mathbf{d}|\mathbf{m})$ 均为高斯函数，后验概率密度函数也是 Gauss 分布。此时，后验概率密度函数最大化（MAP）对应了代价函数

$$S(\mathbf{m})=\frac{1}{2}\left[(g(\mathbf{m})-\mathbf{d}_{obs})^HC_D^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})+\lambda(\mathbf{m}-\mathbf{m}_{prior})^HC_M^{-1}(\mathbf{m}-\mathbf{m}_{prior})\right] \tag{3}$$

的最小化。式中：C_D 为数据协方差矩阵；C_M 为模型的协方差矩阵。

如果正演问题 $g(\mathbf{m})$ 是线性函数，即 $g(\mathbf{m})=G\mathbf{m}$，并且矩阵（$G^H C_D^{-1} G+\lambda C_M^{-1}$）是正定对称的，则 $S(\mathbf{m})$ 是一个正定二次函数，有唯一的极小值点。求解 $S(\mathbf{m})$ 极小值点的收敛速度和解的稳定性取决于矩阵（$G^H C_D^{-1} G+\lambda C_M^{-1}$）条件数的大小，因此可以用 Newton 类的方法快速得到其最优解。如果矩阵（$G^H C_D^{-1} G+\lambda C_M^{-1}$）是不定的，即其特征值有正有负，则要先对 $S(\mathbf{m})$ 做适当的正则化处理（在 $S(\mathbf{m})$ 的构造过程中加入适当的约束条件），再求取其极小值点。

事实上，对先验概率密度函数引入 Gauss 假设，相当于对正问题引入了假设，也可以认为对初始值引入了假设，此时认为在初始值附近正问题能很好地解释实际观测数据，余下的“噪声”成分满足 Gauss 分布，待估计参数与初始值的差也满足 Gauss 分布。下图引用 Tarantola（2005）第三章给出的由正问题（待估计参数与实测数据之间的关系）决定的参数估计问题由弱到强的非线性性示意图说明参数估计问题的强非线性本质上是正问题（待估计参数与实测数据之间的关系）决定的。正问题又是先验的，说明奠基在正问题上的反问题也是具备一定的主观性的。也许正是这种主观性反映了参数反演问题是依据人的认识，依据人对要解决的实际问题的认识程度，决定反问题能解决到的程度。这到底是好事还是坏事？反问题的提出和求解的本质就是如此，无所谓好坏。只有对要解决的具体问题的物理实质有深刻的理解，才可提出能解决实际问题的有效的方法。

至于参数估计问题的思想本质，就是这样了。Bayes 参数估计理论已经把思想讲得很明白了。遗憾的是，即便明白了 Bayes 估计的思想，距离提出能解决实际问题的有效方法还是隔着对实际问题物理本质的深刻理解，也包括对各种算法的深刻把握。

如果正问题 $g(\mathbf{m})$ 是非线性问题，代价函数 $S(\mathbf{m})$ 往往是个非凸的函数，那么 $S(\mathbf{m})$ 的极小值点可能是非常多的。为了让代价函数 $S(\mathbf{m})$ 成为一个凸的函数，可能的做法之一是给定一个非常接近真解的初始解 $\mathbf{m}_0$；或者修改数据和正问题把 $S(\mathbf{m})$ 变成凸问题，但这已经不是在解原问题了，而是在解原问题的一个邻近问题。

具体计算方面，把参数估计问题视为在“巨大”的可能的解空间中探索使后验概率密度函数 $P(\mathbf{m}|\mathbf{d})$ 最大或使误差泛函 $S(\mathbf{m})$ 达到最小的可能解应是体现了求解参数估计问题的本质。

基于上述观点来理解参数估计问题的求解，可以认为，在确定性的观点下利用梯度导引类的方法求解非线性反演问题仅仅是在初始值附近的局部范围内对可能的模型参数进行采样。但勘探地震学中参数估计的解空间规模巨大，初始值决定了梯度导引类算法对解空间的探索范围。事实上，梯度导引类算法给出的都是局部最优解。初始值的选择当然至关重要，但正确的初始值获取也极为困难，如果初始值的选择不能接近真实值，梯度导引类的算法将不能得到有意义的反演结果。另一方面待估计参数的描述方式会直接影响到解空间的规模，要尽可能选择最佳的参数描述方式以使得最终的解空间规模足够小。

另一个在“巨大”的可能的解空间中探索可能解的方法是 Monte Carlo“重要性”采样方法。与梯度导引类方法不同，Monte Carlo 反演方法是基于一定的概率密度函数在“巨大”的可能解空间中进行“随机”采样，而不是像梯度导引类方法在初始值附近对模型空间进行沿梯度方向的采样。Monte Carlo 方法对初始解的依赖性大幅度降低，有可能探索到全局最优解。但是，相比于梯度导引类算法，目前尚无高效的“重要性”采样方法。

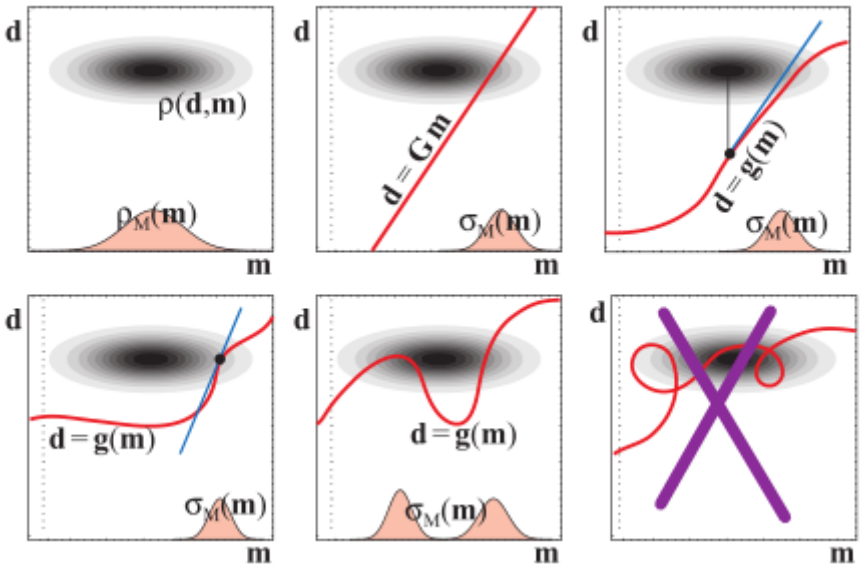


图 1 由正问题决定的参数估计问题由弱到强的非线性性示意图

到目前为止，可以说并不存在求解强非线性参数估计问题的有效解法。只有凸问题的求解数学上是有明确的求解方法的。实际问题中，对复杂的强非线性反问题的求解，只能基于对具体问题物理本质的理解，构造一系列凸问题的组合逼近要求解的强非线性反问题的解。实际反问题的求解并不是一个纯数学问题。

3 Bayes 决策理论与方法

对一个假设或认定进行正确性（合理性）判断、对一个决策行动进行代价评估，在统计数据分析中占有重要地位。这类问题是人工智能领域遇到的核心问题，典型的，譬如分类问题、聚类问题、强化学习问题。可以说人工智能主要是在解决智能决策问题。Bayes 决策理论奠定了智能决策的理论基础。

3.1 Bayes 分类理论框架

理论上讲，Bayes 决策是奠基在假设检验理论基础上的。最基本的假设检验理论，Newman-Pearson 引理，奠定了二分判决问题的原则。很显然，任何统计判决都是基于总体随机向量 $\mathbf{X}^t$ 的样本 $\mathbf{X}=(X_1,X_2,\cdots,X_n)$ 的实测结果 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$，就是总是依赖于数据的。另外，一般地，总假设已知统计模型 $\mathbf{X}\sim\{P_\theta,\theta\in\Theta\}$。分类判决问题和行动决策问题是两类典型的统计决策问题。假设检验理论提供了统计决策的基本原则。

Markov 最佳状态转移决策问题（强化学习的原型）和分类（聚类）问题是 Bayes 决策理论的两个典型应用。下面以 Bayes 分类问题为例讨论 Bayes 决策理论与方法。

分类问题提法：有由原数据中提出的特征属性向量 $\mathbf{x}$，有 $\mathcal{M}$ 个后验条件概率 $p(\omega_i|\mathbf{x}),i=1,2,\cdots,\mathcal{M}$，Bayes 分类问题即用 $\mathcal{M}$ 个后验条件概率密度实现最佳分类。

这显然对应决策性假设检验问题。

Bayes 规则为：$p(\omega_i|\mathbf{x})=\dfrac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(\mathbf{x})},p(\mathbf{x})=\sum_{i=1}^{\mathcal{M}}p(\mathbf{x}|\omega_i)p(\omega_i)$。其中，$p(\omega_i),i=1,2,\cdots,\mathcal{M}$ 代表 $\mathcal{M}$ 类的先验概率密度；$p(\mathbf{x}|\omega_i),i=1,2,\cdots,\mathcal{M}$ 代表 $\mathcal{M}$ 类的先验条件概率密度，是类 ω_i 对应的特征属性向量 $\mathbf{x}$ 的条件概率密度。显然，$p(\mathbf{x})$ 对所有类先验概率密度函数之积求和，不会影响分类结果，因此，解决分类问题时可以忽略分母项 $p(\mathbf{x})$。

对于二分问题，Bayes 分类规则为：若 $p(\omega_1|\mathbf{x})>p(\omega_2|\mathbf{x})$，则 $\mathbf{x}$ 属于 ω_1；若 $p(\omega_1|\mathbf{x})<p(\omega_2|\mathbf{x})$，则 $\mathbf{x}$ 属于 ω_2。

Bayes 分类规则利用后验条件概率密度关系：$p(\mathbf{x}|\omega_1)p(\omega_1) \gtrless p(\mathbf{x}|\omega_2)p(\omega_2)$。事实上，这隐含地对应 MAP 原则。

若取先验概率密度函数 $p(\omega_1)=p(\omega_2)$，可以认为没有关于分类的先验信息，则有 $p(\mathbf{x}|\omega_1) \gtrless p(\mathbf{x}|\omega_2)$。这对应 ML 原则。这显然与似然比假设检验准则是一致的。但是，这样的判决可能发生错误。如图 2 所示。

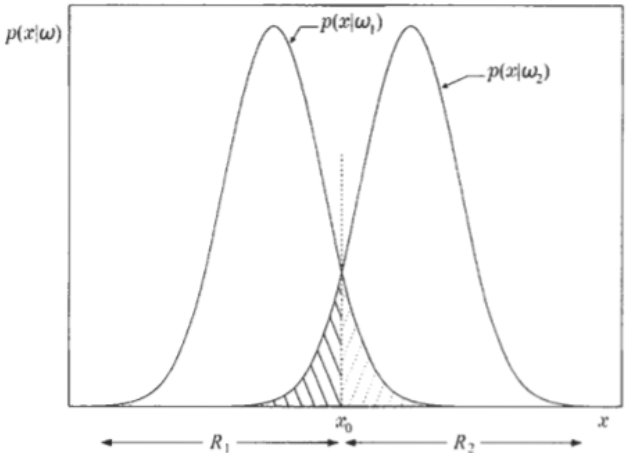


图 2ML 原则下二分分类错误率示意图

由图 2 知，二分分类的错误率可以定义为：

$$p_e = \int_{-\infty}^{\mathbf{x}_0} p(\mathbf{x}|\omega_2)d\mathbf{x} + \int_{\mathbf{x}_0}^{\infty} p(\mathbf{x}|\omega_1)d\mathbf{x} \tag{4}$$

$\mathcal{M}$ 类分类问题中，Bayes 分类原则应是：

$$p(\omega_1|\mathbf{x})p(\omega_1) > p(\omega_j|\mathbf{x})p(\omega_j), \forall j \neq i \text{。} \tag{5}$$

上述分类原则显然不是最佳的，因为不考虑决策错误付出的代价。为此，对二分问题，定义的代价函数为：

$$r = \lambda_{12}p(\omega_1)\int_{R_2} p(\mathbf{x}|\omega_1)d\mathbf{x} + \lambda_{21}p(\omega_2)\int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x} \tag{6}$$

其中，λ_{12} 和 λ_{21} 分别表示把类 1 归为类 2 的代价指标及把类 2 归为类 1 的代价指标。

对于多分问题，同样地有代价函数，

$$r_k = \sum_{i=1}^{\mathcal{M}} \lambda_{ki}p(\omega_k)\int_{R_i} p(\mathbf{x}|\omega_k)d\mathbf{x} \tag{7}$$

Bayes 原则下，引入平均风险函数：

$$r = \sum_{i=1}^{\mathcal{M}} r_k p(\omega_k) = \sum_{i=1}^{\mathcal{M}} \left(\int_{R_i} \lambda_{ki} p(\mathbf{x}|\omega_k)p(\omega_k)d\mathbf{x} \right) \tag{8}$$

很显然，平均风险函数最小应是分类决策的更好准则。

若 $\mathbf{x} \in R_i$，记 $l_i = \sum_{i=1}^{\mathcal{M}} \lambda_{ki} p(\mathbf{x}|\omega_k)p(\omega_k)$，$l_j = \sum_{k=1}^{\mathcal{M}} \lambda_{kj} p(\mathbf{x}|\omega_k)p(\omega_k)$，必有 $l_i < l_j$。

对于二分问题，当然有：

$$\lambda_{11}p(\mathbf{x}|\omega_1)p(\omega_1) + \lambda_{21}p(\mathbf{x}|\omega_2)p(\omega_2) < \lambda_{12}p(\mathbf{x}|\omega_1)p(\omega_1) + \lambda_{22}p(\mathbf{x}|\omega_2)p(\omega_2) \tag{9}$$

$$(\lambda_{11}-\lambda_{12})p(\mathbf{x}|\omega_1)p(\omega_1) < (\lambda_{21}-\lambda_{22})p(\mathbf{x}|\omega_2)p(\omega_2) \tag{10}$$

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} < \frac{(\lambda_{21}-\lambda_{22})p(\omega_2)}{(\lambda_{11}-\lambda_{12})p(\omega_1)} \tag{11}$$

据此，定义似然比：

$$l_{12} = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)}。 \tag{12}$$

对于分类问题的最小错误率，必有

$$p(\omega_i|\mathbf{x}) - p(\omega_j|\mathbf{x}) = 0, i, j = 1,2,\cdots,\mathcal{M}; i \neq j \tag{13}$$

此时，$\mathbf{x}$ 判归 ω_i 或 ω_j 已经无法决定，认为判定错误率最小。很明显，上式定义了一个决策面。

若定义一个单调上升函数 $f(\bullet)$，使 $g_i = f\left(p(\omega_i|\mathbf{x})\right)$，$g_i$ 称为判决函数，则若 $g_i(\mathbf{x}) > g_j(\mathbf{x}), \forall j \neq i$，将 $\mathbf{x}$ 分类到 ω_i。

决策面也可以被描述为，

$$g_{ij}(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0, i, j = 1,2,\cdots,\mathcal{M}, j \neq i \tag{14}$$

上述就是用于解决分类问题的 Bayes 决策理论。根本思想是决策（分类）错误率最小或风险概率最小。理论本身是高度抽象与包容的。但是，在解决实际问题时，$p(\mathbf{x}|\omega_i)$、$p(\omega_i)$ 和 $p(\omega_i|\mathbf{x})$，$i = 1,2,\cdots,\mathcal{M}$，一般是未知的，也可能极为复杂。因此，发展各种方法直接计算决策面可能是更实用的、可行的路线，尽管这样的做法可能会降低分类器的性能。

3.2 Gauss 分布假设下的 Bayes 分类

由于 Bayes 原则中若先验概率密度分布具有一般形式则很难给出实用性的分类算法，类似于前述参数估计理论，对先验概率密度分布引入 Gauss 先验假设是非常必要的。

假设先验概率密度函数 $p(\mathbf{x}|\omega_i)$，$i = 1,2,\cdots,\mathcal{M}$ 是多元正态分布 $\mathcal{N}(\boldsymbol{\mu}_i, \Sigma_i)$，$i = 1,2,\cdots,\mathcal{M}$，即

$$p(\mathbf{x}|\omega_i) = \frac{1}{\sqrt[n/2]{2\pi}\sqrt{|\Sigma_i|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}。 \tag{15}$$

对应的判决函数为：

$$\begin{aligned} g_i(\mathbf{x}) &= \ln\left(p(\mathbf{x}|\omega_i)p(\omega_i)\right) = \ln\left(p(\mathbf{x}|\omega_i)\right) + \ln p(\omega_i) \\ &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln p(\omega_i) + c_i \\ &= -\frac{1}{2}\mathbf{x}^T\Sigma_i^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\Sigma_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\boldsymbol{\mu}_i^T\Sigma_i^{-1}\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_i^T\Sigma_i^{-1}\mathbf{x} + \ln p(\omega_i) + c_i \end{aligned} \tag{16}$$

其中，$c_i = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i|$，$i = 1,2,\cdots,\mathcal{M}$。

若先验概率密度函数 $p(\mathbf{x}|\omega_i)$，$i = 1,2,\cdots,\mathcal{M}$ 符合多元正态分布 $\mathcal{N}(\boldsymbol{\mu}_i, \sigma_i^2 I)$，$i = 1,2,\cdots,\mathcal{M}$，则对应的判决函数修改为：

$$g_i(\mathbf{x}) = \ln\left(p(\mathbf{x}|\omega_i)p(\omega_i)\right) = -\frac{1}{2\sigma_i^2}\left(x_1^2 + x_2^2\right) + \frac{1}{\sigma_i^2}\left(\mu_{i1}x_1 + \mu_{i2}x_2\right) - \frac{1}{2\sigma_i^2}\left(\mu_{i1}^2 + \mu_{i2}^2\right) + \ln p(\omega_i) + c_i \tag{17}$$

显然，由(16)和(17)式知，$g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0$ 定义的决策面是（超）二次曲面。特征属性 $\mathbf{x}(\mathbf{x} \in \mathcal{X})$ 定义的区域由决策面分隔开。这是由决策面定义的非线性分类问题。

很显然，若 $\Sigma_i=\Sigma_j, i,j=1,2,\cdots,\mathcal{M}$，由(16)式知，$g_i(\mathbf{x})-g_j(\mathbf{x})=0$ 定义的决策面中二次项消失，决策面退化为线性的决策超平面。记决策函数为：

$$g_i(\mathbf{x})=\mathbf{w}_i^T\mathbf{x}+w_{i0}\,, \tag{18}$$

其中 $\mathbf{w}_i=\Sigma^{-1}\boldsymbol{\mu}_i$，$w_{i0}=\ln p(\omega_i)-\frac{1}{2}\boldsymbol{\mu}_i^T\Sigma^{-1}\boldsymbol{\mu}_i$。

对于协方差阵 $\Sigma=\sigma^2 I$ 情形，决策函数修改为：

$$g_i(\mathbf{x})=\frac{1}{\sigma^2}\boldsymbol{\mu}_i^T\mathbf{x}+w_{i0}\ 。 \tag{19}$$

根据决策函数，定义超平面：

$$g_{ij}(\mathbf{x})=g_i(\mathbf{x})-g_j(\mathbf{x})=\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0)=0\,, \tag{20}$$

其中 $\mathbf{w}=\boldsymbol{\mu}_i-\boldsymbol{\mu}_j$；$\mathbf{x}_0=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)-\sigma^2\ln\left(\frac{p(\omega_i)}{p(\omega_j)}\right)\frac{\boldsymbol{\mu}_i-\boldsymbol{\mu}_j}{\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\|^2}$。

若 $p(\omega_i)=p(\omega_j)$，则 $\mathbf{x}_0=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)$。超平面是过 $\mathbf{x}_0$ 点的，$p(\omega_j)>p(\omega_i)$ 或 $p(\omega_i)>p(\omega_j)$，超平面会更接近 $\boldsymbol{\mu}_i$ 或 $\boldsymbol{\mu}_j$。

若方差 σ^2 很小，则超平面位置与 $p(\omega_i)$ 和 $p(\omega_j)$ 的关系不大。这是很显然的。小方差代表随机向量聚在均值附近的小范围段，决策超平面的小幅变动不影响分类结果。

对于一般协方差阵 Σ，超平面关系式为：

$$g_{ij}(\mathbf{x})=g_i(\mathbf{x})-g_j(\mathbf{x})=\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0)=0\,, \tag{21}$$

其中，$\mathbf{w}=\Sigma^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$ $\mathbf{x}_0=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)-\ln\left(\frac{p(\omega_i)}{p(\omega_j)}\right)\frac{\boldsymbol{\mu}_i-\boldsymbol{\mu}_j}{\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\|_{\Sigma^{-1}}^2}$

很显然，根据上述讨论，可以基于判别函数直接定义如下最小距离分类器。为此，直接定义判别函数为：

$$g_i(\mathbf{x})=-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) \tag{22}$$

相当于把不显著影响分类结果的常数项忽略。显然这就是 K 均值分类的基本原则。判别函数代表特征属性向量 $\mathbf{x}$ 与类均值的距离。把特征属性向量 $\mathbf{x}$ 与类均值距离的最小化作为分类原则自然是合理的。类内方差最小，类间方差最大是此类分类方法期望特征属性 $\mathbf{x}(\mathbf{x}\in\mathcal{X})$ 应该具备的特征。

这同样分为 $\Sigma=\sigma^2 I$ 和一般协方差阵情形。对于后者，把一般协方差阵进行正交分解 $\Sigma=Q\Lambda Q^T$，则判别函数修改为：

$$\begin{aligned} g_i(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) \\ &= -\frac{1}{2}\left[Q(\mathbf{x}-\boldsymbol{\mu}_i)\right]^T\Lambda^{-1}Q(\mathbf{x}-\boldsymbol{\mu}_i)\,, \\ &= -\frac{1}{2}(\mathbf{x}'-\boldsymbol{\mu}_i')^T\Lambda^{-1}(\mathbf{x}'-\boldsymbol{\mu}_i') \end{aligned} \tag{23}$$

这样就可以把两种协方差阵情形统一。

4 分类问题的非线性性分析

应该意识到特征属性向量 **x** 的分类特征可能是十分复杂的。尤其在原域中，特征属性向量 **x** 的分布在各分类 $\omega_i\,(i=1,2,\cdots,\mathcal{M})$ 间可能犬牙交错。图 3 示意性地展示了由强非线性到线性的分类问题特征。

与强非线性的参数估计问题一样，同样缺乏有效的此类问题的解法。线性超平面分类方法对应与图 3 中最右图。也许二次曲面决策面分类方法能解决图 3 中右 2 图所示的问题。

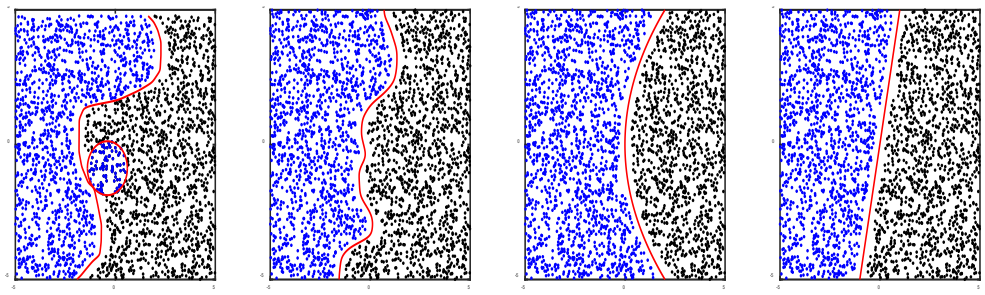


图 3 分类问题的非线性性的示意图

分类问题是否容易，根本地决定于(4)式定义的错误率。而错误率显然取决于各类后验概率密度函数 $P(\omega_i|\mathbf{x}),i=1,2,\cdots,\mathcal{M}$ 的分布情形。各类后验概率密度函数分得越开，分类错误越小。为了改善各类后验概率密度函数的可区分度，主要是从原数据中提取出特征（属性）参数向量 **x**，期望特征属性向量 **x** 对应的各类后验概率密度函数间均值差异大，各类后验概率密度函数的方差尽可能小。这严重依赖于特征表达的程度，即特征属性参数与类 $\omega_i\,(i=1,2,\cdots,\mathcal{M})$ 的关系要非常密切，最好是一对一的、线性的。先验信息 $P(\omega_i),i=1,2,\cdots,\mathcal{M}$ 的引入也是很关键的。

图 4 示意性地展示了特征表达的好坏对可区分度的影响。对于好的可区分度，很粗糙的分类方法都可以实现正确的分类；差的可区分度，高精度的分类方法也会有很高的分类错误率。

由此可见，高端的预处理、合理的先验信息的引入对于降低分类问题的非线性性，提高分类正确度是非常重要的。对比可知，非线性参数估计问题中，我们也利用了这样的解决问题的思路，譬如去除多次波，仅仅进行一次波成像。

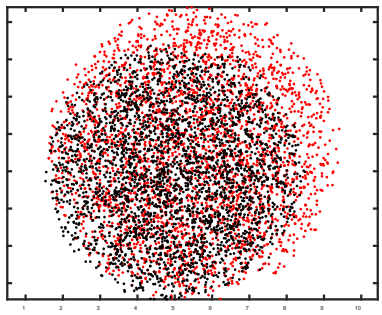


图 4a 没有特征表达的分类问题

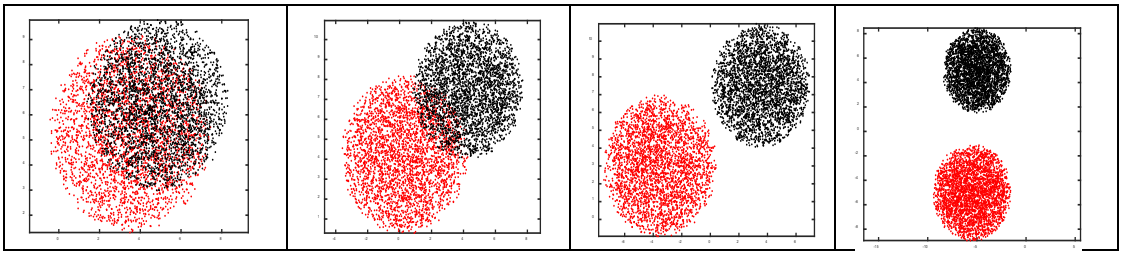


图 4b 各种不同程度的特征表达后的分类问题

5 Bayes 参数估计与 Bayes 决策理论方法的对比分析

由前面的讨论知道，Bayes 参数估计问题与 Bayes 决策问题的基本出发点是一致的，都是基于后验概率密度函数的最大化原则。由于后验概率密度函数在实际问题中难于获取，假设它满足 Gauss 概率分布。据此，Bayes 参数估计问题退化为误差泛函求极小的问题；Bayes 分类决策问题退化为二次型或线性决策面的确定问题。

为清晰起见，下面列表对比二者，以便凸显二者的区别与联系。

表 1 Bayes 参数估计问题与 Bayes 决策问题基本思想列表

$\rho_{\mathbf{M}}(\mathbf{m}\mid\mathbf{d})=\frac{\rho(\mathbf{d},\mathbf{m})}{\rho(\mathbf{d})}=\frac{\rho(\mathbf{d}\mid\mathbf{m})\rho(\mathbf{m})}{\rho(\mathbf{d})}$	$p(\omega_i\mid\mathbf{x})=\frac{p(\mathbf{x}\mid\omega_i)p(\omega_i)}{p(\mathbf{x})},$ $p(\mathbf{x})=\sum_{i=1}^{\mathcal{M}}p(\mathbf{x}\mid\omega_i)p(\omega_i)$
$\hat{\mathbf{m}}_{\mathrm{MAP}}=\underset{\mathbf{m}}{arg\ max}\ \rho_{\mathbf{M}}(\mathbf{m}\mid\mathbf{d})$	$\omega_i=\underset{1<i<\mathcal{M}}{arg\ max}\ P(\omega_i\mid\mathbf{x})$
$\rho(\mathbf{d}\mid\mathbf{m})\sim\mathcal{N}(\mathbf{d}^{obs},C_{\mathbf{d}})$ $\rho(\mathbf{m})\sim\mathcal{N}(\mathbf{m}_{prior},C_{\mathbf{m}})$	$P(\mathbf{x}\mid\omega_i)\sim\mathcal{N}(\boldsymbol{\mu}_i,\Sigma_i)$ $p(\mathbf{x}\mid\omega_i)=\frac{1}{\sqrt[n]{2\pi}\sqrt{\lvert\Sigma_i\rvert}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}$
$\hat{\mathbf{m}}=\underset{\mathbf{m}}{arg\ min}\left\{\begin{matrix}(\mathbf{m}-\mathbf{m}_{prior})^T\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})\\+(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})^T\mathbf{C}_{\mathbf{D}}^{-1}(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})\end{matrix}\right\}$	$p(\omega_i\mid\mathbf{x})-p(\omega_j\mid\mathbf{x})=0,i,j=1,2,\cdots,\mathcal{M};i\neq j$ $g_i=f\left(p(\omega_i\mid\mathbf{x})\right)$ $g_{ij}(\mathbf{x})=-(\frac{1}{2}\mathbf{x}^T\Sigma_i^{-1}\mathbf{x}-\frac{1}{2}\mathbf{x}^T\Sigma_j^{-1}\mathbf{x})$ $+(\frac{1}{2}\mathbf{x}^T\Sigma_i^{-1}\boldsymbol{\mu}_i-\frac{1}{2}\mathbf{x}^T\Sigma_j^{-1}\boldsymbol{\mu}_j)$ $+(\frac{1}{2}\boldsymbol{\mu}_i^T\Sigma_i^{-1}\mathbf{x}-\frac{1}{2}\boldsymbol{\mu}_j^T\Sigma_j^{-1}\mathbf{x})$ $-(\frac{1}{2}\boldsymbol{\mu}_i^T\Sigma_i^{-1}\boldsymbol{\mu}_i-\frac{1}{2}\boldsymbol{\mu}_j^T\Sigma_j^{-1}\boldsymbol{\mu}_j)+c_i-c_j=0$
$\mathbf{d}_{obs}=\mathbf{g}(\mathbf{m}_{\mathrm{B}}+\Delta\mathbf{m})=\mathbf{g}(\mathbf{m}_{\mathrm{B}})+\frac{\partial\mathbf{g}}{\partial\mathbf{m}}\Delta\mathbf{m}+\frac{1}{2}\frac{\partial^2\mathbf{g}}{\partial\mathbf{m}^2}\Delta\mathbf{m}^2+\cdots$ $\Delta\mathbf{d}_{obs}\approx\mathbf{g}(\mathbf{m}_{\mathrm{B}}+\Delta\mathbf{m})-\mathbf{g}(\mathbf{m}_{\mathrm{B}})=\frac{\partial\mathbf{g}}{\partial\mathbf{m}}\Delta\mathbf{m}$ $\Delta\mathbf{d}_{obs}=\mathrm{G}\Delta\mathbf{m}$	$\Sigma_i=\Sigma_j,i,j=1,2,\cdots,\ \mathcal{M}$
$\hat{\mathbf{m}}=\underset{\mathbf{m}}{arg\ min}\left\{\begin{matrix}(\mathbf{m}-\mathbf{m}_{prior})^T\mathbf{C}_{\mathbf{M}}^{-1}(\mathbf{m}-\mathbf{m}_{prior})\\+(\mathrm{G}\Delta\mathbf{m}-\Delta\mathbf{d}_{obs})^T\mathbf{C}_{\mathbf{D}}^{-1}(\mathrm{G}\Delta\mathbf{m}-\Delta\mathbf{d}_{obs})\end{matrix}\right\}$	$g_{ij}(\mathbf{x})=\mathbf{w}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_0)=0$ $\mathbf{w}=\Sigma^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j),$ $\mathbf{x}_0=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)-\ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right)\frac{\boldsymbol{\mu}_i-\boldsymbol{\mu}_j}{\lVert\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\rVert_{\Sigma^{-1}}^2}$

6 Bayes 参数估计与 Bayes 决策理论方法在地震数据分析中的应用

前已述及，统计数据分析主要解决三大类问题：参数估计问题、假设检验问题和回归分析问题。事实上，解决回归分析问题的实质是将参数估计问题和假设检验问题的解决方案联合起来。

地震数据分析，根本目的是基于叠前地震数据和与待估计介质参数相关的先验信息，通过某种形式的波动方程建立起待估计参数和叠前地震数据之间的预测关系，在 Bayes 参数估计理论基础上，提出合适的参数估计反问题，在有效的数值求解方法下，得到期望的参数估计结果。这就是所谓的地震波反演成像结果。目前主要是以背景速度、（角度）反射系数的形式体现的。FWI 和 LS_RTM 就是当前地震波反演成像方法的代表。以 Zoeppritz 方程和（角度）反射系数为基础的 AVA 反演，也是勘探地震领域的一个经典的参数估计问题。广义的地震数据分析，应该包括根据地震波反演成像结果，综合钻井、测井信息、地质知识、岩石物理知识等对油藏的含油气性进行评价，这更多地表现出统计决策问题的特点。

可以认为：油气地震勘探中全过程的数据分析都应建立在统计数据分析的基础上，从地震数据的预处理、地震波反演成像、与储层评价相关的信息提取、油气藏评价、甚至辅助的钻井决策，都要在 Bayes 参数估计和 Bayes 决策理论基础上进行。

把叠前地震数据视为“随机向量”,认为是总体“随机向量” $\mathbf{X}'$ 的样本的具体实现，假设它是满足某种总体分布的。待估计参数也被认为是一种“随机向量”，往往是具有一定的先验认识的。正是在上述观点下，开展 Bayes 参数估计理论基础上的地震波反演成像。正是表 1 中罗列的地震波反演成像的基本理论决定了对地震数据预处理的理论要求，即把一次初至波、反射波和散射波最佳地从实测数据提取出来，而且要恢复成规则的采样。很显然，这是又一个信号分析中的参数估计问题。抽象地看，信号分析领域和地震波成像领域中的参数估计问题本质上没有区别，差异仅仅是具体问题的性质不同或正问题不同而已。因此，信号分析和地震波成像中的核心问题都应该统一在 Bayes 参数估计理论下进行解决。

当前，地震波成像的重点逐渐在向获取更多的先验信息，从而提升参数反演精度的角度转移。从地震成像剖面上，基于各种机器学习的方式进行层位追踪、断层识别、特殊地质体识别，甚至有地质含义的地震相识别是核心工作。判断追踪、识别结果的合理性，广义地是假设检验理论基础上的最佳判决问题。从速度谱上，代替人工自动地拾取合理的 (T_0, V_{rms}) 对也是个典型的最佳判决问题。从强噪音弱初至炮集中，代替人工自动地获取合理的初至波识别与初至拾取也是个典型的最佳判决问题。代替人工基于测井数据、岩石物理知识等进行油藏描述相关的信息提取和决策也是奠基在 Bayes 决策理论基础上的。上述问题不像 Bayes 参数估计问题那样有体系性。但是，这类问题在提高油气勘探效益方面变得越来越重要。

事实上，随着大数据和人工智能时代的到来，基于多源信息的、定量的地震波成像会成为主要的技术发展方向。那种以某种具体算法为核心的技术发展阶段逐渐让位于综合提取各种信息，并把这些信息有效地利用得到最佳的成像结果和油藏描述结果，提高油气勘探的综合效益。

7 总结与讨论

大数据和人工智能（机器学习）正在并将不断地深刻改变各行各业的技术形态及技术发展方向。勘探地震领域，到目前为止，地震数据分析各环节之间的互融互通并没有形成。本质上，地震数据分析的基础信息包括：叠前地震数据（也包括其他非地震场数据）、与地下介质参数相关的数据（也包括地质知识、岩石物理知识这些非数据形态的信息）。地震数据分析的主要目的包括：地下弹性介质参数的反演成像（参数估计）、储层的识别与含油气性评价（甚至包括辅助钻井决策）。地震数据分析应该奠基在 Bayes 参数估计理论和 Bayes 决策理论基础上，以基于多源信息的、定量的地震波成像和综合信息利用的油藏描述与评价作为两个要解决的主要问题，在一个统一的理论框架下开展方法技术研究，最终发展出全新的地震数据分析软件系统（而不是仅仅着眼于地震数据处理系统）。

当前，存在的最大问题是：对 Bayes 参数估计理论下的地震波成像关注过多，对 Bayes 决策理论下的多源信息提取、信息综合利用、知识表达和应用、基于信息综合的最佳决策关注不够，研究积累不足，导致大量有效信息没有被充分利用。由于最佳决策问题缺乏体系性，解决问题需要综合的知识体系，缺乏能把控这种问题的人才，因此不能发展出真正功能强大的技术体系。

但是，大数据和人工智能（机器学习）已经把地震数据分析推到这样的发展阶段。因此，在 Bayes 参数估计理论和 Bayes 决策理论下诱发出以油气藏描述和评价为核心的地震数据分析技术系统和软件系统势所难免。

8 参考文献

[1]. 陈家鼎，郑忠国，2007，概率与统计（第二版），北京大学出版社

[2]. Bertsekas, D. P. and Tsitsiklis, J. N., 2008, Introduction to Probability, Second Edition， Athena Scientific.
[3]. Kay, S., 1993, Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory, Prentice Hall PRT.
[4]. Kay, S., 1998, Fundamentals of Statistical Signal Processing, Volume I: Detectuon Theory, Prentice Hall PRT.
[5]. Moon, T. K., and Stirling, W. C., 2000, Mathematical Methods and Algorithms for Signal Processing, Prentice Hall.
[6]. Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM.
[7]. Theodoridis, S., 2015, Machine Learning: A Bayesian and Optimization Perspective, ELSEVIER.
[8]. Theodoridis, S., and Koutroumbas, K, 2009, Pattern Recognition, ELSEVIER Inc.

几何在地震数据分析中的应用

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】一般高维数据的本质是：高维几何结构飘在满足不同概率分布的随机噪声中。高维地震数据的本质是：局部线性或非线性的同相轴飘在满足不同概率分布的随机噪声中。局部点处，数据是存在相关性的。相关性与数据中蕴含几何结构是对应的。数据分析的本质是：识别出高维数据中的几何结构；近似表达高维数据中的几何结构；利用高维数据中的几何结构。数学分析领域，高维几何结构是用高维函数来表达的。微积分、微分几何、线性代数和泛函分析是研究高维几何结构的基本数学理论。地震数据分析逐步进入到高维统计数据分析、非线性非高斯数据分析、多源数据融合分析的阶段。几何观点是数据分析的结构性思维，对深入认识数据分析的本质有重要价值。本文从几何思维的角度深入讨论地震数据分析的思想，试图把几何思维更多地、更深入地引入到地震数据分析的各个环节中，提升对高维地震数据分析的认识程度。

【关键词】高维数据体，几何结构，高维函数，微积分，Hilbert 空间，基函数族，函数逼近，内积，相关

1 引言

当前，勘探地震学的核心目标依然是：精确地描述与刻画油气藏，解决油气勘探和开发的问题。以地下介质弹性参数估计为目标的地震波反演成像是完成该任务（即精确地描述与刻画油气藏）的最核心技术。

线性成像问题理论上已经定型，值得研究的理论问题很少了，余下的只是某些具体实现中的问题。地震波成像的理论方向已逐渐转到非线性成像，以及多源信息融合下的成像。地震数据中出现强反射（强绕射）同相轴，出现丰富的多次波成分预示着要进行非线性成像。表面上看似能对多次波存在时的地震波进行非线性成像的全波形反演（FWI），算法本质上只是个局部最优化方法，根本不能在稍强的非线性情形下得到有效的结果。有效的非线性成像方法的提出需要全新的思维方式。

当前，地震波反演成像技术发展的重点已逐渐地由对反演算法（Newton 梯度迭代算法）的研究，转到多源信息约束下的提高反演精度的思想方法与技术的研究上来。这与当今大数据、智能化的转型是比较匹配的！油气地震勘探的技术发展也不可能脱离这样的大趋势。暂时撇开非线性反演成像问题不谈，提高成像精度的合理途径应该是引入多源信息约束。抽象后，多源信息来源包括：来自物理系统的高维规则数据体-叠前地震数据体；来自物理或非物理系统的空间散乱数据体（最终也要通过散乱数据插值形成规则数据体）-散乱空间点处的井数据、反射地层数据等。

地震波反演成像的目标是：三维空间中的弹性参数场。勘探地震中，几何结构+层内弹性参数值是基本的参数描述方法。其中合理的几何结构对于保证弹性参数场反演成像结果的正确性是至关重要的。高精度反演成像中，引入对参数场几何结构的约束能起到非常显著的作用。因此，事先获得沉积结构（构造与断层）、沉积相、特殊地质体的几何结构，对于提升反演成像的精度非常重要。研究表明：中波数成分估计精度的提升，除了“两宽一高”+高信噪比数据观测外，引入合理的地质体几何结构的约束，是最重要的措施。

另一方面，从高维数据体角度看：线性结构（平面中直线、空间中的直线、空间中的平面）和非线性结构（平面中的曲线、空间中的曲线、空间中曲面）飘在满足一定概率密度的随机噪声中是高维地震数据体的基本概念模型。从高维地震数据体中，识别并提取出这些线性结构和（或）非线性结构是数据分析及图像分析的根本任务之一。这项工作在机器学习中被称为特征工程，也是当前信号分析和图像分析领域最核心的问题。

地震数据分析、地震波反演成像、油藏描述是一项十分复杂的工作，需要物理的、微积分的思维和知识，但更需要清晰的几何思维和完整的几何知识。物理+几何的思维模型对于构建起高维数据分析本质问题的清晰认识是必不可少的。在形成清晰的认识后，再进入基于微积分、线性代数、泛函分析和偏微分方程的具体算法处理阶段，可以起到事半功倍的效果。物理+几何的思维模式对于提出有效的非线性成像方法也是有极大帮助的。

地震数据分析逐步进入到高维统计数据分析、非线性非高斯数据分析、多源数据融合分析的阶段，几何观点是数据分析的结构性思维，对深入认识数据分析的本质有重要价值。因此，本文从几何思维的角度深入讨论地震数据分析的思想，试图把几何思维更多地、更深入地引入到地震数据分析的各个环节中，提升对高维地震数据分析的认识程度。

2 高维数据分析与几何的相关性

局部线性或非线性的同相轴飘在满足不同概率分布的随机噪声中是对高维地震数据的高度抽象。一般的高维数据也脱不了高维几何结构飘在满足不同概率分布的随机噪声中的抽象模型。

众所周知，平面中直线、空间中的直线、空间中的平面；平面中的曲线、空间中的曲线、空间中曲面首先是几何学的研究对象。解析几何和微分几何是核心研究方法。抽象地，微积分的研究对象就是函数关系表达的“高维”几何结构。Hilbert 空间中对算子和泛函的研究是“高维”几何结构更为抽象和统一的理论。

因此，用几何观点和思维深化高维数据分析是完全必要的。

基本几何知识包括：平面几何；立体几何；解析几何；微分几何。平面几何（平面三角）研究什么？点与直线；平面三角形；平面四边形；平面上的多边形；圆及与圆相关的图形；平面三角学。立体几何研究什么？空间中的直线与平面；空间中的多面体；曲面规定的立体；球面三角学。解析几何研究什么？向量代数（点积 $\rho = cos\,\varphi = \dfrac{\boldsymbol{a}\cdot\boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2\,|\boldsymbol{b}|^2}}$、叉积 $\rho = sin\,\varphi = \dfrac{\boldsymbol{a}\times\boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2\,|\boldsymbol{b}|^2}}$、混合积）；平面解析几何；立体解析几何/空间解析几何；几何变换与坐标变换。微分几何研究什么？平面曲线；空间曲线；曲面。其中，平面曲线的局部性质：①弧微分；②切线和法线；③曲线的凸与凹；④曲率与曲率半径；⑤曲率圆与曲率中心，以及特殊点：①拐点;②顶线;③奇点都是我们特别关心的。平面曲线的定义域;对称性;渐进行为;间断点;与 y 和 x 轴的交点;极大值与极小值;拐点更是具有重要意义。

地震数据分析（也包括一般的数据分析），都已经进入到非平稳和非高斯情形下对非线性结构进行最佳表达、提取的阶段。清晰的几何思维对于开展深入的高维数据分析是非常重要的。

3 勘探地震中的几何尺度问题

尺度，表面上是一个几何概念。在任何物理学科中，尺度都具有至关重要的作用。不同尺度下的物理规律是显著不同的。尺度甚至决定了因果关系是否成立。把研究因果关系的著名学者 Judea Pearl（CAUSALITY，Second Edition，2009）的论断引用如下：Causality – namely our awareness of what causes what in the world. Though it is basic to human thought, causality is a notion shrouded in mystery, controversy, and caution, because scientists and philosophers have had difficulties defining when one event truly causes another.

我们真的知道地震波场与储层含油气性之间的因果关系吗？关于地震勘探的物理，我们能确知的是什么？尺度是个关键！

图 1 展示了不同尺度的勘探地震介质（微观、介观到宏观）；为了对比起见，表 1 展示了材料科学中的宏观、介观与微观的尺度量级。很显然，任何不同领域中，在不同尺度上控制物理规律的方程是相当不同的。

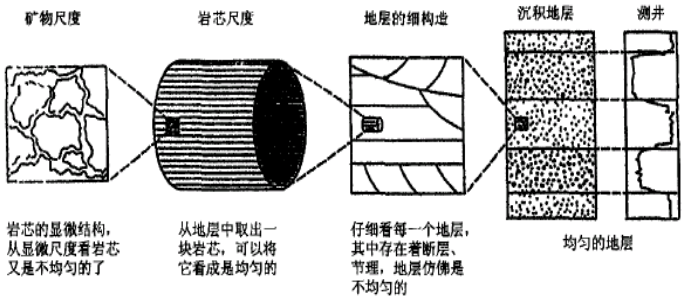


图 1 不同尺度的勘探地震介质（微观、介观到宏观）

尺度	对应空间长度范围 （m）	备注
宏观	10^-2 以上	宏观体系的特点是物理量具有自平均性：即可以把宏观物体看成是由许多的小块所组成，每一小块是统计独立的，整个宏观物体所表现出来的性质是各小块的平均值
介观	10^-4	介观一般指介于宏观和微观之间，尺度在纳米和微米之间，这时量子相干效应起很大作用。（有的学者认为其特征尺度为10-9~10-7 米）。
微观	10^-6	
纳观	10^-9	

表 1 材料科学中的宏观、介观与微观的尺度量级

特征主频	5Hz	15Hz	25Hz	35Hz	45Hz
特征波长	700m	233m	140m	100m	77m

表 2 勘探地震学对应的特征速度、特征主频和特征波长，假定特征速度 3500m/s

按尺度划分的地质体类别	地质特征	尺度量级	数学模型
岩芯	岩石颗粒大小、形态、矿物组分	数毫米到公分（厘米）级别	岩石物理学
储层	岩石颗粒定向排列、节理、裂缝等	数公分到米级别	?
沉积与构造	断层、歼灭、不整合、地震相等	数米、数十米、数百米、甚至数千米的级别	波动方程

表 3 勘探地震中介质、尺度、数学模型及分析方法

在假定特征速度 3500m/s 时，表 2 展示了勘探地震学对应的特征速度、特征主频和特征波长。表 3 展示了勘探地震中，不同的介质、尺度、数学模型及对应的分析方法。

油气勘探的终极目标是对储层的含油气性进行可靠的评价。这依赖于对储层岩性参数（孔隙度、渗透率、饱和度，……）的准确估计。这是测井和岩石物理尺度上研究和解决的问题。储层岩性参数的估计，目前缺乏可靠的数学物理模型。Gassmann 方程描述了孔隙介质岩石物理基础关系。这个方程在什么尺度上对什么样的介质在多大程度上成立并没有明确的物理证明（也很难证明！）。事实上，所有的岩石物理关系都存在很强的假设，普适性缺乏保证。到目前为止，储层评价问题主要还是个“艺术”问题。也可以认为是基于综合信息的最佳推断问题。尺度依然关键制约因素！

地震学是基于波动理论，通过地表激发和观测的叠前地震数据，在 Bayes 估计理论下，估计特征波长尺度上的弹性参数变化（量）。在特征波长尺度上，地震波传播的正问题和相应的弹性参数估计问题都是

科学问题。科学问题的基本特征是整个过程可重复、可检验。勘探地震学解决的是地质体形态的几何刻画问题；特征波长尺度上平均意义下的弹性参数估计问题。基于同相轴上地震子波走时的（大尺度的）背景速度估计问题的数学物理关系是十分明确的，基于同相轴上子波振幅的弹性参数高波数扰动量的估计已经存在很多讲不清楚的物理机制，只能把它们归结为参数估计的精度问题中。也应该注意到不同特征波长的波在介质中传播的物理可能差异很大，超声物理模拟产生的物模地震波场与地震频带内野外地震波场有显著差异。尺度又是个核心问题！

Scientists and philosophers have had difficulties defining when one event truly causes another。在油气勘探领域，这个问题（因果性问题）更为凸显。因为我们试图主要用叠前地震数据解决探区三维空间内的油气藏含油气性的评价问题。但是，地震波场的扰动与油气藏含油气变化之间存在明确的因果关系吗？这是非常令人质疑的！

我对油气勘探的基本观点：

①勘探地震学解决地震特征波长尺度上平均意义下的弹性参数估计问题。广义的地震深度偏移成像解决地质体（包括储层）的几何形态描述问题。核心问题包括：高精度的低中波数成分（各向异性）速度模型建立问题；利用宽带弹性参数估计（主要是宽带波阻抗）解决地质体（包括储层）的弹性参数描述问题。

②测井与岩石物理学结合解决（井点位置处的）储层尺度上岩性参数估计与含油气性评价问题。这是一个基于综合信息的最佳推断问题。

③油气地质学知识综合区域和井点两种信息，给出完整探区的储层含油气评价和钻井决策。这是一个基于综合信息的最佳决策问题。

各个尺度上，控制物理规律（因果关系）的方程不同，导致解决问题的方式迥异!

因果关系和相关关系是数据分析的根本基础。但是，我认为：因果关系是相对成立的；相关关系是普遍成立的。我相信：因果律是存在的！但是否能找到精确的、能用函数表达的因果关系的数学模型很难说！在自然和社会科学中，能找到这种关系的情形是少数的，不能找到这种关系的情形是多数的。勘探地震数据分析和地震波成像中，我们目前所使用的、反映因果关系的数学模型，譬如各种波动方程，都只是在一定的理论假设下、在一定的几何尺度下成立的。对于实际数据和实际介质情形都是近似成立的。因此，有人质疑严格意义下的因果律是否存在不是没有道理的。总体而言，更多地依赖相关关系进行统计推断是会成为今后地震数据分析的主导思想。越是微观尺度上的、与储层有关的数据分析，也更缺乏确定性的函数表达的因果关系，依赖相关关系进行统计推断变得更重要。

尺度，在地震数据分析各环节中，是极为重要的背景因素，它决定着因果关系或是相关关系的存在，当然决定了地震数据分析的基本模式。

4 多信息源的几何意义

当前，地震波反演成像技术发展的重点已逐渐地由对反演算法（Newton 梯度迭代算法）的研究，转到多源信息约束下的提高反演精度的思想方法与技术的研究上来。

首先要解决的问题是不同源数据的信息提取、信息表达。对于信息提取（由数据中提取信息），PCA/ICA/CCA 是线性结构提取的主要方法，可以参考 Machine Learning: A Bayesian & Optimization Perspective 一书中的 19 章前半部分内容，这些是公认的算法。非线性结构提取目前尚无确定性的一套算法，可以参考该书中的 19 章后半部分内容。

数据中蕴含的信息（主要指数据中的线性结构）由特征向量、特征值的线性组合来表示。这种方式是信息表达的典型做法。各种各样的属性提取所做的，本质上，均可归结为 PCA 思想下的特征向量与特征值的获取。

第二、带约束的散乱数据（空间散乱数据）插值成一个规则的数据体，本质上，就是一个多源信息的融合问题。散乱数据插值与一维不规则数据插值很不相同，目前，不存在一个有唯一解的空间散乱数据插值方法。能做的只能是在多源信息约束下，使得散乱数据插值结果更符合物理意义。这是一个定性与定量交织在一起的问题。复杂问题均如此，不能机械地、硬性地去求解一个复杂问题的唯一解，即便找到所谓的“唯一解”，也不是符合或比较符合物理的。

空间散乱数据插值，基本理论奠基在 Laplace 方程/泊松方程数值解，或径向基函数组合进行最佳逼近的理论框架下，这是比较公认的。但是如何引入“合理的”约束是考验研究和应用人员专业水平的。

第三，对于多源信息，假定已得到统一的表达（譬如均已归一化，尺度统一化等等），它们的“共性成分”和对应的“异性成分”体现在什么地方？

计算机并不能自动识别不同物理意义的数据，只有人能区分。至少目前是这样的。数据对于计算机（此处，计算机=算法）而言，本质上是一样的。算法能从数据中看到什么？数据中仅仅包含了线性和非线性结构。各种数据（统一后的各种来源的数据）均如此！这些多源数据的共性成分是什么？我相信，从不同的角度看，会给出不同的结论。但我认为共性成分是数据内和数据间的相关成分，**从几何上看，共性成分是数据间角度一致的部分**。

譬如，$\boldsymbol{a}\in S_1$**和**$\boldsymbol{b}\in S_2$二者相关吗？用$\rho=cos\varphi=\dfrac{\boldsymbol{a}^T\cdot\boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2\,|\boldsymbol{b}|^2}}$来考察。这是提取数据间共性成分最根本的方法，其他方法都是由此变出来的！

譬如，两个二维信息源，如何评价它们之间的几何相关性？先定义它们的梯度向量$\boldsymbol{d}_a,\boldsymbol{a}\in S_1$和$\boldsymbol{d}_b,\boldsymbol{b}\in S_2$，用$\rho=cos\varphi=\dfrac{\boldsymbol{d}_a\cdot\boldsymbol{d}_b}{\sqrt{|\boldsymbol{d}_a|^2\,|\boldsymbol{d}_b|^2}}$（用叉积也可以！）测评二维信息源之间的几何相关性或是共性成分！根本地，数据中内蕴含的几何结构的相似性度量就是它们之间相关性的度量。这种做法与 PCA 方法似乎是不同的，但本质思想都是在利用数据中内蕴含的几何结构。

可以看出，单纯从数据的角度，所谓的多源信息的融合，考虑的主要是不同源数据内在的几何结构的相关性！仅此而已！这显然仅仅是很基础层次上的多源信息融合。

第四、更高一级的信息融合一定是要在知识层次上的！

两个物理量$\boldsymbol{x}\in S_1$和$\boldsymbol{y}\in S_2$，二者是什么关系？一般地，表达为函数关系：$\boldsymbol{y}=f(\boldsymbol{x})$。这个关系是如何建立的？物理量之间函数关系（因果关系）的建立是物理学研究的问题，通过观察、观测、实验、规律总结、验证，然后形成函数关系：$\boldsymbol{y}=f(\boldsymbol{x})$。这是大大小小物理关系形成的基本逻辑。

若有很多$\boldsymbol{x}$与$\boldsymbol{y}$的、确认有关系的数据样本，也可用神经网络学习出一个关系，用网络结构表达的关系。这是当前有监督学习的基本模式。

另外，专家知识也可抽象成几点规则来定性地描述$\boldsymbol{x}$与$\boldsymbol{y}$之间的关系。

第五，多源信息融合的目的是为决策提供素材。从决策的角度反观多源信息融合，可能有助于对多源信息融合提供建议和工作方向。

这是我对多源信息融合的基本观点，地震波成像的理论方向已逐渐转到非线性成像和多源信息融合下的成像。而线性成像问题理论上已经十分定型，值得研究的理论问题很少了！

5 几何在数据分析和图像分析中的应用分析

我首先强调：尽管微积分很重要，无论它多重要，它只是侧重函数微分性、局部性分析的特点，决定了它缺乏全局视野。而几何则不同，它侧重的是研究几何体的全部。尽管它也不放弃几何体的微局部分析，譬如微分几何就是对几何体的微局部进行分析的数学工具。

几何在数据分析和图像分析中的作用主要体现在思维观点的层次上。即用几何思维来看待数据与图像，看待很多有鲜明几何特点的数据与图像分析方法。数据与图像的统计分析思想是更高一个层次的、更抽象的思维逻辑，至少我认为是这样的。尽管不能讲几何思维比概率统计思维低一个层次，但抽象程度上看，概率统计思维一定比几何思维更深刻。几何思维更侧重形象思维。

我眼中的高维数据、高维图像与高维函数（泛函）的几何：高维数据、高维图像与高维函数就是高维空间中的线性或非线性几何结构，就是高维空间中直线、曲线、平面、曲面、几何体。所谓的高维数据分析、图像分析、高维函数（泛函）分析就是对这些几何结构进行分析。

我提出的几个几何（或几何思维）在数据分析和图像分析中应用的范例是：①Hilbert 空间中的坐标系、基函数线性组合的函数表达、基于正交投影的函数逼近。②数据空间噪声空间的正交分解。③向量及函数的自相关与互相关。④函数的凸、极值及最优化。

下面对四个方面做个简述。

①Hilbert 空间中的坐标系、基函数线性组合的函数表达、基于正交投影的函数逼近。这是函数逼近论的核心。这是函数逼近论的理论基础。平方可积的函数空间是 Hilbert 空间，是数据空间，是信号空间，也是图像空间。对高维数据/图像中几何结构的建模预测基本逻辑：选择基函数族（字典）的线性组合

$\sum_{i=1}^{L}\lambda_i\mathbf{e}_i^s$ 表达高维数据/图像中几何结构；基于正交投影实现对几何结构的最佳预测，从而解决各种各样的一大类应用问题。这是对高维数据/图像的建模问题，常被归结为模型驱动的数据分析方法。

②数据空间与噪声空间的正交分解。假设数据与噪声是不相关的、统计正交的。正交的定义：（正交）设 H 是内积空间，$\mathbf{x},\mathbf{y}\in H$，$M,N\in H$。(1) $\mathbf{x}\perp\mathbf{y}\Leftrightarrow\langle\mathbf{x},\mathbf{y}\rangle=0$；(2) $\mathbf{x}\perp M\Leftrightarrow\forall\mathbf{y}\in M$，都有 $\langle\mathbf{x},\mathbf{y}\rangle=0$；(3) $M\perp N\Leftrightarrow\forall\mathbf{x}\in M,\forall\mathbf{y}\in N$，都有 $\langle\mathbf{x},\mathbf{y}\rangle=0$。

数据空间与噪声空间的分解：$H=\{\mathbf{e}_1^s,\mathbf{e}_2^s,\cdots,\mathbf{e}_L^s;\mathbf{e}_1^n,\mathbf{e}_2^n,\cdots,\mathbf{e}_K^n\}$ 是含噪数据分解对应的 Hilbert 空间；$M=\{\mathbf{e}_1^s,\mathbf{e}_2^s,\cdots,\mathbf{e}_L^s\}$ 是信号子空间；$N=\{\mathbf{e}_1^n,\mathbf{e}_2^n,\cdots,\mathbf{e}_K^n\}$ 是对应的噪声子空间。信号子空间与噪声子空间是互补的。因此，含噪数据可以被表达为：

$$\mathbf{d}=\sum_{i=1}^{L}\lambda_i\mathbf{e}_i^s+\sum_{i=1}^{K}\mu_i\mathbf{e}_i^n \tag{1}$$

据此，可以解决各种各样的一大类相关的应用问题。

Chen & Fomel(2015)对数据空间与噪声空间的正交分解的灵活应用的例子。有实测数据：$\mathbf{d}$；有线性（或非线性）结果的预测（即信号预测）：$\hat{\mathbf{s}}_0=P[\mathbf{d}]$，其中 $P[\bullet]$ 代表预测算子。有预测的噪音：$\hat{\mathbf{n}}_0=\mathbf{d}-\hat{\mathbf{s}}_0=\mathbf{d}-P[\mathbf{d}]$。噪音中一般地会存在残余信号，如何把它们提出来，补回到信号预测中？当然是设计合理的滤波器 $\mathbf{w}$！思想是什么？当然是利用噪音和信号的正交性！

滤波器设计的问题提出：

$$\hat{\mathbf{w}}=\underset{\mathbf{w}}{arg\,min}\left\|\hat{\mathbf{n}}_0-\mathbf{w}P[\mathbf{d}]\right\|^2=\underset{\mathbf{w}}{arg\,min}\left\|\mathbf{d}-P[\mathbf{d}]-\mathbf{w}P[\mathbf{d}]\right\|^2 \tag{2}$$

上述最优化问题的解：$\hat{\mathbf{w}}=\dfrac{\hat{\mathbf{n}}_0^T\hat{\mathbf{s}}_0}{\hat{\mathbf{s}}_0^T\hat{\mathbf{s}}_0}$。图 2 是上述思想的应用效果。

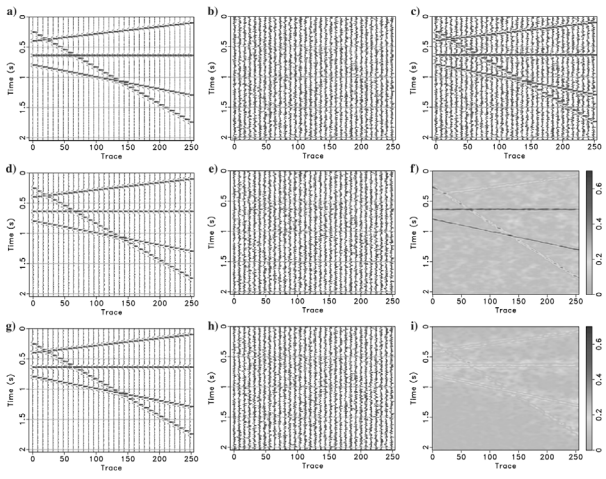


Figure 3. Comparison of using the local orthogonalization-based random noise attenuation approach for denoising test before and after: (a) Clean section, (b) random noise section, (c) noisy section, (d) denoised result using f-x deconvolution, (e) noise section using f-x deconvolution, (f) local similarity map between panels (d and e), (g) denoised result using the proposed method, (h) noise section using the proposed method, and (i) local similarity map between (g and h).

图 2 Sergey Fomel 对数据空间与噪声空间的正交分解的灵活应用

③向量及函数的自相关与互相关。自相关检测的是一个数据体自身包含的线性结构和非线性结构的相关性，原则上，与两个数据体之间的线性结构和非线性结构的相关性检测并没有本质差异。两个数据体之间的线性结构和非线性结构的相关性检测用的是互相关。

抽象地，有“数据向量”：

$$\boldsymbol{a}=\sum_{i=1}^{N}\lambda_i\boldsymbol{\varphi}_i=\left[\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_N\right]\cdot\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{bmatrix},\tag{3}$$

被表达成基函数的线性叠加。此处，所谓的数据向量应该理解为高维数据。譬如，两维数据$\mathbf{d}(x,t)$的频率空间域形式$\mathbf{d}(x,\omega)$。

有数据向量的自相关矩阵及分解：

$$\begin{aligned}\boldsymbol{a}^T\boldsymbol{a}&=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\cdot\begin{bmatrix}\boldsymbol{\varphi}_1\\ \boldsymbol{\varphi}_2\\ \vdots\\ \boldsymbol{\varphi}_N\end{bmatrix}\left[\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_N\right]\cdot\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{bmatrix}\\ &=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\begin{bmatrix}(\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_1) & (\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2) & \cdots & (\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_N)\\ (\boldsymbol{\varphi}_2,\boldsymbol{\varphi}_1) & (\boldsymbol{\varphi}_2,\boldsymbol{\varphi}_2) & \cdots & (\boldsymbol{\varphi}_2,\boldsymbol{\varphi}_N)\\ \vdots & \vdots & \ddots & \vdots\\ (\boldsymbol{\varphi}_N,\boldsymbol{\varphi}_1) & (\boldsymbol{\varphi}_N,\boldsymbol{\varphi}_2) & \cdots & (\boldsymbol{\varphi}_N,\boldsymbol{\varphi}_N)\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{bmatrix}\end{aligned}\tag{4a}$$

（4a）式表明：自相关矩阵元素由$(\varphi_i,\varphi_j),i,j=1,2,\cdots,N$组成，而$cos\,\theta_{i,j}=\dfrac{(\varphi_i,\varphi_j)}{\sqrt{|\varphi_i|^2|\varphi_j|^2}}$。当各基函数正交（不相关）时，自相关矩阵中元素表现为:

$$\begin{aligned}\boldsymbol{a}^T\boldsymbol{a}&=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\cdot\begin{bmatrix}\boldsymbol{\varphi}_1\\ \boldsymbol{\varphi}_2\\ \vdots\\ \boldsymbol{\varphi}_N\end{bmatrix}\left[\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_N\right]\cdot\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{bmatrix}\\ &=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\begin{bmatrix}(\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_1) & 0 & \cdots & 0\\ 0 & (\boldsymbol{\varphi}_2,\boldsymbol{\varphi}_2) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & (\boldsymbol{\varphi}_N,\boldsymbol{\varphi}_N)\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{bmatrix}\end{aligned}\tag{4b}$$

（4）式显然清楚地展示出了高维数据中包含了什么样的正交（不相关）成分，每种成分占比的大小，由特征值的大小决定。可以说两组相同的基函数的组合表达了信号自相关的真正内涵，这就从几何的观点深刻地展示了针对同一个数据体进行数据分析的本质。

同样地，有数据向量的互相关矩阵及分解：

$$\begin{aligned}\boldsymbol{a}^T\boldsymbol{b}&=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\cdot\begin{bmatrix}\boldsymbol{\varphi}_1\\ \boldsymbol{\varphi}_2\\ \vdots\\ \boldsymbol{\varphi}_N\end{bmatrix}\left[\boldsymbol{\psi}_1,\boldsymbol{\psi}_2,\cdots,\boldsymbol{\psi}_N\right]\cdot\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_N\end{bmatrix}\\ &=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\begin{bmatrix}(\boldsymbol{\varphi}_1,\boldsymbol{\psi}_1) & (\boldsymbol{\varphi}_1,\boldsymbol{\psi}_2) & \cdots & (\boldsymbol{\varphi}_1,\boldsymbol{\psi}_N)\\ (\boldsymbol{\varphi}_2,\boldsymbol{\psi}_1) & (\boldsymbol{\varphi}_2,\boldsymbol{\psi}_2) & \cdots & (\boldsymbol{\varphi}_2,\boldsymbol{\psi}_N)\\ \vdots & \vdots & \ddots & \vdots\\ (\boldsymbol{\varphi}_N,\boldsymbol{\psi}_1) & (\boldsymbol{\varphi}_N,\boldsymbol{\psi}_2) & \cdots & (\boldsymbol{\varphi}_N,\boldsymbol{\psi}_N)\end{bmatrix}\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_N\end{bmatrix}\end{aligned}\tag{5a}$$

（5a）式表明：互相关矩阵元素由$\left(\varphi_i,\psi_j\right),i,j=1,2,\cdots,N$组成，而$cos\,\theta_{i,j}=\dfrac{\left(\varphi_i,\psi_j\right)}{\sqrt{\left|\varphi_i\right|^2\left|\psi_j\right|^2}}$。当各基函数正交（不相关）时，互相关矩阵中元素表现为:

$$\begin{aligned}\boldsymbol{a}^T\boldsymbol{b}&=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\cdot\begin{bmatrix}\boldsymbol{\varphi}_1\\ \boldsymbol{\varphi}_2\\ \vdots\\ \boldsymbol{\varphi}_N\end{bmatrix}\left[\boldsymbol{\psi}_1,\boldsymbol{\psi}_2,\cdots,\boldsymbol{\psi}_N\right]\cdot\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_N\end{bmatrix}\\&=\left[\lambda_1,\lambda_2,\cdots,\lambda_N\right]\begin{bmatrix}\left(\boldsymbol{\varphi}_1,\boldsymbol{\psi}_1\right) & 0 & \cdots & 0\\ 0 & \left(\boldsymbol{\varphi}_2,\boldsymbol{\psi}_2\right) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \left(\boldsymbol{\varphi}_N,\boldsymbol{\psi}_N\right)\end{bmatrix}\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_N\end{bmatrix}\end{aligned}\tag{5b}$$

可以说两组不同的基函数的组合表达了信号互相关的真正内涵，这就从几何的观点深刻地展示了针对两个数据体进行数据分析的本质。

事实上，上述自相关和互相关矩阵的分析更多地是对离散数据本身展开的，更多地是基于随机数据的统计分析。此处，我们是先把数据表达成基函数的线性叠加，然后展开自相关和互相关性分析，这与 PCA 的关系更密切，几何意义更凸显。

这是代数问题转化为几何问题的典型。几何含义比代数含义更直观。把测量相似问题转化为测角度问题，可能更适合地震数据与图像的处理。尤其地震数据变化慢，且幅值有正有负，用相关系数ρ测相似性不敏感，用角度θ可能更合适。

Liu & Fomel(2009)对互相关的灵活应用的例子。把一般的互相关$\rho=\dfrac{\boldsymbol{a}^T\cdot\boldsymbol{b}}{\sqrt{\left|\boldsymbol{a}\right|^2\left|\boldsymbol{b}\right|^2}}$写成：

$$\rho^2=\frac{\left(\boldsymbol{a}^T\cdot\boldsymbol{b}\right)^T\left(\boldsymbol{a}^T\cdot\boldsymbol{b}\right)}{\left(\boldsymbol{a}^T\cdot\boldsymbol{a}\right)\left(\boldsymbol{b}^T\cdot\boldsymbol{b}\right)}=\frac{\left(\boldsymbol{b}^T\cdot\boldsymbol{a}\right)\left(\boldsymbol{a}^T\cdot\boldsymbol{b}\right)}{\left(\boldsymbol{b}^T\cdot\boldsymbol{b}\right)\left(\boldsymbol{a}^T\cdot\boldsymbol{a}\right)}=\rho_1\rho_2\tag{6}$$

其中，$\rho_1=\underset{\rho_1}{argmin}\left\|\mathbf{b}-\rho_1\mathbf{a}\right\|=\dfrac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$；$\rho_2=\underset{\rho_2}{argmin}\left\|\mathbf{a}-\rho_1\mathbf{b}\right\|=\dfrac{\mathbf{b}^T\mathbf{a}}{\mathbf{b}^T\mathbf{b}}$；$\rho=\sqrt{\rho_1\rho_2}$。这样重新表达的互相关计算，抗噪性更强。Figure2 清楚地展示了这一点。

事实上，用（5b）的求和式

$$\boldsymbol{a}^T\boldsymbol{b}=\sum_{i=1}^{N}\lambda_i\xi_i\left(\boldsymbol{\varphi}_i,\boldsymbol{\psi}_j\right)\tag{7}$$

应该能进一步提高 Figure2 中叠加的质量，只是计算量会更大一些。

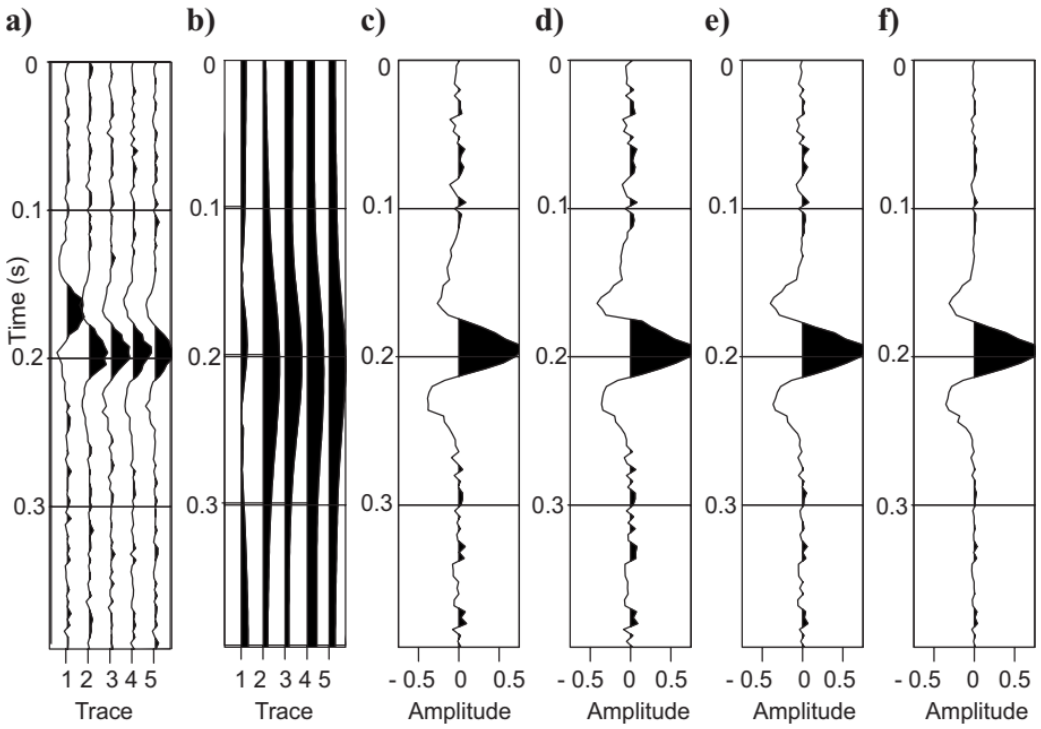


Figure 2. Simple stacking test with fivefold gather. (a) Prestack gather. (b) Weights used in local-correlation weighted stacking. (c) Conventional equal-weight stacking method (S/N = 8.4). (d) Smart stacking method (S/N = 9.2). (e) LMO-based weighted method (S/N = 10.2). (f) Local-correlation weighted stacking (S/N = 13.5).

④函数的凸、极值及最优化。在参数估计的具体计算环节，误差泛函的局部几何性质起着最为关键的作用。

高维函数（泛函）至少局部是凸的（不一定要求是处处可微的），决定了它存在一个局部极值点，形成最优化问题的根本出发点。局部凸的定义是高维函数（泛函）$f(\mathbf{x})$满足：

$$f\left(\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2\right)\le\lambda f(\mathbf{x}_1)+(1-\lambda)f(\mathbf{x}_2) \tag{8}$$

二次型泛函一定是凸的。二次型泛函对应的正问题是线性结构描述的。事实上，在局部点$\mathbf{x}_0$处，对高维函数（泛函）进行 Taylor 展开，

$$f(\mathbf{x})=f(\mathbf{x}_0)+\left.\frac{\partial f}{\partial x}\right|_{\mathbf{x}_0}(\mathbf{x}-\mathbf{x}_0)+\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\left.\frac{\partial^2 f}{\partial x^2}\right|_{\mathbf{x}_0}(\mathbf{x}-\mathbf{x}_0)+\cdots \tag{9}$$

若$Det\left|\left.\frac{\partial^2 f}{\partial x^2}\right|_{\mathbf{x}_0}\right|>0$，高维函数（泛函）在$\mathbf{x}_0$附近存在极值点。

这样的观点，可以解决一大类问题。Newton 迭代法之所以有效果的根本就奠基在高维函数（泛函）至少局部是凸的基础上。

6 总结与讨论

地震数据被认为是随机的、满足一定统计规律的，局部线性或非线性的同相轴飘在满足不同概率分布的随机噪声中被认为是地震数据的本质模型，地震数据分析也被完全建立在统计数据分析的基础上，这是地震数据分析继续深入的基础。

更进一步地，地震数据分析还需要引入很多的几何思维。

所谓高维数据体中蕴含的信息，只是高维数据中的相关结构。无论是空间域（像域）或是时空域（数据域）高维数据体中的相关结构，都是地震数据分析、地震波成像中最基本的信息。提取、表达、描述、分析这些几何结构都是重要的、基础性的工作。几何学知识的补充、完善和深化在当前的地震数据分析中

很有必要。

事实上，特别希望养成建立对所研究问题的几何思维的习惯。用物理+几何思维统领算法层次的思维。这是开展创新研究所必要的思维方式。

越来越不能忍受处处跟在别人的屁股后面！缺乏自己的批判性和创新性思维。今后，一个民族能否在世界上有尊严地生存，极有可能维系于其广大的人民是否有批判性和创新性思维习惯。

7 参考文献

[1]. 李文林等译，2020，数学手册（第十版），布龙施泰因等编，科学出版社。
[2]. 李文林等译，2012，数学指南-实用数学手册，蔡德勒等编，科学出版社。
[3]. Buhlmann,P. and Van De Geer, S.,2011,Statistics for High-Dimensional Data-Methods, Theory and Applications, Springer.
[4]. Liu G, Fomel S, Jin L, et al. Stacking seismic data using local correlation[J]. Geophysics, 2009, 74(3): V43-V48.
[5]. Chen Y, Fomel S. Random noise attenuation using local signal-and-noise orthogonalization[J]. Geophysics, 2015, 80(6): WD1-WD9.

约束非线性优化问题的提出与求解

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】地震波成像问题（归属于广义遥感的一类问题），也包括信号分析、图像分析、机器学习（ML）中很多问题，都是信息不足情形下的决策问题。这一大类问题的共同解决方案都涉及到：多种信息约束下的优化问题的提法；有约束非线性优化问题的解法。待反演参数的目标函数，结合来自各自的物理问题的约束条件，构成了参数估计问题提法的核心。无约束优化问题的求解基本上就是目标函数下降意义下的梯度导引算法，“梯度方向”包括：一般的梯度；共轭梯度；Hessian 逆预条件处理后的梯度。无约束优化问题解存在的必要条件是：目标泛函的一阶导数等于零；充分必要条件是：目标泛函的一阶导数等于零以及二阶导数大于零。同时，局部解附近，目标函数是凸的、连续的、二阶导存在的、紧支撑的也是对目标泛函的基本要求。FWI 和 LS_RTM 基本上就是在求解一个无约束优化问题。但是，FWI 和 LS_RTM 并没有得到广泛地实用。若要继续提升地震波成像算法的精度和数值稳定性，必须引入更多的先验信息，把地震参数估计问题提为多种信息约束下的非线性反演问题。多信息融合、定量的、考虑多次波的地震波成像是地震波成像深化研究的必然方向。但是，勘探地震中的约束非线性优化问题的提法和解法依然存在很多问题。本文试图从波场与介质扰动的非线性关系表达、广义线性化的地震波成像算法、约束非线性反问题的提法及解法、非线性地震波成像的逻辑路线和多次波成像的解决方案的角度展开讨论，促进当前地震波成像问题研究的深化。

【关键词】参数估计问题，反问题提法，凸问题，反问题解法，无约束优化问题，约束非线性优化问题

1 引言

勘探地震理论和实践表明：地震波成像（弹性参数估计）总是一个信息不足情形下的（强）非线性反问题。到目前为止，并不存在一个普适性的求解（强）非线性反问题的数学工具。

勘探地震中的地震波非线性反演成像绝不是一个满足一定理论假设的纯数学问题。勘探地震学近百年来的发展历史，以及最近十几年来基于 Bayes 估计理论的 FWI 和 LS_RTM 在实际应用中的举步维艰，即便发展了各种各样的变种方法依然如此，已经证明了必须要基于勘探地震的介质分布特征、波场特征和数据观测的特点，才能发展出真正能解决勘探地震中地震波非线性反演成像问题的理论与方法。

一定假设下的地震波正演模拟方法可以预测实测地震波场是地震波反演最基本的要求，既然无论如何选择地震波正演模拟方法也无法完全预测实测数据中的波场，转而追求尽可能好地模拟其中的部分特征波场，进行基于特征波场的反演成像是合理的逻辑。通过特征表达参数模型和特征表达数据体得到二者之间存在更密切联系的或近似线性的关系可以把强非线性的参数估计反问题转化为一组较凸的反问题进行求解。这是求解（强）非线性反演问题的正确技术路线。

弹性参数估计问题的（强）非线性根本上是由于参数扰动和波场扰动之间存在（强）非线性所导致的。勘探实践中，地震波场出现多次波是参数扰动与波场扰动之间存在（强）非线性关系的物理实证。Lippmann-Schwinger 方程所描述参数扰动与波场扰动之间的非线性关系与实际波场中多次波存在所展现的非线性之间是存在差异的！散射级数表述的正问题不适合用于参数估计，逆散射级数法进行地震波反演成像至今没有发展出有效的技术。勘探地震中波场扰动与参数扰动之间的非线性关系需要从物理直观和数学表达的有机结合发展出来，不宜把 Lippmann-Schwinger 方程照搬到地震波反演成像中。

正问题具有局部性和因果性，反问题没有局部性和因果性。非同因之果的差异度量是参数估计反问题存在强非线性性的另一个重要原因，FWI 的周期跳现象本质上就是没有满足同因之果的差异度量要求。误差泛函的定义既要考虑有一定的物理意义（譬如能量最低，熵最小），还要考虑同因之果的差异度量。用非同因之果间的差异度量构建反问题是没有意义的。目前，即便是基于最优输运方法定义的误差泛函也没有很好地解决同因之果的差异度量问题。

地震波成像（即弹性参数估计）问题总是一个信息不足情形下的（强）非线性反问题。利用机器学习算法进行参数扰动场的几何结构的提取，进行多信息源中与待估计参数有关的同性信息提取，并把这些先验信息合理地用于从参数估计结果可行解集中决定最终的反演解，可以显著提升非线性反问题解的精度。

复杂油气藏精确描述的高精度地震波反演成像应该推进到宽带波阻抗成像阶段。宽带波阻抗成像对于高精度油藏描述，对于提高油气勘探的效益有重要价值。这是数十年来油气工业界没有得到有效解决的重大问题。众所周知，宽带波阻抗成像比反射系数成像的非线性更强。因此，我们提出特征波反演成像（CWI = Characteristic Wave Inversion Imaging）+宽带波阻抗建模（WBIM = Wide Band Impedance Modeling）

的技术路线，把宽带波阻抗建模提为特征波反演成像基础上的信息融合问题，有效地避开了直接求解宽带波阻抗这个（强）非线性反问题的困难。在“两宽一高”地震数据、高性能计算、机器学习算法的加持下，把地震波反演成像推进到宽带波阻抗成像阶段有了切实的可能。

上述观点体现了我们对地震波成像（参数估计）问题的基本认识。

总体而言，多信息融合、定量的、考虑多次波的地震波成像是地震波成像深化研究的必然方向。约束非线性优化问题的提法和解法会成为地震波成像的核心问题。但是，勘探地震中的约束非线性优化问题的提法和解法依然存在很多问题。本文试图从波场与介质扰动的非线性关系表达、广义线性化的地震波成像算法、约束非线性反问题的提法及解法、非线性地震波成像的逻辑路线和多次波成像的解决方案的角度展开讨论，期望能促进当前地震波成像问题研究的持续深化。

2 波场与介质参数扰动量之间非线性关系的分析

地震波反演（也是所有反问题）的根本逻辑：正问题能（部分地）描述实测的地震数据（地震波场）。这是反演解存在的必要条件。

勘探地震中，物理上地下介质弹性参数的真实分布当然是存在的。但这并不能说明反问题的解一定存在。不能混淆这两个概念。

反问题是奠基在人为构建的正问题基础上，通过类似 Bayes 参数估计准则建立的一套试图估计真实解的算法。至于这套算法能否得到真实解的近似估计结果，能在多大精度上逼近真实解正是参数估计反问题要回答的问题。无论如何，这中间描述待估计参数与（部分）实测数据关系的正问题都起着最核心的作用。

正问题与实测数据的关系决定了反演解的可行解集。反演解对应一个可能非常庞大的解集（解的方差很大！）。引入关于解的先验信息从可行解集中选择一个合理的解。引入各种正则化思想与方法，目的是降低最终反演解的方差或不确定性，提高反演解的精度。

但是，依然存在一个根本的理论问题目前并没有得到很好的说明：波场扰动和介质扰动之间的非线性关系式是否在实际物理和数学表达之间达到一致？我认为：关于波场扰动和介质扰动之间的非线性关系式（即 Lipmann-Schwinger 方程），数学和物理是不一致的，存在两张皮的现象。譬如如何定量地刻画 Born 线性近似的成立条件？如何验证数学上给出的成立条件与物理是否一致？不同尺度介质和不同频率下的波地震波传播的物理机制是非常不同的。当前，只有地震波走时与背景速度的关系是确切的。反射波振幅变化的物理机制是说不清楚的。如果连正问题都存在不确定性，反问题的精确根本无从谈起。但是，任何实际领域中，反问题都不可回避。尽管正确的确切物理讲不清楚，但反问题还是要做，只是我们对反问题解的精度要做出尽可能清楚地评价，做决策时充分意识到解的不精确性。

正问题的本质是：描述待估计参数与（部分）实测波场之间的关系式。

若正问题是线性的，即待估计参数 $\mathbf{x}$ 与（部分）实测波场 $\mathbf{b}^{obs}=\mathbf{b}+\boldsymbol{\eta}$ 之间存在近似的线性关系 $A\mathbf{x}\approx\mathbf{b}^{obs}$。此时在 Gauss 白噪声的假设下、线性系统下反问题的求解理论已经非常成熟了。系数矩阵 A 的性质决定线性系统的解。信号和图像分析领域存在一大类这样的问题。LS_RTM 和 Tomography 也是这样的问题。

但依然存在的一个核心问题：$A\mathbf{x}\approx\mathbf{b}^{obs}$ 包含的物理实质是什么？看起来波动理论及各种变种长期以来为我们所用，也解决了很多问题。但是，深究这个波动理论及各种变种成立的条件，还是有很多令人质疑的地方。如果连正问题都不能保证百分百成立，连正问题的物理含义都不能严格地界定，反问题令人质疑的地方就更多了。

若正问题是非线性的，即待估计参数 $\mathbf{x}$ 与（部分）实测波场 $\mathbf{b}^{obs}=\mathbf{b}+\boldsymbol{\eta}$ 之间存在近似的非线性关系：$\mathbf{b}^{obs}\approx f(\mathbf{x})$ 或写为 $\mathbf{b}^{obs}\approx A(\mathbf{x})$。勘探地震中，数据中存在明显的多次波是地下介质存在强扰动的显式证据。此时待估计参数与（部分）实测波场之间的关系式应该是非线性的。

类似地，若信号与图像领域，数据中蕴含的几何结构的相位和振幅都出现剧烈变化，此时因变量与数据之间存在预测关系式必然也是非线性的。线性结构飘在 Gauss 白噪声中，才能用线性预测关系式进行信号/图像的预测，解决相关的去噪、插值问题。在 AI/ML 领域，输入与输出之间的关系更是极有可能是非线

性的，非线性性越强，训练学习机器的成功率越低，泛化应用效果越差。很多时候，我们不能迷信 AI/ML 能给出更好的处理效果。实际应用表明：输入与输出之间的关系非线性性很强的情形下，AI/ML 算法的应用效果也很差。

线性系统+Gauss 白噪声情形下的（广义）线性问题，理论上已经不存在研究的问题。即便引入 L_0、L_1、$L_p\left(1\le p\le 2\right)$、$L_2$ 以及 TV 正则化约束的反问题，也已经被研究得很深入了。也没有必要在数据偏差项中引入 L_∞ 和 $L_p\left(p\ne 2\right)$ 之类的范数,宁可通过去噪、引入合适的权函数、选择特征数据（特征波场）及合适的正问题，使得数据残差近似满足 Gauss 分布、误差泛函成为一个凸的函数。Tarantola（2005，第四章）提出的在数据偏差项中引入 L_∞ 和 L_1 范数的观点至今没有在地震波反演成像中得到响应。

从学术研究和技术发展方向的角度，今后创新性的研究一定要针对：地震波成像（弹性参数估计）总是一个信息不足情形下的（强）非线性反问题。首先遇到的问题必然是：①描述待估计参数与（部分）实测波场之间的关系式到底该如何表达？②这样的表达方式与实际物理一定要是尽可能相符的，但如何检验它与实际物理是相符的？③这样的表达方式一定要是有利于后续的反演成像的。总之，创新的根本在于提出参数扰动与波场扰动之间关系的新的数学模型。

建立参数扰动与波场扰动之间的数学物理关系，存在如下几种代表性的方法：

① 波动方程。但是波动方程显然没有专注于表达参数扰动与波场扰动之间的关系，是一个笼统的关系式。

② 波场的 Taylor 展开与线性化。

观测波场的 Taylor 展开：

$$\mathbf{u}^{cal}\left(\mathbf{m}\right)=\mathbf{u}\left(\mathbf{m}_0\right)+\left.\frac{\partial\mathbf{u}}{\partial\mathbf{m}}\right|_{\mathbf{m}_0}\Delta\mathbf{m}+\frac{1}{2!}\left.\frac{\partial^2\mathbf{u}}{\partial\mathbf{m}^2}\right|_{\mathbf{m}_0}\Delta\mathbf{m}^2+\frac{1}{3!}\left.\frac{\partial^3\mathbf{u}}{\partial\mathbf{m}^3}\right|_{\mathbf{m}_0}\Delta\mathbf{m}^3+\cdots \tag{1}$$

对上式进行线性化处理，得到：

$$\delta\mathbf{u}\triangleq\mathbf{u}^{cal}\left(\mathbf{m}\right)-\mathbf{u}\left(\mathbf{m}_0\right)\approx\left.\frac{\partial\mathbf{u}}{\partial\mathbf{m}}\right|_{\mathbf{m}_0}\Delta\mathbf{m}\quad\Leftrightarrow\delta\mathbf{u}=A\delta\mathbf{m} \tag{2}$$

但是，这样的做法缺乏明确的物理含义。

③ Lippmann-Schwinger 方程。

来自量子力学的 Lippmann-Schwinger 方程表达地表观测波场，

$$\mathbf{u}^{cal}\left(\mathbf{m}\right)=\mathbf{u}\left(\mathbf{m}_0\right)+G\left(\mathbf{m}_0\right)\Delta\mathbf{m}G\left(\mathbf{m}_0\right)+G\left(\mathbf{m}_0\right)\Delta\mathbf{m}G\left(\mathbf{m}_0\right)\Delta\mathbf{m}G\left(\mathbf{m}_0\right)+\cdots \tag{3}$$

很显然，认为地表记录波场是地下所有散射点各阶波场扰动的叠加。(3)式的物理意义是很明确的。线性化处理，得到：

$$\delta\mathbf{u}\triangleq\mathbf{u}^{cal}\left(\mathbf{m}\right)-\mathbf{u}\left(\mathbf{m}_0\right)\approx G\left(\mathbf{m}_0\right)\Delta\mathbf{m}G\left(\mathbf{m}_0\right)\quad\Leftrightarrow\delta\mathbf{u}=A\delta\mathbf{m} \tag{4}$$

(4)式即波场扰动的一阶 Born 近似表达式。LS_RTM 就是基于上述线性化正问题的。背景速度的层析成像问题也可以用 Lippmann-Schwinger 方程来解释。

④ WRW 模式。该模式的物理含义由图 1 所示，认为地表记录波场是地下所有反射层各阶多次反射波场的叠加。蓝框内 $\mathcal{W}_p^1$ 和 $\mathcal{W}_p^2$ 分别表示一阶和二阶层间多次波，W_0^- 和 W_0^+ 分别代表上下行波 Green 函数。因此，地表记录的全体反射波场自然由下式描述：

$$\mathbf{u}^{cal}=W_0^+RW_0^-+W_0^+RW_p^1RW_0^-+W_0^+RW_p^2RW_0^-+W_0^+RW_p^3RW_0^-+\cdots \tag{5}$$

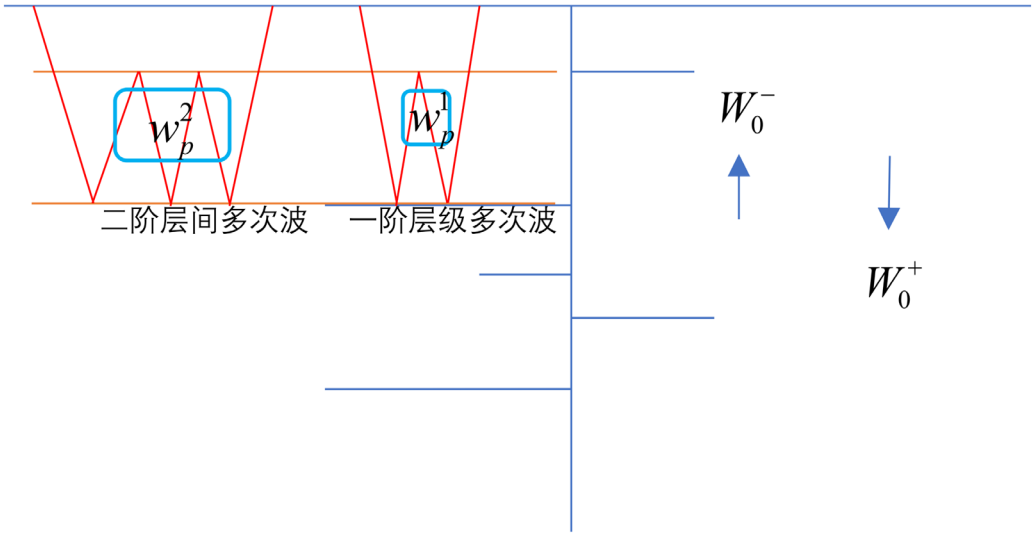


图 1 地表记录波场是地下所有反射层各阶多次反射波场的叠加

很显然，这可以认为是 Lippmann-Schwinger 方程在地下介质为层状情形下的重新表达，也可以认为是 Lippmann-Schwinger 方程的特例或其特征形式。这是 Berkhout 教授带领的研究团队一直坚持使用的描述勘探地震地下介质波传播的数学模型。这个模型的物理意义也是十分清楚的。

(5)式的线性化，是在假设地下介质都是弱反射，各种多次反射都可以被忽略的情况下给出的。得到下式：

$$u(t)=R(t)e^{ik_z 2\Delta z}W(\tau)=R(t)e^{i\omega t}W(\tau)=R(t)*W(t) \tag{6}$$

(6)式中，$W(t)$ 是地震子波，$R(t)$ 是满足线性化要求的反射系数序列。

⑤ 积分表示定理（Aki，P29）。

$$\begin{aligned} u_n(\mathbf{x},t)=&\int_{-\infty}^{\infty}dt\iiint_V f_i(\xi,\tau)G_{ni}^{rigid}(\mathbf{x},t-\tau;\xi,0)dV \\ &-\int_{-\infty}^{\infty}dt\iint_S u_i(\xi,\tau)c_{ijkl}n_j\frac{\partial}{\partial\xi_l}G_{nk}^{rigid}(\mathbf{x},t-\tau;\xi,0)dS \end{aligned} \tag{7}$$

地表观测的波场由封闭面内体积分和封闭面上面积分表达自然是十分合理的。但是，与波动方程表达一样，没有显式地表达参数扰动与波场扰动之间的关系，是一个笼统的关系式。

根据上述分析，一般意义下，勘探地震参数估计中非线性性的物理表现是清楚的，也存在几种常用的表达参数扰动与波场扰动之间关系的数学模型，但非线性问题的求解还是缺乏明确的路线和方法。

由前述分析还可以看出：①背景参数 $\mathbf{m}_0$ 与真解的逼近程度是反问题非线性性强弱的根源；②特征多次波的存在是反问题非线性性强弱的重要影响因素；③反演过程中，确认参数扰动与波场扰动是一对一的、同因之果的，还是十分困难的。正是上述问题，引起了反问题求解极为困难的本质原因。

3 广义线性化反问题的提法与解法

尽管广义线性反演的理论已经非常完善和成熟，但这是目前最常用的，也是行之有效的。它一定是研究非线性成像的基础。当前，相当大部分的研究工作还是聚焦于如何把广义线性问题实现好，解决好技术层面的问题。

线性系统定义：$A\mathbf{x}=\mathbf{b}$ 。非常多的问题最终都归结于一个线性系统，即一个线性方程组。实际问题归纳出来的线性问题往往是超定的或欠定的。超定问题理论上是没有解的，欠定问题理论上是有无穷多解的。实际上，线性问题的求解往往转化为变分问题。当然，也可以在 Bayes 估计理论下理解线性系统下的

参数估计。

估计线性系统的参数 $\mathbf{x}$ 的基本方法是定义如下的变分问题：

$$\mathcal{J}(\mathbf{x})=(A\mathbf{x}-\mathbf{b})^T(A\mathbf{x}-\mathbf{b}) \tag{8}$$

假设系数矩阵 A 是非奇异的，则

$$\hat{\mathbf{x}}=min\mathcal{J}(\mathbf{x})=min\left\{(A\mathbf{x}-\mathbf{b})^T(A\mathbf{x}-\mathbf{b})\right\} \tag{9}$$

就被认为是线性系统的解。其中，解的性质与代价泛函 $\mathcal{J}(\mathbf{x})$ 的几何形态有密切的关系。

下面推导一下代价泛函 $\mathcal{J}(\mathbf{x})$ 的另一种表达形式：

$$\begin{aligned}
&\mathcal{J}(\mathbf{x})=(A\mathbf{x}-\mathbf{b})^T(A\mathbf{x}-\mathbf{b})\\
&=\mathbf{b}^T\mathbf{b}-(A\mathbf{x})^T\mathbf{b}-\mathbf{b}^T(A\mathbf{x})+(A\mathbf{x})^T(A\mathbf{x})\\
&=\mathbf{b}^T\mathbf{b}-\mathbf{x}^TA^T\mathbf{b}-(A^T\mathbf{b})^T\mathbf{x}+\mathbf{x}^TA^TA\mathbf{x}\\
&=\mathbf{b}^T\mathbf{b}-(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})+(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})-\mathbf{x}^TA^T\mathbf{b}-(A^T\mathbf{b})^T\mathbf{x}+\mathbf{x}^TA^TA\mathbf{x}\\
&=\mathbf{b}^T\mathbf{b}-(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})+\left[(A^T\mathbf{b})^T(A^TA)^{-1}-\mathbf{x}^T\right](A^T\mathbf{b})-\left[(A^T\mathbf{b})^T-\mathbf{x}^TA^TA\right]\mathbf{x}\\
&=\mathbf{b}^T\mathbf{b}-(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})+\left[(A^T\mathbf{b})^T(A^TA)^{-1}-\mathbf{x}^T\right](A^T\mathbf{b})-\left[(A^T\mathbf{b})^T(A^TA)^{-1}-\mathbf{x}^T\right](A^TA)\mathbf{x}\\
&=\mathbf{b}^T\mathbf{b}-(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})+\left[(A^T\mathbf{b})^T(A^TA)^{-1}-\mathbf{x}^T\right]\left[(A^T\mathbf{b})-(A^TA)\mathbf{x}\right]\\
&=\mathbf{b}^T\mathbf{b}-(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})+\left[(A^T\mathbf{b})^T(A^TA)^{-1}-\mathbf{x}^T\right]A^TA\left[(A^TA)^{-1}(A^T\mathbf{b})-\mathbf{x}\right]\\
&=\mathbf{b}^T\mathbf{b}-(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})+\left[(A^TA)^{-1}(A^T\mathbf{b})-\mathbf{x}\right]^TA^TA\left[(A^TA)^{-1}(A^T\mathbf{b})-\mathbf{x}\right]\\
&=\mathbf{b}^T\mathbf{b}-(A^T\mathbf{b})^T(A^TA)^{-1}(A^T\mathbf{b})+(\mathbf{x}_o-\mathbf{x})^TA^TA(\mathbf{x}_o-\mathbf{x})\qquad \mathbf{x}_o=(A^TA)^{-1}(A^T\mathbf{b})\\
&=\mathbf{b}^T\mathbf{b}-\hat{\mathbf{b}}^T\hat{\mathbf{b}}+(\mathbf{x}_o-\mathbf{x})^TA^TA(\mathbf{x}_o-\mathbf{x})\\
&=\mathcal{J}_{min}(\mathbf{x}_o)+(\mathbf{x}_o-\mathbf{x})^TA^TA(\mathbf{x}_o-\mathbf{x})
\end{aligned} \tag{10}$$

其中，$\mathbf{x}_o$ 代表 $A\mathbf{x}=\mathbf{b}$ 的解。

对矩阵 A^TA 进行正交分解，可以写出代价泛函 $\mathcal{J}(\mathbf{x})$ 的椭球体解析表达形式：

$$
\begin{aligned}
&\mathcal{J}\left(\mathbf{x}\right)=\mathcal{J}_{min}\left(\mathbf{x}_o\right)+\left(\mathbf{x}_o-\mathbf{x}\right)^T A^T A\left(\mathbf{x}_o-\mathbf{x}\right)\\
&=\mathcal{J}_{min}\left(\mathbf{x}_o\right)+\left(\mathbf{x}_o-\mathbf{x}\right)^T Q^T \Sigma Q\left(\mathbf{x}_o-\mathbf{x}\right)\\
&=\mathcal{J}_{min}\left(\mathbf{x}_o\right)+\left(Q\left(\mathbf{x}_o-\mathbf{x}\right)\right)^{\mathbf{T}} \Sigma Q\left(\mathbf{x}_o-\mathbf{x}\right)\\
&=\mathcal{J}_{min}\left(\mathbf{x}_o\right)+\mathbf{q}^{\mathbf{T}}\Gamma\mathbf{q} \qquad \mathbf{q}=Q\left(\mathbf{x}_o-\mathbf{x}\right)\\
&=\mathcal{J}_{min}\left(\mathbf{x}_o\right)+\sum_{k=1}^{M}\lambda_k\left|q_k\right|^2\\
&=\mathcal{J}_{min}\left(\mathbf{x}_o\right)+\sum_{k=1}^{M}\frac{\left|q_k\right|^2}{1/\lambda_k}
\end{aligned}
\tag{11}
$$

其中，$1/\lambda_k, k=1,2,\cdots,M$ 代表椭球体的半轴长度。图 2 展示了代价泛函 $\mathcal{J}\left(\mathbf{x}\right)$ 的椭球体形状。

可以说(11)式把线性系统求解的本质问题都展示清楚了。代价泛函 $\mathcal{J}\left(\mathbf{x}\right)$ 是凸的，仅存在一个能量最低的点 $\mathcal{J}_{min}\left(\mathbf{x}_o\right)$，一般的反演解就取该点对应的参数估计结果。系数矩阵 A 零空间存在时，反演解是不唯一的。因此需要引入正则化思想和做法。系数矩阵满秩时解是唯一的。系数矩阵条件数小时，收敛是很快的，否则收敛很慢。

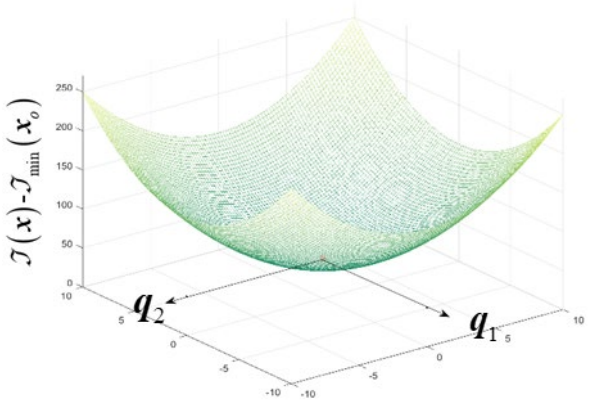


图 2 代价泛函 $\mathcal{J}\left(\mathbf{x}\right)$ 的椭球体形状

广义线性系统反问题的求解，基本上存在如下比较成熟的解法。对于无约束优化问题，由于代价泛函 $\mathcal{J}\left(\mathbf{x}\right)$ 是凸的、二阶导数存在的，因此，梯度导引类算法是主要的，包括最速下降方法；Newton 方法。中间还有一个共轭梯度算法。当然，Gauss-Newton 算法和拟 Newton 算法都是应该了解的。

实际应用中，线性系统可能是无穷多解的，系数矩阵性状不好的，需要引入解的约束。譬如引入：L_0、L_1 以及 TV 正则化约束。此时，整体代价函数不再可以求导。不同情形下，有对应的典型算法。LASSO 问题情形：次梯度导引算法（软门域算法）；TV 正则化情形：主-对偶算法；L_0 范数正则化情形：贪婪算法。

为完整和深刻理解，继续讨论线性系统求解的其他认识。

Beckus-Gilbert 的线性反演观点。

线性系统：$A\mathbf{x}=\mathbf{b}$，可以被重新表达为：

$$A_{M\times N}\mathbf{x}_{N\times 1}=\mathbf{b}_{M\times 1}=\left[\mathbf{a}_1,\mathbf{a}_1,\cdots,\mathbf{a}_N\right]\begin{bmatrix}x_1\\x_2\\\vdots\\x_N\end{bmatrix} \tag{12}$$

反演解被定义为：

$$\hat{\mathbf{x}}_{N\times 1}=C_{N\times M}\mathbf{b}_{M\times 1}=\left[\mathbf{c}_1,\mathbf{c}_2,\cdots,\mathbf{c}_M\right]\begin{bmatrix}b_1\\b_2\\\vdots\\b_M\end{bmatrix} \tag{13}$$

其中，矩阵 $C_{N\times M}$ 被定义为 $A^T_{N\times M}$，也可被定义为 $\left(A^TA\right)^{-1}A^T$。

反演分辨率定义为：

$$\hat{\mathbf{x}}_{N\times 1}=C_{N\times M}\mathbf{b}_{M\times 1}=C_{N\times M}A_{M\times \mathrm{N}}\mathbf{x}_{N\times 1}=R_{N\times \mathrm{N}}\mathbf{x}_{N\times 1} \tag{14}$$

很显然，$R=A^TA$ 或 $R=\left(A^TA\right)^{-1}A^TA$。

SVD 分解的线性反演观点：

线性系统：$A\mathbf{x}=\mathbf{b}$，基于系数矩阵 SVD 分解的线性系统表达：

$$\begin{aligned}&A_{M\times N}\mathbf{x}_{N\times 1}=\mathbf{b}_{M\times 1}\\&\Sigma_{M\times N}V^T_{N\times N}\mathbf{x}_{N\times 1}=\mathbf{b}_{M\times 1}\\&V^T_{N\times N}\mathbf{x}_{N\times 1}=\Sigma^{-1}_{M\times N}U^T_{M\times M}\mathbf{b}_{M\times 1}\\&\Sigma_{M\times N}\mathbf{x}'_{N\times 1}=\mathbf{b}'_{M\times 1}\end{aligned} \tag{15}$$

其中，$\mathbf{x}'_{N\times 1}$ 和 $\mathbf{b}'_{M\times 1}$ 可以认为是原参数和数据向量的特征表达。显然，二者之间的关系是最简单的。线性反演解的性质一目了然。

线性反问题解梯度的物理意义。梯度导引下的线性系统求解算法写为：

$$\mathbf{x}^{k+1}=\mathbf{x}^k+\alpha A^T\left[A\mathbf{x}-\mathbf{d}^{obs}\right] \tag{16}$$

可见，(16)式中梯度项的物理含义是数据残差的直接反投影。梯度的另一个物理含义自然是当前迭代步的参数修正量。已知迭代反演算法每一迭代步的物理含义，正则化处理就可以在物理意义的指导下进行，实现反演结果与实际物理保持一致，同时又提高收敛效率的目的。

4 约束非线性反问题的提法及解法

勘探地震中，信号处理领域、图像处理领域和参数估计领域中的很多具体的反问题都是基于正问题（$A\mathbf{x}\approx\mathbf{b}^{obs}$ 或 $A\left(\mathbf{x}\right)\approx\mathbf{b}^{obs}$）对实测数据的预测误差来定义的。数学本质上，预测误差达到最小并不能保证反演解一定是好的。根本上，这还是依赖于 A 的性态。本质上，首先要求误差泛函是凸的，最好是二次型的。

到目前为止，我们并没有求解非凸反问题的有效解法，更不用讲存在解强非线性反问题的有效解法。

复杂情形下的反问题都可以被归结为：信息不足情形下的决策问题。

求解反问题的重点方向正在逐渐转化为如何尽可能利用先验信息，提高反问题的凸性、反问题解的稳

定性和精度。可以归结为约束非线性反问题的提法和解法是今后反问题研究关注的重点，尤其是针对各种各样的实际问题，如何提一个合适的约束非线性反问题，它的稳定且高效的解法是什么。

4.1 有约束非线性优化问题的提出

下面罗列几种由简单到复杂的反问题提法。

① 无约束优化问题提法：

$$\mathcal{J}(\mathbf{m})=\frac{1}{2}\mathbf{F}^T(\mathbf{m})\mathbf{F}(\mathbf{m})=\frac{1}{2}\left(G(\mathbf{m})-\mathbf{d}^{obs}\right)^T\left(G(\mathbf{m})-\mathbf{d}^{obs}\right) \tag{17}$$

② Tikhonov 正则化问题提法：

$$\mathcal{J}(\mathbf{m})=\frac{1}{2}\mathbf{F}^T(\mathbf{m})\mathbf{F}(\mathbf{m})=\frac{1}{2}\left(G(\mathbf{m})-\mathbf{d}^{obs}\right)^T\left(G(\mathbf{m})-\mathbf{d}^{obs}\right)+\lambda\Omega(\mathbf{m}) \tag{18}$$

③ LASSO（Least absolute shrinkage and selection operator）问题的提法：

$$\hat{\mathbf{m}}=\underset{\mathbf{m}\in\mathcal{M}}{Argmin}\|\mathbf{m}\|_1 \quad s.t.\left(G(\mathbf{m})-\mathbf{d}^{obs}\right)^2\le\delta \tag{19}$$

④ 非常一般的约束非线性问题的提法：

$$\hat{\mathbf{m}}=\underset{\mathbf{m}\in\mathcal{M}}{Argmin}\,F(\mathbf{m}) \quad s.t.\mathbf{m}\in\mathcal{Q} \tag{20}$$

其中，$\mathcal{Q}$ 代表各种先验约束规定的可行解集。一般地，先验约束信息表达为等式约束和不等式约束，上述约束非线性问题重写为：

$$\begin{aligned}&\hat{\mathbf{m}}=\underset{\mathbf{m}\in\mathcal{M}}{Argmin}\,F(\mathbf{m})\\ &s.t.\begin{cases}h_i(\mathbf{m})=0, & i=1,2,\cdots,n\\ g_j(\mathbf{m})\ge 0, & j=1,2,\cdots,p\end{cases}\end{aligned} \tag{21}$$

上述表达方法非常一般化，基本上任何领域的非线性约束问题都可以抽象成上述形式。可以看出，这种表达形式与凸优化问题是非常一致的。期望目标函数和约束函数都是凸的，至少在局部点附近二者都是凸的。

实际具体问题中，合理的约束条件的提出与相应专业领域知识的关系非常密切。只有对本专业领域的广泛知识都有透彻的理解，才能提出合理的约束非线性优化问题。只有提出的反问题本身是合理的，花功夫解决它才是值得的。

4.2 有约束非线性优化问题的求解

关于有约束非线性优化问题的求解，与无约束优化问题求解是类似的，首先还是要理清（局部）解存在的必要和充分必要条件。我认为 Lagrange 条件和 KTT 条件较好地回答了这个问题。几何意义也很明确，基本上是无约束优化问题（局部）解存在的必要和充分必要条件的推广，尽管看起来稍微复杂一些。Lagrange 乘子法基本上就是利用目标泛函的一阶导数等于零的条件导出的。

与无约束优化问题求解不同，有约束非线性优化问题的解法太多了。但根本问题还是要体会千变万化的有约束非线性优化问题解法背后的本质，不能被数学家弄的复杂的表述搞糊涂了。我认为：无论是有约束还是无约束优化，目标函数的下降是本质的要求！目标函数和约束项都局部凸是本质的要求！如果是这样，有约束优化的任何算法设计，都脱不开总体目标函数导引下的梯度，如果总体目标函数不可导，就利用次梯度的概念。另外，投影梯度法是构造算法的基本思想。只要梯度出现方向异常，就用投影算子进行校正。Split-Bregman 算法、LASSO 算法都体系了这样的思想。

梯度的物理含义本身并没有变化，还是当前迭代步的参数扰动量。在此意义下，对梯度进行预处理，

加快收敛速度，逼近有物理意义的解就有了基本的物理依据。忽视几何与物理意义，纯粹通过调参加速有约束非线性优化问题的收敛性，显然不是好的方式。因此，有约束非线性优化问题的求解绝对不是个纯数学问题，必须要与专业领域内的物理/几何理解相结合，按照梯度的物理意义，把约束合理地施加在对梯度的预处理上，从而加速收敛。

5 非线性地震波成像的逻辑路线

由于数学上并不存在一个求解（强）非线性反问题的有效方法，只能有效地求解凸问题的解，因此，把强非线性的参数估计反问题转化为一组较凸的反问题进行求解就是合理的逻辑路线。而把强非线性的参数估计反问题转化为一组较凸的反问题不是纯数学问题，而是依据专业领域知识的实践问题。

勘探地震中，把强非线性的参数估计反问题转化为一组较凸的反问题进行求解的物理依据是：勘探地震的介质特征、波场特征、数据特征。本质上就是用不同的波现象构建不同的反演方法：①透射波（直达、折射、Diving Wave）；②与特征反射层对应的特征反射波；③一次反射和绕射波；④与特征反射层对应的多次反射波。地震波成像的目标：宽波数带的弹性参数场（譬如速度、密度或波阻抗）。可以预计广义遥感的不同专业领域，把强非线性的参数估计反问题转化为一组较凸的反问题的方式是不会相同的，根本上在于：介质特征、波场特征、数据特征在不同的专业领域表现是不同的。

特征波成像（CWI）+多信息融合的宽带波阻抗建模（WBIM）就是 WPI 提出的解决勘探地震中非线性成像问题的具体方案。其中的具体技术环节当然还要不断地深化研究，才能让我们提出的 CWI+多信息融合的宽带波阻抗建模方法技术系列逐渐完善起来，在石油工业界起到大的作用。

5.1 多次波去除与成像问题的解决方案

多次波的出现预示着地下介质参数存在（小尺度的）强扰动，它们与波场之间的关系是（强）非线性的。

海上地震数据成像处理的特殊性就表现在多次波比较显著，而且主要是自由地表多次波比较显著。OBN 数据观测的波场本身就至少包含与自由地表相关的下行多次波场。层间多次波都有可能出现在海上地震勘探和陆上地震勘探的实测波场中。实际地震数据处理中，我们仅仅关注那些特征反射层引起的多次波。

什么是特征反射层？自由地表面（海水面）；（强反射的）海底面；强反射的地下界面。提出特征反射层引起的多次波压制和成像的目的：引入特征反射层可以降低多次波成像的非线性性。单独处理自由表面多次波就是引入特征反射层的实例。鬼波压制、水体相关多次波压制、一般的自由地表多次波压制（SRME 方法）就构成了几十年来海上油气地震勘探的一个特殊的核心问题。这种仅仅关注自由地表相关多次波的思想与方法是非常值得推崇的。这就是用“分而治之”的思想解决一个复杂问题。这中间，如何建立合理的特征反射层与特征多次波之间的数学关系是关键问题。

为什么一定要考虑多次波压制和成像？当前的地震波成像只能解决广义线性成像问题，即只能对一次波进行比较有效的成像。多次波的存在对广义线性成像会产生（严重）干扰。FWI 理论上可以解决非线性成像问题，但是逐步线性化的迭代方法，在实际复杂数据和复杂介质情况下，并不能保证 FWI 会收敛到有意义的反演解。初始迭代阶段，多次波的假像会（严重）干扰低波数参数估计的收敛性，导致迭代不收敛。这充分说明 FWI 方法不能处理包含多次波的非线性成像问题。这些说明根本不存在单一的解决非线性成像问题的数学方法。

去掉自由地表相关多次波（包括鬼波），余下的波场若不包含丰富的层间多次波，更符合广义线性成像的理论假设，一次波成像质量更高。若去掉自由地表相关多次波后，还有丰富的层间多次波，说明地下介质中存在强跃变。定量地估计这些强跃变的介质参数，线性化的方法是做不到的，只能得到近似的解。总之，小尺度、强跃变、或小尺度+强跃变介质分布都是导致参数估计问题非线性性程度提升的物理原因，大尺度（光滑）的、弱扰动的介质分布对应的参数估计问题是弱非线性的。

我重写的 WRW 模式：

$$
\begin{aligned}
&u(\mathbf{x},t)=DW^{-}(v_B)R(\mathbf{x},\theta)W^{+}(v_B)S(\text{所有的一次反射波场})\\
&+DW^{-}(v_B)RR_p^{1}(\mathbf{x},\theta)RW^{+}(v_B)S(\text{所有的一阶层间多次反射波场})\\
&+DW^{-}(v_B)RR_p^{2}(\mathbf{x},\theta)RW^{+}(v_B)S(\text{所有的二阶层间多次反射波场})\\
&+DW^{-}(v_B)RR_p^{3}(\mathbf{x},\theta)RW^{+}(v_B)S(\text{所有的三阶层间多次反射波场})\\
&+\cdots
\end{aligned}
\tag{22}
$$

当前 LS_RTM 对应的 WRW 模式：

$$
u(\mathbf{x},t)\approx DW^{-}(v_B)R(\mathbf{x},\theta)W^{+}(v_B)S(\text{所有的一次反射波场})
\tag{23}
$$

若想处理 Duplex 波的成像，对 $v_B(\mathbf{x})$ 是有特殊要求的，就是说在 $v_B(\mathbf{x})$ 中应包含产生 Duplex 波的正确的反射层位置信息,最好反射系数也是正确的。很显然，这样的要求与反演成像的逻辑是相悖的。但是，实际问题的解决就是这样纠结。实际情况下的反问题本身就不是一个纯数学问题。

(22)式表明，波场被简化为仅仅包含一次反射波，物理上对应假设地下反射系数 $R(\mathbf{x},\theta)$ 都是弱的，多次反射波非常弱，可以被忽略掉。定量上如何认定反射系数都是弱的？物理上看炮集中多次波是否强是一个符合逻辑的选择。

若能压制掉强的层间多次波，就可以用广义线性成像方法进行地震波成像，但是强反射系数的估计结果的误差是比较大的。但是，一般情况下这显然是比较可以接受的做法。油气地震勘探目前就是这样做的。但是，由上述 WRW 模式，可以清楚地看出，仅仅保留一阶反射波，理论上是不能估计地下真实的反射系数的。高精度的强反射系数估计需要进行多次波成像。事实上，利用层间多次波进行基于走时的层析成像，也可以有效地提升背景速度的精度。

由上述 WRW 模式，可以看出多次波层析成像会对中低波数参数估计有贡献。如何产生贡献的？众所周知，界面位置（反射结构形态）在 $v_B(\mathbf{x})$ 估计中的重要性没有得到充分的重视。FWI 迭代法做不到（或很难做到）对界面位置（反射结构形态）的重视。引入机器学习算法能更好地刻画反射结构，引入地质和测井信息能更好地确定反射结构，非常有助于提高 $v_B(\mathbf{x})$ 的估计精度，可以使得估计的 $v_B(\mathbf{x})$ 中包含更多的中波数成分。引入界面位置和它对应的反射系数后，可以模拟多次波，进行层析梯度项对背景速度的修改和高波数扰动梯度项对高波数成分的修改。这是多次波对高精度地震波成像的贡献。

关于所谓的全波场成像方法，基本思想是：把线性化的 LS_RTM 反演成像与反射波 FWI 成像串联，实现对全波场的模拟和反演成像。关键步骤包括：① Kirchhoff 近似下的 LS_RTM 生成保真的零角度反射系数；② 空间数据融合方法把背景速度场与反射系数有机结合形成包含速度跃变的速度场；③ 反射波 FWI 模拟出层间多次波（也可以包括自由表面多次波），进行包含层间多次波的高波数参数扰动的成像。

关键问题：背景速度存在误差问题；高波数参数与背景速度的融合问题；Kirchhoff 近似反射系数反演成像问题；FWI 反演成像的串联迭代问题；成像过程的收敛性问题等。

我认为：所谓的全波场成像方法最好也引入特征反射层的概念，在 $v_B(\mathbf{x})$ 的构建中，仅仅考虑特征反射层。特征反射层的选择依赖于反射系数反演成像结果。还是要引入“分而治之”的思想，对 $v_B(\mathbf{x})$ 的反演中引入特征反射层，引入更多的约束信息。不能指望靠 FWI 迭代算法自动的由低到高把所有波数成分的参数分量都迭代出来。

最后，针对海上数据地震波成像，自由地表是一个“镜”面，充分利用好这个特征面是必须要充分关注的。存在如下几种典型做法。

① 把海上地震数据是视为主震源（点震源）和自由表面二次源（面震源）共同激发的一次波场，进行线性化的 LS_RTM 成像。这虽然不能解决非线性成像问题，提高参数估计的精度，利用二次源照明改善成像质量还是可以的。核心是压制掉最小二乘迭代反演成像过程中的 Crosstalk 假象。

② 假设除自由地表外，其他反射层均为弱反射。自由表面观测的波场是一次波 Green 函数的多次反馈。

根据这样的认识，对于拖缆情形可以提出如下的褶积系统。首先定义$\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)=G(\mathbf{x}_r,\mathbf{x},\mathbf{x}_s;t)*R(\mathbf{x})$为地下介质和 Green 函数构成的一次上行波场。用r代表自由表面反射系数。自由表面和地下介质系统构成了一个反馈体系。如图 3 所示。在该反馈体系中传播的地震波可以用如下褶积系统来描述，

$$\begin{aligned}
&\mathbf{u}^{obs}(\mathbf{x}_r,\mathbf{x}_s;t)=\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)+\{\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\cdot r\}\cdot\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\\
&+\{[\{\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\cdot r\}\cdot\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)]\cdot r\}\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)+\cdots\\
&=\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\{1+\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\cdot r+[\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\cdot r]^2+[\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\cdot r]^2+\cdots\}\\
&=\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\frac{1}{1-\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)\cdot r}
\end{aligned}\tag{24}$$

很显然，(24)式定义了一个非线性的褶积系统。

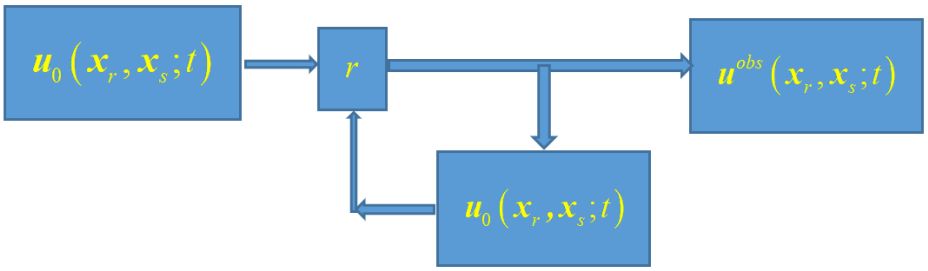


图 3 自由表面、炮检在海面情况项，多次波场生成的反馈机制

已知$\mathbf{u}^{obs}(\mathbf{x}_r,\mathbf{x}_s;t)$和自由界面反射系数，求解上述非线性褶积系统，可以估计出一次反射波场$\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)$。通过迭代法可以求解上述非线性问题，得到一次反射波场$\mathbf{u}_0(\mathbf{x}_r,\mathbf{x}_s;t)$。

③ 如果区分上下行波场，适应 OBS/OBN 观测数据，即利用上下行波场分别估计一次反射波，甚至估计一次波对应的 Green 函数场，同时剥离掉震源子波和自由表面多次波的影响，要引入图 4 表示的反馈褶积模型。

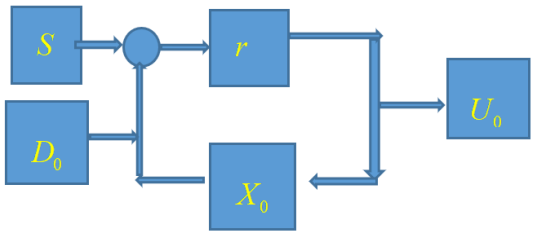


图 4 OBS/OBN 情形下，多次波生成的反馈机制

图 4 中，S代表震源函数；r依然是自由表面反射系数；X_0代表一次波对应的 Green 函数场；D_0是包含多次波的下行波场；U_0是包含多次波的上行波场。根据图 4 展示的反馈机制，可以写出如下的非线性褶积模型。

下行波场对应的非线性褶积模型为：

$$\begin{aligned}
&D_0=S+(X_0r)S+(X_0r)^2S+\cdots=S\left(1+X_0r+(X_0r)^2+\cdots\right)\\
&=S\left(\frac{1}{1-X_0r}\right)
\end{aligned}\tag{25}$$

上行波场对应的非线性褶积模型为：

$$
\begin{aligned}
U_0 &= X_0 S + X_0\left(X_0 r\right) S + X_0\left(X_0 r\right)^2 S + \cdots \\
&= S\left(X_0 + X_0\left(X_0 r\right) + X_0\left(X_0 r\right)^2 + \cdots\right) \\
&= S\left(\frac{X_0}{1 - X_0 r}\right)
\end{aligned} \tag{26}
$$

从(25)和(26)可以看出，

$$
U_0 = X_0\left(\frac{S}{1 - X_0 r}\right) = X_0 D_0 \tag{27}
$$

应该指出，上述(25)-(27)都是在把数据表达在频率空间域导出来的。另外，同样可以看出，由上行波场或下行波场估计一次波对应的 Green 函数场的非线性反褶积的求解，原则上都需要迭代方法。这一点与线性反演理论下的反褶积是显著不同的。

6 结论与讨论

地震波成像（广义遥感），也包括信号分析、图像分析、机器学习（ML）中很多问题都是信息不足情形下的决策问题。这一大类问题的共同解决方案都涉及到：多种信息约束下的优化问题的提法；有约束非线性优化问题的解法。

无约束优化问题基本上就是目标函数下降意义下的梯度导引算法，“梯度方向”包括：一般的梯度；共轭梯度；Hessian 逆预条件处理后的梯度。无约束优化问题解存在的必要条件是：目标泛函的一阶导数等于零；充分必要条件是：目标泛函的一阶导数等于零+二阶导数大于零。局部解附近，目标函数是凸的、连续的、二阶导存在的、紧支撑的。

有约束非线性优化问题的求解，Lagrange 条件和 KTT 条件较好地回答了解存在的充分和必要条件问题。

无论是有约束还是无约束优化，目标函数的下降是本质的要求！目标函数和约束项都是局部凸的是本质的要求！

地震波成像（弹性参数估计）总是一个信息不足情形下的（强）非线性反问题。到目前为止，并不存在一个普适性的求解（强）非线性反问题的数学工具。实际物理系统的参数估计问题不是一个纯数学问题，可以说是 Bayes 估计理论（理念）指导下的、获得有意义的反问题解的实践活动。所以 Hadamard 之前的数学家会认为解反问题是荒谬的！但是，解反问题却又绝对是必要的。对数学和物理问题的透彻理解是解决好反问题的唯一途径。

求解（强）非线性反问题的基本逻辑：把强非线性的参数估计反问题转化为一组较凸的反问题进行求解。

当前，地震波非线性成像的三个重点问题包括:① 基于同因之果原则的特征反射波波动理论层析成像。目的是通过特征反射波数据体的构建克服 Cycle-skipping 引起的非线性问题，核心问题是机器学习算法提取反射层、特殊地质体边界并更好地解决 RMO 拾取问题。② 特征反射层与特征多次波关系约束下的特征多次波压制与成像问题。基本策略是引入特征反射层降低多次波成像的非线性性。其中如何建立合理的特征反射层与特征多次波的数学关系是关键问题。③ 构建合理的途径，实现全波场成像问题。其中特征反射层的层析成像依然是全波场成像中的关键问题。引入反射结构的合理约束是全波场成像能否收敛的核心。

总之，约束非线性优化问题的合理提出与求解构成了地震波非线性反演成像的关键技术环节。

7 参考文献

[1]. Aki, K., and Richards, P., G., 2002, Quantitative seismology, Second edition: University Science Books.
[2]. Berkhout, A. J., 1985, Seismic migration: Imaging of acoustic energy by wave field extrapolation. A. Theoretical aspects: Elsevier Science Publ. Co., Inc.
[3]. Bertsekas, D. P., 1996, Constrained Optimization and Lagrange Multiplier methods, Athena Scientific.

[4]. Edwin K. P. Chong and Stanislaw, H. Zak, 2013, An Introduction to Optimization, Fourth Edition, John Wiley & Sons, Inc.
[5]. Tarantola, A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM.
[6]. Weglein A B. How can the inverse-scattering method really predict and subtract all multiples from a multidimensional earth with absolutely no subsurface information? [J]. The Leading Edge, 1999, 18(1): 132-136.
[7]. Weglein, A. B., Gasparotto, F. A., Carvalho, P. M., and Stolt, R. H., 1997, An inverse-scattering series method for attenuating multiples in seismic reflection data: Geophysics, 62, 1975–1989.

再议地震波成像道集物理实质

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】地震波成像的特殊性表现在：地震数据是通过地表的激发和接收而获得，地下介质是以层状为主，波现象是以一次反射波为主（有时也包括比较明显的多次反射波），因此地震波成像以反射波成像为主，且被分成了背景速度估计和反射系数估计两个部分。尽管两部分有不可分割的紧密联系，但它们还是表现出比较明显的独立性。从广义线性反演成像理论上讲，透射波成像和反射波成像也并没有本质差异，但是发展出的成像方法却是非常不同的。在成像域，基于反射波成像道集的拉平，进行背景速度的估计和反射波最佳聚焦的判断，把基于反射波的背景速度估计和反射系数估计统一在成像道集中，是地震波成像的最独特之处。到目前为止，背景速度建模、角度反射系数估计、AVA 分析与反演依然是服务于油气地震勘探中油藏描述的地震波成像的三个重要环节，其中反射波成像道集居于核心位置。因此，有必要把反射波成像道集的物理意义再次进行深入的讨论。本文重点从叠前地震数据与偏移成像道集的等价性、地震波成像道集的物理实质和生成成像道集的典型方法三个方面再议地震波成像道集的物理实质。为进一步提高对地震波成像的认识水平和应用能力提供基础。

【关键词】地震波成像，反射波，叠前数据，成像道集，背景速度，角度反射系数，角度道集生成

1 前言

理论上，地震波成像就是 Bayes 参数估计理论在勘探地震中的应用实例。但是，实际上，广义的地震波成像从地震数据采集系统设计、与地表相关的复杂因素的消除、精确的背景速度估计或建模，到角度反射系数估计构成了异常复杂的技术链条。在所有广义遥感领域，地震数据处理及分析总体上应该是最为困难的。根本原因当然是地震波在复杂的地下介质中传播，地震波激发接收条件被不可控的野外环境严重地复杂化。

因此，石油工业界中实用的地震波成像并没有解一个线性正问题或非线性正问题所规定的参数反演问题。而是依据地震数据仅仅通过地表的激发和接收获得，地下介质以层状为主，波现象以一次反射波为主（有时也包括比较明显的多次反射波），发展了以反射波成像为核心的地震波成像技术体系。更具体地是，在成像域，基于成像道集的拉平，进行背景速度的估计和反射波最佳聚焦的判断，把基于反射波的背景速度估计和角度反射系数估计统一在成像道集中。

这样做的根本物理依据是：数十 Hz 主频及对应十、数十到百米级波长的地震子波组成的波前面入射到大尺度（数公里、甚至更大尺度）的反射界面上，可以认为是一定方向的局部平面波以一定的角度入射到平反射界面上，产生反射波。背景层速度+角度反射系数应该是地下介质的合理的描述方式。即便发展到了以 FWI + LS_RTM 方法为主的地震波成像阶段，油气地震勘探的上述物理图景依然是占主导地位的。以背景层速度建模、角度反射系数估计和 AVA 分析/反演三个核心技术环节服务于油气勘探及油藏描述为主要框架的模式在当前的石油工业中依然发挥着主导作用。

在当前的地震波成像技术体系中反射波角度成像道集显然居于核心位置。

因此，有必要把反射波成像道集的物理意义再次进行深入的讨论。本文重点从叠前地震数据与偏移成像道集的等价性、地震波成像道集的物理实质和生成成像道集的典型方法三个方面再议地震波成像道集的物理实质。为进一步提高对地震波成像的认识水平和应用能力提供基础。

2 地震波成像的基本逻辑

从不同的角度，地震成像的基本逻辑有不同的描述方法。根本上，叠前地震数据被视为随机数据，地震波成像结果（某种弹性参数）也被视为是随机的，据此导引出 Bayes 参数估计理论下的地震波成像方法技术体系。当 Bayes 原则 $P\left(\boldsymbol{m}\middle|\boldsymbol{d}\right)=\dfrac{P\left(\boldsymbol{d}\middle|\boldsymbol{m}\right)P\left(\boldsymbol{m}\right)}{P\left(\boldsymbol{d}\right)}$ 中的两个先验概率密度函数 $P\left(\boldsymbol{d}\middle|\boldsymbol{m}\right)$ 和 $P\left(\boldsymbol{m}\right)$ 都假设是 Gauss 分布时，条件后验概率密度函数 $P\left(\boldsymbol{m}\middle|\boldsymbol{d}\right)$ 也是 Gauss 的，条件后验概率密度函数 $P\left(\boldsymbol{m}\middle|\boldsymbol{d}\right)$ 的最大化等价于代价函数（误差泛函）的最小化。这就退化为最小二乘意义下地震波成像问题。线性或非线性反演问题的求解可以认为是地震波成像的本质逻辑。

但是，这种纯数学逻辑并不能很好地指导解决实际地震勘探中的地震波成像问题。根本原因是这样的

逻辑脱离了油气地震勘探的物理实际，叠前地震数据采集的物理实际、勘探地震介质的物理实际、地震波场中波现象的物理实际都没有在上述参数估计问题中得到体现。

在地表进行的多次覆盖的激发接收、地下介质层状为主、接收波场以反射波为主的物理实际下，地震波成像的本质是：对来自地下同一反射点的、不同炮检对的反射地震子波进行同相位叠加。

同相位叠加的基本要求是从浅到深背景速度模型的正确。如果不叠加这些来自地下同一反射点的、不同炮检对的反射地震子波，它们应该是深度位置相同的、排齐的（也称拉平的）。如果从浅到深背景速度模型不正确，必然存在剩余深度差。剩余深度差与剩余速度差（当前鳖精速度模型与正确的背景速度之差）存在因果关系，基于此关系，可以构建成像域反射波层析成像速度建模方法技术。这也是当前石油工业界广泛应用的中深层速度建模方法。

如果从浅到深背景速度模型的正确，不叠加这些来自地下同一反射点的、不同炮检对的反射地震子波，就构成了（方位）角度成像道集。它们是 AVA 分析或 AVA 反演的数据基础。

成像道集的后处理是必不可少的，直接进行叠加成像会导致成像质量的严重下降。地表因素引起的反射子波振幅和相位上的不一致、噪音压制不彻底、非规则空间采样、背景速度（尤其是浅层背景速度）建模不准确、偏移算法本身等总会引起问题，导致角度成像道集拉不平、存在噪音、存在成像子波形态个幅值畸变、照明不均匀、振幅不符合 AVA 变化规律。

总之，地震波成像中，成像道集及与成像道集相关的处理是关键问题。这与相关领域的成像方法是非常不同的。

3 叠前地震数据与偏移成像道集的等价性

地震波成像中，Born 近似正问题描述叠前反射（绕射）地震数据，认为它是地下所有绕射点产生的一次绕射波的叠加，即：

$$\delta \boldsymbol{u} \triangleq \boldsymbol{u}^{cal}\left(\boldsymbol{m}\right)-\boldsymbol{u}\left(\boldsymbol{m}_0\right)\approx G\left(\boldsymbol{m}_0\right)\Delta \boldsymbol{m} G\left(\boldsymbol{m}_0\right) \quad \Leftrightarrow \delta \boldsymbol{u}=A\delta \boldsymbol{m} \tag{1a}$$

一般地，上式写为

$$\boldsymbol{d}=\boldsymbol{A}\boldsymbol{m} \tag{1b}$$

线性反演意义下，地震波成像问题被提为如下最小二乘反问题：

$$\mathcal{J}\left(\boldsymbol{m}\right)=\arg\min\frac{1}{2}\left[\boldsymbol{d}^{obs}-A\boldsymbol{m}\right]^T\left[\boldsymbol{d}^{obs}-A\boldsymbol{m}\right] \tag{2}$$

（2）式对应的法方程为：

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{m}}=0\Rightarrow \boldsymbol{m}=\left(A^TA\right)^{-1}A^T\boldsymbol{d}^{obs} \tag{3}$$

上式中的 Hessian 矩阵近似为单位阵 $A^TA\approx I$，（3）式成为一般偏移成像算法得到的成像结果，

$$\boldsymbol{m}\approx A^T\boldsymbol{d}^{obs} \tag{4}$$

（1b）中的 A 和（4）中的 A^T 是相互共轭的算子，即有下式成立，

$$\left(\boldsymbol{m},A^T\boldsymbol{d}\right)=\left(\boldsymbol{d},A\boldsymbol{m}\right) \tag{5}$$

因此，数据空间 $\boldsymbol{d}\left(\boldsymbol{x}_S,\boldsymbol{x}_R,t\right)$ 或 $\boldsymbol{d}\left(\boldsymbol{x}_m,\boldsymbol{h},t\right)$ 与像空间 $\boldsymbol{m}\left(x,\theta,\varphi\right)$ 是共轭空间。其中，θ 是入射张角，φ 是方位角。在数据空间中和在像空间中执行的某些处理是等价的。由于数据空间与地下介质之间不存在明显的局部关联性，在像空间进行某些与地质反射层相关的操作是更为方便的。图 1 展示了这种在两种空间中处理的等价关系。

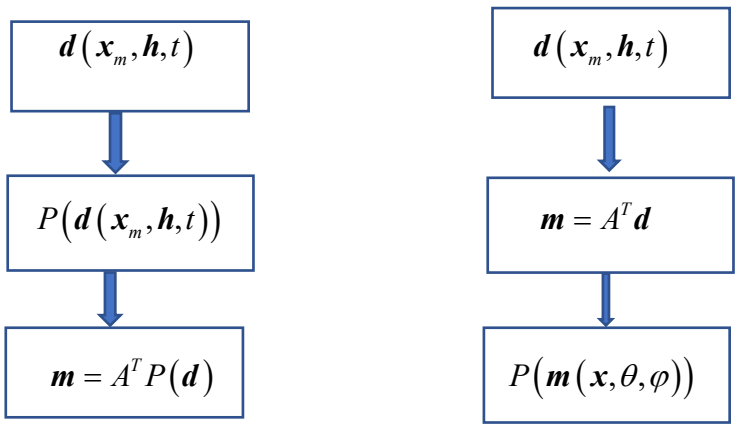


图 1 数据域与成像域成像处理的等价性

在实际的地震数据处理中，在叠前数据域$\boldsymbol{d}(\boldsymbol{x}_m,\boldsymbol{h},t)$进行各种预处理以满足线性化的偏移成像对地震数据的理论要求。但是，往往忽视像空间$\boldsymbol{m}(x,\theta,\varphi)$与叠前数据空间是共轭的、同构的。由于方位张角像$\boldsymbol{m}(x,\theta,\varphi)$的产生需要巨大的计算量和 IO 及存贮量，更容易导致放弃产生$\boldsymbol{m}(x,\theta,\varphi)$，更不用讲花功夫进行$\boldsymbol{m}(x,\theta,\varphi)$的后处理。事实上，对$\boldsymbol{m}(x,\theta,\varphi)$进行一定的处理是提高成像保真度和分辨率的必要措施。

4 地震波成像道集的物理实质

本质上，源端波场被推到成像点(x,y,z)处成为$u_s(x,y,z,t)$，检波点波场被推到成像点(x,y,z)处成为$u_r(x,y,z,t)$，两者在$t=0$（上行波出发时$t_r=0$）时相关，速度正确时，得到最大相关，生成成像结果。这也是波动理论叠前深度偏移的物理真实。很显然，$u_s(x,y,z,t)$和$u_r(x,y,z,t)$中包括了波前面、角度信息，也即包含了 Snell 定理反射关系，形成了产生角度道集的物理基础。RTM 角度道集生成，即把$u_s(x,y,z,t)$和$u_r(x,y,z,t)$中波矢量传播信息提出来，据此生成角度成像道集。

扩展成像道集是引入所谓的地下扩展 Offset，生成所谓的扩展共成像点道集。可以想象，若地震子波无限高频，特定 Offset 的波沿射线传播，$u_s(x,y,z,t)$和$u_r(x,y,z,t)$中仅包含与 Offset 对应的入射角和出射角，必须一对一相关，才能得到正确的角度 CIG。本质上，Kirchhoff 积分 PSDM 生成的 Offset CIG 道集就是这样的“角度 CIG”。

实际上，地震子波带限，空间上存在对应的 Fresnel 带效应，也可以认为波传播方向有了不确定性，或测不准效应。引入所谓的地下扩展 Offset，不能把它直观地理解成距离。物理本质上，它是在通过引入这个量寻找$u_s(x,y,z,t)$和$u_r(x,y,z,t)$中满足 Snell 反射关系的一对入射/反射波的相关，形成相应成像结果。

成像点出的入射/反射关系如图 2 所示。其中，成像点(x,y,z)处的波场满足：$u_s(x,y,z,t)$包括了所有可能的各种角度的入射波；$u_r(x,y,z,t)$包含了所有可能的各种角度的反射波。对于反射界面而言，$u_s(x,y,z,t)$和$u_r(x,y,z,t)$中当然仅包含满足 Snell 定律的一组反射/入射波，只是不知道而已。

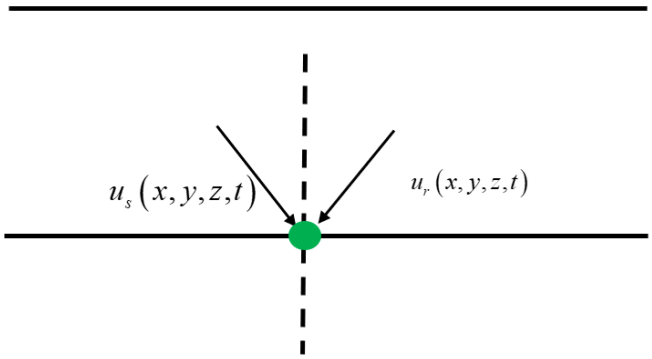


图 2 高频近似局部平面波入射到平界面上

若成像点是一散射点，也没关系，不影响我们对扩展成像道集生成机制的理解。

显式地局部平面相关成像关系式可以表达为：

$$I(x,y,z)=u_s\left(x,p_x^s,y,p_y^s,z,\omega\right)u_s\left(x,p_x^r,y,p_y^r,z,\omega\right) \tag{6}$$

同样地，球面波相关（隐式地局部平面波相关）成像关系式可以表达为：

$$I(x,y,z)=u_s(x,y,z,\omega)u_s(x,y,z,\omega) \tag{7}$$

对于反射而言，仅局部平面波入射/反射关系满足 Snell 定理时，才有相关结果。对于散射而言，各散射角度都会产生相关结果。因此，利用这种物理观察可以发展出区分反射波成像与散射波成像的成像条件。

若不引入扩展地下 Offset，对反射而言，$u_s(x,y,z,t)$和$u_r(x,y,z,t)$中包括有匹配的（即满足 Snell 定律的）局部平面波，相关生成叠加成像结果。对散射而言，$u_s(x,y,z,t)$和$u_r(x,y,z,t)$中包括了各入射、散射匹配后的成像结果。由于散射波聚在(x,y,z)处，分不出角度关系，入射/散射角的成像结果被混在一起。

为了张出成像道集，对于反射而言，需引入扩展地下 Offset，为什么引入扩展地下 Offset 就可以张出反射波成像道集？下面仅讨论 2D 情况。对地下一个反射点，$u_s(x,z,t)$和$u_r(x,z,t)$中一定包含了入射局部平面波和出射局部平面波，但我们不知道方向，也不想用方向检测方法（方向检测是很难的事情！）。此时引入扩展地下 Offset，本质上是把球面波变成局部平面波，它是个与角度相关的量。不能直面地理解扩展地下 Offset 的概念。

扩展地下 Offset 的数学物理逻辑如下：

$$
\begin{aligned}
&u(x,z,t) && \text{球面波}\\
&\downarrow\\
&u(x,p_{h_x},z,t) && \text{局部平面波}\\
&\downarrow\\
&u\left(x,\frac{k_{h_x}}{\omega},z,\omega\right) && (k_{hx},\omega)\text{域}\\
&\downarrow\\
&u(x,z,\omega)e^{\pm ik_{h_x}h_x} && \text{相移}\\
&\downarrow\\
&u(x\pm h_x,z,\omega) && \text{空移}
\end{aligned}
$$

图 3 通过局部平面波表达的扩展地下 Offset 的物理数学含义

因此，有扩展地下 Offset 的成像条件：

$$I(x,h_x,z)=u_s(x\pm h_x,z,\omega)*u_r(x\pm h,z,\omega)$$

实事上，上式公式是不区分反射和散射的，都是成立的。

另外，扩展地下 Offset 生成 CIG 道集只能对单炮道集单向波外推 PSDM 使用。用 RTM 再扩展地下 Offset 张出 CIG 道集，就不可行了，因为仅仅靠扩展 Offset 实现局部平面波的隐式分解，对于复杂构造中的波传播理论上是隐含问题的。这也说明了为什么 RTM 生成角度成像道集不用这种方式的原因。

通过对扩展地下 Offset 生成 CIG 道集的背景理论的分析，已经清楚地知道了角度成像道集的产生，根本的物理就是局部平面波入射到平界面上，满足 Snell 定律。单炮道集中，地下散射点对应的成像道集也可以用这样的思想产生。

引入扩展地下 Offset 生成角度道集的好处是避开了在炮波场和检波波场中估计局部平面波矢量方向的困难。但是，存在为问题也是明显的，即只能用在单向波外推情形和构造不复杂的探区。

5 成像道集的生成方法

前已述及，地震波成像道集在地震波成像中有核心地位。从物理上讲，成像道集的存在决定于：勘探地震中地下介质是广大的沉积层，地震波地表激发接收且多次覆盖，地震波波长与目标反射体尺度相比很小，高频近似成立。因此，不同方向的局部平面波入射到地下反射面的同一深度点上，自然就形成角度成像道集。

既然如此，生成角度成像道集的方法也已经蕴含在上述物理描述中。

对于 Kirchhoff 积分 PSDM，进行射线路径追踪，实现 Common Offset 和 Common Azimuth 数据体的叠前深度偏移，然后重抽成像 Offset CIG 道集，自然就是“角度 CIG”。不过它只是按方位 Offset 排放的，不知道真正角度的“角度 CIG”。如果变换成真正的角度 CIG，实现 Offset 到角度的投影，需要进行射线追踪，把 Offset 映射成角度。由于角度不规则插值运算是少不了的。

对于单程波深度外推，引入扩展 Offset，产生扩展 Offset CIG 道集，然后再利用局部平面分解（局部 Radon 变换），把扩展 Offset CIG 道集转换成角度 CIG 道集。

对于 RTM，需要直接检测炮端波场中的局部平面方向和检端波场中的局部平面波方向，然后利用这两个方向的局部平面波形成角度成像道集。关键问题变成如何检测局部平面波方向。在低信噪比和复杂构造探区，时间空间域局部平面波波前面是十分复杂的，很难得到正确的估计结果，CIG 道集的质量很难得到

保障。此外，计算量也是个大负担。

因此，这三种生成 CIG 道集的方式实际上各有适用性，各有优缺点。理论上最完善的当然是 RTM 中测量局部波前面波矢量方向的方法；实际中最方便的当然是 Kirchhoff 积分 PSDM 实现 Common Offset 和 Common Azimuth 数据体的叠前深度偏移，然后重抽成像 Offset CIG 道集。单程波深度外推，产生扩展 Offset CIG 道集，然后基于局部线性 Radon 变换，把扩展 Offset CIG 道集转换成角度 CIG 道集适用于构造简单的情况。总之，结合具体情况，选择最合适的方法。

6 结论与讨论

地震波成像的特殊性表现在：地震数据是通过地表的激发和接收而获得，地下介质是以层状为主，波现象是以一次反射波为主，因此地震波成像以反射波成像为主，且被分成了反射波背景速度估计和反射系数估计两个部分。

成像道集在地震波成像中具有特殊重要的地位。它把反射波背景速度估计和反射系数估计紧密地连成一体。在成像域，基于反射波成像道集的拉平，进行背景速度的估计和反射波最佳聚焦的判断，把基于反射波的背景速度估计和反射系数估计统一在成像道集中，是地震波成像的最独特之处。到目前为止，背景速度建模、角度反射系数估计、AVA 分析与反演依然是服务于油气地震勘探中油藏描述的地震波成像的三个重要环节。

从物理上讲，成像道集的存在决定于：勘探地震中地下介质是广大的沉积层，地震波地表激发接收且多次覆盖，地震波波长与目标反射体尺度相比很小，高频近似成立。因此，不同方向的局部平面波入射到地下反射面的同一深度点上，自然就形成角度成像道集。

理论上最完善的当然是 RTM 中测量局部波前面波矢量方向的方法；实际中最方便的当然是 Kirchhoff 积分 PSDM 实现 Common Offset 和 Common Azimuth 数据体的叠前深度偏移，然后重抽成像 Offset CIG 道集。单程波深度外推，产生扩展 Offset CIG 道集，然后基于局部线性 Radon 变换，把扩展 Offset CIG 道集转换成角度 CIG 道集适用于构造简单的情况。实际应用中，结合具体情况，选择最合适的方法。

在实际的地震数据处理中，在叠前数据域$\boldsymbol{d}\left(\boldsymbol{x}_m,\boldsymbol{h},t\right)$进行各种预处理以满足线性化的偏移成像对地震数据的理论要求。但是，往往忽视像空间$\boldsymbol{m}\left(x,\theta,\varphi\right)$与叠前数据空间是共轭的、同构的。由于方位张角像$\boldsymbol{m}\left(x,\theta,\varphi\right)$的产生需要巨大的计算量和IO及存贮量，更容易导致放弃产生$\boldsymbol{m}\left(x,\theta,\varphi\right)$，更不用讲花功夫进行$\boldsymbol{m}\left(x,\theta,\varphi\right)$的后处理。事实上，对$\boldsymbol{m}\left(x,\theta,\varphi\right)$进行一定的处理是提高成像保真度和分辨率的必要措施。成像道集的后处理往往能起到事半功倍的作用。

7 参考文献

[1]. Audebert F, Foridevaux P, Racotoarisoa H, et al. Insights into migration in the angle domain[J]. SEG Technical Program Expanded Abstracts, 2002, 21: 1188-1191.

[2]. Biondi B, Symes W W. Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging[J]. Geophysics, 2004, 69(5): 1283-1298.

[3]. Biondi B. Angle-domain common-image gathers from anisotropic migration[J]. Geophysics, 2007, 72(2): S81-S91.

[4]. Cheng JB, Geng JH, Ma ZT. Azimuth-preserved local angle domain imaging and its application to prestack time migration[C]. 73nd EAGE Conference & Exhibition incorporating, 2011.

[5]. Cheng JB, Wang TF, Wang CL, et al, Azimuth-preserved local angle-domain prestack time migration in isotropic, vertical transversely isotropic and azimuthally anisotropic media[J]. Geophysics, 2012, 77(2): S51-S64.

[6]. Dickens A. and Winbow G. A., RTM angle gathers using Poynting vectors. 81st Annual International Meeting, SEG,Expanded Abstracts, 2011: 3109–3113

[7]. Duveneck, E., A pragmatic approach for computing full-volume RTM reflection angle/azimuth gathers. 75th Conference and Exhibition, EAGE, Expanded Abstracts, 2013: Tu-11-01

[8]. Fomel S. Theory of 3-D angle gathers in wave-equation imaging[C]//2004 SEG Annual Meeting. Society of Exploration Geophysicists, 2004.

[9]. Guan H, Williamson P, Denel B, et al. Angle-domain common-image gathers extracted from pre-stack RTM images[M]//SEG Technical Program Expanded Abstracts 2013. Society of Exploration Geophysicists, 2013:

3767-3772.
[10]. Hu J, Wang H, Wang X. Angle gathers from reverse time migration using analytic wavefield propagation and decomposition in the time domain[J]. Geophysics, 2015, 81(1): S1-S9.
[11]. Jin H, McMechan G A, Guan H. Comparison of methods for extracting ADCIGs from RTM[J]. Geophysics, 2014, 79(3): S89-S103.
[12]. Liu LB. Angle-domain Kirchhoff prestack time migration approach in VTI media[J]. Journal of China University of Petroleum, 2012, 36(5): 51-55.
[13]. Liu SY, Wang HZ. Kirchhoff PSDM angle-gather generation based on the traveltime gradient[J]. Applied Geophysics, 2015, 12(1), 64-72.
[14]. Liu TC, Wang HZ, Liu SY. Kirchhoff PSDM angle-gather generation based on the traveltime field[C]. SPG/SEG 2014 Internal Geophysical conference, 2014. (in Chinese).
[15]. Mosher C C, Foster D J, Hassanzadeh S. Common angle imaging with offset plane waves[M]//SEG Technical Program Expanded Abstracts 1997. Society of Exploration Geophysicists, 1997: 1379-1382.
[16]. Patrikeeva N. and Sava P., Comparison of angle decomposition methods for wave equation migration. 83th Annual International Meeting, SEG, Expanded Abstracts, 2013: 3773-3778.
[17]. Prucha M L, Biondi B L, Symes W W. Angle-domain common image gathers by wave-equation migration[M]//SEG Technical Program Expanded Abstracts 1999. Society of Exploration Geophysicists, 1999: 824-827.
[18]. Sava P C, Fomel S. Angle-domain common-image gathers by wavefield continuation methods[J]. Geophysics, 2003, 68(3): 1065-1074.
[19]. Sava P, Fomel S. Time-shift imaging condition in seismic migration[J]. Geophysics, 2006, 71(6): S209-S217.
[20]. Tang B, Xu S, Zhang Y. 3D angle gathers with plane-wave reverse-time migration[J]. Geophysics, 2013, 78(2): S117-S123.
[21]. Tang C. Combining source direction vectors with wavefield decomposition to calculate angle gathers from reverse time migration[M]SEG Technical Program Expanded Abstracts 2014. Society of Exploration Geophysicists, 2014: 3831-3836.
[22]. Vyas M., Du X., Mobley E. and Fletcher R., Methods for computing angle gathers using RTM. 73rd Conference and Exhibition, EAGE, Expanded Abstracts, 2011b, F020
[23]. Vyas M., Nichols D. and Mobley E., Efficient RTM angle gathers using source directions. 81st Annual International Meeting, SEG, Expanded Abstracts, 2011a: 3104–3108
[24]. Whitmore N D, Crawley S. Applications of RTM inverse scattering imaging conditions[M]//SEG Technical Program Expanded Abstracts 2012. Society of Exploration Geophysicists, 2012: 1-6.
[25]. Wu C*, Wang H, Zhou Y, Hu J. Angle-domain common-image gathers in reverse-time migration by combining Poynting vector with local-wavefield decomposition[J]. Geophysical Prospecting, 2019, 67: 2035- 2060.
[26]. Wu RS. Prestack depth migration in angle-domain using beamlet decomposition: local image matrix and local AVA[J]. SEG Technical Program Expanded Abstracts, 2003, 22: 973-976.
[27]. Xie, X., and R. S. Wu, Extracting angle domain information from migrated wavefields: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2002, 1360–1363.
[28]. Xu S, Chauris H, Lambaré G, et al. Common-angle migration: A strategy for imaging complex media[J]. Geophysics, 2001, 66(6): 1877-1894.
[29]. Xu S., Zhang Y. and Tang B., 3D angle gathers from reverse time migration. Geophysics, 2011, 76(2): S77–S92
[30]. Yan R, Xie X B. An angle-domain imaging condition for elastic reverse time migration and its application to angle gather extraction[J]. Geophysics, 2012, 77(5): S105-S115.
[31]. Yan, R., and X. Xie, Angle gather extraction for acoustic and isotropic elastic RTM: 81st Annual International Meeting, SEG, Expanded Abstracts, 2011, 3141–3146.
[32]. Yoon K, Marfurt K J. Reverse-time migration using the Poynting vector[J]. Exploration Geophysics, 2006, 37(1): 102-107.
[33]. Yoon K., Guo M., Cai J. and Wang B., 3D RTM angle gathers from source wave propagation direction and dip of reflector. 81st Annual International Meeting, SEG, Expanded Abstracts, 2011: 3136–3140
[34]. Zhang Q. and McMechan G. A., Common-image gathers in the incident phase-angle domain from reverse time migration in 2D elastic VTI media. Geophysics, 2011a, 76(6): S197–S206
[35]. Zhang Q. and McMechan G. A., Direct vector-field method to obtain angle-domain common-image gathers from isotropic acoustic and elastic reverse-time migration. Geophysics, 2011b, 76(5): WB135–WB149
[36]. Zhang Q. RTM angle gathers and specular filter (SF) RTM using optical flow[M]//SEG Technical Program Expanded Abstracts 2014. Society of Exploration Geophysicists, 2014: 3816-3820.
[37]. Zhang Y., Xu S., Tang B., Bai B., Huang Y. and Huang T., Angle gathers from reverse time migration. The Leading Edge, 2010, 29: 1364–1371

[38]. Zhao L., Feng B. and Wang H., Efficient RTM angle gathers using source propagation direction and reflector normal direction. 82nd Annual International Meeting, SEG, Expanded Abstracts, 2012: doi:10.1190/segam2012-1160.1
[39]. Zou Z, Liu H, Liu HW. Common-angle gathers based on Kirchhoff pre-stack time migration[J]. Chinese Journal of Geophysics, 2010, 53(5): 1207-1214. (in Chinese).
[40]. 王保利, 高静怀, 陈文超, 等. 逆时偏移中用 Poynting 矢量高效地提取角道集[J]. 地球物理学报, 2013 (1): 262-268.
[41]. 王华忠，刘少勇，孔祥宁等. 大规模三维地震数据 Kirchhoff 叠前深度偏移及其并行实现．石油地球物理勘探, 2012, 47(3), 404-410
[42]. 吴成梁*, 王华忠, 冯波, 盛燊. 基于 CLG 光学流和波场分解的逆时偏移角度道集方法研究[J]. 地球物理学报, 2021, 64(4): 1375-1388.
[43]. 吴成梁, 周阳, 胡江涛, 王华忠, 孙荣. 高效逆时偏移角度道集生成方法研究[J]. 石油物探, 2018, 57(3): 404-418.

针对岩性储层描述的定量地震波成像

王华忠[1]，吴成梁[1]，盛燊[1]，许荣伟[1]，雷霆[1]，张如一[1,2]

（1.波现象与智能反演成像研究组（WPI），同济大学海洋与地球科学学院，上海 200092；2.中国石油化工股份有限公司石油物探技术研究院,江苏南京 211103）

【摘要】油气地震勘探目标越来越复杂。碎屑岩储层、火山岩储层、碳酸盐储层、非常规储层都已成为油气勘探的目标。地震波成像技术由定性的、带限反射系数为成像目标转为定量的、宽带波阻抗为成像目标，是面对复杂储层勘探时的必然选择。带限反射系数定性地刻画深度域三维空间地下介质几何结构；宽带波阻抗定量地描述深度域三维空间地下介质岩性变化。后者更适于复杂岩性储层的描述。无论是定性的带限反射系数为目标的地震波成像或是定量的宽带波阻抗为目标的地震波成像都是信息不足情形下的参数估计问题。无论叠前地震数据或是与待估计参数有关的先验信息，在复杂构造和（或）复杂岩性变化情形下，都不满足高精度成像的要求。在大数据和机器学习已经普及的情况下，定量的、信息融合理念下的地震波成像技术路线也是不可回避的。为此提出了一条针对岩性储层的定量地震波成像技术路线--特征波反演成像（CWI=Characteristic Wave Inversion）+信息融合宽带波阻抗建模（WBIM=Wide Band Impedance Modelling）--方法技术系列。本质上，它是把 FWI 宽带波阻抗反演，分解成背景速度建模、背景密度建模、线性化的宽带反射系数估计这三个凸性更好的反问题进行求解。关键是我们并没有把宽带波阻抗估计提成一个非线性的反演问题来求解，而是转化成一个信息融合问题，把已有的、认为最好的背景速度建模、背景密度建模、线性化的宽带反射系数估计结果，通过提出和求解一个约束非线性优化问题，重构出宽带波阻抗模型。这样做完全避开了反演求解宽带波阻抗时的不稳定、不收敛问题，同时在此过程中还可以进一步引入其他先验信息，在信息融合这一步继续提升宽带阻抗建模的精度。这是不同于目前常规波阻抗反演的一种新的宽带波阻抗建模方法技术，建模结果稳定，地质含义更清楚。宽带波阻抗模型各波数带的信息来源清楚，低频（低波数）来自背景波阻抗，高频（高波数）来自高分辨带限反射系数成像。中频成分的补充依赖于背景速度建模和构造成像精度的提高，也依赖于带限反射系数成像包含更充分的低中频信息（期望的带限反射系数是 Gauss 型的反射系数）。首先分析了定量地震波成像的理论基础；在此基础上提出了定量地震波成像的技术路线；把宽带波阻抗分为低波数带、中波数段、高波数带和超高波数带，分析了不同波数带波阻抗成分对宽带波阻抗建模的影响，形成了宽带波阻抗建模的技术流程。指出了定量地震波成像的核心不在于低波数背景阻抗建模结果的定理化，而是宽带反射系数成像结果的定量化。实际数据情况下，定量的宽带反射系数的获取首先要依赖于保真的地震成像，然后是保真成像结果的频带展宽，最后是用测井反射系数标定地震反射系数的量级。理论数据和实际数据的测试结果证明本文所提出的针对岩性储层的定量地震波成像技术系列的可行性、有效性和实用性。

【关键词】岩性储层、定量地震波成像、背景速度估计、宽带反射系数估计、特征波反演成像、基于信息融合的宽带波阻抗建模

1 引言

油气地震勘探最终目的：精确地描述油气藏，进行准确的含油气性分析，做出最佳的钻井决策，得到最高的油气勘探效益。油气地震勘探核心问题：由叠前地震数据及与待估计参数相关的先验信息，进行宽波数带的弹性参数估计（或称广义的高精度地震波成像），与岩石物理知识结合，进行精确的油气藏描述和准确的含油气性评价。宽波数带的弹性参数估计问题：基于叠前地震数据和待估计参数相关的先验信息，基于地震波理论和波动方程（及其各种简化形式），基于 Bayes 参数估计理论，提出并求解一个信息不足情形下的（强）非线性反问题（系统参数反演问题）。高精度地震波成像的信息源：“两宽两高（宽方位、宽频带、高密度、高信噪比）”的地震数据是最主要的信息源。而仅靠大炮数据采集的“两宽两高”叠前地震数据还是远不能满足高精度、定量、针对岩性储层描述的高精度地震波成像的需求，还需要补充待估计的与地下介质岩性变化相关的先验信息。可以说，油气地震勘探的整个发展历史，就是在地震波成像理论的指导下，由地震数据采集技术（也包括其他先验信息收集技术）的不断向前发展的牵引下向前发展的。而地震数据采集技术的进步奠基在装备技术发展的基础上。这些构成了当前油气地震勘探领域的核心技术概念。

地震波成像理论和技术把叠前地震数据及与待估计参数相关的先验信息映射成地下介质弹性参数、物性参数，进而实现高精度的油气藏描述，及进一步的高效油气藏开发。

常规的地震波成像以地下界面的带限反射系数为成像目标，高精度的速度建模也是为保真和高分辨的带限反射系数成像提供更精确的背景速度场。由于震源激发端和检波端各种因素的不可控，地震波成像方法技术并不能完全消除传播过程中的各种影响因素，即便做到保真和高分辨的带限反射系数的成像也是很不容易的。

但是，深水、深层情况下碎屑岩储层、火山岩储层、碳酸盐储层、非常规储层都已成为油气勘探的常规目标。针对这些复杂岩性油藏地震勘探的需求，地震波成像技术由定性的、带限反射系数为成像目标转

为定量的、宽带波阻抗为成像目标成为了必然的选择。

当前，石油工业界开展岩性油藏勘探的地震波成像主要依赖三项关键技术：层析速度反演技术；保真和高分辨方位角度带限反射系数成像方法技术；AVA 分析/反演方法技术。三项技术是串联的，最后一项技术由角度反射系数反演界面两边弹性参数的变化（Δv_p 纵波速度扰动；Δv_s 横波速度扰动；$\Delta\rho$ 密度扰动）。存在的主要问题是 AVA 分析/反演的基本理论基础是奠基在 Zoeppritz 方程基础之上的，而 Zoeppritz 方程是在平面波入射到广大的平反射界面上的假设情况下推导出的。这在碳酸盐岩油藏、火山岩油藏等情形下是不合适的假设。另一方面，保真带限反射系数估计本身就很困难，反演结果存在很大的不确定性。常规 1D 波阻抗反演是当前波阻抗建模的主要方法技术，但是 1D 波阻抗反演仅仅适用于水平层状介质情形[1]。历史上，该项技术主要用于缓横向变化的碎屑岩沉积地层下的油气勘探中，这显然已经完全不适用于目前横向变化剧烈的复杂油气藏的勘探。

Bayes 参数估计理论下的全波形反演（FWI）技术，已经成为业界公认的高精度地震波成像的标志性技术[2]。从理论本质上看， FWI 技术的目标是估计宽波数带的速度变化（或者波阻抗变化），它当然是定量的地震波成像方法。众所周知，当前 FWI 被定义为主要利用透射波估计高精度的背景速度模型，为逆时偏移（RTM）或最小二乘逆时偏移（LS_RTM）提供更好的初始速度模型。即便如此，也需要退化成利用透射波走时信息，并且在海上低频长偏移距数据情形下，才有可能产生实用的效果。根本的困难在于实际观测数据不能满足 FWI 反演的理论假设；在于复杂介质变化（小尺度、强变化）情形下 FWI 反演的高度非线性性。因此，实际数据情形下利用 FWI 反演技术进行宽带波阻抗为目标的成像显然是不可能的。

实际数据的地震波反演成像从来不是一个纯数学问题，即便“两宽两高”的叠前地震数据依然不能满足岩性储层定量反演成像的需求，必须发展出地质信息约束下的地震波成像新理论与新方法。定性的地震波偏移成像的技术研究已经趋于完善。但是，任何更为高端的技术都是在现有技术基础上发展出来的。充分保留现有方法技术的优点，充分借鉴 Bayes 参数估计理论下的 FWI 的理论优势，充分利用大数据和机器学习（ML）中的理念和算法[3]，发展信息融合的深度域真三维宽带波阻抗成像，认为是合理的针对岩性储层的定量地震波成像技术路线。

宽波数带波阻抗可以被分为低波数带（譬如 0-3Hz）、中波数带（譬如 4-8Hz）、高波数带（譬如 9-50Hz）和超高波数带（譬如 51-80Hz）四个部分[4]（注：其中的数字分划是相对的）。背景阻抗的信息来源一定是背景速度场和背景密度场；高波数带阻抗的信息源一定是宽带反射系数（或宽带的速度扰动）。中波数波阻抗的信息来源更多地依赖于与地下介质波阻抗变化有关的（地质）先验信息；超高波数波阻抗的信息源应该是测井数据和岩石物理知识。

分析可知，由不同炮检对地震道中透射波和/或反射波地震子波传播的旅行时估计背景速度/各向异性参数（等于各向异性背景速度 $v(\boldsymbol{x},\theta)$），背景速度估计结果是定量的。因为估计背景速度所用的正问题本质上由 $t=\int_A^B \frac{dl}{v(\boldsymbol{x})}$ 决定，由此正问题可清楚地看出，由 t 求 v 得到的是与 v 的真实量级一致的定量化反演。但是由地震子波的振幅估计带限反射系数或速度扰动，由于地震子波的振幅值量级是相对的（不同类型的震源、震源与介质的耦合；不同类型的检波器、检波器与介质的耦合等使得记录的波场振幅是没有统一标准的；成像方法不能消除所有的波传播效应），导致带限反射系数的估计结果并不能定量化。目前的带限反射系数成像追求的仅仅是保真及高分辨率的结果，并非是定量地描述地下介质的反射系数。实质上，当前反射系数的定量表达式 $R^P=\frac{v_2^P\rho_2-v_1^P\rho_1}{v_2^P\rho_2+v_1^P\rho_1}$ 也仅仅是高度抽象后的数学公式，并非能很好地描述带限子波入射到实际地层时应该表现出的反射系数（就是说目前的褶积模型是存在问题的）。

另外，分析可知，实际数据情形下的 FWI 也是不能得到由低波数，到中波数，再到高波数的波阻抗（或速度）反演结果的。众所周知的关系式 FWI = Tomography + LS_RTM 仅仅在反演骗局（Inverse Crime）意义下是成立的[5]。对于实际数据，这个公式会引起误导。在实际数据情形下，基于反射地震子波振幅的 LS_RTM 方法并不能得到与地下反射系数同等量级的反演结果，因此高波数速度扰动的估计过程是不可能收敛的（连反演结果量级都不能与低波数反演结果保持一致，如何能产生收敛的反演结果！）。

依据上述分析，我们（WPI 研究组）提出了**特征波反演成像（CWI = Characteristic Wave Inversion Imaging）与基于信息融合的宽带波阻抗建模（WBIM = Wide Band Impedance Modeling）**的技术路线[6]，

它是一种针对岩性储层的定量地震波成像方法技术系列。关键技术包括：背景速度模型建立方法、背景密度模型建立方法、保真的宽带反射系数估计方法、基于新融合的宽带波阻抗建模方法。必须说明的是：提高各环节建模精度的先验信息获取方法（包括先验信息的正则化约束方法）也是十分重要的关键方法技术。原因是即便有“两宽两高”的叠前地震数据还是不能很好地满足定量地震波成像的需求，必须引入尽可能多的先验信息，才有可能进一步提升各波数带的波阻抗建模精度。这与当前大数据、人工智能和机器学习越来越多地渗透到勘探地震全流程中的趋势是完全一致的。因此，我们认为基于信息的、定量的地震波成像代表了地震波成像技术的发展趋势。

理论数据和实际数据上的应用实践证明了我们提出的特征波反演成像（CWI）与基于信息融合的宽带波阻抗建模（WBIM）流程及方法技术的有效性。

2 定量地震波成像理论基础

前已述及，油气地震勘探的核心是宽波数带的弹性参数估计提供高精度油藏描述的基础信息，结合井数据、岩石物理和油气地质知识，由高精度油藏描述和含流体性评价提供最佳的井位决策，获得最佳的勘探效益。很显然，油气勘探中宽波数带的弹性参数估计（或称高精度地震波成像）占据核心地位。我们把它定位成：提出并求解一个信息不足情形下的（强）非线性反问题（系统参数反演问题）。

宽波数带的弹性参数估计（或称高精度地震波成像）的信息源包括：（尽可能是“两宽两高”的）叠前地震数据；与待估计参数有关的先验信息。

Bayes 估计理论奠定了宽波数带的弹性参数估计（或称高精度地震波成像）的理论基础。叠前地震数据被记为$\boldsymbol{d}^{obs}\left(\boldsymbol{x}_S,\boldsymbol{x}_R,t\right)$；待估计的宽波数带的弹性参数被记为$\boldsymbol{m}\left(\boldsymbol{x}\right)$。叠前地震数据一般地被认为是五维数据体；反演成像结果是深度域真三维的。叠前地震数据被认为是随机过程，实际野外数据采集被认为是该随机过程的一次实现。认为波动方程（或某种变种）建立起了$\boldsymbol{m}\left(\boldsymbol{x}\right)$与$\boldsymbol{d}^{obs}\left(\boldsymbol{x}_S,\boldsymbol{x}_R,t\right)$之间的一定的预测关系。很显然，这样的预测关系是不确定的，波动方程（或某种变种）并不能完全地预测$\boldsymbol{d}^{obs}\left(\boldsymbol{x}_S,\boldsymbol{x}_R,t\right)$中包含的各种复杂的波现象。实测数据的不确定性，波动方程预测结果的不确定性，使得我们有理由认为估计的结果也是不确定的。这就构成了 Bayes 估计理论的思想基础。

经典的 Bayes 公式为：

$$P\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)=\frac{P\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)P\left(\boldsymbol{m}\right)}{P\left(\boldsymbol{d}^{obs}\right)} \tag{1}$$

其中，$P\left(\boldsymbol{m}\right)$代表先验的表达待估计参数的不确定性的概率密度函数；$P\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)$代表先验的波动方程（或某种变种）预测实测数据时不确定性的概率密度函数（也称似然函数），这里面既有波动方程的不合适引起的预测结果的不确定性，也有实际数据观测中各种复杂因素引起的观测结果的不确定性，震源激发和检波器性能引起的不确定性也都包容在这个概率密度函数中。可以看出，（1）式中包含了实际地震数据观测、波动方程对高精度地震波成像的不满足；也包含了引入与待估计参数有关的先验信息的必要性。仅仅靠波动方程对实际数据的预测误差很小（或似然函数$P\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)$取最大），并不能（远远不能）保证反演结果的可靠性。$P\left(\boldsymbol{d}^{obs}\right)$是描述采集数据过程表现出的不确定性的概率密度函数。一般地，我们很难获取这方面的先验信息，索性假设$P\left(\boldsymbol{d}^{obs}\right)$满足均匀分布，不再引入先验信息。因此，（1）被改写为：

$$P\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)\propto P\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)P\left(\boldsymbol{m}\right) \tag{2}$$

很显然，$P\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)$是后验条件概率密度函数。所谓后验概率是指实际数据观测已经获取的情况下，待估计参数$\boldsymbol{m}\left(\boldsymbol{x}\right)$发生的概率。参数估计的合理逻辑选择自然是：后验条件概率密度函数$P\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)$取最

大时对应的$\hat{\boldsymbol{m}}(\boldsymbol{x})$就是估计结果（或反演结果）。这被称为后验概率最大化（MAP = Maximum A Posteriori）准则，是 Bayes 参数估计理论的基本逻辑。

实际上，$P(\boldsymbol{m})$和$P\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)$都是很难实际计算得到的。因此，先验地假定它们均满足 Gauss 概率分布。很多实际随机过程被假定满足 Gauss 概率分布，这是有大数定理和中心极限定理支持的。因此有，

$$P(\boldsymbol{m})=\frac{1}{\sqrt{(2\pi)^{M}\det(C_{\boldsymbol{m}})}}\exp\left(-\frac{1}{2}\boldsymbol{m}^{T}C_{m}^{-1}\boldsymbol{m}\right) \tag{3a}$$

$$P\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)=\frac{1}{\sqrt{(2\pi)^{N}\det(C_{\boldsymbol{d}})}}\exp\left(-\frac{1}{2}\left(\boldsymbol{d}^{obs}-K(\boldsymbol{m})\right)^{T}C_{\boldsymbol{d}}^{-1}\left(\boldsymbol{d}^{obs}-K(\boldsymbol{m})\right)\right) \tag{3b}$$

其中，$C_{\boldsymbol{m}}$代表待估计参数的协方差矩阵，M是它的维度；$C_{\boldsymbol{d}}$代表待估计参数的协方差矩阵，N是它的维度。$K(\boldsymbol{m})=\boldsymbol{d}$抽象地表示波动方程（或其变种）对实测数据的预测关系，数学上更多用第一类 Fredholm 积分方程来表示，而非用微分形式的波动方程。地震勘探中 Born 近似后的散射场计算式就是典型的第一类 Fredholm 积分方程，它建立起了地下介质参数的扰动量与波场扰动量之间的数学关系式。

把（3）代入（2）式得到

$$\begin{aligned}P\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)&\propto L\cdot\exp\left\{-\frac{1}{2}\left(\left(\boldsymbol{d}^{obs}-K(\boldsymbol{m})\right)^{T}C_{\boldsymbol{d}}^{-1}\left(\boldsymbol{d}^{obs}-K(\boldsymbol{m})\right)+\boldsymbol{m}^{T}C_{\boldsymbol{m}}^{-1}\boldsymbol{m}\right)\right\}\\&=L\cdot\exp\left\{-\frac{1}{2}S(\boldsymbol{m})\right\}\end{aligned} \tag{4}$$

其中，$L=\frac{1}{\sqrt{(2\pi)^{N}\det(C_{\boldsymbol{d}})}}\cdot\frac{1}{\sqrt{(2\pi)^{M}\det(C_{\boldsymbol{m}})}}$。

很显然，根据指数函数的单调性，后验概率密度最大化等价于$S(\boldsymbol{m})$取极小。$S(\boldsymbol{m})$被称为代价函数或误差泛函。Bayes 参数估计问题转化为了误差泛函求极小问题。

$$\hat{\boldsymbol{m}}=\arg\min_{\boldsymbol{m}\in\mathcal{M}}S(\boldsymbol{m})=\arg\min_{\boldsymbol{m}\in\mathcal{M}}\left(\left(\boldsymbol{d}^{obs}-K(\boldsymbol{m})\right)^{T}C_{\boldsymbol{d}}^{-1}\left(\boldsymbol{d}^{obs}-K(\boldsymbol{m})\right)+\boldsymbol{m}^{T}C_{\boldsymbol{m}}^{-1}\boldsymbol{m}\right) \tag{5}$$

这就是最经典的 FWI 反演理论。其目标显然是全频带的、定量的参数估计。理论上看，Bayes 估计理论下的 FWI 似乎是最完美的地震波成像方法。

（5）式求解根本的困难在于实际观测数据不能满足 FWI 反演的理论假设；在于复杂介质变化（小尺度、强变化）情形下它表现出高度非线性性。

经典的勘探地震中，我们并没有硬解（5）式定义的反问题，而是把正问题（波动方程对数据的预测关系）退化成线性的，即

$$\boldsymbol{d}=K(\boldsymbol{m})=K(\boldsymbol{m}_{0})+\frac{\partial K}{\partial\boldsymbol{m}}\Delta\boldsymbol{m}+\frac{1}{2}\frac{\partial^{2}K}{\partial\boldsymbol{m}^{2}}\Delta\boldsymbol{m}^{2}+\cdots\approx\boldsymbol{d}_{0}+\frac{\partial K}{\partial\boldsymbol{m}}\Delta\boldsymbol{m} \tag{6a}$$

$$\Delta\boldsymbol{d}=\boldsymbol{d}-\boldsymbol{d}_{0}=\frac{\partial K}{\partial\boldsymbol{m}}\Delta\boldsymbol{m}=F\Delta\boldsymbol{m} \tag{6b}$$

其中，$\boldsymbol{d}_{0}=K(\boldsymbol{m}_{0})$，$F=\frac{\partial K}{\partial\boldsymbol{m}}$。把（6）式代入（5）式，得到

$$\Delta\hat{\boldsymbol{m}}=\underset{\Delta\boldsymbol{m}\in\mathcal{M}'}{\arg\min}\left(\left(\Delta\boldsymbol{d}^{obs}-F\Delta\boldsymbol{m}\right)^{T}C_{\Delta\boldsymbol{d}}^{-1}\left(\Delta\boldsymbol{d}^{obs}-F\Delta\boldsymbol{m}\right)+\Delta\boldsymbol{m}^{T}C_{\Delta\boldsymbol{m}}^{-1}\Delta\boldsymbol{m}\right) \tag{7}$$

（7）式是地震波层析成像和 LS_PSDM 共同遵循的参数估计反问题提法。因此有了 FWI = Tomography + LS_PSDM 的说法。层析成像解决背景速度估计问题；LS_PSDM 解决带限反射系数（或速度扰动量）估计问题。通过（6）式表示的线性化过程，（5）式定义的非线性反问题被转化为线性的。线性化的反问题理论上具备凸性，存在唯一解。因此（7）式的求解理论上应该更稳定、迭代解的收敛性更好、反演结果更精确[7,8]。实际上，常规的地震波成像过程中的三项关键技术：层析成像、偏移成像、AVA 反演成像都是线性反演理论在勘探地震中的应用实践。可以看出，常规的地震波反演成像目的也是期望得到宽波数带的弹性参数估计。但核心问题出在 Zoeppritz 方程的理论假设（平面波入射到广大的地下介质平反射界面上）不适应当前常见的横向变化剧烈的介质情形；偏移成像给出的反射系数不保真，与地下介质反射界面的“真”反射系数差异可能很大，AVA 反演缺乏可靠的数据基础。常规的 1D 波阻抗反演本质上是地下介质水平层状假设下的 FWI 反演，（5）式定义的 FWI 遇到的问题，1D FWI 同样不能回避。更重要的是，目前油气勘探遇到的复杂介质和复杂油藏情形已经不符合其水平层状介质假设。

因此，针对岩性油藏的定量地震波成像需要新的更合理的技术路线。我们是通过修改（5）式定义的 FWI 形成**由特征波反演成像+基于信息融合的宽带波阻抗建模构成的定量地震波成像路线**。

由前述 Bayes 估计理论可知，波动方程（或各种变种）对实测数据（中的波现象）的预测关系，本质是待估计参数与波现象之间的关系是否接近线性决定了形成的反问题是否凸的，反问题是否有很好的凸性决定了反问题的求解是否能得到稳定可靠的解。先验信息的引入可以促进解的稳定性和精度进一步提升，但过分依赖初始模型或先验信息提高解的精度逻辑上是不合理的。因为先验信息的获取往往比“两宽两高”的地震数据采集更困难，代价更大。比如我们不可能钻很多井来获取关于地下介质弹性参数、物性参数的先验信息。

为此，我们提出了特征波反演成像的概念，（5）式被改写成，

$$\hat{\boldsymbol{m}}=\underset{\boldsymbol{m}\in\mathcal{M}}{\arg\min}\left(\left(\boldsymbol{d}_{C}^{obs}-\boldsymbol{d}_{C}^{cal}\right)^{T}C_{\boldsymbol{d}'}^{-1}\left(\boldsymbol{d}_{C}^{obs}-\boldsymbol{d}_{C}^{cal}\right)+\boldsymbol{m}^{T}C_{\boldsymbol{m}}^{-1}\boldsymbol{m}\right) \tag{8}$$

其中，$\boldsymbol{d}_{C}^{obs}=T\left(\boldsymbol{d}^{obs}\right)$，$T\left(\cdot\right)$代表一个特征波现象提取算子，把实测炮集中的特征波现象$\boldsymbol{d}_{C}^{obs}$提取出来。譬如初至波；特征反射层对应的反射波都可以认为是特征波。而且$\boldsymbol{d}_{C}^{obs}$也可以是特征波的走时。$\boldsymbol{d}_{C}^{cal}\left(=K_{C}\left(\boldsymbol{m}\right)\right)$代表特征波传播算子$K_{C}$模拟出的特征波现象。

我们认为特征波成像主要包括初至波层析成像和特征反射层导引下的特征反射波层析成像。常规的射线理论折射波层析、基于各种反射波时距关系的 NMO 速度分析、共成像点道集射线层析成像也可归于特征波成像的范围。

很显然，特征波成像是基于不同波现象同相轴上地震子波的走时估计背景速度，这是一个定量反演过程。所得的背景速度与地下介质速度量级是一致的。它们可被认为是定量的成像方法。

（7）式规定的另一个线性化反演问题是最小二乘叠前深度偏移（LS_PSDM）或 LS_RTM。常规的地震波偏移成像是 LS_PSDM 或 LS_RTM 的简化实现。（7）式的数学求解可以表示为：

$$\begin{aligned}&-F^{T}C_{\Delta\boldsymbol{d}}^{-1}\left(\Delta\boldsymbol{d}^{obs}-F\Delta\boldsymbol{m}\right)+C_{\Delta\boldsymbol{m}}^{-1}\Delta\boldsymbol{m}=0\\&\left(F^{T}C_{\Delta\boldsymbol{d}}^{-1}F+C_{\Delta\boldsymbol{m}}^{-1}\right)\Delta\boldsymbol{m}=F^{T}C_{\Delta\boldsymbol{d}}^{-1}\Delta\boldsymbol{d}^{obs}\end{aligned} \tag{9}$$

常规的地震波偏移成像的数学表述式为，

$$\Delta\boldsymbol{m}\approx F^{T}C_{\Delta\boldsymbol{d}}^{-1}\Delta\boldsymbol{d}^{obs} \tag{10}$$

实际上，石油工业界常用的偏移成像思想和方法技术并非由上述 Bayes 估计理论退化产生的，而是由物理直觉与波传播算子相结合导出的。认为地表观测的波场是由地表主动源产生的下行波激活地下介质波阻抗异常体，它产生上行绕射波场，所有地下介质中的波阻抗异常体产生的上行波场的叠加形成地表观测

波场。尤其考虑到多次覆盖的地表数据激发接收方式，地震波成像就是把所有地表接收的、由所有炮激发的上行绕射波场推到或收敛到产生它们的波阻抗异常体上。Kirchhoff 积分偏移成像就是直接地按上述物理逻辑实现的；单平方根波动方程偏移是按正向外推炮端激发波场与反向外推接收端波场并在波阻抗异常体处的相关（相关最大对应偏移速度模型是正确的）进行成像的；双平方根波动方程偏移是按炮点和检波点同时沉降的方式，把炮点和检波点外推到波阻抗异常体处按成像条件（上行波出发时刻为零）提取成像值进行成像的。但是，这些具体的实现方法本质上与 Bayes 估计理论退化出的（10）式是统一的。只不过前者物理意义和几何意义更清晰；后者的数学算法更严谨。

很显然，石油工业界常用的偏移成像只不过是对地下波阻抗异常体的位置进行定位，定位精度依赖于背景速度。定位结果的表现是由地震子波的振幅体现出来的，正是“放置或搬家”到成像点处反射子波的振幅体现出地下介质反射系数的相对变化（并期望这种相对变化能比较真实地反映地下介质反射系数真实情况）。直到目前的油气地震地质解释主要还是利用了地震波成像提供的波阻抗异常体的空间定位信息和保真的反射强度信息。事实上，相关成像条件得到定位结果是由下行波场与外推到成像点处的下行波场中子波的相关系数体现出来的，与地震子波的振幅关系更远。如果不强调偏移成像的保真处理，成像结果得到的带限反射系数与地下介质反射系数之间可能相差极大。

实际观测波场中，各炮震源激发产生的地震波能量（振幅）有复杂的影响因素，炮和炮之间差异很大；检波器测震结果的物理意义不同，有位移检波器、速度检波器、加速度检波器和压力检波器，各检波器与介质的耦合相差很大，检波器本身对不同频率响应的一致性很难保证，这都导致检波器检测的地震波振幅是相对的。而且各检波器对同一炮激发的震动的响应差异很大，即便相邻检波器测量的同一反射层的反射子波振幅也很不相同。这些复杂的因素并不能被正问题 $K\left(m\right)=\boldsymbol{d}$ 所描述。地震波反演成像中的正算子往往是很简单的，常密度声波方程是最常用的，而且又被引入了 Born 近似，复杂介质对波场的改造作用不能在成像过程中得到消除。因此，实际数据情况下，即便 LS_RTM 也不能得到收敛的定量成像结果。理论框架下的 FWI 和 LS_RTM 在自身的理论假设下可以得到可靠的定量成像结果。遗憾地，这只是个反演骗局（Inverse Crime）。

当前，西方地球物理公司宣传的 FWI_Imaging[9-11]或全波数带 FWI[12]试图由透射波 FWI 估计低波数带背景速度，然后进到反射波 FWI 估计中波数带的背景速度，再进展到估计高波数带的速度扰动量（这已经是 LS_RTM 解决的问题，这是 LS_RTM 被称为高波数参数反射波 FWI 的原因）。事实上，由于 FWI 反演迭代过程中，由利用波现象同相轴上地震子波的走时信息到利用地震子波的振幅信息，迭代反演中的梯度项由定量变为“定性”，不受控制的 FWI 根本不可能由定量的背景速度的更新自动地转到定量的高波数扰动量的更新。因此，尽管在 Inverse Crime 意义下，FWI 的梯度项由背景速度的更新自动地转到高波数扰动量的更新似乎是理所当然的（事实上，也不是那么地理所当然!）；但对于实际数据，这是不可能自动发生的!

影响地震波振幅的因素非常复杂，基于地震子波振幅的参数估计或定量成像到目前为止并没有完善的解决方案。上述分析清楚地说明：针对岩性储层的定量地震波成像的真正困难在于高波数参数的定量化估计。更明确地，保真与宽带反射系数成像是定量地震波成像的难点。

假定已经获得了定量的背景速度、背景密度、保真与宽带反射系数，基于上述三种成像结果的地震地质解释依然不是深度域真三维空间的油气地质信息的真实反映。深度域真三维空间的宽带波阻抗才是更能反映地下介质岩性变化情况的成像结果。深度域真三维空间的宽带波阻抗自然不能作为 FWI 的成像目标，这只会徒然增加反问题的非线性性，求解过程更不收敛，多解性更严重。这正是我们提出信息融合宽带波阻抗建模的本质原因，信息融合宽带波阻抗建模避开了 FWI 反演宽带波阻抗的强非线性性，可以把更多的与阻抗有关的先验信息融入到宽带波阻抗建模建模过程中，可以充分借用当前机器学习领域发展的新思想和新算法。

定量地震波反演成像是个工业界有急迫需求，理论上也存在待解决问题的研究方向。我们（WPI）提出的 CWI（Characteristic Wave Inversion）+WBIM（Wide Band Impedance Modeling ）策略就是一种定量的地震成像技术路线。

3 定量地震波成像技术路线

前已述及，当前油气地震勘探主要还是利用了地震波成像提供的地下介质波阻抗异常体的定位信息，

以及偏移成像算法附带产生的相对保真的反射强度信息，进行地震地质解释，在井数据和油气地质知识的辅助下进行井位决策和钻井方案决策。

很显然，地震波成像由定性的、主要反映地下介质几何形态附带表现地下介质反射强度变化的带限反射系数成像发展到定量的、反映地下介质岩性变化的宽带波阻抗成像，显然代表了油气工业的重大需求。图 1 展示了由叠前地震数据+与待估计参数相关的先验信息映射成油藏描述需要的定量地震成像结果的三条技术路线。

由前述理论分析知道，Bayes 估计理论下的 FWI 是定量地震波成像最为完善的方法技术。但是，实际数据情况下，FWI 直接反演宽带阻抗是不可能的。这是一个高度非线性的反演问题。

常规的做法包括三项核心技术：背景速度层析成像；带限的角度反射系数成像；AVA 反演（也包括 1D 波动方程波阻抗反演），存在的问题（分析后认为）主要是：AVA 反演的 Zoeppritz 方程理论假设不适应当前复杂的勘探目标；角度反射系数成像结果存在严重的不确定性。另外，1D 波动方程波阻抗反演的理论假设与当前勘探实际差异过大，横向变化剧烈的介质情况下，根本不存在满足 1D 波动方程的 1D 数据。

可以看出，图 1 中所示的、走向定量油藏描述的常规技术路线越来越不适应当前复杂介质与复杂油藏的勘探实际。

HAFFINGER 等[13]提出基于 FWI 反演的 AVA 弹性参数反演流程（WEB-AVO = Wave-equation Based AVO Inversion）试图解决定量储层描述问题。但是，基于 FWI 背景速度及叠前深度偏移成像道集和 1.5D 弹性参数反演这三项关键技术的 WEB-AVO 存在很多逻辑上令人质疑的地方，譬如叠前深度偏移给出的零偏移距地震剖面的是否满足 1.5D 弹性波方程的要求？1.5D 弹性波反演依然是个强非线性问题，如何能得到稳定的反演解？另外，1.5D 弹性参数反演假设地下介质水平层状也不适应当前横向变化复杂的勘探目标。

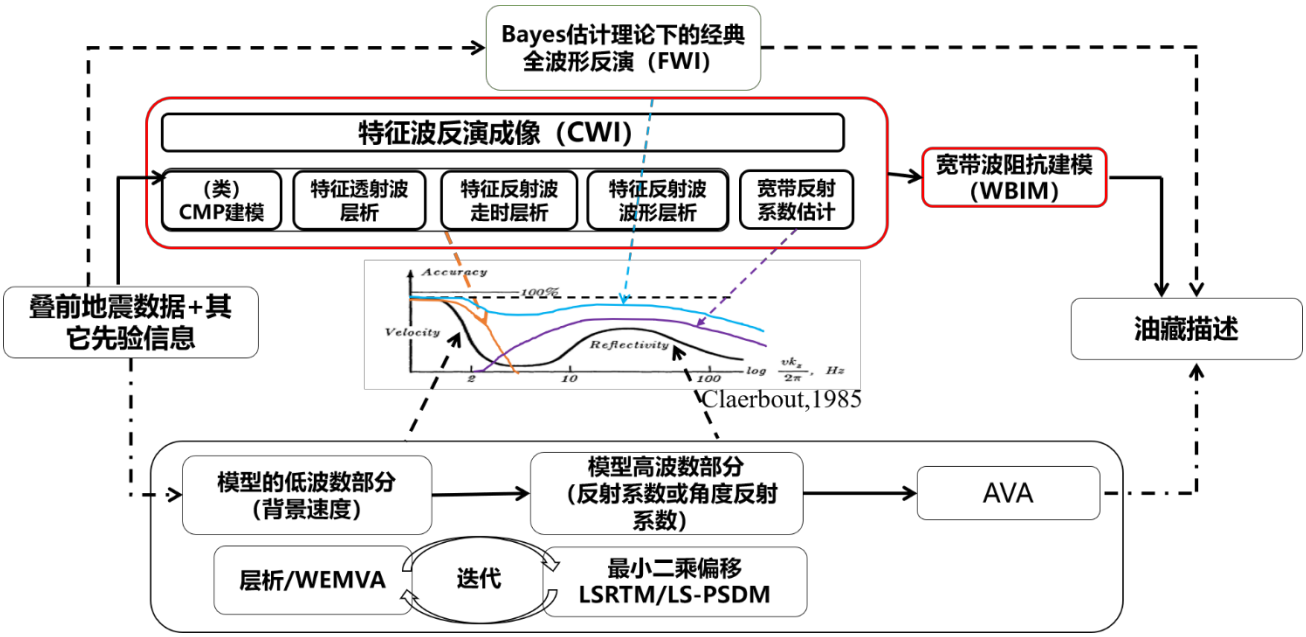


图 1 由叠前地震数据+与待估计参数相关的先验信息映射成油藏描述需要的定量地震成像结果的三条技术路线。

根据上一节的理论分析，我们（WPI）提出了图 2 所示的另外一条定量地震波成像技术路线。

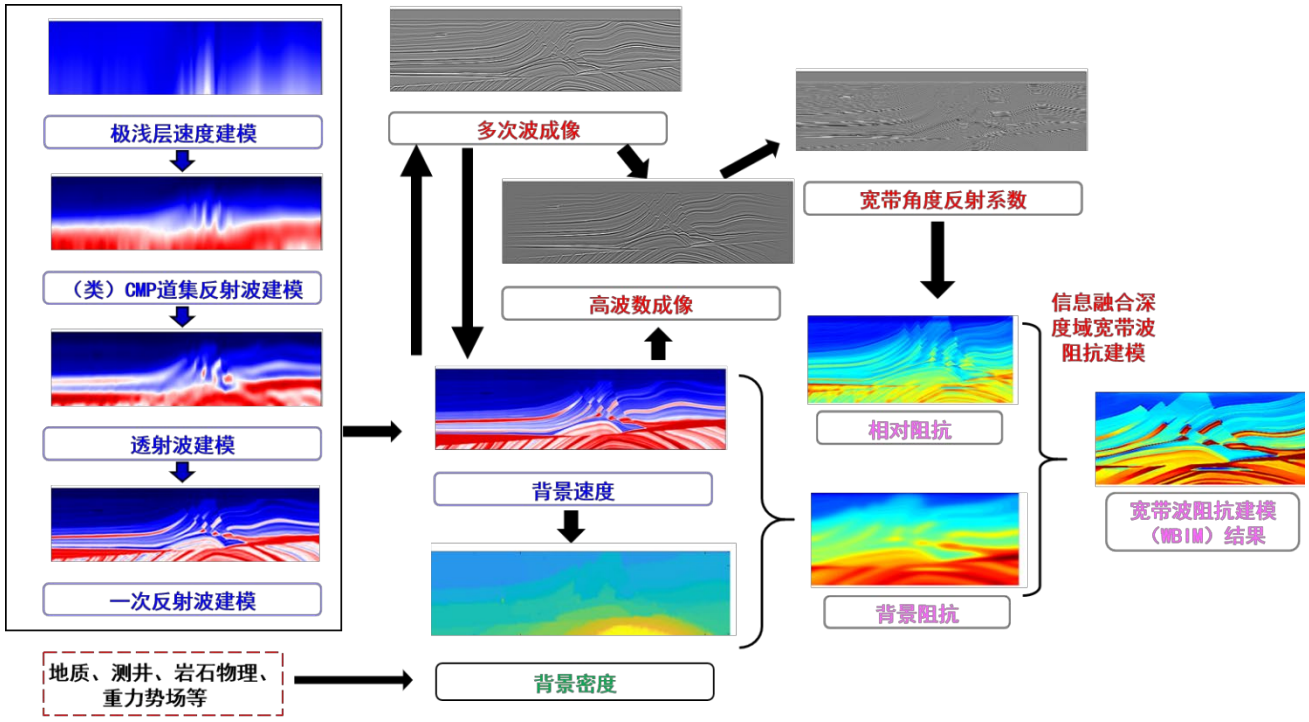


图 2 基于特征波成像+宽带波阻抗建模的定量地震波成像技术路线。

该技术路线中包括背景速度建模、背景密度建模、宽带反射系数（或高波数参数扰动量）估计及信息融合宽带波阻抗建模四大部分。

背景速度+背景密度构成的背景阻抗对于岩性油藏的描述与评价具有重要的意义。背景速度模型的构建包含如下五项关键技术：

① 可靠的初至波识别与初至走时拾取方法技术，我们提出：弱初至埋在强噪声情形下，基于高维、多属性、Markov 最佳状态转移理论框架，充分利用机器学习领域发展出的新算法，实施初至波识别与走时拾取。不能认为初至波 FWI 可以避开可靠的初至波识别与初至走时拾取的困难。事实上，复杂初至情形下初至波 FWI 也不能提供好的背景速度反演结果。除了初始模型差异过大之外，这正是初至波走时测量精度不能保证所引起的。单独处理初至波识别与初至波走时拾取可以通过各种预处理和先验信息的引入解决实际数据（尤其是陆上地震数据）不满足初至波 FWI 理论假设引起的问题。到目前为止，弱初至埋在强噪声情形下的初至波识别与初至走时拾取依然是没有得到很好解决的问题。

② 高效率且适应非水平地表的初至波走时波动理论层析成像方法技术。

③ （类）CMP 道集动校正扫描速度分析方法技术。类 CMP 道集是指 PSTM 或 PSDM 后成像道集反动校正后得到的所谓的 CMP 道集，反动校正用偏移速度。类 CMP 道集受横向变速的影响小，动校正速度分析结果更符合地质实际。

④ 特征反射结构导引下的特征反射波层析成像方法技术。利用反射波建立中深层背景速度不需要所有层位的反射波信息。基本原则是先用一些特征反射结构（特征反射层）对应的反射波同相轴（走时），进行大套层层速度的估计。然后逐渐放入一些相对弱一些的反射层对应的反射波同相轴（走时），提高层速度估计结果的分辨率。

⑤ 反射波 FWI 层析成像方法技术。在步骤④提供的初始速度模型的基础上，进行少许几次反射波 FWI 层析成像迭代。其中的关键是要引入对梯度有地质含义的正则化，并尽可能压制掉偏移响应梯度项的影响。

油气地震勘探中，背景密度模型的构建是比较困难的。原则上，选择包含密度的波动方程（譬如变密度声波方程），利用 FWI 进行多参数反演是可以反演密度的。但这又是一个反演骗局。实际数据情形下，密度变化会影响速度，速度影响到达时，似乎可以由到达时反演密度。但是，密度和到达时的数学关系是什么？目前缺乏这样的数学物理关系式。可能是非线性的异常复杂的关系。密度变化影响同相轴上地震子波的振幅是确定的，密度变化导致波阻抗变化，导致反射子波振幅变化是确定的。但是，影响振幅的因素

太多了。这构成了由叠前地震数据，基于 Bayes 估计理论下的 FWI 反演密度，因为非线性性很强，极难在实用中见到效果的根本原因。据此，我们提出两条构建背景密度模型的技术路线。首先由测井数据提供的密度信息和地震波成像提供的反射结构信息，提出多信息约束下的密度场插值反问题，这是个多约束条件下的非线性性优化问题，解此优化问题，得到背景密度场。如果探区内有重力势场观测数据，再提一个重力势场约束下多信息约束的密度反演问题，把前一步插值方法获得的背景密度场作为初始值，求解该反问题得到所要的背景密度场。由背景速度场和背景密度场就可以得到背景波阻抗。为了方便，也可以借用已有的速度和密度之间的岩石物理关系式，把密度场计算出来，从而产生背景密度场。

宽带反射系数（或高波数参数扰动量）估计在宽带波阻抗建模中提供高波数带（譬如 9-50Hz）的（相对）波阻抗变化信息。理论上，我们需要的是零角度反射系数。实际情况下，一般由角度反射系数的加权叠加作为零角度反射系数的近似。理论上，我们需要的是只有主瓣没有旁瓣的零角度反射系数，主瓣宽度也要尽可能地小。实际上，这也是地震波成像的期望目标。有限频带的震源激发和有限孔径的地表接收下的叠前地震数据不支持我们得到这样的期望成像目标。实际数据情况下，适用于宽带波阻抗建模的反射系数获取方法应该是：① 做好地表一致性处理；② 做好保真的偏移成像（至少要消除球面扩散和吸收衰减）得到最佳的 I_{mig}（保真的偏移成像结果）；③ 用 $H * I_{\text{True}} = I_{\text{mig}}$ 定义一个其它信息约束（主要是三维空间中的构造约束）下的图像反褶积问题（其中 H 是 Hessian 矩阵），求解该反问题得到更宽带的 I_{true}（频带拓宽后的保真的偏移成像结果）；④ 井与震结合标定 I_{true} 的幅值大小。至此，我们可认为得到了定量的宽带反射系数估计结果。

目前，油气地震勘探地震地质解释的主要信息来源就是带限反射系数成像结果。但是，很显然它（$R=\dfrac{I_2-I_1}{I_2+I_1}$）仅仅反映界面处的岩性变化。深度域全三维空间中的大量的岩性变化信息并不包含在带限反射系数成像结果中，而油气地质关心的是深度域全三维空间中介质岩性的变化。这也是为什么一定要追求深度域全三维空间中岩性变化信息完整刻画的地震波成像结果的根源。众所周知，宽波数带的纵波速度 $v_P(x,y,z)$、横波速度 $v_S(x,y,z)$、密度 $\rho(x,y,z)$，波阻抗 $I(x,y,z)$，泊松比 $\sigma(x,y,z)$ 都可以从不同的角度描述深度域全三维空间中的岩性变化，它们之间并非独立的、而是存在紧密关联的。目前，我们的成像目标是：宽波数带的波阻抗 $I(x,y,z)$。事实上，宽波数带的纵波速度 $v_P(x,y,z)$ 也是很合适的选择。我们不是把它提成 Bayes 估计理论下 FWI 方法技术的成像目标，而是在已知背景阻抗和宽带反射系数（或高波数参数扰动量）估计结果的基础上，提一个在其他先验信息约束下的信息融合反问题，进一步融入更多的有地质意义的先验信息（譬如地震相），把背景阻抗和宽带反射系数融合成宽带波阻抗。这样做避开了 Bayes 估计理论下 FWI 方法技术估计宽带波阻抗的强非线性性，得到的结果更稳定，更符合地质逻辑，而且没有降低背景阻抗和宽带反射系数所包含信息的分辨率。

4 宽带波阻抗建模影响因素分析

定义并求解一个约束优化问题实现信息融合宽带波阻抗建模，目的是把背景阻抗和宽带反射系数在其他信息的约束下，有机地融合在一起，得到低波数带（0-3Hz）、中波数带（4-8Hz）和高波数带（9-50Hz）成分都合理存在的宽带波阻抗模型。下面我们通过数值实验在 Marmousi 模型上分析背景阻抗的精度和宽带反射系数的精度对宽带波阻抗建模结果的影响。目的是依据分析结果得到提高宽带波阻抗建模精度的合理做法。

4.1 背景阻抗的生成

用图 3 所示的不同平滑半径的高斯平滑算子施加在 Marmousi 模型上，得到不同平滑程度的背景速度。用描述速度密度之间关系的 Gardner 关系式产生密度场，从而生成平滑程度的背景阻抗。图 4 展示了平滑半径分别为 40m、120m、240m、360m 的背景阻抗。

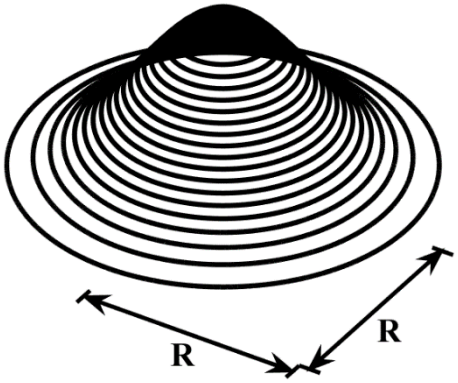
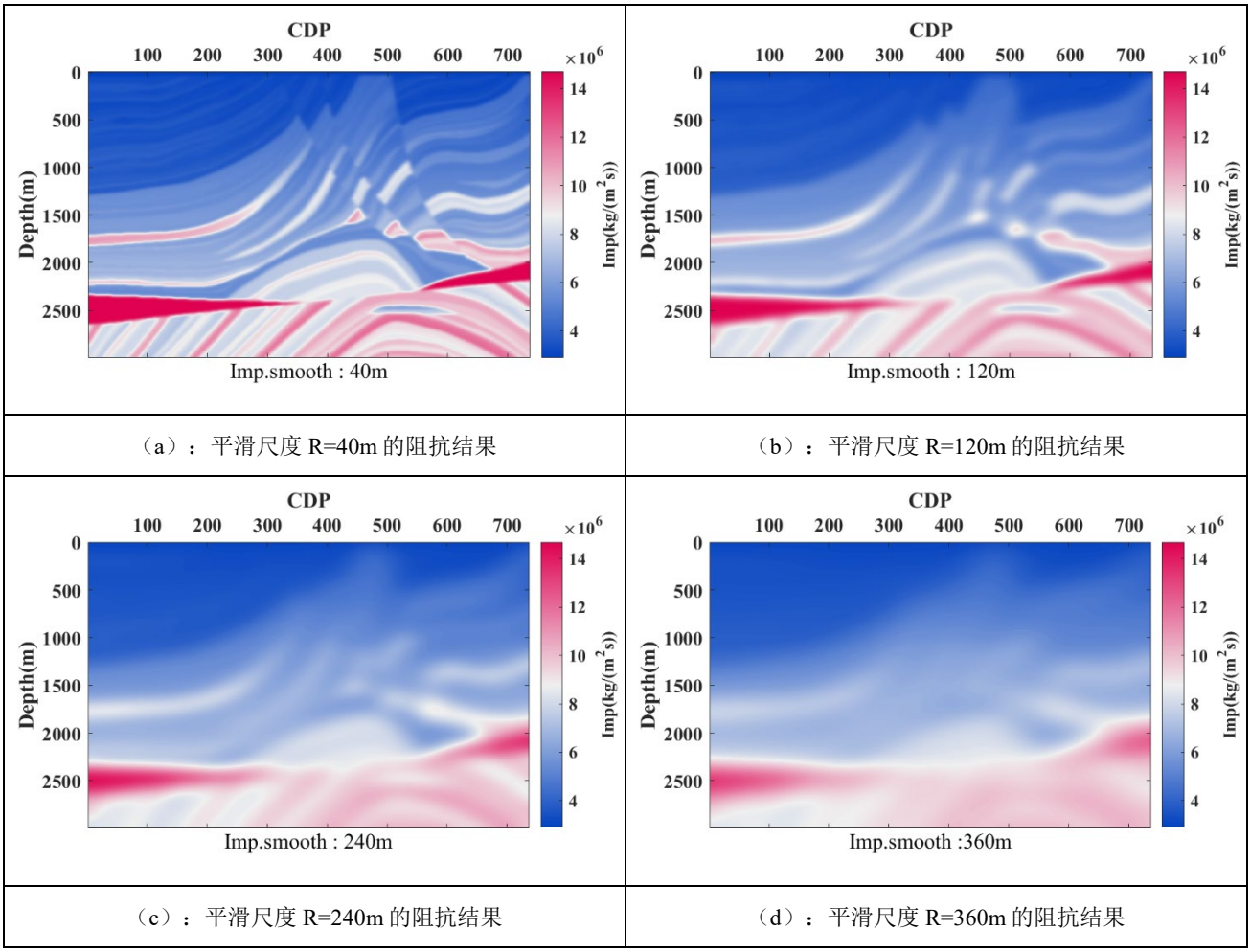


图 3 半径为 R 的 Gauss 平滑算子示意图。



（a）：平滑尺度 R=40m 的阻抗结果

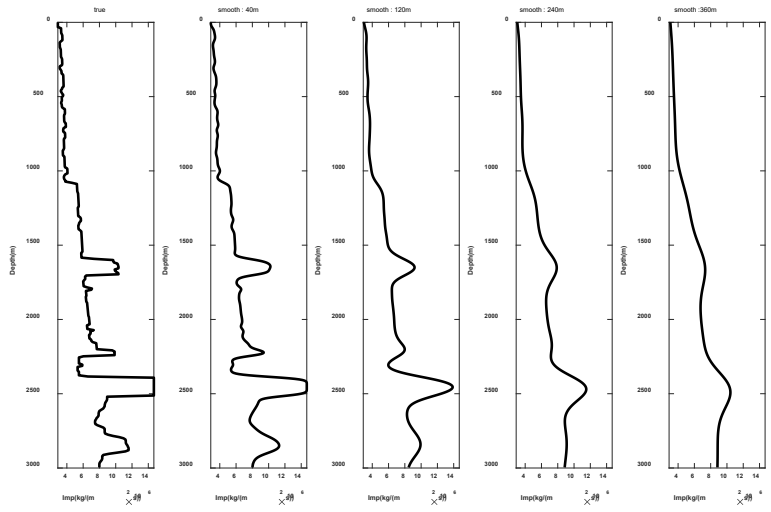
（b）：平滑尺度 R=120m 的阻抗结果

（c）：平滑尺度 R=240m 的阻抗结果

（d）：平滑尺度 R=360m 的阻抗结果

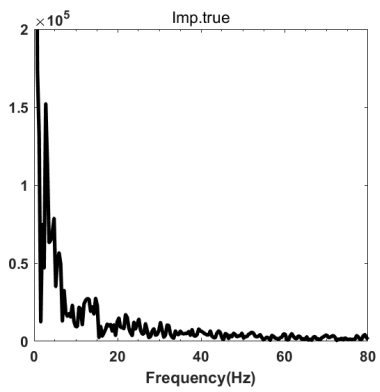
图 4 不同平滑半径的背景阻抗。

为了更清楚地看出平滑算子的影响，从图 4 中各子图抽出第 200 道（模型中间一道）进行展示，如图 5a，图 5b-e 分别是 40 米、120 米、240 米和 360 米平滑后的结果。。

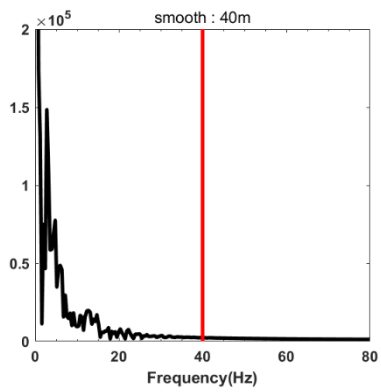


a b c d e

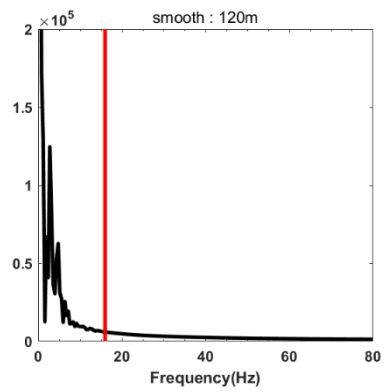
图 5 平滑后的背景阻抗抽线对比(a 是第 200 道的真值，b-e 分别是 40 米、120 米、240 米和 360 米平滑后的结果)。



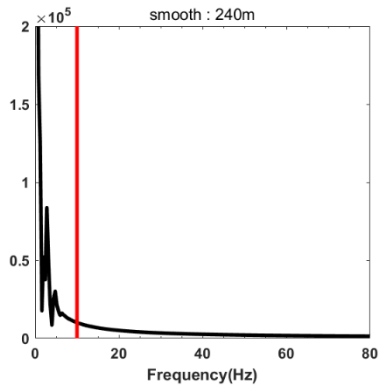
（a）真实模型对应的频谱（波数谱）



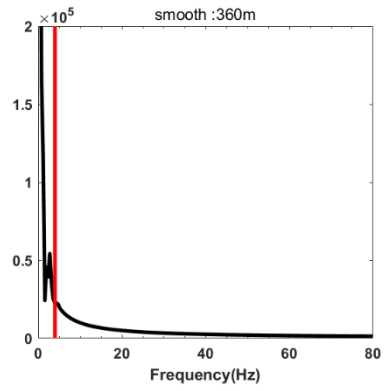
（b）平滑尺度 R=40m 的阻抗结果对应的频谱（波数谱）



（b）平滑尺度 R=120m 的阻抗结果对应的频谱（波数谱）



（b）平滑尺度 R=240m 的阻抗结果对应的频谱（波数谱）



（b）平滑尺度 R=360m 的阻抗结果对应的频谱（波数谱）

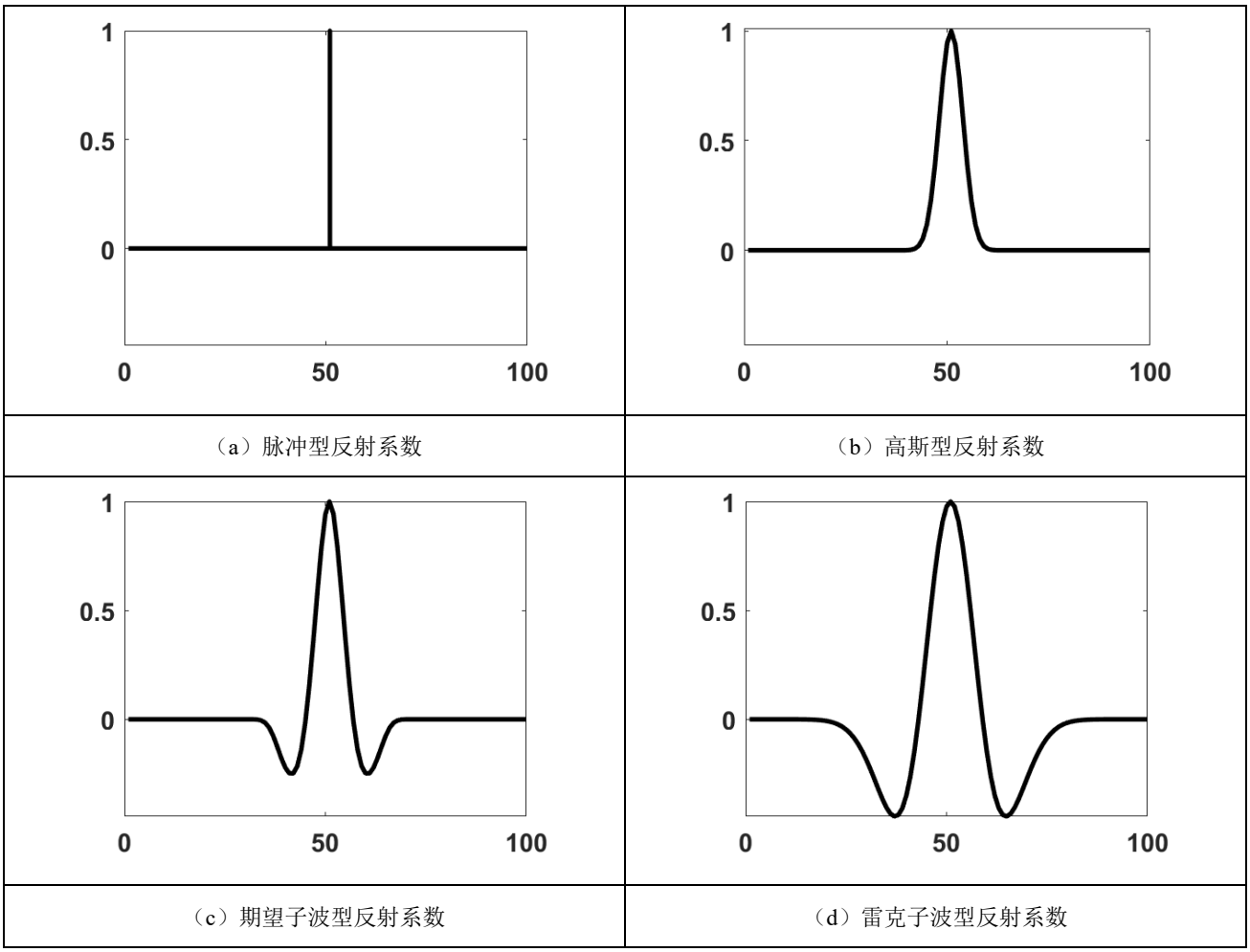
图 6 图 5 中各曲线对应的振幅谱。

由于频率广为大家熟知，通过关系式$\frac{f}{df*nf}=\frac{k}{dk*nk}$把深度域模型 Fourier 变换对应的波数折算成频率。其中，df,nf 分别代表频率域采样间隔以及采样点数；dk,nk 代表波数域采样间隔以及采样点数。图 5 中各曲线对应的振幅谱如图 6 所示。

从图 6 可以看出，随着平滑尺度的增大，高波数的成分逐渐降低。平滑半径 R=40m 时，波阻抗中 40Hz 之上的频率完全衰减；平滑半径 R=120m 时，16Hz 之上的频率完全衰减；平滑半径 R=240m 时 10Hz 之上的频率完全衰减；平滑半径 R=360m 时 4Hz 之上的频率基本完全衰减。

4.2 宽带波阻抗建模需要什么样的反射系数成像结果？

即便最高精度的叠前深度偏移成像，由于受实测地震数据中激发接收的地震子波频带有限和仅在地表观测的观测方式的孔径有限，得到的带限反射系数必然是（我们 WPI 称谓的）雷克（Ricker）子波型反射系数。雷克（Ricker）子波型反射系数的典型特征是：主瓣比较宽、旁瓣比较深。这主要是缺低中频和频带窄导致的。我们认为各种地震子波的分辨率取决于完整的振幅谱，相位谱的影响是次要的，低中频优势成分的存在是让子波旁瓣变浅，从而提高弱层分辨能力的重要信息源。由于雷克（Ricker）子波型反射系数主瓣比较宽、旁瓣比较深，它不是宽带波阻抗建模所期望的反射系数。



（a）脉冲型反射系数　（b）高斯型反射系数

（c）期望子波型反射系数　（d）雷克子波型反射系数

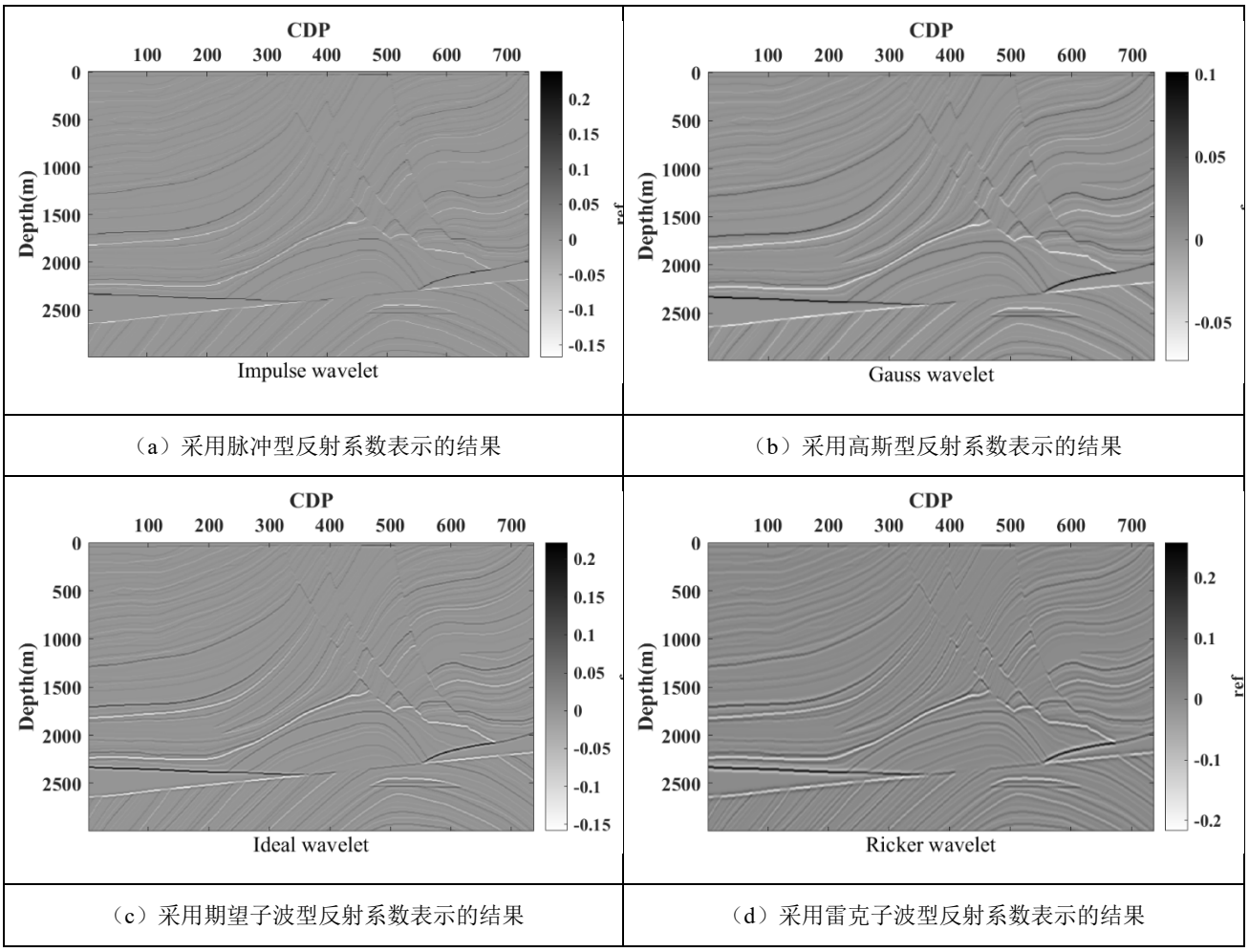
图 7 四种类型的反射系数示意图。

宽带波阻抗建模需要什么样的带限反射系数呢？

为了回答这个问题，我们测试了四种类型的反射系数。四种类型的反射系数见图 7。首先是脉冲型反

射系数，这是我们最想要得到的反射系数，它是全频带的。其次是（我们 WPI 称谓的）高斯型反射系数，高斯型反射系数相对于脉冲型反射系数主要是衰减了高频信息，但低中频信息丰富。然后是（我们 WPI 称谓的）期望子波型反射系数，它具有窄主瓣、低旁瓣的特征。最后是雷克（Ricker）子波型反射系数。这是宽带波阻抗建模中最不想要的。雷克（Ricker）子波型反射系数的低频和高频都大幅度衰减，并且旁瓣的能量比较强，会引入较多的假的波阻抗高频信息。需要说明，期望子波是我们（WPI）根据地质分辨率要求，用子波的主旁瓣宽度比和高度比为参数定义的高分辨子波。

图 8 展示了上述四种反射系数表示的 Marmousi 模型。相当于我们假定用叠前深度偏移成像方法得到了图 8 展示的偏移成像剖面。我们当然知道，脉冲型反射系数和高斯型反射系数表示的 Marmousi 模型成像剖面是理想的、期望的。

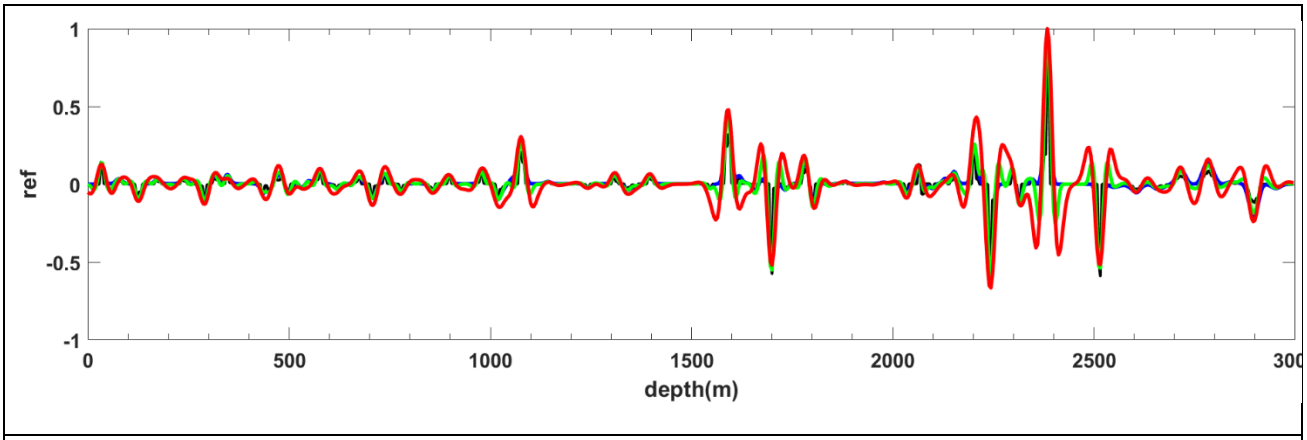


（a）采用脉冲型反射系数表示的结果　　（b）采用高斯型反射系数表示的结果

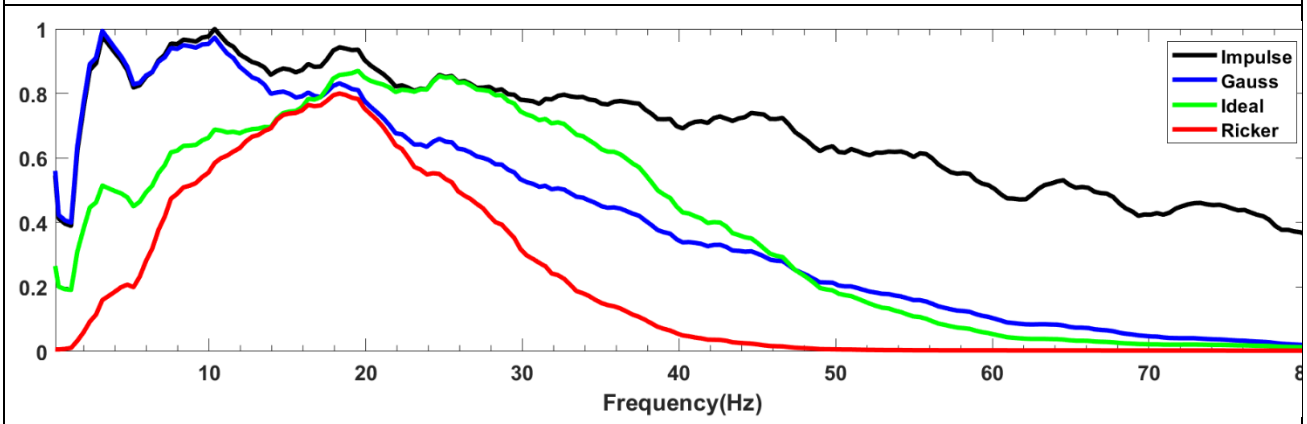
（c）采用期望子波型反射系数表示的结果　　（d）采用雷克子波型反射系数表示的结果

图 8 四种反射系数表示的 Marmousi 模型。

首先对四种反射系数剖面中 CDP200 道进行抽线，并进行频谱分析，结果展示在图 9 中。可以明显看出，脉冲型反射系数的频带展布较广，全频带均有能量；高斯型反射系数在高频端的能量明显衰减，但低中频成分有很好的保留；期望子波型反射系数在低频和高频端均有一定程度的衰减，但有效频带比较宽；雷克子波型反射系数的频带最窄，既缺低频也缺高频。



（a）宽带反射系数结果



（b）宽带反射系数频谱（波数谱）

图 9 宽带反射系数抽线及频谱分析，黑线：脉冲型（Impulse）反射系数；蓝线：Gauss 型反射系数；绿线：期望子波型(Ideal(反射系数；红线：雷克子波型(Ricker)反射系数。

使用这四种反射系数剖面分别与平滑半径为 R=240m 的背景阻抗进行融合，得到的结果如图 10 所示。此处，我们把信息融合宽带波阻抗问题提成一个约束非线性优化问题，具体提法如下：

$$\min\|R\|_1, s.t.\left\|W*R-r_{\text{mig}}\right\|^2\le\varepsilon_1;\left\|R-\nabla_z X\right\|^2\le\delta_1 \qquad (11a)$$

$$\min\|X\|_{TV}, s.t.\left\|L*X-X_b\right\|^2\le\varepsilon_2;\left\|R-\nabla_z X\right\|^2\le\delta_2 \qquad (11b)$$

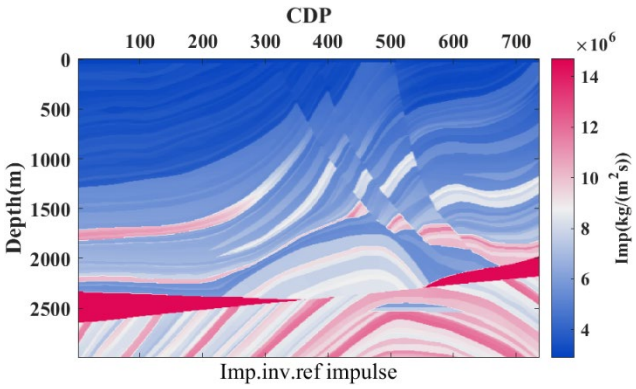
其中，R 是宽带反射系数；r_{mig} 是偏移成像及各种后处理后得到的保真的宽带反射系数；W 是类似地震子波的滤波器；X（$=\frac{1}{2}\ln \boldsymbol{I}$）是宽带波阻抗 $\boldsymbol{I}$ 的半对数表示；X_b 是背景阻抗的半对数表示；L 是一个低通滤波器；$\|\bullet\|_1$ 和 $\|\bullet\|^2$ 分别代表 L_1 和 L_2 范数；$\|\bullet\|_{TV}$ 代表总变差（TV）范数；∇_z 代表沿深度（z）方向的导数。迭代串联求解（11）式得到后续展示的信息融合宽带阻抗结果。约束非线性优化问题的具体求解是一个相当复杂的议题。（11）式的具体解法不是此文的重点，为简单计，略去关于具体解法的冗长讨论。

图 10a 展示不同宽带反射系数与平滑半径 240m 作用下的背景阻抗融合得到的阻抗结果；图 10b 展示从 CDP200 处抽取的各种融合阻抗与对应的背景阻抗和真阻抗的对比。从图中可以看出，脉冲型反射系数与背景阻抗的融合结果趋于真解，与真波阻抗的相关度达到 99.83%；高斯型反射系数由于高频部分受损，融合阻抗的分辨率降低，但与真波阻抗的相关度也能达到 99.47%；期望子波尽管是宽带子波，但由于中低频

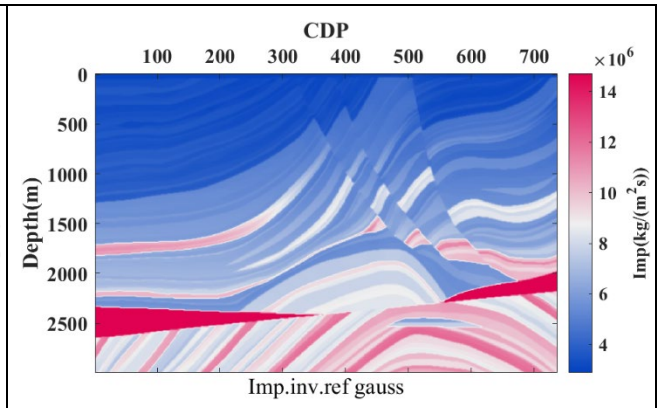
成分的缺失，旁瓣加深，融合结果中的低中频成分难以补全，也已经引入了假的高波数成分，与真波阻抗的相关度只能达到 96.36%。雷克子波型反射系数融合效果最差，融合后波阻抗与真阻抗的相关度只有90.20%，从 10b 中明显可以看出它们之间的差异。

从这个融合过程可以清楚地看出：油气地震勘探中宽带成像子波对于介质岩性刻画的重要性，宽带成像子波的获取最重要的是既要有宽带的激发接收子波，还要有宽方位长偏移距的观测。尤其低中频成分的丢失，成像结果对地下介质岩性描述的误差会很大。高密度和高信噪比的观测对于岩性成像的贡献是从其它角度体现的。

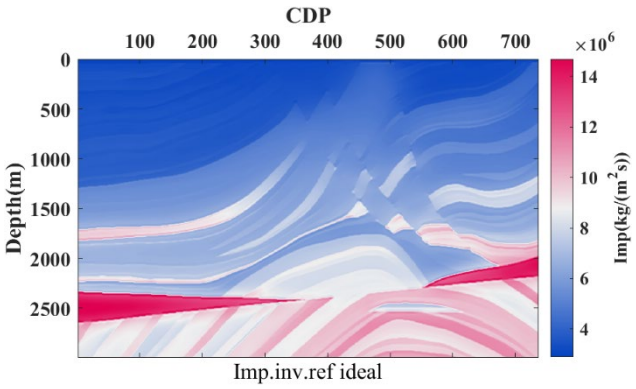
接下来我们再分析一下，背景阻抗建模结果和宽带反射系数成像结果对宽带波阻抗建模结果的不同波数带成分的贡献。图 11 展示了用不同类型宽带反射系数剖面与背景阻抗（R=240m）融合后背景阻抗、融合阻抗和真阻抗的振幅谱对比。



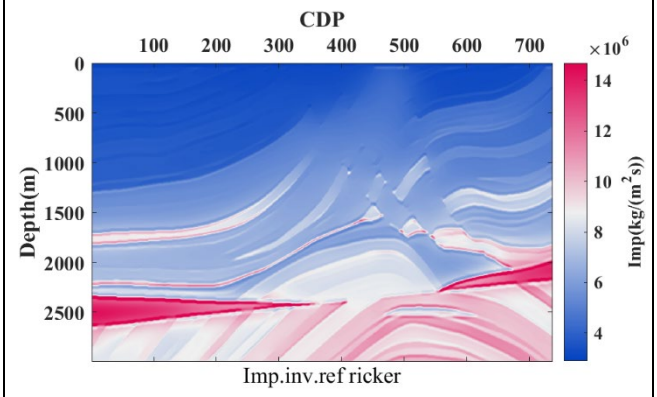
（a）采用脉冲型反射系数融合的波阻抗结果



（b）采用高斯型反射系数融合的波阻抗结果



（c）采用期望子波型反射系数融合的波阻抗结果



（d）采用雷克子波型反射系数融合的波阻抗结果

图 10a 不同宽带反射系数与平滑 240m 的背景阻抗融合得到的阻抗结果。

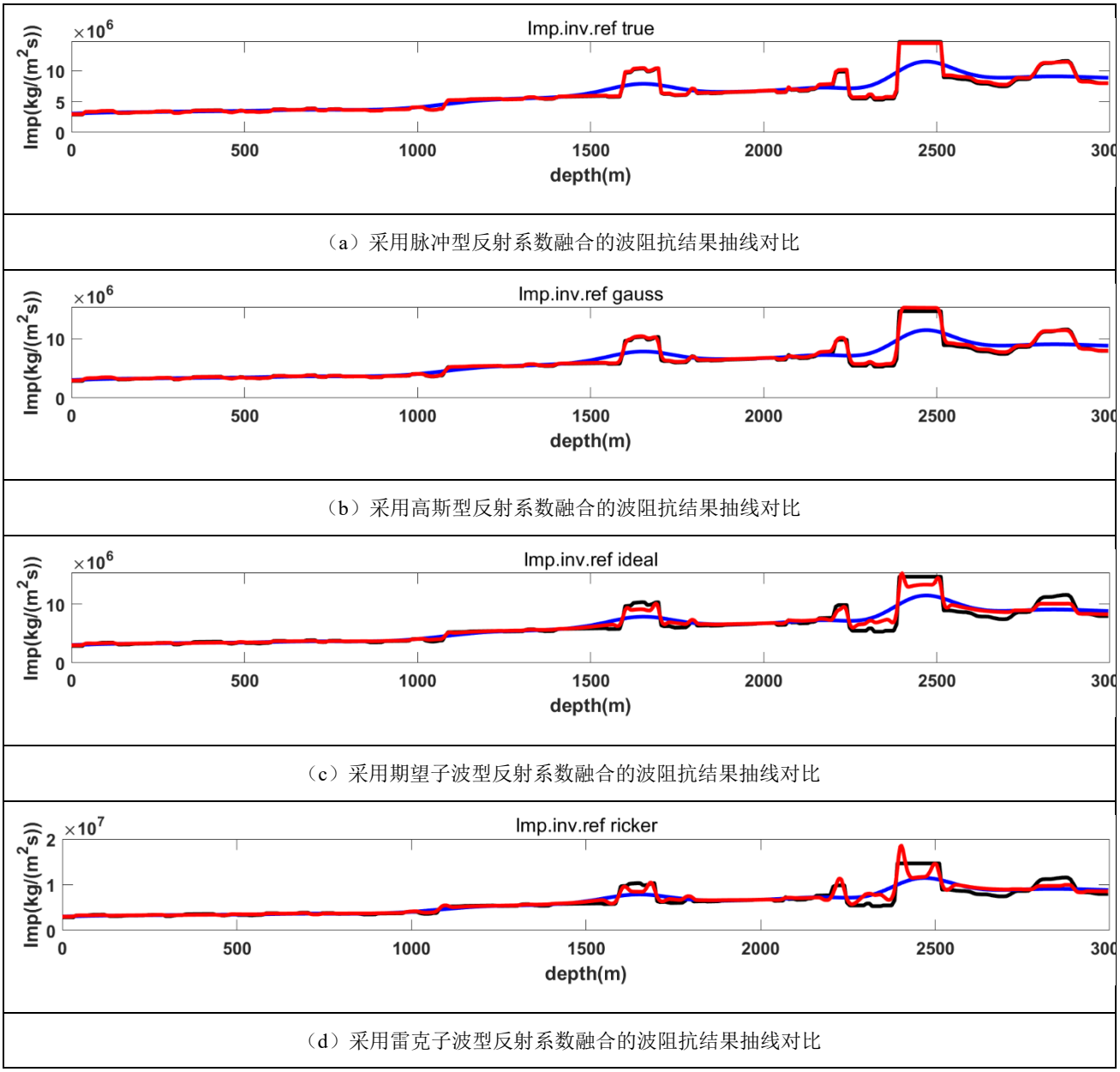
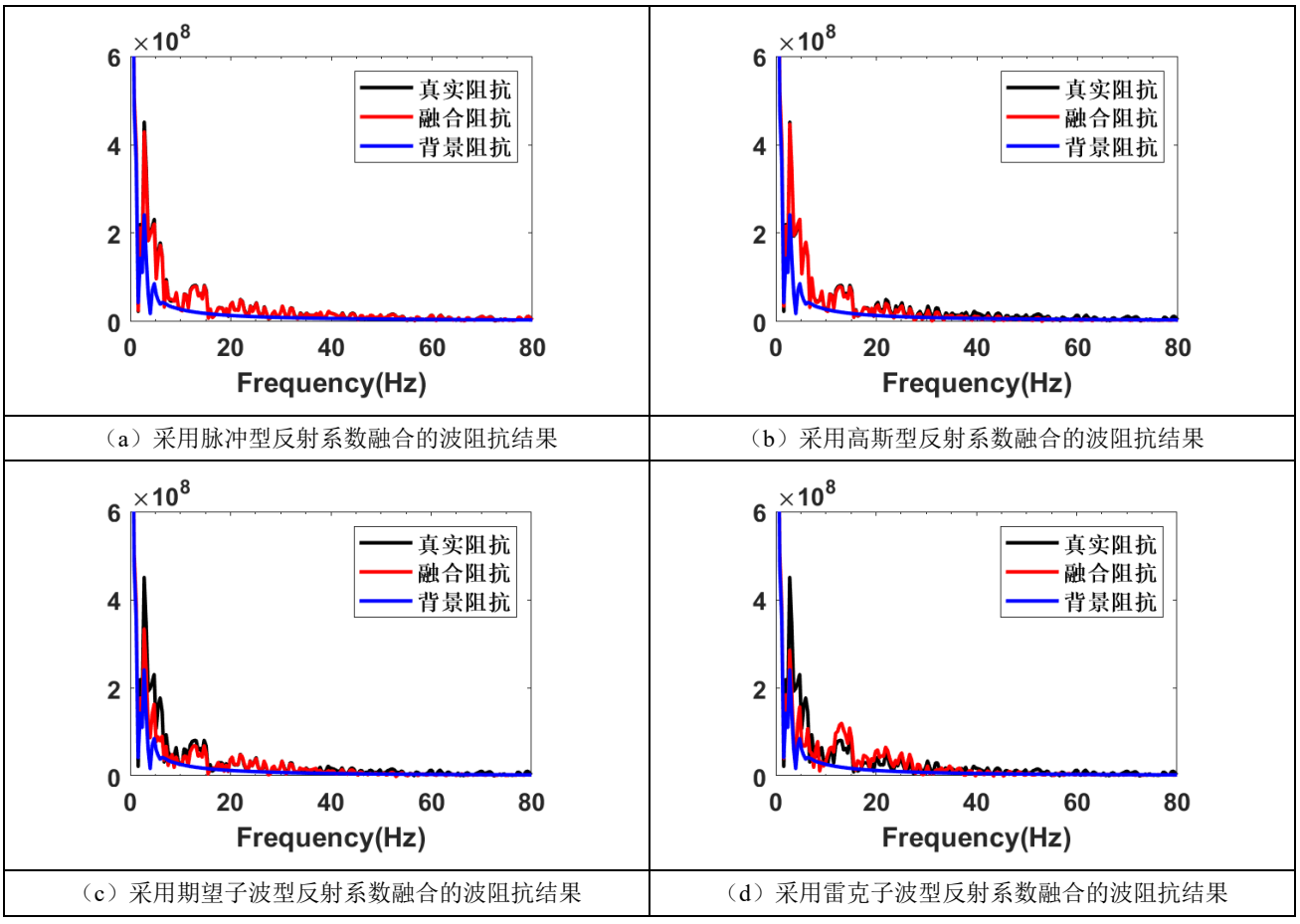


图 10b 不同宽带反射系数融合结果抽线（CDP200）对比。黑线：真实阻抗；红线：融合阻抗；蓝线：背景阻抗。

从图 11 可以清楚地看出：融合阻抗的背景成分都是由背景阻抗贡献出来的，其它部分是由反射系数成像结果贡献出来的。这是很自然的逻辑结果。但是，中波数带（譬如 4-8Hz）也可以由反射系数成像补充出来，只要反射系数成像结果中包含有波阻抗中波数段的信息。 事实上，脉冲型反射系数中包含了除零波数成分外所有波数成分的波阻抗变化信息。因此，道积分方法重建波阻抗时，对于脉冲型反射系数，理论上只要提供任意一层的正确波阻抗就可以得到真波阻抗的恢复。当然这又是一个反演骗局，实际上道积分方法受到的影响很多，根本不可能得到实际可用的波阻抗建模结果[14]。高斯型反射系数成像结果除了超高频部分不能正确贡献外，其他部分的波阻抗贡献结果都是可靠的。应该注意这个结论与高斯函数的方差是由密切联系的。较大方差下的高斯函数也会滤除过多的低中频和超高频成分，从而降低波阻抗建模的精度。理想子波型反射系数成像结果对宽带阻抗的贡献，误差较大的是中波数段。雷克子波型反射系数反射系数对宽带阻抗的贡献，误差较大的既包括中波数段，也包括高波数带，高波数带中明显地引入了假的成分（见图 10b）。

上述融合过程再次告诉我们：地震波成像的真正目标既包括精确的背景速度（背景阻抗），也包括最好是 Gauss 型反射系数呈现出的成像结果（追求脉冲型反射系数成像结果过于理想了）。即便如果得到了

Gauss 型反射系数呈现出的成像结果，成像分辨率也会有本质性提升。但是，Gauss 型反射系数成像结果也是很难得到的，叠前地震数据提供的信息不足以得到这样的成像结果。无论如何，我们知道了地震波叠前偏移成像的目标不是带限的反射系数，尤其不是具有宽主瓣、深（宽）旁瓣特征的雷克（Ricker）子波型的反射系数，期望的偏移成像结果应该是方差较小的 Gauss 型反射系数。常规的偏移成像无论如何是不可能得到 Gauss 型反射系数成像结果的，但是约束反演成像方法是有可能得到实用的 Gauss 型反射系数成像结果的，不过要引入其他的先验信息。



（a）采用脉冲型反射系数融合的波阻抗结果

（b）采用高斯型反射系数融合的波阻抗结果

（c）采用期望子波型反射系数融合的波阻抗结果

（d）采用雷克子波型反射系数融合的波阻抗结果

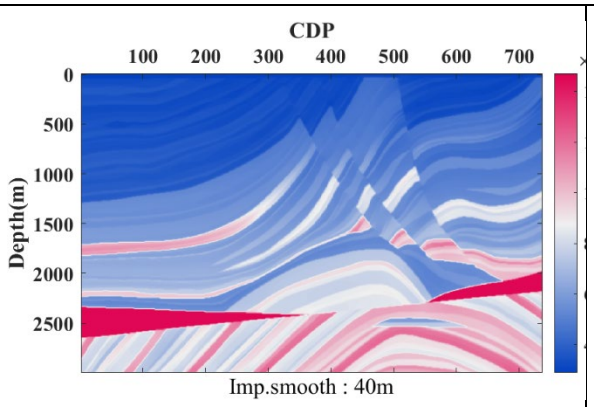
图 11 背景阻抗、融合阻抗和真阻抗的振幅谱。a：对应脉冲型反射系数，b：对应高斯型反射系数；c：对应理想子波型反射系数，d：对应雷克子波型反射系数。黑线真实阻抗；蓝线背景阻抗；红线融合阻抗。

4.3 背景阻抗对宽带波阻抗建模的影响

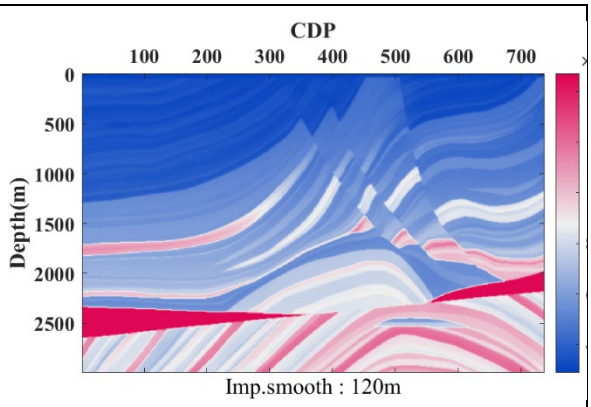
为了分析不同平滑半径获得的背景阻抗（实质上是不同精度的背景阻抗）对于融合阻抗的精度影响，把图 4 中四种背景阻抗与高斯型的宽带反射系数（见图 8）进行融合，得到图 12 展示的融合结果。

从图 12 可以看出，四种不同平滑半径得到的背景阻抗剖面与 Gauss 型反射系数剖面的融合结果，精度都相当高。说明当地震成像剖面表现为 Gauss 型反射系数时，其中包含了波阻抗中的丰富的低中波数信息，即使用很平滑的背景阻抗，也能把宽波数带的波阻抗重构出来。各个波数带中没有显著的成分缺失，重构精度很高。

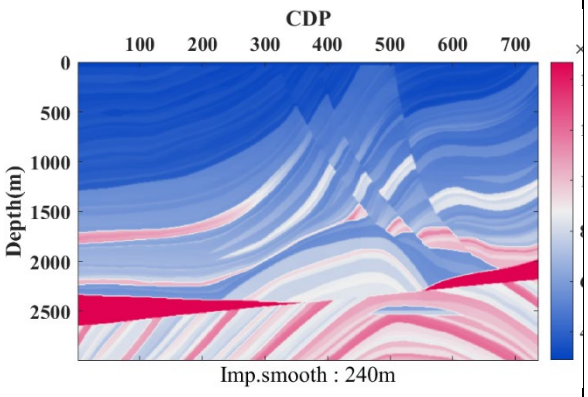
为了看到不同背景阻抗剖面与不同类型反射系数剖面融合结果之间的差异，又做了上述四种不同平滑半径得到的背景阻抗剖面与期望子波型反射系数剖面（见图 8）的融合，结果如图 13 所示。图 13c 还对比了 CDP200 处各种融合阻抗、背景阻抗与真阻抗的振幅谱对比。



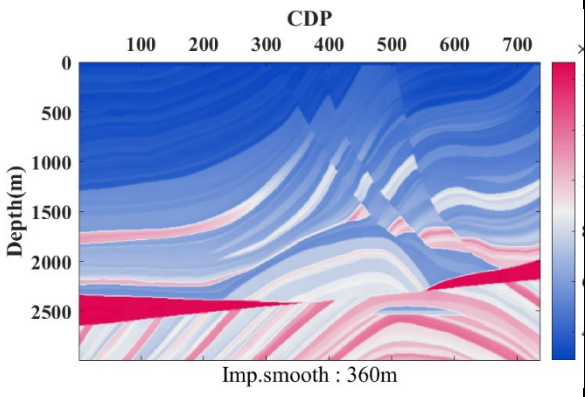
（a）：采用平滑尺度 R=40m 的背景阻抗融合结果



（b）：采用平滑尺度 R=120m 的背景阻抗融合结果

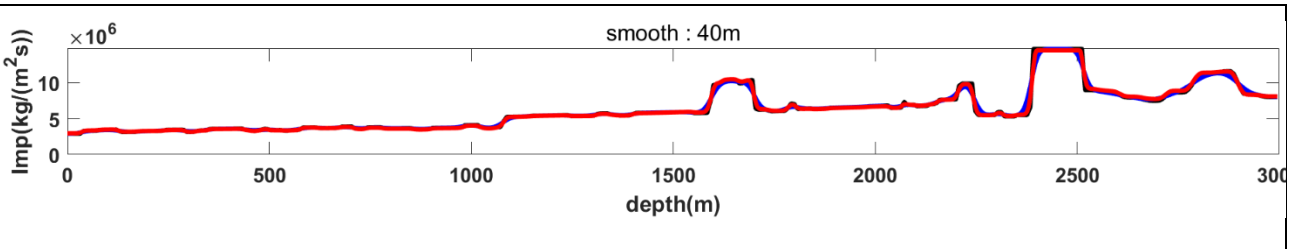


（c）：采用平滑尺度 R=240m 的背景阻抗融合结果

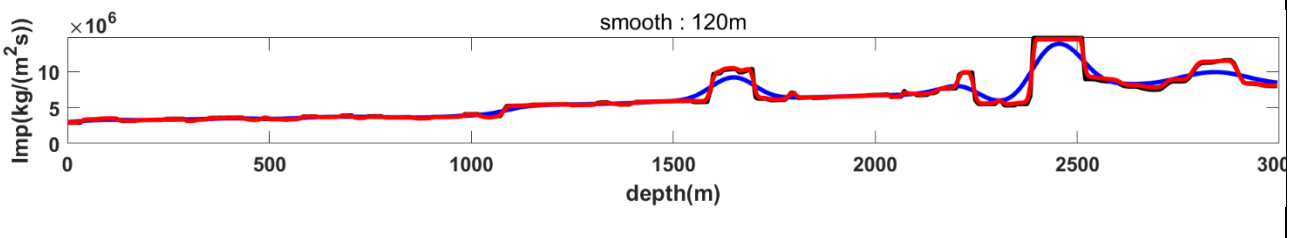


（d）：采用平滑尺度 R=360m 的背景阻抗融合结果

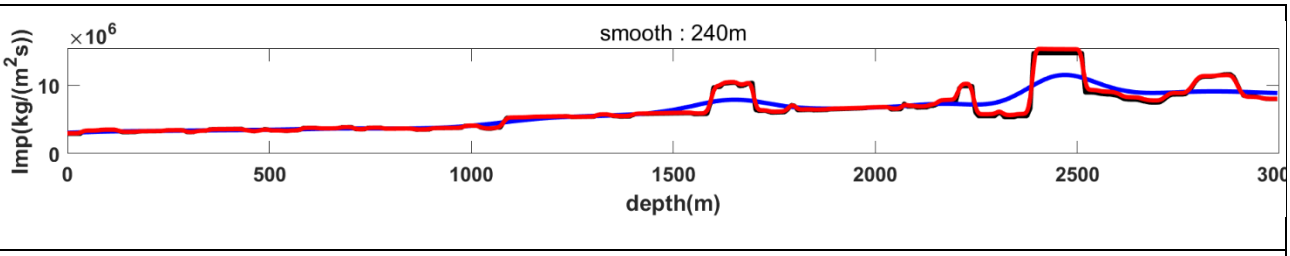
图 12a 不同平滑尺度的背景阻抗与高斯型反射系数融合得到的融合阻抗。



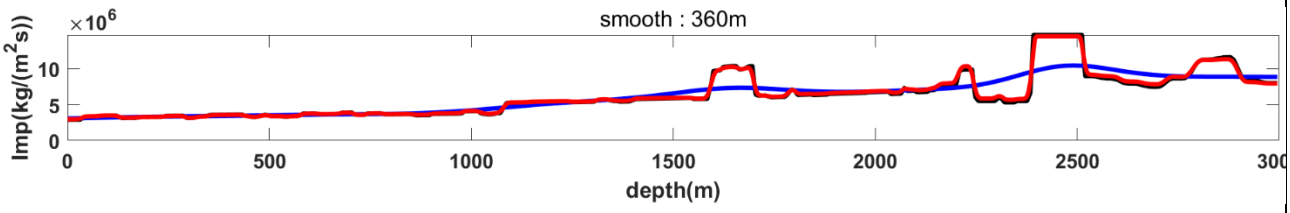
（a）采用平滑尺度 R=40m 的背景阻抗融合结果抽线对比



（b）采用平滑尺度 R=120m 的背景阻抗融合结果抽线对比

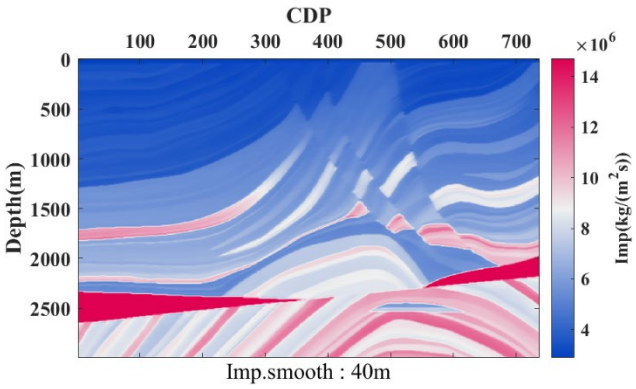


（c）采用平滑尺度 R=240m 的背景阻抗融合结果抽线对比

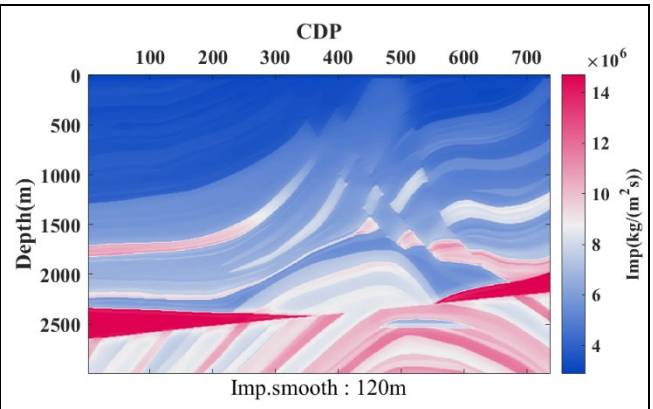


（d）采用平滑尺度 R=360m 的背景阻抗融合结果抽线对比

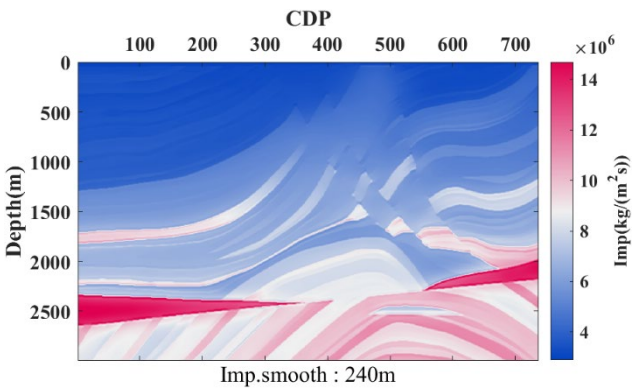
图 12b CDP200 处的抽线对比，黑线真实阻抗，蓝线背景阻抗，红线融合阻抗。



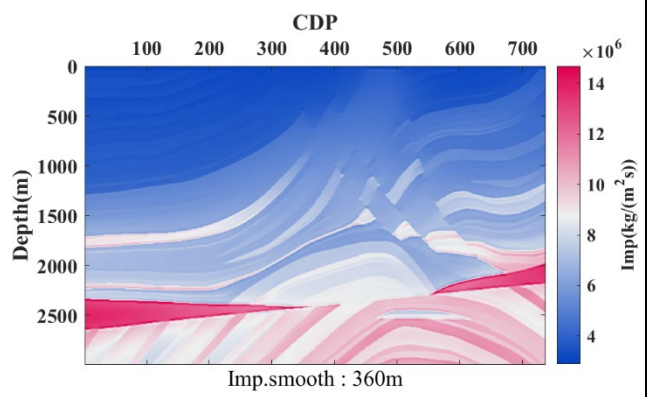
（a）：采用平滑尺度 R=40m 的背景阻抗融合结果



（b）：采用平滑尺度 R=120m 的背景阻抗融合结果

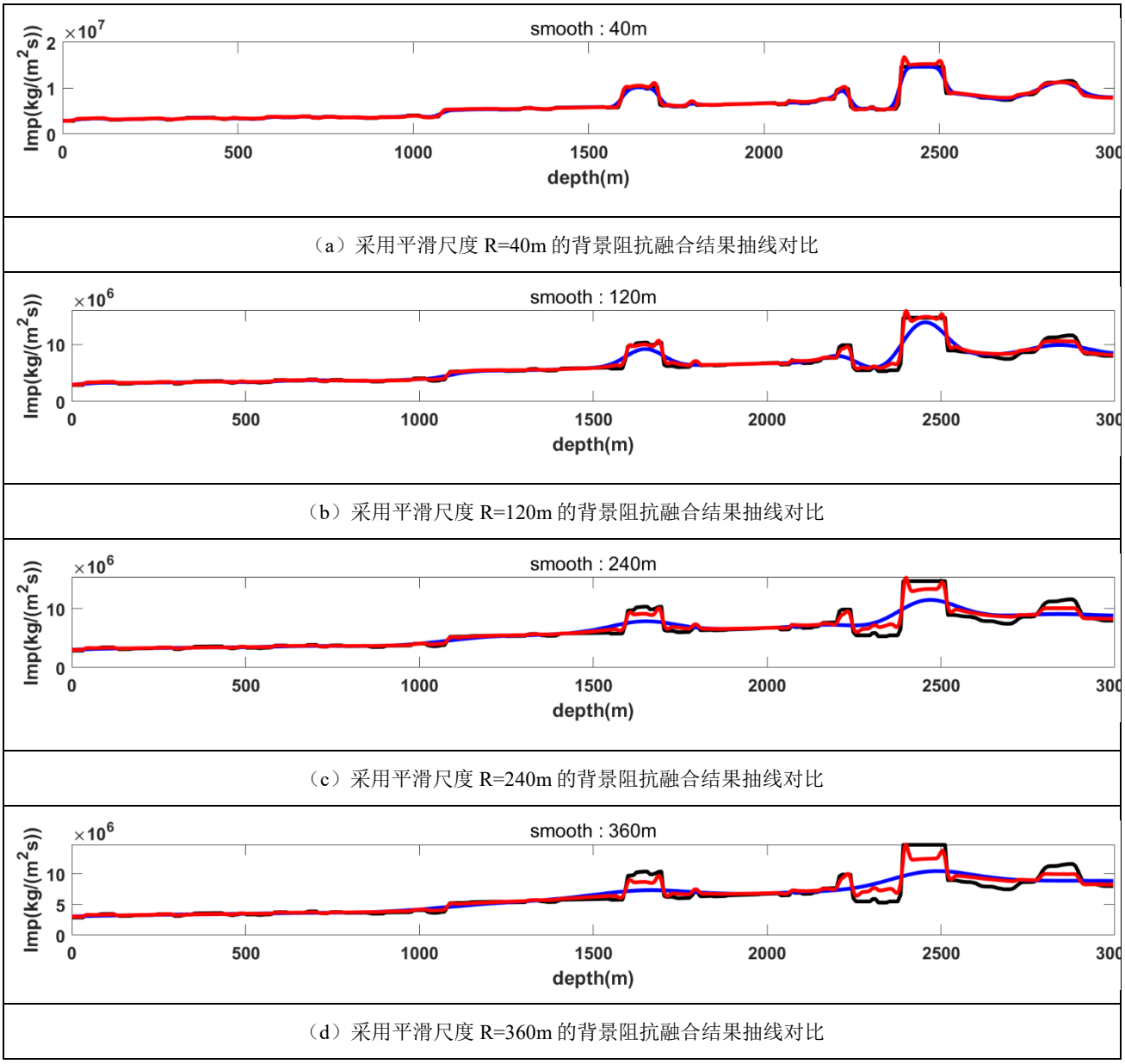


（c）：采用平滑尺度 R=240m 的背景阻抗融合结果



（d）：采用平滑尺度 R=360m 的背景阻抗融合结果

图 13a 不同平滑尺度的背景阻抗与期望子波型反射系数进行融合得到的阻抗。



（a）采用平滑尺度 R=40m 的背景阻抗融合结果抽线对比

（b）采用平滑尺度 R=120m 的背景阻抗融合结果抽线对比

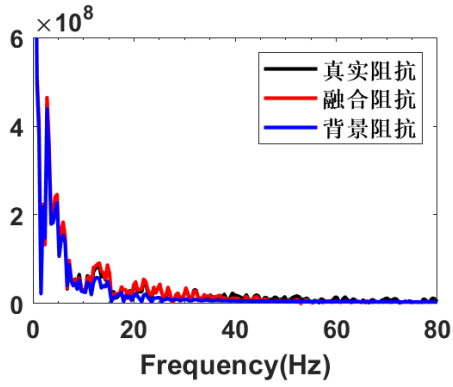
（c）采用平滑尺度 R=240m 的背景阻抗融合结果抽线对比

（d）采用平滑尺度 R=360m 的背景阻抗融合结果抽线对比

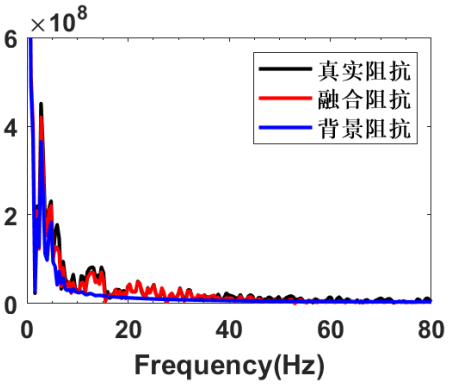
图 13b CDP200 处的抽线对比，黑线真实阻抗，蓝线背景阻抗，红线融合阻抗。

图 13 给出的最主要的结论是：平滑半径加大导致了背景波阻抗中中波数成分损失，理想子波型反射系数剖面也缺乏中波数成分，从而导致中波数阻抗不能被重构，这在图 13b 和图 13c 中可以清楚地看出来。中波数段成分的损失，使得重构的波阻抗与真波阻抗出现明显差异，重构的波阻抗不能真实反映地下介质岩性的变化。很显然，拓宽成像子波的频带，提高背景阻抗的建模精度都是补充中波数阻抗成分的措施。

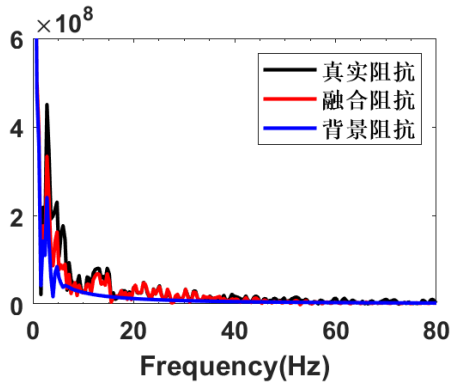
前述分析时，背景阻抗成分是正确的。但是对于实际情况，背景阻抗也会存在一定的误差。因此我们又分析了背景阻抗错误情况下的信息融合宽带波阻抗重构结果。背景阻抗误差引入方式：真实阻抗进行 R=240m 的 Gauss 平滑后，整体减小 15%；整体增加 15%；加入 15%量级的随机扰动。用高斯型反射系数成像结果与它们分别进行信息融合重构，重构结果见图 14。



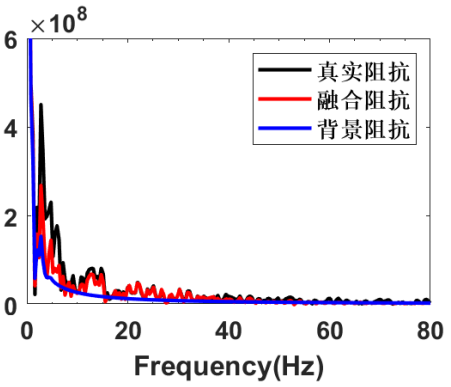
（a）：采用平滑尺度 R=40m 的背景阻抗融合结果



（b）：采用平滑尺度 R=120m 的背景阻抗融合结果

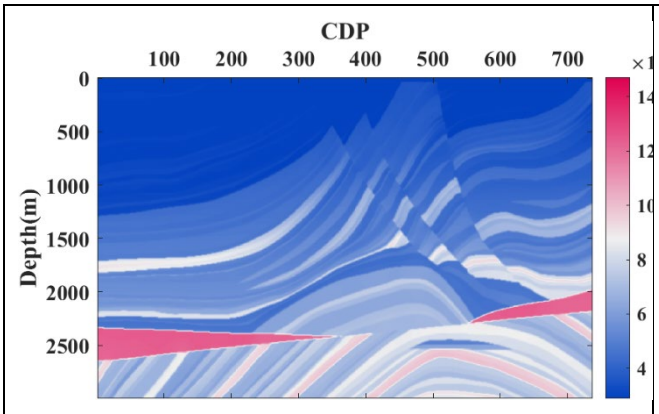


（c）：采用平滑尺度 R=240m 的背景阻抗融合结果

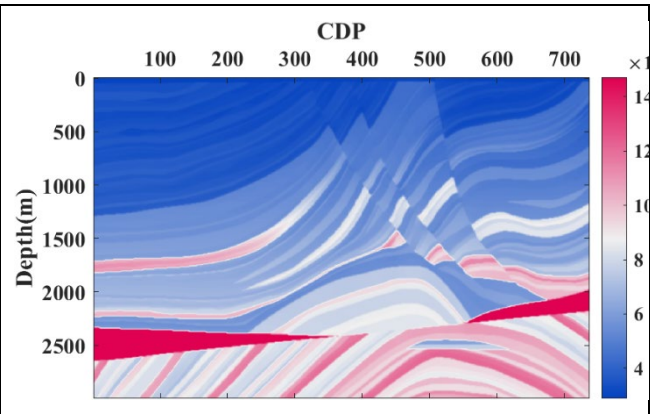


（d）：采用平滑尺度 R=360m 的背景阻抗融合结果

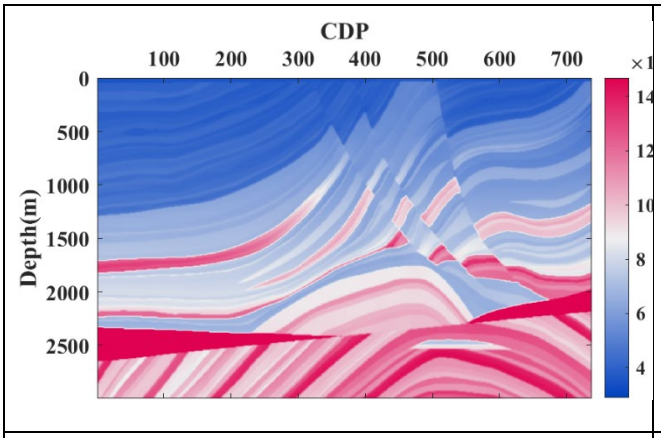
图 13c CDP200 处各种融合阻抗、背景阻抗与真阻抗的振幅谱。a：平滑尺度 R=40m 时；b：平滑尺度 R=120m 时；c：平滑尺度 R=240m 时；d：平滑尺度 R=360m 时的对比情况。黑线真实阻抗；蓝线背景阻抗；红线融合阻抗。



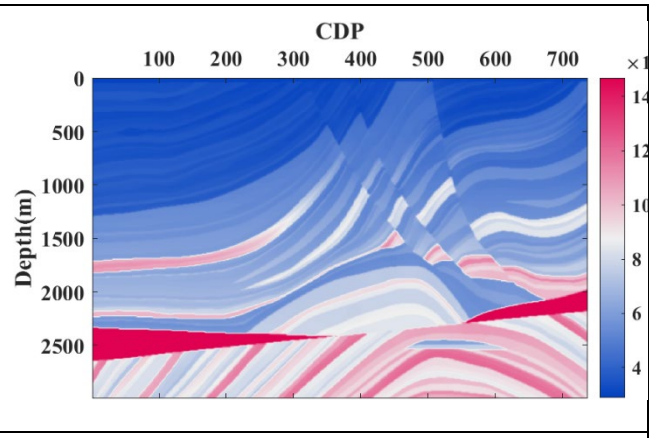
（a）：采用真实波阻抗平滑 240m 后整体减小 15%的背景波阻抗融合的结果



（b）：采用真实波阻抗平滑 240m 后的背景波阻抗融合的结果

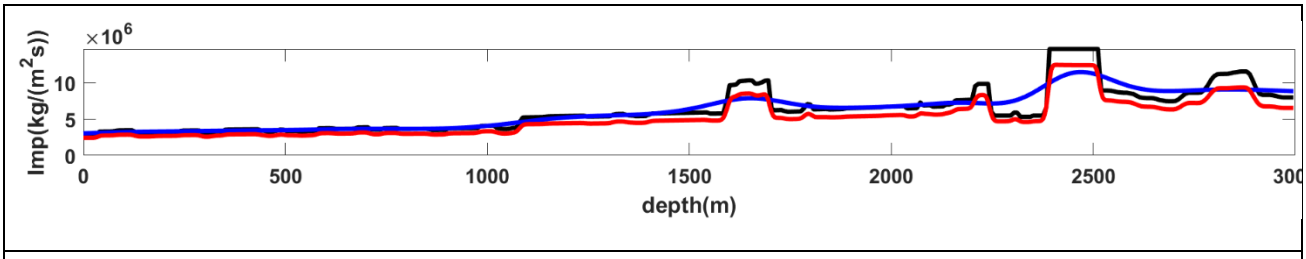


（c）：采用真实波阻抗平滑 240m 后整体增大 15%的背景波阻抗融合的结果

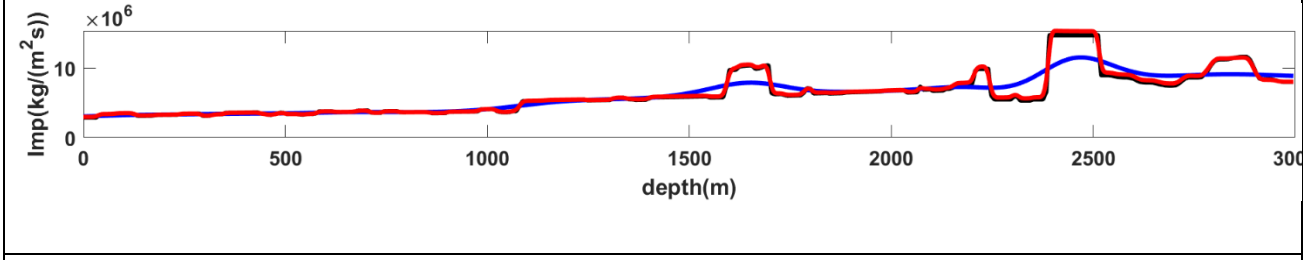


（d）：采用真实波阻抗平滑 240m 后整体加入 15%量级的随机扰动的背景波阻抗融合的结果

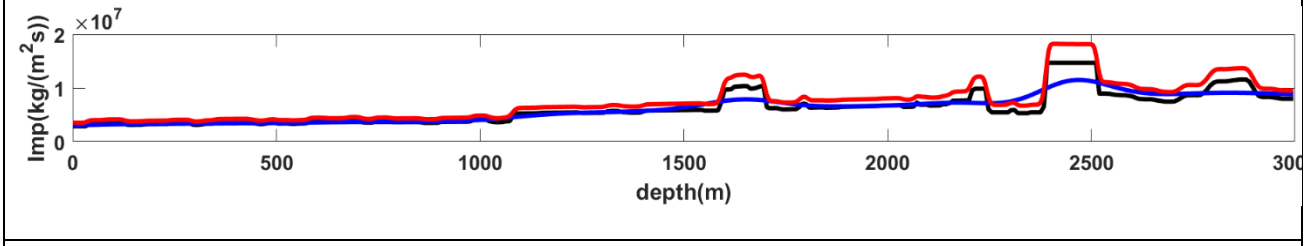
图 14a 错误背景阻抗与高斯型反射系数进行融合得到的阻抗。



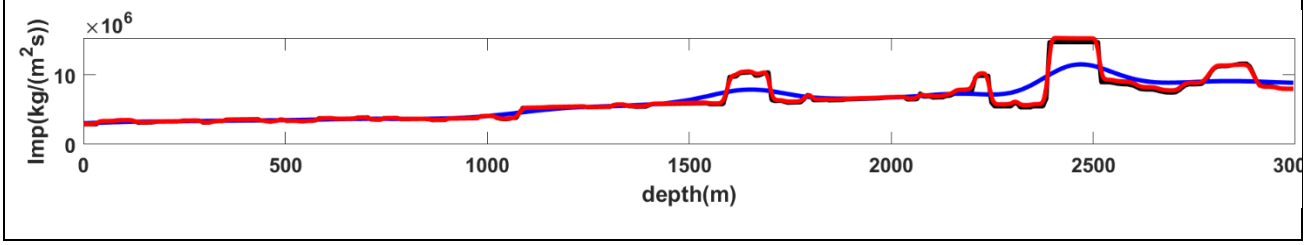
（a）：采用真实波阻抗平滑 240m 后整体减小 15%的背景波阻抗融合的结果抽线对比



（b）：采用真实波阻抗平滑 240m 后的背景波阻抗融合的结果抽线对比



（c）：采用真实波阻抗平滑 240m 后整体增大 15%的背景波阻抗融合的结果抽线对比



（d）：采用真实波阻抗平滑 240m 后整体加入 15%量级的随机扰动的背景波阻抗融合的结果抽线对比

图 14b 错误背景阻抗与高斯型反射系数进行融合得到的阻抗在 CDP200 处的抽线对比。黑线真实阻抗；蓝线背景阻抗；红线融合阻抗。

当背景阻抗为真实阻抗经过半径 240m 的 Gauss 平滑时，重构阻抗与真阻抗匹配度达到 99.27%；但是若此时背景速度由上述背景速度整体减小 15%时，匹配度仅达到 89.28%；若此时背景速度由上述背景速度整体增加 15%时，匹配度仅达到 77.55%；若此时背景速度由上述背景速度加入 15%量级的随机扰动，对重构结果影响不大，匹配度可达 99.26%。庆幸的是，实际情况下建立背景速度模型时，出现系统误差的概率不大。

5 宽带波阻抗建模及定量地震波成像的基本认识

油气地震勘探的真正目的就是通过地震波成像提供的对地下介质弹性参数的准确刻画，结合钻井、测井、岩石物理和油气地质学知识，准确识别与评价油气藏，进行高效益的油气藏开发。

深度域真三维空间中地下介质弹性参数的准确刻画是地震波成像的本质目标，也是油气地震勘探中的核心问题。无论是深度域真三维空间中的宽波数带纵波速度 $v_P(x,y,z)$ 或是宽带波阻抗 $I(x,y,z)$ 都是可以代表地下介质岩性变化的，而带限反射系数仅仅揭示了介质岩性在某些地方（波阻抗跃变的地方）发生了突变，它反映的是介质波阻抗的相对变化量（不是绝对变化量）。很显然，对于准确识别与评价油气藏而言，宽波数带纵波速度 $v_P(x,y,z)$ 或是宽带波阻抗 $I(x,y,z)$ 建模结果是比带限反射系数更合适的基础数据。很长时间以来，地震波成像及后续的地震地质解释之所以总是基于带限反射系数成像结果，首要的原因是带限反射系数的估计（或成像）是一个更稳健的问题，更容易得到可靠的解。以带限反射系数为目标的偏移成像甚至不能称为是一个真正的反问题。偏移成像是油气地震勘探领域独创的方法技术。遗憾的是，理论上它仅仅是一种地下绕（反）射点的空间定位方法技术。

油气地震勘探发展到现在，地震波成像进入定量的、以估计介质弹性参数绝对变化为目标的阶段已不可避免。Bayes 估计理论下的 FWI 技术本质上就是要达到此目标的。实际数据不满足它的理论假设，估计（全）宽波数带的介质弹性参数绝对变化是一个强非线性问题，至少这两个主要原因限制了它的实用化。我们（WPI）提出的特征波成像（CWI）+宽带波阻抗建模（WBIM）也是为了估计（全）宽波数带的介质弹性参数绝对变化。当前的目标是信息融合宽带波阻抗建模（或重构）。用石油工业的术语讲，这是一种针对岩性储层描述的、定量的地震波成像路线与方法技术系列。FWI 的提法侧重于强调算法；针对岩性储层描述的、定量的地震波成像更强调油气勘探的目标和需求。

依据信息融合宽带波阻抗建模（或重构）的思想，**定量地震波成像被分成背景阻抗建模、宽带反射系数估计和信息融合宽带波阻抗建模三个核心环节。**

上一节的分析指出：背景阻抗建模主要利用炮集中透射波（直达波、折射波和 Diving Wave）和反（绕）射波的走时信息，求解线性反演问题（即所谓的层析成像），得到定量的背景速度估计。基于走时信息估计得到的背景速度与介质层速度的背景部分在量级上是一致的，不是特别复杂的情形下，低波数的速度估计结果精度也是很高的。为了给背景阻抗补充进更多的中波数成分，仅仅靠算法的改进（譬如用 FWI 算法）和提高叠前数据的品质（譬如“两宽两高”数据）还是不够的，引入其他先验信息是非常必要的。目前认为，最重要的先验信息就是地下介质的结构信息，首先是大套的反射层位信息、断层信息、不规则地质体信息，如果可能的话，再继续引入地震相这样的更细节的地质信息。这会大幅度补充中波数段速度成分，从而提升速度建模的精度。上述分析知，低中波数波阻抗成分的缺失对于定量地描述岩性变化会引起很大的误差。在当前“两宽两高”地震数据采集逐渐成为常态，高精度的速度建模算法越来越多的情况下，通过机器学习算法提取更多的用于补充速度建模中波数成分的先验信息，进而获得可靠的、包含丰富的低中波数信息的背景阻抗模型，可以保证最终定量的岩性成像结果的可靠性。

前面的分析还指出：在实际数据情况下，宽带反射系数估计结果很难达到定量化。宽带反射系数估计利用的是炮集中反（绕）射波同相轴上地震子波的振幅信息。影响同相轴上地震子波振幅的因素太多了。理论上，应该剥离掉其它因素对振幅的改变，仅仅留下介质阻抗扰动引起的振幅扰动，才能用适当的算法由波场振幅扰动反演地下介质阻抗扰动。我们（WPI）认为：宽带反射系数估计不是一个数学算法可以解

决的，需要一个合理的技术流程，结合合理的关键技术，基本上能得到与地下介质波阻抗扰动量一致的、定量的成像结果。我们提出如下的宽带反射系数定量化估计的技术流程：

① 地表一致性振幅校正，消除激发端和接收端导致的地震子波的振幅变化，余下的振幅变化仅由传播过程和阻抗扰动引起。

② 球面扩散校正。偏移成像过程能处理球面扩散校正时，不必单独处理。

③ 吸收衰减校正。偏移成像过程能处理吸收衰减校正时，不必单独处理。

④ 透射损失校正、薄层干涉校正、散射损失校正。这些因素引起的振幅变化往往被忽略。

⑤ 保真的叠前深度偏移成像。当前，我们的建议是做好常规的叠前深度偏移（PSDM），产生（方位）角度道集，进行照明优选的保真叠加。不建议用 LS_PSDM 和 LS_RTM，因为它们施加的理论假设过多，实际数据很难满足。

⑥ 成像域点扩散函数（PSF）反褶积。目的是校正照明不均匀，进一步提升反射系数成像结果保真度和分辨率。

⑦ 进行测井反射系数与地震成像反射系数的量级标定。我们建议在上述保真成像尽可能做好的情况下，仅仅对大套标志性反射层的反射系数进行一致性标定。

总之，带限反射系数的相对保真性由保真成像处理各环节保障；带限反射系数的频带展宽，甚至扩展成 Gauss 型反射系数，需要提一个合适的约束非线性反问题并进行有效的求解；宽带反射系数成像结果的定量要靠测井反射系数的标定，可以认为井反射系数是一种质控信息。

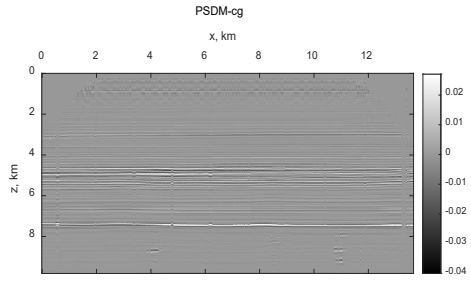
正是在上述认识基础上，我们把特征波成像（CWI）+宽带波阻抗建模（WBIM）流程及技术系列用在理论和实际数据上，基本上验证了方法技术系列的有效性。

6 定量地震波成像数值实验

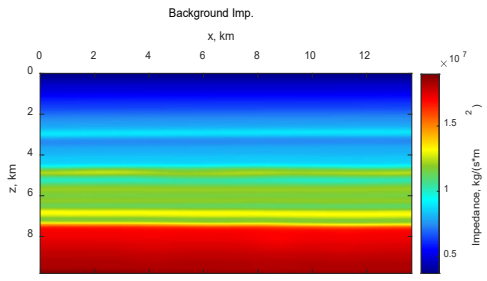
我们在西北某探区抽象出的三维模型数据和三维实际数据上，用特征波成像（CWI）+宽带波阻抗建模（WBIM）技术流程进行了数据处理。

6.1 三维模型数据上宽带波阻抗建模

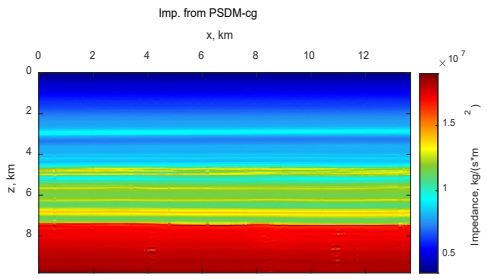
首先利用三维模型来模拟西北某探区实际地质情况。其中，背景速度由模型真速度经过半径 375m 的 Gauss 平滑产生，用 Garder 速度密度关系产生背景密度，从而生成背景阻抗。宽带反射系数 r_{mig} 由 Kirchhoff 叠前深度偏移结果生成。通过（11）式定义的信息融合宽带波阻抗重构方法生成宽带波阻抗建模结果。



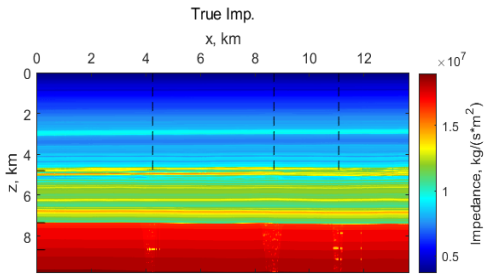
a.反射系数振幅标定后的 Kirchhoff 叠前深度偏移结果剖面



b.基于真速度平滑（半径 375m）后的背景速度产生的背景阻抗

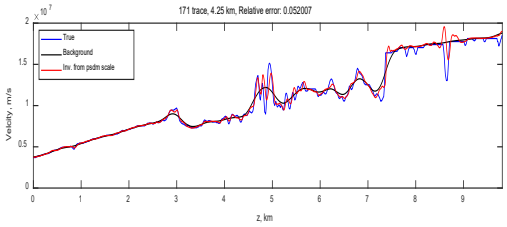


c.信息融合宽带阻抗建模结果

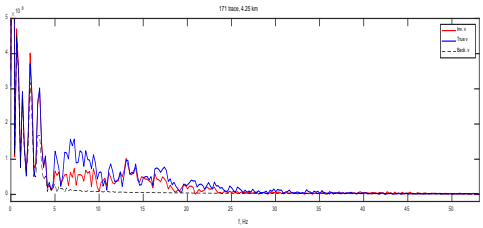


d.真阻抗模型

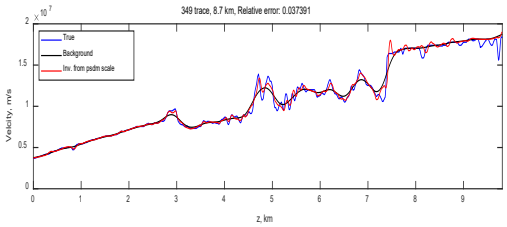
图 15 三维模型数据上宽带波阻抗建模结果的剖面显示。



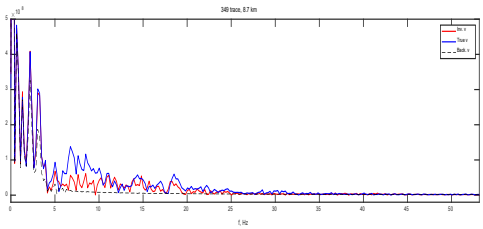
a1. 图 15d 左边虚线位置处背景阻抗、重构阻抗和真阻抗的对比



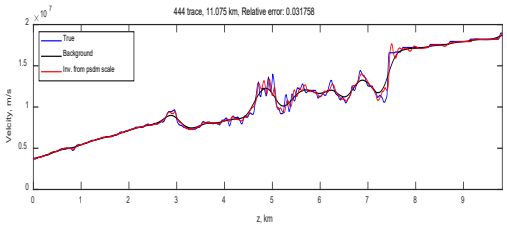
a2. 背景阻抗、重构阻抗和真阻抗的振幅谱对比



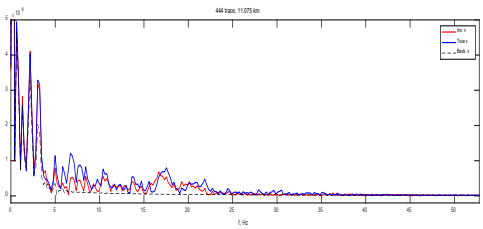
b1. 图 15d 中间虚线位置处背景阻抗、重构阻抗和真阻抗的对比



b2. 背景阻抗、重构阻抗和真阻抗的振幅谱对比

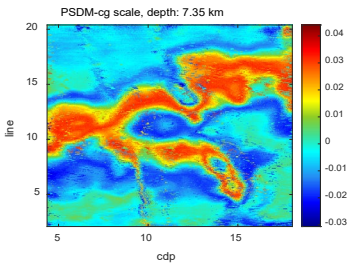


c1. 图 15d 左边虚线位置处背景阻抗、重构阻抗和真阻抗的对比

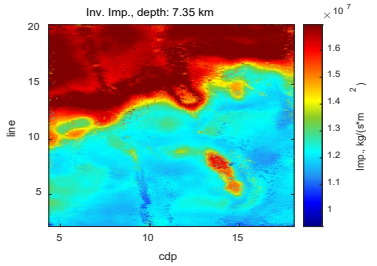


c2. 背景阻抗、重构阻抗和真阻抗的振幅谱对比

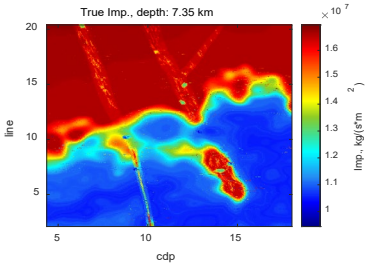
图 16 对应图 15 虚线所示位置处的宽带波阻抗建模结果的抽线对比。



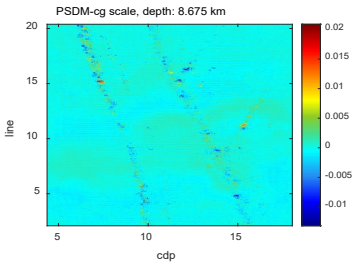
7.35Km 深处标定后的宽带反射系数切片显示



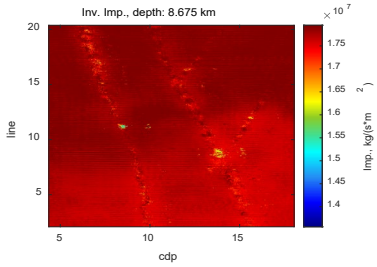
7.35Km 深处重构宽带波阻抗切片显示



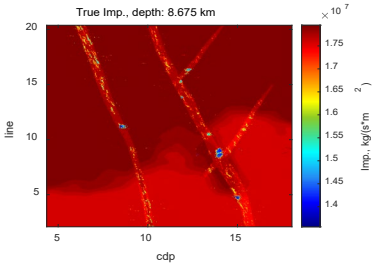
7.35Km 深处真波阻抗切片显示



8.675Km 深处标定后的宽带反射系数切片显示



8.675Km 深处重构宽带波阻抗切片显示



8.675Km 深处重构宽带波阻抗切片显示

图 17 三维模型数据上宽带波阻抗建模结果不同深度的切片显示。

图 15 进行了三维模型数据上宽带波阻抗建模结果的剖面显示。可以看出重构阻抗和真实阻抗由很好的一致性。细致的对比通过图 16 的抽线对比来完成。图 16 显示出：重构的宽带阻抗在低波数段和高波数段都是比较可靠的，5-10Hz 左右的中波数段重构阻抗的振幅谱值与真实阻抗的振幅谱值差异比较大。可以认定这是 Kirchhoff 叠前深度偏移成像结果中，没有提供出中波段的信息。其中，由于带限反射系数成像结果频带有限，反射系数旁瓣太深，导致波阻抗建模中出现假的波阻抗跃变。这在前面的影响因素分析中已经提到过。中波数段部分缺失当然也可以归结为半径 375 米的平滑半径过大，导致背景阻抗中也没有包含这部分中波数成分的信息。

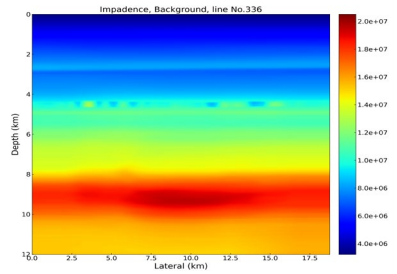
图 17 显示出：振幅标定后的 Kirchhoff 叠前深度偏移成像结果的深度切片并没有很好地展示出地下岩性变化的真实情况。与真实阻抗相比，即便在几何形态上，也与真实的阻抗扰动差异较大。尤其在 7.35Km 的深度切片上，这一点看得很清楚。理论上，这是不应该出现的现象。如果叠前深度偏移做得符合理论预期，叠前深度偏移结果至少在几何形态上能很准确地反映地下介质波阻抗扰动情况。问题当然出在 Kirchhoff 叠前深度偏移成像受偏移速度和成像算法的影响上。应该提及的是，由于三维模型正演模拟数据量巨大，此处用的 Kirchhoff 叠前深度偏移成像结果是由模拟数据一方提供的。即便如此，宽带波阻抗重构结果与真实阻抗的一致性很高，显然更适于岩性油藏的地质解释。

这就从理论模型数据上证明了我们提出的信息融合宽带波阻抗建模路线及方法技术的有效性。

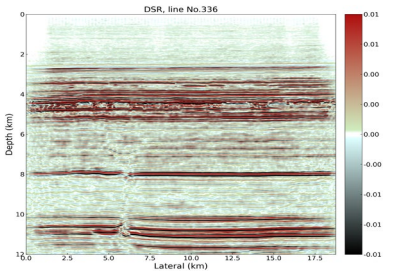
6.2 实际数据上宽带波阻抗建模

选择西北某探区的实际数据进行宽带波阻抗建模处理。关键技术环节包括：①、背景波阻抗模型的建立；② 波动方程叠前深度偏移方法；③ 方位角度成像道集的输出；④ 成像道集后处理，包括道集拉平、最佳照明选择、最优加权叠加；⑤ 成像域反褶积；⑥ 基于信息融合的宽带波阻抗建模。

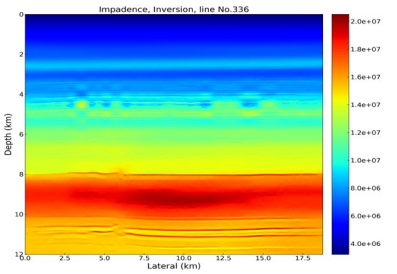
背景阻抗建模中的背景速度来自于叠前深度偏移所用的速度场，背景密度用 Gardner 关系由背景速度转换得到。宽带反射系数由前段中的步骤 2-5 产生。步骤 6 由求解（11）式完成。



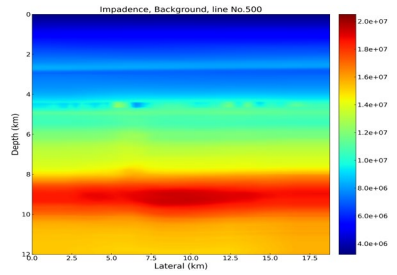
a1 背景阻抗剖面



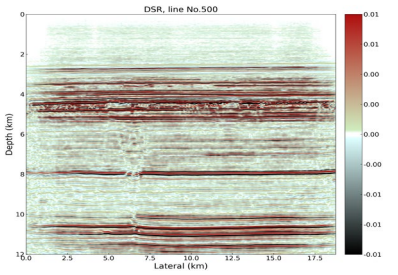
a2 标定后的宽带反射系数剖面



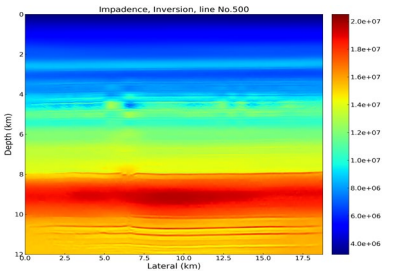
a3 重构的宽带波阻抗剖面



b1 背景阻抗剖面

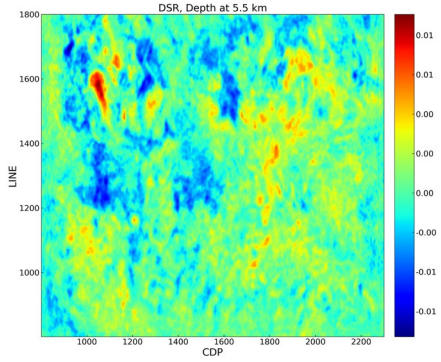


b2 标定后的宽带反射系数剖面

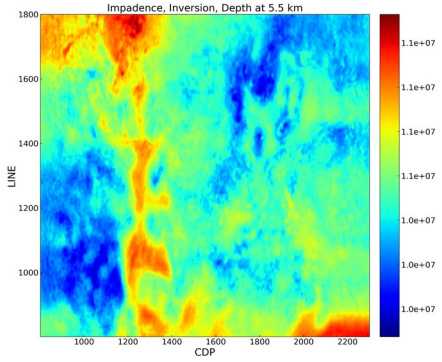


b3 重构的宽带波阻抗剖面

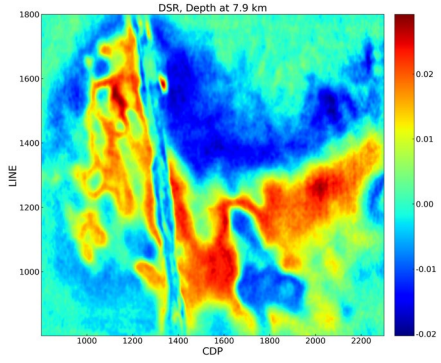
图 18 实际数据上信息融合宽带波阻抗重构的剖面显示。

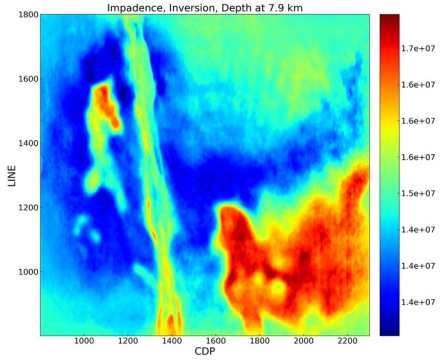


a1 振幅标定后的叠前深度偏移结果深度切片（5.5Km）



a2 重构波阻抗深度切片（5.5Km）





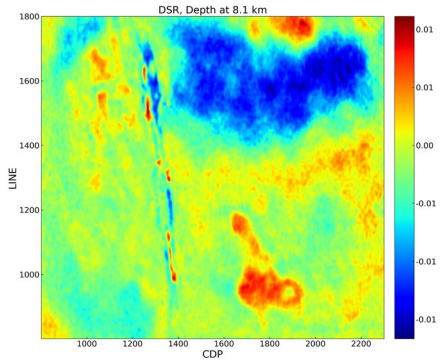
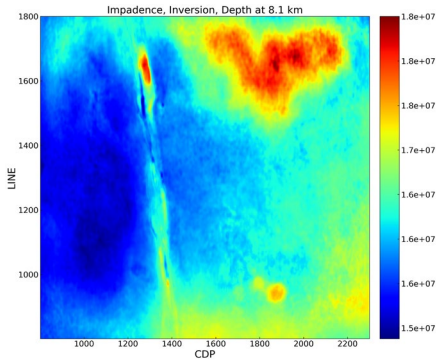
b1 振幅标定后的叠前深度偏移结果深度切片（7.9Km）	b2 重构波阻抗深度切片（7.9Km）
	
c1 振幅标定后的叠前深度偏移结果深度切片（8.1Km）	c2 重构波阻抗深度切片（8.1Km）

图 19 实际数据上信息融合宽带波阻抗重构的深度切片显示。

图 18 按剖面展示了实际数据上信息融合宽带波阻抗重构结果。图 19 按深度切片展示了实际数据上信息融合宽带波阻抗重构结果。由于不知道真实阻抗，信息融合宽带波阻抗重构结果好坏的判断是比较困难的。只能从成像分辨率和地质逻辑合理性来评判。从信息融合的角度，算法不会伤害原有信息的分辨率。理论上，反射系数是绝对阻抗的微分，反射系数剖面的视觉分辨率应该会高于宽带阻抗。但是，从图 19 展示的深度切片看，反射系数成像结果的地质合理性不如重构阻抗，在 5.5Km 处显得十分杂乱。宽带阻抗成像结果的地质合理性还要与该探区的地质解释人员结合进一步判断。

从实际数据处理结果看，我们提出的信息融合宽带波阻抗建模路线及方法技术是可行的。但宽带阻抗成像结果在油藏描述中的有效性，还要进一步检验。

无论如何，实际数据情况下，针对岩性储层的定量地震波成像，要求地震数据采集环节、地震数据预处理各环节、背景速度建模、背景密度建模、宽带反射系数定量估计，都要全面提升质量和精度，才有可能得到能显著提升油藏描述精度和勘探效益的结果。因此我们认为沿着针对岩性储层的定量地震波成像方向提升勘探技术，比用 FWI 技术引领勘探技术的发展更合理。

7 结论与讨论

随着油气勘探的新领域（深水、深层、复杂储层、非常规储层等）逐步扩展，油气勘探的具体目标越来越复杂。但是，精确描述油气藏、确定最佳井位和最佳钻井方案、获得最高油气勘探效益的需求并没有变化；精确描述油气藏需要更精确的地震波成像结果的需求也没有改变。总体上讲，提供宽波数带的、能反映介质岩性变化的、定量的地震波成像结果是石油工业的核心问题。

地震波成像的本质就是利用叠前地震数据和与待估计参数相关的先验信息，利用人为提出的地下介质参数变化和实测数据中波场变化之间的数学物理关系，利用 Bayes 参数估计理论下提出的（非）线性化反演方法，得到介质参数变化的估计结果。FWI 方法、层析成像方法、LS_PSDM/LS_RTM 和一般的偏移成像就是典型地震波成像方法。基于时距关系的动校正和静校正本质上也应归于地震波成像的范畴。宽波数带的参数估计结果是地震波成像的根本目标。带限反射系数成像在石油工业界广为实用，主要是因为对应的偏移成像算法稳定而且物理意义清楚，并非带限反射系数成像结果更能满足精确油藏描述的需求，它仅仅是深度域真三维空间中波阻抗相对变化的体现。实际上，在一个三维油气勘探空间中，仅有很少部位有显著的波阻抗相对变化。当然，我们知道油气勘探主要是对波阻抗发生变化的部位感兴趣。

宽带的、定量的地震波成像结果（譬如纵波速度 $v_P(x,y,z)$ 或波阻抗 $I(x,y,z)$ ）是在深度域真三维空间中对地下介质岩性变化进行准确刻画的物理量，具有明确的油气地质意义，适于进行高精度的、定量的油藏描述。FWI 反演方法的真正目的也就是获得宽带的、定量的参数估计结果，但这是个强非线性反问题，极难得到稳定收敛的、有地质意义的、能实用的解。

本文中，我们提出的特征波成像（CWI）+信息融合宽带波阻抗建模（WBIM）方法技术系列，本质上是把 FWI 宽带波阻抗反演，分解成背景速度建模、背景密度建模、线性化的宽带反射系数估计这三个凸性更好的反问题进行求解。考虑到实际数据的复杂性，这三个关键技术的任何一个都还需要做进一步的变通，目的是在实际数据的限制下还能得到“最佳”的成像结果。关键是我们并没有把宽带波阻抗估计提成一个非线性的反演问题来求解，而是转化成一个信息融合问题，把已有的、认为最好的背景速度建模、背景密度建模、线性化的宽带反射系数估计结果，通过提出和求解一个约束非线性优化问题，重构出宽带波阻抗模型。这样做完全避开了反演求解宽带波阻抗时的不稳定、不收敛问题，同时在此过程中还可以进一步引入其他先验信息，在信息融合这一步继续提升宽带阻抗建模的精度。

本文提出的由背景速度、背景密度、宽带反射系数，结合其他先验信息，融合后得到宽带波阻抗，是不同于目前常规波阻抗反演的一种新的宽带波阻抗建模方法技术。建模结果稳定，地质含义更清楚。宽带波阻抗模型各波数带的信息来源清楚，低频（低波数）来自背景波阻抗，高频（高波数）来自高分辨反射系数成像。中频成分的补充依赖于背景速度建模和构造成像精度的提高（背景速度的波数带向中波数段展宽），也依赖于高分辨反射系数成像精度的提升（宽带反射系数的波数带向中波数段展宽）。

期望的宽带反射系数是高斯型反射系数，其次是理想子波型宽带反射系数。Ricker 子波型反射系数成像结果因为有较宽的主瓣和较为宽深的旁瓣，导致在宽带波阻抗建模中引入很大的误差。Ricker 子波型反射系数成像结果之所以有这样的特点主要是当前很多实际观测数据缺低中频（譬如 2-8Hz）有效成分，且观测方位不够宽和偏移距不够长导致的。数据的低信噪比对成像精度的影响首先体现在影响速度建模的精度上，然后体现在影响偏移成像的精度上。高精度宽带波阻抗建模需要有低中频（中波数）段优势成分的宽带反射系数（譬如方差较小的 Gauss 型反射系数）。宽带反射系数中的低频（包括中频）成分的来源与宽频子波、宽方位长偏移观测又是密切关联的。背景速度的估计精度事实上也严重依赖宽频子波、宽方位长偏移观测。“两宽两高”地震数据观测对定量地震波成像具有核心作用。另外，即便做了保真的地震波成像，宽带反射系数（量级）依然需要测井反射系数的标定，宽带反射系数的量级对波阻抗反演结果影响很大。这说明定量的地震波反演成像需要井与震的结合，这是无法回避的。

尽管“两宽两高”地震数据采集已经成为业界的共识，也已经得到普及应用。但是，满足油藏描述期望的、定量的宽带成像结果依然很难达成。我们一直认为：宽带地震波成像（或宽带参数估计）是（一直是）一个信息不足情况下的最佳估计问题。高精度的地震波成像（宽带参数估计）必须要引入更多的地质、测井、岩石物理信息。针对岩性储层描述的、定量的地震波成像过程与油藏描述过程逐渐融为一体。在大数据、人工智能和机器学习算法的赋能下，测井数据中与岩性变化相关的信息要充分地挖掘出来；地震成像剖面上与地下构造、沉积相关的信息要尽可能地挖掘出来；岩石物理知识，也包括定性的地质知识也要能被计算机表达，所有这些信息都会有助于提升宽带参数估计的精度，在深度域全三维空间中尽可能准确地描述地下介质岩性的状态。因此，今后的地震波成像技术发展，我们认为应该在定量的岩性成像概念引领下向前发展，而不是在 FWI 的理念下向前发展。追求定量的岩性成像，针对的是石油工业的需求；追求 FWI 成像，更多的是算法方面的深究。

尽管文中数值实验显示我们提出的特征波成像（CWI）+信息融合宽带波阻抗建模（BWIM）这种定量的岩性成像方法技术路线是合理的、有效的，但要达到在石油工业界中普及应用的目标，还有很多工作要做。我们认为当前要深化的工作主要不是体现在发展更复杂的反演成像算法上，而是充分利用机器学习算法提取更多的能提升宽带参数估计精度的先验信息，并把这些先验信息合理地融入当前已有的反演成像算法中。对实际应用而言，定量地震波成像对于成像处理各环节是否做到了保真的要求很高，因此质量控制是十分重要的。一项好的方法技术要落地应用，不仅仅是靠理念正确和方法技术的有效，各环节的质控往往起到决定成败的作用。

8 致谢

感谢中石油勘探开发研究院及西北分院、中海油研究院和湛江分公司、中石化物探技术研究院、胜利油田分公司和西北油田分公司对波现象与反演成像研究组（WPI）研究工作的资助与支持。

9 参考文献

1. SHERIFF R E, GELDART L P. Exploration seismology[M]. Cambridge university press, 1995.

2. TARANTOLA A. Inverse problem theory and methods for model parameter estimation[M]. France: Society for industrial and applied mathematics, 2005.

3. THEODORIDIS S. Machine learning: a Bayesian and optimization perspective[M]. Academic press, 2015.

4. 王华忠, 盛燊. 走向精确地震勘探的道路[J]. 石油物探, 2021, 60(5): 693-708

   WANG H Z, SHENG S. Pathway toward accurate seismic exploration. Geophysical Prospecting for Petroleum, 2021, 60(5): 693-708.

5. KAIPIO J, SOMERSALO E. Statistical and computational inverse problems[M]. Springer Science & Business Media, 2006.

6. 王华忠,冯波,王雄文,等.特征波反演成像理论框架[J]. 石油物探, 2017, 56(1):38-49

   WANG H Z, FENG B, WANG X W, et al. The theoretical framework of characteristic wave inversion imaging[J]. Geophysical Prospecting for Petroleum, 2017, 56(1):38-49

7. ASTER R C, BORCHERS B, THURBER C. Parameter Estimation and Inverse Problems[J]. Elsevier, 2004.

8. STOLT R H, WEGLEIN A B. Seismic Imaging and Inversion: Application of Linear Inverse Theory[M]. Cambridge University Press, 2012.

9. ZHANG Z G, Wu Z D, WEI Z Y, et al. FWI Imaging: Full-wavefield imaging through full-waveform inversion[C]. Expanded Abstracts of 90th SEG International Exposition and Annual Meeting. 2020:656-660.

10. WEI Z Y, MEI J W, WU Z D, et al. FWI imaging: Revealing the unprecedented resolution of seismic data[C]. Expanded Abstracts of First International Meeting for Applied Geoscience & Energy. 2021: 682-686

11. HE Y, XING H, HUANG Y, et al. Inversion-based Imaging: FWI beyond Velocity[C]. Expanded Abstracts of 82nd EAGE Annual Conference & Exhibition. 2021:1-5

12. KALINICHEVA T, WARNER M, MANCINI F. Full-bandwidth FWI[C]. Expanded Abstracts of 90th SEG International Exposition and Annual Meeting. 2020:651-655.

13. HAFFINGER P, GISOLF D, PANOS D P, et al. Wave-equation based seismic AVO inversion: A pragmatic approach to use the full elastic wave-field for exploration and production[C]. Expanded Abstracts of Second International Meeting for Applied Geoscience & Energy. 2022:3087-3091

14. 李庆忠.走向精确勘探的道路[M].北京：石油工业出版社，1993

   LI Q Z. The way to obtain a better resolution in seismic prospection[M]. Beijing: Petroleum Industry Press, 1993.

OBN 地震数据成像处理基本逻辑与关键方法技术

王华忠，项健，石聿

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】海洋油气勘探逐渐进入深水深层勘探领域，地下地质构造复杂（横向变速剧烈）、目标油藏复杂（由构造油气藏为主转向构造与地层岩性油气藏并重），同时可能还伴随有海底地形及附近岩性的复杂变化，所有因素促使着海洋油气地震勘探技术的不断变革。提高海洋油气勘探效益的首要问题是发展尽可能满足高精度地震波成像需求的地震数据采集技术及对应的高精度地震波成像技术。当前，无论海上和陆上油气地震勘探，“两宽一高”地震数据采集技术和全波形反演（FWI）/最小二乘逆时偏移（LS_RTM）为代表的地震波成像技术是标志性的领先技术。海上油气地震勘探中，OBN 地震数据采集是目前业界公认的、最有可能真正实现“两宽一高”地震数据采集的技术。与拖缆数据采集相比，海底节点（OBN）数据采集具有宽方位照明、数据信噪比高、无检端鬼波、存在实测的（至少一阶自由表面相关）下行波场、四分量观测等方面的优点。尤其是宽方位照明和存在至少一阶自由表面下行波场的特点，使得 OBN 数据具备了对中深层复杂构造和近海底介质进行高精度成像的能力。着重讨论高精度地震波成像对地震数据采集的要求，指出 OBN 数据采集在海洋油气勘探中的必要性；分析 OBN 数据采集的地震波场的特点，据此提出 OBN 数据地震波成像处理基本逻辑及相应的关键技术；认为海洋油气勘探中地震波成像处理的特殊问题主要由特征反射层引起，海水面、海底面和地下介质中若干强反射层构成了这些特征反射层，提出模型驱动波动理论特征反射层相关多次波预测与压制的技术路线，并对比了几种代表性的多次波预测的基础理论；指出对应当前的线性化偏移成像算子叠前数据域与叠前成像域是等价的，据此以成像道集后处理为中心，给出期望成像道集的定义，把弱旁瓣、定量的反射系数作为保真高分辨地震波成像的目标，在两个域中尽可能完美实现地下同一反（绕/散）射点、不同炮检距反（绕/散）射子波的同相位叠加，尽可能好地实现保真高分辨带限反射系数的成像；提出最好把带限反射系数成像推进到宽带波阻抗成像的技术路线；结合 OBN 数据的特点，给出了 OBN 数据地震波成像处理的基本技术流程，指出各环节的关键方法技术。最后，针对 OBN 数据四分量观测的特点，指出是实际观测的多波地震波场中的波现象（主要是 P_SV 波）与地震波传播及模拟理论不匹配导致了当前多波成像结果达不到预期，建议重点研究实际观测的多波地震波场中的波现象与地震波传播及模拟理论不匹配的物理根源，而不是发展更高端的矢量波成像算法。期望本文的思想观点对 OBN 地震勘探在海洋油气勘探中的进一步应用产生积极的促进作用。

【关键词】海底节点（OBN）地震数据采集及成像处理，特征反射层相关多次波，模型驱动波动理论特征反射层相关多次波预测与压制，海底节点（OBN）地震数据成像处理流程及关键技术。

1 引言

油气地震勘探的核心目标：精确地描述油气藏，进行准确的含油气性分析，做出最佳的钻井决策，得到最高的油气勘探效益。油气地震勘探的核心问题：由叠前地震数据及其它与弹性参数相关的先验信息，进行宽波数带的弹性参数估计（或称广义的高精度地震波成像），基于井数据并与岩石物理知识结合，进行精确的油气藏描述和准确的含油气性评价。对于宽波数带的弹性参数估计或广义的高精度地震波成像而言，成像结果主要表现为保真高分辨方位角度带限反射系数（所谓的高精度背景速度建模结果还不能对精确的油藏描述提供有效的信息），这依旧是当前油气地震勘探地震地质解释最基础的信息。今后可能逐渐发展到宽带波阻抗成像(王华忠, 2021)。

在复杂情况下，譬如深水、深层情形下，地下地质构造复杂（横向变速剧烈）、目标油藏复杂（由构造油气藏为主转向构造与地层岩性油气藏并重），同时可能还伴随有海底地形及海底附近岩性的复杂变化时，保真高分辨方位角度反射系数成像依然是首先要解决的核心问题。这就需要满足地震波成像要求的叠前地震数据体。

海洋油气勘探的技术现状是：依据拖缆和/或 OBS（OBN/OBC）观测的地震数据，利用地震波成像技术（层析反演速度建模+叠前深度偏移成像）对目标地质体进行保真的和高分辨的带限方位角度反射系数成像，基于带限方位角度反射系数成像结果及地质、钻井与岩石物理信息进行油藏描述，进入油气开发阶段(苟量, 2021)。

我国各海域油气地震勘探的关键问题可以大致总结为：1)、南海北部诸探区：莺歌海探区存在地层高温高压、底辟气烟囱导致成像模糊不清和底辟边界刻画不准的问题。琼东南探区存在崎岖海底、凹陷边界及凹陷内构造成像模糊的问题。珠江口盆地（白云凹陷）探区存在陆架坡折带地形复杂、重力流砂体厚度薄、深层存在复杂构造变化，以及中生界潜山内幕成像差、潜山边界成像不清楚的问题。2)、渤海湾探区存在河流相和湖相沉积构造小、薄、碎，潜山边界、潜山内幕成像困难的问题；以及复杂（浅）水体引起的噪声难以压制干净导致成像模糊、波组特征不清楚，难以满足薄/小砂体油藏精确描述的问题。3)、东海海域包括钱塘凹陷、长江口凹陷、苏北坳陷、南黄海坳陷，以砂泥岩沉积为主，陡构造不发育，含油气储

层主要为小尺度砂岩体。存在海底强反射及某些强反射地层引起的长程层间多次波所导致的成像困难的问题。东海海域诸探区，高分辨、高保真成像是主要的目标。针对我国海域油气勘探存在的问题，有必要进一步发展适应性的地震数据采集和高精度地震波成像技术。

与陆上地震勘探相比，海洋地震勘探地震波成像处理的特殊性主要是由海上激发方式、接收方式和包含水体的地下介质中的波传播决定的。海水介质的存在、气枪为主的震源激发及海水压力变化的检测是引起陆海地震数据成像处理差异的根源。水层介质的存在及气枪震源的使用对后续的地震波成像处理带来的最大好处是各炮检之间的地震子波一致性保持较好、没有面波及面波相关的强干扰噪声。但是，海面及海底这两个特殊反射面带来了特殊的问题。总体地，可以罗列为：1)、与海上枪阵激发相关的问题：枪阵组合形式与激发方式、气泡效应。2)、与拖缆相关的问题：拖缆飘移、拖缆定位。3)、与震源和检波器移动相关的问题。4)、与海水速度的动态变化相关的问题。5)、与波传播相关的问题：鬼波（源鬼波、检鬼波、源检鬼波）；海面相关多次波（水体多次波；各种海面相关长程多次波）；硬海底、软海底、崎岖海底、深海底、浅海底相关的波现象。6)、与观测系统相关的问题：OBS（OBN、OBC）引入的相关问题；窄缆观测、多方位（窄缆、宽缆）观测、宽方位观测(Dhelie, 2018)、全方位观测（Coiling）(Moldoveanu, 2015)、沉浸缆观测(Brizard, 2016)引入的相关问题；混叠源观测、连续源观测(Hegnal, 2019)引入的相关问题等。总体而言，最重要地还是体现在：海面相关多次波构成了影响海上油气勘探地震波成像质量的关键因素，海面相关多次波的压制或成像成了海上油气地震勘探的特殊问题。海面、海水和（硬）海底构成的海水体模型，尤其在浅水探区，引起的复杂波现象的消除也是海洋地震勘探的一个特殊问题。

当前海洋油气勘探的基本状况是：勘探目标在向深水、深层、复杂构造与复杂岩性储层方向转化，其主要矛盾依然是地震数据采集和地震波成像不能满足油气藏识别、描述与评价的精度要求。毋容置疑，海上“两宽一高”地震数据观测方式和相应的成像处理方法是解决这个矛盾的针对性方法技术。海面、硬海底面、地下介质特征强反射层引起的多次波是海洋地震数据处理的特殊问题。

首当其冲的依然是地震数据采集问题。海上“两宽一高”地震数据观测是解决海上油气地震勘探中地震波成像不能满足油气藏识别、描述与评价的精度要求问题的根本基础。从目前情况看，OBN 地震数据采集是在海上油气勘探中实现“两宽一高”地震数据观测的合理和可行的技术方案(Hoffe, 2000; Thompson, 2007; Bathellier, 2019; Holloway, 2015)。全方位观测（Coiling）(Moldoveanu, 2015)、沉浸缆观测(Brizard, 2016)，这两种海上地震数据采集方法应该也是海上“两宽一高” 地震数据观测的合理方法，但是至今没有得到推广应用，问题可能出在技术可行性上。OBN 地震数据采集（其他可行的“两宽一高”地震数据采集技术也是可以的）+特征反射层相关多次波压制或成像+初至波 FWI（各向异性）速度建模+成像道集（各向异性）层析速度建模+成像道集为中心的高保真高分辨方位角度带限反射系数成像（最好推进到宽带波阻抗成像），然后在此基础上进入精细油藏描述阶段，应该是今后很长一段时间内海洋油气地震勘探的技术方向。

本文着重讨论高精度地震波成像对地震数据采集的要求，指出OBN数据采集在海洋油气勘探中的必要性；分析 OBN 数据采集的地震波场的特点，据此提出 OBN 数据地震波成像处理基本逻辑及相应的关键技术；提出模型驱动波动理论特征反射层相关多次波预测与压制的技术路线，并对比了几种代表性的多次波预测的基础理论；提出以成像道集后处理为中心，把弱旁瓣、定量的反射系数作为保真高分辨地震波成像的目标，最好推进到宽带波阻抗成像的技术路线；结合 OBN 数据的特点，给出了 OBN 数据地震波成像处理的基本流程，指出各环节的关键方法技术。

2 海洋油气勘探与陆地地震勘探中地震数据采集及成像处理问题的对比

无论海上和陆上油气勘探中的地震波成像处理，首要的问题是震源（震源与周围介质的相互作用）、波在地下介质中的传播（在地表附近介质中的传播、在浅层介质中的传播、在中深层介质中的传播及在深层超深层介质中的传播）及检波器（检波器与周围介质的相互作用）所决定的野外实际采集到的地震波场。

目前来看，只有炮集中的初至（包括早至波）波场（同相轴）和一次反（绕/散）射波场（同相轴）对地震波成像有贡献，而且初至（包括早至波）波同相轴上地震子波走时用于浅表层层析速度反演及建模；一次反（绕/散）射波同相轴上地震子波走时用于中深层速度层析反演及建模；一次反（绕/散）射波同相轴上地震子波振幅用于偏移成像/反演成像估计地下介质的带限反射系数。在偏移速度正确时，棱柱波（Duplex Wave）也可以用于逆时深度偏移（RTM）成像；OBN 数据中一阶下行波场（属于一阶自由表面多次波）也可以通过镜像偏移得到成像结果。一般地，炮集中其他波现象都是要通过预处理压制掉的。即

便是 FWI 技术，也极难把炮集中所有波现象贡献于地震波成像。初至（包括早至波）波场在近地表介质中传播，主要波现象是直达波、折射波和 Diving Wave。事实上，其中可能还有多次折射波、多次 Diving Wave、导波的波场成分，这些波现象的存在使得早至波的利用变得很困难或不可能。初至波（包括早至波）传播路径穿透浅层，最大穿透深度取决于最大偏移距和速度随深度变化情况。一次反（绕/散）射波来自于震源下行波场对反射界面和绕（散）体的有效照明，也取决于检波器对上行反射和绕（散）射波场的有效接收。宽角度和均匀角度增量的有效源端照明和同样的有效检端接收是反射界面及绕（散）体能否可靠地进行地震波成像的根本数据基础。浅层区域（譬如 500ms 以浅区域）反（绕/散）射波采样密度不够，深层超深层反（绕/散）射波照明角度太窄，使得这两个区域中反射界面及绕（散）体成像的精度很难达到地质解释的要求。

上述地震波成像与地震数据观测及地震波传播的关系对海洋和陆上地震勘探都是适用的。依据上述分析可以部分地看出，为什么油气地震勘探领域坚持要开展“两宽一高”的地震数据观测。陆上地震勘探中，（单）节点检波技术已经成为实现“两宽一高”地震数据采集的核心。海洋油气地震勘探中，OBN 地震数据采集技术的情况与陆上地震勘探是一致的。

对比陆上地震勘探的炮集和海洋地震勘探的炮集，二者的主要差异是由震源、近地表介质和检波器造成的。海洋地震勘探情形是：气枪震源（或正在发展中的海洋可控震源(Pramik, 2013)）、水平海面+（可能崎岖的）软/硬海底+速度变化很小的水体+压力检波器；陆上地震勘探情形是：炸药震源或可控震源、（极可能崎岖的）地表面+（极可能）横向速度变化剧烈的风化层+位移（更可能是速度）检波器。面波及面波散射波为主的复杂近地表相关波场+（可能剧变的）道间时差+不同炮检对地震子波振幅和波形的不一致构成了陆上地震数据的主要特点，造成了陆上地震数据处理的主要困难。因此，Cox 认为：（陆上）勘探地震的主要困难是近地表因素引起的(Cox, 1999)。与陆上地震勘探相比，海洋油气勘探中存在两个特殊的强反射面，即海水面和海底面，正是这两个面的存在，才导致了海洋地震数据成像处理的特殊性。这两个面，尤其是海水面的存在，引起的主要波现象：鬼波（源端鬼波、检端鬼波、源检鬼波）；水体相关的多次波；来自海底面下强反射层的自由表面多次波。压制这三种波现象是海洋地震数据成像处理的重点。相比而言，海洋地震勘探炮集中的波现象总体上要比陆上地震勘探炮集中的波现象简单很多。但是，应该注意浅水介质系统（海面+海水+海底及附近介质）下，该系统也会产生复杂的波现象，严重降低浅水探区地震数据的质量，进而降低地震波成像结果的质量。也可以说，海洋油气地震勘探中地震波成像处理的主要困难也是（复杂）水体系统所导致的。客观地讲，正是因为炮集中波现象的简单，海洋地震数据成像处理要比陆上地震数据成像处理简单一些。

除了各自的近地表因素带来的问题，中深层地震波成像的问题是由不同探区的地质情况造成的，与海洋地震勘探或是陆上地震勘探关系不大，各自都有简单地质情况的探区和复杂地质情况的探区。

因此，可以得出如下结论性观点：海洋油气勘探的特殊问题主要是由反射系数接近于 1 的海面+海水+海底及附近介质构成的介质系统所引起的。自由表面（海面）相关多次波的预测、压制或成像处理构成了海洋地震数据处理的特殊问题。其他方面的问题，要么与陆上地震数据处理一致，要么没有明显的特殊性。

据此我们提出了：水体模型（包含海底附近介质）建模+水体相关多次波压制+自由表面（海面）相关（长程）多次波压制+FWI（各向异性）速度建模+成像道集（各向异性）层析速度建模+成像道集为中心的高保真高分辨方位角度带限反射系数成像的海洋地震数据成像处理技术组合。

同时我们认为：海洋地震勘探（也包括陆上地震勘探），多次波主要由特征反射界面引起的，所谓特征反射层包括海面、海底面及地下介质强反射面。据此我们提出特征反射层相关多次波预测与压制的技术路线。

3 OBN 地震数据观测系统及 OBN 数据的特点

前已述及，海洋油气勘探的根本问题：地震数据采集和地震波成像不能满足油气藏识别、描述与评价的精度要求。

理论上讲，地震波成像对地震数据观测技术的基本需求：对地下任意一个成像点进行高精度的成像必须要有不同方位、不同偏移距（不同入射反射张角）的均匀的照明。更准确地讲，均匀采样的方位和偏移距上都有宽带的、形态一致的地震子波，同时方位角展布范围和偏移距展布范围要足够宽。如果地下介质空间中任意一点都有这样的地震波照明，对应的观测系统就是比较理想的。本质上，这就是要求“两宽一

高”的数据观测。应该注意到这是对反（绕/散）射波高精度成像对地震数据采集提出的基本要求。对于高精度地震波层析成像而言，对地震数据采集的要求可以描述为：对地下任一个速度分析单元，要有宽角度的、角度均匀展布的、角度增量尽可能小的波路径穿过。对浅层、中层和深层的任一速度分析单元，都期望有这样的照明。对比可知，目前的地震数据观测方式对于浅层和深层超深层介质的高精度速度建模是十分不友好的。因此，满足高精度地震波成像需求的“两宽一高”地震数据观测的真正实现还有十分大的差距。陆上和海洋油气勘探中，地震波成像对地震数据观测技术的基本需求都是如此，地震数据采集的技术状况也是如此。

目前看来，海洋地震勘探实现“两宽一高”地震数据观测的技术难度比陆上地震勘探更大。首先宽方位、多方位拖缆观测，很难保障有较大的横纵比。全方位观测（Coiling）(Moldoveanu, 2015)和沉浸缆观测系统(Brizard, 2016)似乎是实现海上“两宽一高”数据观测的很好方式，但是目前并没有大规模商业化应用，主要原因应该是当前的技术可实现性存在问题。OBN 观测方式理论上是海洋油气勘探中实现“两宽一高”数据观测的较好方式(Holloway, 2015)，当前海洋地震数据采集的实践也证明了这一点。但是，相比陆上油气地震勘探，密集的 OBN 布设成本太高。宽方位采集是 OBN 数据观测方式的固有优势；宽频带数据主要靠枪阵激发来保障（侧重低频的海洋宽频可控震源的商用化将进一步保障宽频数据的获取(Pramik, 2015)）；高密度采集的实现是用高密度的炮点激发来换取的，这样做还能保证下行照明角度充分且照明能量充分。从石油工业界直至当前的实践看，OBN 数据观测方式实现海上“两宽一高”地震数据采集基本成为共识。

OBN 数据观测方式的特点决定了与其对应的地震波成像处理的特点。为完整起见，对比一下拖缆数据与 OBN 数据的特点及成像处理关键技术。

拖缆数据的技术特点及对应的成像处理关键技术：① 拖缆的漂移引起检波器空间位置的不规则。如果是 Coiling 观测系统，数据规则化更是核心的处理环节。抗假频的数据规则化方法与技术(王华忠等, 2016)是拖缆数据处理的核心技术之一。② 混叠激发的高效采集(Tenghamn, 2018)是海上地震数据采集的常规技术，解混叠（Deblending）方法(Chen, 2014; 宋家文等, 2019)是不可缺少的技术环节。解混叠（Deblending）方法是拖缆数据处理必备的技术。③ 鬼波和水体相关多次波，以及来自深层强反射的自由表面相关多次波是对地震波（一次反（绕/散）射波）成像产生严重干扰的波现象。压制鬼波和水体相关多次波，以及来自深层强反射的自由表面相关多次波是拖缆数据成像处理真正需要的核心技术。应该注意：倾斜缆、变形缆、双层缆压制鬼波都需要对应的方法技术。到底是用野外采集技术压制鬼波或是用平缆观测+鬼波压制算法来压制鬼波还是争论中的问题。

OBN 数据的技术特点及对应的成像处理关键技术：① OBN 节点比较稀疏，炮点位置不规则，导致 OBN 数据空间假频和空间采样不规则现象较为严重。抗假频的数据规则化方法与技术是 OBN 数据处理的核心技术之一。② OBN 数据，与任何海上数据一样，鬼波和水体相关多次波，以及来自深层强反射的自由表面相关多次波是对地震波（一次反（绕/散）射波）成像产生严重干扰的波现象。压制鬼波和水体相关多次波，以及来自深层强反射的自由表面相关多次波是 OBN 数据成像处理的第二项核心技术。③ OBN 数据的另一个特殊性表现为震源和检波器位于两个不同深度面上。OBN 观测到的一定是至少一阶的自由表面多次波。上、下行波分解方法（水陆检合并方法技术(Schalkwijk, 2003)）、炮检不同面的 PSDM 方法（也包括 PSTM 方法）、镜像 PSDM 方法(Ronen, 2005)是必须具备的方法技术。④ OBN 数据的真正优势是宽方位观测。利用宽方位数据的速度建模、方位角度道集的生成及宽带反射系数成像是更为核心的问题。这是 OBN 数据成像处理的第四项核心技术。不过这与拖缆数据和陆上数据的成像处理技术是一致的。⑤ OBN 数据是四分量采集的。水检数据和三分量陆检数据的成像处理方法技术要具备：水检和陆检（垂直分量）合并的上下行波分解技术；水陆和陆检（垂直分量）结合的横波压制技术；三分量陆检数据的成像处理方法技术等。应该注意到：与拖缆数据一样，OBN 数据采集，为了提高效率也会进行混叠激发。因此，同样存在混叠激发数据的解混叠问题。这与拖缆数据的解混叠方法技术是一样的。还应该注意：OBN 原始数据由巨量的炮集组成，不能进行炮集为基础的成像处理，必须基于互易原理，按共接收点道集进行所谓的“单炮处理”，譬如 FWI 和 RTM。这是 OBN 数据成像处理的一个特殊性。关于 OBN 多分量成像处理，P-SV 转换波成像至今没有很成功的实例，P-SV 转换波成像结果没有表现出能提升油藏描述精度及可靠性的潜力。利用压力检波器与陆检垂直分量中纵波成分的高度相关性，压制陆检垂直分量中的 P-SV 波成分的处理是有意义的。压力检波器与陆检垂直分量合并进行上下行波分解的处理效果是明显的。总之，OBN 四分量观测数据合理的成像处理方法技术还要继续探究。

4 特征反射层相关多次波预测与压制

前述分析表明，海洋油气勘探中地震波成像处理的特殊问题主要由特征反射层引起。海水面、海底面和地下介质中若干强反射层构成了这些特征反射层。浅水区域油气勘探时，浅水介质系统（海面+浅水体+海底面及附近介质构成该系统）会产生复杂的干扰波场。深水区域油气勘探时，水体相关多次波和长程自由表面多次波构成主要的干扰波场。地下介质中若干强反射层形成的层间多次波在特定的地质情况下都可能存在，与浅水、深水，甚至陆上探区都没有关系。

正是特征反射层相关多次波预测与压制方法技术构成了海上地震数据成像处理的基本逻辑特点。其他的成像处理方法，譬如子波相关的处理方法、各种去噪方法、数据规则化方法、解混叠方法、偏移速度建模方法和叠前深度偏移成像方法，尽管针对海上拖缆观测数据、海上OBN观测数据及陆上地震数据处理，这些方法技术各有特点和不同，但都不存在本质性差异。

基于上述认识，我们提出：特征反射层相关多次波的预测与压制(王华忠等, 2017)。目的是抓住多次波压制的主要矛盾。原因是一般意义下的、统一的多次波预测与压制理论方法及技术很难提出，提出了也很难具体实现。特征反射层相关多次波预测与压制基本逻辑步骤包括：1)、背景速度+特征反射层构成的速度模型；2)、特征反射层相关的多次波模拟；3)、多次波模拟波场与实测多次波波场的配准；4)、最佳匹配滤波器设计及多次波减去。

水体相关多次波的压制是首先要解决的。因此，我们提出了如下的OBN数据水体速度建模方法流程。首先用下行直达波和初始水体速度进行简单形式的叠前深度偏移成像，得到海底形态（也可以通过其他方式得到初始海底形态）；然后进行一阶水体相关多次波射线走时模拟，基于模拟结果的约束用机器学习算法在实测炮集上检测一阶水体相关多次波走时；根据走时差和射线追踪结果构建层析反演方程，层析反演速度更新量，更新水体速度。同时偏移成像更新海底形态。视情况进入下一轮的水体速度模型反演。最多用到二阶水体相关多次波走时进行水体速度模型层析反演。用简单的方法建立水体速度模型当然也是可以的。已知水体速度模型的基础上，用单向波和双向波数值模拟结合的方法，单向波模拟一次反射波，双向波模拟水体多次波，从双向波模拟波场中减去一次波模拟结果，得到水体中传播的各阶多次波。由于水体速度模型的不准确（水体速度和海底形态可能都不准确），多次波模拟结果与炮集中实测的多次波会存在一定的时差，因此进行多次波模拟波场与实测多次波波场的配准，尽可能消除二者之间的时差。最后进行二者之间最佳匹配滤波器的设计并把水体相关多次波减去。

其它特征反射层之间多次波的消除，也按同样的逻辑进行。特殊之处是特征反射层的拾取。基于机器学习算法发展出了各种层位拾取的方法。有了背景速度+特征反射层构成的速度模型，就可以按照类似水体相关多次波的预测与压制方法进行特定特征反射层的多次波压制。

特征反射层相关多次波预测与压制，应该从浅到深逐层进行。

这是我们提出的具有一般意义的模型驱动海洋多次波预测与压制的方法体系。该方法体系也可以用于陆上地震数据层间多次波压制的处理过程中。

下面从理论上对比一下，当前代表性的多次波压制的方法原理。

4.1 特征反射层相关多次波预测与压制

为了建立准确的多次波预测模型，需要研究波在介质中的传播过程，分析多次波的形成机制。在地震勘探中，常用标量波方程描述地震波的传播。频率空间域声波方程表示为：

$$\left(\nabla^2+\frac{\omega^2}{v^2(\boldsymbol{x})}\right)u\left(\boldsymbol{x},\boldsymbol{x}_s\right)=-\delta(\boldsymbol{x}-\boldsymbol{x}_s)s(\omega) \tag{1}$$

假设地下介质模型由背景速度v_0和速度扰动组成，定义扰动函数$\varepsilon\left(\boldsymbol{x}\right)=v_0^2\left(\boldsymbol{x}\right)/v^2\left(\boldsymbol{x}\right)-1$，带入声波方程有:

$$\begin{aligned}\left(\nabla^2+\frac{\omega^2\left[\varepsilon(\boldsymbol{x})+1\right]}{v_0^2(\boldsymbol{x})}\right)u(\boldsymbol{x},\boldsymbol{x}_s)&=-\delta(\boldsymbol{x}-\boldsymbol{x}_s)S(\omega)\\ \left(\nabla^2+\frac{\omega^2}{v_0^2(\boldsymbol{x})}\right)u(\boldsymbol{x},\boldsymbol{x}_s)&=-\frac{\omega^2}{v_0^2(\boldsymbol{x})}\varepsilon(\boldsymbol{x})u(\boldsymbol{x},\boldsymbol{x}_s)-\delta(\boldsymbol{x}-\boldsymbol{x}_s)S(\omega)\end{aligned} \tag{2}$$

用 $L(\boldsymbol{x},\omega)=\nabla^2+\frac{\omega^2}{v_0^2(\boldsymbol{x})}$ 表示背景介质中波的传播算子，引入逆算子 L^{-1}，格林函数定义为震源是脉冲函数时介质中的波场分布：$G_0(\boldsymbol{x},\boldsymbol{x}_s,\omega)=-L^{-1}\delta(\boldsymbol{x}-\boldsymbol{x}_s)$。将介质扰动视为由一系列散射点组成，在各个散射点处产生背景场的扰动波场，总的地震波场由背景场和扰动场组成，将逆算子 L^{-1} 带入式(2)有：

$$\begin{aligned}u(\boldsymbol{x},\boldsymbol{x}_s)&=L^{-1}\left(-\delta(\boldsymbol{x}-\boldsymbol{x}_s)s(\omega)-k_0^2\varepsilon(\boldsymbol{x})u(\boldsymbol{x},\boldsymbol{x}_s)\right)\\ &=u_0(\boldsymbol{x},\boldsymbol{x}_s)+\int L^{-1}\left(-\delta(\boldsymbol{x}-\boldsymbol{x}')k_0^2\varepsilon(\boldsymbol{x}')u(\boldsymbol{x}',\boldsymbol{x}_s)\right)d\boldsymbol{x}'\\ &=u_0(\boldsymbol{x},\boldsymbol{x}_s)+k_0^2\int G_0(\boldsymbol{x},\boldsymbol{x}')\varepsilon(\boldsymbol{x}')u(\boldsymbol{x}',\boldsymbol{x}_s)d\boldsymbol{x}'\end{aligned} \tag{3}$$

上式为 Lippmann-Schwinger 积分方程，其中 $k_0=\omega/v_0$；u 表示总场；u_0 表示背景波场。将（4）式写为震源与 Green 函数褶积的形式有：

$$\begin{aligned}u=GS&=\left(G_0+G_0VG\right)S=\left(I-G_0V\right)^{-1}G_0S\\ &=\left(G_0+G_0VG_0+G_0VG_0VG_0+\cdots\right)S\end{aligned} \tag{4}$$

式中 G_0 表示背景介质的格林函数，G 表示完整介质的格林函数，V 表示介质扰动。Lippmann-Schwinger积分方程是一个非线性积分方程，要得到总的地震波场需要先知道总场本身，难以直接求解。在 Born 近似下假设介质是弱扰动的，忽略二阶及以上的多次散射，可以得到一次波与介质扰动之间的线性关系。但在实际地震勘探中，波传播介质并不总是弱扰动的，强反射/绕射体的存在导致波场中的多次散射不能直接忽略。

在海上地震勘探中，由于自由表面这一强反射界面的存在，自由表面相关多次波广泛发育。考虑将介质扰动 V 视为由自由表面 r_0 和其它扰动部分 V_0 组成：$V=r_0+V_0$，以自由表面为基准面，地下介质的一次散射响应表示为：

$$X_0(\boldsymbol{x}_r,\boldsymbol{x}_s)=G_0(\boldsymbol{x}_r,\boldsymbol{x})V_0(\boldsymbol{x})G_0(\boldsymbol{x},\boldsymbol{x}_s) \tag{5}$$

式中 $X_0(\boldsymbol{x}_r,\boldsymbol{x}_s)$ 表示在自由表面 $\boldsymbol{x}_s$ 处激发的脉冲震源，经过地下介质散射后在 $\boldsymbol{x}_r$ 的输出，则地表观测数据中的一次波表示为 $P_0=SX_0$。实际地下介质内部存在多次散射，导致层间多次波，这里将其视为 V_0 产生的等效一次波。将 Lippmann-Schwinger 积分方程近似为一次散射和自由表面 r_0 相关的多次散射之和，检波器在自由表面处接收到的上行波场 P 表示为：

$$\begin{aligned}P&=\left(X_0+\left(r_0X_0\right)X_0+\left(r_0X_0\right)^2X_0+\cdots\right)S\\ &=P_0+M=P_0+M_1+M_2+\cdots\end{aligned} \tag{6}$$

式(6)的形式与 Lippmann-Schwinger 积分方程类似，其中的各阶级数表示各阶的自由表面多次波。自由表面多次波的形成过程可以用图 1 所示的反馈环模型表示：到达自由表面的上行波场，经过自由表面的下行反射，形成“二次源”输入地下介质，形成更高一阶的自由表面多次波。根据式(6)可知高一阶的多次波是由低一阶的多次波/一次波褶积 r_0 和 X_0 得到的，所有阶的自由表面多次波可以由地表接收到的上行波场预测：

$$M=\left(r_0X_0\right)P=\left(r_0S^{-1}P_0\right)P \tag{7}$$

式中 S^{-1} 表示子波反褶积算子。式(7)与 SRME 方法(Berkhout, 1997)得到的自由表面多次波预测模型类似，其由信号系统的角度出发，将自由表面多次波的预测描述为观测数据的高维褶积；理论上该方法需要规则完整的观测数据，并通过多轮迭代压制多次波。自由表面多次波预测与波传播模拟之间的关系将通过表示定理说明。

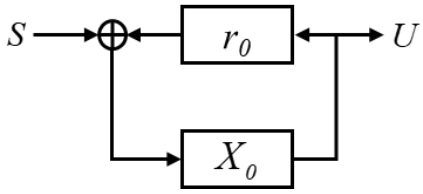


图 1 自由表面多次波的反馈环模型

4.2 Green 表示定理表达的自由表面多次波预测

考虑如下图 2 所示的地震勘探模型(Ikelle, 2005)，震源和检波器放置在一个半球封闭面 $S=S_0+S_R$ 内，S_0 表示下行反射系数为-1 的理想自由表面，S_R 表示图 2 中半圆面。检波点处的波场 $p(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)$ 用如下积分方程表示：

$$\begin{aligned}&p(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)=G(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)s(\omega)\\&+\oint_S\sigma(\boldsymbol{x}')\left[G(\boldsymbol{x}_r,\boldsymbol{x}',\omega)\frac{\partial p(\boldsymbol{x}',\boldsymbol{x}_s,\omega)}{\partial\boldsymbol{n}}-p(\boldsymbol{x}',\boldsymbol{x}_s,\omega)\frac{\partial G(\boldsymbol{x}_r,\boldsymbol{x}',\omega)}{\partial\boldsymbol{n}}\right]d\boldsymbol{x}'\end{aligned} \tag{8}$$

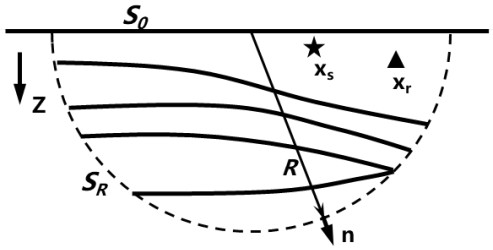


图 2 海上地震勘探物理模型

式中 $p(\boldsymbol{x}',\boldsymbol{x}_s,\omega)$ 表示积分面上点 $\boldsymbol{x}'$ 处的声压波场；$\boldsymbol{n}$ 是封闭面的单位外法线方向向量；$\sigma=1/\rho$ 表示密度倒数；$G(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)$ 是声介质的脉冲响应；方程中的第一项 $G(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)s(\omega)$ 表示 $\boldsymbol{x}_s$ 点震源激发的，在 $\boldsymbol{x}_r$ 处产生的声压波场，该部分波场由封闭面内的介质产生，包括直达波、反射波等波现象；第二项是 Kirchhoff-Helmholtz 积分，其表示封闭面上的波场及其法向导数在介质中任意点处产生的波场。当球面半径 R 趋于无穷大，根据索末菲辐射条件只需要考虑 S_0 上的波场积分。由于自由表面 S_0 处压力值趋于 0，仅考虑声压法向导数产生的波场，由此得到如下含自由表面多次波的表示定理：

$$p(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)=G(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)s(\omega)+\sigma_0\oint_{S_0}G(\boldsymbol{x}_r,\boldsymbol{x}',\omega)\frac{\partial p(\boldsymbol{x}',\boldsymbol{x}_s,\omega)}{\partial\boldsymbol{n}}d\boldsymbol{x}' \tag{9}$$

式中对 S_0 的积分即为自由表面多次波。由于自由表面 S_0 是近似水平的，其法向导数近似为垂向导数，有 $\sigma_0\dfrac{\partial p(\boldsymbol{x}',\boldsymbol{x}_s,\omega)}{\partial z}=i\omega v_z(\boldsymbol{x}',\boldsymbol{x}_s,\omega)$，自由表面多次波的预测关系表示为：

$$m(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)=i\omega\oint_S G(\boldsymbol{x}_r,\boldsymbol{x}',\omega)v_z(\boldsymbol{x}',\boldsymbol{x}_s,\omega)d\boldsymbol{x}' \tag{10}$$

式中 $v_z(\boldsymbol{x}',\boldsymbol{x}_s,\omega)$ 表示自由表面处的垂向速度分量；$G(\boldsymbol{x}_r,\boldsymbol{x}',\omega)$ 是自由表面以下介质的格林函数；

$m(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega)$ 表示预测的多次波。式(10)表明可以将观测数据的垂向分量作为“二次源”输入已知的参数模型，通过波场模拟预测自由表面多次波。

4.3 单向波+双向波数值模拟表达的特征反射层多次波预测

根据(10)式的自由表面多次波模拟预测关系，理论上需要地下介质的完整参数模型才能求解格林函数 $G(\boldsymbol{x}_r,\boldsymbol{x}',\omega)$ 。为了简化多次波预测对模型参数的需求，仅考虑预测地下强阻抗的特征反射层产生的自由表面多次波。在已知背景速度和反射层位置条件下，特征反射层相关的自由表面多次波由输入的“二次源”在自由表面和特征反射层间往返传播产生，其分解为波场的下行延拓和上行延拓过程,为此需要选择合适的波传播算子进行波场模拟。

在 Born 近似下，特征反射界面被视为由一系列散射点组成，有如下的多次波预测关系:

$$m\left(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega\right)=\omega^2\int_x\int_{x'}G_0\left(\boldsymbol{x}_r,\boldsymbol{x},\omega\right)\varepsilon\left(\boldsymbol{x}\right)G_0\left(\boldsymbol{x},\boldsymbol{x}',\omega\right)d\left(\boldsymbol{x}',\boldsymbol{x}_s,\omega\right)d\boldsymbol{x}'d\boldsymbol{x} \tag{11}$$

式中 $d\left(\boldsymbol{x}',\boldsymbol{x}_s,\omega\right)=i\omega v_z(\boldsymbol{x}',\boldsymbol{x}_s,\omega)$ 表示观测数据的在自由表面处的二次源输入； $\varepsilon\left(\boldsymbol{x}\right)$ 表示散射点的扰动强度； $G_0\left(\boldsymbol{x},\boldsymbol{x}',\omega\right)$ 和 $G_0\left(\boldsymbol{x}_r,\boldsymbol{x},\omega\right)$ 分别表示背景介质中的下行和上行传播格林函数。上述方程用时空域的单向波+双向波模拟实现，表示为：

$$\begin{cases} u^+\left(\boldsymbol{x},t\right)=\int_{x'}d\left(\boldsymbol{x}',\boldsymbol{x}_s,t\right)*g_0\left(\boldsymbol{x},t;\boldsymbol{x}',0\right)d\boldsymbol{x}' \\ m\left(\boldsymbol{x}_r,\boldsymbol{x}_s,t\right)=\int_x\varepsilon\left(\boldsymbol{x}\right)\partial_t^2u^+\left(\boldsymbol{x},t\right)*g_0\left(\boldsymbol{x}_r,t;\boldsymbol{x},0\right)d\boldsymbol{x} \end{cases} \tag{12}$$

式中 $d\left(\boldsymbol{x}',\boldsymbol{x}_s,t\right)=FT^{-1}\left(d\left(\boldsymbol{x}',\boldsymbol{x}_s,\omega\right)\right)$ 为时间域的输入数据； g_0 是时空域的背景格林函数； $u^+\left(\boldsymbol{x},t\right)$ 表示输入数据产生的下行波场； $m\left(\boldsymbol{x}_r,\boldsymbol{x}_s,t\right)$ 表示多次波预测模型。上述波场在时空域满足如下声波方程：

$$\begin{cases} \left(\nabla^2+\dfrac{\partial_t^2}{v_0(\boldsymbol{x})}\right)u^+\left(\boldsymbol{x},t\right)=d\left(\boldsymbol{x}',\boldsymbol{x}_s,t\right) \\ \left(\nabla^2+\dfrac{\partial_t^2}{v_0(\boldsymbol{x})}\right)m\left(\boldsymbol{x},t\right)=\varepsilon\left(\boldsymbol{x}\right)\partial_t^2u^+\left(\boldsymbol{x},t\right) \end{cases} \tag{13}$$

式中 $v_0(\boldsymbol{x})$ 表示背景速度模型。自由表面多次波的模拟预测过程如下图 3 所示，图 3(a)为自由表面处的观测数据，其经过自由表面反射形成图 3(b)所示的下行波场，下行波在背景模型中传播，在图 3(c)所示的特征反射层扰动处形成多次散射，散射场在背景模型中上行传播到自由表面，得到图 3(d)所示的多次波预测模型。

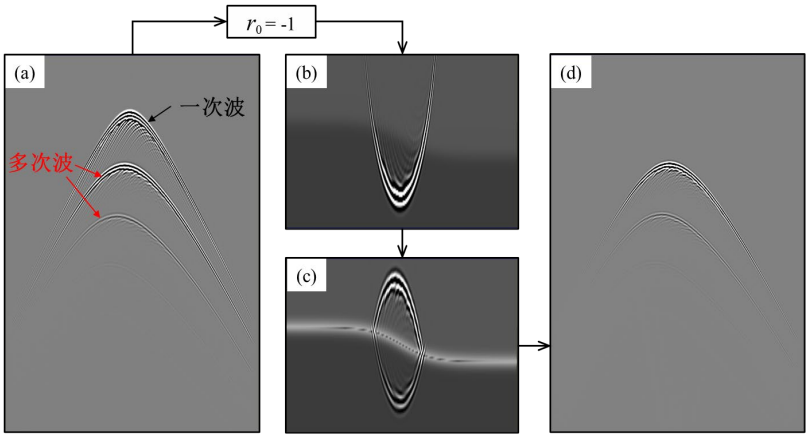


图 3 自由表面多次波的模拟预测过程；(a)自由表面观测数据；(b)下行波场；(c)特征反射层的多次散射场；(d)多次波预测模型

4.4 OBN 上、下行波反褶积多次波压制

在 OBN 观测系统中，震源在海面激发，观测节点位于海底。假设地下介质是水平层状的，OBN 观测的上、下行波数据中，各个方向的平面波独立地满足反褶积关系(Amundsen, 1999)。因此将 OBN 共检波点道集数据进行平面波分解，对各个平面波分量分别进行上下行波的反褶积压制自由表面相关多次波。

以自由表面为基准面，OBN 上、下行波数据的平面波分解表示如下：

$$\begin{aligned} P_{z_r}^{(U)}(k,\omega) &= P_{z_0}^{(U)}(k,\omega)e^{-ik_z z_r} \\ P_{z_r}^{(D)}(k,\omega) &= P_{z_0}^{(D)}(k,\omega)e^{ik_z z_r} \end{aligned} \tag{14}$$

式中 $P_{z_0}^{(U)}$ 和 $P_{z_0}^{(D)}$ 表示自由表面处的上、下行波数据，数据量级都用声压分量标定；z_r 表示 OBN 放置的深度；$e^{\pm ik_z z_r}$ 表示波场深度方向的延拓算子，实现自由表面观测向 OBN 观测的转换；$P_{z_r}^{(U)}$ 和 $P_{z_r}^{(D)}$ 表示 OBN 数据压力分量的上、下行波成分，具有如下反褶积关系：

$$X_0(k,\omega) = \frac{P_{z_r}^{(U)}(k,\omega)}{P_{z_r}^{(D)}(k,\omega)} e^{-2ik_z z_r} \tag{15}$$

式中 $X_0(k,\omega)$ 即为地下介质的一次反射响应，其不包含自由表面多次波。该一维反褶积过程在平面波域逐频率进行。由于 OBN 放置在海底，式(15)的上下行波反褶积结果只包含海底界面以下的介质信息，对海底及浅层构造的照明不足，常用 OBN 的下行波数据进行镜像偏移，补充浅部成像信息。OBN 数据的下行波反褶积表示为：

$$X_0(k,\omega) = \frac{P_{z_r}^{(D)}(k,\omega) - S'}{r_0 \cdot P_{z_r}^{(D)}(k,\omega)} \tag{16}$$

式中 S' 表示下行波数据 $P_{z_r}^{(D)}$ 中的直达波。原下行波数据作为反褶积算子，对去除了直达波的下行波数据反褶积，同样能够求得 X_0，并且其中包含海底面的反射信息。上述 OBN 数据的上下行波反褶积和下行波反褶积方法（UDD 和 DGD）通常在频率域实现，需要考虑其稳定性和抗噪性(Majdanski, 2011)。

上、下行波的褶积关系如图 4 所示。在图 4a 中，直达波 D_1 为一阶下行波；一阶上行波 $U_1 = D_1 X_0$ 表示为以 D_1 为输入源，与地下介质褶积的输出结果，X_0 表示地下介质的一次反射响应；以此类推可知，OBN 数据中的各阶上行波是以对应阶的下行波为输入源，与地下介质系统褶积得到的；与上下行波褶积类似，在下行波中也存在如图 4b 所示的褶积关系，将直达波作为一阶下行波输入，其与介质响应褶积后产生二阶下行波，后续高一阶的下行波都由低一阶的下行波褶积 X_0 得到。根据上述褶积关系，通过上、下行波数据的反褶积能够获得介质的一次反射响应，从而压制自由表面多次波。

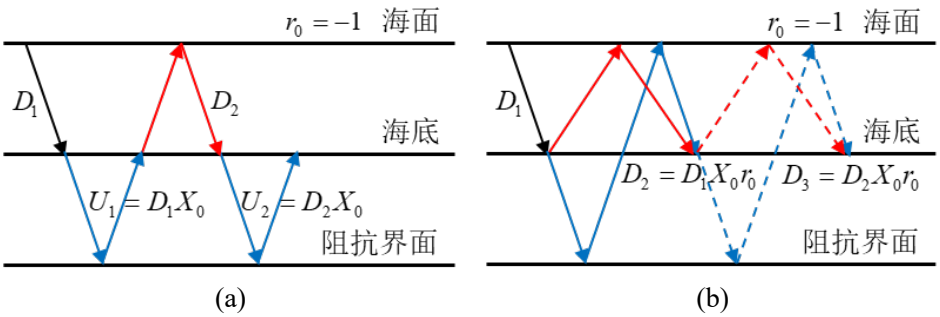


图 4 上、下行波的褶积关系；(a)上、下行波褶积示意；(b)下行波褶积示意

4.5 层间多次波模拟预测

与自由表面多次波类似，地震波会在地下强反射层发生下行反射，导致层间多次波(Jakubowicz, 1998)。通过指定发生下行反射的特征反射层，层间多次波形成的正过程也可以用反馈环模型描述。考虑在 z_i 深度处，界面 V_i 的下行反射形成的层间多次波，定义 V_i 界面以下介质的一次散射响应 X_i 表示为：

$$X_i\left(\boldsymbol{x}_r',\boldsymbol{x}_s',z_i\right)=G_0\left(\boldsymbol{x}_r',\boldsymbol{x}\right)V^*\left(\boldsymbol{x}\right)G_0\left(\boldsymbol{x},\boldsymbol{x}_s'\right) \tag{17}$$

式中 V^* 表示 V_i 界面以下深部的介质扰动； $\boldsymbol{x}_s'$ 和 $\boldsymbol{x}_r'$ 是位于 V_i 界面处虚拟的震源和检波点； $G_0\left(\boldsymbol{x},\boldsymbol{x}_s'\right)$ 和 $G_0\left(\boldsymbol{x}_r',\boldsymbol{x}\right)$ 表示下行和上行传播的背景格林函数。将该散射响应带入 Lippmann-Schwinger 积分方程，得到包含层间多次波的散射级数，表示为：

$$\begin{aligned} u&=G_0S+G_0VG_0S+G_0\left(X_i+\left(X_iV_i\right)X_i+\left(X_iV_i\right)^2X_i+\cdots\right)S_i \\ &=u_0+p_0+m_i \end{aligned} \tag{18}$$

式中 $S_i=V_iX_iS$ 表示在 V_i 界面处发生的下行散射，其作为二次源输入深部介质； m_i 表示由 V_i 界面下行散射导致的层间多次波，形成过程如下图 5 所示。

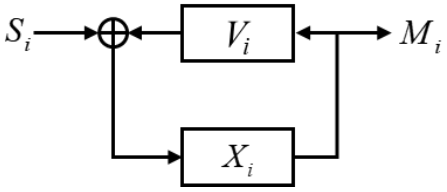


图 5 层间多次波的反馈环模型

波动方程数值模拟是研究波现象的重要方法，常规的有限差分模拟结果难以分离层间多次波和一次波。根据上述含层间多次波的散射级数，通过提取特征反射层，能够实现对层间多次波的直接模拟。在时空域方程表示为：

$$\begin{cases} u_0\left(\boldsymbol{x},t\right)=s\left(\boldsymbol{x}_s,t\right)*g_0\left(\boldsymbol{x},t;\boldsymbol{x}_s,0\right) \\ \delta u_1\left(\boldsymbol{x}',t\right)=\int_{\boldsymbol{x}}\varepsilon^*\left(\boldsymbol{x}\right)\partial_t^2\left(u_0\left(\boldsymbol{x},t\right)+\delta u_2\left(\boldsymbol{x},t\right)\right)*g_0\left(\boldsymbol{x}',t;\boldsymbol{x},0\right)d\boldsymbol{x} \\ \delta u_2\left(\boldsymbol{x},t\right)=\int_{\boldsymbol{x}'}\varepsilon_i\left(\boldsymbol{x}'\right)\partial_t^2\left(\delta u_1\left(\boldsymbol{x}',t\right)\right)*g_0\left(\boldsymbol{x},t;\boldsymbol{x}',0\right)d\boldsymbol{x}' \end{cases} \tag{19}$$

式中 u_0 是震源产生的背景场； ε^* 和 ε_i 分别表示 V^* 和 V_i 的扰动强度； $\boldsymbol{x}'$ 和 $\boldsymbol{x}$ 都表示波场在介质中的空间坐标，仅用于区分 δu_1 和 δu_2； δu_1 和 δu_2 分别是 ε^* 和 ε_i 产生的散射场，其中 δu_2 的下行散射传播到介质 V^* 中产生层间多次波：

$$m_i\left(\boldsymbol{x}_r,t\right)=\int_{\boldsymbol{x}}\varepsilon^*\left(\boldsymbol{x}\right)\partial_t^2\left(\delta u_2\left(\boldsymbol{x},t\right)\right)*g_0\left(\boldsymbol{x}_r,t;\boldsymbol{x},0\right)d\boldsymbol{x} \tag{20}$$

为了求解方程(19)，需要分别提取介质中的反射层 V_i 和 V^*，用波动方程数值模拟背景介质中的波传播。上述波场在时空域满足如下方程：

$$\begin{cases}\left(\nabla^2+\dfrac{\partial_t^2}{v_0(\boldsymbol{x})}\right)u_0(\boldsymbol{x},t)=s(\boldsymbol{x}_s,t)\\ \left(\nabla^2+\dfrac{\partial_t^2}{v_0(\boldsymbol{x})}\right)\delta u_1(\boldsymbol{x},t)=\varepsilon^*(\boldsymbol{x})\partial_t^2\left(u_0(\boldsymbol{x},t)+\delta u_2(\boldsymbol{x},t)\right)\\ \left(\nabla^2+\dfrac{\partial_t^2}{v_0(\boldsymbol{x})}\right)\delta u_2(\boldsymbol{x},t)=\varepsilon_i(\boldsymbol{x})\partial_t^2\left(\delta u_1(\boldsymbol{x},t)\right)\end{cases} \tag{21}$$

方程组(21)每次模拟一个特征反射层V_i相关的层间多次波，模拟的层间多次波都由V_i的下行反射产生，V_i以下的界面V^*只产生上行波。

用图 6(a)所示的模型测试层间多次波模拟方法，该模型从浅到深包含三个界面。考虑第一个反射层相关的层间多次波模拟，反射层的提取和分离如图 6(b)所示。层间多次波模拟过程如下图 7 所示，震源激发如图 7(a)所示的背景波场，背景模型由原模型平滑得到；背景场先在V^*中产生一次散射(图 7(b))，V^*产生的上行散射在V_i处产生下行波场(图 7(c))，将下行波场输入回V^*产生层间多次波(图 7(d))；V^*和V_i波场的相互输入过程是每个时间片进行的，模拟的多次波中包含V_i下行反射产生的所有阶的层间多次波；最后需要单独模拟一次波，将一次波从数据中减去，得到如图 7(e)所示的层间多次波模拟结果。该模拟过程由浅到深逐个特征反射层进行，模拟时清除V_i界面以上的介质扰动。

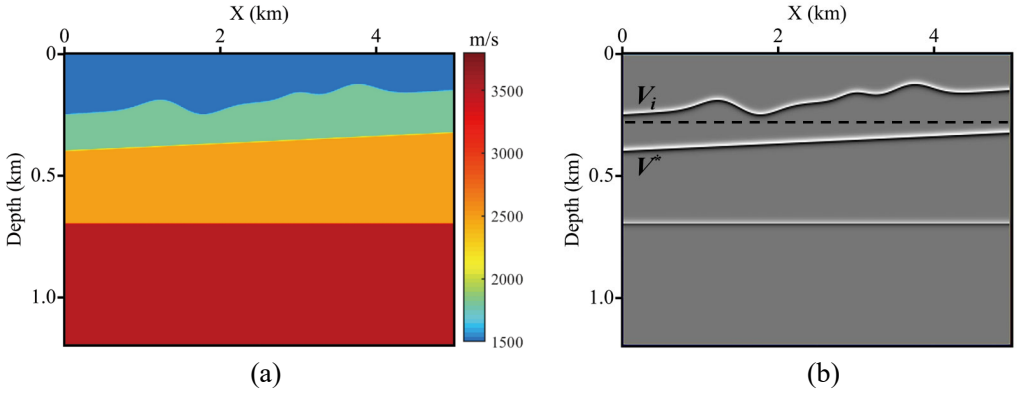


图 6 三层介质速度模型(a)及其反射界面的提取和分离(b)

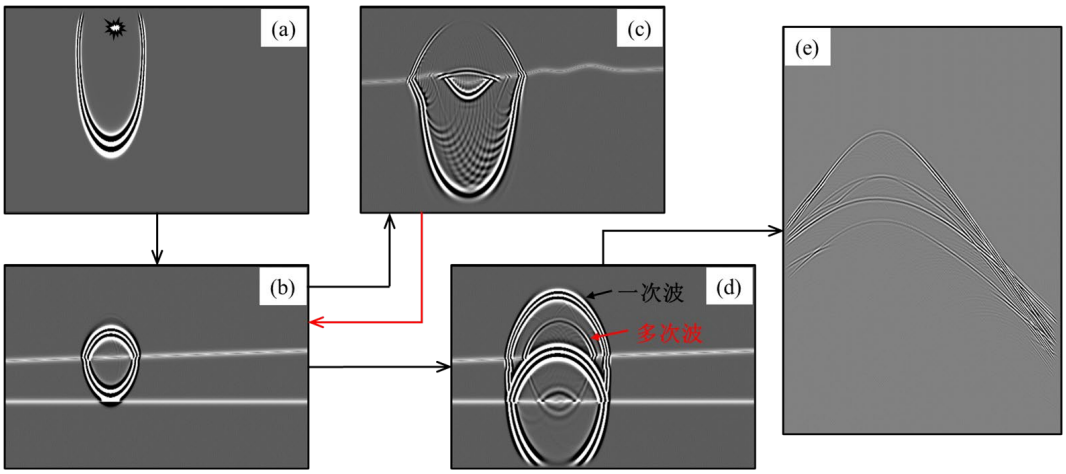


图 7 层间多次波模拟过程；(a)震源下行波场；(b)深部反射层V^*的散射波；(c)特征反射层V_i的下行散射波；(d)包含层间多次波的V^*散射波；(e)分离的层间多次波模拟数据

上述模拟方法假设层间多次波只在一个特征反射层处发生下行反射，会忽略部分二阶及以上的层间多次波，但其已经能够模拟大部分的多次波现象。直接模拟得到的层间多次波模型能够帮助识别和分析实际观测数据中的多次波。

5 OBN 地震数据成像处理流程及关键技术

地下介质的高精度地震波成像是OBN数据成像处理的基本目的，当然也是拖缆和陆上地震数据成像的目的。当前的地震波成像首要目标依然是：方位角度保真高分辨的反射系数，更一般地，可以说是尽可能保真的、带限的相对波阻抗扰动量。

基本的成像公式可以表述为：

$$\delta\mathbf{d}^{obs}=\delta\mathbf{d}^{cal}+\boldsymbol{\eta}=L\left(\mathbf{m}_{\text{initial}}\right)\delta\mathbf{m}+\boldsymbol{\eta} \tag{22}$$

$$\delta\hat{\mathbf{m}}=\arg\min_{\delta\mathbf{m}\in\mathcal{M}}\mathcal{J}\left(\delta\mathbf{m}\right)=\arg\min_{\delta\mathbf{m}\in\mathcal{M}}\left\{\left(\delta\mathbf{d}^{obs}-L\delta\mathbf{m}\right)^{T}C_{D}^{-1}\left(\delta\mathbf{d}^{obs}-L\delta\mathbf{m}\right)+\left(\delta\mathbf{m}\right)^{T}C_{M}^{-1}\left(\delta\mathbf{m}\right)\right\} \tag{23}$$

其中，$\delta\mathbf{d}^{cal}=L\left(\mathbf{m}_{\text{initial}}\right)\delta\mathbf{m}$ 意味着人为选定的波动方程的数值解 $\delta\mathbf{d}^{cal}$ 可以模拟实测波场 $\delta\mathbf{d}^{obs}$ 中特定的波现象，剩余误差 $\boldsymbol{\eta}$ 满足Gauss（白）噪声分布的假设(Tarantola, 2005)。此时，最小二乘意义下的地震波成像结果（或参数估计结果）$\delta\hat{\mathbf{m}}$ 是有可能得到最佳结果的。当然，实际情形与上述地震波成像的线性理论要求相差甚远。$L\left(\mathbf{m}_{\text{initial}}\right)$ 是依赖于初始速度模型 $\mathbf{m}_{\text{initial}}$ 的波传播算子。C_D^{-1} 的引入还是为了满足高精度参数估计（无偏+方差最小的估计）对剩余误差 $\boldsymbol{\eta}$ 满足 Gauss（白）噪声分布的假设。C_M^{-1} 的引入（所引入的是对待估计参数 $\delta\mathbf{m}$ 的先验信息）是为了克服反演结果的不稳定。到目前为止，地震波偏移成像主要用到一次反（绕/散）射波的振幅；地震波层析成像主要用到透射波走时（主要是初至波走时）和一次反（绕/散）射波的走时。期望所选的正问题 $\delta\mathbf{d}^{cal}=L\left(\mathbf{m}_{\text{initial}}\right)\delta\mathbf{m}$ 能很好地描述透射波和一次反（绕/散）射波在实际介质中的传播。应该注意到，（线性化的）地震波偏移成像中存在一个特例，即在 RTM 中 Duplex Wave 能对陡倾角甚至超过 90^{o} 的反射界面进行成像，但必依赖于存在能模拟出 Duplex Wave 的精确的速度结构。这是多次波贡献于陡倾角偏移成像的特例，此时偏移速度模型中背景速度正确且产生 Duplex Wave 的反射界面位置要正确，双向波方程才能产生对 Duplex Wave 传播路径的正确预测，从而对产生 Duplex Wave 的陡倾角界面进行成像。

事实上，炮集中包含的波现象是非常复杂的，海上地震勘探中各种自由表面相关多次波是主要的噪声源。选择一个复杂的正问题尽可能多地模拟炮集中实测的波现象使得剩余噪声 $\boldsymbol{\eta}$ 满足 Gauss（白）噪声分布的假设，或是选择一个简单的正问题模拟炮集中某些特征波现象（所谓特征波现象是指能被所选择的正问题很好模拟的波，譬如前述的透射波和一次反（绕/散）射波），然后把不能模拟的波现象通过预处理手段压制掉，使得剩余噪声 $\boldsymbol{\eta}$ 满足 Gauss（白）噪声分布的假设，这是地震波成像的策略问题。显然，当前石油工业界所用的方法是后者。FWI 方法本质上是属于前者的。但是，FWI 方法在复杂介质和复杂数据情形下的不收敛或很难收敛到有地质意义的解上，也迫使它尽可能利用特征波现象，尽量引入数据预处理压制掉特征波现象之外的不能被所选择的正问题预测的波现象。这是当前地震波成像所遵循的基本逻辑。

基于上述逻辑，海洋/OBN 数据成像处理流程的制定一定要围绕：如何获得仅仅包含透射波和一次反（绕/散）射波的、规则无假频采样的叠前地震数据体；如何获得尽可能准确的偏移速度场（当前重点是保护好初至波，做好初至波 FWI）；如何得到高保真高分辨带限反射系数的成像结果。

另外，高保真高分辨带限反射系数的成像结果依然是目前油藏描述的最主要信息。它是地震波叠前深度偏移提供的。带限反射系数成像的基本理论公式可以写为：

$$\delta\mathbf{m}\approx L\left(\mathbf{m}_{\text{initial}}\right)^{T}\delta\mathbf{d}^{obs} \tag{24}$$

其中，$L\left(\mathbf{m}_{\text{initial}}\right)^{T}$ 是在背景速度场中对地表观测数据进行反传播所使用的算子，高频近似下的射线理论和 Beam 传播理论、单向波方程、双向波方程都可以构造出该算子。背景速度是比较光滑的速度，保证对不同炮检对、不同反（绕/散）射点产生的地震子波走时的预测是正确的，就满足了叠前深度偏移成像对背景速度模型精确性的需求。常密度声波方程构建的RTM算子对复杂构造和强横向变速有最好的适应性，

因此RTM成为了高精度叠前深度偏移成像的代表性方法技术。但是，每种偏移算子都有其适用条件和优缺点。不能认为有了 RTM 就可替代其他偏移方法，RTM 的高精度成像是依赖于高精度的速度模型的（譬如前述的 Duplex Wave 陡倾角成像）。当前，地震波成像算子已经可以处理各向异性介质和吸收衰减介质，但是考虑复杂介质情形带来了各向异性参数建模和 Q 值建模的复杂性。这是当前地震波偏移成像的技术发展现状。大量的文献讨论过偏移算子问题。

(24)式的本质是把地表（海面或 OBN）观测的、不同炮检对的一次反（绕/散射）子波反传到反（绕/散）射点处，并利用合理的成像条件产生成像结果。不同炮检对的一次反（绕/散）射子波的同相位叠加、保幅叠加是高保真成像的基本要求。理论上，地下介质的反射系数只能是正或负的。而实际上，正反射系数出现相伴的负旁瓣或是负反射系数出现相伴的正旁瓣，这是地震波激发与接收的有限频带所导致的。这显然不是地下介质真实反射系数的体现。地震勘探的物理本质某种程度上可以理解为用波前面上的地震子波感知地下介质弹性参数的变化，因此地震子波的分辨率决定了地震波成像的分辨率。地震反演成像的目的就是要剥离掉地震子波对地下介质弹性参数成像的模糊效应，估计真实的地下介质弹性参数的变化。众所周知，地震子波对地下介质弹性参数成像的影响不可能被彻底剥离掉。这是当前地震波成像存在分辨率问题的本质。据此，我们提出了地震波高分辨率叠前偏移成像的期望目标：无旁瓣或弱旁瓣、保幅的、定量的反射系数(王华忠等, 2023)。

勘探地震中，高保真高分辨带限反射系数成像的本质意义就是如此。根据上述认识，结合OBN数据观测方式及所记录波场的特点，可以制定出合理的 OBN 数据成像处理技术流程。

首先，海洋/OBN 数据成像处理技术流程要包括如下关键环节：子波处理环节；水体相关多次波压制环节；自由表面相关多次波压制环节；其他噪声压制环节；数据规则化环节；Deblending 环节（可选）；偏移速度建模环节；叠前深度偏移成像环节；基于成像道集的成像后处理环节。

其中每个处理环节都与上述地震波成像的理论要求有关。事实上，整个海洋/OBN 数据成像处理技术流程围绕前述三个问题展开，即满足线性化的正问题 $\delta\mathbf{d}^{cal}=L\left(\mathbf{m}_{\text{initial}}\right)\delta\mathbf{m}$ 所能预测波现象的要求及基于线性化正问题的成像方法对规则无假频数据的需求；基于线性化正问题所开展的速度层析成像与建模；基于线性化正问题进行的叠前深度偏移成像。

第一、子波处理环节。不同炮检点的激发与接收子波的一致性，且期望地震子波是延续最短的、波形简单的、仅仅是主震源激发产生的。因此，鬼波压制方法(POOLE, 2013; 王永强等, 2021)（基于对鬼波形成与传播进行预测的压制方法；基于缆形变化（陷波点变化）的压制方法）、Designature 方法(Poole, 2015)、Debubble 方法(Parkes, 1984)构成了子波预处理的三项关键技术。

第二、水体相关多次波压制环节(孙维蔷等, 2016; 徐鹏等, 2019)。海面+水体+海底及附近介质构成的系统是海上地震勘探的最主要的干扰噪声源，尤其在浅水油气探区。我们认为：水体速度模型建立、基于水体模型的水体相关多次波编码预测、最佳滤波器设计及水体相关多次波减去是模型驱动压制水体相关多次波的合理方法。目前国内外已经发展了处理水体相关多次波的各种方法技术。我们提出的单程波方程+双程波方程+图像配准进行基于模型的水体相关多次波预测及最佳减去应该是更合理的技术方案。

第三、自由表面相关多次波压制环节。应该知道：是地下介质存在强反射面时，长程自由表面多次波才成为必须重视的干扰波。深层强反射面的长程自由表面多次波由于传播时间长可能超过了记录时间对浅中层反射面成像的影响也不是很严重。因此，机器学习（ML）算法识别浅中层强反射面，构建包含浅中层强反射面的速度模型，开展基于模型的波动理论特征反射层相关多次波预测与减去方法是自由表面相关多次波压制的主要技术方向。另外，我们认为：一次波+自由表面相关多次波联合成像、包含自由表面相关多次波的 FWI 成像可能会是今后海洋/OBN 数据成像处理的重要技术发展方向。

SRME（Surface-Related Multiple Elimination）这类数据驱动的压制自由表面相关多次波的方法，受数据观测方式的制约太大(Dragoset, 2010)，在基于模型和波动方程的 FWI 成像逐渐普及应用的情况下，应该逐渐被基于模型（或模型驱动）的波动理论（特征反射层相关）多次波预测与减去方法所取代(Wang, 2011; Spadavecchia, 2013; Wiggins, 1988)。

第四、其他噪声压制环节。涌浪噪声是海洋勘探的典型噪声、缆本身引起的噪声、外源噪声等这些都是典型的去噪方法技术解决的问题。

第五、数据规则化环节。无论是海底 Node 或是海面炮点都很难保证规则布设。拖缆漂移更是无法避

免的。假设炮检点定位准确的情况下，不规则数据对成像质量的影响是明显的。即便实际地震数据处理中高维数据规则化的计算代价比较大，地震数据的规则化也是不可或缺的。基于 Fourier 变换算子的抗假频数据规则化方法(Xu, 2005)和线性 Radon 谱约束的稀疏反演数据规则化方法是比较典型的、常用的方法(王华忠等, 2016)。基于张量分解的高维数据规则化方法(Zhang, 2022)应该是要发展完善的方法。基于神经网络的深度学习数据规则化方法值得关注，但很难会成为主流的数据规则化方法。另外，值得提及的是：混叠采集数据的解混叠方法，在震源激发时间随机编码的条件下，解混叠方法本质上被转化为在伪解码数据域进行线性信号预测去噪问题。因此，高维数据空间中线性噪声压制、高维数据空间中数据规则化、高维数据空间中解混叠本质上是一个问题，即高维数据中所包含的线性信号的建模预测问题。

第六、精确的速度建模环节。OBN 数据最大的特点是宽方位。提出 OBN 数据采集方法技术的目的主要就是利用宽方位的、更均匀的照明，首先提高复杂介质变化（也可以说是复杂构造变化）情形下偏移速度估计与建模的精度。在速度建模这方面，与陆上油气地震勘探关注点还是有所不同，除了宽方位观测外，陆上还通过节点高密度采集技术的实施，考虑无假频采集到地表相关的低视速度噪声，更好地压制掉这些强噪声，从而提高信噪比满足偏移速度建模的要求。宽方位长偏移距观测提供的初至波走时是提高速度建模精度的重要信息。我们认为：宽方位长偏移距观测数据中的初至波识别（走时检测）+初至波 FWI 是 OBN 数据偏移速度建模必须要开展的工作；构造约束下的成像道集 RMO 射线层析速度建模是必须具有的反射波层析速度建模方法技术，目前中深层速度建模主要还是依赖于这种技术手段。反射波 FWI 还没有证明其在速度建模中的真正效力。我们认为：对反射波 FWI 进行梯度分解，然后对层析梯度项和偏移梯度项分别引入有效的先验约束，是反射波 FWI 逐渐实用化的合理途径。因此，我们提出了特征反射层导引下的特征反射层 FWI 方法及具体的实现方案(吴成梁, 2022)。

第七、叠前深度偏移成像环节。众所周知，OBN 数据观测的是至少一阶自由表面多次波（下行波）和上行波。已知当前的叠前深度偏移成像仅仅对一次反（绕/散）射波进行。因此，OBN 数据预处理必然存在这样一个特殊点：即首先必须进行上、下行波分解(Schalkwijk, 2003)，然后分别对这两个波场进行满足线性化正问题 $\delta\mathbf{d}^{cal}=L\left(\mathbf{m}_{\text{initial}}\right)\delta\mathbf{m}$ 要求的预处理。OBN 数据观测的另一个特殊点是炮检点在分布不同的面上。因此，OBN数据叠前偏移成像方法必然有：针对下行波场的偏移成像方法（镜像偏移方法）(Ronen, 2005)；针对炮检点位于不同面上的 PSTM/PSDM 方法。很显然，镜像偏移方法针对的也是炮检点位于不同面上的叠前数据。幸运地是，各种 PSDM 方法适应炮检点不在同一面上是比较容易的。至于各向异性介质 PSDM、吸收衰减介质 PSDM、以及各向异性+吸收衰减介质 PSDM，甚至弹性波 PSDM 都不是 OBN 数据地震波成像所特有的。大量文献已经研究过这些通用的偏移成像方法。线性化的反演成像方法（LS_RTM 或 LS_PSDM）最近十几年也得到了深入的研究。

第八、基于成像道集的成像后处理环节。首先应该说明，由于地震波成像是基于线性模型 $\delta\mathbf{d}^{cal}=L\left(\mathbf{m}_{\text{initial}}\right)\delta\mathbf{m}$ 的，叠前数据空间与叠前成像空间是等价的。而且，保真和高分辨成像处理的要点是：对来自地下同一点的、不同炮检对的地震子波进行同相位叠加。影响同相位保真叠加的各种因素，既可以在叠前数据域进行处理，也可以在叠前成像域进行处理。事实上，在叠前成像域处理时，由于反射子波已经放在成像点处，对于反射子波不一致、存在噪声、照明不均匀、速度不正确引起的道集不拉平等的各种处理方法的物理意义明确、保真和高分辨成像的目标更有针对性。众所周知：保真和高分辨且满足 AVA 关系的方位角度反射系数或零角度反射系数依然是油藏描述的最基础数据，因此有必要给出期望的成像道集的定义。期望成像道集应该具有如下特点：方位角度或角度方位排放的、零相位子波同相轴体现的、零相位子波中心位置与反射界面深度一致且中心幅值与地下界面真实反射系数一致，另外，各方位角度上零相位子波波形一致、频带一致。若产生了期望的成像道集，就可以认为得到了保真成像结果。实际数据情形下，方位角度共成像点道集与上述期望的共成像点道集相差甚大。这是不能做到保真高分辨成像的根本原因。可以看出，叠前数据空间中与叠前成像空间中的各种处理的本质目的就是为了得到保真和高分辨的成像结果。我们把这样的成像结果定义为：无旁瓣或弱旁瓣的、定量的反射系数。可以看出：成像道集中存在剩余深度差（时间差）、剩余相位差（子波形态不一致）、各方位各角度存在不符合 AVA 关系的振幅变化（包括（有效）照明角度多少也可能非常不一致）是成像道集后处理要解决的三个主要问题。针对性地，我们提出：最佳照明优选振幅保真高分辨叠加成像方法技术(Wu, 2023)。另外，偏移成像得到的带限反射系数是不定量的，即它与地下介质真反射系数即便保真成像情况下量级也是不一致的。做不到定量成像的具体影响因素很多，不再展开讨论。对此，必须引入井数据约束，通过控制地震波成像剖面上标志层反射系数的量级，从而达到偏移成像得到的带限反射系数基本定量化的目的。在此环节中，还可以继续用

各种合理的方法，压制反射系数的旁瓣，最后得到期望的无旁瓣或弱旁瓣、定量的反射系数(Sheng, 2023)。如果再进一步，就可以把背景阻抗与上述定量的无旁瓣或弱旁瓣反射系数融合成宽带波阻抗(Song, 2019; Lei, 2023)，真正实现李庆忠院士(李庆忠, 1993)提出的**波阻抗反演是高分辨率资料处理的最终表达形式**的目标。

据此，我们给出图 8 所示的 OBN 数据地震波成像处理流程。

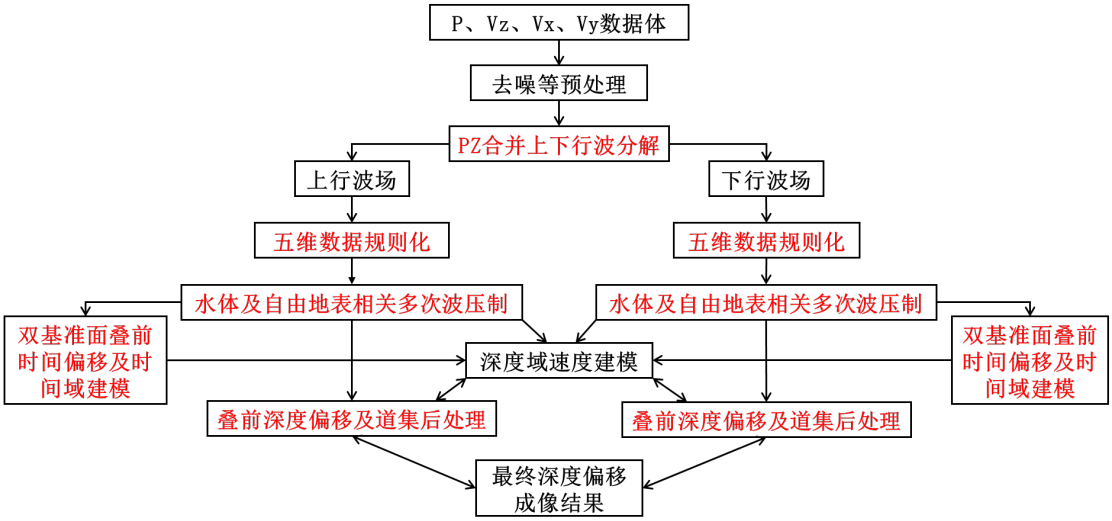


图 8 OBN 数据地震波成像处理流程

关于 OBN 数据的多波成像处理问题，本质上就是 P-P 波和 P-SV 波成像处理，到目前为止，并没有见到能有效地贡献于提升油藏描述精度的成像结果。我们认为：出现这样的结果，主要不是由 P-SV 波成像处理方法导致的，而是因为数据中 P-SV 波的信噪比太低。实质上，这还是地震波传播的物理机制决定的，纵波在实际的孔隙流体介质中为什么不能转换成较强 SV 波，为什么只有局部区域能产生有效的 SV 波而大多数情形下没有有效的 SV 波决定的。如何有效地利用 OBN 数据中的 P-SV 波，还是要继续探索的议题。我们不建议用弹性波 FWI 这类更高端的成像算法进行针对 OBN 数据的矢量波反演成像，而是重点研究波传播的实际物理机制，弄清楚实际数据是否支撑做弹性波 FWI。弹性波方程模拟数据下的弹性波成像能得到很好的成像结果，但这不过是个 Inversion Crime（反演骗局）(Kaipio, 2015)。任何有效的地震波反演成像技术，一定是奠基在实测数据能有效支持的基础上的，理论上还是成像所用的波动方程数值模拟结果与实际波传播结果之间存在高度的一致性。

6 结论与讨论

海上油气勘探的主要矛盾依然是地震数据采集和地震波成像不能满足油气藏识别、描述与评价的精度要求。“两宽一高”地震数据采集方式和以 FWI/LS_RTM 为代表的高精度地震波成像技术是解决海上油气勘探的主要矛盾的、公认的有效方法技术系列。OBN 地震数据采集是基本上得到一致承认的海上“两宽一高”地震数据采集技术。OBN 数据的最大特点是：包含了至少一阶自由地表相关多次波；检波点在海底，与炮点明显地不在一个面上；OBN 数目相对少、纵横向网格间距较大；包含了四分量观测数据。针对复杂构造成像的宽方位、长偏移距观测尽管是开展 OBN 数据采集的主要目的，也是它的典型特点，但并非是 OBN 数据采集所特有。正是因为 OBN 数据观测方式的独特性，才使得 OBN 数据的成像处理有了一定的特殊性。尽管如此，客观地讲，OBN 数据绝大部分的成像处理方法与拖缆和陆上地震数据处理方法差异并不大。另外，应该指出：由于目前 OBN 数目相对少导致纵横向采样间距较大，即便使用了高维数据插值（规则化）技术和上下行波场成像技术，也很难弥补 OBN 空间采样不足带来的成像质量降低问题，有必要发展稀疏 OBN 节点与拖缆数据结合的成像处理方法。

根据 OBN 数据的特点，我们提出：1)、抗假频的数据规则化方法与技术是 OBN 数据处理的核心技术之一；2)、压制鬼波和水体相关多次波，以及来自深层强反射的自由表面相关多次波是 OBN 数据成像处理的第二项核心技术；3)、炮检不同面的 PSDM 方法（也包括 PSTM 方法）、镜像 PSDM 方法是必须具备的方法技术;4)、利用宽方位数据的速度建模、方位角度道集的生成及宽带反射系数成像是更为核心的问题；

5)、发展水检和陆检（垂直分量）合并的上下行波分解技术、水陆和陆检（垂直分量）结合的横波压制技术和三分量陆检数据的成像处理方法技术等是必要的。

针对保真高分辨成像，我们提出：叠前数据空间与叠前成像空间是等价的，保真和高分辨成像处理的要点是对来自地下同一点的、不同炮检对的地震子波进行同相位叠加。影响同相位保真叠加的各种因素，既可以在叠前数据域进行处理，也可以在叠前成像域进行处理。提出期望成像道集应该具有如下特点：方位角度排放的、零相位子波同相轴体现的、零相位子波中心位置与反射界面深度一致的且中心幅值与地下界面真实反射系数一致的，另外，各方位角度上零相位子波波形一致、频带一致。提出保真高分辨成像结果应该是：无旁瓣或弱旁瓣的、定量的反射系数。指出应该把保真高分辨成像推进到宽带波阻抗成像。

我们认为：海洋油气勘探的真正特殊问题主要还是由反射系数接近 1 的海面及海面+海水+海底及附近介质构成的介质系统所引起的、由于地下介质中存在强反射面引起的长程自由表面多次波所引起的。自由表面（海面）相关多次波的预测、压制或成像处理构成了海洋地震数据处理的真正特殊问题。其他方面的问题，要么与陆上地震数据处理一致，要么没有明显的特殊性。

据此我们提出了：水体模型建模+水体相关多次波压制+自由表面（海面）相关（长程）多次波压制+初至波 FWI（各向异性）速度建模+成像道集（各向异性）层析速度建模+成像道集为中心的高保真高分辨方位角度带限反射系数成像（最好推进到宽带波阻抗成像）的海洋地震数据成像处理技术组合。

同时我们认为：海洋/OBN 地震勘探（也包括陆上地震勘探），多次波主要由特征反射界面引起的，所谓特征反射层包括海面、海底面及地下介质强反射面。据此我们提出模型驱动波动理论特征反射层相关多次波预测与压制的技术路线。基本逻辑步骤包括：1)、构建背景速度+特征反射层组成的速度模型；2)、波动理论特征反射层相关的多次波模拟预测；3)、多次波模拟预测波场与实测多次波波场的配准；4)、最佳匹配滤波器设计及多次波减去。在 FWI+LS_RTM 已经成为高精度地震波成像方法技术核心的时代，基于模型的（模型驱动的）波动方程特征反射层相关多次波预测与压制方法技术应该成为海洋/OBN 地震数据处理的核心方法技术。基于数据驱动的（如 SRME 之类的）方法应该逐渐让位于基于模型驱动的多次波预测与压制方法。机器学习算法被引入帮助建立包含特征反射层的速度模型；图像模板之间的时差测量方法（图像配准方法）被借用来检测由于包含特征反射层的速度模型的不准确引起的预测多次波同相轴与实测多次波同相轴之间的时差，帮助设计最佳的匹配滤波器进而实现最佳的多次波减去。这样的理念与方法技术也可以用在无论海洋或陆上地震勘探中的层间多次波压制过程中。

关于OBN数据的多波（即矢量波场）成像处理问题，我们认为：经过几十年的探索实践多波勘探没有得到大规模应用、没有产生明显勘探效益的根本原因，本质上不是地震波成像理论与方法出了问题，而是实际观测的多波地震波场中的波现象（主要是 P_SV 波）与地震波传播及模拟理论不匹配导致了多波成像结果达不到预期。因此，我们不建议发展更高端的矢量波场成像算法，而是重点研究实际观测的多波地震波场中的波现象与地震波传播及模拟理论不匹配的物理根源到底是什么。

参考文献

[1] 王华忠,盛燊. 走向精确地震勘探的道路[J]. 石油物探, 2021, 60(05):693-708.

[2] 苟量,王克斌,王伟,王永明等. 海底节点地震资料处理技术及应用[M]. 石油工业出版社, 2021.

[3] DHELIE P E, DANIELSEN V, LIE J, et al. Improving seismic imaging in the Barents Sea by source-over-cable acquisition[C]. SEG Technical Program Expanded, 2018.

[4] MOLDOVEANU N, KAPOOR J. Quad coil towed streamer marine acquisition[C]. SEG Technical Program Expanded, 2015.

[5] BRIZARD T, GRIMSDALE J, SILIQI R, et al. Will new marine seismic acquisition based on autonomous immersed receiver nodes confirm sustainable production of unrivaled field data quality? [C]. SEG Technical Program Expanded, 2016.

[6] HEGNAL S, KLÜVER T, LIMA J. Continuous wavefields method – a new marine seismic acquisition and processing approach[C]. 81st EAGE Conference and Exhibition, 2019.

[7] HOFFE B H, LINES L R, CARY P W. Applications of OBC recording[J]. Leading Edge, 2000, 19(4):382-391.

[8] THOMPSON M, ARNTSEN B, AMUNDSEN L. Full azimuth imaging through consistent application of ocean bottom seismic[C]. SEG Technical Program Expanded Abstracts, 2007.

[9] BATHELLIER E, HAUMONTÉ L. The Kietta Start-Up: From design to acquisition of a 3D marine full-azimuth, full-offset seismic survey[C]. 81st EAGE Conference and Exhibition 2019.

[10]HOLLOWAY A, GRANT D, WATTS G, et al. The future of deepwater ocean bottom seismic - Are flying nodes the next big step? [C]. SEG Technical Program Expanded, 2015.
[11]PRAMIK, B., Marine Vibroseis: Shaking up the industry[J]. First Break, V31, November, 2013.
[12]COX, MIKE. Static corrections for seismic reflection surveys[M]. 1999.
[13]王华忠,冯波,王雄文,等. 压缩感知及其在地震勘探中的应用[J]. 石油物探, 2016, 55 (04): 467-474.
[14]TENGHAMN R, LONG A, GEO-SERVICES P, et al. Improved operational efficiency using spread spectrum sweeps for marine vibrators[C]. SEG Technical Program Expanded, 2018.
[15]CHEN Y, FOMEL S, HU J. Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization[J]. Geophysics, 2014, 79(5):V179-V189.
[16]宋家文,李培明,王文闯,等. 基于稀疏反演的高效混采数据分离方法[J]. 石油地球物理勘探, 2019, 54 (02): 268-273+235.
[17]SCHALKWIJK K M, WAPENAAR C P A, VERSCHUUR D J. Adaptive decomposition of multicomponent ocean-bottom seismic data into downgoing and upgoing P- and S-waves[J]. Geophysics, 2003, 68(3):1091-1102.
[18]Ronen S, Comeaux L, Miao XG. Imaging downgoing waves from ocean bottom stations[C]. SEG Technical Program Expanded, 2005.
[19]王华忠,冯波,王雄文,等. 特征波反演成像理论框架[J]. 石油物探, 2017, 56(01):38-49.
[20]BERKHOUT A J, VERSCHUUR D J. Estimation of multiple scattering by iterative inversion, Part I: Theoretical considerations[J]. Geophysics, 1997, 62(5):1586-1595.
[21]VERSCHUUR D J, BERKHOUT A J. Estimation of multiple scattering by iterative inversion, Part II: Practical aspects and examples[J]. Geophysics, 1997, 62(5):1596-1611.
[22]VERSCHUUR D J, BERKHOUT A J, WAPENAAR C P A. Adaptive surface-related multiple elimination[J]. Geophysics, 1992, 57(9):1166-1177.
[23]BERKHOUT A J. A unified approach to acoustical reflection imaging, Part I: The forward model[J]. Journal of the Acoustical Society of America, 1993, 93(4):2017-2023.
[24]GABRIEL A, LOPEZ, VERSCHUUR D J. Closed-loop surface-related multiple elimination and its application to simultaneous data reconstruction[J]. Geophysics, 2015, 80(6):V189－V199.
[25]WEGLEIN A B, et al. An inverse-scattering series method for attenuating multiples in seismic reflection data[J]. Geophysics, 1997, 62(6):1975-1989.
[26]IKELLE L T, AMUNDSEN L. Introduction to petroleum Seismology[M]. SEG, 2005.
[27]AMUNDSEN L. Elimination of free-surface related multiples without need of the source wavelet[J]. Geophysics, 1999, 66(1):327-341.
[28]MAJDANSKI M, KOSTOV C, KRAGH E, et al. Attenuation of free-surface multiples by up/down deconvolution for marine towed-streamer data[J]. Geophysics, 2011, 76(6):V129–V138.
[29]JAKUBOWICZ H. Wave equation prediction and removal of interbed multiples[C]. SEG Technical Program Expanded, 1998.
[30]TARANTOLA A. Inverse problem theory and methods for model parameter estimation[M]. SIAM, 2005.
[31]王华忠,吴成梁,盛燊,等. 针对岩性储层的定量地震波成像[J]. 石油物探, 2023, 62(01):1-30.
[32]POOLE G. Pre-migration receiver de-ghosting and re-datuming for variable depth streamer data[C]. SEG Technical Program Expanded Abstracts, 2013.
[33]王永强,王华忠,孙维蔷等. 基于编码解码思想的鬼波预测与压制方法[J]. 地球物理学报, 2021, 64 (04): 1389-1400.
[34]POOLE G, Cooper J, King S, et al. 3D source designature using source-receiver symmetry in the shot tau-px-py domain[C]. 77th EAGE Conference and Exhibition 2015.
[35]PARKES G E, ZIOLKOWSKI A, HATTON L, et al. The signature of an air gun array: Computation from near‐field measurements including interactions - Practical considerations[J]. Geophysics, 1984, 49(2):105-111.
[36]孙维蔷,王华忠. 基于平面波编码的水体相关多次波压制方法研究[J]. 石油物探, 2016, 55 (04): 516-523.
[37]徐鹏,刘亚辉,王华忠. 编码与解码框架下的局部平面波域浅水多次波压制方法[J]. 石油物探, 2019, 58 (03): 346-355+403.
[38]DRAGOSET B, VERSCHUUR E, MOORE I, et al. A perspective on 3D surface-related multiple elimination[J]. Geophysics, 2010, 75(5): 75A245–75A261.
[39]WANG P. Model-based water-layer demultiple[C]. SEG Technical Program Expanded, 2011.
[40]SPADAVECCHIA E, LIPARI V, BIENATI N, et al. Water‐bottom multiple attenuation by Kirchhoff extrapolation[J]. Geophysical Prospecting, 2013, 61(4):725‐734.
[41]WIGGINS J W. Attenuation of complex water-bottom multiples by wave-equation-based prediction and subtraction[J]. Geophysics, 1988, 53(12):1527-1539.

[42]Xu S, Zhang Y, Pham D, et al. Antileakage Fourier transform for seismic data regularization[J]. Geophysics, 2005, 70(4): V87–V95.
[43]ZHANG L, WANG H, WANG B. Vandermonde constrained CANDECOMP/PARAFAC tensor decomposition for high-dimensional seismic data reconstruction[J]. Geophysics, 2022, 87(5): V533–V544.
[44]吴成梁. 特征反射波波动理论走时层析速度建模方法研究[D]. 上海: 同济大学, 2022.
[45]WU C, WANG H, FENG B, et al. Stacking common-image gathers via the flatten-illumination-similarity method. Third International Meeting for Applied Geoscience & Energy. Society of Exploration Geophysicists and American Association of Petroleum Geologists, 2023,1-5.
[46]SHENG S, FENG B, WANG H, et al. Expectation-maximization framework-based image-domain least-squares migration[J]. Geophysics, 2023, 88(4):R545–R558.
[47]Song G, WANG H. Seismic absolute acoustic impedance inversion with L1 norm reflectivity constraint and combined first- and second-order total variation regularizations[J]. Journal of Geophysics and Engineering, 2019, 16(4):773–788
[48]LEI T, WANG H, FENG B. Broadband acoustic impedance building via the fusion of sparsely promoted reflectivity and background acoustic impedance[J]. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 2023, 61:1–9.
[49]李庆忠. 走向精确勘探的道路[M]. 石油工业出版社, 1993.
[50]KAIPIO J, SOMERSALO E. Statistical and computational inverse problems[M]. Springer, 2015.

迭代法解线性系统的基本思想

王华忠，许荣伟

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】地震数据处理和地震波反演成像中的很多问题最后都归结为无约束和有约束条件下的线性系统的求解。非线性系统的求解也是在局部线性假设下通过分步地解一个线性问题最后逼近非线性问题的解。在勘探地震中遇到的基本上都是（超）大规模的线性系统问题的求解，迭代法是基本的解法。另外，这些线性系统往往是欠定的或不满秩的，迭代法寻找一定约束下线性系统的近似解是常见的场景。本文重点讨论迭代法求解线性系统的基本思想，主要简述对迭代法求解线性系统的基本思想的看法，而且重点是投影法求解线性系统的基本思想的看法。

【关键词】地震数据处理，地震波反演成像，线性系统，迭代法，线性系统的求解

1 引言

勘探地震中，地震数据处理和地震波反演成像，逐渐统一在 Bayes 估计的理论框架下。高斯平稳、线性系统是基本的假设。在这样的理论框架下，非常多的问题最后都归结为线性系统的求解。

勘探地震中的实际问题归结出的线性系统，往往是规模巨大的、欠定的。这样的问题必须要用迭代法求解。由于这样的线性系统不存在唯一解，先验约束的引入是必须的。因此，这样的线性系统的求解又常常被提成一个约束优化问题。

迭代法是数值算法的基本模式。线性系统的迭代解是迭代法的基本研究对象。事实上，非线性系统的解法，包括各种神经网络算法、分类算法，也被归结为用迭代法求解。构造迭代法的基本逻辑是压缩映射与不动点原理，这应该是最高层次的逻辑。其他的研究工作是在此逻辑下构造出可行的迭代算法。

典型的迭代算法包括：1、牛顿法或梯度导引类的算法，以目标泛函的下降作为收敛的判断。最速下降法、共轭梯度法都可归结为此类梯度导引类的算法。2、Jacobi、Gauss-Seidel 算法、SOR 算法及 ART 算法应归结为（凸集）投影算法。这两类典型迭代算法又都可认为是在不动点原理与压缩映射逻辑下构造迭代算法的具体实践。

迭代法的本质可以认为是投影法。把线性系统的解投影在一个基函数张成的解空间中，不断地膨胀这个解空间，逼近线性系统的解。

本文重点阐述我们对投影法求解线性系统的基本思想的看法。

2 基本投影法的本质逻辑

投影算子 P：定义为 C^n 或者 R^n 空间的任意的线性组合，且存在 $P=P^2$，即 P 是幂等的。P 的零空间与值域的关系为：

$$\begin{cases} Ker(P)=Ran(I-P)=Ran(Q) \\ Q=I-P \end{cases}, \tag{1}$$

其中，$Ker(P)$ 和 $Ran(P)$ 两个子空间的交集为空集。若 $\boldsymbol{x}\in Ran(P)$则$\boldsymbol{x}=P\boldsymbol{x}$；若$\boldsymbol{x}\in Ker(P)$也成立，必有$\boldsymbol{x}=P\boldsymbol{x}=0$，即 $C^n=Ker(P)\oplus Ran(P)$，C^n 被分为两个不相交的子空间，可记为 $Ran(P)=\mathcal{K},Ker(P)=\mathcal{S}$。

若 P 把 $\boldsymbol{x}\in C^n$ 的任一向量投影成 $\boldsymbol{x}_1,\boldsymbol{x}_1\in\mathcal{K}$，则必存在 $\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2,\boldsymbol{x}_2\in\mathcal{S}$，即：$\begin{cases} P\boldsymbol{x}\in\mathcal{K}\Rightarrow \\ \boldsymbol{x}-P\boldsymbol{x}\in\mathcal{S}\Rightarrow \end{cases}\begin{cases} P\boldsymbol{x}=\boldsymbol{x}_1\in\mathcal{K} \\ Q\boldsymbol{x}=\boldsymbol{x}_2\in\mathcal{S} \end{cases}$，称线性映射 P 把 $\boldsymbol{x}$ 投影到 $\mathcal{K}$，且平行于子空间 $\mathcal{S}$ 把 $\boldsymbol{x}$ 投影到 $\mathcal{K}$。

进一步地，若 P 的秩为 m，则 $I-P=Q$ 的秩为 $n-m$，可通过 $\mathcal{S}$ 的正交补 $\mathcal{L}$ 来定义 S，$\mathcal{S}=\mathcal{L}^{\perp}$或$\mathcal{L}=\mathcal{S}^{\perp}$，即：$\mathcal{L}=Ran(P)$，因此，

$$\boldsymbol{u}=P\boldsymbol{x}, \tag{2a}$$

$$\begin{cases}\boldsymbol{u}\in\mathcal{K}\\ \boldsymbol{x}-\boldsymbol{u}\perp\mathcal{L}\end{cases}, \tag{2b}$$

其中，$\boldsymbol{u}$ 是唯一的。

把上述投影逻辑用于解线性方程组：$A\boldsymbol{x}=\boldsymbol{b}$。假设存在两个 m 维子空间 $\mathcal{K}$ 和 $\mathcal{L}$，解方程组变成：找 $\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}$，满足$\boldsymbol{b}-A\boldsymbol{x}\perp\mathcal{L}$。

令 $\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{\delta}$，则有：$\boldsymbol{b}-A(\boldsymbol{x}_0+\boldsymbol{\delta})=\boldsymbol{b}-A\boldsymbol{x}_0-A\boldsymbol{\delta}=\boldsymbol{r}_0-A\boldsymbol{\delta}$。由（2b）可知：

$$\begin{cases}\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{\delta},\boldsymbol{\delta}\in\mathcal{K}\\ (\boldsymbol{r}_0-A\boldsymbol{\delta},\boldsymbol{w})=0,\forall\boldsymbol{w}\in\mathcal{L}\end{cases}, \tag{3}$$

进一步地，设$\mathcal{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,...,\boldsymbol{v}_m]$构成 $\mathcal{K}$的基函数族，$\mathcal{W}=[\boldsymbol{w}_1,\boldsymbol{w}_2,...,\boldsymbol{w}_m]$构成 $\mathcal{L}$ 的基函数族。则有，

$$\begin{aligned}&\boldsymbol{x}=\boldsymbol{x}_0+\mathcal{V}\boldsymbol{y}\\ &\boldsymbol{b}=A\boldsymbol{x}=A\boldsymbol{x}_0+A\mathcal{V}\boldsymbol{y}\\ &\boldsymbol{b}-A\boldsymbol{x}=\boldsymbol{b}-A\boldsymbol{x}_0-A\mathcal{V}\boldsymbol{y}=\boldsymbol{r}_0-A\mathcal{V}\boldsymbol{y},\\ &(\boldsymbol{r}_0-A\mathcal{V}\boldsymbol{y},\boldsymbol{w}_i)=0,\quad i=1,2,...,m\\ &\boldsymbol{w}_i^{T}(\boldsymbol{r}_0-A\mathcal{V}\boldsymbol{y})=0,\quad i=1,2,...,m\end{aligned} \tag{4}$$

把（4）改成矩阵形式为：$\mathcal{W}^{T}A\mathcal{V}\boldsymbol{y}=W^{T}\boldsymbol{r}_0$，最后有：

$$\tilde{\boldsymbol{x}}=\boldsymbol{x}_0+\mathcal{V}(\mathcal{W}^{T}A\mathcal{V})^{-1}\mathcal{W}^{T}\boldsymbol{r}_0, \tag{5}$$

其中，$(\boldsymbol{r}_0-A\mathcal{V}\boldsymbol{y},\boldsymbol{w}_i)=0,\quad i=1,2,...,m$ 是正交原理。

由 $\boldsymbol{w}_i^{T}(\boldsymbol{r}_0-A\mathcal{V}\boldsymbol{y})=0,\quad i=1,2,...,m$ 可得到 $\boldsymbol{y}=\left(\mathcal{W}^{T}A\mathcal{V}\right)^{-1}\mathcal{W}^{T}\boldsymbol{r}_0$。

因此，由 $\boldsymbol{x}=\boldsymbol{x}_0+\mathcal{V}\boldsymbol{y}$ 可以写出：

$$\begin{aligned}\boldsymbol{x}^{k+1}&=\boldsymbol{x}^{k}+\mathcal{V}(\mathcal{W}^{T}A\mathcal{V})^{-1}\mathcal{W}^{T}\boldsymbol{r}^{k}\\ &=\boldsymbol{x}^{k}+\mathcal{V}(\mathcal{W}^{T}A\mathcal{V})^{-1}\mathcal{W}^{T}(\boldsymbol{b}-A\boldsymbol{x}^{k})\end{aligned}。 \tag{6}$$

应该注意到：$\mathcal{K}$和 $\mathcal{L}$ 的不同选择，可以得到具体的迭代格式。

这样的框架奠定了所有迭代法的理论基础。

最关键的思想体现在式$\begin{cases}P\boldsymbol{x}\in\mathcal{K}\\ \boldsymbol{x}-P\boldsymbol{x}\perp\mathcal{L}\end{cases}$中，其中，投影算子 $P=\mathcal{V}(\mathcal{W}^{T}A\mathcal{V})^{-1}\mathcal{W}^{T}$。其基本思想是：投影算子 $P\boldsymbol{x}$ 把 $\boldsymbol{x}$ 投影到 $\mathcal{K}$空间中，近似地表示 $\boldsymbol{x}$，表达残差 $\boldsymbol{x}-P\boldsymbol{x}$ 要垂直于 $\mathcal{L}$ 空间。$\mathcal{K}$空间与 $\mathcal{L}$ 空间是不同坐标系张成的同一个子空间，也可以是不同的子空间，但两个空间应该是同一个空间中的不同子空间。几何解释见图 1。

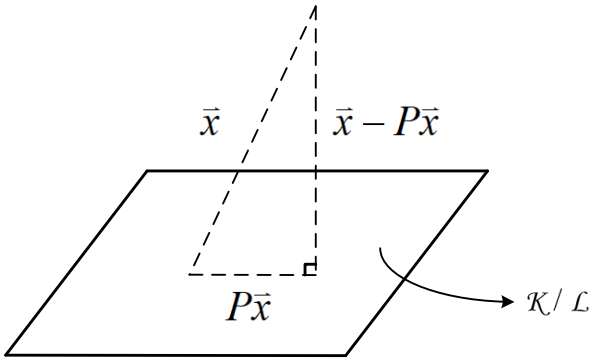


图 1 投影法思想的几何图示

近似表达通过迭代过程中不断调整 $\mathcal{K}$ 空间坐标系及 $\mathcal{L}$ 空间的坐标系实现，但 2（b）或 3 式的关系始终要满足。

实际上，是通过 $\begin{cases}\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{\delta},\boldsymbol{\delta}\in\mathcal{K}\\(\boldsymbol{r}_0-A\boldsymbol{\delta},\boldsymbol{w})=0,\forall\boldsymbol{w}\in\mathcal{L}\end{cases}$，在每个迭代步中，空间 $\mathcal{K}$ 和 $\mathcal{L}$ 不断调整。据此，构建了各种各样的迭代格式。

3 交替投影法的基本思想

交替投影法是一般迭代法的变种。首先，（凸集）（交替）投影也是抽象到一定程度的理论框架，但和不动点原理与压缩映射相比应低一个层次。它首先试图包容基于一般意义下的投影法来构造迭代公式的各种解法。

更具体地，它从凸集投影的角度，为超定、欠定线性系统的求解提供一个构造迭代算法的理论框架。但是，（凸集）（交替）投影仅提供思想逻辑，不负责提供具体算法。具体算法是在它的逻辑指导下由具体的问题导引出来的。因此，透彻的理解（凸集）（交替）投影的基本思想或背后的本质逻辑是重要的。

可以用 ART 这个例子来理解（凸集）（交替）投影的基本思想。

首先，有一个线性系统 $A\boldsymbol{x}=\boldsymbol{b},A_{m\times n}\boldsymbol{x}_{n\times 1}=\boldsymbol{b}_{m\times 1}$，可以是 $m>n$，也可以是 $n>m$ 的。然后，线性系统 A 可分解为一系列的行向量 $\boldsymbol{a}_{n\times i}^T,(i=1,2,...,m)$，即 $A=\left[\boldsymbol{a}_{n\times 1}^T,\boldsymbol{a}_{n\times 2}^T,...,\boldsymbol{a}_{n\times m}^T\right]$，可以认为每一个 $\boldsymbol{a}_{n\times i}^T,(i=1,2,...,m)$ 构成一个子空间。

求解线性系统解的问题可以转化成 $\boldsymbol{b}_{m\times 1}$ 向 m 个凸集（子空间）投影，交替投影的过程。

对于 $A^T\boldsymbol{b}$ 或 $A^T(\boldsymbol{b}-A^T\boldsymbol{x}^k)$，尤其是 $A^T(\boldsymbol{b}-A^T\boldsymbol{x}^k)$ 逐行迭代时，可以认为是在做向凸集的交替投影。每个 $\boldsymbol{a}_{n\times i}^T,(i=1,2,...,m)$ 都是一个凸集。

从此角度看，Jacobi、Gauss-Seidel 算法、SOR 算法、ART 算法及牛顿法的每一步都在做向凸集的投影。不过牛顿法的计算掩盖了凸集交替投影的本质。

此处给出一个二维方程组的 Gauss-Seidel 投影的简单示意：

对于方程组 $\begin{cases}2x_1+x_2=11\\3x_1-x_2=4\end{cases}$，每一个等式可以看作一个子空间，因此，该方程组共有两个子空间，如图 2 中黑线实线（PA 投影算子）与黑色虚线（PB 投影算子）所示，两个空间的 Gauss-Seidel 投影格式分别为：

$$\begin{bmatrix}x_1^{k+1}\\x_2^{k+1}\end{bmatrix}=P_A\left(\begin{bmatrix}x_1^{k}\\x_2^{k}\end{bmatrix}\right)=\begin{bmatrix}x_1^{k}\\(11-x_1^{k})\end{bmatrix},$$

$$\begin{bmatrix} x_1^{k+2} \\ x_2^{k+2} \end{bmatrix} = P_B\left(\begin{bmatrix} x_1^{k+1} \\ x_2^{k+1} \end{bmatrix}\right) = \begin{bmatrix} (4+x_2^{k+1})/3 \\ x_2^{k+1} \end{bmatrix},$$

图 2 中红点表示了每一次投影的结果，给定一个初始的解，通过 PA 将其投影到实线子空间中，然后再经过 PB 投影到虚线子空间中，通过不断的交替投影，最后将迭代到收敛点上。

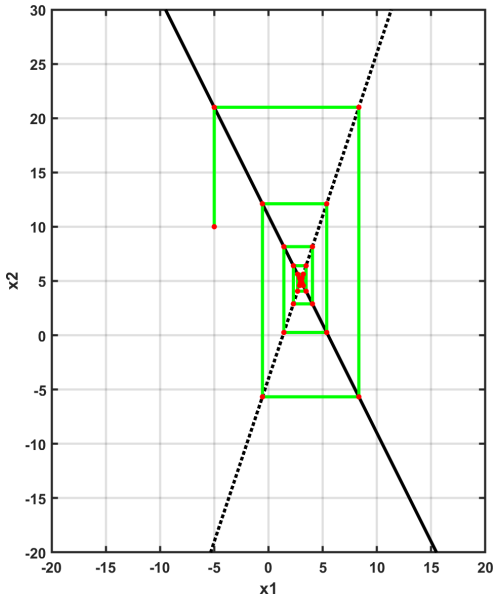


图 2 Gauss-Seidel 交替迭代求解方程组的交替投影法示意

4 关于交替迭代投影的另一种认识

迭代投影的本质思想：在 Hilbert 空间中，存在一个信号 f，这个信号 f 属于子空间 A（信号可以在 A 空间中完全被表达，信号 f 向空间 A 投影为其本身，信号 f 向空间 A 的补空间投影为 0）同时这个信号也属于子空间 B。则我们想求解信号 f，则可以通过不断交替迭代投影来计算 f。$f = P_a P_b P_a P_b f_0$。可以认为，每一次投影都是在舍弃真实信号中不包含的成分。

迭代投影法的基本公式：

$$f^{k+1} = g + Q_a P_b f^k \tag{7}$$

如何理解上式的内在含义？

在 Hilbert 空间中，存在两个子空间$\mathcal{P}_a$，$\mathcal{P}_b$，及其补空间，$\mathcal{P}_a^{\perp}$，$\mathcal{P}_b^{\perp}$。其投影算子分别为P_a，P_b，Q_a，Q_b。

在$\mathcal{P}_b$空间中存在一个信号 f，我们无法在$\mathcal{P}_b$空间中进行观测，只能在$\mathcal{P}_a$空间中进行观测，观测的结果为 g

$$\mathrm{g} = P_a f$$

现在我们想通过观测的 g，来求解信号 f。

信号 f，在向$\mathcal{P}_a$空间投影时，损失了一部分信息，这一部分信息落在$\mathcal{P}_a^{\perp}$空间中，理论上

$$f = P_a f + Q_a f = g + g^{\perp} \tag{8}$$

信号 f 由$\mathcal{P}_a$空间及其补空间共同表示，才可以完全表征信号 f。

但是，由于信号 f 在向$\mathcal{P}_a$空间投影时，损失了 f 在$\mathcal{P}_a$补空间中的信息，我们仅得到了$\mathcal{P}_a$空间中的分量 g。

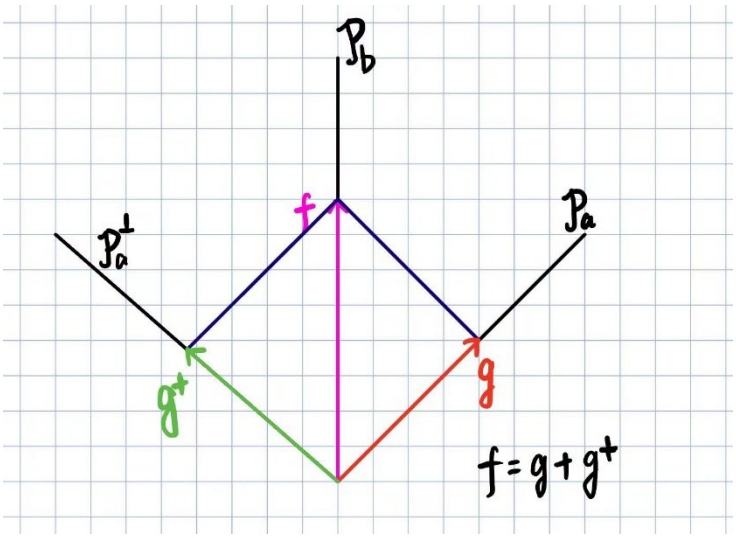


图 3 交替迭代法的基本图示

由于我们仅有$\mathcal{P}_a$空间中的信号 g，因此我们需要计算出 f 在$\mathcal{P}_a$补空间中的分量。如何计算$\mathcal{P}_a$补空间中的分量？

首先我们给定一个初始 $f = f_0$，由于 f 在$\mathcal{P}_b$空间中，但是初始值 f_0 可能不在$\mathcal{P}_b$空间中，因此我们首先要将 f_0 映射到$\mathcal{P}_b$空间中，这一步的主要目的就是舍弃掉 f_0 中的错误分量，因为前提条件是 f_0 是空间$\mathcal{P}_b$中的信号，所以其真解一定是投影到$\mathcal{P}_b$中的分量 $P_b f_0$。

然后我们将 $P_b f_0$ 投影到$\mathcal{P}_a$补空间中，作为真实信号 f 在$\mathcal{P}_a$补空间中的分量

$$g_0^{+} = Q_a P_b f_0 \tag{9}$$

所以

$$f_1 = g + g_0^{+} = g + Q_a P_b f_0 \tag{10}$$

以此类推，则可以得到公式（7）$f_{k+1} = g + Q_a P_b f_k$。

我们认为 f 在$\mathcal{P}_a$空间中的分量 g 是完全准确的，$\mathcal{P}_a$补空间的分量通过不断迭代，也会逐渐接近真实值，最终得到真实的信号 f。

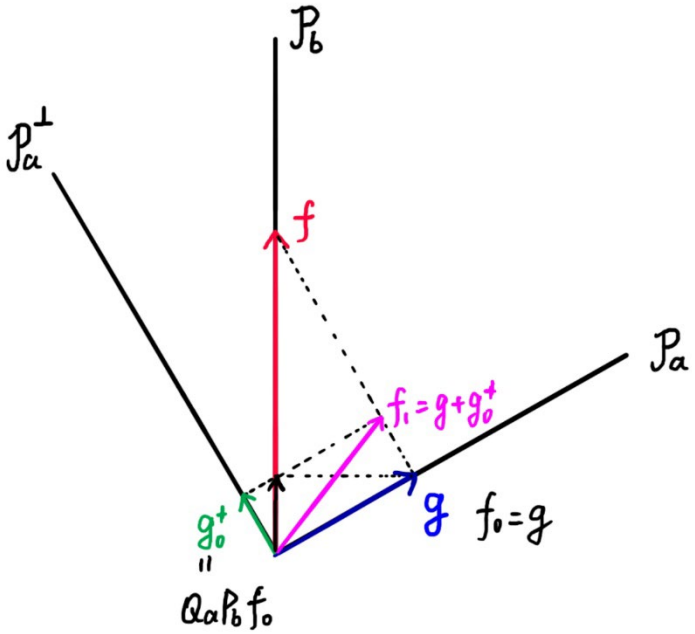


图 4 信号交替投影示意图，**第一次迭代的几何对应关系**

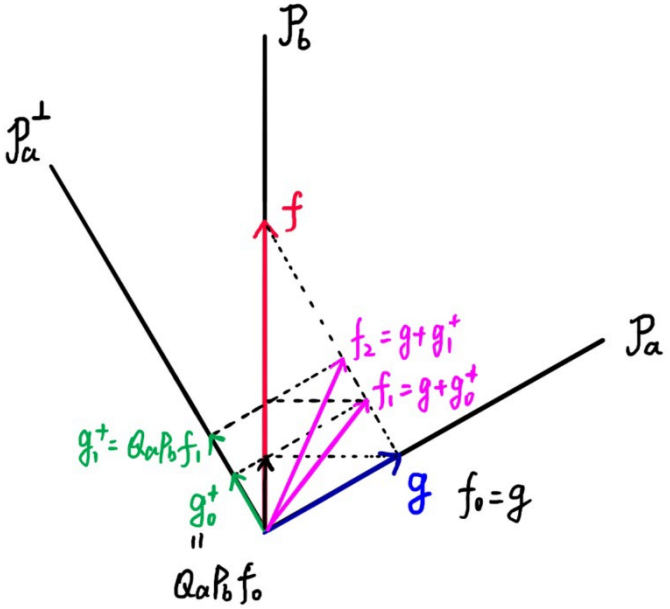


图 5 信号交替投影示意图，**第二次迭代的几何对应关系**

5 关于约束优化问题本质问题认识

1、提出并求解一个最优化问题是人类遇到的最常见，也最有深刻含义的事。人类智能及人工智能的生成最主要的目的是最优地解决生存及各种相关问题。

2、我认为提出反问题/最优化问题的最高形式还是 Bayes 决策。后验概率密度包含了做出最佳推断/决策的所有信息。这个听起来很抽象，但在思想层次上，问题的本质就是这样。

3、MAP 准则建立起了数学分析学家和统计学家解决最优化问题之间的联系。变分原理也是一种转化其它问题为最优化问题，从而解决这些其他问题的思想与方法。Euler 公式的导出等价于找到了变分问题解的必要条件。

4、Bayes 推断的本质就相当于解一个约束优化问题，否则就变成了无约束的 ML 推断问题了。Bayes 推断占尽上风，说明了引入先验信息是求解优化问题的必要措施。

5、ML 推断对应的无约束优化问题的思想，相对而言，是简单的、明晰的，即目标泛函 $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$，$f: \mathbb{R}^n \to \mathbb{R}$，$f(\mathbf{x}) \in C^1$，期望 $f(\mathbf{x})$ 是凸的。

梯度导引（及其变种）类算法是广为熟悉的。

6、最近二十几年来，提目标泛函 $f(\mathbf{x}) \in C^0$，期望它是凸的（也可能非凸，对应 $L_p, 0 \le p < 1$ 的情形），发展出了各种稀疏反演算法。L_0 范数对应一类贪婪算法；L_1 范数对应一类阈值投影算法（LASSO 算法等）。这些算法均应属于投影类算法。

为什么稀疏解一定是好的？一定好于 L2 范数下的光滑解吗？随机反演解会更好吗（更接近真解吗）？

7、约束反演始终是必须的！仅依赖观测数据、正问题、以及逼近误差最小化，反问题/最优化问题的多解性问题得不到解决。

这正是 Bayes 推断战胜或优于 ML 推断的地方。

据此可见，Bayes 推断是约束优化问题的根本思想来源。

尽管数学分析学家可能不这样看，他们也许早就独立地提出约束优化问题，但我还是坚持上述认识。

8、Tikhonov 正则化是 MAP 准则加上 Gauss 分布假设导出的，它是约束优化问题的 Lagrange 乘子法的变种。Tikhonov 正则化问题求解时，λ（超参）是用其他方法确定的，不是 Lagrange 乘子法那样迭代求解的，但这不改变问题背后的本质含义。

9、约束优化问题的求解，与无约束优化问题一样，首先应回答解存在的必要条件是什么。我们现在已知，对于约束优化问题，这个必要条件即 Lagrange 条件。

无约束优化问题存在局部唯一解的条件是很清楚的：$f(\mathbf{x})$ 是局部凸的、$\nabla f(\mathbf{x}) = 0$、$\nabla^2 f(\mathbf{x}) > 0$。$f(\mathbf{x})$ 不可导时，引入次梯度的概念，界定此时存在局部唯一解的必要条件。

对于约束优化问题：

$$\begin{aligned} &\min f(\mathbf{x}): \mathbb{R}^n \to \mathbb{R} \\ &s.t. \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{h} = \{h_1(\mathbf{x}), h_2(\mathbf{x}), \cdots, h_m(\mathbf{x})\}, m < n \end{aligned} \tag{11}$$

约束优化问题解的几何含义应该是：$f(\mathbf{x})$ 是 n 维空间中的曲面，局部凸的；约束项 $\{h_1(\mathbf{x}), h_2(\mathbf{x}), \cdots, h_m(\mathbf{x})\}$ 是 m 个 n 维曲面。隐含地，要求 $f(\mathbf{x})$ 和 $h_1(\mathbf{x}), h_2(\mathbf{x}), \cdots, h_m(\mathbf{x})$ 都是连续的 n 维曲面。

要找的解一定在 m 个 n 维曲面形成的 $n-m$ 维相交曲面上。但此时，还要找到在此 $n-m$ 维相交曲面上令 $f(\mathbf{x})$ 达到最小的点。

这个点在哪里？在

$$\nabla h(\mathbf{x})^T \mathbf{y} = 0 \tag{12a}$$

$$\nabla f\left(\mathbf{x}\right)^{T}\mathbf{y}=0 \tag{12b}$$

成立的地方！

为什么这一定是约束优化问题的必要条件？

首先（12b）式告诉我们：$\mathbf{y}$ 是与梯度方向 $\nabla f\left(\mathbf{x}\right)$ 垂直的 $f\left(\mathbf{x}\right)$ 的切平面上的向量（任一向量）。（12b）式是按**找到在此 $n-m$ 维相交曲面上令 $f\left(\mathbf{x}\right)$ 达到最小的点**的要求导出的。

（12a）式告诉我们：切平面切在什么地方！切在

$$\begin{cases} h_1\left(\mathbf{x}\right)=0 \\ h_2\left(\mathbf{x}\right)=0 \\ \quad\vdots \\ h_m\left(\mathbf{x}\right)=0 \end{cases} \tag{13}$$

成立的地方。

约束优化问题求解的本质含义是：在 n 维空间中沿 $n-m$ 维曲面，在 n 维曲面 $f\left(\mathbf{x}\right)$ 上，找一个点使 $f\left(\mathbf{x}\right)$ 达到极小，即

$$\frac{\partial f\left(\mathbf{x}\right)}{\partial \mathbf{x}\left(\xi\right)}=\nabla f\left(\mathbf{x}\right)^{T}\cdot\dot{\mathbf{x}}\left(\xi\right)=\nabla f\left(\mathbf{x}\right)^{T}\mathbf{y}=0 \tag{14}$$

但上式成立是受约束的，即达到上式成立的点 $\mathbf{x}^{*}$，必在 $n-m$ 维曲面上。在该点 $\mathbf{x}^{*}$ 处，同样地有

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}\left(\xi\right)}=\nabla \mathbf{h}\left(\mathbf{x}\right)^{T}\cdot\dot{\mathbf{x}}\left(\xi\right)=\nabla \mathbf{h}\left(\mathbf{x}\right)^{T}\mathbf{y}=0 \tag{15}$$

此时，$h\left(\mathbf{x}\right)$ 对应 $n-m$ 维曲面。此式的含义是在 $n-m$ 维曲面的某个高度的水平集上找到（12b）成立的点。

在 $\mathbf{x}^{*}$ 处，每个 $h_i\left(\mathbf{x}\right)=0$ 的 n 维曲面都存在自己的切平面及对应的法线方向 $\nabla h_i\left(\mathbf{x}\right)$，它们之间不能方向相同，如果相同，代表两个 n 维曲面在此处是相切的，其中一个要被舍弃（只要一个就够了！）。这是所谓正则性的要求。

因此，有 $\begin{cases}\nabla f\left(\mathbf{x}\right)^{T}\mathbf{y}=0 \\ \nabla h\left(\mathbf{x}\right)^{T}\mathbf{y}=0\end{cases}$ 同时成立时，某点就是约束优化问题的解。

（12b）和（12c）式导出的思想应该是没有问题的。可以对两式的几何含义做更进一步的具体解释。

10、约束优化问题的具体求解算法：

关于约束优化问题解的存在性的必要条件的讨论，并没给出具体的求解算法。对于约束优化问题的具体求解，Lagrange 乘子法是一种具体算法，投影法是更普遍的算法。

对于投影法，在每一步迭代中，基本含义是把上一步的未在可行解中的解投影到可行解中，关键是如何设计具体的投影算法。绝大部分的迭代算法都属于投影类迭代方法，尤其是稀疏反演方法。所以，迭代法中的核心思想是投影法，投影算子的构建是核心。

6 结论与讨论

地震数据处理和地震波反演成像中的很多问题最后都归结为无约束和有约束条件下的线性系统的求解。非线性系统的求解也是在局部线性假设下通过分步地解一个线性问题最后逼近非线性问题的解。由于勘探地震中遇到的基本上都是（超）大规模的线性系统问题的求解，迭代法是基本的解法。另外，这些线性系统往往是欠定的或不满秩的，迭代法寻找一定约束下线性系统的近似解是常见的场景。本文重点讨论迭代法求解线性系统的基本思想，主要是简述对迭代法求解线性系统的基本思想的看法，而且重点是投影法求解线性系统的基本思想的看法。

对于投影法，在每一步迭代中，基本含义是把上一步的未在可行解中的解投影到可行解中，关键是如何设计具体的投影算法。绝大部分的迭代算法都属于投影类迭代方法，尤其是稀疏反演方法。所以，迭代法中的核心思想是投影法，投影算子的构建是核心。

参考文献

[1] Todd K. Moon and Wynn C. Stirling, 2000, Mathematical Methods and Algorithms for Signal Processing. Prentice Hall
[2] Yousef Saad, 2003, Iterative Methods for Sparse Linear Systems, Second Edition, SIAM
[3] Edwin, K. P. Chong, and Stanislaw, H. Żak, 2013, An Introduction to Optimization, Fourth Edition；John Wiley& Sons
[4] Moon T K, Budge S E, Stirling W C, et al. Epistemic decision theory applied to multiple-target tracking[J]. IEEE transactions on systems, man, and cybernetics, 1994, 24(2): 234-245.
[5] Moon T K, Stirling W C. Satisficing negotiation for resource allocation with disputed resources[J]. Negotiation methods for autonomous cooperative systems, 2001: 106-115.

多源（元）信息融合及空间数据分析的基本逻辑

王华忠，雷霆，吴成梁

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】在“两宽一高”叠前地震数据采集和超级算力的支持下，Bayes 参数估计理论下的 FWI 成为地震波成像的核心方法技术基本成为了必然的选择。目前，FWI 的基本计算框架已经定型，虽不能说十分完善，但已经不是 FWI 实用化的主要矛盾。FWI（尤其是 RFWI）方法技术实用化的主要矛盾已经转到如何使其在实际数据地震波成像中得到实用，能得到更好的成像效果，从而提高油藏描述的精度。FWI 是一个信息不足情况下的大规模参数估计问题。尤其在成像目标定位为宽带弹性参数估计时，这是一个很强的非线性反问题。降低强非线性反问题的多解性，提高反演成像结果的精度，重要的措施就是引入与待估计参数相关的先验信息。勘探地震中，这些先验信息应该是与待估计参数一样，都是地下介质空间域的。它们主要包括：与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数据。这些要么是空间散乱点处的，要么是规则化三维空间网格点上的，且是二进制表示的数据。而地质知识和岩石物理关系这些非二进制表达的信息也是非常重要的与待估计参数相关的先验知识。这些先验信息如何统一地表达出来，如何在 FWI 迭代反演过程中合理地施加，成为了当前 FWI 反演成像技术的核心问题。事实上，高维、多属性空间中的初至波走时检测和反射波走时（及走时差）检测也成为了多信息约束下的反问题。三维成像空间中的层位、断层、地震相识别更是基于多种信息的综合判决问题。多信息融合更好地解决 FWI 反演成像问题是当前及今后 FWI 走向实用化的关键。更进一步地，多信息融合更高精度地描述岩性油气藏，并更好地做出含油气的判决是更深层次的问题。在 Bayes 推断（Bayes 判决）理论框架下，统一地讨论上述问题，可以促进上述各种问题研究的深化。本文首先讨论多源（元）信息融合的基本逻辑，然后给出空间随机（数据）集的定义。勘探地震中，与待估计参数或与油气藏描述相关的先验信息主要以地下介质三维空间坐标为标志量来描述，随时间变化的空间信息是次要的。因此，讨论空间数据分析和空间数据融合的基本逻辑，接着对空间数据的特点进行分析，在此基础上，介绍几种典型的空间数据分析方法。最后讨论勘探地震中多信息融合在参数估计反问题中降低反问题多解性的基本思想与方法。

【关键词】Bayes 推断，Bayes 参数估计，FWI 反演成像，先验信息，空间数据表达，空间数据建模，信息融合，降低多解性

1 前言

众所周知，油气地震勘探中，存在两个核心问题：宽带弹性参数估计或广义地震波反演成像；定量的油藏描述和含油气性评价。

但是，这两个问题都是强非线性。前者可被认为是 Bayes 参数估计问题；后者可以说是一个基于信息综合的判决问题（Bayes 判决问题）。后者的解决是依赖于前者的。Bayes 参数估计问题的解决是首要的。

目前，在“两宽一高”叠前地震数据采集和超级算力的支持下，Bayes 参数估计理论下的 FWI 成为地震波成像的核心方法技术已经是不争的实在，尽管它的实用化，尤其针对陆上地震数据的实用化还有很长的路要走，但这并不影响它成为地震波成像的核心方法技术的判断。应该说，经过二十多年的研究，FWI 的基本计算框架已经定型，虽不能说十分完善，但这已经不是 FWI 实用化的主要矛盾。FWI（尤其是 RFWI）方法技术实用化的主要矛盾已经转到如何使其在实际数据地震波成像中得到实用，能得到更好的成像效果，从而提高油藏描述的精度。但是，FWI 是一个信息不足情况下的强非线性反问题，尤其在成像目标定位为宽带弹性参数估计时。

如何降低强非线性反问题的多解性、提高反演成像结果的精度应该说已经形成共识，就是尽可能多地引入与待估计参数相关的先验信息。事实上，在宽带弹性参数估计或广义地震波反演成像的基础上，进行定量的油藏描述和含油气性评价，同样是一个强非线性的反问题，多解性也很强（可能更强）。引入与待描述和评价的油气藏相关的先验信息也是非常必要的。目前，我们远没有达到定量刻画油气藏的程度，基本上是定性地、概略地对油气藏进行描述和评价。即便如此，也需要综合地利用各种与油气藏相关的先验信息。

发展到现在，对于复杂问题的解决，多信息综合、多信息融合的思想和做法已经在所难免。

勘探地震中，这些先验信息应该都是三维（也可以是更高维）空间域的，可以随时间演化，也可以不随时间演化。勘探地震中的先验信息以不随时间演化的空间信息为主。主要应该包括：与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数据。这些要么是空间散乱点处的，要么是规则三维空间网格点上的，且是二进制表示的数据。而地质知识和岩石物理关系这些非二进制表达的信息也是非常重要的与待估计参数相关的先验知识。

很显然，FWI 的真正困难，就在于复杂介质（譬如复杂地表、复杂构造、小尺度和/或强变化的地质体

存在）情形下，它是一个强非线性反问题。强非线性反问题的多解性、不稳定性使得这类方法技术很难实用化。而假设已知正确的偏移速度模型的叠前地震波偏移成像问题，从数学算法上讲，不是一个反问题。因此，它有很强的稳健性，在勘探地震中得到了广泛的应用。但是，给出正确的叠前偏移成像结果的前提是假设已知正确的偏移速度模型，如果把构建偏移速度模型也放在广义的地震波成像过程中，广义的地震波成像就成了一个标准的反问题。在复杂介质情形下，同样很难得到正确的（或合理的）反演解。复杂探区叠前深度偏移成像结果弄不清真实的地下地质结构形态，解决不了相应的地质问题，本质原因就是此时的广义地震波成像是个（强）非线性反问题。如果以宽带弹性参数作为 FWI 的成像目标，该反问题的非线性性更会大幅度提高。这也是石油工业界把地震波成像分为背景速度建模和叠前深度偏移成像两部分的根本原因。毕竟叠前深度偏移成像数学理论上可以不被认为是一个参数估计反问题。

但是，高精度的、定量的油藏描述需要宽带弹性参数成像结果。此时把更多的与待估计参数相关的先验信息（多源（元）信息）提取出来，进行正确的表达，然后用合理的技术路线提升宽带弹性参数成像的精度，就成了当前地震波成像的一个十分关键的、不可回避的问题。

遗憾的是，多源（元）信息存在的异质性、尺度差异、精度差异、模糊性及随机性等导致到目前为止多源（元）信息融合并没有形成统一的、公认的理论框架。基本的认识是：首先统一地表达这些复杂的数据，比较一致的观点是形成随机（数据）集。然后，建立随机（数据）集满足的概率密度。接着对随机（数据）集进行建模。最后，进行基于模型的推断或判决。多源（元）信息融合体现在前三个步骤中，尤其是建立多源（元）信息的联合概率密度函数。这种高度抽象的多源（元）信息融合理论框架，在宽带弹性参数成像中，在油藏描述和含油气性评价中，要适当地变通，然后才能具体地应用。

基于此，本文首先讨论多源（元）信息融合的基本逻辑，然后给出空间随机（数据）集的定义。鉴于勘探地震以不随时间变化的空间数据为主，讨论空间数据分析和空间数据融合的基本逻辑，接着对空间数据的特点进行分析，在此基础上，介绍几种典型的空间数据分析方法。最后讨论勘探地震中多信息融合在参数估计反问题中降低反问题多解性的基本思想与方法。

2 多源（元）信息融合的基本逻辑

多源（元）信息融合首先在军事领域，尤其是战场信息管理及决策领域得到了深入的研究，这是一个最典型的多源（元）信息获取、管理、融合和应用的场景。多源（元）信息融合的方法理论正是在军事领域应用的强力促进下得到长足的发展。众所周知，地球物理学领域中，也存在非常多的多源（元）信息融合的应用场景。因为地球物理场是多种多样的，描述岩性介质的物理参数也是多种多样的、尺度变化也存在很大的跨度。而且各种信息都存在不同程度的不确定性、模糊性和随机性。既有二进制描述的定量化信息，也有非二进制描述的定性信息。多源（元）信息融合在 FWI 宽带弹性参数估计中的应用也是 FWI 发展到现在的必然选择。当然，还有很多需要多源（元）信息融合的应用领域。事实上，由于多源（元）信息的异质性、尺度差异、精度差异、模糊性及随机性等，到目前为止，多源（元）信息融合并没有形成统一的、公认的理论框架。

但是，基于到目前为止的研究已经可以预知多源（元）信息融合首先要解决的是多源（元）信息的抽象表达问题。一般地，多源（元）信息首先可以分为定量的、可二进制表达的及定性的、不可二进制表达的（实际上，计算机处理时，也往往通过某种方式把这些信息转换成可二进制表达的）。第二，多源（元）信息可分为随时间演变的空间数据和不随时间演变的空间数据。第三，多源（元）信息可分为空间散乱的数据和空间规则的数据。最后，还要注意到，多源（元）信息中有很多是一级信息（数据）的特征表达或属性表达结果。

多源（元）信息之间存在的异质性、尺度差异、精度差异、模糊性及随机性更增加了对多源（元）信息进行抽象表达的难度。到目前为止，用随机（数据）集来表达多源（元）信息应该是抽象程度比较高的一种表达形式。

描述随机（数据）集中蕴含的统计规律的最合理方法当然是联合概率密度函数。对随机（数据）集进行建模，本质上就是估计联合概率密度函数中的参数或联合概率密度函数的数字特征。最后基于随机（数据）集的建模结果，进行统计推断，解决相应的问题。对随机（数据）集建模的过程，就是多源（元）信息融合的过程，就是把多源（元）信息统一在所建立的模型中的过程。

我们认为：这是多源（元）信息融合的最深层次的逻辑。

解决具体的应用问题，会对上述逻辑路线进行各种各样的改变。像数据融合、图像融合等都是低一个层次的信息融合解决的问题。一般地，对于随时间演变的空间数据，Kalman 滤波、扩展的 Kalman 滤波，尤其是粒子滤波基本形成了针对此类数据的多源（元）信息融合算法体系。对于不随时间演变的空间数据，随机场理论、（泛/协）Kriging 插值、空间回归建模方法基本形成了针对此类数据的多源（元）信息融合算法体系。

总之，对多源（元）信息构成的随机（数据）集，在 Bayes 推断的理论框架下，进行建模，并基于模型进行统计推断，进而解决各种实际问题，是多源（元）信息融合的深层次思想。

3 空间数据集（随机（数据）集）的定义

由于在勘探地震中，尤其在 FWI 宽带弹性参数估计中，与宽带弹性参数相关的信息仅仅存在于空间域，此处我们给出不随时间变化的空间数据集的定义。我们认为: 与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数据，包括地质知识和岩石物理关系这些定性的、非二进制的数据，都可以统一在这个空间数据集之内。

空间点 $s_i, i=1,2,\cdots,N$ 上，存在物理对象（也可以是非物理对象）的观测值、估计值、或它们的属性等等，这些都是描述（非）物理对象的信息。可以记为 $X_i, i=1,2,\cdots,N$，认为它是随机变量，是满足一定的概率分布 $p_X(x)$ 的。用概率分布来表达先验信息的统计特征，包容了它的模糊性、不确定性。由于描述（非）物理对象的信息可能是多层次、多方面的，可以认为它们是从不同角度、不同尺度以及不同可信度来描述同一个（非）物理对象。因此，在一个空间点上，可以用随机向量 $\mathbf{X}_j, j=1,2,\cdots,M$ 描述多源（元）信息，不同的分量代表不同的信息。

随机变量 $X_i, i=1,2,\cdots,N$ 或随机向量 $\mathbf{X}_i, i=1,2,\cdots,N$ 蕴含了（非）物理对象在一个空间点处统计规律。但是，多源（元）信息融合更关注的是不同空间点处多源（元）信息之间的统计变化关系。此时，需要引入空间随机向量 $\mathbf{Y}=\left(\mathbf{X}_{S_1},\mathbf{X}_{S_2},\cdots,\mathbf{X}_{S_N}\right)$。当然，对于空间信息融合，更需要引入空间随机矩阵 $\mathbf{Y}_{N\times M}=\begin{pmatrix}\mathbf{X}_{S_1}^1 & \mathbf{X}_{S_2}^1 & \cdots & \mathbf{X}_{S_N}^1\\ \mathbf{X}_{S_1}^2 & \mathbf{X}_{S_2}^2 & \cdots & \mathbf{X}_{S_N}^2\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{X}_{S_1}^M & \mathbf{X}_{S_2}^M & \cdots & \mathbf{X}_{S_N}^M\end{pmatrix}$。空间随机矩阵是空间随机场的表达。

可见，随机矩阵 $\mathbf{Y}_{N\times M}$ 包含了（非）物理对象的所有空间变化信息。最基本地，它们包含在随机矩阵满足的（复杂的）联合概率密度函数中。

如果空间随机矩阵（空间随机场）是随时间变化的，存在如下的随时间变化的随机矩阵，即 $\mathbf{Y}_{N\times M}(t)=\begin{pmatrix}\mathbf{X}_{S_1}^1(t) & \mathbf{X}_{S_2}^1(t) & \cdots & \mathbf{X}_{S_N}^1(t)\\ \mathbf{X}_{S_1}^2(t) & \mathbf{X}_{S_2}^2(t) & \cdots & \mathbf{X}_{S_N}^2(t)\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{X}_{S_1}^M(t) & \mathbf{X}_{S_2}^M(t) & \cdots & \mathbf{X}_{S_N}^M(t)\end{pmatrix}$。

我们认为多源（元）信息用这样的数据形式来表达，应该达到了较高的抽象程度。可以包容很多领域中的多源（元）信息融合问题。对这样的数据集进行建模，把其中蕴含的信息统一在一个模型中，就实现了信息融合的目的。

4 空间数据分析或空间信息融合的基本逻辑

FWI 宽带弹性参数估计中，与待估计参数相关的信息都是空间域的。空间数据 $\mathbf{Y}=\left(\mathbf{X}_{S_1},\mathbf{X}_{S_2},\cdots,\mathbf{X}_{S_N}\right)$ 或

$\mathbf{Y}_{N\times M}=\begin{pmatrix}\mathbf{X}_{S_1}^1 & \mathbf{X}_{S_2}^1 & \cdots & \mathbf{X}_{S_N}^1\\ \mathbf{X}_{S_1}^2 & \mathbf{X}_{S_2}^2 & \cdots & \mathbf{X}_{S_N}^2\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{X}_{S_1}^M & \mathbf{X}_{S_2}^M & \cdots & \mathbf{X}_{S_N}^M\end{pmatrix}$ 中蕴含的信息体现在空间位置 $s_i,i=1,2,\cdots,N$ 为标志的随机变量之间存在相关关系上。原则上，空间位置相邻越近，相关关系越强。但是，由于空间随机变量（向量、矩阵、张量）受背后（非）物理规律的制约，空间不相邻的随机变量之间也可能存在很强的相关性。这种先验信息，当然也可以预先表达在空间数据集中。一般地，随机变量（向量、矩阵、张量）满足的概率密度函数对应的数字特征（或其中包含的参数）体现了空间数据集所蕴含的规律。

因此，空间数据分析或空间信息融合的本质应该是：基于已知的空间数据集，假设它满足一定的联合概率密度概率分布，利用 Bayes 推断的理论框架，估计概率分布函数中的参数值（或数字特征），最后用估计的参数值（或数字特征）对整体空间数据集的变化规律实现尽可能准确的推断，从而解决各种相关的问题。其中，空间信息融合统一在对空间数据集的建模过程中。

我们认为：勘探地震中的多源（元）信息融合本质上就是按照上述逻辑进行的。

多源（元）信息约束下的（随机）散乱数据的插值构建规则网格上的参数模型就是一个典型的问题。本质上是在多源（元）信息约束下估计（随机）散乱数据支撑起来的空间数据的均值（期望值）。在此过程中，引入关于期望值的先验信息，进行 Bayes 推断是更为数学化的表达，也是多源（元）信息融合构建规则网格上的参数模型的本意。因为仅仅依赖（随机）散乱数据的统计规律，由于采样点数的不足，若没有先验信息的加入，对全空间上的空间数据均值（期望值）的估计误差可能会很大。

多源（元）信息融合/约束下的统计推断是另一个典型的问题。在同一个空间中，（非）物理对象具有各种不同侧面体现出来的属性。这些属性可以认为是从不同角度展示出来的同一个（非）物理对象的信息。这些信息之间存在或强或弱的关联性。$\mathbf{Y}_{N\times M}=\begin{pmatrix}\mathbf{X}_{S_1}^1 & \mathbf{X}_{S_2}^1 & \cdots & \mathbf{X}_{S_N}^1\\ \mathbf{X}_{S_1}^2 & \mathbf{X}_{S_2}^2 & \cdots & \mathbf{X}_{S_N}^2\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{X}_{S_1}^M & \mathbf{X}_{S_2}^M & \cdots & \mathbf{X}_{S_N}^M\end{pmatrix}$ 就是多源（元）信息的随机矩阵表达。多源（元）信息融合的目的就是提取空间数据矩阵 $\mathbf{Y}_{N\times M}$ 中各不同信息之间的共性信息（相关信息），然后用这些共性信息，提高对空间数据的建模精度，进而提高 Bayes 推断的精度。

另外，还存在一种多信息约束下的系统参数估计问题。一般地，认为存在观测数据 $\mathbf{d}(Z_k),k=1,2,\cdots,K$，它也是随机向量在观测点 $Z_k,k=1,2,\cdots,K$ 的样本值。假定其期望 $E[\mathbf{d}]$ 与空间域点 $s_i,i=1,2,\cdots,N$ 处的系统参数 $\mathbf{m}(s_i),i=1,2,\cdots,N$ 之间满足一定的数学物理方程。把已知系统参数、数学物理方程和初边值条件结合预测观测数据的期望值 $E[\mathbf{d}]$ 的问题称为正问题。由观测数据 $\mathbf{d}(Z_k),k=1,2,\cdots,K$ 估计系统参数 $\mathbf{m}(s_i),i=1,2,\cdots,N$ 的问题称为反问题。还是由于观测数据 $\mathbf{d}(Z_k),k=1,2,\cdots,K$ 的采样不充分和不完美，还包括上述正问题的不精确等，进行多源（元）信息约束下的 Bayes 参数估计。这个问题可以认为是多源（元）信息融合/约束下的统计推断问题的特例。

5 空间数据变化特征分析

空间域的各种先验信息，一般地，都可以视为空间数据。空间数据分析与时间序列分析是不同的。时间序列往往被视为一个随机过程。当然，把空间坐标与时间坐标等价起来，也可以用时间序列分析的观点处理空间数据。但是，空间数据与时间序列数据产生的机制毕竟是不同的。空间相邻的数据不一定存在紧密相关性。因此，对空间数据的建模与分析必须要构建另一套不同于时间序列分析的模式。

首先，我们认为空间数据可以表达为：$\mathbf{d}(\mathbf{x})=\mathbf{d}_{BC}(\mathbf{x})+\mathbf{d}_{DC}(\mathbf{x})+\mathbf{d}_{RC}(\mathbf{x})$。其中，$\mathbf{d}_{BC}(\mathbf{x})$ 是空间数据的背景成分（Background Component）；$\mathbf{d}_{DC}(\mathbf{x})$ 是空间数据的扰动成分（Disturbance Component）；$\mathbf{d}_{RC}(\mathbf{x})$ 是满足一定概率分布的随机成分（Random Component）。$\mathbf{d}_{BC}(\mathbf{x})$ 可以视为是确定性的、（空间）散乱点支撑的函数。这部分的描述可以用径向基函数族或其他基函数族的线性组合的方法；$\mathbf{d}_{DC}(\mathbf{x})$ 用零均值、一定方差

的概率密度函数来描述；$\mathbf{d}_{RC}(\mathbf{x})$ 被视为高斯白噪声。

在上述空间数据表达下，空间数据建模就是在 $\mathbf{d}_{DC}(\mathbf{x})$ 最小的条件下构建对 $\mathbf{d}_{BC}(\mathbf{x})$ 进行最佳表达的模型，同时期望 $\mathbf{d}_{RC}(\mathbf{x})$ 是高斯白噪声。更进一步地，上述建模过程是在多源（元）先验信息的约束下进行的。地震波成像过程中存在很多这样的空间建模问题。

一般地，空间数据 $Z(\mathbf{x})$ 可表达为大尺度变化+小尺度变化的形式，即

$$Z(\mathbf{x})=s(\mathbf{x})+\delta(\mathbf{x}), \tag{1}$$

$s(\mathbf{x})$ 代表空间数据的背景变化，可以认为它是一个连续的函数，用基函数的线性组合来表达，

$$s(\mathbf{x})=\sum_{l=1}^{q}\beta_l f_l(\mathbf{x}),\mathbf{x}\in\mathbf{D}\subset\mathbb{R}^d . \tag{2}$$

则，（1）式变为，

$$Z(\mathbf{x})=s(\mathbf{x})+\delta(\mathbf{x})=\sum_{l=1}^{q}\beta_l f_l(\mathbf{x})+\delta(\mathbf{x}),\ \mathbf{x}\in\mathbf{D}\subset\mathbb{R}^d, \tag{3}$$

其中，$\{f_l(\cdot)|l=1,\cdots,q\}$ 是 q 个基函数的集合，基函数的选择要依据空间数据的先验信息；$\delta(\cdot)$ 是零均值有限方差的随机误差，代表 $Z(\mathbf{x})$ 中不能被 $s(\mathbf{x})$ 表达后的残差。之所以认为 $\delta(\cdot)$ 是随机误差，是因为空间数据 $Z(\mathbf{x})$ 是随机的。$s(\mathbf{x})$ 应该是空间数据 $Z(\mathbf{x})$ 的统计特征（期望）。

空间数据 $Z(\mathbf{x})$ 的最基本特征是：一般意义下，相邻空间位置的随机采样值相关性更大。在高维空间中，这个特点使得高维数据体中蕴含空间结构。正是存在这样的空间结构，说明空间数据 $Z(\mathbf{x})$ 蕴含着某种信息。所谓的对空间数据 $Z(\mathbf{x})$ 的建模就是对这些空间结构信息进行建模表达。

对于 $x\in\mathbb{R}^1$ 的情况，若假设 $Z(x_1),\cdots,Z(x_n)$ 是独立同分布（IID）的高斯分布随机变量，且均值为 μ，方差为 σ_0^2。对观察数据做这样的假设非常合理的，因为每次观测之间是独立进行的。

则由概率论知，均值 μ 的最小方差无偏估计是

$$\hat{\bar{Z}}=\sum_{i=1}^{n}Z(x_i)/n, \tag{4}$$

且估计子 $\hat{\bar{Z}}$ 是高斯的、均值为 μ、方差为 σ_0^2/n。因此，$\hat{\bar{Z}}$ 的 95%置信区间为

$$\left[\hat{\bar{Z}}-1.96\sigma_0^2/n,\ \hat{\bar{Z}}+1.96\sigma_0^2/n\right]. \tag{5}$$

我们反复提及：空间数据的重大特点之一是一般而言相邻点间的随机变量存在更强的相关性。但（非）物理机制又常常使得上述说法不成立。在空间数据分析中，需要引入额外的先验信息。

无论如何，首先讨论空间数据之间相关性表达一定是空间数据分析的基础。假设空间数据正相关且相关性随距离减弱，即

$$\mathrm{cov}\left(Z(x_i),Z(x_j)\right)=\sigma_0^2\cdot\rho^{|x_i-x_j|},\ i,j=1,\cdots,n,\ 0<\rho<1, \tag{6}$$

该相关函数来自于一阶自回归过程 $Z(x_i)=\rho Z(x_{i-1})+\varepsilon(x_i),\ i\geq 1$，其中 $\varepsilon(x_i)$ 是零均值、方差为 $\sigma_0^2(1-\rho^2)$ 的高斯随机变量。据此，可以看出空间数据分析与时间序列分析之间存在一定的类比性。

此时，估计子 $\hat{\bar{Z}}$ 的方差为

$$\begin{aligned}\mathrm{var}\left(\hat{\bar{Z}}\right)&=\frac{1}{n^2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{cov}\left(Z\left(x_i\right),Z\left(x_j\right)\right)\right\}\\&=\frac{\sigma_0^2}{n}\left[1+2\rho/(1-\rho)(1-1/n)-2\left(\rho/(1-\rho)\right)^2\left(1-\rho^{n-1}\right)/n\right],\end{aligned}\tag{7}$$

当 $n=10,\rho=0.26$ 时， $\hat{\bar{Z}}$ 的 95%置信区间为

$$\left[\hat{\bar{Z}}-2.458\sigma_0^2/n,\ \hat{\bar{Z}}+2.458\sigma_0^2/n\right].\tag{8}$$

可见，数据的空间相关性使 $\hat{\bar{Z}}$ 的方差增大。这是空间数据分析的最直观的数据特征。

把（3）写成矩阵方程的形式

$$\mathbf{Z}=\mathbf{f}\boldsymbol{\beta}+\boldsymbol{\delta},\tag{9}$$

其中， $\mathbf{f}\in\mathbb{R}^{n\times q}$ ，第 (i,j) 个元素是 $f_j(\mathbf{x}_i)$ ； $\boldsymbol{\beta}=\left(\beta_1,\cdots,\beta_q\right)^T$ 。 $\boldsymbol{\beta}$ 是基函数族 $\mathbf{f}$ 上对应的坐标向量。 $\boldsymbol{\delta}=\left(\delta(\mathbf{x}_1),\cdots,\delta(\mathbf{x}_n)\right)^T$ 是预测残差向量，假设其协方差矩阵为

$$\mathrm{var}(\boldsymbol{\delta})=\boldsymbol{\Sigma},\tag{10}$$

其中， $\boldsymbol{\Sigma}\in\mathbb{R}^{n\times n}$ 是对称半正定矩阵。

对空间数据建模的主要问题是在最小二乘意义下最佳地估计坐标向量 $\boldsymbol{\beta}$ ，即求解如下加权最小二乘（WLS = Weighted Least Square）问题，

$$\min_{\boldsymbol{\beta}}(\mathbf{Z}-\mathbf{f}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Z}-\mathbf{f}\boldsymbol{\beta}).\tag{11}$$

（11）式估计结果（WLS 估计子）为

$$\hat{\boldsymbol{\beta}}_{wls}=\left(\mathbf{f}^T\boldsymbol{\Sigma}^{-1}\mathbf{f}\right)^{-1}\mathbf{f}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}.\tag{12}$$

当 $\boldsymbol{\Sigma}=\sigma^2\mathbf{I}$ ，（11）退化为

$$\min_{\boldsymbol{\beta}}(\mathbf{Z}-\mathbf{f}\boldsymbol{\beta})^T(\mathbf{Z}-\mathbf{f}\boldsymbol{\beta})/\sigma^2.\tag{13}$$

对应的估计结果为:

$$\hat{\boldsymbol{\beta}}_{ols}=\left(\mathbf{f}^T\mathbf{f}\right)^{-1}\mathbf{f}^T\mathbf{Z}.\tag{14}$$

这即一般的最小二乘（LS）估计子。当 $\boldsymbol{\delta}(\cdot)$ 是高斯白噪声时，估计结果满足 Gauss-Markov 定理，是无偏的、方差最小的估计结果（OLS = Optimal Least Square）。

可见，残差数据 $\boldsymbol{\delta}=\left(\delta(\mathbf{x}_1),\cdots,\delta(\mathbf{x}_n)\right)^T$ 中存在结构特征时，估计结果不可能达到最佳。换个角度就是，对空间数据 $\mathbf{Z}$ 进行最佳建模等价于使得残差数据 $\delta(\cdot)$ 满足高斯白噪声。更进一步地，对空间数据 $\mathbf{Z}$ 进行最佳建模等价于同时估计最佳参数 $\hat{\boldsymbol{\beta}}$ 和使得残差数据 $\delta(\cdot)$ 的协方差满足 $\boldsymbol{\Sigma}=\sigma^2\mathbf{I}$ 的要求。

这体现了空间数据建模的独特之处。

据此，假设残差数据 $\delta(\cdot)$ 的协方差是含参数的 $\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)$ ，代入（12）式得到所谓的估计广义最小二乘解（EGLS = Estimated General Least Square）

$$\hat{\boldsymbol{\beta}}_{egls}=\left(\mathbf{f}^T\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)^{-1}\mathbf{f}\right)^{-1}\mathbf{f}^T\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)^{-1}\mathbf{Z}.\tag{15}$$

$\hat{\boldsymbol{\theta}}$ 是协方差矩阵 $\boldsymbol{\Sigma}$ 中包含的参数 $\boldsymbol{\theta}$ 的估计量。很显然，迭代地估计 $\hat{\boldsymbol{\theta}}$ 和 $\hat{\boldsymbol{\beta}}_{egls}$ 达到最佳的空间数据建模结果。这是空间数据分析的典型方法的基本理论基础。

事实上，由于上述基本理论的假设不太考虑空间数据变化的物理机制，在空间数据建模中引入额外的先验信息就是必然的。这是勘探地震中常见的问题。

6 空间数据分析的典型方法

空间数据统计分析的典型方法来自于 Kriging 提出的思想。残差数据 $\delta(\cdot)$ 的协方差的结构特征为 Kriging 法用半变差函数模型替代协方差阵 $\boldsymbol{\Sigma}$ 奠定了理论基础。空间数据的统计学模型及插值方法，主要指Kriging法及其变体，基本思想是假设空间数据是平稳的，任一点的空间数据可以用周围数据的加权平均（本质是就是统计期望）来预测，基于此定义预测误差泛函。同时把涉及到的空间数据的协方差阵 $\boldsymbol{\Sigma}$ 用半变差函数替代，因为空间数据的协方差阵 $\boldsymbol{\Sigma}$ 这个二阶统计量是未知的，据此形成构建空间数据模型的误差泛函。求解该误差泛函定义的优化问题，就可以得到空间数据的最佳预测模型。

6.1 空间数据分析的 Kriging 方法

一般 Kriging 方法，假设空间数据的均值（趋势）为常数

$$E\left[Z(\mathbf{x})\right]=\mu\,, \tag{16}$$

因此，空间数据可以表达为

$$Z(\mathbf{x})=\mu+\delta(\mathbf{x})\,, \tag{17}$$

其中，$\delta(\mathbf{x})$ 是零均值固有平稳随机过程。不能认为 $\delta(\mathbf{x})$ 是观测噪声，而是（非）物理对象在位置 $\mathbf{x}$ 上的小尺度变化+噪声。

基于（17）式的表达，可以构建如下的预测关系式

$$\hat{Z}(\mathbf{x_u})=\sum_{i=1}^{n}\lambda_i Z(\mathbf{x_i}),\ \ s.t.\ \sum_{i=1}^{n}\lambda_i=1\,, \tag{18}$$

其中，$Z(\mathbf{x_i}),\ i=1,2,\cdots,n$ 是已知观测点处的数据。$\hat{Z}(\mathbf{x_u})$ 是未知空间点处的预测值，约束条件 $\sum_{i=1}^{n}\lambda_i=1$ 使预测值满足无偏估计 $E\left[\hat{Z}(\mathbf{x_u})\right]-E\left[Z(\mathbf{x_u})\right]=0$ 。本质上，这也是求统计期望的基本要求，只是此处不把 $\lambda_i, i=1,2,\cdots,n$ 视为概率密度函数的离散值。（18）式的写出显然利用了 $Z(\mathbf{x_i}),\ i=1,2,\cdots,n$ 满足平稳性的假设，可以认为是时间序列分析中 AR 模型在空间数据分析中的推广应用。

用 Lagrange 乘子法把（18）式定义的约束优化问题转为无约束优化问题，可得目标泛函

$$\min_{\lambda_1,\lambda_2,\ldots\lambda_n} E\left[\left(Z(\mathbf{x_u})-\sum_{i=1}^{n}\lambda_i Z(\mathbf{x_i})\right)^2\right]-2m\left(\sum_{i=1}^{n}\lambda_i-1\right), \tag{19}$$

其中，

$$\begin{aligned} E\left[\left(Z(\mathbf{x_u})-\sum_{i=1}^{n}\lambda_i Z(\mathbf{x_i})\right)^2\right]=&-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\, E\left[\left(Z(\mathbf{x_i})-Z(\mathbf{x_j})\right)^2\right]\Big/2+\\ &2\sum_{i=1}^{n}\lambda_i\, E\left[\left(Z(\mathbf{x_u})-Z(\mathbf{x_i})\right)^2\right]\Big/2. \end{aligned} \tag{20}$$

该式是在 $\sum_{i=1}^{n}\lambda_i=1$ 的条件下导出的。注意，定义（19）式用了均方误差，在 $\sum_{i=1}^{n}\lambda_i=1$ 的条件下，这就与空间数据的协方差联系起来了。

由于协方差函数是空间数据 $Z(\mathbf{x_i})$, $i=1,2,\cdots,n$ 的二阶统计量，而它是未知的。放弃使用协方差函数这个未知的二阶统计量，转为依据实测的、散乱的空间数据定义如下半变差函数

$$\gamma(\mathbf{x},\mathbf{h})=\frac{1}{2}\mathrm{var}\left(Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})\right)=\frac{1}{2}E\left[\left(Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})\right)^2\right]. \tag{21}$$

应该注意到上式中引入了距离向量 $\mathbf{h}$。如果假定空间数据 $Z(\mathbf{x})$ 是广义平稳的，二阶统计量不随空间位置变化，则上式可以简写为

$$\gamma(\mathbf{h})=\frac{1}{2}E\left[\left(Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})\right)^2\right]. \tag{22}$$

若再进一步地假设空间数据变化是各向同性的，即半变差函数 $\gamma(\mathbf{h})$ 与距离向量 $\mathbf{h}$ 的方向无关，上式简化为

$$\gamma(\|\mathbf{h}\|)=\frac{1}{2}E\left[\left(Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})\right)^2\right]. \tag{23}$$

据此，（19）和（20）式定义的目标泛函变为

$$\min_{\lambda_1,\lambda_2,\cdots,m}-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma\left(\|\mathbf{x_i}-\mathbf{x_j}\|\right)+2\sum_{i=1}^{n}\lambda_i\gamma\left(\|\mathbf{x_u}-\mathbf{x_i}\|\right)-2m\left(\sum_{i=1}^{n}\lambda_i-1\right), \tag{24}$$

对（24）式求关于 $\lambda_1,\lambda_2,\cdots,\lambda_n,m$ 的偏导数，并令偏导数为零，可得如下方程组

$$\begin{cases}-2\sum_{j=1}^{n}\lambda_j\gamma\left(\|\mathbf{x_i}-\mathbf{x_j}\|\right)+2\gamma\left(\|\mathbf{x_u}-\mathbf{x_i}\|\right)-2m=0,\ i=1,2,\cdots,n\\ \sum_{i=1}^{n}\lambda_i=1\end{cases}. \tag{25}$$

上式对应的矩阵方程具体地可写为，

$$\begin{bmatrix}\gamma\left(\|\mathbf{x_1}-\mathbf{x_1}\|\right) & \gamma\left(\|\mathbf{x_1}-\mathbf{x_2}\|\right) & \cdots & \gamma\left(\|\mathbf{x_1}-\mathbf{x_n}\|\right) & 1\\ \gamma\left(\|\mathbf{x_2}-\mathbf{x_1}\|\right) & \gamma\left(\|\mathbf{x_2}-\mathbf{x_2}\|\right) & \cdots & \gamma\left(\|\mathbf{x_2}-\mathbf{x_n}\|\right) & 1\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \gamma\left(\|\mathbf{x_n}-\mathbf{x_1}\|\right) & \gamma\left(\|\mathbf{x_n}-\mathbf{x_2}\|\right) & \cdots & \gamma\left(\|\mathbf{x_n}-\mathbf{x_n}\|\right) & 1\\ 1 & 1 & \cdots & 1 & 0\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_n\\ m\end{bmatrix}=\begin{bmatrix}\gamma\left(\|\mathbf{x_u}-\mathbf{x_1}\|\right)\\ \gamma\left(\|\mathbf{x_u}-\mathbf{x_2}\|\right)\\ \vdots\\ \gamma\left(\|\mathbf{x_u}-\mathbf{x_n}\|\right)\\ 1\end{bmatrix}. \tag{26}$$

对应地，（26）式可简写为 $\mathbf{\Gamma\lambda=\gamma}$。数值求解该矩阵方程，可以得到（24）式中的最佳加权系数及超参。然后，用（18）式就可以得到空间数据最佳预测结果。

6.2 空间数据分析中泛 Kriging 方法的基本思想

一般 Kriging 方法假设空间数据是广义平稳的，即一阶和二阶统计量是不随时间变化的。但是，空间数据的变化很难具备二阶平稳性（广义平稳性）。最多只能假设是局部二阶平稳的。

为此，去掉空间数据的二阶平稳性假设，就提出了所谓的泛 Kriging 方法。这与把 Wiener 滤波发展到 Kalman 滤波的扩展从思想上看是完全类似的。

此时，必须考虑空间数据一阶统计量是随空间位置变化的。同时，假设空间数据的背景变化（一阶统计量变化）可以表达为

$$E\left[Z(\mathbf{x})\right]=\mu(\mathbf{x})=\sum_{j=0}^{p}\beta_j f_j(\mathbf{x}). \tag{27}$$

因此，非平稳情形下很一般的空间数据模型就可以写为

$$Z(\mathbf{x})=\sum_{j=1}^{p+1}\beta_{j-1}f_{j-1}(\mathbf{x})+\delta(\mathbf{x}),\tag{28}$$

其中，$\delta(\mathbf{x})$是零均值平稳随机过程。

基于此，泛 Kriging 的基本思想应该是：泛 kriging 方法中，背景部分（$Z(\mathbf{x})$的一阶统计量）是空间变化的，即$E[Z(\mathbf{x})]=\mu(\mathbf{x})$，此时$\mu(\mathbf{x})$称为漂移。在这种情况下，不能用一般 Kriging 法来计算变异函数。如果漂移函数$\mu(\mathbf{x})$是已知的，可以在原始资料中将漂移减去，如剩下的残差$\delta(\mathbf{x})$满足二阶平稳性假定，就可用前面讨论的一般 Kriging 法来对残差$\delta(\mathbf{x})$进行估值，然后再将漂移$\mu(\mathbf{x})$加到对应位置的残差估值上去，其和就是$Z(\mathbf{x})$的估计值。这就构成了两步优化逼近问题。漂移函数$\mu(\mathbf{x})$是需要估计的，$\delta(\mathbf{x})$也要用 Kriging 法进行估计。$\mu(\mathbf{x})$一般可以假设为公式（27）的形式，基函数$f_j(\mathbf{x})$往往是事先给定的，比如多项式函数，需要估计最佳系数β_j值。当然，也可以把两步估计问题合成一步估计问题。

详细的算法，很多参考文献有具体的讨论。

6.3 空间数据分析中协 Kriging 方法的基本思想

空间数据分析中，存在如下情形，$Z_1(\mathbf{x})$数据比$Z_2(\mathbf{x})$数据更容易获取，且$Z_1(\mathbf{x})$数据中包含了$Z_2(\mathbf{x})$数据中的部分信息，也可以说两者存在相关性。因此，把$Z_1(\mathbf{x})$用作对$Z_2(\mathbf{x})$进行估计的先验信息是很正常的考虑。事实上，存在很多这样的思想和做法。此处，讨论的是协 Kriging 方法的基本思想。

在一般 Kriging 方法对观测数据的假设下，写出如下的协 Kriging 方法

$$Z_2^*(\mathbf{x}_0)=\sum_{i=1}^{N_1}\lambda_i^1Z_1(\mathbf{x}_i^1)+\sum_{i=1}^{N_2}\lambda_i^2Z_2(\mathbf{x}_i^2),\tag{29}$$

其中，$Z_2^*(\mathbf{x}_0)$代表$Z_2(\mathbf{x})$在$\mathbf{x}_0$点处的估计值，$Z_1(\mathbf{x}_i^1),i=1,2,\cdots,N_1$代表$Z_1(\mathbf{x})$在各观测点处的观测值；$Z_2(\mathbf{x}_i^2),i=1,2,\cdots,N_2$代表$Z_2(\mathbf{x})$在各观测点处的观测值。$\lambda_i^1,i=1,2,\cdots,N_1$代表在$\mathbf{x}_0$点附近$Z_1(\mathbf{x})$数据的加权系数；$\lambda_i^2,i=1,2,\cdots,N_2$代表在$\mathbf{x}_0$点附近$Z_2(\mathbf{x})$数据的加权系数。

同样地，在两个观测数据都是广义平稳的假设下，基于类似最佳 Wiener 滤波的要求，期望对$Z_2^*(\mathbf{x}_0)$的估计也是无偏和方差最小的。

即有下述式子成立

$$E\left[Z_2(\mathbf{x}_0)-\sum_{i=1}^{N_1}\lambda_i^1Z_1(\mathbf{x}_i^1)-\sum_{i=1}^{N_2}\lambda_i^2Z_2(\mathbf{x}_i^2)\right]=-\mu_1\sum_{i=1}^{N_1}\lambda_i^1-\mu_2\sum_{i=1}^{N_2}(\lambda_i^2-1)=0,\tag{30a}$$

其中，$\mu_1=E[Z_1(\mathbf{x})]$，$\mu_2=E[Z_2(\mathbf{x})]$。上式成立，必有$\sum_{i=1}^{N_1}\lambda_i^1=0$和$\sum_{i=1}^{N_1}\lambda_i^2=1$。

$$\begin{aligned}&\mathrm{Var}[Z_2(\mathbf{x}_0)-Z_2^*(\mathbf{x}_0)]=E\left\{\left[(Z_2(\mathbf{x}_0)-Z_2^*(\mathbf{x}_0))-E[Z_2(\mathbf{x}_0)-Z_2^*(\mathbf{x}_0)]\right]^2\right\}\\&=E\left[(Z_2(\mathbf{x}_0)-Z_2^*(\mathbf{x}_0))^2\right]\\&=E\left[\left(Z_2(\mathbf{x}_0)-\sum_{i=1}^{N_1}\lambda_i^1Z_1(\mathbf{x}_i^1)-\sum_{i=1}^{N_2}\lambda_i^2Z_1(\mathbf{x}_i^2)\right)^2\right]\\&\Rightarrow\min\end{aligned}\quad.\tag{30b}$$

类似一般 Kriging 方法，（30b）式与约束条件$\sum_{i=1}^{N_1}\lambda_i^1=0$和$\sum_{i=1}^{N_1}\lambda_i^2=1$分别结合，可以用 Lagrange 乘子法构建两个无约束优化反问题。求解两个无约束优化反问题的联立方程，可以求出$\lambda_i^1,i=1,2,\cdots,N_1$和

$\lambda_i^2, i=1,2,\cdots,N_2$，以及引入的对应的超参。利用（29）式就可以进行协 Kriging 法的预测。

由于 $Z_1\left(\mathbf{x}_i^1\right), i=1,2,\cdots,N_1$ 和 $Z_2\left(\mathbf{x}_i^2\right), i=1,2,\cdots,N_2$ 两组观测数据不一定在同一点观测，此时，无法计算相应的半变差函数及半协变差函数。据此，又提出了所谓的拟协 Kriging 法的思想。具体的内容，查看相应的参考文献。

很显然，更进一步地会有泛协 Kriging 方法和泛拟协 Kriging 方法的合乎逻辑的扩展。这些内容请查看相关的参考文献。

这些思想，尤其是协 Kriging 方法的思想，对多源（元）信息融合显然是有指导性的。但是，客观地讲，Kriging 方法本质上无非是把时间序列分析的 Wiener 最佳滤波思想，相应地推广到空间数据分析中而已。这不构成多源（元）信息融合的十分抽象的理论框架。

6.4 空间数据分析中的径向基函数建模方法

假如空间数据可以分解为背景+扰动的成分，且背景部分可以基函数族的线性组合进行逼近。Weierstrass 定理下的函数逼近可以作为空间数据分析的另一个理论框架。所不同的是基函数（族）的选择要与空间数据的统计特征关联起来。这就是径向基函数（族）逼近的空间数据建模的基本逻辑。

空间数据的径向基函数（RBF = Radial Basis Function）无约束模型是

$$Z(\mathbf{x})=\sum_{i=1}^{n}\lambda_i\varphi_i(\mathbf{x})=\sum_{i=1}^{n}\lambda_i\varphi\left(\left\|\mathbf{x}-\mathbf{x_i}\right\|\right), \tag{31}$$

其中，$\varphi\left(\left\|\mathbf{x}-\mathbf{x_i}\right\|\right)$ 是以 $\mathbf{x_i}$ 为中心的径向基函数。

空间数据的径向基函数有约束模型是

$$Z(\mathbf{x})=\sum_{i=1}^{n}\lambda_i\varphi\left(\left\|\mathbf{x}-\mathbf{x_i}\right\|\right)+c_0+c_1\mathbf{x},\ s.t.\ \sum_{i=1}^{n}\lambda_i=0,\ \sum_{i=1}^{n}\lambda_i\mathbf{x_i}=0, \tag{32}$$

其中，多项式 $p(\mathbf{x})=c_0+\mathbf{c}_1^T\mathbf{x}$ 是用来确保逼近精度的，约束条件是为了使全局光滑曲面曲率最小。

用 Lagrange 乘子法把（32）式转为无约束优化问题，对目标泛函求极小，最终得到如下矩阵方程

$$\begin{bmatrix}
\varphi\left(\left\|\mathbf{x_1}-\mathbf{x_1}\right\|\right) & \varphi\left(\left\|\mathbf{x_1}-\mathbf{x_2}\right\|\right) & \cdots & \varphi\left(\left\|\mathbf{x_1}-\mathbf{x_n}\right\|\right) & 1 & \mathbf{x_1} \\
\varphi\left(\left\|\mathbf{x_2}-\mathbf{x_1}\right\|\right) & \varphi\left(\left\|\mathbf{x_2}-\mathbf{x_2}\right\|\right) & \cdots & \varphi\left(\left\|\mathbf{x_2}-\mathbf{x_n}\right\|\right) & 1 & \mathbf{x_2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\varphi\left(\left\|\mathbf{x_n}-\mathbf{x_1}\right\|\right) & \varphi\left(\left\|\mathbf{x_n}-\mathbf{x_2}\right\|\right) & \cdots & \varphi\left(\left\|\mathbf{x_n}-\mathbf{x_n}\right\|\right) & 1 & \mathbf{x_n} \\
1 & 1 & \cdots & 1 & 0 & 0 \\
\mathbf{x_1} & \mathbf{x_2} & \cdots & \mathbf{x_n} & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ c_0 \\ c_1
\end{bmatrix}=
\begin{bmatrix}
f(\mathbf{x_1}) \\ f(\mathbf{x_2}) \\ \vdots \\ f(\mathbf{x_n}) \\ 0 \\ 0
\end{bmatrix}, \tag{33}$$

其中，$f(\mathbf{x_i})$ 是已知点的值。用共轭梯度法求解，得到系数 $\lambda_1,\lambda_2,\cdots,\lambda_n$。然后可以基于最佳模型进行空间数据的预测。

空间数据的径向基函数无约束预测是

$$\hat{Z}(\mathbf{x_u})=\sum_{i=1}^{n}\lambda_i\varphi\left(\left\|\mathbf{x_u}-\mathbf{x_i}\right\|\right). \tag{34}$$

空间数据的径向基函数有约束预测是

$$\hat{Z}(\mathbf{x_u})=\sum_{i=1}^{n}\lambda_i\varphi\left(\left\|\mathbf{x_u}-\mathbf{x_i}\right\|\right)+c_0+c_1\mathbf{x_u}. \tag{35}$$

很多文献列出了不同的径向基函数可供选择。但是，很显然径向基的不同在很大程度上决定了空间数据的建模预测效果，最好选择能与空间数据变化特征一致的径向基函数。

7 勘探地震中多信息融合的基本应用

油气地震勘探发展到现在，很多简单问题基本已经得到了很好的或较好的解决。余下的都是复杂问题。其中最核心问题两个当然依然是：宽带弹性参数估计或广义地震波反演成像；定量的油藏描述和含油气性评价。这两个问题都是强非线性的，必须引入多种先验信息才能提高它们的求解精度。

但是，一些先验信息的获取又是一个需要先验信息约束的（强）非线性反问题。譬如强噪音下弱初至波的识别与走时检测；低信噪比情况下的层位识别与断层识别；成像剖面上的地震相识别；进一步地，结合岩心数据和井数据把地震相合理地变成地震岩相；散乱的井数据支撑下的弹性参数建模（生成微井数据）。

所有这些问题，本质上，都可以归结为强非线性情形下的建模和预测问题。引入多源（元）信息的约束是必不可少的。

事实上，最佳模型构建的理论框架是比较明了和确定的，研究也是很深入的。但是，多源（元）信息如何有机地融合起来，有效地约束建模结果，提高建模的质量，进而提高预测结果的精度，反而成了薄弱的问题。

多源（元）信息如何有机地融合起来，真正的困难还是来自于多源（元）信息的异质性、尺度差异、精度差异、模糊性及随机性等。因此，很难构建一个统一的、公认的理论框架来处理这个问题。

我们认为：当前基本的认识是首先要统一地表达这些复杂的数据，比较一致的观点是形成随机（数据）集。然后，建立随机（数据）集满足的概率密度函数。接着对随机（数据）集进行建模，就是估计概率密度函数中包含的参数。最后，基于建立的最佳模型进行推断或判决，从而解决各种场景下的实际问题。

前面罗列的勘探地震中的典型应用场景，本质上，都是依据上述基本认识进行的。多源（元）信息都是融合体现在最佳建模的过程中。把多源（元）信息约束下的最佳建模问题提成一个合理的约束优化反问题是基本的数学操作。具体的应用场景下，多源（元）信息融合有不同的做法。

8 结论与讨论

众所周知，勘探地震中的两个核心问题：宽带弹性参数估计或广义地震波反演成像；定量的油藏描述和含油气性评价，都是一个信息不足情况下的最佳判决问题。前者的目标是得到最佳的弹性参数估计结果，后者是得到尽可能可靠的油藏描述和评级结果。它们都是强非线性。

解决这些强非线性问题，引入相关的先验信息是必须的。勘探地震中这些先验信息主要包括：与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数据。这些要么是空间散乱点处的，要么是规则三维空间网格点上的，且是二进制表示的数据。而地质知识和岩石物理关系这些非二进制表达的信息也是非常重要的与待估计参数相关的先验知识。这些先验信息表现出的异质性、尺度差异、精度差异、模糊性及随机性等等是它们具有的典型特征。因为这些先验信息都是空间域的，空间数据的统计分析不可避免地成了空间信息融合的数学基础。

目前并没有统一地利用这些先验信息的数学框架。但是，首先要统一地表达这些复杂的数据，然后建立统一的数据集所满足的概率密度函数。接着对概率密度函数中包含的参数进行估计。最后利用所建立的最佳模型进行推断，解决各种具体问题。这应该是比较抽象的、也比较统一的多源（元）信息融合的理论逻辑。多源（元）信息融合体现在上述建模过程中。

勘探地震中，多源（元）信息融合的具体应用，目前主要还是把多源（元）信息约束下的最佳建模问题提成一个合理的约束优化反问题，通过求解这个约束优化反问题达到多源（元）信息融合并约束提升建模结果精度的目的。

参考文献

[1] Gamerman, D., 2006, Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference；Chapman & Hall/CRC
[2] Cressie, N. A. C., 1993, Statistics for Spatial Data; John Wiley & Sons, Inc.
[3] Schabenberger, O., and Gotway, C. A., 2005, Statistical Methods for Spatial Data Analysis; Chapman & Hall/CRC

[4] Mahler, R. P. S., 2007, Statistical Mutilsource – multitarget Information Fusion; Artech House, Inc.
[5] Evensen, G., 2009, Data Assimilation: The Ensemble Kalman Filter; Second Edition, Springer

反射波 FWI 在陆上地震数据成像处理中的实用化途径探讨

王华忠，吴成梁，许荣伟，聂振波

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】在“两宽一高”叠前地震数据采集和超级算力的支持下，Bayes 参数估计理论下的 FWI 成为地震波成像的核心方法技术已经逐渐变成现实。但是，由于陆上地震数据的品质差，事实上即便是针对海上地震数据，FWI 也很难取代常规地震数据处理流程，单独地给出有效的成像结果。目前，在稍微复杂的地质情况下，FWI 能为偏移成像提供可靠的速度场已经很不错了。总体而言，利用炮集中初至波（早至波）的透射波 FWI（TFWI = Transmission FWI），如果初至波（早至波）没受到严重的噪音干扰（譬如在深水、长偏移距情形下），给出的速度估计结果还是可以较好地提升偏移成像精度的。但是，利用炮集中的反射波，无论是海上数据还是陆上数据，都没有很好地证明反射波 FWI 能够有效提升偏移成像质量的（各向异性）速度场。反射波 FWI 出现这样的情况，并不是它的理论存在问题，而是它的理论假设没有得到较好的满足所导致。可以肯定地讲，Bayes 参数估计理论下的 FWI 方法理论是完美的，今后的地震波成像必定沿着 FWI 方法理论继续发展下去。今后要做的是，在当前各种条件的限制下，如何最大程度地得到 FWI 方法理论预期的成像结果。因此，本文展开反射波 FWI（RFWI = Reflection FWI）在陆上地震数据成像处理中的实用化途径的探讨。指出反射波 FWI 最核心的问题是克服 Cycle-Skipping（周期跳）现象。但是，从具体实现的角度，我们认为主要应该关注如下三个关键点：1、“同因之果”的时差测量；如果能解决这个问题，Cycle-Skipping 问题基本可以被解决了。2、RFWI 中，层析梯度项和偏移梯度项的有效分解。3、梯度分解基础上的正则化信息提取与施加。我们认为：解决了这三个关键点，反射波 FWI，即便针对陆上数据，也可逐步地得到实用化。我们认为：FWI 的计算框架已经比较明确了，今后 FWI 的方法技术发展重点正在逐渐转到如何让 FWI 在工业界得到有效的应用。针对 FWI 的实用化问题，与待估计参数相关的多源信息如何提取和表达，如何合理地对反演解进行正则化约束，从而提升 FWI 反演解的精度和分辨率，会成为今后 FWI 方法技术发展的焦点。

【关键词】Bayes 参数估计，透射波 FWI，反射波 FWI，周期跳，“同因之果”的时差测量，梯度分解，梯度正则化，RFWI 实用化

1 前言

油气地震勘探经历一百多年的发展，地震数据采集已经进入了“两宽一高”的时代，而地震波成像，自从 Hagdoorn（1954）阐述了地震波成像的内涵，Claerbout（1971）提出波场外推+成像条件的波动方程偏移成像，Tarantola（1984）把地震波成像奠基在 Bayes 参数估计的理论基础上，也逐渐地归结到 FWI 成像方法技术上。这是任何广义遥感领域，即根据系统外的实测数据推测系统内物理参数分布的相关领域，都必然归结出的统一的问题，即系统参数估计问题。尽管有不同的应用领域，不同的应用场景，也发展了看起来很不相同的数据分析方法，但系统参数估计的本质没有变。Bayes 参数估计为它们奠定了统一的理论基础。

可以对 FWI 做这样的定位：它是 Bayes 参数估计理论在油气地震勘探中的一个应用实例。也可以预测到，今后的地震波成像，会沿着 FWI 理论方法不断地深化、拓展，并不断地实用化，这是不可避免的。

众所周知，实际的地下岩石介质的性质是十分复杂的，不同尺度（扫描电镜尺度，实验室薄片尺度、钻井岩心尺度、测井尺度和地震波长尺度）上表现出相当不同的特征。勘探地震中，是波前面上的地震子波在感知地下介质岩性的变化，检波器接收到的同相轴上地震子波的波形变化携带了地下介质岩性变化的信息。在地震波特征波长尺度上，地震子波感知到的是地下介质岩石物性的某种平均的和等价的变化。波动方程的推导过程中，对地下介质做了很多的理想假设。波动方程仅能近似地模拟或解释波在实际地下中的传播过程及结果。所有反问题的根本就在于实测波场与地下介质参数之间的关系是否是精确的、可靠的，还在于实测波场与待估计的地下介质参数之间的关系是否是（强）非线性的。对于实测波场与待估计的地下介质参数的关系，待估计参数的尺度就是个重要的因素。小尺度、强跃变的待估计参数与实测波场之间的关系往往是（强）非线性的。一般地，高波数的待估计参数成分与实测波场之间存在更强的非线性性，光滑部分则反之。这一现象解释了为何地震波反演成像在分辨率上存在局限，尤其是高波数部分的估计精度较低的根本原因所在。FWI 的真正困难，就在于复杂介质（譬如复杂地表、复杂构造、小尺度和/或强变化的地质体存在）情形下，它是一个强非线性反问题。强非线性反问题的多解性、不稳定性使得这类方法技术很难实用化。而假设已知正确的偏移速度模型的叠前地震波偏移成像问题，从数学算法上讲，不是一个严格意义上的反问题。因此，它有很强的稳健性，在勘探地震中得到了广泛的应用。但是，给出正确的叠前偏移成像的前提是假设已知正确的偏移速度模型，如果把构建偏移速度模型也放在广义的地震波成像过程中，广义的地震波成像就成了一个标准的反问题。在复杂介质情形下，同样很难得到正确的（或合理的）反演解。复杂探区叠前深度偏移成像结果解决不了地质问题，弄不清真实的地下地质结构形态，本质

原因就是此时的广义地震波成像是个（强）非线性反问题。

求解（强）非线性反问题的最基本、最合理的做法还是把它转化为一组凸问题来解决，并通过引入待估计参数的更多的先验信息缩小解空间、提高解的稳定性。当然，尽可能采集得到符合 FWI 理论假设需要的叠前地震数据体、仅仅获得待估计参数的光滑部分放弃估计其高波数部分，也是解决（强）非线性反问题的重要途径。

当前关于 FWI 用于（各向异性）速度建模的现状，我们的基本判断是：仅仅在海上（尤其在深水探区）低频长偏移距观测数据情形下，透射波 FWI（TFWI）得到了真正有效的应用，可以有效地改善后续叠前深度偏移成像的质量。TFWI 就是 FWI 的一个凸的子问题；而到目前为止，在实际数据情况下，反射波 FWI（RFWI）给出的（各向异性）速度模型，对叠前深度偏移成像整体的提升效果不显著，也不稳定（有些探区有效果，有些探区效果不明显，也有质量下降的情况出现）。这是 FWI 用于速度建模的基本情况。

理论上，FWI 是把宽波数带弹性参数（譬如宽波数带纵波速度场，宽波数带波阻抗）估计作为目标的。油气地震勘探也需要这种的宽波数带弹性参数估计结果，可以大幅度提高油藏描述和油气预测的精度。

因此，国内外各大石油公司都在推动 FWI 技术向估计宽波数带弹性参数的方向发展，尽管在复杂探区，TFWI 和 RFWI 用于背景速度建模还存在很大的困难。关于 FWI 用于宽带弹性参数估计的现状，可以总结如下：CGG、TGS 和 SLB 公司坚持走时延 FWI（Time Lag FWI）（Zhang 等，2018；Wang 等，2019；Mao 等，2020）+FWI_Imaging（Zhang 等，2020；Wang 等，2021，Wei 等，2021，Rayment 等，2022）的技术路线；PGS 和 DUG 公司坚持背景速度与反射系数同时反演的技术路线（Yang 等，2020，2022；Rayment 等，2022a，2022b；Cheng 等，2023；Roy 等，2023）；我们（波现象与智能反演成像研究组（WPI））提出了特征波反演成像（CWI）（王华忠等，2017；吴成梁，2022）+宽带波阻抗建模/重构（WBIM）(王华忠等，2023；Lei 等，2023）的技术路线。目标都是期望得到宽带的参数估计结果。前两者背景速度和高波数参数（速度扰动或反射系数）都是单独给出的，没有给出宽带参数估计结果。后者给出了宽带阻抗的建模结果。

理论分析知，FWI 在石油工业中的实用化最关键的是反射波 FWI 在估计低中波数（速度）参数的估计中发挥作用，这是目前存在的最关键的问题。高波数参数估计方法，即 LS_RTM，尽管也存在问题，但是在满足 Born 线性化近似假设后，基本上是一个凸问题。

因此，本文集中展开反射波 FWI（RFWI = Reflection FWI）在陆上地震数据成像处理中的实用化途径的探讨。指出反射波 FWI 最核心的问题当然是克服 Cycle-Skipping（周期跳）现象。但是，从具体实现的角度，我们认为主要应该关注如下三个关键点：① “同因之果”的时差测量。如果能解决这个问题，Cycle-Skipping 问题基本可以被解决了。② RFWI 中，层析梯度项和偏移梯度项的有效分解。③ 梯度分解基础上的正则化信息提取与施加。我们认为：解决了这三个关键点，反射波 FWI，即便针对陆上数据，也可逐步地得到实用化。

我们认为：FWI 的计算框架已经比较明确了，今后 FWI 的方法技术发展重点正在逐渐转到如何让 FWI 在工业界得到有效的应用。这也是本文探讨的重点问题。

2 Bayes 参数估计框架下的常规 FWI 的本质

Bayes 参数估计的根本思想是：假设已知实测数据与地下介质弹性参数之间的先验预测关系（即概率密度函数表达的不确定关系）和待估计参数满足的先验概率分布，上述两种先验分布，决定了已知实测数据与待估计参数之间的条件后验概率密度。该条件后验概率密度包含了待估计参数的所有统计特征。一般地，把条件后验概率密度的期望作为待估计参数的结果，其方差作为待估计参数的精度评价。

首先，有实测数据与地下介质弹性参数之间的先验预测关系，

$$\mathcal{M}:\ \boldsymbol{d}^{obs}=H\left(\boldsymbol{m}\right)+\boldsymbol{\varepsilon}\,, \tag{1}$$

其中，模型 $\mathcal{M}$ 包含了实测数据与地下介质弹性参数之间的先验预测关系的不确定性，用 $p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)=p_{\varepsilon}\left(\boldsymbol{d}^{obs}-H\left(\boldsymbol{m}\right)\right)$ 来表达。不确定性既由数据观测过程中的因素引起，也由正问题 $\boldsymbol{d}^{cal}=H\left(\boldsymbol{m}\right)$ 的不准确引起。ε 是正问题不能完全解释实测数据导致的误差。正是 ε 的随机性，刻画了实

测数据与地下介质弹性参数之间的先验预测关系的不确定性。一般地，假设某种概率密度函数刻画了这种不确定性，并假设该概率密度函数是已知的。

单纯地依赖实测数据与地下介质弹性参数之间的先验预测关系的不确定性进行参数估计是最大似然估计（MLE = Maximum Likelihood Estimation）学派的思想。但是，Bayes 参数估计学派认为，仅仅依赖实测数据与地下介质弹性参数之间的先验预测关系不可能得到可靠的估计结果，必须引入待估计参数的先验信息。该先验信息不可能是完全准确的，不确定性用某种先验概率密度函数来描述，可写为 $p\left(\boldsymbol{m}\middle|\mathcal{M}\right)$。很显然，参数 $\boldsymbol{m}$ 是正问题 $\boldsymbol{d}^{cal}=H\left(\boldsymbol{m}\right)$ 中包含的参数，与模型 $\mathcal{M}$ 是相关联的。

根据上述两种先验概率密度函数，依据 Bayes 公式，可以定义如下条件后验概率密度：

$$p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs};\mathcal{M}\right)=\frac{p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)p\left(\boldsymbol{m}\middle|\mathcal{M}\right)}{p\left(\boldsymbol{d}^{obs}\middle|\mathcal{M}\right)},\tag{2}$$

其中，$p\left(\boldsymbol{d}^{obs}\middle|\mathcal{M}\right)$ 是表示实测数据本身满足的不确定的概率密度函数。

概率论指出：条件后验概率密度 $p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs};\mathcal{M}\right)$ 包含了待估计参数 $\boldsymbol{m}$ 的所有统计特征。期望（一阶统计量）可以作为待估计参数 $\boldsymbol{m}$ 的估计结果；方差（二阶统计量）视为待估计参数 $\boldsymbol{m}$ 的估计精度评价。由于勘探地震中，待估计参数 $\boldsymbol{m}$ 有非常高的维度，积分求条件后验概率密度规定的期望和方差基本是不可能的。

因此，条件后验概率密度的最大化准则（MAP = Maximum A Posteriori Estimation）成了常用的参数估计准则。即

$$\hat{\boldsymbol{m}}=\arg\min_{\boldsymbol{m}}p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs};\mathcal{M}\right)=\arg\min_{\boldsymbol{m}}\left\{p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)p\left(\boldsymbol{m}\middle|\mathcal{M}\right)\right\}\tag{3}$$

一般地，两个先验概率密度函数 $p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)$ 和 $p\left(\boldsymbol{m}\middle|\mathcal{M}\right)$ 都被假设是指数型的。已知指数型函数的对数不改变极值点的分布，因此，（3）式被写为：

$$\hat{\boldsymbol{m}}=\arg\min_{\boldsymbol{m}}\left\{-\ln p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)-\ln p\left(\boldsymbol{m}\middle|\mathcal{M}\right)\right\}.\tag{4}$$

显然，（4）式逐渐转化为最优化问题。解析数学家定义如下形式的最优化问题：

$$\mathcal{J}\left(\boldsymbol{m}\right)=Q\left(\boldsymbol{d}^{obs},H\left(\boldsymbol{m}\right)\right)+\lambda\Omega\left(\boldsymbol{m}\right),\tag{5}$$

其中，$Q\left(\boldsymbol{d}^{obs},H\left(\boldsymbol{m}\right)\right)$ 被称为数据偏差项，度量正问题 $\boldsymbol{d}^{cal}=H\left(\boldsymbol{m}\right)$ 预测实测数据 $\boldsymbol{d}^{obs}$ 的误差；$\Omega\left(\boldsymbol{m}\right)$ 定义待估计参数的先验信息。

很显然，若定义（4）式中的对应项为：

$$Q\left(\boldsymbol{d}^{obs},H\left(\boldsymbol{m}\right)\right)=-\ln p\left(\boldsymbol{g}\middle|\boldsymbol{m};\mathcal{M}\right),\tag{6a}$$

$$\Omega\left(\boldsymbol{m}\right)=-\ln p\left(\boldsymbol{m}\middle|\mathcal{M}\right).\tag{6b}$$

可以看出，数学分析学家定义的参数优化反演问题（一般地，称 Tikhonov 正则化问题）与两个先验概率密度函数假设为指数分布的 MAP 估计问题是一致的。

更进一步地，在线性 Gauss 假设下，即存在实测数据与地下介质弹性参数之间的线性预测 $\boldsymbol{d}^{obs}=\boldsymbol{d}^{cal}+\varepsilon=H\boldsymbol{m}+\varepsilon$，并有如下的两个先验概率密度函数满足 Gauss 分布的假设，

对噪声的假设：

$$\varepsilon \sim N\left(\boldsymbol{0},\sigma_{\varepsilon}^{2}\boldsymbol{I}\right) \text{ 即 } p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)\propto \exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left\|\boldsymbol{d}^{obs}-H\boldsymbol{m}\right\|^{2}\right], \tag{7a}$$

和对 $\boldsymbol{m}$ 的假设：

$$\boldsymbol{m} \sim N\left(\boldsymbol{0},\sigma_{\boldsymbol{m}}^{2}\boldsymbol{I}\right) \text{即 } p\left(\boldsymbol{m}\right)\propto \exp\left[-\frac{1}{2\sigma_{\boldsymbol{m}}^{2}}\left\|\boldsymbol{m}\right\|^{2}\right], \tag{7b}$$

则对应的后验概率密度为

$$p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)\propto \exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left\|\boldsymbol{d}^{obs}-H\boldsymbol{m}\right\|^{2}+\frac{\sigma_{\varepsilon}^{2}}{\sigma_{\boldsymbol{m}}^{2}}\left\|\boldsymbol{m}\right\|^{2}\right], \tag{8}$$

类似地，有 MAP 估计下的反演解

$$\hat{\boldsymbol{m}}=\arg\min_{\boldsymbol{m}}\left\{p\left(\boldsymbol{m}\middle|\boldsymbol{d}\right)\right\}=\arg\min_{\boldsymbol{m}}\left\{\mathcal{J}\left(\boldsymbol{m}\right)\right\}, \tag{9}$$

其中，误差泛函定义为

$$\mathcal{J}\left(\boldsymbol{m}\right)=-\frac{1}{2}\left\|\boldsymbol{d}^{obs}-H\boldsymbol{m}\right\|_{\sigma_{\varepsilon}^{2}}^{2}+\lambda\left\|\boldsymbol{m}\right\|^{2}, \tag{10}$$

超参 $\lambda=\dfrac{\sigma_{\varepsilon}^{2}}{\sigma_{\boldsymbol{m}}^{2}}$。

（10）式的解可以写为：$\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\sim\mathcal{N}\left(\hat{\boldsymbol{m}},\hat{\boldsymbol{\Sigma}}\right)$。其中，$\hat{\boldsymbol{m}}=\left(H^{T}H+\lambda I\right)^{-1}H^{T}\boldsymbol{d}^{obs}$，$\hat{\boldsymbol{\Sigma}}=\sigma_{\varepsilon}^{2}\left(H^{T}H+\lambda I\right)^{-1}$。

若实测数据与地下介质弹性参数之间的关系是非线性的，即 $\boldsymbol{d}^{obs}=\boldsymbol{d}^{cal}+\varepsilon=H\left(\boldsymbol{m}\right)+\varepsilon$，则 MAP 估计下的误差泛函定义为

$$\mathcal{J}\left(\boldsymbol{m}\right)=-\frac{1}{2}\left\|\boldsymbol{d}^{obs}-H\left(\boldsymbol{m}\right)\right\|_{\sigma_{\varepsilon}^{2}}^{2}+\lambda\left\|\boldsymbol{m}\right\|^{2} \tag{11}$$

Newton 迭代法用来求解（11）式，

$$\hat{\boldsymbol{m}}^{k+1}=\hat{\boldsymbol{m}}^{k}+\left[\frac{\partial^{2}\mathcal{J}}{\partial\boldsymbol{m}^{2}}\right]_{\hat{\boldsymbol{m}}^{k}}Grad\mathcal{J}\Big|_{\hat{\boldsymbol{m}}^{k}} \tag{12}$$

其中，$\left[\dfrac{\partial^{2}\mathcal{J}}{\partial\boldsymbol{m}^{2}}\right]_{\hat{\boldsymbol{m}}^{k}}$ 是误差泛函 $\mathcal{J}$ 在 $\hat{\boldsymbol{m}}^{k}$ 处的二阶导数项（Hessian 矩阵），$Grad\mathcal{J}\big|_{\hat{\boldsymbol{m}}^{k}}$ 误差泛函 $\mathcal{J}$ 在 $\hat{\boldsymbol{m}}^{k}$ 处的梯度。常规的 FWI 就是求解（12）式。

Bayes 参数估计的基本逻辑框架基本上就是这样。可以看出，炮集波场用算子形式的波动方程 $\boldsymbol{d}^{cal}=H\left(\boldsymbol{m}\right)$ 来预测时，对应的参数估计问题就是 FWI。若引入 Born 近似后的正问题 $\boldsymbol{d}^{cal}=H\boldsymbol{m}$ 预测炮集波场，对应的参数估计问题就是最小二乘逆时偏移（LS_RTM）。

应该注意到，Bayes 参数估计的基本逻辑也为地震信号分析、层析成像奠定了理论基础。实质上，也包含了叠前深度偏移成像的背后理论。

勘探地震发展到现在，发展到“两宽一高”地震数据采集、大数据/AI 和超级算力时代，地震波成像聚集在 Bayes 参数估计框架下的 FWI 方法技术上，具有充分的理论依据。

3 反射波 FWI 实用化存在的关键问题

当前 FWI 在实际应用中存在的问题是明显的。一是如何用好实测炮集中的初至波和早至波，我们称它们为透射波。所谓透射波就是说这些波没有经过反射或散射直接到达检波器。对应地，有透射波 FWI（TFWI）。透射波 FWI 主要利用的是长偏移距情形下的 Diving Wave。它有效穿透的深度，基本上决定了透射波 FWI 建模的深度。一般地，最大偏移距的五分之一到三分之一是 TFWI 建模的深度范围。二是如何用好一次反射波和散射波。很显然，一次反射波和散射波 FWI 实用化的难度依然是如何检测出一次反射波和散射波的震相及对应的走时，尤其是如何定义“同因之果”的实测炮集中和合成炮集中一次反射波和散射波的走时差。所谓“同因之果”指的是实测炮集中和合成炮集中对应的一次反射波和散射波震相是同一个物理源产生的。失去了“同因之果”的走时差测量，反射波 FWI 就没有了数据基础。目前，只是期望提供给 RFWI 的初始模型是非常准确的，据此期望实测炮集中和合成炮集中的一次反射波和散射波震相相差在半个周期之内。这样的要求在复杂介质情况下是非常难以做到的。这应该是 RFWI 很难实用化的首要原因。三是探索应用多次反射波和多次散射波，当然首先是利用多次反射波，而且是特征反射层相关的多次反射波。毕竟把多次波走时信息丢弃，对于高精度速度估计一定会是一种损失。

根据上述分析，我们提出反射波 FWI（RFWI）实用化存在的关键问题。考虑完整性，我们也分析了透射波 FWI（TFWI）实用化遇到的障碍。

首先，对于透射波FWI（TFWI），在陆上（应该也包括海上浅水）油气探区，在超长偏移距情形下，初至波信噪比很低，很难识别出有效的初至波。另一方面，早至波成分异常复杂，浅层折射层产生的一次和多次折射波、水层中的导波等等，很难被透射波 FWI 利用。因此，即便透射波 FWI，在这样的情况下也很难发挥作用。我们的观点是：尤其在长偏移距情况下，尽可能利用目前先进的机器学习（ML）算法识别出强噪音覆盖下的弱初至波并尽可能准确地检测出它的走时，为后续的初至波走时 FWI 提供尽可能可靠的输入数据基础。否则，花巨资采集的宽方位长偏移距数据中的初至波信息就不能被充分地利用起来。可以说，对陆上和浅水区海上地震勘探而言，关于透射波 FWI 的实用化，核心问题也已经转移到如何在长偏移距情形下把强噪音覆盖的弱初至波识别出来并可靠地检测出它们的走时。这看起来是个老问题，但并没有得到有效地解决。

对于反射波 FWI（RFWI），实用化的真正困难，首先也是出在“同因之果”的时差测量上。而对于透射波FWI（TFWI），“同因之果”的时差测量看起来是不存在的，因为合成的透射波走时和实测的透射波走时，都是由同一个背景速度引起的。但是，可以想象到初始背景速度差异过大，同一炮检对对应的 Diving Wave 的波路径相差会很大，这也是非同因之果的体现。因此，透射波 FWI（TFWI）也要给一个比较好的初始速度。从此可以看出，梯度导引类的凸问题求解算法从根本上回避不了对初始模型有一定精度的要求。解决不了“同因之果”的时差测量问题，RFWI 也没有了数据基础。TFWI/RFWI 的基本计算框架，都已经探索很多。虽说不能说已经很完善，但已经不是主要矛盾。目前的主要矛盾，作者认为是缺乏可靠的、满足“同因之果”要求的时差测量结果。反问题中对梯度的正则化约束尽管也很关键，但这是次一级的矛盾。

众所周知，反射波 FWI（包括 TFWI）最核心的问题被归结为存在 Cycle-Skipping 现象。而 Cycle-Skipping 现象是初始速度距离真速度模型太远导致的。但是，从具体实现的角度，除了给出尽可能好的初始速度模型外，作者认为主要应该关注如下三个关键点：① “同因之果”的时差测量。如果能解决这个问题，Cycle-Skipping 问题可以被大大地弱化,收敛速度会大幅提升。② RFWI 中，层析梯度项和偏移梯度项的有效分解。数值实验表明：RFWI 梯度项分解也可以显著地提升 RFWI 的收敛效率和收敛效果。③ 梯度分解基础上的正则化信息提取与施加。继续提升 RFWI 的收敛效率和收敛效果就要对梯度项施加合理的，最好是有地质含义的正则化约束。有效的正则化信息的提取是关键，如何施加还是第二位的。另外，具有地质含义的几何结构信息是最主要的约束信息源。目前，最期望是的是能获取有效的关于地震相（甚至地震岩相）的先验信息。总之，作者认为：解决了这三个关键点，即便针对陆上数据，RFWI 也可逐步实用化。

更进一步，我们提出特征反射层相关多次反射波 FWI（CMRFWI = Characteristic Multiple Reflection FWI）的概念。特征反射层相关特征多次反射波 FWI 的基本逻辑步骤包括：①、初始背景速度+特征反射层构成的速度模型；②、特征反射层相关的特征多次反射波的模拟；③、特征多次反射波模拟波场震相与实测特征多次反射波震相的（非线性）时差测量；④、类似（12）式定义的迭代的特征多次反射波 FWI。然后，再更新初始背景速度+特征反射层构成的速度模型，重新进行下一轮的速度更新。期望这样迭代更

新的轮次不要太多。其中，能否得到有效的速度建模结果，核心依然在于特征多次反射波模拟波场震相与实测特征多次反射波震相的（非线性）时差测量上。

4 反射波 FWI（RFWI）在陆上地震数据成像处理中的实用化途径

FWI 的基本计算框架，都已经探索很多。虽说不能十分完善，但已经不是 FWI 的主要矛盾。目前，FWI（尤其是 RFWI）方法技术的主要矛盾已经转到如何使其在实际数据地震波成像中得到实用，能得到更好的成像效果，从而提高油藏描述的精度。

前已述及，反射波 FWI 的实用化，尤其是在陆上低信噪比数据情形下的实用化，最主要的问题是“同因之果”的时差检测。即实测炮集中和合成炮集中对应的一次反射波或散射波同相轴是同一个反射界面或散射体产生的。常规 FWI 中，无论 TFWI 或 RFWI 都不显式地考虑这个问题，仅仅是泛泛地要求不能存在 Cycle-Skipping 现象。当前，用于背景速度反演的 TFWI 和 RFWI 基本上都是基于地震波走时的 FWI，因此发展了各种走时检测的方法。表面上，大多都是基于相关的时延检测方法。隐含地，假定了检测的时延是符合“同因之果”要求的。

针对“同因之果”的时差检测需求，我们首先提出了基于特征反射结构的特征反射波时差检测方法。在实测炮集数据的叠前深度偏移成像剖面上用机器学习算法检测出特征反射结构（就是反射系数比较大的、主要的反射层），利用反偏移方法产生对应的特征反射波同相轴。类似的方式产生合成炮集数据中的特征反射波同相轴。然后用一般的相关法或 Wiener 最佳匹配方法测量它们之间的时差。从而实现“同因之果”时差检测的目标。

事实上，当前图像处理领域，两个模板之间的差异检测（首先也是它们之间的时空相位差异测量）是其中的一个核心问题。我们认为：应该发展一个统一的、任意模版之间的时差检测策略与对应的算法，并把这样的算法用在各种领域的不同应用场景中，而不是仅仅解决 FWI 中的时差测量问题。很显然，一般意义下的两个任意模版点点之间的时空相位变化是非线性的。这也是一个典型的反问题。我们统称为非线性时差检测问题。在图像处理和机器学习领域，非线性时差检测问题依然是一个没有得到很好解决的关键问题。

可以明显地看出，在成像道集中，用上述非线性时差检测方法，检测出特征反射层的成像深度差，然后恢复成炮集规定的时间差。就可以很好地实现“同因之果”时差检测。因为，在成像道集中，一个典型反射层（或散射体），不同方位偏移距的同相轴没有排齐，存在成像深度差，除了说明偏移速度不正确外，也表明这个深度差是同一个反射层（或散射体）由于偏移速度不正确所产生的处于不同深度的像。把这个深度差用合理的方法转化为时间差，当然是满足“同因之果”要求的。

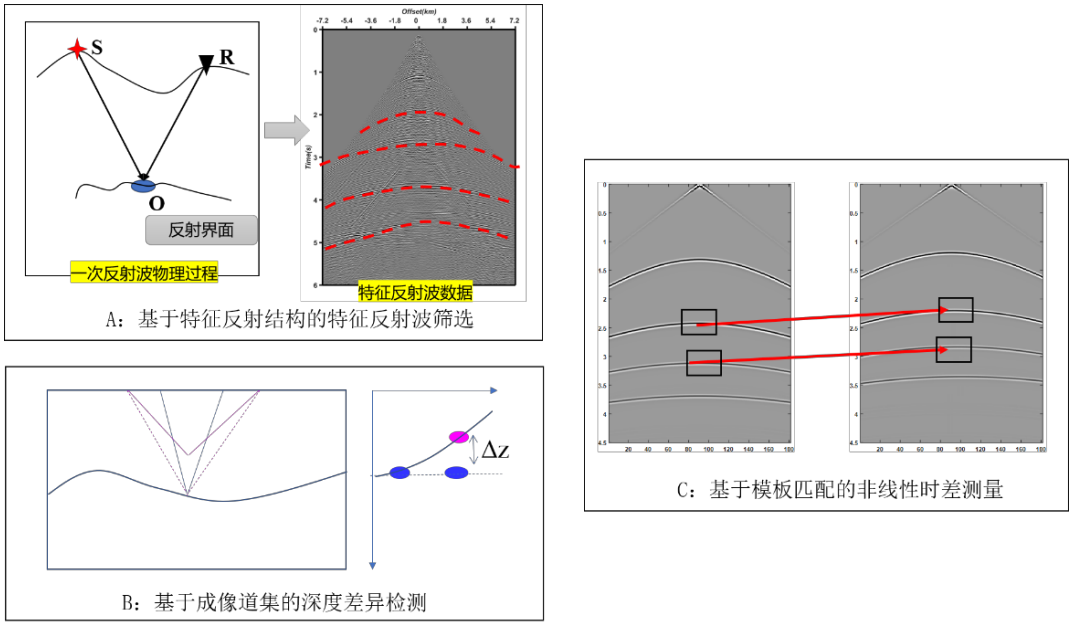


图 1：三种途径解决“同因之果”的时差检测问题示意图

另外，针对两个对应的实测炮集和合成炮集，要用上述非线性时差检测方法检测出满足“同因之果”要求的时差，就必须首先解决特征反射同相轴的对应问题。作者认为：通过目前先进的机器学习算法，自动识别出稀疏的特征强反射同相轴，也可以实现满足“同因之果”要求的时差检测。总之，满足“同因之果”要求的时差检测问题是 RFWI 实用化的首要的关键环节。作者认为：那些认为 RFWI 可以取代常规处理流程，可以不用在数据预处理上花功夫，用一个“黑匣子”的 FWI 算法，就可以得到很好的地震波反演成像结果的想法和做法都是不现实的。

另外，到目前为止，大家都认识到速度模型不正确时，实测炮集和合成炮集中对应的局部同相轴，不仅仅有时间方向上的纵向相位差（时间差），也有空间方向上的横向相位差（时间差）。很显然，纵向相位差和横向相位差的消失，代表了估计出的背景速度模型的正确。理论上，考虑纵向相位差和横向相位差的 FWI，收敛性应该更好。这是 FWI 的一个值得关注的技术发展方向。

假定已经解决了“同因之果”的时差检测问题，提高反射波 FWI 收敛性和收敛效率的第二个要点问题就是对 RFWI 的梯度项进行分解。众所周知，RFWI 的梯度项包含两部分：层析梯度项（“兔子耳朵”项）和偏移梯度项（“偏移画弧”项）。而且后者比前者的幅值强很多。理论上，在迭代更新背景速度时，只能用层析梯度项，偏移梯度项是不能贡献于低中波数速度更新的（当然，形成层析梯度项时需要依赖于偏移成像结果的！），而且它会严重干扰使用层析梯度项进行低中波数速度分量估计的收敛性。因此，必须进行 RFWI 的梯度项的分解，仅仅用层析梯度项更新背景速度。只有背景速度非常准确时，即层析梯度项消失或近于消失时，才可以开始利用偏移梯度项生成对高波数扰动成分的估计。这是反射波 FWI 算法迭代运行时应该遵循的基本逻辑。

当然，RFWI 的梯度项分解不是很容易实现的。首先，如果用偏移+反偏移生成层析梯度项，自然就不存在 RFWI 的梯度项分解问题了。否则，如果用共轭状态法计算 RFWI 的梯度项，就必须考虑用什么样的算法进行梯度项分解。其中最合乎逻辑的方法自然是波传播方向检测方法，因为我们知道层析梯度项的生成是任意空间点处正向传播的主震源波场与逆向传播的伴随源波场相关的结果，而偏移梯度项是在反射（散射）点处入射波场（正向传播的主震源波场）与反（散）射波场（逆向传播的伴随源波场）相关的结果。但是，在复杂介质情形下，要准确地检测出空间任一点处的波传播方向是很不容易的。目前发展了各种各样的方法，譬如 Poynting 矢量方法、光学流方法、结构张量方法、局部 Fourier 分解方法等等。由于层析梯度项是低波数变化的，而偏移梯度项是高波数变化的，因此可以用各种空间波数滤波的方法对二者进行分解。众所周知，在反演迭代算法中，迭代步数是可以作为一个超参使用的，就是说刚开始更新的都是待估计参数场中的低波数成分，随着迭代步数的增加，逐级地加入待估计参数场中的中高低波数成分。这一点可以从梯度是残差波场的反传播得到较好的理解。因此，对层析梯度项的施加，也要遵循这样的基本逻辑。

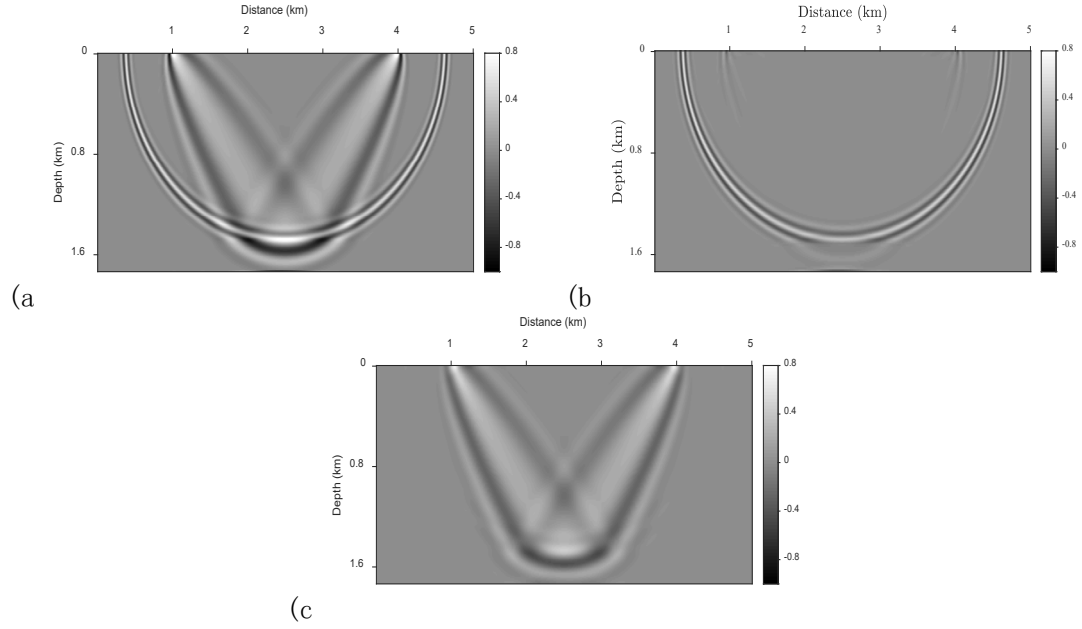


图 2：波数滤波方式分离梯度示意图：（a）原始梯度；（b）：高波数偏移画弧梯度项；（c）：低波数层析梯度项。

我们对地震波反演成像的定位是：地震波反演总是一个信息不足情况下的最佳判决/最佳推断问题。对于一个信息不足的反问题，引入更多的关于反演解的先验信息是必须的。因此，RFWI 中对梯度项引入正则化约束是降低其反演解的非唯一性，提高解的收敛性的必要措施。首先，由梯度导引类的迭代算法的本质可知，步长与梯度项是每一步迭代的更新量，可以认为梯度是归一化的参数更新量。RFWI 中，有单炮道集的梯度、多炮道集的梯度、所有炮的梯度，所有炮的梯度是物理意义最明确的，梯度方向也最正确、最合理的。正则化应该在施加在炮的梯度上。据此可以看出，为了加快计算效率的随机编码多炮 FWI 牺牲了每个迭代步梯度方向的合理性及正确性，这反而又降低了 FWI 的收敛性及收敛效率。这说明随机编码多炮 FWI 的综合效率如何还是值得考究的。对所有炮的梯度进行正则化约束？或是在此基础上对当前步估计的参数场进行参数正则化约束？或是兼而有之？如果正则化信息是可靠的，作者认为最好是兼而有之。与待估计参数场有关的先验信息，哪些信息成分是最重要的？能起到重要的作用？能显著提升反演解的收敛效率和稳定性？计算实践表明：与待估计参数场有关的几何结构信息最重要。有地质意义的与待估计参数场有关的几何结构信息是什么？特征反射结构、特殊地质体边界、断层结构、沉积模式。有地质含义的、最细致的几何结构可以认为是地震相。如何提取这些待估计参数场有关的几何结构信息及如何施加它们是正则化约束方法的核心内容。目前已经有非常多的文献讨论了这个问题。把井数据中蕴含的岩性参数值的信息，融入到地震相中，形成地震岩相，利用先验的地震岩相约束总梯度项或单步反演解，是更深层次的正则化方法与技术。待估计参数场有关的几何结构信息的提取，把井数据中蕴含的岩性参数值的信息融入到几何结构信息中，是一个空间信息提取、表达及融合问题。作者认为：今后 FWI 方法技术实用化发展过程中，这个问题反而会变成核心问题。

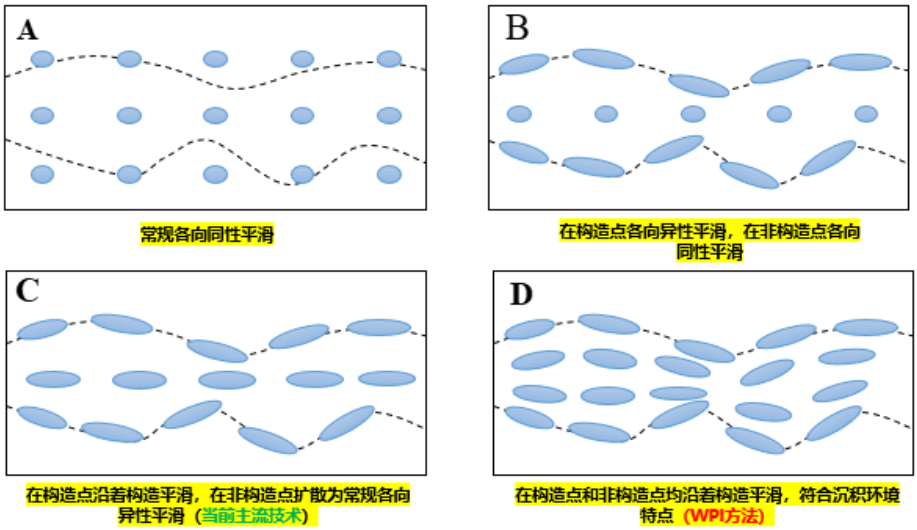


图 3：正则化：反射结构的利用示意图。

总而言之，我们认为，解决了满足“同因之果”要求的时差检测问题、RFWI 梯度分解问题、总梯度项或单步反演解的正则化约束问题，基本上可以解决 FWI（尤其是 RFWI）在陆上地震数据中的实用化。

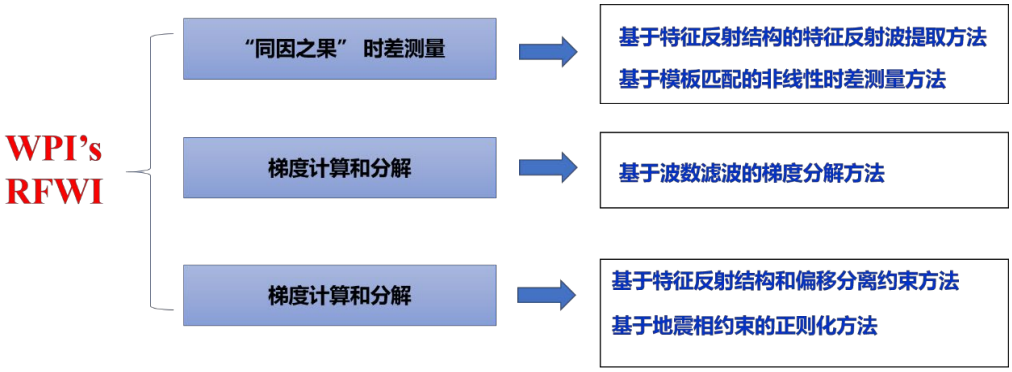


图 4：WPI 提出的针对陆上地震数据 RFWI 的实用化途径。

5 结论与讨论

在“两宽一高”叠前地震数据采集和超级算力的支持下，Bayes 参数估计理论下的 FWI 成为地震波成像的核心方法技术已经逐渐变成现实。但是，FWI 实用化，尤其在陆上和海上浅水探区，并没有达到 FWI 的理论期望。这并不是 FWI 的理论存在问题，而是它的理论假设没有得到较好的满足所导致的。

FWI 是一个信息不足情况下的大规模参数估计问题。信息不足体现在实测地震数据不完美，更体现与待估计参数相关的先验信息不足。

针对 FWI 的实用化问题，我们提出应该聚焦在三个问题上：①“同因之果”的时差测量。如果能解决这个问题，Cycle-Skipping 问题可以被大大地弱化，收敛速度会大幅提升。② RFWI 中，层析梯度项和偏移梯度项的有效分解。数值实验表明：RFWI 梯度项分解也可以显著地提升 RFWI 的收敛效率和收敛效果。③ 梯度分解基础上的正则化信息提取与施加。继续提升 RFWI 的收敛效率和收敛效果就要对梯度项施加合理的，最好是有地质含义的正则化约束。有效的正则化信息的提取是关键，如何施加还是第二位的。另外，具有地质含义的几何结构信息是最主要的约束信息源。目前，最期望是能获取有效的关于地震相（甚至地震岩相）的先验信息。

FWI 的基本计算框架已经定型，虽不能说十分完善，但已经不是 FWI 实用化的主要矛盾。目前，FWI（尤其是 RFWI）方法技术实用化的主要矛盾已经转到如何使其在实际数据地震波成像中得到实用，能得到更好的成像效果，从而提高油藏描述的精度。本文提出的 FWI（尤其是 RFWI）的实用化思想及技术途径，是 WPI 研究团队多年来对 FWI 研究后形成的认识，期望能促进 FWI 在陆上（当然也包括海上）地震勘探中的实际应用。

参考文献

[1] Cheng X, Xu J, Vigh D, et al. Enhanced template matching full-waveform inversion[C]//84th EAGE Annual Conference & Exhibition. European Association of Geoscientists & Engineers, 2023, 2023(1): 1-5.
[2] Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, 36, 467-418
[3] Guust Nolet (Edited), 1987, Seismic Tomography with applications in Global Seismology and Exploration Geophysics, D. Reidel Publishing Company
[4] Hagdoorn, J. G., 1954, A process of seismic reflection interpretation: Geophysical Prospecting, 06, 449-453
[5] Lei T, Wang H, Feng B. Broadband Acoustic Impedance Building via the Fusion of Sparsely Promoted Reflectivity and Background Acoustic Impedance[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-9.
[6] Mao J, Sheng J, Huang Y, et al. Multi-channel dynamic matching full-waveform inversion[M]//SEG Technical Program Expanded Abstracts 2020. Society of Exploration Geophysicists, 2020: 666-670.
[7] Rayment T, Dancer K, McLeman J, et al. High-resolution FWI imaging-an alternative to conventional processing[C]//83rd EAGE Annual Conference & Exhibition. European Association of Geoscientists & Engineers, 2022, 2022(1): 1-5.
[8] Rayment T, Dancer K, McLeman J, et al. High-resolution FWI imaging-an alternative to conventional processing[C]//83rd EAGE Annual Conference & Exhibition. European Association of Geoscientists & Engineers, 2022, 2022(1): 1-5.
[9] Rayment T, McLeman J, Dancer K, et al. Next-generation resolution through multiparameter FWI imaging[C]//Second International Meeting for Applied Geoscience & Energy. Society of Exploration Geophysicists and American Association of Petroleum Geologists, 2022: 922-926.
[10] Roy S, Halliday D F, Elbadry M, et al. Automation in full-waveform inversion: A way forward[C]//84th EAGE Annual Conference & Exhibition. European Association of Geoscientists & Engineers, 2023, 2023(1): 1-5.
[11] Tarantola, A., 1984b. Inversion of seismic reflection data in the acoustic approximation, Geophysics, 49, 1259 –1266.
[12] Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, Geophysics, 51, 10, 1893 – 1903.
[13] Tarantola, A.,1984a. Linearized inversion of seismic reflection data, Geophysical Prospecting, 32, 998 –1015.
[14] Wang B, He Y, Mao J, et al. Inversion-based imaging: from LSRTM to FWI imaging[J]. First Break, 2021, 39(12): 85-93.
[15] Wang P, Zhang Z, Mei J, et al. Full-waveform inversion for salt: A coming of age[J]. The Leading Edge, 2019, 38(3): 204-213.
[16] Wei Z, Mei J, Wu Z, et al. FWI imaging: Revealing the unprecedented resolution of seismic data[C]//SEG

International Exposition and Annual Meeting. SEG, 2021: D011S029R004.
[17] Yang Y, Arasanipalai S, Chemingui N, et al. Simultaneous inversion of velocity and reflectivity for enhanced seismic imaging[C]//Second International Meeting for Applied Geoscience & Energy. Society of Exploration Geophysicists and American Association of Petroleum Geologists, 2022: 832-836.
[18] Yang Y, Ramos-Martinez J, Whitmore N D, et al. FWI using reflections for deep velocity model updates[M]//SEG Technical Program Expanded Abstracts 2020. Society of Exploration Geophysicists, 2020: 691-695.
[19] Zhang Z, Wu Z, Wei Z, et al. FWI Imaging: Full-wavefield imaging through full-waveform inversion[C]//SEG International Exposition and Annual Meeting. SEG, 2020: D031S027R004.
[20] Zhang, Z., J. Mei, F. Lin, R. Huang, and P.Wang, 2018, Correcting for salt misinterpretation with full-waveform inversion: 88th Annual International Meeting, SEG, Expanded Abstracts, 1143–1147.
[21] 王华忠, 冯波, 王雄文, 等. 特征波反演成像理论框架[J]. 石油物探, 2017, 56(1): 38-49.
[22] 王华忠, 吴成梁, 盛燊, 等. 针对岩性储层的定量地震波成像[J]. 石油物探, 2023, 62(1): 1-30.
[23] 吴成梁．特征反射波波动理论走时层析速度建模方法研究[D]．上海:同济大学,2022

高维数据分析的本质问题分析

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】当前，地震数据分析逐渐演化为一般高维数据分析下的一个子问题。所谓的高维数据分析本质上是从高维随机过程的采样数据体中提取信息，然后进行某种推断或判决，从而解决既定的问题。地震数据分析的核心问题是从地表观测的地震波场中推断出地下介质系统弹性参数的分布（即地震波成像），目前的统一称谓是 FWI（Full Waveform Inversion），它也可被认为是 Bayes 参数估计在勘探地震中的应用。因为地震观测数据的不完美，也因为描述地下介质系统弹性参数与地表观测波场之间的关系存在强非线性以及这个关系存在不确定性，地震波反演成像奠基在 Bayes 参数估计理论基础上是不可或缺的。关于待反演参数的先验信息的引入就是必然的，而各种先验信息都是从高维数据体中提取出来的。主要是这个原因导致了地震数据分析逐渐与一般高维数据分析融为一体。事实上，基于概率统计的高维数据分析中已经包含了系统参数估计问题。高维数据分析中的回归分析问题中已经包含了系统参数估计问题。因此，本文对高维数据分析的本质问题展开深入的分析。首先，高维数据被视为一个高维随机过程，它是高维系统的输出。然后，相邻随机变量之间存在相关性，使得高维数据中蕴含一些非线性及线性的结构。实质上，它们是高维空间中的曲线或曲面（局部上，它们是直线或平面）正是这些几何结构的存在，预示着高维数据体中蕴含着一定的信息。接着对高维数据中蕴含的几何结构进行建模。基于构建的最佳模型进行预测（即统计推断），就可以解决实际应用中的各方面的问题。这就是本文要表达的高维数据分析的本质思想。

【关键词】地震数据分析，系统参数估计，地震波反演成像，高维数据分析，高维随机过程，线性及非线性结构，高维数据建模，统计推断

1 引言

大数据和人工智能时代，高维数据（统计）分析（也可以称多元数据（统计）分析）已经成为广泛的议题。为什么会是这样呢？因为无论是物理世界或是非物理世界，最基本的逻辑是：n 个因素共同作用下，“系统”产生响应。基于这样的激励+响应机制，构建一个模型，基于此模型，进行实际过程的预测，然后用于决策，从而解决实际问题。这是一个深层次的、一般规律性的认识世界+改造世界的做法。

n 个因素共同作用下，“系统”产生响应。这个过程可以用一个 n 维函数表达（确定性的表达）；也可以用一个概率模型表达（不确定性表达）。基于“系统”响应的实际观测数据（或采样数据）对这个过程建模，是高维数据分析的根本问题。大数据时代，当然是用数据驱动的方式进行建模。事实上，人类社会认识自然和社会规律的思维模式已经逐渐转向通过数据采集/收集、数据分析、建立模型、统计推断并解决问题的模式上来。通过受控或不受控的（实验）观察/观测，建立因果的函数关系（一般地表现为数学物理方程，而且这个方程表达了一类问题的统一遵循的因果规律。当然这样的因果规律只是在一定的假设下是成立的，能进行准确的预测的）的思维模式逐渐退居幕后。这反映了大数据时代，人类认识世界的思维方式在发生着改变。

很显然，对观测到的高维数据进行分析，就成了关键的问题。

高维数据分析的基本思想与笛卡尔引入坐标系并在其中表达几何结构的思想是密不可分的。可以说，没有笛卡尔引入坐标系的概念，就没有当前高维数据分析的思想与方法。

勒内·笛卡尔 1596 年 3 月 31 日生于法国安德尔-卢瓦尔省的图赖讷，1650 年 2 月 11 日逝世于瑞典斯德哥尔摩，是世界著名的法国哲学家、数学家、物理学家。他对现代数学的发展做出了重要的贡献，因将几何坐标体系公式化而被认为是解析几何之父。他还是西方现代哲学思想的奠基人，是近代唯物论的开拓者且提出了“普遍怀疑”的主张。黑格尔称他为“现代哲学之父”。他的哲学思想深深影响了之后的几代欧洲人，开拓了所谓“欧陆理性主义”哲学。堪称 17 世纪的欧洲哲学界和科学界最有影响的巨匠之一，被誉为“近代科学的始祖”。“我思故我在”！

笛卡尔引入坐标系，支撑起一个抽象空间的概念，是划时代的思想创举。据此构建起了抽象的“高维空间”概念。在此思维模式下，很多问题变得形象且直观！高维空间中到底有什么？数学上，尤其从解析几何看，高维空间中有几何结构。飘在高维空间中的几何结构到底应该是什么？无非是空间曲面、空间曲线。高维空间中的实体由空间曲面封闭形成。

从高维数据的角度看，所谓的高维数据中包含信息就是对应的高维数据空间中蕴含有一定的几何结构。但是，“飘”在高维数据中的几何结构该如何理解？本质上是相邻的高维数据取值大小相近，更准确的讲是相邻的随机变量存在相关性。由相邻的高维数据的（相近）取值体现出来的相关性，指示了高维数据中

包含某种信息。高维数据中包含信息就是其中潜藏着高维几何结构，即所谓的超曲线和超曲面。

高维空间的坐标向量就是高维数据的 n 个潜在影响因素形成的。不要机械地从时间、空间坐标的角度理解支撑起高维空间的基向量。

这就产生了本文关于高维数据分析的总结性观点：高维数据的本质表现形式就是飘在高维空间中的“几何结构”。可以认为“几何结构”是空间曲面、空间曲线。尤其在局部点处，“几何结构”可以是空间平面、空间直线。高维空间的基向量是（应该是）高维数据的特征向量（或潜在影响因素）构成。高维数据分析的本质就是对高维数据中蕴含的几何结构进行建模表达，并基于此进行统计推断，从而解决具体的实际问题。很显然，上述方法可以用在各种各样的应用场景中。

2 空间中的高维函数及分析方法

既然高维数据的本质表现形式就是飘在高维空间中的“几何结构”，而它们又是高维空间中的曲线和曲面。高维数据分析必然要用到高维函数的分析方法。高维函数的微积分提供了高维数据分析的思想和方法。

高维空间中的函数，就是多元自变量的函数，即高维函数或多元函数。数学分析明确指出：多元函数是 n 维笛卡尔空间中的空间曲面、空间曲线。空间直线和空间平面是空间曲线、空间曲面在局部点处的近似表达。

很显然，多元函数的微分学和积分学提供了高维函数的分析方法。

多元函数的微分学是高维函数分析的主体部分。局部点处，多元函数的连续性、可微性、各阶偏导数、方向导数、梯度、偏微分、全微分、极值、局部定义域中连续函数的性质、对函数 Taylor 展开（即一般的级数展开）并进行近似表达等是高维函数分析的主要内容。

高维函数的积分学是奠基在高维函数的微分学的基础上的。沿曲线的积分、曲面下体积的计算、封闭曲面的体积计算等也是高维函数分析的重要内容。但基本思想是明晰的。

进一步的，微分方程、积分方程、变分原理等是高维函数分析的进一步延伸。一般地，这些问题中，函数的维度涉及三维空间+一维时间，

无论如何，高维函数的微积分，构建了对高维函数进行分析的理论基础并提供了具体的分析方法。

更深入地开展高维函数的分析，泛函分析提供了更深刻的思想与方法。

3 高维数据线性分析的思想及分析方法

前已述及，尤其在局部点处，高维数据空间中的“几何结构”可以是空间平面、空间直线。更形象地，高维数据空间中“飘”着一些线性结构。从高维数据空间中“感知”到这些线性结构，并对这些线性结构进行建模表达，构成了高维数据分析的核心内容。

数学上看，高维数据中的线性结构可以被视为高维空间中的若干空间平面。事先选定若干空间平面结构（作为基函数），用这些空间平面结构的线性组合，在最小二乘意义下（也可以在基于高维数据的其他统计特征定义的误差泛函意义下），对实测高维数据（块）进行最佳逼近，就可以“感知”到实测高维数据（块）中的若干空间平面。据此解决相应的问题。

上述过程，也可以从数据矩阵和/或数据张量的角度，基于数据驱动的方式，对数据矩阵和/或数据张量进行特征分解，特征分解的结果可以“感知”到实测高维数据（块）中的若干空间平面。这种数据驱动的方式逐级成为普遍使用的高维数据线性分析方法。

这是高维数据线性分析的基本思想及分析方法。

4 高维数据分析的目的及分析方法

与高维函数分析（确定性分析）相比，高维数据分析的目的、思想与方法是很不相同的。高维数据分析的目的：寻找高维（随机采样）数据中存在的相关关系，对高维空间中的“几何结构”进行建模。基于模型进行预测，然后进行推断决策。从而解决相应的实际问题。

可见，对高维数据中的“非”线性结构进行建模表达是高维数据分析的本质目的。

高维数据分析的基本思想：实际采集到的物理数据及非物理数据，背后的影响因素（潜在自变量）可能是“n维”的。在这个“n维”空间中，“飘”一些空间几何结构（常被称为“流形”，即空间曲线或曲面）。这些空间几何结构或“流形”本质上是相邻的（随机）采样数据取相近的数值而体现出来的。从概率统计意义上，是随机采样数据的统计均值体现出这些空间几何结构。

把随机采样数据视为随机变量或随机过程。随机采样数据满足一定的概率分布函数。随机变量之间的相关性由二阶统计量描述，体现为高维空间的几何结构。因此，高维数据分析的基本思想可以抽象为：视实测数据为随机过程，它受“n维”潜变量的影响，形象地，“n维”空间中存在一些空间几何结构，对“n维”空间中“飘”的几何结构进行广义线性表达，构建最佳逼近的表达式。基于该最佳逼近表达式，进行预测推断。解决实际问题。

高维数据分析的基本方法：

方法 1 视实测高维数据为随机过程，对高维数据进行特征表达。一般地，是在升维后的空间中进行更为特征化的表达，以便抓住高维数据的更本质的特征。然后，直接解释高维数据的这些特征，从而解决实际问题。

认为张量分解高维数据分析方法是典型的方法。

方法 2 视实测高维数据为随机过程，首先对高维数据进行特征表达升维，形成“n维”空间，对“n维”空间中“飘”的几何结构进行广义线性表达。基于标签数据，建立回归分析关系式。利用该关系式，对实际数据进行预测。解决实际问题。这是典型的有监督学习算法。在神经网络算法中，广义的回归分析关系不是显式的，而是体现在网络系统的“权系数” 中。

有监督的神经网络算法，被认为是“非”线性结构表达的典型方法。“非”线性回归分析是常用的方法。

方法 3 视实测高维数据为随机过程，进行特征表达升维，形成“n维”空间，对“n维”空间中“飘”的几何结构，进行聚类分析。解决实际问题。这是典型的无监督学习算法。

Kernel PCA, Local Linear Embedding (LLE，局部线性嵌套)等方法是对高维空间中的“非”线性机构进行建模预测的典型方法。

关于仿生人脑的神经网络算法的学习机制及高维数据分析之间的关系进行如下的分析。首先认为：感知学习+特征表达+监督学习分类是神经网络算法的思想基础。

首先，感知的思想是什么？权系数向量$\boldsymbol{w}$与输入数据向量$\boldsymbol{x}$之间存在相关性是感知的基础（Hebb 学习规则）！感知过程本质上就是特征表达的过程。把实测数据，即输入数据向量$\boldsymbol{x}$，进行升维表达（在“n维”空间中表达）的过程。感知是通过内积、褶积实现的。内积是个匹配滤波器！把相似的、相关的东西过滤出来。基于感知结果的分类，把一个非线性分类问题转化为“拟线性”的分类问题。

通过“感知（由内积、褶积具体实现）”进行特征表达是神经网络算法的本质基础，然后与标签数据进行差异度量，然后反传播调整权系数。不断地用标签数据训练网络达到稳定的学习效果后，开始应用。

差异度量+反传播调整权系数的过程，尽管也很重要，但这是第二位的。

基于神经网络的高维数据分析的适用场景及局限性是什么？它能用于（大规模）参数估计中吗？我的观点是：神经网络算法进行（大规模）参数估计（即勘探地震中的 FWI）逻辑上不通！有些机器学习 ML 算法书上讲：参数估计问题可以被认为是连续分类问题。但是，没有理论证明连续分类可以用目前的 Bayes 分类下发展的各种方法来解决，譬如支持向量机（SVM）用于连续分类似乎是不可思议的。

勘探地震中，参数估计问题是基于因果关系（数学物理方程）进行的。线性系统参数估计的一整套反演解的概念讲得很清楚了。最简单的情形是：由线性正问题$\boldsymbol{Ax}=\boldsymbol{b}$通过求解线性误差泛函（二次型）对应的法方程$\hat{\boldsymbol{x}}=\left(A^TA\right)^{-1}A^T\boldsymbol{b}^{obs}$得到线性系统的解的估计结果。其中$\boldsymbol{b}^{obs}$是实测地震波场。

标签数据（炮集波场与弹性参数模型形成的标签数据）之间存在的大多是相关关系，基于相关关系能进行参数估计吗？如果能，理论依据是什么？到目前为止，勘探地震中，基于深度神经网络及标签数据进

行地震波成像，没有实用化的判例。作者认为：放弃波场与地下介质弹性参数之间的因果关系，进行基于标签数据（炮集波场与弹性参数模型形成的标签数据）和深度神经网络进行参数估计，是丢弃了一个非常有效的先验约束信息，这与 Bayes 推断的思想是背道而驰的。

但是，基于数据间存在的相关关系进行统计推断（统计决策）是没有问题的，目前大数据（大样本数据）/AI 做的就是这样的事情。作者认为：勘探地震数据分析中，凡是具有大数据特征的数据的分析，应该都是可以用目前的神经网络机器学习算法的，譬如:层位识别、初至拾取、典型噪音的去除、速度谱的解释等。

5 地震数据分析基本思想与方法

首先，应该界定：地震数据分析的根本目的包括两个，一是从叠前炮集波场数据和待估计的弹性参数的先验信息中获得尽可能精确的弹性参数估计结果（即地震波成像结果）；二是从弹性参数估计结果、钻井岩芯和测井结果、地质知识和岩石物理知识，对储层含油气性进行评价。

对炮集波场进行预处理，属于信号分析的内容，本质上就是对（最高）五维的观测数据中蕴含的（非）线性结构进行建模预测，达到去噪、数据规则化、解混叠等等目的。显然，这也是高维数据分析的核心内容。

地震数据分析本质上是一般高维数据分析的一个应用分支。它真正特殊的地方在于存在一个炮集波场与地下介质弹性参数之间的因果关系，即波动方程（以及它的各种变种）。

首先，应该认识到任何人为构建的方程都是建立在一定假设基础上的。真实的地下介质可以抽象为如下的岩石的物理模型：骨架+孔隙+流体。图 1 展示了该抽象的物理模型。事实上，骨架和孔隙的形状极为千变万化，孔隙流体也是极为多样。骨架的岩石性质也是多种多样。

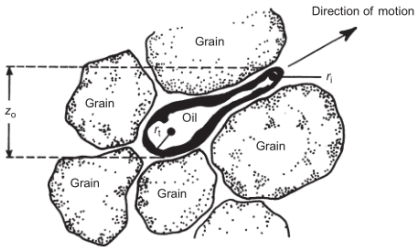


图 1 岩石的抽象物理模型

波动方程导出时，却认为地下介质是均匀的，理想弹性的（可以包含粘弹性）、没有尺度变化特征的。就是基于这样的假设，推导出描述波在地下岩石介质中的传播过程的数学物理关系式。这种因果的关系是地震数据分析的真正基础。Bayes 参数估计理论下，由实测数据估计介质系统的弹性参数形成了地震数据分析的核心目标。

基于波动方程（已知激发与系统对输出的描述波在等效介质中传播的数学表达）及各种变种（对波动方程的进一步简化）对实测波场的预测，在 Bayes 估计理论下，构建反问题，求解反问题，进行弹性参数估计（地震波成像）。

这样的系统参数估计方法有什么质疑的地方？真正令人质疑的是正问题的正确性！勘探地震中，存在三种典型的正问题：1、岩石物理模型---弹性模量与物性参数的关系；2、各种形式的波动方程；3、Zoeppritz 方程及其各种简化形式。它们都存在或多或少的问题，一般地岩石物理模型与地下介质的实际差异极大，Zoeppritz 方程在复杂介质和复杂储层情况下也越来越不适应。正问题的不确定性，使得反问题的解要么根本不存在，要么精度（严重）降低。基于当前所用的波动方程及变种，反演解不存在的情况基本可以排除。

在大数据和人工智能越来越普遍的今天，地震数据分析的一个另一个核心问题变得越来越重要，可以统称为 Learning From Data 的问题，也是 ML 的核心问题。当然，弹性参数估计也是 Learning From Data 的问题，只是它借助于一个确定的波动方程正问题。

它是基于地震数据中体现出的相关关系进行地震数据分析。基本思想是：分析（高维）地震数据的组

成成分，然后进行建模，接下来进行分类/聚类，最后进行推断和判决。

机器学习（ML）的原则是什么？

首先，认为随机采样数据满足缓平稳性假设。所谓缓平稳性假设，本质上是相邻的随机向量之间存在相关性。高维空间中，“飘”着几何结构。几何结构的建模表达方法---基函数的线性叠加。

ARMA/AR/MA 模型是对平稳信号建模的一种典型和特殊的方法。谐波叠加模型、平面波叠加模型是更为直观的模型。由 AR 模型、Kalman 滤波方法发展出来的自适应滤波方法是一般机器学习方法的起源。

ML 能从数据中学到什么？ML 的根本是“感知”到数据中包含什么成分或“感知”到数据中的特征成分（基函数）。

ML 是怎样学习的？深层的逻辑就是用基函数进行相关和匹配，找到数据中包含什么成分的根本逻辑。神经网络的（多层）感知机本质上就是在实现基函数的匹配，是一种数据驱动的基函数的匹配。众所周知，PCA/RPCA/SSA 是数据驱动方式感知数据中的基函数的典型算法。CNN 算法中的褶积，也是在进行“基函数”的匹配，提取数据中蕴含的特征。

目前，基于数据驱动的方式进行地震数据分析已经广泛地用在地震数据预处理（去噪、规则化、Deblending）、地震图像分析（层位拾取、断层识别等）中。

很显然，基于数据驱动的方式，进行数据的特征表达，然后进行分类/聚类，进而解决一定的问题，是合理的做法。但是，用深度神经网络用数据驱动的方式进行（大规模）弹性参数估计（地震波成像），不是合乎逻辑的方法技术选择。这样的做法说不清楚反演解的性质，无法评判反演解的精度，构不成一个理论体系。

6 地震数据分析的发展方向

勘探地震中的弹性参数估计是一个，永远会是一个，信息不足的反问题。在目前复杂介质、复杂储层、深层超深层目标这些情形下，高精度的参数估计，信息不足更为凸显。叠前数据观测方式不满足高精度参数估计的要求、数据品质差、波动方程正问题过于简单（但又不能用过于复杂的方程）等都是信息不足导致反演解进度下降的因素，在这样的情况下，与待估计参数相关的信息的补充显得弥足珍贵。提取待估计参数相关的先验信息，一般地，需要更为高端的和有效的数据驱动的 ML 算法。

作者认为：Bayes 参数估计理论与相关算法及其在勘探地震中的应用（即 FWI），我们已经探索很多，产生了比较深入的了解和认识。但是，Bayes 推断/决策理论及相关算法及其在（散乱或规则的）空间数据/图像数据分析中的应用，尤其是地质知识和岩石物理知识表达为待估计参数相关的先验信息，并把这些信息用来提升弹性参数估计的精度，进而提升油藏描述的精度，在地震数据分析中还要继续深化。总之，把地震波反演成像与油藏描述统一在 Bayes 推断/决策理论下，成为一个统一的问题逐渐地成为了必然要做的事情。

这样的做法，会把各种对油藏描述可能有贡献的相关信息都融汇在油藏描述的过程中。这应该是地震数据分析的正确方向。

这正是大数据和人工智能时代，地震数据分析必然要发生的变革。

7 总结与讨论

20 世纪中叶之前的物理科学，主要是寻找物理世界中的、能用数学物理关系表达的因果关系。目的是解决对物理世界的解释问题。“大”科学家对认识物理世界的规律更感兴趣，认为这才是科学问题。

基于“大”科学家提出的物理定律（数学物理关系），解决改造世界的问题，被认为是技术问题。“大”科学家不屑于做这样的事情。

但是，“大”科学家也无法在高维（n 维）空间中建立可行的数学物理关系，尤其是“n 维”因变量对结果的影响都不怎么占优的情形下，更无法总结出可用的物理规律。

此时，大数据分析就要登场了。

21 世纪以来，技术科学的发展，促进了数据收集/采集技术的极大进步，超大规模算力的计算机系统提

供了大数据分析所需的计算能力，于是，人们认识世界的方式发生了巨大的变化。

把数据视为高维随机向量（随机过程），在高维空间中，由于“相邻”随机变量的相关性，高维空间中呈现出“几何结构”（应该注意到：如果不相邻随机变量呈现出了明显的相关性，高维空间中的“几何结构”会异常复杂。所幸这不是高维数据的普遍特征）。这便是高维空间中的“函数关系”。对此“函数关系”进行回归分析建模或其他形式的建模，就可以对实际数据进行预测，从而解释各种邻域的、各种各样的实际问题。

这就是我理解的高维数据分析的本质。

参考文献

[1] T. W. Anderson, 2003, An Introduction to Multivariate Statistical Analysis; John Wiley & Sons, Inc.
[2] Trevor Hastie, Robert Tibshirani and Martin Wainwright, 2015, Statistcal Learning with Sparsity – The Lasso and Generalizations, CRC Press
[3] Lokenath Debnath and Piotr Mikusiński, 2005, Introduction to Hilbert Spaces with Applications, Third Edition, Elsevier.
[4] I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer
[5] Sergios Thwodoridis, 2015, Machine Learning – A Bayesian and Optimization Perspective; Elsevier
[6] 何晓群编著，2019 年，多元统计分析，第五版，中国人民大学出版社
[7] 伍胜健编著，2009 年，数学分析，北京大学出版社

随机模拟及随机反演的思想与方法

王华忠，许荣伟

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】到目前为止，地震波成像的基本理论框架已经十分清楚了，它在整个石油工业界的应用现状当然也更为大家熟知。随着地震数据采集技术的快速进步，地震波成像基本上逐渐聚焦在 Bayes 参数估计理论框架下。就是假设已知正问题和实际测量数据的不确定性的先验和关于待估计参数（即地震波成像结果）的先验，基于此生成关于待估计参数的后验（概率密度），最后按照该后验（概率密度）产生估计结果（地震波成像结果）。一般地，后验（概率密度）的期望（后验概率加权平均值）作为点估计结果，对应的方差作为估计结果精度的评价。实际上，大规模参数估计中，（超）高维积分求后验（概率密度）的期望是很难做到的。因此，提出用后验（概率密度）的最大值对应的样本值作为估计结果，即 MAP 准则。MAP 准则实际上得到的是众数。后验（概率密度）Gauss 情形下，期望值与众数值是相等的。很显然，这样的估计结果是某种真实待估计参数平滑后的结果，它们分辨率是不够的。后验（概率密度）Gauss 情形下，勘探地震中就产生了 FWI 方法。Newton 梯度导引类算法产生了 MAP 准则下的反演解。很显然，提高参数估计分辨率的方法应该是：在已知后验（概率密度）的条件下，生成满足该分布的随机样本，MAP 附近的样本值，都可以作为参数估计结果。逻辑上，这极大地扩展了参数估计可行解的范围，提供了评价解的精度的途径，进而提高推断和判决的可靠性。事实上，通过随机模拟还可以解决（超）高维函数的积分问题，大规模参数空间的重要性区域探测问题（即全局优化问题）等等。梯度导引类算法之后，在超级算力的推进下，随机模拟算法会逐渐成为核心算法，更好地解决各种各样的实际问题。据此，本文重点讨论 Gauss 假设下，Bayes 参数估计算法存在的问题，随机模拟的基本思想和方法，最后讨论随机模拟在勘探地震中的可能应用。期望这样的讨论能促进随机模拟和随机反演方法在勘探地震中的深化应用。

【关键词】Bayes 参数估计，后验概率密度，期望和方差，MAP，梯度导引类迭代算法，随机模拟算法，随机反演

1 引言

油气地震勘探最终目的：精确地描述油气藏，进行准确的含油气性分析，做出最佳的钻井决策，得到最高的油气勘探效益。

油气地震勘探核心问题：由叠前地震数据及其它相关的先验信息，进行宽波数带的弹性参数估计（或称广义的高精度地震波成像），与井数据和岩石物理知识结合，实现储层岩石物性参数（孔隙度、渗透率等）的推断，开展精确的油气藏描述和准确的含油气性评价。

宽波数带的弹性参数估计归结于 Bayes 参数估计问题。油气藏描述及含油气性评价归结为 Bayes 推断问题。

到目前为止，地震波成像的基本理论框架已经十分清楚了，它在整个石油工业界的应用现状当然更为明确。从上世纪 60 年代开始，多次覆盖的观测方式、数据采集的数字化、计算机的出现，催生了以地震波偏移成像为中心的地震数据处理（流程）。2000 年后，“两宽一高”地震数据采集、超大规模计算机使得逆时叠前深度偏移和 FWI 为代表的地震波成像成为核心。

目前，尽管静校正、CMP 道集叠加、速度谱解释、叠前偏移成像，成像道集层析速度建模依然在石油工业界中处于重要地位，地震波成像基本上逐渐聚焦在 Bayes 参数估计理论框架下。就是假设已知正问题和实际测量数据的不确定性的先验和关于待估计参数（即地震波成像结果）的先验，基于此生成关于待估计参数的后验（概率密度），最后按照该后验（概率密度）生成估计结果（地震波成像结果）。一般地，后验（概率密度）的期望（后验概率加权平均值）作为点估计结果，对应的方差作为估计结果精度的评价。实际上，大规模参数估计中，（超）高维积分求后验（概率密度）的期望是很难做到的。因此，提出用后验（概率密度）的最大值对应的样本值作为估计结果，即 MAP 准则。MAP 准则实际上得到的是类似众数的估计结果，就是以最大概率出现的估计结果。后验（概率密度）Gauss 情形下，期望值与众数值是相等的。很显然，这样的估计结果是某种真实待估计参数平滑后的结果，它们分辨率是不够的。后验（概率密度）Gauss 情形下，勘探地震中就产生了 FWI 方法。Newton 梯度导引类算法产生了 MAP 准则下的反演解。

很显然，提高参数估计分辨率的方法应该是：在已知后验（概率密度）的条件下，生成满足该分布的随机样本，MAP 附近的样本值，都可以作为参数估计结果。逻辑上，这极大地扩展了参数估计可行解的范围，提供了评价解的精度的途径，进而提高推断和判决的可靠性。

事实上，通过随机模拟还可以解决（超）高维函数的积分问题，大规模参数空间的重要性区域探测问题（即全局优化问题）等等。梯度导引类算法之后，在超级算力的推进下，随机模拟算法会逐渐成为核心算法，更好地解决各种各样的实际问题。

据此，本文重点讨论 Gauss 假设下，Bayes 参数估计算法存在的问题，随机模拟的基本思想和方法，最后讨论随机模拟在勘探地震中的可能应用。期望这样的讨论能促进随机模拟和随机反演方法在勘探地震中的深化应用。

2 Bayes 估计的理论框架及解的性质

前已述及，到现在为止地震波成像基本上逐渐聚焦在了 Bayes 参数估计理论框架下。其中的根本原因在于，与任何数据分析领域一样，实际测量的地震数据被视为随机过程（随机向量），从数据中提取相关的信息，进而进行推断，解决各种实际问题。这已经形成了大数据时代解决问题的基本模式。

地震波成像的典型特征是存在实测波场与地下介质弹性参数变化之间的数学物理关系，即波动方程和它的各种变种。从叠前数据中挖掘出关于地下介质弹性参数的信息自然要用到这个强大的、因果的先验知识，尽管我们知道这个数学物理关系并不能很好地反映实测波场与地下介质弹性参数之间的真实的关系。这正是深刻理解地震波反演成像解的性质的关键。所谓的反演解仅仅是对地下介质弹性参数的、精度有限、分辨率有限的估计结果。随着叠前数据质量的提升、关于待估计参数的先验信息的增加、适当提高波动方程逼近真实介质的程度（即考虑弹性、各向异性、粘滞性等），原则上可以提高弹性参数解估计的精度。

2.1 Bayes 参数估计的理论框架

Bayes 参数估计的根本思想是：假设已知实测数据与地下介质弹性参数之间的先验预测关系（即概率密度函数表达的不确定关系）和待估计参数满足的先验概率分布，上述两种先验分布，决定了已知实测数据与待估计参数之间的条件后验概率密度。该条件后验概率密度包含了待估计参数的所有统计特征。一般地，把条件后验概率密度的期望作为待估计参数的结果，其方差作为待估计参数的精度评价。

首先，有实测数据与地下介质弹性参数之间的先验预测关系，

$$\mathcal{M}：\boldsymbol{d}^{obs}=H\left(\boldsymbol{m}\right)+\boldsymbol{\varepsilon}。\tag{1}$$

其中，模型 $\mathcal{M}$ 包含了实测数据与地下介质弹性参数之间的先验预测关系的不确定性，用 $p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)=p_{\varepsilon}\left(\boldsymbol{d}^{obs}-H\left(\boldsymbol{m}\right)\right)$ 来表达。不确定性既由数据观测过程中的因素引起，也由正问题 $\boldsymbol{d}^{cal}=H\left(\boldsymbol{m}\right)$ 的不准确引起。ε 是正问题不能完全解释实测数据导致的误差。正是 ε 的随机性，刻画了实测数据与地下介质弹性参数之间的先验预测关系的不确定性。一般地，假设某种概率密度函数刻画了这种不确定性，并假设该概率密度函数是已知的。

单纯地依赖实测数据与地下介质弹性参数之间的先验预测关系的不确定性进行参数估计是最大似然估计（MLE = Maximum Likelihood Estimation）学派的思想。但是，Bayes 参数估计学派认为，仅仅依赖实测数据与地下介质弹性参数之间的先验预测关系不可能得到可靠的估计结果，必须引入待估计参数的先验信息。该先验信息不可能是完全准确的，不确定性用某种先验概率密度函数来描述，可写为 $p\left(\boldsymbol{m}\middle|\mathcal{M}\right)$。很显然，参数 $\boldsymbol{m}$ 是正问题 $\boldsymbol{d}^{cal}=H\left(\boldsymbol{m}\right)$ 中包含的参数，与模型 $\mathcal{M}$ 是相关联的。

根据上述两种先验概率密度函数，依据 Bayes 公式，可以定义如下条件后验概率密度：

$$p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs};\mathcal{M}\right)=\frac{p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)p\left(\boldsymbol{m}\middle|\mathcal{M}\right)}{p\left(\boldsymbol{d}^{obs}\middle|\mathcal{M}\right)}\tag{2}$$

其中，$p\left(\boldsymbol{d}^{obs}\middle|\mathcal{M}\right)$ 是表示实测数据本身满足的不确定的概率密度函数。

概率论指出：条件后验概率密度 $p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs};\mathcal{M}\right)$ 包含了待估计参数 $\boldsymbol{m}$ 的所有统计特征。期望（一阶统计量）可以作为待估计参数 $\boldsymbol{m}$ 的估计结果；方差（二阶统计量）视为待估计参数 $\boldsymbol{m}$ 的估计精度评价。由于勘探地震中，待估计参数 $\boldsymbol{m}$ 有非常高的维度，积分求条件后验概率密度规定的期望和方差基本是不可能的。

因此，条件后验概率密度的最大化准则（MAP = Maximum A Posteriori Estimation）成了常用的参数估计准则。即

$$\hat{\boldsymbol{m}}=\arg\min_{\boldsymbol{m}} p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs};\mathcal{M}\right)=\arg\min_{\boldsymbol{m}}\left\{p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)p\left(\boldsymbol{m}\middle|\mathcal{M}\right)\right\} \tag{3}$$

一般地，两个先验概率密度函数 $p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)$ 和 $p\left(\boldsymbol{m}\middle|\mathcal{M}\right)$ 都被假设是指数型的。已知指数型函数的对数不改变极小值的分布，因此，（3）式被写为：

$$\hat{\boldsymbol{m}}=\arg\min_{\boldsymbol{m}}\left\{-\ln p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m};\mathcal{M}\right)-\ln p\left(\boldsymbol{m}\middle|\mathcal{M}\right)\right\} \tag{4}$$

显然，（4）式逐渐转化为最优化问题。解析数学家定义如下形式的最优化问题：

$$\mathcal{J}\left(\boldsymbol{m}\right)=Q\left(\boldsymbol{d}^{obs},H\left(\boldsymbol{m}\right)\right)+\lambda\Omega\left(\boldsymbol{m}\right) \tag{5}$$

其中：$Q\left(\boldsymbol{d}^{obs},H\left(\boldsymbol{m}\right)\right)$ 被称为数据偏差项，度量正问题 $\boldsymbol{d}^{cal}=H\left(\boldsymbol{m}\right)$ 预测实测数据 $\boldsymbol{d}^{obs}$ 的误差；$\Omega\left(\boldsymbol{m}\right)$ 定义待估计参数的先验信息。

很显然，若定义（4）式中的对应项为：

$$Q\left(\boldsymbol{d}^{obs},H\left(\boldsymbol{m}\right)\right)=-\ln p\left(\boldsymbol{g}\middle|\boldsymbol{m};\mathcal{M}\right); \tag{6a}$$

$$\Omega\left(\boldsymbol{m}\right)=-\ln p\left(\boldsymbol{m}\middle|\mathcal{M}\right) \tag{6b}$$

可以看出，数学分析学家定义的参数优化反演问题（一般地，称 Tikhonov 正则化问题）与两个先验概率密度函数假设为指数分布的 MAP 估计问题是一致的。

更进一步地，在线性 Gauss 假设下，即存在实测数据与地下介质弹性参数之间的线性预测 $\boldsymbol{d}^{obs}=\boldsymbol{d}^{cal}+\varepsilon=H\boldsymbol{m}+\varepsilon$，并有如下的两个先验概率密度函数满足 Gauss 分布的假设，

对噪声的假设：

$$\varepsilon\sim N\left(\boldsymbol{0},\sigma_{\varepsilon}^{2}\boldsymbol{I}\right)\ \text{即}\ p\left(\boldsymbol{d}^{obs}\middle|\boldsymbol{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left\|\boldsymbol{d}^{obs}-H\boldsymbol{m}\right\|^{2}\right] \tag{7a}$$

和对 $\boldsymbol{m}$ 的假设：

$$\boldsymbol{m}\sim N\left(\boldsymbol{0},\sigma_{\boldsymbol{m}}^{2}\boldsymbol{I}\right)\text{即}\ p\left(\boldsymbol{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\boldsymbol{m}}^{2}}\left\|\boldsymbol{m}\right\|^{2}\right] \tag{7b}$$

则对应的后验概率密度为

$$p\left(\boldsymbol{m}\middle|\boldsymbol{d}^{obs}\right)\propto\exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left\|\boldsymbol{d}^{obs}-H\boldsymbol{m}\right\|^{2}+\frac{\sigma_{\varepsilon}^{2}}{\sigma_{\boldsymbol{m}}^{2}}\left\|\boldsymbol{m}\right\|^{2}\right] \tag{8}$$

类似地，有 MAP 估计下的反演解

$$\hat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m}} \left\{ p\left(\boldsymbol{m}\middle|\boldsymbol{d}\right) \right\} = \arg\min_{\boldsymbol{m}} \left\{ \mathcal{J}\left(\boldsymbol{m}\right) \right\} \tag{9}$$

其中，误差泛函定义为

$$\mathcal{J}\left(\boldsymbol{m}\right) = -\frac{1}{2}\left\| \boldsymbol{d}^{obs} - H\boldsymbol{m} \right\|_{\sigma_\varepsilon^2}^2 + \lambda \left\| \boldsymbol{m} \right\|^2 \tag{10}$$

超参 $\lambda = \dfrac{\sigma_\varepsilon^2}{\sigma_{\boldsymbol{m}}^2}$。

（10）式的解可以写为：$\boldsymbol{m}\middle|\boldsymbol{d}^{obs} \sim \mathcal{N}\left(\hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}}\right)$。其中，$\hat{\boldsymbol{m}} = \left(H^T H + \lambda I\right)^{-1} H^T \boldsymbol{d}^{obs}$，$\hat{\boldsymbol{\Sigma}} = \sigma_\varepsilon^2 \left(H^T H + \lambda I\right)^{-1}$。

Bayes 参数估计的基本逻辑框架基本上就是这样。可以看出，炮集波场用算子形式的波动方程 $\boldsymbol{d}^{cal} = H\left(\boldsymbol{m}\right)$ 来预测时，对应的参数估计问题就是 FWI。若引入 Born 近似后的正问题 $\boldsymbol{d}^{cal} = H\boldsymbol{m}$ 预测炮集波场，对应的参数估计问题就是最小二乘逆时深度偏移（LS_RTM）。

2.2 Bayes 参数估计的解的性质的分析

Gauss 先验分布假设+MAP 准则+误差泛函最小+梯度导引类算法，得到问题的反演解，数学理论上，就是 $\hat{\boldsymbol{m}} = \left(H^T H + \lambda I\right)^{-1} H^T \boldsymbol{d}^{obs}$，$\hat{\boldsymbol{\Sigma}} = \sigma_\varepsilon^2 \left(HH + \lambda I\right)^{-1}$。明显地，这也是真解的 PSF 模糊结果。很显然，Bayes 后验概率密度求期望给出的参数反演解是某种加权平均值，在提高反演可靠性的同时，导致了分辨率降低。目前，FWI 和 LS_RTM 的具体实践中，重点放在了如何获得反演解上，对反演解的精度和分辨率的评价较少涉及。但是，目前已经到了需要花精力评价反演解可靠性和精度的时间了。

众所周知，提高反演解分辨率（精度）的思想与方法是清楚的：期望实测数据尽可能是完美的、满足高精度反演需求的；提高正问题对实测数据的预测精度；期望预测误差是 Gauss 分布（ Gauss 白噪）的；期望正问题关系是尽可能线性的；期望初始值是尽可能接近真解的；施加合理的正则化约束；期望有很多的、尽可能多的先验信息；……。

但是，反问题的多解性、非唯一性是永恒的。反演成像的实践表明：即便做过上述处理，依然达不到期望的精度和分辨率要求。除了继续沿上述已知的路线努力外，还有什么提高反演结果分辨率（精度）的新途径？随机反演可能是值得探索的！

3 随机模拟的基本思想

物理世界的变化规律到底是确定性的？还是随机性的？可以认为宏观上是确定性的（Newton 力学规律），微观上是随机的（量子力学规律）。大量物理粒子的规律性的行为特征也要用概率统计方法描述，譬如分子热力学。

机械的、确定性的、因果的观点描述世界的方法显然是片面的、不完全可取的。基于随机、概率统计基础上的认识论描述和解决客观问题是不可或缺的。

Newton 力学、数学分析（泛函分析）、线性代数、数值分析/数值算法（尤其梯度导引类的数值优化算法）提供了强大的分析问题的逻辑和方法。但是，也固化了我们的思维。根本上看，物理世界、生物世界、人类社会处处存在不确定性。越是微观的、细尺度的、多因素相互作用的情形下，不确定性或随机性的表现越普遍。宏观的、粗尺度的、单因素（少因素）情形下，确定性的表现越明显。由基于确定性的、因果的关系进行预测、做出决策的方式方法要逐渐地让位于由基于随机统计性的、相关性的关系进行预测，并做出决策的方式方法。

随机性是物理世界、生物世界、人类社会的客观特征，它是由样本的取值满足一定的概率分布体现出来的。若样本的取值是确定性的，表明事件都按概率 1 发生，事件没有不确定性。计算机模拟（仿真）确定性关系，从而提升我们对物理及非物理实体的认识，也帮助我们做出估计推断。这是我们所熟悉的。

但是，计算机能完全模拟随机过程吗？随机模拟为什么能帮助我们解决问题？随机模拟能解决哪些问题？

计算机显然不能真正地或完全地仿真物理及非物理世界的随机过程。基本原因是：我们自己都不理解随机过程产生的过程，只知道其结果。数学上，没有理论证明能在计算机上真正地或完全地仿真物理世界的随机过程。数学家可以做到在计算机生成伪随机数。即：给定一个初值，在很长的周期内，不会重复生成相同的随机数序列。伪随机数与物理世界的随机过程是一致的吗？显然是不一致的。

另一方面，科学上，人类的确有必要发展出能模拟物理及非物理世界随机过程的算法。即便做不到，也可以用伪随机模拟的思想与方法帮助我们更好地解决当前遇到的各种各样的问题。

（伪）随机模拟到底模拟什么？模拟满足一定概率密度分布的随机变量的生成过程。本质上，就是在计算机上模拟满足一定概率密度分布的随机变量（随机过程）的样本生成。其中概率分布的形式可以是多种多样的，对应的随机样本的物理及非物理场景下的具体含义也是多种多样的。可以看出，（伪）随机模拟有广泛的应用性。

（伪）随机模拟的关键是：伪随机数的生成。一般地，独立同分布地生成[0,1]空间上均匀分布的随机变量的样本。伪随机数的生成是随机模拟的基础。在计算机上模拟满足任意概率密度函数的随机变量（随机过程）的样本生成，离不开伪随机数产生。伪随机数的生成有很多现有的算法程序。

随机模拟要解决的核心问题：在既定的、巨大的样本空间中，按既定的（后验）概率密度函数，进行重要性采样。这显然是高维参数估计中误差泛函的全局寻优问题。重要性采样也是解决高维积分问题的核心思想。

还有很多其他的应用场景。大数据和人工智能时代，这样的应用场景多种多样。

3.1 随机反演的基本思想

前面已经讨论了 Bayes 参数估计的基本逻辑。FWI 和 LS_RTM 最后归结为梯度导引类的目标泛函求极小的数值优化问题。

Bayes 估计框架下的随机反演，基本思想可以描述为：基于 Bayes 推断的基本逻辑，由后验概率密度对参数进行估计，高维积分求参数的各阶统计量。由于高维情况下积分无法完成，进行基于重要性采样的随机模拟，生成满足后验概率密度的参数随机样本，对生成的参数随机样本进行统计量的估算及精度和分辨率的评价。

随机反演中的随机模拟的核心思想：

①、用辅助的可在计算机上模拟的概率密度函数（一般为均匀分布）的采样（随机数的生成），通过某种映射完成期望的概率密度函数的样本的生成。但这样做法的效率太低。

②、引入 Gibbs 采样算法或 Metropolis-Hasting 算法。这两种算法均在 Markov 链的基础上衍生出来。基本理念是从任意随机采样出发，经过设计的映射，总可以达到期望的概率密度函数及其对应的随机样本。

Gibbs 采样算法和 Metropolis-Hasting 算法物理解释：对于 Boltzmann 分布，从任意随机分布（样本）出发，按设计好的 Markov 链演化规则，总可以达到 Boltzmann 分布及对应的样本。Boltzmann 分布的物理含义是大概率出现的是低能量的样本。这正是重要性采样所追求的目标。

Gibbs 采样，从数学上，由期望分布的 full conditional distribution 构造转移规则，迭代生成最终的期望分布及对应的采样。Gibbs 采样是一种顺序采样（借鉴了 Gauss-seidel 迭代法的思想）。

Metropolis-Hasting 算法改进了①中提及的方法，设计 Markov 链演化规则，迭代达到期望的分布及对应的采样结果。

Markov 链的引入，即为了提高采样的效率，也利用了 Markov 链能从任意分布出发，经过设计的演化规则，迭代收敛到期望分布及对应的采样结果的目标。这就是 Monte Carlo 随机模拟一定要过渡到 MCMC 的根本原因。

可以看出：MCMC 最关键的是如何设计 Markov Chain 的演化规则。至于收敛性，Markov Chain 的基本理论提供了保障。

很显然，生成满足后验概率密度的参数随机样本是随机反演的本质。如何从中选择最后的反演结果及对反演结果的精度和分辨率进行评价，也是重要的问题。目前，这还缺乏一致的方法，各种应用场景下，根据具体要求，发展特定的方法。

3.2 反函数法实现随机模拟的解释

随机模拟的基本问题可以描述为：已知概率密度函数 $f(x)$ 及其累计分布函数 $F(x)=\int_{-\infty}^{x} f(t)dt$，求取一组随机数 $\{x_0,x_1,x_2,.....,x_N\}$ 使其满足概率密度函数 $f(x)$ 的分布。

使用反函数方法来进行随机模拟的基本思想是：令 $y=F(x)\sim U(0,1)$，对 y 进行 0-1 的随机均匀采样，然后计算反函数映射 $x=F^{-1}(y)$ 就可以得到满足 $f(x)$ 概率密度分布的随机数。

为了理解反函数的基本思想，不妨先从简单离散情况下来理解这一问题：

假设随机数 x 的可能取值为 1，2，3，4。$x=\{1,2,3,4\}$。

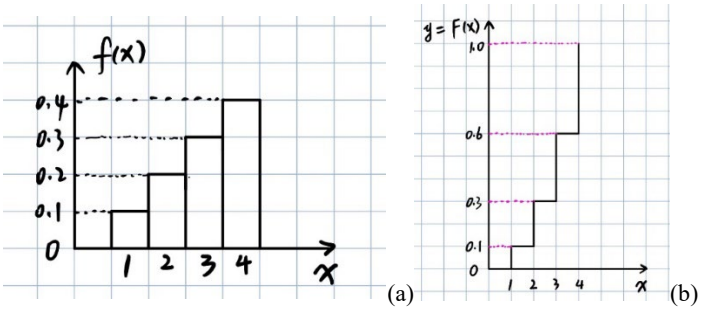
其概率密度函数为

$$f(x)=\begin{cases}0.1 & x=1\\ 0.2 & x=2\\ 0.3 & x=3\\ 0.4 & x=4\end{cases}, \tag{11}$$

则可以计算出其累计分布函数

$$F(x)=\begin{cases}0.1 & x=1\\ 0.3 & x=2\\ 0.6 & x=3\\ 1.0 & x=4\end{cases}。 \tag{12}$$

将公式（11）与公式（12）进行图形化，得到图 1。



(a) (b)

图 1 概率密度图（a）及其累计概率密度图（b）

从图 1b 中可以看出，累计分布函数的范围为 $y\in[0,1]$，并且累积分布函数的取值被分为了四段，每一段的长度正好对应 x 取值的概率密度函数 $f(x)$ 的大小。因此，假设 $y\sim U(0,1)$，即 y 服从 0 到 1 的均匀分布，对 y 进行随机采样，通过采样值 y_i 所在的范围，即可得到 x_i，并且使得到的 x_i 序列满足概率密度函数 $f(x)$。

如果 $y_i\in[0,0.1)$，$x_i=1$；

如果 $y_i\in[0.1,0.3)$，$x_i=2$；

如果 $y_i\in[0.3,0.6)$，$x_i=3$；

如果 $y_i\in[0.6,1.0]$，$x_i=4$。

$F(x)$ 的反函数为

$$x=F^{-1}(y)=\begin{cases}1 & 0.0\le y<0.1\\ 2 & 0.1\le y<0.3\\ 3 & 0.3\le y<0.6\\ 4 & 0.6\le y\le 1.0\end{cases}。\qquad(13)$$

因此，只需要对 y 进行 0-1 的均匀随机采样，然后通过公式（13）进行反变换，就可以得到满足 $f(x)$ 的随机采样序列。

通过离散情况下的简单示例，已基本对通过累积分布函数的反函数进行随机采样有了一个基本的认识，接下来再考虑连续情况下的累计分布函数的反函数法实现随机采样的想法。

问题与前面描述的一样，已知概率密度函数 $f(x)$ 及其累计分布函数 $F(x)=\int_{-\infty}^{x}f(t)dt$，求取一组随机数 $\{x_0,x_1,x_2,.....,x_N\}$ 使其满足概率密度函数 $f(x)$ 的分布，此处 $f(x)$ 满足连续性。

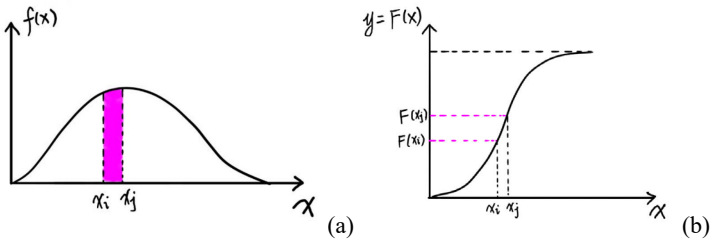


图 2 连续概率密度函数及其累积概率密度函数

为了更好的理解，在这里，引入两个变量，x_i，x_j，图 2a 中桃红色区域表示的是[x_i，x_j]随机数区间内，所有的随机数的采样概率密度相加的大小。在累计分布函数图中（图 2b）中，表示的是 $F(x_j)-F(x_i)$ 长度的大小，当 x_i 趋近于 x_j 时，

$$\lim_{x_i\to x_j}\frac{F(x_j)-F(x_i)}{x_j-x_i}=F'(x_j)=f(x_j)\qquad(14)$$

x_i 逐渐趋近于 x_j，则在累计分布函数中其长度逐渐趋近于累计概率密度函数的导数值，也就是 x_j 的概率密度函数值 $f(x_j)$，因此，与离散的情况相同，也可以认为 x_j 这一点在 $y=F(x)$ 上的长度为其概率密度

函数的大小 $f(x_j)$，同理，对 y 进行 0-1 均匀随机采样，然后通过累积分布函数的反函数就可以得到服从概率密度为 $f(x)$ 的随机采样序列。

首先通过一个简单的正态分布函数的数值示例来进行分析，概率密度函数为均值为 0，方差为 1 的正态分布 $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，如图 3 绿色虚线所示，其累计分布函数为 $y=F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$，如图 3 黑线所示。图 3 展示了累计分布函数值 y 从 0 到 1 间隔 0.01 进行均匀采样（竖轴蓝点所示），及其反函数映射之后对应的采样点 x（横轴红点）的分布，可以看出当 y 均匀分布时，红点的密度和概率密度函数（绿色虚线）的大小成正比例关系，红点密集的地方，概率密度大，红点稀疏的地方，概率密度小。因此可以说累计分布函数改变了累计分布函数值 y 与采样值 x 的数据密度关系，其自动的将均匀分布的累计分布函数值 y，通过累计分布函数的反函数映射成符合概率密度函数 $f(x)$ 的随机采样序列。

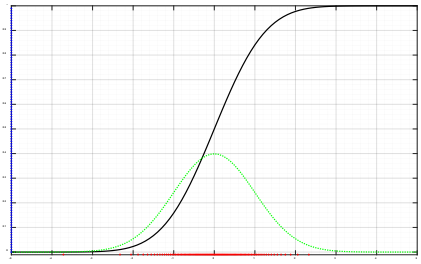


图 3 正态分布函数及其累计分布函数采样值 x 与累计分布函数值 y 的数据密度对应关系

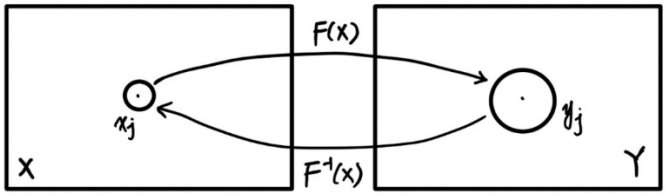


图 4 采样值 x 与累积分布函数值的映射关系图

从前面的示例可以看出，在连续情况下，对应的是数据密度的问题，$y_j=F(x_j)$，在 x_j 点，累积分布函数的导数值 $f(x_j)=F'(x_j)$ 的大小，表示了 y_j 在 x_j 点局部小邻域内的变化快慢，$f(x_j)$ 值越大，则 y_j 的变化的越快，也就是在 x_j 点小的变化引起 y_j 大的变化，x_j 的局部小邻域对应 y_j 的一个局部大邻域，如图 4 所示。反之考虑，如果 y 空间中的采样是均匀采样，也就是 y_j 采样点的邻域内的所有值都会被压缩到 x_j 的小邻域内，从而实现了对 x_j 的重要性采样，而其中的 x 空间的邻域与 y 空间的邻域的缩放与概率密度函数 $f(x_j)$ 有关。

因此，本质上，$f(x)$ 积分得到的累积分布函数 $y=F(x)$ 决定了 x 与 y 的映射关系，并且该映射关系实现了对 y 的均匀采样，反映了对 x 的重要性采样。

3.3 再议重采样算法的思想

重要性采样是随机反演的核心。它是通过重采样算法实现的。对于一个已知的概率密度函数 $g(x)$，其解析式相对复杂，如何对其进行随机采样，使得采样得到的随机数，满足其概率密度 $g(x)$？

假设存在一个概率密度 $f(x)$，在任意 x 位置，$c\cdot f(x)>g(x)$，此处 c 为一比例系数，并且满足概率密度函数 $f(x)$ 的随机数是容易生成的，则可以通过对概率密度函数 $f(x)$ 的随机数进行处理，来实现对 $g(x)$ 的随机采样。如何实现这一过程？

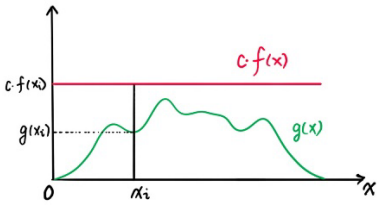


图 5 重采样概率密度函数示意图，红线为 $f(x)$ 概率密度函数，绿线为 $g(x)$ 概率密度函数

如图所示，不妨我们假设 $f(x)$ 为一均匀分布， $g(x)$ 为图中绿线所示，首先 $f(x)$ 进行随机采样，当采样的样本数达到一定数量后，每一个 x 采样的数量应该是相同的。如果对于采样到的所有随机数 x，按照 $\frac{g(x)}{c\cdot f(c)}$ 的概率进行筛选，则保留下来的样本，可以满足 $g(x)$ 的概率分布。更通俗来讲，对 $f(x)$ 的随机采样数是红线的分布，在按照 $\frac{g(x)}{c\cdot f(c)}$ 的概率进行保留后，绿线之上的随机数都被剔除了，只保留下来绿线之下的随机数，使之可以满足 $g(x)$ 的概率分布。

因此，重采样的想法是，

step1：通过 $f(x)$ 随机采样，得到一个随机数 x_i，

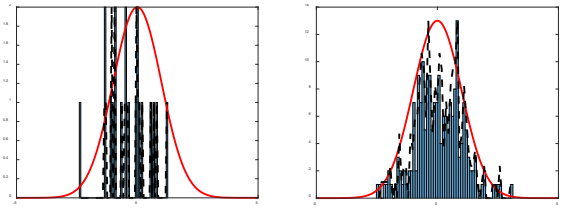
step2：然后按照 $\frac{g(x_i)}{c\cdot f(x_i)}$ 的概率决定是否保留当前的随机数 x_i，这里为了决定是否保留，需要再生成一个随机数 $t_i\sim U(0,1)$，如果 $t_i\le\frac{g(x_i)}{c\cdot f(x_i)}$，则保留 x_i，否则，则舍弃 x_i。

不断重复步骤 1 与步骤 2，则可以得到一组满足 $g(x)$ 概率密度的随机数。

一次采样时，将很难满足 $g(x)$ 的分布，但是当 step1 随机采样的采样随机数数量增加时， x_i 的采样数量也会达到一定数量，然后通过 step2，按照一定比例接受，保留下的 x_i 数量，与步骤 1 生成的 x_i 的数量比将等于 $\frac{g(x_i)}{f(x_i)}$。对于任意 x 均满足这一相应的比例关系时，最终得到的采样序列，将满足 $g(x)$ 的概率分布。

3.4 数值示例

通过重采样对均值为 0，方差为 1 的高斯分布进行随机采样，选择均匀随机采样生成最初始的随机采样数，然后通过重采样获得高斯随机分布的随机数。图 4 展示了随机采样的结果。蓝色直方图为通过重采样得到的采样数，红线为均值为 0，方差为 1 高斯概率密度函数，黑色虚线为统计的随机采样样本的概率密度函数。从图 4 中可以看出，在样本较小时，采样的结果相对较差，随着样本的逐渐增大，红线和黑色虚线逐渐重合，采样得到的随机数满足均值为 0，方差为 1 的高斯分布。



N=20　　N=200

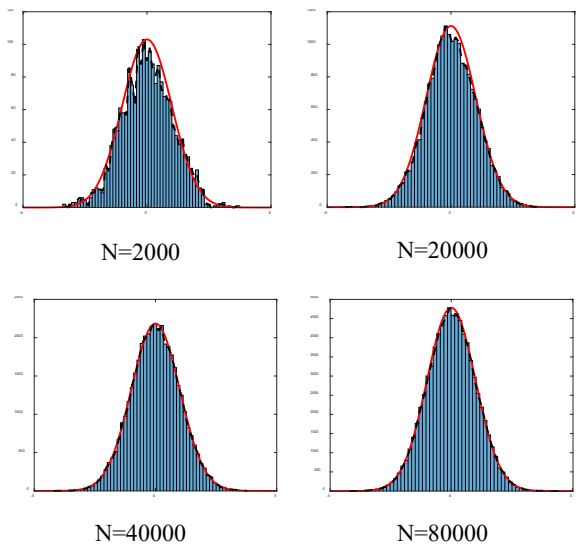


图 6 重采样获取均值为 0，方差为 1 的高斯随机采样数，N 为采样的样本数

4 勘探地震中的随机反演基本思想与做法

前已述及，随机反演（Monte Carlo 反演）是指：在可能的解空间中，进行重要性随机采样，找到临近 MAP 的区域，生成该区域中的随机变量样本，然后用合适的方法产生合理的解并对解进行评价。其中，Monte Carlo 算法的核心是重要性随机采样。Metropolis - Hasting 采样算法是目前典型的重要性采样方法。借助生物进化方法是改善重要性采样效率和性能的必然途径。

勘探地震中，如何利用上述随机反演的思想？

首先，基于叠前地震数据和待估计参数的先验信息进行高精度的、宽带的弹性参数估计，是个高度非线性的问题。目前的做法是 Gauss 分布+非线性（线性）正问题假设+Newton 梯度导引类算法给出的反演解，理论上它是条件后验概率密度加权平均的结果，根据目前的勘探实践，其分辨率很难达到高精度油藏描述的需求。

因此，把随机反演的思想引入到弹性参数估计（或地震波成像）中，是合乎逻辑的选择。尤其，在巨大的计算机算力支撑和大数据驱动的多信息提取用来约束反演解的算法逐渐成为常规的情形下，借用随机反演的思想和算法提升参数估计的精度和分辨率逐渐成为可能。

勘探地震中的随机反演基本思想与做法大致可以总结如下：

① 对于强非线性问题，在可能的解空间中，进行重要性采样，找到临近 MAP 的区域，生成该区域中的随机变量样本，用合适的方法得到合理的解并对解进行评价。这样的做法还可以与梯度导引类方法串联起来使用。

另外，基于类 CMP 道集 Monte Carlo 反演提供初始模型，然后串联梯度导引类层析反演方法也是符合逻辑的选择。

② 对于弱非线性问题，MAP 准则下的梯度导引算法，给出某种加权平均的期望结果。在期望解附近，在 Gauss 后验概率分布控制下，随机生成超分辨率模型参数样本族，在合理的准则下，选择超分辨率反演结果。理论上，这是超分辨率反演一条合理的途径。我们认为：用于储层描述的随机反演应该按此逻辑进行。

③ 对于有可靠的井约束的情形，反射结构约束（最好是地震相约束）井数据插值方法得到伪井结果。以伪井结果作为初始值和/或先验约束值，用梯度导引类算法给出某种加权平均的期望结果。在期望解附近，在 Gauss 分布控制下，随机生成超分辨率模型族，在合理的准则下，选择超分辨率反演结果。这应该是井控情形下，用于储层描述的超分辨率反演的合理逻辑路线。原则上，这样做的超分辨率反演，其结果可以达到采样间隔的尺度。但是，如何选择出合乎地质逻辑的反演解目前依然是没有很好解决的问题！

对勘探地震中的随机反演基本思想与做法的基本评价：随机反演（Monte Carlo 反演）既可以用于解决强非线性反问题的全局寻优，也可以生成后验概率密度对应的期望值附近的一族可行的反演解。前者计算量太大，用于目前的大规模参数估计不一定可行。后者用于得到超分辨率（可达任意采样尺度）的反演解，存在的困难是如何生成一族 MAP 附近的可行解，如何从中选出一个有地质意义的解。当前，FWI 或任何梯度导引类算法给出的反演解，得到满足储层描述要求的、有地质意义的解都是有困难的。

无论如何，随机反演算法可以给出一族 MAP 附近的可行解，提供了选择出更高分辨率可行解的机会，也提供了评价解的精度的随机采样样本，因此引入随机反演应该是值得探索的路线。

5 地质统计学反演与随机反演的关系的认识

作者认为：地质统计学反演是随机反演在勘探地震中的实例，而随机反演是 Monte Carlo Bayes 推断的具体做法。前面已经讨论过随机反演的本质是按照后验概率分布，生成 MAP 附近的随机样本，从这些随机样本中选择或生成反演结果，并对反演结果进行评价。

众所周知，地质统计学的起源来自于 Kriging 插值，其本质是已知散乱的空间数据，推断出规则的空间数据。

Kriging 插值的本质思想是把散乱的空间数据视为随机的，并把它分为三部分，即 $\boldsymbol{d}(\boldsymbol{x})=\boldsymbol{d}_{BC}(\boldsymbol{x})+\boldsymbol{d}_{DC}(\boldsymbol{x})+\boldsymbol{d}_{RC}(\boldsymbol{x})$。其中，$\boldsymbol{d}_{BC}(\boldsymbol{x})$是空间数据的背景成分（Background Component）；$\boldsymbol{d}_{DC}(\boldsymbol{x})$是空间数据的扰动成分（Disturbance Component）；$\boldsymbol{d}_{RC}(\boldsymbol{x})$是满足一定概率分布的随机成分（Random Component）。$\boldsymbol{d}_{BC}(\boldsymbol{x})$可以视为是确定性的、散乱点支撑的函数。这部分的描述可以用径向基函数族的线性组合的方法；$\boldsymbol{d}_{DC}(\boldsymbol{x})$用零均值、一定方差的概率密度函数来描述；$\boldsymbol{d}_{RC}(\boldsymbol{x})$被视为高斯白噪声。

地震统计学反演就是在上述思想下发展出来的。首先，必须引入井约束，加上反射结构约束（最好是地震相约束）井数据插值方法得到伪井结果。这部分就对应$\boldsymbol{d}_{BC}(\boldsymbol{x})$。以伪井结果作为初始值和/或先验约束值，用梯度导引类算法给出某种加权平均的期望结果。在期望解附近，在 Gauss 分布控制下，随机生成超分辨率模型族，在合理的准则下，选择超分辨率反演结果。这部分结果就对应$\boldsymbol{d}_{DC}(\boldsymbol{x})$部分。

高斯白噪声$\boldsymbol{d}_{RC}(\boldsymbol{x})$的加入，为更高分辨的探索解空间增加一些可能性。但这部分基本上是可有可无的。

据此，把地质统计学反演放在 Monte Carlo Bayes 推断的理论框架下，这样可以加深对地质统计学反演的理论认识。

6 随机模拟在勘探地震中的可能应用

基于波动方程数值解法的地震波场确定性模拟，已经广泛地应用在地震数据分析的相关过程中。它可以根据人为设计的、弹性参数表示的地质模型，求解各种形式的波动方程，按既定的观测系统，产生炮集。这样的确定性模拟，可以产生人工合成的炮波场，用于验证地震波成像方法，可以定性地评价地质解释的合理性，可以为深度学习算法提供标签数据，可以分析不同的弹性参数表示的地质模式的地震波响应特征等。

前已述及，随机模拟本质上是生成特定的概率分布对应的随机样本值。理论上，它可以解决高维积分问题，可以对任意的概率分布进行重要性采样（借此做法可以进行全局寻优）。但是，随机模拟在勘探地震中能有什么作用？

作者认为：

① 可以按假定的概率分布随机生成各种形式的地震相模式，从中选择出合理的地震相模式，辅助进行地震解释。

② 可以按假定的概率分布随机生成各种形式的火山岩相模式，与确定性的地震波模拟结合，筛选出符合实际的火山岩相模式。

③ 可以按假定的概率分布随机生成各种形式的层状地层组合，与确定性的地震波模拟结合，筛选出符合实际的薄层地层组合模式。

④ 可以按假定的概率分布随机生成各种形式的层状地层组合，与确定性的地震波模拟结合，按照与实际地震道的匹配程度，生成高分辨的伪井数据。用于其他高分辨反演的先验约束。

⑤ 可以按假定的概率分布随机生成随机噪声，与实际观测的随机噪声对比，用适当的方法减去实际观测的随机噪声。

7 结论与讨论

Bayes 估计理论框架下的地震波成像，以 FWI 为代表，在“两宽一高”地震数据采集和超大规模计算机系统的算力的支撑下，已经成为了油气勘探行业中的核心技术。FWI 的理论核心是在后验概率密度 Gauss 分布假设下，由 MAP 原则+可导的目标泛函+梯度导引类算法产生后验概率密度加权叠加意义下的反演结果。目前的实践表明，这样的反演结果的分辨率达不到实际需求，尤其在针对储层描述的情形下。

很多改善 FWI 成像精度、提高成像分辨率的方法被提出来。效果也是有的，但依然达不到期望的要求。

基于 Monte Carlo 随机模拟的方法是值得探索的新的方法技术路线。基本思想是：在已知后验（概率密度）的条件下，生成满足该分布的随机样本，即 MAP 附近的样本值都可以作为参数估计结果。逻辑上，这极大地扩展了参数估计可行解的范围，并提供了评价解的精度的途径。很显然，这种随机模拟算法产生的反演解族，分辨率原则上可以达到采样间隔的尺度。但是，如何生成有地质意义的反演解族，如何从中选择出有地质意义的反演解，依然是要解决的问题。

无论如何，相对于 FWI，基于 Monte Carlo 随机模拟的参数反演思想与方法为我们提供了一条新的参数反演的技术路线。

事实上，Monte Carlo 随机模拟还可以用在很多不同的应用场景中，譬如模拟随机噪声、地震相、火山岩相、薄互层地层组合等等。

我们相信：对于解决很多复杂的非线性问题，相比于梯度导引类的 Newton 迭代法，在超强算力辅助下，尤其与族群生物进化方法结合的 Monte Carlo 随机模拟方法会起到越来越重要的作用。Monte Carlo 随机模拟与 Bayes 推断会形成重要的研究领域。

参考文献

[1] Gamerman, D. and Lpes, H. F., 2006, Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference, Second Edition, Chapman & Hall/CRC
[2] Robert, C. P. R. and Caella, G., 2004, Monte Carlo Statistical Methods, Second Edition, Springer
[3] Jun S. Liu, 2001, Monte Carlo Strategies in Scientific Computing, Springer
[4] Gentle, J. E., 2003, Random Number Generation and Monte Carlo Methods, Second Edition, Springer
[5] Lomas, S. Luo, M. Irakarama, R. Johnston, M. Vyas, X. Shen, 2023, 3D Probabilistic Full Waveform Inversion: Application to Gulf of Mexico Field Data; 84th EAGE Annual Conference & Exhibition
[6] André Haas and Olivier Dubrule, 1994, Geostatistical inversion-a sequential method of stochastic reservoir modelling constrained by seismic data; FIRST BREAK VOL 12, NO 11
[7] M. Izzatullah, T. Van Leeuwen, D. Peter, 2021, Langevin Dynamics Markov Chain Monte Carlo Solution for Seismic Inversion; 82nd EAGE Annual Conference & Exhibition.
[8] Dehan Zhu and Richard L. Gibson, 2016, Seismic inversion and uncertainty analysis using a trans-dimensional Markov chain Monte Carlo Method; SEG International Exposition and 86th Annual Meeting.
[9] Peng Li, Leandro de Figueiredo, Dario Grana, and Klaus Mosegaard; 2021, Spatially correlated Markov chain Monte Carlo method for petrophysical inversion; 2021 Society of Exploration Geophysicists, First International Meeting for Applied Geoscience & Energy
[10] Georgia K. Stuart, Weihua Yang, Susan E. Minkoff, and Felipe Pereira, 2016, A two-stage Markov chain Monte Carlo method for velocity estimation and uncertainty quantification; SEG International Exposition and 86th Annual Meeting.
[11] Dan Simon, 2013, Evolutionary Optimization Algorithms – Biologically Inspired and Population – Based Approaches to Computer Intelligence; John Wiley & Sons, Inc.

典型的逆矩阵及在统计推断理论与方法中的应用

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】数据是研究对象的外在表现，是可观测与可记录的量，它们蕴含了关于研究对象的相关信息。研究对象是物理或非物理系统，系统可能是线性的或非线性的。数据被认为是：一维的时间序列（当然也可以是其它变量的序列），或高维（>=2D）的空间数据（也可抽象为一般的高维数据体）。为对数据进行分析，把数据视为随机过程（即随机变量按某种顺序排放）。一般地，认为数据对应的随机过程是宽（广义）平稳的、零均值的，存在有对应协方差（阵）的（即数据点与点之间存在某种相关关系的）。联合概率分布函数完全决定了数据的统计特征。但是从应用角度，一般地，不可能给出数据的联合概率分布函数，因此据此定义随机过程的各阶统计量（实质上它们也不方便体现数据蕴含的信息）。因此，发展了各种具体的数据分析方法。数据分析，最本质地，可称 Learning from dataData。首先，弄清楚数据由什么特征成分组成的。然后，对数据进行建模，或对宽（广义）平稳随机过程进行建模，实现对未观测数据的预测。再进一步，建立起可观测量与潜变量之间的关系。一般地，是用回归分析方法建立起可观测量与潜变量之间的关系。参数估计（地震波成像/FWI）也可认为是在已知可观测量和已知可观测量与潜变量之间的关系的条件下，估计潜变量。无论是做上述哪一项工作，本质上，都在求解一个 Bayes 推断问题或 Bayes 参数估计问题。因此，有数据分析的根本逻辑：宽（广义）平稳随机过程是对数据的数学抽象，认为宽（或广义）平稳随机过程是由白噪声线性组合生成的，即 ARMA 模型可以生成宽（广义）平稳随机过程，这种观点是来自于时间序列分析的。但这也构成了空间（散乱）数据分析的理论基础。即空间数据分析时，也要求空间数据是宽（广义）平稳的，把 AR 模型推广到高维空间数据分析中就构成了空间数据分析的理论基础。这应该是数据分析的最基本逻辑。其它所谓的高端数据分析方法均是由此推演出来的！之所以能从数据中挖掘出信息，最根本的逻辑基础是宽（广义）平稳随机过程中各随机变量间存在一定的相关性。从而导致协方差阵存在特定的结构。这些结构，即数据中的蕴含的信息。若自协方差阵是单位阵，宽平稳随机过程就是白噪声，其中不包含什么信息。这是数据分析的根本思想逻辑。本文引入典型的逆矩阵恒等表达式，利用这些恒等表达式，推导数据分析中的统计推断问题所涉及的具体方法的表达式。目的是对线性高斯平稳假设下的统计推断理论做一个完整的表述。

【关键词】实测数据，可观测量，潜变量，随机过程，线性系统与非线性系统，统计推断

1 引言

人类社会已经进入到大数据，以及基于大数据、强算法和高算力发展出的人工智能（Artificial Intelligence，简称 AI）时代。机器学习（Machine Learning，简称 ML）方法成为了这个时代解决科学和社会问题的近乎首选的思想与方法。所谓机器学习是一组方法，它能够自动探测数据中的模式，然后利用所发现的模式预测未来数据，或在不确定情况下执行其他类型的决策。但是，任何事情都是循序渐进发展而来的。AI 和 ML 的发展也是如此。可以说它是基于线性平稳高斯情形下的数据分析思想与方法并架构在人工神经网络的计算框架下发展而来的。

人类早期，对这个世界充满了未知、好奇和恐惧。认识这个世界是首先要解决的问题。认识世界的基础在于观察（即获取数据）和总结（即分析数据并提炼出知识）。（欧洲）中世纪之前，人类所有的知识（主要指自然知识），都是在观测和总结的基础上形成的，还没有自觉地设计理想实验验证所总结知识的正确性。到了文艺复兴初期，物理抽象、物理实验、数学总结成为认识（物理）世界规律的主要思想与方法。伽利略、牛顿为代表的经典物理学是这类知识的代表。19 世纪开始，抽象数学已经起步。20 世纪上半叶，（高等）线性代数、矩阵分析、泛函分析、概率论、随机过程等数学领域已经很成熟。

古希腊到中世纪，数学的发展与实际需求联系非常紧密，数学发展是实际需求推进的。经典数学（微积分、线性代数、概率统计、微分方程等）的发展也深受实用的促进。但是，现代数学基本与应用脱节，按照内在的逻辑概念，独立于客观实体向前发展。

经典数学模型（包括部分现代数学模型）框架预测实际物理过程，进而解决实际问题构成了当前所有学科解决问题的基本思想与方法。这是当前理工科（包括部分文科）教育的主要内容。其主要目的就是理解已有的数学框架，用来解决面对的实际问题。

1900 年前后近一个世纪逐渐发展出完善的概率与统计，提出了基于数据寻找事物潜在规律的理论方法。到目前为止，大数据（标签数据）、人工智能算法、强算力形成了人类认识世界规律，并解决实际问题的新途径。

事实上，接受数学家按一套数学逻辑建立的数学框架解释实际物理（也包括部分非物理）世界的规律，并用来解决各种各样实践问题的观念还是很不容易的。首先，理解深奥的数学框架已经不容易了，再深层

次地弄清楚它为什么可以与实际物理问题相符或者说物理问题为什么与数学框架相符更不容易。

大数据（标签数据）、人工智能算法、强算力框架下产生的不可解释的预测关系更是难以让人理解。但是，人类认识这个世界，解决实际问题的思想与方法，主要就是上述四个模式。到底世界是数学的还是客观物理的？可以说，这个世界的运行是符合特定的逻辑规律的！按严谨的逻辑规律发展出的理论体系，可以（部分地）解释这个世界，也可以帮助我们解决不同领域的各种实际问题。逻辑规律的最严谨表达就是数学了！正是可恶的数学！可爱的数学！可敬的数学！

从基于线性平稳高斯情形下的数据分析思想与方法，到架构在人工神经网络的计算框架，再到如今的 ML/AI 时代，这一技术发展历程是如何演进的呢？

首先数据是研究对象的外在表现，是可观测与可记录的量，它们蕴含了关于研究对象的信息。研究对象是物理或非物理系统，系统可能是线性的或非线性的。数据被认为是：一维的时间序列（当然也可以是其它变量的序列），或高维（>=2D）的空间数据（也可抽象为一般的高维数据体）。为对数据进行分析，把数据视为随机过程（即随机变量按某种顺序排放）。一般地，认为数据对应的随机过程是宽（广义）平稳的、零均值的，存在有对应协方差（阵）的（即数据点之间存在某种相关关系的）。联合概率分布函数完全决定了数据的统计特征。具备完备统计特征的数据样本，包含了系统演化的内在逻辑。只要充分地收集（采样）数据，理论上，通过合适的数学算法，就可以分析把握理解系统本身及演化过程。这是 Learning from Data 可行性的根本逻辑。

首先，一维（时间序列）数据分析，主要包括：谱分析和 AR 模型建立（即由最佳 AR 模型实现对一维数据的未观测部分的预测）。高维数据分析，同样是这两件事。但高维规则数据分析和高维散乱数据分析还是有所不同的，可以说数据采样的规则性对数据分析方法的影响是巨大的。高维规则数据分析依然是：线性结构成分分析和高维 AR 模型建立（即由最佳高维 AR 模型实现对高维数据的未观测部分的预测，譬如高维插值；而对已观测高维数据进行预测，即去噪（这是线性信号预测去噪的理论基础）。高维散乱数据分析的核心任务是：由高维 AR 模型建立实现对高维空间中未观测数据的预测，即散乱数据的插值问题，但做这件事的方法被称为 Kriging 方法。为什么又必须创造一个新名字？因为 Kriging 方法和高维 AR 模型建模方法还是有明显区别的！对于高维散乱空间数据，其协方差阵用可观测数据也是无法估算的，另外，高维数据的协方差是张量，对于实际计算非常不利。因此，引入半变差函数，进一步地，又对半变差函数进行半解释表达，即引入反距离（平方）函数、（各向同性/各向异性）高斯函数等近似地表达半变差函数。这就引入了径向基函数构建基函数族来表达高维空间中任意函数的思想与方法。针对高维数据的非线性结构进行表达或建模预测，这是当前数据分析领域中的一个热点及前沿研究方向。其基本思想还是明了的，依然是借用了微积分的思想，即在局部空间中进行线性特征表达，然后找到一种映射把线性特征表达投影到低维空间中。问题是如何找到这种映射！另外，对高维散乱数据中的非线性结构的特征表达与建模预测，目前似乎讨论不多。无论如何，在这些前沿领域做些研究工作，才有产生更大的创新成果的可能。对高维空间数据中的非线性结构进行分析，人工智能，即神经网络算法，似乎就是在做这个工作，但大量的合格标签如何做？内部的运作机理是什么？非线性结构是如何被神经网络捕捉到的？神经网络如何利用这些非线性结构的？等等，目前似乎都还是一笔糊涂账。另外，多元统计分析是前述数据分析的子集。核心依然是对可观测量进行特征表达及建模，至于可观测量与潜变量之间关系的分析，尤其是因子分析/对应分析/典型相关分析，不过是特征表达基础上的进一步深化应用而已。至此，机器学习中的绝大部分算法基本都涉及到了。把满足平稳线性高斯假设下的数据分析，推进到非平稳、非线性、非高斯假设下的数据分析，是明确的发展方向。把上述思想观点算法用在不同领域中的不用场景下是一个实践问题。但理论至高无上！一定要弄清楚理论。

本文讨论的是平稳线性高斯假设下数据分析中涉及到的统计推断的具体公式的推导，是在典型的逆矩阵的辅助下进行的。通过详细的公式推导，可以更为深刻地理解在 Bayes 框架下，在平稳线性高斯假设下，统计推断中更细节的含义。这是广义数据分析知识框架中十分基础的内容，十分有必要弄清楚。

2 关于最佳决策与最佳推断问题的本质

决策是人类社会中无时无处不在的活动。做出最优决策是期望的。什么是决策？什么是最优决策？决策就是做出决定或做出选择。最优决策是代价最小或收益最大的决策。如何做出决策？如何做出最优决策？数学上如何表达？做决策必定依赖信息，没有信息的决策一定是随意做出的。信息/知识/认知是决策的依据。这些都最终来源于数据（广义的数据）。从数据中提出信息是决策的基础。由信息归纳出知识，广泛

地涉猎知识形成认知，再由认知进行决策是人类从事的决策活动。目前，ML/AI 的决策是（主要是）能定量地估计代价，在代价最小的原则下，由特定的算法所实现的决策。事实上，由数据中提取什么信息本身也是一个（次级的）决策过程。

数据被视为随机过程（对应物理或非物理世界的随机事件）的一次（或多次）具体实现，称为样本。因此，由样本数据估计其中蕴含的信息（潜变量），得到的结果（估计结果）也是随机的，符合某种概率分布的。假若样本数据抽象为向量 $\mathbf{d}$，与其对应的潜变量视为 $\mathbf{m}$，则由样本 $\mathbf{d}$ 估计出潜变量 $\mathbf{m}$，这本质上就是一个统计推断或统计决策过程。此时，潜变量 $\mathbf{m}$ 也是个随机过程（随机向量），应满足后验概率密度 $p(\mathbf{m}|\mathbf{d}^{obs})$。由概率论的知识，$\mathbf{m}$ 的估计（或推断）结果应是条件后验概率密度 $p(\mathbf{m}|\mathbf{d}^{obs})$ 的期望值 $\hat{\mathbf{m}}=\int \mathbf{m}p(\mathbf{m}|\mathbf{d}^{obs})d\mathbf{m}$。这似乎应是最优估计（推断）的最高理论框架！但这样的做法存在两个大问题：（a）后验概率密度 $p(\mathbf{m}|\mathbf{d}^{obs})$ 很难给出，尤其在高维情形下；（b）由于 $\mathbf{m}$ 是高维的，求期望公式中的积分很难计算。

由于由样本数据 $\mathbf{d}$ 估计潜变量 $\mathbf{m}$ 的估计结果 $\tilde{\mathbf{m}}$ 是随机过程，我们自然会期望估计结果 $\tilde{\mathbf{m}}$ 与真实结果 $\mathbf{m}_{true}$ 是无偏的，即 $E[\tilde{\mathbf{m}}]=\mathbf{m}_{true}$ 或偏差 $b(\tilde{\mathbf{m}})=E[\tilde{\mathbf{m}}-\mathbf{m}_{true}]=0$。

但是，当估计结果是渐进无偏时，需要定义均方误差：

$$\begin{aligned}
&E[(\tilde{\mathbf{m}}-\mathbf{m}_{true})^2]=E[(\tilde{\mathbf{m}}-E(\tilde{\mathbf{m}})+E(\tilde{\mathbf{m}})-\mathbf{m}_{true})^2]\\
&=E[(\tilde{\mathbf{m}}-E(\tilde{\mathbf{m}})^2]+2E[(\tilde{\mathbf{m}}-E(\tilde{\mathbf{m}}))(E(\tilde{\mathbf{m}})-\mathbf{m}_{true})]+E[(E(\tilde{\mathbf{m}})-\mathbf{m}_{true})^2]\\
&=E[(\tilde{\mathbf{m}}-E(\tilde{\mathbf{m}})^2]+2E[(\tilde{\mathbf{m}}-E(\tilde{\mathbf{m}}))(E(\tilde{\mathbf{m}})-\mathbf{m}_{true})]+E[(E(\tilde{\mathbf{m}}-\mathbf{m}_{true})^2]\\
&=E[(\tilde{\mathbf{m}}-E(\tilde{\mathbf{m}})^2]+2(E(\tilde{\mathbf{m}})-E(\tilde{\mathbf{m}}))(E(\tilde{\mathbf{m}})-\mathbf{m}_{true})+E[(E(\tilde{\mathbf{m}}-\mathbf{m}_{true})^2]\\
&=Var(\tilde{\mathbf{m}})+b^2(\tilde{\mathbf{m}})
\end{aligned}\tag{1}$$

无偏时，$b(\tilde{\mathbf{m}})=0$，$E(\mathbf{m})=\mathbf{m}_{true}$。否则，均方误差包括方差 $Var(\tilde{\mathbf{m}})$ 项和偏差平方项。在最佳统计推断时，要求均方误差最小，显然存在方差和偏差的折中问题。笔者认为，最佳推断首先期望无偏，然后才是期望方差最小，偏差大的推断失去了意义。但是，具体算法中如何达到上述期望目标，很难简单讲明白，不再赘述。理想的情形是，通过不断地提升数据样本的完备性，通过合适的算法，使得估计（决策）结果达到或逼近无偏，即 $E(\tilde{\mathbf{m}})=\mathbf{m}_{true}$ 的情形，其方差控制在满足要求的程度。这显然是理想的估计/推断/决策结果。由均方误差最小作为估计（决策）是否最佳的依据或准则是很合理的，含义很明确的定量准则。

至此可知为什么均方误差 $E[(\tilde{\mathbf{m}}-\mathbf{m}_{true})^2]$ 最小化成为了最常用的估计或决策判断原则（或定量标准）。

对比一下，Bayes 估计或最大似然（MLE）估计都没有均方误差最小化估计原则明了和实用，尽管它们有更高的理论层次。但是，也应该注意到：由于真值 $\mathbf{m}_{true}$ 未知，也由于求期望要用集平均，均方误差极小化作为最佳估计的标准也不实用。这就导致了最小二乘估计方法在实际应用中的盛行。

最小二乘估计方法的理论假设是什么？即应用于什么情形下才能达到最佳估计结果，或根本达不到最佳估计结果？这是否意味着我们所谓的最佳决策都是理论意义下的，实际上完全做不到最佳决策，至多是次最佳的！Gauss-Markov 定理回答了上述问题。

Gauss-Markov 定理：令 $\mathbf{b}$ 是一个可表示为 $\mathbf{b}=A\boldsymbol{\theta}+\mathbf{e}$ 的随机向量，其中 A 为 $N\times P$ 矩阵（$N>P$），其秩等于 P（即列满秩）；$\boldsymbol{\theta}$ 是一个未知（待估计）向量，而 $\mathbf{e}$ 为一个误差向量，若其均值向量 $E[\mathbf{e}]=0$，方差矩阵 $Var\{\mathbf{e}\}=\sigma^2 I$，$\sigma^2$ 为常数，则在所有线性无偏估计量中，最小二乘估计量具有最小的方差。即若 $\hat{\boldsymbol{\theta}}$ 是最小二乘估计量，$\tilde{\boldsymbol{\theta}}$ 是其它任何线性无偏估计量，有 $Var\{\hat{\boldsymbol{\theta}}\}\le Var\{\tilde{\boldsymbol{\theta}}\}$。

为了给后面基于矩阵的逆讨论最佳估计的方法做出铺垫，下面再分析一下由参数估计的均方误差转化为数据逼近的均方误差的思想。

首先由线性化的正问题 $\mathbf{d}=K\mathbf{m}_{true}+\boldsymbol{\eta}$，其中，$\boldsymbol{\eta}\sim(0,\sigma^2\mathbf{I})$，$K$ 是列满秩的。定义估计结果的均方误

差：$E\left[\left(\hat{\mathbf{m}}-\mathbf{m}_{true}\right)^2\right]$。

如何把上述估计结果的均方误差转化为由数据定义的均方误差：$E\left[\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)^2\right]$？这就要引入线性化的正问题，即：

$$E\left[\left(\hat{\mathbf{m}}-\mathbf{m}_{true}\right)^2\right]\Rightarrow E\left[\left(K\hat{\mathbf{m}}-K\mathbf{m}_{true}\right)^2\right]=E\left[\boldsymbol{\eta}^2\right]=E\left[\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)^2\right] \tag{2}$$

上述逻辑成立的条件是：K 是列满秩的，数据和参数是一对一的映射。这就奠定了基于数据定义的均方误差进行参数估计的理论基础。

勘探地震中，其它实际应用领域也是如此，由于实际数据采集很贵，一般仅仅采样一次，均方误差无法计算。因此，FWI/LS_RTM 成像中，上述均方误差泛函退化为最小二乘误差泛函 $\mathcal{J}\left(\mathbf{m}\right)=\frac{1}{2}\boldsymbol{\eta}^T\boldsymbol{\eta}=\frac{1}{2}\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)^T\left(\mathbf{d}^{cal}-\mathbf{d}^{obs}\right)$。由于 $\boldsymbol{\eta}\sim\left(0,\sigma^2\mathbf{I}\right)$，它是广义平稳的。因此，由各态遍历性，可以用时间平均（最小二乘误差定义）代替集平均（均方误差定义）。总之，时不变线性系统假设且逼近误差满足 $\boldsymbol{\eta}\sim\left(0,\sigma^2\mathbf{I}\right)$ 所导引出的实测数据满足广义平稳性是由均方误差泛函转化为最小二乘误差泛函的理论基础。但是，实际开展 FWI/LS_RTM 成像时，谁真的关心过上述假设是否满足吗？弄清楚地震波反演成像的理论假设真的非常必要。

前已述及，一维和高维数据分析的重要理论基础是 ARMA 模型的构架及基于此模型的最佳预测模型的形成。这是与参数估计问题同等重要的理论框架。

ARMA 模型是对零均值、方差为 Σ 的广义平稳随机过程建模的。其中的 AR 模型也可以被抽象为 $\mathbf{Z}_n=A\mathbf{Z}+\boldsymbol{\eta}_n$ 的形式，其中 $\mathbf{Z}=\left(Z_{n-1},Z_{n-2},\cdots,Z_{n-p}\right)$。假设 $\mathbf{Z}\sim\left(0,\Sigma_{zz}\right)$，$\eta_n\sim\left(0,\sigma^2\right)$。定义均方误差 $E\left[\left(z_n^{obs}-\hat{Z}_n\right)^2\right]$。根据 BLUE 估计子理论，均方误差最小化，导出 Wiener-Hopf方程，可以得到最佳的估计结果。至此，可以把信号处理、地震波参数估计等中遇到的均方误差最小到最小二乘误差泛函最小下的估计问题统一成一个完整的故事。理论上这是很优美的、逻辑上这是严密的。应该充分地注意到最佳估计/推断/决策理论成立的假设条件。

在上述理论问题讨论的铺垫下，就可以进入具体的最佳估计公式的推导过程。

3 一般矩阵的逆及其应用

3.1 一般矩阵逆的基本公式

已知矩阵 A，存在逆矩阵 A^{-1}，加上一个秩为 1 的矩阵后，变为：

$$B=A+\mathbf{x}\mathbf{y}^H \tag{3}$$

其中，$\mathbf{x}\mathbf{y}^H$ 代表（列）向量 $\mathbf{x}$ 和 $\mathbf{y}$ 的外积，其结果是一个秩 1 矩阵。则存在

$$B^{-1}=(A+\mathbf{x}\mathbf{y}^H)^{-1}=A^{-1}-\frac{A^{-1}\mathbf{x}\mathbf{y}^HA^{-1}}{1+\mathbf{y}^HA^{-1}\mathbf{x}} \tag{4}$$

此式称为 Sherman-Morrison 公式或 Woodburg‘s identity 恒等式。

更一般地，若矩阵 A 和 $B\in Mn$，Mn 代表 $n\times n$ 矩阵构成的线性空间。即 A 和 B 都是 $n\times n$ 的矩阵空间中的任意元素，且 $B=A+CRD$，C 是 $n\times r$ 的矩阵，R 是 $r\times r$ 的矩阵，D 是 $r\times n$ 的矩阵，若 B^{-1} 存在的话，则：

$$B^{-1}=(A+CRD)^{-1}=A^{-1}-A^{-1}C(R^{-1}+DA^{-1}C)^{-1}DA^{-1} \tag{5}$$

同时有：

$$B^{-1}=(A-CRD)^{-1}=A^{-1}+A^{-1}C(R^{-1}+DA^{-1}C)^{-1}DA^{-1} \tag{6}$$

该式用于递归最小二乘（RLS=Recursive Least Square）公式的导出过程中！

还存在如下的一个恒等式：

$$(A+CRD)^{-1}CR=A^{-1}C(R^{-1}+DA^{-1}C)^{-1} \tag{7}$$

该式也用在线性系统参数估计公式的推导中。

3.2 应用场景 I：递归最小二乘（RLS）算法公式的导出

关于自适应滤波中的 RLS 算法，其核心思想是：由 AR 模型出发，已知观测数据 $\mathbf{x}(i-1)=(x_{i-1},...,x_{i-m})^T$ 的情况下，假设观测数据是缓平稳变化的，设计出递归的“最佳”滤波器，达到自适应缓平缓假设下“最佳”滤波的期望目标。这与“机器学习”的思想已经有明显的联系，实质上这（自适应滤波）应该是机器学习思想的最初萌芽，只差把此思想及算法与人工神经元搭建起来的神经网络结合起来了。

AR 模型写为：

$$Y(i)=\mathbf{w}^H\mathbf{x} \tag{8}$$

其中，$\mathbf{w}$ 和 $\mathbf{x}$ 分别是滤波器和实测数据的列向量。定义单点预测误差为：

$$e(i)=d(i)-Y(i)=d(i)-\mathbf{w}^H\mathbf{x} \tag{9}$$

据此，有多点预测误差泛函的定义为：

$$E(\mathbf{w})=\frac{1}{2}\sum_{i=1}^{n}\lambda^{n-i}e(i)^2,i=1,2,...,n \tag{10}$$

其中，λ^{n-i} 代表距 n 点（当前点）距离为“$n-i$”的遗忘（遗望）因子且 $\lambda<1$。很显然，“$n-i$”越大，λ^{n-i} 越小。即距离 n 点（当前点）越远，加权系数越小。

显然，对应(10)式的法方程为：

$$\Phi(n)\mathbf{w}(n)=Z(n) \tag{11}$$

其中，

$$\Phi(n)=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{x}(i)\mathbf{x}^H(i) \tag{12}$$

$$Z(n)=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{x}(i)d(i) \tag{13}$$

由(12)和(13)式形式可看出，它可被重写为如下形式，

$$\Phi(n)=\lambda\Phi(n-1)+\mathbf{x}(n)\mathbf{x}^H(n) \tag{14}$$

$$Z(n)=\lambda Z(n-1)+\mathbf{x}(n)d(n) \tag{15}$$

(11)式显然也可重写为：

$$\mathbf{w}(n)=\Phi^{-1}(n)Z(n) \tag{16}$$

由矩阵恒定式：

$$A=B+CDC^H\text{ 及 }A^{-1}=B^{-1}-B^{-1}C(D^{-1}+C^HB^{-1}C)^{-1}C^HB^{-1} \tag{17}$$

知，若令，$A=\Phi(n),B=\lambda\Phi(n-1),C=\mathbf{x}(n),D=\mathbf{I}$，

则有：

$$\Phi^{-1}(n)=\lambda^{-1}\Phi^{-1}(n-1)-\frac{\lambda^{-1}\Phi^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\lambda^{-1}\Phi^{-1}(n-1)}{1+\lambda^{-1}\mathbf{x}^H(n)\Phi^{-1}(n-1)\mathbf{x}(n)} \tag{18}$$

为简洁计，令 $P(n)=\Phi^{-1}(n)$，则(18)式重写为：

$$P(n)=\lambda^{-1}P(n-1)-\lambda^{-1}\mathbf{k}(n)\mathbf{x}^H(n)P(n-1) \tag{19}$$

上式被称 RLS 的 Riccati 方程。其中，

$$\mathbf{k}(n)=\frac{\lambda^{-1}P(n-1)\mathbf{x}(n)}{1+\lambda^{-1}\mathbf{x}^H(n)P(n-1)\mathbf{x}(n)} \tag{20}$$

对上式进行整理，得到：

$$\mathbf{k}(n)+\lambda^{-1}\mathbf{k}(n)\mathbf{x}^H(n)P(n-1)\mathbf{x}(n)=\lambda^{-1}P(n-1)\mathbf{x}(n) \tag{21}$$

$$\begin{aligned}\mathbf{k}(n)&=\lambda^{-1}P(n-1)\mathbf{x}(n)-\lambda^{-1}\mathbf{k}(n)\mathbf{x}^H(n)P(n-1)\mathbf{x}(n)\\&=[\lambda^{-1}P(n-1)-\lambda^{-1}\mathbf{k}(n)\mathbf{x}^H(n)P(n-1)]\mathbf{x}(n)\\&=P(n)\mathbf{x}(n)=\Phi^{-1}(n)\mathbf{x}(n)\end{aligned} \tag{22}$$

由(16)及(15)式，得：

$$\mathbf{w}(n)=\Phi^{-1}(n)Z(n)=P(n)[\lambda Z(n-1)+\mathbf{x}(n)d(n)]=\lambda P(n)Z(n-1)+P(n)\mathbf{x}(n)d(n) \tag{23}$$

把(19)及(22)代入上式，继续推导，可以得到：

$$\begin{aligned}\mathbf{w}(n)&=\lambda\left[\lambda^{-1}P(n-1)-\lambda^{-1}\mathbf{k}(n)\mathbf{x}^H(n)P(n-1)\right]Z(n-1)+\mathbf{k}(n)d(n)\\&=P(n-1)Z(n-1)-\mathbf{k}(n)\mathbf{x}^H(n)P(n-1)Z(n-1)+\mathbf{k}(n)d(n)\\&=\mathbf{w}(n-1)-\mathbf{k}(n)\mathbf{x}^H(n)\mathbf{w}(n-1)+\mathbf{k}(n)d(n)\\&=\mathbf{w}(n-1)+\mathbf{k}(n)[d(n)-\mathbf{x}^H(n)\mathbf{w}(n-1)]\\&=\mathbf{w}(n-1)+\mathbf{k}(n)[d(n)-\mathbf{w}^H(n-1)\mathbf{x}(n)]^H\\&=\mathbf{w}(n-1)+\mathbf{k}(n)\varepsilon^H(n)\end{aligned} \tag{24}$$

其中，$\varepsilon^H(n)=d(n)-\mathbf{w}^H(n-1)\mathbf{x}(n)$ 称为前验估计误差。$\varepsilon^H(n)=d(n)-\mathbf{w}^H(n)\mathbf{x}(n)$ 被称为后验估计误差。$\mathbf{k}(n)$ 被称为递归滤波器增益。

因此，有如下的递归 RLS 解法：

① 计算 RLS 增益向量：$\mathbf{k}(n)$

② 计算前验估计误差：$\varepsilon^H(n)=d(n)-\mathbf{w}^H(n-1)\mathbf{x}(n)$

③ 更新滤波器系数：$\mathbf{w}(n)=\mathbf{w}(n-1)+\mathbf{k}(n)\varepsilon^H(n)$

④ 计算逆矩阵：$P(n)=\lambda^{-1}P(n-1)-\lambda^{-1}\mathbf{k}(n)\mathbf{x}^H(n)P(n-1)$

⑤ 计算当前滤波器输出：$Y(n)=\mathbf{w}^H(n)\mathbf{x}(n)$

⑥ 计算当前后验估计误差：$e(n)=d(n)-\mathbf{w}^H(n)\mathbf{x}(n)$，决定是否计算终止！

这即递归最小二乘法的完整推导。

3.3 应用场景 II：Kalman 滤波公式的导出

Kalman滤波是线性高斯平稳假设下最佳Wiener滤波的修改，以适应缓平稳情形下的信号建模及预测。

Kalman 滤波公式的导出可以在 Bayes 推断的框架下给出，更多地是在引入新息的框架下导出的。后者的推导需要用到前述的矩阵逆公式。

Kalman 滤波的最大概念创新是把缓平稳随机过程随时间的变化视为系统状态随时间的演变，因此引入了状态空间和实测数据空间的概念，定义了状态演变方程和数据观测方程。它们构成了 Kalman 滤波公式推导的基本出发点。

3.3.1 Bayes 推断观点下的 Kalman 滤波公式

从 Bayes 推断观点推导 Kalman 滤波，能给我们一种高屋建瓴的理论视野，能更容易地过渡到深度神经网络下的各种算法。尽管从实际计算的角度，也许基于新息的 Kalman 滤波公式更实用。下面简述一下 Bayes 推断观点下如何导出 Kalman 滤波公式。

3.3.1.1 Bayes 观点下的递归估计（Recursive Estimation）

首先定义状态变量 $\mathbf{x}_t$ 和观测变量 $\mathbf{y}_t$，$t=0,1,2,\cdots,N$。其中，$\mathbf{X}_t=(\mathbf{x}_0,\mathbf{x}_2,\cdots,\mathbf{x}_N)$ 是状态变量空间中的元素；$\mathbf{Y}_t=(\mathbf{y}_0,\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_N)$ 是观测变量空间中的元素。并假设状态演化存在 Markov 性。因此，有：

$$P(\mathbf{x}_{t+1}\mid\mathbf{x}_t,\mathbf{x}_{t-1},\cdots,\mathbf{x}_1,\mathbf{x}_0)=P(\mathbf{x}_{t+1}\mid\mathbf{x}_t) \tag{25}$$

$$P(\mathbf{y}_{t+1}\mid\mathbf{x}_{t+1},\mathbf{y}_t,\mathbf{y}_{t-1},...,\mathbf{y}_1,\mathbf{y}_0)=P(\mathbf{y}_{t+1}\mid\mathbf{x}_{t+1}) \tag{26}$$

Bayes 推断观点下，定义后验概率密度为：

$$P(\mathbf{x}_{t+1}\mid\mathbf{Y}_{t+1})=P(\mathbf{x}_{t+1}\mid\mathbf{Y}_t,\mathbf{y}_{t+1}) \tag{27}$$

即 $\mathbf{Y}_t$ 中增补一个新观测 $\mathbf{y}_{t+1}$ 后，形成 $\mathbf{Y}_{t+1}$，并据此推断 $\mathbf{x}_{t+1}$。很显然，逻辑关系是十分清楚的，关键是如何具体地实现这样的逻辑。由于本质上是依据新增加的数据 $\mathbf{y}_{t+1}$ 定义由 $\mathbf{x}_t$ 演化成 $\mathbf{x}_{t+1}$ 的概率，因此(27)是被称为条件后验概率。按照完整的 Bayes 公式，条件后验概率应有如下关系式：

$$P(\mathbf{x}_{t+1}\mid\mathbf{Y}_{t+1})=P(\mathbf{x}_{t+1}\mid\mathbf{Y}_t,\mathbf{y}_{t+1})=\frac{P(\mathbf{y}_{t+1}\mid\mathbf{x}_{t+1},\mathbf{Y}_t)}{P(\mathbf{y}_{t+1}\mid\mathbf{Y}_t)}P(\mathbf{x}_{t+1}\mid\mathbf{Y}_t) \tag{28}$$

$$P(\mathbf{x}_{t+1}\mid\mathbf{Y}_t)=\int P(\mathbf{x}_{t+1}\mid\mathbf{x}_t)P(\mathbf{x}_t\mid\mathbf{Y}_t)d\mathbf{x}_t \tag{29}$$

按照 Bayes 推断的基本逻辑，$P(\mathbf{y}_{t+1}\mid\mathbf{x}_{t+1},\mathbf{Y}_t)$ 和 $P(\mathbf{x}_{t+1}\mid\mathbf{Y}_t)$ 是两个先验概率密度函数，即(28)式是由两个先验概率密度生成后验概率密度。其中，$P(\mathbf{x}_{t+1}\mid\mathbf{Y}_t)$ 称传播步；$P(\mathbf{x}_{t+1}\mid\mathbf{Y}_{t+1})$ 称更新步。交替（Recursive）执行上述两步，显然可实现状态空间的演化。这就是 Bayes 推断理论下 Kalman 滤波的理论基础。

3.3.1.2 Bayes 观点下的 Kalman 滤波公式

上述理论基础是概念性的，并不能进行实际的运算。有必要把上述概念性理论转化为实际可以操作的算法。首先引入 Kalman 滤波的概念模型。

众所周知，Kalman 滤波概念模型为：

$$\mathbf{x}_{t+1}=A_t\mathbf{x}_t+\mathbf{w}_t \tag{30}$$

$$\mathbf{y}_t=C_t\mathbf{x}_t+\mathbf{v}_t \tag{31}$$

其中，(30)是转态演化方程，$\mathbf{w}_t$ 代表 t 时刻的状态演化（单步）预测误差；(31)是数据观测方程，$\mathbf{v}_t$ 代表 t 时刻的数据观测误差，也可以认为是单步预测实测数据的残余误差。

首先针对转播步展开讨论。$P(\mathbf{x}_t\mid\mathbf{Y}_t)$ 代表由 $\mathbf{Y}_t$ 估计。假定其为 Gauss 分布，有

$$\mathbf{x}_t\mid\mathbf{Y}_t\sim N\left(\hat{\mathbf{x}}_{t|t},P_{t|t}\right) \tag{32}$$

$P(\mathbf{x}_{t+1}\,|\,\mathbf{x}_t)$ 对应状态演化，假定其也满足 Gauss 分步，即成立：

$$\mathbf{x}_{t+1}\,|\,\mathbf{x}_{t+1}\sim N\left(A_t\hat{\mathbf{x}}_{t|t},Q_{t|t}^{\mathbf{ww}}\right) \tag{33}$$

因此，$P(\mathbf{x}_{t+1}\,|\,\mathbf{Y}_t)$ 也满足 Gauss 分布。由 MAP 准则，可以导出：

$$P(\mathbf{x}_{t+1}\,|\,\mathbf{Y}_t)\sim N\left(A_t\hat{\mathbf{x}}_t,P_{t+1|t}\right) \tag{34}$$

其中，协方差 $P_{t+1|t}=A_tP_{t|t}A_t^T+Q_t^{\mathbf{ww}}$ 由 Gauss 线性估计理论给出。

然后，针对更新步进行讨论。由 $P(\mathbf{x}_{t+1}\,|\,\mathbf{Y}_{t+1})$ 满足 Gauss 假设及线性的 Kalman 滤波模型下的估计理论知：

$$\hat{\mathbf{x}}_{t+1|t+1}=E\left[\mathbf{x}_{t+1}\,|\,\mathbf{Y}_t\right]=\hat{\mathbf{x}}_{t+1|t}+R_{\mathbf{xy},\mathbf{Y}_t}R_{yy,\mathbf{Y}_t}^{-1}\left(\mathbf{y}_{t+1}-E\left[\mathbf{y}_{t+1}\right]\right)_{\mathbf{Y}_t} \tag{35}$$

其中，

$$\begin{aligned}R_{\mathbf{xy},\mathbf{Y}_t}&=E\left[\left(\mathbf{x}_{t+1}-E\left[\mathbf{x}_{t+1}\right]\right)\left(\mathbf{y}_{t+1}-E\left[\mathbf{y}_{t+1}\right]\right)^T\right]_{\mathbf{Y}_t}\\&=E\left[\left(\mathbf{x}_{t+1}-\hat{\mathbf{x}}_{t+1}\right)\left(C_{t+1}\left(\mathbf{x}_{t+1}-\hat{\mathbf{x}}_{t+1}\right)+\mathbf{v}_{t+1}\right)^T\right]_{\mathbf{Y}_t}\\&=P_{t+1|t}C_{t+1}^T\end{aligned} \tag{36}$$

$$\begin{aligned}R_{\mathbf{yy},\mathbf{Y}_t}&=E\left[\left(\mathbf{y}_{t+1}-E\left[\mathbf{y}_{t+1}\right]\right)\left(\mathbf{y}_{t+1}-E\left[\mathbf{y}_{t+1}\right]\right)^T\right]_{\mathbf{Y}_t}\\&=E\left[\left(C_{t+1}\left(\mathbf{x}_{t+1}-\hat{\mathbf{x}}_{t+1}\right)+\mathbf{v}_{t+1}\right)\left(C_{t+1}\left(\mathbf{x}_{t+1}-\hat{\mathbf{x}}_{t+1}\right)+\mathbf{v}_{t+1}\right)^T\right]_{\mathbf{Y}_t}\\&=C_{t+1}P_{t+1|t}C_{t+1}^T+R_{t+1}^{\mathbf{vv}}\end{aligned} \tag{37}$$

$$E\left[\mathbf{y}_{t+1}\middle|\mathbf{Y}_t\right]=C_{t+1}\hat{\mathbf{x}}_{t+1} \tag{38}$$

把(36)、(37)和(38)代入(35)式，可以得到：

$$\hat{\mathbf{x}}_{t+1|t+1}=\hat{\mathbf{x}}_{t+1|t}+P_{t+1|t}C_{t+1}^T\left(C_{t+1}P_{t+1|t}C_{t+1}^T+R_{t+1}^{\mathbf{vv}}\right)^{-1}\left(\mathbf{y}_{t+1}-C_{t+1}\hat{\mathbf{x}}_{t+1|t}\right)_{\mathbf{Y}_t} \tag{39}$$

令 $K_{t+1}=\left(C_{t+1}P_{t+1|t}C_{t+1}^T+R_{t+1}^{\mathbf{vv}}\right)^{-1}$ 为 Kalman 增益。则(39)式又可以写为：

$$\hat{\mathbf{x}}_{t+1|t+1}=\hat{\mathbf{x}}_{t+1|t}+K_{t+1}\left(\mathbf{y}_{t+1}-C_{t+1}\hat{\mathbf{x}}_{t+1|t}\right)_{\mathbf{Y}_t} \tag{40}$$

最后，再讨论 $\mathbf{x}_{t+1|t+1}\sim N\left(\hat{\mathbf{x}}_{t+1|t+1},P_{t+1|t+1}\right)$ 中的方差 $P_{t+1|t+1}$。只有给出方差 $P_{t+1|t+1}$ 的递归计算关系式，才能给出 Bayes 推断下 Kalman 滤波的具体实现公式。

理论可知，在 Gauss 假设下由已知显变量 $\mathbf{y}$ Bayes 推断潜变量 $\mathbf{x}$ 的后验协方差为：

$$\mathrm{COV}\left(\mathbf{x}\middle|\mathbf{y}\right)=R_{\mathbf{xx}}-R_{\mathbf{xy}}R_{\mathbf{yy}}^{-1}R_{\mathbf{yx}} \tag{41}$$

据此，可得：

$$\begin{aligned}P_{t+1|t+1}&=P_{t+1|t}-P_{t+1|t}C_{t+1}^T\left(C_{t+1}P_{t+1|t}C_{t+1}^T+R_{t+1}^{\mathbf{vv}}\right)^{-1}C_{t+1}P_{t+1|t}\\&=P_{t+1|t}-K_{t+1}C_{t+1}P_{t+1|t}=\left(I-K_{t+1}C_{t+1}\right)P_{t+1|t}\end{aligned} \tag{42}$$

其中，$P_{t+1|t}=P_{t+1|t}^T$，$R_{\mathbf{xx}}=P_{t+1|t}$，$R_{\mathbf{xy}}=P_{t+1|t}C_{t+1}^T$，$R_{\mathbf{yy}}=\left(C_{t+1}P_{t+1|t}C_{t+1}^T+R_{t+1}^{\mathbf{vv}}\right)^{-1}$。

至此完成了 Bayes 框架下 Kalman 滤波公式的完整推导。

最后再归纳一下 Kalman 滤波公式的算法。

初始条件：$\hat{\mathbf{x}}_{0|-1}$ 及其对应的协方差 $P_{-1|-1}$；已知观测随机向量 $\mathbf{Y}_t=\{\mathbf{y}_0,\mathbf{y}_1,\cdots,\mathbf{y}_t\}$，

Step1：状态传播 $\hat{\mathbf{x}}_{t+1|t}=A_t\hat{\mathbf{x}}_{t|t}$；

Step2：对应的协方差 $P_{t+1|t}=A_tP_{t|t}A_t^T+Q_t^{\mathbf{ww}}$；

Step3：Kalman 增益 $K_{t+1}=\left(C_{t+1}P_{t+1|t}C_{t+1}^T+R_{t+1}^{\mathbf{vv}}\right)^{-1}$；

Step4: 状态更新 $\hat{\mathbf{x}}_{t+1|t+1}=\hat{\mathbf{x}}_{t+1|t}+K_{t+1}\left(\mathbf{y}_{t+1}-C_{t+1}\hat{\mathbf{x}}_{t+1|t}\right)_{\mathbf{Y}_t}$；

Step5：对应的协方差 $P_{t+1|t+1}=\left(I-K_{t+1}C_{t+1}\right)P_{t+1|t}$。

3.3.2 最佳滤波观点下的 Kalman 滤波公式（The innovation approach 新息方法）

把实测数据与新息建立起一对一的关系，是推导出不一定在 Gauss 假设下的 Kalman 滤波公式的关键思路。同时利用矩阵逆的恒等式，也是其中关键的环节。

3.3.2.1 关于新息

Kalman 滤波看待数据的观点：数据由系统输出生成，是对系统输出（即系统状态）的观测结果。系统（状态）是随时间不断演化的。每时刻系统处于特定的状态，转态向量由 $\mathbf{x}$ 表示。系统输出状态的观测结果向量由 $\mathbf{y}$ 表示。因此有状态演化方程：

$$\mathbf{x}_{t+1}=R\mathbf{x}_t+\mathbf{w}_t \tag{43}$$

式中 $\mathbf{w}_t$ 为预测噪声。以及数据观测方程：

$$\mathbf{y}_t=C_t\mathbf{x}_t+\mathbf{v}_t \tag{44}$$

式中 $\mathbf{v}_t$ 为观测噪声。这就完整地建立了系统状态演化及系统输出观测的模型。在数据（或信号）分析领域中，这是个十分经典的模型。

为了在新息观点下推导出 Kalman 滤波公式，对于(43)和(44)式定义的系统状态演化及系统输出观测的模型，需要引入如下假设：

1、所有过程中涉及到的量均是实的。

2、状态噪声（AR 模型预测噪声）过程是零均值的，由如下方差：

$E[\mathbf{w}_t\mathbf{w}_\tau^T]=\mathbf{Q}_t^{\mathbf{ww}}\delta_{t,\tau}$，假设 $\mathbf{w}_t$ 是 Gauss 噪声（不必一定是白噪！）

3、观测噪声 $\mathbf{v}_t$ 是零均值的，有如下方差：

$E[\mathbf{v}_t\mathbf{v}_\tau^T]=\mathbf{R}_t^{\mathbf{vv}}\delta_{t,\tau}$，假设 $\mathbf{v}_t$ 是 Gauss 噪声（不必一定是白噪！）

4、假设状态噪声与观测噪声有相关性，即：$E[\mathbf{w}_t\mathbf{v}_\tau^T]=M_t\delta_{t,\tau},\forall t,\tau$，事实上也可假设二者是不相关的！

5、可以假设初始状态时，随机向量 $\mathbf{x}_0$ 是有均值 $\boldsymbol{\mu}_{\mathbf{x}_0}$ 和方差 Π_0 的，也可设为零均值的。

后续在上述假设的基础上，进行 Kalman 滤波递归公式的推导。Kalman 滤波是线性高斯平稳假设下最佳 Wiener 滤波扩展到缓平稳情形下的对应物。因此，其推导过程必然还会用到下面一些基本概念。

已知观测数据向量序列 $\mathbf{Y}_t=\{\mathbf{y}_0,\mathbf{y}_1,\cdots,\mathbf{y}_t\}$，由 $\mathbf{Y}_t$ 对状态向量 $\mathbf{x}$ 的估计结果记为 $\hat{\mathbf{x}}_t$，而且用 AR 模型进

行建模预测（估计），有：

$$\hat{\mathbf{x}}_t = \sum_{i=0}^{t} a_i \mathbf{y}_i \tag{45}$$

本质上，可以说是 $\hat{\mathbf{x}}_t$ 由状态向量 $\mathbf{x}$ 在 $\mathbf{Y}_t$ 张成的空间中的投影结果定！因此，由正交原理知道：

$$E[(\mathbf{x}-\hat{\mathbf{x}}_t)\mathbf{y}_i^T] = 0, \quad i = 0,1,...,t \tag{46}$$

现在寻找一个正交矢量序列 $\{\boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_t\}$ 张成与观测数据向量序列 $\mathbf{Y}_t$ 所构成的相同的空间，即：

$$Span(\boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_t) = Span(\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_t) \tag{47}$$

满足这种要求的向量 $\boldsymbol{\epsilon}_i$ 被称为 $\mathbf{y}_i$ 的新息。$\{\boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_t\}$ 被称为新息序列（Innovation Sequence）。新息序列由观测数据向量序列 $\mathbf{Y}_t$ 经过 Gram-Schmidt 正交化处理而形成。

另外，假设 $\mathbf{y}_t$ 投影到子空间 $\mathbf{Y}_{t-1}$ 上，得到 $\hat{\mathbf{y}}_{t-1}$，则预测误差为：

$$\boldsymbol{\epsilon}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t-1} \tag{48}$$

这就是新息的定义式。可见，所谓新息，即新观测数据 $\mathbf{y}_t$ 中所包含的不能被 $\mathbf{Y}_{t-1}$ 所表达的那部分新的信息。如果新时刻的数据不包含新的信息，说明系统本身没有演化。也就不必要花功夫进行 Kalman 滤波了。

相似的，$\hat{\mathbf{x}}_{t-1}$ 是 $\mathbf{x_t}$ 投影到 $\mathbf{Y}_{t-1}$ 上的结果；$\hat{\mathbf{v}}_{t-1}$ 是 $\mathbf{v}_t$ 投影到 $\mathbf{Y}_{t-1}$ 上的结果。有：

$$\hat{\mathbf{y}}_{t-1} = C_t \hat{\mathbf{x}}_{t-1} + \hat{\mathbf{v}}_{t-1} \tag{49}$$

显然，$\mathbf{v}_t$ 与 $\mathbf{Y}_{t-1}$ 中各向量正交，因此 $\hat{\mathbf{v}}_{t-1} = 0$。因此，有：

$$\hat{\mathbf{y}}_{t-1} = C_t \hat{\mathbf{x}}_{t-1} \tag{50}$$

令 $\boldsymbol{\epsilon}_t = \{\boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_t\}$，且有：

$$Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \begin{cases} R_{\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j}, & i = j \\ 0, & i \neq j \end{cases} \tag{51}$$

3.3.2.2 基于新息过程的估计

假如已知新息序列，即此时有 $\mathbf{Y}_t = \boldsymbol{\epsilon}_t$，则对 $\hat{\mathbf{x}}_t$ 的均方误差最小的估计有：

$$\hat{\mathbf{x}}_t = \sum_{i=0}^{t} a_i \mathbf{y}_i \tag{52}$$

同时，也应该有：

$$\hat{\mathbf{x}}_t = \sum_{i=0}^{t} b_i \boldsymbol{\epsilon}_i \tag{53}$$

很显然，(52)和(53)应该是等价的。

由于 $\boldsymbol{\epsilon}_t$ 是正交向量序列，则由(53)式有：$E(\hat{\mathbf{x}}_t, \boldsymbol{\epsilon}_j) = \sum_{i=0}^{t} b_i E(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j), \quad j = 0,1,...,t$，其对应的矩阵形式为 $R_{\mathbf{x}_t, \boldsymbol{\epsilon}_i} = b_i R_{\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_i}$，则有：

$$b_i = R_{\mathbf{x}_t, \boldsymbol{\epsilon}_i} R^{-1}_{\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_i} \tag{54}$$

若增加一个新数据 $\mathbf{y}_{t+1}$，则有：

$$\hat{\mathbf{x}}_{t+1}=\sum_{i=0}^{t+1}b_i\boldsymbol{\epsilon}_i=\sum_{i=0}^{t}b_i\boldsymbol{\epsilon}_i+b_{i+1}\boldsymbol{\epsilon}_{i+1}=\hat{\mathbf{x}}_t+R_{\hat{\mathbf{x}}_t,\boldsymbol{\epsilon}_{i+1}}R^{-1}_{\boldsymbol{\epsilon}_{i+1},\boldsymbol{\epsilon}_{i+1}}\boldsymbol{\epsilon}_{i+1} \tag{55}$$

其中，

$$\boldsymbol{\epsilon}_{t+1}=\mathbf{y}_{t+1}-C_t\hat{\mathbf{x}}_t \tag{56}$$

式(56)即为新息更新公式。本质上，这是用新息序列进行基于 AR 模型的滤波，推导出该式的递归滤波公式显然是核心问题。

3.3.2.3 新息用于状态空间模型的处理过程

首先，估计的状态演化过程为：

$$\hat{\mathbf{x}}_{t+1}=R_t\hat{\mathbf{x}}_t+\hat{\mathbf{w}}_t \tag{57}$$

估计误差 $\hat{\mathbf{w}}_t$ 同样地可由新息序列写为：

$$\hat{\mathbf{w}}_t=\sum_{j=0}^{t}R_{\mathbf{w}_j,\boldsymbol{\epsilon}_j}R^{-1}_{\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j}\boldsymbol{\epsilon}_j=\sum_{j=0}^{t}E(\mathbf{w}_j\boldsymbol{\epsilon}_j^T)R^{-1}_{\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j}\boldsymbol{\epsilon}_j \tag{58}$$

其中，

$$E\left(\mathbf{w}_t\boldsymbol{\epsilon}_j^T\right)=E\left(\mathbf{w}_t\left(C_j\mathbf{x}_j+\mathbf{v}_j-C_j\hat{\mathbf{x}}_{j|j-1}\right)_j^T\right)=M_t\delta_{jt} \tag{59}$$

由前述对状态空间系统(43)和(44)的假设，知：$E(\mathbf{w}_t\mathbf{x}_j)=0$，对于 $j\le t$,有 $E(\mathbf{w}_t{\mathbf{v}_j}^T)=M_t\delta_{tj}$。因此，由(58)有：

$$\hat{\mathbf{w}}_{t|t}=M_tR^{-1}_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}\boldsymbol{\epsilon}_t \tag{60}$$

则(57)式重写为：

$$\hat{\mathbf{x}}_{t+1}=R_t\hat{\mathbf{x}}_t+M_tR^{-1}_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}\boldsymbol{\epsilon}_t \tag{61}$$

(61)式定义的方程称传播方程，这一步执行传播过程。

由(55)式及(61)式，可写出：

$$\hat{\mathbf{x}}_{t+1|t+1}=\hat{\mathbf{x}}_{t+1|t}+E\left[\hat{\mathbf{x}}_{t+1}\boldsymbol{\epsilon}_{t+1}^T\right]R^{-1}_{\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}_{t+1}}\boldsymbol{\epsilon}_{t+1} \tag{62}$$

这一步称为更新步。

(62)和(61)合起来构成 Kalman 滤波的核心步骤。

3.3.2.4 Kalman 递归公式的推导

1、关于(62)式中期望 $E\left[\hat{\mathbf{x}}_{t+1}\boldsymbol{\epsilon}_{t+1}^T\right]$及 $R_{\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}_{t+1}}$ 的计算

首先定义状态预测误差：

$$\boldsymbol{\varepsilon}_{t|t-1}=\mathbf{x}_t-\hat{\mathbf{x}}_{t|t-1} \tag{63}$$

它的自协方差矩阵定义为：

$$P_{t|t-1}=E[\boldsymbol{\varepsilon}_{t|t-1}{\boldsymbol{\varepsilon}^T}_{t|t-1}] \tag{64}$$

新息 $\boldsymbol{\epsilon}_t$ 写为：

$$\begin{aligned}\boldsymbol{\epsilon}_t &= \mathbf{y}_t - C_t\hat{\mathbf{x}}_{t|t-1}\\ &= C_t\mathbf{x}_t + \mathbf{v}_t - C_t\hat{\mathbf{x}}_{t|t-1}\\ &= C_t\boldsymbol{\varepsilon}_{t|t-1} + \mathbf{v}_t\end{aligned} \tag{65}$$

新息 $\boldsymbol{\epsilon}_t$ 的自协方差可写为：

$$\begin{aligned}R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t} &= E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T]\\ &= C_tE[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T]C_t^T + C_tE[\boldsymbol{\varepsilon}_{t|t-1}\mathbf{v}_t^T] + E[\mathbf{v}_t\boldsymbol{\varepsilon}_{t|t-1}]C_t^T + E[\mathbf{v}_t\mathbf{v}_t^T]\\ &= C_tP_{t|t-1}C_t^T + \Gamma_{\mathbf{vv}}\end{aligned} \tag{66}$$

其中，$E[\boldsymbol{\varepsilon}_{t|t-1}\mathbf{v}_t^T]=0$，$\Gamma_{\mathbf{vv}}=E[\mathbf{v}_t\mathbf{v}_t^T]$。另外，

$$\begin{aligned}E\left[\mathbf{x}_t\boldsymbol{\epsilon}_t^T\right] &= E\left[\mathbf{x}_t\left(C_t\boldsymbol{\varepsilon}_{t|t-1}+\mathbf{v}_t\right)^T\right]\\ &= E\left[\mathbf{x}_t\boldsymbol{\varepsilon}_{t|t-1}^T\right]C_t^T + E\left[\mathbf{x}_t\mathbf{v}_t^T\right]\\ &= E\left[\left(\hat{\mathbf{x}}_{t|t-1}+\boldsymbol{\varepsilon}_{t|t-1}\right)\boldsymbol{\varepsilon}_{t|t-1}^T\right]C_t^T\\ &= E\left[\hat{\mathbf{x}}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right]C_t^T + E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right]C_t^T\\ &= P_{t|t-1}C_t^T\end{aligned} \tag{67}$$

其中，$E\left[\mathbf{x}_t\mathbf{v}_t^T\right]=0$ 和 $E\left[\hat{\mathbf{x}}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right]C_t^T=0$

2、关于 $P_{t|t-1}$ 的递归计算的推导

首先令：

$$\Pi_t = E[\mathbf{x}_t\mathbf{x}_t^T] \tag{68}$$

$$\Sigma_{t|t-1} = E[\hat{\mathbf{x}}_{t|t-1}\hat{\mathbf{x}}_{t|t-1}^T] \tag{69}$$

已知 $\mathbf{x}_t=\hat{\mathbf{x}}_{t|t-1}+\boldsymbol{\varepsilon}_{t|t-1}$，则：

$$\begin{aligned}\Pi_t &= E\left[\mathbf{x}_t\mathbf{x}_t^T\right] = E\left[\left(\hat{\mathbf{x}}_{t|t-1}+\boldsymbol{\varepsilon}_{t|t-1}\right)\left(\hat{\mathbf{x}}_{t|t-1}+\boldsymbol{\varepsilon}_{t|t-1}\right)^T\right]\\ &= E\left[\hat{\mathbf{x}}_{t|t-1}\hat{\mathbf{x}}_{t|t-1}^T\right] + E\left[\hat{\mathbf{x}}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right] + E\left[\boldsymbol{\varepsilon}_{t|t-1}\hat{\mathbf{x}}_{t|t-1}^T\right] + E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right]\\ &= \Sigma_{t|t-1} + P_{t|t-1}\end{aligned} \tag{70}$$

初始时刻，令 $\Sigma_{0|1}=0, P_{0|1}=\Pi_0$

则由(70)式知：

$$P_{t|t-1} = \Pi_t - \Sigma_{t|t-1} \tag{71}$$

进而：

$$P_{t+1|t} = \Pi_{t+1} - \Sigma_{t+1|t} \tag{72}$$

要形成递归关系，需对 Π_{t+1} 和 $\Sigma_{t+1|t}$ 进行分析。

$$\begin{aligned}\Pi_{t+1} &= E\left[\mathbf{x}_{t+1}\mathbf{x}_{t+1}^T\right] = E\left[\left(R_t\mathbf{x}_t+\mathbf{w}_t\right)\left(R_t\mathbf{x}_t+\mathbf{w}_t\right)^T\right]\\ &= R_tE\left[\mathbf{x}_t\mathbf{x}_t^T\right]R_t^T + R_tE\left[\mathbf{x}_t\mathbf{w}_t^T\right] + E\left[\mathbf{w}_t\mathbf{x}_t^T\right]R_t^T + E\left[\mathbf{w}_t\mathbf{w}_t^T\right]\\ &= R_t\Sigma_{t|t-1}R_t^T + Q_t\end{aligned} \tag{73}$$

其中， $E\left[\mathbf{x}_t\mathbf{w}_t^T\right]=0$。

$$\begin{aligned}\Sigma_{t+1}&=E[\hat{\mathbf{x}}_{t+1|t}\hat{\mathbf{x}}_{t+1|t}^T]\\&=E\left[\left(R_t\hat{\mathbf{x}}_{t|t}+M_tR_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}\boldsymbol{\epsilon}_t\right)\left(R_t\hat{\mathbf{x}}_{t|t}+M_tR_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}\boldsymbol{\epsilon}_t\right)^T\right]\\&=R_tE\left[\hat{\mathbf{x}}_{t|t}\hat{\mathbf{x}}_{t|t}^T\right]R_t^T+R_tE\left[\hat{\mathbf{x}}_{t|t}\boldsymbol{\epsilon}_t^T\right]R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}M_t+M_tR_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}E\left[\boldsymbol{\epsilon}_t\hat{\mathbf{x}}_{t|t}\right]+M_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\right]M_t^T\\&=R_tE\left[\hat{\mathbf{x}}_{t|t}\hat{\mathbf{x}}_{t|t}^T\right]R_t^T+M_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\right]M_t^T\end{aligned}\tag{74}$$

其中，

$$E\left[\hat{\mathbf{x}}_{t|t}\boldsymbol{\epsilon}_t^T\right]=0\tag{75}$$

$$\begin{aligned}E\left[\hat{\mathbf{x}}_{t|t}\hat{\mathbf{x}}_{t|t}^T\right]&=E\left[\left(\hat{\mathbf{x}}_{t|t-1}+E\left[\hat{\mathbf{x}}_t\boldsymbol{\epsilon}_t^T\right]R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}}^{-1}\boldsymbol{\epsilon}_t\right)\left(\hat{\mathbf{x}}_{t|t-1}+E\left[\hat{\mathbf{x}}_t\boldsymbol{\epsilon}_t^T\right]R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}}^{-1}\boldsymbol{\epsilon}_t\right)^T\right]\\&=E\left[\hat{\mathbf{x}}_{t|t-1}\hat{\mathbf{x}}_{t|t-1}^T\right]+E\left[\hat{\mathbf{x}}_t\boldsymbol{\epsilon}_t^T\right]R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}}^{-1}E\left[\hat{\mathbf{x}}_t\boldsymbol{\epsilon}_t^T\right]\end{aligned}\tag{76}$$

注意，其中的交叉期望等于零。把(76)代入(74)式，有：

$$\begin{aligned}\Sigma_{t+1|t}&=R_tE\left[\hat{\mathbf{x}}_{t|t}\hat{\mathbf{x}}_{t|t}^T\right]R_t^T+M_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\right]M_t^T\\&=R_t\left(E\left[\hat{\mathbf{x}}_{t|t-1}\hat{\mathbf{x}}_{t|t-1}^T\right]+E\left[\mathbf{x}_t\boldsymbol{\epsilon}_t^T\right]R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}E\left[\boldsymbol{\epsilon}_t\mathbf{x}_t^T\right]\right)R_t^T+M_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\right]M_t^T\\&=R_t\Sigma_{t|t-1}R_t^T+\left(R_tE\left[\mathbf{x}_t\boldsymbol{\epsilon}_t^T\right]+M_t\right)R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}\left(R_tE\left[\mathbf{x}_t\boldsymbol{\epsilon}_t^T\right]+M_t\right)^T\end{aligned}\tag{77}$$

把(73)和(77)代入(72)，得：

$$\begin{aligned}P_{t+1|t}&=\Pi_{t+1}-\Sigma_{t+1|t}=R_t\Pi_tR_t^T+Q_t-R_t\Sigma_{t|t-1}R_t^T-\left(R_tE\left[\mathbf{x}_t\boldsymbol{\epsilon}_t^T\right]+M_t\right)R_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}\left(R_tE\left[\mathbf{x}_t\boldsymbol{\epsilon}_t^T\right]+M_t\right)^T\\&=R_t\left(\Pi_t-\Sigma_{t|t-1}\right)R_t^T+Q_t-\left(R_tP_{t|t-1}C_t^T+M_t\right)\left(C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{ww}}\right)^{-1}\left(R_tP_{t|t-1}C_t^T+M_t\right)^T\\&=R_tP_{t|t-1}R_t^T+Q_t-\left(R_tP_{t|t-1}C_t^T+M_t\right)\left(C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{ww}}\right)^{-1}\left(R_tP_{t|t-1}C_t^T+M_t\right)^T\end{aligned}\tag{78}$$

规定 $P_{0|-1}=\Pi_0$。这就产生了 $P_{t+1|t}$ 的递归计算公式。

3.3.2.5 Kalman 滤波递归公式的最后归纳

重写传播步和更新步如下：

$$\begin{aligned}\hat{\mathbf{x}}_{t+1|t}&=R_t\hat{\mathbf{x}}_{t|t}+M_tR_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}\boldsymbol{\epsilon}_t\\&=R_t\hat{\mathbf{x}}_{t|t}+M_t\left(C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{ww}}\right)^{-1}\left(\mathbf{y}_t-C_t\hat{\mathbf{x}}_{t|t-1}\right)\end{aligned}\tag{79}$$

$$\begin{aligned}\hat{\mathbf{x}}_{t+1|t+1}&=R_t\hat{\mathbf{x}}_{t+1|t}+M_tR_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}\boldsymbol{\epsilon}_t\\&=R_t\hat{\mathbf{x}}_{t+1|t}+P_{t+1|t}C_{t+1}^T\left(C_{t+1}P_{t+1|t}C_{t+1}^T+\Gamma_{\mathbf{ww}}\right)^{-1}\left(\mathbf{y}_{t+1}-C_{t+1}\hat{\mathbf{x}}_{t+1|t}\right)\end{aligned}\tag{80}$$

为了使公式含义更明了，定义 Kalman 增益（Kalman Gain）：

$$K_t=P_{t|t-1}C_t^TR_{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t}^{-1}=P_{t|t-1}C_t^T\left(C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{ww}}\right)^{-1}\tag{81}$$

则(80)可重写为：

$$\begin{aligned}\hat{\mathbf{x}}_{t+1|t+1}&=R_t\hat{\mathbf{x}}_{t+1|t}+P_{t+1|t}C_{t+1}^T\left(C_{t+1}P_{t+1|t}C_{t+1}^T+\Gamma_{\mathbf{ww}}\right)^{-1}\left(\mathbf{y}_{t+1}-C_{t+1}\hat{\mathbf{x}}_{t+1|t}\right)\\&=R_t\hat{\mathbf{x}}_{t+1|t}+K_{t+1}\left(\mathbf{y}_{t+1}-C_{t+1}\hat{\mathbf{x}}_{t+1|t}\right)\end{aligned}\tag{82}$$

其中，$K_{t+1}=P_{t+1|t}C_{t+1}^T\left(C_{t+1}P_{t+1|t}C_{t+1}^T+\Gamma_{\mathbf{vv}}\right)^{-1}$ 就是 Kalman 增益。

另外，为了进一步简化(78)，使 Kalman 滤波公式显得洁，意义更清楚，前述已定义估计误差协方差为：$P_{t|t-1}=E[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\varepsilon}^T{}_{t|t-1}]$，现在再定义一个滤波后的估计误差协方差：

$$P_{t|t}=E[\boldsymbol{\varepsilon}_{t|t}\boldsymbol{\varepsilon}^T{}_{t|t}] \tag{83}$$

其中，$\boldsymbol{\varepsilon}_{t|t}=\mathbf{x}_t-\hat{\mathbf{x}}_{t|t}$。把(82)代入(83)式得到：

$$\begin{aligned}
P_{t|t}&=E\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right]=E\left[\left(\mathbf{x}_t-\hat{\mathbf{x}}_{t|t}\right)\left(\mathbf{x}_t-\hat{\mathbf{x}}_{t|t}\right)^T\right]\\
&=E\left[\left(\mathbf{x}_t-\hat{\mathbf{x}}_{t|t-1}-K_t\boldsymbol{\epsilon}_t\right)\left(\mathbf{x}_t-\hat{\mathbf{x}}_{t|t-1}-K_t\boldsymbol{\epsilon}_t\right)^T\right]\\
&=E\left[\left(\boldsymbol{\varepsilon}_{t|t-1}-K_t\boldsymbol{\epsilon}_t\right)\left(\boldsymbol{\varepsilon}_{t|t-1}-K_t\boldsymbol{\epsilon}_t\right)^T\right]\\
&=E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right]-E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\epsilon}_t^T\right]K_t^T-K_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\varepsilon}_{t|t-1}^T\right]+K_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\right]K_t^T\\
&=P_{t|t-1}-E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\epsilon}_t^T\right]K_t^T-K_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\varepsilon}_{t|t-1}^T\right]+K_tE\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\right]K_t^T
\end{aligned} \tag{84}$$

其中，

$$\begin{aligned}
\boldsymbol{\epsilon}_t&=\mathbf{y}_t-C_t\hat{\mathbf{x}}_{t|t-1}=C_t\mathbf{x}_t+\mathbf{v}_t-C_t\hat{\mathbf{x}}_{t|t-1}\\
&=C_t\left(\mathbf{x}_t-\hat{\mathbf{x}}_{t|t-1}\right)+\mathbf{v}_t\\
&=C_t\boldsymbol{\varepsilon}_{t|t-1}+\mathbf{v}_t
\end{aligned} \tag{85}$$

进一步地，有：

$$\begin{aligned}
&E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\epsilon}_t^T\right]=E\left[\boldsymbol{\varepsilon}_{t|t-1}\left(C_t\boldsymbol{\varepsilon}_{t|t-1}+\mathbf{v}_t\right)^T\right]=E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right]C_t^T+E\left[\boldsymbol{\varepsilon}_{t|t-1}\mathbf{v}_t\right]\\
&=E\left[\boldsymbol{\varepsilon}_{t|t-1}\boldsymbol{\varepsilon}_{t|t-1}^T\right]C_t^T=P_{t|t-1}C_t^T
\end{aligned} \tag{86}$$

其中，$E[\boldsymbol{\varepsilon}_{t|t-1}\mathbf{v}_t^T]=0$。

把(86)和(66)代入(84)，整理得到

$$P_{t|t}=P_{t|t-1}-P_{t|t-1}C_t^TK_t^T-K_tC_tP_{t|t-1}+K_t(C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{vv}})K_t^T \tag{87}$$

由(81)式，可得：

$$\begin{aligned}
&P_{t|t}=P_{t|t-1}-P_{t|t-1}C_t^TK_t^T-K_tC_tP_{t|t-1}\\
&+P_{t|t-1}C_t^T\left[C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{vv}}\right]^{-1}\left(C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{vv}}\right)K_t^T\\
&=P_{t|t-1}-P_{t|t-1}C_t^TK_t^T-K_tC_tP_{t|t-1}+P_{t|t-1}C_t^TK_t^T\\
&=P_{t|t-1}-K_tC_tP_{t|t-1}=(I-K_tC_t)P_{t|t-1}
\end{aligned} \tag{88}$$

最后，若假设 $M_t\equiv 0$，则(78)式变为：

$$P_{t+1|t}=R_t\left[P_{t|t-1}-P_{t|t-1}C_t[C_tP_{t|t-1}C_t^T+\Gamma_{\mathbf{vv}}]^{-1}C_tP_{t|t-1}\right]R_t^T+Q_t \tag{89}$$

最终，Kalman 递归滤波关系式为：

第一步：时间更新（传播步）：
$$\begin{aligned}
\hat{\mathbf{x}}_{t+1|t}&=R_t\hat{\mathbf{x}}_{t|t}\\
P_{t+1|t}&=R_tP_{t|t}R_t^T+Q_t
\end{aligned}$$

第二步：Kalman 增益：$K_{t+1}=P_{t+1|t}C_{t+1}^{T}\left[C_{t+1}p_{t+1|t}C_{t+1}^{T}+\Gamma_{\mathbf{vv}}\right]^{-1}$

第三步：观测更新步：$\begin{aligned}\hat{\mathbf{x}}_{t+1|t+1}&=\hat{\mathbf{x}}_{t+1|t}+K_{t+1}\left(\mathbf{y}_{t+1}-C_{t+1}\hat{\mathbf{x}}_{t+1|t}\right)\\ p_{t+1|t+1}&=(I-K_{t+1}C_{t+1})P_{t+1|t}\end{aligned}$

这就是递归 Kalman 滤波的详细推导过程。

4 块矩阵的逆及其应用

Gauss 假设下的 Bayes 推断，在统计判决领域具有至关重要的地位。当前，线性 Gauss 平稳（缓平稳）假设下信号与图像领域中的绝大部分数据处理方法，都是基于 Gauss 假设下的 Bayes 推断导出的。线性（也包括弱非线性）系统的参数估计方法（典型的例子就是 FWI 和 LS_RTM）同样是在 Gauss 假设下的 Bayes 推断导出的。这些算法的导出需要引用块矩阵的逆。下面首先介绍块矩阵的逆。

4.1 块矩阵逆的基本公式

在随机向量满足联合高斯分布情形下，据此可进行的一些统计推断。这会用到块矩阵的逆。否则，有些统计推断公式是导不出来的。块矩阵逆的公式的引入是十分必要的。

如果矩阵 A 可分为 $A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}$，则：

$$A^{-1}=\begin{pmatrix}A_{11}^{-1}+A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1}\\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1}\end{pmatrix} \tag{90}$$

其中，$S=A_{22}-A_{21}A_{11}^{-1}A_{12}$。$S$ 称为矩阵 A 的 Schur Complement（Schur 补）。由 $AA^{-1}=I$，可证明(90)式中的 A^{-1} 表达式是正确的！

当 A_{22}^{-1} 存在时，$A_{11}^{-1}+A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1}$ 可以写为 $\left(A_{11}-A_{12}A_{22}^{-1}A_{21}\right)^{-1}$。对比(90)式，若令 $A=A_{11},X=A_{12},R=A_{22}^{-1},Y=A_{21}$，有：

$$\left(A_{11}-A_{12}A_{22}^{-1}A_{21}\right)^{-1}=A_{11}^{-1}-A_{11}^{-1}A_{12}\left(A_{22}+A_{21}A_{11}^{-1}A_{12}\right)^{-1}A_{21}A_{11}^{-1}=A_{11}^{-1}-A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} \tag{91}$$

特殊情形下，如 $A_{21}=0$，则 $A=\begin{pmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{pmatrix}$，有：

$$A^{-1}=\begin{pmatrix}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1}\\ 0 & A_{22}^{-1}\end{pmatrix} \tag{92}$$

另外，还有：

$$\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}^{-1}=\begin{bmatrix}A_{11}^{-1} & 0\\ 0 & 0\end{bmatrix}+\begin{bmatrix}-A_{11}^{-1}A_{12}\\ I\end{bmatrix}S^{-1}\begin{bmatrix}-A_{21}A_{11}^{-1} & I\end{bmatrix} \tag{93}$$

此处，$S=A_{22}-A_{21}A_{11}^{-1}A_{12}$。显然，此式应是(90)式的重新表达。

4.2 应用场景：Gauss 情形下的 Bayes 推断

4.2.1 Gauss 情形下的 Bayes 推断用到的基本公式

设 $\mathbf{X}$ 和 $\mathbf{Y}$ 是两个随机向量，令 $\mathbf{Z}=\left[\mathbf{X},\mathbf{Y}\right]$ 是由它们组成的一个随机向量。则 $\mathbf{Z}=\left[\mathbf{X},\mathbf{Y}\right]$ 的自协方差矩阵为：

$$R_{\mathbf{ZZ}} = Cov(\mathbf{Z}) = E[(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})^T] = \begin{pmatrix} R_{\mathbf{XX}} & R_{\mathbf{XY}} \\ R_{\mathbf{YX}} & R_{\mathbf{YY}} \end{pmatrix} \tag{94}$$

由前面逆矩阵的表达式，有：

$$R_{\mathbf{ZZ}}^{-1} = \begin{pmatrix} R_{\mathbf{XX}}^{-1} + R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}S^{-1}R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1} & -R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}S^{-1} \\ -S^{-1}R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1} & S^{-1} \end{pmatrix} \tag{95}$$

其中，$S = R_{\mathbf{YY}} - R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}$。或者有：

$$R_{\mathbf{ZZ}}^{-1} = \begin{pmatrix} R_{\mathbf{XX}}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}} \\ I \end{pmatrix} S^{-1}\left[-R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}, I\right] \tag{96}$$

下面给出 S^{-1} 和 $(R_{\mathbf{XX}}^{-1} + R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}S^{-1}R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1})^{-1}$ 的具体形式，即：

$$S^{-1} = (R_{\mathbf{YY}} - R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}})^{-1} = R_{\mathbf{YY}}^{-1} + R_{\mathbf{YY}}^{-1}R_{\mathbf{YX}}(R_{\mathbf{XX}} + R_{\mathbf{XY}}R_{\mathbf{YY}}^{-1}R_{\mathbf{YX}})^{-1}R_{\mathbf{XY}}R_{\mathbf{YY}}^{-1} \tag{97}$$

$$(R_{\mathbf{XX}}^{-1} + R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}S^{-1}R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1})^{-1} = R_{\mathbf{XX}} + R_{\mathbf{XY}}R_{\mathbf{YY}}^{-1}R_{\mathbf{YX}} \tag{98}$$

若取 $A = R_{YY}, C = R_{YX}, R = R_{XX}^{-1}, D = R_{XY}$，由(90)式，可直接写出(97)式。若取 $A = R_{XX}^{-1}, C = R_{XX}^{-1}R_{XY}$，，$R = S^{-1}, D = R_{YX}R_{XX}^{-1}$。由(90)式有：

$$\begin{aligned} &(R_{\mathbf{XX}}^{-1} + R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}S^{-1}R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1})^{-1} \\ &= R_{XX} - R_{XX}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}\left(R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{XX}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}} + S\right)^{-1}R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{XX} \\ &= R_{XX} - R_{\mathbf{XY}}\left(R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}} + S\right)^{-1}R_{\mathbf{YX}} \end{aligned} \tag{99}$$

其中，$S = R_{\mathbf{YY}} - R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}$，代入上式，得：

$$\begin{aligned} &(R_{\mathbf{XX}}^{-1} + R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}S^{-1}R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1})^{-1} \\ &= R_{XX} - R_{\mathbf{XY}}\left(R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}} + S\right)^{-1}R_{\mathbf{YX}} \\ &= R_{XX} - R_{\mathbf{XY}}\left(R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}} + R_{\mathbf{YY}} - R_{\mathbf{YX}}R_{\mathbf{XX}}^{-1}R_{\mathbf{XY}}\right)^{-1}R_{\mathbf{YX}} \\ &= R_{XX} - R_{\mathbf{XY}}R_{YY}^{-1}R_{\mathbf{YX}} \end{aligned} \tag{100}$$

至此，(98)式的成立，得到证明。

基于上述块逆矩阵的讨论，可以进行 Gauss 情形下，Bayes 推断本质思想的讨论及具体公式的导出。

4.2.2 Gauss 情形下 Bayes 推断基本理论

已知由随机向量 X 和 Y 形成的联合随机向量 $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$，则其期望为：

$$E(Z) = \boldsymbol{\mu}_Z = \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix} \tag{101}$$

协方差阵：

$$R_{zz} = Cov(Z) = E[(Z-\boldsymbol{\mu}_z)(Z-\boldsymbol{\mu}_z)^T] = \begin{pmatrix} R_{XX} & R_{XY} \\ R_{YX} & R_{YY} \end{pmatrix} \tag{102}$$

联合 Gauss 密度函数：

$$f(Z) = f(X,Y) = \frac{1}{(2\pi)^{p/2}\sqrt{R_{ZZ}}}e^{-\frac{1}{2}(Z-\boldsymbol{\mu}_z)R_{ZZ}^{-1}(Z-\boldsymbol{\mu}_z)^T} \tag{103}$$

其中， $p=m+n, X_{m\times 1}, Y_{n\times 1}$ 向量。

后验条件概率密度函数可以写出：

$$\begin{aligned} f(X\mid Y)&=\frac{f(X,Y)}{f(Y)}==C_1\frac{e^{-\frac{1}{2}\left[(X-\boldsymbol{\mu}_X)^T \quad (Y-\boldsymbol{\mu}_Y)^T\right]\begin{pmatrix}R_{XX} & R_{XY}\\ R_{YX} & R_{YY}\end{pmatrix}^{-1}\begin{bmatrix}X-\boldsymbol{\mu}_X\\ Y-\boldsymbol{\mu}_Y\end{bmatrix}}}{e^{-\frac{1}{2}(Y-\boldsymbol{\mu}_Y)^T R_{YY}^{-1}(Y-\boldsymbol{\mu}_Y)}}\\ &=C_1 e^{-\frac{1}{2}\left\{\left[(X-\boldsymbol{\mu}_X)^T \quad (Y-\boldsymbol{\mu}_Y)^T\right]\begin{pmatrix}R_{XX} & R_{XY}\\ R_{YX} & R_{YY}\end{pmatrix}^{-1}\begin{bmatrix}X-\boldsymbol{\mu}_X\\ Y-\boldsymbol{\mu}_Y\end{bmatrix}-(Y-\boldsymbol{\mu}_Y)^T R_{YY}^{-1}(Y-\boldsymbol{\mu}_Y)\right\}}\\ &=C_1 e^{-\frac{1}{2}E} \end{aligned} \tag{104}$$

$$E=\left[(X-\boldsymbol{\mu}_X)^T \quad (Y-\boldsymbol{\mu}_Y)^T\right]\begin{pmatrix}R_{XX} & R_{XY}\\ R_{YX} & R_{YY}\end{pmatrix}^{-1}\begin{bmatrix}X-\boldsymbol{\mu}_X\\ Y-\boldsymbol{\mu}_Y\end{bmatrix}-\left(Y-\boldsymbol{\mu}_Y\right)^T R_{YY}^{-1}\left(Y-\boldsymbol{\mu}_Y\right)$$

$$=\left[(X-\boldsymbol{\mu}_X)^T \quad (Y-\boldsymbol{\mu}_Y)^T\right]\begin{bmatrix}R_{XX}^{-1}+R_{XX}^{-1}R_{XY}S^{-1}R_{YX}R_{XX}^{-1} & -R_{XX}^{-1}R_{XY}S^{-1}\\ -S^{-1}R_{YX}R_{XX}^{-1} & S^{-1}\end{bmatrix}\begin{bmatrix}X-\boldsymbol{\mu}_X\\ Y-\boldsymbol{\mu}_Y\end{bmatrix} \tag{105}$$

$$-\left(Y-\boldsymbol{\mu}_Y\right)^T R_{YY}^{-1}\left(Y-\boldsymbol{\mu}_Y\right)$$

由于对随机向量 X 进行推断，整理上式，把关于随机向量 X 的项突出出来，

$$\begin{aligned} E&=(X-\boldsymbol{\mu}_X)^T\left[R_{XX}^{-1}+R_{XX}^{-1}R_{XY}S^{-1}R_{YX}R_{XX}^{-1}\right](X-\boldsymbol{\mu}_X)\\ &-2(X-\boldsymbol{\mu}_X)^T R_{XX}^{-1}R_{XY}S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)+Others \end{aligned} \tag{106}$$

上式可进一步写成：

$$\begin{aligned} E&=(X-\boldsymbol{\mu}_X)^T R_{X|Y}^{-1}(X-\boldsymbol{\mu}_X)-2(X-\boldsymbol{\mu}_X)^T\mathbf{b}+Others\\ &=(X-\boldsymbol{\mu}_X)^T A(X-\boldsymbol{\mu}_X)-2(X-\boldsymbol{\mu}_X)^T\mathbf{b}+Others \end{aligned} \tag{107}$$

其中， $A=R_{XX}^{-1}+R_{XX}^{-1}R_{XY}S^{-1}R_{YX}R_{XX}^{-1}, \mathbf{b}=R_{XX}^{-1}R_{XY}S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)$ 。

对(107)极值，令 $\frac{\partial E}{\partial X}=0$ ，有：

$$A\left(X-\boldsymbol{\mu}_X\right)=\mathbf{b} \tag{108}$$

因此，最后关于随机向量 X 最后的推断结果为：

$$X-\boldsymbol{\mu}_X=A^{-1}\mathbf{b}=\left[R_{XX}^{-1}+R_{XX}^{-1}R_{XY}S^{-1}R_{YX}R_{XX}^{-1}\right]^{-1}R_{XX}^{-1}R_{XY}S^{-1}\left(Y-\boldsymbol{\mu}_Y\right) \tag{109}$$

最后：

$$\begin{aligned} X&=\boldsymbol{\mu}_X+\left[R_{XX}^{-1}+R_{XX}^{-1}R_{XY}S^{-1}R_{YX}R_{XX}^{-1}\right]^{-1}R_{XX}^{-1}R_{XY}S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\ &=\boldsymbol{\mu}_X+\left(R_{XX}+R_{XY}R_{YY}^{-1}R_{YX}\right)R_{XX}^{-1}R_{XY}S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\ &=\boldsymbol{\mu}_X+R_{XY}R_{YY}^{-1}\left(Y-\boldsymbol{\mu}_Y\right) \end{aligned} \tag{110}$$

关于(110)式的证明。

$$
\begin{aligned}
X &= \boldsymbol{\mu}_X + \left(R_{XX} + R_{XY}R_{YY}^{-1}R_{YX}\right)R_{XX}^{-1}R_{XY}S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\
&= \boldsymbol{\mu}_X + \left(R_{XY} + R_{XY}R_{YY}^{-1}R_{YX}R_{XX}^{-1}R_{XY}\right)S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\
&= \boldsymbol{\mu}_X + R_{XY}\left(I + R_{YY}^{-1}R_{YX}R_{XX}^{-1}R_{XY}\right)S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\
&= \boldsymbol{\mu}_X + R_{XY}\left(R_{YY}^{-1}R_{YY} + R_{YY}^{-1}R_{YX}R_{XX}^{-1}R_{XY}\right)S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\
&= \boldsymbol{\mu}_X + R_{XY}R_{YY}^{-1}\left(R_{YY} + R_{YX}R_{XX}^{-1}R_{XY}\right)S^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\
&= \boldsymbol{\mu}_X + R_{XY}R_{YY}^{-1}SS^{-1}\left(Y-\boldsymbol{\mu}_Y\right)\\
&= \boldsymbol{\mu}_X + R_{XY}R_{YY}^{-1}\left(Y-\boldsymbol{\mu}_Y\right)
\end{aligned}
\tag{111}
$$

最后，**Gauss** 统计推断结果表示为均值和方差：

$$E\left[X\,|\,Y\right] = \boldsymbol{\mu}_X + R_{XY}R_{YY}^{-1}(Y-\boldsymbol{\mu}_Y) \tag{112}$$

$$Cov\left(X\,|\,Y\right) = A^{-1} = R_{XX} - R_{XY}R_{YY}^{-1}R_{YX} \tag{113}$$

由(107)式可知，后验概率密度函数的协方差阵取为 A^{-1} 。

Tarentola（2005）第三章给出了 L2 范数下参数估计公式的导出。但是，与此对比，上述推导过程更能体现出 Gauss 分布 Bayes 统计推断的本意。

5 线性 Gauss 假设下的 Bayes 统计推断

当前，信号分析（图像分析）领域或是地震波反演成像（弹性参数估计）领域中的很多问题，本质上，都是在进行线性 Gauss 假设下的 Bayes 统计推断。下面详细推断一下涉及到的理论公式。

设 $\mathbf{X}$ 是 Gauss 随机向量，且 $E[\mathbf{X}]=\boldsymbol{\mu}_{\mathbf{X}}, Cov(\mathbf{X})=R_{\mathbf{XX}}$ ， $\boldsymbol{\nu}$ 是零均值的白噪且 $Cov(\boldsymbol{\nu})=R_{\boldsymbol{\nu}\boldsymbol{\nu}}$ 。令：

$$\mathbf{Y} = H\mathbf{X} + \boldsymbol{\nu} \tag{114}$$

则 $\mathbf{Y}$ 的均值为 $\boldsymbol{\mu}_{\mathbf{Y}} = H\boldsymbol{\mu}_{\mathbf{X}}$ 和方差为：

$$R_{\mathbf{YY}} = Cov(Y) = HR_{\mathbf{XX}}H^T + R_{\boldsymbol{\nu}\boldsymbol{\nu}} \tag{115}$$

$\mathbf{Y}$ 和 $\mathbf{Y}$ 的互协方差为：

$$
\begin{aligned}
R_{\mathbf{XY}} &= E[(X-\boldsymbol{\mu}_{\mathbf{X}})(Y-\boldsymbol{\mu}_{\mathbf{Y}})^T] = E[(X-\boldsymbol{\mu}_{\mathbf{X}})(HX - H\boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\nu}\boldsymbol{\nu})^T]\\
&= E[(X-\boldsymbol{\mu}_{\mathbf{X}})(X-\boldsymbol{\mu}_{\mathbf{X}})H^T] = R_{\mathbf{XX}}H^T
\end{aligned}
\tag{116}
$$

推断：已知 $Y=\mathbf{y}$ ，Gauss 统计推断即求 $E[X\,|\,Y=\mathbf{y}]$

已知 $E[X\,|\,Y=\mathbf{y}] = \boldsymbol{\mu}_{\mathbf{X}} + R_{\mathbf{XY}}R_{\mathbf{YY}}^{-1}(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{Y}})$ ，则把(115)和(116)代入，可得

$$E[X\,|\,Y=\mathbf{y}] = \boldsymbol{\mu}_{\mathbf{X}} + R_{\mathbf{XX}}H^T(HR_{\mathbf{XX}}H^T + R_{\boldsymbol{\nu}\boldsymbol{\nu}})^{-1}(\mathbf{y}-H\boldsymbol{\mu}_{\mathbf{X}}) \tag{117}$$

此即为线性模型下 Gauss 统计推断的结果。这是最完整的理论表述。

$$Cov(X\,|\,Y=\mathbf{y}) = R_{\mathbf{XX}} - R_{\mathbf{XY}}R_{\mathbf{YY}}^{-1}R_{\mathbf{YX}} \tag{118}$$

具体地：

$$Cov(X\,|\,Y=\mathbf{y}) = R_{\mathbf{XX}} - R_{\mathbf{XX}}H^T(HR_{\mathbf{XX}}H^T + R_{\boldsymbol{\nu}\boldsymbol{\nu}})^{-1}HR_{\mathbf{XX}} \tag{119}$$

把依据观测结果 $\mathbf{Y}$ 对 $\mathbf{X}$ 的推断写成 $\hat{\mathbf{x}} = E[X\,|\,Y=\mathbf{y}]$ ，这就是更一般的 Bayes 统计推断。

6 提升对统计推断的基本认知

人类已经进入大数据时代，进入了通过数据挖掘产生新的认知的时代。但这并不是要抛弃掉通过实验、

观察、分析形成认知的过程，而是在此基础上的深化。原因是能被一般的实验模拟的、能被人脑的能力所理解的关系还是相对简单的。更加错综复杂的、隐藏在大量数据中的强非线线的关系，需要更强大的算法来挖掘和构建数据中内蕴的模式，进而构建更精确的（应该是更正确的）（映射）关系。架构在人工神经元构建的深度网络上的系列机器学习算法，与大数据+高算力结合，目前看来是形成人类新认知的一个合理的新模式。

这个新的认知模式，有一个基本的假设，就是样本数据是完备的（简而言之就是要大数据），其中包含了构建目标因果关系或目标相关关系的所有统计特征。样本数据的缺失、受污染等就会导致所构建的（映射）关系的精度下降，最终导致推断或决策的正确性降低，或者说人工智能的认知水平不够。

从这样的角度看，AI 的最深层次的问题，依然是统计判决问题。尽管人类的智能不可能是基于大数据的统计推断/判决，但是某种意义下，也是在不断地收集各种信息，与头脑中已有的经验交互，产生（期望产生）最佳的决策。

统计推断似乎始终是人工智能的最根本的逻辑基础，因为统计推断始终追求最佳的推断结果。

以下是笔者关于统计推断的一些基本认识：

1、科学思维的基本逻辑：复杂的实体总是由最基本的元素组合形成。这个基本逻辑似乎真是第一性原理，各个学科，概莫能外。

2、广义平稳随机过程是由白噪声 $\mathrm{WN}\left(0,\ \sigma^2\right)$ 线性组合生成，即由白噪声 $\mathrm{WN}\left(0,\ \sigma^2\right)$ 加权叠加生成。这就是 ARMA 模型生成广义平稳随机过程的基本逻辑。白噪声是最基本的、最简单的广义平稳随机过程！

3、广义平稳随机信号的样本函数（即广义平稳过程的一次具体实现）可由 Fourier 基函数线性叠加生成，这是有严格的数学证明的，Hilbert 空间中的函数逼近理论提供了坚实的基础。但是，Fourier 基函数的线性叠加可以生成广义平稳随机过程的说法是错误的、逻辑上是站不住脚的。遗憾的是，我们往往有这种错觉。严格地讲，Fourier 基函数的线性叠加，仅仅可生成样本函数，并不能生成广义平稳随机过程。但是，大量样本函数的 Fourier 变换系数（表现出随机性）的二阶统计量是广义平稳过程的功率谱。由 Wiener - Khinchin 定理知：功率谱与自相关函数是一对 Fourier 变换。这样，由样本函数（广义平稳随机过程的采样结果）的 Fourier 变换结果（它是随机的）的二阶统计量，即功率谱（其反变换为自相关函数），加上零均值假设（不是必须的！），就决定了一个广义平稳随机过程。这样的做法决定了一个广义平稳随机过程，与基于 ARMA 模型生成一个广义平稳随机过程还是很不一样的。不可以认为：广义平稳随机过程也可由谐波的叠加（加权叠加）生成。这些逻辑概念必须严格地界定清楚！否则会引起思维的混乱！

4、本质上讲，不是由谐波的加权叠加生成广义平稳随机过程，即不是由 Fourier 变换生成广义平稳随机过程，而是由大量样本函数的 Fourier 变换结果（它也是随机的！）的二阶统计量来估计广义平稳随机过程的二阶统计量。即：零均值+确定的功率谱决定了一个广义平稳随机过程。笔者认为：零均值+确定的功率谱也不是唯一地决定了一个广义平稳随机过程（相位没有考虑！），只是统计意义下确定了一个广义平稳随机过程。并不像是 ARMA 模型可以模拟生成一个广义平稳随机平稳过程的采样值。

5、自相关函数与功率谱是 Fourier 变换对。自然地，由样本自相关阵的正交分解（ $R_{\mathbf{xx}}=U\Sigma_{\mathbf{xx}}U^T$ ）可以得到平稳随机信号中所包含的特征成分，其中正交矩阵 U 中的基函数向量就是这些特征向量，这些基函数向量自然是 Fourier 基函数。

6、因此，生成广义平稳随机过程的理论方法就是 ARMA 建模，不要再提用 Fourier 基函数的线性叠加可以生成广义平稳信号的说法。严格地，Fourier 基函数的线性叠加仅仅可以生成样本函数或对样本函数进行逼近。

7、由样本自相关阵的正交分解，提取广义平稳随机信号中的特征量（谐波基函数，平面波基函数）也就顺理成章了。

8、数据驱动来提取广义平稳随机信号中包含的特征量的具体方法：PCA/ICA 方法、K-L 变换方法、小波变换方法及字典学习方法、张量分解方法等等。广义平稳随机信号情形下，各随机变量间存在线性相关关系（还应该是短相关关系），由上述方法得到的特征量，就是 Fourier 基函数（或平面波基函数）。

9、应该注意到：广义平稳随机信号是零均值且具有一定内在结构的方差函数决定的随机过程。

10、应该注意到：生成广义平稳随机过程的方法（即 ARMA 建模方法）与提取广义平稳随机过程特征量的方法是有明显不同的，尽管他们之间存在密切的联系。

11、所谓的数据挖掘，最核心的问题就是对（大）数据中隐含的内在特征量进行感知、表达并应用在构建（映射）关系上。

12、目前，广义平稳随机过程的特征提取和建模理论与算法都已经很成熟了。但是大量实际问题由于偏离了广义平稳随机过程的特征提取和建模理论的假设并没有很好地得到解决。但是，实际数据一定不是严格广义平稳的、更可能是非线性系统产生的，针对这样数据的特征提取和建模，尽管有一些理论进步，但远远没有达到完善的程度。

13、勘探地震中的 FWI 和 LS_RTM 是一个强计算、低（逻辑）推断的问题。理论上，它不是一个十分困难的统计推断、决策问题。储层描述与评价是一个强（逻辑）推断问题。AI 如何很好地解决储层描述与评价问题从上世纪 90 年代开始，就一直是业界关注的，但是，至今也没有很好的解决方案。样本数据不足（不充分）应该是最大的限制因素。

14、架构在人工神经元构建的深度网络上的机器学习算法，是否就是非线性系统产生的、非平稳的、非高斯数据的特征提取、映射关系构建的最合理的方法，目前还没有定论。只是它有仿生人脑的优势，看起来在大数据时代暂时占据了优势。从算法的角度看，把线性问题的迭代解法到非线性问题的迭代解法抽象地视为线性网络下的算法，再进一步地提升到非线性网络下的 ML/AI 算法是很自然的事情，但这依然不能让人相信基于数据样本和人工神经网络架构的 ML/AI 算法就是非线性系统产生的、非平稳的、非高斯数据的特征提取、映射关系构建的终极算法，除非当前的 ML/AI 算法真正进入无监督学习，而且在任何数据条件下都能稳定地产生合理的结果。也许这是个终极的、理想目标。如果到了这种程度，人类的学习还有意义吗？人类的存在还有意义吗？

7 结论

基于大数据，通过数据挖掘产生新的认知，是当今时代科学技术发展的核心思想与方法。众所周知，任何新的思想方法都有其发展而来的逻辑路线。

大数据+强算法+高算力的 AI 时代的到来，本质上，应该是脱胎于统计推断。最基本的统计推断应该是对广义平稳随机过程的建模、预测与推断。

推断的基础当然是要有数据。数据被认为是（广义的）随机过程。数据一定是物理（或非物理）系统的输出。系统是线性的或非线性的。严格的线性系统仅仅是数学假设。实际的物理或非物理系统一定是或强或弱的非线性系统。数据是研究对象（也可以说是系统本身）的外在表现，是可观测与可记录的量，它们蕴含了关于研究对象（系统本身）的信息。一般地，认为数据对应的随机过程是宽（广义）平稳的、零均值的，存在有对应协方差（阵）的（即数据点之间存在某种相关关系的）。

这样的随机过程是由白噪声 $\mathrm{WN}\left(0,\ \sigma^2\right)$ 线性组合生成的，它是可以由 ARMA 模型/MA 模型/AR 模型很好预测的。也可以说用统计推断方法能很好地预测和描述宽（广义）平稳随机过程的变化过程。

延伸到缓平稳情形下，修正的 ARMA 模型也可以较好地预测缓平稳随机过程的变化过程。这就产生了各种自适应滤波（更合理地应称为自适应预测）理论与方法。

众所周知，ARMA 模型是线性的，继续推进到广义线性或（弱）非线性系统情形下，也发展了相应的自适应滤波理论与方法，譬如扩展的 Kalman 滤波方法及更进一步的粒子滤波方法。

上述逻辑继续向前延伸，在借鉴人脑神经元构建出人工神经网络的基础上，发展出了对强非线性系统输出数据进行建模分析的各种算法，这就进入了今天的AI时代，大数据+强算法+高算力产生了能对（较）强非线性系统生成的数据进行深度挖掘，并形成自主推断的能力。在自然语言处理、图像处理和语音处理领域，这样的做法取得了巨大的成功。已经隐隐出现了颠覆目前人类社会生产力和生产关系的趋势。

做任何事情都要追根求源，只有知道既往，才知道如何开来。本文从典型矩阵的逆的角度，讨论了 Bayes 参数估计公式是如何导出的，在什么样的理论假设下导出的；给出了递归最小二乘方法和 Kalman 滤波方法的详细的公式推导过程。事实上，当前勘探地震数据分析的根本基础就是高斯线性平稳假设下的统计推断理论。从典型矩阵的逆的角度，本文把其中的一些重要公式进行了详细推导，据此应该可以对统计

推断的理论、算法及实现有更深入的理解。

参考文献

[1] Moon T I, Stirling W C. Mathematical Methods and Algorithms for Signal Processing [M]. PRENTICE HALL, 2000.
[2] Brockwell P J, Davis R A. Time Series: Theory and Methods [M]. 2nd ed. Springer, 2006.
[3] Theodoridis S. Machine Learning: A Bayesian and Optimization Perspective [M]. Elsevier Ltd., 2015.
[4] Haykin S. Neural Networks and Learning Machines [M]. 3rd ed. Pearson Education, Inc., 2009.
[5] Murphy K P. Machine Learning: A Probabilistic Perspective [M]. The MIT Press, 2012.
[6] Goodfellow I, Bengio Y, Courville A. Deep Learning [M]. MIT Press, 2016.
[7] Sigaud O, Buffet O. Markov Decision Processes in Artificial Intelligence: MDPs, Beyond MDPs and Applications [M]. WILEY, 2010.
[8] Casella G, Berger R L. Statistical Inference [M]. 2nd ed. CENGAGE Learning, 2002.

勘探地震中非线性问题及求解方法分析

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】无论是地震数据分析，还是广义遥感领域的数据分析，包括大数据人工智能/机器学习数据分析，都逐渐地从对线性系统输出数据的分析转到对非线性系统输出数据的分析，其根本原因是非线性系统是物理和非物理世界的更为真实的描述，线性系统不过是非线性系统在特定情况下的近似。非线性系统是物理和非物理世界更为真实的抽象，精准地把握非线性系统随时间的演化规律，并预测其最终演化结果，进而对它进行某种预测或控制，是研究非线性系统的根本目的。线性系统不存在复杂的演化规律，对线性系统的研究重点就集中在不同时刻的演化关系上，不关注其长期的演化过程及演化趋势。非线性问题和线性问题的研究是另一个相关的概念。数学分析的核心研究对象就是高维线性问题（及线性方程组）、多元函数及高维算子（及泛函）。数学分析提供了研究非线性问题和线性问题的思想、方法与工具。非线性问题研究中核心的思想与方法也是由数学分析方法提供的。线性系统相关的问题已经有十分优美、完整的数学理论框架，但是，非线性系统相关问题的研究，由于复杂性和多样性，目前还没有统一的思想与方法。尤其对复杂的非线性系统演化过程及演化结果的建模、预测及描述，目前还处在很初始的阶段。勘探地震系统是一个拟线性系统，但地震数据分析中存在很多重要的非线性问题，譬如乘性噪音问题、强噪声弱信号情形下的去噪问题、复杂情形（即实测波场与地下介质弹性参数之间存在强非线性关系情形）下地震波成像问题、由弹性参数估计岩石性质参数的问题，都是强非线性问题。这些都需要提出新的求解方法。推广到地球科学系统、大气系统、金融系统等，这些强非线性系统的状态演化即演化结果的预测与描述，都具有重要意义。科学问题的研究对象逐渐地转移到非线性问题和非线性系统上，是必然的。此文的目的就是希望 WPI 的研究逐渐由对线性问题和线性系统的研究转移到非线性问题和非线性系统上。

【关键词】线性系统，非线性系统，系统（状态）演化过程，系统演化状态的建模及预测理论与方法，系统演化结果的描述，线性问题与非线性问题，线性算子/泛函及非线性算子/泛函

1 引言

从线性系统的研究转到非线性系统的研究，从线性问题的研究转到非线性问题的研究，随着线性系统（线性问题）研究的逐渐完善，已经不可避免。

对非线性系统的研究，有两个显著的问题：一是对非线性系统的演化过程进行预测；二是研究非线性系统最终演化结果所处的状态（即各种模态的吸引子）。一般地，我们主要关注的是一个非线性系统当前的状态如何变化到未来时刻的其他状态。因此，在相空间中（即状态空间）进行非线性系统状态演化的预测是非线性系统研究最核心的问题。$\boldsymbol{x}$ 代表状态向量，它本身是有结构的，它是一个几何形态（流形），它的结构也是随时间演变的。并且，状态本身又是随时间演化的，这进一步增加了非线性系统的研究难度。将非线性系统演化过程抽象为状态向量的演化过程是非线性系统演化的一个重要概念模型。通常，我们可以用一阶微分方程来描述非线性系统演化，譬如 $\partial \boldsymbol{x} / \partial t = F(\boldsymbol{x}), \boldsymbol{x}(t=0) = \boldsymbol{x}_0$。显然，这是一个单步预测问题。若右端项用状态向量随时间变化的高级微分表示，这涉及到多步预测问题。但是状态向量随时间变化的高级微分的物理含义不很明确，需要在后续的研究中进一步探索。这就是我们对非线性系统演化的数学模型的描述。

目前，用 Markov 链（或隐式 Markov 链）表达非线性系统的演化似乎是比较常用的数学工具。典型的算法是粒子滤波方法。在假定已知演化关系 $\boldsymbol{x}^{k+1} = C(\boldsymbol{x}^k) + \upsilon_k$ 的情况下，用扩展的 Kalman 滤波方法描述非线性系统的演化。非线性系统演化的预测方法目前还在深入研究发展中。事实上，无论哪种方法，都只是非线性系统演化过程的一种近似预测方法。只要预测精度足够好，就是可行的方法。

另外，特别指出，很多非线性问题仅仅是数学分析中的非线性函数或非线性泛函问题，不能与非线性系统的演化问题混在一起。生态系统、经济系统、大气系统等是典型的非线性系统，事实上，物理与非物理领域遇到的众多的非线性系统都是非线性系统，其系统是随时间不断演化的，只是它们的演变程度不是特别剧烈，使得我们往往并没有从非线性系统的角度看待它们，不关心它们演变的最终状态，或者只是把它们近似为线性系统。

目前，线性系统的研究已经非常深入与成熟了，任何科学技术领域和社会科学领域的研究，从线性系统的研究走向非线性系统的研究都是十分必要的。

2 线性系统与非线性系统

2.1 线性系统的定义及性质

线性系统是指满足叠加原理和齐次性质的系统。对于一个系统，若输入信号为$x(t)$，输出为$y(t)$；当输入信号为$x_1(t)+x_2(t)$时，输出为$y_1(t)+y_2(t)$时，系统满足叠加原理；当输入信号为$\alpha x(t)$时，输出为$\alpha y(t)$时，系统具有齐次性性质；当输入信号为$\alpha x_1(t)+\beta x_2(t)$时，输出为$\alpha y_1(t)+\beta y_2(t)$，意味着系统满足线性叠加原理，并具有齐次性质。这样的系统就是线性系统。

对于一个系统，期望其具有短的或长的记忆性、因果性、可逆性、时不变性质，尤其具有线性性质。

线性系统具有如下好的性质。**稳定性**：线性系统在受到扰动后能够恢复到平衡状态；**因果性**：系统的输出仅依赖于当前和过去的输入，与未来的输入无关；**线性相位**：系统的频率响应具有固定的相位差，不会随频率变化而变化，即线性系统不改变输入信号的相位关系，输出信号与输入信号之间是相干信号$y(t)=cx(t)$，其中c是复数。很显然，把一个探区的三维地下介质视为一个系统，它显然是近似线性的。因此，勘探地震领域的信号分析和参数估计显得相对简单。

现代信号分析的研究对象，主要是线性时不变系统，包括其输入、系统本身和系统输出。广义平稳信号应该是线性时不变系统的输出。高斯白噪声冲击线性时不变系统产生广义平稳信号，其输出可以通过 ARMA 模型进行有效的模拟及预测。当系统是最小相位时，此线性系统是可逆的和因果的。这就是应用最佳 Wiener 滤波的理论假设场景。然而，实际情形下，常常偏离上述理论假设。时变系统产生非平稳的输出，即高斯白噪声冲击线性时变系统产生广义非平稳信号。此时，需要进行自适应滤波，实质上就是建立一个非平稳情形下的预测器。LMS（Least-Mean-Square）方法和迭代最小二乘方法是缓平稳情形下的自适应滤波方法。

线性系统也是不断演化的，只是演化关系很简单而已。Kalman 滤波理论框架中，最关键的是提出了状态演化方程$\boldsymbol{x}^{k+1}=C\cdot\boldsymbol{x}^{k}+\upsilon_k$和输出信号的观测方程$\boldsymbol{y}^{k+1}=R\cdot\boldsymbol{x}^{k+1}+\nu_{k+1}$，并在这样的概念下，导出了线性系统的状态演化关系。目标也还是对缓平稳情形下的线性系统（演化）进行建模预测。

另外，静态地，我们常常用线性方程组$\boldsymbol{y}=A\boldsymbol{x}$来表达一个线性系统。此时，$\boldsymbol{y}$是线性系统的输出；$\boldsymbol{x}$是线性系统的激励；$A$是系统响应函数。很多时候，我们研究线性问题及对应的反问题时，往往不关注上述线性方程组与线性系统的关系，仅仅把它理解为线性方程组的求解问题。针对大规模矩阵方程的求解有大量的书籍和文献讨论这个问题。这应该是非常多的领域归纳出的一个共同的问题。勘探地震数据分析中的绝大部分问题最后都归结为线性方程组的求解。但是，我们并没有把这个问题与线性系统联系起来。

无论如何，到目前为止，在高斯、线性、平稳假设下的线性系统的演化关系的研究（包括静态情形下的线性方程组求解问题的现象）已经非常完善，必须要考虑在非高斯、非线性、非平稳假设下开展非线性系统的研究。

2.2 非线性系统的定义及性质

非线性系统是指输出与输入之间不满足叠加原理和齐次性的系统。从数学角度来看，非线性系统的特征是描述系统的方程不能写成未知变量的线性组合，即不再能用一个线性方程组来表示，而是方程中存在变量的高次项、三角函数项、指数函数项等非一次项，或者变量之间存在乘积、复合等非线性关系。例如，一个系统的输入输出关系为$y(t)=x^2(t)$，其中是$x(t)$输入，$y(t)$是输出，当输入$x_1(t)$和$x_2(t)$时，输出分别为$y_1(t)=x_1^2(t)$和$y_2(t)=x_2^2(t)$，而对于输入$x_1(t)+x_2(t)$，输出$y(t)=\left(x_1(t)+x_2(t)\right)^2\neq y_1(t)+y_2(t)$，不满足叠加原理，所以该系统是非线性系统。

非线性系统的性质包括：叠加原理失效：对于非线性系统，两个输入信号单独作用于系统时产生的输出之和，不再等于这两个输入信号同时作用于系统时产生的总输出。**多平衡点**：非线性系统可能存在多个平衡点。例如，在一些机械系统或电路系统中，由于非线性因素的存在，系统可能在不同的位置或状态下达到平衡，且在不同初始条件下，系统会趋向于不同的平衡点，表现出不同的稳态行为。**分岔现象**：当系

统参数发生变化时，平衡点的数量和性质可能会发生突变，这种现象称为分岔。例如，在某些生态系统模型中，随着环境参数（如资源量、捕食者与被捕食者的比例等）的变化，系统可能从一个稳定的平衡状态突然转变为多个平衡状态或出现周期性振荡等不同的行为模式。**混沌行为**：在某些条件下，非线性系统可能出现混沌现象，表现为对初始条件的高度敏感性，即初始条件的微小变化可能会导致系统未来状态的巨大差异，使系统的行为看起来具有随机性和不可预测性，如著名的洛伦兹吸引子就是大气对流模型中的一种混沌现象。**非线性振荡**：非线性系统可能产生复杂的振荡现象，如极限环振荡。即使没有外部输入作用，系统也可能会产生具有一定振幅和频率的持续振荡，例如一些电子振荡电路中的自激振荡现象。**频率响应复杂**：当输入为正弦函数时，非线性系统的输出通常不是同频率的正弦函数，而是包含高次谐波的非正弦周期函数，会产生倍频、分频等现象。例如，在非线性光学系统中，当用特定频率的激光照射非线性光学材料时，会产生频率为入射光频率整数倍的谐波光。可控震源与地表介质视为一个系统的话，它显然是非线性的，使得注入地下的谐波受到畸变，产生低次谐波和高次谐波。**与初始条件和输入大小相关**：非线性系统的稳定性和输出动态过程不仅取决于系统本身的结构和参数，还与系统的初始条件以及输入信号的大小密切相关。例如，一个非线性电路系统在不同的初始电压或电流条件下，可能会进入不同的工作状态，如稳定状态、振荡状态或混沌状态。

很显然，非线性系统的研究是极为复杂和困难的事情。对于强非线性系统，其最终演化结果是很难预测和描述的；中间演化过程也很难建立一个十分有效的模型进行模拟和预测。前已述及，在物理和非物理世界中，所有的系统都具有或强或弱的非线性性，纯粹的线性系统是理论假设的结果。生态系统、经济系统、大气系统等等这些典型的非线性系统目前并没有提炼出统一且有效的数学模型，能对它们的演化过程和演化结果进行较好的预测和描述。股票系统这个典型的强非线性系统尤其缺乏有效的模型进行短期涨跌的预测。巨量的脑神经细胞组成的大脑系统也是非线性系统，人类的创新想法和怪异想法都是这个系统的输出产生的，很难预测或不能预测。类比地，今后巨量的人工神经元组合在一起形成的非线性系统产生超越人脑的创新或邪恶想法，绝非不可能的事情。这是 AI 带给人类社会的巨大不确定性。

针对非线性系统的研究，笔者的观点是：从弱非线性缓平稳过程的研究到中等强度的非线性及较强的非平稳过程的研究再到强非线性强变化过程的研究；从单步演化建模预测到多步演化建模预测再到长时间演化的建模预测；从演化过程建模预测到演化结果的预测和描述。目前，重点还是研究弱非线性缓平稳过程或中等强度非线性且较强非平稳过程。

勘探地震中，波场与介质扰动的关系就代表了一个非线性系统，而且是一个不时变的非线性系统。相对而言，这是个比较简单的非线性系统。介质参数扰动越强，介质参数扰动的尺度越小，波场与介质扰动之间关系的非线性性就越强。波场与地下介质岩性参数之间的关系是更强的非线性关系，因此从参数估计的理论看，不能从已知波场直接推断地下介质的岩性参数，先由观测数据（波场）推断出地下介质弹性参数，然后依据弹性参数和岩性参数之间的关系，进一步推断地下介质岩性参数，似乎是更好的逻辑选择。遗憾的是，岩石介质的弹性参数与岩性参数之间也存在强非线性关系，尽管这种关系主要是静态的。即便这种静态关系，到目前为止，也没有十分确定的、能有效地解决问题的数学模型，仅仅存在一些简单的经验公式，用以描述这一关系。

到目前为止，数学上，远没有发展出能有效地预测非线性系统演化规律的模型，更没有能很好地预测长期演化结果的模型。扩展的 Kalman 滤波方法和粒子滤波方法仅仅是弱非线性情形下单步预测模型；神经网络主要是建立起输入与输出之间的非线性关系，似乎能构建出输出与输入之间的强非线性关系式，这是一种静态关系的构建。笔者认为，非线性系统问题的研究还处于十分初始的阶段，任重道远。

3 线性问题与非线性问题

物理和非物理问题的研究中，我们更多讨论的是线性问题或非线性问题。矩阵方程是线性问题的典型，它实际上是 n 个超平面定义的几何关系。n 维非线性映射（包括非线性泛函）是非线性问题的典型，它实际上是 n 个超曲面定义的几何问题。在研究各种应用场景归结出来的典型线性和非线性问题时，很少把它们与物理或非物理系统联系起来，仅仅专注于用各种数学分析方法求解这两类典型线性和非线性问题。这是目前我们拥有的强大的数学工具。但是，这样做的一个重大缺陷是容易忘记原问题的物理背景。

在解决线性或非线性问题时一定要认识到，这些问题是从所研究的物理或非物理系统中提炼出来的，对于任何一个系统，对其演化过程和演化结果的把握是最重要的两个目标。当线性系统和非线性系统是静态问题时，可以用各种数学分析方法求解典型线性和非线性问题。由于线性系统的演化过程和演化结果的

把握相对地比较简单，对超大规模线性系统 $\boldsymbol{y}=\boldsymbol{Ax}$ 的求解变成核心问题是可以理解的。但是，对于非线性系统的研究而言，必须要关注其动态演化过程的建模、表达和预测，更要关注其中间演化结果和最终演化结果的各种模态及它们的表达形式。

前已述及，各种物理或非物理系统均有或强或弱的非线性性，纯线性系统是特例（也许仅仅是数学假设意义下的存在）。目前，对线性系统进行建模预测已有规范性的解法，包括对线性系统进行特征分析也有完整的，优美的解法。线性系统可以用 $y=Ax$ 或 $b=Ax$，或 $d=Gm$ 这样的矩阵方程来表示，解适定/超定/欠定的线性系统、大规模或中小规模线性系统，都存在对应的理论较完善的方法。从科学研究的角度来看，线性系统相关问题已缺乏可探索的空间，而对非线性系统的研究必会越来越受重视。

即便在静态情形下，非线性系统也不存在矩阵方程那样一个统一的表达式，即不存在一个较为统一的模型来表示或表达非线性系统。首先，偏（常）微分方程一般是非线性系统的表达形式，即输出与系统响应和输入的非线性关系，但很不方便应用，输出与系统响应和输入之间的非线性关系没被表达为明瞭的形式。另外，高维非线性方程组的求解，目前并没有一套完整的数学理论。把它变成变分问题进行求解，即转化为误差泛函求极小的问题，可能存在严重的多解性。勘探地震中 FWI 就是一个典型的例子。

勘探地震中，波场与地下介质参数变化之间的关系是非线性的，地下介质参数变化越强，尺度越小，二者之间关系的非线性性越强。实质上，此时我们假设地下介质参数是静态的，不随时间变化的。FWI 的目标就是由实测的波场去估计地下介质的参数。这是个非线性问题研究的范例。最根本的理论假设是波场 $u(x,y,z;t)$ 与地下介质参数 $\boldsymbol{m}(x,y,z)$ 之间的关系是已知的。即系统输出与输入之间存在的关系 $u(x,y,z;t)=G(\boldsymbol{m}(x,y,z))$ 是已知的。但是，根据勘探地震的实践，勘探地震系统并没有表现出前述非线性系统的典型特征，譬如混沌、叠加原理严重失效、出现不同频率之间的非线性关系等等。如果地下介质参数出现较强的随时间变化的特征，勘探地震预计会变得异常困难。

股票波动、大气系统波动、生态系统波动之所以出现显著的上述非线性特征，系统本身内在参数随时间发生显著变化，输出与输入之间存在强非线性关系，应该是导致非线性系统变化难以建模预测的根源。

据此分析，勘探地震中主要解决的还是静态情形下的非线性问题，即系统参数不随时间变化时的参数估计问题。首先要解决的还是要先构建合理的静态预测模型 $u(x,y,z;t)=G(\boldsymbol{m}(x,y,z))$，然后通过所谓的非线性反演来估计静态的地下介质参数（变化）。

积分形式的描述多次波的方程应该是期望的静态预测模型，仅仅用微分形式的波动方程，很难分析清楚问题。已知，在波动方程领域提出了 Born 级数；在信号分析领域提出了 Volterra 级数，这是（显式地）表达（静态）非线系统最常用的两种模式。

在 Born 级数中，描述了各阶多次散射波。

$$\begin{aligned}&u_{total}(x,y,z;t)=u_0(x,y,z;t)+G(x,y,z;t)\delta\boldsymbol{m}(x,y,z)G(x,y,z;t)\\&+G(x,y,z;t)\delta\boldsymbol{m}_1(x,y,z)G(x,y,z;t)\delta\boldsymbol{m}_2(x,y,z)G(x,y,z;t)\\&+G(x,y,z;t)\delta\boldsymbol{m}_1(x,y,z)G(x,y,z;t)\delta\boldsymbol{m}_2(x,y,z)G(x,y,z;t)\delta\boldsymbol{m}_3(x,y,z)G(x,y,z;t)\\&+\cdots\end{aligned}\tag{1}$$

其中，右端第一项对应透射波项，不经介质扰动作用直达检波器的场；右端第二项对应一次散射波场，常规的偏移成像利用这部分波场；右端第三项对应二次散射波场；右端第三项对应三次散射波场，以此类推。从 Born 级数可以看出，各散射点之间，而且是可能很远的散射点之间产生了联系，这对逆散射反演估计散射强度是极为不利的。但这不是线性系统所具备的特点。这种非局部特性导致逆散射问题的求解极为困难，所以很少人敢碰强非线性问题。

Volterra 级数是信号领域（或时间序列分析领域）线性褶积系统的升级，其基本形式为

$$\begin{aligned}&y(t)=x_0(t)+\int h(\tau)x(t-\tau)d\tau+\iint h(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1d\tau_2\\&+\iiint\iint h(\tau_1,\tau_2,\tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1d\tau_2d\tau_3+\cdots\end{aligned}\tag{2}$$

其中，右端第一项代表常系统；第二项代表一次滑动平均线性子系统（可认为是一次线性相关）；第三项代表二次相关滑动平均线性系统；第三项代表三次相关滑动平均非线性系统，…，依次类推下去。

应该说，用高阶统计量也可描述非线性系统的特性。高阶统计量的定义形式（高阶矩的定义）为，

$$\begin{aligned} \boldsymbol{m}_1 &= \int_{-\infty}^{+\infty} p(\boldsymbol{x})\boldsymbol{x}d\boldsymbol{x} \\ \boldsymbol{m}_2 &= \int_{-\infty}^{+\infty} p(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{m}_1)^2 d\boldsymbol{x} \\ \boldsymbol{m}_3 &= \int_{-\infty}^{+\infty} p(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{m}_1)^3 d\boldsymbol{x} \\ &\cdots \end{aligned} \tag{3}$$

注意，$\boldsymbol{x}$ 是随机向量，上述积分均是多重的。应该注意到：高阶统计量与上述两种模型是存在显著差异的。它仅仅从数据驱动的角度，描述实测的随机向量（它是非线性系统的输出）的数字统计特征。它不是着眼于非线性系统本身，而是着眼于非线性系统的输出。但是，对比高阶矩和 Volterra 级数中高阶项的定义，可知高阶项代表系统的非线性部分。

对于波动理论下的参数估计问题，基于全波动方程正演的梯度导引类的 FWI 并非真正的非线性反演。尽管弹性（声）波方程参数估计问题本身是非线性的，但梯度导引类的 FWI 算法并非针对性的求解非线性反问题的算法。它只能说是一种凸问题的合理解法。理论上，非线性系统参数估计的合理算法应是全局寻优方法。这就是说，目前不存在真正的解决（哪怕静态）非线性系统对应的参数估计问题的针对性方法。

T-Matrix 方法试图解决逆散射问题，但是也不是求解 Born 逆散射问题的合理算法。仅仅是一种尝试而已！笔者还是建议用特征反射层的概念，结合 Green 函数的合理算法（即地震波正演算法），把层间多次波层析成像和高波数参数扰动估计的基本逻辑弄明白。T-Matrix 方法，过于注重矩阵运算，还是没有揭示出背后的波传播物理逻辑。这是地震波反演成像目前没有很好解决的问题。

这进一步说明，在强多次散（反）射情形下，梯度导引类 FWI 方法不是十分针对性的参数估计方法。逐步线性化（每一步解决一个（拟）凸问题）的做法看起来是有效的，但是对于强多次散（反）射情形下的参数估计，未必是有效的方法。此外，全局寻优算法很难保证收敛性，而且计算量巨大。据此可见强多次散（反）射情形下的高精度参数估计还任重道远，需要有新的理论方法。

线性结构飘在具有一定概率分布的随机噪声中是系统输出的概念性表达。线性结构可以认为是系统状态的描述。更一般地，是非线性结构，而且可能是十分复杂的非线性结构飘在具有一定概率分布的随机噪声中。这是高维信号与图像分析及一般多维数据分析的更一般性的概念模型。当前，相当大一部分高维信号和图像分析方法是从含噪的高维数据中检测出其中包含的线性结构或非线性结构。这个问题与非线性系统参数估计问题是十分不同的。重要的差异是参数估计问题假定已知波场与介质参数之间的因果关系；而从高维含噪数据中估计出线性或非线性结构主要依赖高维含噪数据中包含的相关关系。

高维 Wiener 滤波就是假设高维含噪数据中蕴含线性结构；而 Volterra 级数描述的显然是高维实测数据中蕴含非线性结构。已知高维 Wiener 滤波所形成的矩阵方程往往规模庞大，基于 Volterra 级数所形成的矩阵方程规模会更大，而且该矩阵方程的具体含义，尤其与高阶统计量之间的联系，应该弄清楚。尽管有 Volterra 级数试图描述非线性系统输出的内在特征，但是该模型是否真的能刻画非线性系统的本质，并没有定论。

张量分解方法是一种用于表达高维数据中蕴含结构的有效手段，它能够较好的把其中的线性结构预测出来。但能否有效提取高维数据中蕴含的非线性结构，尤其是其中复杂的非线性结构，概念上还难以说清楚。对于提取高维数据中的非线性结构，Volterra 级数方法概念上明确很多，因为它用了数学分析中的非线性泛函表达式。但是，计算依然存在问题。因此，把 Volterra 级数方法与张量分解方法进行对比研究看起来是必要的。

此外，需要注意的是，许多非线性问题仅仅是数学分析中的非线性函数或非线性泛函问题，不能与非线性系统（的演化）混在一起。在高维含噪数据中，非线性结构的建模、预测、插值及去噪等问题，通常可以归结为数学分析中的非线性泛函问题。但是，非线性系统的状态 $\boldsymbol{x}$，可以认为是高维空间中的结构，高维空间可以认为就是状态空间，因此，状态空间中的结构的演化过程，应该就代表了非线性系统的演化过程。所以，数学分析中的非线性函数或非线性泛函问题的研究对于描述非线性系统的演化，不是没有关

系的，而是有密切的关系。相信很多非线性系统演化过程的建模方法的构建还是要利用数学分析的思想与方法。

4 加性噪声与乘性噪声

勘探地震中，乘性噪声问题是一个典型的非线性问题，尤其是强乘性噪声非线性地影响了有效信号。当前，地震数据预处理中，绝大部分的去噪方法都是假设噪声与有效信号是线性叠加的。但是，除了加性噪声，乘性噪声的存在也严重地影响地震数据分析结果的质量，尤其在反演成像逐渐成为主流成像方法的情形下。勘探地震中，乘性噪声主要是震源和检波器以及它们与介质的耦合导致的。

4.1 乘性噪声的一般概念

定义：乘性噪声是指噪声与信号相乘而引入到信号中的噪声。设原始信号为$s(t)$，乘性噪声为$N(t)$，则受到乘性噪声污染的信号可表示为$s_N(t)=s(t)\cdot N(t)$。

与加性噪声$s_N(t)=s(t)+N(t)$不同，乘性噪声的影响与信号本身的强度有关，信号越强，乘性噪声的影响可能越大。乘性噪声是一种噪声水平与信号幅度成正比的噪声类型。

以下是一些典型例子：**通信系统中的衰落信道：**在无线通信中，信号从发射端到接收端常因多径传播而经历衰落。这导致接收到的信号与一个变化的因子相乘，从而产生乘性噪声。**图像处理中的散斑噪声：**常见于超声、雷达和合成孔径雷达（SAR）等相干成像系统中。散斑噪声是乘性的，因为其强度与图像的局部强度成比例变化。**摄影中的胶片颗粒噪声：**在传统摄影中，胶片颗粒噪声可被视为乘性噪声，因为它随图像区域的亮度而变化。**金融数据中的股票价格：**金融时间序列，如股票价格或汇率，常呈现乘性噪声。波动性（噪声）通常与价格序列的水平成正比。**期权定价模型：**像布莱克 - 斯科尔斯模型这样的金融模型假设资产价格遵循具有乘性噪声的随机过程。**生物医学信号处理：**在心电图（ECG）或脑电图（EEG）信号中，噪声水平可能与信号幅度成正比。例如，ECG 信号中的肌肉伪影可能呈现乘性特征。**医学成像：**在 X 射线或磁共振成像（MRI）等成像方式中，噪声可能是乘性的，特别是在低剂量成像中，此时信噪比很低。**环境与自然现象：**由于环境条件的变化，河流流量和降水数据常呈现乘性噪声特征。与天气相关的数据，如降雨量测量，常显示出乘性噪声，因为测量误差与测量值的大小成正比。**恒星光度测量：**由于大气湍流和仪器效应，恒星光线强度的测量常涉及乘性噪声，噪声水平与观测到的光强度成正比。

抑制乘性噪声的基本思路涉及多种旨在在保留原始信号完整性的同时降低噪声的技术和方法。以下是一些基本策略：**变换域处理：小波变换**将信号变换到小波域有助于将噪声与信号分离。然后可以对小波系数进行处理，抑制噪声并保留重要的信号特征。**傅里叶变换：**在频域中，通常可以根据噪声和信号成分的频谱特征将它们区分开来，从而实现有针对性的降噪。**对数变换：**通过应用对数变换，乘性噪声可以转换为加性噪声。这种变换简化了噪声抑制过程，因为加性噪声更容易用传统技术处理。在对数域处理噪声后，应用指数变换将信号恢复到原始域。**滤波技术：中值滤波：**在保留图像边缘的同时有效降低噪声。中值滤波器可应用于对数域来处理乘性噪声。**维纳滤波：**一种自适应滤波技术，可使估计信号与原始信号之间的均方误差最小化。**统计模型与估计方法：最大似然估计（MLE）：**利用统计模型，通过在给定观测到的含噪信号的情况下最大化似然函数来估计潜在信号。**贝叶斯估计：**结合关于信号和噪声的先验知识，通过概率推理来改进估计过程。**自适应方法：自适应滤波器：**这些滤波器根据信号和噪声的局部特性动态调整其参数。例如自适应中值滤波器和卡尔曼滤波器。**自适应阈值处理：**根据局部信噪比调整阈值水平的技术，以有效抑制噪声而不丢失重要的信号细节。**非局部均值（NLM）方法：基于块的处理：**该方法通过对相似图像块进行平均来利用自然图像中的冗余信息以降低噪声。与对数变换结合用于处理乘性噪声时可能特别有效。**正则化技术：总变分（TV）正则化：**通过对图像的梯度进行稀疏性提升，有助于在降噪的同时保留边缘。**Tikhonov 正则化：**对观测数据的保真度和重建信号的平滑度进行加权约束。通过采用这些技术，在图像处理、通信系统和生物医学信号分析等各种应用中，可以比较有效地抑制乘性噪声并提高信号质量。但是，强乘性噪声对有效信号的非线性影响的完整消除显然是非常不容易的。上述方法技术的真实有效性主要还是取决于乘性噪声本身时变空变的强度及尺度。

4.2 对加性和乘性噪声的基本认识

根据上述讨论，可以对加性与乘性噪声的压制进行思想方法上的总结。

首先乘性噪声对信号的影响与加性噪声对信号的影响截然不同。时变且空变剧烈的乘性噪声对信号的影响是非线性的，而后者的影响是线性的。可以看出，若将加性噪声 $N(t)$ 视为一个算子，它不会改变信号 $s(t)$ 的相位，仅改变 $s(t)$ 各采样点的幅值。而乘性噪声算子会改变信号 $s(t)$ 的幅值和相位，且改变幅值的方式也不同，乘性噪声对信号产生更剧烈的改变，以非线性方式改变信号 $s(t)$ 的幅值。

例如，当比较 $d(t)=s(t)\cdot N(t)$ 与 $d(t)=s(t)+N(t)$，且 $N(t)=0.1$ 和 $N(t)=10$ 时，显然乘性噪声对信号 $s(t)$ 的幅值改变很大，且它是以非线性方式改变 $s(t)$ 的幅值，尤其当 $N(t)$ 随时间变化时。要理解乘性噪声对信号 $s(t)$ 相位的改变，可以认为乘性噪声 $N(t)$ 是一个非零相位算子，$N(t)$ 与信号 $s(t)$ 相乘改变了 $s(t)$ 的相位。

乘性噪声在勘探地震学中意味着什么？它对当前的地震波成像有何影响？笔者认为乘性噪声指的是那些并非因地下介质变化而使地震子波发生变换的因素，以及不满足信号与噪声加性关系的影响因素。尽管将这些因素称作噪声似乎显得不够贴切，但我认为这就是其本质。与加性噪声不同，乘性噪声不符合随机噪声的特征，尽管这些因素的变化有时也呈现出随机性，但随机性并非乘性噪声的本质特征。这是笔者个人的观点。基于上述分析，勘探地震学中的乘性噪声是指震源和检波器端诸多复杂因素对地震子波的影响。多次散射及多次反射对地震子波的振幅和相位的影响也表现为乘性噪声的特征。需要强调的是：没有信号（即地震子波）的地方乘性噪声是没有影响的。

勘探地震中，去除或者抑制/减弱乘性噪声的影响是提升地震数据质量的关键。根据前面的分析，首先应该是同相轴上地震子波的排齐，把传播相位的影响尽可能消除，然后假设乘性噪声对初相位的影响是随机的，将信号统一为一致相位或零相位等，这样就能完全消除乘性噪声对信号（地震子波）的影响。假设乘性噪声对相位的影响已经消除，再假设影响幅值的乘性噪声也是随机的，并且继续假设该乘性噪声为高斯零均值噪声，那么可以通过高斯滤波器将其消除。如果该乘性噪声不是高斯零均值噪声，则可以通过非线性（中值）滤波进行抑制。

然而，需要注意的是，噪声对幅值的影响和对相位的影响相互交织，或者可以说，相位的变化往往通过幅值的变化体现出来。这一点在时域中能清晰地观察到。因此，前述那些方法技术到底如何能对乘性噪声起到压制作用，只有在弄清楚区分开乘性噪声和信号（地震子波）的物理机制的条件下，才能得到合理的应用，得到期望的结果。

这就是笔者对勘探地震学中乘性噪声的理解。在全波形反演成像时代，乘性噪声的抑制愈发重要。

5 机器学习中的线性系统与非线性系统对比分析

机器学习是让计算机能够像人一样从数据中学习知识，并且利用知识对新的数据进行预测或者决策的学科。它不是通过手工编程的方式（意思是不再显式地基于各种数值数学算法编程的方式）明确告诉计算机每一个步骤该怎么做，而是让计算机自己从大量的数据中去发现模式和规律。这是豆包给出的机器学习的解释。

机器学习是一组方法，能够自动探测数据中的模式，然后利用所发现的模式预测未来数据，或在不确定情况下执行其他类型的决策（比如规划如何收集更多数据！）。这是 Machine Learning: A Probabilistic Perspective 一书（by Murphy, K. P.,2012）给出的定义。

通过个人经验和代代相传的知识进行学习是人类智慧的核心。同样，任何科学领域的核心任务都在于发展模型（通常称之为理论），以解释每个时期已有的实验证据。换句话说，我们总是从数据中学习，期望归纳出知识，以指导我们的科学技术实践。不同的数据以及对数据的不同关注点催生了不同的科学学科。

机器学习即让计算机从数据中学习，具体而言，旨在探测并揭示与数据生成机制相关的、可能隐藏的结构和规律模式。这些信息不仅有助于我们分析和理解数据的本质，还用于对未来进行预测。除了对潜在结构进行建模，机器学习中一个备受关注的方向是开发高效算法，用于设计模型以及进行分析和预测。在我们所说的大数据时代来临之际，后者正变得愈发重要，因为此时人们必须处理海量数据，这些数据可能存在于维度极高的空间中。针对此类应用进行数据分析，要求算法必须具备在计算上高效，和在性能上稳健，以应对数据在某些情况下受到的大量噪声干扰，并处理在某下情况下可能存在的数据缺失。

几十年来，这些方法和技术一直是众多学科科学研究的核心，比如统计学与统计学习、模式识别、信号与图像处理及分析、计算机科学、数据挖掘、机器视觉、生物信息学、工业自动化以及计算机辅助医学诊断等。尽管名称各异，但所有这些学科都使用一组共同的技术，我们将这些方法统称为机器学习。在过去十几年左右的时间里，这个名称越来越流行。这个名称意味着像大脑学习和预测那样，利用机器 / 计算机进行学习。在某些情况下，这些方法直接受到大脑工作方式的启发，归结出了各种各样的神经网络算法。但是，归根结底机器学习就是 Learning From Data by Computer。

笔者认为，无论从什么样的角度讨论并定义机器学习，其本质已经很清楚了。就是从数据中获得知识，然后帮助我们做出决策。更深层次地，数据一定是从系统中产生的，系统一般是非线性的，理论假设下，系统可以认为是线性的。从线性系统产生的数据中进行学习或是从非线性系统产生的数据中进行学习就成了机器学习的重要分界。

前已述及，大数据时代机器学习算法是目前机器学习领域的核心。从线性系统产生的数据中进行学习的算法已经发展得很成熟。当然，复杂线性系统下的参数估计问题还没有得到很好的解决。机器学习算法真正要着重解决的问题是数学分析手段没有很好解决的非线性问题，并非非线性系统演化问题。众所周知，架构在人工神经网络上的机器学习算法是当前解决非线性问题的利器。

笔者认为：架构在人工神经网络上的机器学习算法是一种新的计算模式，即 ANN/DNN 网络计算模式，它是一种分布并行算法，它模仿人脑计算的功能，是一种智能算法，它的目的是建立起“广义数据”与决策目标之间的非线性映射关系，如下式：

$$F(\mathbf{x},\mathbf{w})=\varphi\left(\sum_k w_{0k}\varphi\left(\sum_j w_{kj}\varphi\left(\cdots\varphi\left(\sum_i w_{li}\tilde{\mathbf{x}}_i\right)\right)\right)\right) \tag{4}$$

其中，$\tilde{\mathbf{x}}=\left(f_\theta^1(\mathbf{x}),\ f_\theta^2(\mathbf{x}),\cdots,\ f_\theta^b(\mathbf{x})\right)$，$\tilde{f}_\theta(\tilde{\mathbf{x}})=\sum_{j=1}^{b}\alpha_j\phi\left(\tilde{\mathbf{x}};\boldsymbol{\beta}_j\right)$。

人工神经网络是由高度抽象的、人工神经元构成的网络。本质上，这是一个计算模型。不过它仿生模拟了人脑神经元及人脑神经网络。事实上，是否真的仿生模拟了人脑神经元及人脑神经网络还很难说。经验地讲，很多仿生优化算法还是有比较好的结果的。既然人脑神经元构成的网络有如此强大的学习、联想、推断和决策能力，可以期望人工神经网络也应该有比较好的类似性能。单个人工神经元的数学模型是描述由输入函数 $\boldsymbol{x}$ 到输出函数 $\boldsymbol{y}$ 之间关系的数学模型，注意不是构建我们熟悉的显式的函数关系，而是由网络节点上的加权系数构成的映射关系。笔者认为这个描述是没有问题的。输出函数 $\boldsymbol{y}$ 与输入函数 $\boldsymbol{x}$ 之间的关系可以用基函数族的线性组合来表示。泛函分析奠定了这方面的坚实基础。这个问题数学上绝对站得住脚。理论和实践也证明了这样的观点。但是输出函数 $\boldsymbol{y}$ 与输入函数 $\boldsymbol{x}$ 之间到底是否有关系，还是由人为选择的合适的输出函数 $\boldsymbol{y}$ 与输入函数 $\boldsymbol{x}$ 决定的，即由标签数据决定的。风马牛不相及的输出函数 $\boldsymbol{y}$ 与输入函数 $\boldsymbol{x}$ 不可能找出什么有用的关系。因此，（完备的）标签数据+（优越的）DNN 算法是深度学习人工智能算法能否成功的关键。即便神经网络算法有强大的非线性映射能力，但针对复杂的非线性系统演化，笔者认为它不一定是最合适的算法。

我们已知非线性系统演化问题与“静态”的非线性问题的求解是息息相关的。神经网络算法建立的非线性系统输入与输出之间的关系式，不是用解析表达式给出的，而是用“巨量的加权系数”表达的，尽管严重缺乏所谓的可解释性、泛化应用的不确定性，但是目前看来，这是人工智能或机器学习的一个可行方向。

是否所有的场景都适用于上述 AI/ML 解决问题的方法，很显然，问题主要出在数据上。只有输入和输出数据的样本集足够大时，即包括了足以建立起二者合理关系的统计特征时，AI/ML 方法就能在这些场景下很好地解决问题。事实上，如果系统是比较线性的，也没有必要杀鸡用牛刀，因为现有的线性系统数据分析方法是相当成熟、有效且廉价。

总之，AI/ML 最核心的作用是在强或较强非线性系统情形下、在“静态”非线性情形下，建立起输入和输出之间的映射关系，用该映射关系解决各种推断和判决问题。至于非线性系统的演化过程的建模预测及最终演化结果的预测和描述，笔者认为还不存在对应的理论与方法。

6 勘探地震中的线性系统与非线性系统对比

勘探地震中，波的激发、传播与接收和地下岩石介质构成一个系统，这个系统基本是线性的，也可以称为拟线性的。这是地震数据预处理、地震波偏移成像与反演成像的最根本的物理基础。

但是，地表接收的波场与地下介质弹性参数（包括岩石性质参数）的变化之间的关系是非线性的（不能再表现为矩阵方程）。弹性参数（值）空变越剧烈（即尺度越小、强度越大），波场越复杂，二者之间的非线性关系越强。直观地，就是炮集中出现更多的多次反射波和多次散射波，还有更多的地震子波振幅和频率的时空变化。另一方面，局部高维数据体中，线性结构飘在高斯随机噪声中的假设此时也非常不符合数据实际，导致各种线性高斯平稳假设下发展出的数据预处理算法越来越失效。地表与地下“双复杂”探区，波场与地下介质弹性参数之间的关系就应该是这样的，导致工业界常用的地震数据成像处理流程在“双复杂”探区得不到希望的结果。

基于上述分析，勘探地震数据分析今后重点要关注的问题应该是：在观测波场与地下介质弹性参数之间出现强或较强非线性关系时，如何进行地震数据预处理和地震波成像。很显然，此时应该发展非线性数据的预处理方法，以及非线性地震波偏移成像与反演成像方法。然而，这两个方面，都没有像线性问题那样存在一个优美严格的数学求解思想与方法体系。

笔者认为：既然勘探地震系统是一个线性或拟线性系统，今后勘探地震数据分析发展的重点放在其中存在的非线性问题上就可以了，而无需在非线性系统演化上投入过多精力。针对油气勘探，时延情形下的地下岩石介质系统演变，应该也不会出现混沌现象和分形现象。

7 结论与讨论

为什么要由对线性问题和线性系统的研究转移到非线性问题和非线性系统上？因为物理和非物理世界中，非线性问题及非线性系统是常态，线性问题及线性系统是数学理论的假设。

真实的世界是复杂的。在各种各样的实际应用中，线性模型的假设往往过于简单。例如，在经济学领域，价格和需求之间的关系通常不是简单的线性关系。当价格上涨时，需求可能会下降，但不是以固定的线性比例下降。如果按照线性模型来处理，可能会对市场情况产生错误的估计。再如，在生态系统中，物种数量的增长与环境资源的关系也不是线性的。一开始，随着资源的增加，物种数量可能会快速增长，但当资源达到一定限度后，由于竞争等因素，增长速度会放慢甚至出现下降，这是典型的非线性关系。线性系统的行为是比较规则的，其输出与输入通常呈比例关系。然而，许多自然现象和工程问题呈现出复杂的、非比例的变化。以流体力学为例，当流体在管道中低速流动时，可能近似符合线性关系，但当流速增大到一定程度后，就会出现湍流等非线性现象，此时线性模型就无法准确描述流体的行为。

非线性问题及非线性系统广泛存在于自然和社会系统中。可以说几乎所有的自然现象和社会现象都包含非线性因素。在物理学中，天体的运动，如行星围绕恒星的运动，考虑到多个星体之间的引力相互作用，这是一个复杂的非线性系统。在生物学领域，生物神经网络的信号传递过程也是非线性的。神经元之间的连接强度和信号的激发与抑制机制导致了复杂的非线性行为，这对于理解大脑的工作原理至关重要。在社会科学方面，交通流量的变化就是非线性的。道路上车辆的密度和车速之间不是简单的线性关系，当车辆密度达到一定程度后，交通就会出现拥堵，车速会急剧下降。

研究非线性问题及非线性系统，可以更好地预测和理解非线性系统行为，捕捉系统随时间的动态变化。例如，在天气预报中，大气系统是高度非线性的。通过对非线性大气模型的研究，可以更准确地预测天气的变化，如暴雨、飓风等极端天气事件。线性模型很难描述这些天气系统中的复杂对流、气旋等现象，而非线性模型可以考虑温度、气压、湿度等多个变量之间的相互作用，从而提供更精准的预报。可以更好地理解非线性系统的内在机制，例如，在化学反应过程中，反应物浓度和反应速率之间的关系往往是非线性的。通过研究这种非线性关系，可以揭示化学反应的微观机制，如反应的活化能、分子碰撞等因素对反应的影响，从而优化化学反应的条件，提高反应效率。

非线性系统的研究对于许多前沿科技的发展有着深远的影响。在人工智能领域，深度学习中的神经网络就是基于非线性函数构建的。神经元之间的非线性激活函数使得神经网络能够拟合复杂的函数关系，从而能够处理图像识别、自然语言处理等复杂任务。在控制工程中，非线性控制理论能够更好地处理具有复杂动态特性的系统，例如机器人的复杂运动控制，使得机器人能够更灵活、精准地完成任务。

尽管勘探地震系统不是一个强非线性系统，但是勘探地震领域存在一些重要的非线性问题，譬如乘性噪音问题、强噪声弱信号情形下的去噪问题、插值问题等、复杂情形（即实测波场与地下介质弹性参数之间存在强非线性关系情形）下地震波反演成像问题、由弹性参数估计岩石性质参数的问题等。

线性问题及线性系统的研究理论上已经十分完善了。必须把研究重点转移到非线性问题及非线性系统的研究上了。

参考文献

[1] Holger Kantz and Thomas Schreiber, 2003, Nonlinear Time Series Analysis, Cambridge University Press.
[2] Jiangqing Fan and Qiwei Yao, 2005, Nonlinear Time Series: Nonparametric and parametric Methods, Springer Science + Business Media, Inc
[3] Peter J. Brockwell and Richard A. Davis, 2006, Time Series: Theory and Methods, Second Edition, Springer
[4] Gonzalo R. Arce, 2005, Nonlinear Signal Processing: A Statistical Approach, A Jogn Wiley & Sons INC., Publication.
[5] Sergios Theodoridis, 2015, Machine Learning: A Bayesian and Optimization Perspective, Elsevier Ltd.
[6] Simon Haykin, 2009, Neural Networks and Learning Machines, Third Edition, Pearson Education, Inc.
[7] Kevin P. Murphy, 2012, Machine Learning: A Probabilistic Perspective, The MIT Press.

勘探地震中需要六分量（包括旋转分量的）观测吗？

王华忠

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】勘探地震是一门基于人工震源激发的震动在地下介质中传播并在地表接收地震波场，通过分析地震波场估计地下介质弹性参数，从而解决地学相关问题的学科。感知并记录地表介质的振动波场是最基础的工作。要完整描述地表介质单元（可变形物体）在微变形情况下形成的运动，需要三个平移分量、六个应变分量以及三个作为矢量的转动分量。但是，地震波引起的转动运动长期以来基本上被忽视，首先是因为人们认为转动效应很小，其次是因为当时没有灵敏的测量设备。实际上，通用转动传感器的现有灵敏度还不足以用于有效的地球物理应用。勘探地震中，震源能量有限，引起的地表介质质点运动很微弱，转动效应更小。到底是否应该引入旋转分量或转动分量观测？理论上应该是必要的！核心问题在于能否有高精度的传感器保真地感知到地表介质质点的 6C 运动并把它们忠实地记录下来。随着最近基于光纤的高灵敏度旋转运动传感器的问世，6C 技术在勘探问题上的广泛应用已指日可待。潜在应用基于多种原理，例如，平移和旋转运动联合观测的扩展数据集，使得利用单台站记录开展诸如波场分离和面波压制等阵列式处理成为可能。此外，测量地震波场的旋转分量（旋度）能够直接分离横波成分，可显著提升横波勘探效果。旋转测量可估算自由表面的空间波场梯度，从而得以开展诸如局部慢度估计和波场重建等分析。再者，旋转运动测量有助于解决因地震传感器与地面耦合不佳而导致的波场失真问题。本文提出勘探地震中是否需要旋转分量或转动分量观测的问题，并依据现有文献展开讨论，目的是希望引起石油工业界相关人员的重视，尽早探索六分量观测在油气勘探中应用的可能性。

【关键词】地表介质质点，平动与旋动，六分量观测，六分量检波器，旋转分量的可能应用

1 引言

在过去几十年里，与对平移型的震动的观测、处理及反演方面的大量研究相比，地震引起的地面运动中的转动部分基本上被忽视了，尽管从理论层面考虑，对这类运动的观测可能确实有用，并且能提供额外信息。过去几年，地震学领域对这一潜在的新观测对象的兴趣有所增加，主要是因为借助光纤或环形激光陀螺仪等现代采集技术，转动运动已能被实际观测到，分辨率在稳步提高，而且这些观测结果与同位置的平移型的震动记录日趋一致。

传统上，地震勘探是通过传感器进行对地球内部结构的探测的，这些传感器最多可输出由地震波场引起的三维平移运动（传感器外壳相对于惯性质量的平移运动）。然后，传感器的运动被解读为对地面运动的估计，地面运动由一个最多包含三维位移、质点速度或质点加速度的矢量场来描述。然而，矢量场的旋转分量，以及安装在地面上的传感器也会受到地震引发的旋转运动影响这一事实，到目前为止在很大程度上被忽视了。与平移运动类似，旋转运动的度量包括角位置（角度）、角速度（旋转速率）和角加速度。旋转运动测量结果可被解读为对位移场旋度的估计。类似地，水听器（或压力）测量结果可被解读为对位移场散度的估计。由于直到最近才出现足够可靠且对全球地震学和地震勘探所需频段应用足够敏感的旋转传感器，因此测量旋转运动的地震实验仍然很少。

旋转地震学这一新兴领域从利用环形激光技术对旋转运动垂直分量的高分辨率测量中受益匪浅（Igel 等，2005，2007）。尽管这些仪器主要用于大地测量以地球的旋转速率测量，但它们对地震和海浪引发的旋转运动也很敏感（Sollberger 等，2006，2009）。环形激光观测引发了多项研究，旨在探索旋转运动观测在解决地震反演问题方面的潜力（Igel 等，2015），以期应用于地震学（Lee 等，2009）。

在勘探地震学中，旋转观测最近也越来越受到关注。这种兴趣源于期望克服诸如相干噪声抑制和采样要求等长期存在的问题。例如，旋转运动数据已被用于辅助横波解释（Sollberger 等，2016a）以及开发单台站噪声抑制方案（Barak 等，2014；Sollberger 等，2018）。此外，自由表面旋转运动测量与地震波场空间梯度的内在联系已被用于实现空间混叠数据的波场重建方案（Muyzert 等，2012）。另外，对未受污染的旋转运动的直接测量可以消除水平分量记录与倾斜（绕水平轴的旋转运动）之间众所周知的模糊性，这种模糊性会降低例如海底记录的数据质量（Lindner 等，2016）。共址平移和旋转运动测量的一个特别有趣之处在于，单台站的六维观测能够提取波场特征，如传播方向、局部慢度和波型，而这些特征原本需要对空间分布广泛的传感器阵列数据进行分析才能得到。

除了测量三维平移运动外，还测量三维旋转运动，这将地震观测范畴扩展到了六维。这样的六维测量有以下潜在优势：（1）例如，由于位移旋度与横波之间的内在联系，能更好地理解局部地面运动。（2）自由表

面的旋转运动单点测量可提供局部空间波场梯度估计，从而提供关于地震波场的空间信息，而这些信息原本只能通过大型密集接收阵列测量得到。例如，地震波场的空间梯度在空间波场滤波器和波场重建算法中有所体现。（3）额外的地震观测数据增强了统计波场分析和滤波技术。（4）旋转运动测量有助于解决因地震仪器与地面耦合不佳而导致的矢量测量失真问题。这种失真可能是由于传感器耦合不佳，在地震波到达时发生倾斜或旋转，或者是由于诸如洋流等外力作用。

本文提出勘探地震中是否需要旋转分量或转动分量观测的问题，并依据现有文献展开讨论，目的是希望引起石油工业界相关人员的重视，尽早探索六分量观测在油气勘探中应用的可能性。

2 六分量地震观测的基础理论

2.1 位移、应变与旋转

勘探地震学和全球地震学依赖于对位移$u(\boldsymbol{x})$、质点速度$\dot{u}(\boldsymbol{x})$或质点加速度$\ddot{u}(\boldsymbol{x})$的空间分量$u_i(\boldsymbol{x})$的记录。我们考虑在时刻t位于$\boldsymbol{x}$处的一个质点，如图 1 所示。示意图 1（右）展示了线段$\delta\boldsymbol{x}$的位移情况。在未变形状态（浅灰色）下，线段端点为$\boldsymbol{x}$和$\boldsymbol{x}+\delta\boldsymbol{x}$，分别因位移$u(\boldsymbol{x})$和$u(\boldsymbol{x}+\delta\boldsymbol{x})$，在变形状态下移动至$\boldsymbol{\xi}$和$\boldsymbol{\xi}+\delta\boldsymbol{\xi}$。由该位移引起的总旋转量对应于$\delta\boldsymbol{x}$与$\delta\boldsymbol{\xi}$之间的夹角$\omega$。

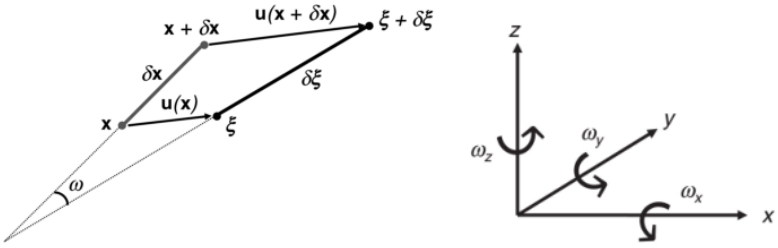


图 1 质点运动关系

在时刻$t+\delta t$，由于受力变形，该质点相对于同一惯性参考系移动到坐标$\boldsymbol{x}+\delta\boldsymbol{x}$。为了描述这种变形，我们考虑$\boldsymbol{x}$处的质点相对于附近$\boldsymbol{x}+\delta\boldsymbol{x}$处一点的相对运动，向量线元$\delta\boldsymbol{x}$定义了这两点之间的线段。对$\boldsymbol{x}+\delta\boldsymbol{x}$处的位移进行 Taylor 展开并保留前两项，得到：

$$u_i(\boldsymbol{x}+\delta\boldsymbol{x})=u_i(\boldsymbol{x})+\partial_j u_i\delta x_j+\mathcal{O}\left(|\delta\boldsymbol{x}|^2\right) \tag{1}$$

其中，∂_j表示在x、y或z方向的空间导数。第一项表示刚体平移，因为$\boldsymbol{x}+\delta\boldsymbol{x}$附近的所有点都具有相同的位移。第二项对应于位移梯度张量（这是一个二阶张量），它显然可以进一步分离为对称部分和反对称部分，忽略高阶项时得到：

$$u_i(\boldsymbol{x}+\delta\boldsymbol{x})=u_i(\boldsymbol{x})+\frac{1}{2}\left(\partial_j u_i+\partial_i u_j\right)+\frac{1}{2}\left(\partial_j u_i-\partial_i u_j\right) \tag{2}$$

假设位移梯度很小（即$\left|\partial_j u_i\right|\ll 1$），位移梯度张量的对称部分和反对称部分可以分别与无穷小应变张量和无穷小刚体旋转张量相关联，有

$$\epsilon_{ij}=\frac{1}{2}\left(\partial_j u_i+\partial_i u_j\right) \tag{3}$$

和

$$\theta_{ij}=\frac{1}{2}\left(\partial_j u_i+\partial_i u_j\right) \tag{4}$$

无穷小刚体旋转只有三个独立分量，可以用向量$\boldsymbol{\Theta}$表示，其元素$\Theta_k=-\frac{1}{2}e_{kij}\theta_{ij}$，其中$e_{kij}$是 Levi-Civita 符号（置换张量）。

已知应变张量$\boldsymbol{\epsilon}$可以分解为体积（主）应变和剪切（偏）应变，

$$\boldsymbol{\epsilon}=\begin{pmatrix}\partial_x u_x & 0 & 0\\ 0 & \partial_y u_y & 0\\ 0 & 0 & \partial_y u_y\end{pmatrix}+\frac{1}{2}\begin{pmatrix}0 & \partial_y u_x+\partial_x u_y & \partial_z u_x+\partial_x u_z\\ \partial_x u_y+\partial_y u_x & 0 & \partial_z u_y+\partial_y u_z\\ \partial_x u_z+\partial_z u_x & \partial_y u_z+\partial_z u_y & 0\end{pmatrix} \tag{5}$$

$\boldsymbol{\epsilon}$的迹对应于相对于初始体积的体积变化（膨胀），它等同于位移场的散度：$\epsilon_{ii}=\nabla\cdot\boldsymbol{u}$。水听器和压力传感器测量$\epsilon_{ii}$。$\boldsymbol{\epsilon}$的非对角元素对应于剪切应变，表示形状的变化而体积不变（（5）式的第二项）。对于无穷小变形，剪切应变等于初始直角的变化，这与旋转相关。

为了确定与变形相关且将由旋转传感器测量的总旋转，我们考虑向量线元（见图 1）。变形将导致$\delta\boldsymbol{x}$的长度和方向发生变化，从而产生一个刚度单元的变形状态$\delta\boldsymbol{\xi}$。定义单位向量$\delta\hat{\boldsymbol{x}}=\delta\boldsymbol{x}\big/|\delta\boldsymbol{x}|$，假设无穷小应变，变形后的总旋转是$\delta\boldsymbol{x}$与$\delta\boldsymbol{\xi}$的叉积，即

$$\boldsymbol{\omega}=\frac{\delta\boldsymbol{x}\times\delta\boldsymbol{\xi}}{|\delta\boldsymbol{x}|^2}=\delta\hat{\boldsymbol{x}}\times\left(\boldsymbol{\epsilon}\cdot\delta\hat{\boldsymbol{x}}\right)+\boldsymbol{\Theta}-\left(\boldsymbol{\Theta}\cdot\delta\hat{\boldsymbol{x}}\right)\delta\hat{\boldsymbol{x}} \tag{6}$$

其分量表达式为：$\omega_i=e_{ijk}\epsilon_{km}\delta\hat{x}_j\delta\hat{x}_m+\Theta_i-\Theta_j\delta\hat{x}_i\delta\hat{x}_i$。

总旋转是由剪切应变引起的应变分量（上式第一项）与刚体旋转（上式第二项和第三项）之和。应变分量是单位向量与相对位移的叉积。应变分量描述了线元围绕其质心的相对旋转，旋转方向遵循右手定则（见图 1），而刚体旋转是相对于其他某个点的。

Aki 和 Richards（2002，P13）在经典弹性理论框架下，进一步假设为微小变形时，给出了与上述讨论类似的关系式，

$$u\left(\boldsymbol{x}+\delta\boldsymbol{x}\right)=u\left(\boldsymbol{x}\right)+\varepsilon\cdot\delta\boldsymbol{x}+\boldsymbol{\omega}\times\delta\boldsymbol{x} \tag{7}$$

其中，ε、$\boldsymbol{\omega}$分别是应变张量和转动二阶张量，并且

$$\boldsymbol{\omega}=\frac{1}{2}\nabla\times u\left(\boldsymbol{x}\right) \tag{8}$$

其中，$\boldsymbol{\omega}$是一个（伪）矢量，它不包含在胡克定律中，代表了扰动所产生的刚体转动角度。这表明，要完整描述点$\boldsymbol{x}$周围介质的变化，需要三个平移分量、六个应变分量以及三个转动分量。

2.2 弹性波动方程与波场的旋度

前面在连续介质力学基础上讨论刚体形变的数学表达。下面讨论波传播的动力学。对于各向同性介质，在无震源区域，应力张量的分量与应变张量（见（3）式）通过各向同性弹性本构方程相关联，

$$\sigma_{ij}=\left(\lambda\delta_{ij}\delta_{kl}+\mu\left(\delta_{ik}\delta_{jl}\right)\right)\epsilon_{kl} \tag{9}$$

其中，λ和μ是拉梅常数，δ_{ij}是克罗内克符号。

由运动方程以及式（3）和式（7），可得弹性波动方程为：

$$\rho\ddot{\boldsymbol{u}} = \boldsymbol{f} + (\lambda + 2\mu)\nabla(\nabla\cdot\boldsymbol{u}) - \mu\nabla\times(\nabla\times\boldsymbol{u}) \tag{10}$$

其中，ρ 是密度，$\boldsymbol{f}$ 表示体力。式（10）使我们能够描述位移扰动在均匀各向同性弹性介质中的传播。

拉梅定理指出，对于位移场 $\boldsymbol{u}$，存在势函数 Φ 和 Ψ，使得（Aki 和 Richard，2002）

$$\boldsymbol{u} = \nabla\Phi + \nabla\times\Psi \tag{11}$$

其中，$\nabla\Phi$ 和 $\nabla\times\Psi$ 分别称为波场 $\boldsymbol{u}$ 的纵波（P 波）和横波（S 波）分量。由于 $\nabla\Phi$ 的旋度为零，因此有，

$$\nabla\times\boldsymbol{u} = \nabla\times\nabla\times\Psi \tag{12}$$

在（局部）均匀各向同性弹性介质中，因此旋转运动仅由横波（S）波表现出来。所以，通过用旋转传感器测量 $\boldsymbol{\omega}$，我们可以提取波场 $\boldsymbol{u}$ 的旋度并分离出其中的 S 波成分，在介质中其表达式为，

$$\boldsymbol{\omega} = \frac{1}{2}\nabla\times\boldsymbol{u} = \frac{1}{2}\begin{pmatrix}\partial_y u_z - \partial_z u_y\\ \partial_z u_x - \partial_x u_z\\ \partial_x u_y - \partial_y u_x\end{pmatrix} \tag{13}$$

在各向异性介质中，这种分离不再成立，甚至 P 波中也可能有旋转分量（Pham 等，2010）。即各向异性介质中不再有纯粹的 P 波和 S 波。

2.3 自由表面的旋转运动

在自由表面，应力张量的垂直分量（$\sigma_{iz}=0, i=x,y,z$）为零（时，），因此也有变张量分量 $\epsilon_{iz}=0, i=x,y,z$。由于自由表面施加的这些约束，式（13）可改写为：

$$\boldsymbol{\omega}^{FS} = \begin{pmatrix}\\ \partial_y u_z\\ -\partial_x u_z\\ \frac{1}{2}(\partial_x u_y - \partial_y u_x)\end{pmatrix} \tag{14}$$

该式将自由表面的旋转测量值（用上标表示）与位移场的空间导数联系起来。注意，式（14）中没有垂直导数，因为自由表面条件使我们能够将垂直导数转换为水平导数。式（14）表明，围绕两个水平轴的旋转测量值提供了自由表面垂直位移空间波场梯度的点估计值，这一量与波场重建和成像相关（Muyzert 等，2012）。考虑到联合位移和位移空间梯度测量能够在远离传感器位置处重建波场，旋转运动提供了大空间尺度上的空间信息，而这些信息原本只能通过大型接收阵列获得。

2.4 体波和面波引起的旋转运动的平面波分析

在自由表面，体波和面波引起的旋转运动估计值可以从单色平面波的加速度测量值推导出来。考虑沿 x 方向传播的波，对于 P 波、SV 波和瑞利波，绕轴的旋转速率和垂直加速度为（Li 等，2004；Langston，2006），

$$\dot{\omega}_y = p_x^a \ddot{u}_z \tag{15}$$

其中，p_x^a 是 P 波、S 波或瑞利波在 x 方向的局部视水平慢度。注意，当 P 波以非垂直入射角撞击自由表面时，会发生 P 波到 S 波的转换，从而产生旋转运动。在频率域，式（15）可写为，

$$\tilde{\dot{\omega}}_y = p_x^a (j2\pi f)\tilde{u}_z \tag{16}$$

从式（16）中我们注意到，旋转振幅随频率和慢度的增加而增大。这些关系表明，与全球地震学相比，在典型的地震勘探频率范围内，对于相似的平移运动振幅，可以预期有显著的旋转能量。此外，与体波相比，慢传播的面波具有更高水平的旋转能量。对于沿 x 方向传播的 SH 波和勒夫波，自由表面围绕垂直轴的诱导旋转运动由下式给出（Igel 等，2005；Lin 等，2011），

$$\dot{\omega}_z = -\frac{1}{2} p_x^b \ddot{u}_y \tag{17}$$

其中，p_x^b 是局部视水平 S 波慢度。式（15-17）表明，旋转速率和加速度由视水平慢度相关联。因此，来自共址六分量（6C）测量的振幅比可用于获取孤立波至的瞬时慢度（速度）信息，而无需使用走时。总之，平移和旋转运动的共址点测量所包含的信息能够直接推断地下地震速度结构和传播方向。因此，有可能从振幅信息中重建接收器附近的局部浅层速度模型，而无需使用走时。

2.5 各向异性

在（局部）均匀弹性三维各向同性介质中，旋度算子将无散度的 S 波与无旋度的 P 波分离（公式 14）。在各向异性介质中，这种分离不再成立，其中 P 波也会对旋度产生贡献，进而影响在各向异性介质中旋转传感器的观测结果。尽管 Pham 等（2010）利用 Christoffel 方程研究了各向异性对 P 波的影响，但是各向异性介质中旋转运动的分析仍有待探索。他们得出结论，对于近震和典型的储层情况，由各向异性引起的准 P 波旋转速率显著且可记录，可用于约束各向异性参数的反演问题。因此，利用旋转运动观测来量化各向异性参数具有潜力，特别是在井下实验中（Horne，2012）。

3 如何测量旋转运动

近期开发用于记录旋转运动的便携式仪器的方法涉及多种技术解决方案。最常见的技术有基于液体的旋转运动传感器、磁流体动力学传感器以及光纤陀螺仪，如图 2 所示，其中（a）是由 Eentec 公司生产的 R1/R2（Eentec，2017）；（b）由应用技术协会（ATA）研发的地震磁流体动力学原型传感器（Pierson 等，2016）。（c）由 Blueseis 公司生产的 BlueSeis3A（BlueSeis，2017）；（d）由 Litef 公司生产的 LCG。

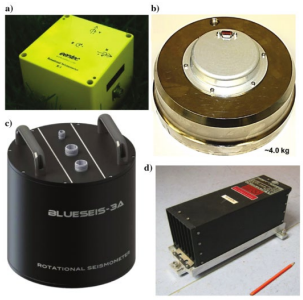


图 2 便携式旋转运动仪器

基于液体的旋转运动传感器通过分子电子转移换能器测量充满液体的圆环内的差异运动。研究发现，这些传感器适用于强震运动测量。Huang 等（2013）和 Egorov 等（2015）对基于液体的旋转运动传感器进行了全面概述。其中一些传感器已成功应用于强震野外考察（Lee 等，2009；Nigbor 等，2009；Wassermann 等，2009），但它们对环境温度变化非常敏感（Bernauer 等，2012）。

磁流体动力学传感器（Pierson 等，2016）是另一种利用稳定外部磁场中导电流体惯性的方法。在外部施加旋转的情况下，流体与磁场之间的相对运动会产生可测量的径向电流。

在光纤陀螺仪中，光纤环内两束反向传播激光束的光程长度差异会产生一种干涉图案，该图案取决于所施加的旋转速率。它利用了 Sagnac 效应（Lefèvre，2024）。关于 Sagnac 效应，见图 3。在外部旋转的情况下，两束反向传播的激光束的相对相位，进而是干涉条纹，会根据角速度（$\dot{\omega}_z$）发生偏移（Igel 等，2005）。

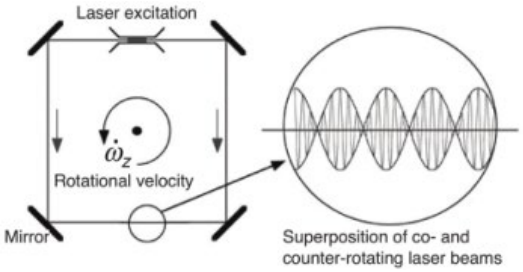


图 3 Sagnac 效应示意图

这种测量原理最重要的优点在于它基于无质量的光子。这使得它本质上对平移运动不敏感，并确保了对纯旋转的测量。已经有多项研究使用光纤陀螺仪来感知旋转运动（Jaroszewicz 等，2006，2012；Velikoseltsev 等，2012；Lindner 等，2016）。最近，首款专门为地震学需求设计的商用光纤陀螺仪已问世（法国公司 iXBlue 制造了 BlueSeis3A；Bernauer 等 2017）。

其他用于测量旋转的技术提议基于微机电系统技术（Alessandro 和 Anna，2014；Liu 和 Pike，2016），或者将传统检波器安装在固定框架上，并通过检波器记录的差分来估计围绕所有三个坐标轴的旋转速率（即旋转检波器；Brokešová 和 Málek，2010），又或者使用包含旋转换能器的仪器来测量粒子的瞬时旋转速率，即所谓的“旋转地震检波器”（Cowles，1981）。

3.1 阵列推导的旋转

由于直到最近才出现可靠的旋转运动仪器，在一些研究中，传统的单分量或多分量传感器阵列被用于通过计算记录的地震波场的水平空间梯度来估计旋转运动（Spudich 等，1995；Huang，2003；Suryanto 等，2006；Langston，2007a，2007b，2007c；Liang 和 Langston，2009；Spudich 和 Fletcher，2009；Sollberger 等，2016b）。利用公式 14 可以估计自由表面围绕所有三个轴的旋转运动。此外，三分量（3C）台站阵列能够估计完整的位移梯度张量，进而估计应变和膨胀（另见“理论基础”部分）（Spudich 和 Fletcher，2009；Donner 等，2017a）。使用有限差分技术估计空间梯度和旋转是基于这样的假设：地下介质是均匀的，并且在阵列区域内变形是线性的。此外，还假设阵列内接收器与地面的耦合以及接收器特性（例如，传递函数、噪声）是均匀的。

通过简单差分相邻台站的记录来计算空间梯度可能特别容易受到噪声影响。因此，通常使用反演方案来稳定空间梯度估计（Liang 和 Langston，2009；Spudich 和 Fletcher，2009）。即使使用反演方案，空间梯度估计的准确性也取决于地震波的视波长和到达方位角与阵列几何形状之间的关系（Schmelzbach 等，2016），这会在梯度估计中引入一种伪各向异性（Ridder 和 Curtis，2017）。这说明用阵列推导旋转存在明显的局限性。发展轻便的、高精度的旋转运动检波器是（看起来是）正确的技术方向。

4 旋转分量观测记录在勘探地震学中的潜在应用

旋转运动测量和六分量（6C）数据在勘探地震学中的应用大致可分为三大类，旋转数据的主要作用包括：（1）对地震波场中横波（S 波）分量的直接测量；（2）对空间波场梯度的估计；（3）作为额外的地震观测数据，扩展用于确定性和统计波场分析的观测数据集。下面具体地讨论各种可能的应用。

4.1 改进 S 波识别

在（局部）各向同性均匀介质中，用势函数表示的波动方程由无旋的 P 波势和无散的 S 波势组成（见式 9）。因此，对这种形式的波场应用旋度算子将分离出 S 波势（见式 10）。由于测量旋转可被理解为测量波场的旋度，旋转测量为波场的 S 波分量提供了直接估计。例如，Igel 等（2016a）表明，从密集高分辨率多分量剖面推导的阵列旋转估计值可用于分离直达 S 波，凸显了旋转数据在改进 S 波勘探方面的潜力。阵列推导的旋转估计值可从以往的纯垂直分量平移数据（例如，提供质点速度数据的检波器数据）中得出，为重新评估现有数据开辟了新途径。

4.2 空间梯度估计与波场重建

在自由表面，围绕两个水平轴的旋转运动对应于垂直质点速度分量的水平梯度（见式 14）。空间梯度的点测量为解决地震勘探中一些长期存在的问题开辟了新途径，例如为避免面波的空间混叠通常需要密集的空间采样。对地震波场及其空间梯度进行共址测量，能够放宽采样要求，并根据香农采样准则从空间混叠的数据中重建波场。这种波场重建技术已成功应用于多分量海洋数据（即测量压力及其在三个方向上的空间梯度；Robertsson 等，2008；Vassallo 等，2010）。Muyzert 等（2012）讨论了将其应用于陆地地震数据，通过结合质点速度和旋转测量来重建混叠的陆地地震数据。Ridder 和 Maddison（2017）提出了一种通过反演波参数和近地表特性来扩展波场重建概念的新方法。Maeda 等（2016）利用地震梯度测量技术，通过日本 Hi-net 地震台网的地震记录来重建地震波场。

4.3 局部速度估计

Langston（2007a，2007b，2007c）以及 Liang 和 Langston（2009）提出的地震梯度测量技术带来了一些有趣的应用，该技术通过结合波场的空间和时间梯度测量，能够估计孤立平面波的局部和瞬时慢度（见式 15 和 17）。

Fichtner 和 Igel（2009），Bernauer 等（2009，2012）提出了一种新颖的层析成像方案，Bernauer 等（2014）对这一概念进行了推广，以适用于最优观测值。受横向加速度与旋转速率之间简单关系（式 17）的启发，“视速度” 这一波场观测值的概念被定义为加速度（或速度）与旋转速率（或旋转角度）的振幅比。将伴随方法应用于此类观测值，可用于计算相应的灵敏度核。直观地说，对震源附近结构的灵敏度为零，只有接收器位置附近的灵敏度才变得重要。这使得能够开发一种基于（例如，频率相关的）视速度观测值的反演方案，而无需使用走时（Bernauer 等，2009）。在没有地震台网的情况下，这种方案可能有助于推导局部地下速度模型（例如，小规模场地效应特征描述或岩石行星物体的地下速度估计）。

4.4 波场分离

Robertsson 和 Muyzert（1999）提议使用三维（体积）接收器布局来估计所有空间导数，以计算所测量地震波场的旋度和散度，从而将波场分离为其 P 波和 S 波成分。仅使用基于地面的多分量接收器，Robertsson 和 Muyzert（1999）通过近似的旋度和散度算子实现了对现场数据的波场分离。Robertsson 和 Curtis（2002）表明，由于自由表面边界条件，在自由表面进行的密集阵列测量足以估计水平和垂直导数（见式 14）。然后，他们推导了一组波场分离滤波器，以估计自由表面的纯上行（入射）P 波和 S 波成分，这需要对空间波场导数进行估计。推导过程涉及泰勒展开，由此产生的一阶项包含波场的空间梯度。

Van Renterghem 等（2016）提出了这些基于梯度的波场分离滤波器的变体，以估计陆地地震数据的纯上行（入射）和/或 P/S 分离的质点速度分量。Van Renterghem 等（2017，2018）将该方法扩展到海底地震数据，例如，去除接收器侧水层多次波（即下行波场）后得到上行 P 波和 S 波成分。对于陆地地震情况，用于分离上行 S 波场的基于梯度的滤波器表达式为（Van Renterghem 等，2017，2018），

$$v_x^{S^u} \approx \frac{1}{2}\left(v_x - \frac{1}{j\upsilon} 2\beta \partial v_z \right) \tag{16}$$

其中，$\upsilon = 2\pi f$ 角频率，β 是接收器位置处的 S 波速度。考虑到 $\partial_x v_z$ 对应于可由旋转传感器测量的绕轴的旋转，式 18 说明了共址速度和旋转测量在波场分离方面的潜力，有助于实现“真实”振幅的 S 波成像。

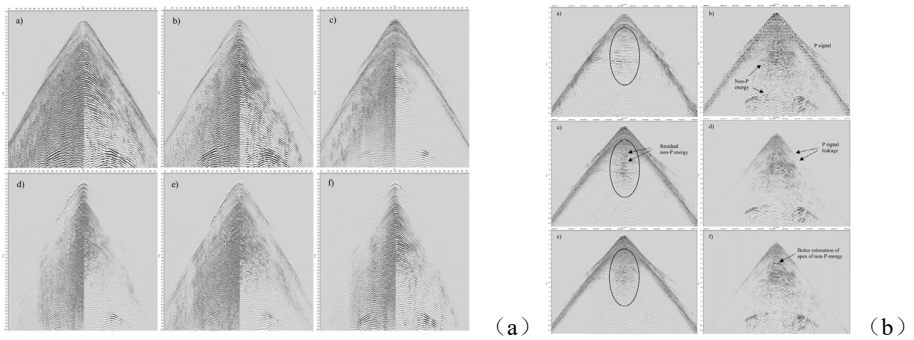
4.5 单台站波场特征描述和面波压制

从不同的角度来看，旋转测量提供了额外的数据，扩展了用于统计波场特征描述和处理的地震观测数据集。Barak 等（2014）提出了一种面波压制方案，该方案基于在一个采样时间窗口内手动提取面波的 6C 统计特征，然后使用这种所谓的“波特征”去除其他时间窗口中的面波。Sollberger 等（2018）通过提供瑞利波的 6C 极化模型，推导出了这些波特征的解析表达式，这使得我们能够以单台站方法识别和压制地滚波。

基于围绕水平轴的旋转运动记录是垂直分量波场面波振幅相对于反射波振幅放大后的慢度缩放版本这一观察结果，Edme 等（2013）提出了一种通过从垂直分量记录中自适应减去时间积分的旋转数据来压制侧向散射面波的技术。在 Edme 和 Muyzert（2013）进行现场测试后，他们使用两个垂直检波器组成的阵列，基于有限差分确定横向旋转运动。

4.6 海底地震仪数据的矢量失真校正

海底地震仪（OBS）会受到由海底洋流以及地震波到达所引起的平移和旋转运动（即传感器的摇晃运动）的影响。此外，OBS 几乎总是放置在未固结的沉积物上（Sutton 等，1981），这导致 OBS-海底系统出现严重的共振。因此，OBS 的水平分量通常看起来噪声较大，并且 OBS 多分量记录存在需要校正的矢量失真问题（Garmany，1984；Dellinger 等，2001）。水平分量的部分干扰源于 OBS 系统的旋转（倾斜）运动（Gaiser，2007）。Lindner 等（2016）最近证明，在海底节点中加入旋转地震仪能够去除 OBS 水平平移运动分量中的倾斜干扰，从而改善 S 波成像。Masoomzadeh 等（2024）尝试在海底节点（OBN）中加入旋转地震仪，利用旋转数据衰减垂直运动传感器上的非压缩波能量。如图 4 所示。图 4A：a）至 c）分别为平移加速度的纵测线、横测线和垂直分量；d）至 f）分别为旋转加速度的相同分量。图 4B：a）水听器数据；b）包含纵波（P 波）和非纵波事件的垂直加速度数据；c）在曲波域减去非 P 波能量后的垂直分量，通过与 a）对比，可识别出一些非 P 波能量；d）使用曲波域方法估计的非 P 波模型，可观察到存在一些 P 波泄漏，且顶点处非 P 波能量不足；e）基于水平旋转分量去噪后的垂直分量，与 a）中的 P 波数据更为相似；f）从旋转数据得到的非 P 波模型，不存在 P 波泄漏，且非 P 波事件的能量比 d）中更强。



（a） （b）

图 4 利用旋转数据衰减垂直运动传感器上的非压缩波能量

5 讨论与结论

利用六分量（6C）数据开展的初步研究突显了直接测量地震波所诱发的三个旋转分量的巨大潜力。可以预期，在过去几年中已取得的旋转传感器技术的快速进展将进一步向前推进，在不久的将来，借助价格合理、便于携带且性能可靠的设备就能实现精确的 6C 测量。

共址平移和旋转（6C）数据的潜在应用可分为以下几类：（1）利用旋转运动与 S 波和面波的直接联系，例如，强化 S 波处理以及压制地滚波；（2）将自由表面的旋转运动数据作为空间波场梯度估计值，用于局部慢度估计、波场重建和波场分离；（3）通过分析旋转和平移数据的相对振幅，提取近接收器的物质属性。6C 测量一个特别吸引人的方面在于，6C 数据能够基于单台站记录进行分析，而传统上这类分析需要大型接收阵列。因此，6C 测量对于那些只能安装少量且分布稀疏的地震传感器的环境应用可能尤为重要，比如在钻孔中、海底，或者用于行星地震学任务。

预计在未来几年，基于 6C 数据，勘探地震学的以下领域将取得重要进展，这些进展对全球地震学也具有重要价值：

1、全弹性波场分解为所有波型：由于 6C 运动是每种波型的特征，通过单台站的旋转直接测量，就可以自动识别和分离波的到达。

2、S 波和转换波成像：由于旋转测量与 S 波之间存在内在联系，预计 S 波成像将从 6C 数据中获益最大。散度传感器的发展可以进一步增强 P 波和 S 波波场的分离效果。

3、无需走时的层析成像进行近地表特征描述和速度模型构建：旋转数据和地震梯度测量能够改进近地表速度模型，通过改进静校正，有助于抑制地震图像中近地表的干扰效应，特别是对 S 波而言。

4、波场重建：自由表面的旋转运动测量与地震波场空间梯度的内在联系，为强大的波场重建技术提供了潜力，这可能促使我们重新思考当前的高道数采集策略。

5、各向异性介质特征描述：与标准的三分量（3C）数据相比，6C 数据提供的三个额外的独立观测值，有助于更好地约束多参数反演问题，例如各向异性参数的估计。例如，P 波在各向异性介质中传播时会表现出旋转分量。

6、微地震监测：鉴于已证明加入旋转数据可增强地震矩张量反演效果，我们预计 6C 数据将有助于更好地约束微地震事件的位置和震源机制。这对于监测水力压裂实验和油藏可能具有重要价值。

7、旋转地震震源：理论上，紧密排列的有源地震震源阵列可以模拟旋转和膨胀震源，这能够进一步改进波场分离技术。

8、海底地震仪（OBS）和海底电缆数据的矢量保真度校正：通过额外考虑旋转数据，可以抑制 OBS 水平分量数据中的倾斜干扰，从而增强海底多分量勘探效果。

总之，随着 6C 检波器技术逐渐地成熟，价格降低到可普及的程度，六分量数据的采集会逐渐在多波勘探中得到广泛应用，成为多波勘探的重要推动力。

参考文献

[1] Aki, K., and P. G. Richards, 2002, Quantitative seismology, 2nd ed.: University Science Book.

[2] Barak, O., F. Herkenhoff, R. Dash, P. Jaiswal, J. Giles, S. de Ridder, R. Brune, and S. Ronen, 2014, Six-component seismic land data acquired with geophones and rotation sensors: Wave-mode selectivity by application of multicomponent polarization filtering: The Leading Edge, 33,1224–1232, doi: 10.1190/tle33111224.1.

[3] Basu, D., A. S. Whittaker, and M. C. Constantinou, 2015, Characterizing rotational components of earthquake ground motion using a surface distribution method and response of sample structures: Engineering Structures, 99, 685–707, doi: 10.1016/j.engstruct.2015.05.029.

[4] Bernauer, F., J. Wassermann, A. Frenois, R. Krissou, A. Bigueur, A. Gaillot, E. de Toldi, D. Ponceau, F. Guattari, and H. Igel, 2017, BlueSeis3A - Full characterization of a 3C broadband rotational ground motion sensor for seismology: 19th EGU General Assembly, EGU2017, Proceedings from the Conference held 23–28 April, 2017 in Vienna, Austria, 15512.

[5] Bernauer, F., J. Wassermann, and H. Igel, 2012, Rotational sensors — A comparison of different sensor types: Journal of Seismology, 16, 595–602, doi: 10.1007/s10950-012-9286-7.

[6] Bernauer, M., A. Fichtner, and H. Igel, 2009, Inferring earth structure from combined measurements of rotational and translational ground motions: Geophysics, 74, no. 6, WCD41–WCD47, doi: 10.1190/1.3211110.

[7] Bernauer, M., A. Fichtner, and H. Igel, 2014, Reducing non uniqueness in finite source inversion using rotational ground motions: Journal of Geophysical Research: Solid Earth, 119, 4860–4875, doi: 10.1002/2014JB011042.

[8] BlueSeis, 2017, www.blueseis.com, accessed 29 March 2018.

[9] Böse, M., J. F. Clinton, S. Ceylan, F. Euchner, M. van Driel, A. Khan, D. Giardini, P. Lognonné, and W. B. Banerdt, 2017, A probabilistic framework for single-station location of seismicity on earth and mars: Physics of the Earth and Planetary Interiors, 262, 48–65, doi: 10.1016/j.pepi.2016.11.003.

[10] Brokešová, J., and J. Málek, 2010, New portable sensor system for rotational seismic motion measurements: Review of Scientific Instruments, 81, 084501, doi: 10.1063/1.3463271.

[11] Castellani, A., and G. Boffi, 1989, On the rotational components of seismic motion: Earthquake Engineering & Structural Dynamics, 18, 785–797, doi: 10.1002/eqe.4290180604.

[12] Cesca, S., F. Grigoli, S. Heimann, A. Gonzalez, E. Buforn, S. Maghsoudi, E. Blanch, and T. Dahm, 2014, The 2013 September-October seismic sequence offshore Spain: A case of seismicity triggered by gas injection?: Geophysical Journal International, 198, 941–953, doi: 10.1093/gji/ggu172.

[13] Cochard, A., H. Igel, B. Schuberth, W. Suryanto, A. Velikoseltsev, U. Schreiber, J. Wassermann, F. Scherbaum, and D. Vollmer, 2006, Rotational motions in seismology: Theory, observation, simulation, in R. Teisseyre, E. Majewski, and M. Takeo, eds., Earthquake source asymmetry, structural media and rotation effects: Springer Verlag Berlin/Heidelberg, 391–411.

[14] Cowles, C. S., 1981, Rotational geophone: U.S. Patent 4,446,541.
[15] Dahm, T., S. Cesca, S. Hainzl, T. Braun, and F. Krüger, 2015, Discrimination between induced, triggered, and natural earthquakes close to hydrocarbon reservoirs: A probabilistic approach based on the modeling of depletion-induced stress changes and seismological source parameters: Journal of Geophysical Research: Solid Earth, 120, 2491–2509, doi: 10.1002/2014JB011778.
[16] D'Alessandro, A., and G. D'Anna, 2014, Retrieval of ocean bottom and downhole Seismic sensors orientation using integrated MEMS gyroscope and direct rotation measurements: Advances in Geosciences, 40, 11–17,doi: 10.5194/adgeo-40-11-2014.
[17] Dellinger, J., R. Clarke, and P. Gutowski, 2001, Horizontal vector infidelity correction by general linear transform:71st Annual International Meeting, SEG, Expanded Abstracts, 865–868.
[18] de Ridder, S. A. L., and B. L. Biondi, 2015, Near-surface Scholte wave velocities at Ekofisk from short noise recordings by seismic noise gradiometry: Seismic noise gradiometry: Geophysical Research Letters, 42, 7031–7038, doi: 10.1002/2015GL065027.
[19] de Ridder, S. A. L., and A. Curtis, 2017, Seismic gradiometry using ambient seismic noise in an anisotropic earth: Geophysical Journal International, 209, 1168–1179, doi: 10.1093/gji/ggx073.
[20] de Ridder, S. A. L., and J. R. Maddison, 2017, Wave field inversion of ambient seismic noise: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.201701687.
[21] Donner, S., M. Bernauer, and H. Igel, 2016, Inversion for seismic moment tensors combining translational and rotational ground motions: Geophysical Journal International, 207, 562–570, doi: 10.1093/gji/ggw298.
[22] Donner, S., M. Bernauer, M. Reinwald, C. Hadziioannou, and H. Igel, 2017c, Improved source inversion from joint measurements of translational and rotational ground motions: AGU Fall Meeting, Abstract S44B–04.
[23] Donner, S., C. Hadziioannou, and H. Igel, 2017a, Retrieval of the seismic moment tensor from joint measurements of translational and rotational ground motions: Sparse networks and single stations, in S. D'Amico, ed., Moment tensor solutions — A useful tool for seismotectonics: Springer Verlag.
[24] Donner, S., R. Joshi, C. Hadziioannou, C. Nunn, M. van Driel, C. Schmelzbach, J. M. Wassermann, and H. Igel,2017b, Benefits of rotational ground motions for planetary seismology: AGU Fall Meeting, Abstract DI21A– 0384.
[25] Edme, P., M. Daly, E. Muyzert, and E. Kragh, 2013, Side scattered noise attenuation using rotation data: 75th Annual International Conference and Exhibition Incorporating SPE EUROPEC 2013, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20130306.
[26] Edme, P., and E. Muyzert, 2013, Rotational data measurement: 75th Annual International Conference and Exhibition Incorporating SPE EUROPEC 2013, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20130382.
[27] Edme,P., and S. Yuan,2016, Local dispersion curve estimation from seismic ambient noise using spatial gradients: Interpretation, 4, no. 3, SJ17–SJ27, doi: 10.1190/INT-2016-0003.1. Eentec, 2017, www.eentec.com, accessed 29 March 2018.
[28] Egorov, E. V., I. V. Egorov, and V. M. Agafonov, 2015, Self-noise of the MET angular motion seismic sensors: Journal of Sensors, 2015, 1–5, doi: 10.1155/2015/512645.
[29] Evans, J. R., F. Followill, C. R. Hutt, R. P. Kromer, R. L. Nigbor, A. T. Ringler, J. M. Steim, and E. Wielandt, 2010, Method for calculating self-noise spectra and operating ranges for seismographic inertial sensors and recorders: Seismological Research Letters, 81, 640–646, doi: 10 .1785/gssrl.81.4.640.
[30] Fichtner, A., and H. Igel, 2009, Sensitivity densities for rotational ground motion measurements: Bulletin of the Seismological Society of America, 99, 1302–1314, doi: 10.1785/0120080064.
[31] Gaiser, J. E., 2007, Detector coupling corrections for vector infidelity of multicomponent OBC data: Geophysics, 72, no. 3, V67–V77, doi: 10 .1190/1.2710192.
[32] Garmany, J. D., 1984, The recovery of true particle motion from three-component ocean bottom seismometer data: Journal of Geophysical Research, 89, 9245–9252, doi: 10.1029/JB089iB11p09245.
[33] Hadziioannou, C., P. Gaebler, U. Schreiber, J. Wassermann, and H. Igel, 2012, Examining ambient noise using collocated measurements of rotational and translational motion: Journal of Seismology, 16, 787–796, doi: 10.1007/s10950-012-9288-5.
[34] Horne, S., 2003, Fracture characterization from walk-around VSPs: Geophysical Prospection, 51, 493–499, doi: 10.1046/j.1365-2478.2003.00391.x.
[35] Horne, S., 2012, Methods and systems for borehole seismic: U.S. Patent 8,255,164.
[36] Huang, B.-S., 2003, Ground rotational motions of the 1999 Chi-Chi, Taiwan earthquake as inferred from dense array observations: Geophysical Research Letters, 30, 1307, doi: 10.1029/2002GL015157.
[37] Huang, H., V. Agafonov, and H. Yu, 2013, Molecular electric transducers as motion sensors: A review: Sensors, 13, 4581–4597, doi: 10.3390/s130404581.

[38] Igel, H., M. Bernauer, J. Wassermann, and K. U. Schreiber, 2015, Rotational seismology: Theory, instrumentation, observations, applications, in R. A. Meyers, ed., Encyclopedia of complexity and systems science: Springer Verlag Berlin/Heidelberg, 1–26.
[39] Igel, H., A. Cochard, J. Wassermann, A. Flaws, U. Schreiber, A. Velikoseltsev, and N. Pham Dinh, 2007, Broad-band observations of earthquake induced rotational ground motions: Geophysical Journal International, 168, 182–196, doi: 10.1111/j.1365-246X.2006.03146.x.
[40] Igel, H., M.-F. Nader, D. Kurrle, A.-M. Ferreira, J. Wassermann, and K. U. Schreiber, 2011, Observations of Earth's toroidal free oscillations with a rotation sensor: The 2011 magnitude 9.0 Tohoku-Oki earthquake: Geophysical Research Letters, 38, doi: 10.1029/2011GL049045.
[41] Igel, H., U. Schreiber, A. Flaws, B. Schuberth, A. Velikoseltsev, and A. Cochard, 2005, Rotational motions induced by the M8.1 Tokachi-oki earthquake, September 25, 2003: Geophysical Research Letters, 32, L08309, doi: 10.1029/2004GL022336.
[42] IRIS-DMC database, 2017, IRIS Data Management Centre – Meta Data Aggregator, http://ds.iris.edu/mda/GS/OKR1, accessed 29 June 2017.
[43] IWGoRS, 2017, International Working Group on Rotational Seismology (IWGoRS), http://www.rotational-seismology.org/, accessed 29 March 2017.
[44] Jaroszewicz, L. R., Z. Krajewski, L. Solarz, and R. Teisseyer, 2006, Application of the fibre optic Sagnac interferometer in the investigation of seismic rotational waves: Measurement Science and Technology, 17, 1186–1193, doi: 10.1088/0957-0233/17/5/S42.
[45] Jaroszewicz, L. R., Z. Krajewski, and K. P. Teisseyre, 2012, Usefulness of AFORS-autonomous fibre-optic rotational seismograph for investigation of rotational phenomena: Journal of Seismology, 16, 573–586, doi: 10.1007/s10950-011-9258-3.
[46] Jost, M. L., and R. B. Herrmann, 1989, A student's guide to and review of moment tensors: Seismological Research Letters, 60, 37–57.
[47] Khan, A., M. van Driel, M. Böse, D. Giardini, S. Ceylan, J. Yan, J. Clinton, F. Euchner, P. Lognonné, N. Murdoch, D. Mimoun, M. Panning, M. Knapmeyer, and W. B. Banerdt, 2016, Single-station and single-event marsquake location and inversion for structure using synthetic Martian waveforms: Physics of the Earth and Planetary Interiors, 258, 28–42, doi: 10.1016/j.pepi.2016.05.017.
[48] Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann, 2015, ObsPy: A bridge for seismology into the scientific Python ecosystem: Computational Science and Discovery, 8, 014003, doi: 10.1088/1749-4699/8/1/014003.
[49] Langston, C. A., 2006, Explosion source strong ground motions in the Mississippi embayment: Bulletin of the Seismological Society of America, 96, 1038–1054, doi: 10.1785/0120050105.
[50] Langston, C. A., 2007a, Wave gradiometry in the time domain: Bulletin of the Seismological Society of America, 97, 926–933, doi: 10.1785/0120060152.
[51] Langston, C. A., 2007b, Spatial gradient analysis for linear seismic arrays: Bulletin of the Seismological Society of America, 97, 265–280, doi: 10.1785/0120060100.
[52] Langston, C. A., 2007c, Wave gradiometry in two dimensions: Bulletin of the Seismological Society of America, 97, 401–416, doi: 10.1785/0120060138.
[53] Lee, W. H. K., M. Celebi, M. I. Todorovska, and H. Igel, 2009, Introduction to the special issue on rotational seismology and engineering applications: Bulletin of the Seismological Society of America, 99, 945–957, doi: 10.1785/0120080344.
[54] Lefèvre, H. C., 2014, The fiber-optic gyroscope, 2nd ed.: Artech House.
[55] Li, H.-N., L.-Y. Sun, and S.-Y. Wang, 2004, Improved approach for obtaining rotational components of seismic motion: Nuclear Engineering and Design, 232, 131–137, doi: 10.1016/j.nucengdes.2004.05.002.
[56] Li, Z., and M. Van der Baan, 2017a, Tutorial on rotational seismology and its applications in exploration geophysics: Geophysics, 82, no. 5, W17–W30, doi: 10.1190/geo2016-0497.1.
[57] Li, Z., and M. Van der Baan, 2017b, Elastic passivesource localization using rotational motion: Geophysical Journal International, 211, 1206–1222, doi: 10.1093/gji/ggx364.
[58] Liang, C., and C. A. Langston, 2009, Wave gradiometry for USArray: Rayleigh waves: Journal of Geophysical Research, 114, B02308, doi: 10.1029/2008JB005918.
[59] Lin, C.-J., H.-P. Huang, N. D. Pham, C.-C. Liu, W.-C. Chi, and W. H. K. Lee, 2011, Rotational motions for teleseismic surfacewaves: Geophysical Research Letters, 38, L15301, doi: 10.1029/2011GL047959.
[60] Lindner, F., J. Wassermann, M. C. Schmidt-Aursch, K. U. Schreiber, and H. Igel, 2016, Seafloor ground rotation observations: Potential for improving signal-to-noise ratio on horizontal OBS components: Seismological Research Letters, 88, 32–38, doi: 10.1785/0220160051.

[61] Liu, H., and W. T. Pike, 2016, A micromachined angular-acceleration sensor for geophysical applications: Applied Physics Letters, 109, 173506, doi:10.1063/1.4966547.
[62] Liu,Y.,andE.Holt,2015,Wave gradiometry and its link with Helmholtz equation solutions applied to USArray in the eastern U.S.: Journal of Geophysical Research: Solid Earth, 120, 5717–5746, doi: 10.1002/2015JB011982.
[63] Maeda, T., K. Nishida, R. Takagi, and K. Obara, 2016, Reconstruction of a 2D seismic wavefield by seismic gradiometry: Progress in Earth and Planetary Science, 33, 31, doi: 10.1186/s40645-016-0107-4.
[64] McLeod, D. P., G. E. Stedman, T. H. Webb, and U. Schreiber, 1998, Comparison of standard and ring laser rotational seismograms: Bulletin of the Seismological Society of America, 88, 1495–1503.
[65] Muyzert, E., A. Kashubin, E. Kragh, and P. Edme, 2012, Land seismic data acquisition using rotation sensors: 74th Annual International Conference and Exhibition Incorporating EUROPEC 2012, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20148780.
[66] Nader, M. F., H. Igel, A. M. G. Ferreira, D. Al-Attar, J. Wassermann, and K. U. Schreiber, 2014, Normal mode coupling observations with a rotation sensor: Geophysical Journal International, 201, 1482–1490, doi: 10.1093/gji/ggv082.
[67] Nigbor, R. L., J. R. Evans, and C. R. Hutt, 2009, Laboratory and field testing of commercial rotational seismometers: Bulletin of the Seismological Society of America, 99, 1215–1227, doi: 10.1785/0120080247.
[68] Pham, N. D., H. Igel, J. de la Puente, M. Käser, and M.A. Schoenberg, 2010, Rotational motions in homogeneous anisotropic elastic media: Geophysics, 75, no. 5, D47–D56, doi: 10.1190/1.3479489.
[69] Pierson, B., D. Laughlin, and R. Brune, 2016, Advances in rotational seismic measurements: 86th Annual International Meeting, SEG, Expanded Abstracts, 2263–2267, doi: 10.1190/segam2016-13956985.1.
[70] Reinwald, M., M. Bernauer, H. Igel, and S. Donner, 2016, Improved finite-source inversion through joint measurements of rotational and translational ground motions: A numerical study: Solid Earth, 7, 1467–1477, doi: 10.5194/se-7-1467-2016.
[71] Robertsson, J. O. A., and A. Curtis, 2002, Wavefield separation using densely deployed three-component single-sensor groups in land surface-seismic recordings: Geophysics, 67, 1624–1633, doi: 10.1190/1.1512809.
[72] Robertsson, J. O. A., I. Moore, M. Vassallo, K. Özdemir, D.-J. van Manen, and A. Özbek, 2008, On the use of multicomponent streamer recordings for reconstruction of pressure wavefields in the crossline direction: Geophysics, 73, no. 5, A45–A49, doi: 10.1190/1.2953338.
[73] Robertsson, J. O. A., and E. Muyzert, 1999, Wavefield separation using a volume distribution of three component recordings: Geophysical Research Letters, 26, 2821–2824, doi: 10.1029/1999GL010472.
[74] Rudziński, L., S. Cesca, and G. Lizurek, 2016, Complex Rupture Process of the 19 March 2013, Rudna Mine (Poland) induced seismic event and collapse in the light of local and regional moment tensor inversion: Seismological Research Letters, 87, 274–284, doi: 10.1785/0220150150.
[75] Salvermoser, J., C. Hadziioannou, S. Hable, L. Krischer, B. Chow, J. Wassermann, U. Schreiber, A. Gebauer, and H. Igel, 2017, An event database for rotational seismology: Seismological Research Letters, 88, 935–941, doi: 10.1785/0220160184.
[76] Schmelzbach, C., M. Häusler, D. Sollberger, C. Van Renterghem, and J. O. A. Robertsson, 2017, Spatial wavefield gradient data in seismic exploration: EAGE/DGG Workshop on Fibre Optics Technology in Geophysics, doi: 10.3997/2214-4609.201700160.
[77] Schmelzbach, C., D. Sollberger, C. Van Renterghem, M. Häusler, J. O. A. Robertsson, and S. A. Greenhalgh, 2016, Seismic spatial gradient and rotation measurements as new observables in land seismic exploration: EGU General Assembly, 18, EGU2016–3787.
[78] Schreiber, K. U., J. Hautmann, A. Velikoseltsev, J. Wassermann, H. Igel, F. Otero, F. Vernon, and P. R. Wells, 2009, Ring laser measurements of ground rotations for seismology: Bulletin of the Seismological Society of America, 99, 1190–1198, doi: 10.1785/0120080171.
[79] Schreiber, K. U., G. E. Stedman, H. Igel, and A. Flaws, 2006, Ring laser gyroscopes as rotation sensors for seismic wave studies, in R. Teisseyre, E. Majewski, and M. Takeo, eds., Earthquake source asymmetry, structural media and rotation effects: Springer Verlag Berlin/Heidelberg, 377–390.
[80] Segall, P., 2010, Earthquake and volcano deformation: Princeton University Press.
[81] Sileny, J., 2004, Regional moment tensor uncertainty due to mismodeling of the crust: Tectonophysics, 383, 133–147, doi: 10.1016/j.tecto.2003.12.007.
[82] Sollberger, D., S. A. Greenhalgh, C. Schmelzbach, C. Van Renterghem, and J. O. A. Robertsson,2018,6-Cpolarisationanalysisusingpoint-measurements of translational and rotational ground-motion: Theory and applications:Geophysical Journal International, 213, 77–97, doi: 10.1093/gji/ggx542.
[83] Sollberger, D., C. Schmelzbach, J. O. A. Robertsson, S. A. Greenhalgh, Y. Nakamura, and A. Khan, 2016b, The shallow elastic structure of the lunar crust: New insights from seismic wavefield gradient analysis: Seismic gradient analysis on the moon: Geophysical Research Letters, 43, 10078–10087, doi: 10.1002/2016GL070883.

[84] Sollberger, D., C. Schmelzbach, J. O. A. Robertsson, H. Maurer, and S. A. Greenhalgh, 2015, Geophone coupling corrections for land-seismic point receiver acquisition: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.201412593.
[85] Sollberger, D., C. Schmelzbach, C. Van Renterghem, J. O. A. Robertsson, and S. A. Greenhalgh, 2016a, Single-component elastic wavefield separation at the free surface using source- and receiver-side gradients: 86th Annual International Meeting, SEG, Expanded Abstracts, 2268–2273, doi: 10.1190/segam2016-13842637.1.
[86] Spudich, P., and J. B. Fletcher, 2009, Software for inference of dynamic ground strains and rotations and their errors from short baseline array observations of ground motions: Bulletin of the Seismological Society of America, 99, 1480–1482, doi: 10.1785/0120080230.
[87] Spudich, P., L. K. Steck, M. Hellweg, J. B. Fletcher, and L. M. Baker, 1995, Transient stresses at Parkfield, California, produced by the M 7.4 Landers earthquake of June 28, 1992: Observations from the UPSAR dense seismograph array: Journal of Geophysical Research: Solid Earth, 100, 675–690, doi: 10.1029/94JB02477.
[88] Stupazzini, M., J. de la Puente, C. Smerzini, M. Kaser, H. Igel, and A. Castellani, 2009, Study of rotational ground motion in the near-field region: Bulletin of the Seismological Society of America, 99, 1271–1286, doi: 10.1785/0120080153.
[89] Suryanto, W., H. Igel, J. Wassermann, A. Cochard, B. Schuberth, D. Vollmer, F. Scherbaum, U. Schreiber, and A. Velikoseltsev, 2006, First comparison of array-derived rotational ground motions with direct ring laser measurements: Bulletin of the Seismological Society of America, 96, 2059–2071, doi: 10.1785/0120060004.
[90] Sutton, G. H., F. K. Duennebier, and B. Iwatake, 1981, Coupling of ocean bottom seismometer to soft bottom: Marine Geophysical Researches, 5, 35–51, doi: 10.1007/BF00310310.
[91] Tanimoto, T., C. Hadziioannou, H. Igel, J. Wassermann, U. Schreiber, and A. Gebauer, 2015, Estimate of Rayleigh-to-Love wave ratio in the secondary microseism by co-located ring laser and seismograph: Geophysical Research Letters, 42, 2650–2655, doi: 10.1002/2015GL063637.
[92] van Driel, M., J. Wassermann, M. F. Nader, B. S. A. Schuberth, and H. Igel, 2012, Strain rotation coupling and its implications on the measurement of rotational ground motions: Journal of Seismology, 16, 657–668, doi: 10.1007/s10950-012-9296-5.
[93] Van Renterghem, C., C. Schmelzbach, and J. O. A. Robertsson, 2016, Wave-field separation of multicomponent land seismic data using spatial wave-field gradients: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.201601357.
[94] Van Renterghem, C., C. Schmelzbach, D. Sollberger, and J. O. A. Robertsson, 2017, Land and ocean bottom spatial gradient-based seismic wave-field separation: 87th Annual International Meeting, SEG, Expanded Abstracts, 2539–2544, doi: 10.1190/segam2017-17704566.1.
[95] Van Renterghem, C., C. Schmelzbach, D. Sollberger, and J.O. A. Robertsson, 2018, Spatial wavefield gradient-based seismic wavefield separation: Geophysical Journal International, 212, 1588–1599, doi: 10.1093/gji/ggx499.
[96] Vassallo, M., A. Özbek, K. Özdemir, and K. Eggenberger, 2010, Crossline wavefield reconstruction from multicomponent streamer data: Part 1: Multichannel interpolation by matching pursuit (MIMAP) using pressure and its crossline gradient: Geophysics, 75, no. 6, WB53–WB67, doi: 10.1190/1.3496958.
[97] Velikoseltsev, A., K. U. Schreiber, A. Yankovsky, J.-P. R. Wells, A. Boronachin, and A. Tkachenko, 2012, On the application of fiber optic gyroscopes for detection of seismic rotations: Journal of Seismology, 16, 623–637, doi: 10.1007/s10950-012-9282-y.
[98] Wassermann, J., S. Lehndorfer, H. Igel, and U. Schreiber, 2009, Performance test of a commercial rotational motions sensor: Bulletin of the Seismological Society of America, 99, 1449–1456, doi: 10.1785/0120080157.
[99] Wassermann, J., A. Wietek, C. Hadziioannou, and H. Igel, 2016, Toward a single-station approach for microzonation: Using vertical rotation rate to estimate love-wave dispersion curves and direction finding: Bulletin of the Seismological Society of America, 106, 1316–1330, doi: 10.1785/0120150250.
[100] Woelz, S., W. Rabbel, and C. Mueller, 2009, Shear waves in near surface 3D media — SH-wavefield separation, refraction time migration and tomography: Journal of Applied Geophysics, 68, 104–116, doi:10.1016/j.jappgeo . 2008.11.004.

深层超深层油气勘探技术路线及针对性地震波成像处理方法技术

王华忠[1]，曾同生[1,2]，吴成梁[1]，许荣伟[1]，项健[1]

①同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

②中国石油勘探开发研究院，北京，100083

【摘要】以塔里木盆地塔河和顺北等探区为代表的深层油气勘探开发取得了许多突破，在 7000 米以深的深层和超深层层系钻遇了高产油气藏。但关于深层超深层油气勘探的关键技术还有待开展系统的研究和总结。虽然中浅层和深层超深层油气勘探的目标本质上都是优质油气藏，但是，在深层超深层油气藏的勘探中，波与介质相互作用的关系变得更为复杂，因而需要变革地震激发、接收及炮检排列方式；与此同时关于深层超深层油气地震勘探的理念和方法都要提升到一个新的层次。以对来自地下介质中同一个反（散）射点不同炮检对的反（散）射地震子波的同相位叠加为深层超深层地震波成像的核心理念，首先分析了深层超深层油气藏探测的物理机制的变化，然后分析这种物理机制的变化对地震数据采集技术提出的新挑战，并据此给出了针对深层超深层油气藏探测的地震数据采集方法技术；针对地震数据预处理，提出了强噪声弱信号情形下的噪声压制方法技术；针对地震波成像处理，提出了关于中层、深层和超深层地质体成像所需的建立速度模型的方法技术体系，最后以保真和高分辨率成像道集为中心，提出了叠前数据域与叠前成像域数据处理的等价性，提出针对深层超深层油气藏高精度地震波成像的合理方法技术组合，使成像结果尽可能满足深层超深层油气藏刻画与评价的需求。

【关键词】深层超深层油气地震勘探，波与介质相互作用的物理机制，保真和高分辨率成像道集，勘探地震介质系统的线性性，同相位叠加，高精度速度建模技术路线，高精度地震波成像技术路线

1 引言

世界范围内，油气地震勘探都在向新的复杂油气勘探场景和新的勘探领域转移，因为简单场景下的油气地震勘探工作已经做得很深入了[1][2]。复杂的油气勘探场景包括：复杂地表（山前带、前山带、黄土塬等地表情形）、复杂构造（推覆构造、逆掩构造等）、隐蔽油气藏、岩性油气藏、非常规油气藏等；拓展出的新的油气勘探领域包括深层超深层、深水、高原、极地等。所有这些变化都必然带来油气勘探思路和勘探技术的变革。

近年来，随着中浅层油气资源勘探程度的增加以及对深部介质地质认识的提高，深层与超深层油气藏勘探受到了国内外业界越来越多的关注[3][4]。回望历史，人们对深层油气的勘探可追溯至 20 世纪 50 年代。1956 年，在美国阿纳达科盆地 Carter-Knox 气田埋深 4663m 的中奥陶统 Simpson 群碳酸盐岩内发现了世界上第一个深层气藏。之后，伴随着深层钻井和完井等技术的突破，油气勘探向超深层（埋深超过 7500m）领域迈进。1977 年，在阿纳达科盆地 Mills Ranch 气田 8097m 深处的寒武系-奥陶系 Arbuckle 群白云岩内发现了气藏。1984 年，在意大利 Villifortuna-Trecate 油田 6400m 深处的三叠系白云岩内发现了油藏。自 1980 年起，深层油气勘探由陆上逐渐向海域拓展。如 1980 年，在阿拉伯盆地 Fateh 气田埋深 4500m 的二叠系 Khuff 组灰岩内发现了气藏。目前，深层油气勘探在墨西哥湾、巴西东部、西非等深水和超深水区已取得重大突破。在国内，仅仅在塔里木盆地，中国石油（CNPC）公司一家就进行了 350 余口井深 5000m～6000m、100 余口井深 6000m～7000m、多于 7 口井深大于 7000m 的油气钻探活动，并且取得了很好的油气勘探效果[5]。

很显然，进行深层与超深层油气地震勘探具有重大意义，但深层超深层油气地震勘探面临以下挑战：深层油气成藏理论不够成熟；深层油气勘探开发难度大；投资回报率低，与此同时还存在以下要解决的关键问题：（1）深层油气藏地质成藏特征缺乏规律性认识；（2）与油气藏精确定位、描述与评价相关的针对性处理方法与技术不够完善；（3）钻/完井等工程技术的进一步改善。

深层超深层与浅中层油气地震勘探的差异究竟体现在何处？本文认为主要体现在以下三个方面：（1）深层超深层油气勘探情形下波与介质相互作用的关系更复杂，接收到的波场也更复杂；（2）深层超深层油气勘探情形下地震波对目标储层的照明更不均匀、照明能量更弱；（3）深层超深层油气勘探情形下储层的几何形态及其中的岩石物性变化更复杂、更难描述。

可以肯定的是：深层超深层油气地震勘探要解决的首要问题一定是满足含油气储层识别与描述要求的、精确的地震波成像，为深层超深层油气地质认识、油藏描述甚至钻井工程提供足够精确的地下介质变化信息。

深层超深层油气地震勘探中地震波成像面临的核心问题包括：（1）目标地质体照明角度变窄；（2）主频降低、频带变窄；（3）深层反射信号弱，信噪比低；（4）介质的各向异性特征明显。

据此可以得出如下判断：由于深层超深层油气地震勘探中波与介质之间的关系更复杂，必然导致深层超深层油气地震勘探在地震数据采集、处理和成像方面需要有针对性的思想和技术。针对性的方法技术主要体现在如下三个方面：（1）满足深层超深层地震波精确成像要求的地震数据采集技术；（2）满足对来自地下同一成像点、不同炮检对反（散）射子波同相叠加要求的背景速度和各向异性参数估计及建模方法与技术；（3）满足深层超深层复杂介质地震波传播规律的叠前深度偏移成像方法与技术。尤其是针对深层超深层油气藏的地震数据采集技术一定要具有先导性。相干或非相干噪声强于深层弱反（散）射信号情形下的地震数据去噪也需要发展针对性的方法技术。

基于上述认识，本文以对来自地下介质中同一个反（散）射点不同炮检对的反（散）射地震子波的同相位叠加为深层超深层地震波成像的核心理念，首先分析了针对深层超深层油气藏探测的物理机制的变化，然后分析了这种物理机制的变化对地震数据采集技术提出的新挑战，并据此给出了针对深层超深层油气藏探测的地震数据采集方法技术；针对地震数据预处理，提出了强噪声弱信号情形下的噪声压制方法技术；针对地震波成像，提出了针对中层、深层超深层地质体成像所需的速度模型建立方法技术体系；最后以保真和高分辨率成像道集为中心，针对叠前数据域与叠前成像域数据处理的等价性，提出了深层超深层油气藏高精度地震波成像的合理方法技术组合，使得成像结果尽可能满足深层超深层油气藏刻画与评价的需求。

2 深层超深层油气勘探的地震地质问题

油气地震勘探的基本思想、方法和技术是从简单地表条件和横向缓变的沉积构造环境下埋深较浅的油气藏勘探逐步发展起来的。简单地表条件是指地表比较平缓、低降速带的厚度及其速度横向变化不甚剧烈，地表一致性静校正能较好地消除近地表条件的影响，地表一致性振幅和相位校正能较好地调整地震子波的一致性。简单地表条件下，叠前炮集中噪声类型比较简单，特征变化不复杂，去噪处理相对容易满足后续地震波成像的需求。动校正（NMO）+倾角时差校正（DMO）+叠前时间偏移（PSTM）技术组合能够满足地震波成像的基本需求。这种简单情形下的油气地震勘探对地震数据采集没有特别的技术要求。

复杂地表条件+简单沉积构造环境下的油气勘探并不常见。按照地质构造理论，复杂地表条件一定是由于中深层强烈的地质构造运动所导致，而中深层强烈的构造运动一定会同时导致中深层复杂的地质构造，进而导致横向速度的剧烈变化。相应地，相对简单的地表条件+复杂的中深层沉积构造环境是常见的勘探场景，海上油气勘探的许多探区都属于这种情形；渤海湾薄、碎、小的地质异常体和复杂古潜山的勘探也属于此类情形；大部分的陆上沉积盆地内的探区基本上也属于此类情形。复杂地表条件+复杂中深层地质构造环境下的油气勘探是前缘盆地、盆山过渡区典型的勘探场景，这种场景下的地震地质问题可归结为：剧变道间时差+（极）低信噪比数据条件下满足复杂构造成像要求的速度模型建立问题，我国西部各油气盆地边缘山前带探区就是该类型的典型场景。

当前，无论是在东部诸探区还是中西部各油气盆地，油气地震勘探的目的层都在向深层超深层领域推进。深层超深层勘探领域一般分为以下四种勘探情形：（1）简单地表条件+横向缓变的沉积构造环境下的深层超深层油气勘探，譬如塔里木盆地深层超深层勘探；（2）简单地表条件+复杂的中深层沉积构造环境下的深层超深层油气勘探，譬如我国东部及海上深层超深层油气勘探；（3）复杂地表条件+相对简单的中深层地质构造环境下的深层超深层油气勘探；譬如四川盆地涪陵地区天然气探区；（4）复杂地表条件+复杂中深层地质构造环境下的深层超深层油气勘探；譬如四川盆地龙门山诸探区和天山南缘库车诸探区等。可以看到，第四种勘探情形是最为复杂的情形。

地震波成像的本质可以归结为：对来自地下同一反（散）射点的、不同炮检对的反（散）地震子波实现同相位叠加。人工震源激发产生的地震波在实际地下介质中传播时，与地下介质相互作用后携带了地下介质变化的信息传播到地表，在地表被检波器接收后获得反（散）射波波场数据。当然，检波器也会接收到诸如折射波、多次波和干扰噪声等其他类型的波场数据。对于勘探地震而言，地下介质系统基本是线性系统，激发地震子波与接收地震子波基本上是相干子波。叠前道集中同相轴上反（散）射子波能产生较好的嵌套效应就证实了这种观点。这也是地震波成像的基本物理逻辑。

从地震波成像的基本物理角度分析，中深层地质目标的地震波成像和深层超深层地质目标的地震波成像并没有本质上的区别，都是对来自地下同一反（散）射点的、不同炮检对的反（散）地震子波实现同相位叠加。但是，对深层超深层地质目标的地震波成像而言，由于目标地质体埋深不同、岩石物性受改造的程度不同、与介质的相互作用不同以及地震波对深层目标地质体的照明情况不同等因素，导致了深层超深层油气地震勘探具有不同于中浅层油气地震勘探的新特征。从根本上来说，在温度和压力的共同作用下，

深层超深层岩石介质的物理化学性质发生了显著变化。这种变化导致岩石介质层内与层间、以及断裂带和裂缝带内外的弹性参数呈现出与中浅层地层岩石介质显著不同的特征。这一现象正是赋予深层超深层油气勘探新内涵的根本原因。从物理角度看，首先是波与介质的相互作用变得更加复杂；其次，噪声与信号的关系发生了新的变化，使得深层超深层的波场特征更复杂化，此外，照明角度变得更狭窄，传统的地震勘探方法和技术在深层超深层的适用性受到挑战，导致深层超深层地震勘探的思想和方法技术需要革新和优化[6]。

深层超深层油气地震勘探中波与介质相互作用的特点可以总结为：（1）波传播路径更长，复杂介质对波传播的影响更严重，譬如吸收衰减、散射、透射损失、转换波损失、多次反射等；（2）深层超深层地质异常体阻抗跃变更小，球面扩散与吸收衰减更严重，导致反射（绕射）信号更弱、信噪比更低；（3）近地表相关随机噪声不随时间减弱，深层超深层弱反（散）射信号的信噪比关系显著不同于中浅层；（4）深层超深层弱反（散）射信号更容易受浅中层的层间多次波污染；（5）照明角度更窄、照明强度更弱；（6）高频损失更严重，主频更低，导致成像分辨率更低；（7）随着偏移距和方位角的增加，介质各向异性（至少VTI）更明显。

所有这些问题促使深层超深层油气勘探的理念、方法和技术需要进行相应的调整与变革。而其中的首要任务是实现能满足深层超深层储层识别与描述需求的精确地震成像。尤其是先要解决深层超深层地质目标的构造成像问题。

3 针对深层超深层油气勘探的地震波成像理念

自开展油气地震勘探以来，地震波成像一直是其核心问题之一。到现在，随着各种地震波成像技术的发展，可以认为地震波成像技术最终归结为全波形反演（FWI）。为什么会这样？原因在于，FWI 是 Bayes 参数估计理论在地震勘探领域的一个应用实例，而 Bayes 参数估计是所有由可观测量估计与可观测量相关的潜变量问题的理论框架[7]。因此，Bayes 参数估计是地震波成像（乃至所有广义遥感成像）的最高理论框架。尽管目前在工业界的实际应用中，地震数据成像处理方法尚未普遍实现基于 FWI 的处理流程，但地震波成像理论研究的重点已经集中于 FWI 方法及其技术路径。具体来说，研究的核心在于探索如何从不同的角度实现 FWI 在实际地震数据处理中的实用化。

Bayes 参数估计的根本思想是：假设已知实测数据（可观测量）与地下介质弹性参数（潜变量）之间的先验预测关系（一般由概率密度函数表达这种预测关系的不确定性）和待估计参数（潜变量）满足的先验概率分布，上述两种先验分布决定了已知实测数据与待估计参数之间的条件后验概率密度。该条件后验概率密度包含了待估计参数的所有统计特征。一般地，将条件后验概率密度的期望作为待估计参数的结果，其方差用来评价待估计参数的估计精度。

首先，有实测数据 $\mathbf{d}^{obs}$ 与地下介质弹性参数 $\mathbf{m}$ 之间的先验预测关系，

$$\mathcal{M}：\ \mathbf{d}^{obs}=G\left(\mathbf{m}\right)+\boldsymbol{\varepsilon} \tag{1}$$

其中，模型 $\mathcal{M}$（即可观测量与相关的潜变量之间的关系表达式，勘探地震中即各种形式的波动方程）包含了实测数据与地下介质弹性参数之间的先验预测关系的不确定性，用来表达。不确定性既由数据观测过程中的因素引起，也由正问题 $\mathbf{d}^{cal}=G\left(\mathbf{m}\right)$ 的不准确引起。ε 是正问题不能完全解释实测数据导致的误差。正是 ε 的随机性刻画了实测数据与地下介质弹性参数之间的先验预测关系的不确定性。一般地，假设某种概率密度函数刻画了这种不确定性，并假设该概率密度函数是已知的。

单纯地依赖实测数据与地下介质弹性参数之间的先验预测关系的不确定性进行参数估计是最大似然估计（ Maximum Likelihood Estimation （MLE））学派的思想。但是，Bayes 参数估计学派认为，仅仅依赖实测数据与地下介质弹性参数之间的先验预测关系不可能得到可靠的估计结果，必须引入待估计参数的先验信息。该先验信息当然不可能是完全准确的，不确定性用某种先验概率密度函数来描述，可写为 $p\left(\mathbf{m}\middle|\mathcal{M}\right)$。很显然，参数 $\mathbf{m}$ 是正问题 $\mathbf{d}^{cal}=G\left(\mathbf{m}\right)$ 中包含的参数，与模型 $\mathcal{M}$ 是相关联的。

根据上述两种先验概率密度函数，依据 Bayes 公式，可以定义如下条件后验概率密度：

$$p\left(\mathbf{m}\middle|\mathbf{d}^{obs};\mathcal{M}\right)=\frac{p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)p\left(\mathbf{m}\middle|\mathcal{M}\right)}{p\left(\mathbf{d}^{obs}\middle|\mathcal{M}\right)} \tag{2}$$

其中，$p\left(\mathbf{d}^{obs}\middle|\mathcal{M}\right)$ 是表示实测数据本身满足的不确定的概率密度函数。

概率论指出：条件后验概率密度 $p\left(\mathbf{m}\middle|\mathbf{d}^{obs};\mathcal{M}\right)$ 包含了待估计参数 $\mathbf{m}$ 的所有统计特征。期望（一阶统计量）可以作为待估计参数 $\mathbf{m}$ 的估计结果；方差（二阶统计量）视为待估计参数 $\mathbf{m}$ 的估计精度评价。由于勘探地震中，待估计参数 $\mathbf{m}$ 有非常高的维度，通过积分求条件后验概率密度规定的期望和方差基本是不可能的。

因此，条件后验概率密度的最大化准则（Maximum A Posteriori Estimation （MAP））成了常用的参数估计准则。即

$$\hat{\mathbf{m}}=\arg\min_{\mathbf{m}}p\left(\mathbf{m}\middle|\mathbf{d}^{obs};\mathcal{M}\right)=\arg\min_{\mathbf{m}}\left\{p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)p\left(\mathbf{m}\middle|\mathcal{M}\right)\right\} \tag{3}$$

一般地，假设两个先验概率密度函数 $p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)$ 和 $p\left(\mathbf{m}\middle|\mathcal{M}\right)$ 都是指数型的。已知指数型函数的对数不改变极大值的分布，因此，（3）式被写为：

$$\hat{\mathbf{m}}=\arg\min_{\mathbf{m}}\left\{-\ln p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)-\ln p\left(\mathbf{m}\middle|\mathcal{M}\right)\right\} \tag{4}$$

显然，（4）式逐渐转化为最优化问题。解析数学家定义如下形式的最优化问题：

$$\mathcal{J}\left(\mathbf{m}\right)=Q\left(\mathbf{d}^{obs},G\left(\mathbf{m}\right)\right)+\lambda\Omega\left(\mathbf{m}\right) \tag{5}$$

其中：$Q\left(\mathbf{d}^{obs},G\left(\mathbf{m}\right)\right)$ 为数据偏差项，度量正问题 $\mathbf{d}^{cal}=G\left(\mathbf{m}\right)$ 预测实测数据 $\mathbf{d}^{obs}$ 的误差；λ 是平衡数据偏差项和正则化项作用大小的超参；$\Omega\left(\mathbf{m}\right)$ 定义待估计参数的先验信息。

很显然，若定义（5）式中的对应项为：

$$Q\left(\mathbf{d}^{obs},G\left(\mathbf{m}\right)\right)=-\ln p\left(\mathbf{g}\middle|\mathbf{m};\mathcal{M}\right) \tag{6a}$$

$$\Omega\left(\mathbf{m}\right)=-\ln p\left(\mathbf{m}\middle|\mathcal{M}\right) \tag{6b}$$

可以看出，数学分析学家定义的参数优化反演问题（一般地，称 Tikhonov 正则化问题）与两个先验概率密度函数假设为指数分布的 MAP 估计问题是一致的。

更进一步地，在线性 Gauss 假设下，即存在实测数据与地下介质弹性参数之间的线性预测 $\mathbf{d}^{obs}=\mathbf{d}^{cal}+\varepsilon=G\mathbf{m}+\varepsilon$，并有如下的两个先验概率密度函数满足 Gauss 分布的假设，

对噪声的假设：

$$\varepsilon\sim N\left(\mathbf{0},\sigma_{\varepsilon}^{2}\mathbf{I}\right)\ \text{即}\ p\left(\mathbf{d}^{obs}\middle|\mathbf{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left\|\mathbf{d}^{obs}-G\mathbf{m}\right\|^{2}\right] \tag{7a}$$

和对 $\mathbf{m}$ 的假设：

$$\mathbf{m}\sim N\left(\mathbf{0},\sigma_{\mathbf{m}}^{2}\mathbf{I}\right)\text{即}\ p\left(\mathbf{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\mathbf{m}}^{2}}\left\|\mathbf{m}\right\|^{2}\right] \tag{7b}$$

则对应的后验概率密度为

$$p\left(\mathbf{m}\middle|\mathbf{d}^{obs}\right)\propto\exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left\|\mathbf{d}^{obs}-H\mathbf{m}\right\|^{2}+\frac{\sigma_{\varepsilon}^{2}}{\sigma_{\mathbf{m}}^{2}}\left\|\mathbf{m}\right\|^{2}\right] \tag{8}$$

类似地，有 MAP 估计下的反演解

$$\hat{\mathbf{m}}=\arg\min_{\mathbf{m}}\left\{p\left(\mathbf{m}\middle|\mathbf{d}^{obs}\right)\right\}=\arg\min_{\mathbf{m}}\left\{\mathcal{J}\left(\mathbf{m}\right)\right\} \tag{9}$$

其中，误差泛函定义为

$$\mathcal{J}\left(\mathbf{m}\right)=-\frac{1}{2}\left\|\mathbf{d}^{obs}-G\mathbf{m}\right\|_{\sigma_{\varepsilon}^{2}}^{2}+\lambda\left\|\mathbf{m}\right\|^{2} \tag{10}$$

其中，超参 $\lambda=\dfrac{\sigma_{\varepsilon}^{2}}{\sigma_{\mathbf{m}}^{2}}$。由于后验概率密度函数 $p\left(\mathbf{m}\middle|\mathbf{d}^{obs}\right)$ 假定是 Gauss 分布的，则（10）式的 MAP 解应表达为：$\mathbf{m}\middle|\mathbf{d}^{obs}\sim\mathcal{N}\left(\hat{\mathbf{m}},\hat{\mathbf{\Sigma}}\right)$。其中，$\hat{\mathbf{m}}=\left(G^{T}G+\lambda I\right)^{-1}G^{T}\mathbf{d}^{obs}$ 是估计结果，$\hat{\mathbf{\Sigma}}=\sigma_{\varepsilon}^{2}\left(G^{T}G+\lambda I\right)^{-1}$ 是对应的方程（阵），反映了估计结果的不精确性/不确定性。

若实测数据与地下介质弹性参数之间的关系是非线性的，即 $\mathbf{d}^{obs}=\mathbf{d}^{cal}+\varepsilon=H\left(\mathbf{m}\right)+\varepsilon$，则 MAP 估计下的误差泛函定义为

$$\mathcal{J}\left(\mathbf{m}\right)=-\frac{1}{2}\left\|\mathbf{d}^{obs}-G\left(\mathbf{m}\right)\right\|_{\sigma_{\varepsilon}^{2}}^{2}+\lambda\left\|\mathbf{m}\right\|^{2} \tag{11}$$

用 Newton 迭代法求解（11）式，

$$\hat{\mathbf{m}}^{k+1}=\hat{\mathbf{m}}^{k}+\left[\frac{\partial^{2}\mathcal{J}}{\partial\mathbf{m}^{2}}\right]_{\hat{\mathbf{m}}^{k}}Grad\mathcal{J}\big|_{\hat{\mathbf{m}}^{k}} \tag{12}$$

其中，$\left[\dfrac{\partial^{2}\mathcal{J}}{\partial\mathbf{m}^{2}}\right]_{\hat{\mathbf{m}}^{k}}$ 是误差泛函 $\mathcal{J}$ 在 $\hat{\mathbf{m}}^{k}$ 处的二阶导数项（Hessian 矩阵）；$Grad\mathcal{J}\big|_{\hat{\mathbf{m}}^{k}}=\left[\dfrac{\partial G}{\partial\mathbf{m}}\right]_{\hat{\mathbf{m}}^{k}}^{T}\left(\mathbf{d}^{obs}-G\left(\hat{\mathbf{m}}^{k}\right)\right)$是误差泛函 $\mathcal{J}$ 在 $\hat{\mathbf{m}}^{k}$ 处的梯度，很显然梯度是残差波场的反传播，反传播算子是 Jacobi 矩阵算子（Frechét 微商算子）的转置算子。常规的 FWI 就是求解（12）式。可以看出，FWI 方法技术的核心在于计算出正确的梯度，使得（12）式收敛到期望的反演解上。收敛性或（12）式收敛与否是 FWI 方法技术最核心的问题。

Bayes 参数估计的基本逻辑框架如上所述。从这一框架可以看出，当炮集波场用算子形式的波动方程 $\mathbf{d}^{cal}=G\left(\mathbf{m}\right)$来预测时，其对应的参数估计问题就是 FWI。若引入 Born 近似后的正问题 $\mathbf{d}^{cal}=G\mathbf{m}$ 预测炮集波场，则对应的参数估计问题就是最小二乘逆时深度偏移（LS_RTM）。LS_RTM 可以视为估计参数高波数成分的 FWI。

应该注意到，Bayes 参数估计的基本逻辑也为地震信号分析、层析成像奠定了理论基础。实质上，也包含了叠前深度偏移成像的背后理论。

上述理论在勘探地震中的具体化讨论时，需要借助于如下的波场 Born 级数表达，

$$\begin{aligned}G\left(\mathbf{x}_{r}\middle|\mathbf{x}_{s}\right)&=G_{0}\left(\mathbf{x}_{r}\middle|\mathbf{x}_{s}\right)-\omega^{2}G_{0}\delta\mathbf{m}G_{0}+\omega^{4}G_{0}\delta\mathbf{m}G_{0}\delta\mathbf{m}G_{0}+\ldots\\&=G_{0}-\omega^{2}\int G_{0}\left(\mathbf{x}_{r}\middle|\xi\right)G_{0}\left(\xi\middle|\mathbf{x}_{s}\right)\delta\mathbf{m}(\xi)d\xi\\&\quad+\omega^{4}\int\int G_{0}\left(\mathbf{x}_{r}\middle|\eta\right)\delta\mathbf{m}(\eta)G_{0}\left(\eta\middle|\xi\right)\delta\mathbf{m}(\xi)G_{0}\left(\xi\middle|\mathbf{x}_{s}\right)d\xi d\eta+\ldots\end{aligned} \tag{13}$$

其中，右端第一项代表炮集中的透射波，直接由炮点 $\mathbf{x}_s$ 传播到检波点 $\mathbf{x}_r$，没有经过散射过程；第二项是一次散射波；第三项代表二次散射波，其余的是更高阶的散射波。$G\left(\mathbf{x}_r \mid \mathbf{x}_s\right)$ 代表总波场；$G_0\left(\boldsymbol{\xi} \mid \boldsymbol{\eta}\right)$ 代表由 $\boldsymbol{\eta}$ 到 $\boldsymbol{\xi}$ 的 Green 函数。

利用透射波、一次散射波和二次散射波计算出各自对应的梯度是 FWI 方法技术的要点。下面列出了透射波梯度、一次散射波和二次散射波对应的梯度项，

$$\frac{\partial \mathcal{J}\left(\mathbf{m}_0\right)}{\partial \mathbf{m}_0}=\left[\frac{\delta G_0\left(\mathbf{x}_r \mid \mathbf{x}_s\right)}{\delta \mathbf{m}_0(\mathbf{x})}\right]^T\left(G^{obs}\left(\mathbf{m}_0\right)-G^{cal}\left(\mathbf{m}_0\right)\right) \tag{14a}$$

是透射波梯度项。Frechét 微商算子的导出需要把透射波传播进一步地写为，

$$\delta G_0\left(\mathbf{x}_r \mid \mathbf{x}_s\right)=\delta G_0^{\mathbf{m}_0^{initial}}\left(\mathbf{x} \mid \mathbf{x}_s\right) \delta \mathbf{m}_0(\mathbf{x}) \delta G_0^{\mathbf{m}_0^{initial}}\left(\mathbf{x}_r \mid \mathbf{x}\right) \tag{14b}$$

对应的 Frechét 微商算子写为，

$$\frac{\delta G_0\left(\mathbf{x}_r \mid \mathbf{x}_s\right)}{\delta \mathbf{m}_0(\mathbf{x})}=\delta G_0^{\mathbf{m}_0^{initial}}\left(\mathbf{x} \mid \mathbf{x}_s\right) \delta G_0^{\mathbf{m}_0^{initial}}\left(\mathbf{x}_r \mid \mathbf{x}\right) \tag{14c}$$

$$\frac{\partial \mathcal{J}(\delta \mathbf{m})}{\partial \delta \mathbf{m}}=\left[\frac{\partial \delta G_2\left(\mathbf{x}_r \mid \mathbf{x}_s\right)}{\partial \delta \mathbf{m}(\mathbf{x})}\right]^T\left(G^{obs}(\delta \mathbf{m})-G^{cal}(\delta \mathbf{m})\right) \tag{15a}$$

是二次散射波梯度项。其中，

$$\begin{aligned}
\frac{\partial \delta G_2\left(x_r \mid x_s\right)}{\partial \delta \mathbf{m}(x)} &= \frac{\partial}{\partial \delta \mathbf{m}(x)}\left(\omega^4 \iint G_0\left(x_r \mid \eta\right) \delta \mathbf{m}(\eta) G_0\left(\eta \mid \xi\right) \delta \mathbf{m}(\xi) G_0\left(\xi \mid x_s\right) d\xi d\eta\right) \\
&= \frac{\partial}{\partial \delta \mathbf{m}(x)}\left(-\omega^2 \int G_0\left(\xi \mid x_s\right) \delta \mathbf{m}(\xi)\left(-\omega^2 \int G_0\left(x_r \mid \eta\right) \delta \mathbf{m}(\eta) G_0\left(\eta \mid \xi\right) d\eta\right) d\xi\right) \\
&= \frac{\partial}{\partial \delta \mathbf{m}(x)}\left(-\omega^2 \int G_0\left(\xi \mid x_s\right) \delta \mathbf{m}(\xi) \delta G_1\left(\xi \mid x_r\right) d\xi\right) \\
&= -\omega^2 G_0\left(x \mid x_s\right) \delta G_1\left(x \mid x_r\right)-\omega^2\left(\int G_0\left(\xi \mid x_s\right) \delta \mathbf{m}(\xi) \frac{\partial \delta G_1\left(\xi \mid x_r\right)}{\partial \delta \mathbf{m}(x)} d\xi\right) \\
&= -\omega^2 G_0\left(x \mid x_s\right) \delta G_1\left(x \mid x_r\right)+\omega^4 \int G_0\left(\xi \mid x_s\right) \delta \mathbf{m}(\xi) G_0\left(\xi \mid x\right) G_0\left(x \mid x_r\right) d\xi \\
&= -\omega^2 G_0\left(x \mid x_s\right) \delta G_1\left(x \mid x_r\right)+G_0\left(x \mid x_r\right) \omega^4 \int G_0\left(\xi \mid x_s\right) \delta \mathbf{m}(\xi) G_0\left(\xi \mid x\right) d\xi \\
&= -\omega^2 G_0\left(x \mid x_s\right) \delta G_1\left(x \mid x_r\right)-\omega^2 G_0\left(x \mid x_r\right) \delta G_1\left(x \mid x_s\right)
\end{aligned} \tag{15b}$$

是对应的 Frechét 微商算子。

$$\frac{\partial \mathcal{J}\left(\mathbf{m}_r\right)}{\partial \mathbf{m}_r}=\left[\frac{\partial \delta G_1\left(\mathbf{x}_r \mid \mathbf{x}_s\right)}{\partial \mathbf{m}_r(\mathbf{x})}\right]^T\left(G^{obs}\left(\mathbf{m}_r\right)-G^{cal}\left(\mathbf{m}_r\right)\right) \tag{16a}$$

是一次散射波梯度项。其中

$$\begin{aligned}
\frac{\partial \delta G_1\left(x_r \mid x_s\right)}{\partial \mathbf{m}_r(x)} &= \frac{\partial}{\partial \mathbf{m}_r(x)}\left(-\omega^2 \int G_0\left(x_r \mid \xi\right) G_0\left(\xi \mid x_s\right) \delta \mathbf{m}_r(\xi) d\xi\right) \\
&= -\omega^2 G_0\left(x_r \mid x\right) G_0\left(x \mid x_s\right)
\end{aligned} \tag{16b}$$

是对应的 Frechét 微商算子。

（14a）（15a）和（16a）对应的梯度响应，如图 1 所示。（14a）（15a）和（16a）定义的梯度项和（12）式定义的迭代公式构成了 FWI 方法技术内核。梯度决定了迭代公式是否收敛。FWI 方法技术的基本理论就是这样的。将上述理论应用于常密度声波方程、变密度声波方程、各向异性拟声波方程、粘声波方程，直至弹性波方程，都是上述理论框架的具体化。很显然，这是个非常完美的、具有广泛普适性的理论框架。

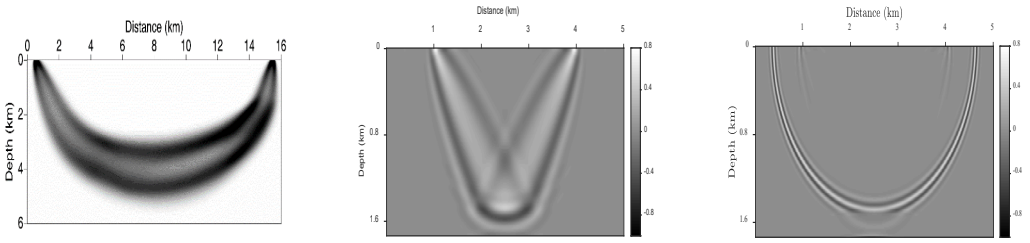


图 1 透射波 FWI、反射波 FWI 和 LS_RTM 的梯度响应

然而，众所周知，实际地下岩石介质的性质极为复杂，不同尺度（扫描电镜尺度，实验室薄片尺度、钻井岩心尺度、测井尺度和地震波长尺度）下，岩石介质表现出显著不同的特征。在勘探地震中，是波前面上的地震子波在感知地下介质岩性的变化，检波器接收到的同相轴上地震子波的波形变化携带了地下介质岩性变化的信息。在地震波特征波长尺度上，地震子波感知到的只是地下介质岩石物性的某种平均化和等价化的变化。在推导波动方程时，对地下介质进行了大量的理想化假设。因此，波动方程只能近似地模拟或解释地震波在实际地下介质中的传播过程。地震波反问题的根本难点在于：人为构建的实测波场与地下介质参数之间的关系是否足够精确且可靠，以及这种关系是否存在强非线性性。此外，待估计参数的尺度是一个重要的影响因素。对于小尺度、强跃变的待估计参数，其与实测波场之间的关系通常呈现出较强的非线性性。一般地，高波数成分的待估计参数与实测波场之间存在更强的非线性性，光滑部分则反之。这也是地震波反演成像分辨率上不去，即高波数部分的估计精确性低的根本原因。

FWI 的核心困难就在于其面对复杂介质（譬如复杂地表、复杂构造、小尺度和/或强变化的地质体）时是一个典型的强非线性反问题。这种强非线性反问题往往表现出多解性、不收敛、不稳定性等特征，使得这类方法技术很难实用化。相比之下，假设已知正确的偏移速度模型，叠前地震波偏移成像问题从数学算法的角度看并不是一个反问题。地震波偏移成像本质上是一种定位方法，保幅成像只是其附带的功能。因此，偏移成像方法具有很强的稳健性，并在勘探地震中得到了广泛的应用。然而，叠前偏移成像的前提是假设已知正确的偏移速度模型，如果把构建偏移速度模型也纳入广义的地震波成像过程中，那么广义的地震波成像就成了一个标准的反问题。在复杂介质情形下，同样很难得到正确的（或合理的）反演解。特别是在复杂探区，叠前深度偏移成像的结果往往难以解决地质问题，也无法准确地揭示真实的地下地质结构形态。其根本原因在于，广义地震波成像在复杂介质条件下同样是一个强非线性反问题。

求解（强）非线性反问题的最基本且最合理的做法是将其转化为一组凸问题来求解，并通过引入更多的有关待估计参数的先验信息来缩小解空间、提高解的稳定性。当然，尽可能采集得到符合 FWI 理论假设的叠前地震数据体、仅估计待估计参数的光滑部分而放弃估计其高波数部分或者仅定位高波数部分出现的空间位置，也是解决（强）非线性反问题的重要途径[8]。

关于当前 FWI 在（各向异性）速度建模中的应用现状，基本判断如下：仅在海上（尤其在深水探区）低频长偏移距观测数据的情形下，透射波 FWI（TFWI）得到了真正有效的应用，能够有效改善后续叠前深度偏移成像的质量，TFWI 实际上是 FWI 的一个凸的子问题。然而，迄今为止，在实际地震数据处理中，反射波 FWI（RFWI）所提供的（各向异性）速度模型，对于提升叠前深度偏移成像质量的效果并不显著且缺乏稳定性（有些探区有效果，而有些探区则效果不明显，甚至会出现质量下降的情况）。这是目前 FWI 在偏移速度建模中应用的基本情况[9][10]。目前，FWI 方法技术的研究重点已逐渐聚焦于如何正确地估计地下介质弹性参数的高波数扰动量[11][12][13]。从 LS_RTM 到 FWI Imaging，再到最近几年发展的 FDR（FWI Derived Reflectivity）[14][15][16][17]，这一技术不断演进并取得了一定进展。为了更好地考虑 AVA 效应，国外许多知名油公司开始转向利用 FDR 生成成像道集的方法技术[18][19][20]。但是，由于高波数扰动量的估计必然依赖于地震子波的振幅，这使得高波数参数估计的收敛性更难以保证，从而给参数反演带来了更大的挑战。

总之，尽管 FWI 已被视为地震波成像的终极理论方法，但是在实际勘探地震中，由于无法完全满足 FWI 的理论假设（如波动方程无法准确预测实际观测的波场，震源函数未知等因素），导致 FWI 在实际数据应用中的反演成像效果远远达不到其理论预期。因此，对于深层超深层地震波成像，亟需针对性的理念和相应的方法技术。

那么，什么是适合深层超深层勘探目标的地震波成像方法技术呢？本文认为，以成像道集为核心的高保真、高分辨率（“双高”）地震波成像理念，正是一个切实可行且针对性强的成像方法技术。事实上，李庆忠院士在《走向精确勘探的道路-高分辨率勘探系统工程剖析》一书中，已经明确阐述了高保真与高分辨率成像处理的思想，并列出了 9 个关键要点：照顾高频、统一波形、对齐时间、抬信压噪、展宽频带、零炮检距（即多项式拟合产生零角度反射系数剖面）、从井出发、零相位化、阻抗反演[21]。然而，遗憾的是，书中并未深入探讨这些思想背后的理论逻辑。

本文认为，“双高”地震波成像处理的根本逻辑是：实现对来自地下同一成像点的、不同炮检对的、形态一致的反射/散射子波进行同相位叠加。能够保证这种同相位叠加处理的物理基础在于：实际勘探中地下介质系统是近似线性的。根据线性系统理论，线性系统的输入信号（激发子波）与输出信号（接收子波）是相干信号（或称为拷贝信号），即 $w_R\left(t\right)=c\cdot w_{\mathrm{S}}\left(t\right)$。其中 $w_{\mathrm{S}}\left(t\right)$ 是震源激发的地震子波，$w_R\left(t\right)$ 是检波器接收到的反（散）子波。可令 $c=\left|A\right|e^{i\phi(t)}$，其中相位函数 $\phi\left(t\right)=\omega t+\phi_0\left(\omega\right)$。若假设传播过程不改变初相位 $\phi_0\left(\omega\right)$，则激发与接收的地震子波仅仅在幅值上有所变化，波形完全一致。这种情况下，则称激发与接收地震子波是完全相干的，其归一化的互相关系数等于 1，它们之间仅存在时间差。地震波偏移速度场决定了这个时间差。因此，高精度的速度建模是“双高”处理的关键技术环节之一。然而，振幅项 $\left|A\right|$ 的影响因素非常复杂，包括球面扩散、吸收衰减、散射损失和透射损失等。但首先要补偿的是球面扩散和吸收衰减。尽管散射损失和透射损失的影响也很严重，但这很难补偿。众所周知，初相位 $\phi_0\left(\omega\right)$ 在实际介质条件下会受地下介质非弹性效应（如吸收衰减效应）的影响而随时间发生变化。因此，在非弹性效应补偿过程中，对地震子波的相位补偿也是很有必要的。上述“双高”处理逻辑是针对单个炮检对的分析，在不同炮检对之间，还有诸如震源和检波器的差异、震源和检波器与介质耦合的不一致、组合激发和组合接收等因素，这些因素也会导致地震子波在振幅和相位（波形）上的不一致。

可以清楚地看出，李庆忠院士列出的“双高”处理的要点，基本上都是为了消除来自地下同一成像点的、不同炮检对的地震子波之间的差异，从而实现真正的同相位叠加。因此，对于深层超深层地震波成像而言，应该将上述逻辑作为发展针对性成像处理流程及相应的方法技术的出发点。

另一个值得关注的逻辑是，叠前数据域（空间）与叠前成像域（空间）之间具有同构性。因为叠前偏移算子 L 和叠前反偏移算子 L^* 是共轭算子，满足 $\left(d,L^*m\right)_{\mathcal{D}}=\left(Ld,m\right)_{\mathcal{M}}$、$\left(Ld_1,Ld_2\right)=\left(m_1,m_2\right)$、$\left(L^*m_1,L^*m_2\right)=\left(d_1,d_2\right)$，意味着在叠前数据空间和成像空间中执行的许多成像处理方法其效果是等价的。换句话说，凡是在数据空间中未能消除的影响“双高”处理效果的因素，仍然可以在成像空间中得到进一步消除。由于叠前数据空间与地下介质之间不存在明显的局部关联性，而在叠前成像空间进行某些与地下反射层相关的处理操作更有针对性，因此，成像道集的生成在深层超深层成像处理中具有核心意义。这意味着，在深层超深层地震波偏移成像过程中，包括 FWI_Imaging 这类反演成像方法，也必须进行成像道集的输出。这也进一步解释了为什么高波数参数的 FWI 一定要发展到 FWI Derived Reflectivity Gather（即 FWI_Imaging 生成成像道集）的阶段。

至此，针对深层超深层地震波成像的、以成像道集为中心的高保真、高分辨率成像处理理念及关键技术可以总结为：（1）实际油气地震勘探中地下介质系统大致可以视为线性系统，激发地震子波与接收地震子波基本是相干信号；（2）叠前数据空间与叠前成像空间是等价的，可以在这两个空间中分别进行各种保真同相位叠加处理；（3）成像道集中地震子波同相位保幅叠加是保真高分辨成像的根本基础；（4）成像道集中的剩余时差校正+剩余形态（剩余相位）校正+最佳照明选择+保幅叠加构成成像道集中“双高”处理的关键技术。为对比起见，将在叠前数据域进行保真同相位叠加“双高”成像处理的关键技术再罗列如下：（1）地表一致性子波振幅与相位校正；（2）地表一致性静校正；（3）高精度速度建模；（4）球面扩散校正；（5）吸收衰减补偿校正。

关于双域（叠前数据域与叠前成像域）中的去噪、成像子波的展宽频带、成像子波的零相位化、成像子波所代表的反射系数的定量化，直至绝对波阻抗（宽带波阻抗的）模型的建立，都应该是深层超深层高精度地震波成像方法中的重要技术环节。

4 针对深层超深层油气地震勘探的地震数据采集技术

高精度地震波成像的基础在于得到能满足成像要求的叠前地震数据体，地震数据采集的根本目的就是提供满足高精度地震波成像需求的叠前地震数据体。目标地质体的深度、地下介质的复杂程度和地震波成像（也包括去噪）方法技术决定了地震数据采集的方式与方法。在深层超深层油气地震勘探中，首要任务是采集到符合深层超深层目标地质体成像所需的叠前地震数据体，然后才是针对性的地震波成像处理流程和相应的方法技术。

从一般意义上讲，高精度地震波成像对叠前地震数据采集技术的基本要求可以概述为以下几点：（1）目标地质体要有宽角度且均匀的照明；（2）目标地质体要有足够的能量和足够的照明；（3）炮与炮、道与道之间的反（散）射子波在振幅、相位和频带上应保持一致；（4）激发地震子波和接收地震子波的频带要足够宽，期望包含有更丰富的低频及邻近低频的信号成分（譬如 1-10Hz 频率成分）；（5）波场和相干噪声的无假频采样，且无论是对炮道集还是共检波点道集，理论上都要求无假频采样。

另外，关于低频及邻近低频的信号成分的激发与接收，需澄清一些关键概念。一般认为，地震分辨率体现在成像子波上。如果宽带阻抗成为未来的主要成像目标，那么宽带阻抗的分辨率应由其结果的数值精度来衡量，具体可通过与测井结果的对比来验证。成像子波的分辨率取决于其完整的振幅谱，同时也与成像子波的相位谱有关。零相位成像子波相较于其他具有不同相位特征的成像子波，能够提供最高的分辨率，这也解释了李庆忠院士强调的成像子波零相位化的原因。高分辨率成像子波具有以下特征：零相位、窄主瓣和弱旁瓣。这些特征对于岩性储层的解释尤为重要，特别是弱旁瓣的成像子波在深层超深层油气地震勘探中的意义尤为突出。窄主瓣的成像子波要求激发子波和接收的地震数据中包含足够的高频成分（譬如 45Hz-80Hz），但是对于深层超深层油气勘探来说，这通常是无法实现的。因此，激发和接收地震子波中需要包含低频及邻近低频的信号成分，且这些低频成分的能量必须足够大（应达到或尽可能接近优势频段频率成分的能量水平），这样才能有效压低成像子波的旁瓣，从而提高深层超深层地震波成像分辨率。

综上分析，就是要实现真正的宽频、宽方位、高密度（“两宽一高”）地震数据采集[22]。

但是，针对深层超深层油气地震勘探，地震数据采集技术应在上述一般性要求的基础上，进一步满足以下针对性要求：（1）更加重视激发低频及邻近低频的信号成分的能量，这是因为低频和邻近低频的信号能量不易被吸收衰减，更容易穿透至深层目标地质体。（2）提高炮密度，由于深层超深层目标地质体的地震波能量弱且波阻抗差小，因此应在每平方千米范围内增加炮密度，以提供足够的能量支持和改善目标地质体的成像质量。（3）长偏移距采集，为了满足深层超深层反射波建模的需求，若无法覆盖较宽的方位角范围，则必须确保采集的偏移距足够长，这有助于提高深层反射波的重建质量和深度模型的精确性。（4）炮检排列尽可能垂直于目标地质体的走向，在长偏移距情形下，这样更有利于利用反射波建立相对准确的深层速度模型。实际上，前述地震数据采集方式也兼顾了透射波 FWI 对观测方式的要求。（5）震源和检波器都应该尽量靠近目标地质体，譬如检波器布置在井中的 VSP 观测、震源和检波器都放置在疏松地表介质下的观测以及检波器布置在海底面上的 OBN/OBC 观测。

一个值得深入研究的问题是，针对深层超深层目标地质体，且其波阻抗扰动较小的情况下，震源需要激发多大的能量，才能使激发的地震波通过深层反射和长偏移的传播后，确保反射（散射）地震波到达检波器时能够达到并超过当前检波器的灵敏度下限而被检波器记录下来。尽管这个问题在上世纪 60-70 年代已有一定的研究，但是对于深层超深层油气地震勘探而言，仍然需要进一步深入探讨其适用性。

与此相关的另一个问题是，如何有效控制近地表相关噪声的能量，以确保在深层长偏移距、弱波阻抗扰动情况下，所产生的（极）弱反（散）射信号的信噪比能够满足当前去噪方法技术的理论假设要求。我们知道，不相干（随机）噪声必须弱于线性信号才符合当前去噪方法技术的基本假设。然而，信噪比低于多少当前的去噪方法即无效，目前尚无明确的定量标准。由此可见，现有的去噪方法技术也应当为野外地震数据采集提供尽可能明确的指导和要求。

5 针对深层超深层油气地震勘探的地震数据预处理

在深层超深层油气地震勘探中，地震数据预处理的目的是使得预处理后的地震数据满足速度估计方法技术和叠前深度成像方法技术的理论假设要求。需要注意的是，每种地震数据预处理方法技术本身也有其相应的理论假设前提。

前面已经分析过，当前的终极地震波成像方法技术是基于 Bayes 框架下的 FWI，其中包括透射波 FWI、反射波 FWI，Born 线性化正问题对应的 LS_RTM 和非线性化正问题对应的 FWI_Imaging 方法技术；此外，退而求其次的地震波成像方法技术包括透射波速度层析、一次反射波成像道集速度层析和一次反（散）射波叠前深度偏移方法技术。当然，上述各种成像方法是与所考虑的介质情形对应的，其中包括变密度或常密度声介质、粘声介质、各向异性声介质、各向异性粘声介质，弹性介质等。显然，在各种成像方法中，任何无法被所选的正问题预测的波，均需要在数据预处理中去除。由非主震源产生的波及不明物理源的随机噪声，更是通过预处理加以去除。凡是不能被所选的正问题考虑的各种波传播现象，也应在地震数据预处理中补偿掉。

因此，地震数据预处理的本质在于：凡是不能被所选的正问题考虑的任何物理因素都应在地震数据预处理中消除掉或压制掉。在地震数据预处理过程中，首先要考虑的是不能被所选的正问题处理的震源激发问题和检波器接收问题，第二是不能被所选的正问题预测的波现象的去除或压制问题，第三是当前正问题的数值解需要无假频规则采样的问题，第四是不能被所选的正问题考虑的各种波传播机制问题。地震数据预处理主要解决上述四个方面的问题。无论是工程或城市地球物理中的浅层地震勘探，或是中深层油气地震勘探，甚至深层超深层油气地震勘探，数据预处理关注的都是上述四个方面的问题。

但是，深层超深层油气勘探中地震数据预处理仍具有其特殊性。下面就深层超深层油气地震勘探中地震数据预处理的上述四个方面的问题展开详细讨论。

首先，波动方程所描述的正问题并未考虑实际震源本身的激发机制、震源与周围介质的耦合效应以及震源组合效应，而是简单地假设存在一个理想的震源函数，譬如 Ricker 子波或其他理论子波。事实上，还期望各炮震源函数是完全相同的。当然，理论上，这个假设对于 FWI 类的反演方法并非必要。然而，如果不同炮震源函数之间存在差异，或者震源函数变化较快，可能会导致反演参数空间的极度膨胀，从而增强反问题的非线性程度，因为震源函数本身也是待反演的参数。事实上，震源函数已知的假设很难在实际应用中满足，尤其是在陆上地震勘探中。同样地，检波器本身的物理系统响应、检波器与周围介质的耦合效应以及检波器组合效应也没有在波动方程表达的正问题中得到考虑。因此，假设每个炮的综合响应一致，一炮所有检波端的综合响应完全一致，显然不符号实际情况。无论是在海上地震数据的子波处理，或是在陆上地震数据的地表一致性子波振幅、相位和频带校正，这些因素对于实现“双高”成像处理都是至关重要的。

第二，众所周知，勘探地震的核心目的是得到深层目标地质体的构造图像，甚至宽带弹性参数估计结果。理论上，所有从地下反射界面或散射体反（散）射回地表的地震波原则上都是有效的，并能够用于目标地质体的成像。噪声主要来自于主动源激发的、近地表传播的波，或者来自于非主动源或未知源的波。总之，凡是所采用的地震波成像算子不能处理的波，都应视为噪声。显然，某一种波现象是否为噪声，与其是否来自地下反射界面和散射点有关，也与地震波成像的理论方法有关。当然，FWI 理论下，凡是能贡献给 FWI 反演成像的波都不能归为噪声。对于被动源数据的成像而言，噪声的概念则更加复杂。去噪的理论框架通常可以归结为：局部高维数据体中线性结构嵌套在符合一定概率分布（一般假设 Gauss 分布）的随机噪声中。在 Bayes 估计理论下，通过采用线性预测算子，能够最佳预测局部高维数据体中线性结构，从而实现去噪的目的。线性结构通常指的是线性同相轴，意味着嵌套在一起的地震子波只有线性相位移，没有振幅或波形上的变化。然而，在非线性、非高斯、非平稳的情形下，去噪依然是任重道远的问题，尤其是在地质构造复杂的区域，譬如山前带探区的地震数据去噪。基于模型数值模拟的噪声并进行减去的策略也是去噪的一种核心思想。然而，假设已知介质参数模型和实际震源函数，再加上复杂的介质情形和边界条件，进行噪声模拟并减去该噪声变得异常困难，因此难以有效去除噪声。尽管如此，预测噪声并减去该预测的噪声仍是去噪的一个重要的方法，尤其是对于多次波的压制。（近）地表相关的波和地表相关的随机噪声，或许也能通过模拟和减去的方法进行有效压制。

第三、波动方程正问题的数值解是实现各种波动方程理论反演成像或偏移成像的基础。一般而言，理论上要求野外地震数据采集具有空间规则性且不存在假频。然而，实际地震数据往往难以满足空间规则和无假频的理论要求。因此，抗假频的数据规则化成为了高精度地震波成像中十分重要的预处理环节之一。

抗假频的数据规则化方法技术在理论框架上与去噪理论相一致，即对非规则局部高维数据体中线性结构进行抗假频的最佳预测，进而实现地震数据的规则化。值得一提的是，在同时源（混叠源）数据采集中，解混叠的主要方法也依赖于对伪解码数据的最佳滤波处理。

第四、不能被所选的正问题考虑的各种波传播机制问题。经过分析可以看出，地震波成像所用的波动方程，无论是在变密度或常密度声介质、粘声介质、各向异性声介质、各向异性粘声介质，还是弹性介质情形下，都是对实际介质中波传播过程的高度抽象和简化后导出的。这些方程能够反映地震波在不同介质中下传播的主要特征。否则，基于这些方程进行的地震波成像将难以产生实际的应用价值。事实上，现有的地震波成像方法主要依赖于炮集中各种波现象对应的同相轴上的地震子波的走时（相位）和振幅信息。Zoeppritz 方程描述了平面波入射到反射界面时，因界面两边弹性参数不同而引起的波场分解机制，进而推导出角度反射（透射）系数与界面两边弹性参数的关系，这也是对波传播机制的高度抽象。球面扩散、聚焦散焦、透射损失、散射损失、多次反射与多次散射、吸收衰减等，都是波传播过程中的主要物理机制。波动方程能处理其中的哪些问题？哪些问题需要留给地震数据预处理方法来解决？对于 FWI 反演成像，理论上不要求进行任何额外的预处理。然而，基于射线理论的成像方法、基于单向波方程的成像方法以及基于双向波方程的成像方法等，都需要与之对应的特定的预处理方法，总之，所有不能被所选的正问题有效处理的各种波传播机制问题，都要通过对应的预处理方法来加以解决。

一般而言，深层超深层油气地震勘探的地震数据预处理应包括以下几个关键环节：地表相关噪声的压制；深层弱信号的保护与增强；层间多次波的压制；地表一致性地震子波的振幅和相位校正；球面扩散校正（可以在 FWI 或波动方程叠前深度偏移过程中进行）以及吸收衰减校正（可以在叠前深度偏移成像环节中进行）。其中，深层弱信号的保护与增强和层间多次波的压制是深层超深层地震波成像处理中需要特别关注和处理的重点问题。

深层超深层油气地震勘探中，深层超深层目标地质体的弱反（散）信号问题是面临的一个重大挑战。其形成的物理原因主要是球面扩散、散射损失、透射/反射损失、吸收衰减以及弱波阻抗差等因素的综合作用。尤其在深层超深层地层，由于温压作用，异常体与围岩之间的波阻抗差异往往较小，这一特征在一定程度上是深层超深层弱信号形成的主要原因之一。此外，深层超深层的弱信号以球面波传播到地表，而地表的随机噪声不随时间衰减，这进一步降低了深层超深层反（散）射信号的信噪比。遗憾的是，针对强随机（不相干）噪声背景下的弱（线性）信号去噪问题，现在尚未有相应的理论与方法。这是深层超深层弱信号增强处理遇到的一个特殊问题。另外，目标地质体上覆地层存在的或强或弱的层间多次波对深层超深层反射波成像会产生显著的干扰，这是深层超深层地震数据预处理面临的另一个回避不了的问题。

用于深层超深层极微弱反（散）射信号增强的方法主要有：1、强随机（不相干）噪声背景下的弱（线性）信号的去噪理论与方法。当前，针对具有保障信噪比的情况下，有许多方法能够有效压制线性噪声。然而，在强随机（不相干）噪声和弱（线性）信号情形下，如何既保持弱（线性）信号的完整性，又能有效地压制强随机（不相干）噪声，依然缺乏理论依据，也缺乏有效的方法技术手段。此外，复杂地表下的震源机制和检波机制（尤其是快速空变）实际上会产生一种乘性噪声，而现在严重缺乏乘性噪声压制的方法理论体系与方法。地表一致性地震波的振幅与相位校正，虽然在地震勘探领域得到广泛应用，但是这并非是乘性噪声压制理论体系下发展的方法，而是一种行业发展的特殊应对方法。因此，上述技术问题进一步增加了深层超深层弱信号增强的难度。2、波场外推近地表噪声压制方法技术。基于近地表光滑模型的波场外推法是一种能够有效压制近地表噪声的方法，但这种方法与深层弱信号无关，理论上应该能在保持深层弱信号的情况下压制近地表传播的噪声。3、不相干（或随机）噪声模拟与减去的方法技术。由于深层超深层反（散）射信号极微弱，因此近地表不相干（或随机）噪声通常会远强于信号本身。在此情况下，依赖高次覆盖数据的同相位叠加来压制这些近地表不相干（或随机）噪声，从而提高深层超深层极微弱反（散）射信号的信噪比，显然不再是一种有效的选择。针对深层超深层极微弱反（散）射信号问题，应发展不相干（或随机）噪声的统计模拟方法（随机模拟方法）及其相应的减去方法。

用于深层超深层层间多次波的识别与压制的方法[23]：特征反射层相关多次波预测与压制方法技术。所谓特征反射层是指波阻抗差异比较大的若干反射层。首先利用测井数据及对应的合成记录，识别特征反射层及相关的层间多次波。然后，融合测井数据、偏移速度建模结果和叠前深度偏移成像结果，生成特征反射层速度模型。基于特征反射层速度模型，采用波动方程正演模拟方法，预测特征反射层相关的各阶多次波。在此基础上，采用最佳减去方法压制特征反射层相关的多次波。此外，可以采用机器学习方法实现上述技术路线。具体而言，首先通过测井数据及对应的合成记录，识别特征反射层及相关的层间多次波，构建机器学习样本数据，训练深度学习网络，然后对实际的深层超深层地震勘探炮集中的特征反射层相关多

次波进行预测与压制。

6 针对深层超深层油气地震勘探的偏移速度建模方法技术

速度估计的本质在于利用地下同一速度估计单元中有宽角度且角度分布相对均匀的透过波的走时，尽可能精确地估计该单元的速度（扰动），这是对层析速度反演本质的描述，无论是透射波反演还是反射波层析反演都是如此。速度估计的另一个本质描述是，当偏移速度正确时，不同炮检距（对）的来自地下同一成像点的反（散）子波应该位于拉平的同相轴上，从而实现同相位叠加。如果道集无法拉平，说明偏移速度存在误差。通过检测不同炮检距（对）的来自地下同一成像点的成像子波的成像深度差，并将其转化为时间差，可以建立起沿该炮检对的波传播路径的速度差与时间差之间的关系，据此形成反（散）波层析矩阵方程，进而进行反射波层析反演并更新偏移速度。FWI，尤其是透射波 FWI（TFWI）和反射波 FWI（RFWI），本质上是波动理论下的上述层析反演的另一种实现方式。此外，还有一些其他形式的初始速度建模方法。例如，NMO+DMO 均方根速度扫描建模方法；PSTM+CIG+NMO-1+NMO 均方根速度扫描建模方法；百分比扫描方法等，这些方法主要适用于横向缓变速介质的情况，都属于利用反射波的初始速度建模方法。

上述提到的速度建模方法都是一般性的方法。而用于深层超深层地震波成像的速度建模方法到底有什么特殊之处？本文认为其特殊性主要体现在以下几个方面：（1）深层超深层地震波成像无法避开浅层介质的影响，为了建立准确的速度模型，必须首先对近地表、浅层和中深层进行足够准确的速度建模，这本身就具有极大的难度。（2）目标层埋深大，严重影响高精度速度估计的基本前提（即地下同一速度估计单元有宽角度的、且角度分布相对均匀的穿透波），深层速度估计精度低的根源在于难以满足这一基本前提。这要求采集足够长偏移距的地震数据，以便于准确建立速度模型并对深部的目标地质体进行精确成像，但实际上，获得这样的数据是非常困难的。（3）深层反（散）信号的信噪比低，在没有足够的信噪比保障的情况下，任何速度估计方法所估计的速度精度都会严重降低。（4）对于深层超深层目标地质体的成像，长偏移距地震数据采集是必不可少的，此时，地震各向异性特征更加显著，至少要考虑 VTI 介质下的速度估计。另外，由于波传播路径长，累积的吸收衰减量大，因此，Q 值的估计也必须加以考虑。

由上述分析可以看出，尽管深层超深层地震波成像的速度建模具有其独特性，但是从方法技术的角度看，深层超深层速度建模方法与中浅层速度建模方法并没有本质上的区别。需要指出的是，由于叠前地震数据往往无法满足深层超深层速度建模的理论要求，因此需要提供更多与深层超深层速度有关的先验信息来提高深层超深层速度建模的精度。然而，这一做法实际上也是在任何复杂情形下提高速度建模精度的常见策略。

7 针对深层超深层油气地震勘探的保真高分辨地震波成像方法技术

可以肯定的是：深层超深层油气藏描述主要还是依赖于高波数参数的地震波偏移成像资料。然而，当前国际上先进的高精度高波数参数估计方法，如 LSRTM、FWI_Imaging、FWI Derived Reflectivity Gather 等方法，这些高波数参数成分 FWI 方法技术在我国西部深层超深层油气地震勘探中尚难以得到实际应用。相比之下，中国东部海上探区，由于地温梯度大，且目标储层的深度不超过 4500 米，因此，可以探索先进的高波数参数成分 FWI 方法在该地区的应用。

我们认为，针对深层超深层油气藏描述的需求，合理的方法技术路线还是应侧重于高保真高分辨的地震数据处理，以尽可能得到保真、弱旁瓣、窄主瓣的叠前深度偏移成像资料。下面简述一下上述观点的理论基础。

地震波高波数参数反演成像可以由（12）式结合（16b）式来实现，这是高波数 FWI（即 LS_RTM）的基本理论框架。如果（12）不进行迭代，仅执行一次，则得到的就是一般的 RTM 成像。RTM 的方法如下：假设已知正确的偏移速度模型和震源子波，利用所选择的波动方程数值模拟炮点到成像点处的波场；然后将检波端记录到的实际波场反推（反传播）到成像点处；接下来，将两个波场进行互相关，互相关系数即为成像结果。显然，在偏移速度和震源子波已知且正确的前提下，在地下介质系统是线性的条件下，两个成像点处相关的子波应该是相干的，并且其归一化互相关系数应该等于 1，这是最理想的情况。在成像子波零相位情形下，主瓣峰值所对应的 RTM 成像反射系数与真实的地下反射系数是相等的。此时成像结果被认为是保真的。

可以对上述观点进行如下的理论分析。如图 2 所示，炮点传播到成像点 $\mathbf{x}$ 处的波场记为 $u^S(\mathbf{x})$，检波端反传到成像点 $\mathbf{x}$ 处的波场记为 $u^R(\mathbf{x})$，偏移成像结果表示为：

$$R_{\text{mig}}(\mathbf{x},\gamma)=\frac{u^R(\mathbf{x})}{u^S(\mathbf{x})} \tag{17}$$

其中，$u^R(\mathbf{x})=R_{true}(\mathbf{x},\gamma)u^S(\mathbf{x})$，$\gamma$ 是反射张角。可以将成像点处的波场 $u^S(\mathbf{x})$ 和 $u^R(\mathbf{x})$ 用 WKBJ 近似(Wentzel-Kramers-Brillouin-Jeffreys approximation)形式进行表达，即 $u^S(\mathbf{x})=A^S(\mathbf{x})e^{i\omega t_S}e^{i\phi_S}$，$u^R(\mathbf{x})=A^R(\mathbf{x})e^{i\omega t_R}e^{i\phi_R}$。WKBJ 形式表达的波场的振幅由波传播过程中各种影响振幅的因素决定；走时（传播相位 ωt）由（各向异性）背景速度（偏移速度）决定；地震子波（初）相位 ϕ 主要由激发接收和吸收衰减因素决定。将两个波场的 WKBJ 近似形式代入（17）式可以得到，

$$\begin{aligned}R_{\text{mig}}(\mathbf{x},\gamma)&=R_{\text{true}}(\mathbf{x},\gamma)\frac{\left|A^R(\mathbf{x})\right|e^{i\omega t_R}e^{i\phi_R}}{\left|A^S(\mathbf{x})\right|e^{i\omega t_S}e^{i\phi_S}}\\&=R_{\text{true}}(\mathbf{x},\gamma)\left|\frac{A^R(\mathbf{x})}{A^S(\mathbf{x})}\right|e^{i\omega(t_R-t_S)}e^{i(\phi_R-\phi_S)}\\&=R_{\text{true}}(\mathbf{x},\gamma)F(\mathbf{x})\end{aligned} \tag{18}$$

其中，$F(\mathbf{x})$ 是一个带通滤波器，期望它是一个零相位、窄主瓣且弱旁瓣的带通滤波器。对（18）式的地震含义的解释，可以说明“双高”成像处理的本质含义。首先，只有当从地表开始到深层超深层的（各向异性）偏移速度模型正确时，走时 t_S 和 t_R 才会相等，即 $e^{i\omega(t_R-t_S)}=e^0=1$ 才成立。注意，t_R 是实测波场反传到成像点的走时，t_S 是炮点到成像点的走时，当偏移速度正确时，在成像点处二者应是相等的。当影响子波形态（初相位）的因素得到彻底的校正时，两个波场中在成像点处的地震子波应该具有相同的相位，即 $\phi_S=\phi_R$，则 $e^{i(\phi_R-\phi_S)}=e^0=1$。最后，当波传播过程中各种影响振幅的因素被彻底消除时，$\left|\frac{A^R(\mathbf{x})}{A^S(\mathbf{x})}\right|=1$。很显然，这样的理想情况下，$F(\mathbf{x})=1$，$R_{\text{mig}}(\mathbf{x},\gamma)=R_{\text{true}}(\mathbf{x},\gamma)$。

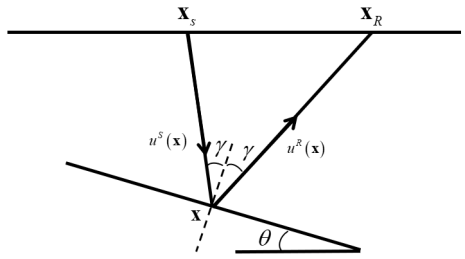


图 2 地震波保真偏移成像分析示意图

由于地震勘探中地震子波和观测孔径是带限的，基于（17）和（18）式表达的成像结果，本质上是两个波场的归一化相关系数，其必然呈现出带限子波的特征。实质上，成像结果以归一化互相关函数的形式呈现出来的，或者说是 $u^S(\mathbf{x})$ 的自相关函数的形式呈现出来的。但是，由于地震子波未知，正演合成传播到成像点处的波场中的地震子波与实测波场反传播到成像点处的波场中的地震子波不可能完全一样，因此，实际的成像结果往往呈现出归一化互相关函数的形式。无论如何，地震波走时 t_S 和 t_R 相等时的零延迟互相关结果应该得到保真的成像结果，即此时期望上述归一化互相关系数尽可能等于 1，在此条件下，保真的反射系数出现在正确的深度位置。需要注意的是，走时相等时，零延迟互相关结果是否等于 1，与 $u^S(\mathbf{x})$

和$u^R(\mathbf{x})$中地震子波的相位特征有密切关系。这表明，高精度的地震波成像高度依赖于接收子波的相位特征。地震波成像的本质目标应该是：零延迟的、保真的偏移成像结果$R_{\text{mig}}(\mathbf{x},\gamma)$出现在正确的深度位置。这是地震波保真成像的真实含义。这说明了：正确的构造成像与反射系数保真的成像是一体的，或者说没有正确的构造成像时，不应该去追求保真成像。

高分辨地震波成像应该是在上述保真成像基础上的外延，即先有保真成像，然后才追求高分辨地震成像。高分辨地震波成像的核心目标是期望上述归一化互相关反射系数是在非零延迟量τ（$\tau=t_R-t_S$）增加时快速衰减至零。勘探地震的专业表述是：在上述保真成像的基础上，主瓣应尽可能窄，旁瓣尽可能浅且窄。很显然，实现高分辨地震波成像的基本逻辑是：$u^S(\mathbf{x})$和$u^R(\mathbf{x})$尽可能波形完全相同，它们本身都是高分辨子波（具有窄主瓣、浅且窄旁瓣的波形）。$u^S(\mathbf{x})$和$u^R(\mathbf{x})$都是脉冲时，只有走时t_S和t_R相等时的零延迟互相关结果，此时分辨率达到极限（成像结果是脉冲，理想情况下成像结果等于真反射系数）。可以看出，$u^S(\mathbf{x})$和$u^R(\mathbf{x})$中地震子波的相位特征对高分率成像的影响很大。

上述分析清楚表明了依赖于（各向异性）偏移速度场的地震波传播相位（ωt）和主要受激发接收因素及吸收衰减因素影响的地震子波初相位（ϕ）对保真和高分辨成像的影响。

（18）式中的振幅影响项$\left|\dfrac{A^R(\mathbf{x})}{A^S(\mathbf{x})}\right|$是如何影响保真和高分辨成像的？首先，$u^S(\mathbf{x})$和$u^R(\mathbf{x})$中地震子波的振幅$A^S(\mathbf{x})$和$A^R(\mathbf{x})$是由波传播过程中影响振幅的因素控制的，主要受球面扩散、振幅的非弹性衰减、散射和透射损失等影响。另外，在（17）和（18）式表达的偏移成像过程中，正演模拟合成$u^S(\mathbf{x})$时，震源子波未知，是人为给定的一个理论子波；另一方面，检波器接收的波场的振幅值必须要经过放大才能被记录下来。检波器（组合）的响应也会伤害$u^R(\mathbf{x})$中地震子波的振幅值$A^R(\mathbf{x})$。当然，波传播算子也（可能）会伤害传播到成像点处地震子波的振幅值。

在深层超深层地震勘探中（浅中层亦如此），如何尽可能使得$\left|\dfrac{A^R(\mathbf{x})}{A^S(\mathbf{x})}\right|$接近于 1，而且在地下任意成像点处均能达到这一目标，就成了保真成像的另一个重要议题。

实际的 RTM 地震波偏移成像，成像点 **x** 处的归一化互相关系数得到的像，具体的表达为，

$$\begin{aligned} R_{\text{mig}}(\mathbf{x},\gamma)&=\int\frac{u^S(\mathbf{x},\gamma)*u^R(\mathbf{x},\gamma)}{u^S(\mathbf{x},\gamma)*u^S(\mathbf{x},\gamma)}dt=\frac{\int u^S(\mathbf{x},\gamma)*R_{true}(\mathbf{x},\gamma)\tilde{u}^S(\mathbf{x},\gamma)dt}{\int u^S(\mathbf{x},\gamma)*u^S(\mathbf{x},\gamma)dt}\\ &=R_{true}(\mathbf{x},\gamma)\frac{\int u^S(\mathbf{x},\gamma)*\tilde{u}^S(\mathbf{x},\gamma)dt}{\int u^S(\mathbf{x},\gamma)*u^S(\mathbf{x},\gamma)dt}\end{aligned}\tag{19}$$

其中，在成像点 **x** 处，$\tilde{u}^S(\mathbf{x})$代表实际震源子波，$u^S(\mathbf{x})$代表震源端波场模拟时给出的震源子波。（19）式的分子项代表零延迟互相关结果（能量），分母项代表成像点处震源端波场模拟时给出的震源子波的零延迟自相关（能量）。由此可见，保真地震波成像时，是（19）式中分子项与分母项的能量比（不是零延迟时刻的地震子波的振幅比）决定是否消除了振幅项的影响。所谓振幅项影响因素的消除，就是在$u^S(\mathbf{x})$和$u^R(\mathbf{x})$中地震子波是相干子波（波形相同的子波或相同相位特征的子波）的假设下，消除波传播过程中的球面扩散、非弹性振幅衰减、透射损失等影响因素。在实际的地震波成像过程中，就是选择更能模拟实际波传播物理过程的波动方程来消除振幅项的影响。

由上述分析可知，假设炮与炮之间的震源子波（相位特征）一致且已知，同时各震源与介质的耦合关系一致，假设检波器之间的激励响应函数一致，且各检波器与介质的耦合一致，在这种条件下，只要选择

更能模拟实际波传播物理过程的波动方程进行炮点到成像点的波场模拟和检波端到成像点的波场反传，振幅对保真成像的影响就可以得到较好的控制。然而，遗憾的是，对于陆上地震勘探而言，要满足这一理想条件困难较大。

对上述分析进行基本总结，可以清楚地看出，影响“双高”地震波成像处理结果的关键因素包括：（1）影响传播相位的（各向异性）偏移速度场；（2）影响接收子波形态（相位特征）的激发接收因素和吸收衰减因素；（3）能够更准确模拟实际波传播物理过程的波动方程（消除波传播机制对震源子波振幅的影响）。

简而言之，子波、偏移速度和波传播算子是“双高”成像处理的核心。在深层超深层油气地震勘探中，由于波传播的地震地质机制更加复杂，更应该做好这三个方面的工作。

同时，必须强调的是，噪声、数据不规则、深层超深层照明角度不足且照明不均匀等因素，对“双高”地震波成像处理的影响同样不可忽视，必须给予充分的重视。

再次强调，“双高”处理需要在叠前数据域和叠前成像域（“双域”）中进行。进行深层超深层高精度地震波成像，首先必须生成与叠前深度偏移对应的方位角度成像道集或倾角域成像道集，在方位角度成像道集或倾角域成像道集的生成过程中尽可能彻底消除影响“双高”处理的各种因素。在方位角度成像道集中进行层析速度反演提高了偏移速度精度，在倾角成像道集进行最佳照明优选提高了偏移成像质量。

最后，针对我国西部深层超深层油气地震勘探的特点，提出了以成像道集为中心的高保真高分率成像处理的基本流程，主要包括：（1） 成像基准面选择；（2） 高频道间时差校正；（3） 近地表综合信息建模；（4） 保振幅保频带去噪声；（5） 5D 数据规则化；（6） 全方位透射波层析速度分析与建模；（7） 全方位（特征）反射波各向异性速度（包括 Q 值）层析反演与建模；（8） 各向异性吸收衰减介质 PSDM 或 LS_PSDM 并生成方位角度道集和倾角域成像道集；（9） 方位张角成像道集和倾角道集中的“双高”处理，尽可能实现保真同相位叠加；（10） 保真弱旁瓣窄主瓣反射系数成像，叠后高维反褶积提高成像子波的分辨率；（11）（如果可能）进一步给出宽带波阻抗成像结果[24][25]。

8 深层超深层油气地震勘探的技术发展方向

由前述分析可知，子波、偏移速度和波传播算子是“双高”成像处理的核心要素，噪声、数据不规则、深层超深层照明角度不足且照明不均匀等因素对“双高”地震波成像有重要影响。

深层超深层油气藏的成像问题首先是对目标油气藏的广角度的、角度均匀分布的有效照明（照明子波形态一致、能量一致）问题、然后是对来自目标油气藏的反（散）射波的同相叠加问题。前者取决于地震数据观测方式，后者取决于背景速度（及各向异性参数和 Q 值）模型。信噪比是地震波成像中的另一个关键问题，尤其是深层超深层地质目标，反射信号通常比较弱，且信噪比较低，这成为深层超深层油气藏精确成像的主要障碍之一。

因此，要实现深层超深层目标地质体的高精度地震波成像，“两宽一高/两宽两高”的地震数据采集是基本要求。此外，低频及邻近低频的频率成分需要以足够的能量激发并被检波器保真地接收，这是深层超深层油气地震勘探的必要需求。由于深层超深层目标地质体的反（散）散射波极为微弱，采集技术必须重点控制非相干噪声和随机噪声的水平。同时，鉴于激发和接收子波的一致性对于“双高”成像处理的重要影响，应该发展相应的方法技术尽量保证不同炮检对之间激发与接收子波的一致性。

进一步地，来自目标地质体反（散）射波的同相叠加是实现精确成像的基础，同相叠加的根本保障首先在于背景速度和各向异性参数的准确性。因此，背景（各向异性）速度建模（包括 Q 值建模）是深层超深层目标地质体高精度成像的关键环节之一。但是，正如前述分析所指出：地震子波的形态（相位）特征对“双高”成像处理具有重大的影响，陆上深层超深层地震勘探中对子波的处理要引起足够的重视。特别是，地表一致性地震子波振幅与相位处理技术已有较长时间未见到创新性的研究成果的报道。

全波形反演(FWI)技术作为当前业界热议的高精度成像方法技术，其实际应用效果并没有预期的那样理想。造成这种结果的主要原因有：实际数据缺失低频数据或低频数据不可靠；缺乏长偏移距的观测数据；地震子波空变且未知；地震数据受各种噪声干扰且（噪声分布）空间不规则；正演算子无法预测所有观测到的波现象（正问题精度）等等。事实上，FWI 本质上是 Bayes 参数估计在勘探地震中的一个应用实例，是一个典型的数学反演问题。但是，勘探地震问题从来也不是一个单纯的数学问题。针对深层超深层油气地震勘探，透射波 FWI 和反射波 FWI 应该可以提供更高精度的偏移速度场，但对高波数成分（角度反射系

数）的估计或成像，我们认为还需要依赖于以成像道集为中心的高保真高分辨率成像处理流程及关键技术，而不能完全依赖 FWI_Imaging 或 FWI_Imaging 生成的角度成像道集这些方法技术[26]。

在以成像道集为中心的高保真高分辨率成像处理流程中，背景（各向异性）速度建模（包括 Q 值建模）的合理方法与技术应包括以下几个方面：（1）数据域基于旅行时和相位的波动理论层析反演与建模，分别考虑透射波和反射波的层析反演；考虑速度、各向异性和 Q 值参数的反演与建模。（2）成像域 Beam 层析成像与建模，同时考虑速度、各向异性和 Q 值反演与建模。（3）提高建模质量的正则化方法，各种现代图像处理方法（机器学习方法）的应用提高建模的地质合理性和精度。而在高精度地震波成像技术中，合理的选择应该包括以下技术：（1）强调照明均衡的 Q 及 VTI/TTI/HTI/ORT 介质特征叠前深度偏移成像技术；（2）复杂介质情况下保真的宽方位角度成像道集和倾角道集的生成及相应的各种后处理技术。

关于上述方法技术，尽管在中浅层油气地震勘探研究过程中已经积累了较为丰富的经验，但针对深层超深层目标地质体的成像，仍有诸多需要完善和改进的地方。主要包括在以下几个方面：（1）在数据采集方面，目前并没有真正实现“两宽一高”的地震数据采集，尤其对陆上地震勘探而言，缺乏有效的低频及邻近低频的地震波成分的激发和接收方法技术。（2）在背景（速度和 Q 值）参数的建模方面，数据域的方法离真正的商业应用还有比较大的差距，成像域中的方法也还有很多值得改进的地方。（3）在叠前深度偏移成像方面，各向异性粘声介质叠前深度偏移成像并输出方位角度和倾角成像道集的波动方程 PSDM（尤其是 RTM）方法技术的商业化应用困难；方位成像道集和倾角道集后处理生成保真、弱旁瓣、窄主瓣反射系数的方法技术商业应用很缺乏。（4）在噪声处理方面，强非相干（随机）噪声加（极）弱线性信号情形下的去噪方法基本缺失。（5）在多次波压制方面，针对深层超深层成像的特征反射层相关层间多次波压制方法技术尚不成熟，在应用中仍存在诸多限制。

总之，深层超深层油气地震勘探的地震波成像技术仍有许多关键问题亟待解决和完善。

9 结论与讨论

深层超深层（包括深水）油气地震勘探已成为国内外油气勘探的重要领域之一，并取得了显著的勘探效果。在这一领域，核心任务仍然是对油气藏进行准确的定位、识别、描述与精细刻画，而实现这一目标最关键之处在于首先要解决深层超深层目标地质体的保真高分辨地震波成像问题。

深层超深层目标地质体地震波成像的困难主要是源于更复杂的地震地质问题，这些难点包括：勘探目标埋藏更深、波传播路径更长、波与介质之间的相互作用更加复杂、反射信号更微弱、优势频带更窄且主频降低，信噪比很低等。因此，深层超深层（包括深水）油气地震勘探亟需更具针对性的地震数据采集技术和地震波成像技术。

“两宽一高”地震数据采集是深层超深层（包括深水）油气地震勘探的基本要求。而业界正在积极发展的 FWI 反演成像方法虽然在理论上符合深层超深层目标地质体成像的需求，但在实际应用中仍存在诸多局限。事实上，针对深层超深层目标地质体的高精度成像，无论是在叠前数据采集还是在地震波成像方面，都需要发展更具针对性的方法与技术。

在叠前数据采集方面，发展低频及邻近低频的频率成分的地震波激发技术，以及被检波器保真接收的方法技术；发展针对提高极微弱反射（散射）信号的信噪比，能够有效控制非相干和随机噪声水平的方法技术；发展尽量保证不同炮检对之间激发与接收子波的一致性的采集技术，使采集得到的地震数据满足深层超深层目标地质体高保真地震成像需求。

在地震数据预处理方面，发展针对强非相干（随机）噪声与（极）弱线性信号并存的复杂情形下的去噪方法，以提高深层超深层数据的信噪比。同时，发展针对性的特征反射层相关层间多次波压制方法技术，为后续地震成像提供更为可靠的基础数据。

在地震波成像方面，发展以方位角度成像道集和倾角域道集为中心的高保真高分辨率成像处理的基本流程和关键技术，尽可能真正实现对来自地下同一成像点、不同炮检对的反（散）射子波进行同相位叠加，最终产生保真、弱旁瓣、窄主瓣的“双高”地震波成像结果。该技术体系是深层超深层（包括深水）目标地质体高精度成像的有效方法，能够有效提升目标地质体的成像精度。

参考文献

[1] 李阳, 薛兆杰, 程喆, 等.中国深层油气勘探开发进展与发展方向[J]. 中国石油勘探, 2020, 25(1): 45-57.

[2] 谢玉洪,袁全社．中国近海深水深层油气地球物理勘探实践与展望[J],石油物探, 2023, 62(2):183-193.
[3] Dyman T S, Crovelli R A, Bartberger C E, et al. Worldwide estimates of deep natural gas resources based on the US Geological Survey World Petroleum Assessment 2000[J]. Natural Resources Research, 2002, 11: 207-218.
[4] 白国平, 曹斌风. 全球深层油气藏及其分布规律[J]. 石油与天然气地质, 2014, 35(1): 19-25.
[5] 马永生, 蔡勋育, 云露, 等.塔里木盆地顺北超深层碳酸盐岩油气田勘探开发实践与理论技术进展. 石油勘探与开发, 2022, 49(1): 1-17.
[6] 王立歆, 李海英, 李弘, 等.复杂地质条件下超深层碳酸盐岩断控缝洞体成像及预测技术. 石油物探, 2022, 61(5): 865-875.
[7] TARANTOLA A. Inverse problem theory and methods for model parameter estimation[M]. Society for industrial and applied mathematics, 2005.
[8] 王华忠, 冯波, 王雄文, 等. 特征波反演成像理论框架[J]. 石油物探, 2017, 56(1): 38-49.
[9] CHENG X, XU J, VIGH D, et al. Enhanced template matching full-waveform inversion[C]. Expanded Abstracts of 84th EAGE Annual Conference & Exhibition. 2023(1): 1-5.
[10] WANG P, ZHANG Z, MEI J, et al. Full-waveform inversion for salt: A coming of age[J]. The Leading Edge, 2019, 38(3): 204-213.
[11] RAYMENT T, DANCER K, MCLEMAN J, et al. High-resolution FWI imaging-an alternative to conventional processing[C]. Expanded Abstracts of 83rd EAGE Annual Conference & Exhibition. 2022(1): 1-5.
[12] YANG Y, ARASANIPALAI S, CHEMINGUI N, et al. Simultaneous inversion of velocity and reflectivity for enhanced seismic imaging[C]. Expanded Abstracts of 2nd IMAGE Annual Conference & Exhibition, 2022: 832-836.
[13] YANG Y, RAMOS-MARTINEZ J, WHITMORE N D, et al. FWI using reflections for deep velocity model updates[C]. Expanded Abstracts of 90th SEG Annual Conference & Exhibition, 2020: 691-695.
[14] WANG B, HE Y, MAO J, et al. Inversion-based imaging: from LSRTM to FWI imaging[J]. First Break, 2021, 39(12): 85-93.
[15] Zhang Z, Wu Z, Wei Z, et al. FWI Imaging: Full-wavefield imaging through full-waveform inversion[C]. Expanded Abstracts of 90th SEG Annual Conference & Exhibition, 2020: 656-660.
[16] WEI Z, MEI J, WU Z, et al. FWI imaging: Revealing the unprecedented resolution of seismic data[C]. Expanded Abstracts of 1st IMAGE Annual Conference & Exhibition, 2021: 682-686.
[17] DAVIES D, BROWN P, AGNISOLA P, et al. Unlocking the Nile Delta subsurface via advanced processing and imaging[J]. The Leading Edge, 2024, 43(9): 596-605.
[18] MAO J, CHENG X, VIGH D, et al. Full-waveform inversion derived prestack reflectivity gathers[C]. Expanded Abstracts of 3rd IMAGE Annual Conference & Exhibition, 2023: 1797-1801.
[19] JIN H, LUO S, LIAO Q, et al. FWI Derived Reflectivity Angle Gathers[C]. Expanded Abstracts of 85th EAGE Annual Conference & Exhibition. 2024(1): 1-5.
[20] CHENG X, FENG Z, DUTTA G, et al. Least-Squares FWI-Derived Reflectivity-a New Way of Seismic Imaging[C]. Expanded Abstracts of 85th EAGE Annual Conference & Exhibition, 2024: 1-5.
[21] 李庆忠. 走向精确勘探的道路-高分辨地震勘探系统工程剖析[M]，北京, 石油工业出版社， 1993: 1-19
[22] 王华忠. “两宽一高”油气地震勘探中的关键问题分析[J]. 石油物探. 2019, 58(3): 313-324.
[23] 王华忠 , 项健 , 石聿. OBN 地震数据成像处理基本逻辑与关键方法技术[J]. 石油物探. 2024, 63(1): 12-29.
[24] LEI T, WANG H, FENG B. Broadband Acoustic Impedance Building via the Fusion of Sparsely Promoted Reflectivity and Background Acoustic Impedance[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-9.
[25] 王华忠, 吴成梁, 盛燊, 等. 针对岩性储层的定量地震波成像[J]. 石油物探, 2023, 62(1): 1-30.
[26] 王华忠,盛燊. 走向精确地震勘探的道路[J].. 石油物探. 2021, 60(5): 693-708.

再论随机模拟及在统计估计中的应用

王华忠，许荣伟，吴成梁，聂振波，韩龙祥

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】到目前为止，地震波成像的基本理论框架已经十分清楚了，它在整个石油工业界的应用现状当然也更为大家熟知。随着地震数据采集技术的快速进步，地震波成像基本上逐渐聚焦在 Bayes 参数估计理论框架下。就是假设已知正问题和实际测量数据的不确定性的先验和关于待估计参数（即地震波成像结果）的先验，基于此生成关于待估计参数的后验（概率密度），最后按照该后验（概率密度）产生估计结果（地震波成像结果）。一般地，后验（概率密度）的期望（后验概率加权平均值）作为点估计结果，对应的方差作为估计结果精度的评价。实际上，大规模参数估计中，（超）高维积分求后验（概率密度）的期望是很难做到的。因此，提出用后验（概率密度）的最大值对应的样本值作为估计结果，即 MAP 准则。MAP 准则实际上得到的是众数。后验（概率密度）Gauss 情形下，期望值与众数值是相等的。很显然，这样的估计结果是某种真实待估计参数平滑后的结果，它们分辨率是不够的。后验（概率密度）Gauss 情形下，勘探地震中就产生了梯度导引类的 FWI 方法，Newton 梯度导引类算法产生了 MAP 准则下的 FWI 反演解。很显然，进一步提高参数估计结果分辨率的方法应该是：在已知后验概率密度的条件下，生成满足该分布的随机样本，即 MAP 附近的样本值都可以作为参数估计结果。逻辑上，这极大地扩展了参数估计可行解的范围（即生成了一个可行解集），提供了评价解的精度的途径，引入其他先验信息从可行解集中选出更符合地质意义的反演解，进而提高推断和判决的可靠性。事实上，通过随机模拟还可以解决（超）高维函数的积分问题，大规模参数空间的重要性区域探测问题（即全局优化问题）等等。可以预计：梯度导引类算法之后，在超级算力的推进下，随机模拟算法会逐渐成为核心算法，更好地解决各种各样的实际问题。据此，本文重点讨论 Gauss 假设下 Bayes 参数估计算法存在的问题、随机模拟的基本思想和方法、最后讨论随机模拟在勘探地震中的可能应用。期望这样的讨论能促进随机模拟和随机反演方法在勘探地震中的深化应用。

【关键词】Bayes 参数估计，后验概率密度，期望和方差，MAP 准则，梯度导引类迭代算法，随机模拟，重要性采样算法，随机反演

1. 引言

油气地震勘探最终目的：精确地描述油气藏，进行准确的含油气性分析，做出最佳的钻井决策，得到最高的油气勘探效益。

油气地震勘探核心问题：由叠前地震数据及其它相关的先验信息，进行宽波数带的弹性参数估计（或称广义的高精度地震波成像），与井数据和岩石物理知识结合，实现储层岩石物性参数（孔隙度、渗透率等）的推断，开展精确的油气藏描述和准确的含油气性评价。

笔者认为：宽波数带的弹性参数估计，由波动方程构建起的实测波场与地下介质弹性参数（变化）之间的数学关系，应归结于 Bayes 参数估计问题，这是个模型驱动的反问题；油气藏描述及含油气性评价，缺乏（宽带）弹性参数与储层岩石物性参数之间的明确的数学关系，可归结为 Bayes 推断/判决问题，可认为这是个数据驱动的反问题。

到目前为止，地震波成像的基本理论框架已经十分清楚了，它在整个石油工业界的应用现状当然更为明确。从上世纪 60 年代开始，多次覆盖的观测方式、数据采集的数字化、计算机的出现，地震波偏移成像理论与方法的提出，催生了以地震波偏移成像为中心的地震数据处理（流程）。2000 年后，“两宽一高”地震数据采集、超大规模计算机使得逆时叠前深度偏移和 FWI 为代表的地震波成像成为核心。

目前，尽管静校正、CMP 道集叠加、速度谱解释、叠前偏移成像，成像道集层析速度建模依然在石油工业界中处于重要地位，地震波成像基本上逐渐聚焦在 Bayes 参数估计理论框架下。就是假设已知正问题和实际测量数据的不确定性的先验和关于待估计参数（即地震波成像结果）的先验，基于此生成关于待估计参数的后验（概率密度），最后按照该后验（概率密度）生成估计结果（地震波成像结果）。一般地，后验（概率密度）的期望（后验概率加权平均值）作为点估计结果，对应的方差作为估计结果精度的评价。实际上，大规模参数估计中，（超）高维积分求后验（概率密度）的期望是很难做到的。因此，提出用后验（概率密度）的最大值对应的样本值作为估计结果，即 MAP 准则。MAP 准则实际上得到的是类似众数的估计结果，就是以最大概率出现的估计结果。后验（概率密度）Gauss 情形下，期望值与众数值是相等的。很显然，这样的估计结果是某种真实待估计参数平滑后的结果，它的分辨率是不够的。后验（概率密度）Gauss 情形下，勘探地震中就产生了 FWI 方法。Newton 梯度导引类算法产生了 MAP 准则下的反演解。

很显然，提高参数估计分辨率的方法应该是：在已知后验（概率密度）的条件下，生成满足该分布的随机样本，MAP 附近的样本值，都可以作为参数估计结果。逻辑上，这极大地扩展了参数估计可行解的范围（即生成了一个可行解集），提供了评价解的精度的途径，引入其他先验信息从可行解集中选出更符合地质意义的反演解，进而提高推断和判决的可靠性。

事实上，通过随机模拟还可以解决（超）高维函数的积分问题，大规模参数空间的重要性区域探测问题（即全局优化问题）等等。梯度导引类算法之后，在超级算力的推进下，随机模拟算法会逐渐成为核心算法，更好地解决各种各样的实际问题。

据此，据此，本文重点讨论 Gauss 假设下 Bayes 参数估计算法存在的问题、随机模拟的基本思想和方法、最后讨论随机模拟在勘探地震中的可能应用。期望这样的讨论能促进随机模拟和随机反演方法在勘探地震中的深化应用。

2. Bayes 估计的理论框架及解的性质

前已述及，到现在为止地震波成像基本上逐渐聚焦在了 Bayes 参数估计理论框架下。其中的根本原因在于，与任何数据分析领域一样，实际测量的地震数据被视为随机过程（随机向量），从数据中提取相关的信息，进而进行推断，解决各种实际问题。这已经形成了大数据时代解决问题的基本模式。

而地震数据分析（简单地可认为等价于（广义的）地震波成像）的典型特征是存在实测波场与地下介质弹性参数变化之间的数学物理关系，即波动方程和它的各种变种。从叠前数据中挖掘出关于地下介质弹性参数的信息自然要用到这个强大的、因果的先验知识，尽管我们知道这个数学物理关系并不能很好地反映实测波场与地下介质弹性参数之间的真实的关系。这正是深刻理解地震波反演成像解的性质的关键。所谓的反演解仅仅是对地下介质弹性参数精度有限的、分辨率有限的估计结果。随着叠前数据质量的提升、关于待估计参数的先验信息的增加、适当提高波动方程逼近真实介质的程度（即考虑弹性、各向异性、粘滞性等），原则上可以提高弹性参数解估计的精度。

3. Bayes 参数估计的理论框架

Bayes 参数估计的根本思想是：假设已知实测数据与地下介质弹性参数之间的先验预测关系（即概率密度函数表达的不确定关系）和待估计参数满足的先验概率分布，上述两种先验分布，决定了已知实测数据与待估计参数之间的条件后验概率密度。该条件后验概率密度包含了待估计参数的所有统计特征。一般地，把条件后验概率密度的期望作为待估计参数的结果，其方差作为待估计参数的精度评价。

首先，有实测数据与地下介质弹性参数之间的先验预测关系，

$$\mathcal{M}:\ \mathbf{d}^{obs}=H\left(\mathbf{m}\right)+\boldsymbol{\varepsilon} \tag{1}$$

其中，模型 $\mathcal{M}$ 包含了实测数据与地下介质弹性参数之间的先验预测关系的不确定性，用 $p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)=p_{\boldsymbol{\varepsilon}}\left(\mathbf{d}^{obs}-H\left(\mathbf{m}\right)\right)$ 来表达。不确定性既由数据观测过程中的因素引起，也由正问题 $\mathbf{d}^{cal}=H\left(\mathbf{m}\right)$ 的不准确引起。任何或复杂或简单的正问题都不可能完整地解释实测数据或解释波在实际地下介质中的传播过程及传播结果，必定存在某种不确定性。$\boldsymbol{\varepsilon}$ 是正问题不能完全解释实测数据导致的误差。正是 $\boldsymbol{\varepsilon}$ 的随机性，刻画了实测数据与地下介质弹性参数之间的先验预测关系的不确定性。一般地，假设某种概率密度函数刻画了这种不确定性，并假设该概率密度函数是已知的。

单纯地依赖实测数据与地下介质弹性参数之间的先验预测关系的不确定性进行参数估计是最大似然估计（MLE = Maximum Likelihood Estimation）学派的思想。但是，Bayes 参数估计学派认为：仅仅依赖实测数据与地下介质弹性参数之间的先验预测关系不可能得到可靠的估计结果，必须引入待估计参数的先验信息，其中原因是实测数据本身是不完备的、不完美的，正问题也是存在不确定性的。但是，该先验信息也不可能是完全准确的（否则就不用反演成像了！），不确定性也要用某种先验概率密度函数来描述，可写为 $p\left(\mathbf{m}\middle|\mathcal{M}\right)$。很显然，参数 $\mathbf{m}$ 是正问题 $\mathbf{d}^{cal}=H\left(\mathbf{m}\right)$ 中包含的参数，与模型 $\mathcal{M}$ 是相关联的。

根据上述两种先验概率密度函数，依据 Bayes 公式，可以定义如下条件后验概率密度：

$$p\left(\mathbf{m}\middle|\mathbf{d}^{obs};\mathcal{M}\right)=\frac{p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)p\left(\mathbf{m}\middle|\mathcal{M}\right)}{p\left(\mathbf{d}^{obs}\middle|\mathcal{M}\right)} \tag{2}$$

其中，$p\left(\mathbf{d}^{obs}\middle|\mathcal{M}\right)$ 是表示实测数据本身满足的不确定的概率密度函数。

概率论指出：条件后验概率密度 $p\left(\mathbf{m}\middle|\mathbf{d}^{obs};\mathcal{M}\right)$ 包含了待估计参数 $\mathbf{m}$ 的所有统计特征。期望（一阶统计量）可以作为待估计参数 $\mathbf{m}$ 的估计结果；方差（二阶统计量）视为待估计参数 $\mathbf{m}$ 的估计精度评价。由于勘探地震中，待估计参数 $\mathbf{m}$ 有非常高的维度，积分求条件后验概率密度规定的期望和方差基本是不可能的。

因此，条件后验概率密度的最大化准则（MAP = Maximum A Posteriori Estimation）成了常用的参数估计准则。即

$$\hat{\mathbf{m}}=\arg\max_{\mathbf{m}}p\left(\mathbf{m}\middle|\mathbf{d}^{obs};\mathcal{M}\right)=\arg\max_{\mathbf{m}}\left\{p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)p\left(\mathbf{m}\middle|\mathcal{M}\right)\right\} \tag{3}$$

一般地，两个先验概率密度函数 $p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)$ 和 $p\left(\mathbf{m}\middle|\mathcal{M}\right)$ 都被假设是指数型的。已知指数型函数的对数不改变极大值的分布，因此，（3）式被写为：

$$\hat{\mathbf{m}}=\arg\min_{\mathbf{m}}\left\{-\ln p\left(\mathbf{d}^{obs}\middle|\mathbf{m};\mathcal{M}\right)-\ln p\left(\mathbf{m}\middle|\mathcal{M}\right)\right\} \tag{4}$$

显然，（4）式逐渐转化为最优化问题。解析数学家定义如下形式的最优化问题：

$$\mathcal{J}\left(\mathbf{m}\right)=Q\left(\mathbf{d}^{obs},H\left(\mathbf{m}\right)\right)+\lambda\Omega\left(\mathbf{m}\right) \tag{5}$$

其中：$Q\left(\mathbf{d}^{obs},H\left(\mathbf{m}\right)\right)$ 被称为数据偏差项，度量正问题 $\mathbf{d}^{cal}=H\left(\mathbf{m}\right)$ 预测实测数据 $\mathbf{d}^{obs}$ 的误差；$\Omega\left(\mathbf{m}\right)$ 定义待估计参数的先验信息。

很显然，若定义（4）式中的对应项为：

$$Q\left(\mathbf{d}^{obs},H\left(\mathbf{m}\right)\right)=-\ln p\left(\mathbf{g}\middle|\mathbf{m};\mathcal{M}\right) \tag{6a}$$

$$\Omega\left(\mathbf{m}\right)=-\ln p\left(\mathbf{m}\middle|\mathcal{M}\right) \tag{6b}$$

可以看出，数学分析学家定义的参数优化反演问题（一般地，称 Tikhonov 正则化问题）与两个先验概率密度函数假设为指数分布的 MAP 估计问题是一致的。

更进一步地，在线性 Gauss 假设下，即存在实测数据与地下介质弹性参数之间的线性预测关系式 $\mathbf{d}^{obs}=\mathbf{d}^{cal}+\varepsilon=H\mathbf{m}+\varepsilon$，并有如下的两个先验概率密度函数满足 Gauss 分布的假设，

对噪声的假设：

$$\varepsilon\sim N\left(\mathbf{0},\sigma_{\varepsilon}^{2}\mathbf{I}\right)\ \text{即}\ p\left(\mathbf{d}^{obs}\middle|\mathbf{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\varepsilon}^{2}}\left\|\mathbf{d}^{obs}-H\mathbf{m}\right\|^{2}\right] \tag{7a}$$

和对 $\mathbf{m}$ 的假设：

$$\mathbf{m}\sim N\left(\mathbf{0},\sigma_{\mathbf{m}}^2\mathbf{I}\right)\text{即 }p\left(\mathbf{m}\right)\propto\exp\left[-\frac{1}{2\sigma_{\mathbf{m}}^2}\left\|\mathbf{m}\right\|^2\right] \tag{7b}$$

则对应的后验概率密度为

$$p\left(\mathbf{m}\middle|\mathbf{d}^{obs}\right)\propto\exp\left[-\frac{1}{2\sigma_{\boldsymbol{\varepsilon}}^2}\left\|\mathbf{d}^{obs}-H\mathbf{m}\right\|^2+\frac{\sigma_{\boldsymbol{\varepsilon}}^2}{\sigma_{\mathbf{m}}^2}\left\|\mathbf{m}\right\|^2\right] \tag{8}$$

类似地，有 MAP 估计下的反演解

$$\hat{\mathbf{m}}=\arg\max_{\mathbf{m}}\left\{p\left(\mathbf{m}\middle|\mathbf{d}\right)\right\}=\arg\max_{\mathbf{m}}\left\{\mathcal{J}\left(\mathbf{m}\right)\right\} \tag{9}$$

其中，误差泛函定义为

$$\mathcal{J}\left(\mathbf{m}\right)=-\frac{1}{2}\left\|\mathbf{d}^{obs}-H\mathbf{m}\right\|_{\sigma_{\boldsymbol{\varepsilon}}^2}^2+\lambda\left\|\mathbf{m}\right\|^2 \tag{10}$$

超参 $\lambda=\dfrac{\sigma_{\boldsymbol{\varepsilon}}^2}{\sigma_{\mathbf{m}}^2}$。（10）式的解可以写为：$\mathbf{m}\middle|\mathbf{d}^{obs}\sim\mathcal{N}\left(\hat{\mathbf{m}},\hat{\boldsymbol{\Sigma}}\right)$。其中，$\hat{\mathbf{m}}=\left(H^TH+\lambda I\right)^{-1}H^T\mathbf{d}^{obs}$，$\hat{\boldsymbol{\Sigma}}=\sigma_{\boldsymbol{\varepsilon}}^2\left(H^TH+\lambda I\right)^{-1}$分别为后验概率密度函数的期望和方差。

Bayes 参数估计的基本逻辑框架基本上就是这样。可以看出，炮集波场用算子形式的波动方程 $\mathbf{d}^{cal}=H\left(\mathbf{m}\right)$来预测时，对应的参数估计问题就是 FWI。若引入 Born 近似后的正问题$\mathbf{d}^{cal}=H\mathbf{m}$ 预测炮集波场，对应的参数估计问题就是最小二乘逆时深度偏移（LS_RTM）。可见，FWI/LS_RTM 有一个十分优美而简洁的理论框架。

4. Bayes 参数估计的解的性质的分析

Gauss 先验分布假设+MAP 准则+误差泛函最小+梯度导引类算法，得到既定问题的反演解，数学理论上，就是 $\hat{\mathbf{m}}=\left(H^TH+\lambda I\right)^{-1}H^T\mathbf{d}^{obs}$，$\hat{\boldsymbol{\Sigma}}=\sigma_{\boldsymbol{\varepsilon}}^2\left(HH+\lambda I\right)^{-1}$。若 $\hat{\mathbf{m}}=\left(H^TH+\lambda I\right)^{-1}H^T\mathbf{d}^{obs}=\left(H^TH+\lambda I\right)^{-1}H^TH\mathbf{m}_{true}=R\mathbf{m}_{true}$，明显地，估计解也是真解的 PSF 模糊结果，$R$ 是模糊算子。很显然，Bayes 后验概率密度求期望给出的参数反演解是某种加权平均值，在提高反演可靠性的同时，导致了分辨率降低。目前，FWI 和 LS_RTM 的具体实践中，重点放在了如何获得反演解上，对反演解的精度和分辨率的评价较少涉及。但是，目前已经到了需要花精力评价反演解可靠性和精度的时间了。

众所周知，提高反演解分辨率（精度）的思想与方法是清楚的：期望实测数据尽可能是完美的、满足高精度反演需求的；提高正问题对实测数据的预测精度；期望预测误差是 Gauss 分布（ Gauss 白噪）的；期望正问题关系是尽可能线性的；期望初始值是尽可能接近真解的；施加合理的正则化约束；期望有尽可能多的先验信息；……。

但是，反问题的多解性、非唯一性是永恒的。反演成像的实践表明：即便使用了上述提高反演解分辨率（精度）的思想与方法，依然可能达不到期望的精度和分辨率要求。除了继续沿上述已知的路线努力外，还有什么提高反演结果分辨率（精度）的新途径？

随机反演可能是值得探索的技术路径！

5. 随机模拟的基本思想

物理世界的变化规律到底是确定性的？还是随机性的？可以认为宏观上是确定性的(Newton 力学规律)，微观上是随机的（量子力学规律）。大量物理粒子的规律性的行为特征也要用概率统计方法描述，譬如分子热力学。

机械的、确定性的、因果的观点描述世界的方法显然是片面的、不完全可取的。基于随机、概率统计基础上的认识论描述客观世界并解决客观问题是不可或缺的。

Newton 力学、数学分析（泛函分析）、线性代数、数值分析/数值算法（尤其梯度导引类的数值优化算法）提供了强大的分析问题的逻辑和方法。但是，也固化了我们的思维。根本上看，物理世界、生物世界、人类社会处处存在不确定性。越是微观地、细尺度地、多因素相互作用、非线性变化的情形下，不确定性或随机性的表现越普遍。宏观的、粗尺度的、单因素（少因素）及线性的情形下，确定性的表现越明显。由基于确定性的、因果的关系进行预测、做出决策的方式方法要逐渐地让位于由基于随机统计性的、相关性的关系进行预测，并做出决策的方式方法。

随机性是物理世界、生物世界、人类社会的客观特征，它是由样本的取值满足一定的概率分布体现出来的。若样本的取值是确定性的，表明事件都按概率 1 发生，事件没有不确定性。计算机模拟（仿真）确定性关系，从而提升我们对物理及非物理实体的认识，也帮助我们做出估计推断。这是我们所熟悉的。

但是，计算机能完全模拟随机过程吗？随机模拟为什么能帮助我们解决问题？随机模拟能解决哪些问题？

计算机显然不能真正地或完全地仿真物理及非物理世界的随机过程。基本原因是：我们自己都不理解随机过程产生的过程，只知道其结果。数学上，没有理论证明能在计算机上真正地或完全地仿真物理世界的随机过程。数学家可以做到在计算机生成伪随机数。即：给定一个初值，在很长的周期内，不会重复生成相同的随机数序列。伪随机数与物理世界的随机过程是一致的吗？显然是不一致的。

另一方面，科学上，人类的确有必要发展出能模拟物理及非物理世界随机过程的算法。即便做不到，也可以用伪随机模拟的思想与方法帮助我们更好地解决当前遇到的各种各样的问题。

（伪）随机模拟到底模拟什么？模拟满足一定概率密度分布的随机变量的生成过程。本质上，就是在计算机上模拟满足一定概率密度分布的随机变量（随机过程）的样本生成。其中概率分布的形式可以是多种多样的，对应的随机样本的物理及非物理场景下的具体含义也是多种多样的。可以看出，（伪）随机模拟有广泛的应用性。

（伪）随机模拟的关键是：伪随机数的生成。一般地，独立同分布地生成[0,1]空间上均匀分布的随机变量的样本。伪随机数的生成是随机模拟的基础。在计算机上模拟满足任意概率密度函数的随机变量（随机过程）的样本生成，离不开伪随机数产生。伪随机数的生成有很多现有的算法程序。

随机模拟要解决的核心问题：在既定的、巨大的样本空间中，按既定的（后验）概率密度函数，进行重要性采样。这显然是高维参数估计中误差泛函的全局寻优问题。重要性采样也是解决高维积分问题的核心思想。

还有很多其他的应用场景。大数据和人工智能时代，这样的应用场景多种多样。

6. 随机反演的基本思想

前面已经讨论了 Bayes 参数估计的基本逻辑。FWI 和 LS_RTM 最后归结为梯度导引类的目标泛函求极小的数值优化问题。但是，$\hat{\mathbf{m}}=\left(H^TH+\lambda I\right)^{-1}H^T\mathbf{d}^{obs}$，$\hat{\boldsymbol{\Sigma}}=\sigma_{\boldsymbol{\varepsilon}}^2\left(H^TH+\lambda I\right)^{-1}$表达的点估计结果是真解的模糊结果，分辨率不高。方差计算比较困难，难以对点估计结果进行有效的精度评价。因此，有必要进行进一步的随机反演来提高最终反演解的分辨率及精度。

Bayes 估计框架下的随机反演，基本思想可以描述为：基于 Bayes 推断的基本逻辑，由后验概率密度对参数进行估计，具体地即通过高维积分求参数的各阶统计量，尤其是一阶二阶统计量。由于高维情况下积分

无法完成，需要进行基于重要性采样的随机模拟，生成满足后验概率密度的参数样本，对生成的参数样本进行统计量的估算及精度和分辨率的评价。

随机反演的目的：生成满足既定后验概率密度的参数样本。

随机反演的核心步骤：

①、进行基于重要性采样的随机模拟，生成满足既定后验概率密度的一族参数样本；

②、进一步引入先验信息，从一族参数样本中挑选出更符合实际意义的参数估计结果。并且评估反演解的精度及可靠性。

7. 典型的随机模拟方法

与求解波动方程定解问题进行波传播过程这种确定性的模拟不同，随机模拟是对既定的概率密度函数 $P(\mathbf{x})$ 进行对应样本 $\mathbf{x}$ 的生成。这是两个完全不同的概念。两种模拟的方式方法也完全不同，应用的领域也非常不同。前者是波动方程的数值求解，已知波动方程、震源函数和初边值条件，波动方程的数值解是确定性的；后者是概率密度函数 $P(\mathbf{x})$ 对应的样本的分布，它是按概率密度函数 $P(\mathbf{x})$ 随机生成的。

随机模拟算法的基本逻辑如下：

1、 直接采样方法

首先，使用计算机可模拟的辅助概率密度函数（通常为均匀分布）进行采样（即生成随机数），然后通过某种映射（如反函数映射）生成目标概率密度函数对应的样本。然而，这种方法的效率通常较低。

2、基于 Markov 链的采样方法

为了提高效率，引入了 Gibbs 采样算法或 Metropolis-Hastings 重要性采样算法。这两种算法均基于 Markov 链设计，其核心思想是从任意初始随机采样出发，通过特定的映射规则，最终收敛到目标概率密度函数，并生成对应的随机样本。

其中，Gibbs 采样的核心思想可以提炼如下：1、分步更新，即每次只更新一个变量，条件于其他变量的当前值。2、条件分布，利用联合分布的条件分布进行采样，逐步逼近目标分布。3、迭代收敛，通过多次迭代，样本逐渐收敛到目标联合分布。4、简化高维问题，将高维采样问题分解为一系列一维采样问题，降低复杂度。总之，Gibbs 采样通过分步更新和条件分布，逐步逼近目标分布，简化高维采样问题。而 Metropolis-Hastings 采样的核心思想可以提炼如下：1、提议（参考）分布，从提议（参考）分布中生成候选样本。2、接受概率，根据接受概率决定是否接受候选样本，确保收敛到目标分布。3、迭代更新，通过迭代生成样本序列，逐步逼近目标分布。4、灵活性，提议（参考）分布的选择灵活，适用于多种复杂分布。总之，Metropolis-Hastings 采样通过提议（参考）分布和接受概率，迭代生成样本，灵活高效地逼近目标分布。

下面具体讨论反函数法和 Metropolis-Hasting 重采样方法的基本思想。

8. 反函数法实现随机模拟的解释

随机模拟的基本问题可以描述为：已知概率密度函数 $f(x)$ 及其累计分布函数 $F(x)=\int_{-\infty}^{x} f(t)dt$，求取一组随机数 $\{x_0, x_1, x_2, ..., x_N\}$ 使其满足概率密度函数 $f(x)$ 的分布。

使用反函数方法来进行随机模拟的基本思想是：令 $y=F(x)\sim U(0,1)$，对 y 进行 0-1 的随机均匀采样，然后计算反函数映射 $x=F^{-1}(y)$ 就可以得到满足 $f(x)$ 概率密度分布的随机数。因此，满足均匀分布 $U(0,1)$ 的随机数生成成了随机模拟的基础。

为了理解反函数的基本思想，不妨先从简单离散情况下来理解这一问题：

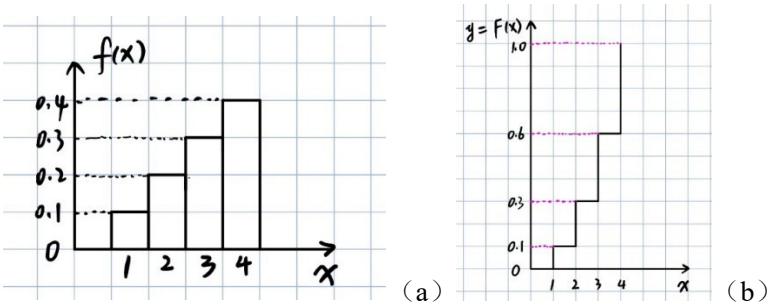
假设随机数 x 的可能取值为 1，2，3，4。其概率密度函数为

$$f(x)=\begin{cases}0.1 & x=1\\ 0.2 & x=2\\ 0.3 & x=3\\ 0.4 & x=4\end{cases} \tag{11}$$

则可以计算出其累计分布函数

$$F(x)=\begin{cases}0.1 & x=1\\ 0.3 & x=2\\ 0.6 & x=3\\ 1.0 & x=4\end{cases} \tag{12}$$

将公式（11）与公式（12）进行图形化，得到图 1。



（a） （b）

图 1 概率密度图（a）及其累计概率密度图（b）

从图 1b 中可以看出，累计分布函数的范围为 $y\in[0,1]$，并且累积分布函数的取值被分为了四段，每一段的长度正好对应 x 取值的概率密度函数 $f(x)$ 的大小。因此，假设 $y\sim U(0,1)$，即 y 服从 0 到 1 的均匀分布，对 y 进行随机采样，通过采样值 y_i 所在的范围，即可得到 x_i，并且使得到的 x_i 序列满足概率密度函数 $f(x)$。即：如果 $y_i\in[0,0.1)$，$x_i=1$；如果 $y_i\in[0.1,0.3)$，$x_i=2$；如果 $y_i\in[0.3,0.6)$，$x_i=3$；如果 $y_i\in[0.6,1.0]$，$x_i=4$。因此有，$f(x)$ 的反函数为

$$x=F^{-1}(y)=\begin{cases}1 & 0.0\le y<0.1\\ 2 & 0.1\le y<0.3\\ 3 & 0.3\le y<0.6\\ 4 & 0.6\le y\le 1.0\end{cases} \tag{13}$$

据此可知，只需要对 y 进行 $[0,1]$ 内的均匀随机采样，然后通过公式（13）进行反变换，就可以得到满足 $f(x)$ 的随机采样序列。

通过离散情况下的简单示例，已基本对通过累积分布函数的反函数进行随机采样有了一个基本的认识，接下来再考虑连续情况下的累计分布函数的反函数法实现随机采样的想法。与前面描述的一样，已知概率密度函数 $f(x)$ 及其累计分布函数 $F(x)=\int_{-\infty}^{x}f(t)dt$，求取一组随机数 $\{x_0,x_1,x_2,.....,x_N\}$ 使其满足概率密度函数 $f(x)$ 的分布，此处 $f(x)$ 满足连续性。

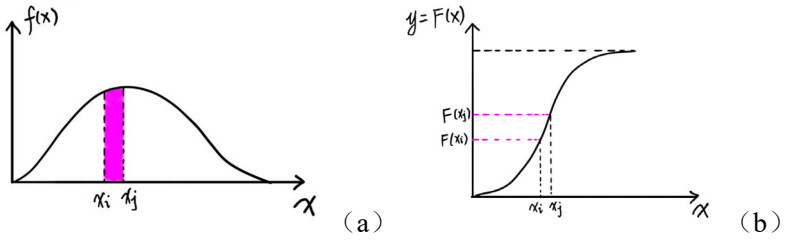


图 2 连续概率密度函数及其累积概率密度函数

为了更好的理解概念，在这里，引入两个变量 x_i 和 x_j，图 2a 中桃红色区域表示的是随机数区间 $[x_i, x_j]$ 内所有的随机数采样对应的概率密度大小。在累计分布函数图中（图 2b）中，$[x_i, x_j]$ 对应的是 $F(x_j)-F(x_i)$ 长度的大小，当 x_i 趋近于 x_j 时，

$$\lim_{x_i \to x_j} \frac{F(x_j)-F(x_i)}{x_j - x_i} = F'(x_j) = f(x_j) \tag{14}$$

即当 x_i 逐渐趋近于 x_j 时，则在累计分布函数中，（14）式的极限逐渐趋近于累计概率密度函数的导数值，也就是 x_j 处的概率密度函数值 $f(x_j)$。这就建立起了 $[0,1]$ 区间均匀分布映射成任意概率密度函数的理论基础。

与离散的情况相同，也可以认为区间长度 $[x_i, x_j]$ 与 $[F(x_i), F(x_j)]$ 之间的映射关系（14）式（极限情况下）定义了 x_j 概率密度函数的大小 $f(x_j)$。因此，同样道理，对 y 进行 $[0,1]$ 内的均匀随机采样，然后通过累积分布函数的反函数就可以得到服从概率密度为 $f(x)$ 的随机采样序列。

首先通过一个简单的正态分布函数的数值示例来进行分析，概率密度函数为均值为 0，方差为 1 的正态分布 $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$，如图 3 绿色虚线所示，其累计分布函数为 $y=F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$，如图 3 黑线所示。图 3 展示了累计分布函数值 y 从 0 到 1 间隔 0.01 进行均匀采样（竖轴蓝点所示），及其反函数映射之后对应的采样点 x（横轴红点）的分布。可以看出当 y 均匀分布时，红点的密度和概率密度函数（绿色虚线）的大小成正比例关系，红点密集的地方，概率密度大，红点稀疏的地方，概率密度小。因此可以说累计分布函数改变了累计分布函数值 y 与采样值 x 的数据密度关系，自动地将均匀分布的累计分布函数值 y，通过累计分布函数的反函数映射成符合概率密度函数 $f(x)$ 的随机采样序列。

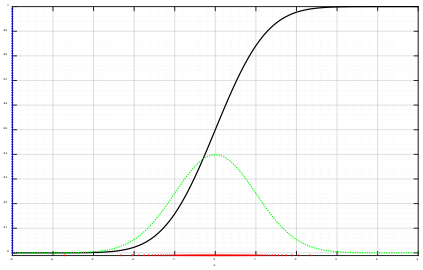


图 3 正态分布函数及其累计分布函数采样值 x 与累计分布函数值 y 的数据密度对应关系

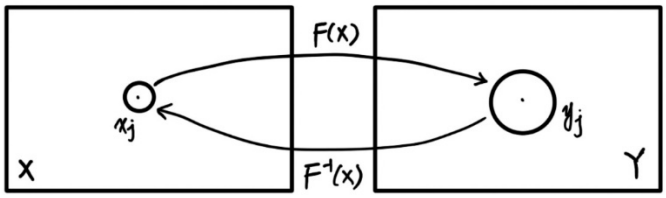


图 4 采样值 x 与累积分布函数值的映射关系图

从前面的示例可以看出，在连续情况下，对应的是数据密度的问题，$y_j = F(x_j)$，在 x_j 点，累积分布函数的导数值 $f(x_j) = F'(x_j)$ 的大小，表示了 y_j 在 x_j 点局部小邻域内的变化快慢，$f(x_j)$ 值越大，则 y_j 变化得越快，也就是在 x_j 点小的变化引起 y_j 大的变化，x_j 的局部小邻域对应 y_j 的一个局部大邻域，如图 4 所示。

如果 y 空间中的采样是均匀采样，当 $\frac{F(x_j)-F(x_i)}{x_j-x_i}$ 比较大时，y_j 采样点的邻域内的所有值都会被压缩到 x_j 的小邻域内，从而实现了对 $f(x)$ 的重要性采样。显然，x 空间中的邻域与 y 空间中的邻域的缩或放取决于概率密度函数 $f(x)$。

因此，本质上，$f(x)$ 积分得到的累积分布函数 $y = F(x)$ 决定了 x 与 y 的映射关系，并且基于对 y 的均匀采样 $U(0,1)$，通过其反函数，映射生成对 x 的采样，最终实现对 $f(x)$ 的重要性采样。遗憾的是，利用反函数法进行随机模拟的限制是：反函数并非都是存在的。理论上，只有函数 $F(x)$ 是一对一的映射时，其反函数才是存在的。

无论其是否很实用，从概念上讲，反函数法清楚地展示出了如何从一个均匀分布 $U(0,1)$ 的采样，通过反函数，映射成既定的、其反函数存在的概率密度函数的采样。

为了发展出更一般和通用的随机模拟方法技术，需要提出不同的逻辑理念。最基本的逻辑出发点依然是通过均匀分布 $U(0,1)$ 的采样（这是一种提议（参考）分布的采样），设计一种算法，逐渐地逼近既定的、（原则上）任意的概率密度函数对应的采样。一般地，我们称这样的算法为 Metropolis-Hastings 重要性采样算法。

9 再议 Metropolis-Hastings 重采样算法的思想

9.1 基本思想

重要性采样是随机反演的核心。它是通过重采样算法实现的。关于重要性采样和重采样的关系需要简要地说明一下。重要性采样（Importance Sampling）和重采样（Resampling）是两个相关但不同的概念，尽管它们在某些情况下可能被混淆或交替使用。重要性采样是一种用于估计既定概率密度函数对应的期望值的统计方法，特别适用于从难以直接采样的概率分布中获取样本。其核心思想是通过引入一个易于采样的提议（参考）分布（proposal distribution），并调整样本的权重来修正偏差，从而估计目标分布的期望值。重采样是指从已有样本中再次抽取样本的过程，常用于调整样本分布或减少方差。在重要性采样中，重采样通常用于根据重要性权重调整样本分布，使其更接近目标分布。为什么重要性采样又叫重采样？在某些情况下，重要性采样和重采样可以结合使用。例如，在粒子滤波（Particle Filter）中，重要性采样用于生成带权重的粒子，而重采样则用于根据这些权重重新抽取粒子，以避免权重退化问题。因此，重采样可以被视为重要性采样的一部分或后续步骤。总而言之，重要性采样是通过调整权重从提议（参考）分布中估计目标分布的期望值；重采样是根据权重重新抽取样本，以调整样本分布。两者结合使用时，重采样作为重要性采样的后续步骤，帮助维持样本的有效性，因此有时会统称为重采样。

对于一个已知的概率密度函数 $g(x)$，其解析式相对复杂，譬如不存在反函数，如何对其进行随机采样，使得采样得到的随机数，满足其概率密度 $g(x)$？

Metropolis-Hastings 重采样算法的核心思想是假设存在一个提议（参考）概率密度 $f(x)$，在任意 x 位置，$c\cdot f(x)>g(x)$。此处 c 为一比例系数（为了用 $f(x)$ 把 $g(x)$ 包起来。应该注意到：概率密度函数都是分布在区间 $[0,1]$ 中，且都是大于零的，找一个参考概率密度函数把目标概率密度函数包起来是很容易的。c 的选择原则是能用 $f(x)$ 把 $g(x)$ 包起来）。还要假设满足概率密度函数 $f(x)$ 的随机数（随机采样值）是容易生成的。这样就则可以通过对概率密度函数 $f(x)$ 的随机数（随机采样值）进行处理，来实现对概率密度函数 $g(x)$ 的随机采样。

如何实现这一过程？

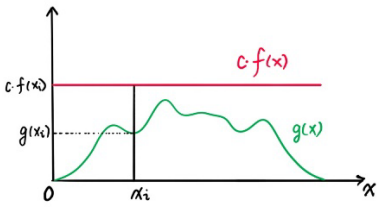


图 5 重采样概率密度函数示意图，红线为 $f(x)$ 概率密度函数，绿线为 $g(x)$ 概率密度函数

如图 5 所示，不妨我们假设概率密度函数 $f(x)$ 对应一均匀分布 $U(0,1)$，$g(x)$ 为图中绿线所示，首先 $f(x)$ 进行随机采样。如果对于采样到的所有随机数（随机样本值）x，按照 $\frac{g(x)}{c\cdot f(x)}$ 比值进行筛选（所谓筛选，就是按一定规则保留或舍弃），则保留下来的样本，可以满足 $g(x)$ 的概率分布。更通俗来讲，对 $f(x)$ 的随机采样数是红线的分布，在按照比值 $\frac{g(x)}{c\cdot f(x)}$ 进行保留后，绿线之上的随机数都被剔除了，只保留下来绿线之下的随机数，使之可以满足 $g(x)$ 的概率分布。可以看出：$g(x)$ 越大，按 $\frac{g(x)}{c\cdot f(x)}$，采样值被保留下来的概率越大；反之，$g(x)$ 越小，按 $\frac{g(x)}{c\cdot f(x)}$，采样值被保留下来的概率也越小。随着按照均匀分布进行采样的采样点数的增加，留下来的采样点逐渐地集中于 $\frac{g(x)}{c\cdot f(x)}$ 较大的区域。很显然，这些区域是 $g(x)$ 比较大的区域。

可以看出，比值 $\frac{g(x)}{c\cdot f(x)}$ 也可以改成含义更为明显的比值 $\frac{cf(x)-g(x)}{cf(x)}$。$\frac{cf(x)-g(x)}{cf(x)}$ 的分子项越大，即比值 $\frac{cf(x)-g(x)}{cf(x)}$ 越大，采样值被拒绝接受的概率越高。反之，被拒绝接受的概率越低。随着按照均匀分布进行采样的采样点数的增加，留下来的采样点逐渐地集中于 $\frac{cf(x)-g(x)}{cf(x)}$ 分子项较小的区域。很显然，这些区域是 $g(x)$ 比较大的区域。

与反函数法思想相比，反函数法是用分布函数的反函数实现对均匀分布 $U(0,1)$ 采样的挤压或拉伸从而完成对相应的概率密度函数的重要性采样。上述重采样的思想是通过引入一个所谓的权重因子

（$w=\dfrac{g(x)}{c\cdot f(x)}$）按照拒绝与接受准则把均匀分布$U(0,1)$采样进行挤压或拉伸从而完成对目标概率密度函数的重要性采样。

总结一下，重采样的想法是，

step1：通过均分分布$U(0,1)$（对应的概率密度函数为$f(x)$）的随机采样，得到一个随机数x_i，

step2：然后按照比值$\dfrac{g(x_i)}{c\cdot f(x_i)}$决定是否保留或拒绝接受当前的随机数$x_i$，这里为了决定是否保留，需要再生成一个随机数$t_i\sim U(0,1)$，如果$t_i\le\dfrac{g(x_i)}{c\cdot f(x_i)}$，则保留$x_i$，否则，则舍弃$x_i$。引入随机数$t_i$的目的是依概率接受或拒绝随机数$x_i$。

不断重复步骤 1 与步骤 2，则可以得到一组满足$g(x)$概率密度的随机数。

很显然，依据均匀分布在$[0,1]$区间内大规模采样后，在上述接收与拒绝原则的作用下，留下来的采样点会逐渐地逼近$g(x)$的样本分布规律。即高概率的地方，样本出现概率大；低概率的地方，样本出现概率低。这就实现了重要性采样的目的。

很显然，Metropolis-Hastings 重采样思想和方法可以解决高维积分问题和推断问题（即高概率出现的样本（均值）是我们推断的目标）。

但是，在高维参数估计问题或全局寻优问题中，通常引入 Monte Carlo 模拟退火方法。很显然，它就是上述 Metropolis-Hastings 重要性采样方法的一种改型。重要的思想是 Monte Carlo 模拟退火方法引入了 Boltzmann 分布。

下面简单地分析一下，在 Monte Carlo 模拟退火方法中引入 Boltzmann 分布的原因。模拟退火方法通过概率接受准则允许接受较差的解，以避免陷入局部最优。Boltzmann 分布提供了这种概率依据，其公式为：$P(\Delta E)=esp\left(-\dfrac{\Delta E}{kT}\right)$，其中，$\Delta E$是能量变化，$k$是 Boltzmann 常数，$T$是温度。很显然，引入 Boltzmann 分布不仅仅提供了概率依据，还通过把实测数据与合成数据逼近误差作为 Boltzmann 分布中的能量ΔE构建起了判断反演解是否可接受的准则。进一步地还引入了温度控制。温度T控制接受较差解的概率。初始高温时，接受概率较高，便于广泛搜索；随着温度降低，接受概率减小，逐渐收敛到全局最优解。模拟退火借鉴了物理退火过程，Boltzmann 分布有明确的物理含义，它描述了系统在热平衡下的状态分布，帮助算法模拟物理退火行为，从而达到全局最优解。通过 Boltzmann 分布，算法在高温时接受较差解的概率较高，增强了全局搜索能力；低温时则倾向于局部优化，逐步收敛。总地来说，引入 Boltzmann 分布是为了提供概率接受准则，控制温度对搜索过程的影响，增强全局搜索能力，并模拟物理退火过程。Monte Carlo 模拟退火方法显然是对上述 Metropolis-Hastings 重采样思想和方法的修改，把它用在解决全局寻优问题上。

众所周知，Monte Carlo 模拟退火方法进一步地还可以与生物遗传算法结合。在 Monte Carlo 模拟退火方法中结合遗传算法的主要原因如下：1、增强全局搜索能力，模拟退火依赖单一解进行迭代，容易陷入局部最优；遗传算法通过种群多样性进行全局搜索，能更有效地探索解空间。2、提高收敛速度，模拟退火收敛速度较慢，尤其在解空间复杂时；遗传算法通过选择、交叉和变异操作，能更快找到较优解。3、平衡探索与开发，模拟退火通过温度控制探索与开发，但效果有限；遗传算法通过种群操作，更好地平衡探索新区域与开发已知区域。4、避免早熟收敛，模拟退火方法容易早熟收敛；遗传算法的种群多样性有助于避免早熟收敛，保持搜索活力。5、适应复杂问题，模拟退火方法适合连续和离散优化，但对复杂问题效果有限；遗传算法能处理多模态、非线性等复杂问题，结合后效果更佳。因此，把二者结合起来是很自然的想法。结合方式包括：1、混合策略，先用遗传算法进行全局搜索，再用模拟退火进行局部优化；2、并行执行，同时运行两种算法，定期交换信息；3、嵌入操作，在遗传算法中嵌入模拟退火，尤其借用其温度控制机制。笔

者实践认为嵌入操作是更合理的做法。总之，结合遗传算法能增强全局搜索能力、提高收敛速度、平衡探索与开发、避免早熟收敛，并更好地适应复杂问题。

9.2 数值示例

通过重采样对均值为 0，方差为 1 的高斯分布进行随机采样，选择均匀随机采样生成最初始的随机采样数，然后通过重采样获得高斯随机分布的随机数。图 4 展示了随机采样的结果。蓝色直方图为通过重采样得到的采样数，红线为均值为 0，方差为 1 高斯概率密度函数，黑色虚线为统计的随机采样样本的概率密度函数。从图 4 中可以看出，在样本较小时，采样的结果相对较差，随着样本的逐渐增大，红线和黑色虚线逐渐重合，采样得到的随机数满足均值为 0，方差为 1 的高斯分布。

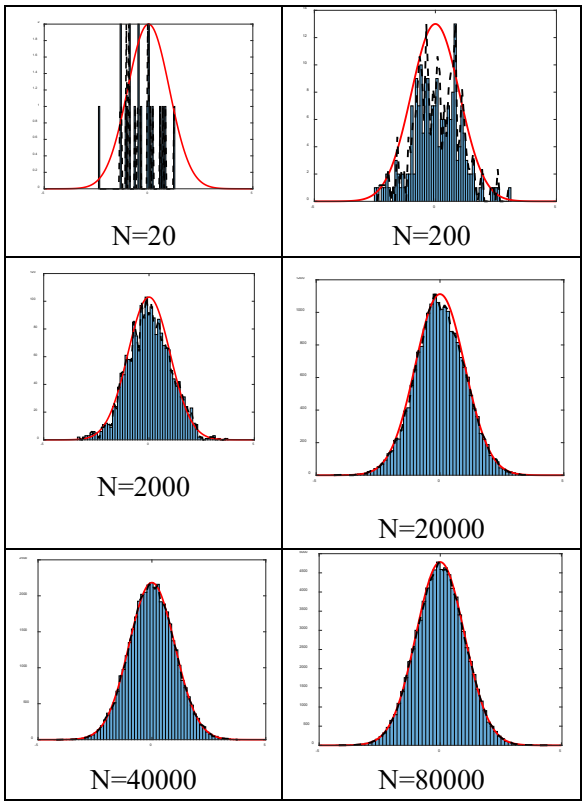


图 6 重采样获取均值为 0，方差为 1 的高斯随机采样数，N 为采样的样本数

10 勘探地震中的随机反演基本思想与做法

前已述及，随机反演（Monte Carlo 反演）是指：在可能的解空间中，进行重要性随机采样，找到临近 MAP 的区域，生成该区域中的随机变量样本，然后用合适的方法生产或挑选出合理的解并对解进行评价。其中，Monte Carlo 算法的核心是重要性随机采样。Metropolis - Hasting 采样算法是目前典型的重要性采样方法。借助生物进化方法是改善重要性采样效率和性能的必然途径。

勘探地震中，如何利用上述随机反演的思想？

首先，基于叠前地震数据和待估计参数的先验信息进行高精度的、宽带的弹性参数估计，是个高度非线性的问题。目前的做法是 Gauss 分布+非线性（线性）正问题假设+Newton 梯度导引类算法给出的反演解，理论上它是条件后验概率密度加权平均的结果，根据目前的勘探实践，其分辨率很难达到高精度油藏描述的需求，且基于后验概率密度函数的协方差阵评价反演解的精度也非常困难。

因此，把随机反演的思想引入到弹性参数估计（或地震波成像）中，是合乎逻辑的选择。尤其，在巨大的计算机算力支撑和大数据驱动的多信息提取用来约束反演解的算法逐渐成为常规的情形下，借用随机反演的思想和算法提升参数估计的精度和分辨率逐渐成为可能。

勘探地震中的随机反演基本思想与做法大致可以总结如下：

① 对于强非线性问题的后验概率密度函数，在其可能的解空间中，进行重要性采样，找到临近 MAP 的区域，生成该区域中的随机变量样本，用合适的方法得到合理的解并对解进行评价。这样的做法还可以与梯度导引类方法串联起来使用。

另外，基于类 CMP 道集 Monte Carlo 反演提供初始模型，然后串联梯度导引类层析反演方法也是符合逻辑的选择。

② 对于弱非线性问题，MAP 准则下的梯度导引算法，给出某种加权平均的期望结果。在期望解附近，在 Gauss 后验概率分布控制下，随机生成超分辨率模型参数样本族，在合理的准则下，选择超分辨率反演结果。理论上，这是超分辨率反演一条合理的途径。我们认为：用于储层描述的随机反演应该按此逻辑进行。

③ 对于有可靠的井约束的情形，反射结构约束（最好是地震相约束）井数据插值方法得到伪井结果。以伪井结果作为初始值和/或先验约束值，用梯度导引类算法给出某种加权平均的期望结果。在期望解附近，在 Gauss 分布控制下，随机生成超分辨率模型族，在合理的准则下，选择超分辨率反演结果。这应该是井控情形下，用于储层描述的超分辨率反演的合理逻辑路线。原则上，这样做的超分辨率反演，其结果可以达到采样间隔的尺度。但是，如何选择出合乎地质逻辑的反演解目前依然是没有很好解决的问题！

对勘探地震中的随机反演基本思想与做法的基本评价：随机反演（Monte Carlo 反演）既可以用于解决强非线性反问题的全局寻优，也可以生成后验概率密度对应的期望值附近的一族可行的反演解。前者计算量太大，用于目前的大规模参数估计不一定可行。后者用于得到超分辨率（可达任意采样尺度）的反演解，存在的困难是如何生成一族 MAP 附近的可行解及如何从中选出一个有地质意义的解。

当前，FWI 或任何梯度导引类算法给出的反演解，得到满足储层描述要求的、有地质意义的解都是有困难的。无论如何，随机反演算法可以给出一族 MAP 附近的可行解，提供了选择出更高分辨率可行解的机会，也提供了评价解的精度的随机采样样本，因此引入随机反演应该是值得探索的路线。

11 地质统计学反演与随机反演的关系的认识

笔者认为：地质统计学反演是随机反演在勘探地震中的实例，而随机反演是 Monte Carlo Bayes 推断的具体做法。前面已经讨论过随机反演的本质是按照后验概率分布，生成 MAP 附近的随机样本，然后从这些随机样本中选择或生成反演结果，并对反演结果进行精度评价。

众所周知，地质统计学的起源来自于 Kriging 插值，其本质是已知散乱的空间数据，推断出规则的空间数据。Kriging 插值的本质思想是把散乱的空间数据视为随机的，并把它分为三部分，即 $\mathbf{d}(\mathbf{x})=\mathbf{d}_{BC}(\mathbf{x})+\mathbf{d}_{DC}(\mathbf{x})+\mathbf{d}_{RC}(\mathbf{x})$。其中，$\mathbf{d}_{BC}(\mathbf{x})$ 是空间数据的背景成分（Background Component）；$\mathbf{d}_{DC}(\mathbf{x})$ 是空间数据的扰动成分（Disturbance Component）；$\mathbf{d}_{RC}(\mathbf{x})$ 是满足一定概率分布的随机成分（Random Component）。$\mathbf{d}_{BC}(\mathbf{x})$ 可以视为是确定性的、散乱点支撑的函数。这部分的描述可以用径向基函数族的线性组合的方法；$\mathbf{d}_{DC}(\mathbf{x})$ 用零均值、一定方差的概率密度函数来描述；$\mathbf{d}_{RC}(\mathbf{x})$ 被视为高斯白噪声。

上述对于空间（随机场）数据的分解，相当于对待反演参数的分解，对于理解地质统计学反演也是很有帮助的。

地质统计学反演就是在上述思想下发展出来的。首先，必须引入井约束，加上反射结构约束（最好是地震相约束）井数据插值方法得到伪井结果。这部分就对应 $\mathbf{d}_{BC}(\mathbf{x})$。以伪井结果作为初始值和/或先验约束值，基于对实测数据的逼近，用梯度导引类算法给出某种加权平均的期望结果。在期望解附近，在 Gauss 分布控制下，用随机模拟算法生成一族超分辨率可行解，进一步引入先验信息在合理的准则下，从中选择出超分辨率反演结果，并对反演解的可靠性进行评价。这就说明了为什么地质统计学反演是随机反演在勘探地震中的实例。

据此，把地质统计学反演放在 Monte Carlo Bayes 推断的理论框架下，与空间数据分析建模相结合，这样可以加深对地质统计学反演的理论认识。

12 随机模拟在勘探地震中的可能应用

基于波动方程数值解法的地震波场确定性模拟，已经广泛地应用在地震数据分析的相关过程中。它可以根据人为设计的、弹性参数表示的地质模型，通过求解各种形式的波动方程，按既定的观测系统，产生炮集。这样的确定性模拟，可以产生人工合成的炮波场，用于验证地震波成像方法，可以定性地评价地质解释的合理性，可以为深度学习算法提供标签数据，可以分析不同的弹性参数表示的地质模式的地震波响应特征等等。

前已述及，随机模拟本质上是生成特定的概率分布对应的随机样本值。理论上，它可以解决高维积分问题，可以对任意的概率分布进行重要性采样（借此做法可以进行全局寻优）。但是，随机模拟在勘探地震中能有什么作用？

作者认为：

① 可以按假定的概率分布随机生成各种形式的地震相模式，从中选择出合理的地震相模式，辅助进行地震解释。

② 可以按假定的概率分布随机生成各种形式的火山岩相模式，与确定性的地震波模拟结合，筛选出符合实际的火山岩相模式。

③ 可以按假定的概率分布随机生成各种形式的层状地层组合，与确定性的地震波模拟结合，筛选出符合实际的薄层地层组合模式。

④ 可以按假定的概率分布随机生成各种形式的层状地层组合，与确定性的地震波模拟结合，按照与实际地震道的匹配程度，生成高分辨的伪井数据。用于其他高分辨反演的先验约束。

⑤ 可以按假定的概率分布随机生成随机噪声，与实际观测的随机噪声对比，用适当的方法减去实际观测的随机噪声。

13 结论与讨论

Bayes 估计理论框架下的地震波成像，以 FWI 为代表，在“两宽一高”地震数据采集和超大规模计算机系统算力的支撑下，已经成为了油气勘探行业中的核心技术。FWI 的理论核心是在后验概率密度 Gauss 分布假设下，由 MAP 原则+可导的目标泛函+梯度导引类算法产生后验概率密度加权叠加意义下的反演结果。目前的实践表明，这样的反演结果的分辨率达不到实际需求，尤其在针对储层描述的情形下。

当然，我们知道当前 FWI 的最核心问题还不是反演精度（分辨率）问题，而是收敛性问题。因此，提出了很多改善 FWI 稳定性和收敛性的方法技术，譬如：改善对实测数据的观测逼近真正的“两宽一高”；提高正问题对实测数据的预测精度；期望预测误差是 Gauss 分布（ Gauss 白噪）的；期望正问题关系尽可能是线性的；期望初始值是尽可能接近真解的；施加合理的正则化约束；期望有尽可能多的先验信息等等。这些措施和方法对于改善 FWI 成像精度和分辨率一定是有效的，但依然达不到期望的要求。

笔者认为：基于 Monte Carlo 随机模拟的方法是值得探索的新的方法技术路线。基本思想是：在已知后验（概率密度）的条件下，生成满足该分布的一族样本，即 MAP 附近的样本值都可以作为参数估计结果。很显然，这种随机模拟算法产生的反演解族，分辨率原则上可以达到采样间隔的尺度。但是，如何生成有地质意义的反演解族，如何从中选择出有地质意义的反演解，依然是要解决的问题。

无论如何，相对于 FWI，基于 Monte Carlo 随机模拟的参数反演思想与方法为我们提供了一条新的参数反演的技术路线。梯度导引类的 FWI 和基于 Monte Carlo 随机模拟的参数反演还可以结合在一起，进一步提高参数反演的精度。

事实上，Monte Carlo 随机模拟还可以用在很多其它不同的应用场景中，譬如模拟随机噪声、地震相、火山岩相、薄互层地层组合等等。

我们相信：对于解决很多复杂的非线性问题，相比于梯度导引类的 Newton 迭代法，在超强算力辅助下，尤其与族群生物进化方法结合的 Monte Carlo 随机模拟方法会起到越来越重要的作用。Monte Carlo 随机模拟与 Bayes 推断结合会形成重要的研究领域。

参考文献

[1] Gamerman, D. and Lpes, H. F., 2006, Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference, Second Edition, Chapman & Hall/CRC

[2] Robert, C. P. R. and Caella, G., 2004, Monte Carlo Statistical Methods, Second Edition, Springer

[3] Randy L. Hanpt and Sue Ellen Hanpt, 2004, Practical Genetic Algorithms, 2th Edition, WILEY-INTERSCIENCE, A John Wiley & Sons, Inc.

[4] Albert Tarantola, 2005, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM.

[5] Jun S. Liu, 2001, Monte Carlo Strategies in Scientific Computing, Springer

[6] Gentle, J. E., 2003, Random Number Generation and Monte Carlo Methods, Second Edition, Springer

[7] Lomas, S. Luo, M. Irakarama, R. Johnston, M. Vyas, X. Shen, 2023, 3D Probabilistic Full Waveform Inversion: Application to Gulf of Mexico Field Data; 84th EAGE Annual Conference & Exhibition

[8] André Haas and Olivier Dubrule, 1994, Geostatistical inversion-a sequential method of stochastic reservoir modelling constrained by seismic data; FIRST BREAK VOL 12, NO 11

[9] M. Izzatullah, T. Van Leeuwen, D. Peter, 2021, Langevin Dynamics Markov Chain Monte Carlo Solution for Seismic Inversion; 82nd EAGE Annual Conference & Exhibition.

[10] Dehan Zhu and Richard L. Gibson, 2016, Seismic inversion and uncertainty analysis using a trans-dimensional Markov chain Monte Carlo Method; SEG International Exposition and 86th Annual Meeting.

[11] Peng Li, Leandro de Figueiredo, Dario Grana, and Klaus Mosegaard; 2021, Spatially correlated Markov chain Monte Carlo method for petrophysical inversion; 2021 Society of Exploration Geophysicists, First International Meeting for Applied Geoscience & Energy

[12] Georgia K. Stuart, Weihua Yang, Susan E. Minkoff, and Felipe Pereira, 2016, A two-stage Markov chain Monte Carlo method for velocity estimation and uncertainty quantification; SEG International Exposition and 86th Annual Meeting.

[13] Dan Simon, 2013, Evolutionary Optimization Algorithms – Biologically Inspired and Population – Based Approaches to Computer Intelligence; John Wiley & Sons, Inc.

[14] Zhu, H., S. Li, S. Fomel, G. Stadler, and O. Ghattas, 2016, A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration: Geophysics, 81, no. 5, R307–R323

[15] Zhao, Z., and M. K. Sen, 2021, A gradient-based Markov chain Monte Carlo method for full-waveform inversion and uncertainty analysis: Geophysics, 86, no. 1, R15–R30

[16] Aleardi, M., 2021, A gradient-based Markov chain Monte Carlo algorithm for elastic pre-stack inversion with data and model space reduction. Geophysical Prospecting, 69(3).

[17] Hu, S., and Z. Zhao, 2023, Uncertainty quantification of full-waveform inversion with adaptive MCMC method: Third International Meeting for Applied Geoscience & Energy, SEG/AAPG, Expanded Abstracts, 665–669

[18] Zhang X, Lomas A, Zhou M, et al., 2023, 3-D Bayesian variational full waveform inversion[J]. Geophysical Journal International, 234(1): 546-561.

[19] Zhang X, Curtis A. 2020, Variational full-waveform inversion[J]. Geophysical Journal International, 222(1): 406-411.

[20] Thurin J, Brossier R, Métivier L. ,2019, Ensemble-based uncertainty estimation in full waveform inversion[J]. Geophysical Journal International, 219(3): 1613-1635.

[21] Hoffmann A, Brossier R, Metivier L, et al. 2024, Uncertainty Quantification for 3D FWI at an Operational Level: an Ensemble Transform Kalman Filter approach[C]//85th EAGE Annual Conference & Exhibition (including the Workshop Programme). European Association of Geoscientists & Engineers, 2024(1): 1-5.

[22] Cen H, Guo K, Wang D. 2024, FWI uncertainty analysis with Stein variational gradient descent[C]//SEG International Exposition and Annual Meeting. SEG, SEG-2024-4101084.

[23] Yao K, Yin H, Duan L, et al., 2024, Uncertainty analysis of full-waveform velocity inversion[C]//Fourth International Meeting for Applied Geoscience & Energy. Society of Exploration Geophysicists and American Association of Petroleum Geologists, 1023-1027

再议空间数据分析及多信息融合思想与方法

王华忠，聂振波，吴成梁，韩龙祥

同济大学海洋与地球科学学院波现象与智能反演成像研究组，上海，200092

【摘要】在“两宽一高”叠前地震数据采集和超级算力的支持下，Bayes 参数估计理论下的 FWI 成为地震波成像的核心方法技术基本成为了必然的选择。目前，FWI 的基本计算框架已经定型，虽不能说十分完善，但已经不是 FWI 实用化的主要矛盾。FWI（尤其是 RFWI）方法技术实用化的主要矛盾已经转到如何使其在实际数据地震波成像中得到实用，能得到更好的成像效果，从而提高油藏描述的精度。FWI 是一个信息不足情况下的大规模参数估计问题，尤其在成像目标定位为宽带弹性参数估计时，这是一个非线性性很强的反问题。降低强非线性反问题的多解性，提高反演成像结果的精度，重要的措施就是引入与待估计参数相关的先验信息。勘探地震中，这些先验信息应该是与待估计参数一样，都是地下介质空间域的。它们主要包括：与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数据。这些要么是空间散乱点处的，要么是规则化三维空间网格点上的，且是二进制表示的数据。而地质知识和岩石物理关系这些非二进制表达的信息也是非常重要的与待估计参数相关的先验知识。这些先验信息如何统一地表达出来，如何在 FWI 迭代反演过程中合理地施加，成为了当前 FWI 反演成像技术能否实用化的核心问题。事实上，高维、多属性空间中的初至波走时检测和反射波走时（及走时差）检测问题；三维成像空间中的层位、断层、地震相识别问题；多种近地表信息综合利用的近地表速度建模问题已经成为了勘探地震领域多信息融合框架下所要解决的典型问题。多信息融合提高 FWI 反演成像问题的精度与稳定性是当前及今后 FWI 走向实用化的关键。更进一步地，基于多信息融合更高精度地描述岩性油气藏，并更好地做出含油气性的判决是更深层次的问题，这应该是 AI/AGI 要在油气勘探领域中针对的终极问题。在 Bayes 推断/判决理论框架下，统一地讨论上述问题，可以促进上述各种问题研究的深化。本文首先给出空间随机（数据）集的定义，讨论多元数据分析的基本思想与方法；基于此讨论空间数据分析和空间数据融合的基本逻辑，接着对空间数据的特点进行分析，并指出多源（元）信息具有异质性、尺度差异、精度差异、模糊性及随机性等特征；在此基础上，介绍几种典型的空间数据分析方法，并提出信息融合的最佳策略是（应该是）提出一个约束优化反问题定义的生成机制产生一个期望的模型，即把多信息融合问题放在约束优化反问题的框架下来解决。最后讨论勘探地震中多信息融合在参数估计反问题中降低反问题多解性的基本思想与方法。

【关键词】Bayes 推断，Bayes 参数估计，FWI 反演成像，先验信息，多源（元）信息，空间数据表达，空间数据建模，信息融合，约束优化反问题，降低反演问题多解性

1 引言

众所周知，油气地震勘探中，存在两个核心问题：宽带弹性参数估计或广义地震波反演成像；定量的油藏描述和含油气性评价。

但是，这两个问题都是强非线性。前者可被认为是 Bayes 参数估计问题；后者可以说是一个基于信息综合的判决问题（Bayes 判决问题）。后者的解决是依赖于前者的。Bayes 参数估计问题的解决是首要的。

目前，在“两宽一高”叠前地震数据采集和超级算力的支持下，Bayes 参数估计理论下的 FWI 成为地震波成像的核心方法技术已经是不争的事实。尽管 FWI 的实用化，尤其针对陆上地震数据的实用化还有很长的路要走，但这并不影响它成为地震波成像的核心方法技术的判断。应该说，经过二十多年的研究，FWI 的基本计算框架已经定型，虽不能说十分完善，但这已经不是 FWI 实用化的主要矛盾。FWI（尤其是 RFWI）方法技术实用化的主要矛盾已经转到如何使其在实际数据地震波成像中得到实用，能得到稳定的、高精度的成像效果，从而提高油藏描述的精度。但是，FWI 是一个信息不足情况下的强非线性反问题，尤其在成像目标定位为宽带弹性参数估计时。

如何降低强非线性反问题的多解性、提高反演成像结果的精度应该说已经形成共识，就是尽可能多地引入与待估计参数相关的先验信息。事实上，在宽带弹性参数估计或广义地震波反演成像的基础上，进行定量的油藏描述和含油气性评价，同样是一个强非线性的反问题，多解性也很强（可能更强）。引入与待描述和评价的油气藏相关的先验信息也是非常必要的。目前，我们远没有达到定量刻画油气藏的程度，基本上是定性地、概略地对油气藏进行描述和评价。即便如此，也需要综合地利用各种与油气藏相关的先验信息。

总而言之，发展到现在，对于复杂问题（譬如陆上地震数据 FWI）的解决，多信息综合、多信息融合的思想和做法已经在所难免。

勘探地震中，这些先验信息（主要是与待估计参数相关的信息）应该都是三维（也可以是更高维）空间域的，可以随时间演化，也可以不随时间演化。勘探地震中的先验信息以不随时间演化的空间信息为主。主要应该包括：与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数

据。这些要么是空间散乱点处的，要么是规则三维空间网格点上的，且是二进制表示的数据。而地质知识和岩石物理关系这些非二进制表达的信息也是非常重要的与待估计参数相关的先验知识。非地震、雷达波和光波等提供的或深层或浅层的与待估计参数相关的信息也逐渐地融入到高精度地震波成像过程中。

很显然，FWI 的真正困难，就在于复杂介质（譬如复杂地表、复杂构造、小尺度和/或强变化的地质体存在）情形下，它是一个强非线性反问题。强非线性反问题的多解性、不稳定性使得这类方法技术很难实用化。而假设已知正确的偏移速度模型的叠前地震波偏移成像问题，从数学算法上讲，不是一个反问题，因此，它有很强的稳健性，在勘探地震中得到了广泛的应用。但是，给出正确的叠前偏移成像结果的前提是假设已知正确的偏移速度模型，如果把构建偏移速度模型也放在广义的地震波成像过程中，广义的地震波成像就成了一个标准的反问题。很显然，FWI=Tomography+PSDM 就是所谓的广义地震波成像。

复杂探区叠前深度偏移成像结果弄不清真实的地下地质结构形态，解决不了相应的地质问题，本质原因就是此时的广义地震波成像（Tomography+PSDM）是个（强）非线性反问题。如果 FWI 的成像目标是宽带弹性参数，该反问题的非线性性更会大幅度提高。这也是石油工业界把地震波成像分为背景速度建模和叠前深度偏移成像两部分的根本原因。毕竟叠前深度偏移成像数学理论上可以不被认为是一个参数估计反问题。

众所周知，高精度的、定量的油藏描述需要宽带弹性参数成像结果。遗憾的是，当前宽带弹性参数成像的精度往往很难满足高精度油藏描述的需求。此时把更多的与待估计参数相关的先验信息（多源（元）信息）提取出来，进行正确的表达，然后用合理的技术路线提升宽带弹性参数成像的精度，就成了当前地震波成像的一个十分关键的、不可回避的问题。

但是，空间规则或空间散乱点上的多源（元）信息存在的异质性、尺度差异、精度差异、模糊性及随机性等，导致到目前为止多源（元）信息融合并没有形成统一的、公认的理论框架。基本的认识是：首先统一地表达这些复杂的多源（元）数据，比较一致的观点是把它们表达为随机（数据）集。然后，基于样本数据，提取随机（数据）集的统计特征。并对随机（数据）集进行统计建模。最后，进行基于模型的推断或判决。事实上，这个基本的认识如何转化成有效的、实用的算法，目前存在很多不同的技术途径。

正是有鉴于此，本文首先给出空间随机（数据）集的定义，讨论多元数据分析的基本思想与方法；基于此讨论空间数据分析和空间数据融合的基本逻辑，接着对空间数据的特点进行分析，并指出多源（元）信息具有异质性、尺度差异、精度差异、模糊性及随机性等特征；在此基础上，介绍几种典型的空间数据分析方法，并提出信息融合的最佳策略是（应该是）提出一个约束优化反问题定义的生成机制产生一个期望的模型，即把多信息融合问题放在约束优化反问题的框架下来解决。最后讨论勘探地震中多信息融合在参数估计反问题中降低反问题多解性的基本思想与方法。

2 空间数据集（随机（数据）集）的定义

由于在勘探地震中，尤其在 FWI 宽带弹性参数估计中，与宽带弹性参数相关的信息仅仅存在于空间域，此处我们给出不随时间变化的空间数据集的定义。我们认为与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数据，包括地质知识和岩石物理关系这些定性的、非二进制的数据，都可以统一在这个空间数据集之内。

空间点 $s_i, i=1,2,\cdots,N$ 上，存在物理对象（也可以是非物理对象）的观测值、估计值、或它们的属性等等，这些都是描述（非）物理对象的信息。可以记为 $X_i, i=1,2,\cdots,N$，认为它是随机变量，是满足一定的概率分布 $p_X(x)$ 的。用概率分布来表达先验信息的统计特征，包容了它的模糊性、不确定性。由于描述（非）物理对象的信息可能是多层次、多方面的，可以认为它们是从不同角度、不同尺度以及不同可信度来描述同一个（非）物理对象。因此，在一个空间点上，可以用随机向量 $\boldsymbol{X}_j, j=1,2,\cdots,M$ 描述多源（元）信息，不同的分量代表不同的信息。

随机变量 $X_i, i=1,2,\cdots,N$ 或随机向量 $\boldsymbol{X}_i, i=1,2,\cdots,N$ 蕴含了（非）物理对象在一个空间点处统计规律。但是，多源（元）信息融合更关注的是不同空间点处多源（元）信息之间的统计变化关系。此时，需要引入空间随机向量 $\boldsymbol{Y}=\left(\boldsymbol{X}_{S_1},\boldsymbol{X}_{S_2},\cdots,\boldsymbol{X}_{S_N}\right)$。当然，对于空间信息融合，更需要引入空间随机矩阵

$$\boldsymbol{Y}_{N\times M}=\begin{pmatrix} X_{S_1}^1 & X_{S_2}^1 & \cdots & X_{S_N}^1 \\ X_{S_1}^2 & X_{S_2}^2 & \cdots & X_{S_N}^2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{S_1}^M & X_{S_2}^M & \cdots & X_{S_N}^M \end{pmatrix}$$。空间随机矩阵是空间随机场的表达。

可见，随机矩阵 $\boldsymbol{Y}_{N\times M}$ 包含了（非）物理对象的所有空间变化信息。最基本地，它们包含在随机矩阵满足的（复杂的）联合概率密度函数中。

如果空间随机矩阵（空间随机场）是随时间变化的，存在如下的随时间变化的随机矩阵，即，

$$\boldsymbol{Y}_{N\times M}(t)=\begin{pmatrix} X_{S_1}^1(t) & X_{S_2}^1(t) & \cdots & X_{S_N}^1(t) \\ X_{S_1}^2(t) & X_{S_2}^2(t) & \cdots & X_{S_N}^2(t) \\ \vdots & \vdots & \ddots & \vdots \\ X_{S_1}^M(t) & X_{S_2}^M(t) & \cdots & X_{S_N}^M(t) \end{pmatrix} \tag{1a}$$

我们认为多源（元）信息用这样的数据形式来表达，应该达到了较高的抽象程度。可以包容很多领域中的多源（元）信息融合问题。对这样的数据集进行建模，把其中蕴含的信息统一在一个模型中，就实现了信息融合的目的。

3 多元数据统计分析的基本思想与方法

多元统计分析的对象是多元数据矩阵 $\boldsymbol{Y}$：

$$\boldsymbol{Y}=\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n1} & \cdots & X_{np} \end{bmatrix}=\begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_p \end{bmatrix} \tag{1b}$$

它是由各种各样的物理问题或非物理问题抽象出来的。对于一个特定的问题，它表示该问题有 p 个待考察的因素，每个因素有 n 个采样样本。对比可知，多元数据矩阵 $\boldsymbol{Y}$ 与空间随机场数据 $\boldsymbol{Y}$ 是十分类似的，仅仅是多元数据矩阵 $\boldsymbol{Y}$ 中某因素的数据样本不是在特定空间位置处获得。空间随机数据不一定是二维（矩阵形式）的，既有可能是高维的。但数据维度不影响数据分析的基本逻辑。

对多元数据矩阵 $\boldsymbol{Y}$，可以认为其 p 个因素对应的列向量支撑起了一个子空间，或构成了一个坐标系。多元统计分析的重点是考察 p 个列向量之间的相互关系。即 p 个考察因素之间的关系。

每个因素可以被视为一个随机变量。一般地，均值和方差描述了该因素的主要特征。而多元统计分析（或多因素统计分析）的要点是分析若干因素之间的相关性。主要方法包括：均值与方差的推断方法；单因变量多自变量线性及非线性回归分析方法和多因变量多自变量线性及非线性回归分析方法；主分量分析方法；因子分析方法；典型相关分析；判别分析与聚类分析方法；对应分析方法等。

很多情形下，我们对 p 个因素的具体含义是很清楚的，可能就是按照 p 个已知的因素进行数据采样的。但是，更多情形是，我们仅仅采集到一维或二维以上的规则或非规则的数据，对其中的潜变量的含义并不十分明瞭。空间数据的采集与分析就是此类情形。这两类数据原则上都可以排成上述多元数据矩阵的形式。实质上，三维以上的数据可表达成张量的形式，也可以通过合适的变换降维表达为矩阵。非规则数据看起来很简单，事实上给数据分析带来了巨大的麻烦。

尽管上述两种情形下采集到的数据的分析方法，思想本质上是一致的，但是具体方法还是差异较大的。一维或二维以上的规则或非规则的数据的分析主要用时间序列分析方法、图像处理方法、高维信号分析方

法；非规则高维数据用空间数据分析方法。而多元数据矩阵的分析方法就是前述的各种多元数据分析方法。总之，多元数据统计分析与高维数据统计分析是数据分析的核心内容。

当前，机器学习的内核本质上依然是在执行多因素分析，只是算法发生了巨大的变化而已。这也是很多人认为机器学习的理论基础就是统计学习的原因。

4 空间数据分析及空间信息融合的基本逻辑

多源（元）信息融合首先在军事领域，尤其是战场信息管理及决策领域得到了深入的研究，这是一个最典型的多源（元）信息获取、管理、融合和应用的场景。多源（元）信息融合的方法理论正是在军事领域应用的强力促进下得到长足的发展。众所周知，地球物理学领域中，也存在非常多的多源（元）信息融合的应用场景。因为地球物理场是多种多样的，描述岩性介质的物理参数也是多种多样的、尺度变化也存在很大的跨度。而且各种信息都存在不同程度的不确定性、模糊性和随机性。既有二进制描述的定量化信息，也有非二进制描述的定性信息。多源（元）信息融合在 FWI 宽带弹性参数估计中的应用也是 FWI 发展到现在的必然选择。当然，还有很多需要多源（元）信息融合的应用领域。事实上，由于多源（元）信息的异质性、尺度差异、精度差异、模糊性及随机性等，到目前为止，多源（元）信息融合并没有形成统一的、公认的理论框架。

但是，基于到目前为止的研究已经可以预知多源（元）信息融合首先要解决的是多源（元）信息的抽象表达问题。一般地，多源（元）信息首先可以分为定量的、可二进制表达的及定性的、不可二进制表达的（实际上，计算机处理时，也往往通过某种方式把这些信息转换成可二进制表达的）。第二，多源（元）信息可分为随时间演变的空间数据和不随时间演变的空间数据。第三，多源（元）信息可分为空间散乱的数据和空间规则的数据。最后，还要注意到，多源（元）信息中有很多是一级信息（所谓一级信息即数据）的特征表达或属性表达结果。

多源（元）信息之间存在的异质性、尺度差异、精度差异、模糊性及随机性更增加了对多源（元）信息进行抽象表达的难度。到目前为止，用随机（数据）集来表达多源（元）信息应该是抽象程度比较高的一种表达形式。

FWI 宽带弹性估计中，与待估计参数相关的信息都是空间域的。空间数据 $\boldsymbol{Y}=\left(\boldsymbol{X}_{S_1},\boldsymbol{X}_{S_2},\cdots,\boldsymbol{X}_{S_N}\right)$ 或

$$\boldsymbol{Y}_{N\times M}=\begin{pmatrix} X_{S_1}^1 & X_{S_2}^1 & \cdots & X_{S_N}^1 \\ X_{S_1}^2 & X_{S_2}^2 & \cdots & X_{S_N}^2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{S_1}^M & X_{S_2}^M & \cdots & X_{S_N}^M \end{pmatrix}$$

中蕴含的信息体现在空间位置 $s_i, i=1,2,\cdots,N$ 为标志的随机变量之间存在相关关系上。原则上，空间位置相邻越近，相关关系越强。但是，由于空间随机变量（向量、矩阵、张量）受背后（非）物理规律的制约，空间不相邻的随机变量之间也可能存在很强的相关性。这种先验信息，当然也可以预先表达在空间数据集中。一般地，随机变量（向量、矩阵、张量）满足的概率密度函数对应的数字特征（或其中包含的参数）体现了空间数据集所蕴含的规律。

因此，空间数据分析或空间信息融合的本质应该是：基于已知的空间数据集，假设它满足一定的联合概率密度概率分布，利用 Bayes 推断的理论框架，估计概率分布函数中的参数值（或数字特征），最后用估计的参数值（或数字特征）对整体空间数据集的变化规律实现尽可能准确的推断，从而解决各种相关的问题。其中，空间信息融合统一在对空间数据集的建模过程中。

我们认为：多源（元）信息融合本质上就是按照上述逻辑进行的。

解决具体的应用问题，会对上述逻辑路线进行各种各样的改变，发展出具体的算法。像数据融合、图像融合等都是低一个层次的信息融合解决的问题。一般地，对于随时间演变的空间数据，Kalman 滤波、扩展的 Kalman 滤波，尤其是粒子滤波基本形成了针对此类数据的多源（元）信息融合算法体系。对于不随时间演变的空间数据，空间随机场理论、（泛/协）Kriging 插值、空间回归建模方法基本形成了针对此类数据的多源（元）信息融合算法体系。勘探地震领域，信息融合的最佳策略是（应该是）提出一个多信息约束优化反

问题定义的生成机制产生一个期望的模型。

总之，对多源（元）信息构成的随机（数据）集，在 Bayes 推断的理论框架下，进行建模，并基于模型进行统计推断，进而解决各种实际问题，是多源（元）信息融合的深层次思想。

下面罗列三个空间数据分析及空间信息融合的典型例子。

多源（元）信息约束下的（随机）散乱数据的插值构建规则网格上的参数模型就是一个典型的问题。本质上是在多源（元）信息约束下估计（随机）散乱数据支撑起来的空间数据的均值（期望值）。在此过程中，引入关于期望值的先验信息，进行 Bayes 推断是更为数学化的表达，也是多源（元）信息融合构建规则网格上的参数模型的本意。因为仅仅依赖（随机）散乱数据的统计规律，由于采样点数的不足，若没有先验信息的加入，对全空间上的空间数据均值（期望值）的估计误差可能会很大。

多源（元）信息融合/约束下的统计推断是另一个典型的问题。在同一个空间中，（非）物理对象具有各种不同侧面体现出来的属性。这些属性可以认为是从不同角度展示出来的同一个（非）物理对象的信息。这些信息之间存在或强或弱的关联性。$\boldsymbol{Y}_{N\times M}=\begin{pmatrix} X_{S_1}^1 & X_{S_2}^1 & \cdots & X_{S_N}^1 \\ X_{S_1}^2 & X_{S_2}^2 & \cdots & X_{S_N}^2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{S_1}^M & X_{S_2}^M & \cdots & X_{S_N}^M \end{pmatrix}$ 就是多源（元）信息的随机矩阵表达。多源（元）信息融合的目的就是提取空间数据矩阵 $\boldsymbol{Y}_{N\times M}$ 中各不同信息之间的共性信息（相关信息），然后用这些共性信息，提高对空间数据的建模精度，进而提高 Bayes 推断的精度。

另外，还存在一种多信息约束下的系统参数估计问题。一般地，认为存在观测数据 $\boldsymbol{d}(Z_k),k=1,2,\cdots,K$，即随机向量 $\boldsymbol{d}$ 在观测点 $Z_k,k=1,2,\cdots,K$ 取得样本值。假定其期望 $E[\boldsymbol{d}]$ 与空间域点 $s_i,i=1,2,\cdots,N$ 处的系统参数 $\boldsymbol{m}(s_i),i=1,2,\cdots,N$ 之间满足一定的数学物理方程。把已知系统参数、数学物理方程和初边值条件情况下预测观测数据的期望值 $E[\boldsymbol{d}]$ 的问题称为正问题。由观测数据 $\boldsymbol{d}(Z_k),k=1,2,\cdots,K$ 估计系统参数 $\boldsymbol{m}(s_i),i=1,2,\cdots,N$ 的问题称为反问题。依然是由于观测数据 $\boldsymbol{d}(Z_k),k=1,2,\cdots,K$ 的采样不充分和不完美，还包括上述正问题的不精确等，必须要进行多源（元）信息约束下的 Bayes 参数估计。这个问题可以认为是多源（元）信息融合/约束下的统计推断问题的特例。

5 空间数据变化特征分析

空间域的各种先验信息，一般地，都可以视为空间数据。时间序列往往被视为一个随时间变化的广义平稳随机过程，已经有了经典的 ARMA 模型来描述它。空间数据分析与时间序列分析理论基础是一致的，都利用了对广义平稳随机过程建模的思想与方法。但是，空间数据与时间序列数据产生的机制毕竟是不同的。空间相邻的数据不一定存在紧密相关性，不平稳性是空间数据固有的特征。另外，空间数据往往是非规则采样的。因此，对空间数据的建模与分析必须要构建另一套不同于时间序列分析的模式。

首先，我们认为空间数据可以表达为：$\boldsymbol{d}(\boldsymbol{x})=\boldsymbol{d}_{BC}(\boldsymbol{x})+\boldsymbol{d}_{DC}(\boldsymbol{x})+\boldsymbol{d}_{RC}(\boldsymbol{x})$。其中，$\boldsymbol{d}_{BC}(\boldsymbol{x})$ 是空间数据的背景成分（Background Component）；$\boldsymbol{d}_{DC}(\boldsymbol{x})$ 是空间数据的扰动成分（Disturbance Component）；$\boldsymbol{d}_{RC}(\boldsymbol{x})$ 是满足一定概率分布的随机成分（Random Component）。$\boldsymbol{d}_{BC}(\boldsymbol{x})$ 可以视为是确定性的、（空间）散乱点支撑的函数。这部分的描述可以用径向基函数族或其他基函数族的线性组合的方法；$\boldsymbol{d}_{DC}(\boldsymbol{x})$ 用零均值、一定方差的概率密度函数来描述；$\boldsymbol{d}_{RC}(\boldsymbol{x})$ 被视为高斯白噪声。

在上述空间数据表达下，空间数据建模的基本逻辑是：用基函数线性叠加形成对 $\boldsymbol{d}_{BC}(\boldsymbol{x})$ 的逼近；用 ARMA 模型逼近 $\boldsymbol{d}_{DC}(\boldsymbol{x})$，假定两个逼近关系之和对实测空间数据的逼近误差是高斯白噪声，求解两个逼近关系中的待定系数，完成对空间数据的建模。

很显然，对$\boldsymbol{d}_{BC}\left(\boldsymbol{x}\right)$的逼近问题是函数逼近问题，对$\boldsymbol{d}_{DC}\left(\boldsymbol{x}\right)$的预测问题是 AR 模型推广到空间数据中的问题。前者是数学分析处理的确定性问题；后者是概率统计处理的随机性问题。

假定空间数据$Z\left(\boldsymbol{x}\right)$是去掉背景成分后的广义平稳随机过程，其最基本特征是：一般意义下，相邻空间位置的随机采样值相关性更大。在高维空间中，这个特点使得高维数据体中蕴含空间结构。正是存在这样的空间结构，说明空间数据$Z\left(\boldsymbol{x}\right)$蕴含着某种信息。所谓的对空间数据$Z\left(\boldsymbol{x}\right)$的建模就是对这些空间结构信息进行建模表达。简单地讲，空间结构信息蕴含在广义平稳随机过程的自相关矩阵中。

下面以一维空间数据为例，简单地讨论一下空间数据的相关性。

对于$x\in\mathbb{R}^1$的情况，若假设$Z\left(x_1\right),\cdots,Z\left(x_n\right)$是独立同分布（IID）的高斯分布随机变量，且均值为μ，方差为σ_0^2。对观察数据做这样的假设非常合理的，因为每次观测之间是独立进行的。

则由概率论知，均值μ的最小方差无偏估计是

$$\hat{Z}=\sum_{i=1}^{n}Z\left(x_i\right)/n \tag{2a}$$

且估计子$\hat{Z}$是高斯的、均值为μ、方差为σ_0^2/n。因此，$\hat{Z}$的 95%置信区间为

$$\left[\hat{Z}-1.96\sigma_0^2/n,\ \hat{Z}+1.96\sigma_0^2/n\right] \tag{2b}$$

我们反复提及：空间数据的重大特点之一是一般而言相邻点间的随机变量存在更强的相关性。但（非）物理机制又常常使得上述说法不成立。在空间数据分析中，需要引入额外的先验信息。

无论如何，空间数据之间相关性表达形式一定是空间数据分析的基础。假设空间数据正相关且相关性随距离减弱，即

$$\mathrm{cov}\left(Z\left(x_i\right),Z\left(x_j\right)\right)=\sigma_0^2\cdot\rho^{\left|x_i-x_j\right|},\ i,j=1,\cdots,n,\ 0<\rho<1 \tag{3}$$

把自相关函数进行参数表达是进行空间数据分析的基本思想，正是引入了自相关函数的参数化表达，导致了与时间序列分析的差异，发展出了所谓的空间数据分析方法。

进一步假设，该相关函数来自于一阶自回归过程$Z\left(x_i\right)=\rho Z\left(x_{i-1}\right)+\varepsilon\left(x_i\right),\ i\geq 1$，其中$\varepsilon\left(x_i\right)$是零均值、方差为$\sigma_0^2\left(1\text{-}\rho^2\right)$的高斯随机变量。此时，估计子$\hat{Z}$的方差为

$$\begin{aligned}\mathrm{var}\left(\hat{Z}\right)&=\frac{1}{n^2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{cov}\left(Z\left(x_i\right),Z\left(x_j\right)\right)\right\}\\&=\frac{\sigma_0^2}{n}\left[1+2\rho/\left(1-\rho\right)\left(1-1/n\right)-2\left(\rho/\left(1-\rho\right)\right)^2\left(1-\rho^{n-1}\right)/n\right]\end{aligned} \tag{4}$$

当$n=10,\rho=0.26$时，$\hat{Z}$的 95%置信区间为

$$\left[\hat{Z}-2.458\sigma_0^2/n,\ \hat{Z}+2.458\sigma_0^2/n\right] \tag{5}$$

可见，数据中的空间相关性使$\hat{Z}$的方差增大。这是空间数据分析的最直观的数据特征。这就奠定了用空间数据的自相关函数（参数表达的、半解析的自相关函数）进行空间数据分析的思想基础。

一般地，空间数据 $Z(\boldsymbol{x})$ 可表达为大尺度变化+小尺度变化的形式，即

$$Z(\boldsymbol{x})=s(\boldsymbol{x})+\delta(\boldsymbol{x}) \tag{6}$$

其中，$s(\boldsymbol{x})$ 代表空间数据的背景变化，可以认为它是一个连续的函数，用基函数的线性组合来表达，

$$s(\boldsymbol{x})=\sum_{l=1}^{q}\beta_l f_l(\boldsymbol{x}),\boldsymbol{x}\in\boldsymbol{D}\subset\mathbb{R}^d \tag{7}$$

则，（6）式变为，

$$Z(\boldsymbol{x})=s(\boldsymbol{x})+\delta(\boldsymbol{x})=\sum_{l=1}^{q}\beta_l f_l(\boldsymbol{x})+\delta(\boldsymbol{x}),\ \boldsymbol{x}\in\boldsymbol{D}\subset\mathbb{R}^d \tag{8}$$

其中，$\{f_l(\cdot)|l=1,\cdots,q\}$ 是 q 个基函数的集合，基函数的选择要依据空间数据的先验信息；$\delta(\cdot)$ 是零均值有限方差的平稳随机过程，代表 $Z(\boldsymbol{x})$ 中不能被 $s(\boldsymbol{x})$ 表达的、剩余的扰动变化部分。

（6）式构成了空间数据分析最基础的概念模型。$s(\boldsymbol{x})$ 是空间数据 $Z(\boldsymbol{x})$ 的统计特征（期望），是确定性的；$\delta(\cdot)$ 被认为是广义平稳随机过程。正是在这样的概念模型的导引下，发展出了常用的空间（随机场）数据分析方法。

把（8）写成矩阵方程的形式，有

$$\boldsymbol{Z}=\boldsymbol{f\beta}+\boldsymbol{\delta} \tag{9}$$

其中，$\boldsymbol{f}\in\mathbb{R}^{n\times q}$，第 (i,j) 个元素是 $f_j(\boldsymbol{x}_i)$；$\boldsymbol{\beta}=(\beta_1,\cdots,\beta_q)^T$。$\boldsymbol{\beta}$ 是基函数族 $\boldsymbol{f}$ 上对应的坐标向量。$\boldsymbol{\delta}=(\delta(\boldsymbol{x}_1),\cdots,\delta(\boldsymbol{x}_n))^T$ 是预测残差向量（空间数据中的广义平稳随机过程部分），假设其协方差矩阵为

$$\mathrm{var}(\boldsymbol{\delta})=\boldsymbol{\Sigma} \tag{10}$$

其中，$\boldsymbol{\Sigma}\in\mathbb{R}^{n\times n}$ 是对称半正定矩阵。

对空间数据建模的主要问题是在最小二乘意义下最佳地估计坐标向量 $\boldsymbol{\beta}$，即求解如下加权最小二乘（WLS = Weighted Least Square）问题，

$$\min_{\boldsymbol{\beta}}(\boldsymbol{Z}-\boldsymbol{f\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{f\beta}) \tag{11}$$

（11）式估计结果（WLS 估计子）为

$$\hat{\boldsymbol{\beta}}_{wls}=(\boldsymbol{f}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{f})^{-1}\boldsymbol{f}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} \tag{12}$$

当 $\boldsymbol{\Sigma}=\sigma^2\boldsymbol{I}$，（11）退化为，

$$\min_{\boldsymbol{\beta}}(\boldsymbol{Z}-\boldsymbol{f\beta})^T(\boldsymbol{Z}-\boldsymbol{f\beta})/\sigma^2 \tag{13}$$

对应的估计结果为:

$$\hat{\boldsymbol{\beta}}_{ols}=(\boldsymbol{f}^T\boldsymbol{f})^{-1}\boldsymbol{f}^T\boldsymbol{Z} \tag{14}$$

这即一般的最小二乘（LS）估计子。当$\boldsymbol{\delta}(\cdot)$是高斯白噪声时，估计结果满足 Gauss-Markov 定理，是无偏的、方差最小的估计结果（OLS = Optimal Least Square）。当$\boldsymbol{\delta}=\left(\delta(\boldsymbol{x}_1),\cdots,\delta(\boldsymbol{x}_n)\right)^T$的自协方差矩阵中存在结构特征时，估计结果不可能达到最佳。这就是说明要么$\boldsymbol{\delta}(\cdot)$满足高斯白噪声假设，要么一致其自相关矩阵，（12）和（14）式的估计结果才是最佳的。

但是，$\boldsymbol{\delta}(\cdot)$的自相关矩阵是未知的，必须要同时估计$\boldsymbol{\beta}$和自相关矩阵，才是空间数据建模的合理做法。尤其是对$\boldsymbol{\delta}(\cdot)$的自相关矩阵进行参数化表达，是空间数据建模的独特之处。

据此，假设残差数据（空间数据中的广义平稳随机过程部分）$\delta(\cdot)$的协方差可写为含参数的$\boldsymbol{\Sigma}\left(\hat{\theta}\right)$，代入（12）式得到所谓的估计广义最小二乘解（EGLS = Estimated General Least Square）

$$\hat{\boldsymbol{\beta}}_{egls}=\left(\boldsymbol{f}^T\boldsymbol{\Sigma}\left(\hat{\theta}\right)^{-1}\boldsymbol{f}\right)^{-1}\boldsymbol{f}^T\boldsymbol{\Sigma}\left(\hat{\theta}\right)^{-1}\boldsymbol{Z} \tag{15}$$

$\hat{\boldsymbol{\theta}}$是协方差矩阵$\boldsymbol{\Sigma}$中包含的参数$\boldsymbol{\theta}$的估计量。很显然，迭代地估计$\hat{\boldsymbol{\theta}}$和$\hat{\boldsymbol{\beta}}_{egls}$达到最佳的空间数据建模结果。**这就是空间数据分析的典型方法的基本理论基础。**

很显然，提一个约束优化反问题对$Z(\boldsymbol{x})$中确定性部分进行建模，用 AR 模型对广义平稳随机过程部分进行建模，并把二者串联起来进行迭代求解（类似 EM 算法），应该是合理的策略。这样做容易考虑空间数据变化的物理机制，在空间数据建模中引入额外的先验信息，得到更好的最终逼近结果。

6 空间数据分析的典型方法

空间数据的最显著特点是：它是空间坐标的函数，它是（常常是）不规则的。但是空间数据分析，本质上，还是利用空间数据的一阶二阶统计量进建模并对未观测（也包括已观测）点进行预测的问题。

最基本的概念模型是：空间随机场（或空间随机数据）$Z(\boldsymbol{x})$被分成均值（背景部分）$\mu(\boldsymbol{x})$+广义平稳随机信号部分$e(\boldsymbol{x})$，再加上不可被建模表达的，或不被所建模型预测的高斯白噪声$\delta(\boldsymbol{x})\sim N(0,\sigma^2 I)$，即$Z(\boldsymbol{x})=\mu(\boldsymbol{x})+e(\boldsymbol{x})+\delta(\boldsymbol{x})$。应该注意到：空间随机数据$Z(\boldsymbol{x})$往往是不平稳的，把一维 AR 模型推广应用在空间数据建模时，必须考虑非平稳的均值对建模问题的影响。

一般地，也必须对不平稳的均值（背景部分）建模估计，去除均值后，才能对广义随机平稳部分$e(\boldsymbol{x})$进行基于 AR 模型的建模，但估计均值$\mu(\boldsymbol{x})$部分时必然会受广义随机平稳部分$e(\boldsymbol{x})$的影响，因此对$Z(\boldsymbol{x})$进行建模本质上必须同时考虑一阶二阶统计量的影响。以前（地震）信号分析时，总是假设一阶统计量为零，这其实回避了很多问题，零均值假设对于零附近波动的信号是合理的，但对于空间数据(非波动变化的空间数据)并不合理。

当$Z(\boldsymbol{x})$均值为常数时，可以把它纳入 AR 模型中，当成一个参数进行估计。当$Z(\boldsymbol{x})$均值空变时，还是要先对均值$\mu(\boldsymbol{x})$进行一般的函数逼近，此时必须同时考虑广义随机平稳部分$e(\boldsymbol{x})$的方差结构对估计$\mu(\boldsymbol{x})$时产生的影响。理论上，对$Z(\boldsymbol{x})$的建模问题，应提成一个 Bayes 参数估计问题，还应该探索在深度学习框架下如何更好地解决该问题。

空间数据统计分析的典型方法来自于 Kriging 提出的思想，目前空间数据统计学建模（插值）方法，主要就是指 Kriging 法及其变体。它显然是把时间序列分析中的 AR 模型用在了空间数据条件分析中。但是 Kriging 用半变差函数模型替代$Z(\boldsymbol{x})$的自协方差阵$\boldsymbol{\Sigma}$，进而由参数化的函数形式表达半变差函数，构成了 Kriging 法（包括其变体）的核心思想。具体地可以描述为：假设空间数据是平稳的，任一点的空间数据可

以用周围数据的加权平均（本质是就是统计期望）来预测，基于此定义预测误差泛函。同时把涉及到的空间数据的自协方差阵 $\boldsymbol{\Sigma}$ 用半变差函数替代，因为空间数据的协方差阵 $\boldsymbol{\Sigma}$ 这个二阶统计量是未知的，据此形成构建空间数据模型的误差泛函。求解该误差泛函定义的优化问题，就可以得到空间数据的最佳预测模型。

6.1 空间数据分析的 Kriging 方法

一般 Kriging 方法，假设空间数据的均值（趋势）为常数

$$E\left[Z\left(\boldsymbol{x}\right)\right]=\mu \tag{16}$$

因此，空间数据可以表达为

$$Z\left(\boldsymbol{x}\right)=\mu+\delta\left(\boldsymbol{x}\right) \tag{17}$$

其中，$\delta\left(\boldsymbol{x}\right)$ 是零均值固有平稳随机过程。不能认为 $\delta\left(\boldsymbol{x}\right)$ 是观测噪声，而是（非）物理对象在位置 $\boldsymbol{x}$ 上的小尺度变化+噪声。

基于（17）式的表达，可以构建如下的预测关系式

$$\hat{Z}\left(\boldsymbol{x_u}\right)=\sum_{i=1}^{n}\lambda_i Z\left(\boldsymbol{x_i}\right),\ \ s.t.\ \sum_{i=1}^{n}\lambda_i=1 \tag{18}$$

其中，$Z\left(\boldsymbol{x_i}\right),\ i=1,2,\cdots,n$ 是已知观测点处的数据。$\hat{Z}\left(\boldsymbol{x_u}\right)$ 是未知空间点处的预测值，约束条件 $\sum_{i=1}^{n}\lambda_i=1$ 使预测值满足无偏估计 $E\left[\hat{Z}\left(\mathbf{x_u}\right)\right]-E\left[Z\left(\mathbf{x_u}\right)\right]=0$。本质上，这也是求统计期望的基本要求，只是此处不把 $\lambda_i, i=1,2,\cdots,n$ 视为概率密度函数的离散值。（18）式的写出显然利用了 $Z\left(\boldsymbol{x_i}\right),\ i=1,2,\cdots,n$ 满足平稳性的假设，可以认为是时间序列分析中 AR 模型在空间数据分析中的推广应用。

用 Lagrange 乘子法把（18）式定义的约束优化问题转为无约束优化问题，可得目标泛函

$$\min_{\lambda_1,\lambda_2,\ldots\lambda_n} E\left[\left(Z\left(\boldsymbol{x_u}\right)-\sum_{i=1}^{n}\lambda_i Z\left(\boldsymbol{x_i}\right)\right)^2\right]-2m\left(\sum_{i=1}^{n}\lambda_i-1\right) \tag{19}$$

其中，

$$\begin{aligned}E\left[\left(Z\left(\boldsymbol{x_u}\right)-\sum_{i=1}^{n}\lambda_i Z\left(\boldsymbol{x_i}\right)\right)^2\right]=-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\, E\left[\left(Z\left(\boldsymbol{x_i}\right)-Z\left(\boldsymbol{x_j}\right)\right)^2\right]\Big/2+\\ 2\sum_{i=1}^{n}\lambda_i\, E\left[\left(Z\left(\boldsymbol{x_u}\right)-Z\left(\boldsymbol{x_i}\right)\right)^2\right]\Big/2\end{aligned} \tag{20}$$

该式是在 $\sum_{i=1}^{n}\lambda_i=1$ 的条件下导出的。注意，定义（19）式用了均方误差，在 $\sum_{i=1}^{n}\lambda_i=1$ 的条件下，这就与空间数据的协方差联系起来了。

由于协方差函数是空间数据 $Z\left(\boldsymbol{x_i}\right),\ i=1,2,\cdots,n$ 的二阶统计量，而它是未知的。放弃使用协方差函数这个未知的二阶统计量，转为依据实测的、散乱的空间数据定义如下半变差函数

$$\gamma\left(\boldsymbol{x},\boldsymbol{h}\right)=\frac{1}{2}\mathrm{var}\left(Z\left(\boldsymbol{x}\right)-Z\left(\boldsymbol{x}+\boldsymbol{h}\right)\right)=\frac{1}{2}E\left[\left(Z\left(\boldsymbol{x}\right)-Z\left(\boldsymbol{x}+\boldsymbol{h}\right)\right)^2\right] \tag{21}$$

应该注意到上式中引入了距离向量$\boldsymbol{h}$。如果假定空间数据$Z(\boldsymbol{x})$是广义平稳的，二阶统计量不随空间位置变化，则上式可以简写为

$$\gamma(\boldsymbol{h})=\frac{1}{2}E\left[\left(Z(\boldsymbol{x})-Z(\boldsymbol{x}+\boldsymbol{h})\right)^2\right] \tag{22}$$

若再进一步地假设空间数据变化是各向同性的，即半变差函数$\gamma(\boldsymbol{h})$与距离向量$\boldsymbol{h}$的方向无关，上式简化为

$$\gamma(\|\boldsymbol{h}\|)=\frac{1}{2}E\left[\left(Z(\boldsymbol{x})-Z(\boldsymbol{x}+\boldsymbol{h})\right)^2\right] \tag{23}$$

据此，（19）和（20）式定义的目标泛函变为

$$\min_{\lambda_1,\lambda_2,\cdots,m}-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\|\boldsymbol{x}_i-\boldsymbol{x}_j\|)+2\sum_{i=1}^{n}\lambda_i\gamma(\|\boldsymbol{x}_u-\boldsymbol{x}_i\|)-2m\left(\sum_{i=1}^{n}\lambda_i-1\right) \tag{24}$$

对（24）式求关于$\lambda_1,\lambda_2,\cdots,\lambda_n,m$的偏导数，并令偏导数为零，可得如下方程组

$$\begin{cases}-2\sum_{j=1}^{n}\lambda_j\gamma(\|\boldsymbol{x}_i-\boldsymbol{x}_j\|)+2\gamma(\|\boldsymbol{x}_u-\boldsymbol{x}_i\|)-2m=0,\ i=1,2,\cdots,n\\ \sum_{i=1}^{n}\lambda_i=1\end{cases} \tag{25}$$

上式对应的矩阵方程具体地可写为，

$$\begin{bmatrix}\gamma(\|\boldsymbol{x}_1-\boldsymbol{x}_1\|) & \gamma(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|) & \cdots & \gamma(\|\boldsymbol{x}_1-\boldsymbol{x}_n\|) & 1\\ \gamma(\|\boldsymbol{x}_2-\boldsymbol{x}_1\|) & \gamma(\|\boldsymbol{x}_2-\boldsymbol{x}_2\|) & \cdots & \gamma(\|\boldsymbol{x}_2-\boldsymbol{x}_n\|) & 1\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \gamma(\|\boldsymbol{x}_n-\boldsymbol{x}_1\|) & \gamma(\|\boldsymbol{x}_n-\boldsymbol{x}_2\|) & \cdots & \gamma(\|\boldsymbol{x}_n-\boldsymbol{x}_n\|) & 1\\ 1 & 1 & \cdots & 1 & 0\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_n\\ m\end{bmatrix}=\begin{bmatrix}\gamma(\|\boldsymbol{x}_u-\boldsymbol{x}_1\|)\\ \gamma(\|\boldsymbol{x}_u-\boldsymbol{x}_2\|)\\ \vdots\\ \gamma(\|\boldsymbol{x}_u-\boldsymbol{x}_n\|)\\ 1\end{bmatrix} \tag{26}$$

对应地，（26）式可简写为$\boldsymbol{\Gamma\lambda}=\boldsymbol{\gamma}$。数值求解该矩阵方程，可以得到（24）式中的最佳加权系数及超参。然后，用（18）式就可以得到空间数据最佳预测结果。

6.2 空间数据分析中泛 Kriging 方法的基本思想

一般 Kriging 方法假设空间数据是广义平稳的，即一阶和二阶统计量是不随时间变化的。但是，空间数据的变化很难具备二阶平稳性（广义平稳性）。最多只能假设是局部二阶平稳的。

为此，去掉空间数据的二阶平稳性假设，就提出了所谓的泛 Kriging 方法。这与把 Wiener 滤波发展到 Kalman 滤波的扩展从思想上看是完全类似的。

此时，必须考虑空间数据一阶统计量是随空间位置变化的。同时，假设空间数据的背景变化（一阶统计量变化）可以表达为

$$E\left[Z(\boldsymbol{x})\right]=\mu(\boldsymbol{x})=\sum_{j=0}^{p}\beta_j f_j(\boldsymbol{x}) \tag{27}$$

因此，非平稳情形下很一般的空间数据模型就可以写为

$$Z(\boldsymbol{x})=\sum_{j=1}^{p+1}\beta_{j-1}f_{j-1}(\boldsymbol{x})+\delta(\boldsymbol{x}) \tag{28}$$

其中，$\delta(\boldsymbol{x})$是零均值平稳随机过程。

基于此，泛 Kriging 的基本思想应该是：泛 kriging 方法中，背景部分（$Z(\boldsymbol{x})$的一阶统计量）是空间变化的，即$E\left[Z(\boldsymbol{x})\right]=\mu(\boldsymbol{x})$，此时$\mu(\boldsymbol{x})$称为漂移。在这种情况下，不能用一般 Kriging 法来计算变异函数。如果漂移函数$\mu(\boldsymbol{x})$是已知的，可以在原始资料中将漂移减去，如剩下的残差$\delta(\boldsymbol{x})$满足二阶平稳性假定，就可用前面讨论的一般 Kriging 法来对残差$\delta(\boldsymbol{x})$进行估值，然后再将漂移$\mu(\boldsymbol{x})$加到对应位置的残差估值上去，其和就是$Z(\boldsymbol{x})$的估计值。这就构成了两步优化逼近问题。漂移函数$\mu(\boldsymbol{x})$是需要估计的，$\delta(\boldsymbol{x})$也要用 Kriging 法进行估计。$\mu(\boldsymbol{x})$一般可以假设为公式（27）的形式，基函数$f_j(\boldsymbol{x})$往往是事先给定的，比如多项式函数，需要估计最佳系数β_j值。当然，也可以把两步估计问题合成一步估计问题。

详细的算法，很多参考文献有具体的讨论。

6.3 空间数据分析中协 Kriging 方法的基本思想

空间数据分析中，存在如下情形，$Z_1(\boldsymbol{x})$数据比$Z_2(\boldsymbol{x})$数据更容易获取，且$Z_1(\boldsymbol{x})$数据中包含了$Z_2(\boldsymbol{x})$数据中的部分信息，也可以说两者存在相关性。因此，把$Z_1(\boldsymbol{x})$用作对$Z_2(\boldsymbol{x})$进行估计的先验信息是很正常的考虑。事实上，存在很多这样的思想和做法。此处，讨论的是协 Kriging 方法的基本思想。

在一般 Kriging 方法对观测数据的假设下，写出如下的协 Kriging 方法

$$Z_2^*(\boldsymbol{x}_0)=\sum_{i=1}^{N_1}\lambda_i^1 Z_1(\boldsymbol{x}_i^1)+\sum_{i=1}^{N_2}\lambda_i^2 Z_2(\boldsymbol{x}_i^2) \tag{29}$$

其中，$Z_2^*(\boldsymbol{x}_0)$代表$Z_2(\boldsymbol{x})$在$\boldsymbol{x}_0$点处的估计值，$Z_1(\boldsymbol{x}_i^1), i=1,2,\cdots,N_1$代表$Z_1(\boldsymbol{x})$在各观测点处的观测值；$Z_2(\boldsymbol{x}_i^2), i=1,2,\cdots,N_2$代表$Z_2(\boldsymbol{x})$在各观测点处的观测值。$\lambda_i^1, i=1,2,\cdots,N_1$代表在$\boldsymbol{x}_0$点附近$Z_1(\boldsymbol{x})$数据的加权系数；$\lambda_i^2, i=1,2,\cdots,N_2$代表在$\boldsymbol{x}_0$点附近$Z_2(\boldsymbol{x})$数据的加权系数。

同样地，在两个观测数据都是广义平稳的假设下，基于类似最佳 Wiener 滤波的要求，期望对$Z_2^*(\boldsymbol{x}_0)$的估计也是无偏和方差最小的。

即有下述式子成立

$$E\left[Z_2(\boldsymbol{x}_0)-\sum_{i=1}^{N_1}\lambda_i^1 Z_1(\boldsymbol{x}_i^1)-\sum_{i=1}^{N_2}\lambda_i^2 Z_2(\boldsymbol{x}_i^2)\right]=-\mu_1\sum_{i=1}^{N_1}\lambda_i^1-\mu_2\sum_{i=1}^{N_2}\left(\lambda_i^2-1\right)=0 \tag{30a}$$

其中，$\mu_1=E\left[Z_1(\boldsymbol{x})\right]$，$\mu_2=E\left[Z_2(\boldsymbol{x})\right]$。上式成立，必有$\sum_{i=1}^{N_1}\lambda_i^1=0$和$\sum_{i=1}^{N_1}\lambda_i^2=1$。

$$\begin{aligned}&\mathrm{Var}\left[Z_2(\boldsymbol{x}_0)-Z_2^*(\boldsymbol{x}_0)\right]=E\left\{\left[\left(Z_2(\boldsymbol{x}_0)-Z_2^*(\boldsymbol{x}_0)\right)-E\left[Z_2(\boldsymbol{x}_0)-Z_2^*(\boldsymbol{x}_0)\right]\right]^2\right\}\\&=E\left[\left(Z_2(\boldsymbol{x}_0)-Z_2^*(\boldsymbol{x}_0)\right)^2\right]\\&=E\left[\left(Z_2(\boldsymbol{x}_0)-\sum_{i=1}^{N_1}\lambda_i^1 Z_1\left(\boldsymbol{x}_i^1\right)-\sum_{i=1}^{N_2}\lambda_i^2 Z_1\left(\boldsymbol{x}_i^2\right)\right)^2\right]\\&\Rightarrow\min\end{aligned}\tag{30b}$$

类似一般 Kriging 方法，（30b）式与约束条件 $\sum_{i=1}^{N_1}\lambda_i^1=0$ 和 $\sum_{i=1}^{N_1}\lambda_i^2=1$ 分别结合，可以用 Lagrange 乘子法构建两个无约束优化反问题。求解两个无约束优化反问题的联立方程，可以求出 $\lambda_i^1, i=1,2,\cdots,N_1$ 和 $\lambda_i^2, i=1,2,\cdots,N_2$，以及引入的对应的超参。利用（29）式就可以进行协 Kriging 法的预测。

由于 $Z_1\left(\boldsymbol{x}_i^1\right), i=1,2,\cdots,N_1$ 和 $Z_2\left(\boldsymbol{x}_i^2\right), i=1,2,\cdots,N_2$ 两组观测数据不一定在同一点观测，此时，无法计算相应的半变差函数及半协变差函数。据此，又提出了所谓的拟协 Kriging 法的思想。具体的内容，查看相应的参考文献。

很显然，更进一步地会有泛协 Kriging 方法和泛拟协 Kriging 方法的合乎逻辑的扩展。这些内容请查看相关的参考文献。

这些思想，尤其是协 Kriging 方法的思想，对多源（元）信息融合显然是有指导性的。但是，客观地讲，Kriging 方法本质上无非是把时间序列分析的 Wiener 最佳滤波思想，相应地推广到空间数据分析中而已。这不构成多源（元）信息融合的十分抽象的理论框架。

7 空间数据分析中的径向基函数建模方法

假如空间数据可以分解为背景+扰动的成分，且背景部分可以基函数族的线性组合进行逼近。Weierstrass 定理下的函数逼近可以作为空间数据分析的另一个理论框架。所不同的是基函数（族）的选择要与空间数据的统计特征关联起来。这就是径向基函数（族）逼近的空间数据建模的基本逻辑。

空间数据的径向基函数（RBF = Radial Basis Function）无约束模型是

$$Z(\boldsymbol{x})=\sum_{i=1}^{n}\lambda_i\varphi_i(\boldsymbol{x})=\sum_{i=1}^{n}\lambda_i\varphi\left(\left\|\boldsymbol{x}-\boldsymbol{x}_i\right\|\right)\tag{31}$$

其中，$\varphi\left(\left\|\boldsymbol{x}-\boldsymbol{x}_i\right\|\right)$ 是以 $\boldsymbol{x}_i$ 为中心的径向基函数。

空间数据的径向基函数有约束模型是

$$Z(\boldsymbol{x})=\sum_{i=1}^{n}\lambda_i\varphi\left(\left\|\boldsymbol{x}-\boldsymbol{x}_i\right\|\right)+c_0+c_1\boldsymbol{x},\ s.t.\ \sum_{i=1}^{n}\lambda_i=0,\ \sum_{i=1}^{n}\lambda_i\boldsymbol{x}_i=0\tag{32}$$

其中，多项式 $p(\boldsymbol{x})=c_0+\boldsymbol{c}_1^T\boldsymbol{x}$ 是用来确保逼近精度的，约束条件是为了使全局光滑曲面曲率最小。

用 Lagrange 乘子法把（32）式转为无约束优化问题，对目标泛函求极小，最终得到如下矩阵方程

$$\begin{bmatrix} \varphi(\|\boldsymbol{x}_1-\boldsymbol{x}_1\|) & \varphi(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|) & \cdots & \varphi(\|\boldsymbol{x}_1-\boldsymbol{x}_n\|) & 1 & \boldsymbol{x}_1 \\ \varphi(\|\boldsymbol{x}_2-\boldsymbol{x}_1\|) & \varphi(\|\boldsymbol{x}_2-\boldsymbol{x}_2\|) & \cdots & \varphi(\|\boldsymbol{x}_2-\boldsymbol{x}_n\|) & 1 & \boldsymbol{x}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi(\|\boldsymbol{x}_n-\boldsymbol{x}_1\|) & \varphi(\|\boldsymbol{x}_n-\boldsymbol{x}_2\|) & \cdots & \varphi(\|\boldsymbol{x}_n-\boldsymbol{x}_n\|) & 1 & \boldsymbol{x}_n \\ 1 & 1 & \cdots & 1 & 0 & 0 \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} f(\boldsymbol{x}_1) \\ f(\boldsymbol{x}_2) \\ \vdots \\ f(\boldsymbol{x}_n) \\ 0 \\ 0 \end{bmatrix} \tag{33}$$

其中，$f(\boldsymbol{x}_i)$是已知点的值。用共轭梯度法求解，得到系数$\lambda_1,\lambda_2,\cdots,\lambda_n$。然后可以基于最佳模型进行空间数据的预测。

空间数据的径向基函数无约束预测是

$$\hat{Z}(\boldsymbol{x}_u)=\sum_{i=1}^{n}\lambda_i\varphi(\|\boldsymbol{x}_u-\boldsymbol{x}_i\|) \tag{34}$$

空间数据的径向基函数有约束预测是

$$\hat{Z}(\boldsymbol{x}_u)=\sum_{i=1}^{n}\lambda_i\varphi(\|\boldsymbol{x}_u-\boldsymbol{x}_i\|)+c_0+c_1\boldsymbol{x}_u \tag{35}$$

很多文献列出了不同的径向基函数可供选择。但是，很显然径向基的不同在很大程度上决定了空间数据的建模预测效果，最好选择能与空间数据变化特征一致的径向基函数。

另外，空间数据分析中的径向基函数建模方法仅仅对$\boldsymbol{Z}(\boldsymbol{x})$中的背景部分建模，并没有对广义平稳部分进行建模。理论上讲，还是要和广义平稳部分的建模串联起来，基于 EM 算法，迭代地求解，更好地对$\boldsymbol{Z}(\boldsymbol{x})$进行建模预测。

8 勘探地震中多信息融合的基本应用

油气地震勘探发展到现在，很多简单问题基本已经得到了很好或较好地解决。余下的都是复杂问题。其中两个最核心问题依然是：宽带弹性参数估计或广义地震波反演成像；定量的油藏描述和含油气性评价。这两个问题都是强非线性的，必须引入多种先验信息才能提高它们的求解精度。

但是，一些先验信息的获取又是一个需要先验信息约束的（强）非线性反问题。譬如强噪音下弱初至波的识别与走时检测；低信噪比情况下的层位识别与断层识别；成像剖面上的地震相识别；进一步地，结合岩心数据和井数据把地震相合理地变成地震岩相；散乱的井数据支撑下的弹性参数建模（生成伪井数据）等等。尽管它们不同于波动方程参数反演问题（它有一个明确的波动方程正问题），所有这些问题（缺乏一个类似的波动方程正问题），本质上，都可以归结为强非线性情形下的建模和预测问题。引入多源（元）信息的约束是必不可少的。

事实上，最佳模型构建的理论框架是比较明了和确定的，研究也是很深入的。但是，多源（元）信息如何有机地融合起来，有效地约束建模结果，提高建模的质量，进而提高预测结果的精度，反而成了薄弱的问题。

多源（元）信息如何有机地融合起来，真正的困难还是来自于多源（元）信息的异质性、尺度差异、精度差异、模糊性及随机性等。因此，很难构建一个统一的、公认的理论框架来处理这个问题。

多元信息融合的基本逻辑应该是这样的。同一个物理实体，有不同方面的物理参数，可以（也十分有必要）从不同的角度来描述该物理实体。这些物理参数之间是有相互制约关系的。它们表现为一定程度的相关关系、甚至因果关系。不同的物理场数据与不同的物理参数联系在一起。对应有各自的正问题。提高不同物

理参数的估计精度，应该利用它们之间存在的相互制约关系。通过联合，通过参数间的相互制约，提升每一个方面的物理参数的反演精度及稳定性。这是**联合参数反演**的本意。如何利用参数间的制约关系，在假定其他方面的物理参数估计结果（譬如重磁电反演结果）已知的条件下，把这些先验信息用于提升某一个特定方面的物理参数（譬如地震波成像）的估计精度及稳定性，这是**多信息约束参数估计/反演**的本意。**多信息约束建模**，是把不同方面的物理参数融合在一起，生成期望的由某个方面的物理参数表达的模型。多信息约束建模与速度估计是不同的，一般地，它不再利用波动方程正问题模拟实测数据，即不再进行参数估计。可以看出，**联合参数反演、多信息约束参数估计/反演、多信息约束建模**三者之间是有紧密联系的。但是，不易把它们合成一个问题来解决。这样做只会增加问题求解的难度。

我们认为：首先统一地表达这些复杂的多源（元）数据，就是把它们表达为随机（数据）集。然后，基于样本数据，提取随机（数据）集的统计特征。并对随机（数据）集进行统计建模。最后，进行基于模型的推断或判决。

这是个抽象的表述，如何转化成不同应用领域的具体算法，是要继续探究的问题。

我们认为：空间（随机场）数据的建模应该可以作为一个基本的理论框架，尤其再继续发展到 Kalman 滤波、粒子滤波理论的情形下。多源（元）信息融合要体现在最佳建模的过程中。把多源（元）信息约束下的最佳建模问题提成一个合理的约束优化反问题是转化成不同应用领域的具体算法的基本的数学操作。勘探地震中的**（多物理场）联合参数反演、多信息约束参数估计/反演、多信息约束建模**等等需要多信息联合解决的问题应该可以在上述理论框架下得到较好的解决。

9 结论与讨论

关于空间数据分析，有如下的观点：

尽管空间数据的确有自己特殊的属性（不同位置的空间随机变量之间的相关关系描述起来更复杂，而且采样更可能不均匀），但是数据分析的本质上，与时间序列分析并无根本的不同（仅仅在一维时间顺序上体现出相关关系，采样基本均匀）。

空间数据被视为趋势项+广义平稳随机过程项+不可预测的白噪声项。时间序列数据也是如此。

趋势项是空间数据（时间序列数据）的均值项。去掉趋势项后，零均值的广义平稳随机过程项就仅仅由（含有结构的）协方差矩阵决定。很显然，空间数据（时间序列数据）应该被视为均值为$\boldsymbol{\mu}(\boldsymbol{x})$、协方差阵为$\Sigma$（假设$\Sigma$是缓变的，不是点点变化的）的广义平稳过程。对空间数据（时间序列数据）的建模，就在建一个预测模型，对均值为$\boldsymbol{\mu}$、协方差阵为Σ的广义平稳过程进行预测。

为了简化问题，把上述预测拆分为两部分：趋势项的建模预测和广义平稳随机过程项的建模预测。前者就是个确定性的函数逼近问题（线性或非线性回归分析）；后者是 ARMA 生产随机过程的建模问题。把结果叠加在一起，构成对均值为$\boldsymbol{\mu}(\boldsymbol{x})$、协方差阵为$\Sigma$的广义平稳过程的预测结果。

可以清楚地看出：对均值项$\boldsymbol{\mu}(\boldsymbol{x})$的估计，需要已知（上述）广义平稳过程项的协方差阵$\Sigma$。而对（上述）广义平稳过程项进行建模时，也要已知均值项。理论上，就是要同时对均值项$\boldsymbol{\mu}(\boldsymbol{x})$和（上述）广义平稳过程项同时建模。

因此一般地是假设均值项为零或常数，对（上述）广义平稳过程项进行 ARMA 建模。把常数均值融在 ARMA 模型中一起估计显然是同时建模的简单特例。但是，这样的假设很不符合实际，常常导致建模效果质量下降。

解决上述问题的基本观点是：提一个约束优化反问题对空间数据中确定性部分进行建模，用 AR 模型对广义平稳随机过程部分进行建模，并把二者串联起来进行迭代求解（类似 EM 算法），应该是合理的策略。这样做容易考虑空间数据变化的物理机制，在空间数据建模中引入额外的先验信息，得到更好的最终逼近结果。

关于多元数据融合，有如下的观点：

众所周知，勘探地震中的两个核心问题：宽带弹性参数估计或广义地震波反演成像；定量的油藏描述和含油气性评价，都是一个信息不足情况下的最佳判决问题。前者的目标是得到最佳的弹性参数估计结果，后者是得到尽可能可靠的油藏描述和评级结果。它们都是强非线性。解决这些强非线性问题，引入相关的先验信息是必须的。勘探地震中这些先验信息主要包括：与钻井岩心相关的数据、与测井相关的数据、与地震成像剖面蕴含的地下反射结构相关的数据。这些要么是空间散乱点处的，要么是规则三维空间网格点上的，且是二进制表示的数据。而地质知识和岩石物理关系这些非二进制表达的信息也是非常重要的与待估计参数相关的先验知识。这些先验信息表现出的异质性、尺度差异、精度差异、模糊性及随机性等等是它们具有的典型特征。因为这些先验信息都是空间域的，空间数据的统计分析不可避免地成了空间信息融合的数学基础。目前并没有统一地利用这些先验信息的数学框架。

但是，首先要统一地表达这些复杂的数据，然后建立统一的数据集所满足的概率密度函数。接着对概率密度函数中包含的参数进行估计。最后利用所建立的最佳模型进行推断，解决各种具体问题。这应该是比较抽象的、也比较统一的多源（元）信息融合的理论逻辑。把空间数据分析的思想进一步提升到多种信息融合生成期望模型是比较具体的，应该可以发展出实用的算法。多种信息表现为空间散乱数据的形式，同一个空间点处，具有同一客观实体的不同方面的特征，在多种信息的约束下，生成描述该客观实体的某种模型，这是空间数据分析下多元数据融合的基本思想。这显然是个 Bayes 决策问题。需要提一个决策代价函数，在决策代价最小的情况下，生成需要的期望模型。

具体到勘探地震中，主要还是把多源（元）信息约束下的最佳建模问题提成一个合理的约束优化反问题，通过求解这个约束优化反问题达到多源（元）信息融合并约束提升建模结果精度的目的。

参考文献

[1] Ronald P. S. Mahler, 2007, Statistical Multisource – Multitarget Information Fusion, Artech House, Inc.
[2] Gamerman, D., 2006, Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference；Chapman & Hall/CRC
[3] Cressie, N. A. C., 1993, Statistics for Spatial Data; John Wiley & Sons, Inc.
[4] Schabenberger, O., and Gotway, C. A., 2005, Statistical Methods for Spatial Data Analysis; Chapman & Hall/CRC
[5] Mahler, R. P. S., 2007, Statistical Mutilsource – multitarget Information Fusion; Artech House, Inc.
[6] Evensen, G., 2009, Data Assimilation: The Ensemble Kalman Filter; Second Edition, Springer
[7] Richard A. Johnson and Dean W. Wichern, 2007, Applied Multivariate Statistical Analysis, 6th Edition, Prentice Hall.

结语

我们知道，勘探地震学的最终目标是得到宽频带（或称宽波数带）的储层弹性参数的反演结果，进而进行含油气储层的定位、识别、描述与刻画。LS_RTM+FWI 是目前勘探地震领域中极受关注的标志性技术，但勘探地震中的实践表明，仅有地表观测的地震数据，不对勘探地震介质的特征和波场特征加以利用，FWI 方法即使仅用透射波进行基于低频长偏移距数据的背景参数的估计也不是很稳定的（视数据情况达到收敛或不收敛）。我们认为，一定假设下的地震波正演方法可以预测实测地震波场是地震波反演最基本的要求。既然无论如何选择地震波正演模拟方法也无法完全预测实测数据中的波场，转而追求尽可能好地模拟其中的部分特征波场，进行基于特征波场的反演成像是合理的逻辑。特征表达参数模型和特征表达数据体，得到二者之间存在更密切联系的关系可以把强非线性的经典 FWI 问题转化为一个较凸的反问题进行求解。因此，宽频带（或称宽波数带）的储层弹性参数的估计问题一定需要一系列的技术组合来实现。我们认为，透射波层析速度反演+Monte Carlo 方法反演宏观背景速度模型+Beam 波角度道集层析速度反演+数据域特征波层析反演+LS_RTM 保真叠前深度偏移+宽带波阻抗建模，结合岩石物理知识，可以实现针对复杂储层的描述和刻画。

在上述技术发展过程中，对我国（包括世界上）重要探区的地下介质分布特征的抽象与描述、波在这样的介质中传播规律及地震数据特征的抽象与表达是 WPI 研究组进行理论与方法研究的出发点。我们一直认为针对实际问题的创新研究才是地震波反演成像研究应该做的，因此必须紧紧抓住波在实际介质中的传播特征这个基本问题。

我们知道，波阻抗成像结果是与油藏描述关系更为密切的。遗憾的是，相对波阻抗成像结果对于地震解释基本没有用处，绝对波阻抗成像必须要有准确的背景部分，否则相对误差会大到对于油藏描述也完全没有帮助的程度。绝对波阻抗反演只有在宽带（尤其是低频）、宽方位数据存在的条件下，波阻抗的背景部分能较为准确地确定的情况才有可能实现。当前，“两宽一高”的地震数据采集已经成为地震数据采集的核心方法技术。在“两宽一高”地震数据逐渐普及应用的情况下，WPI 研究组认为开展绝对波阻抗的研究已经具备了基本的条件。

在地震波偏移成像与反演成像的实践过程中，我们愈来愈深刻地认识到近地表的复杂性及近地表速度模型（包括各向异性参数）和吸收衰减参数模型的正确性对山前带准确成像、东部探区保真成像、深层和超深层成像是非常重要的因素。海上，尤其是我国各浅海探区，与海水面和海底面有关的干扰波场是高精度成像的最主要影响因素。因此，为了得到高质量的目标层成像结果，需要把关注点更多地放在表层和浅层。就是说近地表地球物理方法应受到越来越充分的重视。

非学习类算法（譬如 FWI，LS_RTM，Tomography 等）是 Bayes 框架下一类 Gauss 线性反演的核心内容。正问题（基于物理数学规律建起数据预测器，本质上说因果关系或估计参数与数据的关系是人为定的），误差泛函定义，正则化思想与做法，梯度导引类迭代算法是非学习类算法涉及的几项核心内容。非学习类算法最关键的地方（与学习类算法相比）是它没有所谓的泛化能力，对每一个新数据都要重新运算一遍，没有经验积累，正是从这个意义上讲，学习类算法优于非学习类算法。非常有必要把目前的非学习类算法推广到学习类算法。我并不认为学习类算法的引入就能很好地解决当前的油气地震勘探中的地震数据分析问题（主要是地震波反演成像（弹性参数估计）和储层描述），但如果不研究学习类算法，一定是没认识到学习类算法相对于非学习类算法的优势。我相信现代数据分析的方向一定是基于学习类算法的，发展出各领域的“学习机器人”是真正的目标。

本文集是从 12 卷波现象与反演成像研究组（WPI）年度论文集中把我写的文章按年度收集并排列起来形成的。目的是让 WPI 研究组的学生通过阅读此文集尽快理解 WPI 研究组的学术观点，更快地进入深度研

究状态。很多地方有重复的内容，也有表达不妥的地方。毕竟近供内部使用，存在的问题在日常的相互交流中得到解决。

致谢

感谢中国石化集团公司科技部、中国石化石油物探技术研究院、中国石化勘探开发研究院、中国石化胜利油田分公司物探研究院、中国石油东方地球物理公司、中国石油勘探开发研究院西北分院、中国海油研究总院、中国海油天津分公司勘探开发研究院和中国海油湛江分公司对波现象与反演成像研究组研究工作的不断支持。