

---

# 波现象与智能反演成像研究组（WPI）的哲学思维

王华忠

波现象与智能反演成像研究组（Wave Phenomena and Intellectual Inversion Imaging Group）

同济大学海洋与地球科学学院

众所周知，数据采集和数据分析是信号与图像分析学科以及很多相关学科（以 Remote Sensing 为基本任务的学科）的基本研究内容。勘探地震学也是一个以数据采集和数据分析为基本任务的学科，只不过它的数据类型、数据背后遵循的物理规律、数据分析的对象和目的有其特殊性。

WPI 认为，二十一世纪的勘探地震学已经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。最近几年，大数据分析和人工智能技术在智能数据采集设备、互联网（物联网）和 4G/5G 网络、高性能计算与云计算等技术的促进下有了突飞猛进的发展，迅速地改变了和正在继续改变着人们的生产和生活形态。统计学习与机器学习的理论与方法支撑着大数据分析和人工智能技术的进步。而信号与图像的建模预测理论和 Bayes 框架下的参数估计理论是现代信号与图像分析学科的两项基本内容，也是机器学习理论框架中的核心内容。

油气地震勘探的根本目标是利用实际测量的叠前数据体通过反演成像方法获取地下岩石介质的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，从而满足油气开发的要求。其中的关键问题是如何获取宽波数带的弹性参数场，而弹性参数估计问题是一个典型的、经典的反演问题。

事实上，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着。尤其在大数据分析时代，核心就是在解决这个反问题。

由于可观测量总是有限的（不完备的）、而且是受“噪音”污染的（不完美的），不可能指望从可观测量中获取不可观测量的全部信息，因此需要补充关于不可观测量的先验信息。这是反演问题的根本逻辑。因此，WPI 给反问题的基本定位是：**它总是一个不充分与不完美可观测信息下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性上。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，求解反问题就是为了从这些不确定中得到一些关于不可观测量的有意义的认识。这种表面上的悖论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。

---

完备的、少“噪音”污染的实测数据体是获得高精度反演成像结果的基础。“两宽一高”的地震数据采集技术是目前和今后很长一段时期内高精度反演成像所要求的观测方式。但与震源和检波器相关的硬件技术以及自动化和智能化的采集手段的发展还不能充分地满足“两宽一高”地震数据采集方式的真正实现。而且压缩感知理论下的随机采样技术深受复杂噪音、复杂地表和复杂地下介质的影响难以广泛展开。令人欣慰的是，当今时代先进制造、智能制造技术发展很快，无论是海上或是陆上的地震数据自动化采集技术装备都得到迅速的发展，尤其是海上无人震源、海上可控震源、海上节点检波器技术正逐渐进入实用化阶段，陆上无人“节点”检波器技术也在进行探索。相信不久的将来，自动化和智能化的“两宽一高”采集技术将会逐步实现。

但是，对高精度反演成像而言，数据的完备性不仅仅是要求尽可能采集“两宽一高”的叠前地震数据体。关于要反演参数的先验信息的采集也是非常重要的。譬如精确的近地表高程信息、岩石露头信息、精确的海底高程信息、高光谱信息、各种近地表调查信息、测井信息、录井信息、岩石物理信息等等都是高精度反演所需要的先验信息。

大数据分析时代，对数据的认识、对数据的管理、对数据的加工和预处理、对数据的利用等等要有实质性地提升。

反问题的根本思想逻辑在于反问题如何提出、反问题如何求解。

关于如何提出反问题，WPI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，是一个具有高度抽象性和包容性的反问题提法。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。WPI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。WPI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。

更为实用的、基于代价泛函提出反问题，WPI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 WPI 提出的最基本的概念，在数据残差的度量中必须也只能对同因之果间的差异进行度量。差异度量的方式的确是重要的，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，差异度量方式对度量结果函数的凸性

---

的贡献依然是第二位的。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  $L_\infty$  或  $L_p, p \neq 2$  之类的范数，WPI 认为没有十分的必要，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。总之，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。WPI 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。罚函数的定义主要考虑把待估计参数的结构信息能引入到反演解中，同时对待估计参数施加稀疏性要求。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，WPI 认为不动点理论是最高的理论框架，它简洁、抽象，且有足够的包容性。基于变分原理和 Newton 微积分思想，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。投影算子方法的基本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可行解集中的解收敛到一个不动点上。可行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。

现代信号与图像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与图像分析的知识，譬如高维数据规则化、高维数据去噪音、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应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。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高维局部平面波稀疏反演方法 (或者高维小波变换+稀疏反演方法)。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线性信号的相关性, 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, 这是独立分量分析 (ICA=Independent Component Analysis) 要解决的问题。当然, 最好是用导致信号变化的原因变量来表达信号。但是由数据驱动的方法直接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。总之, 可以认为机器学习理论是现代信号分析理论的深化。由数据驱动, 基于 Bayes 估计理论, 建立实测数据与潜在特征之间的关系, 是机器学习的核心任务。

WPI 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。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。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, 所有的现代信号与图像分析/机器学习 (ML)/人工智能 (AI) 领域的新思想和新进展都可以借用来促进地震波反演成像精度的提升、促进油藏描述精度的提升。

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以 Bayes 估计理论为基础的现代信号与图像分析/机器学习 (ML)/人工智能 (AI) 框架下。在此总框架下, 基于“两宽一高”的叠前地震数据和待估计参数的各种先验信息, 开展 Bayes 估计理论下的参数反演, 得到宽波数带的弹性参数估计结果, 并基于机器学习算法综合各种信息实现高精度的储层描述, 是 WPI 确定的学术和方法技术发展导向。