波现象与智能反演成像研究组

同济大学

# 几何在数据分析及图像分析中的核心作用分析

报告人：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，上海

2022年04月13日

# 目 录

# ◆一、概述

◆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。

◆**但是，……**

# ◆一、概述

◆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。线性成像问题理论上已经定型，值得理论研究的问题很少了。

- ◆ 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。线性化的成像不再满足实际情形。

◆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。

# ◆一、概述

◆撇开非线性问题不谈，提高成像精度的合理途径是引入多源信息约束。

◆多源信息来源：

- ◆来自物理系统的高维规则数据体；
- ◆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。

◆地震波反演成像的目标：

- ◆三维空间中的弹性参数场，勘探地震中，几何结构+层内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的正确性是至关重要的。

# ◆一、概述

◆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。

◆我认为：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（尽管不能谈正确的）地质体几何结构的约束，是最重要的措施。

# ◆一、概述

◆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。

◆这也是信号和图像分析领域最核心的问题。

# ◆一、概述

◆总而言之，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、微积分的思维和知识，更需要清晰的几何思维和完整的几何知识。

◆我个人最喜欢的思维模式：物理+几何。

- ◆在形成一定的认识后，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。

# 目 录

# ◆二、基本的几何知识回顾

◆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

◆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。

# ◆二、基本的几何知识回顾

## ◆基本几何知识包括：

- ◆平面几何
- ◆立体几何
- ◆解析几何
- ◆微分几何

# ◆二、基本的几何知识回顾

## ◆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
- ◆园及和园相关的图形

- ◆平面三角学

# ◆二、基本的几何知识回顾

## ◆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

# ◆二、基本的几何知识回顾

## ◆解析几何研究什么?

- ◆向量代数
  - ◆点积、叉积、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

$$\rho = \cos\varphi = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2\,|\boldsymbol{b}|^2}}$$

$$\rho = \sin\varphi = \frac{\boldsymbol{a}\times\boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2\,|\boldsymbol{b}|^2}}$$

# ◆二、基本的几何知识回顾

## ◆微分几何研究什么？

- ◆平面曲线
- ◆空间曲线
- ◆曲面

◆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。

# ◆二、基本的几何知识回顾

## ◆微分几何研究什么？

### ◆平面曲线

- ◆ 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、曲率园与曲率中心
- ◆ 特殊点：
  - ◆ 1、拐点
  - ◆ 2、顶线
  - ◆ 3、奇点

### ◆平面曲线

- ◆ 关于曲线的研究
  - ◆ 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

地震剖面中的层位和断层、特殊地质体的边界都是平面曲线！

# 目 录

# ◆三、勘探地震中信息的尺度问题

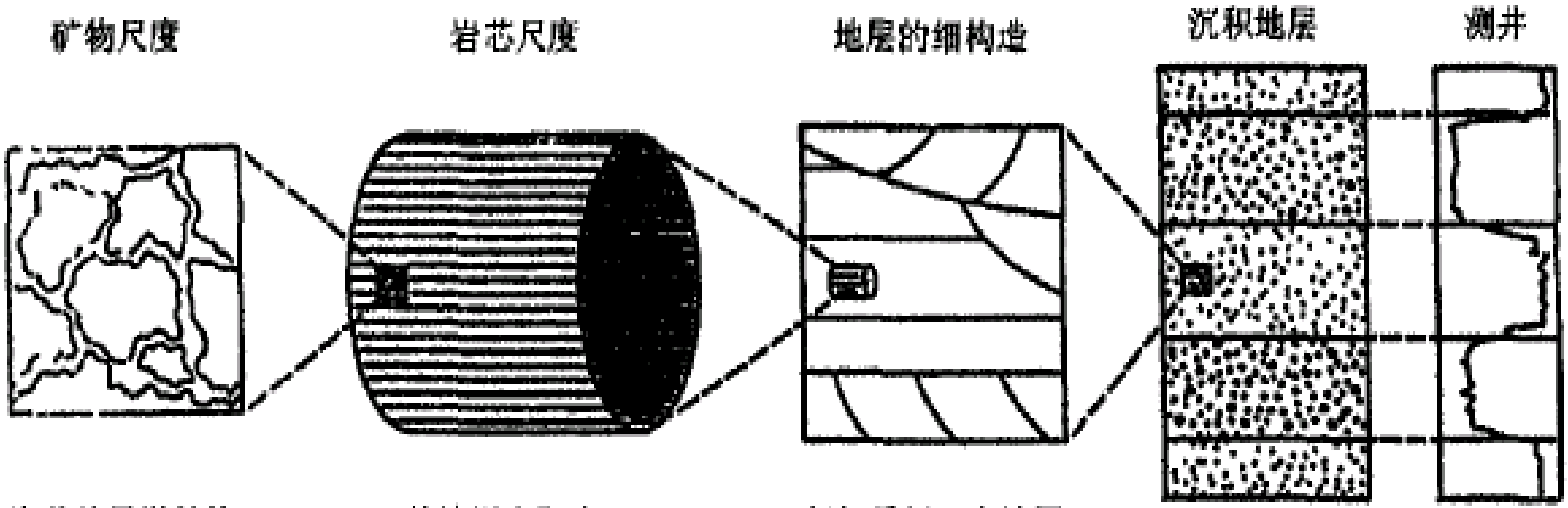
◆Causality – namely our awareness of what causes what in the world. Though it is basic to human thought, causality is a notion shrouded in mystery, controversy, and caution, because scientists and philosophers have had difficulties defining when one event truly causes another.

◆Judea Pearl, CAUSALITY (Second Edition), 2009

◆我们真的知道波场与储层含油气之间的因果关系吗?

◆关于地震勘探的物理，我们能确知的是什么？尺度是个关键！

# ◆三、勘探地震中信息的尺度问题




岩芯的显微结构，从显微尺度看岩芯又是不均匀的了

从地层中取出一块岩芯，可以将它看成是均匀的

仔细看每一个地层，其中存在着断层、节理，地层仿佛是不均匀的

均匀的地层

**微米以下级别**　　**厘米级别**　　**米级别**　　**米级之上的级别**

## 不同尺度的勘探地震介质（微观、介观到宏观）

# ◆三、勘探地震中信息的尺度问题

| 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 | |

**材料科学中的宏观、介观与微观的尺度量级**

勘探地震中的尺度如何划分？具体量级如何界定?

# ◆三、勘探地震中信息的尺度问题

## 假定特征速度3500m/s

| 特征主频 | 5Hz | 15Hz | 25Hz | 35Hz | 45Hz | 55Hz | 65Hz | 75Hz |
|---|---|---|---|---|---|---|---|---|
| 特征波长 | 700m | 233m | 140m | 100m | 77m | 63m | 53m | 47m |

**勘探地震学对应的特征速度、特征主频和特征波长**

# ◆三、勘探地震中信息的尺度问题

| 按尺度划分的地质体类别 | 地质特征 | 尺度量级 | 数学模型 | 分析方法 | 分析精度 |
|---|---|---|---|---|---|
| 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 | 岩芯观察与统计分析 | 半定量 |
| 储层 | 岩石颗粒定向排列、节理、裂缝等 | 数公分到米级别 | ? | 多源信息综合与评价 | 定性 |
| 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 | 地震勘探 | 定量 |

**勘探地震中介质、尺度、数学模型及分析方法**

# ◆三、勘探地震中信息的尺度问题

◆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。
- ◆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！

# ◆三、勘探地震中信息的尺度问题

◆勘探地震学是由波动理论，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。

◆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。

# ◆三、勘探地震中信息的尺度问题

◆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?

# ◆三、勘探地震中信息的尺度问题

这是一个Bayes参数估计问题

这两个尺度上，弹性参数的有机融合的理论基础还存在问题。井点上的信息不能与三维体上的信息有机地融为一体。

## ◆我对油气勘探的基本观点：

- ◆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：高精度的低中波数成分（各向异性）速度模型建立问题
  - ◆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。
- ◆测井与岩石物理学解决（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。
- ◆油气地质学知识弥合上述两种方法技术，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。

结合岩石、矿物、沉积、构造，各种先验信息和知识，进行综合评估，给出钻井决策。

# ◆三、勘探地震中信息的尺度问题

◆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**的数学模型都只是在一定的理论假设下（精确）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。

◆更多地依赖**相关关系**进行统计推断是会成为地震数据分析的主导思想。

# 目 录

## ◆四、多源信息的几何意义

◆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。

# ◆四、多源信息的几何意义

◆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Machine Learning: A Bayesian & Optimization Perspective 一书中的19章后半部分内容。

◆数据中蕴含的信息（即数据中的线性结构）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。

# ◆四、多源信息的几何意义

◆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。

# ◆四、多源信息的几何意义

◆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），“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。数据对于计算机而言，本质上是一样的！此处，计算机=算法。

◆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，是数据间角度一致的部分。

◆譬如，$\boldsymbol{a} \in S_1$ $\boldsymbol{b} \in S_2$ 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，其他都是由此变出来的！

# ◆四、多源信息的几何意义

◆譬如，两个二维信息源，如何评价他们的相关性？先定义它们的梯度向量 $\boldsymbol{d}_1 \in S_1$ 和 $\boldsymbol{d}_2 \in S_2$，用 $\rho = \cos\varphi = \dfrac{\boldsymbol{d}_1 \cdot \boldsymbol{d}_2}{\sqrt{|\boldsymbol{d}_1|^2 |\boldsymbol{d}_2|^2}}$ 测评二维信息源之间的相关性或是共性成分！

◆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。

# ◆四、多源信息的几何意义

◆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 $x \in S_1$ $y \in S_2$，二者是什么关系？。 一般地，表达为函数关系：$y = f(x)$

◆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: $y = f(x)$。这是大大小小物理关系形成的基本逻辑，参考“物理手册”，“岩石物理手册”。
- ◆ 若有很多$x$ 与 $y$ 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。
- ◆ 另外，专家知识也可抽象成几点规则来定性地描述 $x$ 与 $y$ 之间的关系。

# ◆四、多源信息的几何意义

◆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，多源信息融合下的成像，线性成像问题理论上已经定型，值得理论研究的问题很少了！

# 目 录

# ◆五、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。

# ◆五、几何在数据分析和图像分析中的应用分析

◆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。

# ◆五、几何在数据分析和图像分析中的应用分析

◆我眼中的高维数据、高维图像与高维函数（泛函）：

◆此处不包括算子。算子是一种映射，几何上，就是一种坐标变换。

◆我喜欢思维上的提纲挈领，不喜欢一开始就面对一堆乱呼呼的细节！

◆高维数据、高维图像与高维函数（泛函）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。

# ◆五、几何在数据分析和图像分析中的应用分析

◆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：

◆1、Hilbert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。

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆1、Hilbert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。

- ◆平方可积的函数空间是Hilbert空间，是数据空间，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）
  - ◆基函数（字典）的线性组合 $\sum_{i=1}^{L} \lambda_i e_i^s$
  - ◆基于正交投影的几何结构的建模预测
- ◆解决各种各样的一大类应用问题

- ◆这是对数据/图像的建模问题。常被归结为模型驱动的数据分析方法。

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆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。

**定义5** (正交) 设$H$是内积空间,$x,y\in H$, $M,N\subset H$.

(1) $x\perp y \Leftrightarrow \langle x,y\rangle=0$;

(2) $x\perp M \Leftrightarrow \forall y\in M$, 都有$\langle x,y\rangle=0$;

(3) $M\perp N \Leftrightarrow \forall x\in M$, $\forall y\in N$,都有$\langle x,y\rangle=0$.

◆数据空间与噪声空间的分解：

$$H=\left\{e_1^s,e_2^s,\cdots,e_L^s;e_1^n,e_2^n,\cdots,e_K^n\right\} \quad M=\left\{e_1^s,e_2^s,\cdots,e_L^s\right\} \quad N=\left\{e_1^n,e_2^n,\cdots,e_K^n\right\}$$

$$d=\sum_{i=1}^{L}\lambda_i e_i^s+\sum_{i=1}^{K}\mu_i e_i^n$$

◆解决各种各样的一大类应用问题

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆2、数据空间与噪声空间的正交分解。

### ◆Sergey Fomel对数据空间与噪声空间的正交分解的灵活应用：

- ◆有实测数据：$\boldsymbol{d}$
- ◆有线性（或非线性）结果的预测（即信号预测）：$\hat{\boldsymbol{s}}_0 = P[\boldsymbol{d}]$
- ◆有预测的噪音：$\hat{\boldsymbol{n}}_0 = \boldsymbol{d} - \hat{\boldsymbol{s}}_0 = \boldsymbol{d} - P[\boldsymbol{d}]$
  - ◆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 $\boldsymbol{w}$ ！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：$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \left\| \hat{\boldsymbol{n}}_0 - \boldsymbol{w}P[\boldsymbol{d}] \right\|^2 = \arg\min_{\boldsymbol{w}} \left\| \boldsymbol{d} - P[\boldsymbol{d}] - \boldsymbol{w}P[\boldsymbol{d}] \right\|^2$

$$\hat{\boldsymbol{w}} = \frac{\hat{\boldsymbol{n}}_0^T \hat{\boldsymbol{s}}_0}{\hat{\boldsymbol{s}}_0^T \hat{\boldsymbol{s}}_0}$$

这个滤波器的本质显然是提取噪音中那些与预测的信号相关性最大的成分。滤波器系数很小，说明噪声中没有或很少有残余信号了。

# ◆五、几何在数据

◆2、数据空间与噪声空间的正交分解。

◆ Sergey Fomel对数据空间与噪声空间的正交分解的灵活应用：

Wiener滤波并不本质，相关才是深入数据体内的本质思想！

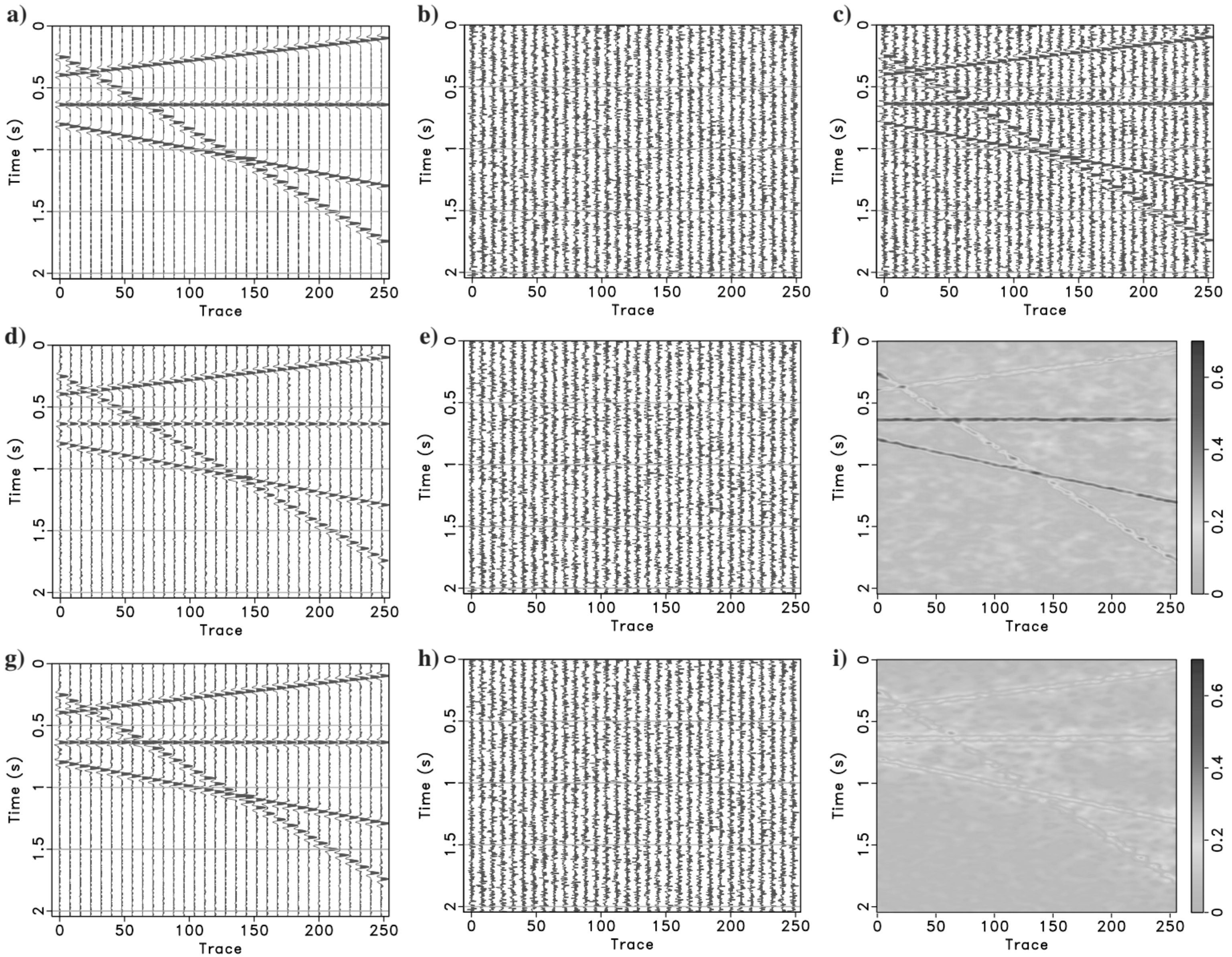



Figure 3. Comparison of using the local orthogonalization-based random noise attenuation approach for denoising test before and after: (a) Clean section, (b) random noise section, (c) noisy section, (d) denoised result using *f-x* deconvolution, (e) noise section using *f-x* deconvolution, (f) local 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).

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆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：$$\boldsymbol{a}=\sum_{i=1}^{N}\lambda_i\boldsymbol{\varphi}_i=[\boldsymbol{\varphi}_1,\ \boldsymbol{\varphi}_2,\cdots,\ \boldsymbol{\varphi}_N]\cdot\begin{bmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{bmatrix}$$

认为数据中的几何结构由基函数来表达

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆3、向量及函数的自相关与互相关。

$$a^T a = [\lambda_1, \ \lambda_2, \cdots, \ \lambda_N] \cdot \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_N \end{bmatrix} [\varphi_1, \ \varphi_2, \cdots, \ \varphi_N] \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix}$$

PCA分析要解决的问题

◆有数据向量的自相关矩阵及分解：

$$= [\lambda_1, \ \lambda_2, \cdots, \ \lambda_N] \begin{bmatrix} (\varphi_1, \varphi_1) & (\varphi_1, \varphi_2) & \cdots & (\varphi_1, \varphi_N) \\ (\varphi_2, \varphi_1) & (\varphi_2, \varphi_2) & \cdots & (\varphi_2, \varphi_N) \\ \vdots & \vdots & \ddots & \vdots \\ (\varphi_N, \varphi_1) & (\varphi_N, \varphi_2) & \cdots & (\varphi_N, \varphi_N) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix}$$

基函数可以由数据自相关矩阵分解出来

$$= [\lambda_1, \ \lambda_2, \cdots, \ \lambda_N] \begin{bmatrix} (\varphi_1, \varphi_1) & 0 & \cdots & 0 \\ 0 & (\varphi_2, \varphi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\varphi_N, \varphi_N) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix}$$

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆3、向量及函数的自相关与互相关。

$$a^T b = [\lambda_1, \ \lambda_2, \cdots, \ \lambda_N] \cdot \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_N \end{bmatrix} [\psi_1, \ \psi_2, \cdots, \ \psi_N] \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix}$$

CCA分析要解决的问题

### ◆有数据向量的互相关矩阵及分解：

$$= [\lambda_1, \ \lambda_2, \cdots, \ \lambda_N] \begin{bmatrix} (\varphi_1, \psi_1) & (\varphi_1, \psi_2) & \cdots & (\varphi_1, \psi_N) \\ (\varphi_2, \psi_1) & (\varphi_2, \psi_2) & \cdots & (\varphi_2, \psi_N) \\ \vdots & \vdots & \ddots & \vdots \\ (\varphi_N, \psi_1) & (\varphi_N, \psi_2) & \cdots & (\varphi_N, \psi_N) \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix}$$

$$= [\lambda_1, \ \lambda_2, \cdots, \ \lambda_N] \begin{bmatrix} (\varphi_1, \psi_1) & 0 & \cdots & 0 \\ 0 & (\varphi_2, \psi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\varphi_N, \psi_N) \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix}$$

两数据间共性信息就体现在矩阵方程的元素中。即各个基函数之间的内积关系。

基函数可以由数据自相关矩阵分解出来

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆3、向量及函数的自相关与互相关。

### ◆观测两个向量之间互相关的不同角度。

◆两个向量之间的互相关测量的是角度关系。归一化相关系数 $\rho$ 即两个矢量的点积 $\cos\varphi$。

$$\rho = \cos\varphi = \frac{\boldsymbol{a}^T \cdot \boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2 |\boldsymbol{b}|^2}}$$

◆这是代数转化为几何的典型。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。地震数据变化慢（幅值有正有负），用 $\rho$ 测相似性不敏感，用 $\varphi$ 可能更合适。

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆3、向量及函数的自相关与互相关。

### ◆Sergey Fomel对互相关的灵活应用：

$$\rho = \frac{\boldsymbol{a}^T \cdot \boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2 |\boldsymbol{b}|^2}} \Rightarrow \rho^2 = \frac{\left(\boldsymbol{a}^T \cdot \boldsymbol{b}\right)^T \left(\boldsymbol{a}^T \cdot \boldsymbol{b}\right)}{\left(\boldsymbol{a}^T \cdot \boldsymbol{a}\right)\left(\boldsymbol{b}^T \cdot \boldsymbol{b}\right)} = \frac{\left(\boldsymbol{b}^T \cdot \boldsymbol{a}\right)\left(\boldsymbol{a}^T \cdot \boldsymbol{b}\right)}{\left(\boldsymbol{b}^T \cdot \boldsymbol{b}\right)\left(\boldsymbol{a}^T \cdot \boldsymbol{a}\right)} = \rho_1 \rho_2$$

红色的 $\rho$ 好于蓝色的 $\rho$。

$$\rho_1 = \underset{\rho_1}{\operatorname{argmin}} \left\| \boldsymbol{b} - \rho_1 \boldsymbol{a} \right\| = \frac{\boldsymbol{a}^T \boldsymbol{b}}{\boldsymbol{a}^T \boldsymbol{a}}$$

$$\rho_2 = \underset{\rho_2}{\operatorname{argmin}} \left\| \boldsymbol{a} - \rho_1 \boldsymbol{b} \right\| = \frac{\boldsymbol{b}^T \boldsymbol{a}}{\boldsymbol{b}^T \boldsymbol{b}}$$

$$\Rightarrow \rho = \sqrt{\rho_1 \rho_2}$$

# ◆五、几何在数据分析和图像分析中的应用分析

## ◆3、向量及函数的自相关与互相关。

### ◆Sergey Fomel对互相关的灵活应用：

$$\rho = \frac{a^T \cdot b}{\sqrt{|a|^2 |b|^2}} \Rightarrow \rho^2 = \rho_1 \rho_2$$

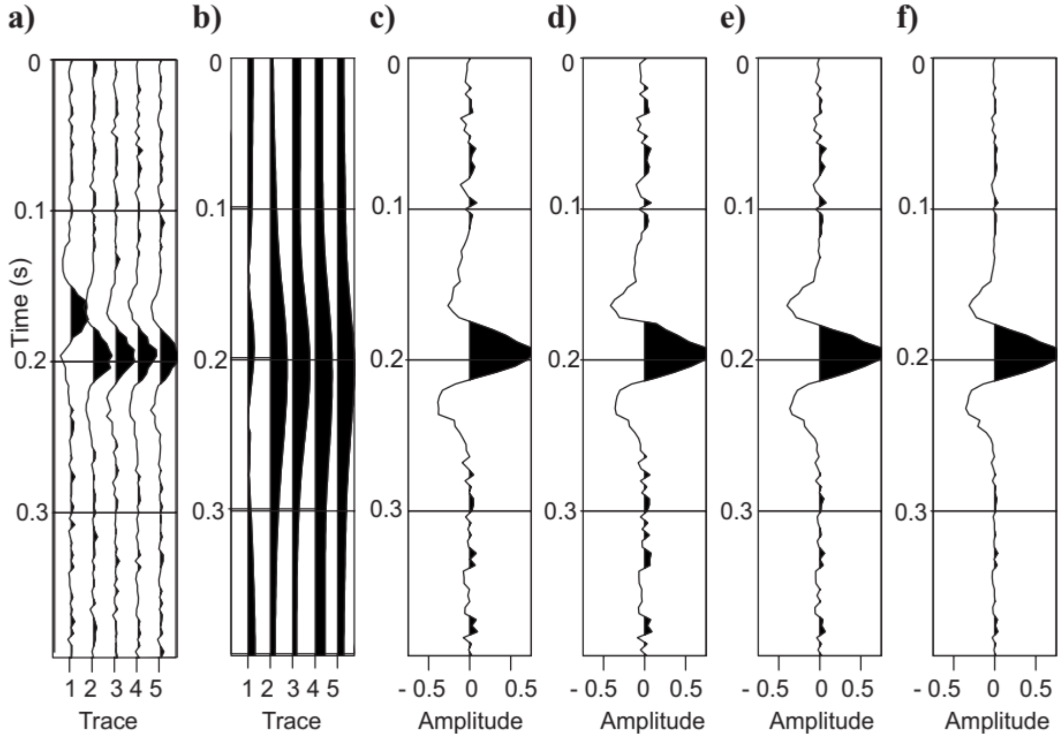



Figure 2. Simple stacking test with fivefold gather. (a) Prestack gather. (b) Weights used in local-correlation weighted stacking. (c) Conventional equal-weight stacking method (S/N = 8.4). (d) Smart stacking method (S/N = 9.2). (e) LMO-based weighted method (S/N = 10.2). (f) Local-correlation weighted stacking (S/N = 13.5).

新思想用于老问题，可以发表很多文章，关键是要有独立的、新颖的思想。

## ◆五、几何在数据分析和图像分析中的应用分析

### ◆4、函数的凸、极值及最优化。

◆高维函数（泛函）是凸的，决定了存在一个极值点，形成最优化的根本出发点。

$$f\left(\lambda x_1+\left(1-\lambda\right)x_2\right)\leq \lambda f\left(x_1\right)+\left(1-\lambda\right)f\left(x_2\right)$$

◆二次型泛函是凸的。二次型泛函对应的正问题是线性结构描述的。

◆在局部点上，对高维函数（泛函）进行Taylor展开，

$$f\left(x\right)=f\left(x_0\right)+\left.\frac{\partial f}{\partial x}\right|_{x_0}\left(x-x_0\right)+\frac{1}{2}\left(x-x_0\right)^T\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0}\left(x-x_0\right)+\cdots$$

线性结构与局部Taylor展开是对应的。

◆若 $Det\left|\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0}\right|>0$，高维函数（泛函）在 $x_0$ 附近存在极值点。

◆这样的观点，可以解决一大类问题。Newton迭代法的根本就奠基于此！

# ◆五、几何在数据分析和图像分析中的应用分析

◆5、评价一下神经网络算法。

- ◆（深度）神经网络算法缺乏几何思维，很难理解其背后的算法机制。如果没有改变，很难继续深化下去。

# 目 录

# ◆六、总结与讨论

◆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（非）线性成像。成像目标也由面向构造的成像（宽带反射系数）转移到面向储层的成像（宽带波阻抗）。

◆所谓数据中蕴含的信息，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。尤其是解析几何和微分几何。

# ◆六、总结与讨论

◆我特别希望是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。

◆我特别强调一下微分几何，希望把微分几何充分地用在当前的地震数据分析中，做出与众不同的研究。而不是处处抄别人的作业。

◆我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。

# ◆六、总结与讨论

◆给中石油勘探院建议的前瞻性课题（地震波成像方面）：

◆1、自动与智能的复杂近地表速度建模技术

◆1.1 基于空间散乱多源信息融合的极浅层速度建模技术

◆1.2 强噪声弱初至情形下智能化的初至波识别与走时检测方法

◆1.3 复杂地形条件下的高效的透射波波动理论走时层析成像技术

◆1.4 复杂地形条件下的高效的透射波波动理论波形层析成像技术

◆2、多信息约束下的反射波波动理论高精度建模技术

◆2.1 用于建模的断层及层位智能识别技术

◆2.2 用于建模的特殊地震相智能识别技术

◆2.3 特征反射波波动理论走时层析成像技术

◆2.4 特征反射波波动理论波形层析成像技术

# ◆六、总结与讨论

## ◆给中石油勘探院建议的前瞻性课题（地震波成像方面）：

### ◆3、宽带波阻抗建模方法技术

- ◆ 3.1 井、震、地质、岩石物理知识的表达方法及融合方法理论
- ◆ 3.2 基于多信息融合散乱数据插值的背景密度建模方法
- ◆ 3.3 低波数、中波数、高波数地震波波动理论反演的方法理论
- ◆ 3.4 先验信息约束下的高波数地震波波动理论反演方法及技术
- ◆ 3.5 基于多信息融合的宽带波阻抗建模方法技术

### ◆4、特征反射层相关的多次波压制与成像方法技术

- ◆ 4.1 基于特征反射层引起的波现象的水体建模技术
- ◆ 4.2 基于复杂水体模型的水体相关多次波预测与压制技术
- ◆ 4.3 基于特征反射层的长程及层间多次波预测及压制技术
- ◆ 4.4 自由表面多次波与一次波联合反演成像技术
- ◆ 4.5 基于特征反射层的长程及层间多次波层析成像技术
- ◆ **注：特征反射层特指海面、强反射海底面和强反射地层。**

# 谢谢
# 欢迎批评指正