



波现象与智能反演成像研究组



# 信息 智能学习与Bayes估计

## FWI的迷思

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

2020年11月24号

波现象与智能反演成像研究组Seminar



# 提 纲

- ◆一、系统参数估计反问题的本质到底是什么？
- ◆二、构建系统参数估计反问题的正确逻辑是什么？
- ◆三、如何进行信息的综合利用？
- ◆四、Bayes估计理论的再认识
- ◆五、学习类算法的根本逻辑是什么？
- ◆六、总结与认识

# ◆一、系统参数估计反问题的本质到底是什么？

◆宏观的实际物理系统是确定的。

◆宏观的实际物理系统是非常复杂的。

◆实际物理系统的输出可以用半随机模型“信号（波现象） + （广义）高斯噪音”来描述。

◆系统参数估计反问题的本质到底是什么？我们的目标是估计实际物理系统还是估计代表真实物理系统的模型！

# ◆一、系统参数估计反问题的本质到底是什么？

◆估计实际物理系统显然是不可能的！

◆我们能做的只能是对宏观的实际物理系统提炼出物理模型、数学模型或数学物理模型，并用这个数学物理模型来代表要估计的实际物理系统。

◆勘探地震中是用控制方程中的弹性参数来代表实际物理系统！

◆我认为：这就是系统参数估计反问题的本质。

◆系统参数估计反问题的根本基础在于代表要估计的实际物理系统的数学物理模型是否是合适的！



# 提 纲

- ◆一、系统参数估计反问题的本质到底是什么？
- ◆二、构建系统参数估计反问题的正确逻辑是什么？
- ◆三、如何进行信息的综合利用？
- ◆四、Bayes估计理论的再认识
- ◆五、学习类算法的根本逻辑是什么？
- ◆六、总结与认识



## ◆二、构建系统参数估计反问题的正确逻辑是什么？

### ◆系统参数估计反问题似乎是一个很主观的问题！

◆实际物理系统的数学物理模型是按照人为构建的物理理论建立的，存在一系列的假设。

◆参数估计理论也是人为构建的，反问题的解往往是无穷多的！

◆为什么反问题能为我们提供有用的反演解，能一定程度上解决实际问题？



## ◆二、构建系统参数估计反问题的正确逻辑是什么？

◆我认为最根本的原因是：依据基本的物理和数学逻辑建立了**系统物性参数**与**实际物理系统状态变量**之间的函数（映射）关系。

◆这是一种因果关系，不是相关关系。

◆实测数据=实际物理系统状态变量+（广义）Gauss噪声。

$$d(x, y, z=0; t) = u(x, y, z=0; t) + \eta(x, y, z=0; t)$$

对实际物理系统状态变量的采样

实际物理系统状态变量

$$A(m(x))u(x, t) = f(x_s, t)$$

对实际物理系统状态变量的预测



## ◆二、构建系统参数估计反问题的正确逻辑是什么？

◆系统物性参数与实际物理系统状态变量之间的函数（映射）关系决定了反演解的有效性。也决定了反问题的非线性程度。

◆反问题解的存在性，（部分）实测数据落在正问题的值域中。

◆非线性性不是反问题的本质属性！某种程度上，这是人为自找的！反问题的非线性取决于所选择的系统物性参数、状态方程和要预测的状态变量的成分！这些都是人为选择的！





## ◆二、构建系统参数估计反问题的正确逻辑是什么？

◆Rockafellar (1993) put forward :“In fact, the great watershed in optimization is not between linearity and non-linearity, but convexity and non-convexity”。

◆ Rockafellar, R. T., 1993, Lagrange multipliers and optimality, SIAM Reviews, 35:183-283.

◆对于非线性优化问题，重点不是考虑线性或非线性，而是凸或非凸！

◆构建凸的反问题才是构建系统参数估计反问题的正确逻辑，也是更好地解决系统参数估计反问题的先决条件。然后才是算法层次上的问题。





# 提 纲

- ◆一、系统参数估计反问题的本质到底是什么？
- ◆二、构建系统参数估计反问题的正确逻辑是什么？
- ◆三、如何进行信息的综合利用？
- ◆四、Bayes估计理论的再认识
- ◆五、学习类算法的根本逻辑是什么？
- ◆六、总结与认识

# ◆三、如何进行信息的综合利用？

## ◆什么是信息？

### ◆狭义的信息：

◆ Rochester,1996: 信息是一个有组织的事实和数据集合。

◆ 数据转化为信息；信息转变为知识；知识转变为智慧（决策）。

◆ Hartley,1928; Ursul,1971: 信息是被消除的不确定性。

### ◆广义的信息：

◆ 我的理解：可被用于交流的、完整的知识。

### ◆三、如何进行信息的综合利用？

#### ◆为什么要从信息的角度认识反问题？

- ◆1、求解反问题的过程也是消除不确定的过程。
- ◆2、叠前数据中所包含的信息，不足以消除估计反演解的不确定性。
- ◆3、引入关于反演解的先验信息是必须的。

## ◆三、如何进行信息的综合利用？

### ◆地震波成像中，先验信息包括什么？

- ◆ 1、从统计的观点，先验信息包括  $m_{\text{prior}}$   $C_m$
- ◆ 2、从泛函分析的观点，先验信息由不同的范数来定义。
  - ◆ 1和2都是软约束！
- ◆ 3、待估计参数一定是蕴含结构的，尤其是弹性参数。结构信息是它的重要特征。现代图像分析的本质就是识别图像中包含的结构信息。引入这个核心特征可以大幅提升反演成像精度、降低方差。
- ◆ 4、从测井、地质、岩石物理等途径也可以获取待估计参数的先验信息。它们可视为空间随机的。对空间随机过程进行建模是利用它们的基础。主要研究工作包括：散乱数据插值、数据同化、数据融合、多元空间数据的统计等。利用空间散乱数据的建模结果来约束反演成像是提高反演成像精度的又一个主要方向。



### ◆三、如何进行信息的综合利用？

#### ◆如何理解先验信息在反演成像中的应用？

- ◆正问题：系统物性参数与实际物理系统状态变量之间的函数（映射）关系。
- ◆反问题解的存在性：观测数据（部分地）落在正问题映射算子的值域内。
- ◆正问题规定了反演解的可行集。但仅仅依赖数据逼近只能确定可能的反演解，这样的反演解方差很大。
- ◆引入先验信息可以认为是为了进一步缩小了反演解的可行集，降低解的方差。
- ◆也可以说，引入先验信息是求解强非线性反问题的必要条件！

### ◆三、如何进行信息的综合利用？

## ◆如何理解先验信息在反演成像中的应用？

- ◆当前地震波反演成像中，正问题映射算子的伪逆研究居于核心地位。
- ◆误差泛函的最优化求解，对于梯度类算法，核心依然是伪逆！

$$\hat{m} = m_{prior} + \left( G^t C_D^{-1} G + C_M^{-1} \right)^{-1} G^t C_D^{-1} \left( d^{obs} - G m_{prior} \right)$$

- ◆正确的伪逆中已经包含了数据和模型参数的先验约束。但是，由于缺乏对  $C_D$  和  $C_M$  的认识，当前的反演实践中基本上都没有用正确的伪逆。

### ◆三、如何进行信息的综合利用？

#### ◆数据残差中的结构和模型中残差的结构该如何理解？

◆数据误差： $e_D = Gm - d^{obs}$

◆参数误差： $e_M = m - m_{prior}$

◆L2范数下，数据误差和参数误差都蕴含假设满足Gauss分布。尤其期望它们是Gauss白噪声。



### ◆三、如何进行信息的综合利用？

◆线性系统、Gauss白噪假设下，系统参数估计结果是无偏和方差最小的。Gauss-Markov定理。

◆此时，自协方差/自相关矩阵： $E\{e_D^T e_D\} = \sigma_D^2 I$

◆说明数据误差向量各分量之间是不相关的，甚至是正交的。

◆实际上，数据误差  $e_D = Gm - d^{obs}$  很难满足Gauss白噪的要求。自协方差/自相关矩阵  $E\{e_D^T e_D\} = V$ 。这说明误差向量各分量之间是相关的。

◆此时，系统参数估计结果达不到无偏和方差最小这种最佳状态。如果是迭代算法，收敛速度降低！极端情况可能不收敛！

◆引入加权最小二乘： $\min \|e_D^T W_D e_D\|^2$  可以显著改善解的性质！

### ◆三、如何进行信息的综合利用？

#### ◆加权最小二乘的统计意义：

- ◆对于数据误差非高斯白噪声情形下的解估计，为了达到数据误差高斯白噪声的理论要求，需要对非高斯白噪声误差向量引入（正交）变换矩阵，解开各分量之间的相关性。**K-L变换就是这样做的！**

◆即： $\varepsilon_D = P e_D (= P G m - P d^{obs}) \Rightarrow E\{\varepsilon_D^T \varepsilon_D\} = \sigma_D^2 I \quad E\{e_D^T e_D\} = V = P^T P$

$$\min \|e_D^T W e_D\|^2 = \min \|\varepsilon_D^T \varepsilon_D\|^2 = \min \left\{ (P G m - P d^{obs})^T (P G m - P d^{obs}) \right\}$$

$$= \min \left\{ (G m - d^{obs})^T P^T P (G m - d^{obs}) \right\}$$

$$= \min \left\{ (G m - d^{obs})^T W_D (G m - d^{obs}) \right\} = \min \|e_D^T W_D e_D\|^2$$

$$W = P^T P$$

### ◆三、如何进行信息的综合利用？

#### ◆加权最小二乘的统计意义：

- ◆把  $W_D$  选为非Gauss白噪数据误差向量自相关矩阵的逆，就可以对非Gauss白噪数据误差向量的各分量实现解相关，使得非Gauss白噪数据误差向量变成的Gauss白噪的，满足Gauss-Markov定理的要求。
- ◆这样做可以大幅提升解估计的精度！
  - ◆迭代法可以提升收敛效率。
- ◆等价的做法：数据预条件；摧毁数据误差向量中的结构信息；数据正则化。

### ◆三、如何进行信息的综合利用？

#### ◆引入参数先验信息的统计意义：

◆参数误差： $\mathbf{e}_M = \mathbf{m} - \mathbf{m}_{prior}$

◆自协方差/自相关矩阵： $E\{\mathbf{e}_M^T \mathbf{e}_M\} = \sigma_M^2 \mathbf{I}$ 。同样期望参数误差向量满足Gauss分布。尤其是期望满足Gauss白噪声分布。

◆这说明，估计结果与先验信息是最接近的。

◆解估计就是：与先验信息最接近条件下，基于解估计的正演数据最接近实测数据。

◆这是当前参数反演的本质含义！

### ◆三、如何进行信息的综合利用？

#### ◆引入参数先验信息的统计意义：

- ◆同样地，参数误差向量不会满足Gauss白噪的理论假设，需要引入合理的变换，解开参数误差向量各分量之间的相关性，使得：

$$\boldsymbol{\varepsilon}_M = \mathbf{P} \mathbf{e}_M \left( = \mathbf{P} \mathbf{m} - \mathbf{P} \mathbf{m}_{prior} \right)$$

$$\mathbf{E} \left\{ \boldsymbol{\varepsilon}_M^T \boldsymbol{\varepsilon}_M \right\} = \sigma_M^2 \mathbf{I}$$

$$\mathbf{E} \left\{ \mathbf{e}_M^T \mathbf{e}_M \right\} = \mathbf{E} \left\{ \left( \mathbf{m} - \mathbf{m}_{prior} \right)^T \left( \mathbf{m} - \mathbf{m}_{prior} \right) \right\} = \mathbf{V}_M = \mathbf{P}_M^T \mathbf{P}_M$$

$$\mathbf{E} \left\{ \mathbf{e}_M^T \mathbf{W}_M \mathbf{e}_M \right\} = \mathbf{E} \left\{ \left( \mathbf{m} - \mathbf{m}_{prior} \right)^T \mathbf{P}_M^T \mathbf{P}_M \left( \mathbf{m} - \mathbf{m}_{prior} \right) \right\} = \mathbf{E} \left\{ \boldsymbol{\varepsilon}_M^T \boldsymbol{\varepsilon}_M \right\}$$

### ◆三、如何进行信息的综合利用？

#### ◆引入参数先验信息的统计意义：

- ◆把  $W_M$  选为非Gauss白噪参数误差向量自相关矩阵的逆，就可以对非Gauss白噪参数误差向量的各分量实现解相关，使得非Gauss白噪参数误差向量变成的Gauss白噪的。满足对参数误差向量所做的高斯假设的要求。
- ◆这样做同样可以大幅提升解估计的精度！
- ◆等价的做法：参数预条件；摧毁参数误差向量中的结构信息（Laplace滤波）；各种各样的参数正则化。
- ◆多参数反演中，不同参数之间的耦合，也体现在参数误差向量中，还要考虑引入各参数之间的协方差矩阵的逆！

## ◆三、如何进行信息的综合利用？

### ◆如何理解先验信息在反演成像中的应用？

- ◆反问题研究发展到现在，对数据逼近的理解还有必要进一步加深。对先验信息的利用，更有必要深化。
- ◆先验信息的引入和利用，最根本地还是要从降低解的不确定性。最好从概率统计的角度来了解，并尽可能从统计量的角度来施加。
- ◆现有的一些变通的方法是可行的、也是有效的。它们是最好的吗？
- ◆缺乏理论指导的实践，往往是盲目的！



# 提 纲

- ◆一、系统参数估计反问题的本质到底是什么？
- ◆二、构建系统参数估计反问题的正确逻辑是什么？
- ◆三、如何进行信息的综合利用？
- ◆四、Bayes估计理论的再认识
- ◆五、学习类算法的根本逻辑是什么？
- ◆六、总结与认识



## ◆四、Bayes估计理论的再认识

◆所有的反演成像问题本质上都是在信息不充分情况下的决策问题（估计反演解的问题）！

◆这是反问题的本质含义！

◆解决这样一个决策问题，首先要制定出正确的、或者说合理的决策逻辑。

◆Bayes估计理论就提供了这样一个决策逻辑！



## ◆四、Bayes估计理论的再认识

### ◆Bayes估计理论奠定了解决非线性反问题的逻辑基础吗？

#### ◆勘探地震反问题是一个强非线性问题！

◆这个定位是没有错的！

#### ◆Bayes估计理论告诉我们如何求解（强）非线性反问题吗？

#### ◆我认为基本没有！

#### ◆概念上讲，基于Bayes估计理论，假若得到了后验概率密度函数，反问题的解的统计特征已经都知道了！

◆但是，这样的理论并不能很好地指导我们求解（强）非线性的反问题！

## ◆四、Bayes估计理论的再认识

### ◆Bayes估计理论的适用性是什么？

◆从根本上讲， Bayes估计理论还是**要求或希望**后验概率密度函数是单峰的！

◆更进一步， 后验概率密度函数是满足广义Gauss分布的。

◆如果后验概率密度函数是多峰的， 反问题的解的意义在Bayes估计理论下也是不明确的。

◆因此， Bayes估计理论的适用范围与凸问题的求解是一致的！

## ◆四、Bayes估计理论的再认识

### ◆非线性反问题的本质是什么？

- ◆实际物理系统只有一个解（真解）。依据不充分的数据和先验信息，获得它的近似解是合理的，也是可能的。
  - ◆估计子是无偏的+方程最小的！
  - ◆这是合乎物理的、对解估计的期望！
- ◆Bayes估计理论的提出，目的并不是得到无偏的+方程最小的解。只是给出参数分布的后验条件概率密度函数。利用后验概率密度函数并不能有效地求解非线性反问题。

## ◆四、Bayes估计理论的再认识

### ◆非线性反问题的本质是什么？

- ◆实际物理系统只有一个解（真解）；反问题是基于人为构建的正问题的，而非线性反问题有很多的解；仅仅性态良好的凸问题才有唯一的解。
- ◆合理的逻辑结论是：反问题的非线性性是人为的！
- ◆因为我们构建了一个非线性的系统物性参数与实际物理系统状态变量之间的函数（映射）关系，对应的代价泛函是非凸的、多峰的，导致了反问题的非线性性！

## ◆四、Bayes估计理论的再认识

### ◆变分问题或最优化问题是非线性反问题的合理提法吗？

#### ◆参数反演问题对应的变分问题的提法：

$$\hat{\mathbf{m}} = \arg \min_{\mathbf{m}} \{S(\mathbf{m})\} = \arg \min_{\mathbf{m}} \{(\mathbf{G}(\mathbf{m}) - \mathbf{d}_{obs})^T \mathbf{C}_D^{-1} (\mathbf{G}(\mathbf{m}) - \mathbf{d}_{obs})\}$$

#### ◆对于非线性反问题，这样的提法也是不合理的。它本质上与非线性正问题是不相容的！仅仅在 $G(m)=d$ 为线性或近似线性时它才是有意义的。若 $G(m)=d$ 是强非线性的，上述提法没有意义！

#### ◆本质上，变分问题是求系统对应的稳态解。非线性正问题时，有非常多的局部稳态解。解如何评价？

#### ◆这清楚地说明了：最大似然估计理论也存在的缺陷，仅仅依赖数据逼近对非线性情形完全无法评价估计结果。也许最大似然估计根本就不适用于参数估计。

## ◆四、Bayes估计理论的再认识

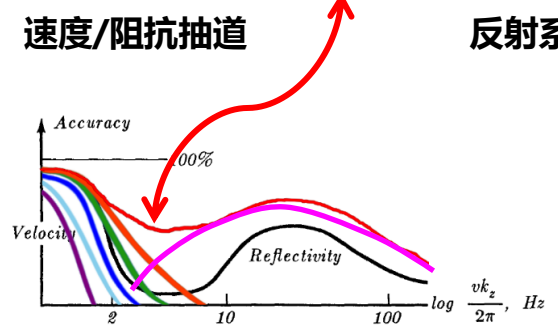
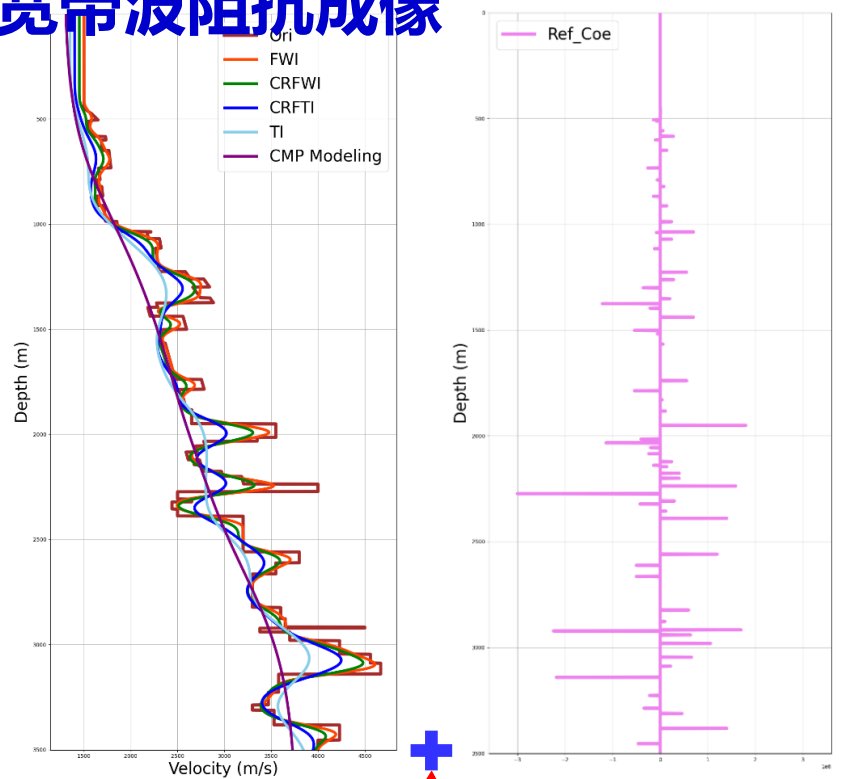
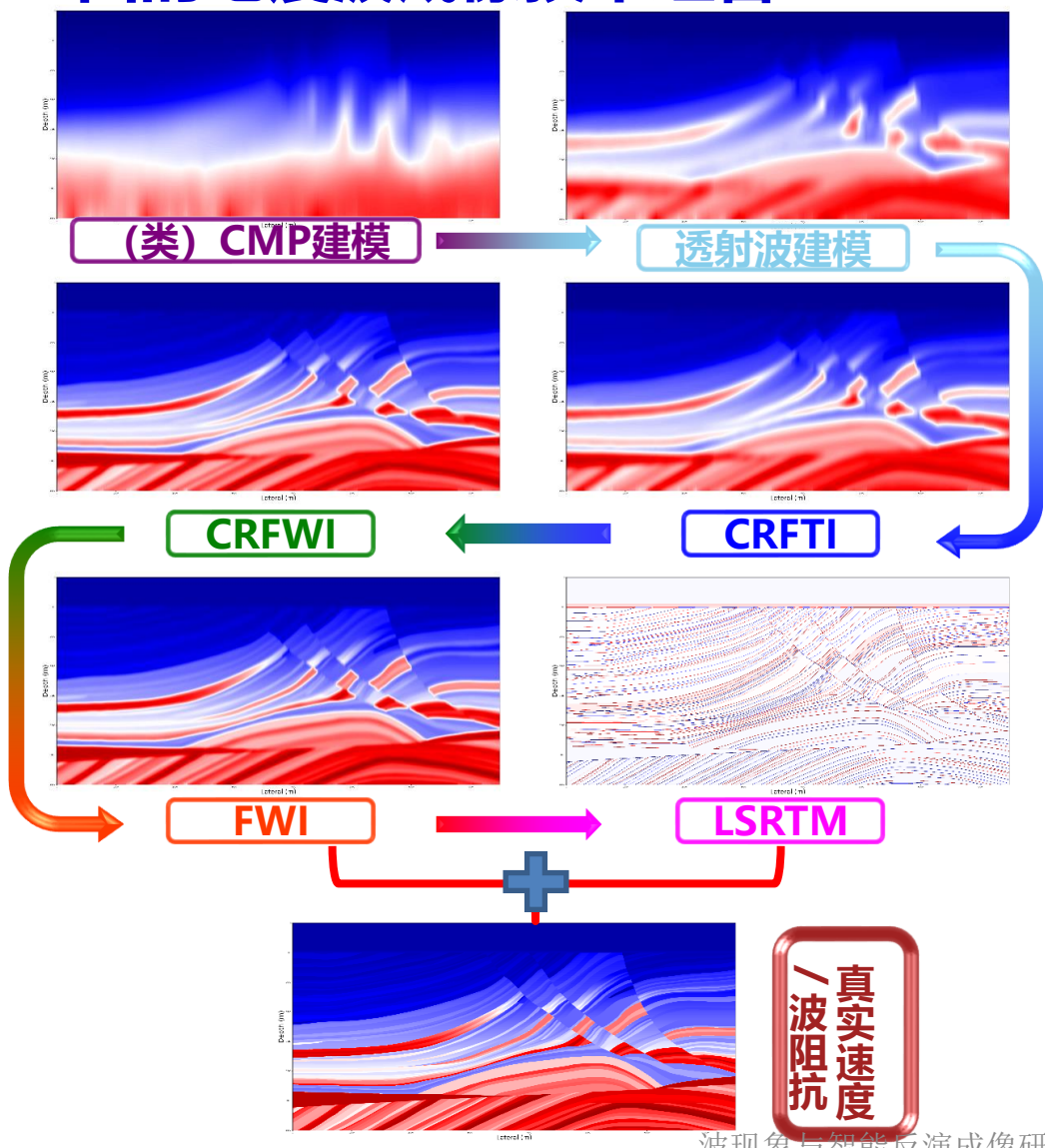
### ◆总结性观点：

- ◆ 当前，没有非线性反问题求解的严谨的数学理论。
- ◆ Bayes估计仅仅在条件后验密度函数是单峰的，才能给出反演解的明确意义。
- ◆ 变分问题或优化问题只有在代价函数为凸的情形下，解的意义才是明确的。
- ◆ **非线性反问题是人造的！** 构建一个强非线性反问题，然后努力去求解。这不是一个正确的路线。
- ◆ 要努力构建凸的反问题，把凸的反问题求解好。
- ◆ 不对正问题进行深入地分析，不细心选择正问题，而是构建一个强非线性反问题进行求解，导致反演结果不能收敛或很难收敛到期望的解。**FWI** 就是这样做的典型的例子！



# ◆四、Bayes估计理论的再认识

## ◆凸的地震波成像技术组合---CWI+宽带波阻抗成像



(Jon F. Claerbout, 1984)





# 提 纲

- ◆一、系统参数估计反问题的本质到底是什么？
- ◆二、构建系统参数估计反问题的正确逻辑是什么？
- ◆三、如何进行信息的综合利用？
- ◆四、Bayes估计理论的再认识
- ◆五、学习类算法的根本逻辑是什么？
- ◆六、总结与认识

## ◆五、学习类算法的根本逻辑是什么？

◆所有的（参数估计）反问题本质上都是在信息不充分情况下的决策问题（估计反演解的问题）！

### ◆学习类算法的本质：

◆ 伪逆的训练：  $\hat{B} = \arg \min_{B \in R^{n \times m}} E \left( m^{Index-data} - B d^{Training-data} \right)$

◆ 伪逆的推广应用：  $\hat{m} = \hat{B} d^{obs}$

◆ 当  $Km = d$  是线性系统时，  $\hat{B} = \left( K^T C_D^{-1} K + C_M^{-1} \right) K^T C_D^{-1}$  。

◆ 当  $d^{obs}$  是缓平稳变化时，伪逆的推广应用效果应该是可以的，精度是有保障的。

◆ 上述过程与Kalman滤波和自适应滤波是类似的，滤波器设计是估计最佳滤波器系数。

## ◆五、学习类算法的根本逻辑是什么？

### ◆学习类算法的本质：

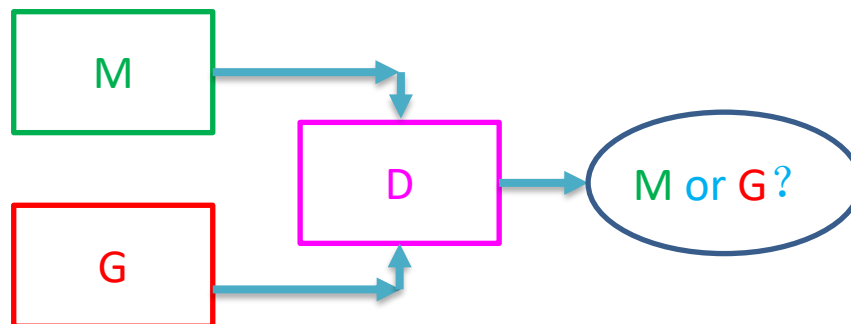
- ◆学习类算法与参数估计类算法的估计重点不一样。
- ◆有监督学习类算法基于已知标签和训练数据估计伪逆算子中的基函数。而且是不依赖于系统控制方程的。
  - ◆把系统控制方程约束下的估计作为初始估计应该是合适的。
- ◆参数估计算法基于实测数据和系统控制方程估计系统参数。

## ◆五、学习类算法的根本逻辑是什么？

### ◆学习类算法的本质：

#### ◆无监督学习算法GAN（生成对抗网络）的含义：

- ◆ 假设有一种概率分布M，它相对于我们是一个黑盒子。为了了解这个黑盒子中的东西是什么，我们构建了两个东西G和D，G是另一种我们完全知道的概率分布，D用来区分一个事件是由黑盒子中那个不知道的东西产生的还是由我们自己设的G产生的。
- ◆ 不断的调整G和D，直到D不能把事件区分出来为止。在调整过程中，需要：
  - ◆ 1、优化G，使它尽可能的让D混淆。
  - ◆ 2、优化D，使它尽可能的能区分出假冒的东西。
- ◆ 当D无法区分出事件的来源的时候，可以认为，G和M是一样的。从而，我们就了解到了黑盒子中的东西。
- ◆ 学习数据中的概率分布！



## ◆五、学习类算法的根本逻辑是什么？

### ◆学习类算法的本质：

- ◆ 我认为，学习类算法的本质就是**估计伪逆算子**。
- ◆ 伪逆算子的估计和参数估计从反问题求解上看，本质上是类似的。
- ◆ 伪逆算子的估计一般地是对应非线性正问题的。因为此时的正问题不是人为给定的，而是由设计的算法所隐含的。理论上，非线性正问题对应的“伪逆算子”目前没有数学理论来表达。
- ◆ 深度网络如何实现伪逆算子的估计就是一个黑匣子！缺乏严谨的数学理论的支持！
- ◆ 概率图模型方法也是数学上不明朗的！
- ◆ 正是这种数学上的不明朗，导致不愿意花功夫深入研究“**黑匣子类的**”**学习类算法**！

## ◆五、学习类算法的根本逻辑是什么？

### ◆学习类算法的本质：

- ◆事实上，非线性反问题情形下的参数估计，解的评价也是说不清楚的。但是，在凸问题情形下，参数估计解至少概念上是清楚的。可以借助勘探地震的物理来明确非线性反演时反演成像问题解的意义。
- ◆但是，一般意义下的数据分析，缺乏物理问题的辅助，伪逆算子估计又是一个黑匣子，泛化应用问题完全失去了判断依据！



# 提 纲

- ◆一、系统参数估计反问题的本质到底是什么？
- ◆二、构建系统参数估计反问题的正确逻辑是什么？
- ◆三、如何进行信息的综合利用？
- ◆四、Bayes估计理论的再认识
- ◆五、学习类算法的根本逻辑是什么？
- ◆六、总结与认识

## ◆六、总结与认识

- ◆ Bayes估计理论本身是比较完美的、Learning from data的框架性、指导性理论。
- ◆ 但是，基于Bayes估计理论的实践，还要依据各应用领域的特点提出具体的做法。尤其是凸问题的提法。不能是数学意义下的生搬硬套。
- ◆ FWI就是一种容易产生误导的勘探地震反问题提法，偏向于算法研究，而不是偏向于解决实际问题。
- ◆ 宽带波阻抗反演应该是更能导引出有用的实用化技术的地震波反演成像问题的提法。



A large, faint, light blue background logo of WPI is centered on the slide. It features a stylized mountain range with several peaks of varying heights, enclosed within a semi-circular border. Below the mountains, the letters 'WPI' are written in a large, bold, sans-serif font. The entire logo is rendered in a light blue color, making it a subtle background element.

***Thanks!***