反问题笔记

Inverse Problem Notes

讲学笔记

Research Note, at University of Houston
github.com/cainmagi/Research-Notebooks

First release, February 13, 2018

目 录

1. 普适离散反问题

反问题 *(inverse problem)*与**正问题** *(forward problem, modelization problem, simulation problem)*相对, 是基于已经存在的**观测结果**, 逆推出产生如此观测结果的**模型**。一个最大的不同是, 正问题里确定模型时, 观测结果有唯一解; 反问题里确定观测结果时, 模型没有唯一解。

故而, 为了得到有效的解、或**解集** *(collection of models)*, 往往须要对问题施加约束条件, 这种约束条件可以看成是对模型的一种先验信息, 它定义了解集可能存在的概率分布 (因而我们可以探索到最大概率的解)。

当数据有较高的采样频率时, 可以描述一个**离散反问题** *(discrete inverse problem)*来代替原问题。离散反问题从数学上更容易解, 但它不能完全等同于原问题, 因为离散化的过程中引入了一些简化, 造成一些信息因为离散化被隐藏了起来。

本章的核心是**参数集** *(parameter set)*上的 “**状态信息** *(state of information)*”, 一般被描述为前述的先验信息/概率密度。这种方法可以用在完全非线性问题里, 但计算开销甚大。且它的解一般少于传统方法得到的解。

1.1 模型空间与数据空间

研究一个物理系统 $\mathfrak{G}$ 一般分为三步:

(1) **系统参数化 *(parameterization of system)***: 确定一个最小的参数集, 使得通过这些参数可以完全描述物理系统;

(2) **正向模型 *(forward modeling)***: 预测并给定一些参数, 通过物理定律, 建立参数 → 观测结果之间的关系;

(3) **反向模型 *(inverse modeling)***: 使用这些预测的参数, 推测出实际模型参数的解。

可见, 前两步是**归纳** *(inductive)*, 最后一步是**演绎** *(deductive)*, 前两步的物理建模可能更困难; 至于最后一步, 有成熟、有效的数学方法来解。

首先让我们考量**系统参数化**的问题:

■ **例 1.1** 考虑一个弹性系统 $\mathfrak{G}$, 对于空间上任意的点 $\mathbf{x}$, 有压力 $\sigma^{ij}(\mathbf{x})$, 拉力 $\varepsilon^{ij}(\mathbf{x})$, 则可以定义弹性硬度张量 $c_{kl}^{ij}(\mathbf{x})$ 为:

$$\sigma^{ij}(\mathbf{x}) = c^{ij}_{kl}(\mathbf{x})\varepsilon^{kl}(\mathbf{x}). \tag{1.1}$$

同时还可以定义弹性柔度张量 $s^{ij}_{kl}(\mathbf{x})$ 为:

$$\varepsilon^{ij}(\mathbf{x}) = s^{ij}_{kl}(\mathbf{x})\sigma^{kl}(\mathbf{x}). \tag{1.2}$$

这里, 张量 $\mathbf{s}$ 是张量 $\mathbf{c}$ 的逆, 亦即 $c^{ij}_{kl}s^{kl}_{mn} = \delta^i_m\delta^j_n$。故而选择这两个参数中的任何一个都是完全等价 *(completely equivalent)*的。 ■

选择任意一组符合上述要求的特定参数集就是系统参数化, 对于两个等价的参数集, 其参数之间可以一一映射。

1.1.1 模型空间

与系统参数化无关, 我们可以引入一个抽象空间, 或称为流形 *(manifold)*, 用其中的每一个点代表一个模型 $\mathbf{m}$, 于是这样的流形记为模型空间 *(model space)* $\mathfrak{M} = \{\mathbf{m}_1, \mathbf{m}_2, \ldots\}$。

一个模型空间可以被不同的 (等价) 系统参数描述。例如, 设若 $\mathfrak{M}$ 的参数有两维, 记为 $\{m^1, m^2\}$, 则 m^1 和 m^2 可以选用不同的参数进行测量。由此定义的模型, 或记为点$\mathbf{m}$, 一般来说不可以看成是一个 (线性空间内的) 向量, 因为其坐标轴可能是非线性的。即使坐标轴都是线性的, 也一般不是笛卡尔坐标系。

■ **例 1.2** 考虑一个模型空间$\mathfrak{M}$ 已经确定了系统参数为 $\{\kappa, \mu\}$(分别为体积模量和剪切模量), 可以定义两个模型 $\mathbf{m}_1$, $\mathbf{m}_2$ 之间的距离为

$$d = \sqrt{\left(\log\frac{\kappa_2}{\kappa_1}\right)^2 + \left(\log\frac{\mu_2}{\mu_1}\right)^2}. \tag{1.3}$$

若以对数化的系统参数 $\{\kappa^* = \log\kappa, \mu^* = \log\mu\}$ 来代替(1.3)的参数, 有:

$$d = \sqrt{(\kappa^*_2 - \kappa^*_1)^2 + (\mu^*_2 - \mu^*_1)^2}. \tag{1.4}$$

可见, (1.4)是的参数是笛卡尔坐标系, 但(1.3)的不是。 ■

性质 1.1.1 — 离散性. 须知的是, 一个实际物理模型, 其参数系 *(set of quantities)*(亦即坐标轴)往往是连续的, 记为 $m(\mathbf{x})$; 但本章考虑的模型是经过抽样、离散化的模型, 其参数系为 $m^\alpha = m^\alpha(\mathbf{x})$。坐标轴的离散性满足以下条件:

(1) 参数系由选取的系统参数决定, 一个模型可以选用不同的参数系;
(2) 参数系维数 α 有限, 本章不考虑无限的情况;
(3) 参数系的抽样应该足够多, 能不失真地反映出原始模型的特征;
(4) 参数系的值本身可以取连续也可以取离散, 为了方便表达, 可以默认认为取值总是连续的, 因为离散的取值可以用狄拉克 *(Dirac)*函数表达。

■

性质 1.1.2 – 线性. 设任意两个同属一个模型空间 $\mathfrak{M}$, 使用同一系统参数 m^α 的模型 $\mathbf{m}_1, \mathbf{m}_2$, 其任意第 i 维的值均满足:

$$(\mathbf{m}_1 + \mathbf{m}_2)^\alpha = (\mathbf{m}_1)^\alpha + (\mathbf{m}_2)^\alpha, \tag{1.5}$$

$$(\lambda \mathbf{m})^\alpha = \lambda (\mathbf{m})^\alpha. \tag{1.6}$$

这样的的 $\mathfrak{M}$ 称为线性模型空间 *(linear model space)*, 记为 $\mathbb{M}$

例 1.2使用的对数化系统参数 $\{\kappa^*, \mu^*\}$ 是可以构成一个线性模型空间。因此, 此例中的距离 d 相当于该线性空间的*Euclidean* 范数。

为了处理一个反问题, 设模型空间 $\mathfrak{M}$ 为全集, 其任意一个子集记为 $\mathcal{A}$, 则可以令全集概率 $P(\mathfrak{M}) = 1$, 且其子集概率 $P(\mathcal{A}) \in [0, 1]$。此处定义的概率, 是通过一种特定的关联关系定义的, 这种关系不受到空间是否是线性的影响。并且, 能相应地根据该概率求出每个模型的概率密度分布。

1.1.2 数据空间

对于一个模型的观测结果, 由于观测仪器的不同也可能产生诸多不同的结果, 设全集 $\mathfrak{D}$ 为所有可观测仪器产生的观测结果的集合, 那么任何一组来自一个仪器的测量结果都可以看成它的元素 $\mathbf{d}$; 与 $\mathfrak{M}$ 相似, $\mathfrak{D}$ 同样也可以看成是一个流形, $\mathbf{d}$ 可以看成是流形中的一个点。

与性质 1.1.1和性质 1.1.2相似, 我们他讨论的数据空间也有类似的性质。离散性此处从略, 在此我们介绍线性:

性质 1.1.3 – 线性. 在一个线性数据空间 $\mathbb{D}$ 内, 任意两组测量结果 (来自两个仪器)$\mathbf{d}_1, \mathbf{d}_2$, 其任意第 i 维的值均满足:

$$(\mathbf{d}_1 + \mathbf{d}_2)^\alpha = (\mathbf{d}_1)^\alpha + (\mathbf{d}_2)^\alpha, \tag{1.7}$$

$$(\lambda \mathbf{d})^\alpha = \lambda (\mathbf{d})^\alpha. \tag{1.8}$$

1.1.3 合并流形

有时候, 一个物理系统的输入和输出参数可以清楚地分开, 这时使用两个流形 $\mathfrak{M}, \mathfrak{D}$ 来描述并无问题。但某些情况下 (例如有反馈系统), 不能做到这一点。因此有时我们也用一个合并的流形$\mathfrak{X}$ 来表示整个参数空间 *(parameter space)*。这种情况下, 其元素 (或点)$\mathbf{x}$ 被称为一组参数。

1.2 信息态

本节讨论了基于概率论定义的信息态 *(states of information)*。

1.2.1 概率的定义

考虑一个有限维流形 $\mathfrak{X}$, 正如在节 1.1.1末尾提到的那样, 可以从中定义任意的子集 $\mathcal{A}, \mathcal{B}, \ldots$, 我们称 $\mathfrak{X}, \mathcal{A}, \mathcal{B}, \ldots, \varnothing$ 所属的为一个域 F, 且 F 是一个σ 域 *(σField)*。

定义 1.2.1 — σ 域. 由集合 Ω 的众子集 (例如 $\mathcal{A},\mathcal{B},\ldots$) 构成的超集 *(super set)*, 在满足以下条件时称为一个域 *(Field)*。

(1) $\Omega \in F$;

(2) 若 $\mathcal{A} \in F$, 则 $\bar{\mathcal{A}} \in F$;

(3) 若 $\mathcal{A},\mathcal{B} \in F$, 则 $\mathcal{A}\cup\mathcal{B} \in F$。

特别地, 如果 F 内所有满足两两互斥的集合 $\mathcal{A}_i, i=1\to\infty$, 其交集 $\mathcal{A}$ 满足

$$\mathcal{A}=\bigcup_{i=1}^{\infty}\mathcal{A}_i,\ \ \mathcal{A}\in F, \tag{1.9}$$

则这样的域F 称为σ 域。

由于满足σ 域, 若对每个子集关联到一个概率算子 $P(\cdot)$, 则该算子满足性质 1.2.1。

性质 1.2.1 — 概率算子的基本性质.

(1) 互斥的集合 $\mathcal{A},\mathcal{B}$ 满足: $P(\mathcal{A}\cup\mathcal{B})=P(\mathcal{A})+P(\mathcal{B})$;

(2) 趋向空集 $\varnothing$ 的集合序列 $A_1\supseteq A_2\supseteq\ldots$ 对应地满足 $\lim\limits_{i\to\infty}P(A_i)=0$;

(3) 空集 $\varnothing$ 满足 $P(\varnothing)=0$;

(4) 任意的两集合并集的概率: $P(\mathcal{A}\cup\mathcal{B})=P(\mathcal{A})+P(\mathcal{B})-P(\mathcal{A}\cap\mathcal{B})$.

于是, 考量模型 $\mathbf{x}\in\mathfrak{X}$(同时也是一组参量), 则可以定义

定义 1.2.2 — 模型集的概率. 定义一个多维参量的实函数 $f(\mathbf{x})$, 则集合 $\mathcal{A}$ 的概率定义为:

$$P(\mathcal{A})=\int_{\mathbf{x}\in\mathcal{A}}f(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{1.10}$$

若 $\forall\,\mathbf{x},\ f(\mathbf{x})>0$ 且 $P(\mathfrak{X})=1$, 这样的实函数 f 称为概率密度函数, P 称为概率或对标准化测度 *(normalized measurement)*的概率。

若以上条件不满足, 这样的实函数称为相对概率密度函数 *(relative probability density function)*, P 称为相对概率或对非标准化测度 *(nonnormalized measurement)*的概率。

例如, 对定义域 $\{x>0\}$ 的单变量实区间定义 $f=1/x$, 显然该函数对全集的积分 P 为 ∞, 故其不能代表概率, 但可以用来比较两个区间 $\mathcal{A},\mathcal{B}$ 上对应的积分 P, 这就是其称为相对概率的原因。

由于模型 $\mathbf{x}$ 往往有实际的物理意义 (例如每个变量都代表一个物理量, 具有一个特定的物理量纲 *(physical dimensions)*), 由于 P 无量纲, 通常 f 需要是一个有量纲的函数, 并且其量纲可以抵消模型 $\mathbf{x}$ 的量纲。

若对模型 $\mathbf{x}$ 能进行一个坐标变换, 使得新坐标系下的模型 $\mathbf{x}^*=\mathbf{x}^*(\mathbf{x})$, 于是有

$$P(\mathcal{A})=\int_{\mathbf{x}\in\mathcal{A}}f(\mathbf{x})\mathrm{d}\mathbf{x}=\int_{\mathbf{x}^*\in\mathcal{A}}f^*(\mathbf{x}^*)\mathrm{d}\mathbf{x}^*=\int_{\mathbf{x}^*\in\mathcal{A}}f(\mathbf{x})\left|\frac{\partial\mathbf{x}}{\partial\mathbf{x}^*}\right|\mathrm{d}\mathbf{x}^*. \tag{1.11}$$

1.2.2 概率的解释

1.2.3 概率的性质