



无限维 Bayes 反演理论与算法

贾骏雄^{1*}, 孟德宇¹, 张远祥²

1. 西安交通大学数学与统计学院, 西安 710049;

2. 兰州大学数学与统计学院, 兰州 730000

E-mail: jjx323@xjtu.edu.cn, dymeng@mail.xjtu.edu.cn, zhangyuanxiang@lzu.edu.cn

收稿日期: 2024-04-05; 接受日期: 2024-08-05; 网络出版日期: 2025-04-21; * 通信作者

国家自然科学基金 (批准号: 12322116, 12271428, 12226004 和 12326606) 和国家重点研发计划 (批准号: 2023YFC3503400) 资助项目

摘要 反问题是一个重要的数学研究领域, 其在医学成像、地震勘探成像、图像处理和天气预报等众多工程技术领域有着广泛的应用. 基于反问题的不适定性, 人们引入正则化思想求解这类问题, 得到参数的一个近似估计. 随着计算能力的提升, 在医学成像和勘探成像等领域, 人们不再满足于获取待估参数的一个合理估计, 而是试图综合经验知识和观测数据的不确定性信息, 给出待估参数不确定性的完整刻画. 为了实现这一目标, 反问题被转化为 Bayes 统计推断问题, 进而发展出了 Bayes 反演理论与数值算法. 不同于经典的统计学研究, 反问题研究中待估参数与观测数据是由复杂的数学模型 (如偏微分方程) 联系起来的, 因而需要引入新的思路和新的数学理论. 本文聚焦于针对无限维反问题建立的无限维 Bayes 反演理论, 从先验测度构造、Bayes 适定性、有限元离散、统计抽样算法和统计大样本理论等方面梳理现有的研究工作, 旨在阐明无限维 Bayes 反演方法的基本研究思路、核心研究问题、已有结果和方法以及未来可能的研究方向.

关键词 反问题 无限维 Bayes 方法 离散不变算法 变分推断 后验收缩率估计

MSC (2020) 主题分类 65L09, 49N45, 62F15

1 绪论

医学成像、地震勘探成像和雷达成像等工程技术问题在医疗和国防等领域中有着重要的作用, 因而受到了研究者的广泛关注. 自 20 世纪 50 年代, 伴随着地球物理勘探等工程领域的需求, 反问题获得了重要的发展, 逐步成为应用数学领域的一个主要研究方向 (参见文献 [41]). 令 X 和 Y 是赋范线性空间, 反问题通常是指从观测数据 $y \in Y$ 推断模型参数 $u \in X$. 这里模型参数与观测数据之间的关系可如下表示:

$$y = \mathcal{G}(u), \quad (1.1)$$

英文引用格式: Jia J X, Meng D Y, Zhang Y X. Infinite-dimensional Bayesian inversion theory and algorithms (in Chinese). Sci Sin Math, 2025, 55: 1649–1688, doi: 10.1360/SSM-2024-0100

其中 $\mathcal{G}: X \rightarrow Y$ 称为参数到观测的映射 (或称为正演算子). 正演算子是描述参数与数据之间关系的数学模型, 例如, 在地震勘探成像中可以建模为声波和黏弹波动方程^[74], 在数值天气预报中建模为流体方程^[45], 在数据分析中建模为神经网络^[22, 137]. 在实际问题中, 观测总是含有噪声的 (模型噪声、测量噪声或数值误差), 其具体形式通常难以确定, 但其统计性质经常可以获得. 这里考虑如下的加性噪声模型:

$$y = \mathcal{G}(u) + \eta, \quad (1.2)$$

其中 η 表示观测噪声. 关于其他形式的噪声模型, 可进一步参见文献 [121].

在 20 世纪早期, 基于对物理问题的直觉认识, 研究者普遍认为, 描述物理问题的一个正确数学模型必须是适定的, 即满足数学家 Hadamard^[86] 在 1923 年给出的适定性概念, 其包含 3 个条件, 对于上述问题具体可表述如下:

- (1) (问题的解是存在的) 对于所有的 $y \in Y$, 至少存在一个 $u \in X$, 使得 $y = \mathcal{G}(u)$;
- (2) (问题的解是唯一的): 对于所有的 $y \in Y$, 至多存在一个 $u \in X$, 使得 $y = \mathcal{G}(u)$;
- (3) (问题的解连续依赖于数据) 参数 u 连续地依赖于测量数据 y .

然而, 地震勘探成像和医学成像等典型的反问题并不满足这 3 个条件: 当噪声不在 $\mathcal{G}$ 的值域中时, 解可能不存在; 当考虑有限测量数据或边界测量数据时, 解可能不唯一; 正演算子即使是可逆的, $\mathcal{G}^{-1}$ 大多也是无界算子. 关于这 3 个方面的简明阐述, 可参见文献 [79, 第 2 小节]. 正是因为这些原因, 不适定的反问题在很长的一段时间内没有引起学术界的充分重视. 苏联科学院 Tikhonov 院士奠基性地提出了正则化方法^[69, 204], 其基本思想是, 考虑不适定问题的一个带有紧约束的问题, 这是原问题的一个近似适定问题, 利用这个近似适定问题获得反问题的稳定近似解. 经过几十年的发展, 正则化方法形成了丰富的理论体系, 在一般的 Banach 空间上建立正则化方法是不适定问题研究的一个重要领域 (参见文献 [193]). 关于正则化方法的近期发展, 文献 [17] 也给出了丰富的阐述.

正则化方法通常致力于在可用数据的基础上给出待估参数的一个合理估计. 与这一思路不同, 在 1970 年的文献 [75] 中, Franklin 针对线性反问题将其转化为 Bayes 统计推断问题, 在可分 Hilbert 空间上详细推导了 Gauss 先验、Gauss 噪声情形下的 Bayes 后验. 限于当时计算能力的限制, Bayes 反演方法并没有得到学术界的重视. 不同于正则化方法, Bayes 反演方法将未知量、观测噪声都建模为随机变量, 随机性反映了观测者对其数值的不确定性. 从 Bayes 反演方法的角度来看, 反问题的解不是参数的单个近似估计, 而是当所有可用信息都被融入模型中时得到的后验概率分布, 这一概率分布蕴含了先验与观测不确定性, 是参数信息的完整描述. 基于后验概率分布, 借由统计决策理论, 我们可以给出合理的单个参数估计 (如后验均值和最大后验估计), 并利用后验分布评估单个估计的可靠性. 基于后验分布提供的信息, 我们可进一步引入统计分析方法, 在实际问题中辅助决策, 如医学成像中的伪影检测^[223].

自 1970 年 Franklin 的工作之后, 针对 Bayes 反演方法的发展主要集中在有限维 Bayes 推断方法. 围绕着地震勘探成像问题, Tarantola 在其专著 [202] 中扩展了有限维空间中的 Bayes 反演方法, 特别地, 该专著的第 5 章从无限维空间中最大后验估计 (数学理论方面并不完全严谨) 的角度阐述了 Bayes 方法与对偶方法. 在专著 [121] 中, Kaipio 和 Somersalo 完整阐述了有限维 Bayes 反演理论及其在 X 射线断层成像、电阻抗成像、光学断层成像等诸多领域中的应用. 在专著 [36] 中, Calvetti 和 Somersalo 详细地介绍了 Bayes 反演领域中的科学计算方法. 然而, 偏微分方程反问题一般都是无限维的^[41], 包括未知参数、微分方程系统本身和反演输入的数据. 为了应用有限维的 Bayes 反演方法, 上述文献中采取的一般做法是, 先将反问题离散, 得到离散的反问题; 然后利用有限维 Bayes 反演方

法求解离散的反问题, 称为“先离散、再 Bayes”. “先离散、再 Bayes”的思路取得了广泛的应用, 但也面临着一些问题, 例如, (1) 在 2004 年, Lassas 和 Samuli^[145] 的论证表明, 基于有限维空间的全变差先验导出的 Bayes 方法无法给出离散不变的后验均值估计, 即后验均值保持边缘的性质会随着离散维数的增加而消失, 这说明全变差先验不是无限维空间中良定的概率测度. (2) 有限维空间中构建的随机游走 Markov 链 Monte Carlo (Markov chain Monte Carlo, MCMC) 算法的抽样效率对于离散维数的增加并不一致, 随着离散维数的增加抽样效率会迅速下降 (参见文献 [48]), 这限制了 Bayes 反演方法在高维问题中的应用.

为了解决这些问题, Cotter 等^[45] 在 2009 年开创性地提出了无限维 Bayes 反演的适定性理论. 事实上, 1970 年 Franklin 的文献 [75] 中已经有了无限维 Bayes 反演适定性理论的雏形, 但其研究仅针对 Gauss 分布假设下的线性反问题, 且方法难以直接推广到非线性、非 Gauss 等更一般的情形. 随后, Stuart^[195] 详细阐述了无限维 Bayes 反演的适定性理论、新的抽样效率不随离散维数下降的 MCMC 算法, 以及这一新的理论框架如何应用在扩散系数反演、波动方程波速反演、逆时热传导、天气预报、地下介质渗流系数反演等问题中. 无限维 Bayes 反演方法采用了与有限维理论不一样的角度, 试图先在无限维空间构建 Bayes 适定性理论、后验抽样算法, 进而将离散推迟到可能的最后一步, 称为“先 Bayes、再离散”. 相较于“先离散、再 Bayes”的思路, “先 Bayes、再离散”在数学理论上更为复杂, 需要运用无限维可分 Banach 空间上的概率测度理论、随机分析理论等, 但其也带来了如下很多好处. (1) 构建了 Bayes 适定性, 即后验概率测度存在、唯一且连续依赖于观测数据. 无限维 Bayes 反演的适定性理论将偏微分方程的适定性理论与 Bayes 反演更紧密地联系了起来. (2) 无限维 Bayes 反演理论给出了无限维空间中的理想解 (无限维空间上的后验概率测度), 从而易于引入数值偏微分方程分析理论, 给出有限维逼近解的收敛速率估计. (3) 无限维 Bayes 反演理论使得我们可以更明确地建立起 Bayes 理论与无限维空间上的正则化方法^[69] 之间的联系. (4) 直接在无限维空间上构建 Bayes 反演理论可以更好地利用无限维空间上反问题 (如偏微分方程反问题) 的性质, 从而构造抽样效率不依赖于网格离散的高效算法. 在带有梯度、Hessian 信息的抽样算法构造中, 更容易引入“先优化、后离散”的优化方法. 事实上, 先分析无限维空间上的问题, 将离散推迟到可能的最后一步的研究思路, 广泛地出现在各个领域, 如波动方程的可控性研究^[225] 和机器学习研究^[66, 151].

经过十来年的快速发展, 无限维 Bayes 反演理论与算法已经逐步成为反问题研究中的一个重要前沿领域, 本文旨在从 Bayes 适定性、有限元离散逼近、后验点估计、统计抽样算法、近似抽样算法、统计理论等方面梳理现有的研究工作, 阐明无限维 Bayes 反演方法的基本研究思路、核心研究问题、已有结果和方法以及未来可能的研究方向, 为想要了解这一领域的学者从我们的研究视角提供一个较为系统的总结. 虽然我们尽了最大的努力, 但限于我们所知, 一些有意思的主题并没有被包含进来, 如最优实验设计^[10] 和替代模型研究^[136, 196, 214].

本文余下内容的安排如下. 第 2 节分 4 个小节梳理 Bayes 反演的适定性理论和数值逼近方法. 具体而言, 第 2.1 小节简单地介绍无限维空间上的 Gauss 概率测度, 并给出先验概率测度的一些最新进展. 第 2.2 小节从直观的角度给出无限维 Bayes 公式, 并简述 Bayes 反演的 (P, d) 适定性. 第 2.3 小节围绕有限元离散, 简述有限维离散逼近的基本思路. 第 2.4 小节对无限维 Bayes 理论的最大后验估计、统计决策理论进行简要介绍, 进而阐述 Bayes 方法与正则化方法的关系. 第 3 节分 4 个小节梳理 Bayes 反演的后验抽样算法. 第 3.1 小节回顾 Metropolis-Hastings 算法的一般框架, 并基于一般框架简述常见的 4 种离散不变抽样算法. 第 3.2 小节对集合 Kalman 滤波进行简明的介绍, 并给出最新的研究进展. 第 3.3 小节从 3 个典型的例子出发简单介绍源自机器学习研究的变分推断方法. 第 3.4 小节聚焦于 Darcy 流反演问题, 从数值实践的角度探讨离散不变性, 从而阐明在无限维空间构造算法的

必要性. 最后, 第 4 节围绕后验收缩率估计、Bernstein-von Mises 定理介绍无限维 Bayes 反演统计理论的近期发展.

2 Bayes 反演理论

本节先介绍几种常见的定义在可分 Banach 空间上的概率测度. 然后, 在对常见的概率测度有了基本的了解后, 对 Bayes 适定性理论框架进行简要介绍. 类似于偏微分方程的适定性理论, Bayes 反演的适定性理论是讨论 Bayes 后验计算、数值逼近理论、后验统计性质分析的数学理论基础. 接下来, 简要介绍离散逼近的一般理论分析框架. 本节最后简要回顾最大后验估计理论. 有关最大后验估计的讨论揭示了 Bayes 反演方法与正则化方法之间的联系.

2.1 先验概率测度

Bayes 反演方法研究中涉及的先验概率测度、噪声概率测度、后验概率测度一般而言都定义在可分 Banach 空间上, 因而可分 Banach 空间上的概率测度理论对 Bayes 反演方法的研究具有重要的意义. 无限维可分空间上概率测度理论已经有了很长的研究历史, 关于这方面的一般讨论, 感兴趣的读者可参见文献 [25, 27, 140, 181]. 下面以 Gauss 测度为例, 对其主要思想进行简要介绍.

令 $\mathcal{H}$ 表示可分的 Hilbert 空间, 其上的内积和范数分别记为 $\langle \cdot, \cdot \rangle$ 和 $\|\cdot\|$. 一般而言, 可以取 $\mathcal{H} = L^2(D; \mathbb{R})$, 即定义在 Lipschitz 边界的开区域 $D \subset \mathbb{R}^d$ 上的平方可积实值函数, 其中 d 表示空间的维数. 在可分 Hilbert 空间 $\mathcal{H}$ 上, 我们以如下方式定义随机函数:

$$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \xi_j \phi_j, \quad (2.1)$$

其中 $\{\phi_j\}_{j=1}^{\infty}$ 表示空间 $\mathcal{H}$ 上的一组标准正交基, $\xi = \{\xi_j\}_{j=1}^{\infty}$ 是定义在概率空间 $(\Omega = \mathbb{R}^{\infty}, \mathcal{B}(\Omega), \mathbb{P})$ 上的独立同分布的随机变量序列, $\gamma = \{\gamma_j\}_{j=1}^{\infty}$ 表示具有衰减性的无限序列, $u_0 \in \mathcal{H}$ 表示随机函数的均值. 一般而言, 称表达式 (2.1) 为 Karhunen-Loève 展开, 特征函数族 $\{\phi_j\}_{j=1}^{\infty}$ 为 Karhunen-Loève 基. 令 $\mathcal{H}^t$ 表示 Hilbert 尺度空间, 其上的范数可定义如下:

$$\|u\|_{\mathcal{H}^t} = \left(\sum_{j=1}^{\infty} j^{\frac{2t}{d}} |\langle u, \phi_j \rangle|^2 \right)^{1/2}.$$

定义如下空间:

$$L_{\mathbb{P}}^2(\Omega; \mathcal{H}^t) := \{v : D \times \Omega \rightarrow \mathbb{R} \mid \mathbb{E}(\|v\|_{\mathcal{H}^t})^2 < \infty\},$$

其中 $\mathbb{E}$ 表示求期望. 若假设 $\xi_1 \sim \mathcal{N}(0, 1)$, $\gamma_j \asymp j^{-\frac{s}{d}}$ 且 $t < s - \frac{d}{2}$, 则随机函数 u 属于 Hilbert 空间 $L_{\mathbb{P}}^2(\Omega; \mathcal{H}^t)$, 称为可分 Hilbert 空间上的 Gauss 随机元, 并称 u 服从定义在可分 Hilbert 空间 $\mathcal{H}$ 上的 Gauss 测度 $\mathcal{N}(u_0, \mathcal{C})$. 对于空间 $\mathcal{H}$ 上的概率测度 μ , 可以定义其方差算子如下:

$$\mathcal{C} = \int_{\mathcal{H}} u \otimes u \mu(du),$$

其中 $\otimes$ 表示 Hilbert 空间上的张量积^[182]. 对于 Gauss 测度 $\mathcal{N}(m_0, \mathcal{C})$, 通过简单的推导, 可得

$$\mathcal{C} = \mathbb{E}u \otimes u = \sum_{j=1}^{\infty} \gamma_j^2 \phi_j \otimes \phi_j. \quad (2.2)$$

通常这里的方差算子是一个正定、对称的迹算子, 且容易看出 $\{\gamma_j^2, \phi_j\}_{j=1}^\infty$ 是方差算子 $\mathcal{C}$ 的特征值 - 特征向量对.

选取 Gauss 测度作为先验测度, 随机函数通常属于 $\mathcal{H}^t$ 或 C^t ($0 < t < s - \frac{d}{2}$), 即随机函数较为光滑, 因而无法刻画不连续的函数, 如分片常值函数. 然而, 在图像重建等反问题^[187]中提出的能够保持边缘的全变差 (total variation, TV) 正则获得了广泛的应用 (参见文献 [17]). 基于 TV 正则化方法中取得的成功, 人们试图提出 TV 先验测度. 在有限维空间, TV 先验测度的定义容易给出, 但文献 [145] 证明了有限维有意义的 TV 先验测度无法直接推广到无限维空间, 特别地, 后验均值不具备保持边缘的性质且不具备离散不变性. 为了解决这一问题, Lasso 等^[144]构造了离散不变的 Besov 测度. 更进一步地, Dashti 等^[52]在无限维可分 Hilbert 空间上基于展开式 (2.1) 给出了 Besov 测度的定义, 简单来讲, 当 ξ_1 服从 q -指数分布且 γ 满足一定的衰减性时, 随机函数 u 即可以看作是从 Besov 测度采样得到的. 随后, 文献 [108] 进一步推广了 Besov 测度的定义, 利用变指标 Besov 函数空间理论, 给出了变指标 Besov 先验测度的定义, 使得 Besov 类先验测度可以在局部区域刻画函数的可积性和可导性.

针对 Besov 类型的先验测度, 文献 [31, 92, 133] 中所展示的数值结果表明 Besov 先验可以保持边缘, 但保持边缘的性质依赖于 Haar 小波基, 因而不连续点的位置会倾向于在 Haar 小波基的二分格点处. 针对这些不足, Markkanen 等^[161]提出了 Cauchy 差分先验; Yao 等^[216]提出了 TV-Gauss 先验测度; Hosseini 和 Nigam^[97]系统研究了尾部指数衰减的先验测度, 给出了 Bayes 适定性分析.

作为这一小节的结束, 我们有必要说明无限维空间上的 Gauss 随机元与 Gauss 随机场有着深刻的联系, 在先验测度的构造、后验统计理论分析的研究中, 这一联系具有重要的作用. 区域 $D \subset \mathbb{R}^d$ 上的可测映射 $u: D \times \Omega \rightarrow \mathbb{R}^n$ 称为一个随机场, 是指: 一方面, 对于任意固定的 $x \in D$, 映射 $u(x; \cdot)$ 是取值于 $\mathbb{R}^n$ 上的随机变量; 另一方面, 对于任意的 $\omega \in \Omega$, $u(\cdot; \omega): D \rightarrow \mathbb{R}^n$ 是一个向量场. 对于区域 D 上的任意的点集 $\{x_k\}_{k=1}^K$ (K 是任意的正整数), 若随机向量 $(u(x_1; \cdot), \dots, u(x_K; \cdot))$ 是 Gauss 随机向量, 则称 u 是一个 Gauss 随机场. 对于 Gauss 随机场, 可以分别定义均值函数和协方差函数如下:

$$u_0(x) = \mathbb{E}u(x), \quad c(x, y) = \mathbb{E}(u(x) - u_0(x))(u(y) - u_0(y)). \quad (2.3)$$

由协方差函数, 可以定义方差算子如下:

$$(\mathcal{C}\phi)(x) = \int_D c(x, y)\phi(y)dy. \quad (2.4)$$

专著 [81, 第二章] 从无限维统计模型研究的角度对 Gauss 过程 (Gauss 随机场) 进行了细致的讨论, 详细地介绍了 Gauss 过程的聚集性质. 在专著 [80] 的第二、三章以及附录中, 作者从非参 Bayes 统计相合性研究的角度对 Gauss 过程进行了细致的介绍, 并且简述了 Gauss 过程与无限维空间上的 Gauss 随机元之间的关系. 事实上, Gauss 过程与无限维空间上的 Gauss 随机元在很弱的条件下都是等价的, 参见如下定理 (参见专著 [80, 引理 I.7]):

定理 2.1 对于一个 Gauss 过程 $u = (u(x; \omega) : x \in [0, 1]^d)$, 若其样本轨道属于空间 $C^\beta([0, 1]^d)$, 则这个 Gauss 过程是空间 $C^\alpha([0, 1]^d)$ ($\alpha < \beta$) 上的 Gauss 随机元.

关于 Gauss 过程与 Gauss 随机元的介绍以及这两种不同观点在 Bayes 反演中的应用, 感兴趣的读者可以参见专著 [168]. 受到深度神经网络研究的启发, 深度 Gauss 过程获得了广泛关注. 最近, Dunlop 等^[61]从遍历性的角度对深度 Gauss 过程进行了细致探讨. Abraham 和 Deo^[2]将深度 Gauss 过程用于 Bayes 反演研究, 给出了后验测度相合性估计, 且在 Darcy 流反演渗透率问题上得到了较经典 Gauss 先验更快的收缩速率估计.

2.2 适定性理论

在本小节中, 我们并不打算从严谨的数学理论推导的角度阐述无限维 Bayes 适定性理论, 而是从有限维 Bayes 公式出发, 尽量从直观的角度说明无限维 Bayes 适定性理论的核心思想. 在经典书籍 [121] 中, Kaipio 和 Somersalo 对有限维 Bayes 反演理论进行了详细的讨论. 令 X 和 Y 表示可分 Banach 空间. 为了简洁, 考虑如下模型:

$$y = \mathcal{G}(u) + \eta, \quad (2.5)$$

其中 $y \in Y$ 表示观测到的数据, $u \in X$ 表示带估计的参数, $\eta \in Y$ 表示观测噪声.

在有限维空间的假设下, 令 N_u 和 N_y 为正整数, 取 $X = \mathbb{R}^{N_u}$, $Y = \mathbb{R}^{N_y}$, 可以得到如下的 Bayes 公式:

$$\rho^y(u) = \frac{1}{Z^y} \rho(y - \mathcal{G}(u)) \rho_0(u), \quad (2.6)$$

其中 $\rho_0(u)$ 表示先验概率密度, $\rho(y - \mathcal{G}(u))$ 表示似然函数, $\rho^y(u)$ 表示后验概率密度, Z^y 表示归一化常数, 具有如下表达式:

$$Z^y := \int_{\mathbb{R}^{N_u}} \rho(y - \mathcal{G}(u)) \rho_0(u) du.$$

这一 Bayes 公式成立的条件是 $Z^y > 0$, 详细的论证, 可参见文献 [121, 定理 3.1].

将 Bayes 公式 (2.6) 推广到无限维空间会遇到的问题是, 无限维空间上不存在通常理解的密度函数 (概率测度关于 Lebesgue 测度的 Radon-Nikodym 导数^[42, 60]). 具体而言, 我们有如下周知的定理 2.2 (参见文献 [11, 定理 10.41]).

定理 2.2 令 $(X, \|\cdot\|)$ 是无限维可分 Banach 空间, μ 是其上局部有限的平移不变测度, 则 $\mu = 0$.

如何扩展有限维的 Bayes 定理到无限维空间呢? 一个最直接的思路是, 将与无限维随机元相关的密度函数改写为概率测度. 具体而言, 若 X 和 Y 是无限维可分 Banach 空间, 令 μ_0 和 μ^y 分别表示先验概率测度和后验概率测度, 取 $A \subset \mathcal{B}(X)$, 其中 $\mathcal{B}(X)$ 表示空间 X 上的所有 Borel 可测集构成的集合, 则直观来讲可将 (2.6) 进行如下改写:

$$\int_A \mu^y(du) = \frac{1}{Z^y} \int_A \rho(\mathcal{G}(u); y) \mu_0(du), \quad (2.7)$$

其中

$$Z^y := \int_X \rho(\mathcal{G}(u); y) \mu_0(du).$$

需要注意的是, 这里的积分区域是无限维可分 Banach 空间上的 Borel 可测集 (例如, 集合 A 可以由某个 Sobolev 空间中的函数构成), 关于这类积分的详细解释参见文献 [51, 181]. 一般而言, Bayes 公式应当与积分区域 A 无关, 因而可以进一步给出无限维可分 Banach 空间上的 Bayes 公式:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z^y} \exp(-\Phi(u; y)). \quad (2.8)$$

这里遵循文献 [45, 55, 195] 中的习惯, 将似然函数 $\rho(\mathcal{G}(u); y)$ 写成了上面的形式, 其中位势函数 $\Phi(u; y) = -\log \rho(\mathcal{G}(u); y)$ 通常称为负对数似然函数.

从有限维的 Bayes 公式直接改写过来的 Bayes 公式 (2.8) 是否有明确的数学定义? 回答是肯定的, 这一形式的 Bayes 公式在统计文献中早已出现, 且针对经典的统计学问题有着丰富的研究成果 (参见

文献 [81, 222]). 不同于经典的统计学研究, 在反问题的研究中, 位势函数 $\Phi(u; y)$ 里蕴含着复杂的非线性算子 (如 Darcy 流方程^[195]、Navier-Stokes 方程^[45] 和扩散方程^[110, 111]), 因而经典的统计学研究方法与理论成果难以直接应用于含无限维函数参数的反问题 (如偏微分方程反问题^[41]). 早在 1970 年, Franklin^[75] 在 Gauss 先验、Gauss 噪声的假设下, 针对线性反问题细致讨论了如何构建无限维 Bayes 反演理论. Tarantola 在专著 [201, 202] 中针对线性反问题在无限维空间的框架下探讨了 Bayes 反演理论 (非线性问题的探讨集中于最大后验估计). 直到 2009 年, Cotter 等^[45] 首次提出了适用于偏微分方程反问题的 Bayes 反演适定性理论 (适用于非线性问题), 并论证了二维不可压缩 Navier-Stokes 方程的数据同化问题的 Bayes 适定性. Stuart^[195] 进一步以逆扩散系数、波动方程波速反演、反向热传导、天气预报等问题为例, 完整地阐明了这一理论框架, 极大地推动了 Bayes 反演理论的发展. 随后, Dashti 和 Stuart^[55] 弱化了文献 [45, 195] 中的条件, 给出了更为一般的 Bayes 反演理论框架.

令 $\text{Prob}(X, \mu_0)$ 表示可分 Banach 空间 X 上关于测度 μ_0 绝对连续的概率测度构成的集合, $d_{\text{Hell}}(\cdot, \cdot)$ 表示测度的 Hellinger 距离, 则文献 [55] 中所阐述的 Bayes 适定性可如下定义.

定义 2.1 对于 Bayes 反演问题 (2.8), 称 Bayes 反演是 (Lipschitz, Hellinger) 适定的, 如果如下三点成立:

- (1) $\mu^y \in \text{Prob}(X, \mu_0)$ 存在 (后验测度存在);
- (2) μ^y 在 $\text{Prob}(X, \mu_0)$ 中是唯一确定的 (后验测度唯一);
- (3) $d_{\text{Hell}}(\mu^y, \mu^{\tilde{y}}) \leq C\|y - \tilde{y}\|_Y$ (后验测度关于观测数据局部 Lipschitz 连续).

上述 Bayes 反演的 (Lipschitz, Hellinger) 适定性获得了广泛关注. 众多研究者针对不同的偏微分方程反问题证明了其 Bayes 反演的 (Lipschitz, Hellinger) 适定性, 如 Darcy 流扩散系数反演问题^[54, 103]、基于水平集方法的反问题^[104]、多频逆介质散射问题^[112, 114]、非局部微分方程反演问题^[59, 221]、具有 Dirac 源项的 Helmholtz 方程反源问题^[68] 和刻画肿瘤生长的 Cahn-Hilliard 模型^[120]. 为了证明 Bayes 反演的 (Lipschitz, Hellinger) 适定性, 一般而言, 我们需要正演算子 (函数参数到观测数据的映射) 在合适的空间上是局部有界且连续的, 并且对观测噪声的概率分布也需要进行适当的假定. 这些条件有时难以验证 (如机器学习领域的分类问题和噪声为退化分布的反问题), 因而, Latz^[146, 147] 近期扩展了 Bayes 反演的 (Lipschitz, Hellinger) 适定性, 提出了 Bayes 反演的 (P, d) 适定性.

定义 2.2 令 P 表示可分 Banach 空间 X 上的概率测度构成的集合, (P, d) 表示 X 上的概率测度构成的以 d 为度量的度量空间, 则 Bayes 反问题称为 (P, d) 适定的, 是指 Bayes 反问题满足如下 3 条性质:

- (1) $\mu^y \in P$ 存在 (后验测度存在);
- (2) μ^y 在 P 中是唯一确定的 (后验测度唯一);
- (3) 映射 $y \rightarrow \mu^y$ 关于度量 d 是连续的 (后验测度的稳定性).

需要指出, 定义 2.1 与 2.2 的主要区别在于后验测度的稳定性. 在定义 2.1 中需要验证后验测度关于观测数据在 Hellinger 度量下是局部 Lipschitz 连续的, 这在 Hadmard 所提出的适定性定义里是不需要的. 定义 2.2 弱化了局部 Lipschitz 连续, 仅要求后验测度关于观测数据在合适的度量下是连续的. 在这个定义下, 我们给出如下的 4 个条件 (严谨论述参见文献 [147]):

- (1) 似然函数关于所有的数据 $y \in Y$ 和几乎所有的 u 是正的;
- (2) 对于给定的 $y \in Y$, 似然函数关于先验测度 μ_0 可积;
- (3) 对于任意的数据 $y \in Y$, 似然函数一致的被一个关于先验测度 μ_0 可积的函数控制;
- (4) 对于几乎所有的 u , 似然函数关于数据 y 连续.

基于这些条件, Latz^[146, 147] 验证了 $(\text{Prob}(X), d_{\text{Prok}})$, $(\text{Prob}(X), d_{\text{TV}})$ 和 $(\text{Prob}(X, \mu_0), d_{\text{Hell}})$ 适定性, 这

里 $\text{Prob}(X)$ 表示 X 上的概率测度构成的集合, d_{Prok} 表示 Prokhorov 度量, d_{TV} 表示概率测度的全变差度量. 这一扩展极大放宽了对于正演算子和噪声分布的假设, 拓宽了无限维 Bayes 反演理论的适用范围. 有关 Bayes 反演适定性最新的研究进展, 可以进一步参见文献 [29, 95, 124, 143, 148].

注 2.1 在常见的关于无限维 Bayes 反演的研究中 (参见文献 [55, 195]), 我们都会考虑可分的 Banach 空间或可分的 Hilbert 空间. 当需要在不可分的空间上考虑问题时, 我们会寻求在不可分空间的一个可分子空间上构建 Bayes 反演的相关理论 (参考文献 [55, 第2 和 3章]). 根据我们所知, 这其中的原因可能在于:

- 在不可分的无限维空间上 Borel σ - 代数与空间上由开球生成的 σ - 代数不一致, 然而在很多问题的分析中, 这两者的一致性会带来极大的方便. 感兴趣的读者可参见文献 [181, 第 1.1 小节], 其中许多基础的定理都依赖于两个 σ - 代数是一致的;
- 当考虑不可分的无限维数量空间时, 通常所定义的概率测度的支撑集的概率可能不是 1, 甚至可能概率是 0 (参考文献 [80, 第 A.4 小节]), 这违背直觉认知;
- 当无限维空间缺乏可分性时, 上文中 (2.7) 的意义会不太明确, 其难以通过 Bochner 积分来理解 (参见文献 [181, 第 1.1 和 1.2 小节]), 可能需要引入新的数学工具.

2.3 离散逼近

类比偏微分方程的适定性理论, Bayes 反演的适定性理论为数值求解奠定了理论基础. 基于适定性理论, 可以较容易地对离散后验分布的逼近速率进行分析, 得到离散的后验分布随着离散的加密收敛到无限维空间上后验测度的速率估计. 首先简要回顾 Bayes 反演的离散计算方法, 主要聚焦于基于有限元的数值逼近方法 [32, 79, 208]. 关于基于有限差分、谱方法等的离散逼近, 可参见文献 [4, 63].

令 $D \subset \mathbb{R}^d$ 是有界连通的开区域, 考虑 V_h 是 $L^2(D)$ 空间的基于连续 Lagrange 基函数 $\{\phi_j\}_{j=1}^\infty$ 的有限元离散空间. 记与基函数相关的节点为 $\{\mathbf{x}_j\}_{j=1}^n$, 从而有 $\phi_j(\mathbf{x}_i) = \delta_{ij}$ 其中 $i, j \in \{1, \dots, n\}$. 在这样的假设下, 我们需要统计推断的是函数参数 $u \in L^2(D)$ 的 n 维逼近 $u_h = \sum_{j=1}^n u_j \phi_j \in V_h$. 更具体地, 需要统计推断的是向量 $\mathbf{m} = (m_1, \dots, m_n)^T$. 在本小节以及后面的叙述中, 黑体字母都用来表示向量和矩阵.

实际离散计算中很重要的一点是如何计算函数空间 $L^2(D)$ 上的内积. 对于 $u_1, u_2 \in L^2(D)$, 记有限维逼近函数为 $u_{1h}, u_{2h} \in V_h$, 从而易知

$$(u_1, u_2)_{L^2(D)} \approx (u_{1h}, u_{2h}) \approx (\mathbf{u}_1, \mathbf{u}_2)_{\mathbf{M}} = \mathbf{u}_1^T \mathbf{M} \mathbf{u}_2, \quad (2.9)$$

其中 $\mathbf{u}_1$ 和 $\mathbf{u}_2$ 分别是有限维函数 u_{1h} 和 u_{2h} 对应的向量, 矩阵 $\mathbf{M} = (M_{ij})_{i,j=1,\dots,n}$ 定义如下:

$$M_{ij} = \int_D \phi_i(\mathbf{x}) \phi_j(\mathbf{x}) d\mathbf{x}.$$

从上面的简要叙述可知, $L^2(D)$ 的离散逼近空间是带权重 $\mathbf{M}$ 内积的 Euclid 空间 $\mathbb{R}_{\mathbf{M}}^n$, 而不是通常的 Euclid 空间 $\mathbb{R}^n$. 在 Bayes 反演方法中, 我们会遇到从空间 $L^2(D)$ 到空间 $L^2(D)$ 的算子 (无限维空间到无限维空间), 从 $L^2(D)$ 到 q 维空间 $\mathbb{R}^q$ 的算子 (无限维空间到有限维空间), 以及 r 维空间 $\mathbb{R}^r$ 到 $L^2(D)$ 的算子 (有限维空间到无限维空间), 以及这些映射的对偶算子. 如何在 $L^2(D)$ 的离散逼近空间 $\mathbb{R}_{\mathbf{M}}^n$ 上给出算子及其对偶算子的离散形式是离散计算的关键. 通过简单的推导 [32], 可以得知如下结论.

(1) 对于算子 $\mathcal{B}: L^2(D) \rightarrow L^2(D)$, 其离散近似算子 $\mathbf{B}: \mathbb{R}_M^n \rightarrow \mathbb{R}_M^n$ 具有如下形式:

$$\mathbf{B} = \mathbf{M}^{-1} \mathbf{K},$$

其中 $\mathbf{K} = (K_{ij})_{i,j=1,\dots,n}$ 且 $K_{ij} = \int_D \phi_i \mathcal{B} \phi_j d\mathbf{x}$. 进一步地, 算子 $\mathbf{B}$ 的对偶算子 $\mathbf{B}^* = \mathbf{M}^{-1} \mathbf{B}^T \mathbf{M}$.

(2) 对于算子 $\mathbf{F}: \mathbb{R}_M^n \rightarrow \mathbb{R}^q$, 其对偶算子 $\mathbf{F}^\natural: \mathbb{R}^q \rightarrow \mathbb{R}_M^n$ 可如下表示:

$$\mathbf{F}^\natural = \mathbf{M}^{-1} \mathbf{F}^T.$$

(3) 对于算子 $\mathbf{V}: \mathbb{R}^r \rightarrow \mathbb{R}_M^n$, 其对偶算子 $\mathbf{V}^\diamond: \mathbb{R}_M^n \rightarrow \mathbb{R}^r$ 可如下表示:

$$\mathbf{V}^\diamond = \mathbf{V}^T \mathbf{M}.$$

对于 Gauss 先验测度 $\mathcal{N}(u_0, C_0)$, 取 $C_0 = \mathcal{A}^{-2}$, 其中 $\mathcal{A}$ 是定义域为 $D(\mathcal{A}) = \{u \in H^2(D) : \text{在区域边界 } \partial D \text{ 上满足 } \alpha \nabla u \cdot \mathbf{n} = 0\}$ ($\mathbf{n}$ 表示外法线方向) 的微分算子. 具体而言, 对于 $u \in D(\mathcal{A})$, $f \in L^2(D)$, 有 $\mathcal{A}u = f$ 满足

$$\begin{cases} -\alpha \Delta u + \beta u = f, & x \in D, \\ \alpha \nabla u \cdot \mathbf{n} = 0, & x \in \partial D, \end{cases} \quad (2.10)$$

其中 $\alpha, \beta > 0$. 对于算子 $\mathcal{A}$, 可以得到其离散算子 $\mathbf{A} = \mathbf{M}^{-1} \mathbf{K}$, 其中

$$K_{ij} = \int_D (\alpha \nabla \phi_i(\mathbf{x}) \cdot \nabla \phi_j(\mathbf{x}) + \beta \phi_i(\mathbf{x}) \phi_j(\mathbf{x})) d\mathbf{x}.$$

从而无限维空间上的 Gauss 测度 $\mathcal{N}(u_0, C_0)$ 可以用具有如下密度函数的有限维 Gauss 分布进行逼近:

$$\pi(\mathbf{u}) \propto \exp \left(-\frac{1}{2} \|\mathbf{A}(\mathbf{u} - \mathbf{u}_0)\|_M^2 \right), \quad (2.11)$$

这里 $\propto$ 表示正比关系, 即 $a \propto b$ 表示存在常数 C 使得 $a = Cb$. 基于这样的离散方式, 从 Gauss 测度中进行采样的逼近计算公式^[32]如下:

$$\mathbf{u} = \mathbf{u}_0 + \mathbf{K}^{-1} \mathbf{M}^{1/2} \boldsymbol{\xi}, \quad (2.12)$$

其中 $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathbf{I})$ (这里的 $\mathbf{I}$ 表示通常的 Euclid 空间 $\mathbb{R}^n$ 上的恒等算子). 需要说明的是这里的矩阵 $\mathbf{M}$ 一般而言并不是对角矩阵, 因此如何计算 $\mathbf{M}^{1/2} \boldsymbol{\xi}$ 并不是平凡的 (一般而言, 矩阵 $\mathbf{M}^{1/2}$ 不再是稀疏矩阵, 因而一般仅计算 $\mathbf{M}^{1/2} \boldsymbol{\xi}$, 并不显示计算出 $\mathbf{M}^{1/2}$), 感兴趣的读者可参见文献 [40]. 基于这样的离散方式, 我们也很容易计算 (2.4) 中方差算子对应的 Gauss 核函数的逼近核函数 $c_h(\cdot, \cdot)$, 其计算公式如下:

$$c_h(\mathbf{x}, \mathbf{y}) = \boldsymbol{\Phi}(\mathbf{x})^T \mathbf{K}^{-1} \mathbf{M} \mathbf{K}^{-1} \boldsymbol{\Phi}(\mathbf{y}), \quad (2.13)$$

其中 $\boldsymbol{\Phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_n(\mathbf{x}))^T$.

基于所述的离散方法, 进一步可以得到正算子 $\mathcal{G}$ 的离散 $\mathcal{G}^n$ 逼近, 从而得到位势函数 $\Phi(u; y)$ 的离散形式 $\Phi^n(\mathbf{u}; \mathbf{y})$. 如果假设观测数据属于有限维空间, 且取噪声分布为 Gauss 分布 $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$, 则可以得到有限维离散近似 Bayes 公式如下:

$$\rho^{\mathbf{y}}(\mathbf{u}) \propto \exp \left(-\frac{1}{2} \|(\mathcal{G}^n(\mathbf{u}) - \mathbf{y})\|_{\mathbf{\Gamma}}^2 - \frac{1}{2} \|\mathbf{A}(\mathbf{u} - \mathbf{u}_0)\|_M^2 \right). \quad (2.14)$$

基于如上的离散方法, Bayes 统计反演方法被用于求解一系列大规模反问题, 如全球地层介质的反演 [32]、冰流问题中底部边界条件反演 [177]、多频散射数据逆源问题 [115]、地震勘探全波形反演问题 [224].

注 2.2 在如上离散逼近的构造中, 关于正演算子的离散依赖于正演算子的具体性质. 特别地, 在第 3 节中将对统计计算方法进行简要回顾, 在很多的统计计算方法中, 我们需要计算位势函数 $\Phi(u; y)$ 关于参数 u 的梯度和 Hessian 算子等. 如同绪论中所述, 类似于偏微分方程约束下的优化问题中“先离散、后优化”和“先优化、后离散”的讨论, 我们同样面临“先离散、后 Bayes”和“先 Bayes、后离散”两种分析计算方式.

本文聚焦于“先 Bayes、后离散”的无限维 Bayes 反演理论与计算方法. 这两种分析与计算方式各有优缺点, 难以从十分宽泛的角度进行阐述. 在偏微分方程约束优化问题中, 关于这两种计算方式的讨论, 我们推荐文献 [56, 79, 93]. 文献 [33, 46, 107] 从 Bayes 反演方面对这两种计算方式进行了讨论. 但总体而言, 关于这两种计算方式的讨论还很有限, 但如果涉及计算梯度、Hessian 算子等, 偏微分方程约束优化问题中细致的研究在 Bayes 反演研究中同样具有重要价值.

本小节的第一部分简要地介绍了基于有限元的离散, 下面对离散逼近速率估计进行简要的阐述 (参见文献 [46, 54]). 令 μ_n^y 和 $\Phi^n(u; y)$ 分别是对后验测度 μ^y 和位势函数 $\Phi(u; y)$ 在 X 的 n 维子空间上的逼近, 则有

$$\frac{d\mu_n^y}{d\mu_0}(u) = \frac{1}{Z_n^y} \exp(-\Phi^n(u; y)), \quad (2.15)$$

其中

$$Z_n^y = \int_X \exp(-\Phi^n(u; y)) d\mu_0(u),$$

这里 $\Phi^n(u; y)$ 被看作是由正演算子 $\mathcal{G}$ 的离散导出的位势函数的离散逼近. 文献 [46] 证明了如下关键定理.

定理 2.3 假设 Φ 和 Φ^n 满足合适的条件 (本文中不再赘述, 所需条件保证了 Bayes 反演原问题和离散逼近问题的适定性). 对于任意的 $\epsilon > 0$, 存在常数 $K = K(\epsilon) > 0$ 使得

$$|\Phi(u; y) - \Phi^n(u; y)| \leq K \exp(\epsilon \|u\|_X^2) \psi(n), \quad (2.16)$$

其中 $\lim_{n \rightarrow \infty} \psi(n) = 0$. 在这些条件下, 我们有

$$d_{\text{Hell}}(\mu^y, \mu_n^y) \leq C \psi(n), \quad (2.17)$$

其中 C 是与离散维数 n 无关的常数. 基于 Hellinger 距离的性质, 我们得到了后验均值和后验方差算子的离散逼近估计.

当假设观测值 $\mathbf{y} \in \mathbb{R}^{N_y}$, 即观测空间是有限维的时, 假设噪声服从 Gauss 分布 $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ 时, 位势函数和离散逼近位势函数具有如下形式:

$$\Phi(u; \mathbf{y}) = \frac{1}{2} \|\mathcal{G}(u) - \mathbf{y}\|_{\mathbf{\Gamma}}^2, \quad \Phi^n(u; \mathbf{y}) = \frac{1}{2} \|\mathcal{G}^n(u) - \mathbf{y}\|_{\mathbf{\Gamma}}^2. \quad (2.18)$$

如果假设

$$\|\mathcal{G}(u) - \mathcal{G}^n(u)\| \leq K(\epsilon) \exp(\epsilon \|u\|_X^2) \psi(n), \quad (2.19)$$

则可以得到

$$|\Phi(u; \mathbf{y}) - \Phi^n(u; \mathbf{y})| \leq K(2\epsilon) \exp(2\epsilon \|u\|_X^2) \psi(n). \quad (2.20)$$

从而由定理 2.3 可以得到后验逼近测度 $\mu_n^{\mathbf{y}}$ 逼近真实后验测度 $\mu^{\mathbf{y}}$ 的速率估计.

众所周知, 统计抽样算法 (如下文中提及的 MCMC 抽样算法) 在估计统计量 (如均值) 时误差通常是以 $\sqrt{N}$ (N 表示样本个数) 的速度衰减的. 定理 2.3 中的离散误差估计与统计抽样误差估计结合起来使得我们对计算误差有了更深入的理解, 从而在解决实际问题中更好地平衡离散精度和抽样精度 (计算资源有限). 上述离散逼近理论被广泛应用于众多反问题的分析求解中, 如基于 Navier-Stokes 方程的数据同化问题^[46]、Darcy 流方程渗透率反演问题^[54]、逆形状声场散射问题^[30]、双曲守恒律反演初值问题^[65]、工业水力旋流器的时变状态估计问题^[171] 和并行抽样算法的分析^[44].

2.4 最大后验估计

Bayes 方法将反问题转化为统计推断问题, 从而给出了反问题不确定性分析的完整理论分析框架. 为了更好地理解 Bayes 反演理论, 我们有必要准确理解 Bayes 反演方法与经典正则化方法之间的联系与区别. 在有限维 Bayes 反演方法的讨论中, 这一联系是显然的, 即 Bayes 反演的最大后验估计对应了某些正则优化问题, 经典的正则化方法计算得到的解可以认为是 Bayes 后验概率最大的点, 通常称为最大后验点估计^[121, 121]. 但在无限维可分 Banach 空间上, 这一联系并不是显然的, 甚至最大后验点估计该如何定义都不再是显然的. 下面以 Gauss 概率先验和 Besov 先验的研究为例阐述最大后验 (maximum a posteriori, MAP) 估计研究的核心结论.

令 X 表示无限维可分 Hilbert 空间, 考虑 X 上的 Gauss 先验概率测度 $\mu_0 = \mathcal{N}(0, C_0)$, E 表示 Gauss 测度 μ_0 的 Cameron-Martin 空间 (其上的范数定义为 $\|C_0^{-1/2} \cdot\|_X$). 定义如下泛函:

$$I(u) = \begin{cases} \Phi(u) + \frac{1}{2} \|u\|_E^2, & \text{如果 } u \in E, \\ +\infty, & \text{如果 } u \notin E. \end{cases} \quad (2.21)$$

令 $z \in E$, $B(z, \delta) \subset X$ 表示中心在 z 、半径为 δ 的开球.

定理 2.4 令位势函数满足如下条件:

- (1) 对于任意的 $\epsilon > 0$, 存在与 ϵ 有关的常数 $M \in \mathbb{R}$ 使得对于任意的 $u \in X$, 有 $\Phi(u; \mathbf{y}) \geq M - \epsilon \|u\|_X^2$;
- (2) 位势函数 $\Phi(u; \mathbf{y})$ 关于 u 是局部有上界的;
- (3) 位势函数 $\Phi(u; \mathbf{y})$ 关于 u 局部 Lipschitz 连续.

进一步假设 $\mu_0(X) = 1$, 对于任意的 $z_1, z_2 \in E$, 有

$$\lim_{\delta \rightarrow 0} \frac{\mu^{\mathbf{y}}(B(z_1, \delta))}{\mu^{\mathbf{y}}(B(z_2, \delta))} = \exp(I(z_2) - I(z_1)). \quad (2.22)$$

这个定理在文献 [53] 中被证明, 且这一结论被用于分析与 Navier-Stokes 方程相关的数据同化问题. 定理 2.4 中利用中心在 $z \in E$ 的小球处的后验概率的比值来定义 MAP 估计显然与有限维空间中 MAP 估计的定义相一致, 并且等式 (2.22) 说明求解 MAP 估计等价于求解泛函 $I(u)$ 的最小值.

注 2.3 定理 2.4 中小球的球心取值于先验 Gauss 测度 μ_0 的 Cameron-Martin 空间 E 值得我们特别注意, 这正反映了无限维情形与有限维情形的核心区别. 在无限维的情形下, Cameron-Martin 空

间 E 在空间 X 上是稠密的, 但同时 E 是一个零测集, 即 $\mu_0(E) = 0$. 在有限维空间 $\mathbb{R}^n$ 中, 全空间 $X = E = \mathbb{R}^n$, 因此 $\mu_0(E) = 1$. 这一核心区别导致有限维空间上的分析并不能仅令空间维数 $n \rightarrow \infty$ 就得到无限维空间的准确结果, 例如, 文献 [90] 分析了 MAP 估计, 但其证明仅对有限维情形成立.

在定理 2.4 中所定义的 MAP 估计的意义下, 根据所知难以得到带有 Besov 先验、分层先验情形下其与相应泛函极值问题对应关系的严格刻画. 基于测度的 Fomin 可微性理论, Helin 和 Burger^[91] 提出了弱最大后验 (weak MAP, wMAP) 这一新的概念, 定义如下.

定义 2.3 令 $B(z, \delta) \subset X$ 表示以 z 为心、 δ 为半径的球, 如果存在 $u \in \text{supp}(\mu^y)$ 且对于所有 $h \in E$ 有

$$\lim_{\delta \rightarrow 0} \frac{\mu^y(B(u-h, \delta))}{\mu^y(B(u, \delta))} \leq 1, \quad (2.23)$$

则称 u 是 wMAP 估计.

上述定义中的空间 E 不必是先验的 Cameron-Martin 空间 (对于非 Gauss 测度也没有这一概念), 而是满足如下要求的空间:

- (1) 空间 E 在空间 X 中是拓扑稠密的;
- (2) 记 $d_h \mu^y : \mathcal{B}(X) \rightarrow \mathbb{R}$ 是测度 μ^y 沿着 h 的 Fomin 导数^[26], 假设对于所有的 $h \in E$ 有

$$\frac{dd_h \mu^y}{d\mu^y} \in C(X),$$

即 $d_h \mu^y$ 关于 μ^y 的 Radon-Nikodym 导数是连续的.

基于定义 2.3, 文献 [5, 91] 得到了 Besov 先验下 wMAP 估计对应于如下泛函极值问题:

$$\arg \min_u \frac{1}{2} \|\mathcal{G}(u) - y\|_{\Gamma}^2 + \|u\|_{B_{p,p}^s}^p. \quad (2.24)$$

为了给读者一个整体的印象不至于陷入复杂的数学细节, 这里关于 Besov 先验下 wMAP 估计与泛函极值关系的说明并没有给出完整的条件. 关于 Bayes 理论与正则化理论的联系, 感兴趣的读者可以进一步参见文献 [58, 64, 108, 128, 138, 159, 211].

后验最大估计联系了 Bayes 反演方法与正则化方法, 对其研究具有重要的意义. 另外, 最大后验估计是后验测度的一种点估计, 特别地, 在 Bayes 统计中后验均值 (*posteriori* mean, CM) 估计

$$\hat{u} = \int_X u \mu^y(du) \quad (2.25)$$

是更加常用的一种点估计. 如果需要基于后验测度给出一个点估计 (Bayes 决策理论), 哪种点估计是更好的选择? 事实上, 这两种估计都有可能成为很糟糕的估计. 类似于文献 [121, 第三章] 中所展示的图 3.1, 本文在图 1 中给出了一维的一个简单的示意图. 从图 1(a) 中可以看到 CM 估计不是一个好的估计, 图 1(b) 说明 MAP 估计在图示的情形下不是一个好的估计.

在 Bayes 统计决策理论中, 人们定义了 Bayes 损失 (Bayes cost, BC) 用以刻画期望损失, 从而对 CM 和 MAP 估计提供更深刻的理解. 如果记 $C(\cdot, \cdot)$ 为损失函数, 则 Bayes 损失函数可定义如下:

$$\text{BC}(\hat{u}) = \int_X C(u, \hat{u}) \mu^y(du), \quad (2.26)$$

其中 $\hat{u}$ 是后验点估计. 显然希望寻找使得 Bayes 损失最小的点估计, 即

$$\hat{u}_C = \arg \min_{\hat{u}} \text{BC}(\hat{u}). \quad (2.27)$$

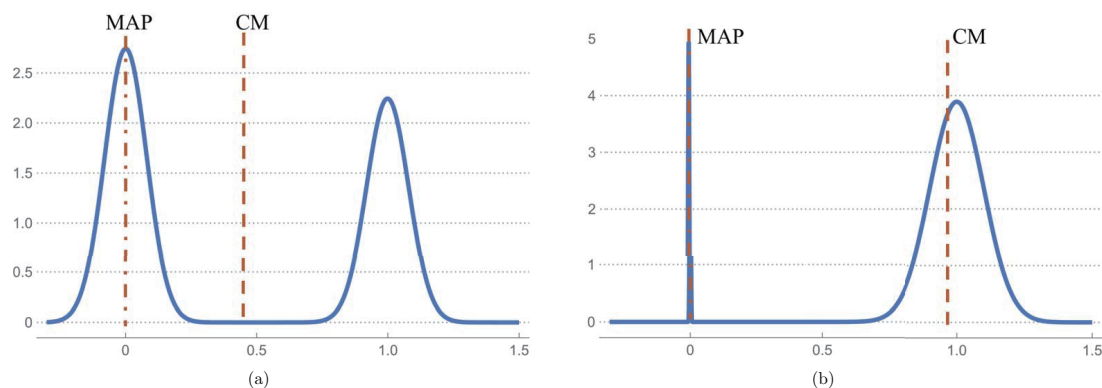


图 1 (网络版彩图) 两个一维密度函数说明最大后验估计与后验均值估计都有可能不是一个好的估计. (a) 后验分布有两个类似的峰组成, 后验均值估计落在了低概率区域; (b) 后验分布由一个很尖的峰与一个平缓的峰构成, 最大后验估计落在了概率较小的尖峰处

经典的统计理论 (限于篇幅, 下面给出的结论是有限维空间的结论, 即 u 是有限维向量) 告诉我们, 当 $C(u, \hat{u}) := \|u - \hat{u}\|_X^2$ 时, 问题 (2.27) 的解即为 CM 估计. 然而, 我们需要取损失函数为

$$C(u, \hat{u}) = \begin{cases} 0, & \|u - \hat{u}\|_\infty \leq \delta, \\ 1, & \|u - \hat{u}\|_\infty > \delta, \end{cases} \quad (2.28)$$

并考虑 $\delta \rightarrow 0$ 的极限情形时才能得到 MAP 估计. 可以看到, CM 估计中损失函数的选取更为自然, 因此在统计学的文献中人们对 CM 估计更为偏爱. 但文献 [34] 指出, 在高维 (有限维) 问题的最新研究中, CM 估计计算量大难以获得, 同时 MAP 估计计算量小且在很多情形下可以给出很好的估计结果. 为了更深刻地理解 MAP 估计, 该文献证明了基于 Bregman 距离所给出的损失函数, MAP 估计是使得 Bayes 损失最小的点估计, 从而在一定程度上说明了 MAP 估计是一个合适的 Bayes 估计. 文献 [91, 175] 就 CM 和 MAP 估计在无限维空间的情形下作了细致的讨论, 得到了类似的结论, 同样说明了 MAP 估计可以看作 Bregman 距离损失函数推导出的 Bayes 估计. 关于点估计的进一步讨论 (如带超参模型的点估计、点估计对应的极小化泛函的收敛性分析), 感兴趣的读者可以进一步参见文献 [14, 15, 62, 142].

3 后验统计计算

上一节回顾了无限维 Bayes 反演的适定性、逼近计算、后验点估计等理论, 本节聚焦于后验概率测度信息的计算. 第 3.1 小节简要介绍无限维空间上的 Metropolis-Hastings 方法, 给出常见的 MCMC 抽样算法. 第 3.2 小节对无限维 Bayes 反演的集合 Kalman 滤波算法进行介绍, 并回顾一些最新进展. 第 3.3 小节对变分推断方法进行简要介绍, 特别地, 对近期在无限维空间发展出的理论方法进行回顾. 最后, 第 3.4 小节对无限维空间所构造算法的离散不变性进行简单探讨, 从而试图说明为什么要引入无限维的方法.

3.1 统计抽样算法

根据第 2.2 小节的介绍, 我们了解到无限维 Bayes 公式具有如下形式:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z^y} \exp(-\Phi(u; y)). \quad (3.1)$$

下面基于文献 [48, 55, 195] 对 MCMC 算法的思想进行简要介绍. 令 $P(u, dv)$ 表示 Markov 转移核, 即对任意的 $u \in X$, $P(u, \cdot)$ 是空间 $(X, \mathcal{B}(X))$ 上的概率测度. 为了从后验概率测度中抽样, 从而基于大量样本计算统计量 (如后验均值和后验方差), 我们需要 Markov 转移核 $P(\cdot, \cdot)$ 关于后验测度是不变的, 即

$$\int_X \mu^y(du) P(u, \cdot) = \mu^y(\cdot). \quad (3.2)$$

容易看到, 如果 Markov 转移核 P 和后验概率测度 μ^y 满足细致平衡 (detailed balance) 条件

$$\mu^y(du) P(u, dv) = \mu^y(dv) P(v, du), \quad (3.3)$$

则 Markov 转移核关于后验测度是不变的. 基于 Markov 转移核 P 就能按照如下方式构造 Markov 链:

$$u_0 \rightarrow P(u_0, \cdot) \rightarrow u_1 \rightarrow P(u_1, \cdot) \rightarrow u_2 \rightarrow P(u_2, \cdot) \rightarrow u_3 \rightarrow \cdots.$$

由于转移核 P 关于后验测度是不变的, 所以能够实现从后验测度中抽取大量的样本. 可以看到构造 Markov 链的关键在于如何构造转移核 $P(\cdot, \cdot)$.

在算法 1 中, 我们展示了 Metropolis-Hastings 方法的一般计算框架, 其通过接受或拒绝 Markov 核 Q (基于随机偏微分方程构造, 下文详细介绍) 给出的建议函数参数, 构造我们期望的 Markov 转

算法 1 Metropolis-Hastings 方法的一般计算框架

对于给定的 $a: X \times X \rightarrow [0, 1]$, 按如下步骤生成样本 $\{u^{(k)}\}_{k \geq 0}$:

1. 设置 $k = 0$, 选取 $u^{(0)} \in X$;
 2. 生成新的样本 $v^{(k)} \sim Q(u^{(k)}, dv)$;
 3. 依概率 $a(u^{(k)}, v^{(k)})$, 取 $u^{(k+1)} = v^{(k)}$, 否则取 $u^{(k+1)} = u^{(k)}$;
 4. 令 $k \rightarrow k + 1$, 返回步骤 2.
-

移核 P . 从算法 1 接受与拒绝的计算过程中, 容易知道转移核 P 具有如下的形式:

$$P(u, dv) = Q(u, dv)a(u, v) + \delta_u(dv) \int_X (1 - a(u, w))Q(u, dw). \quad (3.4)$$

这一形式初看起来可能不容易理解, 可以考虑 $A \subset \mathcal{B}(X)$, 则直观上有

$$P(u, A) = \begin{cases} \int_A a(u, v)Q(u, dv) + \int_X (1 - a(u, v))Q(u, dv), & u \in A, \\ \int_A a(u, v)Q(u, dv), & u \notin A. \end{cases}$$

基于上面的公式, 可以较为容易理解公式 (3.4). 通过简单的推导, 可知细致平衡条件 (3.3) 等价于如下公式成立:

$$\mu^y(du)Q(u, dv)a(u, v) = \mu^y(dv)Q(v, du)a(v, u). \quad (3.5)$$

沿着文献 [48] 所述的思路, 我们在空间 $(X \times X, \mathcal{B}(X) \otimes \mathcal{B}(X))$ 上定义

$$\nu(du, dv) = \mu^y(du)Q(u, dv), \quad \nu^T(du, dv) = \mu^y(dv)Q(v, du),$$

其中 $\mathcal{B}(X) \otimes \mathcal{B}(X)$ 表示乘积空间上的乘积 σ 域. 文献 [55, 定理 21] 证明了如下定理.

定理 3.1 假设 ν 和 ν^T 是空间 $(X \times X, \mathcal{B}(X) \otimes \mathcal{B}(X))$ 上的等价测度, 并且有

$$\nu(du, dv) = r(u, v)\nu^T(du, dv).$$

对于在 (3.4) 中定义的 Markov 转移核 P , 其满足细致平衡条件 (3.3) 或 (3.5) 的等价条件是

$$r(u, v)a(u, v) = a(v, u), \quad \nu\text{-a.s.}$$

特别地, 如果选取

$$a(u, v) = \min\{1, r(v, u)\} = \min\left\{1, \frac{d\nu^T}{d\nu}(u, v)\right\},$$

则细致平衡条件自然满足.

可以看到, 构造 MCMC 算法最后的一个关键步骤是选取 Q . 在无限维空间理论中, 人们主要基于无限维 Langevin 方程构建 Q . 具体而言, 无限维 Langevin 方程具有如下形式:

$$\frac{du}{dt} = -\mathcal{K}(\mathcal{C}_0^{-1}u - \gamma D\Phi(u; y)) + \sqrt{2\mathcal{K}}\frac{dB}{dt}, \quad u(0) = u_0, \quad (3.6)$$

其中 $\mathcal{C}_0$ 是先验测度 $\mu_0 = \mathcal{N}(0, \mathcal{C}_0)$ 的方差算子, $\mathcal{K}$ 是预条件算子, B 是空间 X 上的方差算子为恒等算子的 Brownian 运动, $D\Phi(u; y)$ 表示函数参数 u 的 Fréchet 导数. 一般而言, $\mathcal{C}_0^{-1}$ 是微分算子, 从而方程 (3.6) 一般是一个偏微分方程. 基于 Crank-Nicolson 逼近格式, 可以得到

$$v = u - \frac{1}{2}\delta\mathcal{K}\mathcal{C}_0^{-1}(u + v) - \delta\gamma\mathcal{K}D\Phi(u; y) + \sqrt{2\delta\mathcal{K}}\xi_0, \quad (3.7)$$

其中 ξ_0 是 Gauss 白噪声. 下面所述的方法, 当 $\gamma = 1$ 时, 条件扩散 (conditioned diffusion) 问题在文献 [197] 中被提出. 随后, 文献 [21] 扩展到了 $\gamma = 0$ 的情形, 并且所发展的算法进一步被用于数据同化领域 [47] 中. 需要说明一下, 参数 $\gamma = 0$ 时所发展的算法, 在 1998 年的文献 [18] 中已经被提及 (推导的过程并不是基于如上的随机微分方程 (3.6), 关注的也不是反问题的无限维 Bayes 反演方法). 下面分 4 种情形分别加以介绍.

情形 1 $\gamma = 0, \mathcal{K} = \mathbf{I}$.

在这种情形下, 无限维 Langevin 方程具有如下形式:

$$\frac{du}{dt} = -\mathcal{C}_0^{-1}u + \sqrt{2}\frac{dB}{dt}, \quad u(0) = u_0. \quad (3.8)$$

上述方程的 Crank-Nicolson 逼近格式为

$$\left(I + \frac{1}{2}\delta\mathcal{C}_0^{-1}\right)v = \left(I - \frac{1}{2}\delta\mathcal{C}_0^{-1}\right)u + \sqrt{2\delta}\xi_0, \quad (3.9)$$

称为 Crank-Nicolson (CN) 建议. 若 $(I + \frac{1}{2}\delta\mathcal{C}_0^{-1})^{-1}$ 容易计算, 则可以通过 (3.9) 产生新的样本, 即算法 1 的步骤 2. 如果 $\mathcal{C}_0 = \mathcal{A}^{-2}$, 其中 $\mathcal{A}$ 由 (2.10) 定义, 则计算 $(I + \frac{1}{2}\delta\mathcal{C}_0^{-1})^{-1}$ 大致等价于计算两次偏微分方程 (2.10).

情形 2 $\gamma = 0, \mathcal{K} = \mathcal{C}_0$.

在这种情形下, 无限维 Langevin 方程具有如下形式:

$$\frac{du}{dt} = -u + \sqrt{2}\frac{dW}{dt}, \quad u(0) = u_0, \quad (3.10)$$

其中 W 是 $\mathcal{C}_0$ -Wiener 过程. 其 Crank-Nicolson 逼近格式为

$$(2 + \delta)v = (2 - \delta)u + \sqrt{8\delta}w, \quad (3.11)$$

其中 $w \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_0)$. 令 $\beta = \sqrt{8\delta/(2 + \delta)^2}$, 则 (3.11) 可写为

$$v = (1 - \beta^2)^{1/2}u + \beta w. \quad (3.12)$$

通常, 称 (3.11) 或 (3.12) 为预条件 CN (preconditioned CN, pCN) 建议. 基于第 2.3 小节中所介绍的离散方法, 可知 w 的离散逼近可以通过 (2.12) 快速计算. 对于情形 1 和 2, 借助于定理 3.1, 可以计算得到函数 a 如下:

$$a(u, v) = \min\{1, \exp(\Phi(u; y) - \Phi(v; y))\}. \quad (3.13)$$

关于函数 a 的详细推导过程, 可参见文献 [55, 第 5.2 小节]. 将 (3.9) 和 (3.13) 代入到算法 1 中, 得到 CN 算法. 将 (3.12) 和 (3.13) 代入到算法 1 中, 得到 pCN 算法.

注 3.1 在情形 1 和 2 中, 由于选取了 $\gamma = 0$, 在 Markov 核 Q 中不含有似然函数所提供的信息, 我们可以证明随机微分方程 (3.8) 和 (3.10) 是关于先验测度不变的, 进而可以得到 (参见文献 [55, 例 7 和 8])

$$\int_X \mu_0(du)Q(u, dv) = \mu_0(dv).$$

情形 3 $\gamma = 1, \mathcal{K} = \mathbf{I}$.

在这种情形下, 无限维 Langevin 方程形式如下:

$$\frac{du}{dt} = -\mathcal{C}_0^{-1}u - \gamma D\Phi(u; y) + \sqrt{2}\frac{dB}{dt}, \quad u(0) = u_0. \quad (3.14)$$

基于这一随机偏微分方程, 可以得到 Crank-Nicolson Langevin (CNL) 建议如下:

$$(2\mathcal{C}_0 + \delta)v = (2\mathcal{C}_0 - \delta)u - 2\delta\mathcal{C}_0 D\Phi(u; y) + \sqrt{8\delta\mathcal{C}_0}w, \quad (3.15)$$

其中 $w \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_0)$. 借助于定理 3.1, 可以计算得到函数 a 如下:

$$a(u, v) = \Phi(u; y) + \frac{1}{2}\langle v - u, D\Phi(u; y) \rangle + \frac{\delta}{4}\langle \mathcal{C}_0^{-1}(u + v), D\Phi(u; y) \rangle + \frac{\delta}{4}\|D\Phi(u; y)\|^2. \quad (3.16)$$

将 CNL 建议 (3.15) 与函数 (3.16) 代入到算法 1 中, 得到 CNL 算法. 关于 (3.15) 和 (3.16) 的详细推导过程, 可参见文献 [21, 第 4 节, 附录].

情形 4 $\gamma = 1, \mathcal{K} = \mathcal{C}_0$.

在这种情形下, 无限维 Langevin 方程形式如下:

$$\frac{du}{dt} = -u - \mathcal{C}_0\gamma D\Phi(u; y) + \sqrt{2}\frac{dW}{dt}, \quad u(0) = u_0, \quad (3.17)$$

其中 W 是 $\mathcal{C}_0$ -Wiener 过程. 进而, 可以得到预条件 Crank-Nicolson Langevin (preconditioned Crank-Nicolson Langevin, pCNL) 建议

$$(2 + \delta)v = (2 - \delta)u - 2\delta\mathcal{C}_0 D\Phi(u; y) + \sqrt{8\delta}w, \quad (3.18)$$

其中 $w \sim \mu_0 = \mathcal{N}(0, C_0)$. 进一步地, 借助于定理 3.1, 可以计算得到函数 a 如下:

$$a(u, v) = \Phi(u; y) + \frac{1}{2} \langle v - u, D\Phi(u; y) \rangle + \frac{\delta}{4} \langle u + v, D\Phi(u; y) \rangle + \frac{\delta}{4} \|C_0^{1/2} D\Phi(u; y)\|^2. \quad (3.19)$$

将 pCNL 建议 (3.18) 和函数 (3.19) 代入到算法 1 中, 得到 pCNL 算法. 关于 (3.18) 和 (3.19) 的详细推导过程, 可参见文献 [21, 第 4 节, 附录].

注 3.2 不同于情形 1 和 2, 在情形 3 和 4 中, 由于选取了 $\gamma = 1$, 在 Markov 核 Q 中含有似然函数所提供的信息, 所以可以证明随机微分方程 (3.14) 和 (3.17) 是关于后验测度 μ^y 不变的, 感兴趣的读者可参见文献 [55, 88, 89] 了解详细的假设与证明过程. 这里需要说明的是, 文献 [55, 88, 89] 对位势函数 $\Phi(u; y)$ 作了如下假设:

$$\Phi(u; y) \leq M_1(1 + \|u\|)^{N_1}, \quad \|D\Phi(u; y)\| \leq M_2(1 + \|u\|)^{N_2}, \quad (3.20)$$

其中 N_1 和 N_2 是两个正常数, M_1 和 M_2 是与 u 无关的正常数, $\|\cdot\|$ 表示合适的范数 (文献中并不统一, 这里并不想进行细致的严格数学论证, 仅作一些粗略讨论). 这些条件对于很多非线性反问题是不适用的 (不论用哪种范数), 如稳态 Darcy 流渗透率反演问题 (参见文献 [55] 中的第 1.3 小节和第 334 页例 3). 据我们所知, 文献 [55, 第 5.6.1 小节] 指出, 或许可以通过引入停时 (stopping time) 等证明方法弱化条件 (3.20) 使之可以应用于稳态 Darcy 流渗透率等反演问题, 但截至 2024 年 8 月仍然没有出现相关的研究.

注 3.3 关于如上所述的无限维空间 MCMC 类抽样算法, 理论研究分为两个方面: 缩放极限 (scaling limit) 和谱间隙 (spectral gap) 分析. 在缩放极限方面, 文献 [20, 162, 178] 中的分析表明, 标准的 MCMC 算法建议分布的方差需要根据离散维数的某个倍数减小, 为了准确估计统计量, 所需的抽样量会随着离散维数的增加而增大; 无限维空间 MCMC 类算法建议分布的方差无需根据离散维数进行调整, 因而为了准确估计统计量, 所需的抽样量不会随着离散维数的增加而增大. “先优化再离散”的研究中也有类似于缩放极限的讨论, 感兴趣的读者可参见文献 [107] 中的附录 (其中给出了一个简单的数值算例) 以及偏微分方程约束优化问题的专著 [56]. 在谱间隙分析方面, 文献 [87] 中的分析表明, 标准的 MCMC 算法的谱间隙会随着离散维数的增加而减小, pCN 算法的谱间隙不依赖于离散维数. 最近, 无限维 MCMC 谱间隙的分析理论也被扩展到了非 Gauss 先验测度的情形 (参见文献 [96]).

注 3.4 不同于反问题求解的经典正则化方法, Bayes 反演方法得到的解并不仅仅是函数参数的单个估计 (当然, 我们可以通过 Bayes 决策理论给出满足一定需求的点估计, 参见第 2.4 小节), 而是整个后验测度, 因而我们不能用后验均值等点估计的性质来判定抽样算法的收敛性、评判抽样算法的抽样效率高. 关于抽样算法的抽样是否充分、建议分布方差是否选取合理、如何判断不同抽样算法的效率等, 我们可以绘制样本的迹 (trace), 计算样本的自相关性和有效样本数等. 这里限于篇幅, 不再作细致的介绍, 可参见文献 [37].

关于无限维空间 MCMC 类的抽样算法, 还有很多值得探讨的内容 (如非 Gauss 先验下抽样算法的构造、Metropolis-within-Gibbs 算法和 Hamilton Monte Carlo 算法等), 感兴趣的读者可进一步参见文献 [7, 19, 20, 50, 63, 67, 94, 100, 174, 188, 199, 203, 210, 212].

3.2 集合 Kalman 滤波

本小节的主要目的是阐述利用集合 Kalman 滤波求解无穷维 Bayes 反问题的基本原理和算法. 为此, 先回顾基本的 Kalman 滤波及相关算法. 为了便于理解相关算法的基本思想和底层逻辑, 下面先从有限维情形着手论述, 相关算法的无穷维推导和应用 [24, 191, 192] 与有限维情形是平行的.

Kalman 滤波是由出生于匈牙利的数学家 Kalman 为了解决太空航行中的导航及控制问题而提出的, 可参见早期的文献 [122, 123]. Kalman 滤波的基本思想是利用系统的动态模型和观测数据, 通过递归的方式估计系统状态, 并使得估计的状态具有最小均方差, 从而有效地将噪声和不确定性考虑在内. 具体地, 考虑随机动力学模型

$$v_{k+1} = \psi(v_k) + \xi_k, \quad 0 \leq k \leq K-1, \quad v_0 \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(m_0, \mathcal{C}_0), \quad \xi_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma) \quad (3.21)$$

及数据观测模型

$$y_{k+1} = h(v_{k+1}) + \eta_{k+1}, \quad 0 \leq k \leq K-1, \quad \eta_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma), \quad (3.22)$$

其中 v_0 与序列 $\{\xi_k\}$ 和 $\{\eta_k\}$ 相互独立, 且对所有的 j 和 k , ξ_j 与 η_k 也相互独立. 在实际应用中, (3.21) 可以理解为控制状态演化的随机微分方程的离散化, 即使潜在信号是由确定性映射 ψ 所控制, 随机动力学模型 (3.21) 中的随机量 ξ_k 也可以认为是确定性映射的模型误差. 记

$$Y_k = \{y_1, y_2, \dots, y_k\},$$

则 (3.21) 和 (3.22) 的联立即为数据同化领域^[149, 189] 的通用数学问题, 其主要任务可以简述为给出或逼近滤波分布 (filtering distribution) $\mathbb{P}(v_k | Y_k)$ 或平滑分布 (smoothing distribution) $\mathbb{P}(v_k | Y_K)$, 这两种情形的区别在于状态 v_k 是基于前 k 个时刻的观察还是基于整个时间跨度的数据来进行推断的, 下文中主要针对前者进行阐述.

假设 (3.21) 和 (3.22) 均为线性模型, 即假设成立

$$\begin{cases} v_{k+1} = \Psi v_k + \xi_k, \\ y_{k+1} = H v_{k+1} + \eta_{k+1}, \end{cases} \quad (3.23)$$

其中 v_0 及序列 $\{\xi_k\}$ 及 $\{\eta_k\}$ 的假设如前所述. 由于 v_{k+1} 和 y_{k+1} 都是 Gauss 随机变量的仿射变换, 因此, $\mathbb{P}(v_k | Y_k)$ 也是 Gauss 分布, 可由其均值及协方差完全刻画. 关于 $\mathbb{P}(v_k | Y_k)$ 的均值及协方差的显示迭代更新公式即为所谓的 Kalman 滤波 (Kalman filter, Kf) 算法.

记

$$\begin{aligned} (\text{预测分布}) \quad \hat{\pi}_{k+1} &= \mathbb{P}(v_{k+1} | Y_k) = \mathcal{N}(\hat{m}_{k+1}, \hat{\mathcal{C}}_{k+1}), \\ (\text{分析分布}) \quad \pi_{k+1} &= \mathbb{P}(v_{k+1} | Y_{k+1}) = \mathcal{N}(m_{k+1}, \mathcal{C}_{k+1}). \end{aligned}$$

顾名思义, 预测分布 $\hat{\pi}_{k+1}$ 是结合状态演化方程和已有数据 Y_k 对状态 v_k 的变化趋势进行预测, 而分析分布 π_{k+1} 则是在测得新数据 y_{k+1} 的情形下利用数据 Y_{k+1} 对预测结果进行校正. 下面的定理给出了 Kf 算法的具体刻画.

定理 3.2 (参见文献 [189, 定理 8.3]) 对于所有的 $1 \leq k \leq K-1$, 由 (3.23) 所确定的滤波分布 $\mathbb{P}(v_k | Y_k)$ 的协方差 $\mathcal{C}_k$ 是正定的, 且

$$\begin{cases} \hat{m}_{k+1} = \Psi m_k, \\ \hat{\mathcal{C}}_{k+1} = \Psi \mathcal{C}_k \Psi^T + \Sigma, \\ \mathcal{C}_{k+1}^{-1} = (\Psi \mathcal{C}_k \Psi^T + \Sigma)^{-1} + H^T \Gamma^{-1} H, \\ \mathcal{C}_{k+1}^{-1} m_{k+1} = (\Psi \mathcal{C}_k \Psi^T + \Sigma)^{-1} \Psi m_k + H^T \Gamma^{-1} y_{k+1}. \end{cases}$$

这里省略如上定理的证明, 感兴趣的读者可参见文献 [189, 定理 8.3]. 值得注意的是, 上面定理中协方差的更新没用到数据信息, 且预测步骤中的协方差是以仿射变换的方式依赖于前一状态的协方差, 而分析步骤的协方差则是以非线性的方式依赖前一状态的协方差. 借助于 Woodbury 公式 (参见文献 [189, 引理 8.6]), Kf 算法可以等价地改写为算法 2 的形式, 其中的 K_k 又称为 Kalman 增益 (Kalman gain). 可以证明 Kf 算法给出了状态均值的最优化估计 (参见文献 [189, 定理 8.7]). 对于 Kf 算法的其他推导方式或更新公式, 可参见文献 [13, 149, 183, 190] 及其中的参考文献.

算法 2 Kf 算法

给定初始分布 $\pi_0 = \mathcal{N}(m_0, \mathcal{C}_0)$.

1. 设置 $k = 0$.

2. 预测:

$$\hat{m}_{k+1} = \Psi m_k, \quad \hat{\mathcal{C}}_{k+1} = \Psi \mathcal{C}_k \Psi^T + \Sigma.$$

3. 分析:

$$m_{k+1} = \hat{m}_{k+1} + K_{k+1}(y_{k+1} - H\hat{m}_{k+1}),$$

$$K_{k+1} = \hat{\mathcal{C}}_{k+1} H^T (H \hat{\mathcal{C}}_{k+1} H^T + \Gamma)^{-1},$$

$$\mathcal{C}_{k+1} = (I - K_{k+1} H) \hat{\mathcal{C}}_{k+1}.$$

4. 令 $k \rightarrow k+1$, 若 $k < K$, 返回步骤 2; 否则输出预测分布 $\hat{\pi}_K = \mathcal{N}(\hat{m}_K, \hat{\mathcal{C}}_K)$ 和滤波分布 $\pi_K = \mathcal{N}(m_K, \mathcal{C}_K)$.

需要指出的是, Kf 算法只能处理线性随机动力学模型和线性观察数据模型的情形, 若随机动力学模型为非线性的且数据观察模型为线性的情形, 则成立

$$\begin{cases} v_{k+1} = \psi(v_k) + \xi_k, \\ y_{k+1} = H v_{k+1} + \eta_{k+1}, \end{cases} \quad (3.24)$$

其中 v_0 及序列 $\{\xi_k\}$ 和 $\{\eta_k\}$ 的假设如 (3.21) 和 (3.22) 所述. 此时, 若状态 v_k 视为其均值 m_k 的小扰动, 即成立

$$\psi(v_k) \approx \psi(m_k),$$

则有

$$\hat{m}_{k+1} = \mathbb{E}[v_{k+1}|Y_k] = \mathbb{E}[\psi(v_k) + \xi_k|Y_k] = \psi(v_k) \approx \psi(m_k)$$

及

$$\begin{aligned} \hat{\mathcal{C}}_{k+1} &= \mathbb{E}[(v_{k+1} - \hat{m}_{k+1}) \otimes (v_{k+1} - \hat{m}_{k+1})|Y_k] \\ &\approx \mathbb{E}[(\psi(v_k) - \psi(m_k) + \xi_k) \otimes (\psi(v_k) - \psi(m_k) + \xi_k)|Y_k] \\ &= \mathbb{E}[(\psi(v_k) - \psi(m_k)) \otimes (\psi(v_k) - \psi(m_k))|Y_k] + \Sigma \\ &\approx D\psi(m_k) \mathbb{E}[(v_k - m_k) \otimes (v_k - m_k)|Y_k] D\psi(m_k)^T + \Sigma \\ &= D\psi(m_k) \mathcal{C}_k D\psi(m_k)^T + \Sigma, \end{aligned}$$

其中 $D\psi(m_k)$ 为 $\psi(\cdot)$ 在 m_k 处的 Fréchet 导数, 且 $\hat{m}_{k+1}$ 和 $\hat{\mathcal{C}}_{k+1}$ 的计算过程中用到了 ξ_k 与 Y_k 及 ξ_k 与 v_k 的相互独立性. 只需在 Kf 算法中将预测均值 $\hat{m}_{k+1}$ 和协方差 $\hat{\mathcal{C}}_{k+1}$ 作相应的替换即得到扩展

Kalman 滤波 (extended Kalman filter, ExKf) 算法. 可以看出 ExKf 算法是通过线性化的方式来逼近预测协方差的, 从而其计算代价主要在于非线性映射 ψ 的 Fréchet 导数的计算. 关于 ExKf 算法的发展及系统性理论可参见文献 [106], ExKf 算法在天气预报中的应用可参见文献 [43]. 大多数地球物理应用中的状态空间维数都很高, 这也使得 ExKf 算法的应用变得不切实际.

从前面的论述可知, 在高维情形的应用中, 无论是 Kf 算法还是 ExKf 算法, 预测协方差 (特别是 Fréchet 导数 $D\psi(\cdot)$) 的估计和存储都将变得低效和昂贵, 集合 Kalman 滤波 (ensemble Kalman filter, EnKf) 算法正是为了克服这一困难而产生的, 其基本思想是将一个粒子集合的演化与 Kalman 型算法相结合, 并在更新过程中使用该粒子集合的经验协方差来逼近预测步骤的协方差 $\hat{C}_{k+1}$. 关于 EnKf 算法的早期发展可参见文献 [71, 73, 98, 99], 也可参见近期的综述性文献 [72, 149, 189] 等. 目前在 EnKf 算法的基础上发展出了一系列的相关算法, 大部分都可纳入广义非线性最小二乘与 Gauss-Newton 迭代或 Levenberg-Marquardt 方法相结合的一致优化框架内 (参见文献 [189, 第三部分, Kalman 反演]), 我们将这一类方法统称为 EnKf 算法.

下面回到本节的主题, 即阐述如何利用 EnKf 算法来求解 Bayes 反问题

$$y = \mathcal{G}(u) + \eta, \quad u \in X, \quad y \in Y, \quad \eta \sim \mathcal{N}(0, \Gamma), \quad (3.25)$$

并给出一般的算法框架. 为此, 先以数据增广的方式构造人工动力系统:

$$\begin{cases} z_{k+1} = \Xi(z_k) + s\zeta_k, & s \in \{0, 1\}, \\ d_{k+1} = \tilde{H}z_{k+1} + \tilde{\eta}_{k+1}, \end{cases} \quad (3.26)$$

其中

$$\Xi(z_k) = \begin{pmatrix} u_k \\ G(u_k) \end{pmatrix} =: \begin{pmatrix} u_k \\ p_k \end{pmatrix},$$

$\tilde{H} = (0, I) : X \times Y \rightarrow Y$, $\{\zeta_k\}$ 和 $\{\tilde{\eta}_k\}$ 为独立同分布序列 ($\zeta_0 \sim \mathcal{N}(0, \tilde{\Sigma})$, $\tilde{\eta}_0 \sim \mathcal{N}(0, \tilde{\Gamma})$), $\{u_0^{(j)}\}_{j=1}^J$ 一般选取为未知量 u 的先验概率分布的随机采样. 若正演映射 $\mathcal{G} : X \rightarrow Y$ 是非线性的, 则 (3.26) 中的状态模型 $z_{k+1} = \Xi(z_k) + s\zeta_k$ 即为非线性的, 但数据观测模型 $d_{k+1} = \tilde{H}z_{k+1} + \tilde{\eta}_{k+1}$ 是线性的, 这与 (3.24) 的情形一致.

假设第 k 步使用的粒子集合记为

$$\{z_k^{(j)}\}_{j=1}^J = \left\{ \begin{pmatrix} u_k^{(j)} \\ \mathcal{G}(u_k^{(j)}) \end{pmatrix} \right\}_{j=1}^J,$$

则 $\{z_k^{(j)}\}_{j=1}^J$ 的经验协方差即为

$$\hat{C}_k := \frac{1}{J} \sum_{j=1}^J (z_k^{(j)} - \bar{z}_k) \otimes (z_k^{(j)} - \bar{z}_k) = \begin{pmatrix} C_k^{uu} & C_k^{up} \\ (C_k^{up})^\top & C_k^{pp} \end{pmatrix},$$

其中

$$\bar{z}_k = \frac{1}{J} \sum_{j=1}^J z_k^{(j)} = \begin{pmatrix} \frac{1}{J} \sum_{j=1}^J u_k^{(j)} \\ \frac{1}{J} \sum_{j=1}^J \mathcal{G}(u_k^{(j)}) \end{pmatrix} =: \begin{pmatrix} \bar{u}_k \\ \bar{p}_k \end{pmatrix},$$

$$\begin{aligned}
C_k^{uu} &= \frac{1}{J} \sum_{j=1}^J (u_k^{(j)} - \bar{u}_k) \otimes (u_k^{(j)} - \bar{u}_k), \\
C_k^{up} &= \frac{1}{J} \sum_{j=1}^J (u_k^{(j)} - \bar{u}_k) \otimes (\mathcal{G}(u_k^{(j)}) - \bar{p}_k), \\
C_k^{pp} &= \frac{1}{J} \sum_{j=1}^J (\mathcal{G}(u_k^{(j)}) - \bar{p}_k) \otimes (\mathcal{G}(u_k^{(j)}) - \bar{p}_k).
\end{aligned}$$

在 Kf 算法中, 只需要将均值的更新替换为粒子集合 $\{z_k^{(j)}\}_{j=1}^J$ 的更新, 而将预测步协方差的更新替换为粒子集合的经验协方差的更新, 就可以得到求解反问题 (3.25) 的 EnKf 算法 (算法 3) [102].

算法 3 EnKf 算法

给定初始粒子 $\{z_0^{(j)}\}_{j=1}^J$.

1. 设置 $k = 0$.

2. 预测:

$$\begin{aligned}
\zeta_k^{(j)} &\stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tilde{\Sigma}), \quad j = 1, 2, \dots, J, \quad \hat{z}_{k+1}^{(j)} = \Xi(z_k^{(j)}) + \zeta_k^{(j)}, \\
\hat{m}_{k+1} &= \frac{1}{J} \sum_{j=1}^J \hat{z}_{k+1}^{(j)}, \quad \hat{C}_{k+1} = \begin{pmatrix} C_k^{uu} & C_k^{up} \\ (C_k^{up})^T & C_k^{pp} \end{pmatrix}.
\end{aligned}$$

3. 分析:

$$\begin{aligned}
\tilde{\eta}_k^{(j)} &\stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tilde{\Gamma}), \quad j = 1, 2, \dots, J, \quad d_{k+1}^{(j)} = d_{k+1} + \tau \tilde{\eta}_k^{(j)}, \quad \tau \in \{0, 1\}, \\
z_{k+1}^{(j)} &= \hat{m}_{k+1} + K_{k+1}(d_{k+1}^{(j)} - \tilde{H} \hat{m}_{k+1}), \quad K_{k+1} = \hat{C}_{k+1} \tilde{H}^T (\tilde{H} \hat{C}_{k+1} \tilde{H}^T + \Gamma)^{-1}.
\end{aligned}$$

4. 令 $k \rightarrow k+1$, 若 $k < K$, 返回步骤 2; 否则输出粒子集合 $\{z_k^{(j)}\}_{j=1}^J, k = 1, 2, \dots, K$.

这里将 EnKf 算法写成算法 3 的形式, 目的只是明确其与传统 Kf 算法之间的联系. 实际应用中, 在绝大多数情形下, 我们往往只关注未知量 u 的信息, 因此可以将算法 3 中的经验协方差 $\hat{C}_{k+1}$ 的更新简化为部分经验协方差 C_k^{uu} 和 C_k^{pp} 的更新, 而将粒子集合 $\{z_k^{(j)}\}$ 的更新简化为未知量粒子集合 $\{u_k^{(j)}\}$ 的更新, 从而较大地提升 EnKf 算法的计算效率, 具体细节可参见文献 [101, 算法 1].

事实上, EnKf 算法可以看作是滤波分布 $\pi(v_k | Y_k)$ 的同权序列 Monte Carlo (sequential Monte Carlo, SMC) 逼近 [24, 189, 191]. 记

$$\pi_k^J(z_k) = \frac{1}{J} \sum_{j=1}^J \delta(z_k - z_k^{(j)}), \quad (3.27)$$

当随机动力学模型中的非线性关系较弱或数据噪声较小时, 滤波分布可近似看成是 Gauss 分布; 当 J 充分大时, $\pi_k^J(z_k)$ 可作为滤波分布的较好逼近, 从而 EnKf 算法输出的粒子集合可视为滤波分布的逼近采样 (此时, (3.26) 中的参数 $s = 1$, 其目的是使得粒子集合更快达到混合时间) [77, 78]. 反之, 当随机动力学模型中的非线性关系较强或数据噪声较大时, $\pi_k^J(z_k)$ 逼近滤波分布的效果较差, 即便如此, EnKf 算法输出的粒子集合的均值仍能很好地逼近状态变量, 只不过这种情形最好将 EnKf 算法理解为序列优化方法 [189] (此时, (3.26) 中的参数 $s = 0$) 而不是当作逼近采样方法.

EnKf 算法在反问题、深度学习及数据同化等领域获得了广泛应用, 概括起来大致有如下 3 个方面的原因 (参见文献 [189]). (1) EnKf 算法可以在不估计随机动力学模型 (在反问题求解中则为正演

映射 g) 的导数的情形下进行计算, 其本质是使用统计线性化来逼近相关导数, 这对于模型导数难以计算或者黑箱模型就显得尤为重要; (2) 当粒子集合的规模 J 小于未知量的维数时, 使用经验协方差而不是模型协方差可以显著降低计算成本; (3) 当随机动力学模型的非线性关系不是太强时, 粒子集合的分布包含了未知量不确定性的有用信息, 从而粒子集合本身可以作为滤波分布 (在 Bayes 反问题情形则为后验概率分布) 的逼近随机采样.

最后需要指出的是, 对 EnKf 算法的分析是困难的, 相关理论研究还处于开始阶段. 对于线性情形, 当粒子规模 $J \rightarrow \infty$ 时, EnKf 算法将收敛于 Kalman 滤波分布^[141, 150, 160]; 而对于非线性情形, 粒子集合则不能很好地逼近相应的滤波分布 (参见文献 [70]). 在一致平均场框架下, 关于 EnKf 算法的综述可参见文献 [35], 该框架为分析 EnKf 算法逼近真实滤波分布的准确性提供了理论基础. 对于粒子规模不大或者固定粒子规模情形, EnKf 算法仍能提供良好的状态估计, 这也被视为 EnKf 算法无可争议的优点之一, 相关的理论分析可参见文献 [9, 24, 83, 127, 191, 192, 205, 206] 等.

总之, 由于 EnKf 算法有着方法简单、计算量小且允许使用部分、低秩、经验相关信息等优点, 随着数据的快速增长和人工智能技术的迅速发展, 可以预见 EnKf 算法将会有着更加广阔的应用前景和发展空间.

3.3 变分推断

以 pCN 算法为代表的抽样算法需要大量抽样从而计算后验统计信息, 针对稳态 Darcy 流反渗透率问题, 文献 [48] 抽取了 10^6 个样本, 即计算了 10^6 个偏微分方程. 根据我们的计算经验, 对于稍复杂的反问题, 大约需要抽取 10^5 – 10^7 个样本才可以有效地估计后验统计信息. 求解 10^6 个偏微分方程是计算量极其庞大的计算任务, 因而以 pCN 为代表的无限维空间抽样算法还是难以应用于全波形反演^[184] 等大规模反演问题.

2012 年, AlexNet^[139] 在 ImageNet 比赛中获得了突破性进展, 准确率远超第二名 (第一名的前五误差率 (Top-5 error rate) 错误率为 15.3%, 第二名为 26.2%). 自此, 神经网络获得了广泛关注, 相关研究渗透到了众多领域. 特别地, 深度学习方法在反问题求解中取得了令人瞩目的效果 (参见文献 [12]). 然而现实世界中的学习问题面临众多不确定性信息, 因此有必要对深度学习的不确定性进行分析, Bayes 统计在其中扮演了重要的角色 (参见文献 [1]). 神经网络中包含大量的可学习参数, 进而训练神经网络可以看作高维空间上的统计推断问题, 面临着与无限维 Bayes 反演类似的困难: 如何从高维 (甚至是无穷维) 空间上的概率分布中快速计算后验统计信息.

为了处理高维空间带来的计算困难, 在机器学习领域发展出了丰富的变分推断理论与算法. 变分推断的发展可以追溯到 20 世纪 80 年代, 基于平均场假设的变分推断方法很早就被用来研究神经网络^[173, 176]. 随后, 在 20 世纪 90 年代, 变分推断被众多学者用于概率图模型的研究 (参见文献 [105, 119, 172]). 至今, 变分推断伴随着机器学习 (特别是深度学习) 的研究获得了极大的关注, 理论与算法也越来越丰富. 关于这方面的近期发展, 我们推荐两篇综述文献 [23, 219]. 如前文所述, 无限维 Bayes 反演与机器学习研究面临着相似的核心困难, 那么一个显而易见的问题是, 变分推断方法能否用于无限维 Bayes 反演领域? 沿着“先离散, 再 Bayes”的思路, 针对有限维反问题, 有一些很有意思的工作. 文献 [117] 研究了带有超参数的分层 Bayes 反演, 基于平均场逼近 (假设反演参数与超参数是独立的随机变量)、Gauss 先验噪声假设, 给出了变分推断的迭代求解格式, 每一步迭代都具有解析表达式, 因而计算速度相较抽样算法得到了极大提升. 特别地, 文献 [117] 给出了算法收敛性的初步分析, 为算法参数初值的选取提供了一定的指导. 随后, 这一工作被推广到了带有厚尾分布、偏态 $-t$ 分布观

测噪声假设下的有限维反问题^[85,116]以及多孔介质流中的随机介质反演问题^[215]. 近期, 有限维空间上的变分推断算法也被用于求解地学反问题^[220]和非局部微分方程反问题^[194], 从而快速地近似计算后验统计信息. 下面从无限维 Bayes 反演理论的角度, 对变分推断在反问题领域的发展作一些回顾.

首先, 简略介绍什么是变分推断. 本小节假设 X 是一个可分 Hilbert 空间. 简单来讲, 变分推断将后验计算问题转化为了一个关于概率分布的优化问题

$$\arg \min_{\nu \in \mathcal{A}} D(\nu \| \mu^y), \quad (3.28)$$

其中 $D(\cdot \| \cdot)$ 表示概率测度间的某个度量, $\mathcal{A}$ 是 $\mathcal{B}(X)$ 上的概率测度构成的集合. 例如, 可取 $\mathcal{A}$ 表示 Gauss 测度的集合:

$$\mathcal{A} = \{\mathcal{N}(\bar{u}, \mathcal{C}) : \bar{u} \in X, \mathcal{C} \text{ 是 } X \text{ 上的对称、正定的迹算子}\}. \quad (3.29)$$

由变分推断这一宽泛的定义可以看出, 不同度量、概率测度集合的选择可以给出不同的算法. 这里仅考虑度量 $D(\cdot \| \cdot)$ 取为如下 Kullback-Leibler (KL) 散度的情形:

$$D_{KL}(\nu \| \mu^y) = \int_X \log \left(\frac{d\nu}{d\mu^y}(u) \right) \frac{d\nu}{d\mu^y}(u) \mu^y(du). \quad (3.30)$$

在上式中, 采用通常的规定 $0 \log 0 = 0$.

注 3.5 需要注意的是, 这里所采用的 KL 散度其实并不是真正意义上的度量, 其一般而言不满足对称性, 即

$$D_{KL}(\nu \| \mu^y) \neq D_{KL}(\mu^y \| \nu).$$

然而 KL 散度是非负的, 且具有度量的一些好的性质 (参见文献 [179]), 例如, 如果 $D_{KL}(\nu \| \mu^y) = 0$, 则有 $\nu = \mu^y$. KL 散度描述了从概率测度 μ^y 到概率测度 ν 的信息增益 (information gain), 因而在有限维空间上的变分推断研究中, 经常选取 KL 散度来度量两个概率测度的距离 (参见文献 [22]).

当固定了度量 $D(\cdot \| \cdot)$ 之后, 集合 $\mathcal{A}$ 如何选取是构建变分推断方法的关键.

(1) 如果 $\mathcal{A}$ 选取得过小, 仅包含较为简单的概率测度, 则优化问题会变得易于求解, 但对后验测度的近似必然会不精确.

(2) 如果 $\mathcal{A}$ 选取得很大, 对后验测度的近似会更为精确, 但计算会变得困难. 例如, 取 $\mathcal{A}$ 为对先验测度绝对连续的所有可能概率测度, 则最优解就是后验测度, 计算并没有得到简化.

不同的变分推断方法会选取不同的集合 $\mathcal{A}$, 从而在求解精度与可计算性之间寻求平衡. 下面对 3 种典型的情形进行简要介绍.

假设 3.1 $\mathcal{A}$ 是 Gauss 概率测度构成的集合.

当假设集合 $\mathcal{A}$ 是 Gauss 概率测度构成的集合时, 文献 [179,180] 给出了详细的分析. 具体而言, 文献 [179] 证明了如下定理.

定理 3.3 令 X 是可分 Hilbert 空间, 先验概率测度 $\mu_0 = \mathcal{N}(u_0, \mathcal{C}_0)$, 其中 $\mathcal{C}_0$ 是 X 上的正定、对称的迹算子, $u_0 \in \mathcal{H}$ ($\mathcal{H}$ 表示先验概率测度 μ_0 的 Cameron-Martin 空间). 考虑如下集合 $\mathcal{A}$:

- (1) $\mathcal{A}_1 = \{\text{空间 } X \text{ 上的 Gauss 概率测度}\};$
- (2) $\mathcal{A}_2 = \{\text{空间 } X \text{ 上等价于 } \mu_0 \text{ 的 Gauss 概率测度}\};$
- (3) 对于空间 X 上的对称、正定的迹算子 $\hat{\mathcal{C}}$, 取 $\mathcal{A}_3 = \{\text{空间 } X \text{ 具有方差算子 } \hat{\mathcal{C}} \text{ 的 Gauss 概率测度}\};$

(4) 对于固定的 $\hat{u} \in X$, 取 $\mathcal{A}_4 = \{\text{空间 } X \text{ 具有均值 } \hat{u} \text{ 的 Gauss 概率测度}\}$.

对于这 4 种情形, 如果存在 $\nu \in \mathcal{A}_i$ ($i = 1, 2, 3, 4$) 使得 $D_{KL}(\nu \parallel \mu^y) < \infty$, 则变分优化问题

$$\arg \min_{\nu \in \mathcal{A}_i} D_{KL}(\nu \parallel \mu^y)$$

的最优解存在, 并且最优概率测度 ν 与先验概率测度 μ_0 是等价的.

这一定理给出了变分优化问题解的存在性, 并且告诉我们最优逼近概率测度 ν 与先验概率测度 μ_0 等价, 进而什么样的 Gauss 测度等价于 μ_0 就是一个很重要的问题. 令 $\mathcal{HS}(\mathcal{H})$ 表示空间 $\mathcal{H}$ 上的 Hilbert-Schmidt 算子构成的空间, 文献 [179, 180] 进一步证明了, 对于 Gauss 概率测度 $\nu = \mathcal{N}(u, \mathcal{C})$, 如果 $u \in \mathcal{H}$ 且方差算子 $\mathcal{C}^{-1} = \mathcal{C}_0^{-1} + \Gamma$ 满足

$$\|\mathcal{C}_0^{1/2} \Gamma \mathcal{C}_0^{1/2}\|_{\mathcal{HS}(\mathcal{H})} < \infty,$$

则 Gauss 概率测度 ν 与先验概率测度 μ_0 等价. 基于等价 Gauss 概率测度的刻画, 文献 [180] 给出了有限秩参数化和 Schrödinger 参数化方法, 进而基于 Robbins-Monro 算法构造了新的 Gauss 逼近变分推断算法.

基于 Gauss 概率测度逼近后验概率测度的另一种方法是 Laplace 逼近方法, 即对正演算子 $\mathcal{G}$ 进行线性化. 在 Gauss 先验和 Gauss 噪声的假设下, 对于线性问题, 可以显式推导出后验概率测度的表达式, 从而快速计算协方差函数等统计量 [32].

假设 3.2 平均场假设, 即带求参数的分量之间互相独立.

关于平均场假设下的变分推断, 文献 [115] 在较为宽泛的一般假设下构建了无限维空间上的平均场变分推断理论, 并将其运用于带有 Gauss 噪声、Laplace 噪声的多频数据 Helmholtz 方程逆源反问题. 这里, 举个例子说明其核心思想, 考虑如下问题:

$$y = \mathcal{G}u + \eta, \quad (3.31)$$

其中 $y \in Y \subset \mathbb{R}^{n_y}$, $\eta \sim \mathcal{N}(0, \tau^{-1}I)$, $\mathcal{G} : X \rightarrow Y$ 是一个线性紧算子, $u \in X$ (X 是一个无限维可分 Hilbert 空间) 且

$$u \sim \mu_0^u = \mathcal{N}(0, \mathcal{C}_0^K(\lambda)), \quad \mathcal{C}_0^K(\lambda) = \sum_{k=1}^K \lambda^{-1} \alpha_k e_k \otimes e_k + \sum_{k=K+1}^{\infty} \alpha_k e_k \otimes e_k,$$

$$\lambda \sim \mu_0^\lambda = \text{Gamma}(\alpha_0, \beta_0), \quad \tau \sim \mu_0^\tau = \text{Gamma}(\alpha, \beta).$$

这里 $\{\alpha_k, e_k\}_{k=1}^{\infty}$ 是某个对称、正定的迹算子的特征系统, $\text{Gamma}(\alpha, \beta)$ 表示参数为 α 和 β 的 Gamma 分布. 在这个问题中, 假设了先验测度方差算子的参数 λ 与观测噪声分布的参数 τ 是未知的参数, 因而在 Bayes 反演的框架中, 假设它们服从 Gamma 分布 (Gamma 分布在这个问题中是共轭分布, 即先验与后验属于同样的分布族). 在现有的假设下, 我们所考虑的反问题不再是反演 u , 而是反演一组参数 (u, λ, τ) , Bayes 公式具有如下形式:

$$\frac{d\mu^y}{d\mu_0}(u, \lambda, \tau) = \frac{1}{Z^y} \tau^{\frac{n_y}{2}} \exp\left(-\frac{\tau}{2} \|\mathcal{G}u - y\|^2\right), \quad (3.32)$$

其中 $\mu_0 = \mu_0^u \otimes \mu_0^\lambda \otimes \mu_0^\tau$ 是乘积概率测度, Z^y 为归一化常数. 所谓平均场假设, 即假设反演参数 u , λ 和 τ 是互相独立的 3 个参数. 在这个例子中, 我们引入参考测度

$$\mu_r = \mu_r^u \otimes \mu_r^\lambda \otimes \mu_r^\tau = \mathcal{N}(0, \mathcal{C}_0) \otimes \mu_0^\lambda \otimes \mu_0^\tau,$$

其中

$$\mathcal{C}_0 = \sum_{k=1}^{\infty} \alpha_k e_k \otimes e_k.$$

容易看到, 先验测度 μ_0 与参考测度 μ_r 是等价的, 从而有

$$\frac{d\mu_0}{d\mu_r}(u, \lambda, \tau) \propto \exp(-\Phi^0(u, \lambda, \tau)). \quad (3.33)$$

下面假设逼近测度 ν 具有如下表达式:

$$\frac{d\nu}{d\mu_r}(u, \lambda, \tau) \propto \exp(-\Phi_u^r(u) - \Phi_\lambda^r(\lambda) - \Phi_\tau^r(\tau)), \quad (3.34)$$

进而有

$$\frac{d\nu^u}{d\mu_r^u} \propto \exp(-\Phi_u^r(u)), \quad \frac{d\mu^\lambda}{d\mu_r^\lambda} \propto \exp(-\Phi_\lambda^r(\lambda)), \quad \frac{d\mu^\tau}{d\mu_r^\tau} \propto \exp(-\Phi_\tau^r(\tau)). \quad (3.35)$$

上述定义式 (3.34) 其含义是, 逼近测度关于参考测度的密度函数关于参数 u , λ 和 τ 具有独立的 3 个分量. 粗略来讲, 定义变分推断的约束集合 $\mathcal{A} = \mathcal{A}_u \times \mathcal{A}_\lambda \times \mathcal{A}_\tau$, 其中

$$\mathcal{A}_u = \{\nu^u \in \mathcal{B}(X) : \nu^u \text{ 与 } \mu_r^u \text{ 等价, 且 } \Phi_u^r \text{ 满足一些有界、可积条件}\},$$

集合 $\mathcal{A}_\lambda$ 和 $\mathcal{A}_\tau$ 可类似定义. 限于篇幅, 这里不列出位势函数所满足的具体有界、可积性条件, 准确的数学定义参见文献 [115]. 基于这些假设, 文献 [115] 证明了逼近概率测度定义 (3.34) 中的位势函数具有如下形式:

$$\begin{aligned} \Phi_u^r(u) &= \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} (\Phi^0(u, \lambda, \tau) + \Phi(u, \lambda, \tau)\nu^\lambda)(d\lambda)\nu^\tau(d\tau) + \text{Const}, \\ \Phi_\lambda^r(\lambda) &= \int_{\mathbb{R}^+} \int_X (\Phi^0(u, \lambda, \tau) + \Phi(u, \lambda, \tau)\nu^u)(du)\nu^\tau(d\tau) + \text{Const}, \\ \Phi_\tau^r(\tau) &= \int_{\mathbb{R}^+} \int_X (\Phi^0(u, \lambda, \tau) + \Phi(u, \lambda, \tau)\nu^u)(du)\nu^\lambda(d\lambda) + \text{Const}, \end{aligned}$$

其中 Φ 是 Bayes 公式中的负对数似然函数, Const 表示无关的常数. 在上述的假设下, 我们其实可以进一步计算得到 Φ_u^r , Φ_λ^r 和 Φ_τ^r 的解析表达式. 当然, 我们需要说明这里得到的解析表达式并不能直接计算, 因为任何一个位势函数的解析表达式依赖于其他两个参数的近似后验测度, 例如, Φ_u^r 的解析表达式依赖于 ν^λ 和 ν^τ . 然而基于 Φ_u^r , Φ_λ^r 和 Φ_τ^r 的形式, 我们容易构造迭代算法: 给定位势函数中参数 (可能是函数参数) 的初值, 然后依次循环更新位势函数 Φ_u^r , Φ_λ^r 和 Φ_τ^r 直到逼近测度中的参数变化小于某个阈值.

注 3.6 基于平均场逼近的一般变分推断理论不仅适用于上述 Gauss 先验、Gauss 噪声的情形, 同样适用于带有 Laplace 噪声的情形 (参见文献 [115, 第 3.2 小节]). 针对多参数反演问题 [38, 118], 基于平均场假设 (假设各个参数是独立的随机变量) 构造变分推断算法值得进一步研究.

注 3.7 在上述介绍中, 假设了参数 u 服从先验分布 $\mu_0 = \mathcal{N}(0, \mathcal{C}_0^K(\lambda))$, 其方差算子具有如下形式:

$$\mathcal{C}_0^K(\lambda) = \sum_{k=1}^K \lambda^{-1} \alpha_k e_k \otimes e_k + \sum_{k=K+1}^{\infty} \alpha_k e_k \otimes e_k.$$

这里的超参数 λ 仅能够调整方差算子的前 K 项, 如果似然函数含有较多的信息, 提前固定参数 K 会限制模型的有效性, 如何突破这一限制是一个很有意思的问题. 文献 [198] 采用了非中心参数化的方式发展了新的平均场变分推断方法, 从而在一定程度上克服了这一问题.

注 3.8 基于平均场逼近的变分推断理论在机器学习中有广泛的应用, 例如, 基于平均场变分推断理论, 在噪声是非独立同分布的情形时, 文献 [217] 构建了变分去噪模型, 进一步文献 [218] 构造了变分超分模型. 借助无限维的平均场变分推断理论, 文献 [113] 在非独立同分布噪声假设的情形下, 发展了新的变分逆深度生成模型, 在多频数据 Helmholtz 方程逆源问题的计算中取得了很好的效果.

假设 3.3 $\mathcal{A} := \{\nu : \nu = \mu_0 \circ T^{-1}, T \text{ 是某种变换}\}.$

在假设 3.3 中的变换 T 的不同选取可以导出很多不同类型的变分推断方法, 如基于标准化流 (normalizing flow) 的变分推断^[220] 和 Stein 变分梯度下降 (Stein variational gradient descent, SVGD)^[156] 等. 这里, 简要介绍 SVGD 算法及其无限维版本 iSVGd (infinite-dimensional SVGD)^[107].

SVGD 是 Liu 和 Wang^[156] 提出的 (针对有限维空间), 根据我们的理解, 其基本的想法是将变换 T 分解成一系列的变换, 如图 2 所示, 其中红色的点表示采样得到的样本粒子. 初始的样本粒子一般从先验测度 μ_0 中抽取, 经过一系列的非线性变换 (恒等算子的微小扰动) 最终得到的样本粒子就可以认为是从后验测度中抽取出来的. 更具体地, 图 2 中每个变换具有如下形式:

$$T_i = I + \epsilon \phi_i, \quad i = 1, 2, \dots, N, \quad (3.36)$$

其中 ϵ 是一个很小的正数, I 表示恒等变换, $\{\phi_i\}_{i=1}^N$ 是待确定的变换. 容易观察到, 如果 $N \rightarrow \infty$ 且 ϕ_i 的选取不加以限制, 则复合变换 $T_1 \circ T_2 \circ \dots$ 可以逼近任意变换, 因而计算得不到简化. 在 SVGD 方法中, 变换函数 ϕ_i 被限制在一个再生核 Hilbert 空间 (reproducing kernel Hilbert spaces, RKHS) 中. 记 RKHS 为 $\mathcal{H}_K$, 其中 K 表示核函数. 关于 RKHS 的详细介绍, 可参见文献 [49, 163]. 如果将 ϕ_i 限制在 $\mathcal{H}_K$ 中并且记 $\nu_i = \mu_0 \circ (T_i \circ \dots \circ T_1)^{-1}$, 则直观上来讲应当通过如下方式确定 ϕ_i :

$$\phi_i^* = \arg \max_{\phi_i \in \mathcal{H}_K, \|\phi_i\|_{\mathcal{H}_K} \leq 1} \left\{ -\frac{d}{d\epsilon} D_{KL}(\nu_{i-1} \cdot T_i^{-1} \parallel \mu) \right\}_{\epsilon=0}. \quad (3.37)$$

文献 [156] 通过计算得到了问题 (3.37) 的解析解表达式如下:

$$\phi_i^*(\cdot) \propto \mathbb{E}_{u \sim \nu_{i-1}} [K(u, \cdot) D_u \log p(u | y) + D_u K(u, \cdot)], \quad (3.38)$$

其中 $q(u | y)$ 表示后验概率密度函数 (这里的简介考虑有限维空间), D_u 表示 u 的梯度. 通过样本粒子的平均值取代最优解解析表达式 (3.38) 中的期望算子, 得到可计算的样本粒子迭代公式

$$u_j^k \leftarrow u_j^{k-1} + \frac{\epsilon_{k-1}}{L} \sum_{\ell=1}^L [K(u_\ell^{k-1}, \cdot) D_{u_\ell^{k-1}} \log p(u_\ell^{k-1} | y) + D_{u_\ell^{k-1}} K(u_\ell^{k-1}, \cdot)], \quad (3.39)$$

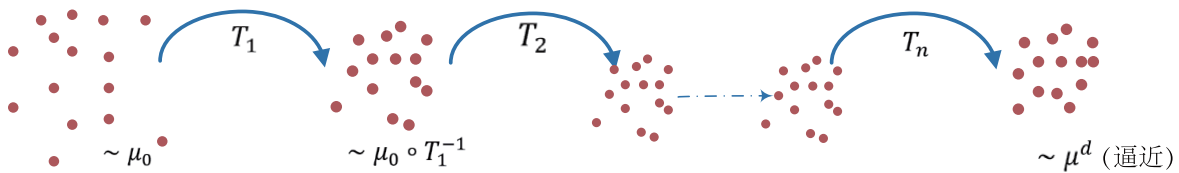


图 2 (网络版彩图) SVGD 基本思路示意图

其中 k 表示迭代步数, $j = 1, \dots, L$ (L 表示总的样本数目). SVGD 算法在机器学习领域获得了广泛的关注 (参见文献 [57, 84, 185]), 其背后蕴涵着 Riemann 流形上的梯度流分析等深刻的数学理论 (参见文献 [134, 154, 155, 158]).

沿着“先 Bayes, 后离散”的研究思路, 一个直接的问题是, SVGD 的理论算法能否在无限维空间建立? 事实上, 即使在深度神经网络的研究中, 研究者也希望在无限维空间上构建基于粒子演化的近似抽样算法, 从无限维空间这一观点出发可以避免深度神经网络众多不同参数对应同一拟合函数从而使得 SVGD 失效的难题 (参见文献 [213]). 文献 [107] 针对无限维 Bayes 反演问题, 引入了算子值核函数, 进而在无限维可分 Hilbert 空间上建立了 SVGD 迭代格式 (称为 iSVGd). 进一步地, 该文献中引入了核函数变换的性质, 从而给出了带有预条件算子的 iSVGd. 通过无限维空间上的分析, 文献 [107] 给出了 iSVGd 迭代有意义的严格条件, 同时揭示了: 相较于有限维的理论^[154], 无限维空间中预条件算子不能随意取, 须保障核函数中不包含先验 Gauss 测度的 Cameron-Martin 空间范数从而保障样本粒子不过度集中. 针对 iSVGd, 还有很多理论问题有待解决, 例如, 如何将文献 [158] 中的理论推广到 iSVGd 的情形, 从而建立 iSVGd 中经验分布趋于后验测度的速率估计.

3.4 离散不变

本小节对离散不变性进行一些探讨, 试图通过具体的数值算例揭示为何要在函数空间上直接构造 MCMC、Kalman 滤波和变分推断等算法.

具体而言, 考虑稳态 Darcy 流模型^[55] 如下:

$$\begin{cases} -\nabla \cdot (e^u \nabla w) = f, & x \in D, \\ w = 0, & x \in \partial D, \end{cases} \quad (3.40)$$

其中 $f = 1$ 是一个常值函数, $D = (0, 1)^2 \subset \mathbb{R}^2$. 考虑如下反问题: 基于方程的解 w 在区域内部离散的 100 个点上进行测量所得到的测量数据, 推断函数参数 u . 在这一问题中, 前文中的空间 $X = L^2(D)$. 当 $u \in L^\infty(D)$ 时, 易知方程的解 $w \in H_0^1(D)$. 针对这一具体问题, 取两种先验测度: $\mu_0^1 = \mathcal{N}(0, \mathcal{C}_0)$ (其中 $\mathcal{C}_0 = (0.5I - 0.1\Delta)^{-2}$) 和 $\mu_0^2 = \mathcal{N}(0, I)$ (其中 I 为空间 X 上的单位算子). 显然, 先验 μ_0^1 的方差算子是迹算子, 而 μ_0^2 的方差算子不是迹算子, 因而先验测度 μ_0^2 的选取不符合无限维 Bayes 理论的要求. 在文献中, 研究者更加强调的是, 相较于经典的随机游走, pCN 算法 (第 3.1 小节) 的建议分布相较于经典随机游走有微小差别, 进而经典的算法只需要经过较小的改动就可以获得离散不变的性质. 需着重指出, pCN 算法构建于无限维空间, 这要求其先验测度必须在无限维空间中具有明确数学意义, 不可随意选定. 因此, 当从有限维空间视角探讨 pCN 算法时^[37], 还需强调先验 Gauss 分布的方差矩阵不仅要满足正定、对称条件, 其特征值也需要满足特定的衰减性条件 (具体条件可参见文献 [55, 假设 2]).

在图 3 中, 针对稳态 Darcy 流问题, 我们展示了不同算法在离散维数为 $\{50^2, 150^2, 300^2, 450^2\}$ 时的行为. 图中的横坐标是 (3.12) 中的参数 β , 其大致可理解为步长, β 越小, MCMC 抽样中前后的样本就越相似. 图中的纵坐标表示 MCMC 算法的接收率. 这里展示的计算结果是基于第 2.3 小节介绍的离散方法, 利用有限元开源计算软件 FEniCSx^[16, 157] 实现的, 程序可以在网址 <https://github.com/jjx323/IllustrateDimIndependenceMCMC/tree/main> 上下载. 图 3(a) 展示了经典的随机游走算法^[48] 的结果. 可以看到, 当 β 增加时, 接收率会下降, 而且接收率下降的速率与离散维数相关, 离散维数越大, 接受率在 β 相同时越小. 图 3(b) 展示的是第 3.1 小节所介绍的 pCN 算法 (采用了符合无限维理论的先验测度 μ_0^1) 的结果. 可以看到, 不论离散维数如何变化, β 与接受率的变化关系都一致. 图 3(c)

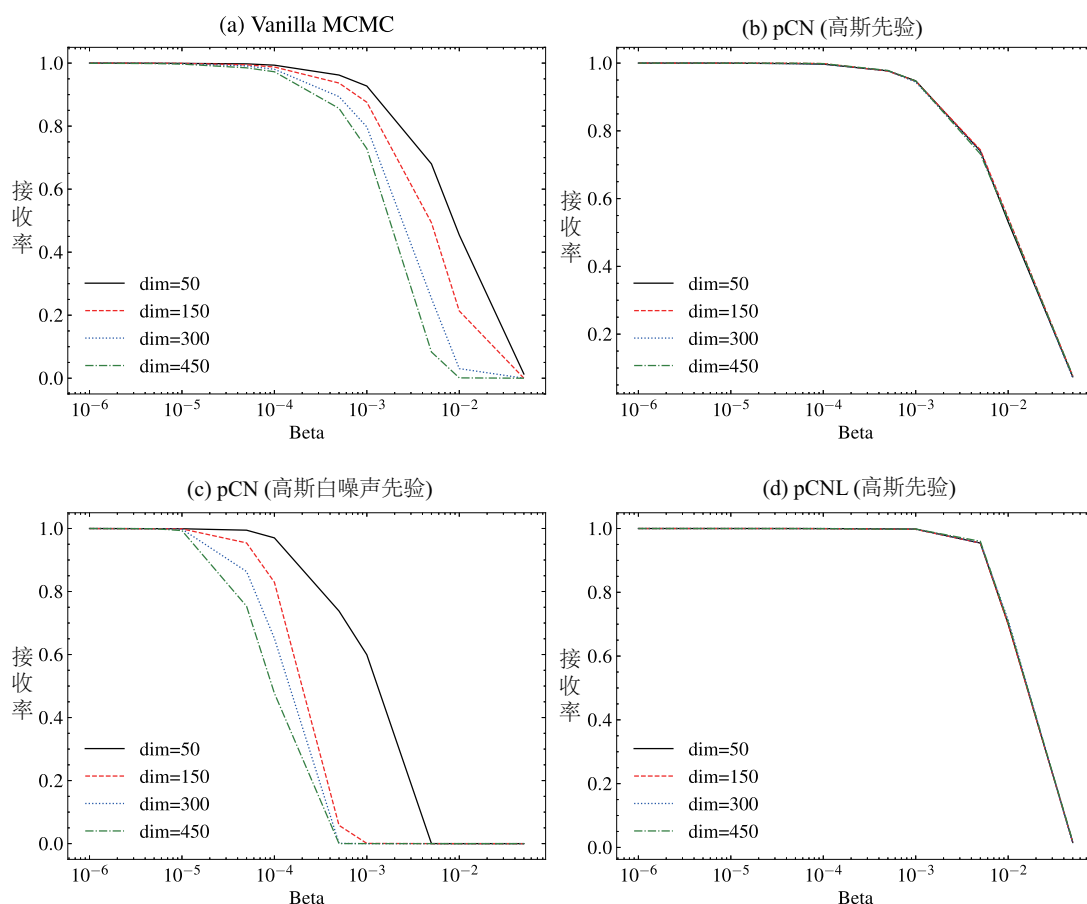


图 3 (网络版彩图) 不同算法中, 步长与接受率之间的关系. (a) 经典的基于随机游走的 MCMC 算法 (Gauss 先验 $\mathcal{N}(0, (0.5I - 0.1\Delta)^{-2})$); (b) 带有 Gauss 先验 $\mathcal{N}(0, (0.5I - 0.1\Delta)^{-2})$ 的 pCN 算法 (先验测度符合理论要求); (c) 带有先验测度 $\mathcal{N}(0, I)$ 的 pCN 算法 (先验测度不符合理论要求); (d) 基于“先 Bayes、再离散”的思路 (梯度通过对偶方法计算) 得到的 pCNL 算法 (Gauss 先验 $\mathcal{N}(0, (0.5I - 0.1\Delta)^{-2})$)

展示的是第 3.1 小节所介绍的 pCN 算法 (采用了不符合无限维理论的先验测度 μ_0^2) 的结果. 可以看到, β 与接受率的变化关系并不一致, 离散维数越大, 算法的抽样效率越低. 图 3(d) 展示的是第 3.1 小节所介绍的 pCNL 算法 (采用了符合无限维理论的先验测度 μ_0^1) 的结果, 可以看到, 不论离散维数如何变化, β 与接受率的变化关系都一致, 即抽样算法的抽样效率关于离散维数是不变的.

图 3(d) 展示了 pCNL 算法的结果. 在 pCNL 算法中, 我们需要计算梯度信息. 这时类似于“先优化, 后离散”的研究, 我们需要用对偶方法 (adjoint method) 计算导数信息才能得到图中所展示的不同离散维数下, 参数 β 与接受率一致的变化关系 (一致的抽样效率). 在简单的常微分方程的例子中, 可以观察到“先优化、后离散”与“先离散、后优化”通常会给出不同的算法格式, 关于 pCNL 等需要计算梯度信息的抽样算法同样会有“先优化、后离散”思路下的优点与缺点, 感兴趣的读者可以进一步参见文献 [79, 第 496–497 页] 和 [33].

在无限维空间上的变分推断系列研究中, 文献 [107, 113, 115, 198] 在数值实验中同样验证了算法的离散不变性. 文献 [198, 图 7] 对比了无限维空间中构造的基于平均场假设的变分推断算法与基于“先离散、后 Bayes”思路给出的变分推断算法 [117], 同样观察到了“先 Bayes、后离散”所带来的离散不变性. 文献 [113] 将无限维变分推断理论与 Fourier 神经算子 [135, 152] 学习相结合, 进而构造了具有离

散不变性的基于机器学习的快速反演方法.

4 统计理论

广泛的应用研究表明, 无限维 Bayes 反演方法是对反问题进行不确定性分析的有效工具, 对其的深入理解将有助于我们设计更加有效的反演方法. 下面从一个特殊的情形入手, 说明统计理论保障研究的重要性. 参考文献 [131] 的理论结果, 考虑线性模型如下:

$$y = \mathcal{G}u + \frac{1}{\sqrt{n}}\eta, \quad (4.1)$$

其中 $n \in \mathbb{N}^+$, $\mathcal{G} : X \rightarrow Y$ 是有界线性自伴算子 (很多问题中可以假设是紧算子), 其余符号的意义与第 2.2 小节相同. 分别取先验和噪声的分布如下:

$$u \sim \mu_0 = \mathcal{N}(0, \mathcal{C}_0), \quad \eta \sim \mathcal{N}(0, I). \quad (4.2)$$

根据 Bayes 理论, 可知后验概率测度 $\mu^y = \mathcal{N}(u_p, \mathcal{C}_p)$, 其中

$$u_p = \mathcal{C}_0 \mathcal{G}(\mathcal{G} \mathcal{C}_0 \mathcal{G} + nI)^{-1}y, \quad \mathcal{C}_p = \mathcal{C}_0 - \mathcal{C}_0 \mathcal{G}(\mathcal{G} \mathcal{C}_0 \mathcal{G} + nI)^{-1} \mathcal{G} \mathcal{C}_0. \quad (4.3)$$

从上面的表达式, 显然有如下直观认识.

(1) 后验均值函数 u_p 的正则性 (光滑性) 依赖于先验方差算子 $\mathcal{C}_0$.

(2) 后验方差 $\mathcal{C}_p$ 蕴涵了偏离均值的程度, 反映了我们对均值是参数准确估计的信任程度. 算子 $\mathcal{C}_0 \mathcal{G}(\mathcal{G} \mathcal{C}_0 \mathcal{G} + nI)^{-1} \mathcal{G} \mathcal{C}_0$ 显然是非负定的, 因此相较于先验, 依据后验测度所给出的信息, 我们会更加相信后验均值给出了准确的估计.

有这两点直观说明后, 我们考虑一种情形: 真实参数 $u^\dagger \in L^2(D) \setminus H^1(D)$, 即函数参数不一阶可导 (正则性较低); 先验方差算子 $\mathcal{C}_0 = \alpha(I - \Delta)^{-5}$, 其中 $\alpha \in \mathbb{R}^+$; 空间 $Y = L^2(D)$. 在这种情形下, 由后验均值的表达式容易看出 $u_p \in H^{10}(D)$ (这里并不是严格说明, 因为 Δ 算子的边界选取, u_p 严谨来讲并不在 $H^{10}(D)$ 中而是在一个 Hilbert 尺度空间中, 但总而言之 u_p 是一个正则性较好的函数). 后验均值 u_p 正则性很高, 然而真值是一个一阶弱导数都不在 L^2 空间的函数, 显然后验均值不是一个可信的估计. 但如果选取 α 足够小, 则后验方差会告诉我们后验均值估计十分可信, 并且这一结论不会随着噪声强度的减小 ($n \rightarrow \infty$) 而改变. 这个简单的例子说明, 后验测度是否提供了准确的、有意义的不确定性分析有赖于先验测度的选取和真实参数的性质, 因而有必要发展无限维 Bayes 反演方法的统计理论, 为应用研究提供理论保障.

在无限维 (非参) Bayes 推断的研究中 (针对经典统计问题), 已经有丰富的理论成果 (参见文献 [186]), 这里仅聚焦于无限维 Bayes 反演方法 (针对反问题) 的统计理论, 简述其主要定理及近期发展. 我们以第 3.4 小节所介绍的 Darcy 流问题为例来介绍无限维 Bayes 统计理论的基本定理, 具体而言, 考虑

$$y_i = \mathcal{G}(u)(x_i) + \eta_i, \quad (4.4)$$

其中 $i = 1, 2, \dots, N$. 非线性算子 $\mathcal{G}$ 表示方程 (3.40) 所确定的解算子, 其将参数 u 映射到解 w 在区域 D 内部的稀疏测量

$$w(x_i) = \mathcal{G}(u)(x_i), \quad x_i \in D.$$

在统计理论研究中, 考虑随机测量, 即 x_i 是空间 D 上随机采样得到, 记 (x_i, y_i) 所服从的分布是 P_u , 进而有 $D_N = \{(x_i, y_i)\}_{i=1}^N \sim P_u^N$. 由于我们考虑了随机测量, 所以这里更换前文中所使用的后验测度的记号 μ^y 为 μ^{D_N} . 关于随机测量情形下的 Bayes 公式其严谨的数学论证参见文献 [81, 168].

选取 $\beta > 1 + d/2$, 令 $\mathcal{R}$ 是 L^2 空间的子空间且满足 $\|\cdot\|_{H^\beta} \leq C\|\cdot\|_{\mathcal{R}}$. 定义 $B_{\mathcal{R}}(M) = \{u \in \mathcal{R} : \|u\|_{\mathcal{R}} \leq M\}$, 选取参数空间 $\mathcal{U} \subset H^\beta(D)$. 容易证明, Darcy 流问题的正算子 $\mathcal{G}$ 满足有界性与局部 Lipschitz 连续性:

$$\sup_{u \in \mathcal{U} \cap B_{\mathcal{R}}(M)} \sup_{x \in D} |\mathcal{G}(u)(x)| \leq C_1 \|f\|_{L^\infty(D)} < +\infty,$$

$$\|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_{L^2(D)} \leq C_2 \|u_1 - u_2\|_{H^{-1}(D)}, \quad \forall u_1, u_2 \in \mathcal{U} \cap B_{\mathcal{R}}(M),$$

其中 C_1 和 C_2 是与 u, u_1 和 u_2 无关的常数. 进一步地, 在对 f 和 $\mathcal{U}$ 以及方程的解 w 给出合适的假设条件后, 可以得到如下条件稳定性估计:

$$\|u_1 - u_2\|_{L^2(D)} \leq C \|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_{L^2}^\eta, \quad \eta = \frac{\beta - 1}{\beta + 1}.$$

粗略来讲, 在正算子满足局部有界、局部 Lipschitz 连续性条件及条件稳定性后, 如果假设先验分布是一定尺度变换后的 Gauss 概率测度, 先验 Gauss 概率测度的再生核 Hilbert 空间 $\mathcal{H} \subset H^\alpha(D)$, 其中 $\alpha > \beta + d/2$, 且真实参数 $u^\dagger \in \mathcal{H}$, 则当 $N \rightarrow \infty$ 时, 可以得到如下估计:

$$P_{u^\dagger}^N(\mu^{D_N}(\{u : \|\mathcal{G}(u) - \mathcal{G}(u^\dagger)\|_{L^2} \leq m\delta_N, \|u\|_{H^\beta} \leq m\}) \leq 1 - e^{-bN\delta_N^2}) = o(1), \quad (4.5)$$

其中 $b > 0, m = m(b)$ 是足够大的常数, 参数 $\delta_N = N^{-(\alpha+1)/(2\alpha+2+d)}$, $u^\dagger$ 表示真实参数. 进一步地, 对于足够大的常数 $M > 0$, 利用条件稳定性估计可以得到

$$P_{u^\dagger}^N(\mu^{D_N}(\{u : \|u - u^\dagger\|_{L^2} > M\delta_N^\eta\}) \geq e^{-bN\delta_N^2}) = o(1). \quad (4.6)$$

上式意味着有如下关于后验均值 u_p 的估计:

$$\|u_p - u^\dagger\|_{L^2} = O_{P_{u^\dagger}^N}(\delta_N^\eta), \quad (4.7)$$

其中 $O_{P_{u^\dagger}^N}(\delta_N^\eta)$ 表示依概率 δ_N^η 的速度收敛 [207]. 估计 (4.5) 表明, 当数据无限增多 $N \rightarrow \infty$ 时, 预测误差 $(\mathcal{G}(u) - \mathcal{G}(u^\dagger))$ 的误差) 按照速率 δ_N 趋向于 0. 类似地, 估计 (4.6) 表明, 当数据无限增多 $N \rightarrow \infty$ 时, 后验概率测度按照速率 δ_N^η 集中在真值附近. 一般而言, 称估计 (4.6) 为后验收缩率估计, 满足后验收缩率估计则称 Bayes 后验测度是相合的 (consistency). 需要说明, 上述后验收缩率估计考虑了数据无限增多时后验概率测度的极限性质. 也可如同问题 (4.1), 针对函数测量数据展开研究, 分析噪声水平趋于 0 时后验概率测度的极限性质, 进而得到后验收缩率估计.

在上一段中, 我们以 Darcy 流渗流系数反演这一非线性反问题为例说明了什么是后验收缩率估计. 事实上, 关于后验收缩率估计的研究是从 Gauss 先验和 Gauss 噪声假设下的线性问题开始的, 因为在这一情形下, 可以计算出后验概率测度的具体表达式 (4.3). 基于后验概率测度的具体表达式, 可以进行偏差方差分析, 从而得到后验收缩率估计. 基于这一思路, 文献 [131] 首次对线性反问题, 在先验方差算子与正算子可以同时对角化的假设下进行了详细的分析, 给出了后验收缩率的最优估计, 并且更进一步对后验置信域进行了详细的分析, 指出在极限理论中, 相较于真值过光滑的先验会导致难以给出合适的置信域估计, 欠光滑的先验会给出保守的置信域估计. 随后, 文献 [131] 中的方法被推广

用来研究严重不适定问题^[8]和逆向热传导问题^[132]. 对于偏微分方程反问题, 先验方差算子和正演算子可同时对角化是一个比较强的假设条件, 文献 [6] 引入了偏微分方程的理论分析方法, 进而在不要求同时对角化的条件下给出了几乎最优的收敛速率估计. 针对正问题含有超椭圆算子的情形, 文献 [126] 扩展了文献 [125] 中的分析方法, 引入了拟微分算子等工具, 给出了后验收缩率估计和置信域分析. 不同于这些研究, 文献 [153] 给出了线性反问题的 Oracle 型后验收缩率估计. 最近, 文献 [39] 通过运用扇形算子等工具, 克服了缺失偏差方差分解的困难, 在可分 Banach 空间上给出了后验收缩率估计. 在以上的研究中, 先验概率测度的选取都不依赖于数据, 但为了得到好的收缩率估计, 通常需要先验概率测度与真值之间的正则性匹配. 为了处理这一问题, 在先验方差算子、正算子和噪声方差算子可同时对角化的假设下, 文献 [130, 200] 详细地研究了经验 Bayes 方法的后验收缩率估计, 详细分析了 Bayes 置信区域与频率置信区域之间的关系. 随后, 文献 [109] 在不可同时对角化的假设下给出了经验 Bayes 方法的后验收缩率估计, 从而扩展了理论的适用范围. 文献 [129] 引入了连续模, 将预测误差的收缩率估计转换为了关于参数的后验收缩率估计, 进而避免了引入正算子的奇异值分解, 从而适用于更为广泛的问题.

据我们所知, 针对非线性反问题后验收缩率估计的最早研究是 2013 年的文献 [209]. 该文献将后验收缩率的估计分为了两个步骤: 第一步, 估计回归问题 (估计函数 $\mathcal{G}(u)(\cdot)$) 的后验收缩率估计; 第二步, 借助反问题的条件稳定性将回归函数 $\mathcal{G}(u)(\cdot)$ 的后验收缩率估计转换为函数参数 u 的后验收缩率估计. 这一证明思路的一个关键点在于, 如何将函数参数 u 的先验概率测度性质转换为回归函数 $\mathcal{G}(u)(\cdot)$ 的先验概率测度的性质, 进而可以方便地应用非参 Bayes 已有的研究结果 (参见文献 [186]). 文献 [209] 提出了这一重要的思路, 给出了初步的探索, 但限于该文献中的证明技术, 该文献结论应用范围有限, 例如, 无法应用于稳态 Darcy 流 Gauss 先验假设的情形. 自 2013 年的文献 [209], 针对非线性反问题后验收缩率估计的研究十分有限, 直到 2019 年才有了系列重要新结果. 针对 X-ray 变换^[164, 166]、Schödinger 方程反位势函数^[167, 169]、Darcy 流反演问题^[82, 165, 169]和 Caldéron 问题^[3, 28], 研究者引入了假设检验、经验过程估计等重要的非参 Bayes 后验分析理论, 进而得到了更广泛的先验概率测度假设下的系列收缩率估计.

除了后验收缩率估计, 另一个重要的无限维 Bayes 反演统计理论是 Bernstein-von Mises (BvM) 定理. 引入记号 $\mathbb{I}_u[h] := D\mathcal{G}_u[h]$, 这里 $D\mathcal{G}_u$ 表示非线性正演算子在 u 处的 Fréchet 导数. 这里用记号 $\mathbb{I}_u$ 的原因在于, $\mathbb{I}_u^* \mathbb{I}_u$ 通常称为信息算子, 对应了统计学中的 Fisher 信息 (information) 矩阵的概念. BvM 定理给出了类似如下收敛性结论:

$$\sqrt{N} \langle u - \bar{u}_N, \psi \rangle_{L^2} \mid D_N \xrightarrow{d} N(0, \|\mathbb{I}_{u^\dagger} \bar{\psi}_{u^\dagger}\|_{L^2}^2), \quad \text{按照 } P_{u^\dagger}^N \text{ 概率}, \quad (4.8)$$

其中 $\bar{u}_N$ 表示后验均值估计, $\bar{\psi}_{u^\dagger}$ 满足 $\mathbb{I}_{u^\dagger}^* \mathbb{I}_{u^\dagger} \bar{\psi}_{u^\dagger} = \psi$ (通常称这一方程为信息方程). 上述收敛性是指按照概率 $P_{u^\dagger}^N$ 依分布收敛. 具体而言, 如果随机变量 $Z_N \sim \mu_N$, $Z \sim \mu$ 且 $\mu_N = \mu_N \mid D_N$ (概率测度 μ_N 依赖于随机变量 $D_N \sim P_{u^\dagger}^N$), 则称 Z_N 按照概率 $P_{u^\dagger}^N$ 依分布收敛到 Z , 是指

$$d_{\text{weak}}(\mu_N, \mu) \xrightarrow{P_{u^\dagger}^N} 0, \quad N \rightarrow 0,$$

其中 d_{weak} 是概率测度的弱收敛推导出的距离. 事实上, 可以证明, 如果要在真值 $u^\dagger$ 附近估计 $\langle u, \psi \rangle_{L^2}$, 则上述极限分布的方差 $\|\mathbb{I}_{u^\dagger} \bar{\psi}_{u^\dagger}\|_{L^2}^2$ 达到了 Cramér-Rao 下界 (逆 Fisher 信息), 即给出了最小的极限方差. 简洁地, BvM 定理表明了, 在测量数据无限增多时, 后验概率分布在极限的意义下可以被一个 Gauss 概率分布逼近, 这个 Gauss 概率分布的均值是后验均值, 方差是逆信息算子. 近期发表的专

著 [168] 对 Bayes 反演中的 BvM 定理作了基本介绍. BvM 定理是否成立与信息方程 $\mathbb{I}_{u^\dagger}^* \mathbb{I}_{u^\dagger} \bar{\psi}_{u^\dagger} = \psi$ 的可解性关系紧密, 对于上文所述的 Darcy 流问题, 容易得知

$$\mathbb{I}_u[h] = -\mathfrak{L}_u^{-1}[\nabla \cdot (e^u h \nabla w_u)], \quad \mathbb{I}_u^*[g] = e^u \nabla w_u \cdot \nabla \mathfrak{L}_u^{-1}[g],$$

其中 w_u 是参数为 u 时方程的解, $\mathfrak{L}_u := -\nabla \cdot (e^u \nabla \cdot)$. 文献 [165, 168] 构造了反例, 说明算子 $\mathbb{I}_u$ 可能不是单射、也可能没有闭值域, 从而对于光滑的函数 ψ 信息方程难以求解, 即 BvM 定理对于 Darcy 流方程反演渗流系数的问题是不成立的. 但对于 Schrödinger 方程反演位势函数等问题 (参见文献 [165, 167, 168]), 可以在很一般的条件下证明 BvM 定理成立, 从而其后验概率分布在渐近的意义下可以被 Gauss 概率分布近似, 进而在渐近的意义下 Bayes 置信区域可借由 Gauss 分布进行刻画.

从关于后验收缩率估计、BvM 定理的简单回顾中可以看到, 针对非线性反问题的探讨依赖于经验过程、假设检验等统计理论, 同时在分析中需要推导新的条件稳定性估计、分析信息方程 (某些偏微分方程) 的性质, 这些都依赖于对具体非线性问题的细致分析, 因而类似于非线性偏微分方程, 我们难以期待如同线性问题一样构建统一的理论体系. 除了统计反问题中经常考虑的 Darcy 流反演、逆时扩散等问题, 针对多维扩散问题的漂移向量场反演^[170]、Caldéron 问题^[28] 和反散射问题^[76], 近期也出现了有关后验收缩率估计和 BvM 定理的研究.

参考文献

- 1 Abdar M, Pourpanah F, Hussain S, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf Fusion*, 2020, 76: 243–297
- 2 Abraham K, Deo N. Deep Gaussian process priors for Bayesian inference in nonlinear inverse problems. *arXiv: 2312.14294*, 2023
- 3 Abraham K, Nickl R. On statistical Calderón problems. *Math Stat Learn*, 2020, 2: 165–216
- 4 Agapiou S, Bardsley J M, Papaspiliopoulos O, et al. Analysis of the Gibbs sampler for hierarchical inverse problems. *SIAM/ASA J Uncertain Quantif*, 2014, 2: 511–544
- 5 Agapiou S, Burger M, Dashti M, et al. Sparsity-promoting and edge-preserving maximum *a posteriori* estimators in non-parametric Bayesian inverse problems. *Inverse Problems*, 2018, 34: 045002
- 6 Agapiou S, Larsson S, Stuart A M. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Process Appl*, 2013, 123: 3828–3860
- 7 Agapiou S, Papaspiliopoulos O, Sanz-Alonso D, et al. Importance sampling: Intrinsic dimension and computational cost. *Statist Sci*, 2017, 32: 405–431
- 8 Agapiou S, Stuart A M, Zhang Y X. Bayesian posterior contraction rates for linear severely ill-posed inverse problems. *J Inverse Ill-Posed Probl*, 2014, 22: 297–321
- 9 Al-Ghaddas O, Sanz-Alonso D. Non-asymptotic analysis of ensemble Kalman updates: Effective dimension and localization. *Inf Inference-J IMA*, 2024, 13: iaad043
- 10 Alexanderian A. Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review. *Inverse Probl*, 2021, 37: 043001
- 11 Aliprantis C D, Border K C. *Infinite Dimensional Analysis*. Berlin: Springer, 2006
- 12 Arridge S, Maass P, Öktem O, et al. Solving inverse problems using data-driven models. *Acta Numer*, 2019, 28: 1–174
- 13 Asch M, Bocquet M, Nodet M. *Data Assimilation: Methods, Algorithms, and Applications*. Philadelphia: SIAM, 2016
- 14 Ayanbayev B, Klebanov I, Lie H C, et al. Γ -convergence of Onsager-Machlup functionals: I. With applications to maximum *a posteriori* estimation in Bayesian inverse problems. *Inverse Problems*, 2022, 38: 025005
- 15 Ayanbayev B, Klebanov I, Lie H C, et al. Γ -convergence of Onsager-Machlup functionals: II. Infinite product measures on Banach spaces. *Inverse Problems*, 2022, 38: 025006
- 16 Baratta I A, Dean J P, Dokken J S, et al. DOLFINx: The next generation FEniCS problem solving environment. <https://orbilu.uni.lu/handle/10993/59327>, 2023

- 17 Benning M, Burger M. Modern regularization methods for inverse problems. *Acta Numer*, 2018, 27: 1–111
- 18 Bernardo J, Berger J, Dawid A, et al. Regression and classification using Gaussian process priors. *Bayesian Statist*, 1998, 6: 475–502
- 19 Beskos A, Girolami M, Lan S, et al. Geometric MCMC for infinite-dimensional inverse problems. *J Comput Phys*, 2017, 335: 327–351
- 20 Beskos A, Pinski F J, Sanz-Serna J M, et al. Hybrid Monte Carlo on Hilbert spaces. *Stochastic Process Appl*, 2011, 121: 2201–2230
- 21 Beskos A, Roberts G, Stuart A, et al. MCMC methods for diffusion bridges. *Stoch Dyn*, 2008, 8: 319–350
- 22 Bishop C M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006
- 23 Blei D M, Kucukelbir A, McAuliffe J D. Variational inference: A review for statisticians. *J Amer Statist Assoc*, 2017, 112: 859–877
- 24 Blömker D, Schillings C, Wacker P, et al. Well posedness and convergence analysis of the ensemble Kalman inversion. *Inverse Problems*, 2019, 35: 085007
- 25 Bogachev V I. *Gaussian Measures*. Providence: Amer Math Soc, 1998
- 26 Bogachev V I. *Differentiable Measures and the Malliavin Calculus*. Providence: Amer Math Soc, 2010
- 27 Bogachev V I. *Measure Theory*. Heidelberg-Berlin: Springer, 2017
- 28 Bohr J. A Bernstein-von-Mises theorem for the Calderón problem with piecewise constant conductivities. *Inverse Probl*, 2023, 39: 015002
- 29 Bohra P, Pham T A, Dong J, et al. Bayesian inversion for nonlinear imaging models using deep generative priors. *IEEE Trans Comput Imag*, 2022, 8: 1237–1249
- 30 Bui-Thanh T, Ghattas O. An analysis of infinite dimensional Bayesian inverse shape acoustic scattering and its numerical approximation. *SIAM/ASA J Uncertain Quantif*, 2014, 2: 203–222
- 31 Bui-Thanh T, Ghattas O. A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors. *Inverse Probl Imaging*, 2015, 9: 27–53
- 32 Bui-Thanh T, Ghattas O, Martin J, et al. A computational framework for infinite-dimensional Bayesian inverse problems part I: The linearized case, with application to global seismic inversion. *SIAM J Sci Comput*, 2013, 35: A2494–A2523
- 33 Bui-Thanh T, Nguyen Q P. FEM-based discretization-invariant MCMC methods for PDE-constrained Bayesian inverse problems. *Inverse Probl Imaging*, 2016, 10: 943–975
- 34 Burger M, Lucka F. Maximum *a posteriori* estimates in linear inverse problems with log-concave priors are proper Bayes estimators. *Inverse Problems*, 2014, 30: 114004
- 35 Calvello E, Reich S, Stuart A M. Ensemble Kalman methods: A mean field perspective. *arXiv:2209.11371*, 2022
- 36 Calvetti D, Somersalo E. *Introduction to Bayesian Scientific Computing*. New York: Springer, 2007
- 37 Calvetti D, Somersalo E. *Bayesian Scientific Computing*. Cham: Springer, 2023
- 38 Cen S, Jin B, Liu Y, et al. Recovery of multiple parameters in subdiffusion from one lateral boundary measurement. *Inverse Problems*, 2023, 39: 104001
- 39 Chen D H, Li J, Zhang Y. A posterior contraction for Bayesian inverse problems in Banach spaces. *Inverse Problems*, 2024, 40: 045011
- 40 Chen J, Anitescu M, Saad Y. Computing $f(A)b$ via least squares polynomial approximations. *SIAM J Sci Comput*, 2011, 33: 195–222
- 41 Cheng J, Liu J J, Zhang B. Inverse problems for PDEs: Models, computations and applications (in Chinese). *Sci Sin Math*, 2019, 49: 643–666 [程晋, 刘继军, 张波. 偏微分方程反问题: 模型、算法和应用. 中国科学: 数学, 2019, 49: 643–666]
- 42 Cheng S H. *Foundations of Measure Theory and Probability Theory* (in Chinese). Beijing: Peking Univ Press, 2004 [程士宏. 测度论与概率论基础. 北京: 北京大学出版社, 2004]
- 43 Cohn S, Ghil M, Isaacson E. Applications of estimation theory to numerical weather prediction. In: *Dynamic Meteorology: Data Assimilation Methods*. New York: Springer, 1981, 36: 139–224
- 44 Conrad P R, Davis A D, Marzouk Y M, et al. Parallel local approximation MCMC for expensive models. *SIAM/ASA J Uncertain Quantif*, 2018, 6: 339–373
- 45 Cotter S L, Dashti M, Robinson J C, et al. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 2009, 25: 115008
- 46 Cotter S L, Dashti M, Stuart A M. Approximation of Bayesian inverse problems for PDEs. *SIAM J Numer Anal*,

- 2010, 48: 322–345
- 47 Cotter S L, Dashti M, Stuart A M. Variational data assimilation using targetted random walks. *Internat J Numer Methods Fluids*, 2012, 68: 403–421
- 48 Cotter S L, Roberts G O, Stuart A M, et al. MCMC methods for functions: Modifying old algorithms to make them faster. *Statist Sci*, 2013, 28: 424–446
- 49 Cucker F, Smale S. On the mathematical foundations of learning. *Bull Amer Math Soc (NS)*, 2002, 39: 1–49
- 50 Cui T, Law K J H, Marzouk Y M. Dimension-independent likelihood-informed MCMC. *J Comput Phys*, 2016, 304: 109–137
- 51 Da Prato G. *Introduction to Stochastic Analysis and Malliavin Calculus*. Pisa: Edizioni della Normale, 2014
- 52 Dashti M, Harris S, Stuart A M. Besov priors for Bayesian inverse problems. *Inverse Probl Imag*, 2012, 6: 183–200
- 53 Dashti M, Law K J H, Stuart A M, et al. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 2013, 29: 095017
- 54 Dashti M, Stuart A M. Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J Numer Anal*, 2011, 49: 2524–2542
- 55 Dashti M, Stuart A M. The Bayesian approach to inverse problems. In: *Handbook of Uncertainty Quantification*. Cham: Springer, 2017, 311–428
- 56 De los Reyes J C. *Numerical PDE-Constrained Optimization*. Cham: Springer, 2015
- 57 Detommaso G, Cui T, Marzouk Y, et al. A Stein variational Newton method. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates, https://proceedings.neurips.cc/paper_files/paper/2018/hash/fdaa09fc5ed18d3226b3a1a00f1bc48c-Abstract.html, 2018
- 58 Ding L, Lu S, Cheng J. Weak-norm posterior contraction rate of the 4DVAR method for linear severely ill-posed problems. *J Complexity*, 2018, 46: 1–18
- 59 Ding M H, Zheng G H. Determination of the reaction coefficient in a time dependent nonlocal diffusion process. *Inverse Problems*, 2021, 37: 025005
- 60 Dudley R M. *Real Analysis and Probability*. Cambridge: Cambridge Univ Press, 2002
- 61 Dunlop M M, Girolami M A, Stuart A M, et al. How deep are deep Gaussian processes? *J Mach Learn Res*, 2018, 19: 54, 46
- 62 Dunlop M M, Helin T, Stuart A M. Hyperparameter estimation in Bayesian MAP estimation: Parameterizations and consistency. *SMAI J Comput Math*, 2020, 6: 69–100
- 63 Dunlop M M, Iglesias M A, Stuart A M. Hierarchical Bayesian level set inversion. *Stat Comput*, 2017, 27: 1555–1584
- 64 Dunlop M M, Stuart A M. MAP estimators for piecewise continuous inversion. *Inverse Problems*, 2016, 32: 105003
- 65 Duong D L. *Inverse problems for hyperbolic conservation laws: A Bayesian approach*. PhD Thesis. Brighton: Univ of Sussex, 2021
- 66 E W N. The dawning of a new era in applied mathematics. *Notices Amer Math Soc*, 2021, 68: 565–571
- 67 Engel M, Kanjilal O, Papaioannou I, et al. Bayesian updating and marginal likelihood estimation by cross entropy based importance sampling. *J Comput Phys*, 2023, 473: 111746
- 68 Engel S, Hafemeyer D, Münch C, et al. An application of sparse measure valued Bayesian inversion to acoustic sound source identification. *Inverse Problems*, 2019, 35: 075005
- 69 Engl H W, Hanke M, Neubauer A. *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic Publ, 1996
- 70 Ernst O G, Sprungk B, Starkloff H J. Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems. *SIAM/ASA J Uncertain Quantif*, 2015, 3: 823–851
- 71 Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res*, 1995, 99: 10143–10162
- 72 Evensen G. *Data Assimilation: The Ensemble Kalman Filter*. Berlin: Springer-Verlag, 2009
- 73 Evensen G, van Leeuwen P J. Assimilation of geosat altimeter data for the agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon Wea Rev*, 1996, 124: 85–96
- 74 Fichtner A. *Full Seismic Waveform Modelling and Inversion*. Berlin-Heidelberg: Springer-Verlag, 2011
- 75 Franklin J N. Well-posed stochastic extensions of ill-posed linear problems. *J Math Anal Appl*, 1970, 31: 682–716
- 76 Furuya T, Kow P Z, Wang J N. Consistency of the Bayes method for the inverse scattering problem. *Inverse Problems*, 2024, 40: 055001
- 77 Garbuno-Inigo A, Hoffmann F, Li W, et al. Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler. *SIAM J Appl Dyn Syst*, 2020, 19: 412–441

- 78 Garbuno-Inigo A, Nüsken N, Reich S. Affine invariant interacting Langevin dynamics for Bayesian inference. *SIAM J Appl Dyn Syst*, 2020, 19: 1633–1658
- 79 Ghattas O, Willcox K. Learning physics-based models from data: Perspectives from inverse problems and model reduction. *Acta Numer*, 2021, 30: 445–554
- 80 Ghosal S, Vaart A V D. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge: Cambridge Univ Press, 2017
- 81 Giné E, Nickl R. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. New York: Cambridge Univ Press, 2016
- 82 Giordano M, Nickl R. Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Problems*, 2020, 36: 085001
- 83 Gottwald G A, Majda A J. A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlinear Processes Geophys*, 2013, 20: 705–712
- 84 Grathwohl W, Wang K C, Jacobsen J H, et al. Learning the Stein discrepancy for training and evaluating energy-based models without sampling. In: *Proceedings of the 37th International Conference on Machine Learning*. International Conference on Machine, vol. 119. Ann Arbor: PMLR, 2020, 3732–3747
- 85 Guha N, Wu X, Efendiev Y, et al. A variational Bayesian approach for inverse problems with skew-t error distributions. *J Comput Phys*, 2015, 301: 377–393
- 86 Hadamard J. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. New Haven: Yale Univ Press, 1923
- 87 Hairer M, Stuart A M, Vollmer S J. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann Appl Probab*, 2014, 24: 2455–2490
- 88 Hairer M, Stuart A M, Voss J. Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann Appl Probab*, 2007, 17: 1657–1706
- 89 Hairer M, Stuart A M, Voss J, et al. Analysis of SPDEs arising in path sampling. Part I: The Gaussian case. *Commun Math Sci*, 2005, 3: 587–603
- 90 Hegland M. Approximate maximum *a posteriori* with Gaussian process priors. *Constr Approx*, 2007, 26: 205–224
- 91 Helin T, Burger M. Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 2015, 31: 085009
- 92 Herrmann L, Keller M, Schwab C. Quasi-Monte Carlo Bayesian estimation under Besov priors in elliptic inverse problems. *Math Comp*, 2021, 90: 1831–1860
- 93 Hinze M, Pinnau R, Ulbrich M, et al. *Optimization with PDE Constraints*. New York: Springer, 2008
- 94 Hoang V H, Schwab C, Stuart A M. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 2013, 29: 085010
- 95 Holden M, Pereyra M, Zygalakis K C. Bayesian imaging with data-driven priors encoded by neural networks. *SIAM J Imag Sci*, 2022, 15: 892–924
- 96 Hosseini B, Johndrow J E. Spectral gaps and error estimates for infinite-dimensional Metropolis-Hastings with non-Gaussian priors. *Ann Appl Probab*, 2023, 33: 1827–1873
- 97 Hosseini B, Nigam N. Well-posed Bayesian inverse problems: Priors with exponential tails. *SIAM/ASA J Uncertain Quantif*, 2017, 5: 436–465
- 98 Houtekamer P L, Derome J. Methods for ensemble prediction. *Mon Wea Rev*, 1995, 1237: 2181–2196
- 99 Houtekamer P L, Mitchell H L. Data assimilation using an ensemble Kalman filter technique. *Mon Wea Rev*, 1998, 126: 796–811
- 100 Huang D Z, Huang J, Reich S, et al. Efficient derivative-free Bayesian inference for large-scale inverse problems. *Inverse Problems*, 2022, 38: 125006
- 101 Iglesias M A. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Probl*, 2016, 32: 025002
- 102 Iglesias M A, Law K J H, Stuart A M. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 2013, 29: 045001
- 103 Iglesias M A, Lin K, Stuart A M. Well-posed Bayesian geometric inverse problems arising in subsurface flow. *Inverse Problems*, 2014, 30: 114001
- 104 Iglesias M A, Lu Y, Stuart A M. A Bayesian level set method for geometric inverse problems. *Interfaces Free Bound*, 2016, 18: 181–217

- 105 Jaakkola T, Saul L K, Jordan M I. Fast learning by bounding likelihoods in sigmoid type belief networks. In: Advances in Neural Information Processing Systems. Cambridge: MIT Press, 1995, 528–534
- 106 Jazwinski A H. Stochastic Processes and Filtering Theory. New York: Dover, 2007
- 107 Jia J, Li P, Meng D. Stein variational gradient descent on infinite-dimensional space and applications to statistical inverse problems. SIAM J Numer Anal, 2022, 60: 2225–2252
- 108 Jia J, Peng J, Gao J. Bayesian approach to inverse problems for functions with a variable-index Besov prior. Inverse Problems, 2016, 32: 085006
- 109 Jia J, Peng J, Gao J. Posterior contraction for empirical Bayesian approach to inverse problems under non-diagonal assumption. Inverse Probl Imaging, 2021, 15: 201–228
- 110 Jia J, Peng J, Gao J, et al. Backward problem for a time-space fractional diffusion equation. Inverse Probl Imaging, 2018, 12: 773–799
- 111 Jia J, Peng J, Yang J. Harnack’s inequality for a space-time fractional diffusion equation and applications to an inverse source problem. J Differential Equations, 2017, 262: 4415–4450
- 112 Jia J, Wu B, Peng J, et al. Recursive linearization method for inverse medium scattering problems with complex mixture Gaussian error learning. Inverse Problems, 2019, 35: 075003
- 113 Jia J, Wu Y, Li P, et al. Variational inverting network for statistical inverse problems of partial differential equations. J Mach Learn Res, 2023, 24: 9597–9656
- 114 Jia J, Yue S, Peng J, et al. Infinite-dimensional Bayesian approach for inverse scattering problems of a fractional Helmholtz equation. J Funct Anal, 2018, 275: 2299–2332
- 115 Jia J, Zhao Q, Xu Z, et al. Variational Bayes’ method for functions with applications to some inverse problems. SIAM J Sci Comput, 2021, 43: A355–A383
- 116 Jin B. A variational Bayesian method to inverse problems with impulsive noise. J Comput Phys, 2012, 231: 423–435
- 117 Jin B, Zou J. Hierarchical Bayesian inference for Ill-posed problems via variational method. J Comput Phys, 2010, 229: 7317–7343
- 118 Jing X, Yamamoto M. Simultaneous uniqueness for multiple parameters identification in a fractional diffusion-wave equation. Inverse Probl Imaging, 2022, 16: 1199–1217
- 119 Johnson M J, Willsky A S. Stochastic variational inference for Bayesian time series models. In: Proceedings of the 31st International Conference on International Conference on Machine Learning. International Conference on Machine Learning, vol. 32. San Diego: JMLR, 2014, 1854–1862
- 120 Kahle C, Lam K F, Latz J, et al. Bayesian parameter identification in Cahn-Hilliard models for biological growth. SIAM/ASA J Uncertain Quantif, 2019, 7: 526–552
- 121 Kaipio J, Somersalo E. Statistical and Computational Inverse Problems. New York: Springer-Verlag, 2005 [Kaipio J, Somersalo E. 统计与计算反问题. 刘逸侃, 徐定华, 程晋, 译. 北京: 科学出版社, 2018]
- 122 Kalman R E. A new approach to linear filtering and prediction problems. J Basic Eng, 1960, 82: 35–45
- 123 Kalman R E, Bucy R S. New results in linear filtering and prediction theory. J Basic Eng, 1961, 83: 95–108
- 124 Kawakami H. Stabilities of shape identification inverse problems in a Bayesian framework. J Math Anal Appl, 2020, 486: 123903
- 125 Kekkonen H, Lassas M, Siltanen S. Analysis of regularized inversion of data corrupted by white Gaussian noise. Inverse Problems, 2014, 30: 045009
- 126 Kekkonen H, Lassas M, Siltanen S. Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators. Inverse Problems, 2016, 32: 085005
- 127 Kelly D T B, Law K J H, Stuart A M. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. Nonlinearity, 2014, 27: 2579–2603
- 128 Klebanov I, Wacker P. Maximum *a posteriori* estimators in ℓ^p are well-defined for diagonal Gaussian priors. Inverse Problems, 2023, 39: 065009
- 129 Knapik B, Salomond J B. A general approach to posterior contraction in nonparametric inverse problems. Bernoulli, 2018, 24: 2091–2121
- 130 Knapik B T, Szabó B T, van der Vaart A W, et al. Bayes procedures for adaptive inference in inverse problems for the white noise model. Probab Theory Related Fields, 2016, 164: 771–813
- 131 Knapik B T, van der Vaart A W, van Zanten J H. Bayesian inverse problems with Gaussian priors. Ann Statist, 2011, 39: 2626–2657
- 132 Knapik B T, van der Vaart A W, van Zanten J H. Bayesian recovery of the initial condition for the heat equation.

- Comm Statist Theory Methods, 2013, 42: 1294–1313
- 133 Kolehmainen V, Lassas M, Niinimäki K, et al. Sparsity-promoting Bayesian inversion. *Inverse Problems*, 2012, 28: 025005
- 134 Korba A, Salim A, Arbel M, et al. A non-asymptotic analysis for Stein variational gradient descent. In: *Proceedings of the 34th Conference on Neural Information Processing Systems*. Red Hook: Curran Associates, 2020, 4672–4682
- 135 Kovachki N, Lanthaler S, Mishra S. On universal approximation and error bounds for Fourier neural operators. *J Mach Learn Res*, 2021, 22: 1–76
- 136 Kovachki N, Li Z, Liu B, et al. Neural operator: Learning maps between function spaces with applications to PDEs. *J Mach Learn Res*, 2023, 24: 89, 97
- 137 Kovachki N B, Stuart A M. Ensemble Kalman inversion: A derivative-free technique for machine learning tasks. *Inverse Problems*, 2019, 35: 095005
- 138 Kretschmann R. Are minimizers of the Onsager-Machlup functional strong posterior modes? *SIAM/ASA J Uncertain Quantif*, 2023, 11: 1105–1138
- 139 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. La Jolla: NIPS, 2012, 1–9
- 140 Kukush A. *Gaussian Measures in Hilbert Space*. Great Britain: ISTE, 2019
- 141 Kwiatkowski E, Mandel J. Convergence of the square root ensemble Kalman filter in the large ensemble limit. *SIAM/ASA J Uncertain Quantif*, 2015, 3: 1–17
- 142 Lambley H, Sullivan T J. An order-theoretic perspective on modes and maximum *a posteriori* estimation in Bayesian inverse problems. *SIAM/ASA J Uncertain Quantif*, 2023, 11: 1195–1224
- 143 Lanthaler S, Mishra S, Weber F. On Bayesian data assimilation for PDEs with ill-posed forward problems. *Inverse Problems*, 2022, 38: 085012
- 144 Lassas M, Saksman E, Siltanen S. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl Imaging*, 2009, 3: 87–122
- 145 Lassas M, Samuli S. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems*, 2004, 20: 1537–1563
- 146 Latz J. On the well-posedness of Bayesian inverse problems. *SIAM/ASA J Uncertain Quantif*, 2020, 8: 451–482
- 147 Latz J. Bayesian inverse problems are usually well-posed. *SIAM Rev*, 2023, 65(3): 831–865
- 148 Laumont R, Bortoli V D, Almansa A, et al. Bayesian imaging using plug & play priors: When Langevin meets tweedie. *SIAM J Imag Sci*, 2022, 15: 701–737
- 149 Law K J H, Stuart A M, Zygalakis K C. *Data Assimilation: A Mathematical Introduction*. Cham: Springer, 2015
- 150 Le Gland F, Monbet V, Tran V D. Large sample asymptotics for the ensemble Kalman filter. PhD Thesis. Rennes: Institut National de Recherche en Informatique et en Automatique, 2009
- 151 Li Q, Chen L, Tai C, et al. Maximum principle based algorithms for deep learning. *J Mach Learn Res*, 2017, 18: 1–29
- 152 Li Z, Kovachki N B, Azizzadenesheli K, et al. Fourier neural operator for parametric partial differential equations. In: *International Conference on Learning Representations*. <https://openreview.net/forum?id=c8P9NQVtmnO>, 2021
- 153 Lin K, Lu S, Mathé P. Oracle-type posterior contraction rates in Bayesian inverse problems. *Inverse Probl Imaging*, 2015, 9: 895–915
- 154 Liu C, Zhuo J, Cheng P, et al. Understanding and accelerating particle-based variational inference. In: *Proceedings of the International Conference on Machine Learning*. Ann Arbor: PMLR, 2019, 4082–4092
- 155 Liu Q. Stein variational gradient descent as gradient flow. In: *Advances in Neural Information Processing Systems*. La Jolla: NIPS, 2017, 1–9
- 156 Liu Q, Wang D. Stein variational gradient descent: A general purpose Bayesian inference algorithm. In: *Advances in Neural Information Processing Systems*. <https://proceedings.neurips.cc/paper/2016/hash/b3ba8f1bee1238a2f37603d90b58898d-Abstract.html>, 2016
- 157 Logg A, Mardal K A, Wells G N. *Automated Solution of Differential Equations by the Finite Element Method*. Berlin-Heidelberg: Springer, 2012
- 158 Lu J, Lu Y, Nolen J. Scaling limit of the Stein variational gradient descent: The mean field regime. *SIAM J Math Anal*, 2019, 51: 648–671
- 159 Lu S, Niu P, Werner F. On the asymptotical regularization for linear inverse problems in presence of white noise. *SIAM/ASA J Uncertain Quantif*, 2021, 9: 1–28

- 160 Mandel J, Cobb L, Beezley J D. On the convergence of the ensemble Kalman filter. *Appl Math*, 2011, 56: 533–541
- 161 Markkanen M, Roininen L, Huttunen J M, et al. Cauchy difference priors for edge-preserving Bayesian inversion. *J Inverse Ill-Posed Probl*, 2019, 27: 225–240
- 162 Mattingly J C, Pillai N S, Stuart A M. Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann Appl Probab*, 2012, 22: 881–930
- 163 Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of Machine Learning*. Cambridge: MIT Press, 2012
- 164 Monard F, Nickl R, Paternain G P. Efficient nonparametric Bayesian inference for X-ray transforms. *Ann Statist*, 2019, 47: 1113–1147
- 165 Monard F, Nickl R, Paternain G P. Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. *Ann Statist*, 2021, 49: 3255–3298
- 166 Monard F, Nickl R, Paternain G P. Consistent inversion of noisy non-Abelian X-ray transforms. *Comm Pure Appl Math*, 2021, 74: 1045–1099
- 167 Nickl R. Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation. *J Eur Math Soc (JEMS)*, 2020, 22: 2697–2750
- 168 Nickl R. *Bayesian Non-Linear Statistical Inverse Problems*. Berlin: EMS Press, 2023
- 169 Nickl R, Paternain G P. On some information-theoretic aspects of non-linear statistical inverse problems. In: *ICM—International Congress of Mathematicians*, vol. VII. Berlin: EMS Press, 2023, 5516–5538
- 170 Nickl R, Ray K. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann Statist*, 2020, 48: 1383–1408
- 171 Oates C J, Cockayne J, Aykroyd R G, et al. Bayesian probabilistic numerical methods in time-dependent state estimation for industrial hydrocyclone equipment. *J Amer Statist Assoc*, 2019, 114: 1518–1531
- 172 Oppor M, Saad D. *Advanced Mean Field Methods: Theory and Practice*. Cambridge: MIT Press, 2001
- 173 Oppor M, Winther O. A mean field algorithm for Bayes learning in large feed-forward neural networks. In: *Proceedings of the 10th International Conference on Neural Information Processing Systems*. Denver: MIT Press, 1996, 225–231
- 174 Pavliotis G A, Stuart A M, Vaes U. Derivative-free Bayesian inversion using multiscale dynamics. *SIAM J Appl Dyn Syst*, 2022, 21: 284–326
- 175 Pereyra M. Revisiting maximum-a-posteriori estimation in log-concave models. *SIAM J Imaging Sci*, 2019, 12: 650–670
- 176 Peterson C, Anderson J R. A mean field theory learning algorithm for neural networks. *Complex Systems*, 1987, 1: 995–1019
- 177 Petra N, Martin J, Stadler G, et al. A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems. *SIAM J Sci Comput*, 2014, 36: A1525–A1555
- 178 Pillai N S, Stuart A M, Thiéry A H. Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann Appl Probab*, 2012, 22: 2320–2356
- 179 Pinski F J, Simpson G, Stuart A M, et al. Kullback-Leibler approximation for probability measures on infinite dimensional spaces. *SIAM J Math Anal*, 2015, 47: 4091–4122
- 180 Pinski F J, Simpson G, Stuart A M, et al. Algorithms for Kullback-Leibler approximation of probability measures in infinite dimensions. *SIAM J Sci Comput*, 2015, 37: A2733–A2757
- 181 Prato G D, Zabczyk J. *Stochastic Equations in Infinite Dimensions*. Cambridge: Cambridge Univ Press, 2014
- 182 Reed M, Simon B. *Functional Analysis I: Methods of Modern Mathematical Physics*. Singapore: Elsevier, 2003
- 183 Reich S, Cotter C. *Probabilistic Forecasting and Bayesian Data Assimilation*. New York: Cambridge Univ Press, 2015
- 184 Richter M. *Inverse Problems—Basics, Theory and Applications in Geophysics*. Cham: Birkhäuser/Springer, 2020
- 185 Rothfuss J, Fortuin V, Josifoski M, et al. PACOH: Bayes-optimal meta-learning with PAC-guarantees. In: *International Conference on Machine Learning*. Ann Arbor: PMLR, 2021, 9116–9126
- 186 Rousseau J. On the frequentist properties of bayesian nonparametric methods. *Annu Rev Stat Appl*, 2016, 3: 211–231
- 187 Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Phys D*, 1992, 60: 259–268
- 188 Rudolf D, Sprungk B. On a generalization of the preconditioned Crank-Nicolson metropolis algorithm. *Found Comput Math*, 2018, 18: 309–343
- 189 Sanz-Alonso D, Stuart A M, Taeb A. *Inverse Problems and Data Assimilation*. Cambridge: Cambridge Univ Press, 2023

- 190 Särkkä S, Svensson L. Bayesian Filtering and Smoothing. Cambridge: Cambridge Univ Press, 2013
- 191 Schillings C, Stuart A M. Analysis of the ensemble Kalman filter for inverse problems. *SIAM J Numer Anal*, 2017, 55: 1264–1290
- 192 Schillings C, Stuart A M. Convergence analysis of ensemble Kalman inversion: The linear, noisy case. *Appl Anal*, 2018, 97: 107–123
- 193 Schuster T, Kaltenbacher B, Hofmann B, et al. Regularization Methods in Banach Spaces. Berlin: Walter de Gruyter, 2012
- 194 Song X, Zheng G H, Jiang L. Variational Bayesian inversion for the reaction coefficient in space-time nonlocal diffusion equations. *Adv Comput Math*, 2021, 47: 31
- 195 Stuart A M. Inverse problems: A Bayesian perspective. *Acta Numer*, 2010, 19: 451–559
- 196 Stuart A M, Teckentrup A L. Posterior consistency for Gaussian process approximations of Bayesian posterior distributions. *Math Comp*, 2018, 87: 721–753
- 197 Stuart A M, Voss J, Wilberg P. Conditional path sampling of SDEs and the Langevin MCMC method. *Commun Math Sci*, 2004, 2: 685–697
- 198 Sui J, Jia J. Non-centered parametric variational Bayes' approach for hierarchical inverse problems of partial differential equations. *Math Comp*, 2024, 93: 1715–1760
- 199 Sun Z, Zheng G H. Solving linear Bayesian inverse problems using a fractional total variation-Gaussian (FTG) prior and transport map. *Comput Statist*, 2023, 38: 1811–1849
- 200 Szabó B, van der Vaart A W, van Zanten J H. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann Statist*, 2015, 43: 1391–1428
- 201 Tarantola A. Inverse Problem Theory. Amsterdam: Elsevier, 1987
- 202 Tarantola A. Inverse Problem Theory and Methods for Model Parameter Estimation. Philadelphia: SIAM, 2005
- 203 Teh Y W, Thiery A H, Vollmer S J. Consistency and fluctuations for stochastic gradient Langevin dynamics. *J Mach Learn Res*, 2016, 17: 193–225
- 204 Tikhonov A N, Arsenin V Y. Methods for Solving Ill-Posed Problems. Beijing: Geological Publishing House, 1979 [Tikhonov A N, Arsenin V Y. 不适定问题的解法. 王秉忱, 译. 北京: 地质出版社, 1979]
- 205 Tong X T, Majda A J, Kelly D. Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation. *Commun Math Sci*, 2015, 14: 1283–1313
- 206 Tong X T, Majda A J, Kelly D. Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity*, 2016, 29: 131–140
- 207 van der Vaart A W. Asymptotic Statistics. Cambridge: Cambridge Univ Press, 1998
- 208 Villa U, Petra N, Ghattas O. hIPPYlib: An extensible software framework for large-scale inverse problems. *JOSS*, 2018, 3: 940
- 209 Vollmer S J. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Probl*, 2013, 29: 125011
- 210 Vollmer S J. Dimension-independent MCMC sampling for inverse problems with non-Gaussian priors. *SIAM/ASA J Uncertain Quantif*, 2015, 3: 535–561
- 211 Wacker P, Knabner P. Wavelet-based priors accelerate maximum-*a-posteriori* optimization in Bayesian inverse problems. *Methodol Comput Appl Probab*, 2020, 22: 853–879
- 212 Wang K, Bui-Thanh T, Ghattas O. A randomized maximum *a posteriori* method for posterior sampling of high dimensional nonlinear Bayesian inverse problems. *SIAM J Sci Comput*, 2018, 40: A142–A171
- 213 Wang Z, Ren T, Zhu J, et al. Function space particle optimization for Bayesian neural networks. In: Proceedings of the International Conference on Learning Representations. <https://openreview.net/forum?id=BkgtDsCcKQ>, 2019, 1–24
- 214 Yan L, Shin Y, Xiu D. Sparse approximation using $\ell_1 - \ell_2$ minimization and its application to stochastic collocation. *SIAM J Sci Comput*, 2017, 39: A214–A239
- 215 Yang K, Guha N, Efendiev Y, et al. Bayesian and variational Bayesian approaches for flows in heterogeneous random media. *J Comput Phys*, 2017, 345: 275–293
- 216 Yao Z, Hu Z, Li J. A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementations. *Inverse Problems*, 2016, 32: 075006
- 217 Yue Z, Yong H, Zhao Q, et al. Variational denoising network: Toward blind noise modeling and removal. In: Advances in Neural Information Processing Systems. <https://proceedings.neurips.cc/paper/2019/hash/6395ebd0f4b>

- 478145ecfbaf939454fa4-Abstract.html, 2019
- 218 Yue Z, Yong H, Zhao Q, et al. Deep variational network toward blind image restoration. *IEEE Trans Pattern Anal Mach Intell*, 2024, 46: 7011–7026
- 219 Zhang C, Bütetpage J, Kjellström H, et al. Advances in variational inference. *IEEE Trans Pattern Anal Mach Intell*, 2017, 41: 2008–2026
- 220 Zhang X, Nawaz M A, Zhao X, et al. An introduction to variational inference in geophysical inverse problems. *Adv Geophys*, 2021, 62: 73–140
- 221 Zhang Y X, Jia J, Yan L. Bayesian approach to a nonlinear inverse problem for a time-space fractional diffusion equation. *Inverse Problems*, 2018, 34: 125002
- 222 Zhao L C, Wang Z F. Introduction to Advanced Statistics (in Chinese). Beijing: Higher Ed Press, 2016 [赵林成, 王占峰. 高等统计学概论. 北京: 高等教育出版社, 2016]
- 223 Zhou Q, Yu T, Zhang X, et al. Bayesian inference and uncertainty quantification for medical image reconstruction with Poisson data. *SIAM J Imag Sci*, 2020, 13: 29–52
- 224 Zhu H, Li S, Fomel S, et al. A Bayesian approach to estimate uncertainty for full-waveform inversion using *a priori* information from depth migration. *Geophysics*, 2016, 81: R307–R323
- 225 Zuazua E. Propagation, observation, and control of waves approximated by finite difference methods. *SIAM Rev*, 2005, 47: 197–243

Infinite-dimensional Bayesian inversion theory and algorithms

Junxiong Jia, Deyu Meng & Yuanxiang Zhang

Abstract Inverse problems constitute a significant area of mathematical research, with extensive applications across various engineering and technical fields such as medical imaging, seismic exploration imaging, image processing, and weather forecasting. Owing to the ill-posedness of inverse problems, the concept of regularization is introduced to solve these problems, resulting in an approximate estimation of the parameters. With the advancement of computational capabilities, people in fields like medical and exploration imaging are no longer satisfied with obtaining a reasonable estimate of the parameters to be estimated. Instead, they attempt to integrate empirical knowledge and uncertainty information of observational data to provide a complete characterization of the uncertainty of the parameters to be estimated. To achieve this goal, people transform inverse problems into Bayesian statistical inference problems, leading to the development of Bayesian inversion theory and numerical algorithms. Unlike classical statistical research, in inverse problem research, the parameters to be estimated and the observational data are connected by complex mathematical models (e.g., partial differential equations), thus necessitating the introduction of new ideas and mathematical theories. In this paper, we focus on the infinite-dimensional Bayesian inversion theory established for inverse problems and organize existing research work from aspects such as prior measure construction, Bayesian well-posedness, finite element discretization, statistical sampling algorithms, and statistical large-sample theory. The aim is to clarify the basic research ideas, core research issues, existing results and methods of infinite-dimensional Bayesian inversion methods, and potential future research directions.

Keywords inverse problems, infinite-dimensional Bayesian methods, discretization-invariant algorithms, variational inference, posteriori contraction estimates

MSC(2020) 65L09, 49N45, 62F15

doi: 10.1360/SSM-2024-0100