中国科学: 数学 2024 年 第 54 卷 第 ? 期: 1 ∼ 43
SCIENTIA SINICA Mathematica
综 述

无限维贝叶斯反演理论与算法

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, 西安市咸宁西路28号, 陕西西安, 710049;
2. 兰州大学, 数学与统计学院, 兰州市天水南路222号, 甘肃兰州, 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: 2024-XX-XX; * 通信作者
国家自然科学基金(批准号：12322116, 12271428, 12226004, 12326606);国家重点研发计划(批准号：2023YFC3503400)

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、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、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、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，反问题被转化为贝叶斯统计推断问题，进而发展出了贝叶斯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(例如：偏微分方程)联系起来的，因而需要引入新的思路、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贝叶斯反演理论，从先验测度构造、贝叶斯适定性、有限元离散、统计抽样算法、统计大样本理论等方面梳理现有的研究工作，旨在阐明无限维贝叶斯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 无限维贝叶斯 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、雷达成像等工程技术问题在医疗、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[1]. 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

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

英文引用格式: Jia J J, Meng D Y, Zhang Y X. Infinite-Dimensional Bayesian Inversion (in Chinese). Sci Sin Math, 2024, 54: 1–43, doi: 10.1360/SSM-2024-XXXX

其中$\mathcal{G}: X \to Y$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、粘弹波动方程[2]；在数值天气预报中建模为流体方程[3]；在数据分析中建模为神经网络[4,5]. 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)，其具体形式我们通常难以确定，但其统计性质经常是可以获得的. 这里我们考虑如下的加性噪音模型

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

其中η表示观测噪音. 关于其他形式的噪声模型，可进一步参考专著[6].

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[7]在1923年给出的适定性概念，其包含三个条件，对于上述问题具体可表述为：

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、医学成像等典型的反问题并不满足这三个条件：当噪音不在$\mathcal{G}$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的时候，解可能不唯一；正演算子即使是可逆的，$\mathcal{G}^{-1}$大多是无界算子. 关于这三个方面的简明阐述，推荐阅读文献[8]第二小节. 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. 前苏联科学院Tikhonov院士奠基性地提出了正则化方法[9,10]，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[11]. 关于正则化方法的近期发展，文献[12]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年的论文[13]中，Franklin针对线性反问题将其转化为贝叶斯统计推断问题，在可分Hilbert空间上详细推导了高斯先验、高斯噪声情况下的贝叶斯后验. 限于当时计算能力的限制，贝叶斯反演方法并没有得到学术界的重视. 不同于正则化方法，贝叶斯反演方法将未知量、观测噪音都建模为随机变量，随机性反映了观测者对其数值的不确定性. 从贝叶斯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(例如：后验均值，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，例如：医学成像中的伪影检测[14].

自1970年Franklin的工作之后，针对贝叶斯反演方法的发展主要集中在有限维贝叶斯推断方法. 围绕着地震勘探成像问题，Tarantola在其专著[15]中扩展了有限维空间中的贝叶斯反演方法，特别地，在专著的第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)的角度阐述了贝叶斯方法与对偶方法. 在专著[6]中，Kaipio与Somersalo完整阐述了有限维贝叶斯反演

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[16]中，Calvetti与Somersalo详细地介绍了贝叶斯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[1]，包括未知参数、微分方程系统本身以及反演输入的数据. 为了应用有限维的贝叶斯反演方法，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，然后利用有限维贝叶斯反演方法求解离散的反问题，我们称为“先离散、再贝叶斯”. “先离散、再贝叶斯”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年的论文[17]中，Lassas与Samuli的论证表明：基于有限维空间的全变差先验导出的贝叶斯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马尔柯夫链蒙特卡洛算法(Markov chain Monte Carlo, MCMC)的抽样效率对于离散维数的增加并不一致，随着离散维度的增加抽样效率会迅速下降[18]，这限制了贝叶斯反演方法在高维问题中的应用.

为了解决这些问题，Cotter, Dashti, Robinson与Stuart在2009年开创性的论文[3]中提出了无限维贝叶斯反演的适定性理论. 事实上，1970年Franklin的论文[13]中已经有了无限维贝叶斯反演适定性理论的雏形，但其研究仅针对高斯分布假设下的线性反问题，且方法难以直接推广到非线性、非高斯等更一般的情况. 随后，Stuart在论文[19]中详细阐述了无限维贝叶斯反演的适定性理论、新的抽样效率不随离散维度下降的马尔柯夫链蒙特卡洛算法，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. 无限维贝叶斯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贝叶斯适定性理论、后验抽样算法，进而将离散推迟到可能的最后一步，我们称为“先贝叶斯、再离散”. 相较于“先离散、再贝叶斯”的思路，“先贝叶斯、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，但其也带来了很多好处：(1) 构建了贝叶斯适定性，即：后验概率测度存在、唯一且连续依赖于观测数据. 无限维贝叶斯反演的适定性理论将偏微分方程的适定性理论与贝叶斯反演更紧密地联系了起来. (2) 无限维贝叶斯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) 无限维贝叶斯反演理论使得我们可以更明确地建立起贝叶斯理论与无限维空间上的正则化方法[10]之间的联系. (4) 直接在无限维空间上构建贝叶斯反演理论可以更好地利用无限维空间上反问题（例如：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中，例如：波动方程的可控性研究[20]，机器学习研究[21, 22].

经过十来年的快速发展，无限维贝叶斯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，本文旨在从贝叶斯适定性、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，阐明无限维贝叶斯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，例如：最优实验设计[23]和替代模型研究[24–26].

本文的主要安排如下：在第2节中，我们分四个小节梳理贝叶斯反演的适定性理论以及数值逼

近方法. 具体而言，在第2.1小节，我们简单地介绍无限维空间上的高斯概率测度，并给出先验概率测度的一些最新进展. 在第2.2小节，我们从直观的角度给出无限维贝叶斯公式，并简述贝叶斯反演的(P,d)适定性. 在第2.3小节，围绕着有限元离散，我们简述有限维离散逼近的基本思路. 在第2.4小节，我们对无限维贝叶斯理论的最大后验估计、统计决策理论进行了简要介绍，进而阐述了贝叶斯方法与正则化方法的关系. 在第3节中，我们分四个小节梳理贝叶斯反演的后验抽样算法. 在第3.1小节，我们回顾了Metropolis-Hastings算法的一般框架，并基于一般框架简述了常见的四种离散不变抽样算法. 在第3.2小节，我们对集合Kalman滤波进行了简明的介绍，并给出了最新的研究进展. 在第3.3小节，我们从三个典型的例子出发简单介绍了源自机器学习研究的变分推断方法. 在第3.4小节，我们聚焦于Darcy流反演问题，从数值实践的角度探讨了离散不变性，从而阐明在无限维空间构造算法的必要性. 最后，我们在第4节中，围绕着后验收缩率估计、Bernstein-von Mises定理介绍了无限维贝叶斯反演统计理论的近期发展.

2 贝叶斯反演理论

在这一小节，我们先来介绍几种常见的定义在可分Banach空间上的概率测度. 在对常见的概率测度有了基本的了解后，我们将对贝叶斯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，贝叶斯反演的适定性理论是讨论贝叶斯后验计算、数值逼近理论、后验统计性质分析的数学理论基础. 接下来，我们将简单的介绍离散逼近的一般理论分析框架. 在本章最后，我们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贝叶斯反演方法与正则化方法之间的联系.

2.1 先验概率测度

贝叶斯反演方法研究中涉及到的先验概率测度、噪音概率测度、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贝叶斯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，感兴趣的读者可以参考[27–30]. 下面我们以高斯测度为例，对其主要思想进行简要介绍.

令$\mathcal{H}$表示可分的Hilbert空间，其上的内积和范数记为$\langle\cdot,\cdot\rangle$和$\|\cdot\|$. 一般而言，我们可以取$\mathcal{H}=L^2(D;\mathbb{R})$，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

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

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，我们称表达式(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 scale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

$$L^2_{\mathbb{P}}(\Omega;\mathcal{H}^t) := \left\{v : D\times\Omega\to\mathbb{R}\,|\,\mathbb{E}(\|v\|_{\mathcal{H}^t})^2<\infty\right\},$$

其中$\mathbb{E}$表示求期望. 若我们假设$\xi_1\sim\mathcal{N}(0,1)$，$\gamma_j\asymp j^{-\frac{s}{d}}$ 且$t<s-\frac{d}{2}$，则随机函数u属于Hilbert空间$L^2_{\mathbb{P}}(\Omega;\mathcal{H}^t)$，称为可分Hilbert空间上的高斯随机元，并称$u$服从定义在可分Hilbert空间$\mathcal{H}$上的高斯测度$\mathcal{N}(u_0,\mathcal{C})$. 对于空间$\mathcal{H}$上的概率测度$\mu$，我们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[31]. 对于高斯测度$\mathcal{N}(m_0,\mathcal{C})$，通过简单的推导，我们可得

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

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.

选取高斯测度作为先验测度，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，例如分片常值函数. 然而，在图像重建等反问题[32]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[12]. 基于TV正则在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先验测度的定义是容易给出的，但在论文[17]中作者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，特别的，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，Lassas等构造了离散不变的Besov测度[33]. 更进一步，Dashti等在无限维可分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，在论文[34]中，作者进一步推广了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、可导性.

针对Besov类型的先验测度，论文[35–37]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，Huttunen等提出了Cauchy差分先验[38]；Yao等提出了TV-Gauss先验测度[39]；Hosseini等系统研究了尾部指数衰减的先验测度，给出了贝叶斯适定性分析[40].

作为这一小节的结束，我们有必要说明一下无限维空间上的高斯随机元与高斯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$u:D\times\Omega\to\mathbb{R}^n$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$\omega\in\Omega$，$u(\cdot;\omega):D\to\mathbb{R}^n$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$(u(x_1;\cdot),\ldots,u(x_K;\cdot))$是高斯随机向量，则我们称$u$是一个高斯随机场. 对于高斯随机场我们可以定义均值函数，协方差函数如下

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

由协方差函数，我们可以定义方差算子如下

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

在专著[41]的第二章，作者从无限维统计模型研究的角度对高斯过程（高斯随机场）进行了细致的讨论，详细的介绍了高斯过程的聚集性质. 在专著[42] 的第二、三章以及附录里，作者从非参贝叶斯统计相合性研究的角度对高斯过程进行了细致的介绍，并且简述了高斯过程与无限维空间上的高斯随机元之间的关系. 事实上，高斯过程与无限维空间上的高斯随机元在很弱的条件下都是等价的，参见如下定理2.1（摘自专著[42]中引理I.7）：

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

关于高斯过程与高斯随机元的介绍以及这两种不同观点在贝叶斯反演中的应用，感兴趣的读者可以进一步参考最近出版的专著[43]. 受到深度神经网络研究的启发，深度高斯过程获得了广泛关注. 最近，Dunlop等[44]从遍历性的角度对深度高斯过程进行了细致探讨. Abrahamd等[45]将深度高斯过程用于贝叶斯反演研究，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高斯先验更快的收缩速率估计.

2.2 适定性理论

在这一小节中，我们并不打算从严谨的数学理论推导的角度阐述无限维贝叶斯适定性理论，而是从有限维贝叶斯公式出发，尽量从直观的角度说明无限维贝叶斯适定性理论的核心思想. 在经典书籍[6]中，Kaipio与Somersalo对有限维贝叶斯反演理论进行了详细的讨论. 令X, Y表示可分Banach空间. 为了简洁，我们考虑如下的模型：

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

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

在有限维空间的假设下，即：令N_u, N_y为正整数，取$X = \mathbb{R}^{N_u}$, $Y = \mathbb{R}^{N_y}$，我们可以得到如下的贝叶斯公式：

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

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表达式:

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

这一贝叶斯公式成立的条件是$Z^y > 0$. 详细的论证，读者可以参考专著[6]的定理3.1 （或其中文译本[46]）.

将贝叶斯公式(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[47, 48]）. 具体而言，我们有如下周知的定理2.2(感兴趣的读者参考专著[49]中的定理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$\mu = 0$.

那么如何扩展有限维的贝叶斯定理到无限维空间呢？一个最直接的思路就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, Y是无限维可分Banach空间，令μ_0, μ^y分别表示先验概率测度与后验概率测度，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，则直观来讲可将公式(2.6)进行如下改写：

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

其中$Z^y := \int_X \rho(\mathcal{G}(u); y)\mu_0(du)$. 需要注意，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[27, 50]. 一般而言，贝叶斯公式应当是与积分区域A无关的，因而我们可以进一步给出无限维可分Banach空间上的贝叶斯公式：

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

这里我们遵循文献[3, 19, 51]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.

从有限维的贝叶斯公式直接改写过来的贝叶斯公式(2.8)是否有明确的数学定义？回答是肯定的，这一形式的贝叶斯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[41, 52]. 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(例如：Darcy流方程[19]、Navier-Stokes方程[3]、扩散方程[53, 54])，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(例如：偏微分方程反问题[1]). 早在1970年，Franklin[13]在高斯先验、高斯噪音的假设下，针对线性反问题细致讨论了如何构建无限维贝叶斯反演理论. Tarantola在他的专著[15, 55]中针对线性反问题在无限维空间的框架下探讨了贝叶斯反演理论（非线性问题的探讨集中于最大后验估计）. 直到2009年，Cotter等在论文[3]中首次提出了适用于偏微分方程反问题的贝叶斯反演适定性理论（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贝叶斯适定性. 在论文[19]中，Stuart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，完整的阐明了这一理论框架，极大的推动了贝叶斯反演理论的发展. 随后，Dashti与Stuart在[51]中弱化了论文[3, 19]中的条件，给出了更为一般的贝叶斯反演理论框架.

令$\mathrm{Prob}(X, \mu_0)$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，$d_{\mathrm{Hell}}(\cdot, \cdot)$表示测度的Hellinger距离，则文献[51] 中所阐述的贝叶斯适定性可如下定义.

定义2.1　对于贝叶斯反演问题(2.8)，如果下面三点成立

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

则我们称贝叶斯反演是(Lipschitz, Hellinger)适定的.

上述贝叶斯反演的(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贝叶斯反演的(Lipschitz, Hellinger)适定性，例如：Darcy流扩散系数反演问题[56, 57]，基于水平集方法的反问题[58]，多频逆介质散射问题[59, 60]，非局部微分方程反演问题[61, 62]，具有Dirac源项的Helmholtz方程反源问题[63]，刻画肿瘤生长的Cahn-Hilliard模型[64]. 为了证明贝叶斯反演的(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，并且对观测噪音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（例如：机器学习领域的分类问题，噪音为退化分布的反问题），因而Latz近期扩展了贝叶斯反演的(Lipschitz, Hellinger)适定性[65, 66]，提出了贝叶斯反演的(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，则贝叶斯反问题称为(P, d)适定的是指贝叶斯反问题满足如下三条性质：

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$y \to \mu^y$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与定义2.2的主要区别在与后验测度的稳定性. 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，我们给出如下的四个条件(严谨论述见[66])：

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[65, 66]验证了$(\mathrm{Prob}(X), d_{\mathrm{Prok}})$，$(\mathrm{Prob}(X), d_{TV})$，$(\mathrm{Prob}(X, \mu_0), d_{Hell})$适定性. 这里$\mathrm{Prob}(X)$表示$X$上的概率测度构成的集合，$d_{\mathrm{Prok}}$表示Prokhorov度量，$d_{TV}$表示概率测度的全变差度量. 这一扩展极大的放宽了对于正演算子与噪音分布的假设，拓宽了无限维贝叶斯反演理论的适用范围. 有关贝叶斯反演适定性最新的研究进展，可以进一步参考[67–71].

注2.1 在常见的关于无限维贝叶斯反演的研究中[19, 51]，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贝叶斯反演的相关理论(参考文献[51]的第2-3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-代数是不一致的，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. 感兴趣的读者可以参考专著[27]第1.1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度量空间时，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[42]的第A.4小节)，这违背直觉认知；
- 当无限维空间缺乏可分性的时候，上文中公式(2.7)的意义就会不太明确，其难以通过Bochner积分来理解(参考文献[27]的第1.1-1.2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，贝叶斯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，我们可以较容易的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，我们先简要的回顾一下贝叶斯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[8, 72, 73]. 关于基于有限差分、谱方法等的离散逼近，感兴趣的读者可以参考[74, 75].

令$D \subset \mathbb{R}^d$是有界连通的开区域，我们考虑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$\{\boldsymbol{x}_j\}_{j=1}^n$，从而我们有$\phi_j(\boldsymbol{x}_i) = \delta_{ij}$其中$i, j \in \{1, \ldots, n\}$. 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$u_h = \sum_{j=1}^n u_j\phi_j \in V_h$. 更具体的，需要统计推断的是向量$\boldsymbol{m} = (m_1, \ldots, m_n)^{\mathrm{T}}$. 在这一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$L^2(D)$上内积. 对于$u_1, u_2 \in L^2(D)$，我们记有限维逼近函数为$u_{1h}, u_{2h} \in V_h$，从而易知

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

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

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

从上面的简要叙述，我们可知$L^2(D)$的离散逼近空间是带权重$\boldsymbol{M}$内积的欧几里得空间$\mathbb{R}^n_{\boldsymbol{M}}$，而不是通常的欧几里得空间$\mathbb{R}^n$. 在贝叶斯反演方法中，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$\mathbb{R}^n_{\boldsymbol{M}}$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[72]，我们可以得知

1. 对于算子$\mathcal{B} : L^2(D) \to L^2(D)$，其离散近似算子$\boldsymbol{B} : \mathbb{R}^n_{\boldsymbol{M}} \to \mathbb{R}^n_{\boldsymbol{M}}$具有如下形式

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

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

2. 对于算子$\boldsymbol{F} : \mathbb{R}^n_{\boldsymbol{M}} \to \mathbb{R}^q$，其对偶算子$\boldsymbol{F}^\natural : \mathbb{R}^q \to \mathbb{R}^n_{\boldsymbol{M}}$可如下表示

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

3. 对于算子$\boldsymbol{V}:\mathbb{R}^r\to\mathbb{R}^n_{\boldsymbol{M}}$，其对偶算子$\boldsymbol{V}^\diamond:\mathbb{R}^n_{\boldsymbol{M}}\to\mathbb{R}^r$可如下表示

$$\boldsymbol{V}^\diamond=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{M}.$$

对于高斯先验测度$\mathcal{N}(u_0,\mathcal{C}_0)$，我们取$\mathcal{C}_0=\mathcal{A}^{-2}$，其中$\mathcal{A}$是定义域为$D(\mathcal{A})=\{u\in H^2(D):$在区域边界$\partial D$上满足$\alpha\nabla u\cdot\boldsymbol{n}=0\}$（$\boldsymbol{n}$表示外法线方向）的微分算子. 具体而言，对于$u\in D(\mathcal{A}),f\in L^2(D)$, 我们有$\mathcal{A}u=f$满足

$$\begin{cases}-\alpha\Delta u+\beta u=f, & x\in D,\\ \alpha\nabla u\cdot\boldsymbol{n}=0, & x\in\partial D.\end{cases}\tag{2.10}$$

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

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

从而无限维空间上的高斯测度$\mathcal{N}(u_0,\mathcal{C}_0)$可以用具有如下密度函数的有限维高斯分布进行逼近

$$\pi(\boldsymbol{u})\propto\exp\left(-\frac{1}{2}\|\boldsymbol{A}(\boldsymbol{u}-\boldsymbol{u}_0)\|^2_{\boldsymbol{M}}\right).\tag{2.11}$$

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，从高斯测度中进行采样的逼近计算公式[72]如下

$$\boldsymbol{u}=\boldsymbol{u}_0+\boldsymbol{K}^{-1}\boldsymbol{M}^{1/2}\boldsymbol{\xi},\tag{2.12}$$

其中$\boldsymbol{\xi}\sim\mathcal{N}(0,\boldsymbol{I})$(这里的$\boldsymbol{I}$表示通常的欧几里得空间$\mathbb{R}^n$上的恒等算子). 需要说明的是这里的矩阵$\boldsymbol{M}$一般而言并不是对角矩阵，因此如何计算$\boldsymbol{M}^{1/2}\boldsymbol{\xi}$并不是平凡的(一般而言，矩阵$\boldsymbol{M}^{1/2}$不再是稀疏矩阵，因而我们一般仅计算$\boldsymbol{M}^{1/2}\boldsymbol{\xi}$，并不显示计算出$\boldsymbol{M}^{1/2}$)，感兴趣的读者可以参考[76]. 基于这样的离散方式，我们也很容易计算公式(2.4)中方差算子对应的高斯核函数的逼近核函数$c_h(\cdot,\cdot)$，其计算公式如下：

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

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

基于所述的离散方法，我们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$\Phi^n(\boldsymbol{u};\boldsymbol{y})$. 如果我们假设观测数据属于有限维空间，且取噪音分布为高斯分布$\mathcal{N}(\boldsymbol{0},\boldsymbol{\Gamma})$，就可以得到有限维离散近似贝叶斯公式如下：

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

基于如上的离散方法，贝叶斯统计反演方法被用于求解一系列大规模反问题，例如：全球地层介质的反演[72]、在冰层流动问题里反演边界条件[77]、多频散射数据逆源问题[78]、地震勘探全波形反演问题[79].

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. 特别的，在第3节中我们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、Hessian算子等. 如同绪论中所述，类似于偏微分方程(PDE)约束下的优化问题中“先离散、后优化”与“先优化、后离散”的讨论，我们同样面临“先离散、后贝叶斯”与“先贝叶斯、后离散”两种分析计算方式.

在本文中，我们聚焦于“先贝叶斯、后离散”的无限维贝叶斯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PDE约束优化问题中，关于这两种计算方式的讨论，我们推荐[8,80,81]. 在论文[82–84]中，从贝叶斯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，PDE约束优化问题里细致的研究在贝叶斯反演研究中同样具有重要价值.

在这一小节的第一部分我们简明的介绍了基于有限元的离散，下面我们对离散逼近速率估计进行简要的阐述[82,85]. 令μ_n^y，$\Phi^n(u;y)$是对后验测度μ^y以及位势函数$\Phi(u;y)$在X的n维子空间上的逼近，则我们有

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

其中$Z_n^y=\int_X\exp\left(-\Phi^n(u;y)\right)d\mu_0(u)$. 这里$\Phi^n(u;y)$被看做是由正演算子$\mathcal{G}$的离散导出的位势函数的离散逼近. 在论文[82]中证明了如下关键定理.

定理2.3　假设Φ与Φ^n满足合适的条件（本文中不再赘述，所需条件保证了贝叶斯反演原问题与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

$$|\Phi(u;y)-\Phi^n(u;y)|\leqslant K\exp(\epsilon\|u\|_X^2)\psi(n), \tag{2.16}$$

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

$$d_{\mathrm{Hell}}(\mu^y,\mu_n^y)\leqslant C\psi(n), \tag{2.17}$$

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

当我们假设观测值$\boldsymbol{y}\in\mathbb{R}^{N_y}$，即观测空间是有限维的，假设噪音服从高斯分布$\mathcal{N}(\mathbf{0},\boldsymbol{\Gamma})$时，位势函数与离散逼近位势函数具有如下形式：

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

如果我们假设

$$|\mathcal{G}(u)-\mathcal{G}^n(u)|\leqslant K(\epsilon)\exp\left(\epsilon\|u\|_X^2\right)\psi(n), \tag{2.19}$$

则我们立刻可以得到

$$|\Phi(u;\boldsymbol{y})-\Phi^n(u;\boldsymbol{y})|\leqslant K(2\epsilon)\exp\left(2\epsilon\|u\|_X^2\right)\psi(n). \tag{2.20}$$

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

周知，统计抽样算法(例如下文中提及的马尔柯夫链蒙特卡洛(MCMC)抽样算法)在估计统计量（如：均值）的时候误差通常是以$\sqrt{N}$(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，从而在解决实际应用问题中更好的平衡离散精度与抽样精度（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，例如：基于Navier-Stokes方程的数据同化问题[82]，Darcy流方程渗透率反演问题[85]，逆形状声场散射问题[86]，双曲守恒律反演初值问题[87]，工业水力旋流器的时变状态估计问题[88]，并行抽样算法的分析[89].

2.4 最大后验估计

贝叶斯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. 为了更好的理解贝叶斯反演理论，我们有必要准确的理解贝叶斯反演方法与经典正则化方法之间的联系与区别. 在有限维贝叶斯反演方法的讨论中，这一联系是显然的，即：贝叶斯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贝叶斯后验概率最大的点，通常称为最大后验点估计[6, 46]. 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，我们以高斯概率先验、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，我们考虑X上的高斯先验概率测度$\mu_0 = \mathcal{N}(0, \mathcal{C}_0)$，$E$表示高斯测度$\mu_0$的Cameron-Martin空间(其上的范数定义为$\|\mathcal{C}_0^{-1/2} \cdot \|_X$). 定义如下泛函

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

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$u \in X$，我们有$\Phi(u; y) \geqslant M - \epsilon\|u\|_X^2$.
2. 位势函数$\Phi(u; y)$关于u是局部有上界的.
3. 位势函数$\Phi(u; y)$关于u局部Lipschitz连续.

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

$$\lim_{\delta \to 0} \frac{\mu^y(B(z_1, \delta))}{\mu^y(B(z_2, \delta))} = \exp\left(I(z_2) - I(z_1)\right). \tag{2.22}$$

这个定理在论文[90]中被证明，且这一结论被用于分析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高斯测度μ_0的Cameron-Martin空间E值得我们特殊注意，这正反映了无限维情况与有限维情况的核心区别. 在无限维的情况下，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仅令空间维数$n\to\infty$就得到无限维空间准确的结果，例如在论文[91]中，作者分析了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，根据我们所知难以得到带有Besov先验、分层先验情况下其与相应泛函极值问题对应关系的严格刻画. 在论文[92]中，基于测度的Fomin可微性理论，作者提出了弱最大后验(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$u\in\mathrm{supp}(\mu^y)$且对于所有$h\in E$有

$$\lim_{\delta\to 0}\frac{\mu^y(B(u-h,\delta))}{\mu^y(B(u,\delta))}\leqslant 1, \tag{2.23}$$

则我们称u是wMAP估计.

上述定义中的空间E不必是先验的Cameron-Martin空间（对于非高斯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$d_h\mu^y:\mathcal{B}(X)\to\mathbb{R}$是测度$\mu^y$沿着$h$的Fomin导数[93]，我们假设对于所有的$h\in E$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，在论文[92,94]中得到了Besov先验下wMAP估计对应于如下泛函极值问题：

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

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. 关于贝叶斯理论与正则化理论的联系，感兴趣的读者可以进一步参考[34,95–100].

后验最大估计联系了贝叶斯反演方法与正则化方法，对其研究具有重要的意义. 另一方面，最大后验估计是后验测度的一种点估计，特别的，在贝叶斯统计中后验均值(posterior mean, CM)估计

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

是更加常用的一种点估计. 如果我们需要基于后验测度给出一个点估计（贝叶斯决策理论），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[6]中第三章所展示的图3.1，我们在图1中给出了一维的一个简单的示意图. 从图1左边的子图中可以看到CM估计不是一个好的估计，同样的右边的子图说明MAP估计在图示的情况下不是一个好的估计. 在贝叶斯统

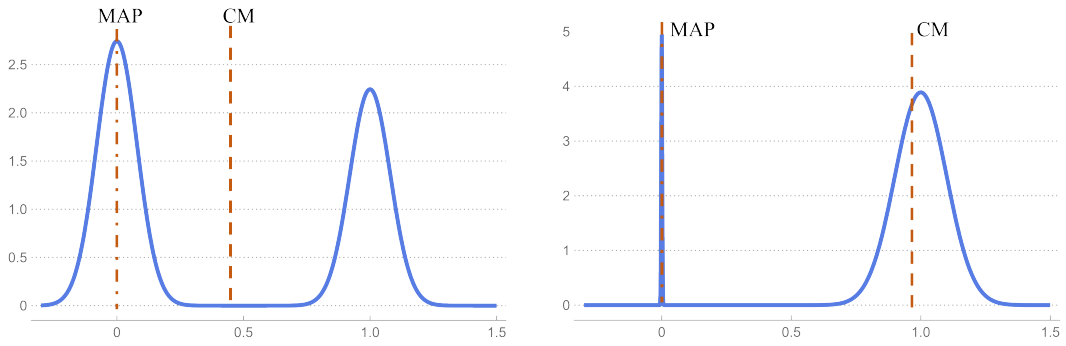

图 1 两个一维的密度函数说明最大后验估计与后验均值估计都有可能不是一个好的估计. 图中MAP表示后验最大估计(maximum a posteriori, MAP)，CM表示后验均值估计(conditional mean, CM)

计决策理论中，人们定义了贝叶斯损失(Bayes cost)用以刻画期望损失，从而对CM、MAP估计提供更深刻的理解. 如果我们记$C(\cdot,\cdot)$为损失函数，那么贝叶斯损失(记为$BC(\cdot)$)可定义如下：

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

其中$\hat{u}$是后验点估计. 我们显然希望寻找使得贝叶斯损失最小的点估计，即：

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

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

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

并考虑$\delta \to 0$的极限情况时才能得到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. 但论文[101]中指出，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况下可以给出很好的估计结果. 为了更深刻的理解MAP估计，作者证明了：基于Bregman距离所给出的损失函数，MAP估计是使得贝叶斯损失最小的点估计，从而在一定程度上说明了MAP估计是一个合适的贝叶斯估计. 在论文[92, 102]中，作者就CM、MAP估计在无限维空间的情形下做了细致的讨论，得到了类似的结论，同样说明了MAP估计可以看做Bregman距离损失函数导出的贝叶斯估计. 关于点估计的进一步讨论(例如：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考[103–106].

3 后验统计计算

在上一节中我们回顾了无限维贝叶斯反演的适定性、逼近计算、后验点估计等理论，这一节我们聚焦于后验概率测度信息的计算. 在第一小节，我们简要的介绍无限维空间上的Metroplis-Hastings方法，给出常见的马尔柯夫链蒙特卡洛(MCMC)抽样算法. 在第二小节，我们对无限维贝

叶斯反演的集合Kalman滤波算法进行介绍，并回顾一些最新进展. 在第三小节，我们对变分推断方法进行简要介绍，特别的，对近期在无限维空间发展出的理论方法进行回顾. 最后，我们对无限维空间所构造算法的离散不变性进行简单的探讨，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，我们了解到无限维贝叶斯公式具有如下形式

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

下面我们基于文献[18, 19, 51]对MCMC算法的思想进行简要的介绍. 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（例如：后验均值，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：

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

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

$$\mu^y(du)P(u,dv) = \mu^y(dv)P(v,du), \tag{3.3}$$

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

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

由于转移核P关于后验测度是不变的，我们就能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$P(\cdot,\cdot)$?

算法 1 Metropolis-Hastings方法的一般计算框架

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

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$k\to k+1$，返回步骤2.

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，下文中详细介绍）给出的建议函数参数，构造我们期望的Markov转移核P. 从算法1接受与拒绝的计算过程中，我们容易知道转移核P具有如下的形式：

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

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

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

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

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

沿着论文[18]所述的思路，我们在空间$(X \times X, \mathcal{B}(X) \otimes \mathcal{B}(X))$上定义

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

其中$\mathcal{B}(X) \otimes \mathcal{B}(X)$表示乘积空间上的乘积$\sigma$域. 在文献[51](定理21)中证明了如下定理.

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

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

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

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

特别的，如果我们选取

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

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：

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

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，$D\Phi(u;y)$ 表示对函数参数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，我们可以得到

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

其中ξ_0是高斯白噪声. 下面所述的方法，当$\gamma = 1$时，针对条件扩散(conditioned diffuion)问题在论文[107]中提出. 随后，在论文[108]中被扩展到了$\gamma = 0$的情况，并且所发展的算法进一步被用于数据同化领域[109]中. 需要说明一下，参数$\gamma = 0$时所发展的算法，在1998年的论文[110]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贝叶斯反演方法). 下面我们分四种情况分别加以介绍.

情况1：$\gamma=0,\mathcal{K}=\mathrm{I}$.

在这种情况下，无限维Langevin方程具有如下形式：

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

上述方程的Crank-Nicolson逼近格式为：

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

称为Crank-Nicolson(CN)建议. 若$\left(I+\frac{1}{2}\delta\mathcal{C}_0^{-1}\right)^{-1}$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$\left(I+\frac{1}{2}\delta\mathcal{C}_0^{-1}\right)^{-1}$大致等价于计算两次偏微分方程(2.10).

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

在这种情况下，无限维Langevin方程具有如下形式：

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

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

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

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

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

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

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

关于函数a的详细推导过程，读者可以参考文献[51]的第5.2节. 将公式(3.9)与(3.13)代入到一般算法1中，我们就得到了Crank-Nicolson(CN)算法. 将公式(3.12)与(3.13)代入到一般算法1中，我们就得到了预条件Crank-Nicolson(pCN)算法.

注3.1 在情况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，进而我们可以得到(文献[51]中例7与例8)：

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

情况3：$\gamma=1,\mathcal{K}=\mathrm{I}$.

在这种情况下，无限维Langevin方程形式如下：

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

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

$$(2\mathcal{C}_0+\delta)v=(2\mathcal{C}_0-\delta)u-2\delta\mathcal{C}_0D\Phi(u;y)+\sqrt{8\delta\mathcal{C}_0}w, \tag{3.15}$$

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

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

将CNL建议(3.15)与函数(3.16)代入到一般算法1中，我们就得到了Crank-Nicolson Langeivn(CNL)算法. 关于(3.15)与(3.16)的详细推导过程，读者可以参考论文[108]的第4节以及附录.

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

在这种情况下，无限维Langevin方程形式如下：

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

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

$$(2+\delta)v=(2-\delta)u-2\delta\mathcal{C}_0D\Phi(u;y)+\sqrt{8\delta}w, \tag{3.18}$$

其中$w\sim\mu_0=\mathcal{N}(0,\mathcal{C}_0)$. 进一步，借助于定理3.1，我们可以计算得到函数a如下：

$$a(u,v)=\Phi(u;y)+\frac{1}{2}\langle v-u,D\Phi(u;y)\rangle+\frac{\delta}{4}\langle u+v,D\Phi(u;y)\rangle+\frac{\delta}{4}\|\mathcal{C}_0^{1/2}D\Phi(u;y)\|^2. \tag{3.19}$$

将pCNL建议(3.18)与函数(3.19)代入到一般算法1中，我们就得到了预条件Crank-Nicolson Langeivn (pCNL)算法. 关于(3.18)与(3.19)的详细推导过程，读者可以参考论文[108]的第4节以及附录.

注3.2 不同于情况1和情况2，在情况3和情况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，我们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，感兴趣的读者可以参考[51, 111, 112]了解详细的假设与证明过程. 这里需要说明的是，在文献[51, 111, 112]中，作者对位势函数$\Phi(u;y)$做了如下假设：

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

其中N_1,N_2是两个正常数，M_1,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，仅做一些粗略讨论）. 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），例如：稳态Darcy流渗透率反演问题(文献[51]的第1.3节与第334页例3). 据我们所知，在文献[51]的第5.6.1节，作者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，但至今（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)与谱间隙(spectral gap)分析. 在缩放极限方面，论文[113–115]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度的某个倍数减小，为了准确估计统计量，所需的抽样量会随着离散维度的增加而增大；无限维空间MCMC类算法建议分布的方差无需根据离散维度进行调整，因而为了准确估计统计量，所需的抽样量不会随着离散维度的增加而增大. “先优化

再离散”的研究中也有类似于缩放极限的讨论，感兴趣的读者可以参考论文[84]的附录(其中给出了一个简单的数值算例)以及PDE约束优化问题的专著[80]. 在谱间隙分析方面，论文[116]中的分析表明：标准的MCMC算法的谱间隙会随着离散维度的增加而减小，pCN算法的谱间隙不依赖于离散维度. 最近，无限维MCMC谱间隙的分析理论也被扩展到了非高斯先验测度的情形[117].

注3.4　不同于反问题求解的经典正则化方法，贝叶斯反演方法得到的解并不仅仅是函数参数的单个估计(当然，我们可以通过贝叶斯决策理论给出满足一定需求的点估计，参考第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低. 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，我们可以画出样本的迹(trace)，计算样本的自相关性，计算有效样本数等. 这里限于篇幅，不再做细致的介绍，读者可以参考专著[118].

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(例如：非高斯先验下抽样算法的构造，Metropolis-within-Gibbs算法，Hamiltonian Monte Carlo算法等)，感兴趣的读者推荐进一步阅读[74, 115, 119–130].

3.2　集合卡尔曼滤波

这一节的主要目的是阐述利用集合卡尔曼滤波求解无穷维贝叶斯反问题的基本原理和算法. 为此, 我们先回顾基本的卡尔曼滤波及相关算法. 为了便于理解相关算法的基本思想和底层逻辑, 下面我们先从有限维情形着手论述, 相关算法的无穷维推导和应用[131–133]与有限维情形是平行的.

卡尔曼滤波是由出生于匈牙利的数学家卡尔曼(Rudolf Emil Kalman, 1930 - 2016) 为了解决太空航行中的导航及控制问题而提出的, 早期的文献有 [134, 135] 等. 卡尔曼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

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

及数据观测模型

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

其中v_0 与序列$\{\xi_k\}$ 及$\{\eta_k\}$ 相互独立, 且对所有的j 和k, ξ_j 和η_k 也相互独立. 在实际应用中, (3.21) 式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

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

则 (3.21) 及 (3.22) 的联立即为数据同化领域 [136, 137] 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$\mathbb{P}(v_k|Y_k)$ (filtering distribution) 或平滑分布$\mathbb{P}(v_k|Y_K)$ (smoothing distribution), 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.

当 (3.21) 及 (3.22) 均为线性模型, 即假设成立

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

其中v_0 及序列$\{\xi_k\}$ 及$\{\eta_k\}$ 的假设如前所述. 由于v_{k+1} 及y_{k+1} 都是高斯随机变量的仿射变换, 因此, $\mathbb{P}(v_k|Y_k)$ 也是高斯分布, 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卡尔曼滤波算法(Kalman filter, Kf).

记

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

顾名思义, 预测分布$\widehat{\pi}_{k+1}$ 就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} 的情况下利用数据Y_{k+1} 对预测结果进行校正. 下面的定理给出了Kf 的具体刻画:

定理3.2 [137, 定理8.3] 对于所有的$1 \leqslant k \leqslant K-1$, 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

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

这里我们略去如上定理的证明, 感兴趣的读者可参见[137, 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[137, 引理8.6], Kf 可以等价地改写为算法2 的形式, 其中的K_k 又称为卡尔曼增益(Kalman gain). 可以证明Kf 给出了状态均值的最优化估计[137, 定理8.7]. 对于Kf 的其他推导方式或更新公式, 可参见[136, 138–140] 及其中的参考文献.

需要指出的是, Kf 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, 当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, 即成立

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

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

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

从而有

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

算法 2 卡尔曼滤波算法 (Kf)

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

1. 设置$k = 0$;
2. 预测:

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

3. 分析:

$$\begin{aligned} m_{k+1} &= \widehat{m}_{k+1} + K_{k+1}(y_{k+1} - H\widehat{m}_{k+1}), \\ K_{k+1} &= \widehat{\mathcal{C}}_{k+1} H^{\mathrm{T}} (H\widehat{\mathcal{C}}_{k+1} H^{\mathrm{T}} + \Gamma)^{-1}, \\ \mathcal{C}_{k+1} &= (I - K_{k+1}H)\widehat{\mathcal{C}}_{k+1}; \end{aligned}$$

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

及

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

其中$D\psi(m_k)$ 为$\psi(\cdot)$ 在m_k 处的Fréchet 导数, 且$\widehat{m}_{k+1}$ 和$\widehat{\mathcal{C}}_{k+1}$ 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 中将预测均值$\widehat{m}_{k+1}$ 和协方差$\widehat{\mathcal{C}}_{k+1}$ 作相应的替换即得到扩展卡尔曼滤波算法(Extended Kalman filter, ExKf). 可以看出ExKf 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 的发展及系统性理论, 可参见文献[141], ExKf 在天气预报中的应用可参见[142], 由于大多数地球物理应用中的状态空间维数都很高, 这也就使ExKf 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 还是ExKf, 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, 集合卡尔曼滤波(Ensemble Kalman filter, EnKf) 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卡尔曼型算法相结合, 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$\widehat{C}_{k+1}$. 关于EnKf 的早期发展可参见[143–146], 近期的综述性文献有[136, 137, 147] 等. 目前在EnKf 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高斯-牛顿迭代或Levenberg-Marquardt方法相结合的一致优化框架内[137, Part III Kalman Inversion], 我们将这一类方法统称为EnKf 方法.

下面我们将回到这一节的主题, 即阐述如何利用EnKf 来求解贝叶斯反问题

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

并给出一般的算法框架. 为此, 我们先以数据增广的方式构造人工动力系统:

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

其中$\Xi(z_k) = \begin{pmatrix} u_k \\ G(u_k) \end{pmatrix} := \begin{pmatrix} u_k \\ p_k \end{pmatrix}$, $\widetilde{H} = \begin{pmatrix} 0, & I \end{pmatrix} : \ X \times Y \to Y$, $\{\zeta_k\}$ 与$\{\widetilde{\eta}_k\}$ 为i.i.d. 序列($\zeta_0 \sim \mathcal{N}(0, \widetilde{\Sigma})$, $\widetilde{\eta}_0 \sim \mathcal{N}(0, \widetilde{\Gamma})$), $\{u_0^{(j)}\}_{j=1}^J$ 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$\mathcal{G} : X \to Y$ 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$d_{k+1} = \widetilde{H} z_{k+1} + \widetilde{\eta}_{k+1}$ 是线性的, 这与(3.24) 式的情形是一致的.

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:

$$\widehat{\mathcal{C}}_k : = \frac{1}{J} \sum_{j=1}^J (z_k^{(j)} - \overline{z}_k) \otimes (z_k^{(j)} - \overline{z}_k) = \begin{pmatrix} \mathcal{C}_k^{uu} & \mathcal{C}_k^{up} \\ (\mathcal{C}_k^{up})^{\mathrm{T}} & \mathcal{C}_k^{pp} \end{pmatrix},$$

其中

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

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

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) [148].

这里我们将EnKf 写成算法3 的形式, 目的只是明确其与传统Kf 之间的联系. 实际应用中, 在极大多数情形下, 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$\widehat{\mathcal{C}}_{k+1}$ 的更新简化为部分经验协方差$\mathcal{C}_k^{up}$ 及$\mathcal{C}_k^{pp}$ 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 的计算效率, 具体细节可参见[149, Algorithm 1].

事实上, EnKf 可以看作是滤波分布$\pi(v_k|Y_k)$ 的同权序列蒙特卡洛(sequential Monte Carlo, SMC)逼近[131, 133, 137]. 记

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

算法 3 集合卡尔曼滤波算法 (EnKf)

给定初始粒子$\{z_0^{(j)}\}_{j=1}^J$;

1. 设置$k=0$;

2. 预测:

$$\zeta_k^{(j)} \sim \mathcal{N}(0,\widetilde{\Sigma}),\ \ i.i.d.,\ \ j=1,2,\cdots,J, \qquad \widehat{z}_{k+1}^{(j)} = \Xi(z_k^{(j)}) + \zeta_k^{(j)},$$
$$\widehat{m}_{k+1} = \frac{1}{J}\sum_{j=1}^{J}\widehat{z}_{k+1}^{(j)}, \qquad \widehat{\mathcal{C}}_{k+1} = \begin{pmatrix} \mathcal{C}_k^{uu} & \mathcal{C}_k^{up} \\ (\mathcal{C}_k^{up})^{\mathrm{T}} & \mathcal{C}_k^{pp} \end{pmatrix};$$

3. 分析:

$$\widetilde{\eta}_k^{(j)} \sim \mathcal{N}(0,\widetilde{\Gamma}),\ \ i.i.d.,\ \ j=1,2,\cdots,J, \qquad d_{k+1}^{(j)} = d_{k+1} + \tau\widetilde{\eta}_k^{(j)},\ \ \tau\in\{0,1\},$$
$$z_{k+1}^{(j)} = \widehat{m}_{k+1} + K_{k+1}(d_{k+1}^{(j)} - \widetilde{H}\widehat{m}_{k+1}), \qquad K_{k+1} = \widehat{\mathcal{C}}_{k+1}\widetilde{H}^{\mathrm{T}}(\widetilde{H}\widehat{\mathcal{C}}_{k+1}\widetilde{H}^{\mathrm{T}}+\Gamma)^{-1};$$

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

当随机动力学模型中的非线性关系较弱或数据噪音较小时, 滤波分布可近似看成是高斯分布, 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 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) [150, 151]. 反之, 当随机动力学模型中的非线性关系较强或数据噪音较大时, $\pi_k^J(z_k)$ 逼近滤波分布的效果较差, 即便如此, EnKf 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 理解为序列优化方法[137] (此时, (3.26) 中的参数$s=0$) 而不是当作逼近采样方法.

EnKf方法现在在反问题、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三个方面的原因[137]: (1) EnKf 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$\mathcal{G}$) 的导数的情况下进行计算, 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) 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(在贝叶斯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 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$J\to\infty$ 时, EnKf 将收敛于卡尔曼滤波分布[152–154]; 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[155]. 在一致平均场框架下, 关于EnKf 方法的综述可参见[156], 该框架为分析EnKf 方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 方法仍能提供良好的状态估计, 这也被视为EnKf 方法无可争议的优点之一, 相关的理论分析可参见[131–133, 157–161] 等.

总之, 由于EnKf 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以及人工智能技术的迅速发展, 可以预见EnKf 方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，论文[18] 中抽取了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$10^5 - 10^7$个样本才可以有效的估计后验统计信息. 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[162]等大规模反演问题.

2012年，AlexNet[163]在ImageNet比赛中获得了突破性进展，准确率远超第二名(Top 5错误率为15.3%，第二名为26.2%). 自此，深度神经网络获得了广泛关注，相关研究渗透到了众多领域. 特别的，深度学习方法在反问题求解中取得了令人瞩目的效果[164]. 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，贝叶斯统计在其中扮演了重要的角色[165]. 深度神经网络中包含大量的可学习参数，进而训练深度神经网络可以看做高维空间上的统计推断问题，面临着与无限维贝叶斯反演类似的困难：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1980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[166, 167]. 随后，在1990年代，变分推断被众多学者用于概率图模型的研究[168–170]. 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，我们推荐两篇综述论文[171, 172]. 如前文所述无限维贝叶斯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贝叶斯反演领域？沿着“先离散，再贝叶斯”的思路，针对有限维反问题，有一些很有意思的工作. 在论文[173]中研究了带有超参数的分层贝叶斯反演，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)、高斯先验噪声假设，作者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. 特别的，在论文[173] 中作者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、skew-t分布观测噪音假设下的有限维反问题[174, 175]以及多孔介质流中的随机介质反演问题[176]. 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[177]、非局部微分方程反问题[178]，从而快速的近似计算后验统计信息. 下面，从无限维贝叶斯反演理论的角度，我们对变分推断在反问题领域的发展做一些回顾.

首先，我们来简略的介绍一下什么是变分推断. 在这一小节，我们假设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

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

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：最简单的我们可取$\mathcal{A}$表示高斯测度的集合

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

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就可以给出不同的算法.

这里，我们仅考虑度量$D(\cdot||\cdot)$取为如下Kullback – Leibler(KL)散度的情况

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

在上式中，我们采用通常的规定$0\log 0=0$.

注3.5　需要注意，这里的所采用的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[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[4].

当我们固定了度量$D(\cdot||\cdot)$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$\mathcal{A}$选取的过小，仅包含较为简单的概率测度，那么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$\mathcal{A}$选取的很大，对后验测度的近似会更为精确，但计算就会变得困难. 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，我们对三种典型的情况进行简单的介绍.

假设1：$\mathcal{A}$是高斯概率测度够成的集合.

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

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、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. $\mathcal{A}_1=\{$空间X上的高斯概率测度$\}$.

2. $\mathcal{A}_2=\{$空间X上等价于μ_0的高斯概率测度$\}$.

3. 对于空间X上的对称、正定、迹算子$\hat{\mathcal{C}}$，取$\mathcal{A}_3=\{$空间X具有方差算子$\hat{\mathcal{C}}$的高斯概率测度$\}$.

4. 对于固定的$\hat{u}\in X$，取$\mathcal{A}_4=\{$空间X具有均值$\hat{u}$的高斯概率测度$\}$.

对于这四种情形，如果存在$\nu\in\mathcal{A}_i(i=1,2,3,4)$使得$D_{KL}(\nu||\mu^y)<\infty$，则变分优化问题

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

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是等价的，进而什么样的高斯测度等价于μ_0 就是一个很重要的问题. 在论文[179, 180]中，作者进一步证明：对于高斯概率测度$\nu=\mathcal{N}(u,\mathcal{C})$，如果$u\in\mathcal{H}$且方差算子$\mathcal{C}^{-1}=\mathcal{C}_0^{-1}+\Gamma$满足

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

其中$\mathcal{HS}(\mathcal{H})$表示空间$\mathcal{H}$上的Hilbert-Schmidt算子构成的空间. 基于等价高斯概率测度的刻画，在论文[180]中，作者给出了有限秩参数化与Schrödinger参数化方法，进而基于Robbins-Monro算法构造了新的高斯逼近变分推断算法.

基于高斯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. 在高斯先验、高斯噪音的假设下，对于线性问题，我们就可以显示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[72].

假设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，在论文[78]中，作者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，并将其运用于带有高斯噪音、Laplace噪音的多频数据Helmholtz方程逆源反问题. 这里，我们举个例子说明其核心思想，考虑如下问题：

$$y = \mathcal{G}u + \eta, \tag{3.31}$$

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

$$u \sim \mu_0^u = \mathcal{N}(0, \mathcal{C}_0^K(\lambda)), \quad \mathcal{C}_0^K(\lambda) = \sum_{k=1}^{K} \lambda^{-1}\alpha_k e_k \otimes e_k + \sum_{k=K+1}^{\infty} \alpha_k e_k \otimes e_k,$$
$$\lambda \sim \mu_0^\lambda = \mathrm{Gamma}(\alpha_0, \beta_0), \quad \tau \sim \mu_0^\tau = \mathrm{Gamma}(\alpha, \beta).$$

这里$\{\alpha_k, e_k\}_{k=1}^{\infty}$是某个对称、正定、迹算子的特征系统，$\mathrm{Gamma}(\alpha, \beta)$表示参数为$\alpha, \beta$的Gamma分布. 在这个问题中，我们假设了先验测度方差算子的参数λ与观测噪音分布的参数τ是未知的参数，因而在贝叶斯反演的框架中，我们假设他们服从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)，贝叶斯公式具有如下形式：

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

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$u, \lambda, \tau$是互相独立的三个参数. 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$\mu_0$与参考测度$\mu_r$是等价的，从而有

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

下面我们假设逼近测度ν具有如下表达式：

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

进而我们有

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

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,τ具有独立的三个分量. 粗略来讲，我们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

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

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论文[78]. 基于这些假设，在论文[78]中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是贝叶斯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$\Phi_u^r,\Phi_\lambda^r,\Phi_\tau^r$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$\Phi_u^r,\Phi_\lambda^r,\Phi_\tau^r$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与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高斯先验、高斯噪音的情况，同样适用于带有Laplace噪音的情况(参见论文[78]的3.2节). 针对多参数反演问题[181, 182]，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，我们假设了参数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个很有意思的问题. 在论文[183]中，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着广泛的应用，例如：基于平均场变分推断理论，在噪音是非独立同分布的情况时，在论文[184] 中构建了变分去噪模型，进一步在论文[185]中构造了变分超分模型. 借助无限维的平均场变分推断理论，在论文[186]中，在非独立同分布噪音假设的情况下，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：$\mathcal{A}:=\{\nu\,:\,\nu=\mu_0\circ T^{-1},\ T\text{是某种变换}\}$.

在假设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，例如：基于标准化流(normalizing flow)的变分推断[177]，Stein变分梯度下降(Stein variational gradient descen-

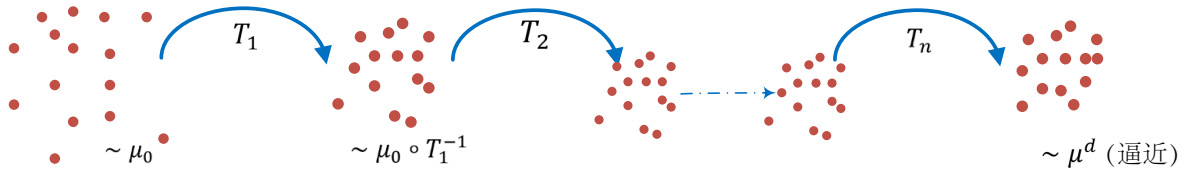

图 2 Stein变分梯度下降(Stein variational gradient descent, SVGD)基本思路示意图

t, SVGD)[187]等. 这里，我们简要的介绍一下SVGD算法以及其无限维版本iSVGD(infinite-dimensional SVGD)[84].

SVGD是在论文[187]中提出的(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，其中的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. 更具体的，图2中每个变换具有如下形式：

$$T_i = I + \epsilon\phi_i, \quad i = 1, 2, \ldots, N, \tag{3.36}$$

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. 我们容易观察到：如果$N \to \infty$且ϕ_i的选取不加以限制，那么复合变换$T_1 \circ T_2 \circ \cdots$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(RKHS)空间中. 下面，我们记RKHS空间为$\mathcal{H}_K$，其中K表示核函数. 关于RKHS的详细介绍，我们推荐[188, 189]. 如果将ϕ_i限制在$\mathcal{H}_K$中并且记$\nu_i = \mu_0 \circ (T_i \circ \cdots \circ T_1)^{-1}$，则直观上来讲应当通过如下方式确定$\phi_i$：

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

论文[187]中通过计算得到了问题(3.37)的解析解表达式如下：

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

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表示对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，我们就得到了可计算的样本粒子迭代公式：

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

其中k表示迭代步数，$j = 1, \ldots, L$(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[190–192]，其背后蕴含着黎曼流形上的梯度流分析等深刻的数学理论[193–196].

沿着“先贝叶斯，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，即使是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[197]. 在论文[84]中，针对无限维贝叶斯反演问题，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，论文中引入核函数变换的性质，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，论文[84]中

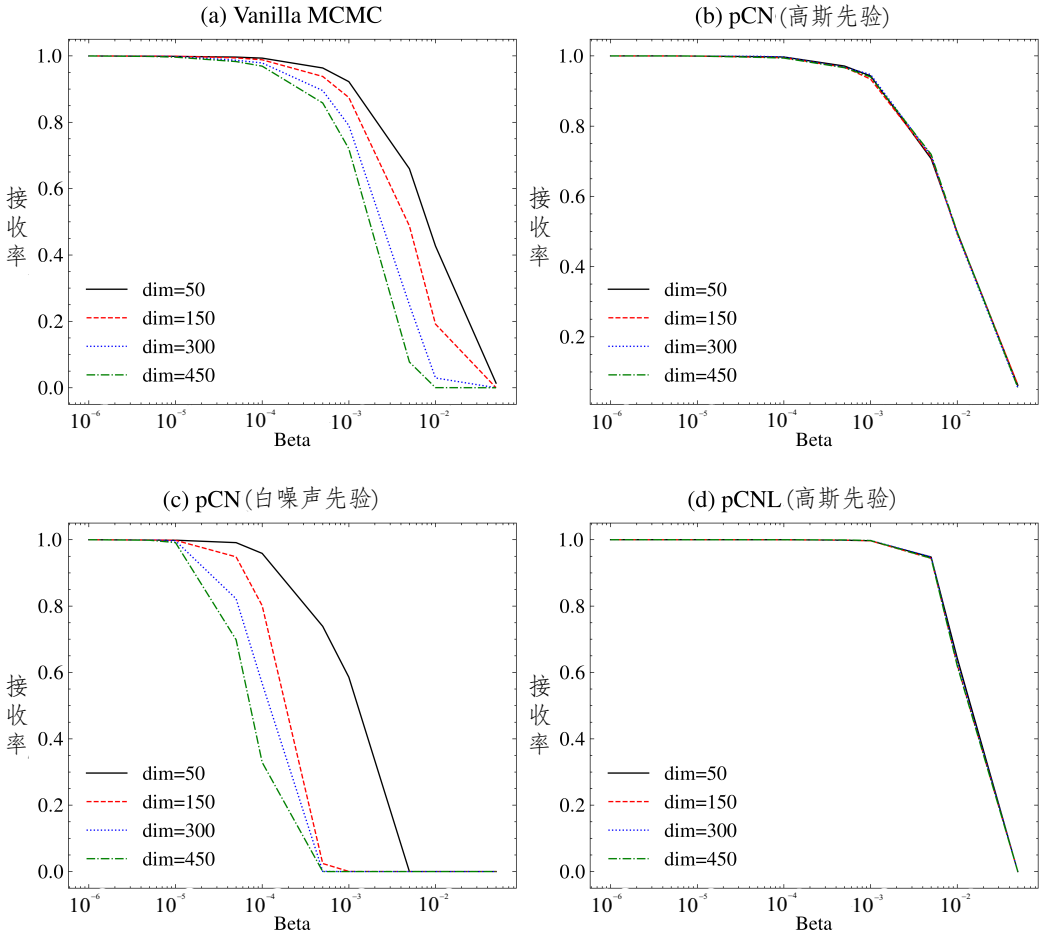

图 3　不同算法中，步长与接受率之间的关系. (a)：经典的基于随机游走的MCMC算法(高斯先验$\mathcal{N}(0,(0.5I-0.1\Delta)^{-2})$). (b)：带有高斯先验$\mathcal{N}(0,(0.5I-0.1\Delta)^{-2})$的pCN算法(先验测度符合理论要求). (c)：带有先验测度$\mathcal{N}(0,I)$的pCN算法(先验测度不符合理论要求). (d)：基于“先贝叶斯、再离散”的思路(梯度通过对偶方法计算)得到的pCNL算法(高斯先验$\mathcal{N}(0,(0.5I-0.1\Delta)^{-2})$).

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[198]，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高斯测度的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：如何将论文[195]中的理论推广到iSVGD的情况，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

在这一小节，我们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，变分推断等算法. 具体而言，我们考虑稳态Darcy流模型[51]如下:

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

其中$f = 1$是一个常值函数，$D = (0,1)^2 \subset \mathbb{R}^2$. 我们考虑的反问题：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$X = L^2(D)$. 在$u \in L^\infty(D)$时，易知方程的解$w \in H_0^1(D)$. 针对这一具体问题，我们取两种先验测度：$\mu_0^1 = \mathcal{N}(0, \mathcal{C}_0)$其中$\mathcal{C}_0 = (0.5I - 0.1\Delta)^{-2}$ 与$\mu_0^2 = \mathcal{N}(0, I)$其中$I$为空间$X$上的单位算子. 显然，先验$\mu_0^1$的方差算子是迹算子，而$\mu_0^2$的方差算子不是迹算子，因而先验测度$\mu_0^2$的选取不符合无限维贝叶斯理论的要求. 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经过较小的改动就可以获得离散不变的性质. 这里我们还想强调：pCN在无限维空间构造，其先验测度需要是在无限维空间上有意义的测度，而不能随便选取，因此如果我们从有限维的角度探讨pCN时[118]也需要强调先验高斯分布的方差矩阵不仅是正定、对称的，其特征值需要满足一定的衰减性条件(具体条件参考文献[51]中的假设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度为$\{50^2, 150^2, 300^2, 450^2\}$时的行为. 图中的横坐标是公式(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[199, 200]实现的，程序可以在网址https://github.com/jjx323/IllustrateDimIndependenceMCMC/tree/main 上下载. 图3的子图(a)中展示了经典的随机游走算法[18]的结果. 可以看到当β增加的时候接收率会下降，而且接收率下降的速率是和离散维度相关的，离散维度越大接受率在β相同时越小. 图3的子图(b)中展示的是第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，不论离散维度如何变化，β与接受率的变化关系都是一致的. 图3的子图(c)中展示的是第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，离散维度越大算法的抽样效率越低. 图3的子图(d)中展示的是第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，不论离散维度如何变化，β与接受率的变化关系都是一致的，即抽样算法的抽样效率关于离散维度是不变的.

在图3的子图(d)中，我们展示了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度下，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，我们就可以观察到“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考文献[8](第496-497页)与[83].

在无限维空间上的变分推断系列研究中[78, 84, 183, 186]，作者在数值实验中同样验证了算法的离散不变性. 在文献[183]的图7中，作者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、后贝叶斯”思路给出的变分推断算法[173]，同样观察到了“先贝叶斯、后离散”所带来的离散不变性. 在文献[186]中，将无限维变分推断理论与Fourier神经算子[201, 202]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，无限维贝叶斯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，我们从一个特殊的情况入手，说明统计理论保障研究的重要性. 参考论文[203]的理论结果，我们考虑线性模型如下：

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

其中$n \in \mathbb{N}^+$，$\mathcal{G} : X \to Y$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. 我们取先验、噪音的分布如下：

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

根据贝叶斯理论，我们可知后验概率测度$\mu^y = \mathcal{N}(u_p, \mathcal{C}_p)$其中

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

从上面的表达式，我们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$\mathcal{C}_p$蕴含了偏离均值的程度，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，我们考虑一种情况：真实参数$u^\dagger \in L^2(D)\backslash H^1(D)$，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$Y = L^2(D)$. 在这种情况下，我们由后验均值的表达式容易看出$u_p \in H^{10}(D)$ (这里并不是严格的说明，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 scale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. 但如果我们选取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音强度的减小($n \to \infty$)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，真实参数的性质，因而有必要发展无限维贝叶斯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）贝叶斯推断的研究中(针对经典统计问题)，已经有了丰富的理论成果[204]，这里仅聚焦于无限维贝叶斯反演方法(针对反问题)的统计理论，简述其主要定理以及近期发展. 我们以第3.4小节所介绍的Darcy流问题为例来介绍无限维贝叶斯统计理论的基本定理，具体而言我们考虑

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

其中$i = 1, 2, \ldots, N$. 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，我们考虑随机测量，即x_i是空间D上随机采样得到，我们记(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，我们更换前文中所使用的后验测度的记号μ^y为μ^{D_N}. 关于随机测量情况下的贝叶斯公式其严谨的数学论证参见[41, 43].

选取$\beta > 1 + d/2$，令$\mathcal{R}$是L^2空间的子空间且满足$\|\cdot\|_{H^\beta} \leqslant C\|\cdot\|_{\mathcal{R}}$. 定义$B_{\mathcal{R}}(M) = \{u \in \mathcal{R} : \|u\|_{\mathcal{R}} \leqslant M\}$，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:

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

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

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

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

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

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

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，对于足够大的常数$M > 0$，利用条件稳定性估计可以得到

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

上式意味着，我们有如下关于后验均值u_p的估计

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

其中$O_{P_{u^\dagger}^N}(\delta_N^\eta)$表示依概率$\delta_N^\eta$的速度收敛[205]. 估计(4.5)表明：当数据无限增多$N \to \infty$时，预测误差($\mathcal{G}(u)$与$\mathcal{G}(u^\dagger)$的误差)按照速率δ_N趋向于0. 类似的，估计(4.6)表明：当数据无限增多$N \to \infty$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，我们称估计(4.6)为后验收缩率估计，满足后验收缩率估计则我们称贝叶斯后验测度是相合的(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. 我们也可以如问题(4.1)一样考虑函数测量数据，但噪音水平趋于0时后验概率测度的极限性质，得到类似的后验收缩率估计.

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高斯先验、高斯噪音假设下的线性问题开始的，因为在这一情况下，我们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，就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，论文[203]中首次对线性反问题，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，论文[203]中的方法被推广用来研究严重不适定问题[206]与逆向热传导问题[207]. 对于偏微分方程反问题，先验方差算子与正演算子可同时对角化是一个比较强的假设条件，在论文[208]中作者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，在论文[209]中，作者扩展了论文[210]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、置信域分析. 不同于这些研究，在论文[211]中，作者给出了线性反问题的Oracle型后验收缩率估计. 最近，在论文[212]中作者通过运用扇形算子等工具，克服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. 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，噪音方差算子可同时对角化的假设下，在论文[213, 214]中详细的研究了经验贝叶斯方法的后验收缩率估计，详细分析了贝叶斯置信区域与频率置信区域之间的关系. 随后，在论文[215]中在不可同时对角化的假设下给出了经验贝叶斯方法的后验收缩率估计，从而扩展了理论的适用范围. 在论文[216]中作者引入了连续模，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年的论文[217]. 作者在论文中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，进而可以方便的应用非参贝叶斯已有的研究结果[204]. 在论文[217]中作者提出了这一重要的思路，给出了初步的探索，但限于论文中的证明技术，论文的结论应用范围有限，例如：无法应用于稳态Darcy流高斯先验假设的情形. 自2013年的这篇论文，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[218, 219]、Schödinger方程反位势函数[220, 221]、Darcy流反演问题[221–223]、Caldéron问题[224, 225]，研究者引入了假设检验、经验过程估计等重要的非参贝叶斯后验分析理论，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贝叶斯反演统计理论是Bernstein-von Mises(BvM)定理. 我们引入记号$\mathbb{I}_u[h] := D\mathcal{G}_u[h]$，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(**I**nformation)矩阵的概念. BvM定理给出了类似如下收敛性结论

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

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$\mu_N = \mu_N|D_N$ (概率测度μ_N依赖于随机变量$D_N \sim P_{u^\dagger}^N$)，我们称Z_N按照概率$P_{u^\dagger}^N$依分布收敛到Z是指

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

其中d_{weak}是概率测度的弱收敛导出的距离. 事实上，可以证明：如果我们要在真值$u^\dagger$附近估计$\langle u,\psi\rangle_{L^2}$，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. 简洁的来叙述，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高斯概率分布逼近，这个高斯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[43]对贝叶斯反演中的BvM定理做了基本的介绍. 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，我们容易得知

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

其中w_u是参数为u时方程的解，$\mathfrak{L}_u:=-\nabla\cdot(e^u\nabla\cdot)$. 在文献[43, 222]中，作者构造了反例，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[43, 220, 222]，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进的意义下可以被高斯概率分布近似，进而在渐进的意义下贝叶斯置信区域可借由高斯分布进行刻画.

从关于后验收缩率估计、BvM定理的简单回顾中，我们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[226]、Caldéron问题[225]、反散射问题[227]，近期也出现了有关后验收缩率估计、BvM定理的研究.

参考文献

1 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, 494: 643–666
2 Fichtner A. Full Seismic Waveform Modelling and Inversion, Berlin Heidelberg: Springer-Verlag, 2011
3 Cotter S L, Dashti M, Robinson J C, et al. Bayesian inverse problems for functions and applications to fluid mechanics. Inverse Probl, 2009, 2511: 115008
4 Bishop C M. Pattern Recognition and Machine Learning, Springer, New York, 2006
5 Kovachki N B, Stuart A M. Ensemble Kalman inversion: a derivative-free technique for machine learning tasks. Inverse Probl, 2019, 359: 095005, 35
6 Kaipio J, Somersalo E. Statistical and Computational Inverse Problems, New York: Springer-Verlag, 2005
7 Hadamard J. Lectures on Cauchy's Problem in Linear Partial Differential Equations, New York: Dover Publications, 1953
8 Ghattas O, Willcox K. Learning physics-based models from data: Perspectives from inverse problems and model reduction. Acta Numer, 2021, 30: 445–554
9 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
10 Engl H W, Hanke M, Neubauer A. Regularization of Inverse Problems, Kluwer Academic Publishers Group, Dordrecht, 1996
11 Schuster T, Kaltenbacher B, Hofmann B, et al. Regularization Methods in Banach Spaces, Walter de Gruyter GmbH & Co. KG, Berlin, 2012
12 Benning M, Burger M. Modern regularization methods for inverse problems. Acta Numer, 2018, 27: 1–111
13 Franklin J N. Well-posed stochastic extensions of ill-posed linear problems. J Math Anal Appl, 1970, 31: 682–716
14 Zhou Q, Yu T, Zhang X, et al. Bayesian inference and uncertainty quantification for medical image reconstruction with Poisson data. SIAM J Imaging Sci, 2020, 131: 29–52
15 Tarantola A. Inverse problem theory and methods for model parameter estimation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005

16 Calvetti D, Somersalo E. Introduction to Bayesian Scientific Computing, Springer, New York, 2007
17 Lassas M, Samuli S. Can one use total variation prior for edge-preserving Bayesian inversion? Inverse Probl, 2004, 205: 1537–1563
18 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, 283: 424–446
19 Stuart A M. Inverse problems: A Bayesian perspective. Acta Numer, 2010, 19: 451–559
20 Zuazua E. Propagation, observation, and control of waves approximated by finite difference methods. SIAM Rev, 2005, 472: 197–243
21 E W. The dawning of a new era in applied mathematics. Notices Amer Math Soc, 2021, 684: 565–571
22 Li Q, Chen L, Tai C, et al. Maximum principle based algorithms for deep learning. J Mach Learn Res, 2017, 18: 165, 29
23 Alexanderian A. Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: a review. Inverse Probl, 2021, 374: 043001, 31
24 Stuart A M, Teckentrup A L. Posterior consistency for Gaussian process approximations of Bayesian posterior distributions. Math Comp, 2018, 87310: 721–753
25 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
26 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, 391: A214–A239
27 Prato G D, Zabczyk J. Stochastic Equations in Infinite Dimensions, Cambridge: Cambridge University Press, 2014
28 Bogachev V I. Measure Theory, Heidelberg: Springer Berlin, 2017
29 Bogachev V I. Gaussian Measures, Providence, RI: American Mathematical Society, 1998
30 Kukush A. Gaussian Measures in Hilbert Space, Great Britain: ISTE Ltd, 2019
31 Reed M, Simon B. Functional Analysis I: Methods of Modern Mathematical Physics, Elsevier (Singapore) Pte Ltd, 2003
32 Rudin L, Oscher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Phys D: Nonlinear Phenomena, 1992, 60: 259–268
33 Lassas M, Saksman E, Siltanen S. Discretization-invariant Bayesian inversion and Besov space priors. Inverse Probl Imag, 2009, 31: 87–122
34 Jia J, Peng J, Gao J. Bayesian approach to inverse problems for functions with a variable-index Besov prior. Inverse Probl, 2016, 328: 085006
35 Kolehmainen V, Lassas M, Niinimäki K, et al. Sparsity-promoting Bayesian inversion. Inverse Probl, 2012, 282: 025005
36 Bui-Thanh T, Ghattas O. A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors. Inverse Probl Imag, 2015, 91: 27–53
37 Herrmann L, Keller M, Schwab C. Quasi-Monte Carlo Bayesian estimation under Besov priors in elliptic inverse problems. Math Comput, 2021, 90330: 1831–1860
38 Markkanen M, Roininen L, Huttunen J M, et al. Cauchy difference priors for edge-preserving Bayesian inversion. J Inverse Ill-Pose P, 2019, 272: 225–240
39 Yao Z, Hu Z, Li J. A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementations. Inverse Probl, 2016, 327: 075006
40 Hosseini B, Nigam N. Well-posed Bayesian inverse problems: Priors with exponential tails. SIAM/ASA Journal on Uncertainty Quantification, 2017, 51: 436–465
41 Giné E, Nickl R. Mathematical Foundations of Infinite-Dimensional Statistical Models, New York: Cambridge University Press, 2016
42 Ghosal S, Vaart A v d. Fundamentals of Nonparametric Bayesian Inference, Cambridge: Cambridge University Press, 2017
43 Nickl R. Bayesian Non-Linear Statistical Inverse Problems, Berlin: EMS Press, 2023
44 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
45 Abraham K, Deo N. Deep gaussian process priors for bayesian inference in nonlinear inverse problems. arXiv preprint arXiv:2312.14294, 2023

46 Kaipio J, Somersalo E . 统计与计算反问题, 北京: 科学出版社, 2018
47 程士宏. 测度论与概率论基础, 北京: 北京大学出版社, 2004
48 Dudley R M. Real Analysis and Probability, Cambridge: Cambridge University Press, 2002
49 Aliprantis C D, Border K C. Infinite Dimensional Analysis, Springer, Berlin, 2006
50 Da Prato G. Introduction to Stochastic Analysis and Malliavin Calculus, Edizioni della Normale, Pisa, 2014
51 Dashti M, Stuart A M. The Bayesian approach to inverse problems. 2017: 311–428
52 赵林成，王占峰. 高等统计学概论, 北京: 高等教育出版社, 2016
53 Jia J, Peng J, Yang J. Harnack's inequality for a space-time fractional diffusion equation and applications to an inverse source problem. J Differ Equations, 2017, 2628: 4415–445
54 Jia J, Peng J, Gao J, et al. Backward problem for a time-space fractional diffusion equation. Inverse Probl Imag, 2018, 123: 773–799
55 Tarantola A. Inverse Problem Theory, Elsevier Science Publishers, B.V., Amsterdam, 1987
56 Dashti M, Stuart A M. Uncertainty quantification and weak approximation of an elliptic inverse problem. SIAM J Numer Anal, 2011, 496: 2524–2542
57 Iglesias M A, Lin K, Stuart A M. Well-posed Bayesian geometric inverse problems arising in subsurface flow. Inverse Probl, 2014, 3011: 114001, 39
58 Iglesias M A, Lu Y, Stuart A. A Bayesian level set method for geometric inverse problems. Interfaces Free Bound, 2016, 182: 181–217
59 Jia J, Wu B, Peng J, et al. Recursive linearization method for inverse medium scattering problems with complex mixture gaussian error learning. Inverse Probl, 2019, 357: 075003
60 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, 2759: 2299–2332
61 Zhang Y X, Jia J, Yan L. Bayesian approach to a nonlinear inverse problem for a time-space fractional diffusion equation. Inverse Probl, 2018, 3412: 125002, 19
62 Ding M H, Zheng G H. Determination of the reaction coefficient in a time dependent nonlocal diffusion process. Inverse Probl, 2021, 372: 025005
63 Engel S, Hafemeyer D, Münch C, et al. An application of sparse measure valued Bayesian inversion to acoustic sound source identification. Inverse Probl, 2019, 357: 075005, 33
64 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, 72: 526–552
65 Latz J. On the well-posedness of Bayesian inverse problems. SIAM/ASA J Uncertain Quantif, 2020, 81: 451–482
66 Latz J. Bayesian inverse problems are usually well-posed. SIAM Rev, 2023, 653: 831–865
67 Bohra P, Pham T A, Dong J, et al. Bayesian inversion for nonlinear imaging models using deep generative priors. IEEE Trans Comput Imaging, 2022, 8: 1237–1249
68 Holden M, Pereyra M, Zygalakis K C. Bayesian imaging with data-driven priors encoded by neural networks. SIAM J Imaging Sci, 2022, 152: 892–924
69 Laumont R, De Bortoli V, Almansa A, et al. Bayesian imaging using plug & play priors: when Langevin meets Tweedie. SIAM J Imaging Sci, 2022, 152: 701–737
70 Lanthaler S, Mishra S, Weber F. On Bayesian data assimilation for PDEs with ill-posed forward problems. Inverse Probl, 2022, 388: 085012, 44
71 Kawakami H. Stabilities of shape identification inverse problems in a Bayesian framework. J Math Anal Appl, 2020, 4862: 123903, 15
72 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, 356: A2494–A2523
73 Villa U, Petra N, Ghattas O. hIPPYlib: an Extensible Software Framework for Large-scale Deterministic and Bayesian Inverse Problems. J Open Sour Soft, 2018, 330: 940
74 Dunlop M M, Iglesias M A, Stuart A M. Hierarchical Bayesian level set inversion. Stat Comput, 2017, 276: 1555–1584
75 Agapiou S, Bardsley J M, Papaspiliopoulos O, et al. Analysis of the Gibbs sampler for hierarchical inverse problems. SIAM/ASA J Uncertain Quantif, 2014, 21: 511–544
76 Chen J, Anitescu M, Saad Y. Computing $f(A)b$ via least squares polynomial approximations. SIAM J Sci Comput,

2011, 331: 195–222
77 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, 364: A1525–A1555
78 Jia J, Zhao Q, Xu Z, et al. Variational Bayes' method for functions with applications to some inverse problems. SIAM J Sci Comput, 2021, 431: A355–A383
79 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 check for updates on crossmark. Geophysics, 2016, 815: R307–R323
80 De los Reyes J C. Numerical PDE-constrained optimization, Springer, Cham, 2015
81 Hinze M, Pinnau R, Ulbrich M, et al. Optimization with PDE Constraints, New York: Springer, 2008
82 Cotter S L, Dashti M, Stuart A M. Approximation of Bayesian inverse problems for PDEs. SIAM J Numer Anal, 2010, 481: 322–345
83 Bui-Thanh T, Nguyen Q P. FEM-based discretization-invariant MCMC methods for PDE-constrained Bayesian inverse problems. Inverse Probl Imaging, 2016, 104: 943–975
84 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, 604: 2225–2252
85 Dashti M, Stuart A M. Uncertainty quantification and weak approximation of an elliptic inverse problem. SIAM J Numer Anal, 2011, 496: 2524–2542
86 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, 21: 203–222
87 Duong D L. Inverse problems for hyperbolic conservation laws: A Bayesian approach, University of Sussex, 2020
88 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, 114528: 1518–1531
89 Conrad P R, Davis A D, Marzouk Y M, et al. Parallel local approximation MCMC for expensive models. SIAM/ASA J Uncertain Quantif, 2018, 61: 339–373
90 Dashti M, Law K J H, Stuart A M, et al. MAP estimators and their consistency in Bayesian nonparametric inverse problems. Inverse Probl, 2013, 299: 095017, 27
91 Hegland M. Approximate maximum a posteriori with Gaussian process priors. Constr Approx, 2007, 262: 205–224
92 Helin T, Burger M. Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems. Inverse Probl, 2015, 318: 085009, 22
93 Bogachev V I. Differentiable measures and the Malliavin calculus, American Mathematical Society, Providence, RI, 2010
94 Agapiou S, Burger M, Dashti M, et al. Sparsity-promoting and edge-preserving maximum *aposteriori* estimators in non-parametric Bayesian inverse problems. Inverse Probl, 2018, 344: 045002, 37
95 Dunlop M M, Stuart A M. MAP estimators for piecewise continuous inversion. Inverse Probl, 2016, 3210: 105003, 50
96 Wacker P, Knabner P. Wavelet-based priors accelerate maximum-a-posteriori optimization in Bayesian inverse problems. Methodol Comput Appl Probab, 2020, 223: 853–879
97 Klebanov I, Wacker P. Maximum *a posteriori* estimators in ℓ^p are well-defined for diagonal Gaussian priors. Inverse Probl, 2023, 396: 065009, 27
98 Kretschmann R. Are minimizers of the Onsager-Machlup functional strong posterior modes? SIAM/ASA J Uncertain Quantif, 2023, 114: 1105–1138
99 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, 91: 1–28
100 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
101 Burger M, Lucka F. Maximum *a posteriori* estimates in linear inverse problems with log-concave priors are proper Bayes estimators. Inverse Probl, 2014, 3011: 114004, 21
102 Pereyra M. Revisiting maximum-a-posteriori estimation in log-concave models. SIAM J Imaging Sci, 2019, 121: 650–670
103 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

104 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 Probl, 2022, 382: 025005, 32

105 Ayanbayev B, Klebanov I, Lie H C, et al. Γ-convergence of Onsager-Machlup functionals: II. Infinite product measures on Banach spaces. Inverse Probl, 2022, 382: 025006, 35

106 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, 114: 1195–1224

107 Stuart A M, Voss J, Wiberg P. Fast communication conditional path sampling of SDEs and the Langevin MCMC method. Commun Math Sci, 2004, 24: 685–697

108 Beskos A, Roberts G, Stuart A, et al. MCMC methods for diffusion bridges. Stoch Dyn, 2008, 83: 319–350

109 Cotter S L, Dashti M, Stuart A M. Variational data assimilation using targetted random walks. Internat J Numer Methods Fluids, 2012, 684: 403–421

110 Bernardo J, Berger J, Dawid A, et al. Regression and classification using Gaussian process priors. Bayesian Statistics, 1998, 6: 475–502

111 Hairer M, Stuart A M, Voss J, et al. Analysis of SPDEs arising in path sampling. I. The Gaussian case. Commun Math Sci, 2005, 34: 587–603

112 Hairer M, Stuart A M, Voss J. Analysis of SPDEs arising in path sampling. II. The nonlinear case. Ann Appl Probab, 2007, 175-6: 1657–1706

113 Mattingly J C, Pillai N S, Stuart A M. Diffusion limits of the random walk Metropolis algorithm in high dimensions. Ann Appl Probab, 2012, 223: 881–930

114 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, 226: 2320–2356

115 Beskos A, Pinski F J, Sanz-Serna J M, et al. Hybrid Monte Carlo on Hilbert spaces. Stochastic Process Appl, 2011, 12110: 2201–2230

116 Hairer M, Stuart A M, Vollmer S J. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. Ann Appl Probab, 2014, 246: 2455–2490

117 Hosseini B, Johndrow J E. Spectral gaps and error estimates for infinite-dimensional Metropolis-Hastings with non-Gaussian priors. Ann Appl Probab, 2023, 333: 1827–1873

118 Calvetti D, S E. Bayesian Scientific Computing, Springer, Cham, 2023

119 Vollmer S J. Dimension-independent MCMC sampling for inverse problems with non-Gaussian priors. SIAM/ASA J Uncertain Quantif, 2015, 31: 535–561

120 Teh Y W, Thiery A H, Vollmer S J. Consistency and fluctuations for stochastic gradient Langevin dynamics. J Mach Learn Res, 2016, 177: 7, 33

121 Cui T, Law K J H, Marzouk Y M. Dimension-independent likelihood-informed MCMC. J Comput Phys, 2016, 3041: 109–137

122 Beskos A, Girolami M, Lan S, et al. Geometric MCMC for infinite-dimensional inverse problems. J Comput Phys, 2017, 33515: 327–351

123 Huang D Z, Huang J, Reich S, et al. Efficient derivative-free Bayesian inference for large-scale inverse problems. Inverse Probl, 2022, 3812: 125006, 40

124 Agapiou S, Papaspiliopoulos O, Sanz-Alonso D, et al. Importance sampling: intrinsic dimension and computational cost. Statist Sci, 2017, 323: 405–431

125 Hoang V H, Schwab C, Stuart A M. Complexity analysis of accelerated MCMC methods for Bayesian inversion. Inverse Probl, 2013, 298: 085010, 37

126 Rudolf D, Sprungk B. On a generalization of the preconditioned Crank-Nicolson metropolis algorithm. Found Comput Math, 2018, 182: 309–343

127 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, 20

128 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, 401: A142–A171

129 Pavliotis G A, Stuart A M, Vaes U. Derivative-free Bayesian inversion using multiscale dynamics. SIAM J Appl Dyn Syst, 2022, 211: 284–326

130 Sun Z, Zheng G H. Solving linear Bayesian inverse problems using a fractional total variation-Gaussian (FTG) prior and transport map. Comput Statist, 2023, 384: 1811–1849

131 Schillings C, Stuart A M. Analysis of the ensemble Kalman filter for inverse problems. SIAM J Numer Anal, 2017, 553: 1264–1290
132 Schillings C, Stuart A M. Convergence analysis of ensemble Kalman inversion: the linear, noisy case. Appl Anal, 2018, 971: 107–123
133 Blömker D, Schillings C, Wacker P, et al. Well posedness and convergence analysis of the ensemble Kalman inversion. Inverse Probl, 2019, 358: 085007
134 Kalman R E. A new approach to linear filtering and prediction problems. Trans ASME Ser D J Basic Engrg, 1960, 821: 35–45
135 Kalman R E, Bucy R S. New results in linear filtering and prediction theory. Trans ASME Ser D J Basic Engrg, 1961, 83: 95–108
136 Law K J H, Stuart A M, Zygalakis K C. Data Assimilation: A Mathematical Introduction, Cham: Springer, 2015
137 Sanz-Alonso D, Stuart A M, Taeb A. Inverse Problems and Data Assimilation, Cambridge: Cambridge University Press, 2023
138 Asch M, Bocquet M, Nodet M. Data Assimilation: Methods, Algorithms, and Applications, Philadelphia: Society for Industrial and Applied Mathematics (SIAM), 2016
139 Reich S, Cotter C. Probabilistic Forecasting and Bayesian Data Assimilation, New York: Cambridge University Press, 2015
140 Särkkä S, Svensson L. Bayesian Filtering and Smoothing, Cambridge: Cambridge University Press, 2013
141 Jazwinski A H. Stochastic processes and filtering theory. Courier Corporation, 2007
142 Cohn S, Ghil M, Isaacson E. Applications of estimation theory to numerical weather prediction. In Dynamic Meteorology: Data Assimilation Methods, Springer, 1981, 36: 139–224
143 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
144 Evensen G, Van Leeuwen P J. Assimilation of geosat altimeter data for the agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. Mon Wea Rev, 1996, 1241: 85–96
145 L H P, Derome J. Methods for ensemble prediction. Mon Wea Rev, 1995, 1237: 2181–2196
146 Houtekamer P L, Mitchell H L. Data assimilation using an ensemble Kalman filter technique. Mon Wea Rev, 1998, 1263: 796–811
147 Evensen G. Data Assimilation: The Ensemble Kalman Filter, Berlin: Springer-Verlag, 2009
148 Iglesias M A, Law K J H, Stuart A M. Ensemble Kalman methods for inverse problems. Inverse Probl, 2013, 294: 045001
149 Iglesias M A. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. Inverse Probl, 2016, 322: 025002
150 Garbuno-Inigo A, Hoffmann F, Li W, et al. Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler. SIAM J Appl Dyn Syst, 2020, 191: 412–441
151 Garbuno-Inigo A, Nikolas N, Reich S. Affine invariant interacting Langevin dynamics for Bayesian inference. SIAM J Appl Dyn Syst, 2020, 193: 1633–1658
152 Le Gland F, Monbet V, Tran V D. Large sample asymptotics for the ensemble Kalman filter. PhD Thesis, 2009
153 Kwiatkowski E, Mandel J. Convergence of the square root ensemble Kalman filter in the large ensemble limit. SIAM-ASA J Uncertain, 2015, 31: 1–17
154 Mandel J, Cobb L, Beezley J D. On the convergence of the ensemble kalman filter. Applications of Mathematics, 2011, 566: 533–541
155 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, 2015, 3: 823–851
156 Calvello E, Reich S, Stuart A M. Ensemble Kalman methods: A mean field perspective, 2022
157 Al-Ghattas O, Sanz-Alonso D. Non-asymptotic analysis of ensemble Kalman updates: effective dimension and localization. Inf Inference, 2024, 131: Paper No. iaad043, 66
158 Gottwald G A, Majda A J. A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. Nonlinear Proc Geoph, 2013, 205: 705–712
159 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, 2710: 2579–2603
160 Tong X T, Majda A J, Kelly D. Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation.

Commun Math Sci, 2015, 145: 1283 – –1313
161 Tong X T, Majda A J, Kelly D. Nonlinear stability and ergodicity of ensemble based Kalman filters. Nonlinearity, 2016, 292: 131–140
162 Richter M. Inverse Problems—Basics, Theory and Applications in Geophysics, Birkhäuser/Springer, Cham, 2020
163 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. 2012
164 Arridge S, Maass P, Öktem O, et al. Solving inverse problems using data-driven models. Acta Numer, 2019, 28: 1–174
165 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
166 Opper M, Winther O. A mean field algorithm for Bayes learning in large feed-forward neural networks. In: Neural Information Processing Systems. 1996
167 Peterson C, Anderson J R. A mean field theory learning algorithm for neural networks. Complex Syst, 1987, 1: 995–1019
168 Jaakkola T, Saul L K, Jordan M I. Fast learning by bounding likelihoods in sigmoid type belief networks. In: Neural Information Processing Systems. 1995
169 Johnson M J, Willsky A S. Stochastic variational inference for bayesian time series models. In: International Conference on Machine Learning. 2014
170 Opper M, Saad D. Advanced Mean Field Methods: Theory and Practice, USA: MIT Press: Cambridge, MA, 2001
171 Zhang C, Bütepage J, Kjellström H, et al. Advances in variational inference. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 41: 2008–2026
172 Blei D M, Kucukelbir A, McAuliffe J D. Variational inference: a review for statisticians. J Amer Statist Assoc, 2017, 112518: 859–877
173 Jin B, Zou J. Hierarchical Bayesian inference for ill-posed problems via variational method. J Comput Phys, 2010, 22919: 7317–7343
174 Jin B. A variational Bayesian method to inverse problems with impulsive noise. J Comput Phys, 2012, 2312: 423–435
175 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
176 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
177 Zhang X, Nawaz M A, Zhao X, et al. An introduction to variational inference in geophysical inverse problems. In: Inversion of Geophysical Data. Elsevier, 2021, 73–140
178 Song X, Zheng G H, Jiang L. Variational Bayesian inversion for the reaction coefficient in space-time nonlocal diffusion equations. Adv Comput Math, 2021, 473: 31, 28
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, 476: 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, 376: A2733–A2757
181 Jing X, Yamamoto M. Simultaneous uniqueness for multiple parameters identification in a fractional diffusion-wave equation. Inverse Probl Imaging, 2022, 165: 1199–1217
182 Cen S, Jin B, Liu Y, et al. Recovery of multiple parameters in subdiffusion from one lateral boundary measurement. Inverse Probl, 2023, 3910: 104001, 31
183 Sui J, Jia J. Non-centered parametric variational Bayes' approach for hierarchical inverse problems of partial differential equations. Math Comp, 2024, 93348: 1715–1760
184 Yue Z, Yong H, Zhao Q, et al. Variational denoising network: toward blind noise modeling and removal. In: NeurIPS. 2019
185 Yue Z, Yong H, Zhao Q, et al. Deep variational network toward blind image restoration. IEEE T Pattern Anal, 2024: 1–16
186 Jia J, Wu Y, Li P, et al. Variational inverting network for statistical inverse problems of partial differential equations. J Mach Learn Res, 2023, 24201: 1–60
187 Liu Q, Wang D. Stein variational gradient descent: A general purpose Bayesian inference algorithm. In: NeurIPS.

2016
188 Cucker F, Smale S. On the mathematical foundations of learning. Bull Amer Math Soc (NS), 2002, 391: 1–49
189 Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning, MIT Press, Cambridge, MA, 2012
190 Grathwohl W, Wang K C, Jacobsen J H, et al. Learning the Stein discrepancy for training and evaluating energy-based models without sampling. In: International Conference on Machine Learning. PMLR, 2020, 3732–3747
191 Rothfuss J, Fortuin V, Josifoski M, et al. Pacoh: Bayes-optimal meta-learning with PAC-guarantees. In: International Conference on Machine Learning. PMLR, 2021, 9116–9126
192 Detommaso G, Cui T, Marzouk Y, et al. A Stein variational Newton method. In: Advances in Neural Information Processing Systems. 2018
193 Liu Q. Stein variational gradient descent as gradient flow. In: Advances in neural information processing systems. 2017
194 Liu C, Zhuo J, Cheng P, et al. Understanding and accelerating particle-based variational inference. In: International Conference on Machine Learning. PMLR, 2019, 4082–4092
195 Lu J, Lu Y, Nolen J. Scaling limit of the Stein variational gradient descent: the mean field regime. SIAM J Math Anal, 2019, 512: 648–671
196 Korba A, Salim A, Arbel M, et al. A non-asymptotic analysis for Stein variational gradient descent. In: NeurIPS. 2020
197 Wang Z, Ren T, Zhu J, et al. Function space particle optimization for Bayesian neural networks. In: ICLR. 2019
198 Liu C, Zhuo J, Cheng P, et al. Understanding and accelerating particle-based variational inference. In: ICML. 2019, 4082–4092
199 Logg A, Mardal K A, Wells G N. Automated Solution of Differential Equations by the Finite Element Method, Springer, 2012
200 Baratta I A, Dean J P, Dokken J S, et al. DOLFINx: the next generation FEniCS problem solving environment. preprint, 2023. doi:10.5281/zenodo.10447666
201 Li Z, Kovachki N B, Azizzadenesheli K, et al. Fourier neural operator for parametric partial differential equations. In: ICLR. 2020
202 Kovachki N, Lanthaler S, Mishra S. On universal approximation and error bounds for Fourier neural operators. J Mach Learn Res, 2021, 22: 1–76
203 Knapik B T, van der Vaart A W, van Zanten J H. Bayesian inverse problems with Gaussian priors. Ann Statist, 2011, 395: 2626–2657
204 Rousseau J. On the frequentist properties of bayesian nonparametric methods. Annu Rev Stat Appl, 2016, 31: 211–231
205 van der Vaart A W. Asymptotic Statistics, Cambridge University Press, Cambridge, 1998
206 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, 223: 297–321
207 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, 427: 1294–1313
208 Agapiou S, Larsson S, Stuart A M. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. Stochastic Process Appl, 2013, 12310: 3828–3860
209 Kekkonen H, Lassas M, Siltanen S. Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators. Inverse Probl, 2016, 328: 085005, 31
210 Kekkonen H, Lassas M, Siltanen S. Analysis of regularized inversion of data corrupted by white Gaussian noise. Inverse Probl, 2014, 304: 045009, 18
211 Lin K, Lu S, Mathé P. Oracle-type posterior contraction rates in Bayesian inverse problems. Inverse Probl Imaging, 2015, 93: 895–915
212 Chen D H, Li J, Zhang Y. A posterior contraction for Bayesian inverse problems in Banach spaces. Inverse Probl, 2024, 404: 045011
213 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, 1643-4: 771–813
214 Szabó B, van der Vaart A W, van Zanten J H. Frequentist coverage of adaptive nonparametric Bayesian credible sets. Ann Statist, 2015, 434: 1391–1428
215 Jia J, Peng J, Gao J. Posterior contraction for empirical Bayesian approach to inverse problems under non-diagonal

assumption. Inverse Probl Imag, 2021, 152: 201–228
216 Knapik B, Salomond J B. A general approach to posterior contraction in nonparametric inverse problems. Bernoulli, 2018, 243: 2091–2121
217 Vollmer S J. Posterior consistency for Bayesian inverse problems through stability and regression results. Inverse Probl, 2013, 2912: 125011, 32
218 Monard F, Nickl R, Paternain G P. Efficient nonparametric Bayesian inference for X-ray transforms. Ann Statist, 2019, 472: 1113–1147
219 Monard F, Nickl R, Paternain G P. Consistent inversion of noisy non-Abelian X-ray transforms. Comm Pure Appl Math, 2021, 745: 1045–1099
220 Nickl R. Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. J Eur Math Soc (JEMS), 2020, 228: 2697–2750
221 Nickl R, Paternain G P. On some information-theoretic aspects of non-linear statistical inverse problems. In: ICM—International Congress of Mathematicians. Vol. VII. Sections 15–20. EMS Press, Berlin, 2023, 5516–5538
222 Monard F, Nickl R, Paternain G P. Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. Ann Statist, 2021, 496: 3255–3298
223 Giordano M, Nickl R. Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. Inverse Probl, 2020, 368: 085001, 35
224 Abraham K, Nickl R. On statistical Calderón problems. Math Stat Learn, 2019, 22: 165–216
225 Bohr J. A Bernstein–von-Mises theorem for the Calderón problem with piecewise constant conductivities. Inverse Probl, 2023, 391: 015002, 18
226 Nickl R, Ray K. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. Ann Statist, 2020, 483: 1383–1408
227 Furuya T, Kow P Z, Wang J N. Consistency of the Bayes method for the inverse scattering problem. http://www.math.ntu.edu.tw/ jnwang/pub/resources/Bayes-IS.pdf, 2023

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. This paper focuses on the infinite-dimensional Bayesian inversion theory established for inverse problems and organizes 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, posterior contraction estimates**

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

doi: 10.1360/SSM-2024-XXXX