arXiv:2302.04518v1 [stat.ML] 9 Feb 2023

贝叶斯逆问题中的高斯过程回归：关于加权误差度量的实验设计新结果

T. Helin、A.M. Stuart、A.L. Teckentrup、K.C. Zygalakis

摘要 现代应用中产生的贝叶斯后验分布往往因评估数据似然的计算成本过高而难以计算。例如，气候建模和地下流体流动中的偏微分方程模型所涉及的逆问题便属于此类。为缓解似然评估成本高昂的问题，一种自然的方法是使用高斯过程回归为似然构建代理模型，从而得到在实践中便于计算的近似后验分布。本文旨在介绍高斯过程回归，特别是其在为逆问题构建代理模型这一背景下的应用；我们还针对训练点的恰当选择提出了新的见解，这些见解源于高斯过程在近似贝叶斯反演中的应用。我们证明，真实后验分布与近似后验分布之间的误差可以由真实似然与近似似然之间的误差界定，该误差是在以真实后验分布加权的 L^2-范数下测得的；此外，我们还证明，在此范数下最小化真实似然与近似似然之间的误差，意味着应依据真实后验分布来选择高斯过程代理模型中的训练点。

Tapio Helin LUT大学，工程学院，Yliopistokatu 4，邮政信箱20，FI-53851 Lappeenranta，芬兰，电子邮箱：tapio.helin@lut.fi Andrew M Stuart 加州理工学院，计算与数学科学系，1200 E. California Blvd.，MC 305-16，Pasadena，CA 91125，美国，电子邮箱：astuart@caltech.edu Aretha L Teckentrup 爱丁堡大学，数学学院，詹姆斯·克拉克·麦克斯韦大楼，Peter Guthrie Tait路，EH9 3FD，苏格兰，电子邮箱：a.teckentrup@ed.ac.uk Konstantinos C Zygalakis 爱丁堡大学，数学学院，詹姆斯·克拉克·麦克斯韦大楼，Peter Guthrie Tait路，EH9 3FD，苏格兰，电子邮箱：k.zygalakis@ed.ac.uk

1 引言

科学和工程领域的数学模型中的参数往往无法完全确定，必须根据观测数据进行估计。准确重建参数并估计重建结果的不确定性，对于可靠的预测和风险评估至关重要。随着大规模科学实验的增加以及传感器技术的发展，可用数据近来激增，这意味着亟须开发新的算法，以便将观测数据与复杂的数学模型无缝集成。

现代应用中涉及的模型（例如基于偏微分方程的模型）复杂度很高，导致计算成本巨大，使得许多求解逆问题的方法在实践中难以实现。解决这一问题的一种广泛采用的方法，是用计算成本更低的代理模型近似原模型，从而获得计算速度快得多的近似似然；这种方法在实践中可用于模拟，并且理想情况下，在解决所关注的参数学习问题时，能够在需要准确的地方达到足够的精度 [19, 25, 4, 24, 11, 42]。

本文探讨了贝叶斯方法在逆问题中使用代理模型的情况。在这种方法中，我们需要求得以观测数据为条件的未知参数的后验分布。本文特别关注随机代理模型和高斯过程回归。我们给出了新的结果，表明在设计代理模型时，应考虑代理模型所处的背景，即贝叶斯后验分布；具体而言，应在参数空间中后验分布具有显著概率质量的区域选择训练点。这会带来一个有些循环的问题，因为我们使用高斯过程回归，正是为了能够计算后验分布。然而，实践中已有多种计算方法可用于绕过这一问题，包括一种只需访问近似后验分布的序贯设计策略[40], ，以及在使用马尔可夫链蒙特卡罗等抽样方法探索后验分布的过程中更新训练点 [12, 6]。

本文其余部分的结构如下。第 2 节介绍逆问题中的贝叶斯后验分布，讨论使用代理模型进行计算近似，并证明在由后验分布加权的范数下，真实后验分布与近似后验分布之间的新误差界。主要创新成果是定理 1 和定理 2。第 3 节介绍将高斯过程回归用作代理模型，并证明了逆问题背景下高斯过程回归精度及训练点恰当选取方面的新结果。主要新成果是推论1、推论2和定理 3。第 4 节给出一些简单的数值示例，以说明理论发现。

2 贝叶斯逆问题及其近似

本节将建立贝叶斯逆问题，并介绍如何使用随机代理模型对其进行近似。第 2.1 小节将介绍贝叶斯反演的框架，

并讨论相应的马尔可夫链蒙特卡罗方法。第 2.2 小节包含两个大型示例，说明了代理建模的必要性。第 2.3 小节将详细介绍代理建模，并在第 2.4 小节给出误差估计，总结代理模型中的误差对贝叶斯逆问题解的影响。

2.1 贝叶斯逆问题

我们关注如何求解一个逆问题：根据含噪声的间接数据 $y \in \mathbb{R}^{d_y}$，确定未知参数 $u \in U$，其中数据由下式给出

$$y = \mathcal{G}(u) + \eta, \tag{1}$$

其中，观测算子为 G： $U \to \mathbb{R}^{d_y}$。为便于说明，我们假设噪声 η 是一个取值于 $\mathbb{R}^{d_y}$ 的高斯随机变量 N (0， Γ) 的一次实现，其中 Γ是已知的正定协方差矩阵；不过，也可以用类似的方法处理 η 上的其他分布 [23]。

我们采用贝叶斯统计视角，将二元组 (u， y) 视为一个随机变量（在有限维情形下为 u ）或一个随机过程（在无限维情形下为 u ）。目标是求条件随机变量$u|y$的分布。这种方法使问题适定，符合 Hadamard 的意义：存在唯一的条件分布 $u|y$，且它随 y [41, 20] 连续变化。

在没有数据的情况下，我们假设 u 服从先验测度μ_0。因此，方程 (1) 定义了 $y|u$ 的条件分布，并假设 η 与 u相互独立。 先验测度的适当选择取决于具体应用。接下来，我们关注条件随机变量 $u|y$上的后验分布 μ^y，可通过贝叶斯定理将其刻画如下。这给出了后验分布相对于先验分布的 Radon-Nikodym 导数（例如参见 [41]）。

命题1 假设 U 是可分*Banach*空间，$\mathcal{G}: U \to \mathbb{R}^{d_y}$ 连续，且 $\mu_0(U) = 1$。那么，关于条件随机变量 $u|y$ 的后验分布 μ^y 相对于 μ_0 绝对连续，并由贝叶斯定理给出：

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\exp\big(-\Phi(u)\big),$$

其中

$$\Phi(u) = \frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2, \qquad and \qquad Z = \mathbb{E}_{\mu_0}\Big(\exp\big(-\Phi(u)\big)\Big).$$

在前文中，我们采用记号约定 $\|\cdot\|_A = \|A^{-\frac{1}{2}}\cdot\|_2$，如 [41], 所示, -其中 $\|\cdot\|_2$ 为欧几里得范数， A 为任意对称正矩阵。在有限维情形下，$u \in U \subseteq \mathbb{R}^{d_u}$，且我们要推断有限个

未知参数时，贝叶斯定理可以用先验分布和后验分布的概率密度函数（pdf）来表示，分别记作 π_0 和 π^y[17]。其形式为

$$\pi^y(u) = \frac{1}{Z}\exp\big(-\Phi(u)\big)\pi_0(u),$$

其中 Φ 和 Z 的定义见命题1。

术语 $\exp\big(-\Phi(u)\big)$ 被称为数据似然，它来自于 $y|u$的分布。换言之，它刻画了在给定参数 u的特定取值时，观测到数据 y 的可能性。由于 $y = \mathrm{G}(u) + \eta$ 和 $\eta \sim \mathrm{N}(0,\Gamma)$，我们有 $y|u \sim \mathrm{N}(\mathrm{G}(u),\Gamma)$，因此 $y|u$ 的概率密度函数与 $\exp\big(-\Phi(u)\big)$成正比。归一化常数 Z 确保后验分布 μ^y 是一个概率分布，其中 $\mu^y(U) = 1$。根据贝叶斯定理，$Z = Z(y)$是数据 y的边缘概率密度函数，因此刻画了在给定观测模型G的情况下观测到数据 y 的可能性。Z 因此通常被称为模型证据。Z 的解析值通常未知，而对 Z进行数值计算则 notoriously 困难（参见例如 [47]）。

在实际应用中，通常需要计算与后验分布 μ^y 相关的某个感兴趣量。例如，可以计算期望值 $\mathbb{E}_{\mu^y}[g(u)]$，其中 $g\colon U \to U$ 取为条件均值 $\mathbb{E}[u|y] = \mathbb{E}_{\mu^y}[u]$ 的恒等函数；或者 $g\colon U \to \{0,1\}$ 取为指示函数 $\mathrm{I}_{u\in A}$，用于计算事件概率 $\mathbb{P}[u|y \in A] = \mathbb{E}_{\mu^y}[\mathrm{I}_{u\in A}]$。从后验分布中采样，以便计算期望值及其他感兴趣量，通常首选的方法是马尔可夫链蒙特卡罗（MCMC）[14, 27, 35, 9,13, 10]。一个典型例子是 Metropolis-Hastings 算法，该算法从有限维参数空间 $U \subseteq \mathbb{R}^{d_u}$ 上的目标密度 π 中采样，步骤如下：

1. 选择 $u^{(1)}$，使得 $\pi(u^{(1)}) > 0$。2. 在状态 $u^{(i)}$ 下，从密度 $q(u' \mid u^{(i)})$ 中抽取一个候选样本 u'。3. 以概率接受样本 u'

$$\alpha(u' \mid u^{(i)}) = \min\left(1, \frac{\pi(u')\, q(u^{(i)} \mid u')}{\pi(u^{(i)})\, q(u' \mid u^{(i)})}\right),$$

即 $u^{(i+1)} = u'$以概率 $\alpha(u' \mid u^{(i)})$；否则保持在 $u^{(i+1)} = u^{(i)}$。

步骤2和步骤3会重复执行，直到生成所需数量的样本。在逆问题和贝叶斯后验分布的背景下，我们特别注意到，无需知道目标密度 π 的归一化常数，因为该常数会在 α中的比值里约去。该算法的关键要素是步骤2中的 提议密度 q，其选项范围十分广泛，从简单的随机游走，到使用目标函数的（高阶）导数以及参数空间的几何结构的方法（例如参见 [35, 10, 13]）。Metropolis-Hastings 算法背后有一套丰富的理论支撑；特别是，在温和的假设下，可以保证 $u^{(i)}$ 的分布 $i \to \infty$ 收敛于目标密度π （例如参见 [35]）。

马尔可夫链蒙特卡罗方法通常需要反复计算数据似然。在上述 Metropolis-Hastings 算法中，这一点体现在步骤3：对于每个提议 u'，都需要计算 $\exp\left(-\Phi(u')\right)$。在现代应用中，似然的计算成本非常高，这很快就会变得难以实施。例如，地球物理学中的逆问题（地下流动模型）和气候（大气环流模型）中的逆问题都属于此类，其中 G 涉及求解一个或多个耦合的偏微分方程（偏微分方程）。下一小节将给出两个这样的例子。

2.2 大规模复杂贝叶斯逆问题示例

地下流动模型

稳态地下流体流动的一个简单模型如下

$$-\nabla \cdot (k(x)\nabla p(x)) = g(x), \qquad x \in D, \tag{2}$$

其中，k 表示地下介质的渗透率（或水力传导率），p 表示流体的压力水头，g 包含所有源项或汇项，并在计算域 D [46, 36]上施加适当的边界条件。该模型来源于多孔介质中单相流体流动的达西定律，$q(x) = -k(x)\nabla p(x)$，以及质量守恒定律，$\nabla \cdot q = g$。

在此背景下，一个典型的逆问题是：根据在固定位置 $x_i \in D$处对压力水头 $y = \{p(x_i) + \eta_i\}_{i=1}^{d_y}$ 或达西通量 $y = \{k(x_i)\nabla p(x_i) + \eta_i\}_{i=1}^{d_y}$ 进行的含噪测量，重建渗透率 k。尽管我们通常希望重建函数 $k \in L^2(D)$，但在计算中往往会选择对 k进行参数化。于是，逆问题就转化为学习参数化中的系数$u \in \mathbb{R}^{d_u}$。例如，我们可以选择一个分片常数模型

$$k(x;u) = \sum_{j=1}^{d_u} u_j \mathrm{I}_{x\in D_j}(x),$$

其中，D_j 是对区域 D 的划分，表示不同类型岩石的各个层，而 u_j 是层 D_j 中的渗透率 k 值。由于渗透率始终为正，且不同类型岩石的渗透率可能相差多个数量级，因此，关于 $u = \{u_j\}_{j=1}^{d_u}$ 的先验分布 μ_0 应体现这些特性，通常会使用对数正态分布。

参数到观测映射 $\mathsf{G} : \mathbb{R}_{>0}^{d_u} \to \mathbb{R}^{d_y}$ 的评估，其中 $\mathsf{G}(u) = \{p(x_i;\, u)\}_{i=1}^{d_y}$ 或 $\mathsf{G}(u) = \{k(x_i;\, u)\nabla p(x_i;\, u)\}_{i=1}^{d_y}$，需要针对每个 u 使用不同的系数 k 求解偏微分方程 (2)，因此通常代价高昂。

大气环流模型

大气环流模型 (GCM) 通常由一组三维空间、随时间变化的偏微分方程离散化而成，其空间维度分别表示地球经度、纬度和垂直维度。未知参数出现在子网格尺度模型中，这些模型试图捕捉云尺度上不可解析的物理过程。原则上，可以利用数据学习这些参数，数据形式为对地球大气进行时间平均的卫星测量数据。不确定性量化以及贝叶斯方法至关重要，因为对气候模型中的不确定性进行估计（并降低不确定性）是现代气候建模研究的核心目标 [38]。本文将考虑 [31]中的一个特定 GCM。用于构建计算模型的守恒定律经离散化后，包含两个具有以下形式的方程

$$\begin{aligned}\text{Moisture Conservation:}\quad & \frac{\partial q}{\partial t} + v\cdot\nabla q = -\frac{q - q_{\text{ref}}(T;\theta)}{\tau_q(q,T;\theta)}\\ \text{Energy Conservation:}\quad & \frac{\partial T}{\partial t} + v\cdot\nabla T = \frac{T - T_{\text{ref}}(q,T;\theta)}{\tau_T(q,T;\theta)} + \cdots,\end{aligned}$$

并与质量和动量的其他守恒定律相耦合。该模型包含两个未知参数： θ_{RH}，即参考相对湿度；以及 θ_τ 松弛时间尺度； q_{ref}、 T_{ref}， τ_q 和 τ_T 的函数形式均已知。未知参数为 $\theta = (\theta_{RH},\ \theta_\tau)$。逆问题旨在利用气候统计量，了解 θ 以及 θ 中的不确定性；论文 [12] 采用自由对流层相对湿度、降水率以及极端降水频率测度的 30 天平均值。因此，参数到数据映射需要对离散化的守恒定律进行模拟。这会产生 $O(10^2)$ 个带噪声的间接观测值，据此可以了解这两个未知参数。类似的相关逆问题见 [15][第 5.10], 小节，其中未知参数的数量为 4。

2.3 代理模型

为使计算在实践中可行，通常会用代理模型（也称为仿真器、元模型或降阶模型）来近似对数似然 Φ或正向模型 G。现有多种适用的代理模型，参见例如 [42, 23]及其中的参考文献。近年来，人们尤其关注随机代理模型的应用。研究表明，当代理模型的误差相对于观测值中的噪声较大时，这类模型能够更好地近似后验分布 [8, 2],，并提高马尔可夫链蒙特卡罗方法的统计效率 [11]。随机代理模型的例子包括概率数值学领域的方法 [8, 22],、随机投影方法 [21, 29], 以及高斯过程仿真器 [19, 42]。

给定用于对数似然 Φ 或正向模型 G 的随机代理模型，可以通过不同方式利用它来近似后验分布μ^y。下文将统一采用针对 Φ的代理模型，因为根据似然的具体形式，针对 G 的代理模型可以转换为针对 Φ的代理模型。将随机代理模型记为 Φ_N，其中 N 表示离散化参数，例如高斯过程模拟器中的训练点数量，或概率偏微分方程求解器所用的网格点数量。利用代理模型的均值 $\mathbb{E}[\Phi_N]$ 来近似 Φ，即可得到基于均值的近似

$$\frac{d\mu^{y,N}_{\text{mean}}}{d\mu_0}(u) = \frac{1}{Z_N^{\text{mean}}}\exp\big(-\mathbb{E}[\Phi_N(u)]\big), \tag{3}$$
$$Z_N^{\text{mean}} = \mathbb{E}_{\mu_0}\Big(\exp\big(-\mathbb{E}[\Phi_N(u)]\big)\Big).$$

或者，我们可以直接使用随机代理模型 Φ_N 来近似 Φ，并取由此得到的似然随机近似的期望值。这便得到边缘近似

$$\frac{d\mu^{y,N}_{\text{marginal}}}{d\mu_0}(u) = \frac{1}{Z_N^{\text{marginal}}}\mathbb{E}\Big(\exp\big(-\Phi_N(u)\big)\Big), \tag{4}$$
$$Z_N^{\text{marginal}} = \mathbb{E}_{\mu_0}\left(\mathbb{E}\Big(\exp\big(-\Phi_N(u)\big)\Big)\right).$$

直观而言，边缘近似会在近似后验分布中引入额外的不确定性，这种不确定性源于代理模型的随机性。对代理模型随机性的一种解释是：由于构建模型时使用的信息量/资源有限，因此模型本身仍存在一定的不确定性（或误差）。在近似后验分布中纳入这种不确定性，便可以在推断参数 u时考虑代理模型中的误差。在实际应用中，由于计算资源有限，代理模型的精度可能受到限制，其不确定性（或误差）可能较大，甚至与观测值 y中的不确定性相当。因此，务必将其纳入考虑，以避免产生过度自信且有偏的预测 [8, 2]。

在特殊情形下，可以通过解析推导证明，边缘近似会导致似然发生某种形式的方差膨胀。其中一种情形是，正向模型 G 可用高斯过程 $\mathsf{G}_N \sim \mathrm{GP}(m_N^{\mathcal{G}}, K_N)$ 进行近似，其均值为 $m_N^{\mathcal{G}}$，协方差核为 K_N（更多详情见第 3 节）。基于均值的近似假设观测模型为 $y = m_N^{\mathcal{G}}(u) + \eta$，而边缘近似则使用

$$y = \mathcal{G}_N(u) + \eta = m_N^{\mathcal{G}}(u) + \xi(u) + \eta,$$

其中 $\xi(u) \sim \mathrm{N}(0,\ K_N(u,\ u))$。由此得到的近似似然与以下表达式成正比

$$\frac{1}{\sqrt{\det(\Gamma)}} \exp\left(-\frac{||y - m_N^{\mathcal{G}}(u)||_\Gamma^2}{2}\right)$$

对于基于均值的近似，以及

$$\frac{1}{\sqrt{\det(\Gamma + K_N(u,u))}} \exp\left(-\frac{||y - m_N^{\mathcal{G}}(u)||^2_{(\Gamma+K_N(u,u))}}{2}\right)$$

对于边缘近似，参见例如 [5, 7, 2]。因此，两种似然之间的差异取决于 Γ（观测值 y 的不确定性）与 K_N（代理模型 Φ_N 的不确定性）之间的关系。如果 K_N 相较于 Γ 较小，那么两种近似后验将会相似。

从决策理论视角 [40, 16] 出发，也可以为近似后验分布提供动机。假设 U 是 $\mathbb{R}^{d_u}$ 的紧子集，并且我们用高斯过程 $\Phi_N \sim \mathrm{GP}(m_N^\Phi, k_N)$ 近似负对数似然 Φ。那么，（未归一化的）边缘近似可使期望 L^2-损失最小化：

$$\mathbb{E}\Big(\exp\big(-\Phi_N(u)\big)\Big)\pi_0(u) = \underset{f\in L^2(U)}{\arg\min}\ \mathbb{E}\int_U \big(\exp\big(-\Phi_N(u)\big)\pi_0(u) - f(u)\big)^2\,\mathrm{d}u.$$

而（未归一化的）基于均值的近似则使期望 L^1-损失最小化：

$$\exp\big(-m_N^\Phi(u)\big)\pi_0(u) = \underset{f\in L^1(U)}{\arg\min}\ \mathbb{E}\int_U \Big|\exp\big(-\Phi_N(u)\big)\pi_0(u) - f(u)\Big|\,\mathrm{d}u.$$

注 *1*（其他近似后验）除了上述近似之外，我们还可以考虑基于样本的近似

$$\frac{d\mu^{y,N}_{\text{sample}}}{d\mu_0}(u) = \frac{1}{Z_N^{\text{sample}}}\exp\big(-\Phi_N(u)\big),$$
$$Z_N^{\text{sample}} = \mathbb{E}_{\mu_0}\left(\exp\big(-\Phi_N(u)\big)\right),$$

这会得到后验分布 [42, 23] 的随机近似。如果随机代理模型 Φ_N 存在较大的不确定性，那么与 Φ_N 的不同样本对应的后验分布近似可能会大相径庭。

另一种边缘近似可以通过计算上述归一化随机似然的期望值来定义，

$$\frac{d\mu^{y,N}_{\text{marginal}'}}{d\mu_0}(u) = \mathbb{E}\left(\frac{1}{Z_N^{\text{sample}}}\exp\big(-\Phi_N(u)\big)\right).$$

然而，这种近似并不容易适用于（伪边缘）马尔可夫链蒙特卡罗等抽样方法，因为 Z_N^{sample} 的计算难以处理。 ■

2.4 代理模型加速的后验分布中的误差

为了证明近似后验分布 $\mu^{y,N}_{\text{mean}}$ 和 $\mu^{y,N}_{\text{边缘}}$在实践中的适用性，我们希望证明，当 $N \to \infty$ 时，它们收敛到真实后验 μ^y。我们将使用 Hellinger距离来度量误差

$$d_{\text{Hell}}(\mu^y, \mu^{y,N}) = \left(\frac{1}{2} \int_U \left(\sqrt{\frac{d\mu^y}{d\nu}} - \sqrt{\frac{d\mu^{y,N}}{d\nu}} \right)^2 d\nu \right)^{1/2},$$

其中， ν 是任意测度，使得 μ^y 和 $\mu^{y,N}$ 均相对于 ν 绝对连续（且距离的取值与 ν.的选择无关）。在 [23, 42, 43] 中，根据问题的正则性，分别针对 p、q_1 和 q_2，以$\|\Phi - \mathbb{E}[\Phi_N]\|_{L^p_{\mu_0}(U)}$和 $\left\| \mathbb{E}\left[|\Phi - \Phi_N|^{q_1} \right]^{1/q_1} \right\|_{L^{q_2}_{\mu_0}(U)}$表示，得到了关于 $d_{\text{Hell}}(\mu^y, \mu^{y,N}_{\text{mean}})$ 和 $d_{\text{Hell}}(\mu^y, \mu^{y,N}_{\text{边缘}})$ 的界。这些结果表明，在贝叶斯推断中，只要 Φ_N 在先验分布 μ_0赋予显著质量的所有区域内都能很好地近似 Φ，就足以实现准确的代理建模。然而，我们实际上认为， Φ_N 只需在后验分布 μ^y 赋予显著质量的区域内很好地近似 Φ。以下结果通过在选择 $\nu = \mu^y$的情况下分析 Hellinger距离得出，体现了这一点。

定理1（基于均值近似的收敛性）假设存在与 N无关的标量 C_1、 $C_2 \geq 0$，使得对于赫尔德共轭指数对（p_1, p_1'），有[1]

(i) $\|\exp((\Phi - \mathbb{E}[\Phi_N]))\|_{L^{p_1}_{\mu^y}(U)} \leq C_1(p_1)$;
(ii) $C_2^{-1} \leq Z_N^{\text{mean}} \leq C_2$.
那么存在常数 $C_{\text{Thm1}} = C_{\text{Thm1}}(C_1, C_2, Z)$，与 N无关，使得

$$d_{\text{Hell}}(\mu^y, \mu^{y,N}_{\text{mean}}) \leq C_{\text{Thm1}} \|\Phi - \mathbb{E}[\Phi_N]\|_{L^{2p_1'}_{\mu^y}(U)}.$$

定理2（边际近似的收敛性）假设存在与 N 无关的标量 C_1、 $C_2 \geq 0$，使得对于赫尔德共轭指数对 (p_1, p_1') 和 (p_2, p_2')，有[1]

(i) $\left\| \mathbb{E}\left[(\exp(\Phi - \Phi_N))^{p_2} \right]^{1/p_2} \right\|_{L^{2p_1}_{\mu^y}(U)} \leq C_1(p_1, p_2)$

[1] 对于 $p_1 = \infty$，假设(i)要求对标准 $L^\infty(U)$范数进行界定，这是由于相应的赫尔德不等式 $\int_U f\, g \mu^y(\mathrm{d}u) \leq \sup_{u \in U} |f(u)| \int_U |g| \mu^y(\mathrm{d}u) = \|f\|_{L^\infty(U)} \|g\|_{L^1_{\mu^y}(U)}$。

(ii) $C_2^{-1} \leq Z_N^{\text{marginal}} \leq C_2$.

那么存在常数 $C_{\text{Thm2}} = C_{\text{Thm2}}(C_1, C_2, Z)$，且该常数与 N 无关，使得

$$d_{\text{Hell}}\left(\mu, \mu^{y,N}_{\text{marginal}}\right) \leq C_{\text{Thm2}} \left\| \mathbb{E}\left[(\Phi - \Phi_N)^{p_2'} \right]^{1/p_2'} \right\|_{L^{2p_1'}_{\mu^y}(U)}.$$

请注意，这些结果适用于一般情形，即参数空间 U是巴拿赫空间，并不要求 U 是有限维的。这些结果的证明与 [42, 23], 中的证明类似，可见附录。上述定理中的假设通常适用于任意随机代理模型 Φ_N，事实上也适用于噪声 η 上的一般分布。这些假设需要针对具体情况逐一验证。第 3 节针对高斯噪声以及基于高斯过程回归的代理模型进行了验证。

定理 1 和定理 2 中的假设（i）与代理模型 Φ_N的可积性有关。两个定理中都出现的指数 p_1与 u尾部的衰减率有关。指数 p_2与 Φ_N分布尾部的衰减率有关，仅定理 2 需要该指数，因为基于均值的近似 $\mu^{y,N}_{\text{mean}}$不利用 Φ_N中的随机性。定理 1 和定理 2 中的假设（ii）确保归一化常数 Z_N^{mean}和 Z_N^{marginal} 有定义，从而保证近似后验有定义。

3 高斯过程回归

现在，我们希望使用高斯过程回归为数据似然构建一个代理模型。高斯过程回归（又称高斯过程仿真或克里金法）是一种基于在选定的一组训练点（又称设计点）上对 f 进行有限次评估，从而构建函数 f 近似的方法。

这里我们将考虑对参数到观测映射 $\mathsf{G}\colon U \to \mathbb{R}^{d_y}$ 或负对数似然 $\Phi\colon U \to \mathbb{R}$ 进行仿真。为简单起见，我们将重点讨论标量值函数的仿真，并假设在 $d_y > 1$ 的情况下，通过分别仿真 G 的每个分量来构建其仿真器。不过，我们在此指出，在实践中，纳入 G 不同分量之间的相关性，可以显著提高近似后验分布 [2] 的质量。

在第 3.1 小节中，我们建立高斯过程回归的框架；随后在第 3.2 小节中，讨论如何在代理建模中使用这种回归技术。第 3.3 小节回到第 2.2 小节中的两个示例，以说明与原始模型 f相比，代理模型 f_N 有望降低计算成本。第 3.4 小节专门讨论近似后验分布的误差估计，该估计以用于近似后验分布中似然贡献的 GP 代理模型的误差为依据。第 3.5 小节则给出高斯过程回归的误差估计，这些估计可用于上一小节的分析。

3.1 设置

设 $U \subseteq \mathbb{R}^{d_u}$ 为有限维空间，并设 $f\colon U \to \mathbb{R}$ 为任意函数。高斯过程回归是一种贝叶斯方法，其出发点是为函数 f设定一个高斯过程先验。换言之，我们将 f建模为高斯过程的一个样本

$$f_0 \sim \mathrm{GP}(m(u), k(u,u')), \tag{7}$$

其中，$m : U \to \mathbb{R}$ 选定的均值函数，给出 $\mathbb{E}[f_0(u)] = m(u)$，以及$k : U \times U \to \mathbb{R}$ 选定的对称正定协方差核，给出$\mathbb{C}\mathrm{ov}(f_0(u), f_0(u')) = k(u, u')$。这意味着，对于任意一组点 $\{\tilde{u}^m\}_{m=1}^M \subseteq U$，向量 $[f_0(\tilde{u}^1); \ldots; f_0(\tilde{u}^M)] \in \mathbb{R}^M$服从多元高斯分布，其均值为 $[m(\tilde{u}^1); \ldots; m(\tilde{u}^M)] \in \mathbb{R}^M$，协方差矩阵为 $K \in \mathbb{R}^{M\times M}$，其元素为$K_{i,j} = k(\tilde{u}^i, \tilde{u}^j)$。

均值函数m的典型选择包括零函数和多项式[33]。应用中常用的一类协方差函数是 Matèrn协方差函数 [26],，其定义为

$$k_{\nu,\lambda,\sigma_k^2}(u,u') = \sigma_k^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\|u-u'\|_2}{\lambda}\right)^{\nu} B_\nu \left(\frac{\|u-u'\|_2}{\lambda}\right), \tag{8}$$

其中，Γ 表示伽马函数，B_ν 表示第二类修正贝塞尔函数，ν、λ和 σ_k^2为正参数。参数 λ称为相关长度，决定了 $f_0(u)$ 与 $f_0(u')$ 之间产生相关性的长度尺度。参数 σ_k^2是边际方差 $\sigma_k^2 = k_{\nu,\lambda,\sigma_k^2}(u,u) = \mathbb{V}[f_0(u)]$, ，并决定了 $f_0(u)$ 的典型幅度。最后，参数 ν 称为光滑度参数，决定了 f_0 关于u的样本路径的正则性。对于任意 $\epsilon > 0$，f_0 的样本路径几乎必然属于索伯列夫空间 $H^{\nu-\epsilon}(U)$，例如参见 [18]。

在 $\nu \to \infty$ 的极限下，我们得到高斯协方差函数

$$k_{\infty,\lambda,\sigma_k^2}(u,u') = \sigma_k^2 \exp\left(-\frac{\|u-u'\|_2^2}{2\lambda^2}\right),$$

也称为平方指数函数或径向基函数（RBF）协方差函数。当 ν 为半整数时，Matèrn 协方差函数的公式还会进一步简化，常见的选择包括

$$k_{\frac{1}{2},\lambda,\sigma_k^2}(u,u') = \sigma_k^2 \exp\left(-\frac{\|u-u'\|_2}{\lambda}\right),$$
$$k_{\frac{3}{2},\lambda,\sigma_k^2}(u,u') = \sigma_k^2 \left(1+\sqrt{3}\frac{\|u-u'\|_2}{\lambda}\right)\exp\left(-\sqrt{3}\frac{\|u-u'\|_2}{\lambda}\right).$$

选择 $\nu = \frac{1}{2}$ 会得到指数协方差函数，也称为拉普拉斯协方差函数；在一维输入变量 u 的情况下，它是Ornstein-Uhlenbeck过程的协方差函数。

图1展示了高斯过程 f_0 with$m = 0$ 的5条独立样本路径，以及Matérn 协方差函数中参数的4种不同取值。图中还以黑色实线表示均值 $\mathbb{E}[f(u)] = m(u) = 0$，并以均值周围的浅灰色区间表示边缘标准差 $\sqrt{k(u,u)} = \sigma = 1$。上排的两幅图展示了$\nu = \frac{1}{2}$，此时样本路径连续但不可微；下排的两幅图展示了 $\nu = \infty$，此时样本路径无限光滑。左列显示较长的相关长度 λ，因此主要呈现大尺度变化；右列显示较短的相关长度，因此小尺度波动更多。

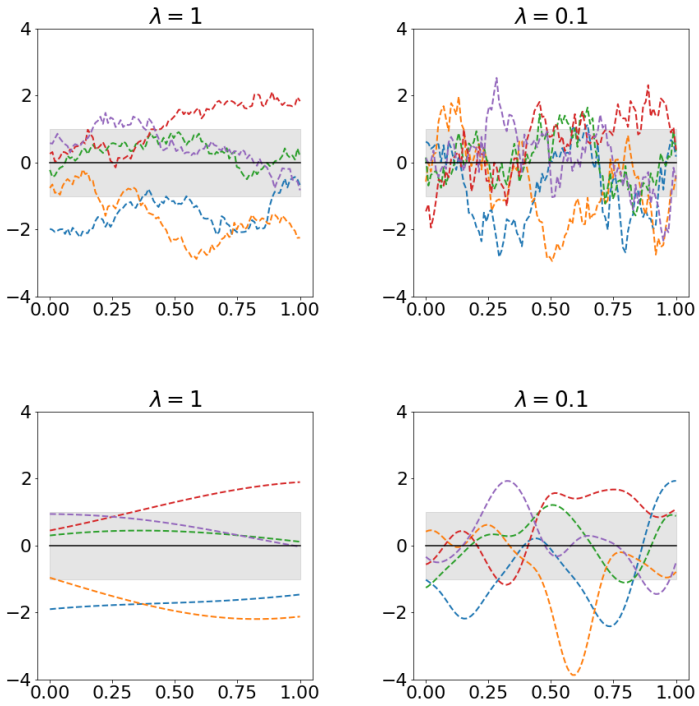

图1：高斯过程 f_0 with Matérn 协方差函数$k_{\nu,\lambda,\sigma_k^2}$ 的样本路径。左上：$\{\nu, \lambda, \sigma_k^2\} = \{1\,2, 1, 1\}$。右上：$\{\nu, \lambda, \sigma_k^2\} = \{1\,2, 0.1, 1\}$。左下：$\{\nu, \lambda, \sigma_k^2\} = \{\infty, 1, 1\}$。右下：$\{\nu, \lambda, \sigma_k^2\} = \{\infty, 0.1, 1\}$。

现在假设给定一组互不相同的训练点$D_N := \{u^n\}_{n=1}^N \subseteq U$以及相应的函数值

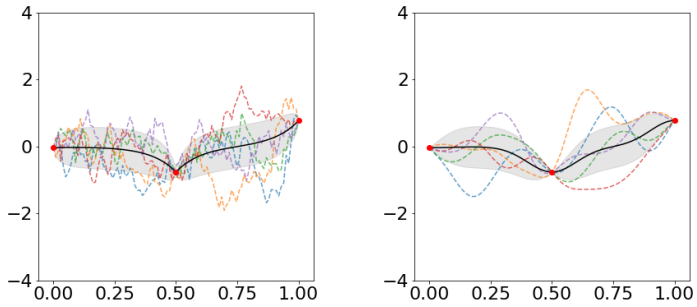

图 2：具有 Matérn 协方差函数的高斯过程 f_3 的样本路径 $k_{\nu,\ \lambda,\ \sigma_k^2}$，其中 $f(x) = \sin(x-2.5)^2$。左：$\{\nu,\ \lambda,\ \sigma_k^2\} = \{1/2,\ 0.1,\ 0.63^2\}$ 右：$\{\nu,\ \lambda,\ \sigma_k^2\} = \{\infty,\ 0.1,\ 0.63^2\}$。

$$f(D_N) := [f(u^1); \ldots; f(u^N)] \in \mathbb{R}^N. \tag{9}$$

由于 f_0 是一个高斯过程，对于任意一组测试点 $\{\tilde{u}^m\}_{m=1}^M \subseteq U$，向量 $[f_0(u^1); \ldots; f_0(u^N); f_0(\tilde{u}^1); \ldots, f_0(\tilde{u}^M)] \in \mathbb{R}^{N+M}$ 服从多元高斯分布。给定值 $f_0(u^1) = f(u^1), \ldots, f_0(u^N) = f(u^N)$ 后，$f_0(\tilde{u}^1), \ldots, f_0(\tilde{u}^M)$ 的条件分布仍为高斯分布，其均值和协方差由高斯随机变量条件分布的标准公式给出 [33]。因此，基于已知值 $f(D_N)$ 对高斯过程（7）进行条件化，可得到另一个高斯过程 f_N，称为预测过程。我们有

$$f_N \sim \mathrm{GP}(m_N^f(u), k_N(u,u')), \tag{10}$$

其中，预测均值 $m_N^f\colon U \to \mathbb{R}$ 和预测协方差 $k_N\colon U \times U \to \mathbb{R}$ 均有显式表达式，且取决于式 (7) 中所作的建模选择。从现在起，我们将聚焦于常用的选择 $m \equiv 0$；非零均值的情形将在备注 2 中讨论。当 $m \equiv 0$ 时，我们有

$$m_N^f(u) = k(u, D_N)^T K(D_N, D_N)^{-1} f(D_N), \tag{11}$$

$$k_N(u,u') = k(u,u') - k(u, D_N)^T K(D_N, D_N)^{-1} k(u', D_N), \tag{12}$$

其中，$k(u,\ D_N) = [k(u,\ u^1); \ldots; k(u,\ u^N)] \in \mathbb{R}^N$ 以及 $K(D_N,\ D_N) \in \mathbb{R}^{N \times N}$ 是一个矩阵，其 ij^{th} 个元素等于 $k(u^i,\ u^j)$ [33]。

图2与图1类似，展示了在 Matèrn 协方差核中选择不同超参数时预测过程 f_N（使用 $N = 3$）的样本。我们还展示了更新后的均值 $\mathbb{E}[f_N(u)] = m_N^f(u)$ 以及更新后的边缘标准差 $\sqrt{k_N(u,\ u)}$。我们看到，超参数的选择会对预测过程的行为产生显著影响。良好的取值

超参数通常并非先验已知，而是必须根据式（9）中的观测值 $f(D_N)$ 进行估计（例如参见 [33] 及其中的参考文献）。图2中使用 scikit-learn [32] 对 λ 和 σ^2 进行了估计。

由 (11) 可知，m_N^f 在训练点 D_N 处插值函数 f，因为向量 $k(u^n, D_N)$ 是矩阵 $K(D_N, D_N)$ 的第 n^{th} 行。换言之，对于所有 $n = 1, \ldots, N$，我们有 $m_N^f(u^n) = f(u^n)$。对于预测协方差 k_N，注意到对于所有 $u \in U$，$k_N(u,u) < k(u,u)$，因为根据假设，$K(D_N, D_N)$ 是正定的。此外，我们还注意到，对于 $n = 1, \ldots, N$，$k_N(u^n, u^n) = 0$，因为 $k(u^n, D_N)^T K(D_N, D_N)^{-1} k(u^n, D_N) = k(u^n, u^n)$。

备注 *2*（非零均值先验）如果在 (7) 中使用非零均值 $m(\cdot)$，预测均值 m_N^f 的公式将变为

$$m_N^{f,m}(u) = m(u) + k(u, D_N)^T K(D_N, D_N)^{-1}(f(D_N) - m(D_N)),$$

其中 $m(D_N) := [m(u^1); \ldots; m(u^N)] \in \mathbb{R}^N$。预测协方差 $k_N(u, u')$ 与 (11) 中相同。与 $m \equiv 0$ 的情况一样，对于 $n = 1, \ldots, N$，我们有 $m_N^f(u^n) = f(u^n)$，并且 m_N^f 是 f 的插值函数。在对 m 作出适当假设的条件下，在 $m \equiv 0$ 情况下推导出的任何误差界都可以推广到一般情况，参见例如 [43]。■

备注 *3*（含噪函数值）如果我们观测到的不是像 (9) 中那样的精确函数值，而是含噪函数值

$$d := [f(u^1) + \varepsilon_1; \ldots; f(u^N) + \varepsilon_N] \in \mathbb{R}^N,$$

其中 $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$ 为独立同分布，则上述给出的预测过程 f_N 的公式仍然成立，只需将 $K(D_N, D_N)$ 替换为 $K(D_N, D_N) + \sigma^2 \mathrm{I}$。■

3.2 高斯过程回归作为代理模型

高斯过程回归主要有两种应用场景：(i) 生成输入输出对 $\{u^n, f(u^n)\}_{n=1}^N$ 的真实函数 f 未知；或者 (ii) 生成 $\{u^n, f(u^n)\}_{n=1}^N$ 的函数 f 已知，但计算开销极大。在场景 (i) 中，我们希望根据观测到的输入输出对，学习函数 f 是什么。这通常基于真实数据，因此，假设观测到的函数值包含噪声是很常见的。在场景 (ii) 中，我们希望构建一种代理模型（又称降阶模型、元模型或仿真器），其计算成本低于原始数学模型 f。这通常基于通过在计算机上模拟数学模型而获得的数据 $\{u^n, f(u^n)\}_{n=1}^N$，因此，假设函数值中不含噪声可能更为恰当。

为了为 f 构建代理模型，我们在输入 $\{u^n\}_{n=1}^N$ 处模拟 N 次模型运行，以获得输出 $\{f(u^n)\}_{n=1}^N$，然后遵循上一节概述的贝叶斯方法，得到预测过程 $f_N \sim \mathrm{GP}(m_N^f(u), k_N(u, u'))$。因此，它与许多其他基于模型运行或快照的代理模型类似。

高斯过程 f_N 是 f 的随机代理模型。f_N 中的不确定性由预测方差 k_N 表示，它刻画了在观测到有限信息量 $\{u^n,\ f(u^n)\}_{n=1}^N$ 后，关于 f 尚存的不确定性；换言之，它是在假设 f 是高斯过程先验 (7) 的一个样本的基础上，对 f 与 m_N^f 之间误差的建模。在训练点 D_N 处，我们已观测到 f，因此对于 f 取何值不存在不确定性，m_N^f 中也不存在误差（即 $k_N(u^n,\ u^n) = 0$ 且 $m_N^f(u^n) = f(u^n)$，参见图 2）。远离训练点时，由于我们不知道 f 的值，因此仍存在一定的不确定性/误差（一般而言，即 $k_N(u,\ u) > 0$ 且 $m_N^f(u) \neq f(u)$）。

接下来需要说明，f_N 确实是一个代理模型，也就是说，相比原始模型 f，对其进行评估的成本要低得多。为此，请注意，在第 2.3 节的近似后验中，我们通常需要计算预测均值 m_N^f和预测方差 k_N。从式 (11) 可以看出，预测均值 m_N^f 是核函数评估的线性组合：

$$m_N^f(u) = \sum_{n=1}^{N} \alpha_n k(u, u^n), \qquad \alpha = K(D_N, D_N)^{-1} f(D_N) \in \mathbb{R}^N.$$

因此，要在未观测输入 $u \in U \setminus D_N$处计算 $m_N^f(u)$，我们只需计算上述表达式右侧的和。首先，求和项的数量 N 通常较少。这对应于我们需要运行原始模型 f的次数；在许多实际应用中，这一次数会受到计算资源的限制。其次，系数 α 的计算通常可能需要一定成本，因为这涉及对 $K(D_N,\ D_N)$进行 Cholesky分解（或类似运算），其计算成本为 $O(N^3)$；不过，如前所述，在本文所考虑的代理建模应用中，N 通常较小。此外，系数 α只需计算一次，而不必针对每个 u都重新计算。第三，核 $k(u,\ u^n)$的计算成本较低，因为这通常只需计算多项式和指数函数等标准函数。对预测方差 $k_N(u,\ u)$的计算也有类似的论证。

3.3 高斯过程回归作为代理模型的示例

我们回到第 2.2 小节介绍的两个示例，以说明采用高斯过程代理建模所带来的潜在加速效果。

地下流动示例中的计算计时

考虑第 2.2 小节中的地下流动示例。下表给出了具有代表性的计算计时，对比了偏微分方程解 $f(u)$ 的求值成本与高斯过程

代理建模中的两项主要成本，即确定均值 α 的核表示系数，以及计算均值 m_N^f。更多细节和更全面的数值实验见 [2]。给定参数 u 后，对代表真实模型 f 的偏微分方程进行数值求解，该过程在 Firedrake [34] 中实现。

设置	正在计算 $f(u)$	正在计算 $m_N^f(u)$	正在计算 α
$d_u = 2, d_y = 10, D =$ （0,1），$N = 10$	2.6×10^{-1}秒	3.6×10^{-5}秒	3.2×10^{-3}s
$d_u = 2, d_y = 10, D = (0,1), N = 20$	2。6×10^{-1}秒	4。0×10^{-5}秒	1.2×10^{-2}秒
$d_u = 10, d_y = 10, D = (0,1), N = 10$	2.6×10^{-1}秒	5.6×10^{-5}秒	3.2×10^{-3}秒

GCM示例的计算动机

在论文 [12] 中，高斯过程回归被用于加速对气候建模中一个理想化大气环流模型（GCM）参数进行的贝叶斯推断，该模型在第 2.2 小节中有所介绍。参数到数据的映射计算成本高昂，而本文概述的高斯过程代理模型可用于加速计算。具体而言，[12] 中考虑的示例包含 2 个未知参数，观测值的数量级为 $O(10^2)$，而计算模型中的变量数量级为 $O(10^5)$；通过使用高斯过程代理模型，可以避开在 $O(10^5)$ 维空间中进行计算，从而显著降低计算成本。衡量计算成本的自然指标是：为执行马尔可夫链蒙特卡罗（MCMC）并求解贝叶斯逆问题而需要进行的 GCM 评估次数。不使用仿真时，该次数为 $O(10^5)$；使用高斯过程仿真，则只需 $O(10^2)$ 次正向模型评估 [12]，便可获得精度相同的贝叶斯逆问题解。这些评估用于为高斯过程提供训练数据，尤其是采用基于集合卡尔曼的校准-模拟-采样方法来设计这些训练数据时所需的评估 [6]。为了给校准-模拟-采样方法奠定坚实的理论基础，我们在本文中研究了贝叶斯逆问题背景下高斯过程代理建模的设计点最优选择问题。

3.4 高斯过程加速的后验分布误差

现在，假设我们在近似后验分布（3）和（4）中使用高斯过程回归作为代理模型。应用定理 1 和定理

2 可得以下误差界。我们用 $H_k(U)$ 表示先验协方差核 k （见例如 [44]）对应的再生核希尔伯特空间（RKHS），该核见（7）。

推论1 假设 $U \subseteq \mathbb{R}^{d_u}$, $\Phi \in H_k(U)$，并通过对 Φ 应用高斯过程回归构造随机代理模型，得到 $\Phi_N \sim GP(m_N^{\Phi}(u),\ k_N(u,u'))$。那么存在常数 C_{Cor1}, $C'_{\mathrm{Cor1}} > 0$，它们与 N 无关，使得

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y,N}_{\mathrm{mean}}) \leq C_{\mathrm{Cor1}} \left\| \Phi - m_N^{\Phi} \right\|_{L^2_{\mu^y}(U)},$$

并且，在附加假设 U 有界且具有利普希茨边界，以及 上确界$_{u\in U}\, k_N(u,u) \to 0$ （当 $N \to \infty$ 时）的条件下，

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y,N}_{\mathrm{marginal}}) \leq C'_{\mathrm{Cor1}} \left(\left\| \Phi - m_N^{\Phi} \right\|_{L^2_{\mu^y}(U)} + \left\| k_N^{1/2}(\cdot,\cdot) \right\|_{L^2_{\mu^y}(U)} \right).$$

推论2 假设 $U \subseteq \mathbb{R}^{d_u}$, $\mathcal{G}^j \in H_k(U)$ 对于 $j = 1, \ldots, d_y$ 均成立，并通过对 $\mathcal{G}$ 的各个分量分别应用高斯过程回归，构造随机代理模型，得到 $\mathcal{G}^j_N \sim GP(m_N^{\mathcal{G}^j}(u),\ k_N(u,u'))$ 以及 $\Phi_N(u) = \frac{1}{2}\|y - \mathcal{G}_N(u)\|^2_{\Gamma}$。那么存在常数 C_{Cor2}, $C'_{\mathrm{Cor2}} > 0$，它们与 N 无关，使得

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y,N}_{\mathrm{mean}}) \leq C_{\mathrm{Cor2}} \sum_{j=1}^{d_y} \left\| \mathcal{G}^j - m_N^{\mathcal{G}^j} \right\|_{L^2_{\mu^y}(U)},$$

并且，在附加假设 U 有界且具有利普希茨边界，以及 $\sup_{u\in U} k_N(u,u) \to 0$ 当 $N \to \infty$时，

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y,N}_{\mathrm{marginal}}) \leq C'_{\mathrm{Cor2}} \left(\sum_{j=1}^{d_y} \left\| \mathcal{G}^j - m_N^{\mathcal{G}^j} \right\|_{L^2_{\mu^y}(U)} + \left\| k_N^{1/2}(\cdot,\cdot) \right\|_{L^2_{\mu^y}(U)} \right).$$

上述结果的证明与 [42] 中类似，后者使用由先验测度 μ_0 加权的 L^2-范数 $L^2_{\mu_0}(U)$。证明概要见附录。然而，与 [42], 不同，本文推导的新结果中的所有误差度量均使用由真实后验测度 μ^y 加权的 L^2-范数$L^2_{\mu^y}(U)$。这使我们能够推断，随机代理模型Φ_N在参数空间中 μ^y 未赋予显著质量的部分区域内的表现，并不会对近似后验分布的准确性产生重大影响。

对于基于均值的近似（3），结果表明，我们只需分别控制预测均值 m_N^{Φ} 和 $m_N^{\mathcal{G}^j}$ 的误差。对于边缘近似（4），我们还需要控制预测方差 k_N。

3.5 加权空间中高斯过程回归的误差界

现在，我们研究当 $N \to \infty$时，预测均值 m_N^f和预测方差 k_N的行为。对于我们希望模拟的一般函数 $f\colon\ U \to \mathbb{R}$，利用表示定理（[33, 39]），可以证明式（11）中定义的预测均值 m_N^f是先验协方差核 k的再生核希尔伯特空间 H_k（U）中的最小范数插值函数：

$$m_N^f = \underset{\substack{g\in H_k(U)\ \text{s.t.}\\ g(u^{(n)})=f(u^{(n)}),\,1\le n\le N}}{\arg\min} \|g\|_{H_k(U)}. \tag{13}$$

对于式 (8) 中定义的 Matèrn 核，其再生核希尔伯特空间作为向量空间等于索伯列夫空间 $H^{\nu+d_u/2}(U)$，且范数 [44] 等价。因此存在常数 $C_{\text{low}}(U), C_{\text{up}}(U) > 0$，使得对于所有 $g \in H_k(U)$，我们有

$$C_{\text{low}}(U)\|g\|_{H^{\nu+d_u/2}(U)} \le \|g\|_{H_k(U)} \le C_{\text{up}}(U)\|g\|_{H^{\nu+d_u/2}(U)}. \tag{14}$$

给定训练点 D_N，我们引入映射 $m_N^{(\cdot)}\colon\ H_k(U) \to H_k(U)$，该映射基于式 (11) 中给出的预测均值 m_N^f。令 $g(D_N) := [g(u^1); \ldots; g(u^N)] \in \mathbb{R}^N$ 如式 (9)，我们定义

$$g(u) \mapsto m_N^g(u) = k(u, D_N)^T K(D_N, D_N)^{-1} g(D_N). \tag{15}$$

预测方差 k_N 满足命题 2 中的等式，由此可将关于 m_N^f 的收敛结果转化为关于 k_N[37, 42] 的收敛结果。

命题 2 假设 k_N和 $m_N^{(\cdot)}$分别由 (12)和 (15)给出。那么对于任意 $u \in U$，我们有

$$k_N(u,u)^{\frac{1}{2}} = \sup_{\substack{g\in H_k(U)\ s.t.\\ \|g\|_{H_k(U)}=1}} |g(u) - m_N^g(u)|.$$

˜对于任意有界集合 $U \subseteq U$，我们定义填充距离 $h_{D_N,\tilde{U}}$ 为

$$h_{D_N,\tilde{U}} := \sup_{u\in\tilde{U}} \inf_{u^n\in D_N\cap\tilde{U}} \|u - u^n\|_2.$$

˜填充距离是集合 U 中任意一点到某个训练点 $u^n \in D_N \cap \tilde{U}$的最大距离。我们要求训练点 D_N构成空间填充点集（位于 $\tilde{U}$中），以确保当 $N \to \infty$时，填充距离趋于零。以下结果由散乱数据逼近中的标准结果 [44, 28, 1]推得。

˜**命题 3** 假设 $U \subseteq U$ 是一个满足角度为 θ的内部锥条件的利普希茨区域，且 $\tilde{U}$ 包含于立方体 $B(u_c, R_c) = \{u \in \mathbb{R}^{d_u} : \|u - u_c\|_\infty \le R_c\}$中，其中 $u_c \in \mathbb{R}^{d_u}$和 $0 < R_c < \infty$为某些值。进一步假设，再生核希尔伯特空间 $H_k(\tilde{U})$与索伯列夫空间 $H^\tau(\tilde{U})$及 $f \in H^\tau(\tilde{U})$同构。那么

存在常数 $C_{\text{Prop3}}, C'_{\text{Prop3}} > 0$，它们与 f、D_N和 N无关，使得

$$\|f - m_N^f\|_{H^\beta(\tilde{U})} \le C_{\text{Prop3}}(\theta) h_{D_N,\tilde{U}}^{\tau-\beta} \|f\|_{H^\tau(U)}, \qquad \text{for any } \beta \le \tau,$$

对于所有满足 D_N且 $h_{D_N,\tilde{U}} \le h_0(\tilde{U}) = C'_{\text{Prop3}} R_c$ 的集合。

命题 3 使我们能够在子域 $\tilde{U}$ 中局部地界定误差 $f - m_N^f$，也就是说，可以根据局部填充距离$h_{D_N,\tilde{U}}$来界定 $\|f - m_N^f\|_{H^\beta(\tilde{U})}$。证明见附录。

对于许多点集 D_N，填充距离 $h_{D_N,\tilde{U}}$ 关于 N 的变化都可以明确刻画，参见例如 [43, 45] 及其中的参考文献。当 $N \to \infty$ 时，最快可能的衰减率为 $h_{D_N,\tilde{U}} = O(N^{-\frac{1}{d_u}})$，例如均匀张量网格即可达到该速率。如下文命题4所示，随机采样点集的填充距离也以（几乎）最优速率衰减。

命题4 假设

(i) $\tilde{U} \subseteq \mathbb{R}^{d_u}$ 是一个满足内部锥条件的有界利普希茨区域，且 $\tilde{U}$包含于立方体 $B(u_c, R_c) = \{u \in \mathbb{R}^{d_u} : \|u - u_c\|_\infty \le R_c\}$ 中，其中 $u_c \in \mathbb{R}^{d_u}$ 和 $0 < R_c < \infty$为某些常数；*(ii)* $g : [0,\infty) \to [0,\infty)$ 连续且单调递增，并满足 $g(0) = 0$ 以及 $\lim_{x \downarrow 0} g(x) \exp(x^{-3d_u}) = \infty$；*(iii)* 训练点 $D_N \subseteq U$ 是从测度 ν 中独立同分布抽取的，该测度的密度 ρ满足 $\rho(u) \ge \rho_{\min} > 0$，对所有 $u \in \tilde{U}$ 均成立。

那么存在常数 C_{Prop4}、$C'_{\text{Prop4}} > 0$ 和 $0 < C''_{\text{Prop4}} \le 1$，它们与 N、u_c 和 R_c无关，使得对于任意 $\varepsilon > 0$，有

$$\mathbb{E}_\nu[g(h_{D_N,\tilde{U}})] \le C_{\text{Prop4}} R_c\, g(N^{-\frac{1}{d_u}+\varepsilon}),$$

$$\mathbb{P}_\nu[h_{D_N,\tilde{U}} > h] \le C'_{\text{Prop4}} \left(\frac{h}{R_c}\right)^{-d_u} \left(1 - C''_{\text{Prop4}} \left(\frac{h}{R_c}\right)^{d_u}\right)^N, \quad \text{for any } 0 \le h \le R_c.$$

特殊情形 $\tilde{U} \subseteq [0,1]^{d_u}$ 下命题4的证明见 [30]。一般情形可通过简单的变换 $u \mapsto R_c u + u_c$得出。常数C_{Prop4} 和 C''_{Prop4} 分别趋于 ∞ 和 0（当 $\rho_{\min} \to 0$时）。命题4的结果还可进一步推广至训练点并非从 ν独立同分布（i.i.d.）抽样，而是由一个以ν 为平稳分布 [30]的一致遍历马尔可夫链生成的情形。

受推论 1 和 2 中结果的启发，我们现在研究量值$\mathbb{E}_\nu\left[\|f - m_N^f\|_{L^2_{\mu^y}(U)}\right]$和 $\mathbb{E}_\nu\left[\|k_N^{1/2}(\cdot,\cdot)\|_{L^2_{\mu^y}(U)}\right]$，其中假设设计点 D_N是从 ν中抽样得到的。我们尤其关注抽样测度 ν（密度为 ρ）与后验分布 μ^y（密度为π^y）之间的相互作用。直观而言，我们预期 $\nu \approx \mu^y$，因为在训练点密度较高的区域，高斯过程回归会更加准确。

定理 3 假设 U 是一个有界利普希茨区域，$H_k(U)$ 与索伯列夫空间 $H^\tau(U)$同构，且 $f \in H^\tau(U)$。进一步假设，对于所有 $N \in \mathbb{N}$，

(i) $U_N \subseteq \mathbb{R}^{d_u}$ 是紧的，且 $U_N \subseteq \left\{u \in \mathbb{R}^{d_u} : \pi^y(u) \le C_1^2 N^{-\frac{2\tau}{d_u}}\right\}$；*(ii)* 训练点 D_N 是从测度 ν_N 中独立同分布地抽取的，该测度的密度 ρ_N满足 $\rho_N(u) \ge \rho_{\min} > 0$，对所有 $u \in U \setminus U_N$均成立，并且 $\rho_N(u) = 0$ 在其他情况下成立；*(iii)* $U\setminus U_N$ 是满足角度为 θ的内部锥条件的利普希茨区域，且 $U\setminus U_N$ 包含于立方体 $B(u_c, R_c^N) = \{u \in \mathbb{R}^{d_u} : \|u - u_c\|_\infty \le R_c^N\}$中，其中某个 $u_c \in \mathbb{R}^{d_u}$ 和 $0 < R_c^N < C_2 \log N$。

于是存在常数 $C_{\text{Thm3}} > 0$，它与 f 和 N无关，使得对所有$0 \le \beta \le \tau$ 和 $\varepsilon > 0$，都有

$$\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U)}\right] \le C_{\text{Thm3}}\, N^{-\frac{\tau-\beta}{d_u}+\varepsilon}\, \|f\|_{H^\tau(U)}.$$

此外，对于任意划分 $U \setminus U_N \subseteq \cup_{i=1}^r B_i$，其中每个 B_i 都是满足内部锥条件且锥角为 θ'的有界利普希茨区域，都存在常数 C'_{Thm3}，使得对所有 $0 \le \beta \le \tau$，都有

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U)} \mathrm{I}_{\{h_{D_N,B_i} \le h_0(B_i), 1\le i\le n\}}\right]\\ &\le C'_{\text{Thm3}}\left(\left(\sup_{u\in U_N}\pi^y(u)\right)^{\frac12} + \sum_{i=1}^r\left(\sup_{u\in B_i}\pi^y(u)\right)^{\frac12}\mathbb{E}_{\nu_N}\left[h_{D_N,B_i}^{\tau-\beta}\right]\right).\end{aligned}$$

定理 3 中 $\beta = 0$ 的取值给出了 $\|f - m_N^f\|_{L^2_{\mu^y}(U)}$的一个界，而对于任意 $\delta > 0$，$\beta = d_u/2 + \delta$通过命题 2 和索伯列夫嵌入定理（如 [42]中所述）给出了 $\|k_N^{1/2}(\cdot,\cdot)\|_{L^2_{\mu^y}(U)}$ 的一个界。只有由于假设（*ii*）中需要 $\rho_{\min} > 0$，才要求 U 有界。假设（*iii*）中的$R_c^N \le C_2 \log N$可以去掉，因为U有界，因此 R_c^N在 N中一致有界。不过，R_c^N确实是 C_{Thm3} 中的一个因子，因此对 R_c^N的不同选择可能会产生渐近前效应，从而减缓收敛。在假设（*iii*）中，我们假设角度 θ 与 N无关，这确保了命题 3 中的常数$C_{\text{Prop3}}(\theta)$也与 N无关。

定理 3 的结果让我们得以洞察训练点 D_N 的选择与后验分布 μ^y之间的相互作用。我们将集合 U_N 定义为参数空间 U 中不包含显著后验质量的部分。随后，我们根据抽样测度 ν_N，仅在 $U\setminus U_N$中布置训练点。第一项结论表明，即使忽略集合 U_N，只在后验密度足够大的区域布置训练点，我们仍能在 N 中获得（几乎）最优的收敛速率。这里的最优性是指，在 d_u 维空间中填充距离所能达到的最快收敛速率 N^{-1/d_u}，以及用 N函数值在 $L^2(U)$中逼近 $f \in H^\tau(U)$所能达到的最快收敛速率 $N^{-\tau/d_u}$（例如参见 [43]及其中的参考文献）。对于$U\setminus U_N$上的任意抽样测度 ν_N，第一项结论均成立；但一般而言，我们预计当 $\rho_{\min}$ 趋近于零时，C_{Thm3} 会变差。

第二项结论表明，抽样测度 ν_N 的选择应使得在后验质量 $\sup_{u\in B_i}\pi^y(u)$ 较大的区域，局部填充距离 h_{D_N,B_i} 较小。换言之，我们希望在后验密度最高的区域布置训练点。我们还注意到，第二项结论中的界表明，应在后验密度 π^y 的尾部过采样，因为填充距离应与平方根 π^y 相平衡。这或许与外推通常比插值困难得多这一事实有关。

最后，我们指出，上述观察结果与该主题的既有研究一致。在 [40], 中，作者针对边缘近似（4）提出了一种序贯策略，以优化训练点 D_N 的选择。实际上，这会使训练点落在后验密度较高的区域。类似地， [6, 12],中校准-模拟-采样方法所选取的设计点，会使训练点落在 GCM 示例中求解贝叶斯逆问题时后验密度较高的区域，该示例见第 2.2 小节。本文的结果首次从近似后验分布的准确性角度，为这种训练点选择提供了理论依据。

贝叶斯求积中训练点 D_N 的最优选择是一个相关问题，即我们关注的是对误差进行界定

$$\left|\int_U f(u)\pi(u)\mathrm{d}u - \int_U m_N^f(u)\pi(u)\mathrm{d}u\right|,$$

对于某个目标测度 π，这一问题已在 [3]中通过数值方法进行研究。他们观察到，根据 π的略微扩张版本来选择设计点 D_N，也就是对 π的尾部进行过采样，似乎能得到最小误差。

4 数值示例

我们以两个简单的说明性示例结束本文。在第一个示例中，我们的函数为 $f(u) = u$，后验测度为 $\mu \sim \mathrm{N}(1, 1)$。接着，我们取设计测度为 $\nu \sim \mathrm{N}(1, \sigma^2)$。我们感兴趣的是分析以下误差量

$$e(N,\nu) = \int_{-\infty}^{\infty} \mathbb{E}_\nu |m_N^f(u) - f(u)|^2 \mu(du) \tag{16}$$

既考察设计测度的方差 σ^2 ν，也考察高斯过程回归中所使用的点数 N。在所有计算中，我们都使用了用于设计点的 10^3 个实现，以近似公式 (16) 中的期望。请注意，此示例并不满足定理 3 的所有假设，因为 U 无界，并且没有通过舍弃 U_N而对参数空间进行先验截断，但其行为仍如该定理所预测。

如图3所示，结果与定理 3一致。特别是，正如预期的那样，我们观察到，随着设计点数量 N 增加 $e(N,\nu)$

误差会降低，并且对于 σ 的不同选择，其降低速率相同。此外，也许更有意思的是，我们观察到，随着设计测度的方差增加，误差总体上会降低，直到在接近真实后验方差的值处达到最小值，随后缓慢增加。

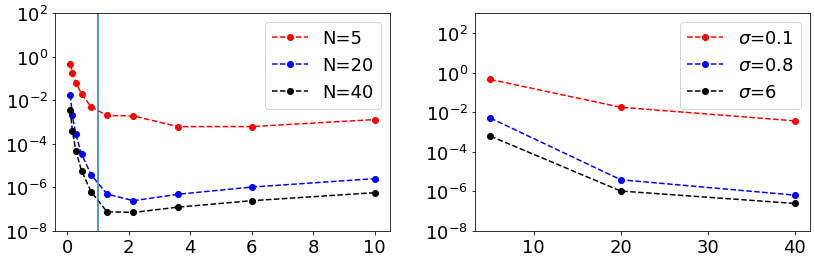

图 3： $e(N, \nu)$ 随 σ 和 N 的变化情况；此处采用高斯后验分布 μ 和设计测度 ν，并使用高斯协方差核 k。

在第二次数值实验中，我们再次进行第一次实验，仍然采用 $f(u) = u$，但这次使用真实后验测度 $\mu \sim U[-1, 1]$ 和设计测度 $\nu \sim U[-\epsilon, \epsilon]$。从图 4 可以看出，结果与之前研究的高斯情形相似；在 ϵ固定时，误差随 N 减小。

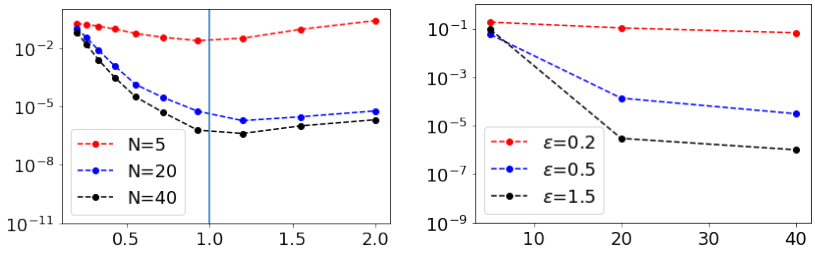

图 4： $e(N, \nu)$ 随 ϵ 和 N 的变化情况；此处采用均匀后验 μ 和设计测度 ν，并使用高斯协方差核 k。

致谢

ALT 谨此感谢剑桥的艾萨克·牛顿数学科学研究所，在开展本论文相关工作的项目期间给予的支持与热情接待。未来数据驱动工程的数学与统计基础 本研究获得英国工程与物理科学研究委员会资助，项目编号 EP/R014604/1。AMS 也感谢美国国家科学基金会的资助（项目编号 AGS-1835860）。

参考文献

1. R. Arcangéli、*M. C. López de Silanes* 和 *J. J. Torrens*，《将采样不等式推广到分数阶索博列夫半范数及导数数据》，《数值数学》，*121 (2012)*，第 *587–608* 页。 *2. T. Bai*、*A. L. Teckentrup* 和 *K. C. Zygalakis*，《与线性偏微分方程相关的贝叶斯逆问题中的高斯过程》。准备中，*2023*。 *3. F.-X. Briol*、*C. J. Oates*、*J. Cockayne*、*W. Y. Chen* 和 *M. Girolami*，《核求积中的采样问题》，载于 *ICML17*，机器学习研究论文集，*2017*，第 *586–595* 页。 *4. T. Bui-Thanh*、*K. Willcox* 和 *O. Ghattas*，《高维参数输入空间下大型系统的模型降阶》，*SIAM. J. Sci. Comput.*，*30 (2008)*，第 *3270–3288* 页。 *5. D. Calvetti*、*M. Dunlop*、*E. Somersalo* 和 *A. Stuart*，《贝叶斯反演中的模型误差迭代更新》，《反问题》，*34 (2018)*，第 *025008* 页。 *6. E. Cleary*、*A. Garbuno-Inigo*、*S. Lan*、*T. Schneider* 和 *A. M. Stuart*，《校准、仿真、采样》，《计算物理学杂志》，*424 (2021)*，第 *109716* 页。 *7. J. Cockayne*、*C. Oates*、*T. Sullivan* 和 *M. Girolami*，《偏微分方程约束贝叶斯逆问题的概率数值方法》，载于美国物理联合会会议论文集 *1853*，*2017*，第 *060001* 页。 *8. P. R. Conrad*、*M. Girolami*、*S. Särkkä*、*A. Stuart* 和 *K. Zygalakis*，《微分方程的统计分析：为数值解引入概率测度》，《统计计算》，*27 (2017)*，第 *1065–1082* 页。 *9. P. R. Conrad*、*Y. M. Marzouk*、*N. S. Pillai* 和 *A. S mith*，《通过局部近似加速计算密集型模型的渐近精确马尔可夫链蒙特卡罗》，*J. Am. Stat. Assoc.*，*111 (2016)*，第 *1591–1607* 页。 *10. S. L. Cotter*、*G. O. Roberts*、*A. M. Stuart* 和 *D. White*，《函数的马尔可夫链蒙特卡罗方法：改进旧算法以提升速度》，*Stat. Sci.*，*28 (2013)*，第 *424–446* 页。 *11. T. Cui*、*C. Fox* 和 *M. O'Sullivan*，《采用新型自适应延迟接受 *Metropolis–Hastings* 算法对大型地热储层模型进行贝叶斯校准》，*Water Resour. Res.*，*47 (2011)*。 *12. O. R. Dunbar*、*A. Garbuno-Inigo*、*T. Schneider* 和 *A. M. Stuart*，《理想化 *GCM* 中对流参数的校准与不确定性量化》，*J. Adv. in Model. Earth Sy.*，*13 (2021)*，第 *e2020MS002454* 页。 *13. M. Girolami* 和 *B. Calderhead*，《黎曼流形朗之万方法与哈密顿蒙特卡罗方法》，《皇家统计学会杂志*B*辑》，*73 (2011)*，第 *123–214* 页。 *14. W. Hastings*，《使用马尔可夫链的蒙特卡罗采样方法及其应用》，*Biometrika*，*57 (1970)*，第 *97–109* 页。 *15. D. Z. Huang*、*T. Schneider* 和 *A. M. Stuart*，《用于逆问题的迭代卡尔曼方法》，《计算物理学杂志》，*463 (2022)*，第 *111262* 页。 *16. M. Järvenpää*、*M. U. Gutmann*、*A. Vehtari* 和 *P. Marttinen*，《针对含噪似然评估的并行高斯过程代理贝叶斯推断》，*Bayesian Anal.*，*16 (2021)*，第 *147–178* 页。

17. J. P. Kaipio 和 *E. Somersalo*，《统计与计算反问题》，施普林格，*2005*。 *18. M. Kanagawa*、*P. Hennig*、*D. Sejdinovic* 和 *B. K. Sriperumbudur*，《高斯过程与核方法：联系与等价性的综述》，*arXiv* 预印本 *arXiv:1807.02582*，*(2018)*。 *19. M. C. Kennedy* 和 *A. O'Hagan*，《计算机模型的贝叶斯校准》，《皇家统计学会杂志*B*辑》，*63 (2001)*，第 *425–464* 页。 *20. J. Latz*，《贝叶斯逆问题的适定性》，*SIAM/ASA* 不确定性量化期刊，*8 (2020)*，第 *451–482* 页。 *21. E. B. Le*、*A. Myers*、*T. Bui-Thanh* 和 *Q. P. Nguyen*，《一种用于求解大规模偏微分方程约束反问题的数据可扩展随机失配方法》，《反问题》，*33 (2017)*，第 *065003* 页。 *22. H. C. Lie*、*A. M. Stuart* 和 *T. J. Sullivan*，《常微分方程概率积分器的强收敛速率》，《统计计算》，*29 (2019)*，第 *1265–1283* 页。 *23. H. C. Lie*、*T. J. Sullivan* 和 *A. L. Teckentrup*，《贝叶斯逆问题中的随机正向模型与对数似然》，*SIAM/ASA* 不确定性量化期刊，*6 (2018)*，第 *1600–1629* 页。 *24. Y. Marzouk* 和 *D. Xiu*，《逆问题贝叶斯推断的随机配点方法》，《计算物理通讯》，*6 (2009)*，第 *826–847* 页。 *25. Y. M. Marzouk*、*H. N. Najm* 和 *L. A. Rahn*，《用于高效贝叶斯求解逆问题的随机谱方法》，《计算物理学杂志》，*224 (2007)*，第 *560–586* 页。 *26. B. Matérn*，《空间变异》，施普林格，*2013*。 *27. N. Metropolis*、*A. Rosenbluth*、*M. Rosenbluth*、*A. Teller* 和 *E. Teller*，《快速计算机上的状态方程计算》，《化学物理学杂志》，*21 (1953)*，第 *1087* 页。 *28. F. Narcowich*、*J. Ward* 和 *H. Wendland*，《具有散布零点函数的索博列夫界及其在径向基函数曲面拟合中的应用》，《数学计算》，*74 (2005)*，第 *743–763* 页。 *29. A. Nemirovski*、*A. Juditsky*、*G. Lan* 和 *A. Shapiro*，《随机规划的稳健随机逼近方法》，《*SIAM*优化杂志》，*19 (2009)*，第 *1574–1609* 页。 *30. C. J. Oates*、*J. Cockayne*、*F.-X. Briol* 和 *M. Girolami*，《一类基于 *Stein* 方法的估计量的收敛速率》，《伯努利》，*25 (2019)*，第 *1141–1159* 页。 *31. P. A. O'Gorman* 和 *T. Schneider*，《理想化 *GCM* 模拟的广泛气候条件下的水文循环》，《气候杂志》，*21 (2008)*，第 *3815–3832* 页。 *32. F. Pedregosa*、*G. Varoquaux*、*A. Gramfort*、*V. Michel*、*B. Thirion*、*O. Grisel*、*M. Blon- del*、*P. Prettenhofer*、*R. Weiss*、*V. Dubourg*、*J. Vanderplas*、*A. Passos*、*D. Cournapeau*、*M. Brucher*、*M. Perrot* 和 *E. Duchesnay*，《*Scikit-learn*：*Python* 机器学习》，《机器学习研究杂志》，*12 (2011)*，第 *2825–2830* 页。 *33. C. E. Rasmussen* 和 *C. K. Williams*，《机器学习中的高斯过程》，麻省理工学院出版社，*2006*。 *34. F. Rathgeber*、*D. A. Ham*、*L. Mitchell*、*M. Lange*、*F. Luporini*、*A. T. McRae*、*G.-T. Bercea*、*G. R. Markall* 和 *P. H. Kelly*，《*Firedrake*：通过组合抽象实现有限元方法的自动化》，《*ACM*数学软件汇刊》，*43 (2016)*，第 *1–27* 页。 *35. C. Robert* 和 *G. Casella*，《蒙特卡罗统计方法》，施普林格，*1999*。 *36. Y. Rubin*，《应用随机水文地质学》，牛津大学出版社，*2003*。 *37. M. Scheuerer*、*R. Schaback* 和 *M. Schlather*，《空间数据插值——随机问题还是确定性问题？》，《欧洲应用数学杂志》，*24 (2013)*，第 *601–629* 页。 *38. T. Schneider*，《地球系统建模 *2.0*：迈向数据驱动且不确定性量化的气候模型》，《美国物理学会通报》，*65 (2020)*。 *39. B. Schoelkopf* 和 *A. Smola*，《核方法学习》，麻省理工学院出版社，*2002*。 *40. M. Sinsbeck* 和 *W. Nowak*，《用于求解贝叶斯逆问题的计算机实验序贯设计》，*SIAM/ASA* 不确定性量化期刊，*5 (2017)*，第 *640–664* 页。 *41. A. M. Stuart*，《逆问题》，第 *19* 卷，载于《数值数学学报》，剑桥大学出版社，*2010*，第 *451–559* 页。 *42. A. M. Stuart* 和 *A. L. Teckentrup*，《贝叶斯后验分布的高斯过程近似的后验一致性》，《数学计算》，*87 (2018)*，第 *721–753* 页。

43. A. L. Teckentrup，具有估计超参数的高斯过程回归的收敛性及其在贝叶斯逆问题中的应用，SIAM/ASA 不确定性量化期刊，8（2020），第 1310–1337 页。44. H. Wendland，散乱数据逼近，剑桥大学出版社，2004。45. G. Wynne、F.-X. Briol 和 M. Girolami，设定错误的似然函数和光滑性下高斯过程均值的收敛保证，《机器学习研究杂志》，22（2021）。46. D. Zhang，多孔介质流动的随机方法：应对不确定性，爱思唯尔，2001。47. Y. Zhou、A. M. Johansen 和 J. A. Aston，迈向自动模型比较：一种自适应序贯蒙特卡洛方法，计算与图形统计学期刊，25（2016），第 701–726 页。

附录：结果证明

（定理***1***的证明）首先，注意到基于均值的后验近似 $\mu^{y,N}_{\text{mean}}$ 相对于真实后验分布 μ^y 绝对连续，因为 $\exp\left(-\mathbb{E}[\Phi_N(u)]\right)$ 和 $\exp\left(-\Phi(u)\right)$ 均为正，且 μ_0 是一个概率测度：

$$\mu^y(A) = \int_A \exp\left(-\Phi(u)\right)\mu_0(\mathrm{d}u) = 0 \Rightarrow \mu^{y,N}_{\text{mean}}(A) = \int_A \exp\left(-\mathbb{E}[\Phi_N(u)]\right)\mu_0(\mathrm{d}u) = 0.$$

出于相同的原因，先验分布 μ_0 相对于真实后验分布 μ^y 绝对连续。因此，我们得到 Radon-Nikodym导数

$$\frac{d\mu^{y,N}_{\text{mean}}}{d\mu^y}(u) = \frac{d\mu^{y,N}_{\text{mean}}}{d\mu_0}(u)\frac{d\mu_0}{d\mu^y}(u) = \frac{d\mu^{y,N}_{\text{mean}}}{d\mu_0}(u)\left(\frac{d\mu^y}{d\mu_0}(u)\right)^{-1} = \frac{\frac{1}{Z^{\text{mean}}_N}\exp\left(-\mathbb{E}[\Phi_N(u)]\right)}{\frac{1}{Z}\exp\left(-\Phi(u)\right)}.$$

利用 Hellinger距离的定义以及不等式 $\nu = \mu^y$

$$\left(1-\frac{ab}{cd}\right)^2 = \frac{1}{c^2d^2}(cd-ab)^2 = \frac{1}{c^2d^2}(cd-cb+cb-ab)^2 \le \frac{2}{c^2d^2}\left((cd-cb)^2+(cb-ab)^2\right),$$

对于实数 $a,\ b,\ c,\ d \in \mathbb{R}$，我们得到

$$\begin{aligned} 2\, d^2_{\text{Hell}}(\mu^y, \mu^{y,N}_{\text{mean}}) &= \int_U \left(1 - \sqrt{\frac{\frac{1}{Z^{\text{mean}}_N}\exp\left(-\mathbb{E}[\Phi_N(u)]\right)}{\frac{1}{Z}\exp\left(-\Phi(u)\right)}}\right)^2 \mu^y(\mathrm{d}u) \\ &\le 2\int_U \left(1-\exp\left(\frac{1}{2}(\Phi(u)-\mathbb{E}[\Phi_N(u)])\right)\right)^2 \mu^y(\mathrm{d}u) + \\ &\qquad 2\,Z\left(Z^{-1/2} - (Z^{\text{mean}}_N)^{-1/2}\right)^2 \int_U \exp\left((\Phi(u)-\mathbb{E}[\Phi_N(u)])\right)\mu^y(\mathrm{d}u) \\ &=: I + II. \end{aligned}$$

利用指数函数的局部利普希茨连续性、共轭指数（p_1, p_1'）下的 Hölder 不等式，以及不等式（$a+b$）$^2 \le 2$（a^2+b^2），对于

$a, b \in \mathbb{R}$， $L^{p_1}_{\mu^y}$中的三角等式，以及关于 $x \to x^2$的 Jensen不等式，我们得到

$$\begin{aligned} I &\leq \frac{1}{2}\int_U \left(1+\exp\left(\frac{1}{2}(\Phi(u)-\mathbb{E}[\Phi_N(u)])\right)\right)^2 (\Phi(u)-\mathbb{E}[\Phi_N(u)])^2\,\mu^y(\mathrm{d}u) \\ &\leq \frac{1}{2}\left\|\left(1+\exp\left(\frac{1}{2}(\Phi-\mathbb{E}[\Phi_N])\right)\right)^2\right\|_{L^{p_1}_{\mu^y}} \left\|(\Phi-\mathbb{E}[\Phi_N])^2\right\|_{L^{p_1'}_{\mu^y}} \\ &\leq (1+C_1(p_1))\,\|\Phi-\mathbb{E}[\Phi_N]\|^2_{L^{2p_1'}_{\mu^y}} \end{aligned}$$

对于 II，利用不等式（$a^{-1/2}-b^{-1/2}$）$^2 \leq$（$a-b$）$^2 \max\{a^{-3}, b^{-3}\}$，对于 $a, b>0$，我们有

$$II \leq 2Z\max(Z^{-3},(Z_N^{\mathrm{mean}})^{-3})\left(Z-(Z_N^{\mathrm{mean}})\right)^2\int_U \exp\left((\Phi(u)-\mathbb{E}[\Phi_N(u)])\right)\mu^y(\mathrm{d}u).$$

然后，与 I 的界类似，并对 $x \to x^2$ 使用 Jensen不等式，我们得到

$$\begin{aligned} \left(Z-(Z_N^{\mathrm{mean}})\right)^2 &= \left(\int_U \exp\big(-\Phi(u)\big)-\exp\big(-\mathbb{E}[\Phi_N(u)]\big)\mu_0(\mathrm{d}u)\right)^2 \\ &= \left(\int_U \left(1-\exp\left(\frac{1}{2}(\Phi(u)-\mathbb{E}[\Phi_N(u)])\right)\right)\mu^y(\mathrm{d}u)\right)^2 \\ &\leq (1+C_1(p_1))\,\|\Phi-\mathbb{E}[\Phi_N]\|^2_{L^{2p_1'}_{\mu^y}} \end{aligned}$$

因此

$$II \leq 2Z\ \max\left\{Z^{-3}, C_2^3\right\}(1+C_1(p_1))\,C_1(1)\ \|\Phi-\mathbb{E}[\Phi_N]\|^2_{L^{2p_1'}_{\mu^y}}.$$

至此证明完成，其中

$$C_{\mathrm{Thm1}} = \sqrt{(1+C_1(p_1))\left(\frac{1}{2}+Z\max\left\{Z^{-3}, C_2^3\right\}C_1(1)\right)}.$$

证明（定理*2*）与定理1的证明类似，我们计算

$$\frac{d\mu^{y,N}_{\mathrm{marginal}}}{d\mu^y}(u) = \frac{\frac{1}{Z_N^{\mathrm{marginal}}}\mathbb{E}\Big(\exp\big(-\Phi_N(u)\big)\Big)}{\frac{1}{Z}\exp\big(-\Phi(u)\big)},$$

并利用 Hellinger距离 的定义及 $\nu=\mu^y$，我们得到

$$
\begin{aligned}
2\, d^2_{\mathrm{Hell}}(\mu^y, \mu^{y,N}_{\mathrm{marginal}}) = &= \int_U \left(1 - \sqrt{\frac{\frac{1}{Z_N^{\mathrm{marginal}}}\mathbb{E}\Big(\exp\big(-\Phi_N(u)\big)\Big)}{\frac{1}{Z}\exp\big(-\Phi(u)\big)}}\right)^2 \mu^y(\mathrm{d}u) \\
&\le 2\int_U \left(1 - \sqrt{\mathbb{E}\big[\exp\big(\Phi(u) - \Phi_N(u)\big)\big]}\right)^2 \mu^y(\mathrm{d}u) + \\
&\qquad 2\,Z\left(Z^{-1/2} - (Z_N^{\mathrm{marginal}})^{-1/2}\right)^2 \int_U \mathbb{E}\big[\exp\big(\Phi(u) - \Phi_N(u)\big)\big]\mu^y(\mathrm{d}u) \\
&=: I + II.
\end{aligned}
$$

利用不等式 $(a-b)^2 = \left(\frac{a^2-b^2}{a+b}\right)^2 \underline{\le \frac{(a^2-b^2)^2}{a^2+b^2}}$，对于 $a,\ b \in \mathbb{R}$，结合共轭指数为 q_1 和 q_1' 的赫尔德不等式，我们得到

$$
I \le 2\left\|\left(1 - \mathbb{E}\big[\exp\big(\Phi - \Phi_N\big)\big]\right)^2\right\|_{L^{q_1'}_{\mu^y}} \left\|\left(1 + \mathbb{E}\big[\exp\big(\Phi - \Phi_N\big)\big]\right)^{-1}\right\|_{L^{q_1}_{\mu^y}}.
$$

我们按照 [23,定理 3.1] 证明中的方法，对上文右侧的第二个因子进行估计：

$$
\left\|\left(1 + \mathbb{E}\big[\exp\big(\Phi - \Phi_N\big)\big]\right)^{-1}\right\|_{L^{q_1}_{\mu^y}} \le \min\left\{\|1\|_{L^{q_1}_{\mu^y}}, \left\|\mathbb{E}\big[\exp\big(\Phi - \Phi_N\big)\big]^{-1}\right\|_{L^{q_1}_{\mu^y}}\right\} = 1
$$

这对任意 $q_1 \in [1,\infty]$ 都成立，因此我们可以选择 $q_1 = \infty$ 和 $q_1' = 1$。

对于第一个因子，利用期望的线性性、指数函数的局部利普希茨连续性、共轭指数为 p_2, p_2'（相对于 $\mathbb{E}$）和 p_1, p_1'（相对于 μ^y）的赫尔德不等式、不等式 $|a+b|^{p_2} \le 2^{p_2-1}(|a|^{p_2} + |b|^{p_2})$，以及 $L^{2p_1}_{\mu^y}$ 中的三角不等式，可得

$$
\begin{aligned}
I &\le 2\left\|\left(1 - \mathbb{E}\big[\exp\big(\Phi - \Phi_N\big)\big]\right)^2\right\|_{L^1_{\mu^y}} \\
&\le 2\int_U \left(\mathbb{E}\big[\,(1 + \exp(\Phi(u) - \Phi_N(u)))\,(\Phi(u) - \Phi_N(u))\,\big]\right)^2 \mu^y(\mathrm{d}u) \\
&\le 2\left\|\mathbb{E}\big[\,(1 + \exp(\Phi - \Phi_N))^{p_2}\,\big]^{1/p_2}\right\|^2_{L^{2p_1}_{\mu^y}} \left\|\mathbb{E}\big[\,(\Phi - \Phi_N)^{p_2'}\,\big]^{1/p_2'}\right\|^2_{L^{2p_1'}_{\mu^y}} \\
&\le 2^{p_2}(1 + C_1(p_1,p_2))\left\|\mathbb{E}\big[\,(\Phi - \Phi_N)^{p_2'}\,\big]^{1/p_2'}\right\|^2_{L^{2p_1'}_{\mu^y}}.
\end{aligned}
$$

对于 II，利用不等式 $(a^{-1/2} - b^{-1/2})^2 \le (a-b)^2 \max\{a^{-3},\ b^{-3}\}$，对于 $a,\ b > 0$，可得

$$
II \le 2\,Z\max(Z^{-3}, (Z_N^{\mathrm{marginal}})^{-3})\left(Z - (Z_N^{\mathrm{marginal}})\right)^2 \int_U \mathbb{E}\big[\exp\big(\Phi(u) - \Phi_N(u)\big)\big]\mu^y(\mathrm{d}u).
$$

接着，与 I 的上界推导相同，并对 $x \to x^2$ 使用 Jensen不等式，可得

$$\begin{aligned}\left(Z - (Z_N^{\text{marginal}})\right)^2 &= \left(\int_U \exp\big(-\Phi(u)\big) - \mathbb{E}\big[\exp\big(-\Phi_N(u)\big)\big]\mu_0(\mathrm{d}u)\right)^2 \\ &= \left(\int_U \left(\mathbb{E}\big[1 - \exp\big(\Phi(u) - \Phi_N(u)\big)\big]\right)^2 \mu^y(\mathrm{d}u)\right)^2 \\ &\le 2^{p_2-1}(1 + C_1(p_1,p_2))\left\|\mathbb{E}\big[\,(\Phi - \Phi_N)^{p_2'}\,\big]^{1/p_2'}\right\|^2_{L^{2p_1'}_{\mu^y}}\end{aligned}$$

因此

$$II \le Z\ \max\left\{Z^{-3}, C_2^3\right\}\ 2^{p_2}(1 + C_1(p_1,p_2))C_1(1,1)\left\|\mathbb{E}\big[\,(\Phi - \Phi_N)^{p_2'}\,\big]^{1/p_2'}\right\|^2_{L^{2p_1'}_{\mu^y}}.$$

由此完成证明，其中

$$C_{\text{Thm2}} = \sqrt{2^{p_2-1}(1 + C_1(p_1,p_2))\left(1 + Z\max\left\{Z^{-3}, C_2^3\right\}C_1(1,1)\right)}.$$

证明（推论***1***）我们验证定理 1 和定理 2 的假设。对于假设 (*ii*)，我们可以分别参照 [42], 中引理 4.1 和引理 4.7 的证明。注意，我们不必假设 $\sup_{u\in U}\left|\Phi(u) - m_N^{\Phi}(u)\right|$ 在 N 趋于无穷时收敛到 0，且 $\sup_{u\in U}\|\mathsf{G}(u)\| \le C_{\mathcal{G}} < \infty$；利用核 k 的再生核希尔伯特空间 $H_k(U)$ 的定义、柯西-施瓦茨不等式以及最小范数插值性质 (13)，我们可以直接对 $\sup_{u\in U} m_N^{\Phi}(u)$ 给出一个不依赖于 N 的上界：

$$\sup_{u\in U}\left|m_N^{\Phi}(u)\right| = \sup_{u\in U}\left|\langle m_N^{\Phi}(\cdot), k(\cdot,u)\rangle_{H_k(U)}\right| \le \|\Phi\|_{H_k(U)}\sup_{u\in U}\sqrt{k(u,u)}.$$

我们可以像在 [42, Lemma4.1], 的证明中那样界定 Z_N^{mean}，并像在 [42, Lemma 4.7] 的证明中那样界定 Z_N^{marginal}。对于定理1中的假设 (i)，由 $p_1 = \infty$ 可立即得到结论，因为如上所述

$$\|\exp\left((\Phi - \mathbb{E}[\Phi_N])\right)\|_{L^\infty(U)} \le \exp\left(\|\Phi\|_{H_k(U)}(1 + \sup_{u\in U}\sqrt{k(u,u)})\right).$$

对于定理2中的假设 (i)，类似地，对于 $p_1 = \infty$，我们有如下量

$$\sup_{u\in U}\mathbb{E}\big[\,(\exp(\Phi(u) - \Phi_N(u)))^{p_2}\,\big]^{1/p_2} \le \mathbb{E}\big[\exp\big(p_2\sup_{u\in U}(\Phi(u) + \Phi_N(u))\big)\big]^{1/p_2}$$

对于任意 $p_2 < \infty$，该量都可以被界定。取 $p_2 = 2$ 即可得到所需结果。 □

证明（推论***2***）这与推论1的证明非常相似，只需作相同的修改，即可界定 $\sup_{u\in U}\left|m_N^{\mathcal{G}^j}(u)\right|$，使其不依赖于 N。□

（命题 ***3*** 的证明）由 [1, 定理 3.2]可知

$$\|f - m_N^f\|_{H^\beta(\tilde{U})} \le C(\theta) h_{D_N,\tilde{U}}^{\tau-\beta} \|f - m_N^f\|_{H^\tau(\tilde{U})},$$

对于任意 $\beta \le \tau$，以及所有满足 $h_{D_N,\,\tilde{U}} \le h_0(\tilde{U}) = C'_{\text{Prop3}} R_c$ 的集合 D_N。利用 $\tilde{U} \subseteq U$、$H^\tau(U)$ 中的三角不等式、由 (14) 得到的 $H_k(U)$ 与 $H^\tau(U)$ 之间的范数等价性，以及由 (13) 得到的 $\|m_N^f\|_{H_k(U)} \le \|f\|_{H_k(U)}$，可得

$$\begin{aligned}\|f - m_N^f\|_{H^\tau(\tilde{U})} &\le \|f\|_{H^\tau(U)} + \|m_N^f\|_{H^\tau(U)} \\ &\le \left(1 + C_{\text{low}}(U)^{-1} C_{\text{up}}(U)\right) \|f\|_{H^\tau(U)}.\end{aligned}$$

至此，证明完成，$C_{\text{Prop3}}(\theta) = C(\theta)\left(1 + C_{\text{low}}(U)^{-1} C_{\text{up}}(U)\right)$。 □

证明（定理 ***3***）步骤 1：我们将误差拆分为

$$\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U)}\right] \le \mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U\setminus U_N)}\right] + \mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U_N)}\right].$$

Step 2: For $\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U_N)}\right]$, we have by assumption (i) and $\beta \le \tau$ that

$$\begin{aligned}\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U_N)}\right] &\le \left(\sup_{u\in U_N} \pi^y(u)\right)^{\frac{1}{2}} \mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(U_N)}\right] \\ &\le C_1 N^{-\frac{\tau}{d_u}} \mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\tau(U_N)}\right] \\ &\le C_1 \left(1 + C_{\text{low}}(U)^{-1} C_{\text{up}}(U)\right) N^{-\frac{\tau}{d_u}} \|f\|_{H^\tau(U)},\end{aligned}$$

其中，我们按照命题 3 的证明对 $\|f - m_N^f\|_{H^\tau(U_N)}$ 进行了界定。

Step 3: For $\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta_{\mu^y}(U\setminus U_N)}\right]$, the linearity of expectation gives

$$\begin{aligned}\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(U\setminus U_N)}\right] = \mathbb{E}_{\nu_N}&\left[\|f - m_N^f\|_{H^\beta(U\setminus U_N)} \mathrm{I}_{\{h_{D_N,U\setminus U_N} \le h_0(U\setminus U_N)\}}\right] \\ &+ \mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(U\setminus U_N)} \mathrm{I}_{\{h_{D_N,U\setminus U_N} > h_0(U\setminus U_N)\}}\right].\end{aligned}$$

对于第一项，应用命题 3 和命题 4，并结合共轭指数为 $p = \infty$ 和 $q = 1$ 的赫尔德不等式，可得

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(U\setminus U_N)} \mathrm{I}_{\{h_{D_N,U\setminus U_N} \le h_0(U\setminus U_N)\}}\right] \\ &\le C_{\text{Prop3}}(\theta)\, \mathbb{E}_{\nu_N}\left[h_{D_N,U\setminus U_N}^{\tau-\beta}\right] \sup_{D_N} \|f - m_N^f\|_{H^\tau(U\setminus U_N)} \\ &\le C_{\text{Prop3}}(\theta)\, C_{\text{Prop4}}\, C_2 \left(1 + C_{\text{low}}(U)^{-1} C_{\text{up}}(U)\right) N^{-\frac{\tau-\beta}{d_u}+\varepsilon} \|f\|_{H^\tau(U)},\end{aligned}$$

对于任意 $\varepsilon > 0$，如步骤2中所示，我们已对 $\|f - m_N^f\|_{H^\tau(U\setminus U_N)}$ 进行了界定。

对于第二项，我们使用共轭指数 $p = \infty$ 和 $q = 1$ 的赫尔德不等式，得到

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(U\setminus U_N)}\mathrm{I}_{\{h_{D_N,U\setminus U_N)} > h_0(U\setminus U_N))\}}\right]\\&\leq \sup_{D_N}\|f - m_N^f\|_{H^\beta(U\setminus U_N)}\mathbb{P}_{\nu_N}\left[h_{D_N,U\setminus U_N} > h_0(U\setminus U_N)\right].\end{aligned}$$

第二个因子可由命题4界定。对于第一个因子，我们再次使用步骤2中的 $\|f - m_N^f\|_{H^\beta(U\setminus U_N))}$ 界。将两项相加，得到

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(U\setminus U_N)}\right] \leq \left(1 + C_{\text{low}}(U)^{-1}C_{\text{up}}(U)\right)\|f\|_{H^\tau(U)}\\&\left(C_{\text{Prop3}}(\theta)\, C_{\text{Prop4}}\, C_2\, N^{-\frac{\tau-\beta}{d_u}+\varepsilon} + C'_{\text{Prop4}}\,(C'_{\text{Prop3}})^{-d_u}\left(1 - C''_{\text{Prop4}}(C'_{\text{Prop3}})^{d_u}\right)^N\right).\end{aligned}$$

由于当 N 足够大时，几何衰减项有界于代数收敛项，且两项均随 N单调递减，因此可得

$$\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(U\setminus U_N)}\right] \leq CN^{-\frac{\tau-\beta}{d_u}+\varepsilon}\|f\|_{H^\tau(U)},$$

其中 $C = \left(1 + C_{\text{low}}(U)^{-1}C_{\text{up}}(U)\right)\max\{C_{\text{Prop3}}(\theta)C_{\text{Prop4}}\, C_2, C'_{\text{Prop4}}\,(C'_{\text{Prop3}})^{-d_u}\}$。这就证明了第一个结论，其中 $C_{\text{Thm3}} = C + C_1\left(1 + C_{\text{low}}(U)^{-1}C_{\text{up}}(U)\right)$。

步骤5：对于第二个断言，我们进一步将 $\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H_\mu^\beta(U\setminus U_N)}\right]$ 按子域 $\overline{U\setminus U_N} \subseteq \{B_i\}_{i=1}^r \subseteq U$ 进行拆分，从而得到

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H_{\mu^y}^\beta(U\setminus U_N)}\mathrm{I}_{\{h_{D_N,B_i} \leq h_0(B_i),1\leq i\leq n\}}\right]\\&\leq \sum_{i=1}^r\left(\sup_{u\in B_i}\pi^y(u)\right)^{\frac{1}{2}}\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(B_i)}\mathrm{I}_{\{h_{D_N,B_i} \leq h_0(B_i)\}}\right].\end{aligned}$$

然后，与步骤3中一样，我们有

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\|f - m_N^f\|_{H^\beta(B_i)}\mathrm{I}_{\{h_{\bar{D}_N,B_i} \leq h_0(B_i)\}}\right]\\&\leq C_{\text{Prop3}}(\theta')\,\mathbb{E}_{\nu_N}\left[h_{D_N,B_i}\right]^{\tau-\beta}\left(1 + C_{\text{low}}(U)^{-1}C_{\text{up}}(U)\right)\|f\|_{H^\tau(U)}.\end{aligned}$$

至此，第二个断言的证明完成，其中

$$C'_{\text{Thm3}} = \left(1 + C_{\text{low}}(U)^{-1}C_{\text{up}}(U)\right)\|f\|_{H^\tau(U)}\left(1 + rC_{\text{Prop3}}(\theta')\right).$$