arXiv:2601.09007v1 [math.ST] 13 Jan 2026

通过广义稳定性实现统计非线性逆问题的全局多项式时间估计

作者：Sven Wang

洛桑联邦理工学院（*EPFL*）数学研究所
sven.wang@epfl.ch

非线性统计逆问题给统计分析和计算都带来了重大挑战。基于似然的估计量通常会导致非凸且可能具有多峰结构的优化景观，而马尔可夫链蒙特卡罗（MCMC）方法的混合速度可能呈指数级缓慢。我们提出一类计算上易于处理的估计量——插件估计量和偏微分方程惩罚 M 估计量——用于通过算子方程 $L_f u = g$ 定义的逆问题，其中 f 是未知参数，u 是观测解。核心思路是用弱约束松弛取代精确的偏微分方程约束，从而得到条件凸问题；在许多偏微分方程实例中，这些问题是嵌套的二次优化问题，无需计算正向映射 $G(f)$，也不需要偏微分方程求解器。对于由椭圆型偏微分方程产生的典型非线性逆问题，包括达西流模型 $L_f u = \nabla \cdot (f \nabla u)$ 和一个定态薛定谔模型，我们证明这些估计量能够达到目前已知的最佳统计收敛速率，同时可在全局范围内以多项式时间计算。在达西模型中，我们得到一个显式的次二次 $o(N^2)$ 算术运行时间上界，用于根据 N 个含噪样本估计 f。我们的分析基于新的广义稳定性估计，将经典稳定性拓展到正向算子的值域之外，并结合了非参数 M 估计的工具。我们还为达西问题推导出自适应速率，为设计适用于广泛非线性逆问题的、具有多项式时间保证的统计算法提供了蓝图。我们的估计量还可为多项式时间贝叶斯计算提供有理论依据的热启动初始化。

目录

1 主要思想与概述.2 1.1 广义 M 估计量.4 1.2 达西模型的全局 $o(N^2)$-时间定理.5 1.3 广义稳定性估计.6 1.4 马尔可夫链蒙特卡罗的热启动初始化.6 1.5 讨论及与相关文献的比较.8 2 偏微分方程模型的主要结果.10 2.1 达西流：无限维估计量.11 2.2 达西流：高维估计量.13 2.3 多项式时间贝叶斯计算.16 2.4 对未知光滑度的自适应.18 2.5 薛定谔模型的结果.19 3 广义 M 估计量的抽象定理.20 4 进一步讨论与未来方向.22

*2020*年数学学科分类学科分类：主要分类 62G05、62G20；次要分类 35R30、68Q25、62C20。
关键词与短语：非线性统计逆问题、偏微分方程逆问题、非参数 M-估计、广义稳定性估计、多项式时间算法、凸松弛、达西流、薛定谔方程。

参考文献 23
A 辅助偏微分方程结果.26A.1 达西流.26A.2 薛定谔模型.29B 第 2 节的证明.29B.1 定理2.1 的证明.29B.2 定理2.4 的证明.34B.3 定理2.6 的证明.37B.4 定理2.8 的证明.38B.5 定理2.9 的证明.41C 定理3.1 的证明.42D 两个辅助引理.45D.1 经验范数的集中性.45D.2 基本尾部不等式.45

1. 主要思想与概述。 近年来，非线性逆问题在现代统计学、应用数学与纯数学以及科学计算领域引起了广泛关注[58, 43, 49, 6, 29, 28]。每当未知统计参数 $f \in \mathcal{F}$ 只能间接地 通过非线性前向映射观测到时，就会出现这类问题

$$\mathcal{G} : \mathcal{F} \to \mathcal{H},$$

其中， $\mathcal{F}$ 表示高维或无限维参数空间， $\mathcal{H}$ 是可获取观测的空间，观测会受到测量噪声的影响。典型例子见于医学影像、地球物理流体动力学以及天气与气候建模，其中 f 编码了物理系统的未知参数。在这类情形下，观测空间是希尔伯特空间 $\mathcal{H} \subseteq L^2(\mathcal{O})$，由定义在空间域或时空域 $\mathcal{O} \subseteq \mathbb{R}^d$上的函数构成。本文考虑统计非线性逆问题，其中数据由含噪测量构成

$$Y_i = \mathcal{G}(f_0)(X_i) + \varepsilon_i, \qquad i = 1, \dots, N, \tag{1}$$

其中， $f_0 \in \mathcal{F}$是某个未知的真实值数据生成参数， $X_i \in \mathcal{O}$ 是设计点， $\varepsilon_i \in \mathbb{R}$ 是相互独立的（例如高斯）噪声变量， $N \geq 1$ 表示统计样本量。

在本文研究的这类关键逆问题中， $\mathcal{G}$ 由描述观测系统的偏微分方程（PDE）的非线性参数到解映射加以描述。参数 f 在偏微分方程中起系数函数的作用，而$\mathcal{G}(f) \in L^2(\mathcal{O})$ 表示可观测解 [58, 43, 29, 28]——下面我们将介绍几个具体示例。主要的推断任务是根据数据恢复 f_0

$$Z^{(N)} = (Y_i, X_i)_{i=1}^N$$

（由1给出）。次要目标是估计 $\mathcal{G}(f_0)$，这构成了在由 $\mathcal{G}, \{\mathcal{G}(f) : f \in \mathcal{F}\} \subseteq L^2(\mathcal{O})$的值域所定义的（通常为非线性的）函数流形上的一个受偏微分方程约束的回归问题。上述框架的重要性推动了应用数学和统计学领域的方法学取得重大进展——其中尤其包括广泛使用的最大似然估计量和基于正则化的估计量 [25, 31]，以及贝叶斯反演技术 [58, 43]。

设计具有严格算法保证的统计重建方法已被证明颇具挑战。在实践中，最大似然和贝叶斯方法依赖于

对（带惩罚项的）非凸对数似然泛函进行最小化——对于高斯噪声而言，噪声等于

$$-\ell_N(f) = \frac{1}{2N}\sum_{i=1}^{N}[Y_i - \mathcal{G}(f)(X_i)]^2, \quad f \in \mathcal{F}, \tag{2}$$

或使用 MCMC 算法从高维非对数凹贝叶斯后验分布中采样$d\Pi(f|Z^{(N)}) \propto \exp(\ell_N(f))\, d\Pi(f)$ 使用 MCMC 算法 [14, 55, 38]。这些任务的难点在于负对数似然（2）的非凸性，而这又是由 $\mathcal{G}$的非线性所导致的。因此，经典凸优化理论 [10] 和对数凹采样 [16, 18] 均不再适用，由此，上述估计量的计算事先可能是指数级困难，甚至属于“NP”难题 [5]。尽管在建立统计一致性结果和收敛速率方面 [47, 39, 1, 22] 以及伯恩斯坦-冯·米塞斯型定理方面 [42, 40], 已有显著进展，但关于算法或计算保证——即如何通过数值方法求得统计估计量——的文献仍相对有限 [48, 8, 43, 23]，且依赖于较强的假设（下文将进一步讨论）。

本文的核心指导问题是开发既具有可证明的统计保证 又具备全局多项式时间可计算性的方法，具体可概括如下。此处考虑的计算成本基于基本算术运算（例如实数的加法和乘法）次数，稍后将对其进行形式化定义。

多项式时间估计方法的存在性问题。 是否存在一个估计量序列，即数据的可测函数 $\hat{f}_N = \hat{f}_N(Z^{(N)})$，使得

(i) 具有多项式统计收敛速率，

$$\forall N \geq 1: \ \mathbb{E}_{f_0}\|\hat{f}_N - f_0\|^2 \leq CN^{-\gamma} \quad \text{for some} \quad C > 0, \gamma > 0, \quad \text{and}$$

(ii) 计算成本 $c_N > 0$ 至多随 $N \to \infty$ 多项式增长。

我们的主要贡献是提出并发展了两种统计程序的理论，分别称为“偏微分方程惩罚 M 估计量”和“插件式 M 估计量”。它们扩展了经典的 M 估计，计算上切实可行，且无需调用偏微分方程求解器。这些方法基于弱形式施加偏微分方程约束。作为一种典型的非线性逆问题，我们考虑达西流 模型；该模型一直是应用数学和统计学新理论与方法的重要试验平台 [58, 14, 43, 58, 43, 1,22, 47, 13, 35]。据我们所知，目前尚无可证明具有多项式时间复杂度的统计反演方法；因此，我们的方法为达西模型提供了新的多项式时间可行性保证（见下文定理 1.1 ），并给出了明确的次二次 $o(N^2)$算术运行时间保证。另一个例子是薛定谔方程的非线性逆问题，例如该问题与光声层析成像相关 [4, 42, 48]。

本文在概念上的一项创新，是引入一种新的广义稳定性概念，将经典稳定性扩展到正向算子的值域之外。这些新的估计是我们证明的关键，并为证明以下两项提供了统一的方法：(i) 统计收敛速率；(ii) 所提出的广义 M 估计量的全局多项式时间可计算性。另一项推论是，我们的多项式时间估计量为高维贝叶斯计算提供了有原则的热启动初始化，从而能够针对所考虑的模型，以多项式时间算法近似后验期望并进行后验抽样；见第1.4节。

1.1. 广义 M 估计量。 在概述主要结果之前，我们先说明所提出的估计程序背后的直觉，以下称之为广义 M 估计量。这些估计量的动机来自两种非常简单的思路，用于改造非凸（可能难以处理）的似然景观（2）。这两种思路包括：(i) 扩充参数；以及 (ii) 对定义映射的偏微分方程约束采用某种弱形式 $\mathcal{G}$。我们将考虑由算子方程隐式定义的正向映射 $\mathcal{G}$

$$L_f[u] = g \quad + \quad \text{boundary conditions}, \tag{3}$$

其中， g 是一个已知的源项，而$(L_f\colon f \in \mathcal{F})$ 表示一族（微分）算子。假设对于每个 $f \in \mathcal{F}$， $\mathcal{G}(f)$ 都是（3）的唯一解 u。这种设定涵盖了许多典型的偏微分方程参数推断问题，例如达西流问题，其中 $L_f u = \nabla \cdot (f \nabla u)$，以及定态薛定谔模型，其中 $L_f u = \Delta u/2 - fu$；此外还包括下文将进一步讨论的若干其他非线性偏微分方程实例，如麦基恩-弗拉索夫方程和反应-扩散方程。

我们不再考虑似然曲面 $f \mapsto \ell_N(f)$，而是将回归函数 u （偏微分方程的观测解）与统计参数 f分离，并通过在目标函数中加入一项，将二者(u, f) 联系起来。对这一原则的初步（朴素）应用将得到如下最小化问题

$$\text{minimize} \;\; J(f,u) = \frac{1}{N}\sum_{i=1}^{N}[Y_i - u(X_i)]^2 + \lambda^2 \|u - \mathcal{G}(f)\|^2_{L^2(\mathcal{O})} + \mathcal{R}(u,f). \tag{4}$$

下文中， $\lambda > 0$ 表示某个正则化参数，而 $\mathcal{R}$ 表示某种（例如 Tikhonov 类型的）复杂度惩罚项。本文选择 L^2（$\mathcal{O}$）-范数，是因为它类似于信息论中的 Kullback–Leibler 散度 KL（f_1， f_2） $\approx \frac{1}{2}\|\mathcal{G}$（$f_1$） $-\mathcal{G}$（f_2）$\|^2_{L^2}$（当均匀采样 X_i 和 $\varepsilon_i \sim N$（0，1） 时等号成立）。这种解耦使得我们能够进一步将作用于回归函数层面的“偏微分方程惩罚项” $\|u - \mathcal{G}$（f）$\|^2_{L^2}$修改为 $\|\!|L_f u - g|\!\|$，其中 $\|\!|\cdot|\!\|$为某个适当的范数。

方法 I：*PDE* 惩罚 M 估计量。

$$\text{minimize} \;\; J(f,u) = \frac{1}{2}\sum_{i=1}^{n}(Y_i - u(X_i))^2 + \lambda^2 \|\!|L_f u - g|\!\|^2 + \mathcal{R}(u,f). \tag{5}$$

通过最小化偏微分方程惩罚项来规避可能代价高昂的偏微分方程求解器评估，这一思路已出现在一些关键方法中，尤其是物理信息神经网络（PINNs） [50], 、逆问题中所谓的“一次性”表述 [30], 、最优控制[62] 和偏微分方程约束回归 [26]， 以及非参数统计学中的“平缓化”技术[57], ； 更多讨论见下文。

本文研究的第二种估计程序，是对（5）的另一种简化，我们称之为插件式 M 估计。这里进一步将最小化任务分解为两个嵌套问题：

方法二：插件式 M 估计量。

$$\begin{aligned} &\hat{u}_N \in \arg\min_{u \in \mathcal{U}} J_1(u), \quad J_1(u) := \frac{1}{2}\sum_{i=1}^{n}(Y_i - u(X_i))^2 + \mathcal{R}_1(u), \\ &\hat{f}_N \in \arg\min_{f \in \mathcal{F}} J_2(f), \quad J_2(f) := \lambda^2 \|\!|L_f \hat{u}_N - g|\!\|^2 + \mathcal{R}_2(f). \end{aligned} \tag{6}$$

当然，我们完全可以不考虑上述动机，独立地看待插件式 M 估计量。事实上，文献中已有“插件原则”：先进行纯回归或“数据平滑”，再进行反演，参见关于 [7] 线性反问题的论文，以及下文进一步讨论的论文 [32]。关键在于，非线性反演本身被表述为偏微分方程约束的可计算最小化问题。鉴于此，我们上文给出的动机提供了一种统一的视角，可将方法 I 和方法二视为广义 M 估计量。我们还注意到，初步估计量 $\hat{u}_N$ 的 $\mathcal{G}(f_0)$ 不一定要通过惩罚最小二乘法获得，也可以通过其他非参数回归方法获得。

公式 (5) 和 (6) 从多个角度来看都具有较高的计算吸引力。首先，显而易见的是，(5)–(6) 中的最小化不再需要计算$\mathcal{G}(f)$（基于似然的方法需要进行此类计算），从而避免了可能代价高昂的偏微分方程求解器调用。其次，在许多关键示例中，例如 Darcy 模型和薛定谔模型，$(f,u) \mapsto L_f u$ 是双线性的，因此目标函数 (5) 是条件二次的。代入法 (6) 变成了一个嵌套的二次问题。第三个原因涉及允许进行优化的函数 f 的值域。通常，映射 $\mathcal{G}(f)$——其中包含对微分算子 L_f的求逆——仅在一个受非线性约束的子集 $\mathcal{F}$ 上有定义（例如非负函数）。相比之下，微分算子 $L_f u$ 通常对任意足够光滑的函数 f都有定义，这使得能够将 (5) 表述为一个无约束最小化问题。这将为本文所研究的非线性问题提供便利，参见第 1.2 节和第 2 节。

1.2. 适用于达西模型的全局 $o(N^2)$ 运行时间定理。 我们以达西模型为例说明本文贡献的范围。对于任意足够光滑且严格为正的电导率函数 $f\colon \mathcal{O} \to (f_{min}, \infty)$，$f_{min} > 0$，正向映射 $\mathcal{G}(f)$ 定义为边值问题的解$u = \mathcal{G}(f)$：

$$\nabla \cdot (f \nabla u) = g \quad \text{on } \mathcal{O}, \qquad u|_{\partial\mathcal{O}} = 0, \tag{7}$$

其中， g 表示一个固定且已知的源项。我们的目标是根据 u的含噪数据识别 f。为方便起见，我们始终假设 g 是无穷次可微的。于是，对于任意$f \in C^{1+\eta}(\mathcal{O})$ 且满足 $\eta > 0$的情形，都存在唯一的经典解 $u \in C^{2+\eta}(\mathcal{O}) \cap C(\bar{\mathcal{O}})$，参见 [20]中的定理 6.14。达西流问题一直是典型的逆问题，也是马尔可夫链蒙特卡罗方法 [14] 、贝叶斯反演 [58, 43, 1, 22]、非参数 M 估计 [47]、“参数化偏微分方程” [13] 和科学机器学习[35, 36, 52], 的试验平台，同时也是研究确定性噪声下逆问题的经典理论的试验平台 [31]。据我们所知，目前尚无已知的多项式时间统计一致估计量（无论是在统计设定还是确定性设定下）。

下述定理表明，逆问题可以在次二次$o(N^2)$ 运行时间内求解，同时保持此前通过最大似然和贝叶斯方法所达到的相同统计收敛速率 [47, 22]。为正式陈述该定理，令 $\mathcal{O} \subseteq \mathbb{R}^d (d \geq 1)$ 为一个光滑有界区域。对于整数 $\alpha \geq 1$，令 $H^\alpha(\mathcal{O})$ 表示通常的 L^2阶索об列夫空间，该空间由在 α上具有 $\mathcal{O}$阶弱导数的函数组成。我们假设回归测量 $Z^{(N)} = (Y_i,\ X_i)_{i=1}^N$ 由（1）产生，其中

$$Y_i = \mathcal{G}(f_0) + \varepsilon_i, \ \text{with } X_i \sim^{i.i.d.} \text{Unif}(\mathcal{O}) \ \text{ and } \ \varepsilon_i \sim^{i.i.d.} N(0,1),$$

其中 Unif$(\mathcal{O})$ 表示在 $\mathcal{O}$ 上的均匀分布。

定理 1.1. 固定整数 $\alpha > (d/2+1)$ 和常数 $f_{min}, R > 0$，并考虑 具有某个已知光滑源项 $g\colon \mathcal{O} \to (0,\infty)$的达西流模型(7)。 那么存在一些常数 $C > 0$, $\gamma \in (0,2)$ 以及一列统计估计量 $(f_N\colon N \geq 1)$，使得

ˆf_N 是某个算法的输出，该算法的确定性运行时间为 $O(N^\gamma)$，并且满足统计收敛速率

$$\mathbb{E}^N_{f_0}\left[\|\hat{f}_N - f_0\|^2_{L^2}\right] \leq CN^{-\frac{2\alpha-2}{2\alpha+2+d}},$$

对于所有 $f_0 \in H^\alpha(\mathcal{O})$ 满足 $\inf_{x\in\mathcal{O}} f_0(x) \geq f_{min}$ 和 $\|f_0\|_{H^\alpha} \leq R$ 的情形，其中 $\mathbb{E}^N_{f_0}$ 表示在 $Z^{(N)}$ 的分布下的期望。

ˆ定理1.1 由下文的定理 2.4 推出，其中给出了 f_N 的显式构造。该估计量是对式（6）中的插件式 M 估计量进行高维小波离散化，并采用了精心选择的（二次）Sobole夫惩罚项。后文将证明，对该过程进行修改还可以实现自适应收敛速率，见定理 2.8。与此前针对达西模型中高维马尔可夫链蒙特卡罗的计算保证（参见 [43]中的定理 5.3.6）相比，这里的保证不依赖任何形式的热启动初始化，因此是“全局的”。此外，这里给出的运行时间保证并不依赖高概率论证来断言似然“可能是对数凹的”，因此我们能够显著放宽构造多项式时间统计一致估计量所需的正则性和维数假设。此前的结果要求 $\alpha \geq 22$ 和 $d = \{2, 3\}$，而定理 1.1 只要满足 $\alpha > d/2 + 1$即可成立——进一步讨论见第 1.5 节。图 1展示了该方法。

1.3. 广义稳定性估计。 现有的大多数非线性逆问题统计收敛速率证明策略 [47, 39, 43, 67]，都是通过证明前向风险 $\|\mathcal{G}(f_N) - \mathcal{G}(f_0)\|_{L^2}$ 中的一个ˆ一致性结果来实现的；该风险等价于统计海林格距离 $h(f_1, f_2)$ [22]，，随后再利用稳定性估计 （例如达西流的情形，参见[53, 33, 9]），其形式如下：

$$\|f - f_0\|_{L^2} \lesssim |||\mathcal{G}(f) - \mathcal{G}(f_0)|||, \tag{8}$$

ˆ对于某个合适的范数 $|||\cdot|||$，以证明 f_N 收敛至 f_0。然而，这种证明策略似乎有一个根本局限：估计量 $\mathcal{G}(f_0)$ 必须取值于非线性值域 $\{\mathcal{G}(f) \colon f \in \mathcal{F}\}$ 之内，而这可能难以在多项式时间内实现。证明广义 M 估计量的统计收敛保证的一个关键要素，是所谓的广义稳定性估计 。我们在此引入这一概念——它描述定义方程 $L_f u = g$ 而非算子 $\mathcal{G}$的稳定性性质；即使 $f \notin \mathcal{F}$和 $u \notin \{\mathcal{G}(f) \colon f \in \mathcal{F}\}$ 分别位于 $\mathcal{G}$的区域和值域之外，也仍可对其进行表述。简化而言，这些稳定性估计意味着，对于任意一对足够光滑的函数(f, u)，

$$\|f - f_0\|_{L^2} \lesssim |||u - \mathcal{G}(f_0)|||_1 + |||L_f u - g|||_2,$$

对于合适的范数， $|||\cdot|||_1$， $|||\cdot|||_2$。事实证明，通过对现有偏微分方程证明技术进行相对简单的扩展，可以在达西流模型（命题 2.2）和薛定谔模型（命题2.10）中获得此类估计。请注意，只要偏微分方程约束 $L_f u = g$ 得到精确满足，第二项就会消失，从而恢复 8）中关于 $\mathcal{G}$的经典稳定性形式。

1.4. *MCMC* 的热启动初始化。 我们的理论对马尔可夫链蒙特卡罗算法的多项式时间初始化具有启示意义，这些算法用于从高维后验分布中采样 [14]

$$d\Pi(f|Z^{(N)}) \propto \exp(\ell_N(f))d\Pi(f)$$

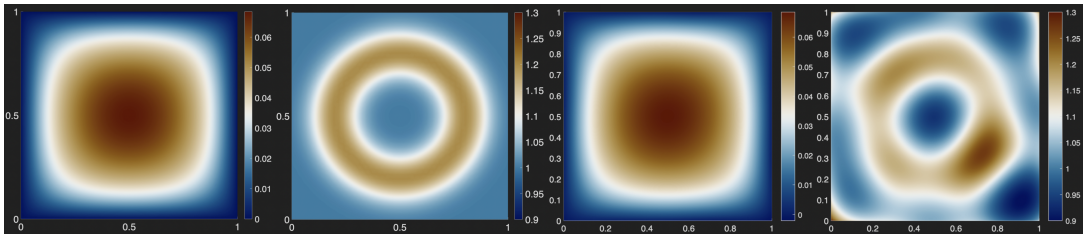

图 1. *Darcy*流逆问题中所提出的插件估计过程的数值示例。从左到右：真实解 $u_0 = \mathcal{G}(f_0)$ 和系数 f_0，通过平滑回归步骤获得的初步估计量 $\hat{u}_N$，以及基于 $N = 3000$ 个含噪测量的插件估计量 $\hat{} f_N$。

近年来，人们对“贝叶斯反演”进行了广泛研究 [58, 43]。在一系列研究工作中 [48, 2, 8] ，基于 [48],中提出的思想，人们给出了一种策略，为朗之万型马尔可夫链蒙特卡罗算法提供多项式时间混合保证。具体而言，研究证明，在达西流和薛定谔模型等模型中，可以预期对数似然（因而也包括对数后验）曲面在真实值 $\mathcal{I}$周围的某个区域 f_0 内局部对数凹（见图1.4）。因此，当后验分布以足够快的速率集中于 f_0 附近时，其大部分概率质量将包含在一个子区域中 $\mathcal{C} \subseteq I$——在许多模型中，可以通过非参数后验收缩技术验证此类条件 [19, 48]。相反，远离真实值 f_0时，对数凹性通常无法得到保证（图1.4）。

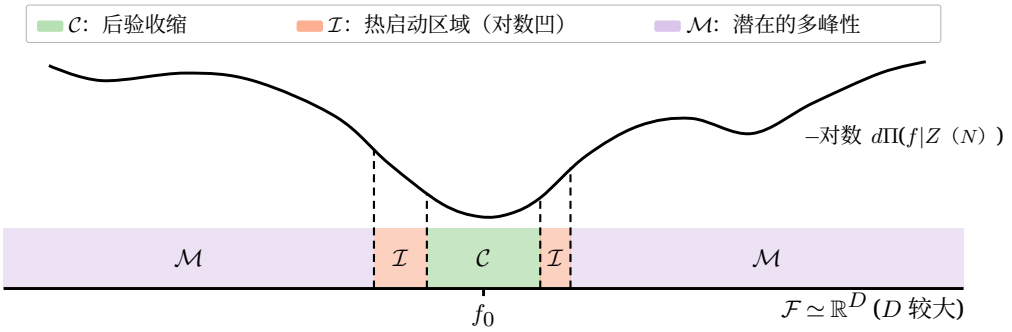

图22。对数后验景观示意图。通常（例如在达西流和薛定谔模型中），负对数后验以高概率在包含 f_0 的某个区域 $\mathcal{I}$ 内为凸函数 *[48, 43]*，而在 $\mathcal{M}$ 中可能呈现多峰性。因此，冷启动*MCMC*算法可能会因多峰性或“自由能垒” *[5]* 而遭遇计算困难。

在此类情形下，可以将后验计算问题分解为：(i) 全局初始化问题，即在 $\mathcal{I}$ 内为 MCMC链寻找一个起始点；以及 (ii) 成功进行“热启动”初始化后的局部计算问题。在达西问题中，初始化问题 (i) 迄今仍未解决，而局部计算问题 (ii) 已在 [43] 中得到处理——尽管是在较强的正则性假设下，详见第 2.3 节。事实上，[5] 表明，在冷启动初始化下，由于相关回归模型中所谓的“自由能垒”，高维非对数凹采样可能会慢到呈指数级。在达西流中，当 $f_0 \in H^\alpha(\mathcal{O})$ 足够光滑且严格为正时，[43]（见该文献的第 5.3 节）表明，后验分布以高概率在半径为 $D^{-8/d}$ 的区域内呈对数凹性（忽略对数因子）。这里采用了参数 D 维的离散化表示 $\mathbb{R}^D \ni \theta \mapsto f_\theta$

其维度随 $N \to \infty$ 增长为 $D \lesssim N^{\frac{d}{2\alpha+d}}$。因此，粗略地说，可以将达西流的初始化任务表述为：在多项式时间内找到一个足够好的估计量 $\vartheta_{init} \in \mathbb{R}^D$，使得

$$P_{f_0}^N(\vartheta_{init} \in \mathcal{I}) \xrightarrow{N\to\infty} 1, \text{ with } \mathcal{I} := \left\{\theta : \|\theta - \theta_{0,D}\|_{\mathbb{R}^D} \leq N^{-\frac{s}{2\alpha+d}} / \log N\right\},$$

其中，$\theta_{0,\ D}$ 表示 f_0的截断级数表示。详情请参见第2.3 节。注意，区域 $\mathcal{I}$ 会随着 N， D 趋于无穷而缩小。

定理 1.2（达西流的后验计算，非正式表述）。考虑 $d \leq 3$对应的达西问题，其中 $f_0 \geq f_{min} > 0$ 和 $f_0 \in H^\alpha(\mathcal{O})$ 充分大。则存在一个多项式时间初始化器 ϑ_{init}，可将其初始化于区域 $\mathcal{I}$（N， D）中。此外，对于任意多项式精度水平 $\varepsilon \geq N^{-P}$，以趋于1的概率，由 $\theta \in \mathbb{R}^D$ 上的高斯 *Matérn* 型级数先验（详见第2.3节）得到的后验均值 $\bar{\theta}_N :=$ $\int \theta d\Pi(\theta|Z^{(N)})$ 可通过多项式数量的 $O(N^\gamma)$ $\gamma > 0$ 次高维 *MCMC* 算法迭代计算至精度 ε，该算法以 ϑ_{init} 为初始值。

请注意，后验均值的计算 $\bar{\theta}_N \in \mathbb{R}^D$ 是一项非平凡的任务，因为这相当于计算关于非对数凹后验分布的一个 D维积分。在第2.3节中，我们将详细解释可用于后验计算的马尔可夫链蒙特卡罗算法，该算法是“未校正朗之万算法”（ULA）的一个修改版本，其构造遵循 [48, 43]。随后，将我们的理论与 [48, 43]中发展的局部计算结果相结合，即可证明该结果。类似地，也可以为最大后验（MAP）估计量的计算给出多项式时间保证，但本文不对此展开。

1.5. 讨论及与相关文献的比较。

1.5.1. 统计收敛保证。 已有多项研究分别探讨了基于正则化的方法和贝叶斯推断程序在非线性逆问题中的统计收敛性质。文献 [47]和 [39] 给出了推导最大后验估计量收敛速率以及高斯先验下后验收缩速率的一般方法，相关结果还可参见 [1, 22, 67] 。另一类研究考察了非参数*Bernstein-vonMises*现象，旨在建立后验分布的渐近正态性；更多参考文献可参见 [42, 40]以及 [43]。尽管这类结果在薛定谔方程模型中已得到充分理解，但对于达西模型，与“信息算子”$D\mathcal{G}_{f_0}^* D\mathcal{G}_{f_0}$ [43]的值域性质相关的根本性困难仍然存在。对于达西模型，上述工作所得到的已知最佳收敛速率为

$$\mathbb{E}[\|\hat{f}_N - f_0\|_{L^2}^2] \lesssim N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}} \qquad \text{(inverse problem)},$$

$$\mathbb{E}[\|\mathcal{G}(\hat{f}_N) - \mathcal{G}(f_0)\|_{L^2}^2] \lesssim N^{-\frac{2(\alpha+1)}{2(\alpha+1)+d}} \qquad \text{(forward problem)},$$

其中 α 表示 f_0的光滑性。对于正问题，可以证明该速率是极小极大最优的 [47]。本文提出的所有方法均达到了这些速率——见下文定理 1.1 和 2.1、2.4 。对于薛定谔模型，我们的方法在正问题和逆问题中均达到了极小极大最优收敛速率，见下文定理 2.9 。

论文 [32] 在贝叶斯框架下采用了相关的插件原则：首先推导正问题量的后验收缩（特别是 u 及其导数），然后通过基于经典数值方法的明确反演步骤恢复 f，而在实践中

这相当于求解一个输运偏微分方程。其研究主要聚焦于 $d=2$，并依赖于诸如 $\mathcal{G}(f_0)$ 属于先验协方差算子的迭代特征空间这一结构性假设。相比之下，我们的方法避免了通过输运/特征线进行显式反演：我们通过偏微分方程惩罚项来构建 f 的恢复问题，并利用广义稳定性，从而得到嵌套的二次优化问题。这为离散化估计量提供了明确的非渐近 $o(N^2)$运行时间保证（参见定理2.4）；而要为基于输运的反演获得可比的端到端复杂度保证，则需要对涉及双曲型偏微分方程的完整数值流程进行单独分析。

1.5.2. 现有的计算保证与 MCMC 混合时间。近年来，高维贝叶斯计算领域涌现了大量研究。尤其是，人们对高维对数凹采样 [16, 18]以及 MCMC 算法 [23]的无限维混合产生了浓厚兴趣。然而，在本文讨论的非线性逆问题背景下，“吉布斯测度” $\Pi(\cdot|Z^{(N)})$ 是非凸的，计算保证仍然是一项核心挑战。如第 1.4 节所述，一系列始于论文 [48]（参见 [2, 8, 43]）的研究，在热启动条件下为朗之万型马尔可夫链推导出了若干理论计算保证。在薛定谔模型中，[48]甚至证明了全局多项式时间计算保证，不过所用证明技术与我们的方法截然不同。该研究中的技术依赖于对正向算子线性化 $D\mathcal{G}$ 的所谓“梯度稳定性”论证，以断言期望似然函数的局部对数凹性，随后利用经验过程理论中的测度集中论证 [59]，以断言 ℓ_N 以高概率具有局部对数凹性。尽管这种方法首次取得了定性的多项式时间结果，但要明确给出运行时间保证中的指数似乎相当困难。此外，计算可行性似乎依赖于某个高概率事件的发生，而该事件可能只有在 N较大时才具有显著概率。第三，由于前述测度集中论证，需要极强的正则性假设（例如 $d \le 3$ 和 $\alpha \ge 22$）——这一点可以从我们的多项式时间 MCMC 定理 2.7 中看出，该定理建立在这些参考文献的基础上。本研究的主要目标之一是放宽这些严格的假设。与此一致，我们的统计估计量的计算运行时间具有明确的（次二次）指数，且为确定性运行时间，从而朝着更实用的算法保证迈进一步。不过，我们的估计量能否实现（半参数）有效的不确定性量化 [65, 43], 仍不明确；更多讨论见第 4 节。

我们还要提及经典的基于正则化的文献，这些文献研究了（带惩罚项的）最小二乘方案——通常是在确定性噪声的背景下，参见例如 [27,41, 25, 24, 31]。与本文的统计设定类似，这些文献中提出的算法保证通常受限于相关最小二乘泛函的非凸性，因此仅能提供局部的“热启动”保证；综述请参见 [31] 。

1.5.3. 物理信息神经网络、偏微分方程正则化方法及“一次性”公式。 通过惩罚偏离定义方程所规定的“物理规律”来解决问题的思路 $L_f = g$ 如今已广泛应用于应用数学和科学计算中的许多重要方法。与我们的方法 I（5）相关的是逆问题的“一次性”公式， [30] 它们同样采用了扩展优化空间的思路。近年来，最引人注目的此类方法是物理信息神经网络（PINN） [51, 17]。这些方法最初旨在求解（可能是高维的）偏微分方程，但近来也被拓展用于求解逆问题，综述请参见 [15] 。在物理信息神经网络出现之前，类似思路已见于偏微分方程最优控制 [62] 以及

偏微分方程正则化回归 [56], 中，尽管通常并非以逆问题为目标。近年来，针对上述方法，统计一致性理论逐渐发展起来，例如参见 [3], [17], [37]。然而，至少对于 PINN 类方法而言，严格的统计-算法保证似乎仍遥不可及——这是因为其中函数参数采用高度非线性的神经网络参数化——这也是一个极具挑战性的未来研究方向。

1.5.4. 非线性偏微分方程。 最后，我们简要讨论本理论是否适用于更复杂的非线性逆问题。从方法论上说，我们的方法可应用于由 $\mathcal{L}_f[u] = g$ 给出的任何逆问题，包括定义方程为非线性的情形，即 L_f 是一个非线性微分算子。如果映射 $(f, u) \mapsto L_f u$ 对 f 变量仍然是线性的，那么插件估计量的两个步骤都构成二次优化问题，因此在计算上是可处理的。这包括推断 McKean–Vlasov 方程 [45], 中的相互作用势，以及推断源于物理学、化学和生物学的反应扩散型方程中的非线性“反应函数”。在这些模型中，传统的基于似然的方法和贝叶斯方法比本文讨论的椭圆型偏微分方程示例中的方法成本更高，因此，通过广义稳定性估计为这些非线性模型中的插件方法推导统计保证，将具有重要意义。相比之下，还有一类非线性逆问题，其中待推断的参数 f 并未定义在与回归测量相同的空间（或时空）区域上，因此似乎难以直接应用我们的思路。值得注意的是，这类问题包括确定性动力系统 [46, 44] 的数据同化中对初始条件的推断问题，以及具有边界测量的非线性 X射线变换问题 [49] 。

一些记号。 我们先介绍全文通用的一些记号。对于开集 $\mathcal{O} \subseteq \mathbb{R}^d$，记 $H^\alpha(\mathcal{O})$ 为标准的 $L^2(\mathcal{O})$–Sobolev空间，即弱导数直至 α 阶均平方可积的函数空间。此外， $C(\mathcal{O})$ 和 $C(\mathcal{O})$ 分别表示定义在 $\mathcal{O}$ 和 $\mathcal{O}$ 上的连续函数空间，并赋予上确界范数。对于整数 $k \geq 1$，$C^k(\mathcal{O})$ 表示 k 次连续可微函数空间，其中所有不超过 k 阶的偏导数的上确界范数均有限，

$$\|u\|_{C^k(\mathcal{O})} := \max_{|\beta| \leq k} \sup_{x \in \mathcal{O}} |\partial^\beta u(x)|.$$

对于非整数 $\alpha > 0$， $C^\alpha(\mathcal{O})$ 表示通常的 Hölder空间。最后，记 ∇、 $\nabla\cdot$ 和 Δ 分别为梯度、散度和拉普拉斯算子。

2. 偏微分方程模型的主要结果。 本节介绍我们针对两个典型偏微分方程模型示例所得的主要结果：达西问题 和稳态薛定谔方程。在这两种情形下，正向映射 $\mathcal{G}$ 均作为椭圆型偏微分方程的参数到解算子出现，其中无限维统计参数空间 $\mathcal{F}$ 假设由 Sobolev正则函数构成。全文中，令 $\mathcal{O} \subseteq \mathbb{R}^d$ 表示一个有界光滑（开）区域，其边界 $\partial\mathcal{O}$ 光滑；不失一般性，假设其 Lebesgue测度为 1，即 $\mathrm{vol}(\mathcal{O}) = 1$。对于整数 $\alpha > 0$、 $f_{min} > 0$ 和 $R > 0$，记

$$\mathcal{F} = \mathcal{F}(\alpha, f_{min}, R) := \left\{ f \in H^\alpha(\mathcal{O}) : f \geq f_{min},\ \|f\|_{H^\alpha(\mathcal{O})} \leq R \right\}. \tag{9}$$

在达西模型中，下界 $f_{min} > 0$ 确保一致椭圆性，而光滑度等级 α 将决定所能达到的收敛速率。后续理论要求$\alpha > d/2$ ，以保证偏微分方程分析所需的充分正则性——在此情形下，对于任意 $\eta \in (0, \alpha - d/2)$ ，索伯列夫嵌入 $H^\alpha(\mathcal{O}) \hookrightarrow C^\eta(\mathcal{O})$ 确保函数属于 $\mathcal{F}$

存在逐点代表元。本节中，我们始终考虑随机设计回归模型

$$Y_i = \mathcal{G}(f)(X_i) + \varepsilon_i, \qquad X_i \sim^{i.i.d.} \mathrm{Unif}(\mathcal{O}), \quad \varepsilon_i \sim^{i.i.d.} N(0,1), \quad 1 \le i \le N, \tag{10}$$

其中 $(X_i)_{i=1}^N$ 和 $(\varepsilon_i)_{i=1}^N$ 相互独立。尽管我们的方法可推广至更一般的采样分布 $X_i \sim P_X$ 以及更广泛的噪声类别（例如次高斯 $\varepsilon_i \in \mathbb{R}$），为清晰起见，我们将讨论范围限定在当前设定中。

受第 1.1 节所阐述直觉的启发，我们为式（5）中定义的偏微分方程惩罚 M 估计量及式（6）中相应的插件估计量建立了理论保证；参见定理 2.1。核心结果之一是定理 2.4，该定理表明，插件估计量的高维小波离散化在计算时间低于二次的同时，达到了目前已知的最优统计收敛速率。第 2.3 节讨论这些结果带来的进一步计算意义，包括多项式时间采样保证；第2.4 节则推导自适应速率。上述结果聚焦于达西模型。受篇幅限制，第2.5 节较简要地讨论了稳态薛定谔方程的类似结论。

2.1. 达西流：无限维估计量。 固定 $g \in C^\infty(\mathcal{O})$。对于任意满足 $\inf_{x\in\mathcal{O}} f(x) > 0$ 且 $f \in C^{1+\eta}(\bar{\mathcal{O}})$ 的导电率 $f\colon \mathcal{O} \to (0, \infty)$，其中 $\eta > 0$，狄利克雷问题

$$\begin{cases} \nabla \cdot (f \nabla u) = g & \text{in } \mathcal{O}, \\ u = 0 & \text{on } \partial\mathcal{O}, \end{cases} \tag{11}$$

根据标准椭圆理论 [20,定理 6.14]，存在唯一的经典解 $u_f \in C^2(\mathcal{O}) \cap C(\bar{\mathcal{O}})$。我们将相应的参数到解映射记为 $\mathcal{G}(f) = u_f$。每当 $\alpha > d/2 + 1$ 且 $f \in \mathcal{F}(\alpha, f_{\min}, R)$ 时，索伯列夫嵌入 $H^\alpha(\mathcal{O}) \hookrightarrow C^{1+\eta}(\mathcal{O})$ 保证对于所有 $f \in \mathcal{F}$，$\mathcal{G}(f)$ 在$\eta \in (0, \alpha - d/2 - 1)$ 上都有定义。

下面的定理讨论达西模型的两种估计程序。第一种是文献（5）中引入的 PDE 惩罚 M 估计量，并采用适当选择的 Tikhonov 惩罚项。记作 $L_f u = \nabla \cdot (f \nabla u)$，我们将 PDE 惩罚 M 估计量定义为泛函的任意可测极小化解

$$J(u,f) = \frac{1}{N}\sum_{i=1}^N \big[Y_i - u(X_i)\big]^2 + \lambda_N^2 \|L_f u - g\|_{L^2(\mathcal{O})}^2 + \mu_N^2 \Big(\|u\|_{H^{\alpha+1}(\mathcal{O})}^2 + \|f\|_{H^\alpha(\mathcal{O})}^2 \Big), \tag{12}$$

在$(u,f) \in H^{\alpha+1}(\mathcal{O}) \times H^\alpha(\mathcal{O})$上，其中正则化参数为

$$\lambda_N = N^{-\frac{2}{2(\alpha+1)+d}}, \qquad \mu_N = N^{-\frac{\alpha+1}{2(\alpha+1)+d}}. \tag{13}$$

第二种估计量是文献（6）中的两阶段插件 M 估计量。具体而言，它定义为

$$\begin{aligned} \hat{u}_N &\in \operatorname*{arg\,min}_{u \in H^{\alpha+1}(\mathcal{O})} \Big\{ \frac{1}{N}\sum_{i=1}^N \big[Y_i - u(X_i)\big]^2 + \mu_N^2 \|u\|_{H^{\alpha+1}(\mathcal{O})}^2 \Big\}, \\ \hat{f}_N &\in \operatorname*{arg\,min}_{f \in H^\alpha(\mathcal{O})} \Big\{ \|\nabla \cdot (f \nabla \hat{u}_N) - g\|_{L^2(\mathcal{O})}^2 + \nu_N^2 \|f\|_{H^\alpha(\mathcal{O})}^2 \Big\}, \end{aligned} \tag{14}$$

其中 μ_N与（13）中的相同，而$\nu_N = N^{-\frac{\alpha-1}{2(\alpha+1)+d}}$.注意，在（14）中，两个优化问题都是严格凸的（事实上是二次的），因此都存在唯一解。

重要的是，这两种表述都允许 $u_0 = \mathcal{G}(f_0)$ 的估计量落在正向算子 $\mathcal{G}$的值域之外，并且分别在 线性空间 $H^\alpha(\mathcal{O})$ 和 $H^{\alpha+1}(\mathcal{O})$ 上构造为无约束问题；特别是，它们既不施加 $f>0$的正性约束，也不施加值域约束 $u \in \{\mathcal{G}(f) : f \in \mathcal{F}\}$。先验来看，这种方法能否达到与基于似然的方法 [47, 43, 22]相同的统计保证并不明显。正如第 1.3 节所述，这是由于达西问题具有广义稳定性性质，我们将在引理 2.2 中陈述这些性质。现在我们给出这两种估计量的主要收敛结果，并将在第 B 节证明。

定理 *2.1.* 设 $d \geq 1$，并假设 $\mathcal{O} \subseteq \mathbb{R}^d$ 是一个光滑有界区域。设 $\alpha > d/2+1$ 且 $\alpha \geq 3$。固定常数 $f_{min} > 0$、 $R > 0$。假设数据 *(10)* 来自某个 $\hat{f}_0 \in \mathcal{F}(\alpha, f_{min}, R)$，并令 $(\hat{u}_N, \hat{f}_N)$ 表示 *(12)* 中的广义 *M* 估计量或 *(14)* 中的插件式 *M* 估计量。存在仅依赖于 $\mathcal{O}, d, \alpha, f_{min}, R$ 的 $C > 0$，使得对所有 $N \geq 1$,

$$\mathbb{E}^N_{f_0}\left[\|\hat{u}_N - \mathcal{G}(f_0)\|^2_{L^2(\mathcal{O})}\right] \leq C N^{-\frac{2(\alpha+1)}{2(\alpha+1)+d}}, \quad \textit{and} \tag{15}$$

$$\mathbb{E}^N_{f_0}\left[\|\hat{f}_N - f_0\|^2_{L^2(\mathcal{O})}\right] \leq C N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}}. \tag{16}$$

定理 2.1 建立了两种收敛速率。第一种涉及估计 $\mathcal{G}(f_0)$ 的正问题，其速率与估计 $(\alpha+1)$-光滑回归函数的极小极大速率一致。文献 [47] 证明了，在极小极大意义下，该速率在模型类 $\{u_f\colon f \in \mathcal{F}(\alpha, f_{\min}, R)\}$ 上是最优的。第二种速率针对恢复电导率 f_0 的逆问题。从启发式角度看，该速率对应于观测信号为 $\alpha+1$-光滑时，一个 2-病态逆问题的极小极大估计速率。事实上，已知逆映射 $\mathcal{G}^{-1}: u \mapsto f$ 满足一个 2-病态稳定性估计：对于足够光滑且一致椭圆的系数 f_1, f_2，有

$$\|f_1 - f_2\|_{L^2(\mathcal{O})} \lesssim \|u_{f_1} - u_{f_2}\|_{H^2(\mathcal{O})},$$

例如，参见 [47]。类似的稳定性估计也见于早期研究，例如 [53]针对 L^∞–C^2 界，以及 [33]针对一维 L^2 估计。定理 2.1 背后的核心分析洞见是，即便偏微分方程约束仅近似满足，恰当的稳定性估计仍然成立。以下引理将这一广义稳定性现象精确地表述出来。

引理 2.2。 设 $\mathcal{O} \subseteq \mathbb{R}^d$ 是一个光滑有界区域，并令 f_{min}， $g_{min} > 0$。假设对于 $i \in \{1,2\}$，我们有函数

$$u_i \in H^2(\mathcal{O}), \quad f_i \in C^1(\mathcal{O}), \quad g_i \in L^2(\mathcal{O}) \quad \textit{with} \quad \nabla \cdot (f_i \nabla u_i) = g_i \quad \textit{on } \mathcal{O}.$$

假设第一个三元组 $(u_1,\ f_1,\ g_1)$ 还满足 $u_1 \in C^2(\mathcal{O}) \cap C(\bar{\mathcal{O}})$， $g_1 \in C(\mathcal{O})$以及

$$\inf_{x \in \mathcal{O}} f_1(x) \geq f_{min} \quad \textit{and} \quad \inf_{x \in \mathcal{O}} g_1(x) \geq g_{min}, \quad \textit{and} \quad u_1|_{\partial\mathcal{O}} = 0.$$

那么，对于每个 $B > 0$，都存在一个常数 $C > 0$，它仅依赖于 f_{min}， g_{min}， $\mathcal{O}$ 和 B，使得只要 $\|f_1\|_{C^1} \vee \|u_1\|_{C^2} \leq B$，就有

$$\|f_1 - f_2\|_{L^2} \leq C\big(\|f_2\|_{C^1}\|u_1 - u_2\|_{H^2} + \|g_1 - g_2\|_{L^2}\big).$$

引理 2.2 的一个关键特点是其假设条件具有 不对称性。主要结构性假设（统一椭圆性、狄利克雷条件以及关于 g_1的下界）仅施加于

参考三元组 (u_1, f_1, g_1)，而 (u_2, f_2, g_2) 除了基本正则性外不受其他限制。由于正向映射 $\mathcal{G}$ 具有有界性，常数 $B \geq \|f_1\|_{C^1} \vee \|u_{f_1}\|_{C^2}$ 可以在三元组 $g_1 = g, u_1 = \mathcal{G}(f_0)$ 以及 $f_1 = f_0 \in \mathcal{F}$ 上统一选取，其中 $\alpha > d/2 + 1$ 表示数据生成真实值。此外，我们注意到，定理 1.1的证明实际上给出了针对适当的惩罚经验风险泛函的更强（次高斯）集中不等式——详情请参见证明。

2.2. 达西流：高维估计量。 本节介绍支撑定理 1.1 的高维估计量。该估计量源自对无限维插件方案（14）进行的高维小波框架离散化。与狄利克雷拉普拉斯特征函数等其他表示系统相比，本方法选用小波，以便准确逼近 $H^{\alpha+1}(\mathcal{O})$ 直到边界处 $\partial\mathcal{O}$ 的函数。本节的主要结果是定理 2.4，该定理表明，该估计量可在次二次运行时间内计算，同时具有与定理2.1中的无限维对应估计量相同的统计收敛性质。

在陈述主要结果之前，我们先澄清本节所采用的计算模型。

备注 2.3（计算模型）。定理 2.4中的显式复杂度界统计与矩阵—向量乘积和线性系统求解相关的浮点运算次数。我们假设可以通过预言机访问点求值以及 $L^2(\mathcal{O})$ 小波基函数及其导数的内积，这些量可以预先以高精度计算。此外，我们以上界估计将计算成本控制在稠密线性代数的计算成本之内；利用 Φ 或 Ψ 的额外稀疏结构，可能会进一步缩短运行时间，但本文不对此展开讨论。该计算模型细化了以往所采用的计算时间概念，例如 [48] 中以及下文第 2.3节所使用的马尔可夫链蒙特卡罗迭代复杂度；这些概念假定能够精确地通过预言机访问正向映射的求值 $\mathcal{G}(f)$ 。

小波框架与预备知识。 我们介绍一些记号和基本事实，这些内容将在全文中用于小波框架。关于该理论的全面介绍，请参阅 [61]。与前文一样，令 $\mathcal{O} \subseteq \mathbb{R}^d$ 表示一个光滑有界区域。对于某个 $u > \alpha + 1$，我们记

$$\{\phi_{lk} : l \geq 0, 1 \leq k \leq N_l\},$$

一个由光滑性至少为 $C^u(\mathcal{O})$ 的小波构成的 $L^2(\mathcal{O})$ 小波框架，其中光滑性的含义见 [61] 中的定义 5.25； [61] 中的定理 5.51 保证了此类小波系统的存在。在“分辨率层级” $l \geq 0$ 上，至多有 $N_l \lesssim 2^{ld}$ 个元素。给定任意多尺度向量

$$v = (v_{lk} : l \geq 0,\ 1 \leq k \leq N_l),$$

于是我们可以定义以下序列范数

$$\|v\|_{h^\alpha}^2 := \sum_{l \geq 0} \sum_{k \leq N_l} 2^{2l\alpha} v_{lk}^2, \quad \alpha \geq 0,$$

其中 h^α 表示所有范数有限的序列构成的空间。上述小波框架的一个关键性质是，$\|\cdot\|_{h^\alpha}$ 刻画了经典的索伯列夫范数。事实上，对于任意 $f \in H^\alpha(\mathcal{O})$ ，都有

$$\inf_v \|v\|_{h^\alpha} \simeq \|f\|_{H^\alpha(\mathcal{O})}, \tag{17}$$

下确界取遍所有可容许表示 $f=\sum_{l,\ k} v_{lk}\phi_{lk}$，参见 [61] 中的定理 5.51（其中 $p=q=2$）。左侧的下确界是必要的，因为（ϕ_{lk}）可能不构成基，因此 f 的表示可能并不唯一。不过，存在一个有界线性算子 $S:\ H^\alpha(\mathcal{O})\to h^\alpha$，它可选取一种表示，使得

$$\|S(f)\|_{h^\alpha}\simeq\|f\|_{H^\alpha(\mathcal{O})},\qquad f=\sum_{l,k}S(f)_{lk}\phi_{lk},$$

见 [61] 中的定理 5.51。我们将第 $(l,\ k)$ 个系数的选择算子记为

$$S_{lk}:H^\alpha(\mathcal{O})\to\mathbb{R},\ \ S_{lk}(f)=S(f)_{lk}.$$

在整个过程中，我们固定一个这样的选择算子。我们还需要上述小波空间的有限维截断。对于分辨率层级 $J\geq 1$，我们将频率截至层级 J 的子空间记为

$$L^2_J(\mathcal{O})=\operatorname{span}(\phi_{lk},\ l\leq J,\ k\leq N_l)\subseteq L^2(\mathcal{O}),$$

其维数为

$$\dim(L^2_J(\mathcal{O}))\leq p_J:=\sum_{l\leq J}N_l\lesssim 2^{Jd},\quad J\geq 1.$$

稍微滥用记号，我们将用 $v\in\mathbb{R}^{p_J}$ 表示有限分辨率多尺度向量所在的欧几里得空间，并且

$$\|v\|^2_{h^\alpha_J}:=\sum_{l\leq J,k\leq N_l}2^{2l\alpha}v^2_{lk},\quad \alpha\geq 0,\ J\geq 1,\ v\in\mathbb{R}^{p_J}.$$

2.2.1. 构造 $\hat{}u_N$。 令 $\mu_N=N^{-\frac{\alpha+1}{2\alpha+1+d}}$ 表示 $(\alpha+1)$ 光滑函数的极小极大估计速率。当样本量为 $N\geq 1$ 时，我们选择分辨率层级 $J=J_N$，使得

$$2^{Jd}\simeq N\mu_N^2=N^{\frac{d}{2(\alpha+1)+d}}.$$

这一选择平衡了小波投影的逼近误差与估计 $\alpha+1$-正则函数的 $p_J\asymp 2^{Jd}$ 个系数时产生的随机误差。估计量 $\hat{}u_N$ 将表示为

$$\hat{u}_N:=\sum_{l\leq J,k\leq N_l}\hat{\eta}_{lk}\phi_{lk}\in H^{\alpha+1}(\mathcal{O}),\tag{18}$$

其中 $\hat{}\eta_N$ 是二次泛函的（唯一）极小化器

$$\frac{1}{N}\sum_{i=1}^N\Big(Y_i-\sum_{l\leq J,k\leq N_l}\eta_{lk}\,\phi_{lk}(X_i)\Big)^2+\mu_N^2\|\eta\|^2_{h_J^{\alpha+1}}\quad\text{over}\quad\eta\in\mathbb{R}^{p_J}.$$

用矩阵形式表示，我们可以将其写成一个高维惩罚最小二乘问题

$$\hat{\eta}_N=\operatorname*{arg\,min}_{\eta\in\mathbb{R}^{p_J}}\Big(\frac{1}{N}\big\|Y-\Phi\eta\big\|^2_{\mathbb{R}^N}+\mu_N^2\eta^T\Lambda^{\alpha+1}\eta\Big),\tag{19}$$

其中，$\|\cdot\|_{\mathbb{R}^N}$ 表示欧几里得范数，$\Phi\in\mathbb{R}^{N\times p_J}$ 是元素为$\Phi_{i,\ (l,k)}=\phi_{lk}(X_i)$ 的“设计矩阵”，而 Λ 是表示 $h^{\alpha+1}$-范数的 $p_J\times p_J$ 对角矩阵，

$$\Lambda:=\operatorname{diag}\big(1\ldots\underbrace{2^{2l}\ldots 2^{2l}}_{N_l\text{ times}}\ldots 2^{2J}\ldots 2^{2J}\big).\tag{20}$$

从该公式可以立即看出，(19)确实存在唯一的极小化器$\hat{}\eta_N$——因为 Λ 是严格正定的——其表达式为

$$\hat{\eta}_N = \left(\Phi^T\Phi + N\mu_N^2\Lambda^{\alpha+1}\right)^{-1}\Phi^T Y. \tag{21}$$

由于 $p_J \asymp 2^{Jd} \asymp N\mu_N^2 = N^{\frac{d}{2\alpha+1}+d}$，构造 $\Phi^\top\Phi$ 并求解由对$(\Phi^T\Phi + \mu_N^2\Lambda^{\alpha+1})$ 求逆所涉及的相应线性方程组（假设采用稠密线性代数）的成本上界为

$$O(Np_J^2 + p_J^3) = O\left(N^{1+\frac{2d}{2(\alpha+1)+d}} + N^{\frac{3d}{2(\alpha+1)+d}}\right) = O\left(N^{1+\frac{2d}{2(\alpha+1)+d}}\right),$$

其余矩阵-向量运算的计算成本阶数更低 r.

$\hat{}$2.2.2. 构造 f_N。 在给定$\hat{}u_N$的条件下，计算$\hat{}f_N$ 的插件估计步骤仍可表示为基展开

$$\hat{f}_N := \sum_{l\le J, k\le N_l} \hat{\theta}_{lk}\phi_{lk}, \tag{22}$$

$\hat{}$其中 θ_N 是以下高维惩罚最小二乘问题的解。按照之前的方式选取 $J = J_N$ 后，我们将源项 g的框架系数记为 $\gamma \in \mathbb{R}^{p_J}$，

$$\gamma_{lk} = S_{lk}(g), \quad \forall\, l \le J,\ k \le N_l.$$

$\hat{}$估计量 θ_N 定义为

$$\hat{\theta}_N = \arg\min_{\theta\in\mathbb{R}^{p_J}} \|\gamma - \Psi\theta\|^2_{\mathbb{R}^{p_J}} + \nu_N^2\theta^T\Lambda^\alpha\theta,$$

其中 $\nu_N = N^{-\frac{}{(\alpha-1)\,2(\alpha+1)+d}}$ 按照无限维情形中的方式选取， Λ 由式（20）给出，矩阵 $\Psi \in \mathbb{R}^{p_J\times p_J}$ 由以下系数给出

$$\begin{aligned}\Psi_{(lk),(l'k')} &= S_{l'k'}\left(\nabla\cdot(\phi_{lk}\nabla\hat{u}_N)\right)\\ &= S_{l'k'}(\nabla\phi_{lk}\cdot\nabla\hat{u}_N) + S_{l'k'}(\phi_{lk}\Delta\hat{u}_N)\\ &= \sum_{m\le J, j\le N_m} (\hat{\eta}_N)_{mj}\left[S_{l'k'}(\nabla\phi_{lk}\cdot\nabla\phi_{mj}) + S_{l'k'}(\phi_{lk}\Delta\phi_{mj})\right].\end{aligned}$$

这里，$\hat{}\eta_N$ 是由式（21）给出的估计量。请注意，构造 Ψ 需要获取基函数及其导数乘积的小波框架系数，例如$S_{l'k'}(\nabla\phi_{lk}\cdot\nabla\phi_{mj})$ 和$S_{l'k'}(\phi_{lk}\Delta\phi_{mj})$；在我们的计算模型中，我们将这些视为可通过预言机获取的量（或可预先离线高精度计算的量）。给定 Ψ，θ_N 具有闭式表达式

$$\hat{\theta}_N = \left(\Psi^T\Psi + \nu_N^2\Lambda^\alpha\right)^{-1}\Psi^T\gamma, \quad \hat{\theta}_N \in \mathbb{R}^{p_J}. \tag{23}$$

定理 *2.4.* 考虑有界光滑区域 $\mathcal{O}\subseteq\mathbb{R}^d$上的达西流问题，其中 $g \in C^\infty(\mathcal{O})$光滑且为正。假设 $\alpha > d/2+1$ 且 $f_0 \in \mathcal{F}(\alpha, f_{min}, R)$，其中 $f_{min}, R > 0$。存在常数 $C_1, C_2 > 0$，它们至多依赖于 $\alpha, d, \mathcal{O}, f_{min}, R\hat{}\,\hat{}$，以及指数 $\kappa = \kappa(d,\alpha) < 2$，使得对于所有 $N \ge 1$，由式（*18*）–（*23*）定义的估计量$\hat{}\eta_N$、$\hat{}u_N$、θ_N、f_N满足以下收敛速率

$$\mathbb{E}^N_{f_0}[\|\hat{u}_N - u_0\|^2_{L^2}] \le C_1 N^{-\frac{2(\alpha+1)}{2(\alpha+1)+d}}, \quad \mathbb{E}^N_{f_0}[\|\hat{f}_N - f_0\|^2_{L^2}] \le C_1 N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}}, \tag{24}$$

$\hat{}$且高维向量 $\hat{}\eta_N, \theta_N \in \mathbb{R}^{p_J}$可在 C_2N^κ 次浮点运算内计算得到。

定理 2.4特别蕴含定理1.1。事实上，明确的运行时间指数为

$$\kappa = \max\left\{1 + \frac{2d}{2(\alpha+1)+d}, \frac{3d}{2(\alpha+1)+d}\right\} = 1 + \frac{2d}{2(\alpha+1)+d}$$

由于 $\alpha > d/2+1$，我们有 $\kappa < 2$，从而得到一致的次二次计算复杂度。具体而言，尽管统计收敛速率会随着物理维数 d 的增加而下降，但达到这些速率（在正问题中为极小极大最优速率）并不会带来超二次计算成本。

2.3. 多项式时间贝叶斯计算。 我们现在给出两个关于高维后验抽样多项式时间计算保证的结果。下文使用的多项式时间计算概念不同于第 2.2 节采用的概念。贝叶斯文献通常以马尔可夫链蒙特卡罗算法达到预定精度所需的迭代次数来衡量计算复杂度，并假设每次迭代都可以通过预言机访问来评估正向映射 $\mathcal{G}$（以及可能的导数）。因此，本节的多项式时间保证关注的是迭代复杂度，而非总算术成本。

我们简要回顾达西模型及相关非线性逆问题的一种标准贝叶斯非参数表述，例如 [48, 43] 中所考虑的表述。对于 $k \geq 1$，令 $(e_k, \lambda_k) \in L^2(\mathcal{O}) \times (0, \infty)$ 表示 $\mathcal{O}$ 上负狄利克雷拉普拉斯算子的 $L^2(\mathcal{O})$-正交特征基。按照 [47, 43, 1], ，我们通过该基上的截断级数展开来参数化非负扩散系数。具体而言，对于 $\theta \in \mathbb{R}^D$，定义

$$f_\theta := \Phi \circ F_\theta, \quad \text{with} \quad F_\theta(x) := \sum_{k=1}^{D} \theta_k e_k(x).$$

这里，$\Phi: \mathbb{R} \to (f_{min}, \infty)$ 是一个光滑且严格递增的“链接函数”，满足 $\Phi' > 0$，且其各阶导数均有界，上确界$_{x\in\mathbb{R}} |\Phi^{(k)}| < \infty$，并且下确界$_{x\geq z} \Phi'(z) > 0$对所有 $z \in \mathbb{R}$均成立。我们在 θ 上采用如下形式的高斯对角先验：

$$\Pi = N(0, N^{-\frac{d}{2\alpha+d}}\Sigma), \qquad \Sigma = \text{diag}(\lambda_j^{-\alpha} : 1 \leq j \leq D) \in \mathbb{R}^{D\times D}.$$

将数据记为 $Z^{(N)} = ((Y_i, X_i): 1 \leq i \leq N)$，则后验密度除一个乘法常数外为

$$\pi(\theta|Z^{(N)}) \propto \exp\Big(-\frac{1}{2}\sum_{i=1}^{N}[Y_i - \mathcal{G}(f_\theta)(X_i)]^2) - \frac{1}{2}\theta^T\Sigma^{-1}\theta\Big), \quad \theta \in \mathbb{R}^D. \tag{25}$$

分别令$\theta_0 \in \ell^2(\mathbb{N})$ $\theta_0^{(D)} \in \mathbb{R}^D$ 表示 F_0 的级数系数及其截断对应项；

$$\theta_{0,k} = \langle F_0, e_k\rangle_{L^2}, \qquad \theta_0^{(D)} = (\theta_{0,k} : 1 \leq k \leq D).$$

˜令 $H^\alpha(\mathcal{O})$ 表示由狄利克雷拉普拉斯算子谱定义的索об列夫空间

$$\tilde{H}^\alpha(\mathcal{O}) := \Big\{f = \sum_{k=1}^{\infty} a_k e_k \ \text{ for } \ \sum_{k=1}^{\infty} a_k^2 \lambda_k^\alpha < \infty\Big\}.$$

以下假设概括了第 5.3 节所需的正则性假设[43], 其中局部化的达西后验分布的采样问题（25）采用类似于 [48]的“梯度稳定性”方法进行研究。

~假设2.5。令 $f_0 = \Phi \circ F_0$ 对于某个函数 $F_0 \in H^\alpha(\mathcal{O})$，对于某个 $\alpha \geq 22$，其中 $d \geq 3$，且 $\mathcal{O} \subseteq \mathbb{R}^d$ 是一个有界光滑区域。对于 $N \geq 1$，令维数为 $D = D_N := aN^{\frac{d}{2\alpha+d}}$，其中 $a > 0$。

对光滑性的高要求 $\alpha \geq 22$ 继承自 [43]，并非我们方法本身所固有。在该文献的定理 5.3.5 中，证明了在假设2.5成立时，后验密度 $\pi(\theta|Z^{(N)})$ 在1 （关于 $Z^{(N)}$的分布）趋近于 1 的概率下，满足以下两个性质。首先，它在大小为 $\{\theta : \|\theta - \theta_{0\text{的区域上是对数凹的}, D}\|_{\mathbb{R}^D} \leq r\}$其中 $r = cD^{-8/d}$，且 $c > 0$ 是一个足够小的常数。其次，$\Pi(\cdot|Z^{(N)})$ 的绝大部分质量集中在该区域内。因此，成功进行热启动初始化的一个充分条件是（当 $N \to \infty$时渐近成立）：在多项式时间内找到一点 $\vartheta_{init} \in \mathbb{R}^D$，使得当 $N \to \infty$时，1趋近于 $\|\vartheta_{init} - \theta_{0, D}\|_{\mathbb{R}^D} \leq D^{-8/d}/\log N$.

定理 2.6. 考虑满足假设 2.5 的达西问题。那么，存在一个可在多项式时间内计算的估计量 $\vartheta_{init} = \vartheta_{init}(Z^{(N)}) \in \mathbb{R}^D$，使得 $\mathbb{P}_{f_0}^N\left(\|\vartheta_{init} - \theta_{0,D}\|_{\mathbb{R}^D} \leq D^{-8/d}/\log N\right) \to 1$ 当 $N \to \infty$ 时成立。

定理2.6 的证明见第B.3节。事实证明，$\alpha \geq 12$ 实际上足以达到 $D^{-8/d}/\log N$ 的收敛速率。由定理2.6 和 [43,定理 5.3.6],可知，高维后验均值可通过马尔可夫链蒙特卡罗在多项式次迭代内计算；这是下一项结果的主题。对于“高斯创新” $\xi_k \sim^{i.i.d.} N(0, I_{D\times D})$，我们在 $\mathbb{R}^D$ 上定义如下具有步长 $\delta > 0$ 的时间离散朗之万型马尔可夫链蒙特卡罗算法，

$$\begin{cases} \vartheta_0 & = \vartheta_{init}, \\ \vartheta_{k+1} & = \vartheta_k + \delta[\nabla \tilde{\ell}_N(\vartheta_k) + \Sigma^{-1}\vartheta_k] + \sqrt{2\gamma}\xi_{k+1}, \quad k \geq 0, \end{cases} \tag{26}$$

~其中，ℓ_N 是根据 [43,式（5.8）] 构造的全局对数凹“替代似然”函数，并且在 ϑ_{init}的某个邻域内与真实对数似然一致；详见[48] 中的详细构造。记后验均值为

$$\bar{\theta}_N = \bar{\theta}_N(Z^{(N)}) := \mathbb{E}[\theta|Z^{(N)}] = \int_{\mathbb{R}^D} \theta d\pi(\theta|Z^{(N)}) \in \mathbb{R}^D,$$

我们有如下定理。

定理 2.7. 假设假设2.5 成立。固定 $P > 0$。对于任意 $N \geq 1$，令 $D = D_N$如上选择，并令 $\rho = \rho_N \geq N^{-P}$ 表示所需的精度水平。则存在适当选择的步长 $\delta = \delta(N, D, \varepsilon)$ 和指数 $\gamma > 0$，使得经过多项式数量的 $J = O(N^\gamma)$ 次迭代后，遍历平均

$$\bar{\theta}_J^{MCMC} = \frac{1}{J}\sum_{k=J_{in}}^{J+J_{in}} \vartheta_k \in \mathbb{R}^D$$

*MCMC*链 $(\vartheta_k : k \geq 0)$ 从 (26) 满足

$$\mathbb{P}\left(\|\bar{\theta}_J^{MCMC} - \mathbb{E}[\theta|Z^{(N)}]\|_{\mathbb{R}^D} \leq \rho_N\right) \xrightarrow{N\to\infty} 1.$$

这里，$\mathbb{P}$ 表示数据分布 P_0^N与 *MCMC* 链的随机性的乘积概率$(\vartheta_k : k \geq 0)$。

2.4. 对未知光滑度的自适应。 我们在本节最后介绍插件式 M 估计量的一个自适应版本，该版本能够适应真实值的未知光滑度。事实证明，可实现的收敛速率取决于 $u_0 = \mathcal{G}(f_0)$ 的光滑度，而不是 f_0 本身的光滑度。下一个定理在无限维情形下表述；相应的高维版本可仿照附录B中的构造类似地推导。为简单起见，并与全文采用的偏微分方程理论保持一致，我们只考虑取整数值的光滑度指数。我们将自适应插件式估计程序定义如下。对于整数 ,$b \geq 0$，记 $[a: \quad b] = \{a, a+1, \ldots, b\}$。

1. 构造初步估计量

$$(\hat{\beta}, \hat{u}_N) \in \underset{\beta \in [\beta_{\min}:\beta_{\max}],\ u \in H^{\beta}(\mathcal{O})}{\arg\min} \frac{1}{N}\|Y - u(X)\|^2_{\mathbb{R}^N} + \mu^2_{N,\beta}\big(\|u\|^2_{H^{\beta}(\mathcal{O})} + A^2\big), \tag{27}$$

对于合适的常数 $A > 0$，并且当 $\mu_{N,\ \beta}: = N^{-\frac{\beta}{2\beta+d}}$。

2. 对于正则化参数 $\nu_{N,\ \hat{\beta}} := N^{-\frac{\hat{\ }}{\ \beta - 2_2\hat{\ }\beta + d}}$，计算插件式估计

$$\hat{f}_N \in \underset{f \in H^{\alpha_{\min}}(\mathcal{O})}{\arg\min} \|L_f \hat{u}_N - g\|^2_{L^2(\mathcal{O})} + \nu^2_{N,\hat{\beta}}\|f\|^2_{H^{\alpha_{\min}}(\mathcal{O})}. \tag{28}$$

我们强调，自适应性完全通过步骤 1 中的初步估计量 $\hat{\ }u_N$ 实现，而式（28）中的逆步骤则在固定的索об列夫空间 $H^{\alpha_{\min}}(\mathcal{O})$上进行；自适应性仅通过正则化参数 $\nu_{N,\ \hat{\ }\beta}$的数据驱动的选择体现出来。

$\hat{\ }$定理 2.8. 设 $(\hat{\ }u_N,\ f_N)$ 为（(27)–(28)）中定义的估计量。设 $\alpha_{\min} > d/2 + 1$，并假设 $f_0 \in H^{\alpha_{\min}}(\mathcal{O})$。设 $2 \leq \beta_{\min} < \beta_{\max} < \infty$ 为整数，并定义

$$\beta_0 := \max\big\{\beta \in [\beta_{\min} : \beta_{\max}] : u_0 \in H^{\beta}(\mathcal{O})\big\}.$$

那么，存在常数 C_1，$C_2 > 0$ 使得

$$\mathbb{P}^N_{f_0}(\hat{\beta} \geq \beta_0) \xrightarrow{N\to\infty} 1, \tag{29}$$

$$\mathbb{P}^N_{f_0}\big(\|\hat{u}_N - u_0\|_{L^2(\mathcal{O})} \geq C_1 N^{-\frac{\beta_0}{2\beta_0+d}}\big) \xrightarrow{N\to\infty} 0, \tag{30}$$

$$\mathbb{P}^N_{f_0}\big(\|\hat{f}_N - f_0\|_{L^2(\mathcal{O})} \geq C_2 N^{-\frac{\beta_0-2}{2\beta_0+d}}\big) \xrightarrow{N\to\infty} 0. \tag{31}$$

常数 C_1 可在函数类 $\{f_0: \ \|\mathcal{G}(f_0)\|_{H^{\beta_0}(\mathcal{O})} \leq R\}$上统一选取，而 C_2可在 $\{f_0: \ \|\mathcal{G}(f_0)\|_{H^{\beta_0}(\mathcal{O})} \leq R,\ \ \|f_0\|_{C^1(\mathcal{O})} \leq R\}$上统一选取。

定理 2.8 的证明见第 B.4 节。定理 2.8 表明，过程（27）–（28）适应的是 $u_0 = \mathcal{G}(f_0)$ 的未知光滑度，而非 f_0 的未知光滑度。该自适应过程以非参数 M 估计的形式表述，主要是为了便于阐述。其他自适应策略，例如 Lepskii 型方法 [34],，也可能获得相近的结果。本方法的一个优点是，步骤 1 同时给出光滑度指数的估计，从而能够针对所有 $\gamma \in [0, \beta_0]$ 实现关于索об列夫损失 $H^{\gamma}(\mathcal{O})$ 的自适应，尤其适用于广义稳定性估计中相关的 $H^2(\mathcal{O})$-范数。据我们所知，此前达西模型中唯一的自适应速率是 [22], 所述的结果：自适应仅针对正问题实现，而未针对扩散率 f_0 实现，参见 [22] 中的定理 8–10。值得注意的是，由于这里采用了 L^2-损失函数，无需在收敛速率中引入额外的对数因子即可实现自适应，参见 [21] 第 8 章。

2.5. 薛定谔模型的结果。 现在我们转向第二个主要的偏微分方程示例，即稳态薛定谔方程的一个逆问题，例如 [42, 48, 1] 中研究的情形。本节的主要结果是定理 2.9，它为偏微分方程惩罚估计量和插件式 M 估计量都确立了极小极大最优收敛速率。正向映射 $\mathcal{G}(f)$ 通过其唯一解 $u = u_f$ 定义为

$$\begin{cases} \frac{1}{2}\Delta u - fu = 0 & \text{on } \mathcal{O}, \\ u = g & \text{on } \partial\mathcal{O}, \end{cases} \tag{32}$$

其中 $g \in C^\infty(\partial\mathcal{O})$ 是已知的严格正边界数据。每当 $f \geq 0$ 和 $f \in C^\eta(\mathcal{O})$ 对于某个 $\eta > 0$成立时，椭圆理论可得唯一的经典解 $u \in C^{2+\eta}(\mathcal{O}) \cap C(\mathcal{O})$。关于薛定谔型方程的偏微分方程理论综述，见 [12]；关于该问题在光声层析成像（PAT）中的出现，见 [4]。对于 $f\colon \mathcal{O} \to \mathbb{R}$，记

$$L_f u := \frac{1}{2}\Delta u - fu.$$

我们考虑以下两种估计程序，它们是为达西模型引入的估计程序的直接对应形式。对于 $(u, f) \in H^{\alpha+2}(\mathcal{O}) \times H^\alpha(\mathcal{O})$，定义目标函数

$$J(u,f) := \frac{1}{N}\sum_{i=1}^{N}\big(Y_i - u(X_i)\big)^2 + \lambda_N^2 \big\|L_f u - g\big\|_{L^2(\mathcal{O})}^2 + \mu_N^2\Big(\|u\|_{H^{\alpha+2}(\mathcal{O})}^2 + \|f\|_{H^\alpha(\mathcal{O})}^2\Big). \tag{33}$$

PDE 惩罚 M 估计量定义为任意一个可测极小化解

$$(\hat{u}_N, \hat{f}_N) \in \operatorname*{arg\,min}_{(u,f)\in H^{\alpha+1}(\mathcal{O})\times H^\alpha(\mathcal{O})} J(u,f), \tag{34}$$

其中正则化参数为

$$\lambda_N = N^{-\frac{2}{2(\alpha+2)+d}}, \qquad \mu_N = N^{-\frac{\alpha+2}{2(\alpha+2)+d}}.$$

或者，插件式 M 估计量通过以下两阶段程序定义

$$\hat{u}_N = \operatorname*{arg\,min}_{u\in H^{\alpha+1}(\mathcal{O})} \left\{ \frac{1}{N}\sum_{i=1}^{N}\big(Y_i - u(X_i)\big)^2 + \mu_N^2 \|u\|_{H^{\alpha+1}(\mathcal{O})}^2 \right\}, \tag{35}$$

随后

$$\hat{f}_N = \operatorname*{arg\,min}_{f\in H^\alpha(\mathcal{O})} \left\{ \big\|L_f \hat{u}_N - g\big\|_{L^2(\mathcal{O})}^2 + \nu_N^2 \|f\|_{H^\alpha(\mathcal{O})}^2 \right\}, \tag{36}$$

其中 $\mu_N = N^{-2\frac{\alpha+2}{\alpha+2\quad()+d}}$, $\nu_N = N^{-2\frac{\alpha}{2\alpha+2()+d}}$。

为避免过多的技术细节，我们推导出以下高概率收敛结果，而非均方误差结果。

定理 *2.9.* 设 $d \geq 1$，并假设 $\mathcal{O} \subseteq \mathbb{R}^d$ 是一个光滑有界区域。设 $\alpha > d/2 + 1$ 和 $\alpha \geq 3$，并固定常数 $f_{\min} \geq 0$ 和 $R > 0$。假设数据 *(10)* 源自某个 $f_0 \in \mathcal{F}(\alpha, f_{\min}, R)$，并令 $(\hat{u}_N, f_N)$ 表示在 *(34)* 中定义的 *PDE* 惩罚 *M* 估计量或在 *(35)–(36)* 中定义的插件式 *M* 估计量。存在一个仅依赖于 $\mathcal{O}, d, \alpha, f_{\min}, R$ 的 $C > 0$，使得

$$\mathbb{P}_{f_0}^N\Big(\|\hat{u}_N - \mathcal{G}(f_0)\|_{L^2(\mathcal{O})}^2 \geq C N^{-\frac{2(\alpha+2)}{2(\alpha+2)+d}}\Big) \xrightarrow{N\to\infty} 0, \tag{37}$$

$$\mathbb{P}_{f_0}^N\Big(\|\hat{f}_N - f_0\|_{L^2(\mathcal{O})}^2 \geq C N^{-\frac{2\alpha}{2(\alpha+2)+d}}\Big) \xrightarrow{N\to\infty} 0. \tag{38}$$

证明见第 B.5 节。与达西模型类似，证明的一个关键要素是薛定谔问题满足以下广义稳定性估计。

引理 2.10. 设 $\mathcal{O} \subseteq \mathbb{R}^d$ 为光滑有界区域，且 $c_{\min} > 0$。假设对于 $i \in \{1,2\}$，函数 $u_i \in H^2(\mathcal{O})$、$f_i \in C(\mathcal{O})$ 和 $h_i \in L^2(\mathcal{O})$ 满足

$$L_{f_1} u_1 = h_1 \qquad \textit{and} \qquad L_{f_2} u_2 = h_2 \quad \textit{in } \mathcal{O}.$$

此外，假设(u_1, f_1, h_1) 满足 $u_1 \in C^2(\mathcal{O})$ 和 $h_1 \in L^\infty(\mathcal{O})$，并且

$$\inf_{x \in \mathcal{O}, i=1,2} u_i(x) \geq c_{\min}.$$

那么存在常数 $C > 0$，它仅依赖于 $c_{\min}$ 以及$\|u_1\|_{C^2(\mathcal{O})}$和$\|h_1\|_{L^\infty(\mathcal{O})}$的上界，使得

$$\|f_1 - f_2\|_{L^2(\mathcal{O})} \leq C\big(\|u_1 - u_2\|_{H^2(\mathcal{O})} + \|h_1 - h_2\|_{L^2(\mathcal{O})}\big).$$

采用与达西流情形相同的技术，还可以推导出薛定谔问题高维离散化在次二次时间内可计算的类似结果，以及关于自适应估计的结果。由于这些结果与第 2.2-2.4 节的结果非常相似，因此这里不再探讨这一推广。

3. 广义 M 估计量的一个抽象定理。 现在，我们基于经典思想，在 [64], 中提出的一般结果之上，进一步阐述关于“增广 M 估计量”收敛速率的一般结论，这些经典思想贯穿于证明之中。此处的表述有意采用略显技术性的形式，以便灵活应用于本文讨论的各种估计程序。

设 $d \geq 1$，并令 $\mathcal{O} \subseteq \mathbb{R}^d$ 为一个非空的博雷尔可测子集。令 $X_i \in \mathcal{O}$ 为我们获得回归测量值的设计点

$$Y_i = u_0(X_i) + \varepsilon_i, \quad i = 1, ..., N,$$

其中 $u_0 :\ \mathcal{O} \to \mathbb{R}$ 是某个回归函数，$\varepsilon_i \in \mathbb{R}$ 是相互独立的高斯测量误差，其方差至多为 $\sigma^2 > 0$,

$$\varepsilon_i \sim N(0, \sigma_i^2), \quad \sup_{1 \leq i \leq N} \sigma_i \leq \sigma.$$

本节的结果不对 X_i 是确定性的还是随机的作任何假设。

我们考虑关于两个变量 $\eta \in H$ 和 $\theta \in \Theta$ 的 M 估计量，其中 $\eta \in H$ 应理解为回归函数的参数化变量，而 $\theta \in \Theta$ 是另一个我们感兴趣的统计参数。全文中，我们固定在 $\mathcal{O}$ 上的一种回归函数参数化，

$$\eta \mapsto u_\eta \text{ for all } \eta \in H, \ u_\eta : \mathcal{O} \to \mathbb{R}.$$

我们将考虑以下形式的泛函的最小化

$$J : H \times \Theta \to \mathbb{R}, \quad J(\eta, \theta) = \frac{1}{N} \sum_{i=1}^{n} (Y_i - u_\eta(X_i))^2 + \mathcal{R}(\eta, \theta),$$

其中 $\mathcal{R}:\ H \times \Theta \to [0, \infty)$ 是某个非负惩罚项，在本节的ˆ语境下可以任意选取。下文假设极小化解 $(\hat{\eta},\ \theta)$ J 存在，且存在一种选择 $(\hat{\eta},\ \hat{\theta})$，它是数据 $(Y_i,\ X_i)_{i=1}^N$ 的可测函数，该选择

在与上述定理相关的情形下，可以利用众所周知的可测选择定理（例如 [11]）加以验证。我们将经验范数记为

$$\frac{1}{N}\sum_{i=1}^{N} h(X_i)^2 = \|h\|_N^2, \quad h : \mathcal{O} \to \mathbb{R}.$$

对于任意 $(\eta, \theta) \in H \times \Theta$ 和函数 $u : \mathcal{O} \to \mathbb{R}$，我们定义如下“差异泛函”

$$\tau^2((\eta,\theta), u) := \|u_\eta - u\|_N^2 + \mathcal{R}(\eta,\theta).$$

利用该差异泛函，我们引入如下局部化回归函数类。对于任意“中心”元素 $\eta^* : \mathcal{O} \to \mathbb{R}$ 和 $R > 0$，令

$$\mathcal{U}(\eta^*, R) := \big\{u = u_\eta \text{ for some } (\eta,\theta) \in H \times \Theta \text{ with } \tau^2((\eta,\theta), u_{\eta^*}) \le R^2\big\}.$$

直观地说，$\mathcal{U}(\eta^*, R)$ 是由参数对 (η, θ) 表示的所有回归函数 u_η 的集合，这些函数根据 τ^2 泛函的度量，与参考点 η^*“接近”。我们可以将 η^* 视为真实值的“最佳逼近”。$\mathcal{U}(\eta^*, R)$ 关于经验范数的度量熵积分记为

$$\mathcal{I}(\eta^*, R) := \sigma R + \int_0^{\sigma R} H^{1/2}(\rho, \mathcal{U}(\eta^*, R), \sigma\|\cdot\|_N)\, d\rho.$$

定理 3.1。对于 $(\eta^*, \theta^*) \in H \times \Theta$ 任意元素，设

$$\Psi(R) \ge \mathcal{I}(\eta^*, R) \tag{39}$$

是一个上界，使得 $R \mapsto \Psi(R)/R^2$ 单调不增。存在一个普适常数 $C > 0$，使得对所有这类 (η^*, θ^*) 以及 $\Psi(\cdot)$，任意满足条件的 $N \ge 1, \delta_N > 0$ 都有

$$\sqrt{N}\delta_N^2 \ge C\Psi(\delta_N),$$

且对任意 $R \ge \delta_N$，有

$$\mathbb{P}_0^N\big(\tau^2((\hat\eta,\hat\theta), u_0) \ge 2\big(\tau^2((\eta^*,\theta^*), u_0) + R^2\big)\big) \le C\exp(-NR^2/C). \tag{40}$$

此外，对于所有 $k \ge 1$，我们还有如下矩界：

$$\mathbb{E}_0^N\big[\tau^{2k}((\hat\eta,\hat\theta), u_0)\big] \le C_k\Big(\tau^{2k}((\eta^*,\theta^*), u_0) + \delta_N^{2k} + \frac{\delta_N^{2k-2}\sigma^2}{N}\Big), \quad \textit{some } C_k > 0. \tag{41}$$

证明见第C节，它将惩罚最小二乘估计量的经典经验过程论证推广到当前的增广参数设定，特别是 [64], 。其核心思想是基于差异泛函 τ^2 进行局部化论证，并结合针对局部化类 $\mathcal{U}(\eta^*, R)$ 的高斯过程集中不等式。随后，关于 $\Psi(R)/R^2$ 的单调性假设使我们能够采用标准的“剥离论证”。

注 3.2. 虽然需要处理额外的技术细节，但我们的结果可以推广到独立次高斯噪声变量的情形，只要这些变量满足某个 $\sigma > 0$

$$\mathbb{E}\big[e^{\lambda\varepsilon_i}\big] \le \exp(\lambda^2\sigma^2/2), \qquad \forall\lambda \in \mathbb{R},\ \forall i \ge 1. \tag{42}$$

在这种情况下，证明中使用的 Borell–Sudakov–Tsirelson 不等式可以替换为次高斯过程相应的尾部不等式；例如，参见 [64] 中的推论 8.3。我们不再探讨这一推广。

注 3.3（白噪声）。类似于定理 3.1的集中结果也可以在白噪声观测模型中表述；该模型在文献中也很常见，参见例如 [47, 32]。这里， u_0 可以属于一般的可分希尔伯特空间 $\mathbb{H}$，并且观测到高斯过程的一次实现

$$Y_N = u_0 + N^{-1/2}\mathbb{W},$$

其中， $\mathbb{W}$ 是以 $\mathbb{H}$为指标、协方差为$\mathbb{E}[\mathbb{W}(g)\,\mathbb{W}(h)] = \langle g,\ h\rangle_{\mathbb{H}}$的等距正态中心化白噪声过程。在这种情况下，考虑参数化 $\eta \mapsto u_\eta \in \mathbb{H}$，目标函数定义为

$$(\hat{\eta}, \hat{\theta}) \in \operatorname*{arg\,min}_{(\eta,\theta)\in H\times\Theta} -2\langle Y_N, u_\eta\rangle + \|u_\eta\|_{\mathbb{H}}^2 + \mathcal{R}(\eta,\theta).$$

证明的思路与离散设计情形基本相同，只需将全文中的经验范数 $\|\cdot\|_N$替换为$\|\cdot\|$范数。由于本文的应用针对的是有限个 $N \geq 1$ 测量的离散化情形，而计算复杂度问题在这种情形下最为自然，因此我们在此不再进一步探讨白噪声表述。

4. 进一步讨论与未来方向。

*边界及其他惩罚项。*我们的目标函数不包含用于强制满足边界条件的显式惩罚项。相比之下，在物理信息神经网络（PINNs）中，通常会将边界失配项纳入损失函数，以促使其满足偏微分方程约束 [50]。对于本文研究的基于偏微分方程的逆问题，这并非必要：下文建立的广义稳定性估计表明，即使偏微分方程约束仅以弱形式施加，单凭内部信息也足以控制参数误差。尽管如此，对于其他逆问题或不同的算子方程，额外的惩罚项——包括边界项或其他结构正则化项——可能有所帮助，而我们的框架也能轻易扩展以纳入这些项。

*不确定性量化（UQ）。*一个自然的问题是，当前这些非似然方法能否不仅达到最优收敛速率，还能实现与统计不确定性量化（UQ）相关的更精细的最优性概念，例如克拉美–罗意义下的半参数效率 [65]。具体而言，可以考察基于我们的代理目标函数的估计量，能否为如下形式的正则线性泛函实现尽可能小的渐近方差

$$\sqrt{N}\,\langle \hat{f}_N - f_0, \psi\rangle,$$

对于适当的检验函数 ψ。在满足伯恩斯坦–冯·米塞斯定理的情形下，已知贝叶斯方法对这类泛函具有渐近最优性。然而，在达西反问题中，由于正向映射 $\mathcal{G}$ 所导致的值域限制，沿任意光滑方向 ψ 都无法实现 $\sqrt{N}$ 一致估计；参见 [43] 及 [65] 第25章。本文研究的目标函数基于易于处理的似然替代形式，因此不再与费舍尔信息算子 $D\mathcal{G}_{f_0}^* D\mathcal{G}_{f_0}$ 直接关联。因此，尽管这些目标函数能达到最优速率，原则上仍可能导致次优的渐近协方差结构。对渐近分布和ˆ效率的详细分析——包括ˆu_N 和 f_N 的极限分布——超出了本文的范围，也是未来研究中一个有意义的方向。关于 M 估计量极限分布的相关结果，另请参见 [57] 。

参考文献

[1] AGAPIOU, S. 和 WANG, S.（2024）。贝叶斯反问题中的拉普拉斯先验与空间非均匀性。Bernoulli 30 878–910。[2] ALTMEYER, R.（2022）。高维广义线性模型中基于采样的后验推断的多项式时间保证。arXiv 预印本 arXiv:2208.13296。 [3] ARNONE, E., KNEIP, A., NOBILE 和 SANGALLI, L. M.（2022）。偏微分方程正则化空间回归一致性的一些初步结果。《统计学报》32 209–238。 [4] BAL, G. 和 UHLMANN, G.（2010）。光声学的逆扩散理论。Inverse Problems 26 085010。[5] BANDEIRA, A. S., MAILLARD, A., NICKL 和 WANG, S.（2023）。高斯先验中的自由能垒及冷启动马尔可夫链蒙特卡罗在高维单峰分布中的失效。《英国皇家学会哲学汇刊 A 辑》381 20220150。[6] BENNING, M. 和 BURGER, M.（2018）。逆问题的现代正则化方法。《数值学报》27 1–111。[7] BISSANTZ, N., HOHAGE, T., MUNK 和 RUYMGAART, F.（2007）。统计逆问题中一般正则化方法的收敛速率及其应用。SIAM 数值分析杂志 45 2610–2636。 [8] BOHR, J. 和 NICKL, R.（2024）。高维后验测度的对数凹近似及非线性逆问题中的稳定性。《亨利·庞加莱研究所年刊（B辑）：概率与统计》60 2619–2667。亨利·庞加莱研究所。[9] BONITO, A., COHEN, A., DEVORE, R., PETROVA 和 WELPER, G.（2017）。椭圆偏微分方程扩散系数估计。《SIAM 数学分析杂志》49 1570–1592。https://doi.org/ 10.1137/16M1094476 MR3639575[10] BOYD, S. 和 VANDENBERGE, L.（2004）。凸优化。剑桥大学出版社。 [11] BROWN, L. D. 和 PURVES, R.（1973）。极值的可测选择。统计年刊 902–912。[12] CHUNG, K. L. 和 ZHAO, Z. X.（1995）。从布朗运动到薛定谔方程。施普林格出版社，柏林。https://doi.org/ 10 1007 978 3 642 57856 4 1329992 . / - - - - [13] MR COHEN, A., DEVORE 和 SCHWAB, C.（）。参数和随机椭圆偏微分方程的解析正则性与多项式逼近。《分析与应用》9 11–47。 2011 [14] COTTER, S. L., ROBERTS, G. O., STUART, A. M. 和 WHITE, D.（2013）。函数的马尔可夫链蒙特卡洛方法：改进旧算法以提升速度。《统计科学》28 424-446。 [15] CUOMO, S., DI COLA, V. S., GIAMPAOLO, F., ROZZA, G., RAISSI, M. 和 PICCIALLI, F.（2022）。通过物理信息神经网络实现科学机器学习：现状与未来方向。《科学计算杂志》92 88。[16] DALALYAN, A. S.（2017）。从光滑对数凹密度进行近似采样的理论保证。《英国皇家统计学会杂志 B辑：统计方法学》79 651–676。https://doi.org/10.1111/rssb.12183 MR3641401[17] DOUMÈCHE, N., BIAU, G. 和 BOYER, C.（2025）。关于物理信息神经网络的收敛性。Bernoulli 31 2127–2151。[18] DURMUS, A. 和 MOULINES, E.（2018）。通过未校正朗之万算法进行高维贝叶斯推断。arXiv 预印本。[19] GHOSAL, S. 和 VAN DER VAART, A.（2007）。非独立同分布观测下后验分布的收敛速率。《统计年刊》35 192–223。[20] GILBARG, D. 和 TRUDINGER, N. S.（1998）。二阶椭圆偏微分方程。施普林格出版社，柏林—纽约。[21] GINÉ, E. 和 NICKL, R.（2016）。无限维统计模型的数学基础。剑桥大学出版社，纽约。https://doi.org/ 10 1017 9781107337862 3588285 . / [22] CBO MR GIORDANO, M. 和 NICKL, R.（）。椭圆逆问题中采用高斯过程先验的贝叶斯推断的一致性。arXiv 预印本。[23] HAIRER, M., STUART, A. M. 和 VOLLMER, S. J.（2014）。无限维空间中 Metropolis–Hastings 算法的谱隙。《应用概率年刊》24 2455-2490。[24] HANKE, M.（2017）。不适定问题的共轭梯度型方法。Chapman and Hall/CRC。[25] HANKE, M., NEUBAUER, A. 和 SCHERZER, O.（1995）。非线性不适定问题中 Landweber 迭代的收敛性分析。《数值数学》72 21–37。[26] HINZE, M., PINNAU, R., ULBRICH, M. 和 ULBRICH, S.（2008）。带偏微分方程约束的优化 23。施普林格科学与商业媒体。[27] HOFMANN, B., KALTENBACHER, B., POESCHL, C. 和 SCHERZER, O.（2007）。巴拿赫空间中非光滑算子的 Tikhonov 正则化收敛速率结果。Inverse Problems 23 987。[28] ISAKOV, V.（2006）。偏微分方程的逆问题。施普林格。[29] KAIPIO, J. P. 和 SOMERSALO, E.（2005）。统计与计算逆问题。施普林格。

[30] KALTENBACHER, B. (2016). 基于一体化表述的逆问题正则化。SIAM 数值分析杂志 54 2594–2618. [31] KALTENBACHER, B., NEUBAUER, A. and SCHERZER, O. (2008). 非线性不适定问题的迭代正则化方法。de Gruyter，柏林。https://doi.org/

10 1515 9783110208276 2459012 .

[32] 2024

/ MR KOERS, G., SZABO, B. and VAN DER VAART, A. ( ). 非线性逆问题的线性方法。arXiv 预印本 arXiv:2411.19797. [33] KUNISCH, K. and ITO, K. (1994). 关于椭圆边值问题中系数到解映射的单射性与线性化。数学分析与应用杂志 188 1040-1066.[34] LEPSKII, O. (1991). 高斯白噪声中的自适应估计问题。概率论及其应用 35 454–466. [35] LI, Z., KOVACHKI, N., AZIZZADENESHELI, K., LIU, B., BHATTACHARYA, K., STUART, A. and ANAND- KUMAR, A. (2020). 用于参数化偏微分方程的傅里叶神经算子。arXiv 预印本 arXiv:2010.08895.[36] LI, Z., ZHENG, H., KOVACHKI, N., JIN, D., CHEN, H., LIU, B., AZIZZADENESHELI, K. and ANANDKU- MAR, A. (2024). 用于学习偏微分方程的物理信息神经算子。ACM/IMS 数据科学杂志 1 1–27. [37] LU, Y., CHEN, H., LU, J., YING, L. and BLANCHET, J. (2021). 椭圆型偏微分方程的机器学习：快速收敛率泛化界、神经缩放定律与极小极大最优性。arXiv 预印本 arXiv:2110.06897. [38] MEYN, S. P. and TWEEDIE, R. L. (2009). 马尔可夫链与随机稳定性。剑桥大学出版社，剑桥。[39] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2021). 含噪非阿贝尔X射线变换的一致反演。纯粹与应用数学通讯 74 1045–1099. [40] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2021). 高斯过程先验下非线性逆问题贝叶斯不确定性量化的统计保证。统计年刊 49 3255 3298– .[41] NEUBAUER, A. (1989). 非线性不适定问题的Tikhonov正则化：最优收敛速率与有限维近似。Inverse problems 5 541. [42] NICKL, R. (2020). 统计逆问题的伯恩斯坦–冯·米泽斯定理 I：薛定谔方程。欧洲数学学会杂志 22 2697–2750. [43] NICKL, R. (2023). 贝叶斯非线性统计逆问题。苏黎世高级数学讲座。EMS出版社.[44]NICKL, R. (2024). 时间演化方程的伯恩斯坦–冯·米泽斯定理。arXiv 预印本 arXiv:2407.14781.[45] NICKL, R., PAVLIOTIS, G. A. and RAY, K. (2025). 麦基恩–弗拉索夫模型中的贝叶斯非参数推断。统计年刊 53 170–193. [46] NICKL, R. and TITI, E. S. (2024). 关于高斯过程先验下数据同化的后验一致性：二维纳维–斯托克斯方程。统计年刊 52 1825–1844. [47] NICKL, R., VAN DE GEER, S. and WANG, S. (2018). 偏微分方程约束回归问题中惩罚最小二乘估计量的收敛速率。arxiv 预印本。[48] NICKL, R. and WANG, S. (2022). 通过朗之万型算法进行高维后验测度的多项式时间计算。欧洲数学学会杂志 26 1031–1112. [49] PATERNAIN, G. P., SALO, M. and UHLMANN, G. (2023). 几何逆问题 204。剑桥大学出版社。[50] RAISSI, M., PERDIKARIS, P. and KARNIADAKIS, G. E. (2019). 物理信息神经网络：用于求解涉及非线性偏微分方程的正问题与逆问题的深度学习框架。计算物理学杂志 378 686–707. [51] RAISSI, M., PERDIKARIS, P. and KARNIADAKIS, G. E. (2019). 物理信息神经网络：用于求解涉及非线性偏微分方程的正问题与逆问题的深度学习框架。计算物理学杂志 378 686–707. [52] REINHARDT, N., WANG, S. and ZECH, J. (2024). 神经算子的统计学习理论。arXiv 预印本 arXiv:2412.17582. [53] RICHTER, G. R. (1981 41 210 221 10 1137 0141016). 稳态扩散方程的一个逆问题。SIAM 应用数学杂志 – . https://doi.org/ . / MR 628945

[54] 1981 RICHTER, G. R. ( ). 空间变化扩散系数的数值识别。计算数学 36 375–386. [55] ROBERT, C. P., CASELLA, G. and CASELLA, G. (1999). 蒙特卡罗统计方法 2。施普林格。[56] SANGALLI, L. M. (2021). 带偏微分方程正则化的空间回归。国际统计评论 89 505–531.

[57] SPOKOINY, V. (2023). 非线性回归估计中的有限样本推断。arXiv 预印本 arXiv:2305.08193.[58] STUART, A. M. (2010). 逆问题：贝叶斯视角。Acta Numer. 19 451–559. [59] TALAGRAND, M. (2014). 随机过程的上界和下界。施普林格。[60] TRIEBEL, H. (1978). 插值理论、函数空间、微分算子。North-Holland Publishing Company。[61] TRIEBEL, H. (2008). 区域上的函数空间与小波。EMS数学专著丛书 7。欧洲数学会（EMS），苏黎世。https://doi.org/
10 4171 019 2455724 . / MR
[62]. 2010
TRÖLTZSCH, F. ( ). 偏微分方程最优控制：理论、方法与应用 112。American Mathematical Soc.[63] VAN DE GEER, S. (2001). 带复杂度惩罚的最小二乘估计。统计数学方法 10 355-374.[64] VAN DE GEER, S. A. (2000). 经验过程理论的应用。剑桥大学出版社，剑桥。[65] VAN DER VAART, A. W. (1998). 渐近统计。剑桥大学出版社，剑桥。https: //doi.org/
10 1017 9780511802256 1652247 . /CBO MR VAN DER VAART, A. W. and
[66] 1996
WELLNER, J. A. ( ). 弱收敛与经验过程。施普林格统计丛书。施普林格出版社，纽约。MR1385671 (97g:60035) [67] VOLLMER, S. J. (2013). 通过稳定性与回归结果实现贝叶斯逆问题的后验一致性。Inverse Problems 29 125011.

附录 A：辅助偏微分方程结果

我们介绍两个主要偏微分方程示例——达西模型和薛定谔模型——的关键解析性质。所需性质包括：首先，正向映射 $\mathcal{G}$ 在适当的（索 Sobolev）光滑性球之间的有界性；其次，针对 $\mathcal{G}^{-1}$ 的广义稳定性估计（如上文所述）。请注意，正向映射 $\mathcal{G}$ 的有界性要求弱于贝叶斯估计量和最大后验估计量收敛证明中所需的条件 [47, 22, 43]，后者要求 $\mathcal{G}$ 满足局部利普希茨估计。我们不要求此类利普希茨估计的技术原因在于，u 的正则性已被显式惩罚。

A.1. 达西流。以下引理断言，达西算子将 H^α-有界集映射到 $H^{\alpha+1}$-有界集。

引理 *A.1.* 设达西算子 $\mathcal{G}$ 由式 *(11)* 给出，其中 $g \in C^\infty(\mathcal{O})$。设 $\mathcal{F} = \mathcal{F}(\alpha, f_{min}, R)$ 由式 *(9)* 给出，其中 $f_{min} > 0$、$R > 0$ 和 $\alpha > d/2 + 1$。此外，假设 $g \in H^{\alpha-1}(\mathcal{O})$。存在常数 $C = C(d, \mathcal{O}, f_{min}, g)$，使得对所有满足 $f \geq f_{min}$ 的 $f \in H^\alpha(\mathcal{O})$，均有

$$\|\mathcal{G}(f)\|_{H^{\alpha+1}} \leq C(1 + \|f\|_{H^\alpha}^\gamma).$$

特别地，有

$$\sup_{f \in \mathcal{F}(\alpha, f_{min}, R)} \|\mathcal{G}(f)\|_{H^{\alpha+1}} \vee \|\mathcal{G}(f)\|_{C^2} =: M(\alpha, f_{min}, R) < \infty.$$

证明。第一个结论由 [47]，的引理 23 得出；注意，$\alpha > d/2+2$ 的正则性假设（由 [47] 中的引理 22 继承而来）可以无代价地减弱为 $\alpha > d/2+1$。[47] 的引理 23 证明了 $\gamma > 0$ 可以取为 $\gamma = \alpha^2 + \alpha$。第二个结论由索伯列夫嵌入 $C^2 \subseteq H^{\alpha+1}$ 得出，这得益于假设 $\alpha > d/2 + 1$。 □

接下来，我们来证明达西模型的广义稳定性，并为方便读者再次陈述该性质。

引理 2.2. 设 $\mathcal{O} \subseteq \mathbb{R}^d$ 是一个光滑有界区域，且令 f_{min}，$g_{min} > 0$。假设对于 $i \in \{1, 2\}$，我们有函数

$$u_i \in H^2(\mathcal{O}), \quad f_i \in C^1(\mathcal{O}), \quad g_i \in L^2(\mathcal{O}) \quad \textit{with} \quad \nabla \cdot (f_i \nabla u_i) = g_i \quad \textit{on } \mathcal{O}.$$

假设第一个三元组 (u_1, f_1, g_1) 还满足 $u_1 \in C^2(\mathcal{O}) \cap C(\bar{\mathcal{O}})$，$g_1 \in C(\mathcal{O})$以及

$$\inf_{x \in \mathcal{O}} f_1(x) \geq f_{min} \quad \textit{and} \quad \inf_{x \in \mathcal{O}} g_1(x) \geq g_{min}, \quad \textit{and} \quad u_1|_{\partial\mathcal{O}} = 0.$$

那么，对于每个 $B > 0$，都存在一个仅依赖于 $f_{min}, g_{min}, \mathcal{O}$ 和 B的常数 $C > 0$，使得只要 $\|f_1\|_{C^1} \vee \|u_1\|_{C^2} \leq B$，便有

$$\|f_1 - f_2\|_{L^2} \leq C\big(\|f_2\|_{C^1}\|u_1 - u_2\|_{H^2} + \|g_1 - g_2\|_{L^2}\big).$$

证明 引理2.2。我们将证明分为三个步骤。我们的证明基于 [47, 引理 24] 中使用椭圆型偏微分方程理论的方法。

步骤 *1*。 对于 f_1, f_2 如假设条件中所述，记为 $h = f_1 - f_2$。根据假设， h 满足 es

$$\begin{aligned}\nabla \cdot (h\nabla u_1) &= \nabla \cdot (f_1 \nabla u_1) - \nabla \cdot (f_2 \nabla u_2) - \nabla \cdot (f_2 \nabla (u_1 - u_2)) \\ &= \nabla \cdot (f_2 \nabla (u_2 - u_1)) + (g_1 - g_2).\end{aligned}$$

右侧的 L^2-范数还可以进一步由以下量给出上界

$$\|\nabla \cdot (f_2 \nabla (u_1 - u_2))\|_{L^2} + \|g_1 - g_2\|_{L^2} \le 2\|f_2\|_{C^1}\|u_2 - u_1\|_{H^2} + \|g_1 - g_2\|_{L^2}. \tag{43}$$

步骤 *2*：通过格林公式得到乘子恒等式。 总体目标是为左侧的第一项（43）建立下界。为此，我们证明以下恒等式：对于 $v = e^{-\lambda u_1} h$（其中 $\lambda > 0$ 稍后选取，此处暂取任意值），

$$\langle \nabla \cdot (h\nabla u_1), h e^{-\lambda u_1} \rangle_{L^2} = \langle \frac{1}{2}\Delta u_1 + \lambda \|\nabla u_1\|^2, v^2 \rangle_{L^2} + \frac{1}{2}\int_{\partial \mathcal{O}} v^2 \frac{\partial u_1}{\partial n}, \tag{44}$$

其中 $n = n(x)$ 表示 $x \in \partial\mathcal{O}$上的外法向量。注意，由于$\|u_1\|_{C^2(\mathcal{O})} \le B$且 $f_1 - f_2 \in C^1(\mathcal{O})$ （其 $C^1(\mathcal{O})$-范数有限），函数 v、 u_1和 ∇u_1 在 $\mathcal{O}$ 上满足利普希茨连续，并且可以唯一连续延拓至 $\mathcal{O}$；边界项按此意义理解。为证明（44），我们使用格林恒等式（分部积分）得到

$$\langle \Delta u_1, v^2 \rangle + \langle \nabla u_1, \nabla (v^2) \rangle = \int_{\mathcal{O}} \nabla \cdot (v^2 \nabla u_1) = \int_{\partial \mathcal{O}} \frac{\partial u_1}{\partial n} v^2,$$

然后还有

$$\langle \Delta u_1, v^2 \rangle + \frac{1}{2}\langle \nabla u_1, \nabla (v^2) \rangle = \frac{1}{2}\langle \Delta u_1, v^2 \rangle + \frac{1}{2}\int_{\partial \mathcal{O}} \frac{\partial u_1}{\partial n} v^2.$$

此外，根据定义 $v = e^{-\lambda u_1} h$，我们有

$$\nabla v = e^{-\lambda u_1}\nabla h - \lambda e^{-\lambda u_1} h \nabla u_1 = e^{-\lambda u_1}(\nabla h - \lambda h \nabla u_1).$$

由于 $\nabla (v^2) = 2v\nabla v$，我们得到

$$\begin{aligned}\frac{1}{2}\langle \nabla u_1, \nabla (v^2) \rangle_{L^2} &= \langle \nabla u_1, v \nabla v \rangle_{L^2} \\ &= \left\langle \nabla u_1,\ v e^{-\lambda u_1}(\nabla h - \lambda h \nabla u_1) \right\rangle_{L^2} \\ &= \langle \nabla h \cdot \nabla u_1,\ v e^{-\lambda u_1} \rangle_{L^2} - \lambda \langle \|\nabla u_1\|^2,\ v^2 \rangle_{L^2}.\end{aligned}$$

利用前述方程和 $h e^{-2\lambda u_1} = v e^{-\lambda u_1}$，可得

$$\begin{aligned}\langle \nabla \cdot (h\nabla u_1), h e^{-2\lambda u_1} \rangle_{L^2} &= \langle h\Delta u_1 + \nabla h \cdot \nabla u_1, h e^{-2\lambda u_1} \rangle_{L^2} \\ &= \langle h\Delta u_1, h e^{-2\lambda u_1} \rangle_{L^2} + \frac{1}{2}\langle \nabla u_1, \nabla (v^2) \rangle_{L^2} + \langle \lambda \|\nabla u_1\|^2, v^2 \rangle_{L^2} \\ &= \langle \Delta u_1, v^2 \rangle_{L^2} + \frac{1}{2}\langle \nabla u_1, \nabla (v^2) \rangle_{L^2} + \langle \lambda \|\nabla u_1\|^2, v^2 \rangle_{L^2} \\ &= \langle \frac{1}{2}\Delta u_1 + \lambda \|\nabla u_1\|^2, v^2 \rangle_{L^2} + \frac{1}{2}\int_{\partial \mathcal{O}} v^2 \frac{\partial u_1}{\partial n}.\end{aligned}$$

步骤 *3*：下界 (44)。我们证明，(44) 的右端有一个正数倍的 $\|h\|^2_{L^2(\mathcal{O})}$作为下界，只需在 v 的定义中适当选取 $\lambda = \lambda(g_{\min}, B) > 0$。

（*i*）边界项的正性。由于 $g_1 \geq g_{\min} > 0$ 和 $u_1|_{\partial\mathcal{O}} = 0$，强最大值原理表明

$$u_1(x) < 0, \qquad x \in \mathcal{O}. \tag{45}$$

因此，根据 Hopf边界点引理（例如，参见 [20, 引理 3.4]），我们有

$$\frac{\partial u_1}{\partial n}(x) > 0, \qquad x \in \partial\mathcal{O}, \tag{46}$$

因此，

$$\frac{1}{2}\int_{\partial\mathcal{O}} v^2 \frac{\partial u_1}{\partial n} \geq 0. \tag{47}$$

（*ii*）乘子的逐点下界。回忆一下，u_1 满足 $\nabla\cdot(f_1\nabla u_1) = g_1$，因此

$$g_1(x) = f_1(x)\Delta u_1(x) + \nabla f_1(x)\cdot\nabla u_1(x), \qquad x \in \mathcal{O}. \tag{48}$$

利用 $f_1 \geq f_{\min}$ 和 $\|\nabla f_1\|_\infty \leq \|f_1\|_{C^1} \leq B$，我们推得对每个 $x \in \mathcal{O}$，

$$g_{\min} \leq f_1(x)\Delta u_1(x) + \|\nabla f_1\|_\infty \|\nabla u_1(x)\| \leq B\Delta u_1(x) + B\|\nabla u_1(x)\|.$$

因此，对每个 $x \in \mathcal{O}$，以下两个不等式中至少有一个成立：

$$\Delta u_1(x) \geq \frac{g_{\min}}{2B} \qquad \text{or} \qquad \|\nabla u_1(x)\| \geq \frac{g_{\min}}{2B}. \tag{49}$$

由于 $\|u_1\|_{C^2} \leq B$，我们有 $\|\Delta u_1\|_\infty \leq c(B)$，其中某个 $c(B) < \infty$成立。因此，取足够大的$\lambda = \lambda(g_{\min}, B) > 0$，可得逐点下界

$$\frac{1}{2}\Delta u_1(x) + \lambda\|\nabla u_1(x)\|^2 \geq c_0, \qquad x \in \mathcal{O}, \tag{50}$$

其中某个常数 $c_0 = c_0(g_{\min}, B) > 0$。

(iii) 结论。结合式（44）、式（47）和式（50）可得

$$\left\langle \nabla\cdot(h\nabla u_1),\, he^{-2\lambda u_1}\right\rangle_{L^2} \geq c_0\|v\|^2_{L^2(\mathcal{O})}.$$

另一方面，由柯西–施瓦茨不等式以及界 $\|e^{-2\lambda u_1}\|_\infty \leq \exp(2\lambda B) < \infty$，

$$\left|\left\langle \nabla\cdot(h\nabla u_1),\, he^{-2\lambda u_1}\right\rangle_{L^2}\right| \leq \|\nabla\cdot(h\nabla u_1)\|_{L^2}\,\|he^{-2\lambda u_1}\|_{L^2} \leq \exp(2\lambda B)\|\nabla\cdot(h\nabla u_1)\|_{L^2}\,\|h\|_{L^2}.$$

因此，

$$\|\nabla\cdot(h\nabla u_1)\|_{L^2}\,\|h\|_{L^2} \geq c\,\|v\|^2_{L^2(\mathcal{O})}, \tag{51}$$

对于某个 $c = c(g_{\min},\ f_{\min},\ B,\ \mathcal{O},\ d) > 0$。由于 u_1 有界且 $v = e^{-\lambda u_1}h$，我们有 $\|v\|_{L^2} \asymp \|h\|_{L^2}$ 其中常数仅依赖于 λ 和 B。将其代入（51）得

$$\|\nabla\cdot(h\nabla u_1)\|_{L^2} \gtrsim \|h\|_{L^2},$$

将此界与（43）结合，即可得到所声称的稳定性估计。 □

A.2. 薛定谔模型。以下引理断言，薛定谔参数到解的映射将 H^{α}有界集映射到 $H^{\alpha+2}$有界集。这是对[47, 引理 27] 的直接重述（关于边界数据 g的假设相同）；列出此引理仅为方便起见。我们还注意到，第二个结论可由索伯列夫嵌入$H^{\alpha+2} \subseteq C^2$（$\mathcal{O}$）对 $\alpha > d/2$ 得出。

引理 A.2. 考虑式（*32*）中的薛定谔算子 $\mathcal{G}$，其中 $g \in C^{\infty}(\partial\mathcal{O})$ 固定。则对于任意 $\alpha > d/2$，存在常数 $C > 0$，使得对所有满足 $f \geq 0$ 的 $f \in H^{\alpha}(\mathcal{O})$

$$\|\mathcal{G}(f)\|_{H^{\alpha+2}} \leq C(1 + \|f\|_{H^{\alpha}}^{\gamma}).$$

特别地，对于任意 $R > 0$，有

$$\sup_{f \geq 0,\ \|f\|_{H^{\alpha}} \leq R} \|\mathcal{G}(f)\|_{H^{\alpha+2}} \vee \|\mathcal{G}(f)\|_{C^2} =: M(\alpha, R) < \infty.$$

接下来，我们给出薛定谔方程广义稳定性性质的证明，并为方便读者再次陈述该性质。

引理 2.10. 设 $\mathcal{O} \subseteq \mathbb{R}^d$ 是一个光滑有界区域，且令 $c_{\min} > 0$。假设对于 $i \in \{1, 2\}$，函数 $u_i \in H^2(\mathcal{O})$、$f_i \in C(\mathcal{O})$ 和 $h_i \in L^2(\mathcal{O})$ 满足

$$L_{f_1} u_1 = h_1 \qquad \textit{and} \qquad L_{f_2} u_2 = h_2 \quad \textit{in } \mathcal{O}.$$

此外，假设 (u_1, f_1, h_1) 满足 $u_1 \in C^2(\mathcal{O})$ 和 $h_1 \in L^{\infty}(\mathcal{O})$，并且

$$\inf_{x \in \mathcal{O},\, i=1,2} u_i(x) \geq c_{\min}.$$

因此，存在一个常数 $C > 0$，其仅取决于 $c_{\min}$ 以及 $\|u_1\|_{C^2(\mathcal{O})}$和 $\|h_1\|_{L^{\infty}(\mathcal{O})}$的上界，使得

$$\|f_1 - f_2\|_{L^2(\mathcal{O})} \leq C\big(\|u_1 - u_2\|_{H^2(\mathcal{O})} + \|h_1 - h_2\|_{L^2(\mathcal{O})}\big).$$

引理2.10的证明。对于 $i = 1,\ 2$，有 $f_i = (\Delta u_i/2 - h_i)/u_i$。映射 $z \mapsto z^{-1}$ 的利普希茨连续常数在区间 $[c_{min}, \infty)$上有界于 c_{min}^{-2}。因此，所需的稳定性估计如下得到：

$$\begin{aligned}
\|f_1 - f_2\|_{L^2} &= \Big\|\frac{\Delta u_1/2 - h_1}{u_1} - \frac{\Delta u_2/2 - h_2}{u_2}\Big\|_{L^2} \\
&\leq \Big\|\frac{\Delta u_1 - \Delta u_2}{2u_2}\Big\|_{L^2} + \Big\|\Delta u_1\big(\frac{1}{2u_1} - \frac{1}{2u_2}\big)\Big\|_{L^2} + \Big\|\frac{h_2 - h_1}{u_2}\Big\|_{L^2} + \Big\|h_1\big(\frac{1}{u_1} - \frac{1}{u_2}\big)\Big\|_{L^2} \\
&\leq c_{min}^{-1}\|u_1 - u_2\|_{H^2} + c_{min}^{-2}\|u_1\|_{C^2}\|u_1 - u_2\|_{L^2} + c_{min}^{-1}\big\|h_2 - h_1\big\|_{L^2} \\
&\qquad + c_{min}^{-2}\|h_1\|_{\infty}\|u_1 - u_2\|_{L^2}. \\
&\leq C(\|u_1 - u_2\|_{H^2} + \|h_1 - h_2\|_{L^2}),
\end{aligned}$$

其中，常数 $C < \infty$ 仅取决于 c_{min}、$\|u_1\|_{C^2}$ 和 $\|h_1\|_{\infty}$。 □

附录 B：第2节的证明

我们在第2节给出主要结果的证明。全文中，经验范数记作 $\|h\|_N := \big(N^{-1}\sum_{i=1}^{N} h^2(X_i)\big)^{1/2}$。

B.1. 定理2.1的证明。

B.1.1. *PDE* 惩罚 *M* 估计量。 我们首先讨论 PDE 惩罚 M 估计量的情形。

步骤 *1*：正则化经验风险的集中不等式。我们将定理3.1 中的 $\eta \in H$ （采用“恒等参数化”）替换为 $\sigma = 1, u \in H^{\alpha+1}(\mathcal{O})$，将 $\theta \in \Theta$ 替换为 $f \in H^{\alpha}(\mathcal{O})$，并取正则项

$$\mathcal{R}(u,f) = \lambda_N^2 \|L_f u - g\|_{L^2(\mathcal{O})}^2 + \mu_N^2 \Big(\|u\|_{H^{\alpha+1}(\mathcal{O})}^2 + \|f\|_{H^{\alpha}(\mathcal{O})}^2 \Big), \tag{52}$$

这里回顾一下， $\lambda_N = N^{-\frac{2}{2}\frac{\alpha+1}{()+d}}$，且 $\mu_N = N^{-\frac{2}{2}\frac{\alpha+1}{\alpha+1()+d}}$。由于 $f_0 \in \mathcal{F}$，根据引理 A.1， $u_0 = \mathcal{G}(f_0)$ 具有有限的 $H^{\alpha+1}(\mathcal{O})$-范数，

$$\sup_{f \in \mathcal{F}(\alpha, f_{min}, R)} \|\mathcal{G}(f)\|_{H^{\alpha+1}(\mathcal{O})} =: M < \infty.$$

令 $(f^*, u^*) = (f_0, u_0)$。根据 $\mathcal{R}$ 的定义，局部回归函数类满足

$$\mathcal{U}(\eta^*, R) \subseteq \{u : \|u\|_{H^{\alpha+1}} \leq R/\mu_N\}.$$

因此，利用有界光滑区域上的索博列夫类的标准熵界（参见 [60] 中的定理 4.10.3），我们可以用经验范数将熵积分上界估计为

$$\int_0^R \sqrt{H(\mathcal{U}(\eta^*, R), \|\cdot\|_\infty, \rho)} d\rho \lesssim \int_0^R \Big(\frac{R}{\mu_N \rho} \Big)^{\frac{d}{2(\alpha+1)}} d\rho \lesssim R \mu_N^{-\frac{d}{2(\alpha+1)}}. \tag{53}$$

这里我们利用了 $\alpha + 1 > d/2$ （根据假设）。因此，对于适当的常数 $C_1 > 1$，我们可以设

$$\Psi(R) := C_1 R \mu_N^{-\frac{d}{2(\alpha+1)}} \geq R + \int_0^R \sqrt{H(\mathcal{U}(\eta^*, R), \|\cdot\|_\infty, \rho)} d\rho, \tag{54}$$

其中我们利用了 $\mu_N \leq 1$ 对所有 $N \geq 1$均成立。注意， $\Psi(R)/R^2$ 关于 $R > 0$单调递减。根据我们对 μ_N的选择，有 $\mu_N^{-d/(2(\alpha+1))} = N^{d/(4(\alpha+1)+2d)}$。于是，对于 $C > 0$ 作为定理 3.1中的常数，我们可以验证条件

$$\sqrt{N}\delta_N^2 \geq C\Psi(\delta_N) = CC_1 \delta_N N^{\frac{d}{4(\alpha+1)+2d}}$$

对 $\delta_N = C_2 N^{-\frac{2}{2}\frac{\alpha+1}{()+d}}$成立，其中 $C_2 > 0$ 取充分大的值。令

$$\tau^2((u,f), u_0) = \|u - u_0\|_N^2 + \mathcal{R}(u,f).$$

有

$$\tau^2((u^*, f^*), u_0) = \mu_N^2 (\|f_0\|_{H^{\alpha}(\mathcal{O})}^2 + \|u_0\|_{H^{\alpha+1}(\mathcal{O})}^2) \leq (R^2 + M^2)\mu_N^2 \lesssim N^{-\frac{\alpha+1}{2(\alpha+1)+d}},$$

定理 3.1 因此可得，对所有 $L \geq L_0 > 0$ (L_0 充分大的值）以及某个 $C_3 > 0$,

$$\mathbb{P}_0^N \Big(\tau^2((\hat{u}_N, \hat{f}_N), u_0) \geq L^2 N^{-\frac{2(\alpha+1)}{2(\alpha+1)+d}} \Big) \leq C_3 \exp\big(- L^2 N^{\frac{d}{2(\alpha+1)+d}} / C_3 \big), \tag{55}$$

以及矩界

$$\mathbb{E}_0^N \big[\tau^{2k}((\hat{u}_N, \hat{f}_N), u_0) \big] \lesssim \delta_N^{2k}, \quad \text{any } k \geq 1.$$

步骤 2：的收敛 L^2-风险。为完成式（15）的证明，我们需要说明可以用 L^2-范数替代经验范数。定义 L^2-对应的τ-泛函如下：

$$\mathcal{T}^2((u,f),u_0)=\|u-u_0\|_{L^2}^2+\mathcal{R}(u,f).$$

取足够大的 $K\geq L_0^2$ 和 $K\geq 16$，并考虑事件

$$A_N=\big\{\mathcal{T}^2((\hat{u}_N,\hat{f}_N),u_0)\geq K^2\mu_N^2,\ \tau^2((\hat{u}_N,\hat{f}_N),u_0)\leq K\mu_N^2\big\}.$$

在该事件上，我们有 $\|\hat{u}_N\|_{H^{\alpha+1}}\leq\sqrt{K}$，并且不失一般性地取足够大的 $L_0\geq M\geq\|u_0\|_{H^{\alpha+1}}$，可得

$$\begin{aligned}\|\hat{u}_N-u_0\|_{L^2}^2&\geq\|\hat{u}_N-u_0\|_{L^2}^2-\|\hat{u}_N-u_0\|_N^2\\&=\mathcal{T}^2((\hat{u}_N,\hat{f}_N),u_0)-\tau^2((\hat{u}_N,\hat{f}_N),u_0)\\&\geq(K^2-K)\mu_N^2\geq K^2\mu_N^2/4\geq K\|\hat{u}_N-u_0\|_N^2/4,\end{aligned}$$

同时

$$\|\hat{u}_N-u_0\|_{H^{\alpha+1}}\leq L_0+\sqrt{K}\leq 2\sqrt{K}.$$

因此，有

$$A_N\subseteq\Big\{\exists v\in H^{\alpha+1}(\mathcal{O}):\|v\|_{H^{\alpha+1}}\leq 2\sqrt{K},\ \|v\|_{L^2}\geq K\mu_N/2,\ \|v\|_{L^2}\geq\sqrt{K}\|v\|_N/2\Big\}.$$

由于 $\alpha+1>d/2$，索伯列夫嵌入表明存在某个 $M_K>0$，上确界 $\sup_{w:\|w\|_{H^{\alpha+1}}\leq 2\sqrt{K}}\|w\|_\infty\leq M_K\lesssim\sqrt{K}$。利用引理 D.1 以及 $R=K\mu_N/2$ 和 $M=M_K\asymp\sqrt{K}$（注意到$\sqrt{K}/2\geq 2$），可知存在常数 $C>0$,

$$\mathbb{P}_0^N(A_N)\leq 2\exp(-N\mu_N^2K/C),\quad K>0\ \text{large enough}.$$

将此结果与经验范数的集中结果（55）相结合，我们得出，对于某个 $C>0$ 以及所有 $K\geq K_0$ 足够大的值,

$$\begin{aligned}\mathbb{P}_0^N\big(\mathcal{T}^2((\hat{u}_N,\hat{f}_N),u_0)\geq K^2\mu_N^2\big)&\leq\mathbb{P}_0^N\big(A_N\big)+\mathbb{P}_0^N\big(\tau^2((\hat{u}_N,\hat{f}_N),u_0)\geq K\mu_N^2\big)\\&\leq C\exp(-N\mu_N^2K/C).\end{aligned}$$

为得到相应的矩界，为简便起见，记 $Z=\mathcal{T}^2((\hat{u}_N,\hat{f}_N)\ u_0)$。由前面的显示不等式可知,

$$\mathbb{P}_0^N(Z^k\geq t)=\mathbb{P}_0^N(Z\geq t^{1/k})\lesssim\exp\big(-N\mu_Nt^{1/2k}/C\big),\qquad\forall k\geq 1,\ \forall t\geq(K_0\mu_N)^{2k}.$$

利用引理 E.2，取 $\gamma=1/(2k)$、$L=N\mu_N/C$和 $a=(K_0\mu_N)^{2k}$（注意，$La^\gamma=(N\mu_N/C)\cdot(K_0\mu_N)\simeq K_0N\mu_N^2\geq 1$，因为 $K_0\geq 1$），我们推得

$$\begin{aligned}\mathbb{E}_0^N\big[Z^k\big]&\lesssim(K_0\mu_N)^{2k}+\int_{(K_0\mu_N)^{2k}}^\infty\exp\big(-N\mu_Nt^{1/2k}/C\big)\,dt\\&\lesssim\mu_N^{2k}+\frac{\mu_N^{2k(1-\frac{1}{2k})}}{N\mu_N}\exp\big(-K_0N\mu_N^2/C\big)\\&\leq\mu_N^{2k}+\frac{\mu_N^{2k-2}}{N}=O(\mu_N^{2k}),\end{aligned}$$

其中最后一步用到了 $\mu_N=o(\sqrt{N})$。 —

步骤 3：逆问题的收敛速率。利用 $\mathcal{R}$ 和 $\mathcal{T}^2$的定义，前述界意味着，对于任意 $k \geq 1$，都存在 $C_k > 0$，使得

$$\mathbb{E}_0^N[\|\hat{f}_N\|_{H^\alpha}^{2k}] + \mathbb{E}_0^N[\|\hat{u}_N\|_{H^{\alpha+1}}^{2k}] \leq C_k, \quad \forall N \geq 1, \tag{56}$$

以及

$$\mathbb{E}_0^N[\|\hat{u}_N - u_0\|_{L^2}^{2k}] \leq C_k \mu_N^{2k}, \quad \forall N \geq 1. \tag{57}$$

由标准的索博列夫插值论证可得

$$\begin{aligned} \|\hat{u}_N - u_0\|_{H^2} &\lesssim \|\hat{u}_N - u_0\|_{L^2}^{\frac{\alpha-1}{\alpha+1}} \|\hat{u}_N - u_0\|_{H^{\alpha+1}}^{\frac{2}{\alpha+1}} \\ &\lesssim \|\hat{u}_N - u_0\|_{L^2}^{\frac{\alpha-1}{\alpha+1}} \left(\|\hat{u}_N\|_{H^{\alpha+1}}^{\frac{2}{\alpha+1}} + \|u_0\|_{H^{\alpha+1}}^{\frac{2}{\alpha+1}} \right). \end{aligned} \tag{58}$$

接下来验证，我们可以对 $(u_1, f_1, g_1) = (u_0, f_0, g)$ 和 $(u_2, f_2, g_2) = (\hat{u}_N, f_N, L_{\hat{f}_N} \hat{u}_N)$ 使用引理 2.2。由于 $f_0 \in \mathcal{F}(\alpha, f_{\min}, R)$，我们有 $f_0 \geq f_{\min} > 0$ on $\mathcal{O}$；并且根据索伯列夫嵌入 $H^\alpha(\mathcal{O}) \hookrightarrow C^1(\mathcal{O})$（回忆 $\alpha > d/2 + 1$），还有 $\|f_0\|_{C^1(\mathcal{O})} \lesssim \|f_0\|_{H^\alpha(\mathcal{O})} \leq R$。此外，根据引理 A.1，相应的解 $u_0 = \mathcal{G}(f_0)$ 满足 $u_0 \in H^{\alpha+1}(\mathcal{O}) \subseteq C^2(\mathcal{O})$ 和 $\|u_0\|_{C^2(\mathcal{O})} \leq B$，其中某个 $B = B(\alpha, f_{\min}, R, \mathcal{O})$。最后，边界条件 $u_0|_{\partial\mathcal{O}} = 0$ 根据 $\mathcal{G}$ 的定义成立。因此，引理 2.2 得出

$$\|\hat{f}_N - f_0\|_{L^2}^2 \lesssim \|\hat{f}_N\|_{C^1}^2 \|\hat{u}_N - u_0\|_{H^2}^2 + \|L_{\hat{f}_N} \hat{u}_N - g\|_{L^2}^2 =: I + II$$

我们分别对这两项取期望并进行界定。对于第二项，根据 $\mathcal{R}$ 的定义以及从 λ_N 式（13）中选取

$$\mathbb{E}_0^N[II] \lesssim \lambda_N^{-2} \mathbb{E}_0^N[\mathcal{R}(\hat{u}_N, \hat{f}_N)] \lesssim \lambda_N^{-2} \mathbb{E}_0^N[\mathcal{T}^2((\hat{u}_N, \hat{f}_N), u_0)] \lesssim N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}}.$$

由于 $\alpha > d/2 + 1$，由索伯列夫嵌入可得 $H^\alpha(\mathcal{O}) \hookrightarrow C^1(\mathcal{O})$。由（56）可知，$\sup_{N \geq 1} \mathbb{E}_0^N \|\hat{f}_N\|_{C^1(\mathcal{O})}^{2k} < \infty$ 对任意 $k \geq 1$均成立。利用这一点、柯西–施瓦茨不等式以及（58）和（56），

$$\begin{aligned} \mathbb{E}_0^N[I] &\leq \left(\mathbb{E}_0^N \|\hat{f}_N\|_{C^1}^4\right)^{1/2} \left(\mathbb{E}_0^N \|\hat{u}_N - u_0\|_{H^2}^4\right)^{1/2} \\ &\lesssim \left(\mathbb{E}_0^N \|\hat{f}_N\|_{C^1}^4\right)^{1/2} \left(\mathbb{E}_0^N \|\hat{u}_N - u_0\|_{L^2}^{\frac{4(\alpha-1)}{\alpha+1}} \left(\|\hat{u}_N\|_{H^{\alpha+1}}^{\frac{8}{\alpha+1}} + \|u_0\|_{H^{\alpha+1}}^{\frac{8}{\alpha+1}} \right) \right) \right)^{1/2} \\ &\lesssim \left(\mathbb{E}_0^N \left[\|\hat{u}_N - u_0\|_{L^2}^{\frac{4(\alpha-1)}{\alpha+1}} \|\hat{u}_N\|_{H^{\alpha+1}}^{\frac{8}{\alpha+1}} \right] \right)^{1/2} + \|u_0\|_{H^{\alpha+1}}^{\frac{4}{\alpha+1}} \left(\mathbb{E}_0^N \|\hat{u}_N - u_0\|_{L^2}^{\frac{4(\alpha-1)}{\alpha+1}} \right)^{1/2} \\ &=: I_a + I_b. \end{aligned} \tag{59}$$

项 I_b 利用詹森不等式和（57）进行估计，

$$I_b \lesssim \mathbb{E}_0^N \left[\|\hat{u}_N - u_0\|_{L^2}^4 \right]^{\frac{\alpha-1}{2(\alpha+1)}} \lesssim \mu_N^{\frac{2(\alpha-1)}{\alpha+1}} \lesssim N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}}.$$

对于项 I_a，我们使用赫尔德不等式 $\|vw\|_{L^2} \lesssim \|v\|_{L^p} \|w\|_{L^q}$，其中

$$p = \frac{2(\alpha+1)}{\alpha-1}, \quad q = \alpha + 1, \quad \text{with } \frac{1}{p} + \frac{1}{q} = 1/2,$$

以及（56）–（57），可得

$$I_a \lesssim \left(\mathbb{E}_0^N \|\hat{u}_N - u_0\|_{L^2}^4\right)^{\frac{\alpha-1}{2(\alpha+1)}} \left(\mathbb{E}_0^N \|\hat{u}_N\|_{H^{\alpha+1}}^4\right)^{\frac{1}{\alpha+1}} \lesssim \mu_N^{\frac{2(\alpha-1)}{\alpha+1}} \lesssim N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}}.$$

将所有结果合并，可得

$$E_0^N \left[\|\hat{f}_N - f_0\|_{L^2}^2 \right] = O(N^{-\frac{2\alpha-2}{2\alpha+2+d}}),$$

由此完成了 PDE 惩罚 M 估计量收敛速率的证明。

B.1.2. 插件估计量。现在我们证明插件估计量的类似收敛保证。

步骤 *1*：收敛性：$\hat{}u_N$。我们再次应用定理 3.1，这次只考虑一个参数$\eta = u \in H^{\alpha+1}$。[形式上，我们可以取 $\Theta = \{0\}$ 使得该最小化问题与 θ 无关。] 正则化泛函定义为

$$\mathcal{R}(u) = \mu_N^2 \|u\|_{H^{\alpha+1}}^2.$$

定义泛函

$$\tau^2(\hat{u}_N, u_0) := \|\hat{u}_N - u_0\|_N^2 + \mu_N^2 \|u\|_{H^{\alpha+1}}^2,$$

并令 $u^* = u_0$。根据 $\mathcal{R}$ 的定义，有

$$\mathcal{U}(\eta^*, R) \subseteq \{u : \|u\|_{H^{\alpha+1}} \le R/\mu_N\}, \quad R > 0.$$

因此，与广义 M 估计量证明中相同的论证（53）-（54）表明定理 3.1 的假设条件得到满足，其中 $\Psi(R) := C_1 R \mu_N^{2(\alpha+1)}$，且 $\delta_N := C_2 N^{-\frac{\alpha+1}{\alpha+1+d}}$，当 C_1, $C_2 > 0$足够大时。这意味着我们有如下集中不等式

$$\mathbb{P}_0^N\Big(\tau^2(\hat{u}_N, u_0) \ge L^2 N^{-\frac{2(\alpha+1)}{2(\alpha+1)+d}}\Big) \le C_3 \exp\big(-L^2 N^{\frac{d}{2(\alpha+1)+d}}/C_3\big), \quad L \ge L_0, \tag{60}$$

其中 $L_0 > 0$ 是某个足够大的常数，并且有如下矩界

$$\mathbb{E}_0^N\big[\tau^{2k}(\hat{u}_N, u_0)\big] \lesssim \delta_N^{2k}, \quad \text{any } k \ge 1.$$

前述界也适用于 $L^2(\mathcal{O})$-范数。仿照 PDE 惩罚 M 估计量的证明，定义

$$\mathcal{T}^2(\hat{u}_N, u_0) =: \|\hat{u}_N - u_0\|_{L^2}^2 + \mu_N^2 \|u\|_{H^{\alpha+1}}^2,$$

并应用引理 D.1 可得，对于所有 $K \ge K_0$ 充分大的，

$$\mathbb{P}_0^N\big(\mathcal{T}^2(\hat{u}_N, u_0) \ge K^2 N^{-\frac{2(\alpha+1)}{2(\alpha+1)+d}}\big) \lesssim \exp(-K N^{\frac{d}{2(\alpha+1)+d}}/C),$$

$$\mathbb{E}_0^N\big[\mathcal{T}^{2k}(\hat{u}_N, u_0)\big] \lesssim \mu_N^{2k} = N^{-\frac{2k(\alpha+1)}{2(\alpha+1)+d}},$$

这尤其意味着，存在某个 $C_k > 0$,

$$\mathbb{E}_0^N\big[\|\hat{u}_N - u_0\|_{L^2}^{2k}\big] \le C_k \mu_N^{2k}, \quad \text{and} \quad \mathbb{E}_0^N\big[\|\hat{u}_N\|_{H^{\alpha+1}}^{2k}\big] \le C_k, \quad \forall N \ge 1. \tag{61}$$

$\hat{}$ $\hat{}$步骤 *2*： 的收敛。根据 f_N 的定义，有

$$\begin{aligned}
\|\nabla \cdot (\hat{f}_N \nabla \hat{u}_N) - g\|_{L^2}^2 + \nu_N^2 \|\hat{f}_N\|_{H^\alpha}^2 &\le \|\nabla \cdot (f_0 \nabla \hat{u}_N) - g\|_{L^2}^2 + \nu_N^2 \|f_0\|_{H^\alpha}^2 \\
&\le \|\nabla \cdot (f_0 \nabla (\hat{u}_N - u_0))\|_{L^2}^2 + \nu_N^2 \|f_0\|_{H^\alpha}^2 \\
&\lesssim \|f_0\|_{C^1}^2 \|\hat{u}_N - u_0\|_{H^2}^2 + \nu_N^2 \|f_0\|_{H^\alpha}^2 \\
&\lesssim \|\hat{u}_N - u_0\|_{H^2}^2 + \nu_N^2.
\end{aligned}$$

注意，右端始终有限，因为根据假设，$\hat{}u_N \in H^{\alpha+1}$ 和 $\alpha \ge 1$。前述不等式尤其意味着

$$\|\hat{f}_N\|_{H^\alpha}^2 \lesssim 1 + \nu_N^{-2} \|\hat{u}_N - u_0\|_{H^2}^2.$$

与 PDE 惩罚估计量的证明一样，引理 2.2 中的广义稳定性估计的假设条件对 $(u_1, f_1, g_1) = (u_0, f_0, g)$ 成立，因为 $f_0 \ge f_{\min}$ 和 $u_0 \in$

$C^2(\overline{\mathcal{O}})$ 其中 $\|f_0\|_{C^1} \vee \|u_0\|_{C^2} \le B$。利用引理 2.2、索伯列夫嵌入 $H^\alpha(\mathcal{O}) \hookrightarrow C^1(\mathcal{O})$（因为 $\alpha > d/2+1$），以及柯西 - 施瓦茨不等式，我们得到

$$\begin{aligned}\mathbb{E}_0^N\left[\|\hat{f}_N - f_0\|_{L^2}^2\right] &\lesssim \mathbb{E}_0^N\left[\|\hat{f}_N\|_{C^1}^2\|\hat{u}_N - u_0\|_{H^2}^2 + \|\nabla\cdot(\hat{f}_N\nabla\hat{u}_N) - g\|_{L^2}^2\right]\\ &\lesssim \mathbb{E}_0^N\left[(1+\|\hat{f}_N\|_{C^1}^2)\|\hat{u}_N - u_0\|_{H^2}^2\right]\\ &\lesssim \mathbb{E}_0^N\left[(1+\|\hat{f}_N\|_{H^\alpha}^2)\|\hat{u}_N - u_0\|_{H^2}^2\right]\\ &\lesssim \left(\mathbb{E}_0^N\left[1+\eta_N^{-4}\|\hat{u}_N - u_0\|_{H^2}^4\right]\right)^{1/2}\left(\mathbb{E}_0^N\left[\|\hat{u}_N - u_0\|_{H^2}^4\right]\right)^{1/2}. \end{aligned} \tag{62}$$

使用紧随（59）之后的相同插值论证，以及（61）中的矩界，我们现在可以估计

$$\begin{aligned}&\left(\mathbb{E}_0^N\left[\|\hat{u}_N - u_0\|_{H^2}^4\right]\right)^{1/2}\\ &\lesssim \left(\mathbb{E}_0^N\left[\|\hat{u}_N - u_0\|_{L^2}^{\frac{4(\alpha-1)}{\alpha+1}}\|\hat{u}_N\|_{H^{\alpha+1}}^{\frac{8}{\alpha+1}}\right]\right)^{1/2} + \|u_0\|_{H^{\alpha+1}}^{\frac{4}{\alpha+1}}\left(\mathbb{E}_0^N\|\hat{u}_N - u_0\|_{L^2}^{\frac{4(\alpha-1)}{\alpha+1}}\right)^{1/2}\\ &\lesssim N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}} = \nu_N^2.\end{aligned}$$

将上述结果代入（62）右端的两个因子，可得

$$\mathbb{E}_0^N\left[\|\hat{f}_N - f_0\|_{L^2}^2\right] \lesssim N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}},$$

从而完成定理的证明。 □

B.2. 定理2.4的证明。

步骤 *1*：$\hat{}$ 的收敛。令 $\eta_0 = S(u_0) \in h^{\alpha+1}$ 表示满足以下条件的 u_0的框架系数

$$\|u_0\|_{H^{\alpha+1}(\mathcal{O})} \simeq \|\eta_0\|_{h^{\alpha+1}},$$

并将 $\eta_{0,\ J} \in \mathbb{R}^{p_J}$ 在分辨率层级 J处的截断记为。对于

$$u_{0,J} := \sum_{l\le J, k\le N_l}(\eta_{0,J})_{lk}\phi_{lk} \in L_J^2(\mathcal{O}),$$

C^u正则小波框架的标准逼近性质给出

$$\|u_0 - u_{0,J}\|_{L^2} \lesssim 2^{-J(\alpha+1)}\|u_0\|_{H^{\alpha+1}}. \tag{63}$$

我们将对高维参数 3.1 应用定理 $\eta \in H = \mathbb{R}^{p_J}$，其参数化为

$$u_\eta = \sum_{l\le J, k\le N_l}\eta_{lk}\phi_{lk},$$

其中 $\eta^* = \eta_{0,\ J}$且惩罚项为 $\mathcal{R}(\eta) = \mu_N^2\|\eta\|_{h_J^{\alpha+1}}^2$。回忆定义

$$\tau^2(\eta, u_0) = \|u_0 - u_\eta\|_N^2 + \mu_N^2\|\eta\|_{h_J^{\alpha+1}}^2.$$

与无限维情形类似，局部化函数类包含在有界索об列夫球中。对于某个常数 $c>0$，我们有

$$\mathcal{U}(\eta^*, R) \subseteq \{u : \|u\|_{H^{\alpha+1}} \le cR/\mu_N\},\ \ R>0.$$

因此，再次利用定理 4.10.3 [60], 中的熵界，定理 3.1 的假设条件在 Ψ（R）下得到满足：$= C_1 R\mu_N^{-\frac{2}{d}(\alpha+1)}$，并且在 δ_N：$= C_2 N^{-\frac{\alpha+1}{2\alpha+1-()+d}}$下，对于某个$C_1$，$C_2>0$。根据定理 3.1，对于所有 $R \geq \delta_N$ 以及某个 $C>0$,

$$\mathbb{P}_{f_0}^N\Big(\tau^2(\hat{\eta}_N, u_0) \geq 2\big(\|u_0 - u_{0,J}\|_N^2 + \mu_N^2 \|\eta_{0,J}\|_{h_J^{\alpha+1}}^2 + R^2\big)\Big) \leq C\exp(-NR^2/C). \tag{64}$$

对中心化随机变量应用伯恩斯坦不等式（回忆一下，vol（$\mathcal{O}$） $=1$）

$$\begin{aligned} Z_i &= (u_0 - u_{0,J})^2(X_i) - \|u_0 - u_{0,J}\|_{L^2}^2, \\ |Z_i| &\leq \|u_0 - u_{0,J}\|_\infty^2 \lesssim 2^{-2J(\alpha+1-d/2)}\|u_0\|_{H^{\alpha+1}}^2 \lesssim \mu_N^2 2^{Jd}, \\ \mathbb{E}[Z_i^2] &\lesssim \|u_0 - u_{0,J}\|_{L^2}^2 \|u_0 - u_{0,J}\|_\infty^2 \lesssim \mu_N^4 2^{Jd}, \end{aligned}$$

可得，对于任意 $L \geq 1$，以及某个 C，C'，$C''>0$,

$$\begin{aligned} \mathbb{P}_0^N\big(\|u_0 - u_{0,J}\|_N^2 - \|u_0 - u_{0,J}\|_{L^2}^2 \geq L^2\mu_N^2\big) &\leq 2\exp\Big(-\frac{NL^4\mu_N^4}{C(1+L^2)2^{Jd}\mu_N^4}\Big) \\ &\lesssim \exp(-NL^2/2^{Jd}C') \\ &\lesssim \exp(-\mu_N^{-2}L^2/C'') \\ &\lesssim \exp(-N\mu_N^2 L^2/C''). \end{aligned}$$

这里，在最后一步中，我们还利用了 $\mu_N = N^{-\frac{\alpha+1}{2\alpha+1-()+d}} \leq N^{-1/4}$，这是由我们的假设$\alpha > d/2+1$得出的。

将其与（64）结合，并注意到$\|\eta_{0,\ J}\|_{h_J^{\alpha+1}} \lesssim \|u_0\|_{H^{\alpha+1}}$有一个常数上界，可知对所有 $K \geq K_0$ 以及某个 $C>0$,

$$\mathbb{P}_{f_0}^N\big(\tau^2(\hat{\eta}_N, u_0) \geq K^2\mu_N^2\big) \leq C\exp(-NK^2\mu_N^2/C).$$

B按照定理 2.1的证明中的论证方式，特别可得矩界

$$\mathbb{E}_0^N\big[\|\hat{u}_N - u_0\|_{L^2}^{2k}\big] \leq C_k\mu_N^{2k}, \quad \text{and} \quad \mathbb{E}_0^N\big[\|\hat{u}_N\|_{H^{\alpha+1}}^{2k}\big] \leq C_k, \quad \forall N \geq 1. \tag{65}$$

现在定义 τ^2-泛函的 L^2-对应项,

$$\mathcal{T}^2(\eta, u_0) = \|u_0 - u_\eta\|_{L^2}^2 + \mu_N^2\|\eta\|_{h^{\alpha+1}}^2.$$

如定理 2.1 的证明中步骤 2 所示，同样应用引理 D.1可得，对于所有$K \geq K_0 > 0$ 充分大时，以及某个 $C>0$,

$$\mathbb{P}_0^N\big(\mathcal{T}^2((\hat{u}_N, \hat{f}_N), u_0) \geq K^2\mu_N^2\big) \leq C\exp(-N\mu_N^2 K/C),$$

并且，对于某个 $C_k>0$,

$$\mathbb{E}_0^N\big[\|\hat{u}_N - u_0\|_{L^2}^{2k}\big] \leq C_k\mu_N^{2k}, \quad \text{and} \quad \mathbb{E}_0^N\big[\|\hat{u}_N\|_{H^{\alpha+1}}^{2k}\big] \leq C_k, \quad \forall N \geq 1. \tag{66}$$

ˆ步骤 2： 的收敛。令 $\theta_0 = S(f_0) \in h^\alpha$ 和 $\theta_{0,\ J} \in \mathbb{R}^{p_J}$ 表示 f_0 的框架系数（及其截断对应项），并令

$$f_{0,J} := \sum_{l \leq J, k \leq N_l} (\theta_{0,J})_{lk}\phi_{lk} \in L_J^2(\mathcal{O}).$$

于是我们有

$$\|f_0\|_{H^\alpha(\mathcal{O})} \simeq \|\theta_0\|_{h^\alpha}, \quad \|f_0 - f_{0,J}\|_{L^2}^2 \lesssim 2^{-2J\alpha}\|f_0\|_{H^\alpha}^2, \quad \|f_0 - f_{0,J}\|_{H^1}^2 \leq 2^{-2J(\alpha-1)}\|f_0\|_{H^\alpha}^2.$$

ˆ根据 θ_N 的定义，我们有“基本不等式”

$$\|\Psi\hat{\theta}_N-\gamma\|_{\mathbb{R}^{p_J}}^2+\nu_N^2\|\hat{\theta}_N\|_{h^\alpha}^2\le\|\Psi\theta_{0,J}-\gamma\|_{\mathbb{R}^{p_J}}^2+\nu_N^2\|\theta_{0,J}\|_{h^\alpha}^2. \tag{67}$$

右侧第二项的上界为

$$\nu_N^2\|\theta_{0,J}\|_{h^\alpha}^2\le\nu_N^2\|\theta_0\|_{h^\alpha}^2\simeq\nu_N^2\|f_0\|_{H^\alpha}^2\lesssim N^{-\frac{2(\alpha-1)}{2(\alpha+1)+d}}.$$

对于第一项，记截至层级 J 的系数选择算子为

$$S^{(J)}:H^\alpha(\mathcal{O})\to\mathbb{R}^{p_J},$$

并引入相应的“伪投影算子”

$$P^{(J)}:L^2(\mathcal{O})\to L_J^2(\mathcal{O}),\quad P^{(J)}(f)=\sum_{l\le J,k}S(f)_{lk}\phi_{lk}.$$

于是我们估计

$$\begin{aligned}\|\Psi\theta_{0,J}-\gamma\|_{\mathbb{R}^{p_J}}&=\|S^{(J)}(\nabla\cdot(f_{0,J}\nabla\hat{u}_N)-g)\|_{\mathbb{R}^{p_J}}\\&\le\|S(\nabla\cdot(f_{0,J}\nabla\hat{u}_N)-g)\|_{\ell^2(\mathbb{N})}\\&\simeq\|\nabla\cdot(f_{0,J}\nabla\hat{u}_N)-g\|_{L^2(\mathcal{O})}\end{aligned}$$

利用 $L_{f_0}u_0=g$、 $C^2(\mathcal{O})$-界引理 A.1，并回想 $2^{-J(\alpha-1)}\simeq\nu_N=_{\alpha-1}N^{-\frac{}{2(\alpha+1)+d}}$，我们进一步将其估计为

$$\begin{aligned}&\|\nabla\cdot(f_{0,J}\nabla\hat{u}_N)-\nabla\cdot(f_0\nabla u_0)\|_{L^2(\mathcal{O})}\\&\quad\le\|\nabla\cdot(f_{0,J}\nabla(\hat{u}_N-u_0))\|_{L^2}+\|\nabla\cdot((f_{0,J}-f_0)\nabla u_0)\|_{L^2}\\&\quad\lesssim\|f_{0,J}\|_{C^1}\|\hat{u}_N-u_0\|_{H^2}+\|f_{0,J}-f_0\|_{H^1}\|u_0\|_{C^2}\\&\quad\lesssim\|\hat{u}_N-u_0\|_{H^2}+2^{-J(\alpha-1)}\|f_0\|_{H^\alpha}\\&\quad\lesssim\|\hat{u}_N-u_0\|_{H^2}+\nu_N.\end{aligned}$$

Usin结合前述步骤，以及 $H^{\alpha-1}$ 是满足该不等式的乘法代数这一事实 ty

$$\|vw\|_{H^{\alpha-1}}\lesssim\|v\|_{H^{\alpha-1}}\|w\|_{H^{\alpha-1}}\quad\forall v,w\in H^{\alpha-1}.$$

以及 Sobole夫插值不等式，我们现在得出结论：

$$\begin{aligned}\|L_{\hat{f}_N}\hat{u}_N-g\|_{L^2}&\lesssim\|S^{(J)}(L_{\hat{f}_N}\hat{u}_N-g)\|_{\mathbb{R}^{p_J}}+2^{-J(\alpha-1)}\|L_{\hat{f}}\hat{u}_N-g\|_{H^{\alpha-1}}\\&=\|\Psi\hat{\theta}_N-\gamma\|_{\mathbb{R}^{p_J}}+2^{-J(\alpha-1)}\|L_{\hat{f}}\hat{u}_N-g\|_{H^{\alpha-1}}\\&\lesssim\|\Psi\hat{\theta}_N-\gamma\|_{\mathbb{R}^{p_J}}+\nu_N\big(\|\hat{f}_N\|_{H^\alpha}\|\hat{u}_N\|_{H^{\alpha+1}}+\|g\|_{H^{\alpha-1}}\big)\\&\lesssim\|\hat{u}_N-u_0\|_{H^2}+\nu_N\big(1+\|\hat{f}_N\|_{H^\alpha}\|\hat{u}_N\|_{H^{\alpha+1}}\big)\\&\lesssim\big(1+\|\hat{u}_N\|_{H^{\alpha+1}}+\|\hat{f}_N\|_{H^\alpha}\|\hat{u}_N\|_{H^{\alpha+1}}\big)\Big(\|\hat{u}_N-u_0\|_{L^2}^{\frac{\alpha-1}{\alpha+1}}+\nu_N\Big).\end{aligned}$$

ˆ为了进一步估计 $\|f_N\|_{H^\alpha}$，注意到上述论证和“基本不等式”（67）还意味着

$$\nu_N^2\|\hat{f}_N\|_{H^\alpha}^2\le\|\nabla\cdot(f_{0,J}\nabla\hat{u}_N)-g\|_{L^2}^2+\nu_N^2\|f_{0,J}\|_{H^\alpha}^2\lesssim\|\hat{u}_N-u_0\|_{H^2}^2+\nu_N^2.$$

应用广义稳定性引理2.2 ，取三元组f_0， u_0， g）ˆ(f， ˆu_N， Lˆf_Nˆu_N)，这是合理的，因为 f_0， u_0， g满足所需的“椭圆性”条件。综合以上结果，我们得到

$$\begin{aligned}\|\hat{f}_N - f_0\|_{L^2} &\lesssim \|\hat{f}_N\|_{C^1}\|\hat{u}_N - u_0\|_{H^2} + \|L_{\hat{f}_N}\hat{u}_N - g\|_{L^2}\\ &\lesssim \|\hat{f}_N\|_{H^\alpha}\|\hat{u}_N - u_0\|_{H^2}\\ &\quad + \big(1+\|\hat{u}_N\|_{H^{\alpha+1}}\big)\big(1+\nu_N^{-1}\|\hat{u}_N - u_0\|_{H^2}\big)\Big(\|\hat{u}_N - u_0\|_{L^2}^{\frac{\alpha-1}{\alpha+1}} + \nu_N\Big)\\ &=: I + II.\end{aligned}$$

然后，利用矩界（66），我们可以对每一项 I和 II分别进行论证，类似于在式（61 ）之后关于第B.1.2 节无限维情形证明中步骤 2 的论证，从而完成证明并得到

$$\mathbb{E}_0^N\big[\|\hat{f}_N - f_0\|_{L^2}^2\big] \lesssim \nu_N^2,$$

由此完成定理2.4的证明。 □

ˆ**B.3. 定理2.6的证明**。令 f_N 表示定理2.4中的插件估计量，该估计量可在多项式时间内计算。由定理2.4 的证明可知，当 $L \geq 1$充分大时，有

$$\mathbb{P}_0^N\big(\|\hat{f}_N\|_{H^\alpha} \leq L\big) \xrightarrow{N\to\infty} 1.$$

ˆ此外，由于 $\alpha > d/2+1$，我们有，在事件 $\{\|f_N\|_{H^\alpha} \leq L\}$ 上，Sobolev 嵌入 $H^{\alpha-1} \hookrightarrow L^\infty$ 和插值可推出

$$\|\hat{f}_N - f_0\|_\infty \lesssim \|\hat{f}_N - f_0\|_{H^{\alpha-1}} \lesssim \|\hat{f}_N - f_0\|_{L^2}^{\frac{1}{\alpha}}\|\hat{f}_N - f_0\|_{H^\alpha}^{\frac{\alpha-1}{\alpha}} \leq L\|\hat{f}_N - f_0\|_{L^2}^{\frac{1}{\alpha}}.$$

根据定理2.4，可得

$$\|\hat{f}_N - f_0\|_\infty \xrightarrow{N\to\infty} 0 \text{ in } \mathbb{P}_0^N\text{-probability}.$$

特别地，令 $\rho :=$ 下确界$_x f_0(x) - f_{min} > 0$，并取足够大的 L，则事件

$$A_N := \{\inf_x \hat{f}_N(x) \geq f_{min} + \rho/2\} \cap \{\|\hat{f}_N\|_{H^\alpha} \leq L\} \cap \{\|\hat{f}_N - f_0\|_{L^2} \leq LN^{-\frac{\alpha-1}{2(\alpha+1)+d}}\}.$$

其概率趋于1， $\mathbb{P}_0^N(A_N) \to 1$，当 $N \to \infty$时。在关于链接函数 Φ 的假设下，其逆函数 Φ^{-1} 限制在区间 $[f_{min}+\rho/2, \infty)$上是一致利普希茨连续的，且各阶导数均一致有界，因此我们有如下索伯列夫估计

$$\sup_{\|f\|_{H^\alpha}\leq L,\ f\geq f_{min}+\rho/2} \|\Phi^{-1}\circ f\|_{H^\alpha} \leq M < \infty$$

其中某个有限常数 $M = M(\alpha, \Phi, f_{min}, \rho, L, \mathcal{O})$。例如，参见引理 [47，30]，其证明用到了法阿·迪·布鲁诺公式以及加利亚多–尼伦伯格插值不等式。令 ˆ $F_N :$ $= \Phi^{-1}\circ$ ˆf_N，并令 $F_0 :$ $= \Phi^{-1}\circ f_0$（在事件 A_N 之外，我们可以用可测的方式任意设定 ˆ ˆF_N）。那么在事件 A_N 上，有 $\|F_N\|_{H^\alpha} \leq M$，此外还有

$$\|\hat{F}_N - F_0\|_{L^2} \lesssim \|\hat{f}_N - f_0\|_{L^2} \leq LN^{-\frac{\alpha-1}{2(\alpha+1)+d}}.$$

最后，我们定义 θ_{init} 为$L^2(\mathcal{O})$在$(e_k : k \geq 1)$-基下的前 D 个系数，

$$\theta_{init,k} = \langle \hat{F}_N, e_k\rangle_{L^2} \quad \forall 1 \leq k \leq D, \quad \theta_{init} \in \mathbb{R}^D.$$

根据正交归一基$(e_k : k \geq 1)$的Parseval恒等式，我们于是有

$$\|\theta_{init} - \theta_{0,D}\|_{\mathbb{R}^D} = \Big(\sum_{k=1}^{D} \big|\langle \hat{F}_N - F_0, e_k\rangle_{L^2}\big|^2\Big)^{1/2}$$
$$\leq \|\hat{F}_N - F_0\|_{L^2} \lesssim L N^{-\frac{\alpha-1}{2(\alpha+1)+d}}.$$

对于任意 $\alpha \geq 12$（特别是对于假设中的任意 $\alpha \geq 22$），后一个速率的阶为$o($ $D^{-8/d}/\log D)$，因为 $D^{-8/d} = N^{-\frac{8}{2\alpha+d}}$且

$$\frac{\alpha-1}{2(\alpha+1)+d} \geq \frac{\alpha-3}{2\alpha+d} > \frac{8}{2\alpha+d},$$

至此，证明完成。

□

B.4. 定理2.8的证明。

步骤 *1*。 我们首先使用定理 3.1 证明经验范数下的（30），其中 $\eta = (\beta, u) \in H$，参数化为 $\eta = (\beta, u) \mapsto u$，其中

$$H = \big\{(\beta, u) : \beta \in [\beta_{\min}, \beta_{\max}],\ u \in H^{\beta}(\mathcal{O})\big\}.$$

令 $\eta^* = (\beta_0, u_0)$，并考虑 $R > 0$ 的局部化函数类

$$\mathcal{U}^*(\eta^*, R) = \{u : \|\hat{u} - u_0\|_N^2 + \mu_{N,\beta}^2(\|u\|_{H^\beta}^2 + A^2) \leq R^2 \ \text{ for some } \ \beta \in [\beta_{\min}, \beta_{\max}]\}.$$

对于满足 $\mu_{N,\beta}^2 > R^2/A^2$ 的任意 β，不可能有取值为 β 的配对 $(\beta, u) \in H$ 作出贡献。因此

$$\beta(R) := \inf\{\beta : \mu_{N,\beta} \leq R/A\},$$

$\|\cdot\|_\infty$ 的 $\mathcal{U}^*$ 熵(η^*, R)，它是经验熵的上界，并且满足某个 $C > 0$

$$H(\mathcal{U}^*(\eta^*, R), \|\cdot\|_\infty, \rho) \leq C \sum_{\beta \geq \beta(R)} \Big(\frac{R}{\rho \mu_{N,\beta}}\Big)^{\frac{d}{\beta}} \lesssim \Big(\frac{R}{\rho \mu_{N,\beta(R)}}\Big)^{\frac{d}{\beta(R)}}.$$

这里，我们用到了 $H^\beta(\mathcal{O})$ $\beta > d/2$ 的标准熵界（见 [60] 中的定理 4.10.3）、求和项数量固定且有限这一事实，以及

$$\mu_{N,\beta}^{d/\beta} = N^{\frac{d}{2\beta+d}} \leq N^{\frac{d}{2\beta(R)+d}}, \quad \text{for} \quad \beta \geq \beta(R).$$

注意到 $\beta_{min} > d/2$，于是我们可以将 $\mathcal{U}^*(\eta^*, R)$ 的度量熵积分上界为

$$\int_0^R H^{1/2}(\mathcal{U}^*(\eta^*, R), \|\cdot\|_\infty, \rho) d\rho \lesssim \int_0^R \Big(\frac{R}{\rho \mu_{N,\beta(R)}}\Big)^{\frac{d}{2\beta(R)}} d\rho$$
$$\lesssim R \mu_{N,\beta(R)}^{-\frac{d}{2\beta(R)}}$$
$$= R\sqrt{N}\mu_{N,\beta(R)} \leq R^2\sqrt{N}/A.$$

选取常数 $A > 0$ 足够大（仅依赖于 $C > 0$ 以及熵条件中的常数（39）），即可得出（39）对所有 $R > 0$均成立。此后我们固定这样一个 $A > 0$ 取值。

应用定理3.1，可得存在某个 $C_1>0$，使得对所有 $R>0$

$$\mathbb{P}_{f_0}^N\Big(\|\hat{u}_N-u_0\|_N^2+\mu_{N,\hat{\beta}}^2\|\hat{u}_N\|_{H^{\hat{\beta}}}^2\geq 2\mu_{N,\beta_0}^2(\|u_0\|_{H^{\beta_0}}^2+A^2)+2R^2\Big)$$
$$\lesssim C_1\exp\big(-\frac{NR^2}{C_1}\big), \tag{68}$$

以及

$$\mathbb{E}_{f_0}^N\big[\|\hat{u}_N-u_0\|_N^2+\mu_{N,\hat{\beta}}^2\|\hat{u}_N\|_{H^{\hat{\beta}}}^2\big]\lesssim\mu_{N,\beta_0}^2(\|u_0\|_{H^{\beta_0}}^2+A^2)+\frac{1}{N}\simeq\mu_{N,\beta_0}^2.$$

利用上一步推导出的集中不等式，$R=\mu_{N,\ \beta_0}$ 可知以下事件的概率趋于1，

$$A_N:=\Big\{\|\hat{u}_N-u_0\|_N^2+\mu_{N,\hat{\beta}}^2\|\hat{u}_N\|_{H^{\hat{\beta}}}^2\leq 2\mu_{N,\beta_0}^2(\|u_0\|_{H^{\beta_0}}^2+A^2+1)\Big\},\quad \mathbb{P}_{f_0}^N(A_N)\xrightarrow{N\to\infty}1.$$

因此，在证明的剩余部分中，我们可以将讨论限制在此事件上。

步骤 *2*：的收敛 L^2风险。 现在我们想要证明 L^2-风险具有相应的收敛速率。固定某个 $\beta\in[\beta_{min}:\ \ \beta_{max}]$ 并假设事件 $A_N\cap\{\beta=\hat{}\beta\}$发生。对于某些常数 $B,\ B'>0$，有

$$\|\hat{u}_N\|_\infty\leq B\|\hat{u}_N\|_{H^{\hat{\beta}}}\leq B'\mu_{N,\beta_0}/\mu_{N,\hat{\beta}}=:M_N,$$

我们希望对函数类应用引理D.1 。

$$\mathcal{U}=\{u-u_0:\|u\|_{H^\beta}\leq M_N/B\},$$

其熵界为

$$H(\mathcal{U},\|\cdot\|_\infty,\rho)\lesssim\big(\frac{M_N}{\rho}\big)^{\frac{d}{\beta}}.$$

i) 情形 $\beta\geq\beta_0$。 在此情形下，我们有 $M_N\to\infty$，使得

$$\sup_{v\in\mathcal{U}}\|v\|_\infty\leq 2M_N$$

对所有 N 充分大的情形。对于这样的 N，引理 D.1 中的熵条件变为

$$N\delta_N^2\gtrsim M_N^2\Big(\frac{M_N}{\delta_N}\Big)^{\frac{d}{\beta}}\quad\Longleftrightarrow\quad\delta_N\gtrsim N^{-\frac{\beta}{2\beta+d}}M_N\simeq\mu_{N,\beta_0}.$$

然后，我们可以应用引理 D.1，取 $M=2M_N$和 $\delta_N\simeq\mu_{N,\ \beta_0}$，从而得到，对于常数$C,\ C',\ C''>0$,

$$\begin{aligned}&\mathbb{P}_{f_0}^N\big(A_N\cap\{\hat{\beta}=\beta\},\ \|\hat{u}_N-u_0\|_{L^2}\geq 2\|\hat{u}_N-u_0\|_N,\ \|\hat{u}_N-u_0\|_{L^2}\geq C\mu_{N,\beta_0}\big)\\&\leq\mathbb{P}_{f_0}^N\big(\exists v\in\mathcal{U}:\ \|v\|_{L^2}\geq 2\|v\|_N,\ \|v\|_{L^2}\geq C\mu_{N,\beta_0}\big)\\&\lesssim\exp\Big(-\frac{N\mu_{N,\beta_0}^2}{C'M_N^2}\Big)\xrightarrow{N\to\infty}0.\end{aligned}$$

对于充分大的常数 $L>0$，且由于 $\mathbb{P}_{f_0}^N(A_N)\to 1$，我们得到

$$\mathbb{P}_{f_0}^N\big(\{\hat{\beta}=\beta\}\cap\{\|\hat{u}_N-u_0\|_{L^2}\geq L\mu_{N,\beta_0}\}\big)\xrightarrow{N\to\infty}0.$$

ii）情形 $\beta<\beta_0$。在这种情况下，我们有 $M_N\to 0$，因此 L^∞ 界的阶为 1，

$$\sup_{v\in\mathcal{U}}\|v\|_\infty\leq M_N+\|u_0\|_\infty\lesssim 1.$$

因此，引理中的熵条件D.1 变为

$$N\delta_N^2 \gtrsim \delta_N^{-\frac{d}{\beta}} \quad \Longleftrightarrow \quad \delta_N \gtrsim \mu_{N,\beta}.$$

然后，我们可以应用引理 D.1，取某个固定的 $M<\infty$（其阶为常数），并取 $\delta_N \simeq \mu_{N,\beta}$，从而得到：对于常数 $C, C', C'' > 0$,

$$\begin{aligned}
&\mathbb{P}_{f_0}^N\big(A_N \cap \{\hat\beta = \beta\},\ \|\hat u_N - u_0\|_{L^2} \geq 2\|\hat u_N - u_0\|_N,\ \|\hat u_N - u_0\|_{L^2} \geq C\mu_{N,\beta_0}\big) \\
&\qquad \leq \mathbb{P}_{f_0}^N\big(\exists v \in \mathcal{U}:\ \|v\|_{L^2} \geq 2\|v\|_N,\ \|v\|_{L^2} \geq C\mu_{N,\beta}\big) \\
&\qquad \xrightarrow{N\to\infty} 0.
\end{aligned}$$

取足够大的常数 $L>0$，且由于 $\mathbb{P}_{f_0}^N(A_N) \to 1$，我们得到

$$\mathbb{P}_{f_0}^N\big(\{\hat\beta = \beta\} \cap \{\|\hat u_N - u_0\|_{L^2} \geq L\mu_{N,\beta}\}\big) \xrightarrow{N\to\infty} 0.$$

注意，速率 $\mu_{N,\beta}$ 慢于目标速率 μ_{N,β_0}；我们将在下一步看到，事件 $\hat{} \{\beta < \beta_0\}$ 的概率趋于零。

$\hat{}$ $\hat{}$步骤 *3*：的行为. 固定 $\beta < \beta_0$，并假设 $\beta = \beta_\circ$ 在 A_N 上，我们有

$$\|\hat u_N\|_{L^2} \leq \|\hat u_N\|_{H^\beta} \lesssim \mu_{N,\beta_0}/\mu_{N,\beta} \to 0.$$

此外，根据步骤 2 ii)，以趋于 1 的概率，有 $\|\hat{} u_N - u_0\|_{L^2} \leq L\mu_{N,\beta_\circ}$ 在此情形成立的事件上，

$$\|\hat u_N - u_0\|_{L^2} \geq \|u_0\|_{L^2} - \|\hat u_N\|_{L^2} - O(\mu_{N,\beta_0}/\mu_{N,\beta}) \xrightarrow{N\to\infty} \|u_0\|_{L^2} > 0,$$

这与 $\|\hat{} u_N - u_0\|_{L^2} = O(\mu_{N,\beta}) \to 0$ 矛盾。因此我们注意到 $P_{f_0}^N(\hat{} \beta = \beta) \to 0$，于是

$$\mathbb{P}_0^N(\hat\beta \geq \beta_0) \xrightarrow{N\to\infty} 1.$$

步骤 *4*：逆问题。回忆插件估计量

$$\hat f_N \in \operatorname*{arg\,min}_{f \in H^{\alpha_{\min}}} J(f), \quad J(f) = \|L_f \hat u_N - g\|_{L^2}^2 + \nu_{N,\hat\beta}^2 \|f\|_{H^{\alpha_{\min}}}^2, \tag{69}$$

$^{\beta-2}$其中 $\nu_{N,\ \beta} = N^{-\ 2\beta+d}$。我们在该事件上

$$B_N := A_N \cap \{\hat\beta \geq \beta_0\} \cap \{\|\hat u_N - u_0\|_{L^2} \leq L\mu_{N,\beta_0}\},$$

对某个充分大的 $L>0$，其概率趋于1，这是由前述证明步骤得到的。根据 J 的定义并利用 $L_{f_0} u_0 = g$，可得

$$\begin{aligned}
J(f_0) &= \|L_{f_0}(\hat u_N - u_0)\|_{L^2}^2 + \nu_{N,\hat\beta}^2 \|f\|_{H^{\alpha_{\min}}}^2 \\
&\leq \|f_0\|_{C^1}^2 \|\hat u_N - u_0\|_{H^2}^2 + \nu_{N,\beta_0}^2 \|f_0\|_{H^{\alpha_{\min}}}^2,
\end{aligned}$$

其中我们还用到了 $\nu_{N,\ \hat{}\beta} \leq \nu_{N,\ \beta_{0\circ}}$ 由于 $f_0 \in H^{\alpha_{min}}$，第二项 $\nu_{N,\ \beta_0}^2 \|f_0\|_{H^{\alpha_{\min}}}^2 = O(\nu_{N,\ \beta_0}^2)$ 具有所需的阶。

为估计第一项，令 $u_{0,\ J}$ 表示 u_0的高维小波截断，其中 $2^J \simeq N^{\frac{1}{2\beta_0+d}}$，如证明定理 2.4中所示。于是

$$\|\hat u_N - u_0\|_{H^2} \leq \|\hat u_N - u_{0,J}\|_{H^2} + \|u_0 - u_{0,J}\|_{H^2} =: I + II.$$

由于 $u_0 \in H^{\beta_0}(\mathcal{O})$，第二项可估计为

$$II = \|u_0 - u_{0,J}\|_{H^2} \lesssim 2^{J(\beta_0-2)}\|u_0\|_{H^{\beta_0}} \lesssim N^{-\frac{\beta_0-2}{2\beta_0+d}}.$$

我们通过插值论证来估计第一项。首先，注意到

$$\|u_{0,J}\|^2_{H^{\hat\beta}} \simeq \sum_{l\le J,k} S(f)^2_{lk} 2^{2l\hat\beta} \lesssim 2^{2J(\hat\beta-\beta_0)}\|u_0\|^2_{H^{\beta_0}} \lesssim N^{\frac{2(\hat\beta-\beta_0)}{2\beta_0+d}} \le \mu^2_{N,\beta_0}/\mu^2_{N,\hat\beta},$$

ˆ因为 $\beta \ge \beta_0$。于是，

$$\begin{aligned}
\|\hat u_N - u_{0,J}\|_{H^2} &\lesssim \|\hat u_N - u_{0,J}\|^{2/\hat\beta}_{H^{\hat\beta}} \|\hat u_N - u_{0,J}\|^{(\hat\beta-2)/\hat\beta}_{L^2} \\
&\lesssim \big(\|\hat u_N\|_{H^{\hat\beta}} + \|u_{0,J}\|_{H^{\hat\beta}}\big)^{2/\hat\beta}\big(\|\hat u_N - u_0\|_{L^2} + \|u_0 - u_{0,J}\|_{L^2}\big)^{(\hat\beta-2)/\hat\beta} \\
&\lesssim \big(\mu_{N,\beta_0}/\mu_{N,\hat\beta}\big)^{2/\hat\beta}\mu^{(\hat\beta-2)/\hat\beta}_{N,\beta_0} \\
&= N^{-\frac{\beta_0}{2\beta_0+d}+\frac{2}{2\hat\beta+d}} \le N^{-\frac{\beta_0-2}{2\beta_0+d}}.
\end{aligned}$$

证明至此完成。 □

B.5. 定理 2.9 的证明与定理 2.1的证明类似。回顾

$$\mu_N := N^{-\frac{\alpha+2}{2(\alpha+2)+d}} \qquad \text{and} \qquad \nu_N := N^{-\frac{\alpha}{2(\alpha+2)+d}}.$$

我们先证明插件式 M 估计量的情形。根据引理 A.2，存在 $M = M(\alpha, f_{\min}, R) > 0$ 使得 $\|u_0\|_{H^{\alpha+2}} \le M$。因此，我们可以将定理 3.1 应用于 $H = \mathcal{U} = H^{\alpha+2}(\mathcal{O})$、$\eta^* = u_0$ 和熵上界

$$R + \int_0^R \Big(\frac{R}{\mu_N \rho}\Big)^{\frac{d}{2(\alpha+2)}} d\rho \lesssim R\mu_N^{-\frac{d}{2(\alpha+2)}} =: \Psi(R).$$

这里，我们利用了与定理3.1 相关的局部回归函数类满足 $\mathcal{U}^*(u_0, R) \subseteq \{u : \|u\|_{H^{\alpha+2}} \le R/\mu_N\}$这一事实，同时还利用了定理 4.10.3 中索博列夫类的 L^∞ 熵界，其中 [60], 适用，因为 $\alpha + 2 > d/2$。要求$\sqrt{N}\delta_N^2 \ge \Psi(\delta_N)$ 随后给出了以下收敛速率：

$$\delta_N \gtrsim N^{-\frac{\alpha+2}{2(\alpha+2)+d}} = \mu_N$$

对于ˆu_N，我们得到

$$\mathbb{P}^N_{f_0}\Big(\|\hat u_N - \mathcal{G}(f_0)\|^2_N + \mu^2_N\|\hat u_N\|^2_{H^{\alpha+2}(\mathcal{O})} \ge C\mu^2_N\Big) \lesssim \exp\Big(-\frac{N\mu^2_N}{C}\Big) \xrightarrow{N\to\infty} 0,$$

对于某个充分大的常数 $C > 0$。应用引理 D.1 ，正如定理2.4 的证明中那样，还可得到

$$\mathbb{P}^N_{f_0}\Big(\|\hat u_N - \mathcal{G}(f_0)\|^2_{L^2} + \mu^2_N\|\hat u_N\|^2_{H^{\alpha+2}(\mathcal{O})} \ge C\mu^2_N\Big) \xrightarrow{N\to\infty} 0,$$

对于某个（可能更大的）常数 $C > 0$。这就证明了定理的第一条结论。

令 A_N 表示前一个显示公式中的事件。为证明第二个结论，我们始终可以在事件 A_N 上进行。在 A_N 上，根据插值，我们有

$$\|\hat u_N - u_0\|_{H^2} \lesssim \|\hat u_N - u_0\|^{\frac{\alpha}{\alpha+2}}_{L^2}\|\hat u_N - u_0\|^{\frac{2}{\alpha+2}}_{H^{\alpha+2}} \lesssim N^{-\frac{\alpha}{2(\alpha+2)+d}}$$

此外，对于任意 $\beta \in (d/2, \alpha+2)$，索博列夫嵌入定理和插值论证表明

$$\begin{aligned}\|\hat{u}_N - u_0\|_\infty &\lesssim \|\hat{u}_N - u_0\|_{H^\beta} \\ &\lesssim \|\hat{u}_N - u_0\|_{L^2}^{\frac{\alpha+2-\beta}{\alpha+2}} \|\hat{u}_N - u_0\|_{H^{\alpha+2}}^{\frac{\beta}{\alpha+2}} \\ &\lesssim \mu_N^{\frac{\alpha+2-\beta}{\alpha+2}} \xrightarrow{N\to\infty} 0.\end{aligned}$$

设 $u_{\min} := \inf_{x\in\mathcal{O}} u_0(x)$。在我们的常规假设 $\alpha > d/2$, $g_{\min} > 0$ 以及 $g \in C^\infty(\partial\mathcal{O})$ 满足 $\inf_{\partial\mathcal{O}} g \geq g_{\min}$的条件下，Feynman–Kac论证表明存在一致正性界 $\inf_{x\in\mathcal{O}} u_f(x) \geq g_{\min} e^{-c\|f\|_\infty} > 0$ 对所有 $f \in \mathcal{F}$均成立，见 [47]中的公式 (5.36)。特别地，有 $u_{\min} > 0$。因此

$$\mathbb{P}_{f_0}^N\big(\inf_{x\in\mathcal{O}} \hat{u}_N \geq u_{\min}/2\big) \xrightarrow{N\to\infty} 1.$$

$\hat{}$根据 f_N的定义，有

$$\begin{aligned}\|L_{\hat{f}_N}\hat{u}_N\|_{L^2}^2 + \nu_N^2\|\hat{f}_N\|_{H^\alpha}^2 &\leq \|L_{f_0}\hat{u}_N\|_{L^2}^2 + \nu_N^2\|f_0\|_{H^\alpha}^2 \\ &\leq \|L_{f_0}\hat{u}_N - L_{f_0}u_0\|_{L^2}^2 + \nu_N^2\|f_0\|_{H^\alpha}^2 \\ &\lesssim \|\hat{u}_N - u_0\|_{H^2}^2 + \nu_N^2\|f_0\|_{H^\alpha}^2 \\ &\lesssim \nu_N^2,\end{aligned}$$

$\hat{}$ $\hat{}$特别地，这也意味着对于 $\beta \in (0, \alpha - d/2)$，在事件 A_N上，$\|f_N\|_{C^\beta} \lesssim \|f_N\|_{H^\alpha} \leq C''$对某个常数 C''成立。因此，将引理2.10 中的广义稳定性估计应用于第一个三元组$(u_0, f_0, 0)$ 和第二个三元组$\hat{}(\hat{}u_N, f_N, L_{\hat{}f_N}\hat{}u_N)$，可得在事件 A_N上，

$$\|\hat{f}_N - f_0\|_{L^2} \lesssim \|\hat{u}_N - u_0\|_{H^2} + \|L_{\hat{f}_N}\hat{u}_N\|_{L^2} \lesssim \nu_N,$$

这就证明了第二个结论。 □

附录 C：定理3.1 的证明

证明通过将 [64] 中的思路调整到带有“增广参数” (η, θ) 的当前情形来完成。任取 (θ^*, η^*)，记

$$\hat{u} := u_{\hat{\eta}}, \qquad u^* := u_{\eta^*}.$$

$\hat{}$根据 $(\hat{}\eta_N, \theta_N)$ 的定义，有

$$\frac{1}{N}\sum_{i=1}^N (Y_i - u_{\hat{\eta}}(X_i))^2 + \mathcal{R}(\hat{\eta}, \hat{\theta}) \leq \frac{1}{N}\sum_{i=1}^N (Y_i - u^*(X_i))^2 + \mathcal{R}(\eta^*, \theta^*),$$

这与下式相同

$$\tau^2((\hat{\eta}, \hat{\theta}), u_0) - \tau^2((\eta^*, \theta^*), u_0) \leq \frac{2}{N}\sum_{i=1}^N \varepsilon_i(\hat{u}(X_i) - u^*(X_i)). \tag{70}$$

接下来，我们注意到，泛函 τ^2 满足关于η变量的以下“类三角”不等式：对于任意的 $\eta, \eta', \eta'' \in H$，其对应的回归函数为 u, u', u''，以及任意

$\theta,\theta'\in\Theta$，利用经验范数的不等式 $\|a-b\|_N^2\le 2\|a-c\|_N^2+2\|c-b\|_N^2$ 以及惩罚项 $\mathcal{R}\ge 0$ 的非负性，我们有

$$\tau^2((\eta,\theta),u_0)\le 2\tau^2((\eta,\theta),u')+2\tau^2((\eta',\theta'),u_0),$$
$$\tau^2((\eta,\theta),u'')\le 2\tau^2((\eta,\theta),u')+2\tau^2((\eta'',\theta'),u'),$$

这意味着

$$\tau^2((\hat\eta,\hat\theta),u^*)\ge\frac{1}{2}\tau^2((\hat\eta,\hat\theta),u_0)-\tau^2((\eta^*,\theta^*),u_0), \tag{71}$$

$$\tau^2((\hat\eta,\hat\theta),u_0)\ge\frac{1}{2}\tau^2((\hat\eta,\hat\theta),u^*)-\tau^2((\eta^*,\theta^*),u_0). \tag{72}$$

将我们要界定其概率的事件记为

$$A_N:=\big\{\tau^2((\hat\eta,\hat\theta),u_0)\ge 2\big(\tau^2((\eta^*,\theta^*),u_0)+R^2\big)\big\}. \tag{73}$$

关于 A_N，利用（71），可得

$$\tau^2((\hat\eta,\hat\theta),u^*)\ge R^2.$$

此外，根据 A_N 的定义并利用（72），我们得到

$$\begin{aligned}\tau^2((\hat\eta,\hat\theta),u_0)-\tau^2((\eta^*,\theta^*),u_0)&=\frac{1}{3}\tau^2((\hat\eta,\hat\theta),u_0)+\frac{2}{3}\tau^2((\hat\eta,\hat\theta),u_0)-\tau^2((\eta^*,\theta^*),u_0)\\&\ge\frac{1}{3}\tau^2((\hat\eta,\hat\theta),u_0)+\frac{1}{3}\tau^2((\eta^*,\theta^*),u_0)\\&\ge\frac{1}{6}\tau^2((\hat\eta,\hat\theta),u^*).\end{aligned}$$

结合（70），进一步可得

$$\frac{1}{N}\sum_{i=1}^N\varepsilon_i(\hat u(X_i)-u^*(X_i))\ge\frac{1}{12}\tau^2((\hat\eta,\hat\theta),u^*).$$

上述论证使我们能够采用“切片法”来对 A_N 的概率进行上界估计，

$$\begin{aligned}\mathbb{P}_0^N(A_N)&\le\sum_{s=1}^\infty\mathbb{P}_0^N\big(A_N\text{ and }\tau^2((\hat\eta,\hat\theta),u^*)\in\big[2^{2(s-1)}R^2,2^{2s}R^2\big]\big)\\&\le\sum_{s=1}^\infty\mathbb{P}_0^N\Big(\sup_{u\in\mathcal{U}(\eta^*,2^sR)}\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^N\varepsilon_i(u(X_i)-u^*(X_i))\Big|\ge\frac{1}{48}\sqrt{N}2^{2s}R^2\Big)\\&:=\sum_{s=1}^\infty P_s\end{aligned}$$

我们记作 $\mathcal{U}_s=\mathcal{U}(\eta^*,\ 2^sR)$。对于每个 $s\ge 1$，在给定设计点 $(X_i)_{i=1}^N$ 的条件下，该过程

$$X_u=\frac{1}{\sqrt{N}}\sum_{i=1}^N\varepsilon_i(u(X_i)-u^*(X_i)),\quad u\in\mathcal{U}_s,$$

是以协方差度量为协方差的中心高斯分布

$$d^2(u,u')=\frac{1}{N}\sum_{i=1}^N\sigma_i^2(u(X_i)-u'(X_i))^2\le\sigma^2\|u-u'\|_N^2,$$

相对于该度量，集合 $\mathcal{U}_s$ 的直径为

$$D_s := \sup_{u,u' \in \mathcal{U}_s} d(u,u') \le 2\sigma 2^s R.$$

次高斯过程上确界的 Dudley不等式（例如，参见 [21] 中的定理 2.3.6 和 2.3.7，或 [66]中的定理 2.2.4）表明，对于任意 $v \in \mathcal{U}_s$

$$\begin{aligned}\mathbb{E}_0^N\big[\sup_{u\in\mathcal{U}_s}|X_u|\big] &\le \mathbb{E}_0^N|X_v| + 4\sqrt{2}\int_0^{D_s/2}\sqrt{H(\mathcal{U}_s,\sigma\|\cdot\|_N,\rho)}d\rho \\ &\le \sigma 2^s R + 4\sqrt{2}\int_0^{\sigma 2^s R}\sqrt{H(\mathcal{U}_s,\sigma\|\cdot\|_N,\rho)}d\rho \\ &\le 4\sqrt{2}\Psi(2^s R) \le \frac{1}{96}\sqrt{N}2^{2s}R^2,\end{aligned}$$

在假设（39）中将常数 $C>0$ 取足够大，即可确保最后一个不等式成立。利用这一点以及 Borell-Sudakov-Tsirelson 不等式（参见 [21] $\sup_{u\in\mathcal{U}_s}|X_u|$ 中的定理 2.5.8），注意到一 $\sigma^2 2^{2s}R^2$ 维方差的上界为 supVar(),

$$P_s \le \mathbb{P}_0^N\Big(\sup_{u\in\mathcal{U}_s}|X_u| - \mathbb{E}_0^N\sup_{u\in\mathcal{U}_s}|X_u| \ge \frac{1}{96}\sqrt{N}2^{2s}R^2\Big) \le \exp\Big(-\frac{N2^{2s}R^2}{2\cdot 96^2\sigma^2}\Big).$$

将前面的不等式相加可得

$$P_0^N(A_N) \le \sum_{s=1}^{\infty} P_s \le C\exp\big(-NR^2/C\sigma^2\big),$$

这里还用到了：在假设（39）下，存在 $\sqrt{N}\delta_N^2 \ge C\sigma\delta_N$使得$N\overline{R^2}/(2*96^2\sigma^2)$对于任意 $R \ge \delta_N$，都可以由某个普适常数从下方界定。这就证明了（40）。

为证明（41）关于 $k=1$，我们像 [63] 中的引理 2.2 一样，对集中不等式（40）进行积分。记 $\hat{}\tau^2 = \tau^2((\hat{}\eta,\ \hat{}\theta)\ u_0)$。对于满足（39）的 δ_N，令

$$t_N^* = 2(\tau^2((\eta^*,\theta^*),u_0) + \delta_N^2).$$

Then

$$\begin{aligned}\mathbb{E}_0^N[\hat{\tau}^2] &\le t_N^* + \int_{t_N^*}^{\infty}\mathbb{P}_0^N(\hat{\tau}^2 \ge t)dt \le t_N^* + C\int_{\delta_N}^{\infty}\exp\Big(-\frac{Nu^2}{C\sigma^2}\Big)du^2 \\ &\le t_N^* + \frac{C'\sigma^2}{N}\exp\Big(-\frac{N\delta_N^2}{C\sigma^2}\Big) \le t_N^* + \frac{C'\sigma^2}{N},\end{aligned}$$

由此得证。

现在我们证明关于$\hat{}\tau$的高阶矩的结论。对于整数 $k \ge 2$，有

$$\mathbb{E}_0^N[\hat{\tau}^{2k}] \le (t_N^*)^k + \int_{(t_N^*)^k}^{\infty}\mathbb{P}_0^N(\hat{\tau}^2 \ge t^{1/k})dt \le (t_N^*)^k + C\int_{\delta_N^{2k}}^{\infty}\exp\Big(-\frac{Nt^{1/k}}{C\sigma^2}\Big)dt \tag{74}$$

由于 $\Psi(R) \ge \sigma R$，假设 $\sqrt{N}\delta_N^2 \ge C\overline{\Psi}(\delta_N)$ 蕴含 $N\delta_N^2/(C\sigma^2) \ge 1$，因此引理 D.2 适用于 $\gamma = 1/k$、$a = \delta_N^{2k}$和 $L = N/(C\sigma^2)$。利用该引理，我们得到

$$\int_{\delta_N^{2k}}^{\infty}\exp\Big(-\frac{Nt^{1/k}}{C\sigma^2}\Big)dt \lesssim \frac{\sigma^2\delta_N^{2k-2}}{N}\exp\Big(-\frac{N\delta_N^2}{C\sigma^2}\Big) \le \frac{\sigma^2\delta_N^{2k-2}}{N},$$

如所需。 □

附录 D：两个辅助引理

D.1. 经验范数的集中性。

引理 D.1（经验范数的集中性）。设 $\mathcal{O}\subseteq\mathbb{R}^d$ 是某个具有有限 *Lebesgue*测度的开子集，且 $X_i \sim^{i.i.d.} \mathit{Unif}(\mathcal{O})$。设 $\mathcal{U}$ 是定义在 $\mathcal{O}$ 上的一类实值函数，并具有一致上界

$$\sup_{u\in\mathcal{U}}\|u\|_{L^\infty(\mathcal{O})} =: M < \infty,$$

其中 $\|u\|_{L^\infty(\mathcal{O})} = \sup_{x\in\mathcal{O}}|u(x)|$。存在一个普适常数 $C>0$，使得对于任何满足条件的$(\delta_N : N\geq 1)$，

$$N\delta_N^2 \geq CM^2 H(\mathcal{U}, \|\cdot\|_\infty, \delta_N), \tag{75}$$

以及任何 $R \geq C(\delta_N + M/\sqrt{N})$，均有

$$\mathbb{P}\Big(\|u\|_{L^2(\mathcal{O})} \geq 2\|u\|_N \text{ for some } u\in\mathcal{U} \text{ with } \|u\|_{L^2(\mathcal{O})}\geq R\Big) \leq 2\exp\left(-\frac{NR^2}{CM^2}\right).$$

Proof。引理 D.1 是对 [52] 中的引理 A.5 的直接特化，适用于均匀设计 $X_i \sim \mathrm{Unif}(\mathcal{O})$ 以及具有一致包络 $M=\sup_{u\in\mathcal{U}}\|u\|_\infty$ 的函数类 $\mathcal{U}$。

□

对于属于充分规则的索博列夫球的函数类 $\mathcal{U}$

$$\mathcal{U} \subseteq \{u\in H^\alpha(\mathcal{O}) : \|u\|_{H^\alpha}\leq M\}, \quad \alpha > d/2,$$

前述引理蕴含如下结论。根据索伯列夫嵌入 $\|\cdot\|_{H^\alpha}$ H $\mathcal{U}$ $\|\cdot\|_\infty$ $\delta_N \lesssim M/\delta_N$ $^{d/\alpha}$ [60, $\|\cdot\|_\infty \lesssim$ 4 10 3] 并利用熵界 (,,) ()(例如定理 . .)，我们看到条件 (75) 对以下情形成立

$$\delta_N \gtrsim MN^{-\frac{\alpha}{2\alpha+d}},$$

其中 $\gtrsim$ 表示一个与 M,N 无关的常数。由此可知，对于某个常数 $C'>0$,

$$\mathbb{P}\Big(\|u\|_{L^2(\mathcal{O})} \geq 2\|u\|_N \text{ for some } \|u\|_{H^\alpha}\leq M,\ \|u\|_{L^2(\mathcal{O})}\geq R\Big) \leq 2\exp\left(-\frac{NR^2}{C'M^2}\right),$$

其中某个常数为 C'。

D.2. 一个基本的尾不等式。 以下尾界将在我们的证明中反复使用。

引理 D.2. 对于任意 $\gamma\in(0,1)$，存在常数 $C_\gamma>0$，使得对于所有满足 $La^\gamma\geq 1$ 的 $L,a>0$,

$$\int_a^\infty e^{-Lt^\gamma}\,dt \leq C_\gamma \frac{a^{1-\gamma}}{L}\,e^{-La^\gamma}.$$

证明。将 $u=Lt^\gamma$ 代入即可看出，因此

$$t=\left(\frac{u}{L}\right)^{1/\gamma}, \qquad dt=\frac{1}{\gamma}L^{-1/\gamma}u^{\frac{1}{\gamma}-1}\,du.$$

Then

$$\int_a^\infty e^{-Lt^\gamma}\,dt = \frac{1}{\gamma}L^{-1/\gamma}\int_{La^\gamma}^\infty u^{\frac{1}{\gamma}-1}e^{-u}\,du = \frac{1}{\gamma}L^{-1/\gamma}\Gamma\left(\frac{1}{\gamma}, La^\gamma\right),$$

其中 Γ 是上不完全伽马函数,

$$\Gamma(s,z)=\int_z^\infty u^{s-1}e^{-u}\,du.$$

对于固定的 $s>0$, Γ 的尾部满足

$$\Gamma(s,z)\le C_s\,z^{s-1}e^{-z}\qquad \text{for } z\ge 1.$$

将此应用于 $s=\frac{1}{\gamma}$ 和 $z=La^\gamma$, 可得

$$\int_a^\infty e^{-Lt^\gamma}\,dt\le C_\gamma\,\frac{a^{1-\gamma}}{L}\,e^{-La^\gamma},$$

其中 $C_\gamma>0$ 仅取决于 γ。 □