arXiv:2303.01512v1 [stat.ML] 2 Mar 2023

基于积分概率度量的贝叶斯后验扰动分析

ALFREDO GARBUNO-I ˜NIGO *、Tapio Helin †、Franca Hoffmann ‡、和Bamdad Hosseini §

摘要。近年来，科学、工程和机器学习领域中的大规模逆问题所涉及的贝叶斯推断受到了广泛关注。本文通过分析后验测度相对于似然势和先验测度扰动的稳定性，考察贝叶斯方法的稳健性。我们借助一族类似于最优传输中对偶问题的积分概率度量（散度），提出了新的稳定性结果。与以往研究相比，我们的成果有三个方面的突出之处：（1）构造了适用于当前问题的新型积分概率度量族；（2）这些新度量使我们能够以简便的方式研究似然和先验扰动；（3）我们的分析适用于仅局部利普希茨连续的似然势函数，因此可用于广泛的非线性逆问题。我们还通过具体且新颖的例子进一步印证了理论发现，在这些例子中，我们针对不同类型的扰动得到了后验测度的逼近率，并为分析近期应用于贝叶斯逆问题的机器学习技术的收敛性指明了方向，例如数据驱动先验和神经网络替代模型。

关键词。逆问题，贝叶斯，适定性，积分概率度量，最优传输

1. 引言。 贝叶斯逆问题/推断问题（BIPs）广泛存在于科学和工程应用中，在这些应用中，不确定性量化（UQ）至关重要[24]。近年来，随着机器学习和数据科学的兴起，贝叶斯逆问题在这些领域找到了新的应用，尤其是由于支持不确定性量化的学习算法正日益受到欢迎 [1, 3, 11, 21, 23, 33, 38]。贝叶斯逆问题的这些新应用领域提出了令人兴奋的新理论问题，涉及贝叶斯逆问题的适定性和稳定性，而这反过来又需要开发新的工具。本文旨在探讨这些问题，重点关注贝叶斯后验测度相对于先验测度、似然和数据的扰动性质。许多学习问题涉及高维参数空间，例如图像；图像可以自然地视为函数型数据，因此函数空间贝叶斯逆问题框架成为一种自然选择。

考虑定义在可分巴拿赫空间 X 上的贝叶斯逆问题，其中 $\mathbb{P}(\mathrm{X})$ 表示 X 上的拉东概率测度空间。我们考虑后验测度 $\nu \in \mathbb{P}(\mathrm{X})$，其密度是相对于底层先验 $\mu \in \mathbb{P}(\mathrm{X})$ 定义的：

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(u) = \frac{1}{Z(y)} \exp(-\Phi(u;y)), \quad Z(y) := \mu\left(\exp(-\Phi(u;y))\right), \tag{1.1}$$

其中我们使用了记号 $\mu(f) := \int_{\mathcal{X}} f(y)\mathrm{d}\mu(u)$。我们将 $\Phi : \mathrm{X} \times \mathrm{Y} \mapsto \mathbb{R}$ 称为似然势，Y 是观测数据 y 所在的空间，而 $Z(y)$ 是确保 ν 为概率测度的归一化常数（参见 [46]，了解函数空间框架下贝叶斯公式的推导）。

本文关注后验测度 ν 关于似然函数 Φ、数据 y 和先验 μ 的扰动性质。更

*墨西哥城墨西哥自治技术学院，墨西哥城，墨西哥，（alfredo.garbuno@itam.mx）
†拉彭兰塔-拉赫蒂理工大学，拉彭兰塔，芬兰，（tapio.helin@lut.fi）
‡加利福尼亚州理工学院，帕萨迪纳，加利福尼亚州，美国，franca.hoffmann@caltech.edu）§华盛顿大学，西雅图，华盛顿州，美国，（bamdadh@uw.edu）

确切地说，设 $\mu_\varepsilon \in \mathbb{P}(\mathsf{X})$、 Φ_ε 和 y_ε 表示先验、似然函数和数据的扰动，并由 $\varepsilon \in [0, +\infty)$ 参数化，从而通过（1.1）得到后验测度 ν_ε。那么，我们希望证明，当 $\varepsilon \to 0$ 时，后验分布 $\nu_\varepsilon \to \nu$ 以适当的意义收敛。具体而言，我们将选择一个统计散度 $\mathsf{D} : \mathbb{P}(\mathsf{X}) \times \mathbb{P}(\mathsf{X}) \to \mathbb{R}_{\geq 0}$，并建立形如 $\mathsf{D}(\nu, \nu_\varepsilon) \lesssim \psi(\varepsilon)$ 的不等式，其中 ψ 为某个函数，从而也得到收敛速率。

此类扰动结果与贝叶斯逆问题在高维或函数型参数推断中的应用密切相关。例如，在函数空间推断中，似然扰动会自然出现，因为在进行任何模拟之前都应采用离散化方案；或者在使用多项式混沌展开[28, 42, 27, 26], 高斯过程回归 [45, 48] 或神经网络 [15, 25] 等替代模型时，也会出现似然扰动，此时似然（或底层前向模型）会被计算成本低廉的近似所替代。事实上，似然扰动是贝叶斯逆问题适定性理论中研究较为充分的主题[7, 46, 18, 16, 22, 47, 37, 43]。然而，除 [43], 之外，以往的研究忽略了先验测度 μ的扰动。这在很大程度上源于对 D 的选择，因为全变差（TV）或海林格度量意义下的收敛要求后验分布ν 与扰动 ν_ε 相互等价，才能得到有意义的收敛速率。反过来，这也要求先验 μ 与扰动 μ_ε 相互等价，但这一要求过于严格，排除了现代应用中出现的一些有趣情形，例如数据驱动的先验构造 [3, 23, 38], ，即从训练数据集学习先验的情形。例如，我们可以将 μ_ε 视为由一组独立样本得到的 μ 的经验近似。另一种常见方法是对$\mu_\varepsilon = T^\varepsilon_\sharp \eta$进行参数化（即通过 T^ε将 η推前测度），其中 T^ε 是一个离线学习得到的映射（例如神经网络），而 η是定义在抽象潜在空间上的通用参考测度。为说明这种方法为何容易导致奇异先验，考虑一个简单例子： μ 是高斯分布，而 $\mu_\varepsilon = \mu(\cdot - \varepsilon v)$（传输映射就是 $T^\varepsilon : x \mapsto x + \varepsilon v$），其中元素 $v \in \mathsf{X}$ 位于μ [5, 的卡梅隆-马丁空间之外，见第 2 节4]；在这种情况下， μ 与 μ_ε 变得相互奇异。

的一个创新之处 [43] 是将$\mathcal{D}$ 定义为 Wasserstein 度量$\mathcal{W}_p$。这些度量足够灵活，可以根据先验扰动来控制后验扰动；但仍需假设似然 Φ 是全局利普希茨连续的，这排除了许多非线性逆问题。在这里，我们将通过把$\mathcal{D}$ 定义为积分概率度量（IPM） [1] 来解决这一问题：

$$\mathcal{D}(\eta_1, \eta_2) = \sup_{\psi \in \Gamma} \eta_1(\psi) - \eta_2(\psi), \qquad \forall \eta_1, \eta_2 \in \mathbb{P}(\mathcal{X}), \tag{1.2}$$

其中 Γ 是定义在 $\mathcal{X}$ [2]上的实值函数子集。其思想是，如果 Γ 足够大，那么可以选择一个 ψ 来区分 $\mathbb{P}(\mathcal{X})$ 中的任意一对不同测度，从而通过最大化右侧使得 $\mathcal{D}(\eta_1, \eta_2) = 0$ 当且仅当 $\eta_1 = \eta_2$；这在 [4]的术语中被称为 D 的 散度性质。我们注意到，尽管散度性质在理论上很有吸引力，但在许多不确定性量化应用中并非必需。例如，设 $\psi^\star$ 是一个关注量，其关于

[1]我们将这些积分概率“度量”称为度量，是因为这是文献中沿用的术语，尽管我们的积分概率度量通常只是散度或半度量，并不一定满足度量的所有度量公理。[2]我们主要关注这样的情况： Γ 是关于代价函数 c 的利普希茨函数类，因为这足以满足我们的应用需求；不过，我们的结果也可以推广到更广泛的选择。

相对于 ν 的期望值是所需的。那么，我们只需确保集合 Γ包含 $\psi^\star$，这样就可以直接界定

$$|\nu(\psi^\star) - \nu_\varepsilon(\psi^\star)| \le \mathcal{D}(\nu, \nu_\varepsilon),$$

这使我们能够控制关注量的误差。

积分概率度量已广泛应用于统计理论 [31] ，近年来也应用于机器学习 [44, 4]；常见的例子包括最大均值差异（MMD） [30, 2] 以及 f散度 [32]。积分概率度量还自然地出现在最优传输（OT）以及 Wasserstein 度量对偶形式的研究中 [49]；这一性质也将在本文中加以利用。与此同时，与之密切相关的最优传输半度量已用于随机偏微分方程和函数空间马尔可夫链蒙特卡罗算法的收敛性分析[13, 14, 19, 17]；这些研究为我们的分析提供了重要启发。在本文中，我们利用积分概率度量的灵活性，尤其是根据似然势 Γ调整空间的选取 Φ ，并在比以往研究更一般的情形下获得扰动界。根据具体似然调整积分概率度量的思路，是简化我们许多论证的关键，也使我们能够得到灵活的界。

1.1. Main Contributions. Our main contributions are summarized below:

- 我们使用抽象积分概率度量来衡量后验分布的近似速率。积分概率度量是概率测度差异的一大类，其中包括 Wasserstein度量等。我们的方法根据具体问题的似然来调整积分概率度量的选择，并不要求对数似然满足任何全局利普希茨稳定性假设。定理 3.2 和 3.8 分别分析了似然扰动和先验扰动对后验分布的影响，而推论 3.6 和 3.7 则考察了数据扰动。

- 第 4 节中的多个具体应用展示了我们的扰动分析的强大能力。在此，我们获得了由先验近似产生的后验测度的定量收敛速率，包括第 4.1节中基于一组样本对先验进行的经验近似；第 ´节中具有不同超参数的 Matérn 核高斯测度等乘积先验的谱扰动；第 4.2节中被识别为参数化传输映射推前测度的数据驱动先验；以及第 4.3节中似然的深度神经网络（DNN）代理模型。4.4。

1.2. 概述。 本文其余部分安排如下：第 2 节讨论数学预备知识和积分概率度量。我们在第 3 节推导稳定性结果，并在第 4节概述应用。

2. 类距离函数与积分概率度量。 这里汇集一些关于积分概率度量的基本性质和定义，用于推导第 3节中的扰动结果。我们称函数 $c : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ 在类距离的意义下 [13] 具有类距离性质，如果它是正的、对称的、下半连续的，并且满足 $c(u, v) = 0$ 当且仅当 $u = v$。给定这样的类距离 c ，我们定义 $\mathbb{P}(\mathcal{X})$的以下子空间：

$$P_1(\mathcal{X}; c) := \left\{ \eta \in \mathbb{P}(\mathcal{X}) \;\middle|\; \eta(c(0, u)) < +\infty \right\},$$

以及相对于 c 定义在 X 上的 1-Lipschitz函数空间，

$$\mathrm{Lip}_1(\mathcal{X}; c) := \left\{ \psi : \mathcal{X} \mapsto \mathbb{R} \;\middle|\; \sup_{u \neq v} \frac{|\psi(u) - \psi(v)|}{c(u, v)} \le 1 \right\}.$$

我们也用 $\mathrm{Lip}_1^0(\mathcal{X};\ c)$ 表示 $\mathrm{Lip}_1(\mathcal{X};\ c)$ 中在原点取值为零的函数所构成的子空间。最后，我们定义如下一个类似于康托罗维奇-鲁宾斯坦泛函的积分概率度量 [6]:

$$\begin{aligned}\mathcal{D}(\eta_1,\eta_2;c) &:= \sup_{\psi\in \mathrm{Lip}_1(\mathcal{X};c)} \eta_1(\psi)-\eta_2(\psi),\\ &\equiv \sup_{\psi\in \mathrm{Lip}_1^0(\mathcal{X};c)} \eta_1(\psi)-\eta_2(\psi), \qquad \eta_1,\eta_2\in\mathbb{P}(\mathcal{X}),\end{aligned} \tag{2.1}$$

其中，两个定义等价，仅仅是因为右侧泛函的值在将 ψ 平移一个常数后不会改变。众所周知（例如参见 [49, 特例 5.4]），如果 c 满足三角不等式，那么泛函 D 关于一个最优传输代价函数满足对偶定理，即

$$\mathcal{D}(\eta_1,\eta_2;c)=\mathcal{W}(\eta_1,\eta_2;c):=\inf_{\pi\in\Pi(\eta_1,\eta_2)}\int_{\mathcal{X}\times\mathcal{X}} c(u,v)\mathrm{d}\pi(u,v), \tag{2.2}$$

其中，$\Pi(\eta_1,\eta_2)$ 是 η_1 与 η_2 之间所有耦合构成的空间，也就是 $\mathbb{P}(\mathrm{X}\times\mathrm{X})$ 中由边缘分布为 η_1,η_2 的测度组成的子空间。这进一步意味着，如果 c 是 X 上的度量，那么 D 在 $P_1(\mathrm{X};\ c)$ [6, 49] 上定义了一个度量。在特定情形下，当 $c(u,v)=\|u-v\|^p$ 对于 $p\geq 1$ 成立时，泛函 $\mathrm{D}^{1/p}$ 与概率测度的著名 Wasserstein-p度量一致。

在第 3 节中，我们将需要使用代价函数 c，这些函数不是度量，而是半度量，因为它们满足弱三角不等式（三角不等式中的常数大于一）。在这种情况下，泛函 $\mathcal{D}$ 可能不再满足对偶性（2.2），但我们仍可利用以下不等式

$$\mathcal{D}(\eta_1,\eta_2;c)\leq\mathcal{W}(\eta_1,\eta_2;c), \tag{2.3}$$

得益于关于 ψ 的 Lipschitz 假设。该恒等式在第 4 节中将非常有用；在该节中，我们使用耦合论证或 Wasserstein距离的已知界来界定 D。此外，由于我们假设 c 类似于距离，因此可以验证 $\mathrm{W}(\cdot,\cdot;c)$ 满足 [4], 术语中的散度性质，即 $\mathrm{W}(\eta_1,\eta_2;\ c)=0$ 当且仅当 $\eta_1=\eta_2$。于是，根据 (2.3)，当 $\eta_1=\eta_2$ 时，$\mathrm{D}(\eta_1,\eta_2;\ c)=0$。然而，要证明逆命题，进而证明 D 的散度性质，我们还需要对 c 作出额外假设。

Proposition 2.1. 假设 c 类似于距离，并且存在一个度量 ρ，使得 $\rho(u,v)\leq c(u,v)$，则 $\mathrm{D}(\cdot,\cdot;\ c)$ 满足散度性质。

证明。根据上文的讨论，我们只需证明 $\eta_1\neq\eta_2$ 蕴含 $\mathrm{D}(\eta_1,\eta_2;\ c)>0$。我们有 $\mathrm{Lip}_1(\mathrm{X};\ \rho)\subset\mathrm{Lip}_1(\mathrm{X};\ c)$，结合 $\mathrm{W}(\cdot,\cdot;\rho)$ 的对偶性结果，可得 $\mathrm{W}(\cdot,\cdot;\rho)\leq\mathrm{D}(\cdot,\cdot;\ c)$。结论成立，因为 $\mathrm{W}(\cdot,\cdot;\rho)$ 是一个度量，因此满足散度性质。 □

3. 后验测度扰动理论。本节概述我们的主要理论贡献，并展示在适当的积分概率度量下，后验分布关于似然和先验扰动的利普希茨稳定性。更确切地说，给定式（ν（1.1）中的真实后验分布，我们可以考虑以下测度

$$\frac{\mathrm{d}\nu_*}{\mathrm{d}\mu_*}(u)=\frac{1}{Z_*(y)}\exp\left(-\Phi(u;y)\right),\quad\text{and}\quad\frac{\mathrm{d}\nu'_*}{\mathrm{d}\mu_*}(u)=\frac{1}{Z'_*(y)}\exp\left(-\Phi'(u;y)\right),$$

其归一化常数 $Z_*(y)$ 和$Z'_*(y)$ 的定义与式（1.1）类似。于是，若$\mathcal{D}$ 是满足弱三角不等式的积分概率度量，我们便可直接写出

$$\mathcal{D}(\nu,\nu'_*) \lesssim \mathcal{D}(\nu,\nu_*) + \mathcal{D}(\nu_*,\nu'_*),$$

因此，右侧第一项仅涉及先验扰动，而第二项涉及似然扰动。因此，在本节中分别考察这些扰动是很自然的。我们首先在第 3.1 节概述关于似然的主要假设，随后在第 3.2 节讨论似然扰动，包括扰动数据的情形，并在第 3.3 节讨论先验扰动。

3.1. 似然的假设。下面汇总对似然势 Φ 所作的一组假设，这些假设将在本文余下部分中始终使用。这些是贝叶斯逆问题 [18, 16, 22, 43, 46, 47] 适定性理论中的标准假设，可确保所得后验分布定义良好。

假设 1。似然势 $\Phi : \mathrm{X} \times \mathrm{Y} \mapsto \mathbb{R}$ 满足以下一个或多个假设：

(i)（先验可测性）对于任意固定的 $y \in \mathrm{Y}$，函数 $\Phi(\cdot; y) : \mathrm{X} \mapsto \mathbb{R}$ 关于先验 μ 可测。(ii)（局部有界的上界和下界）存在一个下半连续函数 $f : \mathrm{X} \to \mathbb{R}_{>0}$ 以及局部有界函数 $g : \mathrm{Y} \to \mathbb{R}_{>0}$ 和 $h : \mathrm{X} \times \mathrm{Y} \to \mathbb{R}_{>0}$，使得 $\forall (u, y) \in \mathrm{X} \times \mathrm{Y}$ 成立，即

$$-\log f(u) - \log g(y) \le \Phi(u; y) \le -\log h(u, y).$$

根据上述假设，我们可以立即保证，对于固定的数据 $y \in \mathrm{Y}$，后验测度 ν 存在且唯一。

Proposition 3.1 ([16, 定理 4。3])。假设 Φ 满足假设*1* ，其中$f \in L^1(\mathcal{X}\ \mu)$。那么，对于任意固定的 $y \in \mathcal{Y}$，我们有 $0 < Z(y) < +\infty$，这意味着 ν是定义良好的概率测度。

需要强调的是，虽然我们的假设足以确保后验分布定义良好，但绝非必要条件。事实上， [22] 在弱得多的假设下也能得到相同的结果；但为了构造积分概率度量并获得稳定性结果，我们采用了更强的假设，尤其是条件 (ii)。与此同时，条件 (ii) 足够一般，能够涵盖许多有趣的应用，例如第4节所述的应用。

3.2. 似然的扰动。本节研究似然 Φ发生扰动时，后验测度 ν 所受到的影响。这类扰动自然出现在函数空间贝叶斯逆问题的离散化以及计算开销较大的似然的代理建模中 [28, 42, 27, 26, 45, 48, 15, 25]。

考虑修正后的似然势 $\Phi'(u; y)$，进而定义扰动后的后验分布 $\nu' \in \mathbb{P}(\mathrm{X})$：

$$\frac{\mathrm{d}\nu'}{\mathrm{d}\mu} = \frac{1}{Z'(y)} \exp\left(-\Phi'(u; y)\right), \qquad Z'(y) := \int_{\mathcal{X}} \exp(-\Phi'(u; y)) \mathrm{d}\mu(u). \tag{3.1}$$

我们在这一一般设定下发展扰动理论，并在后文考察两种具体设定：一种是数据 y 受到扰动，另一种是似然势 Φ 由于近似而本身发生改变。

THEOREM 3.2. 设 Φ和 Φ' 满足假设 *1* ，且具有相同的函数f， g， h，并设 p，$q \in [1,\infty]$ 满足 $1/p + 1/q = 1$。则有

$$\mathcal{D}(\nu,\nu';c) \le \frac{2g^2(y)\|fc(\cdot,0)\|_{L^p(\mu)}\|f\|_{L^p(\mu)}}{\|h(\cdot,y)\|^2_{L^1(\mu)}}\|\Phi(\cdot;y)-\Phi'(\cdot;y)\|_{L^q(\mu)}, \quad \textit{for all } y \in \mathcal{Y}.$$

注 3.3。注意，要求 Φ' 与 Φ 使用相同的函数 f、g、h 满足假设 1 并无妨碍。事实上，如果 Φ'使用不同的函数 f'、g'、h'满足该假设，那么只需将 f替换为 $f \vee f'$，将 g替换为$g \vee g'$，并将 $h(\cdot,y)$ 替换为 $h(\cdot,y) \wedge h'(\cdot,y)$ [3]，定理 3.2 仍然成立。此外，由于 ν和 ν'都关于 μ 绝对连续，因此可以将假设 1(ii)放宽为仅要求其对 μ-几乎处处的 u 成立，而非对 X 中所有点成立。

证明。我们的证明步骤与 [43, 定理 14]的证明类似。根据 D 的定义以及测度 ν、ν'，我们有

$$\begin{aligned}\mathcal{D}(\nu,\nu';c) = &\sup_{\psi\in \mathrm{Lip}_1^0(\mathcal{X};c)} \frac{1}{Z(y)}\int_{\mathcal{X}}\psi(u)\exp(-\Phi(u;y))\mathrm{d}\mu(u)\\ &\qquad - \frac{1}{Z'(y)}\int_{\mathcal{X}}\psi(u)\exp(-\Phi'(u;y))\mathrm{d}\mu(u)\\ \le &\frac{1}{Z'(y)}\sup_{\psi\in \mathrm{Lip}_1^0(\mathcal{X};c)}\int_{\mathcal{X}}\psi(u)\left[\exp(-\Phi(u;y))-\exp(-\Phi'(u;y))\right]\mathrm{d}\mu(u)\\ &+\frac{|Z(y)-Z'(y)|}{Z(y)Z'(y)}\sup_{\psi\in \mathrm{Lip}_1^0(\mathcal{X};c)}\int_{\mathcal{X}}\psi(u)\exp(-\Phi(u;y))\mathrm{d}\mu(u) =: T_1 + T_2.\end{aligned}$$

首先，我们利用指数函数的中值定理以及假设 1(ii)由 Φ 和 Φ'满足这一事实，对 T_1 进行界定。

$$\begin{aligned}Z'(y)T_1 &\le \sup_{\psi\in \mathrm{Lip}_1^0(\mathcal{X};c)}\int_{\mathcal{X}}|\psi(u)|\exp\left(-\left[\Phi(u;y)\wedge\Phi'(u;y)\right]\right)|\Phi(u;y)-\Phi'(u;y)|\mathrm{d}\mu(u)\\ &\le g(y)\sup_{\psi\in \mathrm{Lip}_1^0(\mathcal{X};c)}\int_{\mathcal{X}}|\psi(u)|f(u)|\Phi(u;y)-\Phi'(u;y)|\mathrm{d}\mu(u).\end{aligned}$$

现在利用 $\mathrm{Lip}_1^0(X;c)$ 中的函数由 $c(\cdot,0)$ 有界这一事实以及 H¨older 不等式，我们可以进一步如下估计 T_1：

$$\begin{aligned}Z'(y)T_1 &\le g(y)\int_{\mathcal{X}}c(u,0)f(u)|\Phi(u;y)-\Phi'(u;y)|\mathrm{d}\mu(u)\\ &\le g(y)\|fc(\cdot,0)\|_{L^p(\mu)}\|\Phi(\cdot;y)-\Phi'(\cdot;y)\|_{L^q(\mu)}.\end{aligned}$$

通过对 T_2 应用类似的论证，我们可以写出

$$\begin{aligned}Z(y)Z'(y)T_2 &\le |Z(y)-Z'(y)|g(y)\int_{\mathcal{X}}c(u,0)f(u)\mathrm{d}\mu(u)\\ &\le |Z(y)-Z'(y)|g(y)\|fc(\cdot,0)\|_{L^p(\mu)}.\end{aligned}$$

[3]给定 a， $b \in \mathbb{R}$ 我们用 $a \vee b$ 表示它们的最大值，用 $a \wedge b$ 表示它们的最小值。

接下来，我们对 $|Z(y) - Z'(y)|$ 进行界定。

$$\begin{aligned}
|Z(y) - Z'(y)| &= \left| \int_{\mathcal{X}} \exp(-\Phi(u;y)) - \exp(-\Phi'(u;y)) \mathrm{d}\mu(u) \right| \\
&\leq \left| \int_{\mathcal{X}} \exp\left(-\left[\Phi(u;y) \wedge \Phi'(u;y)\right]\right) |\Phi(u;y) - \Phi'(u;y)| \mathrm{d}\mu \right| \\
&\leq g(y) \left| \int_{\mathcal{X}} f(u) |\Phi(u;y) - \Phi'(u;y)| \mathrm{d}\mu(u) \right| \\
&\leq g(y) \|f\|_{L^p(\mu)} \|\Phi(\cdot;y) - \Phi'(\cdot;y)\|_{L^q(\mu)} .
\end{aligned}$$

将此界与我们此前关于 T_2 的界结合起来，得到：

$$T_2 \leq \frac{g^2(y)}{Z(y)Z'(y)} \|f\|_{L^p(\mu)} \|fc(\cdot,0)\|_{L^p(\mu)} \|\Phi(\cdot;y) - \Phi'(\cdot;y)\|_{L^q(\mu)}.$$

结合关于 T_1、T_2 的界，我们便得到

$$\mathcal{D}(\nu,\nu';c) \leq \frac{g(y)\|fc(\cdot,0)\|_{L^p(\mu)} \left[Z(y) + g(y)\|f\|_{L^p(\mu)} \right]}{Z(y)Z'(y)} \|\Phi(\cdot;y) - \Phi'(\cdot;y)\|_{L^q(\mu)}. \tag{3.2}$$

然而，由于 Φ 满足假设 1（ii），我们有

$$Z(y) \leq g(y) \int_{\mathcal{X}} f(u) \mathrm{d}\mu(u) \leq g(y)\|f\|_{L^1(\mu)} \leq g(y)\|f\|_{L^p(\mu)}, \tag{3.3}$$

并且由于假设 1（ii）也由 Φ' 满足，我们得到下界

$$Z(y)Z'(y) \geq \left(\int_{\mathcal{X}} h(u,y) \mathrm{d}\mu(u) \right)^2 = \|h(\cdot,y)\|^2_{L^1(\mu)}.$$

将这些界代回（3.2）即可得到所需结果。 □

上述证明直接得出以下推论。

Corollary 3.4. 定理 *3.2* 的结论在界 (3.2)下同样成立，该界由常数 $Z(y)$和 $Z'(y)$ 显式表示。

注 3.5. 方程（3.2）还揭示了这些界的一个微妙性质：当 $Z(y)$或 $Z'(y)$ 趋于零时，这些界中的 Lipschitz 常数会趋于无穷大。文献 [43] 也观察到了这一现象，它体现了贝叶斯逆问题的脆弱性 [37,35, 36], ：也就是说，如果模型设定有误，具体表现为似然集中在先验概率较低的区域，那么后验测度就会对似然、数据和先验的扰动非常敏感。

我们还注意到，定理 3.2 还通过在上界中采用 $L^q(\mu)$-范数，扩展了以往的研究，而不是采用混合的 $L^1(\mu)$ 和 $L^2(\mu)$ 范数，见[43]。尤其是，L^∞-情形为研究神经网络替代模型所产生的扰动打开了大门，例如文献 [15] 中研究并在第 4.4 节讨论的这类模型。

3.2.1. 数据扰动。 本节讨论定理3.2 的一些推论，这些推论与以下应用相关：似然势中的扰动 Φ 由数据扰动 y引起。证明后验分布对数据 y 的连续依赖性，是贝叶斯逆问题[46, 18, 16, 47] 适定性理论中的一个核心问题，旨在确立数据到后验映射的连续性。

为此，我们考虑 y，$y' \in \mathcal{Y}$，将 y' 视为 y的扰动。然后应用定理 3.2，取 $\Phi'(u;\ y) \equiv \Phi(u;\ y + (y' - y)$ 可立即得到以下结果：

Corollary 3.6. 设 Φ 满足假设 *1*，并令 p，$q \in [1,\infty]$ 满足$1/p + 1/q = 1$。对于固定的 $y \in \mathcal{Y}$和 $r > 0$，考虑后验测度

$$\frac{d\nu'}{d\mu}(u) = \frac{1}{Z(y')} \exp(-\Phi(u; y')), \qquad \forall y' \in B_r^{\mathcal{Y}}(y),$$

其中，$B_r^{\mathcal{Y}}(y) \subset \mathrm{Y}$ 表示以 y 为中心、半径为 r 的球。于是有

$$\mathcal{D}(\nu, \nu'; c) \le \frac{2\left(\sup_{z \in B_r^{\mathcal{Y}}(y)} g^2(z)\right) \|fc(\cdot, 0)\|_{L^p(\mu)} \|f\|_{L^p(\mu)}}{\|\inf_{z \in B_r^{\mathcal{Y}}(y)} h(\cdot, z)\|_{L^1(\mu)}^2} \|\Phi(\cdot; y) - \Phi(\cdot; y')\|_{L^q(\mu)}.$$

在许多应用中，似然势 Φ 满足更强的正则性约束，使我们能够根据数据扰动的大小来控制 $\mathrm{D}(\nu, \nu'; c)$。

Corollary 3.7. 假设推论 *3.6* 中的条件成立，且另外，$\forall y' \in B_r^{\mathcal{Y}}(y)$ 存在一个函数 $b : \mathcal{X} \to \mathbb{R}_{\ge 0}$ 使得对于 μ-几乎处处 u，

$$|\Phi(u; y) - \Phi(u; y')| \le b(u) \|y - y'\|_{\mathcal{Y}}, \qquad \forall y' \in B_r^{\mathcal{Y}}(y).$$

于是有

$$\mathcal{D}(\nu, \nu'; c) \le \frac{2\left(\sup_{z \in B_r^{\mathcal{Y}}(y)} g^2(z)\right) \|fc(\cdot, 0)\|_{L^p(\mu)} \|f\|_{L^p(\mu)} \|b\|_{L^q(\mu)}}{\|\inf_{z \in B_r^{\mathcal{Y}}(y)} h(\cdot, z)\|_{L^1(\mu)}^2} \|y - y'\|_{\mathcal{Y}}.$$

特别地，上述结果表明，若 $\|\inf_{z \in B_r(y)} h(\cdot, z)\|_{L^1(\mu)} > 0$, $f, fc(\cdot, 0) \in L^p(\mu)$，且 $b \in L^q(\mu)$，则映射 $y \mapsto \nu$ 关于 D 局部利普希茨连续。

3.3. 先验扰动。这里我们将注意力转向后验分布相对于先验测度 μ 的扰动性质 ν。更确切地说，我们考虑先验测度 $\mu_* \in \mathbb{P}(\mathrm{X})$，并定义后验测度 $\nu_* \in \mathbb{P}(\mathrm{X})$:

$$\frac{\mathrm{d}\nu_*}{\mathrm{d}\mu_*} = \frac{1}{Z_*(y)} \exp\left(-\Phi(u; y)\right). \tag{3.4}$$

我们再次在这一一般设定下发展扰动理论，随后在各种应用的背景下具体阐述我们的主要定理。与似然扰动相比，先验扰动的主要区别在于，这里我们需要针对定义 D 所依据的函数 c 进行调整，以界定后验测度之间的距离。更确切地说，应当针对通过相关类距离函数定义的不同积分概率度量，分别度量先验扰动和后验扰动。

Theorem 3.8. 假设 Φ 满足假设 *1*，此外还存在一个$L : \mathcal{X} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_{\ge 0}$，使得

$$|\Phi(u; y) - \Phi(v; y)| \le L(u, v; y) c(u, v)$$

对于所有 u、$v \in \mathrm{X} \times \mathrm{X}$ 以及 $y \in \mathrm{Y}$，且对于任意固定的 y 值，$L(\cdot, \cdot; y)$ 都是下半连续的。则有

$$\mathcal{D}(\nu, \nu_*; c) \le \frac{g^2(y)\left[\|f\|_{L^1(\mu)} + \|fc(\cdot, 0)\|_{L^1(\mu)}\right]}{\|h(\cdot, y)\|_{L^1(\mu)} \|h(\cdot, y)\|_{L^1(\mu_*)}} \mathcal{D}(\mu, \mu_*; c_y),$$

其中，新的类距离函数 $c_y : \mathbf{X} \times \mathbf{X} \to \mathbb{R}_{\geq 0}$ 定义为

$$c_y(u,v) := \big[1 \vee c(u,0) \vee c(v,0)\big] \cdot \big[f(u) \vee f(v)\big] \cdot \big[1 \vee L(u,v;y)\big] \cdot c(u,v). \tag{3.5}$$

注释3.9。我们强调一下关于新代价函数 c_y 的两个重要事实：(i) 下标 y 表示代价通过 Φ 的局部利普希茨常数 $L(\cdot,\cdot;y)$ 依赖于数据 y。当然，在先验扰动的情况下，这种对 y 的依赖是无关紧要的，因为假设数据是固定的。或者，可以用 $\sup_{y \in B_r^{\mathcal{Y}}(0)} L(\cdot,\cdot;y)$ 替换 $L(\cdot,\cdot;y)$，从而得到一个类似的结果，其代价对于所有 $y \in B_r^{\mathcal{Y}}(0)$ 都是一致的。事实上，任何上界为 $c_y(u,v)$ 的代价也都满足该结果。(ii) 函数 f 以及 L 出现在 c_y 的定义中，表明新代价 c_y 适用于似然 Φ。注意，f 控制着 $\exp(-\Phi(u;y))$ 的增长速率，因此由表达式 (3.5) 可清楚看出，c_y 的增长速率也与似然及其利普希茨常数和原始代价 c 的增长速率相关。直观地说，似然增长得越慢且越规整，c_y 就越接近 c。

证明。根据 D 的定义，我们可以写成

$$\begin{aligned}
\mathcal{D}(\nu,\nu_*;c) &= \sup_{\psi \in \mathrm{Lip}_1^0(\mathcal{X};c)} \frac{1}{Z}\int_{\mathcal{X}} \psi(u)\exp(-\Phi(u;y))\mathrm{d}\mu(u) \\
&\qquad\qquad - \frac{1}{Z_*}\int_{\mathcal{X}} \psi(u)\exp(-\Phi(u;y))\mathrm{d}\mu_*(u) \\
&\leq \left[\frac{1}{Z_*}\sup_{\psi \in \mathrm{Lip}_1^0(\mathcal{X};c)} \int_{\mathcal{X}} \psi(u)\exp(-\Phi(u;y))\mathrm{d}\mu(u)\right. \\
&\qquad\qquad \left. - \int_{\mathcal{X}} \psi(u)\exp(-\Phi(u;y))\mathrm{d}\mu_*(u)\right] \\
&\quad + \left[\frac{|Z - Z_*|}{Z Z_*}\sup_{\psi \in \mathrm{Lip}_1^0(\mathcal{X};c)} \int_{\mathcal{X}} \psi(u)\exp(-\Phi(u;y))\mathrm{d}\mu(u)\right] \\
&=: T_1 + T_2.
\end{aligned}$$

我们首先对 T_1 进行界定。对于任意 $\psi \in \mathrm{Lip}_1^0(\mathbf{X};\ c)$ 以及任意 u、$v \in \mathbf{X}$，有

$$\begin{aligned}
&|\psi(u)\exp(-\Phi(u;y)) - \psi(v)\exp(-\Phi(v;y))| \\
&\quad \leq |\psi(u)||\exp(-\Phi(u;y)) - \exp(-\Phi(v;y))| + \exp(-\Phi(u;y))|\psi(u;y) - \psi(v;y)| \\
&\quad \leq c(u,0)|\exp(-\Phi(u;y)) - \exp(-\Phi(v;y))| + \exp(-\Phi(u;y))c(u,v).
\end{aligned}$$

现在应用指数函数的中值定理，以及假设 1(ii) 和关于 $\Phi(u;\ y)$ 的局部利普希茨假设，我们可以进一步写出

$$\begin{aligned}
&|\psi(u)\exp(-\Phi(u;y)) - \psi(v)\exp(-\Phi(v;y))| \\
&\quad \leq c(u,0)\exp\left(-[\Phi(u;y) \wedge \Phi(v;y)]\right)|\Phi(u;y) - \Phi(v;y)| + \exp(-\Phi(u;y))c(u,v) \\
&\quad \leq g(y)c(u,0)\left[f(u) \vee f(v)\right]L(u,v;y)c(u,v) + g(y)f(u)c(u,v) \\
&\quad \leq g(y)\Big[\big[1 \vee c(u,0) \vee c(v,0)\big]\big[f(u) \vee f(v)\big]\big[1 \vee L(u,v;y)\big]\Big]c(u,v) \\
&\quad = g(y)c_y(u,v).
\end{aligned}$$

因此， $g(y)^{-1}\psi(\cdot)\exp(-\Phi(\cdot;\,y)) \in \mathrm{Lip}_1(\mathrm{X};\,c_y)$，由此立即得到界

$$Z_*(y)T_1 \le g(y)\mathcal{D}(\mu,\mu_*;c_y).$$

接下来对 T_2进行界定时，我们再次使用假设 1(ii)，得到

$$\begin{aligned} Z(y)Z_*(y)T_2 &\le \left[g(y)\int_{\mathcal{X}} c(u,0)f(u)\mathrm{d}\mu(u)\right]|Z(y)-Z_*(y)| \\ &= g(y)\|fc(\cdot,0)\|_{L_1(\mu)}|Z(y)-Z_*(y)|. \end{aligned}$$

此外，根据假设 1(ii)、关于 Φ 的 Lipschitz 假设以及 $c_y(u,v)$ 的定义，我们可以直接验证 $g(y)^{-1}\exp(-\Phi(\cdot;\,y)) \in \mathrm{Lip}_1(\mathcal{X};\,c_y)$ 因此

$$|Z(y)-Z_*(y)| \le g(y)\mathcal{D}(\mu,\mu_*;c_y),$$

由此可得最终的界

$$T_2 \le \frac{g^2(y)\|fc(\cdot,0)\|_{L^1(\mu)}}{Z(y)Z_*(y)}\mathcal{D}(\mu,\mu_*;c_y).$$

结合对 T_1、T_2 的界，进而得到

$$\mathcal{D}(\nu,\nu_*;c) \le \frac{g(y)\left[Z(y)+g(y)\|fc(\cdot,0)\|_{L^1(\mu)}\right]}{Z(y)Z_*(y)}\mathcal{D}(\mu,\mu_*;c_y)\,. \tag{3.6}$$

根据假设 1(ii) 以及 Z、Z_* 的定义，我们得到下界

$$Z(y)Z_*(y) \ge \left(\int_{\mathcal{X}} h(u,y)\mathrm{d}\mu(u)\right)\left(\int_{\mathcal{X}} h(u,y)\mathrm{d}\mu(u)\right) = \|h(\cdot,y)\|_{L^1(\mu)}\|h(\cdot,y)\|_{L^1(\mu_*)}.$$

将此界和（3.3）代回（3.6），证明即告完成。 □

类似于似然扰动的情形，我们给出定理3.8的一个对应结论，其界以模型证据 Z（y）和 Z_*（y）表示，并突出了与备注3.5中讨论的类似现象，即当 Z（y）和 Z_*（y）趋于零时，我们的界会发散。这表明，在模型设定错误的情况下，后验分布对先验扰动具有敏感性。

Corollary 3.10. 定理*3.8* 的结果对于界（3.6）同样成立，该界以常数 Z（y）和 Z_*（y）表示。

4. 应用。 本节将定理 3.2 和 3.8 应用于不同的情形，以展示其广泛的适用性，并说明如何以不同方式利用这些结果，获得各种情况下的定量收敛速率；在这些情况下，后验分布会因先验或似然项的变化而发生扰动。在第 4.1小节中，我们考虑一个典型的逆问题，旨在量化这样一种贝叶斯后验分布中的扰动：先验通过使用有限个样本得到的经验测度进行近似。在第 4.2小节中，我们考虑使用 Mat´ern 核的高斯过程回归，这是一个针对无限维参数空间的逆问题。在第 4.3小节中，我们考虑这样产生的先验扰动：将推前映射扰动到一个固定的参考测度；例如，这种设置在生成模型中十分常见。最后，第 4.4 小节讨论用于贝叶斯反演的深度神经网络替代模型。

4.1. 经验先验回归。考虑一个前向模型为 $y = G(u) + \xi$ 的回归问题，其中 $G : \mathbb{R}^d \to \mathbb{R}^d$ 是映射 $(u_1, \dots, u_d) \to (\tanh(u_1) \dots, \tanh(u_d))$，$y \in \mathbb{R}^d$ 且 $\xi \sim \mathsf{N}(0, \sigma^2 I_d)$。该模型导出的似然势为

$$\Phi(u; y) = \frac{1}{2\sigma^2}|G(u) - y|^2,$$

并且我们可以验证，该势函数满足假设 1，其中 $f = g = 1$ and$h(u, y) = \exp(-(d + |y|^2)/\sigma^2)$，因为 $|G(u)| \le \sqrt{d}$。现在考虑定义在 u 上的先验分布 μ，以及由 N 个独立同分布样本得到的 μ 的经验近似 μ_N。我们的目标是控制由此得到的真实后验分布 ν 与后验分布 ν_N 之间的误差：前者由先验分布 μ 和贝叶斯公式 (1.1) 定义，后者则由 μ_N 得出。

令代价函数 $c(u, v) = |u - v|$，此时 $\mathsf{D}(\cdot, \cdot; c)$ 与一阶瓦瑟斯坦距离 $\mathsf{W}_1(\cdot, \cdot)$ 一致。注意，$G \in \mathsf{Lip}_1(\mathbb{R}^d; c)$，因此对于任意 $y \in \mathsf{Y}$，$\Phi(\cdot; y)$ 都是全局利普希茨连续的，其利普希茨常数为 $L(y) = (\sqrt{d} + |y|)/\sigma^2$。现在我们可以选择

$$c_y(u, v) = [1 \vee L(y)][1 \vee |u| \vee |v|]\,|u - v|.$$

并应用定理 3.8以及恒等式 (2.2) 得到

$$\mathcal{W}_1(\nu, \nu_N) \le \frac{1 + \mu(|\cdot|)}{\exp\left(-\frac{2}{\sigma^2}(d + |y|^2)\right)}\mathcal{W}(\mu, \mu_N; c_y).$$

利用柯西–施瓦茨不等式，对于任意 $\pi \in \Pi(\mu, \mu_N)$，可写为

$$\begin{aligned}
\pi(c_y(u, v)) &\le L(y)\left(\pi\left([1 \vee |u| \vee |v|]^2\right)\right)^{1/2}\left(\pi((|u - v|)^2\right)^{1/2} \\
&\le L(y)\left(\pi\left(1 + |u|^2 + |v|^2\right)\right)^{1/2}\left(\pi((|u - v|)^2\right)^{1/2} \\
&= L(y)\left(1 + \mu(|\cdot|^2) + \mu_N(|\cdot|^2)\right)^{1/2}\left(\pi((|u - v|)^2\right)^{1/2}.
\end{aligned}$$

对 π 取下确界，我们得到

$$\mathcal{W}(\mu, \mu_N; c_y) \le L(y)\left(1 + \mu(|\cdot|^2) + \mu_N(|\cdot|^2)\right)^{1/2}\mathcal{W}_2(\mu, \mu_N),$$

其中 W_2 表示二阶瓦瑟斯坦距离。由此得到以下界

$$\mathcal{W}_1(\nu, \nu_N) \le \frac{L(y)\,[1 + \mu(|\cdot|)] \cdot \left(1 + \mu(|\cdot|^2) + \mu_N(|\cdot|^2)\right)^{1/2}}{\exp\left(-\frac{2}{\sigma^2}(d + |y|^2)\right)}\mathcal{W}_2(\mu, \mu_N). \tag{4.1}$$

现在，针对生成 μ_N 的 i.i.d. 经验样本，对上述表达式取期望，可得

$$\begin{aligned}
\mathbb{E}\mathcal{W}_1(\nu, \nu_N) &\le \mathbb{E}\left[\frac{L(y)\,[1 + \mu(|\cdot|)] \cdot \left(1 + \mu(|\cdot|^2) + \mu_N(|\cdot|^2)\right)^{1/2}}{\exp\left(-\frac{2}{\sigma^2}(d + |y|^2)\right)}\mathcal{W}_2(\mu, \mu_N)\right] \\
&\le \frac{L(y)(1 + \mu(|\cdot|))}{\exp\left(-\frac{2}{\sigma^2}(d + |y|^2)\right)}\left[1 + \mu(|\cdot|^2) + \mathbb{E}\mu_N(|\cdot|^2)\right]^{1/2}\left[\mathbb{E}\mathcal{W}_2^2(\mu, \mu_N)\right]^{1/2}.
\end{aligned}$$

现在可以应用 [10, 定理 1]，该定理为 $\mathbb{E}\mathsf{W}_2^2(\mu, \mu_N)$给出了精确的速率。特别地 ,

假设 $\mu(|\cdot|^3) < +\infty$，我们得到界

$$\left[\mathbb{E}\mathcal{W}_1(\nu,\nu^N)\right]^2 \le C(\mu)\frac{(\sqrt{d}+|y|)^2\exp\left[\frac{4}{\sigma^2}\left(d+|y|^2\right)\right]}{\sigma^4}\times\begin{cases}N^{-1/2}+N^{-1/3} & \text{if } d<4,\\ N^{-1/2}\log(1+N)+N^{-1/3} & \text{if } d=4,\\ N^{-2/d}+N^{-1/3} & \text{if } d>4,\end{cases}$$

其中 $C(\mu) < +\infty$ 是一个取决于 μ前三个矩的常数。

4.2. Mat´ern 核的高斯过程回归。这里我们考虑一个非参数回归问题，其高斯过程先验属于 Mat´ern 类。后验分布的扰动由定义该高斯过程先验的超参数控制。令 $\Omega = [0,1]^d$，并令 $\kappa_i \in C_c^\infty(\Omega)$ 对于 $i = 1, \ldots, m$ 为固定函数（滤波器），并定义前向映射

$$G : u \mapsto (\langle\kappa_1, u\rangle, \ldots, \langle\kappa_m, u\rangle)^\top,$$

对于函数 $u \in L^2(\Omega)$，并用 $\langle\cdot,\cdot\rangle$ 表示 L^2内积。我们再次考虑模型 $y = G(u) + \xi$，其中 $\xi \sim \mathsf{N}(0, \sigma^2 I_m)$，由此得到似然

$$\Phi(u; y) = \frac{1}{2\sigma^2}|G(u) - y|^2.$$

我们取 $\mathsf{X} = L^2(\Omega)$，并注意到 $G : \mathsf{X} \to \mathbb{R}^m$ 有界且线性。先验则选取

$$\mu = \mathsf{N}(0, \gamma^2(\Delta + \tau\,\mathrm{id})^{-2\alpha}),$$

其中，Δ 和 id 分别表示 X 上的拉普拉斯算子和恒等算子，γ、$\tau > 0$ 为常数，$\alpha > 0$ 为整数。上述高斯先验在 Ω 上定义了一个高斯场，其协方差函数为 Mat´ern 协方差函数 [51]。在实践中，超参数γ、τ通常通过经验贝叶斯、最大似然或交叉验证 [52] 等各种技术进行调优。参数 α 也可以视为超参数，但为方便起见，此处将其视为固定参数。本文不讨论超参数调优策略，而是希望控制这些超参数取不同值时所得后验分布之间的距离。为此，设 $\gamma_\star$、$\tau_\star$ 为第二组超参数，其对应的 Mat´ern 先验为 $\mu_\star$，并令 ν 和$\nu_\star$ 表示分别由 μ 和 $\mu_\star$ 产生的后验分布，其中使用上述相同的似然势。下面的目标是界定 $\mathsf{W}_1(\nu, \nu_\star)$。

由于 G 有界且线性，我们可以验证 Φ 满足假设 1，其中 $f = g = 1$ 且 $h(u, y) = \exp(-\frac{1}{\sigma^2}\left[\|G\| \cdot \|u\|_{\mathcal{X}}^2 + |y|^2\right])$；这里 $\|G\|$ 表示 G的算子范数。根据 Fernique 定理，函数 h 在任意高斯先验下均可积。现在令 $c(u, v) = \|u - v\|_{\mathcal{X}}$。通过直接计算，我们可以验证Φ 满足定理 3.8 的假设，其中 $L(u, v; y) = \frac{\|G\|}{\sigma^2}(\|u\|_{\mathcal{X}} + \|v\|_{\mathcal{X}} + |y|)$。因此，根据 Fernique 定理和定理 3.8 可得 $\mathsf{W}_1(\nu, \nu_\star) \le C\mathsf{D}(\mu, \mu_\star; c_y)$，其中我们可以取

$$c_y(u, v) = [1 \vee \|u\|_{\mathcal{X}} \vee \|v\|_{\mathcal{X}}] \cdot [1 \vee L(u, v; y)]\|u - v\|_{\mathcal{X}}.$$

我们可以将来自 — $L\ u\ v\ y$ C $_1\ \nu\ \nu_\star\ \le(,;)$ 的系数 $\frac{\|G\|}{\sigma^2}$ 并入常数，C 得到 $\mathsf{W}(,)\,\mathsf{D}(\mu,\mu_\star;\ c_y'),$ $u\ v$ 其中 $c_y'(,)\,(1 + \|u\|_{\mathcal{X}} + \|v\|_{\mathcal{X}} + |y|)^2\|u - v\|_{\mathcal{X}}$。对于固定的 $y \in \mathsf{Y}$，这

个代价等价于 c_y，但使用后者更方便。因此，我们接下来需要对 $\mathsf{D}(\mu,\mu_\star;\, c_y')$ 进行界定，并通过耦合论证来完成。

由于 μ、$\mu_\star$ 是高斯测度，它们具有 Karhunen-Lo´eve 展开

$$\mu = \mathrm{Law}\left\{\sum_{j=1}^{\infty}\sqrt{\lambda_j}\xi_j x_k\right\}, \qquad \xi_j \overset{i.i.d.}{\sim} \mathsf{N}(0,1),$$
$$\mu_\star = \mathrm{Law}\left\{\sum_{j=1}^{\infty}\sqrt{\lambda_j^\star}\xi_j x_k\right\}, \qquad \xi_j \overset{i.i.d.}{\sim} \mathsf{N}(0,1),$$

其中 x_j 是 Ω 上 Δ 的特征函数，并且

$$\lambda_j = \gamma^2\left(\tilde{\lambda}_j+\tau\right)^{-2\alpha}, \qquad \lambda_j^\star = \gamma_\star^2\left(\tilde{\lambda}_j+\tau_\star\right)^{-2\alpha},$$

˜其中 λ_j 表示 Δ 的特征值。利用下面的 4.1 引理，我们得到界

$$\mathcal{D}(\mu,\mu_\star;c_y') \le C\sum_{j=1}^{\infty}\left|\sqrt{\lambda_j}-\sqrt{\lambda_j^\star}\right|^2, \tag{4.2}$$

对于某个常数 $C>0$，其取决于 μ 和 $\mu_\star$ 的矩。现在，通过代数运算即可将误差控制在超参数方面。我们可以写成

$$\left|\sqrt{\lambda_j}-\sqrt{\lambda_j^\star}\right| \le (\tilde{\lambda}_j+\tau)^{-\alpha}|\gamma-\gamma_\star| + \gamma_\star\left|\left(\tilde{\lambda}_j+\tau\right)^{-\alpha}-\left(\tilde{\lambda}_j+\tau_\star\right)^{-\alpha}\right|.$$

我们还可以利用中值定理进一步估计第二项，得到

$$\left|\sqrt{\lambda_j}-\sqrt{\lambda_j^\star}\right| \le (\tilde{\lambda}_j+\tau)^{-\alpha}|\gamma-\gamma_\star| + \alpha\gamma_\star\left((\tilde{\lambda}_j+\tau)^{-\alpha-1}\vee(\tilde{\lambda}_j+\tau_\star)^{-\alpha-1}\right)|\tau-\tau_\star|.$$

代入（4.2）可得界

$$[\mathcal{W}_1(\nu,\nu^\star)]^2 \le C\left(|\gamma-\gamma_\star|+|\tau-\tau_\star|\right)^2,$$

˜其中，常数 $C>0$ 取决于 α、γ、$\gamma_\star$、τ、$\tau_\star$ 以及 λ_j 的选择；虽然可以显式推导出该常数，但为简洁起见，我们不采用这一路径。

现给出导出上述界（4.2）的技术引理，该引理本身也颇具独立意义。该结果为利用简单的耦合论证，对广泛一类代价函数 $\mathcal{W}(\cdot,\cdot;c)$ 下乘积测度之间的距离进行界定，提供了一种通用且便捷的方法 c。

Lemma 4.1. 设 $\mathbf{X}$ 为一个可分希尔伯特空间，且 μ、$\mu_* \in \mathbb{P}(\mathbf{X})$ 为如下形式的乘积测度

$$\mu = \mathrm{Law}\left\{\sum_{j=1}^{\infty}\sqrt{\lambda_j}\xi_j x_j\right\}, \quad \xi_j \overset{i.i.d.}{\sim} \eta,$$
$$\mu_* = \mathrm{Law}\left\{\sum_{j=1}^{\infty}\sqrt{\lambda_j^*}\xi_j^* x_j^*\right\}, \quad \xi_j^* \overset{i.i.d.}{\sim} \eta_*,$$

其中 $\{\lambda_j\}_{j=1}^\infty$ 和 $\{\lambda_j^*\}_{j=1}^\infty$ 是固定的实值序列，$\{x_j\}_{j=1}^\infty$ 和 $\{x_j^*\}_{j=1}^\infty$ 是 $\mathbf{X}$ 中的标准正交基，且 η，$\eta_* \in \mathbb{P}(\mathbb{R})$ 均值为零且具有单位方差。假设

两个随机和都几乎处处收敛于 X，因此 μ 和 μ_* 有定义，并将$c:\mathrm{X}\times\mathrm{X}\to\mathbb{R}_{\geq 0}$ 定义为如下形式的下半连续半度量

$$c(u,v)=\Big(\|u\|_{\mathcal{X}}+\|v\|_{\mathcal{X}}\Big)^{s}\|u-v\|_{\mathcal{X}},$$

其中 $s\geq 0$。那么有

$$\begin{aligned}\mathcal{D}(\mu,\mu_*;c)\leq{}& 2^{\frac{1\vee(2s-1)}{2}+1}\left(\mathbb{E}_{\mu}\|u\|_{\mathcal{X}}^{2s}+\mathbb{E}_{\mu_*}\|u_*\|_{\mathcal{X}}^{2s}\right)^{\frac{1}{2}}\\&\times\left(\mathcal{W}_2^2(\eta,\eta_*)\sum_{j=1}^{\infty}\lambda_j+\sum_{j=1}^{\infty}\lambda_j\|x_j-x_j^*\|_{\mathcal{X}}^2+\sum_{j=1}^{\infty}\left(\sqrt{\lambda_j}-\sqrt{\lambda_j^*}\right)^2\right)^{\frac{1}{2}}.\end{aligned}$$

证明。设 ϖ_0 是 η、η_* 之间达到 $\mathrm{W}_2(\eta,\eta_*)$ 的最优耦合。若这样的耦合不存在，则令 $\mathrm{W}_2(\eta,\eta_*)=+\infty$，此时该界显然成立，因此下文假设 $\mathrm{W}_2(\eta,\eta_*)<+\infty$。现在令 π_0 为 μ 和 μ_* 之间的一个耦合，其定义为随机变量 $(u,u_*)\in\mathrm{X}\times\mathrm{X}$ 的分布，构造如下：

$$u=\sum_{j=1}^{\infty}\sqrt{\lambda_j}\xi_j x_j,\qquad u_*=\sum_{j=1}^{\infty}\sqrt{\lambda_j^*}\xi_j^* x_j^*,\qquad (\xi_j,\xi_j^*)\overset{i.i.d.}{\sim}\varpi_0.$$

根据对偶性（2.3），我们有

$$\begin{aligned}\mathcal{D}(\mu,\mu_*;c)&\leq\int_{\mathcal{X}\times\mathcal{X}}c(u,u_*)\mathrm{d}\pi_0(u,u_*)\\&=\mathbb{E}_{\pi_0}c\left(\sum_{j=1}^{\infty}\sqrt{\lambda_j}\xi_j x_j,\sum_{j=1}^{\infty}\sqrt{\lambda_j^*}\xi_j^* x_j^*\right)\\&\leq 2^{\frac{1\vee(2s-1)}{2}}\left(\mathbb{E}_{\mu}\|u\|_{\mathcal{X}}^{2s}+\mathbb{E}_{\mu_*}\|u_*\|_{\mathcal{X}}^{2s}\right)^{1/2}\left(\mathbb{E}_{\pi_0}\|u-u_*\|_{\mathcal{X}}^2\right)^{1/2}.\end{aligned}$$

此外，根据帕塞瓦尔恒等式，我们有

$$\begin{aligned}\mathbb{E}_{\pi_0}\|u-u_*\|_{\mathcal{X}}^2&=\mathbb{E}_{\pi_0}\left\|\sum_{j=1}^{\infty}\sqrt{\lambda_j}\xi_j x_j-\sqrt{\lambda_j^*}\xi_j^* x_j^*\right\|_{\mathcal{X}}^2\\&\leq 4\,\mathbb{E}_{\pi_0}\left[\sum_{j=1}^{\infty}(\xi_j-\xi_j^*)^2\lambda_j+\left\|\sum_{j=1}^{\infty}\xi_j^*\sqrt{\lambda_j}\left(x_j-x_j^*\right)\right\|_{\mathcal{X}}^2+\sum_{j=1}^{\infty}(\xi_j^*)^2\left(\sqrt{\lambda_j}-\sqrt{\lambda_j^*}\right)^2\right]\\&=4\left(\sum_{j=1}^{\infty}\mathbb{E}_{\varpi_0}|\xi_j-\xi_j^*|^2\lambda_j+\mathbb{E}_{\eta_*}\left\|\sum_{j=1}^{\infty}\xi_j^*\sqrt{\lambda_j}\left(x_j-x_j^*\right)\right\|_{\mathcal{X}}^2+\sum_{j=1}^{\infty}(\sqrt{\lambda_j}-\sqrt{\lambda_j^*})^2\right)\\&\leq 4\left(\mathcal{W}_2^2(\eta,\eta_*)\sum_{j=1}^{\infty}\lambda_j+\mathbb{E}_{\eta_*}\left(\sum_{j=1}^{\infty}\xi_j^*\sqrt{\lambda_j}\left\|x_j-x_j^*\right\|_{\mathcal{X}}\right)^2+\sum_{j=1}^{\infty}(\sqrt{\lambda_j}-\sqrt{\lambda_j^*})^2\right)\\&=4\left(\mathcal{W}_2^2(\eta,\eta_*)\sum_{j=1}^{\infty}\lambda_j+\sum_{j=1}^{\infty}\lambda_j\left\|x_j-x_j^*\right\|_{\mathcal{X}}^2+\sum_{j=1}^{\infty}(\sqrt{\lambda_j}-\sqrt{\lambda_j^*})^2\right).\end{aligned}$$

将其代回 D 的上界，即得所需结果。 □

4.3. 推前先验。下面我们考虑一个抽象示例，其中先验测度 μ 被确定为某个固定参考测度的推前测度，而μ_* 则是同一参考测度经过扰动后的推前测度

映射。这一问题是数据驱动技术的核心，这类技术旨在从数据中学习先验，近来颇受欢迎 [38, 11, 23] ，这得益于生成式机器学习模型（如生成对抗网络（GANs））的巨大成功[12] 以及归一化流 [20]。

设 $\mathbf{X} = [0,1]^d$，再次考虑具有线性前向映射 $G : \mathbf{X} \to \mathbb{R}^m$ 的逆问题，由此得到似然势

$$\Phi(u;y) = \frac{1}{2\sigma^2}|G(u) - y|^2.$$

注意，与第4.2节中一样， Φ 满足假设 1，其中 $f \equiv g \equiv 1$ 和 $h(u,y) = \exp\left(-\frac{1}{\sigma^2}\left[\|G\|_\infty^2 \cdot |u|^2 + |y|^2\right]\right)$。事实上，由于 $\mathbf{X}$ 是紧的，我们可以推知， Φ 的利普希茨常数可取为 $L(u,v;y) = C(1+|y|)$ ，其中 $C > 0$ 为某个常数。然后，重复导出 (4.1) 的相同计算可知，在此例中该界无需任何修改即可成立。

现在假设 $\mu = T_\sharp \varrho$，即 ϱ 通过映射 $T : \mathbf{X} \to \mathbf{X}$ 的推前测度，其中 $\varrho \in \mathbb{P}(\mathbf{X})$ 是参考测度。类似地，设 $\mu_* = T_\sharp^* \varrho$，其中 $T^* : \mathbf{X} \to \mathbf{X}$ 是第二个传输映射。在上述应用中，我们将 T^* 视为 T 的近似。因此，自然会尝试用 T 与 T^* 之间的距离来界定 $\mathrm{W}_1(\nu,\nu_*)$。要得到这样的结果，关键在于依据映射之间的 L^p 距离给出Wasserstein距离的稳定性估计。我们回顾以下结果：

Proposition 4.2（[39, 定理 2]）。设 $\mathcal{X} = [0,1]^d$ 并假设 T， $T^* \in L^p(\varrho) \cap C(\mathcal{X})$ 对于某个 $p \geq 1$成立。那么 $\mathcal{W}_p(\mu, \mu_*) \leq \|T - T^*\|_{L^p(\varrho)}$.

将上述结果与 $p = 2$ 应用于界（4.1），即可立即得到

$$\mathcal{W}_1(\nu,\nu_*) \leq \frac{C(1+|y|)(1+\mu(|\cdot|))}{\|h(\cdot,y)\|_{L^1(\mu)}\|h(\cdot,y)\|_{L^1(\mu_*)}}\left(1+\mu(|\cdot|^2)+\mu_*(|\cdot|^2)\right)^{1/2}\|T-T^*\|_{L^2(\varrho)}.$$

量 $\|T - T^*\|_{L^2(\varrho)}$ 可利用现成的逼近理论结果进行控制。例如，可以令 T^* 为多项式 [9], 小波 [29],神经网络 [8], 或核逼近 [50] ，以逼近 T ；在正则性充分的假设下，可得到定量收敛速率。

4.4. DNN替代模型。近年来，偏微分方程的贝叶斯反演在统计学中受到广泛关注，尤其是在以随机场输入进行不确定性量化的数值处理领域。这是由科学与工程领域对大规模计算求解的需求所推动的。输入数据的随机级数展开会产生一个无限维参数化贝叶斯逆问题，针对该问题，人们已研究出高效的求积方法 [42]。例如，在假设观测数据服从加性高斯噪声模型的情况下，研究表明，由椭圆型和抛物型偏微分方程所产生的逆问题中的似然势可全纯延拓至复域的某个子集 [40, 41]。对该子集进行量化，可得出似然和后验分布的显式稀疏谱逼近策略。

这里考虑具有全纯延拓至 $\mathbb{C}^d$ 的某个子集的似然势函数 $\Phi(;y) \{[0,1]^d \to \mathbb{R}$ 的前馈深度神经网络代理模型。前馈深度神经网络的架构是仿射变换的迭代复合，随后接一个非线性激活函数。更准确地说，逼近似然势函数的深度神经网络 $\Phi^N : \mathbb{R}^d \to \mathbb{R}$

似然势 Φ 可以表示为

$$\Phi^N(u;y) = W_L\sigma(\dots\sigma(W_2\sigma(W_1u+b_1)+b_2)\dots)+b_L, \quad u\in\mathbb{R}^d, \tag{4.3}$$

其中 $W_j\in\mathbb{R}^{d_{j+1}\times d_j}$ 且 $b_j\in\mathbb{R}^{d_{j+1}}$，适用于 $j=1,\dots,L$，其中 $d_1=d$ 且 $d_{L+1}=1$。请注意，此处我们固定了 y，而一般而言，网络的参数可能依赖于数据 y [4]。函数 $\sigma:\mathbb{R}^d\to\mathbb{R}^d$ 是基于单调激活函数 $\sigma:\mathbb{R}\to\mathbb{R}$为任意 d 定义的，该函数逐元素作用于任意向量。请注意，方程 (4.3) 中 σ 的输入维度会随层而变化。隐藏层的数量 $L-1$ 称为深度，而 N，即 W_j和 b_j, $j=1,\dots,L$中的非零分量总数，称为深度神经网络的规模。我们将进一步把研究范围限制在修正线性单元 (ReLU) 激活函数$\sigma(u)=\max\{0,u\}$，因为这保证了可以应用 [15] 的结果（见下文命题 4.3）。

为简便起见，下面假设 $\Phi\geq 0$ 从而假设 1 在 $f\equiv g\equiv 1$下得到满足。此外，考虑一个固定的观测数据集 $y\in\mathbb{R}^m$。根据 [34], 我们假设 Φ 具有到伯恩斯坦多椭圆的全纯延拓

$$\mathcal{E}_{\boldsymbol{\rho}} := \mathcal{E}_{\rho_1}\times\dots\times\mathcal{E}_{\rho_d}\subset\mathbb{C}^d, \tag{4.4}$$

其中 $\boldsymbol{\rho}=(\rho_1,\dots,\rho_d)$，$\mathrm{E}_{\rho_j}\subset\mathbb{C}$表示一个包含 $[-1,1]$且半轴之和为 $\rho_j>1$的椭圆。首先，我们回顾深度神经网络的一项存在性结果，该结果表明，存在具有所需深度和规模的深度神经网络，能够随着网络规模的增大，以指数速度逼近全纯似然函数：

Proposition 4.3（[34, 定理 3。6]）。固定 $y\in\mathcal{Y}$，并假设 $\Phi(\cdot;y):[-1,1]^d\to\mathbb{R}$具有到 $\mathcal{E}_{\boldsymbol{\rho}}$（见（4.4））中给出的 $M\in\mathbb{N}$ 全纯延拓，则存在一个 ReLU 深度神经网络 $\Phi^N:[0,1]^d\to\mathbb{R}$ 使得

$$\|\Phi(\cdot;y)-\Phi^N(\cdot;y)\|_{W^{1,\infty}([-1,1]^d)}\leq C\exp(-\kappa M^{\frac{1}{d+1}}),$$

其中 $\kappa=\kappa(\boldsymbol{\rho},d)>0$和 $C=C(\Phi,\boldsymbol{\rho},d)>0$是常数，并且

$$L\leq CM^{\frac{1}{d+1}}\log_2 M \quad and \quad N\leq M.$$

现在我们希望证明，上述指数收敛速率可以扩展为关于 $\mathrm{D}(\nu,\nu^N;c)$ 的界，其中 ν^N 是通过将 Φ 替换为Φ^N而得到的后验分布。为此，我们现在考虑先验 $\mu\in\mathbb{P}([0,1]^d)$ 和类距离函数c，使得 $c(\cdot,0)\in L^1(\mu)$，并令 $h(\cdot,y)$ 为假设 1 中的函数，令 Φ^N 为上文命题 4.3 中指定的深度神经网络。于是我们可以计算

$$\begin{aligned}\Phi^N(u;y)&\leq|\Phi^N(u;y)-\Phi(u;y)|+\Phi(u;y)\\&\leq C\exp(-\kappa M^{\frac{1}{d+1}})-\log h(u;y)\\&\leq-\log\left(\widetilde{C}h(u;y)\right).\end{aligned}$$

同时，$\Phi^N(u;y)\geq -C\exp(-\kappa M^{\frac{1}{d+1}})$，因此 Φ^N满足假设1，其中函数 h^N 与 h 成正比，且函数 f^N 为常数[5]。于是

[4]还可以将该网络参数化为函数 $\Phi:\mathrm{X}\times\mathrm{Y}\to\mathbb{R}$，但为简单起见，我们在此不这样做[5]注意，由于假定数据 y 固定，因此函数 g 在此处也是固定的

定理的应用 3.2给出了以下界

$$\mathcal{D}(\nu,\nu^N;c)\le C\|c(\cdot,0)\|_{L^1(\mu)}\exp\left(-\kappa M^{\frac{1}{d+1}}\right),$$

其中常数 $C=C(\Phi,\kappa,y)>0$。

5. 结论。我们提出了新的理论结果，用于控制贝叶斯后验测度 ν 相对于似然势函数Φ 的扰动，以及先验测度 μ 相对于积分概率度量的扰动。积分概率度量是一类广泛的散度，其中包括 Wasserstein 度量等。所提出的结果在理论上便于使用，因为它们允许我们根据具体问题调整积分概率度量，因此能够适用于大类非线性且局部利普希茨的似然势函数。

在讨论先验扰动时，积分概率度量尤为适用，因为这些散度不要求扰动后的先验相对于原始先验仍然绝对连续。这种灵活性为分析一些有趣的近似方法打开了大门，例如由一组样本构建的经验先验、协方差核迥异的高斯先验，以及推前先验，正如我们的应用所展示的那样。这些结果还为分析贝叶斯逆问题中现代机器学习技术的收敛性铺平了道路，例如数据驱动先验或深度神经网络近似。

在未来的研究中，将本文所展示的先验与似然扰动结果扩展到频率学派框架，并研究相对于某个“真实”数据生成分布的平均稳定性，将会很有意义 $y|u^\dagger$。此外，第4节所讨论的应用还引出了许多有趣的后续问题，例如非参数情形下的回归，或层次高斯场之间的距离。

6. 致谢。AGI 得到了 Asociaci´on Mexicana de Cultura, A.C. 的支持，并获得墨西哥国家研究人员系统的部分支持。TH 的工作得到了芬兰科学院（决定编号 326961）和 LUT 大学研究基金会的支持。FH 得到了加州理工学院的启动资金以及德国研究联合会（DFG）通过项目 390685813 - GZ 2047/1 - HCM 提供的支持。BH 得到了美国国家科学基金会 DMS-208535 项目的资助。

参考文献

[1] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya 等。《深度学习中的不确定性量化综述：技术、应用与挑战》。Information Fusion，76：243–297，2021。[2] M. Arbel, A. Korba, A. Salim 和 A. Gretton。《最大均值差异梯度流》。《神经信息处理系统进展》，32，2019。[3] S. Arridge, P. Maass, O. ¨Oktem 和 C.-B. Sch¨onlieb。《使用数据驱动模型求解逆问题》。Acta Numerica，28：1–174，2019。[4] J. Birrell, P. Dupuis, M. A. Katsoulakis, Y. Pantazis 和 L. Rey-Bellet。《(f,Γ)-散度：在 f-散度与积分概率度量之间插值》。Journal of Machine Learning Research，23(39)：1–70，2022。[5] V. I. Bogachev。《高斯测度》。美国数学学会，普罗维登斯，1998。[6] V. I. Bogachev 和 A. V. Kolesnikov。《Monge-Kantorovich 问题：成果、联系与展望》。Russian Mathematical Surveys，67(5)：785–890，2012。[7] M. Dashti 和 A. M. Stuart。《逆问题的贝叶斯方法》。载于 R. Ghanem、D. Higdon 和 H. Owhadi 编，Handbook of Uncertainty Quantification，第1–118页。Springer International Publishing，2016。[8] R. DeVore, B. Hanin 和 G. Petrova。《神经网络近似》。Acta Numerica，30：327–444，2021。

[9] R. A. DeVore 和 G. G. Lorentz。《构造逼近》。Springer Science & Business Media，1993。[10] N. Fournier 和 A. Guillin。《经验测度在Wasserstein距离下的收敛速率》。《概率论及相关领域》，162(3)：707–738，2015。[11] M. Gonz´alez, A. Almansa 和 P. Tan。《通过结合自编码先验的联合后验最大化求解逆问题》。《SIAM成像科学杂志》，15(2)：822–859，2022。[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville 和 Y. Bengio。《生成对抗网络》。《ACM通讯》，63 144，2020 11 139( ): – .[13] M. Hairer, J. C. Mattingly 和 M. Scheutzow。《渐近耦合与 Harris定理的一般形式及其在随机时滞方程中的应用》。《概率论及相关领域》，149(1-2)：223–259，2011。[14] M. Hairer, A. M. Stuart 和 S. J. Vollmer。《无限维空间中Metropolis–Hastings算法的谱隙》。《应用概率年刊》，24(6)：2455–2490，2014。[15] L. Herrmann, C. Schwab 和 J. Zech。《贝叶斯偏微分方程反演中后验期望的深度神经网络表达》。《逆问题》，36(12)：125011，2020。[16] B. Hosseini。《具有无限可分和重尾先验测度的适定贝叶斯逆问题》。《SIAM/ASA不确定性量化杂志》，5：1024–1060，2017。[17] B. Hosseini 和 J. E. Johndrow。《具有非高斯先验的无限维Metropolis–Hastings算法的谱隙与误差估计》。《应用概率年刊》，2022。即将刊出。[18] B. Hosseini 和 N. Nigam。《适定贝叶斯逆问题：指数尾先验》。《SIAM/ASA不确定性量化杂志》，5：436–465，2017。[19] J. E. Johndrow 和 J. C. Mattingly。《用于贝叶斯采样的马尔可夫链近似误差界》。arXiv预印本：1711.05382，2017。[20] I. Kobyzev, S. J. Prince 和 M. A. Brubaker。《归一化流：导论与当前方法综述》。《IEEE模式分析与机器智能汇刊》，43(11)：3964–3979，2020。[21] N. Kovachki, R. Baptista, B. Hosseini 和 Y. Marzouk。《使用单调生成对抗网络进行条件采样》。arXiv预印本：2006.06755，2020。[22] J. Latz。《贝叶斯逆问题的适定性》。《SIAM/ASA不确定性量化杂志》，8(1)：451–482，2020。[23] R. Laumont, V. D. Bortoli, A. Almansa, J. Delon, A. Durmus 和 M. Pereyra。《使用即插即用先验进行贝叶斯成像：当朗之万遇上Tweedie》。《SIAM成像科学杂志》，15(2)：701–737，2022。[24] O. Le Maˆıtre 和 O. M. Knio。《不确定性量化的谱方法：及其在计算流体动力学中的应用》。Springer Science & Business Media，2010。[25] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart 和 A. Anandku- mar。《用于参数化偏微分方程的傅里叶神经算子》。arXiv预印本 arXiv:2010.08895，2020。[26] X. Ma 和 N. Zabaras。《基于自适应稀疏网格配置法求解逆问题的高效贝叶斯推断方法》。《逆问题》，25(3)：035013，2009。[27] Y. M. Marzouk 和 H. N. Najm。《贝叶斯逆问题推断中的降维与多项式混沌加速》。《计算物理学杂志》，228(6)：1862–1902，2009。[28] Y. M. Marzouk, H. N. Najm 和 L. A. Rahn。《高效求解贝叶斯逆问题的随机谱方法》。《计算物理学杂志》，224(2)：560–586，2007。[29] Y. Meyer。《小波与算子》。剑桥大学出版社，1992。[30] K. Muandet, K. Fukumizu, B. Sriperumbudur, B. Sch¨olkopf 等。《分布的核均值嵌入：综述与展望》。《机器学习基础与趋势》，10(1-2)：1–141，2017。[31] A. M¨uller。《积分概率度量及其生成函数类》。《应用概率进展》，29(2)：429–443，1997。[32] S. Nowozin, B. Cseke 和 R. Tomioka。《 f-GAN：通过变分散度最小化训练生成式神经采样器》。《神经信息处理系统进展》，29，2016。[33] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis 和 R. Willett。《成像逆问题的深度学习技术》。《IEEE信息论选刊》，1(1)：39–56，2020。[34] J. A. Opschoor, C. Schwab 和 J. Zech。《高维全纯映射的指数型ReLU深度神经网络表达》。《构造逼近》，55(1)：537–582，2022。[35] H. Owhadi 和 C. Scovel。《贝叶斯推断的脆弱性与新的塞尔伯格公式》。

《数学科学通信》14(1):83–145，2016。[36] H. Owhadi、C. Scovel 和 T. Sullivan. 连续世界中有限信息条件下贝叶斯推断的脆弱性。《电子统计学杂志》9:1–79，2015。[37] H. Owhadi、C. Scovel 和 T. Sullivan. 关于贝叶斯推断的脆弱性。《SIAM评论》57(4):566–582，2015。[38] D. V. Patel、D. Ray 和 A. A. Oberai. 使用深度生成先验求解基于物理的贝叶斯逆问题。《计算机在应用力学与工程中的方法》400:115428，2022。[39] A. Sagiv. 推前测度之间的 Wasserstein距离及其在不确定性量化中的应用。arXiv 预印本 arXiv:1902.05451，2019。[40] C. Schillings 和 C. Schwab. 用于贝叶斯逆问题的稀疏自适应 Smolyak 求积。《逆问题》29(6):065011，2013。[41] C. Schillings 和 C. Schwab. 参数算子方程贝叶斯反演中的稀疏性。《逆问题》30(6):065007，2014。[42] C. Schwab 和 A. M. Stuart. 贝叶斯逆问题的稀疏确定性近似。《逆问题》28(4):045003，2012。[43] B. Sprungk. 贝叶斯逆问题的局部Lipschitz稳定性。《逆问题》36(5):055015，2020。[44] B. K. Sriperumbudur、K. Fukumizu、A. Gretton、B. Sch¨olkopf 和 G. R. Lanckriet. 关于积分概率度量、ϕ $0901\ 2698$, -散度与2009二元分类。arXiv 预印本 arXiv: . .[45] A. Stuart 和 A. Teckentrup. 高斯过程近似的贝叶斯后验分布的后验一致性。《计算数学》87(310):721–753，2018。[46] A. M. Stuart. 逆问题：贝叶斯视角。《Acta Numerica》19:451–559，2010。[47] T. J. Sullivan. 适定贝叶斯逆问题与重尾稳定拟Banach空间先验。《逆问题与成像》11(5):857–874，2017。[48] A. L. Teckentrup. 具有估计的超参数的高斯过程回归的收敛性及其在贝叶斯逆问题中的应用。《SIAM/ASA不确定性量化杂志》8(4):1310–1337，2020。[49] C. Villani. 最优传输：古今论述。Springer，纽约，2009。[50] H. Wendland. 散乱数据近似。剑桥大学出版社，2004。[51] P. Whittle. 多维随机过程。《国际统计学会会刊》40(2):974–994，1963。[52] C. Williams 和 C. E. Rasmussen. 机器学习中的高斯过程。麻省理工学院出版社，2006。