arXiv:2303.02554v1 [math.NA] 5 Mar 2023

面向集中概率密度的自增强多项式近似方法

TIANGANG CUI*、SERGEY DOLGOV†、和 OLIVIER ZAHM‡

摘要。传输映射方法提供了一种强大的统计学习工具，能够利用可逆变换将目标高维随机变量与某个参考随机变量耦合起来。本文提出了基于一般可分函数构造 Knothe–Rosenblatt（KR）重排的新计算技术。我们首先介绍一种新的 KR 重排构造方法：通过使用张量积谱多项式和向下封闭稀疏指标集来近似目标随机变量的密度，并保证其数值实现中的可逆性。与其他基于多线性逼近或非线性优化的 KR 重排构造方法相比，我们的新构造仅依赖于加权最小二乘逼近过程。随后，受近期提出的深度张量列车方法（Cui 和 Dolgov，*Found. Comput. Math.* 22:1863–1922，2022）的启发，我们通过使用映射复合对密度逼近问题进行预条件化，从而增强稀疏多项式的近似能力。这尤其适用于许多应用中常见的高维集中概率密度。我们通过 *self-reinforced* 个 KR 重排的复合来近似复杂的目标密度，其中，基于相同近似假设构造的 KR 重排被用于对密度逼近问题进行预条件化，以构建每一个新的 KR 重排。我们在若干由常微分方程（ODEs）和偏微分方程（PDEs）控制的逆问题上展示了所提方法的效率，以及使用复合映射的重要性。

关键词。稀疏多项式、最优加权最小二乘、传输映射、深度逼近、不确定性量化、逆问题

MSC 分类号。65D15, 65D32, 65C05, 65C60, 62F15, 65N21, 65L09

1. 引言。自从第一台通用数字计算机 ENIAC 发明以来，模拟复杂高维随机变量的一种常见策略，是对参考随机变量（例如高斯分布）进行确定性或随机变换，使其转化为目标随机变量 [33]。经典示例包括马尔可夫链蒙特卡罗（MCMC）和重要性采样。前者生成一个随机变量的马尔可夫链，并使其渐近收敛于目标随机变量；后者则利用参考随机变量和权函数刻画目标随机变量。最优传输 [50]及相关变换方法的发展，为这一任务提供了新的思路。无需构造马尔可夫链或使用权函数，可以通过映射 $\mathcal{T}$ 将取值于 $\mathcal{X} \subseteq \mathbb{R}^d$ 的目标随机变量 $\boldsymbol{X}$ 与取值于 $\mathcal{U} \subseteq \mathbb{R}^d$ 的参考随机变量 $\boldsymbol{U}$ 耦合起来：$\mathcal{U} \to \mathcal{X}$ 使得 $\mathcal{T}(\boldsymbol{U}) = \boldsymbol{X}$。分别以 $\nu_{\boldsymbol{X}}$和 $\nu_{\boldsymbol{U}}$表示 $\boldsymbol{X}$ 和 $\boldsymbol{U}$ 的分布，则 $\nu_{\boldsymbol{U}}$ 在映射 $\mathcal{T}$下的推前（记为 $\mathcal{T}_\sharp \nu_{\boldsymbol{U}}$）等价于 $\nu_{\boldsymbol{X}}$。本文聚焦于 KR 重排 [29, 42], ，这是一种以三角形式构造此类映射的特殊方法

$$\mathcal{T}(\boldsymbol{U}) = \begin{pmatrix} \mathcal{T}_1(U_1) \\ \mathcal{T}_2(U_1, U_2) \\ \vdots \\ \mathcal{T}_d(U_1, U_2, \ldots, U_d) \end{pmatrix}.$$

这里，每个项 $\mathcal{T}_k : \mathbb{R}^k \to \mathbb{R}$ 都是一个标量值多元函数，依赖于前 k 个输入变量，因此，前 k 个 KR 重排项对 $(U_1, \ldots U_k)$ 的映像

*莫纳什大学数学学院，维多利亚州3800，澳大利亚 tiangang.cui@monash.edu
†巴斯大学数学科学系，英国巴斯s.dolgov@bath.ac.uk
‡格勒诺布尔阿尔卑斯大学、Inria、法国国家科学研究中心、格勒诺布尔国立综合理工学院、LJK，法国格勒诺布尔 olivier.zahm@inria.fr

与目标边缘变量 $(X_1, \ldots X_k)$ 具有相同的分布。

一种用于计算 KR 重排的方法采用基于样本的映射方法，利用从目标测度中抽取的一组样本来估计映射 $\nu_{\boldsymbol{X}}$。这种方法通常用于无法直接获取目标测度，但可以获得其实现值 $\boldsymbol{X}$ 的情形。值得注意的例子包括 [2, 41], 中介绍的方法，它们通过最小化 $\mathcal{T}$ 相对于 $\nu_{\boldsymbol{X}}$ 的 Kullback–Leibler（KL）散度，在由多项式基参数化的三角变换族中确定映射 $\mathcal{T}_\sharp \nu_{\boldsymbol{U}}$ 。在这种情形下，KL 散度$D_{\mathrm{KL}}(\nu_{\boldsymbol{X}} \| \mathcal{T}_\sharp \nu_{\boldsymbol{U}})$通过蒙特卡洛方法，利用从 $\nu_{\boldsymbol{X}}$中抽取的样本进行近似。正规化流方法（例如 [7, 9, 30, 40], ）采用类似的 KL 散度最小化问题，但映射不一定是三角变换，且通常由神经网络参数化。总体而言，基于样本的映射方法易于灵活实现，因为它只需要一组从 $\nu_{\boldsymbol{X}}$中抽取的样本。不过，由于 KL 散度是通过蒙特卡洛估计的，该方法的误差率为 $N^{-1/2}$，其中 N 为样本量。有关分析，请参阅 [51] 及其中的参考文献。

另一类方法采用从密度映射 的方法，通过对 $\boldsymbol{X}$的未归一化密度进行点值评估来估计映射。这适用于目标密度可以计算、但难以从中采样的情况。值得注意的例子包括 [5, 36, 49]的工作，该工作通过最小化反向 KL 散度 $D_{\mathrm{KL}}(\mathcal{T}_\sharp \nu_{\boldsymbol{U}} \| \nu_{\boldsymbol{X}})$来求得三角映射$\mathcal{T}$。尽管这种方法可能不会受到从样本映射 方法缓慢的 $N^{-1/2}$误差收敛率影响，但其实践中的训练过程通常相当复杂——目标函数存在许多局部极小值，并且每次优化迭代都需要在候选映射作用下，对变换后的参考变量反复计算目标密度。

除了最小化反向 KL 散度之外，人们还开发了函数近似方 法，用于从未归一化密 度构造 KR 重排。其思想是计算目标随机变量的某个近似 密度的精确 KR 重排。例如， [14, 15, 21]使用张量方法来构造该近似密度。通过这种方式，该方法避开了高度非线性的优化过程，相关误差分析通常也较为成熟。然而，在许多应用中，人们关注的高维随机变量具有集中的密度函数，并表现出复杂的非线性交互作用。在这种情况下，使用标准逼近工具直接近似目标密度函数通常在计算上不可行。遵循 [14], ，规避这一挑战的一种方法是引入分层构造技术，构建 KR 重排的复合，即 $\mathcal{T} = \mathcal{Q}^{(1)} \circ \cdots \circ \mathcal{Q}^{(L)}$. 每个映射 $\mathcal{Q}^{(\ell)}$都通过贪心方式构造为从参考随机变量$\boldsymbol{U}$到预条件化桥接随机变量 the law $(\mathcal{Q}^{(1)} \circ \cdots \circ \mathcal{Q}^{(\ell-1)})^\sharp \nu_{\boldsymbol{X}^{(\ell)}}$的 KR 重排。其中，桥接随机变量 $\boldsymbol{X}^{(\ell)}$ 的选取使其密度函数具有较低的复杂度。预条件化是降低 KR 重排训练复杂度的关键。

贡献与概要。本文对用于根据密度构造 KR 重排的函数逼近方法进行了推广和改进。不同于开创性论文中使用张量方法的做法 [14, 15],，我们在第2 节提出通过最小二乘形式，使用稀疏张量积基来近似目标密度。具体而言，我们对未归一化目标密度的平方根进行近似，再将所得函数平方，以确保密度近似的非负性。非负性对 KR 重排的数值稳定性至关重要。与平方张量

方法相比，使用稀疏张量积基的 KR 重排计算复杂度不受变量排序的影响。

原则上可以使用任意张量积基，但我们在第 3 节表明，使用正交基函数会在计算所得 KR 重排时产生额外的稀疏模式。此外，本文考虑正交多项式基，因为它们可提供解析表达式，从而能够快速计算 KR 重排。借鉴 [34, 35], 中的思路，我们还展示了如何使用向下封闭指标集和最优加权最小二乘法 [11, 26,35, 38] ，在高维情形下自适应地搜索合适的多项式基。

在第 4节中，我们采用 [14]的分层构造技术，设计了一种自增强稀疏多项式逼近流程。该流程利用前几层构建的 KR 重排的复合，对下一个密度近似问题进行预条件处理，从而生成新一层的 KR 重排。与单层构造相比，这种方法在精度和复杂度方面都能提供更高效的 KR 重排。它还为如何增强稀疏多项式方法的逼近能力提供了启示。我们给出了实用方案，通过自适应方法和各种形式的桥接密度来构建复合 KR 重排。在第 5节中，我们通过若干由常微分方程和偏微分方程支配的贝叶斯逆问题，展示了所提方法的效率。

2. 通过可分函数近似构建 KR 重排。

2.1. 符号说明。 我们考虑相对于勒贝格测度绝对连续的概率测度。随机变量 $\boldsymbol{X}$ 的分布记为$\nu_{\boldsymbol{X}}$，其密度记为 $f_{\boldsymbol{X}}$。给定另一个密度为 $f_{\boldsymbol{Y}}$的随机变量 $\boldsymbol{Y} \sim \nu_{\boldsymbol{Y}}$， $\nu_{\boldsymbol{X}}$ 与 $\nu_{\boldsymbol{Y}}$之间的海林格距离定义为

$$D_{\mathrm{H}}(\nu_{\boldsymbol{X}}, \nu_{\boldsymbol{Y}}) = \left(\frac{1}{2}\int \left(\sqrt{f_{\boldsymbol{X}}(\boldsymbol{x})} - \sqrt{f_{\boldsymbol{Y}}(\boldsymbol{x})}\right)^2 \mathrm{d}\boldsymbol{x}\right)^{\frac{1}{2}}. \tag{2.1}$$

给定一个微分同胚 $\mathcal{S}: \mathcal{X} \to \mathcal{U}$， $\mathcal{S}(\boldsymbol{X})$的分布称为 $\nu_{\boldsymbol{X}}$在 $\mathcal{S}$下的推前，记为 $\mathcal{S}_\sharp\, \nu_{\boldsymbol{X}}$。类似地， $\mathcal{S}^{-1}(\boldsymbol{U})$的分布称为 $\nu_{\boldsymbol{U}}$ 在 $\mathcal{S}$ 下的拉回测度，记为 $\mathcal{S}^\sharp\, \nu_{\boldsymbol{U}}$。 $\mathcal{S}_\sharp\, \nu_{\boldsymbol{X}}$ 和 $\mathcal{S}^\sharp\, \nu_{\boldsymbol{U}}$ 均具有由变量替换公式给出的密度

$$\mathcal{S}_\sharp\, f_{\boldsymbol{X}}(\boldsymbol{u}) = \left(f_{\boldsymbol{X}} \circ \mathcal{S}^{-1}\right)(\boldsymbol{u}) \left|\nabla_{\boldsymbol{u}} \mathcal{S}^{-1}(\boldsymbol{u})\right| \tag{2.2}$$

$$\mathcal{S}^\sharp\, f_{\boldsymbol{U}}(\boldsymbol{x}) = \left(f_{\boldsymbol{U}} \circ \mathcal{S}\right)(\boldsymbol{x}) \left|\nabla_{\boldsymbol{x}} \mathcal{S}(\boldsymbol{x})\right|, \tag{2.3}$$

对所有 $\boldsymbol{u}$， $\boldsymbol{x} \in \mathbb{R}^d$均成立，其中 $|\cdot|$ 表示矩阵的行列式。

我们用 $\mathbb{N}$ 表示所有正整数的集合，并用 $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ 表示。对于 $d \in \mathbb{N}$，我们用黑体字母表示 d 维向量和多重指标。例如，$\boldsymbol{k} = (k_1, k_2, \ldots\ldots k_d) \in \mathbb{N}_0^d$，以及 $\boldsymbol{x} = (x_1, x_2, \ldots\ldots x_d) \in \mathbb{R}^d$。此外，令 $\mathbf{0} = (0, 0, \ldots\ldots, 0) \in \mathbb{N}_0^d$，$\mathbf{1} = (1, 1, \ldots\ldots, 1) \in \mathbb{N}^d$，并令 $\boldsymbol{e}_j = (0, \ldots\ldots, 1, \ldots\ldots, 0)$ 为 $\mathbb{R}^d$ 中第 j 个单位向量。集合 $\mathcal{K}$ 的基数记为 $|\mathcal{K}|$。

给定一个 d 维空间，我们用 $[d] = \{1, \ldots\ldots d\} \subset \mathbb{N}$ 表示完整的坐标索引集。给定坐标索引子集 $\mathfrak{u} = (u_1, \ldots\ldots u_s) \subseteq [d]$，其基数为 $|\mathfrak{u}| = s \le d$，我们将其补集记为 $\mathfrak{u}^c = [d] \setminus \mathfrak{u}$。据此，我们还可以定义 $\boldsymbol{x}_{\mathfrak{u}} = (x_{u_1}, \ldots\ldots x_{u_s}) \in \mathbb{R}^{|\mathfrak{u}|}$，$\mathcal{X}_{\mathfrak{u}} = \mathcal{I}^{|\mathfrak{u}|}$。对于任意标量值函数$h$： $\mathbb{R} \to \mathbb{R}$，记 $h(\boldsymbol{x}_{\mathfrak{u}}) = \prod_{i\in\mathfrak{u}} h(x_i)$。

给定矩阵 $\mathsf{A} \in \mathbb{R}^{n\times d}$ 和坐标索引集 $\mathfrak{u} \subseteq [n]$、$\mathfrak{v} \subseteq [d]$，其基数分别为 $|\mathfrak{u}| = s \le n$、$|\mathfrak{v}| = t \le d$。我们将 $\mathsf{A}_{\mathfrak{u}:} \in \mathbb{R}^{s\times d}$ 和 $\mathsf{A}_{:\mathfrak{v}} \in \mathbb{R}^{n\times t}$ 分别记为由 A 中行索引集为 $\mathfrak{u}$ 的行子集和列索引集为 $\mathfrak{v}$的列子集构成的矩阵。类似地，$\mathsf{A}_{\mathfrak{u}\mathfrak{v}} \in \mathbb{R}^{s\times t}$ 表示 A 的子矩阵，其行

的索引为 $\mathfrak{u}$，列的索引为 $\mathfrak{v}$。

对于多重指标集 $\mathcal{K} \subset \mathbb{N}_0^d$，我们定义函数矩阵$(\mathcal{K})$，将 $\mathcal{K}$ 转换为矩阵 $\mathbf{K}$ =矩阵$(\mathcal{K})$，使得 $\mathbf{K} \in \mathbb{N}_0^{|\mathcal{K}| \times d}$。我们还引入映射$\sigma\colon \mathcal{K} \to [|\mathcal{K}|]$，其中 $[|\mathcal{K}|] = \{1, 2, \dots |\mathcal{K}|\}$，使得 $\sigma(\boldsymbol{k}) \in [|\mathcal{K}|]$ 给出有序集 $\mathcal{K}$ 中多重指标 $\boldsymbol{k}$ 的标量索引。按照此记法，对于任意 $\boldsymbol{k} \in \mathcal{K}$，行 $\mathbf{K}_{\sigma(\boldsymbol{k}):} = (k_1, \dots..., k_d)$ 包含多重指标 $\boldsymbol{k}$ 的整数。

本文中使用的所有多维空间都是配备乘积形式归一化权函数的笛卡尔积空间。例如，对于 $\mathcal{X} \subseteq \mathbb{R}^d$，我们有 $\mathcal{X} = \mathcal{I}^d$，其中 $\mathcal{I} \subseteq \mathbb{R}$，且相应的权函数形式为 $\lambda(\boldsymbol{x}) = \prod_{i=1}^d \lambda(x_i)$，使得 $\int_{\mathcal{I}} \lambda(x_i)\mathrm{d}x_i = 1$，其中 $i = 1, \dots d$。在区间 $\mathcal{I}$ 上，我们将内积和加权内积记为

$$\langle f, g\rangle = \int_{\mathcal{I}} f(x)g(x)\mathrm{d}x \quad \text{and} \quad \langle f, g\rangle_\lambda = \int_{\mathcal{I}} f(x)g(x)\lambda(x)\mathrm{d}x,$$

分别地。由这些内积诱导的范数记为$\|f\|_{L^2(\mathcal{X})} := \langle f, f\rangle^{1/2}$和$\|f\|_{L^2_\lambda(\mathcal{X})} := \langle f, f\rangle_\lambda^{1/2}$。

2.2. KR 重排。 KR 重排背后的基本思想源于分布函数的一项基本性质：给定一个具有密度 f_X 的标量随机变量 X，其分布函数 $\mathcal{F}_X(x) = \int_{-\infty}^x f_X(x')\mathrm{d}x'$ 会将随机变量 X 映射为一个在区间 $[0, 1]$ 上均匀分布的随机变量。要将这一性质推广到向量值随机变量，我们将多元密度函数 $f_{\boldsymbol{X}}$ 分解为一维条件边缘分布的乘积

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) \dots f_{X_d|\boldsymbol{X}_{[d-1]}}(x_d|\boldsymbol{x}_{[d-1]}), \tag{2.4}$$

其中，条件边缘分布和边缘分布分别定义为

$$f_{X_t|\boldsymbol{X}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) = \frac{f_{\boldsymbol{X}_{[t]}}(\boldsymbol{x}_{[t-1]}, x_t)}{f_{\boldsymbol{X}_{[t-1]}}(\boldsymbol{x}_{[t-1]})}, \tag{2.5}$$

$$f_{\boldsymbol{X}_{[t]}}(\boldsymbol{x}_{[t]}) = \int f_{\boldsymbol{X}}(\boldsymbol{x}_{[t]}, \boldsymbol{x}'_{[t]^c})\mathrm{d}\boldsymbol{x}'_{[t]^c}. \tag{2.6}$$

接下来 [42], 令 $\mathcal{F}_{\boldsymbol{X}}\colon \quad \mathcal{X} \mapsto [0, 1]^d$ 定义为

$$\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{pmatrix} \mathcal{F}_{X_1}(x_1) \\ \mathcal{F}_{X_2|X_1}(x_2|x_1) \\ \vdots \\ \mathcal{F}_{X_d|\boldsymbol{X}_{[d-1]}}(x_d|\boldsymbol{x}_{[d-1]}) \end{pmatrix}, \tag{2.7}$$

其中每个分量都是单变量条件分布函数

$$\mathcal{F}_{X_t|\boldsymbol{X}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) = \int_{-\infty}^{x_t} f_{X_t|\boldsymbol{X}_{[t-1]}}(x'_t|\boldsymbol{x}_{[t-1]})\mathrm{d}x'_t. \tag{2.8}$$

由条件边缘分布的非负性可知，每个分量 $\mathcal{F}_{X_t|\boldsymbol{X}_{[t-1]}}$ 都是关于第 t 个变量的非递减函数，即 $\partial_t \mathcal{F}_{X_t|\boldsymbol{X}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) \geq 0$。此外，$\mathcal{F}_{\boldsymbol{X}}$ 具有下三角性质，即第 t 个分量是一个标量值函数，仅依赖于前 t 个变量 $\boldsymbol{x}_{[t]}$。映射 $\mathcal{F}_{\boldsymbol{X}}$ 满足

$$\mathcal{F}_{\boldsymbol{X}}^\sharp f_{\boldsymbol{\Xi}}(\boldsymbol{x}) = \big(f_{\boldsymbol{\Xi}} \circ \mathcal{F}_{\boldsymbol{X}}\big)(\boldsymbol{x}) \left|\nabla_{\boldsymbol{x}} \mathcal{F}_{\boldsymbol{X}}(\boldsymbol{x})\right| = \left|\nabla_{\boldsymbol{x}} \mathcal{F}_{\boldsymbol{X}}(\boldsymbol{x})\right| = f_{\boldsymbol{X}}(\boldsymbol{x}),$$

其中，$\boldsymbol{\Xi}$ 是在 $[0, 1]^d$ 上均匀分布的随机变量。这表明 $\boldsymbol{\Xi} = \mathcal{F}_{\boldsymbol{X}}(\boldsymbol{X})$，这等价于 $\boldsymbol{X} = \mathcal{F}_{\boldsymbol{X}}^{-1}(\boldsymbol{\Xi})$。反映射 $\mathcal{F}_{\boldsymbol{X}}^{-1}$ 也是下三角的，且 $\boldsymbol{x} = \mathcal{F}_{\boldsymbol{X}}^{-1}(\boldsymbol{\xi})$可计算为

$$\mathcal{F}_{\boldsymbol{X}}^{-1}(\boldsymbol{\xi}) = \big[\mathcal{F}_{X_1}^{-1}(\xi_1),\ \mathcal{F}_{X_2|X_1}^{-1}(\xi_2|x_1),\ \cdots,\ \mathcal{F}_{X_d|\boldsymbol{X}_{[d-1]}}^{-1}(\xi_d|\boldsymbol{x}_{[d-1]})\big]^\top, \tag{2.9}$$

其中， $\mathcal{F}^{-1}_{X_t|\boldsymbol{X}_{[t-1]}}(\cdot|\boldsymbol{x}_{[t-1]})$表示 $x_t \mapsto \mathcal{F}_{X_t|\boldsymbol{X}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]})$的反函数。我们假设条件边缘分布 $f_{X_t|\boldsymbol{X}_{[t-1]}}$ 几乎处处为正，因此$x_t \mapsto \mathcal{F}_{X_t|\boldsymbol{X}_{[t-1]}}(x_t|\boldsymbol{x}_{t-1})$严格单调，从而存在反函数。

备注 2.1. 上述构造得到一个 KR 重排 $\mathcal{T}_{\boldsymbol{X}}(\boldsymbol{U}) = \boldsymbol{X}$，其参考变量为均匀参考变量 $\boldsymbol{U}$。若要使用一般参考随机变量 $\boldsymbol{U}$，我们可以定义复合映射 $\mathcal{T}_{\boldsymbol{X}} = \mathcal{F}^{-1}_{\boldsymbol{X}} \circ \mathcal{F}_{\boldsymbol{U}}$，其中 $\mathcal{F}_{\boldsymbol{U}}$是式（2.4）–（2.8）中定义的下三角映射。所得映射 $\mathcal{T}_{\boldsymbol{X}}$ 仍然是下三角的。

为高维随机变量构造 KR 重排的一个瓶颈，是计算边缘密度（2.6）中的积分。我们的策略是使用结构化近似类来近似密度 $f_{\boldsymbol{X}}$，该近似类允许对近似密度的边缘分布进行解析计算。这样，相应的近似 KR 重排便可显式求得。

2.3. 近似目标密度。 我们提出用如下形式的密度来近似目标密度函数 $f_{\boldsymbol{X}}$

$$f_{\hat{\boldsymbol{X}}}(\boldsymbol{x}) = \frac{1}{\hat{z}} \left(\gamma + g(\boldsymbol{x})^2\right) \lambda(\boldsymbol{x}), \quad \hat{z} = \gamma + \int_{\mathcal{X}} g(\boldsymbol{x})^2 \lambda(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{2.10}$$

这里， $\lambda(\boldsymbol{x}) = \lambda(x_1) \ldots \lambda(x_d)$ 是一个由正权函数构成的乘积函数，使得$\int_{\mathbb{R}} \lambda(x)\mathrm{d}x = 1$， $\gamma > 0$ 为标量，且 $g \in L^2_\mu(\mathcal{X})$ 是平方可积函数。根据构造，式（2.10）中的密度处处为正 $f_{\hat{\boldsymbol{X}}}(\boldsymbol{x}) > 0$。在实际应用中，我们按如下方式构造 $f_{\hat{\boldsymbol{X}}}$。首先，我们选择具有足够重尾部的 λ，使得

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\lambda(\boldsymbol{x})} < \infty, \tag{2.11}$$

有界。特别地，这意味着密度之比 $f_{\boldsymbol{X}}(\boldsymbol{x})/f_{\hat{\boldsymbol{X}}}(\boldsymbol{x})$ 有上界，这是确保 $f_{\hat{\boldsymbol{X}}}$ 可用作稳健的重要性抽样密度的充分条件，参见 [39, 第 9] 章。给定这样的权函数，我们用 $\Phi(\boldsymbol{x})$ 表示唯一的（相差一个加性常数）势函数，使得目标密度函数可写为

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{z} \exp(-\Phi(\boldsymbol{x})) \lambda(\boldsymbol{x}), \quad z = \int_{\mathcal{X}} \exp(-\Phi(\boldsymbol{x})) \lambda(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{2.12}$$

这里， $\boldsymbol{x} \mapsto \Phi(\boldsymbol{x})$ 可以逐点求值，但归一化常数 z 可能事先未知。接下来，我们通过求解以下最小二乘问题，将 g 构造为 $\exp(-\Phi(\boldsymbol{x}))$ 平方根的近似

$$g = \operatorname*{arg\,min}_{h\in\mathcal{V}_{\mathcal{K}}} \left\|\exp(-\tfrac{1}{2}\Phi) - h\right\|_{L^2_\lambda(\mathcal{X})}. \tag{2.13}$$

这里， $\mathcal{V}_{\mathcal{K}} \subset L^2_\mu(\mathcal{X})$ 是由张量积多元基函数张成的函数子空间，即，

$$\mathcal{V}_{\mathcal{K}} = \operatorname{span}\{\psi_{\boldsymbol{k}} : \boldsymbol{k} \in \mathcal{K}\}, \quad \psi_{\boldsymbol{k}} : \boldsymbol{x} \mapsto \prod_{i=1}^{d} \psi_{k_i}(x_i), \tag{2.14}$$

其中， $\mathcal{K} \subset \mathbb{N}_0^d$ 是一个多重指标集，且每个 ψ_{k_i} 都是由$k_i \in \mathbb{N}_0$索引的一元函数。如下一节2.4所述，张量积结构（2.14）使得能够解析地计算$f_{\hat{\boldsymbol{X}}}$的任意边缘分布，从而可以求得相应的 KR 重排 $\mathcal{F}_{\hat{\boldsymbol{X}}}$。我们将 ψ_k 的定义推迟到第3.1 节，并将指标集 $\mathcal{K}$ 的构造推迟到第3.2节。最后，我们令 $\gamma > 0$为满足$\sqrt{\gamma} \leq \|\exp(-\frac{1}{2}\Phi) - g\|_{L^2_\lambda(\mathcal{X})}$的任意正常数。以下命题见于 [14, 第 3.3], 节，¯ 建立了以下两者之间的联系：

近似 $\exp(-\frac{1}{2}\Phi)$ 时产生的误差，以及随机变量 $\hat{\boldsymbol{X}}$ 作为目标随机变量 $\boldsymbol{X}$ 的近似时产生的误差。

PROPOSITION 2.2. 令 $\lambda: \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ 满足 $\int_{\mathbb{R}^d} \lambda(\boldsymbol{x})d\boldsymbol{x} = 1$。假设近似 $g \in L^2_\lambda(\mathbb{R}^d)$ 满足 $\|\exp(-\frac{1}{2}\Phi) - g\|_{L^2_\lambda(\mathcal{X})} \leq \epsilon$，其中某个 $\epsilon \geq 0$，$\hat{}$，且常数 $\gamma \geq 0$ 满足 $\sqrt{\gamma} \leq \epsilon$. 。那么，密度为 $f_{\hat{\boldsymbol{X}}}$、定义于 (2.10) 中的随机变量 $\boldsymbol{X}$ 满足 $D_{\mathrm{H}}(\nu_{\hat{\boldsymbol{X}}}, \nu_{\boldsymbol{X}}) \leq 2\epsilon/\sqrt{z}$。

由于 $\sqrt{z} = \|\exp(-\frac{1}{2}\Phi)\|_{L^2_\lambda(\mathcal{X})}$，对于相对误差为

$$\frac{\|\exp(-\frac{1}{2}\Phi) - g\|_{L^2_\lambda(\mathcal{X})}}{\|\exp(-\frac{1}{2}\Phi)\|_{L^2_\lambda(\mathcal{X})}} \leq \tau$$

对于某个 $\tau \geq 0$，并且 $\gamma \leq \tau^2 \|\exp(-\frac{1}{2}\Phi)\|^2_{L^2_\lambda(\mathcal{X})}$，命题 2.2还表明，近似随机变量与目标随机变量之间的海林格距离满足 $D_{\mathrm{H}}(\nu_{\hat{\boldsymbol{X}}}, \nu_{\boldsymbol{X}}) \leq 2\tau$。

2.4. 计算近似 KR 重排 $\mathcal{F}_{\hat{\boldsymbol{X}}}$。 计算近似随机变量 $\boldsymbol{X}$ 的 KR 重排需要 $\boldsymbol{X}$的边缘分布。一个简单的方法是计算近似密度（ $g(\boldsymbol{x})^2$ 2.10）中对相应变量的积分，但计算效率可能较低。在下面的命题中，我们利用基 $\psi_{\boldsymbol{k}}$ 的乘积结构，推导 $\boldsymbol{X}$边缘分布的简洁表达式。

PROPOSITION 2.3. 令 $g \in \mathcal{V}_{\mathcal{K}}$ 具有如下形式

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathcal{K}} c_{\boldsymbol{k}} \psi_{\boldsymbol{k}}(\boldsymbol{x}), \tag{2.15}$$

其中系数 $(c_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathcal{K}}$为实值。对于 $1 \leq i \leq d$，我们将质量矩阵 $\mathsf{M}^{(i)} \in \mathbb{R}^{|\mathcal{K}| \times |\mathcal{K}|}$ 定义为

$$\mathsf{M}^{(i)}_{\sigma(\boldsymbol{j})\sigma(\boldsymbol{k})} = \int_{\mathcal{I}} \psi_{j_i}(x_i)\psi_{k_i}(x_i)\lambda(x_i)\mathrm{d}x_i, \quad \boldsymbol{j}, \boldsymbol{k} \in \mathcal{K}. \tag{2.16}$$

给定坐标索引集 $\mathfrak{q} = (q_1, \ldots q_s) \subseteq [d]$，其基数为 $s \leq d$，定义累积质量矩阵 $M^{(\mathfrak{q})} = M^{(q_1)} \circ \cdots \circ M^{(q_s)}$，其中 $\circ$ 表示哈达玛乘积。考虑秩揭示分解 $M^{(\mathfrak{q})}_{\sigma}(\boldsymbol{j})\sigma(\boldsymbol{k}) = \sum_{\ell=1}^{r(\mathfrak{q})} L^{(\mathfrak{q})}_{\sigma}(\boldsymbol{j})\ell L^{(\mathfrak{q})}_{\sigma}{}_{\boldsymbol{k}}(\,)\ell$，其中 $r(\mathfrak{q})$ 是 $M^{(\mathfrak{q})}$的秩，则边缘随机变量 $\hat{\boldsymbol{X}}_{\mathfrak{q}^c}$ 的密度为

$$f_{\hat{\boldsymbol{X}}_{\mathfrak{q}^c}}(\boldsymbol{x}_{\mathfrak{q}^c}) = \int_{\mathcal{X}_{\mathfrak{q}}} f_{\hat{\boldsymbol{X}}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}_{\mathfrak{q}} = \frac{1}{\hat{z}}\big(\gamma + p(\boldsymbol{x}_{\mathfrak{q}^c})\big)\lambda(\boldsymbol{x}_{\mathfrak{q}^c}). \tag{2.17}$$

其中

$$p(\boldsymbol{x}_{\mathfrak{q}^c}) = \sum_{\ell=1}^{r(\mathfrak{q})} \Big(\sum_{\boldsymbol{k} \in \mathcal{K}} c_{\boldsymbol{k}} \mathsf{L}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell} \prod_{i \in \mathfrak{q}^c} \psi_{k_i}(x_i) \Big)^2, \quad \hat{z} = \gamma + \sum_{\boldsymbol{j} \in \mathcal{K}, \boldsymbol{k} \in \mathcal{K}} c_{\boldsymbol{j}}\, c_{\boldsymbol{k}}\, \mathsf{M}^{([d])}_{\sigma(\boldsymbol{j})\sigma(\boldsymbol{k})}. \tag{2.18}$$

证明。由于近似目标密度为 $f_{\hat{\boldsymbol{X}}}(\boldsymbol{x}) = {}^1\hat{z}\,(\gamma + g(\boldsymbol{x})^2)\,\lambda(\boldsymbol{x})$，且权函数 $\lambda(\boldsymbol{x})$ 已归一化并具有乘积形式，因此 $\boldsymbol{X}_{\mathfrak{q}^c}$ 的密度为

$$f_{\hat{\boldsymbol{X}}_{\mathfrak{q}^c}}(\boldsymbol{x}_{\mathfrak{q}^c}) = \frac{1}{\hat{z}} \int_{\mathcal{X}_{\mathfrak{q}}} \big(\gamma + g(\boldsymbol{x})^2\big)\,\lambda(\boldsymbol{x})\mathrm{d}\boldsymbol{x}_{\mathfrak{q}} = \frac{1}{\hat{z}}\Big(\gamma + \int_{\mathcal{X}_{\mathfrak{q}}} g(\boldsymbol{x})^2 \lambda(\boldsymbol{x}_{\mathfrak{q}})\mathrm{d}\boldsymbol{x}_{\mathfrak{q}}\Big)\lambda(\boldsymbol{x}_{\mathfrak{q}^c}).$$

边缘函数 $p(\boldsymbol{x}_{\mathfrak{q}^c}) = \int_{\mathcal{X}_{\mathfrak{q}}} g(\boldsymbol{x})^2 \lambda(\boldsymbol{x}_{\mathfrak{q}})\mathrm{d}\boldsymbol{x}_{\mathfrak{q}}$ 可表示为

$$p(\boldsymbol{x}_{\mathfrak{q}^c}) = \int_{\mathcal{X}_{\mathfrak{q}}} \Big(\sum_{\boldsymbol{j} \in \mathcal{K}} c_{\boldsymbol{j}} \prod_{i=1}^{d} \psi_{j_i}(x_i) \Big) \Big(\sum_{\boldsymbol{k} \in \mathcal{K}} c_{\boldsymbol{k}} \prod_{i=1}^{d} \psi_{k_i}(x_i) \Big) \prod_{i \in \mathfrak{q}} \lambda(x_i)\mathrm{d}\boldsymbol{x}_{\mathfrak{q}}$$

$$
\begin{aligned}
&= \sum_{\boldsymbol{j},\boldsymbol{k}\in\mathcal{K}} c_{\boldsymbol{j}}\, c_{\boldsymbol{k}} \Big(\prod_{i\in\mathfrak{q}^c} \psi_{j_i}(x_i)\psi_{k_i}(x_i) \Big) \Big(\prod_{i\in\mathfrak{q}} \int_{\mathcal{I}} \psi_{j_i}(x_i)\psi_{k_i}(x_i)\lambda(x_i)\mathrm{d}x_i \Big) \\
&= \sum_{\boldsymbol{j},\boldsymbol{k}\in\mathcal{K}} c_{\boldsymbol{j}}\, c_{\boldsymbol{k}} \Big(\prod_{i\in\mathfrak{q}^c} \psi_{j_i}(x_i)\psi_{k_i}(x_i) \Big) \Big(\sum_{\ell=1}^{r(\mathfrak{q})} \mathsf{L}^{(\mathfrak{q})}_{\sigma(\boldsymbol{j})\ell} \mathsf{L}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell} \Big) \\
&= \sum_{\ell=1}^{r(\mathfrak{q})} \Big(\sum_{\boldsymbol{k}\in\mathcal{K}} c_{\boldsymbol{k}} \mathsf{L}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell} \prod_{i\in\mathfrak{q}^c} \psi_{k_i}(x_i) \Big)^2 .
\end{aligned}
$$

常数 $\hat{}z$ 是对所有坐标上的 $(\gamma + g(\boldsymbol{x})^2)\lambda(\boldsymbol{x})$ 进行积分所得，因此也可以使用上述公式得到其表达式。由此得证。 □

注2.4（变量排序）。如命题 2.3所示，我们可以获得任意变量子集的边缘密度。因此，给定任意变量排序 $\mathfrak{p} = (p_1, \ldots p_d)$，我们可以使用上述定义的过程来计算与重排序变量 $\boldsymbol{X}_{\mathfrak{p}} = (X_{p_1}, \ldots X_{p_d})$ 相关联的 $\hat{}$ KR 重排。计算成本与变量排序无关。这比 [14, 15], 中的平方张量列车方法更加灵活，因为按任意顺序对变量进行边缘化可能会显著增加计算复杂度。

3. 多项式空间中的自适应最小二乘逼近。上一节概述的过程可以与任意乘积形式基函数结合使用。例如，带有乘积形式核的高斯过程和径向基函数 [6, 52, 53] 、多元小波和傅里叶基 [17, 18, 19, 20, 25, 32],以及稀疏谱多项式 [46, 45, 54, 55]。这里我们考虑多元基函数 $(\psi_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{N}_0^d}$，它们通过张量化正交归一单变量多项式 $(\psi_k)_{k=0}^{\infty}$ 获得，并满足 $\langle\psi_j,\psi_k\rangle_\lambda = \delta(j,\ k)$，其中 $\delta(j,\ k)$ 是克罗内克δ。这样，张量积基函数 $(\psi_{\boldsymbol{k}}) = (\psi_{k_1}) \otimes\cdots\otimes (\psi_{k_d})$ 是正交归一的，因为

$$
\langle\psi_{\boldsymbol{j}},\psi_{\boldsymbol{k}}\rangle_\lambda = \prod_{i=1}^{d}\langle\psi_{j_i},\psi_{k_i}\rangle_\lambda = \prod_{i=1}^{d}\delta(j_i,k_i) = \delta(\boldsymbol{j},\boldsymbol{k}). \tag{3.1}
$$

3.1. 正交多项式基。我们回顾一个关于 λ的众所周知的条件，该条件可确保多项式在 $L^2_\lambda(\mathbb{R})$中稠密，见 [1]。

PROPOSITION 3.1. 若存在常数$a > 0$ 使得，多项式在 $L^2_\lambda(\mathbb{R})$ 中稠密，

$$
\int_{\mathbb{R}} \exp(a|x|)\lambda(x)\mathrm{d}x < \infty. \tag{3.2}
$$

正交归一多项式$(\psi_k)_{k=0}^{\infty}$ （其中次数为$(\psi_k) = k$ ）构成$L^2_\lambda(\mathbb{R})$的希尔伯特基，因此任意多元函数 $u \in L^2_\lambda(\mathcal{X})$ 都可以表示为

$$
u(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{N}_0^d} u_{\boldsymbol{k}}\psi_{\boldsymbol{k}}(\boldsymbol{x}), \quad u_{\boldsymbol{k}} = \int u(\boldsymbol{x})\psi_{\boldsymbol{k}}(\boldsymbol{x})\mathrm{d}\lambda(\boldsymbol{x}), \quad \psi_{\boldsymbol{k}} : \boldsymbol{x} \mapsto \prod_{i=1}^{d}\psi_{k_i}(x_i).
$$

此外，序列 $u_{\boldsymbol{k}}$ 是平方可和的，并且对于任意 $\mathcal{K} \subset \mathbb{N}_0^d$，我们有

$$
\Big\|u - \sum_{\boldsymbol{k}\in\mathcal{K}} u_{\boldsymbol{k}}\psi_{\boldsymbol{k}}\Big\|^2_{L^2_\lambda(\mathcal{X})} = \sum_{\boldsymbol{k}\notin\mathcal{K}} u^2_{\boldsymbol{k}}. \tag{3.3}
$$

(3.2) 中的假设本质上是关于 λ 尾部的一个条件，许多测度都满足该条件。分解式 (3.3) 是分析基于正交基的近似的基石。在我们的设定中，目标

概率密度函数必然可积，因此将目标密度的平方根投影到正交基上，可得到平方可和系数。若目标函数满足额外的正则性假设——通常表现为全纯性或混合光滑性——则存在一个指标集序列 $\mathcal{K}_n$ ，其$|\mathcal{K}_n| = n$ 使截断误差 (3.3) 快速衰减（例如按代数速度衰减）。关于一般多项式近似的详情，请参阅 [22, 45, 46] ；关于其在逆问题中的应用，请参阅 [44, 43]; 关于 KR 重排的分析，请参阅 [57] 。此外，我们利用 (3.3) 指导第 3.2 节中指标集的构造。下面介绍一类众所周知的正交单变量多项式基。

DEFINITION 3.2. 回忆一下 *Beta*分布的密度函数，

$$\lambda^{\alpha,\beta}(x) = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}(1-x)^{\alpha}(1+x)^{\beta},$$

其中 $\alpha, \beta > -1$，适用于 $x \in \mathcal{I} = (-1,\ 1)$。*Jacobi*多项式 $(J^{\alpha,\beta}_k)_{k=0}^{\infty}$ 是 *Sturm–Liouville*问题的特征函数，

$$-\frac{1}{\lambda^{\alpha,\beta}(x)}\frac{\partial}{\partial x}\Big(\lambda^{\alpha+1,\beta+1}(x)\frac{\partial}{\partial x}J^{\alpha,\beta}_k(x)\Big) = \sigma^{\alpha,\beta}_k J^{\alpha,\beta}_k(x),$$

其特征值为 $\sigma^{\alpha,\beta}_k = k(k+\alpha+\beta+1)$。归一化多项式 $\psi^{\alpha,\beta}_k(x) = J^{\alpha,\beta}_k(x)/w^{\alpha,\beta}_k$，其中 $w^{\alpha,\beta}_k = \|J^{\alpha,\beta}_k\|_{\lambda^{\alpha,\beta}\mathrm{d}x} = \big(\int_{-1}^{1} J^{\alpha,\beta}_k(x)^2\lambda^{\alpha,\beta}(x)\mathrm{d}x\big)^{1/2}$，满足 $\langle\psi^{\alpha,\beta}_j, \psi^{\alpha,\beta}_k\rangle_{\lambda^{\alpha,\beta}\mathrm{d}x} = \delta_{jk}$。*Jacobi*多项式有一些值得注意的特殊情形。当 $\alpha = \beta = 0$ 时，我们得到勒让德多项式。当 $\alpha = \beta = -\frac{1}{2}$ 和 $\alpha = \beta = \frac{1}{2}$ 时，我们分别得到第一类切比雪夫多项式和第二类切比雪夫多项式。

以下命题表明，在使用正交基函数时，命题2.3中引入的质量矩阵具有简单的表达式。

PROPOSITION 3.3. 假设我们有如 *(3.1)* 中的正交基函数 $\{\psi_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}_0^d}$ 、一个多重指标集 $\mathcal{K} \subset \mathbb{N}_0^d$ 及其对应的矩阵形式 $K = matrix(\mathcal{K})$。设 $\mathfrak{q} \subseteq [d]$ 为一个坐标索引集。令 $\mathfrak{u}(\mathfrak{q}) = (u_1(\mathfrak{q}), \ldots\ u_m(\mathfrak{q})) \subseteq [|\mathcal{K}|]$ 为 $K_{:\mathfrak{q}}$ 的行指标集，使得子矩阵 $K_{\mathfrak{u}(\mathfrak{q})\mathfrak{q}}$ 包含 $K_{:\mathfrak{q}}$ 的所有唯一行。那么矩阵 $M^{(\mathfrak{q})}$ 可表示为

$$\mathsf{M}^{(\mathfrak{q})}_{jk} = \delta(\mathsf{K}_{j\mathfrak{q}}, \mathsf{K}_{k\mathfrak{q}}) = \prod_{i\in\mathfrak{q}}\delta(\mathsf{K}_{ji}, \mathsf{K}_{ki}),$$

对所有 $1 \le j, k \le |\mathcal{K}|$ 均成立。此外，$\mathsf{M}^{(\mathfrak{q})}$ 的秩为 $\mathrm{rank}(\mathsf{M}^{(\mathfrak{q})}) = m$，即 $\mathsf{K}_{:\mathfrak{q}}$ 的唯一行数。此外，$\mathsf{M}^{(\mathfrak{q})}$ 可分解为

$$\mathsf{M}^{(\mathfrak{q})} = \mathsf{P}^{(\mathfrak{q})}(\mathsf{P}^{(\mathfrak{q})})^{\top},\ \ \mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell} = \delta(\sigma(\boldsymbol{k}), u_\ell(\mathfrak{q})),\ \ \boldsymbol{k}\in\mathcal{K},\ \ \ell = 1, \ldots, |\mathfrak{u}(\mathfrak{q})|. \tag{3.4}$$

最后，对于 $\mathfrak{q} = [d]$，矩阵 $\mathsf{M}^{([d])}$ 是 $|\mathcal{K}| \times |\mathcal{K}|$ 单位矩阵。

证明。 给定坐标索引集 $\mathfrak{q} \subseteq [d]$，我们将基函数集合$\{\psi_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathcal{K}}$ 限制到 $\boldsymbol{x}_{\mathfrak{q}}$，由此得到一个多重集

$$\{\psi_{\boldsymbol{k}_{\mathfrak{q}}} : \boldsymbol{k}\in\mathcal{K}\},\quad \psi_{\boldsymbol{k}_{\mathfrak{q}}} : \boldsymbol{x}_{\mathfrak{q}} \mapsto \prod_{i\in\mathfrak{q}}\psi_{k_i}(x_i),$$

其基数为 $|\mathcal{K}|$。 $\{\psi_{\boldsymbol{k}_{\mathfrak{q}}}$ 中的元素： $\boldsymbol{k}\in\mathcal{K}\}$ 可能并不唯一。利用映射$\sigma : \mathcal{K} \to [|\mathcal{K}|]$，我们还可以用标量索引表示 $\{\psi_{\boldsymbol{k}_{\mathfrak{q}}} : \boldsymbol{k}\in\mathcal{K}\}$ ，从而得到

$$\{\phi_{\sigma(\boldsymbol{k})} : \boldsymbol{k}\in\mathcal{K}\},\quad \phi_{\sigma(\boldsymbol{k})}(\boldsymbol{x}_{\mathfrak{q}}) := \psi_{\boldsymbol{k}_{\mathfrak{q}}}(\boldsymbol{x}_{\mathfrak{q}}).$$

给定矩阵 $\mathsf{K} = \mathsf{matrix}(\mathcal{K})$，回想一下， $\{\boldsymbol{k}_{\mathfrak{q}}$ 中的每个元素： $\boldsymbol{k}\in\mathcal{K}\}$ 都对应于子矩阵 $\mathsf{K}_{:\mathfrak{q}}$ 的第 $\sigma(\boldsymbol{k})$ 行，而指标集 $\mathfrak{u}(\mathfrak{q}) \subseteq [|\mathcal{K}|]$ 的构造方式是——

使得子矩阵 $\mathrm{K}_{\mathfrak{u}(\mathfrak{q})\mathfrak{q}}$ 包含 $\mathrm{K}_{\cdot\mathfrak{q}}$的所有唯一行。这样，基函数集合 $\{\phi_j : \ j \in \mathfrak{u}(\mathfrak{q})\}$ 就包含多重集 $\{\phi_{\sigma(\boldsymbol{k})} : \ \boldsymbol{k} \in \mathcal{K}\}$中的唯一元素。多重集 $\{\phi_{\sigma(\boldsymbol{k})} : \ \boldsymbol{k} \in \mathcal{K}\}$ 中的每个元素都可以表示为

$$\phi_{\sigma(\boldsymbol{k})}(\boldsymbol{x}_{\mathfrak{q}}) = \sum_{\ell=1}^{|\mathfrak{u}(\mathfrak{q})|} \mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell}\, \phi_{u_\ell(\mathfrak{q})}(\boldsymbol{x}_{\mathfrak{q}}).$$

其中 $\mathsf{P}^{(\mathfrak{q})}$ 定义于（3.4）中。根据命题 2.3 及上述等式，有

$$\begin{aligned}
\mathsf{M}^{(\mathfrak{q})}_{\sigma(\boldsymbol{j})\sigma(\boldsymbol{k})} &= \int_{\mathcal{X}_{\mathfrak{q}}} \phi_{\sigma(\boldsymbol{j})}(\boldsymbol{x}_{\mathfrak{q}})\phi_{\sigma(\boldsymbol{k})}(\boldsymbol{x}_{\mathfrak{q}})\lambda(\boldsymbol{x}_{\mathfrak{q}})\mathrm{d}\boldsymbol{x}_{\mathfrak{q}} \\
&= \sum_{\ell=1}^{|\mathfrak{u}(\mathfrak{q})|}\sum_{\ell'=1}^{|\mathfrak{u}(\mathfrak{q})|} \mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{j})\ell}\mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell'} \int_{\mathcal{X}_{\mathfrak{q}}} \phi_{u_\ell(\mathfrak{q})}(\boldsymbol{x}_{\mathfrak{q}})\phi_{u_{\ell'}(\mathfrak{q})}(\boldsymbol{x}_{\mathfrak{q}})\lambda(\boldsymbol{x}_{\mathfrak{q}})\mathrm{d}\boldsymbol{x}_{\mathfrak{q}} \\
&= \sum_{\ell=1}^{|\mathfrak{u}(\mathfrak{q})|}\sum_{\ell'=1}^{|\mathfrak{u}(\mathfrak{q})|} \mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{j})\ell}\mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell'}\, \delta(\ell,\ell') = \sum_{\ell=1}^{|\mathfrak{u}(\mathfrak{q})|} \mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{j})\ell}\mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell},
\end{aligned}$$

其中倒数第二个等式源于集合 $\{\phi_j : j \in \mathfrak{u}(\mathfrak{q})\}$ 中的元素关于 λ 已归一化且相互正交。证明完毕。 □

在命题 2.3 中应用（2.18），并结合命题3.3中的分解（3.4），对 $g(\boldsymbol{x})^2$, 的边缘函数 $g(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{K}} c_{\boldsymbol{k}}\psi_{\boldsymbol{k}}(\boldsymbol{x})$ 在对由 $\mathfrak{q} \subseteq [d]$ 所标记的坐标积分后为

$$p(\boldsymbol{x}_{\mathfrak{q}^c}) = \int_{\mathcal{X}_{\mathfrak{q}}} g(\boldsymbol{x})^2\lambda_{\mathfrak{q}}(\boldsymbol{x}_{\mathfrak{q}})\mathrm{d}\boldsymbol{x}_{\mathfrak{q}} = \sum_{\ell=1}^{|\mathfrak{u}^{\mathfrak{q}}|}\Big(\sum_{\boldsymbol{k}\in\mathcal{K}} c_{\boldsymbol{k}}\mathsf{P}^{(\mathfrak{q})}_{\sigma(\boldsymbol{k})\ell}\prod_{i\in\mathfrak{q}^c}\psi_{k_i}(x_i)\Big)^2. \tag{3.5}$$

算法 3.1 KR 重排的求值。输入为权函数 λ、正交基函数系 $(\psi_k)_{k=0}^{\infty}$、多重指标集 $\mathcal{K}$、近似系数 $(c_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{K}}$、常数 γ、变量排序 $\mathfrak{p} = (p_1, \ldots\ldots\ p_d)$，以及参考随机变量 $\boldsymbol{U}$ 的一个实现 $\boldsymbol{u}$。

1: 函数 EvaluateKR(λ, $(\psi_k)_{k=0}^{\infty}$, $\mathcal{K}$, $(c_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{K}}$, γ, $\mathfrak{p}$, $\boldsymbol{u}$) 2: $\hat{z} \leftarrow \gamma + \sum_{\boldsymbol{k}\in\mathcal{K}} c_{\boldsymbol{k}}^2$ ▷ $\hat{z}$ 是归一化常数 3: $\boldsymbol{\xi} \leftarrow \mathcal{R}(\boldsymbol{u})$ ▷ $\boldsymbol{\xi}$ 是 $[0,1]^d$ 中均匀随机变量的一次实现 4: 对于 $t = 1, 2, \ldots\ d$，执行 5: $\mathfrak{q}_t \leftarrow \mathfrak{p}_{[d]\setminus[t]}$ 6: 找出指标集 $\mathfrak{u}^{\mathfrak{q}_t}$，并使用 (3.4) 计算 $\mathsf{P}^{(\mathfrak{q}_t)}$ 7: 按照 (3.6) 组装 $q(x_t)$ 8: 如果 $t = 1$，则 ▷ 由 $(\mathsf{P}^{(\mathfrak{q}_1)}, (c_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{K}}, \gamma, \hat{z})$ 定义的边缘分布 9: $x_{p_1} \leftarrow \mathcal{F}^{-1}_{\bar{X}_{p_1}}(\xi_{p_1})$ 10: 否则 ▷ 由 $(\mathsf{P}^{(\mathfrak{q}_{t-1})}, \mathsf{P}^{(\mathfrak{q}_t)}, (c_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{K}}, \gamma)$ 定义的条件分布 11: $x_{p_t} \leftarrow \mathcal{F}^{-1}_{\hat{X}_{p_t}|\hat{\boldsymbol{X}}_{\mathfrak{p}_{[t-1]}}}(\xi_{p_t}|\boldsymbol{x}_{\mathfrak{p}_{[t-1]}})$ ▷ $\boldsymbol{x}_{\mathfrak{p}_{[t-1]}} = (x_{p_1}, \ldots, x_{p_{t-1}})$ 12: 返回 $\boldsymbol{x} = (x_1, x_2, \ldots\ x_d)$，使得 $\boldsymbol{x} = \bar{\mathcal{F}}^{-1}_{\hat{\boldsymbol{X}}} \circ \mathcal{F}_{\boldsymbol{U}}(\boldsymbol{u})$

算法 3.1 概述了计算 KR 重排的过程。请注意，可以在离线阶段预先计算算法 3.1 的步骤 2 和步骤 6，以降低复杂度。在算法 3.1 的步骤 9 和步骤 11 中计算逆分布函数时，需要使用几种求根方法（例如假位法和牛顿法）进行若干次迭代，以达到接近机器精度的准确度。这需要针对固定的 $\boldsymbol{x}_{[t-1]}$ 多次计算条件密度 $x_t \mapsto f_{\hat{X}_t|\hat{\boldsymbol{X}}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]})$。为高效执行这些操作，可以预先进行一些计算

，然后再求逆条件分布函数。根据 $\hat{}_{\boldsymbol{X}}$ 的定义（2.10），在变量 $\boldsymbol{x}_{[t-1]}$ 固定时，可以将条件密度写成

$$f_{\hat{X}_t|\hat{\boldsymbol{X}}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) = \frac{1}{\zeta} q(x_t)\lambda(x_t), \tag{3.6}$$

其中

$$q(x_t) = \gamma + \sum_{\ell=1}^{r} \Big(\sum_{\boldsymbol{k}\in\mathcal{K}} \psi_{k_t}(x_t) \mathsf{C}_{\boldsymbol{k}\ell} \Big)^2, \quad \mathsf{C}_{\boldsymbol{k}\ell} = c_{\boldsymbol{k}} \mathsf{P}^{([d]\setminus[t])}_{\sigma(\boldsymbol{k})\ell} \prod_{i=1}^{t-1} \psi_{k_i}(x_i), \quad \zeta = \gamma + p(\boldsymbol{x}_{[t-1]}).$$

此外，由于 ψ_k 是多项式，函数 $q(x_t)$ 在 (3.6) 中是一个次数不超过 $2n_t$的多项式，其中 n_t 是坐标 x_t上的单变量多项式的最高次数，即 $n_t = \max\{k_t : \boldsymbol{k} \in \mathcal{K}\}$。因此，我们有

$$f_{\hat{X}_t|\hat{\boldsymbol{X}}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) = \Big(\sum_{j=0}^{2n_t} a_j \psi_j(x_t) \Big) \lambda(x_t),$$

其中系数 $a_j = a_j(\boldsymbol{x}_{[t-1]})$ 可通过在一维情况下采用任意精确配点法，并使用 $2n_t$ 个点预先计算得到。示例见附录 A.2 。与直接计算 (3.6) 相比，配点法的计算复杂度可能更低，因为在求根过程中无需对集合 $\mathcal{K}$ 上的平方项求和。对条件密度积分可得分布函数

$$\mathcal{F}_{\hat{X}_t|\hat{\boldsymbol{X}}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) = \sum_{j=0}^{2n_t} a_j \Big(\int_{-\infty}^{x_t} \psi_j(y)\lambda(y)\mathrm{d}y \Big). \tag{3.7}$$

在附录 A中，我们证明，许多常用正交多项式对应的 (3.7) 中的积分都可通过解析方法求得，其计算复杂度与 $2n_t$ 呈线性关系，包括区间 $(-1,1)$ 上的切比雪夫多项式、区间 $(-1,1)$ 上的勒让德多项式、定义在 $[0,\infty)$ 上的拉盖尔多项式，以及区间 $(-\infty,\infty)$ 上的厄米多项式。

备注 3.4. 记所有坐标上的单变量基函数的最大基数为 n。矩阵 $\mathsf{P}^{([d]\setminus[t])}$ 具有 $|\mathcal{K}|$ 个非零元素，如（3.4）所示。因此，在算法 3.1的每次迭代中，在配点上计算条件密度需要 $O(n|\mathcal{K}|)$ 次浮点运算（flops）。使用快速傅里叶变换或类似的快速变换获取（a_j）中的系数（$_{j=1}^{m}$ （3.7）还需要额外的 $O(n\log(n))$ 次浮点运算。算法 3.1 中计算 KR 重排的总体计算复杂度为 $\mathcal{O}(dn|\mathcal{K}| + dn\log(n))$。

备注 3.5. 尽管本小节重点讨论正交多项式，但命题 3.3 和算法3.1通常适用于任意正交基函数。

算法 3.2 使用正交基函数构造 KR 重排。输入包括目标密度的势函数 Φ、权函数λ、正交基函数系（ψ_k）$_{k=0}^{\infty}$以及相对误差阈值 τ。

1: **过程** ConstructKR(Φ, λ, $(\psi_k)_{k=0}^{\infty}$, τ)2: 找到多重指标集 $\mathcal{K}$ 和系数 $(c_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{K}}$，使得

$$\Big\| \sum_{\boldsymbol{k}\in\mathcal{K}} c_{\boldsymbol{k}} \psi_{\boldsymbol{k}}(\boldsymbol{x}) - \exp(-\tfrac{1}{2}\Phi) \Big\|_{L^2_\lambda(\mathcal{X})} \le \tau \big\| \exp(-\tfrac{1}{2}\Phi) \big\|_{L^2_\lambda(\mathcal{X})}.$$

3: 选择一个常数 $\gamma \le \tau^2 \| \exp(-\frac{1}{2}\Phi) \|^2_{L^2\lambda(\mathcal{X})}$

4: **返回** $\mathcal{K}$，$(c_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{K}}$和 γ，它定义了映射 $\mathcal{F}_{\hat{}_{\boldsymbol{X}}}$。

3.2. 自适应最小二乘逼近。 我们详细介绍如何构造指标集 $\mathcal{K}$ 以及如何基于加权最小二乘法计算 $g \in \mathcal{V}_{\mathcal{K}}$ 式（2.13）。从某个初始指标集 $\mathcal{K}_1 \subset \mathbb{N}_0^d$出发，我们采用贪心算法自适应地扩展当前指标集 $\mathcal{K}_n$，使其满足 $\mathcal{K}_{n+1} = \mathcal{K}_n \cup \{\boldsymbol{k}_n^1, \dots, \boldsymbol{k}_n^{m_n}\}$，其中 $\boldsymbol{k}_n^1$，…，$\boldsymbol{k}_n^{m_n} \in \mathbb{N}_0^d$是 m_n下文将定义的多重指标。借鉴 [34, 35], 中的思想，我们将贪心算法中的所有集合 $\mathcal{K}_n$ 限制为向下封闭 [10, 12], ，即

$$\forall \boldsymbol{k} \in \mathcal{K}_n \quad \text{and} \quad \forall \boldsymbol{k}' \in \mathbb{N}_0^d, \quad \boldsymbol{k}' \le \boldsymbol{k} \quad \Longrightarrow \quad \boldsymbol{k}' \in \mathcal{K}_n, \tag{3.8}$$

其中，$\boldsymbol{k}' \le \boldsymbol{k}$是一个偏序，使得对于所有 $i = 1, \dots, d$，$k_i' \le k_i$。向下封闭性质是设计可处理的 $\mathcal{K}_n$构造的关键，因为它不允许指标集 $\mathcal{K}_n$包含任何“空洞”，见 [12]。当且仅当所有 $\boldsymbol{k}_n^1, \dots, \boldsymbol{k}_n^{m_n}$都属于 $\mathcal{K}_n$ 的缩减边界时，新指标集 $\mathcal{K}_{n+1}$仍然向下封闭，该边界定义为

$$\mathcal{K}_n^{\mathrm{RM}} = \{\boldsymbol{k} \notin \mathcal{K}_n : \boldsymbol{k} - \mathbf{e}_i \in \mathcal{K}_n \text{ for all } 1 \le i \le d \text{ such that } k_i > 0\}, \tag{3.9}$$

其中，$\mathbf{e}_i$ 表示 $\mathbb{N}_0^d$的第 i个标准基向量。我们采用整体追踪策略[35] ，从 $\mathcal{K}_n^{\mathrm{RM}}$中选取多重指标。对于给定参数 $\theta \in (0, 1]$，我们选取最少数量 m_n的多重指标 $\boldsymbol{k}_n^1, \dots, \boldsymbol{k}_n^{m_n} \in \mathcal{K}_n^{\mathrm{RM}}$，使得

$$\sum_{\boldsymbol{k} \in \{\boldsymbol{k}_n^1, \dots, \boldsymbol{k}_n^{m_n}\}} e_n(\boldsymbol{k}) \ge \theta \sum_{\boldsymbol{k} \in \mathcal{K}_n^{\mathrm{RM}}} e_n(\boldsymbol{k}),$$

其中

$$e_n(\boldsymbol{k}) = \Big(\int \Big(\exp(-\tfrac{1}{2}\Phi(\boldsymbol{x})) - g_n(\boldsymbol{x}) \Big) \psi_{\boldsymbol{k}}(\boldsymbol{x}) \lambda(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \Big)^2,$$

$$g_n(\boldsymbol{x}) = \min_{h \in \mathcal{V}_{\mathcal{K}_n}} \int \Big(\exp(-\tfrac{1}{2}\Phi(\boldsymbol{x})) - h(\boldsymbol{x}) \Big)^2 \lambda(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

当 e_n 和 g_n 已知时，上述过程实际上会从缩减边界中选出使 L^2 误差衰减最快的候选指标集，参见（3.3）。在实践中，e_n 和 g_n 通过离散 L^2 投影和蒙特卡洛样本进行近似。我们采用最优加权最小二乘法 [11, 26, 35, 38] ，引入重要性抽样方案以提高离散 L^2投影的稳定性。该重要性抽样方案使用最优密度

$$\Lambda_n(\boldsymbol{x}) = \frac{1}{|\mathcal{K}_n|} \sum_{\boldsymbol{k} \in \mathcal{K}_n} (\psi_{\boldsymbol{k}}(\boldsymbol{x}))^2 \lambda(\boldsymbol{x}),$$

以及似然权函数 $w(\boldsymbol{x}) = \lambda(\boldsymbol{x})/\Lambda_n(\boldsymbol{x})$，由此得到以下估计量

$$\hat{e}_n(\boldsymbol{k}) = \Big(\frac{1}{N} \sum_{i=1}^{N} \big(\exp(-\tfrac{1}{2}\Phi(\boldsymbol{X}^{(i)})) - \hat{g}_n(\boldsymbol{X}^{(i)}) \big) \psi_{\boldsymbol{k}}(\boldsymbol{X}^{(i)}) w(\boldsymbol{X}^{(i)}) \Big)^2,$$

$$\hat{g}_n(\boldsymbol{x}) = \min_{h \in \mathcal{V}_{\mathcal{K}_n}} \frac{1}{N} \sum_{i=1}^{N} \big(\exp(-\tfrac{1}{2}\Phi(\boldsymbol{X}^{(i)})) - h(\boldsymbol{X}^{(i)}) \big)^2 w(\boldsymbol{X}^{(i)}),$$

其中 $\boldsymbol{X}^{(1)}$、……、$\boldsymbol{X}^{(N)}$是 $\boldsymbol{X} \sim \Lambda_n$ 的独立副本。与经典蒙特卡洛方法相比，最优加权最小二乘法保证了经验投影 $\hat{g}_n$ 的准最优性。

3.3. 定义域为 $\mathbb{R}^d$ 的密度的区域变换。与高斯权函数 $\lambda(\boldsymbol{x}) \propto \exp(-\frac{1}{2}\|\boldsymbol{x}\|_2^2)$相对应的厄米多项式，是在无界区域 $\mathbb{R}^d$上近似密度的自然候选方法；详情参见附录

A.3。然而，使用厄米多项式——或者当支撑集为$(0,\infty)^d$时使用拉盖尔多项式——来近似随机变量时，有几点需要注意。首先，在无界域上对分布函数求逆，在分布尾部附近可能会遇到数值上的困难。更重要的是，无界域上的权函数可能不满足式(2.11)中的尾部条件。事实上，命题3.2中的条件(3.3 断言，权函数 λ 的尾部衰减速度不得慢于指数函数。对于许多重尾随机变量，这不满足尾部条件(2.11)。这里我们讨论几种可用于规避这一问题的方法。

一种自然的选择是根据目标密度向尾部的先验衰减率进行区域截断。然而，这样做会引入额外的近似步骤，可能降低该方法的准确性。

区域变换方法 [4, 47] 是处理定义在$\mathbb{R}^d$ 且满足尾部条件（2.11）的密度的另一种选择。为简洁起见，我们定义定义在 $\mathbb{R}$上的单变量密度的域映射。通过取一维权函数的张量积，该方法可以推广到多维情形。作为起点，将实线 $\mathbb{R}$ 映射到区间（-1, 1）上，所用映射为微分同胚$z:\ \mathbb{R}\to$（-1，1）。给定定义在（ λ^{ref}上的参考权函数（-1，1）及其对应的正交归一基函数（ψ_i）$_{i=0}^{\infty}$，扩展基函数

$$(\phi_i)_{i=0}^{\infty},\quad \phi_i(x)=\psi_i(z(x)), \tag{3.10}$$

构成了关于权函数、定义在 $L^2_\mu(\mathbb{R})$ 上的一组正交归一基函数

$$\lambda(x)=\lambda^{\mathrm{ref}}(z(x))z'(x). \tag{3.11}$$

这可以通过变量替换 $x=x(z)$ 来证明，

$$\delta(j,k)=\int_{-1}^{1}\psi_j(z)\psi_k(z)\lambda^{\mathrm{ref}}(z)\mathrm{d}z=\int_{-\infty}^{\infty}\phi_j(x)\phi_k(x)\lambda(x)\mathrm{d}x.$$

表 1 列出了一些常用的域映射及其诱导权函数 $\lambda(x)$，它们基于 $\lambda^{\mathrm{ref}}(z)=\frac{1}{\pi}(1-z^2)^{-1/2}$，该函数是第一类切比雪夫多项式的权函数。

表 1：域映射、逆映射以及对于权函数 $\lambda^{\mathrm{ref}}(z)=\frac{1}{\pi}(1-z^2)^{-1/2}$，其诱导权函数 $\lambda(x)$ 在 $(-1,1)$ 上的示例。

	$z(x)$	$z'(x)$	$x(z)$	$x'(z)$	$\lambda(x)$
对数型	双曲正切(x)	$1-$ 双曲正切$(x)^2$	$\frac{1}{2}$对数$\left(\frac{1+z}{1-z}\right)$	$\frac{1}{1-z^2}$	$\frac{\text{双曲正割}(x)}{\pi}$
代数的	$\frac{x}{\sqrt{1+x^2}}$	$\frac{1}{(1+x^2)^{3/2}}$	$\frac{z}{\sqrt{1-z^2}}$	$\frac{1}{(1-z^2)^{3/2}}$	$\frac{1}{\pi(1+x^2)}$

我们发现，尾部条件（2.11）可以通过定义域变换来解决。例如，使用表1中的代数映射，目标密度与权函数之比满足

$$\log\Big(\frac{f_X(x)}{\lambda(x)}\Big)=\mathcal{O}\big(\log(1+x^2)+\log f_X(x)\big)=\mathcal{O}(\log|x|+\log f_X(x)).$$

因此，对于对数密度函数衰减速度快于对数速率的任意充分光滑目标密度，我们有尾部条件 $\sup_{x\in\mathbb{R}} f_X(x)/\lambda(x)<\infty$。事实上，柯西分布的均值和方差未定义，在这种情况下构成一个边界情形。

使用定义在 $\mathbb{R}$ 上的映射基函数（ϕ_j）$_{j=0}^{\infty}$ 以及诱导权函数 $\lambda(x)$，可以

直接对目标密度$f_X(x)$应用平方根近似过程。更有趣的是，如附录 B 所示，我们也可以使用逆映射 $x(z)$等价地拉回目标密度 $f_X(x)$，然后使用参考权重 $\lambda^{\mathrm{ref}}(z)$和定义在 $(-1,1)$ 上的基函数 $(\psi_j)_{j=0}^{\infty}$，对拉回密度 $f_Z(z) = f_X(x(z))x'(z)$应用平方根近似过程。这样一来，我们便可以复用针对有界域定义完善的实现过程，同时，基于定义在 $\mathbb{R}$上的映射基函数 $(\phi_j)_{j=0}^{\infty}$ 的收敛分析（参见 [47]及其中的参考文献）仍然适用。

4. 自强化近似。对于集中在较小子域内或具有复杂相关结构的目标密度，可能需要使用高阶指标集，才能一步构造出目标密度的近似 $f_{\boldsymbol{X}}$。以往的研究 [8, 43], 表明，适当的线性预条件化／缩放可以显著降低稀疏多项式近似的复杂度。本文提出了一个通用的非线性预条件框架，利用正交基函数近似集中概率密度。我们的方法具有自增强特性，因为它使用由已有函数近似定义的 KR 重排来预条件化下一步的函数近似，以处理复杂度不断增加的新密度。我们还给出了一些具体的实现策略。

4.1. 分层构造。给定一个密度为$f_{\boldsymbol{X}}$的目标随机变量 $\boldsymbol{X}$，我们最终构造出传输映射的复合

$$\mathcal{T}_L = \mathcal{Q}_1 \circ \mathcal{Q}_2 \circ \ldots \circ \mathcal{Q}_L,$$

该复合由一系列桥接密度引导

$$f_{\boldsymbol{X}_1}, f_{\boldsymbol{X}_2}, \ldots, f_{\boldsymbol{X}_L} := f_{\boldsymbol{X}}.$$

构造结束时，参考随机变量 $\boldsymbol{U}$经复合映射 $\mathcal{T}_L$ 推前后得到的分布近似于 $\boldsymbol{X}$。

复合映射中的层 $(\mathcal{Q}_\ell)_{\ell=1}^{L}$ 是递归构建的。在第 ℓ 层，复合映射 $\mathcal{T}_\ell = \mathcal{Q}_1 \circ \ldots \circ \mathcal{Q}_\ell$ 通过 $f_{\boldsymbol{X}_\ell} := (\mathcal{T}_\ell)_\sharp f_{\boldsymbol{U}}$近似桥接密度 $f_{\boldsymbol{X}_\ell}$。在新一轮迭代中，当前映射 $\mathcal{T}_\ell$ 将下一个桥接密度预处理为拉回测度 $\mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_{\ell+1}}$. 我们构造一个新的多项式近似来近似函数 $\sqrt{\mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_{\ell+1}}}$（更准确地说，是其未归一化版本），然后推导出相应的 KR 重排 $\mathcal{Q}_{\ell+1}$，使得 $(\mathcal{Q}_{\ell+1})_\sharp f_{\boldsymbol{U}}$ 按照第 2 节中的详细说明近似于 $\mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_{\ell+1}}$。这将复合映射更新为 $\mathcal{T}_{\ell+1} = \mathcal{T}_\ell \circ \mathcal{Q}_{\ell+1}$，该映射通过 $f_{\boldsymbol{X}_{\ell+1}} := (\mathcal{T}_{\ell+1})_\sharp f_{\boldsymbol{U}}$近似$f_{\boldsymbol{X}_{\ell+1}}$。在温和的假设下，以下命题说明了自强化构造过程的收敛性。

PROPOSITION 4.1. 假设存在一个常数 $\eta \in (0,1)$，使得桥接密度满足 上确界 $_{1\le\ell<L} D_{\mathrm{H}}\big(f_{\boldsymbol{X}_\ell}, f_{\boldsymbol{X}_{\ell+1}}\big) \le \eta$。假设每个映射 $\mathcal{Q}_\ell$ 都具有误差缩减性质，

$$D_{\mathrm{H}}\big((\mathcal{Q}_{\ell+1})_\sharp f_{\boldsymbol{U}}, \mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_{\ell+1}}\big) \le \omega D_{\mathrm{H}}\big(f_{\boldsymbol{U}}, \mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_{\ell+1}}\big) \tag{4.1}$$

其中某个常数为 $\omega \in [0,1)$。设初始映射 $\mathcal{Q}_1$ 满足 $D_{\mathrm{H}}((\mathcal{Q}_1)_\sharp f_{\boldsymbol{U}}, f_{\boldsymbol{X}_1}) \le \omega\eta$。那么，所得复合映射 $\mathcal{T}_L$ 满足

$$D_{\mathrm{H}}\big((\mathcal{T}_L)_\sharp f_{\boldsymbol{U}}, f_{\boldsymbol{X}_L}\big) \le \frac{\omega}{1-\omega}\eta.$$

证明。 由于海林格距离满足三角不等式，我们有

$$\begin{aligned} D_{\mathrm{H}}\big(f_{\hat{\boldsymbol{X}}_{\ell+1}}, f_{\boldsymbol{X}_{\ell+1}}\big) &= D_{\mathrm{H}}\big((\mathcal{T}_\ell \circ \mathcal{Q}_{\ell+1})_\sharp f_{\boldsymbol{U}}, f_{\boldsymbol{X}_{\ell+1}}\big) \\ &= D_{\mathrm{H}}\big((\mathcal{Q}_{\ell+1})_\sharp f_{\boldsymbol{U}}, \mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_{\ell+1}}\big) \end{aligned}$$

$$
\begin{aligned}
&\le \omega D_{\mathrm{H}}\big(f_{\boldsymbol{U}}, \mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_{\ell+1}}\big)\\
&= \omega D_{\mathrm{H}}\big(f_{\hat{\boldsymbol{X}}_\ell}, f_{\boldsymbol{X}_{\ell+1}}\big)\\
&\le \omega\Big(D_{\mathrm{H}}\big(f_{\hat{\boldsymbol{X}}_\ell}, f_{\boldsymbol{X}_\ell}\big) + D_{\mathrm{H}}\big(f_{\boldsymbol{X}_\ell}, f_{\boldsymbol{X}_{\ell+1}}\big)\Big)\\
&\le \omega\Big(D_{\mathrm{H}}\big(f_{\hat{\boldsymbol{X}}_\ell}, f_{\boldsymbol{X}_\ell}\big) + \eta\Big),
\end{aligned}
$$

对于任意 $0 \le \ell < L$。当有 $D_{\mathrm{H}}\big(f_{\hat{\boldsymbol{X}}_1}, f_{\boldsymbol{X}_1}\big) \le \omega\eta$ 时，我们有

$$
D_{\mathrm{H}}\big(f_{\hat{\boldsymbol{X}}_L}, f_{\boldsymbol{X}_L}\big) \le \omega^L\eta + \omega^{L-1}\eta + \ldots + \omega\eta \le \frac{\omega}{1-\omega}\eta,
$$

至此，证明完成。 □

式（4.1）中的误差缩减性质是一个自然的假设，因为它本质上表明，我们需要构造一个新映射 $\mathcal{Q}_{\ell+1}$，使其比恒等映射好 $\omega < 1$ 倍。我们还指出，通过局部逼近增强稀疏多项式逼近能力还有其他方法，参见 [13, 56] 及其中的参考文献。本文介绍的基于多项式的KR重排有望与这些局部逼近相结合，从而进一步提高精度。

4.2. 桥接密度与自适应。 这里我们在贝叶斯逆问题的背景下讨论桥接密度的构造。我们将观测数据记为 $\boldsymbol{y} \in \mathbb{R}^m$，将前向模型记为 $G : \mathcal{X} \to \mathbb{R}^m$，并假设观测噪声是独立同分布（i.i.d.）的零均值高斯随机变量，方差为 σ_n^2。似然函数的形式为

$$
\mathcal{L}^y(\boldsymbol{x}) \propto \exp(-\Phi^y(\boldsymbol{x})), \quad \Phi^y(\boldsymbol{x}) = \frac{1}{2\sigma_n^2}\|\boldsymbol{y} - G(\boldsymbol{x})\|^2,
$$

并将先验密度表示为 $\pi_0(\boldsymbol{x}) \propto \exp(-\Phi_0(\boldsymbol{x}))\,\lambda(\boldsymbol{x})$。这样，与 $\mathcal{L}^y(\boldsymbol{x})\pi_0(\boldsymbol{x})$ 成正比的目标后验密度可表示为

$$
f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{z}\,\exp(-\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x}))\,\lambda(\boldsymbol{x}), \quad z = \int_{\mathcal{X}} \exp(-\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x}))\,\lambda(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.
$$

下面给出一些桥接密度的示例。

示例 4.2. 温度调整。参照 [24], 的研究，可以使用一系列温度参数 $0 < \beta_1 \le \beta_1 \le \cdots \le \beta_L = 1$ 来构造桥接密度

$$
f_{\boldsymbol{X}_\ell} = \frac{1}{z_\ell}\exp(-\beta_\ell\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x}))\,\lambda(\boldsymbol{x}), \quad z_\ell = \int \exp(-\beta_\ell\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x}))\,\lambda(\boldsymbol{x})\mathrm{d}\boldsymbol{x},
$$

对于 $\ell = 1, \ldots L$，通过调整似然函数的温度。对于贝叶斯逆问题，温度 $\beta_\ell < 1$ 对应于方差逐渐增大的似然函数。

示例 4.3. 数据分批。考虑一个基数为 m的数据集，可以构造 L 个互不相交的索引集 $\mathcal{J}_1, \ldots \mathcal{J}_L$，使得对于所有$i \ne j, \quad i, \quad j \in [L]$ 以及 $\cup_{i=1}^L \mathcal{J}_i = [m]$均成立。对于每个 $\mathcal{J}_i$，我们可以定义一个数据失配函数

$$
\Phi^y(\boldsymbol{x}; \mathcal{J}_i) = \frac{1}{2\sigma_n^2}\sum_{j\in\mathcal{J}_i}(y_j - G_j(\boldsymbol{x}))^2,
$$

其中 $G_j : \mathcal{X} \to \mathbb{R}$ 是第 i个可观测量的前向模型。由此，可以定义一系列桥接密度如下

$$
f_{\boldsymbol{X}_\ell} = \frac{1}{z_\ell}\exp\Big(-\sum_{i=1}^{\ell}\Phi^y(\boldsymbol{x}; \mathcal{J}_i) - \Phi_0(\boldsymbol{x})\Big)\,\lambda(\boldsymbol{x}), \quad \ell = 1, \ldots, L,
$$

其中 $z_\ell = \int\exp(-\sum_{i=1}^{\ell}\Phi^y(\boldsymbol{x}; \mathcal{J}_i) - \Phi_0(\boldsymbol{x}))\,\lambda(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$. 这种桥接密度特别有用

当模拟前向模型的计算成本与可观测量的基数成正比增长时。

上述示例中的桥接密度可推广至具有非高斯分布观测噪声的问题，以及其他非贝叶斯推断问题。注意，这两类桥接密度可以结合使用。例如，可以在每个数据批次内进行温度调整。正如命题4.1所指出的，控制相邻桥接密度之间的海灵格距离，对于控制复合映射的总体海灵格误差至关重要。在我们的示例中，可以通过自适应地选择下一个温度或下一个数据批次来实现这一点。

借助示例 4.2，我们讨论序贯蒙特卡罗文献中提出的自适应退火方法，例如 [3, 28, 37], ，以便精确控制桥接密度中的海灵格距离。我们希望控制桥接密度之间的海灵格距离，使其满足 $D_{\mathrm{H}}\big(f_{\boldsymbol{X}_\ell},\ f_{\boldsymbol{X}_{\ell+1}}\big) \approx \eta$其中 $\eta \in (0,\ \mathbf{1})$。定义 $\Delta = \beta_{\ell+1} - \beta_\ell$后，海灵格距离的平方可表示为

$$D_{\mathrm{H}}(f_{\boldsymbol{X}_\ell}, f_{\boldsymbol{X}_{\ell+1}})^2 = \frac{1}{2}\int \Big(\sqrt{\frac{f_{\boldsymbol{X}_{\ell+1}}}{f_{\boldsymbol{X}_\ell}}} - 1\Big)^2 f_{\boldsymbol{X}_\ell}\mathrm{d}\boldsymbol{x} = 1 - \frac{I}{\sqrt{z_\ell z_{\ell+1}}}. \tag{4.2}$$

其中

$$I = \int \exp\Big(-\big(\frac{\Delta}{2} + \beta_\ell\big)\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x})\Big)\lambda(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.$$

我们可以利用当前的传输映射复合 $\mathcal{T}_\ell$ 来估计平方海林格距离，以选择温度增量 Δ。推前密度 $f_{\hat{\boldsymbol{X}}_\ell} = (\mathcal{T}_\ell)_\sharp f_{\boldsymbol{U}}$ 给出了一个重要性采样公式，用于估计各项

$$z_{\ell+1} = \int \frac{\exp(-(\Delta+\beta_\ell)\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x}))\lambda(\boldsymbol{x})}{f_{\hat{\boldsymbol{X}}_\ell}(\boldsymbol{x})} f_{\hat{\boldsymbol{X}}_\ell}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{4.3}$$

$$z_\ell = \int \frac{\exp(-\beta_\ell\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x}))\lambda(\boldsymbol{x})}{f_{\hat{\boldsymbol{X}}_\ell}(\boldsymbol{x})} f_{\hat{\boldsymbol{X}}_\ell}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{4.4}$$

$$I = \int \frac{\exp(-(\Delta/2+\beta_\ell)\Phi^y(\boldsymbol{x}) - \Phi_0(\boldsymbol{x}))\lambda(\boldsymbol{x})}{f_{\hat{\boldsymbol{X}}_\ell}(\boldsymbol{x})} f_{\hat{\boldsymbol{X}}_\ell}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{4.5}$$

在（4.2）中。可以抽取样本 $(X_\ell^i)_{i=1}^N \sim f_{\hat{\boldsymbol{X}}_\ell}$ 并计算

$$F_\ell^i = \Phi^y(\boldsymbol{X}_\ell^i), \tag{4.6}$$

$$K_\ell^i = \beta_\ell\Phi^y(\boldsymbol{X}_\ell^i) + \Phi_0(\boldsymbol{X}_\ell^i) - \log\lambda(X_\ell^i) + \log f_{\hat{\boldsymbol{X}}_\ell}(X_\ell^i) \tag{4.7}$$

然后，由（4.2）–（4.5）中的公式可得估计的平方海林格距离

$$D_{\ell,+}(\Delta) := 1 - \frac{\sum_{i=1}^N \exp(-\frac{\Delta}{2}F_\ell^i - K_\ell^i)}{\sqrt{\sum_{i=1}^N \exp(-K_\ell^i)}\sqrt{\sum_{i=1}^N \exp(-\Delta F_\ell^i - K_\ell^i)}}. \tag{4.8}$$

根据柯西-施瓦茨不等式，它是非负的（当 $\Delta \geq 0$ 时）。

为了在数值上满足命题 4.1 中的误差缩减性质，我们还希望估计桥接密度 $f_{\boldsymbol{X}_\ell}$及其近似 $f_{\hat{\boldsymbol{X}}_\ell}$之间的（平方）海林格距离。按照与上述类似的推导，我们有

$$D_{\mathrm{H}}(f_{\hat{\boldsymbol{X}}_\ell}, f_{\boldsymbol{X}_\ell})^2 \approx D_\ell := 1 - \frac{\sum_{i=1}^N \exp(-\frac{1}{2}K_\ell^i)}{\sqrt{\sum_{i=1}^N \exp(-K_\ell^i)}}, \tag{4.9}$$

其中， K_ℓ^i 定义见（4.7）。根据詹森不等式，估计量 D_ℓ 为非负的。估计量 $D_{\ell,+}(\Delta)$ 和 D_ℓ 关于 K_ℓ^i具有平移不变性。

算法 4.1 具有自适应退火的自强化 KR 重排。

过程 LayeredKR(Φ^y, Φ_0, λ,$(\psi_k)_{k=0}^{\infty}$, β_1, ω, η)
  $\ell \leftarrow 1$
  $\mathcal{F}_{\hat{\boldsymbol{X}}_\ell} \leftarrow$ ConstructKR(Φ', λ, $(\psi_k)_{k=0}^{\infty}$, τ) ▷ $\Phi' = \beta_\ell \Phi^y + \Phi_0$, $\tau = \omega\eta$
  $\mathcal{T}_\ell \leftarrow \mathcal{F}_{\hat{\boldsymbol{X}}_\ell}^{-1} \circ \mathcal{F}_{\boldsymbol{U}}$ ▷ $\mathcal{Q}_1 \equiv \mathcal{T}_1$
  当 $\beta_\ell < 1$ **时执行**
    $\ell \leftarrow \ell+1$
    $(\beta_\ell, \varepsilon_{\ell-1}) \leftarrow$ NextBeta(Φ^y, Φ_0, $\beta_{\ell-1}$, $\mathcal{T}_{\ell-1}$, η)
    $\mathcal{F}_{\hat{\boldsymbol{X}}_\ell} \leftarrow$ ConstructKR(Φ', λ, $(\psi_k)_{k=0}^{\infty}$, τ) ▷ Φ'=对数 $\lambda-$ 对数 $\mathcal{T}_\ell^\sharp f_{\boldsymbol{X}_\ell}$, $\tau=\omega\varepsilon_{\ell-1}$
    $\mathcal{Q}_\ell \leftarrow \mathcal{F}_{\hat{\boldsymbol{X}}_\ell}^{-1} \circ \mathcal{F}_{\boldsymbol{U}}$, $\mathcal{T}_\ell \leftarrow \mathcal{T}_{\ell-1} \circ \mathcal{Q}_\ell$
  $L \leftarrow \ell$ 和 **返回** $\mathcal{T}_L$

过程 NextBeta(Φ^y, Φ_0, β_ℓ, $\mathcal{T}$, η)
  抽取样本 $\boldsymbol{X}_\ell^i = \mathcal{T}(\boldsymbol{U}_i)$，其中 $\boldsymbol{U}_i \sim f_{\boldsymbol{U}}$对于 $i = 1, \ldots N$。
  $\varepsilon_\ell \leftarrow \sqrt{D_\ell}$ 其中 D_ℓ 由式（4.9）给出。 ▷ 误差估计
  求 Δ 使得 $D_{\ell,+}(\Delta) = \eta^2$，其中 $D_{\ell,+}(\Delta)$ 由式（4.8）给出。
  返回 $\beta' = \min(\beta + \Delta, 1)$ 和 ε_ℓ。

5. 数值示例。 我们通过三个由常微分方程和偏微分方程控制的逆问题，展示所提出方法的效率。我们比较了使用稀疏多项式构建的自增强 KR 重排（参见第3节）与 的张量列车（TT）构造，比较自增强构造与单层构造，并测试各种桥接密度的性能。最后，我们还给出一些结论，以讨论数值结果。请注意，加权最小二乘逼近以及算法 4.1 中使用的自适应过程均基于随机估计。因此，在所有示例中，我们重复进行 9 次实验，以报告性能结果的均值和标准差。

5.1. 易感-感染-移除模型。我们首先将所提出的方法应用于估计分舱易感-感染-移除（CSIR）模型的参数。这里考虑的 CSIR 模型是[23]中模型的简化版本。我们首先将一个地理区域划分为 $K \in \mathbb{N}$ 个分舱，并将给定时刻 t 第 k个分舱中易感个体、感染个体和移除个体的数量分别记为 $S_k(t)$、$I_k(t)$ 和 $R_k(t)$。随后，每个分舱内部个体之间以及不同分舱之间的相互作用由以下微分方程组建模

$$\begin{cases} \dot{S}_k &= -\theta_k S_k I_k + \frac{1}{2}\sum_{j\in\mathcal{J}_k}(S_j - S_k), \\ \dot{I}_k &= \theta_k S_k I_k - \nu_k I_k + \frac{1}{2}\sum_{j\in\mathcal{J}_k}(I_j - I_k), \\ \dot{R}_k &= \nu_k I_k + \frac{1}{2}\sum_{j\in\mathcal{J}_k}(R_j - R_k), \end{cases} \tag{5.1}$$

其中，$\mathcal{J}_k$ 是包含第 k 个隔室所有邻居的指标集。上述微分方程组的行为由参数 $\theta_k \in \mathbb{R}$ 和 $\nu_k \in \mathbb{R}$ 控制，适用于 $k = 1, \ldots K$；它们分别表示第 k 个隔室的感染率和康复率。

求解 (5.1) 中的微分方程，时间区间为 $t \in [0, 5]$，固定的非齐次初始状态为 $S_k(0) = 99 - K + k$、$I_k(0) = K + 1 - k$ 和 $R_k(0) = 0$，其中 $k = 1, \ldots K$。我们旨在根据在 6 个等间距时间点对 $I_k(t)$ $k = 1$、$\ldots$ K 的含噪观测，估计未知参数$\boldsymbol{x} = (\theta_1, \nu_1, \ldots \theta_K, \nu_K) \in \mathbb{R}^{2K}$。对于 θ_k 和 ν_k 中的每一个，我们在定义域 $[0, 2]$ 上指定均匀先验，由此得到 $\pi_0(\boldsymbol{x}) = \prod_{k=1}^{2K} 1_{[0,2]}(x_k)$。在本实验中，我们使用一个“真实的”

参数 $\boldsymbol{x}_{\mathrm{true}} = [0.1, 1, \ldots 0.1, 1]$, 来模拟合成观测值

$$\boldsymbol{y}_{k,j} = I_k\Big(\frac{5j}{6}; \boldsymbol{x}_{\mathrm{true}}\Big) + \eta_{k,j}, \quad k = 1, \ldots, K, \quad j = 1, \ldots, 6,$$

其中 $\eta_{k,\ j}$是均值为零的标准高斯测量噪声 $\mathcal{N}(0,\ 1)$。由此得到似然函数

$$\mathcal{L}^y(\boldsymbol{x}) \propto \exp(-\Phi^y(\boldsymbol{x})), \quad \Phi^y(\boldsymbol{x}) = \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{6}\Big\{I_k\Big(\frac{5j}{6}; \boldsymbol{x}\Big) - \boldsymbol{y}_{k,j}\Big\}^2. \tag{5.2}$$

我们考虑定义在一维格点上的隔室模型，其中 k号分舱仅与索引相邻的分舱相连 $k-1$和 $k+1$。我们施加周期边界条件，使得 $Z_{K+1} = Z_1$ 且$Z_0 = Z_K$ 对于 $Z \in \{S, I, R\}$成立。通过改变分舱数量，进而改变参数维度，我们可以测试所提方法的可扩展性。在时间区间（5.1）内求解微分方程，采用固定的非齐次初始状态 $t \in [0, 5]$ $S_k(0) = 99-K+k,$ $I_k(0) = K+1-k,$ 以及 $R_k(0) = 0$ 对于 $k = 1,$ ..... K。采用带自适应时间步长的显式龙格–库塔方法求解微分方程，并控制绝对误差和相对误差均不超过 10^{-6}。

单层构造与自增强构造的比较。我们展示了在含一个分舱的 CSIR 模型（即 $K = 1$）中使用多层构造的优势。我们采用自适应退火、每个坐标上最高阶为 30 的勒让德多项式，以及停止容差为 0。05 的最小二乘近似，构建自增强 KR 重排。在构造过程中，算法4.1 生成了三层稀疏多项式近似，其基数为 $\{293 \pm 55, 268 \pm 18, 189 \pm 2\}$。构建复合映射总共需要 2420 ± 348次密度评估。所得海灵格误差为0.0181 ± 0.0046。为构建单层 KR 重排，我们采用每个坐标上最高阶为 60 的勒让德多项式。在此示例中，使用基数为 3721 的全部多项式基（加权最小二乘过程中需要进行 14884 次密度评估）后，所得 KR 重排的海灵格误差为 0.375 ± 0.023。与单层构造相比，自增强构造的误差约为 5%，但使用的基函数总数仅为 20%。值得一提的是，对于分舱数量更多的问题，我们无法应用单层 KR 重排。图1 展示了通过自增强构造和单层构造得到的近似后验密度之间的比较。

稀疏多项式与张量列车。 我们测试了自增强稀疏多项式构造 KR 重排的维度可扩展性，并将其性能与 [14]的基于TT的构造进行比较。我们测试了包含 $K \in \{1, 2, 3, 4\}$ 个分舱的 CSIR 模型。对于基于稀疏多项式和基于TT的构造，我们都采用均匀参考测度，并在每个坐标上使用最高阶数为 30 的勒让德多项式。为通过自适应退火构造自增强 KR 重排，算法4.1 使用初始温度调整参数$\beta_1 = 10^{-3}$ 和自适应阈值 $\eta = 0.1$。我们首先设置基于TT的构造，其中算法 4.1 中的基于多项式的 KR 重排替换为基于TT的 KR 重排。每一层的 TT 近似通过 2 次 TT-Cross 迭代完成，SVD截断容差为 10^{-2} ，最大 TT秩为 10。在基于稀疏多项式的构造中，我们使每层的密度评估次数与基于TT的构造相匹配，而不对估计误差设置停止容差。所得映射的海林格距离、

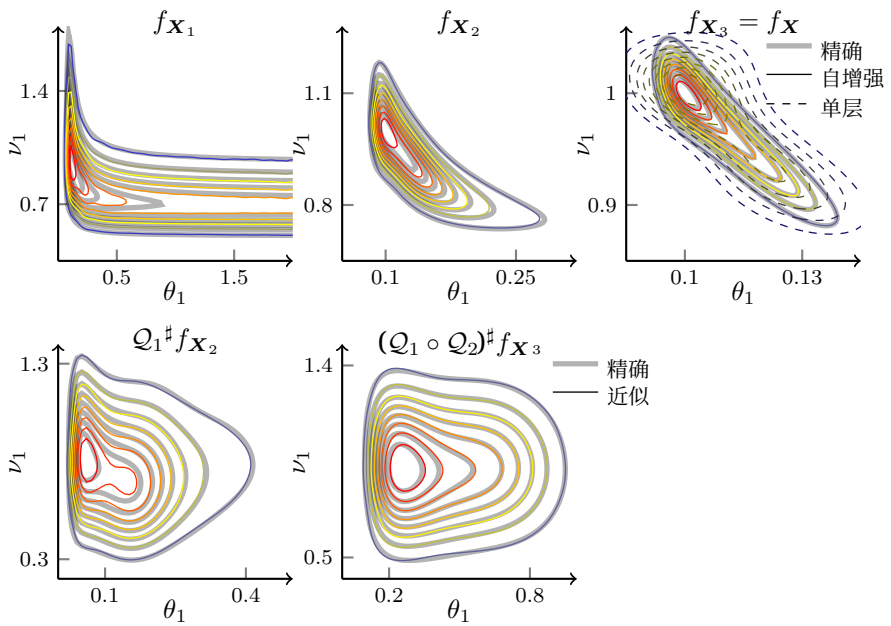


图1：具有 $K = 1$ 的 CSIR 示例。上排：粗灰色等高线表示桥接密度，实线等高线表示自强化近似，右侧虚线等高线表示单层近似。下排：用于自增强构造的拉回密度（粗灰色）及相应的稀疏多项式近似（实线）。

所用密度评估次数以及各参数维度下复合映射中的层数（$d = 2K$）列于图 2。与基于TT的构造相比，基于多项式的构造在较低维设置下精度略低；随着维度增加，精度差距逐渐缩小。

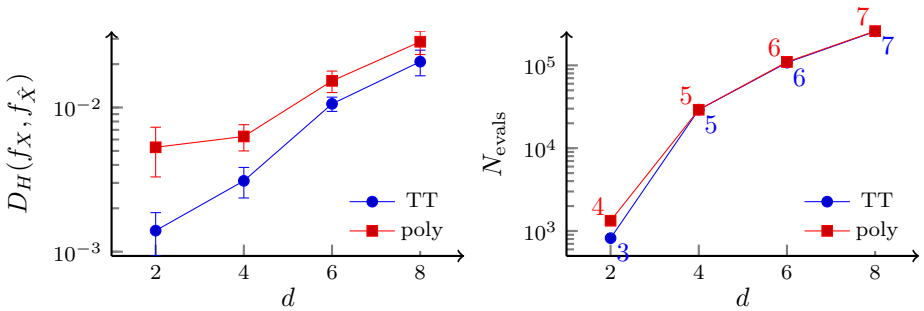


图2：CSIR 示例。各参数维度（$d = 2K$）下的海林格距离（左）和密度评估次数（右）。右图中的标签表示复合映射中的层数。

5.2. 椭圆型偏微分方程。接下来，我们考虑地下水建模中常见的椭圆型偏微分方程参数估计问题。给定问题域$D = [0, 1]^2$，我们希望根据地下水位 $u(s, \boldsymbol{X})$的部分观测，估计未知的扩散率场 $\kappa(s, \boldsymbol{X})$ $s \in D$；该地下水位是满足此偏微分方程的函数

$$-\nabla \cdot (\kappa(s, \boldsymbol{X}) \nabla u(s, \boldsymbol{X})) = 0, \quad s \in (0, 1)^2, \tag{5.3}$$

并施加狄利克雷边界条件 $u|_{s_1=0} = 1 + s_2/2$，并在水平方向施加 $u|_{s_1=1} = -\sin(2\pi s_2) - 1$，在垂直方向施加无通量边界条件 $\partial u/\partial s_2|_{s_2=0} = \partial u/\partial s_2|_{s_2=1} = 0$。对于 $\boldsymbol{X}$的每个实现，我们采用连续双线性有限元伽辽金方法，对方程（5.3）进行数值求解。有限元解 u_h 在 D 上均匀矩形网格的每个 $h = 1/64$坐标方向上以网格尺寸 D进行计算。

我们使用均值为零的均匀先验随机场对扩散率场的对数进行建模，其协方差算子由 Mat´ern 协方差函数定义

$$C(s,t) = \frac{2^{1-\nu}}{\Gamma(\nu)}\Big(\sqrt{2\nu}\frac{\|s-t\|_2}{\ell}\Big)^{\nu} K_\nu\Big(\sqrt{2\nu}\frac{\|s-t\|_2}{\ell}\Big), \quad s,t \in D,$$

其中 $\nu = 2$ and $\ell = 1$。利用 Karhunen–L´oeve (KL) 展开，log $\kappa(s, \boldsymbol{X})$ 得到一个有限维近似

$$\log \kappa(s, \boldsymbol{X}) \approx \sum_{k=1}^{d} X_k \sqrt{\omega_k}\varphi_k(s),$$

其中 $\{\varphi_k(s)\ \omega_k\}$ 是按特征值降序排列的协方差算子的第 k 个特征对，每个随机系数 X_k 均服从均值为零、方差为一的均匀先验，即 $X_k \sim \text{uniform}(-\sqrt{3}, \sqrt{3})$。

为建立观测模型，我们在 $D = [0,1]^2$ 上均匀笛卡尔网格的顶点处定义的 $m = 15\times15$ 个位置测量地下水位 $u(s, X)$，网格大小为 $1/(\sqrt{m}+1)$。测量值受到均值为零、方差为 $\sigma_n^2 = 10^{-2}$的独立同分布高斯噪声的干扰。由此得到参数到观测量映射

$$y_i = G_i(\boldsymbol{x}) + \eta_i, \quad G_{i_1+(i_2-1)\sqrt{m}}(\boldsymbol{x}) = u_h\Big(\Big(\frac{i_1}{\sqrt{m}+1}, \frac{i_2}{\sqrt{m}+1}\Big), \boldsymbol{x}\Big) \tag{5.4}$$

对于 i_1，$i_2 = 1$，...，15，其中 $\eta_i \sim \mathcal{N}(0,\ \sigma_n^2)$。合成数据生成为 $\boldsymbol{y} = G(\boldsymbol{x}_{\text{true}}) + \eta$，其中 $\boldsymbol{x}_{\text{true}} = (0.5,\ \ldots\ 0.5)$。

稀疏多项式与张量列车的比较。我们采用数量不同的 KL基函数，$d \in \{6, 8, 10, 12, 14, 16\}$，以测试自增强稀疏多项式 KR 重排构造的维度可扩展性，并将其性能与 [14] 的张量列车（TT）构造进行比较。对于基于稀疏多项式的构造和基于TT的构造，我们均采用均匀参考测度，并在每个坐标上使用最高阶数为 20 的勒让德多项式。为通过自适应退火构建自增强 KR 重排，算法 4.1 采用初始退火参数 $\beta_1 = 10^{-3}$ 和自适应阈值 $\eta = 0.5$。在每一层中，TT近似通过 2 次 TT-Cross 迭代完成，SVD截断容差为 10^{-2}，最大 TT秩为 5。在基于稀疏多项式的构造中，我们将每层的密度评估次数与基于TT的构造相匹配，而不设置估计误差的停止容差。图 3 报告了每个 KL 维度 d 对应的所得映射的海林格散度、所用密度评估次数以及复合映射中的层数。与基于TT的构造相比，基于多项式的构造在所有 d 情况下都更准确，并且在 $d \le 14$ 时所需的密度评估次数更少。当 $d = 16$ 时，基于TT的构造所需的密度评估次数更少，且与基于多项式的构造之间的精度差距也更小。

温度调整与数据分批。 我们还比较了基于稀疏多项式构建复合映射的退火方法（参见示例 4.2）和数据分批方法（参见示例 4.3）。表中报告了所得映射的海林格散度、所用的密度评估次数以及复合映射中的层数

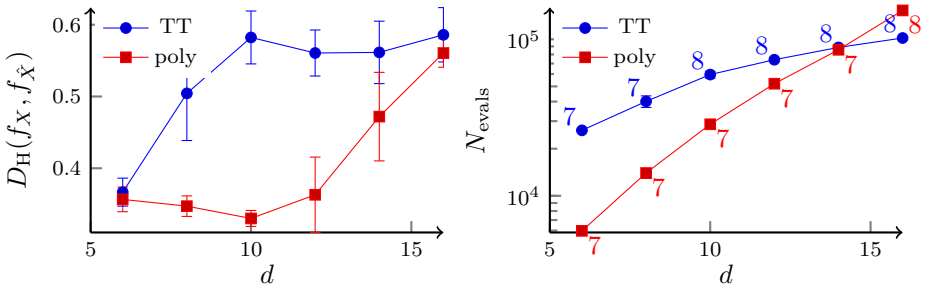


图 3：椭圆示例。不同参数维度下的海林格距离（左）和密度评估次数（右）。右图中的标签表示复合映射中的层数。

对于 $d \in \{6, 12\}$，相关结果列于表 2。尽管数据分批方法的精度略高，但它使复合映射中的层数增加一倍，所需的密度评估次数也增加一倍。

表 2：椭圆示例。不同参数维度和不同桥接密度下的海林格距离、密度评估次数以及复合映射中的层数。

d	温度调整			数据分批		
	$D_H(f_X, f_{\hat{X}})$	N_{evals}	$\#\{\beta\}$	$D_H(f_X, f_{\hat{X}})$	N_{evals}	$\#\{\beta\}$
6	0.36 ± 0.017	5972 ± 139	7	0.30 ± 0.023	14378 ± 568	16
12	0.36 ± 0.053	52019 ± 2209	7	0.33 ± 0.017	138756 ± 4557	16

单层构造与自增强构造的比较。 为了使单层构造可行，我们仅使用 9 个观测值，位置为 $m = 3 \times 3$（5.4）中指定的位置。为了展示后验集中度和参数维数变化时的比较，我们考虑三个实验：一个基准实验，其中 $d = 6$，且观测噪声方差较大 $\sigma_{\mathrm{n}}^2 = 10^{-1}$；一个维度递增的实验 $d = 12$，方差保持不变 $\sigma_{\mathrm{n}}^2 = 10^{-1}$；以及一个维度为 $d = 6$、方差降低的实验 $\sigma_{\mathrm{n}}^2 = 10^{-2}$。对于两种构造，我们将最小二乘过程的停止容差设为 0.01，并将多项式基函数的最大基数设为 3.2×10^4。表 3报告了两种构造所使用的海林格距离、密度评估次数以及基函数总数。在此情形下，自增强构造在基准实验中的效率略高于单层构造。无论是维数增加还是观测噪声方差降低（因此后验更加集中），自增强构造的计算效率都远高于单层构造。对于 $d = 12$ 和 $\sigma_{\mathrm{n}} = 10^{-2}$ 实验，单层构造都用尽了基函数的最大基数。对于 $d = 12$，它使用多出 6.5 倍的基函数，达到了与自增强构造相当的精度。对于 $\sigma_{\mathrm{n}} = 10^{-2}$，其误差约为自增强构造的十倍。

5.3. 扳手的线弹性分析。 最后，我们考虑一个更复杂的线弹性问题 [31, 48], ，其中物理体 $\mathcal{D} \subset \mathbb{R}^2$ 的位移场 $u : \mathcal{D} \to \mathbb{R}^2$在外力 f作用下处于平衡状态。给定应变场 $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^\top)$，位移 u 满足偏微分方程

$$\operatorname{div}(K : \varepsilon(u)) = f \quad \text{on} \quad \mathcal{D} \subset \mathbb{R}^2,$$

表3：椭圆示例。单层和自增强构造的海林格距离、密度评估次数以及基函数总数。

$(d,\ \sigma_n^2)$	自增强 $D_H(f_X, \hat{f}_X)$	N_{evals}	# 基	$\#\{\beta\}$	单层 $D_H(f_X, \hat{f}_X)$	N_{evals}	# 基
$(6, 10^{-1})$	4.1e-3±8.6e-4	2.5e3±2.8e2	8.6e2±8.4e1	2	4.9e-3±1.4e-3	9.5e3±8.7e2	3.2e3±2.3e2
$(12,\ 10^{-1})$	4.8e-3±1.1e-4	1.4e4±2.4e3	4.9e3±8.0e2	2	6.6e-3±4.6e-4	9.6e4	3.2e4
$(6, 10^{-2})$	4.3e-3±8.5e-4	2.5e4±3.3e3	8.1e3±7.9e2	3	7.6e-2±9.4e-3	1.28e5	3.2e4

其中， K 为胡克张量，满足

$$K : \varepsilon(u) = \frac{\exp(\theta)}{1+\nu}\varepsilon(u) + \frac{\nu\exp(\theta)}{1-\nu^2}\mathrm{trace}(\varepsilon(u))\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},$$

$\nu = 0.3$为泊松比，$\exp(\theta)$： $\mathcal{D} \to \mathbb{R}_{>0}$ 为随空间变化的杨氏模量。我们的目标是利用位移的部分观测，估计杨氏模量的对数 θ ： $\mathcal{D} \to \mathbb{R}$。

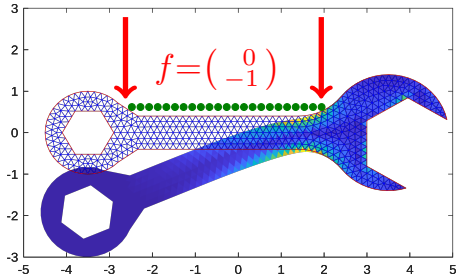


图 4：线弹性示例的设置。

我们假设未知的杨氏模量对数服从高斯随机场先验$\theta \sim \mathcal{N}(0, C)$，其中 $C(s,t) = \exp(-\|s-t\|_2^2)$是定义域$\mathcal{D}\times\mathcal{D}$上的高斯协方差函数。经过有限元离散化后， θ 被替换为分片常数场，其各单元上的取值汇集为向量$\theta \in \mathbb{R}^{d_\theta}$，其中 $d_\theta = 925$ 为网格中的单元数。因此，参数 θ 服从均值为零、协方差矩阵为 $\Sigma_{ij} = C(s_i, s_j)$的高斯先验，其中 $s_i \in \mathcal{D}$ 为第 i个单元的中心。我们通过将 $u = u(\theta)$ Galerkin投影到三角形网格上的连续分片仿射函数空间，计算数值解 $u^h = u^h(\theta)$。定义域 $\mathcal{D}$、网格、边界条件以及解的一个冯·米塞斯应力样本如图 4 所示。我们观测沿绿色线分布的 $m = 26$ 个关注点的竖向位移 u_2^h，力 f 施加于该线上，见图 4。受扰观测值为 $y = u_2^h + e$，其中 e 为均值为零的$\mathbb{H}^1$正态噪声，信噪比为 10。

我们不直接对高维离散参数 θ进行运算，而是采用 [15, 16] 的数据无关降维方法来压缩参数维度。我们首先通过对先验分布下的费舍尔信息矩阵进行积分，构造敏感性矩阵 $H \in \mathbb{R}^{d_\theta \times d_\theta}$，然后计算矩阵铅笔 $(H,\ C^{-1})$ 的特征对 $(\varphi_i,\ \nu_i)$，并按特征值 ν_i降序排列。由此得到重新参数化形式 $\theta = \sum_{i=1}^{d_\theta} \varphi_i x_i$，其中每个 x_i 都服从零均值标准高斯先验。根据特征值 ν_i截断重新参数化形式，即 $\theta = \sum_{i=1}^{d} \varphi_i x_i$，其中 $d < d_\theta$，从而得到随机场的降维参数化。

然后，我们构造自增强 KR 重排来近似

降维后验分布。我们将高斯先验截断在 ± 5个标准差处，从而得到截断的近似域 $[-5,5]^d$。对于基于稀疏多项式和基于 TT 的构造，我们采用与椭圆型偏微分方程示例完全相同的设置。所得映射的海灵格距离、所用密度评估次数，以及每个降维参数维度 d 下复合映射的层数，报告于表 4中。与基于 TT 的构造相比，基于多项式的构造精度相近，且在 $d=7$下所需的密度评估次数相近。对于 $d=11$，基于多项式的构造精度略低于基于 TT 的构造。

表4：线弹性示例。海林格距离、密度评估次数以及复合映射中的层数。

d	TT			稀疏多项式		
	$D_H(f_X, \hat{f}_X)$	N_{evals}	$\#\{\beta\}$	$D_H(f_X, \hat{f}_X)$	N_{evals}	$\#\{\beta\}$
7	0.21±0.024	44667±2083	7	0.22±0.020	45863±1077	7
11	0.41±0.083	92155±385	8	0.49±0.033	108162±6473	8

5.4. 结论。在数值示例中，使用稀疏多项式构造的自增强 KR 重排，与基于张量列车（TT）的方法性能相近。自增强构造显著提升了稀疏多项式近似的近似能力。在 CSIR 模型中，如果不采用自增强构造，甚至无法扩展到两个以上的分舱。对于椭圆型偏微分方程，单层构造仅适用于后验密度较不集中的情况——这可以通过增大观测噪声并减少观测值数量实现；而自增强构造则不受这些限制。我们还证明，数据分批是构建桥接密度的有效方法。尽管在椭圆型偏微分方程示例中，数据分批方法使用的层数多于退火方法，但对于似然评估成本与数据维数成正比的问题，它可能更具优势。此外，数据分批方法天然更适用于序贯实验设计，即依次观测数据的情形；详情请参见 [27] 及其中的参考文献。

致谢。 TC 感谢澳大利亚研究委员会通过资助项目 DP210103092 提供支持。SD 感谢工程与物理科学研究委员会新研究者奖 EP/T031255/1 提供支持。OZ 感谢 ANR JCJC 项目 MODENA (ANR-21-CE46-0006-01) 提供支持。

附录 A. 多项式基与分布函数。为实现分布函数的代数精确计算，我们在此给出一套方法，用于计算一系列多项式和权函数的形式为 $\int \psi_j(y)\lambda(y)\mathrm{d}y$ 的不定积分。下文中，我们令权函数归一化，即 $\int_{\mathcal{I}} \lambda(x)\mathrm{d}x = 1$，并令基函数也归一化，即 $\langle \psi_j, \psi_k \rangle_\lambda = \delta(j, k)$。

A.1. 区间 $\mathcal{I} = (-1, 1)$ 上的切比雪夫多项式。归一化权函数和归一化的第一类切比雪夫多项式定义如下：

$$\lambda(x) = \frac{1}{\pi\sqrt{1-x^2}}, \quad \psi_j(x) = w_j \cos(j \cos^{-1}(x)), \quad j = 0, 1, \ldots, \tag{A.1}$$

其中 $w_0 = 1$ 和 $w_j = \sqrt{2}$ 适用于 $j \geq 1$。我们有以下不定积分

$$\int \psi_j(x)\lambda(x)\mathrm{d}x = \begin{cases} -\frac{1}{\pi}\cos^{-1}(x), & j = 0 \\ -\frac{\sqrt{2}}{j\pi}\sin(j\cos^{-1}(x)), & j \geq 1 \end{cases}.$$

归一化权函数和归一化第二类切比雪夫多项式定义为

$$\lambda(x) = \frac{2\sqrt{1-x^2}}{\pi}, \quad \psi_j(x) = \frac{\sin((j+1)\cos^{-1}(x))}{\sin(\cos^{-1}(x))}, \quad j = 0, 1, \ldots. \tag{A.2}$$

我们有如下不定积分

$$\int \psi_j(x)\lambda(x)\mathrm{d}x = \begin{cases} \frac{1}{\pi}\left(\frac{1}{2}\sin(2\cos^{-1}(x)) - \cos^{-1}(x)\right), & j = 0 \\ \frac{1}{\pi}\left(\frac{1}{j+2}\sin((j+2)\cos^{-1}(x)) - \frac{1}{j}\sin(j\cos^{-1}(x))\right), & j \geq 1 \end{cases}.$$

A.2. 区间上的勒让德多项式 $\mathcal{I} = (-1,1)$. 对于用勒让德多项式表示的密度函数，使用配点法和第二类切比雪夫多项式来计算其分布函数更为简单。以式(3.6)中的条件密度为例，我们有

$$f_{\hat{X}_t|\hat{\boldsymbol{X}}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) = \frac{1}{2\zeta}q(x_t), \quad q(x_t) = c + \sum_{\ell=1}^{r}\Big(\sum_{\boldsymbol{k}\in\mathcal{K}}\psi_{k_t}(x_t)\mathsf{C}_{\boldsymbol{k}\ell}\Big)^2,$$

其中，(ψ_{k_t})为最高阶数为 n_t的勒让德多项式。我们将 $\varphi_{2n_t+1}(x)$的根定义为配点，其中 φ_j是第二类切比雪夫多项式（A.2）。由此得到 $x^j = \cos\left(\frac{j\pi}{2n_t+2}\right)$, $j = 1, \ldots$ $2n_t + 1$。这样，通过在配点处计算函数 $q(x_t)$，即可应用配点法，使用切比雪夫基表示$\frac{1}{2\zeta}q(x_t)$，即 $\frac{1}{2\zeta}q(x_t) = \sum_{j=0}^{2n_t} a_j\,\varphi_j(x_t)$，其中系数$(a_j)_{j=1}^{2n_t}$ 可通过快速傅里叶变换以$\hat{}$ $\hat{}O(n_t\log(n_t))$次运算计算得到。于是， $X_t|\boldsymbol{X}_{[t-1]}$ 的分布函数为

$$\mathcal{F}_{\hat{X}_t|\hat{\boldsymbol{X}}_{[t-1]}}(x_t|\boldsymbol{x}_{[t-1]}) = \sum_{j=0}^{2n_t}\frac{a_j}{j+1}\big(\cos((j+1)\cos^{-1}(x_t)) + (-1)^j\big).$$

A.3. 定义在 $\mathcal{I} = (-\infty,\infty)$ **上的厄米多项式**。对于厄米多项式，归一化权函数为 $\lambda(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$，归一化多项式为 $\psi_j(x) = (j!)^{-1/2}p_j(x)$ $j = 0,1,\ldots$，其中 $p_j(x)$ 是由三项递推关系定义的概率论厄米多项式

$$p_{j+1}(x) = xp_j(x) - jp_{j-1}(x), \quad p_1(x) = x, \quad p_0(x) = 1.$$

由于该递推关系也可表示为 $p_{j+1}(x) = xp_j(x) - p_j'(x)$，我们有 $(p_j(x)\lambda(x))' = -p_{j+1}(x)\lambda(x)$，由此得到以下不定积分

$$\int \psi_j(x)\lambda(x)\mathrm{d}x = \begin{cases} \frac{1}{2}\mathrm{erf}(x/\sqrt{2}), & j = 0 \\ -\frac{w_j}{w_{j-1}}\psi_j(x)\lambda(x), & j \geq 1 \end{cases}.$$

A.4. 定义于 $\mathcal{I} = [0,\infty)$**上的拉盖尔多项式**。对于拉盖尔多项式，归一化权函数是指数分布的密度， $\lambda(x) = \exp(-x)$，归一化多项式由以下三项递推关系给出

$$(j+1)\psi_{j+1}(x) = (2j+1-x)\psi_j(x) - j\psi_{j-1}(x), \quad \psi_1(x) = 1-x, \quad \psi_0(x) = 1.$$

由于该递推关系也可表示为 $x\psi_j'(x) = j(\psi_j(x) - \psi_{j-1}(x))$，且对于 $j \geq 1$，恒等式 $\psi_j'(x) = -\sum_{k=0}^{j-1} \psi_k(x)$ 成立，因此我们得到以下不定积分

$$\int \psi_j(x)\lambda(x)\mathrm{d}x = \begin{cases} -\exp(-x), & j = 0 \\ \frac{1}{j} x \exp(-x) \sum_{k=0}^{j-1} \psi_k(x), & j \geq 1 \end{cases}.$$

附录 B. 在第 3.3 节中实现域映射。假设我们有目标密度 $f_X(x)$，$x \in \mathbb{R}$，以及一个可逆变换 $z(x)$（参见表 1）。我们要证明，使用域映射近似目标密度的以下两种方法是等价的：(i) 对 f 应用第 2.3 节的平方根近似过程，使用式 (3.11) 中定义的诱导权函数 $\lambda(x)$ 和式 (3.10) 中定义的扩展基函数 $(\phi_j)_{j=0}^{\infty}$；以及 (ii) 对拉回密度 $f_Z(z) := f_X(x(z))x'(z)$应用平方根近似过程，使用权函数 $\lambda^{\mathrm{ref}}(z)$ 和定义在 $(-1, 1)$ 上的多项式 $(\psi_j)_{j=0}^{\infty}$。

首先介绍直接近似方法 (i)。使用平方根近似过程，首先将目标密度表示为

$$f_X(x) = p(x)\lambda(x), \quad p(x) = \frac{f_X(x)}{\lambda(x)}, \tag{B.1}$$

然后将函数 $\sqrt{p(x)}$投影到扩展基函数 $(\phi_j)_{j=0}^{\infty}$ 上，并采用诱导权函数 $\lambda(x)$，从而得到正交展开

$$\sqrt{p(x)} = \sum_{j=0}^{\infty} a_j \phi_j(x), \quad a_j = \int_{-\infty}^{\infty} \sqrt{p(x)}\phi_j(x)\lambda(x)\mathrm{d}x. \tag{B.2}$$

这样，近似就由系数 $(a_j)_{j=0}^{\infty}$确定。

现在考虑拉回测度方法 (ii)。为应用平方根近似过程，我们将拉回密度 $f_Z(z) = f_X(x(z))x'(z)$ 表示为

$$f_Z(z) = q(z)\lambda^{\mathrm{ref}}(z), \quad q(z) = \frac{f_Z(z)}{\lambda^{\mathrm{ref}}(z)}. \tag{B.3}$$

然后，利用基函数 $(\psi_j)_{j=0}^{\infty}$ 以及区间 $(-1, 1)$ 上的权函数 $\lambda^{\mathrm{ref}}(z)$，函数 $\sqrt{q(z)}$ 可作正交展开

$$\sqrt{q(z)} = \sum_{j=0}^{\infty} b_j \psi_j(z), \quad b_j = \int_{-1}^{1} \sqrt{q(z)}\psi_j(z)\lambda^{\mathrm{ref}}(z)\mathrm{d}z. \tag{B.4}$$

注意，函数 $q(z)$ 也可以表示为

$$q(z) = \frac{f_Z(z)}{\lambda^{\mathrm{ref}}(z)} = \frac{f_X(x(z))x'(z)}{\lambda^{\mathrm{ref}}(z)} \frac{\lambda(x(z))}{\lambda(x(z))} = \frac{p(x(z))}{\lambda^{\mathrm{ref}}(z)} \lambda(x(z))x'(z)$$

其中，p 在式 (B.1) 中定义。根据诱导权函数 $\lambda(x)$ 的定义以及映射 $z'(x)$ 的可逆性，有 $\lambda^{\mathrm{ref}}(z) = \lambda(x(z))x'(z)$，因此 $q(z) = p(x(z))$。这样，系数 $(a_j)_{j=0}^{\infty}$ 可表示为

$$a_j = \int_{-\infty}^{\infty} \sqrt{p(x)}\psi_j(z(x))\lambda^{\mathrm{ref}}(z(x))z'(x)\mathrm{d}x = \int_{-1}^{1} \sqrt{q(z)}\psi_j(z)\lambda^{\mathrm{ref}}(z)\mathrm{d}z = b_j.$$

因此，$(a_j)_{j=0}^{\infty}$ 在 (B.2) 中的系数与 $(b_j)_{j=0}^{\infty}$ 在 (B.4) 中的系数相同。

REFERENCES

[1] N. I. Akhiezer and N. Kemmer, *The classical moment problem: and some related questions*

in analysis，第5卷，Oliver & Boyd，爱丁堡，1965年。[2] R. Baptista, Y. Marzouk, and O. Zahm, *On the representation and learning of monotone triangular transport maps* 2009 10303, 2020，arXiv预印本 arXiv:。()。

[3] A. Beskos, A. Jasra, N. Kantas, A. Thiery, et al., *On the convergence of adaptive sequential monte carlo methods* 26 2016 1111 1146，《应用概率年刊》，()，第–页。[4] J. P. Boyd, *Chebyshev and Fourier spectral methods*，Courier Corporation，2001年。[5] M. Brennan, D. Bigoni, O. Zahm, A. Spantini, and Y. Marzouk, *Greedy inference with structure-exploiting lazy maps* 33 2020 8330 8342，NeurIPS，()，第–页。

[6] M. D. Buhmann, *Radial basis functions: theory and implementations*，第12卷，剑桥大学出版社，2003年。[7] A. Caterini, R. Cornish, D. Sejdinovic, and A. Doucet, *Variational inference with continuously-indexed normalizing flows* 2021, 44 53，载于 UAI，第–页。

[8] P. Chen, U. Villa, and O. Ghattas, *Hessian-based adaptive sparse quadrature for infinite-dimensional bayesian inverse problems* 327 2017，《计算方法与应用力学工程》，()，第147–172页。[9] R. T. Q. Chen, J. Behrmann, D. K. Duvenaud, and J.-H. Jacobsen, *Residual flows for invertible generative modeling* 32, 2019，载于 NeurIPS，第卷。

[10] A. Chkifa, A. Cohen, and C. Schwab, *Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs*，《纯粹与应用数学杂志》，103（2015），第400–428页。[11] A. Cohen and G. Migliorati, *Optimal weighted least-squares methods*，《SMAI计算数学期刊》，3（2017），第181–203页。[12] A. Cohen and G. Migliorati, *Multivariate approximation in downward closed polynomial spaces*，载于《当代计算数学——庆祝 Ian Sloan 80岁生日》，Springer，2018年，第233–282页。[13] P. R. Conrad, Y. M. Marzouk, N. S. Pillai, and A. Smith, *Accelerating asymptotically exact mcmc for computationally intensive models via local approximations*，《美国统计学会杂志》，111（2016），第1591–1607页。[14] T. Cui and S. Dolgov, *Deep composition of tensor-trains using squared inverse rosenblatt transports* 22 2022 1863 1922，《计算数学基础》，()，第–页。[15] T. Cui, S. Dolgov, and O. Zahm, *Scalable conditional deep inverse rosenblatt transports using tensor-trains and gradient-based dimension reduction* 2106 04170, 2023，arXiv:。()。

[16] T. Cui and O. Zahm, *Data-free likelihood-informed dimension reduction of bayesian inverse problems* 37 2021 045009，《反问题》，()，第页。

[17] I. Daubechies, *Ten lectures on wavelets*，SIAM，1992年。[18] C. de Boor, R. A. DeVore, and A. Ron, *On the construction of multivariate (pre) wavelets*，《构造性近似》，9（1993），第123–166页。[19] R. A. DeVore and V. A. Popov, *Interpolation of besov spaces*，《美国数学学会汇刊》，305（1988），第397–414页。[20] M. Döhler, S. Kunis, and D. Potts, *Nonequispaced hyperbolic cross fast fourier transform*，《SIAM数值分析杂志》，47（2010），第4415–4428页。[21] S. Dolgov, K. Anaya-Izquierdo, C. Fox, and R. Scheichl, *Approximation and sampling of multivariate probability distributions in the tensor train decomposition* 30，《统计计算》，(2020)，第603–625页。[22] D. Dũng, V. Temlyakov, and T. Ullrich, *Hyperbolic cross approximation*，Springer，2018年。[23] R. Dutta, S. N. Gomes, D. Kalise, and L. Pacchiardi, *Using mobility data in the design of optimal lockdown strategies for the COVID-19 pandemic* 17 2021，《PLoS计算生物学》，()，第1–25页。[24] A. Gelman and X.-L. Meng, *Simulating normalizing constants: From importance sampling to bridge sampling to path sampling* 1998 163 185，《统计科学》，()，第–页。[25] V. Gradinaru, *Fourier transform on sparse grids: Code design and the time dependent schrödinger equation* 80 2007 1 22，《计算》，()，第–页。[26] J. Hampton and A. Doostan, *Coherence motivated sampling and convergence analysis of least squares polynomial chaos regression* 290，《计算方法与应用力学工程》，(2015)，第73–97页。[27] X. Huan, *Numerical approaches for sequential Bayesian optimal experimental design*，博士论文，麻省理工学院，2015年。[28] N. Kantas, A. Beskos, and A. Jasra, *Sequential monte carlo methods for high-dimensional inverse problems: A case study for the navier–stokes equations*，《SIAM/ASA不确定性量化期刊》，2（2014），第464–489页。[29] H. Knothe, *Contributions to the theory of convex bodies.*，《密歇根数学杂志》，4（1957），第39–52页。

[30] J. Kruse、 G. Detommaso、 U. Köthe、 and R. Scheichl、 *HINT: Hierarchical invertible neural transport for density estimation and bayesian inference* 35 2021 8191 8199 , （）, 第–页。

[31] R. Lam、 O. Zahm、 Y. Marzouk、 and K. Willcox、 *Multifidelity dimension reduction via active subspaces* 42 2020 929 956 , 《SIAM科学计算杂志》,

（）, 第A–A页。[32] S. G. Mallat、 *Multiresolution approximations and wavelet orthonormal bases of L^2（ℝ）*, 《美国数学学会汇刊》, 315（1989）, 第69–87页。[33] N. Metropolis、 *The beginning of the monte carlo method*, 《洛斯阿拉莫斯科学》, 15（1987）, 第125–130页。

[34] G. Migliorati、 *Adaptive polynomial approximation by means of random discrete least squares*, 载于《数值数学与高级应用》, Springer, 2015, 第547–554页。[35] G. Migliorati、 *Adaptive approximation by optimal weighted least-squares methods*, 《SIAM数值分析杂志》, 57（2019）, 第2217–2245页。[36] T. Moselhy and Y. Marzouk、 *Bayesian inference with optimal maps*, 《计算物理学杂志》, 231（2012）, 第7815–7850页。[37] A. Myers、 A. H. Thiéry、 K. Wang、 and T. Bui-Thanh、 *Sequential ensemble transform for bayesian inverse problems* 427 2021 110055 , 《计算物理学杂志》,

（）, 第页。[38] A. Narayan、 J. Jakeman、 and T. Zhou、 *A christoffel function weighted least squares algorithm for collocation approximations* 86 2017 1913 1947 , 《数学与计算》, （）,

第–页。[39] A. B. Owen、 *Monte Carlo theory、 methods and examples*, 2013年。[40] G. Papamakarios、 E. Nalisnick、 D. J. Rezende、 S. Mohamed、 and B. Lakshminarayanan、 *Normalizing flows for probabilistic modeling and inference* 22 2021 , 《机器学习研究杂志》,

（）, 第1–64页。[41] M. D. Parno and Y. M. Marzouk、 *Transport map accelerated markov chain monte carlo*, 《SIAM/ASA不确定性量化杂志》, 6（2018）, 第645–682页。[42] M. Rosenblatt、 *Remarks on a multivariate transformation*, 《数学统计年刊》, 23（1952）, 第470–472页。[43] C. Schillings and C. Schwab、 *Scaling limits in computational bayesian inversion*, 《ESAIM数学建模与数值分析》, 50（2016）, 第1825–1856页。[44] C. Schwab and A. M. Stuart、 *Sparse deterministic approximation of Bayesian inverse problems* 28 2012 045003 , 《逆问题》, （）, 第页。

[45] J. Shen、 T. Tang、 and L.-L. Wang、 *Spectral methods: algorithms、 analysis and applications*, 第41卷, Springer Science & Business Media, 2011年。[46] J. Shen and L.-L. Wang、 *Sparse spectral approximations of high-dimensional problems based on hyperbolic cross* 48 2010 1087 1109 , 《SIAM数值分析杂志》,

（）, 第–页。[47] J. Shen、 L.-L. Wang、 and H. Yu、 *Approximations by orthonormal mapped chebyshev functions for higher-dimensional problems in unbounded domains* 265 , 《计算与应用数学杂志》,

（2014）, 第264–275页。[48] K. Smetana and O. Zahm、 *Randomized residual-based error estimators for the proper generalized decomposition approximation of parametrized problems* , 《国际数值方法工程杂志》,

121（2020）, 第5153–5177页。[49] A. Spantini、 D. Bigoni、 and Y. Marzouk、 *Inference via low-dimensional couplings*, 《机器学习研究杂志》, 19（2018）, 第2639–2709页。[50] C. Villani、 *Optimal transport: old and new*, 第338卷, Springer Science & Business Media, 2008年。

[51] S. Wang and Y. Marzouk、 *On minimax density estimation via measure transport*, arXiv预印本 arXiv:2207.10231, （2022）。[52] H. Wendland、 *Scattered data approximation*, 第17卷, 剑桥大学出版社, 2004年。[53] C. K. Williams and C. E. Rasmussen、 *Gaussian processes for machine learning*, 第2卷, 麻省理工学院出版社, 剑桥（马萨诸塞州）, 2006年。[54] D. Xiu、 *Stochastic collocation methods: a survey*, 《不确定性量化手册》, （2015）, 第1–18页。[55] D. Xiu and G. E. Karniadakis、 *The wiener–askey polynomial chaos for stochastic differential equations* 24 2002 619 644 , 《SIAM科学计算杂志》,

（）, 第–页。[56] J. Yan and T. Zhou、 *Adaptive multi-fidelity polynomial chaos approach to bayesian inference in inverse problems* 381 2019 110 128 , 《计算物理学杂志》,

（）, 第–页。[57] J. Zech and Y. Marzouk、 *Sparse approximation of triangular transports part i: The finite-dimensional case* 2022 1 68 , 《构造逼近》, （）, 第–页。