使用生成模型混合估计具有低维结构的分布

Rong Tang 和 Yun Yang

伊利诺伊大学厄巴纳-香槟分校

arXiv:2301.00890v1 [stat.ME] 2 Jan 2023

摘要

对于满足所谓流形假设的数据进行统计推断，近年来受到越来越多的关注。该假设认为，高维环境空间中的数据点位于维度低得多的子流形附近。在机器学习中，基于编码器-解码器对的生成建模方法，通过显式施加低维流形结构，成功学习了图像和文本等复杂的高维分布。在这项工作中，我们提出一种新方法，通过生成模型混合来估计未知子流形上的分布。我们证明，使用单个编码器-解码器对的传统生成建模方法通常无法捕捉流形假设下的数据分布，除非底层流形具有全局参数化；然而，采用一组编码器-解码器对，分别学习数据支撑流形的不同局部片，即可解决这一问题。我们通过严谨的理论分析证明，所提出的估计量在隐式估计具有流形结构的数据分布时，能够达到极小极大最优收敛速率。实验表明，通过利用参数共享，所提出的方法只需付出极少的额外计算开销，便能显著提升传统基于自编码器的生成建模方法的性能。

K关键词： 自编码器；分布估计；生成模型；流形；极小极大速率 e.

1 引言

利用低维结构对复杂的高维分布进行建模和估计，仍然是现代统计学习面临的主要挑战之一。假设我们观测到 n 个独立同分布的样本$\{X_1, X_2, \ldots, X_n\}$，它们根据某个未知分布$\mu^*$分布在环境欧几里得空间 $\mathbb{R}^D$ 中。我们希望基于这些样本估计 μ^*，以开展统计推断并生成新的样本。分布估计中最流行的非参数方法之一是核密度估计（KDE）。已有研究表明，当 μ^* 相对于 $\mathbb{R}^D$的勒贝格测度具有密度函数，且该密度函数是 β阶光滑的时，KDE能够以最优速率$n^{-\frac{\beta}{2\beta+D}}$ 恢复 $\mathbb{R}^D$ 中任意点处的密度值（Silverman, 2018年；Tsybakov，2009年）。然而，非参数速率 $n^{-\frac{\beta}{2\beta+D}}$ 会受到维数灾难的影响，因为环境维度 D 出现在速率指数中；在涉及图像和文本的机器学习应用中，该维度可能非常高（Brock 等，2018年；Oord 等，2016年）。为了避免维度呈指数级增长，通常的做法是假设数据具有某种额外结构，使数据空间的有效维度相对较低。

近年来备受关注的一种结构是所谓的流形假设，该假设认为数据存在于一个 d维子流形 $\mathcal{M}$ 中，而该子流形嵌入在可能具有高维度的环境空间 $\mathbb{R}^D$中。尽管子流形的几何结构比

传统欧几里得空间更为复杂，但流形假设是许多科学与技术领域中一种自然的假设。例如，在计算机视觉和医学影像领域，数据通常是以向量化像素强度表示的图像。尽管图像可能包含数百万个像素，但通常由相对少量的全局特征决定，例如相机投影、光照条件、纹理、物体位置和方向。其他具有低维流形结构的高维复杂数据还见于自然语言处理（Luo et al., 2020; Ling et al., 2017）、蛋白质-蛋白质相互作用检测（You et al., 2010; Terradot et al., 2004），以及天文学和形状分析（Mardia, 1999; Jupp & Mardia, 2009 ）。

用于对流形值数据建模的统计理论和方法已在各种情境中得到发展（Lin et al., 2020, 2017; Zhang et al., 2022; Lan et al., 2021; Divol, 2022; Tang & Yang, 2022; Berenfeld et al., 2022）。具体而言，近期已有多项研究探讨了如何估计位于未知低维黎曼子流形上的概率测度。例如，Divol（2022）考虑了一种卷积核密度型估计量，该估计量首先利用局部多项式估计技术估计子流形的体积测度。他们证明，所提出的估计量在 Wasserstein损失下能够达到极小极大最优误差界。Tang & Yang（2022）构造了一种两步估计量：第一步估计数据所支撑的子流形；第二步基于小波型估计量，在估计得到的子流形上恢复分布。他们还表明，这种估计在某些对抗损失函数下是极小极大最优的。Berenfeld et al.（2022）提出了一种基于高斯位置-尺度混合的贝叶斯程序，用于估计分布在未知子流形附近的数据的密度，并提供了理论保证。然而，尽管这些现有方法在理论上颇具吸引力，但它们通常难以随环境维度和样本量的增加实现良好的计算可扩展性，因此在对图像和文本等海量高维真实数据进行建模时，实施成本较高。

机器学习文献中的基于自编码器的深度生成建模方法，例如变分自编码器（VAE）（Kingma 和 Welling，2013；Rezende 等人，2014；Kingma 等人，2016）、Wasserstein自编码器（WAE）（Tolstikhin 等人，2019）、InfoVAE（Zhao 等人，2019）和推断型Wasserstein生成对抗网络（iWGAN）（Chen 等人，2022），在生成逼真的合成图像和文本方面取得了巨大成功，而且通常易于高效实现。然而，尽管这些方法在经验上取得了成功，目前仍缺乏一个通用的理论框架来解释这些生成建模方法是否以及如何受益于低维流形结构；这些现有方法在极小极大意义上是否理论最优，也尚不明确。例如，自编码器的关键步骤是提取 d ($d \ll D$) 潜在特征（通过编码器 Q: $\mathbb{R}^D \to \mathbb{R}^d$），以便利用这些特征准确重建原始数据（通过解码器$G$： $\mathbb{R}^d \to \mathbb{R}^D$）。换言之，这些基于自编码器的方法隐含地假设数据 X 具有低维结构，从而能够被准确重建，即 $X \approx G(Q(X))$。此外，真实世界的数据往往落在不具备全局参数化的流形上。例如，数据空间可能是无边界流形（如球面），也可能是非连通的（Khayatkhoei 等人，2018）。由于缺乏全局参数化，采用单一编码器/解码器对的传统自编码器方法无法在不产生失真的情况下恢复整个数据空间。我们的实证结果（参见图 1）也表明，传统的基于自编码器的生成建模方法往往会生成偏离真实流形、外观不真实的样本。

本文提出了一种新的生成建模方法，用于学习流形支撑分布。该方法在理论上达到极小极大最优，计算效率高，并且在生成复杂但逼真的数据方面展现出良好潜力。与大多数依赖数据空间存在全局参数化这一强假设的现有生成建模方法不同，我们采用多个编码器/解码器对，其中每一对都对应于数据支撑流形某个局部

片的参数化。此外，我们利用单位分解技术，将各局部片中估计得到的局部概率测度拼接起来，从而得到流形上概率测度的全局估计。另外，大多数现有方法在构建用于定义生成对抗网络（Goodfellow 等人，2014年）类型估计量的目标函数时，只是简单地代入数据的经验分布；由于没有考虑目标分布的光滑性，这可能导致理论上的不足。我们则提出代入一个平滑性正则化版本，并证明它能够提高估计精度。具体而言，我们证明，当目标分布为 α-光滑，且位于 $\mathbb{R}^D$中的一个 d维、β-光滑子流形上时，相应的估计量 $\widehat{\mu}$ 基于 n 个数据点，在 1-瓦瑟斯坦距离下达到阶为 $O\left(\frac{\log n}{\sqrt{n}} \vee n^{-\frac{\alpha \wedge}{\alpha \wedge \left(\frac{}{2}\beta-1+1\right)\left(\left(\beta-1\right)\right)+d}}\right)$ 的非渐近误差界（这里 $a \wedge b$表示 $\min\{a, b\}$）；当 $\alpha \leq \beta - 1$时，该误差界对应于仅相差一个对数因子的极小极大速率。由此得到的收敛速率不受“维数灾难”的影响，仅取决于数据的内在维数 d 。我们的数值结果还表明，对于具有低内在维数结构的目标分布，所提方法往往比传统的基于自编码器的生成建模方法和经典核密度估计器更准确。

1.1 符号说明

这里总结一些必要的符号和定义。对于任意正整数 k，我们采用简写 $[k] := \{1, 2, \cdots, k\}$。我们用 $\|\cdot\|_p$ 表示通常的向量 ℓ_p 范数，并用 $\|\cdot\|$ 表示 ℓ_2 范数。我们用 $\mathbb{S}_1^d = \{x \in \mathbb{R}^{d+1} : \|x\| = 1\}$ 表示 $\mathbb{R}^{d+1}$中的 d 维单位球面。对于概率测度 μ，我们用 $\mathrm{supp}(\mu)$ 表示其支撑集。对于任意测度 ν 和映射 G，推前测度 $\mu = G_{\#}\nu$ 定义为满足以下条件的唯一测度：对于任意可测集 A，均有 $\mu(A) = \nu(G^{-1}(A))$。对于两个概率测度 μ、ν，μ 与ν之间的 1-瓦瑟斯坦距离定义为 $W_1(\mu, \nu) = \inf\left\{\int \|x - T(x)\|_1 \mathrm{d}\mu(x) : T_{\#}\mu = \nu\right\} = \sup\left\{\int f(x)\mathrm{d}(\mu - \nu) : \mathrm{Lip}(f) \leq 1\right\}$，其中 $\mathrm{Lip}(f)$ 表示 f的最小利普希茨常数。在不存在歧义时，对于绝对连续概率测度 ν，我们也可以用 ν 表示其密度函数。我们用 $\mathcal{P}(\Omega)$ 表示 Ω 上概率测度的集合。我们用 $\mathbb{B}_r(x)$ 表示在 ℓ_2 距离下，以 x为中心、半径为 r 的闭球。我们用 $C_r^\alpha(\Omega)$ 表示所有 α-赫尔德光滑函数构成的集合，其 Hölder 范数$\|\cdot\|_{C^\alpha}(\Omega)$ 有界于 r （例如参见 Evans (2010a)）。类似地，我们用$C_r^\alpha(\Omega; \mathbb{R}^D) = \left\{f = (f_1, \ldots, f_D) : \Omega \to \mathbb{R}^D \,\middle|\, \forall j \in [D], f_j \in C_r^\alpha(\Omega)\right\}$ 表示相应的向量值函数空间。

1.2 组织结构

本文其余部分的组织结构如下。第2节简要介绍基于自编码器的生成建模方法，并定义流形上的光滑分布。第3节介绍我们提出的模型，其实现和理论性质分别在第 4 节和第 5 节中描述。第 6 节和第 7 节分别给出模拟和真实数据应用。

2 背景

2.1 基于自编码器的生成建模方法

假设有从数据空间 $\mathcal{X} \subset \mathbb{R}^D$ 上的未知目标分布 μ^* 中采样得到的独立同分布数据 $X^{(n)} = \{X_1, X_2, \cdots, X_n\}$。在生成建模文献中，目标分布 μ^*由其采样方案隐式指定，并由生成模型表示。数学上，一个

生成模型定义为一对 $(\nu_0,\ G)$，其中 ν_0 是定义在低维潜在空间$\mathcal{Z} \subset \mathbb{R}^d$上的分布，称为生成分布，易于从中采样；而 $G\colon \mathcal{Z} \to \mathcal{X}$ 是从 $\mathcal{Z}$ 到$\mathcal{X}$的映射，称为生成映射，因此若 $Z \sim \nu_0$，则 $G(Z) \sim \mu^*$。生成建模的目标是拟合一个生成模型，使其指定的随机过程所产生的模拟数据看起来与真实数据无法区分。具体而言，基于自编码器的生成建模方法引入一族编码器Q，将数据 X 映射到（低维）潜变量 Z；并引入一族解码器 G，从潜变量重建数据，使二者共同最小化以下目标函数：

$$\frac{1}{n}\sum_{i=1}^{n} c\big(X_i, G(Q(X_i))\big) + \text{Penalty term},$$

其中，$c(\cdot,\ \cdot)$ 是一个代价函数。目标函数的第一项 $n^{-1}\sum_{i=1}^{n} c(X_i,\ G(Q(X_i)))$ 对应于重构代价：一种常见的选择是平方损失 $c(x,\ y) = \|x-y\|_2^2$。该重构代价会促使 $(G\circ Q)_{\#}\mu^*$ 的推前测度尽可能接近 μ^*。另一方面，第二项包含一个用于正则化编码器/解码器对的惩罚项。例如，在变分自编码器中（Kingma 和 Welling，2013年；Rezende 等人，2014年；Kingma 等人，2016年），惩罚项取为由（概率）编码器诱导的潜变量分布与先验分布 ν_0 之间的平均 Kullback-Leibler（KL）散度；在 iWGAN 中（Chen 等人，2022年），惩罚项取为对重构数据分布 $(G\circ Q)_{\#}\mu^*$ 与使用先验 ν_0 的生成模型 $G_{\#}\nu_0$ 所诱导分布之间的 1-瓦瑟斯坦距离的逼近。其他方法则选择惩罚项，直接匹配编码器诱导的潜变量分布与潜在空间中的给定先验（例如，WAE，Tolstikhin 等人（2019年）；infoVAE，Zhao 等人（2019年）；切片WAE，Kolouri 等人（2018年）），由此得到以下训练目标：

$$\frac{1}{n}\sum_{i=1}^{n} c\big(X_i, G(Q(X_i))\big) + \lambda \cdot \mathcal{D}(Q_{\#}\widehat{\mu}_{\mathrm{em}}, \nu_0), \tag{1}$$

其中，$\widehat{\mu}_{\mathrm{em}} := n^{-1}\sum_{i=1}^{n}\delta_{X_i}$ 表示数据 $X^{(n)}$的离散经验分布，而 $\mathcal{D}$ 是用于刻画潜在空间中分布之间接近程度的一般性差异度量。为便于后文指称，我们将训练目标具有式（1）形式的任何方法称为潜分布匹配自编码器（LDMAE）。理想情况下，通过最小化（1）学习得到的解码器 G具有如下性质：$\widehat{\mu} = G_{\#}\nu_0 \approx \mu^*$。与直接处理环境空间上分布的方法（Goodfellow 等人，2014年；Arjovsky 等人，2017年；Khayatkhoei 等人，2018年）相比，潜分布匹配方案在计算上有多项优势。首先，在潜在空间 $\mathcal{Z}$ 中量化分布差异时，$\mathcal{D}$ 的选择更加灵活，因为这些分布通常相对于 $\mathcal{Z}$上的勒贝格测度具有密度函数；相比之下，许多常用的差异度量，例如全变差距离、海灵格距离和KL散度，已知并不适合刻画环境空间上近乎奇异的测度之间的接近程度（Li 等人，2017年；Xu 等人，2018年）。其次，在维度相对较低的潜在空间中计算分布之间的差异效率高得多，不受维数灾难的影响，从而使训练过程更加稳定、耗时更少。此外，$\mathcal{D}$ 具有更大的灵活性，因此可以采用计算公式简单且明确的度量，例如平方最大均值差异（MMD）（Tolstikhin 等人，2019年；Zhao 等人，2019年）和切片瓦瑟斯坦距离（Kolouri 等人，2018年）。

在本小节末尾，我们将描述 LDMAE 的两个重要局限，这些局限促使我们提出将在第3节介绍的方法。具体而言，根据前述对训练目标 (1) 的分解视角，$\widehat{\mu}$ 相对于目标分布 μ^* 的估计误差取决于两个项：(1) μ^* 与 $(G\circ Q)_{\#}\mu^*$之间的“距离”；(2) $Q_{\#}\mu^*$ 与 ν_0之间的“距离”，

$\widehat{}$其中，Q 表示通过最小化 (1) 得到的学习到的编码器。令 $\mathcal{M}$ 表示真实数据生成分布 μ^* 的支撑集，它是嵌入在 $\mathbb{R}^D$中的一个 d维子流形。要控制第一个距离，$\mathcal{M}$ 需要具有全局参数化；也就是说，我们可以找到一些连续映射 $G^*: \mathbb{R}^d \to \mathbb{R}^D$和 $Q^*: \mathbb{R}^D \to \mathbb{R}^d$，使得对于任意 $x \in \mathcal{M}$，$G^*(Q^*(x)) = x$。许多常见流形都不满足这一全局参数化条件，例如非连通流形，以及球面和环面这类无边界流形。第二个距离取决于经验数据分布 $\widehat{\mu}_{\mathrm{em}}$ 诱导的$\widehat{}$ $\widehat{}$经验潜在分布 $Q_{\#}\widehat{\mu}_{\mathrm{em}}$ 对总体分布 $Q_{\#}\mu^*$的逼近程度。然而，即便假设流形 $\mathcal{M}$ 具有全局参数化，使得对于某个 G^*，$\mu^* = G^*_{\#}\nu_0$ 和$Q^* = (G^*)^{-1}$ 分别属于解码器族和编码器族，离散经验分布$\widehat{\mu}_{\mathrm{em}}$ 在逼近光滑测度时仍可能存在统计不足（Liang, 2020; Tang 与 Yang, 2022）。因此，直接将 $Q_{\#}\widehat{\mu}_{\mathrm{em}}$ 代入惩罚项可能会导致过拟合。

2.2 单位分解与流形上的分布

直观来说，流形是一种局部类似于欧几里得空间的拓扑空间。形式上，流形的数学定义如下。

定义 1. d维流形 $\mathcal{M}$ 定义为满足以下条件的拓扑空间：*(1)* 存在一个覆盖 $\mathcal{M}$ 的图册，该图册由一组 d 维坐标图 $\mathscr{A} = \{(U_\lambda, \varphi_\lambda)\}_{\lambda\in\Lambda}$ 组成，即 $\mathcal{M} = \bigcup_{\lambda\in\Lambda} U_\lambda$。*(2)* 图册 $\mathscr{A}$中的每个坐标图 [1](U, φ) 都包含一个同胚映射 $\varphi: U \to \widetilde{U}$，称为坐标映射，它将一个开集 $U \subset \mathcal{M}$ 映射到一个开集 $\widetilde{U} \subset \mathbb{R}^d$。

如果 $\mathcal{M} \subset \mathbb{R}^D$，并且将每个坐标图中的坐标映射 φ 及其逆 φ^{-1} 视为定义在欧几里得空间子集上的函数时，它们都是 β光滑映射，那么我们称流形 $\mathcal{M}$ 是嵌入于 $\mathbb{R}^D$ 中的（β光滑）子流形。与流形相关的另一个有用概念是单位分解。

定义 2。流形 $\mathcal{M}$ 的单位分解是满足以下条件的一组函数 $\{\rho_\lambda\}_{\lambda\in\Lambda}$

1. $0 \le \rho_\lambda \le 1$ 对所有 $\lambda \in \Lambda$都成立，并且 $\sum_{\lambda\in\Lambda} \rho_\lambda(x) = 1$ 对所有 $x \in \mathcal{M}$都成立。

2. 每个点 $x \in \mathcal{M}$ 都有一个邻域，它只与有限多个 $\lambda \in \Lambda$ 的$\mathrm{supp}(\rho_\lambda)$ 相交 .

利用单位分解，可以将局部坐标图中的构造粘合起来，从而在流形上形成全局构造。可以从流形的任意开覆盖 $\{U_\lambda\}_{\lambda\in\Lambda}$ 构造一个单位分解，其中分解 $\{\rho_\lambda\}_{\lambda\in\Lambda}$ 以同一集合为指标集，并且对任意 $\lambda \in \Lambda$，都有 $\mathrm{supp}(\rho_\lambda) \subset U_\lambda$。这样的单位分解称为从属于 开覆盖 $\{U_\lambda\}_{\lambda\in\Lambda}$。

对于具有图册 $\mathscr{A} = \{(U_\lambda, \varphi_\lambda)\}_{\lambda\in\Lambda}$ 的流形 $\mathcal{M}$，假设 Λ 是有限的，并将其记为 $\Lambda = [K]$。给定从属于开覆盖 $\{U_k\}_{k\in[K]}$ 的单位分解 $\{\rho_k\}_{k\in[K]}$，可以将 $\mathcal{M}$ 上的任意分布 μ^* 分解为

$$\mu^* = \sum_{k\in[K]} \rho_k\mu^* = \sum_{k\in[K]} (\varphi_k^{-1})_{\#}\big[(\varphi_k)_{\#}(\rho_k\mu^*)\big], \tag{2}$$

其中，第一个不等式利用了对所有 $x \in \mathcal{M}$ 均有 $\sum_{k\in[K]} \rho_k(x) = 1$ 这一点；第二个不等式则利用了 $\mathrm{supp}(\rho_k) \subset U_k$，且 φ_k 是 $U_k \to \varphi_k(U_k)$ 上的同胚这一事实。[2] 那么，如果我们记 $\nu_k^* = (\varphi_k)_{\#}\big(\frac{\rho_k\mu^*}{\mathbb{E}_{\mu^*}[\rho_k]}\big)$，则 μ^* 可表示为以下生成模型混合：

$$\mu^* = \sum_{k\in[K]} \mathbb{E}_{\mu^*}[\rho_k] \cdot (\varphi_k^{-1})_{\#}\nu_k^*, \tag{3}$$

[1]为使符号说明简洁，省略下标 λ。[2]如果 $(\varphi_k)_{\#}(\rho_k\mu^*)$ 对每个 $k \in [K]$ 都存在相对于 $\mathbb{R}^d$ 上勒贝格测度的 α-光滑密度函数，那么称 $\mathcal{M}$ 上的 μ^* 为 α-光滑分布。

分解 (3) 表明，任何支撑在嵌入于 $\mathbb{R}^D$ 上的 d 维子流形上的分布 μ^*，只要其图册由至多 K 个坐标图组成，就属于以下生成模型混合类：

$$\mathcal{S}^* = \Big\{ \mu = \sum_{k\in[K]} p_k \cdot (G_k)_{\#}\nu_k \;\Big|\; G_k : \mathbb{R}^d \to \mathbb{R}^D, \nu_k \in \mathcal{P}(\mathbb{R}^d), 0 \le p_k \le 1, \sum_{k\in[K]} p_k = 1 \Big\}.$$

该空间构成了我们的模型空间，用于表示流形上的分布。

3 潜分布匹配自编码器混合

根据第2节中的讨论，我们可以看到，当待估计的目标分布位于一个没有全局参数化的一般子流形上时，传统的基于自编码器的生成建模方法可能表示能力不足。然而，任何以流形为支撑集的分布都可以表示为生成模型的混合形式（参见分解式 (3)）这一性质，促使我们采用多个编码器/解码器对，并使用单位分解为它们赋予适当权重，从而将它们拼接起来。

回顾一下，我们有一组 n 个独立同分布的观测值 $X^{(n)} = \{X_1, \dots, X_n\}$，它们从目标分布 μ^* 中采样，且该分布位于嵌入在 $\mathbb{R}^D$ 中的 d 维子流形 $\mathcal{M}$ 上，其中 $d \le D$。令 $\{S_k\}_{k\in[K]}$ 为 $\mathcal{M}$ 的一个适当选取的开覆盖 $\mathrm{supp}(\mu^*) \subset \mathbb{R}^D$，并固定一个从属于 $\{S_k\}_{k\in[K]}$ 的单位分解 $\{\rho_k\}_{k\in[K]}$，[3] 可在不了解 $\mathcal{M}$ 的情况下选取（参见第 4 节）。对于任意一般的逼近族 $\mathcal{G}$，其由 (G, Q, v) 构成，其中 $\mathbf{G} = (G_1, G_2, \cdots, G_K)$ 且 $G_k \colon \mathbb{R}^d \to \mathbb{R}^D$，$\mathbf{Q} = (Q_1, Q_2, \cdots, Q_K)$ 且 $Q_k \colon \mathbb{R}^D \to \mathbb{R}^d$，而 $\mathbf{v} = (\nu_1, \nu_2, \cdots, \nu_K)$ 且 $\nu_k \in \mathcal{P}(\mathbb{R}^d)$，我们定义以下估计量，并称其为潜分布匹配自编码器混合（MLDMAE）估计量：

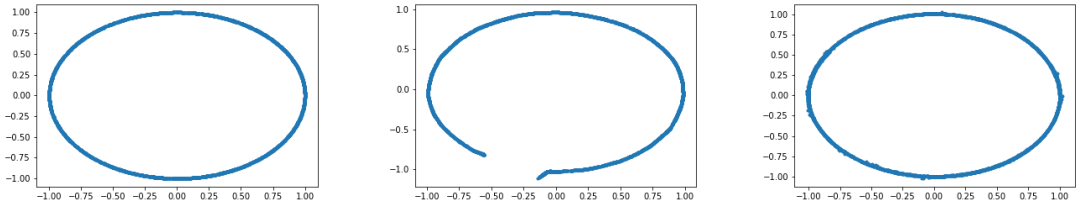
$$\begin{aligned} \widehat{\mu} &= \sum_{k\in[K]} \widehat{p}_k \cdot (\widehat{G}_k)_{\#}\widehat{\nu}_k, \qquad \text{with} \quad \widehat{p}_k = \frac{1}{n}\sum_{i=1}^n \rho_k(X_i) \quad \text{and} \\ (\widehat{\mathbf{G}}, \widehat{\mathbf{Q}}, \widehat{\mathbf{v}}) &= \operatorname*{arg\,min}_{(\mathbf{G},\mathbf{Q},\mathbf{v})\in\mathcal{G}} \sum_{k=1}^K \Big\{ \frac{1}{n}\sum_{i=1}^n c\big(X_i, G_k(Q_k(X_i))\big) \cdot \rho_k(X_i) + \lambda_k \cdot \mathcal{D}\big(\widetilde{\nu}_{k,Q_k}, \nu_k\big) \Big\}, \end{aligned} \tag{4}$$

其中，回顾一下，$c(\cdot,\ \cdot)$ 是代价函数，$\widetilde{\nu}_{k,\ Q_k}$ 是对 $(Q_k)\#(\frac{\mu^*\cdot\rho_k}{\mathbb{E}_{\mu^*}[\rho_k]})$ 密度的（光滑性正则化）估计量（精确定义见附录C），而 $\mathcal{D}(\cdot,\ \cdot)$ 是潜在空间中分布之间的一种通用差异度量。不同于传统的 LDMAE 估计量，MLDMAE 还可以使用经验贝叶斯方法为局部潜变量选择依赖于数据的先验分布，这提高了建模的灵活性，并可能降低逼近误差。我们在定理 1 中证明，对于经过精心选择的逼近族 $\mathcal{G}$、代价函数 c、差异度量 $\mathcal{D}$ 和光滑性正则化估计量 $\widetilde{\nu}_{k,\ Q_k}$，所得估计量 $\widehat{\mu}$ 在 1-瓦瑟斯坦距离下达到极小极大收敛速率，当 $W_1(\widehat{\mu},\ \mu^*) \le C\, n^{-\frac{\alpha\wedge}{2(\beta-1)+d}\vee \text{对数}\sqrt{n}}$，其中 μ^* 是未知的 β-光滑、d 维子流形上的 α- 光滑分布。

MLDMAE 的目标函数可分解为两部分：重构代价 $n^{-1}\sum_{k=1}^K\sum_{i=1}^n c\big(X_i,\ G_k(Q_k(X_i))\big)\cdot\rho_k(X_i)$ 和惩罚项 $\sum_{k=1}^K \lambda_k \cdot \mathcal{D}\big(\widetilde{\nu}_{k,\ Q_k},\ \nu_k\big)$。我们还可以允许不同的 $k \in [K]$ 上的编码器/解码器对具有不同的潜在维度 d。重构代价旨在学习 μ^* 的支撑流形的局部参数化，具体做法是强制编码器/解码器对 $(\widehat{Q}_k,\ \widehat{G}_k)$ 表示分解（3）中 $\mathcal{M}$ 的局部片 $U_k = \mathcal{M}\cap S_k$ 的某个坐标系 $(\varphi_k,\ \varphi_k^{-1})$。因此，它对应于 μ^* 的支撑恢复。通过采用多个

[3] 这里我们可以考虑任意函数 $\rho_k \colon \mathbb{R}^D \to [0,1]$，使得 $\mathrm{supp}(\rho_k) \subset S_k \subset \mathbb{R}^D$ 且 $\sum_{k\in[K]}\rho_k(x) = 1$，对于任意 $x \in \cup_{k\in[K]} S_k$；注意，$\{\rho_k|_{S_k\cap\mathcal{M}}\}_{k\in[K]}$ 将构成 $\mathcal{M}$的单位分解。

编码器/解码器对，MLDMAE 估计量避免了传统 LDMAE 估计量中隐含采用的限制性全局参数化假设。因此，MLDMAE 适用于更广泛的问题（示意见图 1）。另一方面，惩罚项旨在强制重加权后的局部潜在分布 $(\widehat{Q}_k)_{\#}\left(\frac{\mu^* \cdot \rho_k}{\mathbb{E}_{\mu^*}[\rho_k]}\right)$ 与预先指定的先验族中的某个成员 $\widehat{\nu}_k$ 相匹配，从而使 $\widehat{\nu}_k$ 接近分解（3）中的 ν_k^*。



(a) 真实数据　(b) LDMAE　(c) MLDMAE

图 1：当目标分布为球面上的均匀分布时，LDMAE 与 MLDMAE 的比较。图 (a) 展示真实数据，图 (b)、(c) 分别展示基于 10000 个训练样本，由 MLDMAE 和 LDMAE 估计器生成的随机生成样本。MLDMAE 中选用的单位分解是第 4 节中描述的光滑单位分解，包含 $K = 10$和 $\gamma = 10$。差异度量 $\mathcal{D}(\cdot,\cdot)$ 选为采用高斯核的最大均值差异。可以看到，LDMAE 无法捕捉球面的正确形状。原因在于，球面（或任何无边界流形）都需要至少两个覆盖图才能构成描述它的图册。LDMAE 模型使用单个编码器/解码器对，因此生成的曲线具有起点和终点。相比之下，我们的估计量能够学习无法进行全局参数化的一般流形。

在实践中，我们也可以不选择最佳的数据依赖型先验分布，而是将先验分布固定为简单的分布 ν_0，例如各向同性高斯分布。此外，$\mathcal{D}(\cdot,\cdot)$ 可以选为某种平方最大均值差异（MMD）损失[4]，其可通过闭式公式高效计算。式（4）中的第 k 个平滑性正则化分布 $\widetilde{\nu}_{k,Q_k}$，可以通过对其（加权）经验对应分布 $(Q_k)_{\#}\widehat{\mu}_n^k$ 进行核平滑来构造，其中 $\widehat{\mu}_n^k = (n\widehat{p}_k)^{-1}\sum_{i=1}^n \rho_k(X_i)\,\delta_{X_i}$，从而得到 $\widetilde{\nu}_{k,Q_k}(z) = (n\widehat{p}_k)^{-1}\sum_{i=1}^n k(z, Q_k(X_i))\rho_k(X_i)$，其中核函数 k 取适当值。注意，当核函数 k 为高斯核 $\widetilde{k}(x,y) = (2\pi h)^{-\frac{d}{2}}\exp(-\frac{\|x-y\|^2}{2h})$，且带宽参数为 h 时，$\widetilde{\nu}_{k,Q_k}$ 对应于高斯平滑分布 $(\widetilde{Q}_k)_{\#}\widehat{\mu}_n^k$，其中 $\widetilde{Q}_k$ 是由 $\widetilde{Q}_k(X) = Q_k(X) + \sqrt{h}\cdot\mathcal{N}(0, I_d)$ 定义的随机扰动编码器。[5] 采用这种平滑性正则化分布，可视为应用随机数据增强来增加编码后训练样本的变异性，从而缓解对数据的潜在过拟合，并提高所得估计量的泛化能力。

在我们的估计量中引入编码器-解码器结构带来了多项优势。在计算方面，编码器将高维数据转换为低维潜变量，因此我们只需对低维分布计算惩罚项。因此，与直接处理环境空间中分布的生成建模方法（例如 iWGAN）相比，MLDMAE框架的计算负担更小。在理论方面，当 $d \ll D$ 时，数据分布 μ^* 在 $\mathbb{R}^D$ 中成为奇异测度。因此，借助由编码器-解码器对明确诱导的、关于 μ^* 支撑流形的信息，我们可以运用非参数密度估计的经典技术（例如小波截断）构造极小极大最优估计量。具体而言，生成模型混合（3）中定义的潜变量真实分布 $(Q_k)_{\#}\left(\frac{\mu^* \cdot \rho_k}{\mathbb{E}_{\mu^*}[\rho_k]}\right)$ 以高概率相对于 $\mathbb{R}^d$ 上的勒贝格测度绝对连续（参见附录C中的引理 4），这使我们能够开发光滑性正则化

[4] 对于正定再生核 k，MMD损失定义为 $\mathrm{MMD}^2(\mu_1, \mu_2) = \mathbb{E}_{X,\, X' \in \mu_1}[k(X,\, X')] + \mathbb{E}_{Y,\, Y' \in \mu_2}[k\ Y\ Y'\ ] - 2\mathbb{E}_{X \in \mu_1\ Y \in \mu_2}[k\ X\ Y\ ](,\ ),\ (,\ )$ 此处，对于随机映射 Q，推前测度 $Q_{\#}\mu$ 定义为满足以下条件的测度：对于任意可测函数 f，$\int f(x)\,\mathrm{d}[Q_{\#}\mu] = \mathbb{E}\left[\int f(Q(x))\,\mathrm{d}\mu\right]$ 其中，期望是相对于 Q 的随机性而言的。

估计量，借鉴 Liang（2020）和 Singh等（2018）的方法。事实上，正如 Liang（2020）和 Tang & Yang (2022) 所指出的，当 $\alpha \leq \beta - 1$ 时，MLDMAE估计器达到的速率 $O\big(n^{-\frac{\alpha+1}{2\alpha+d}} \vee \sqrt[n]{\text{对数}}\big)$ 相对于1-瓦瑟斯坦距离而言，在对数因子范围内是极小极大最优的。

4 计算

本节讨论所提出方法的一些重要计算方面。

单位分解的选择：我们需要解决的一个问题是如何选择合适的单位分解 $\{\rho_k\}_{k\in[K]}$。为此，我们可以先使用足够大的簇数 K，对数据运行一种聚类算法，例如（小批量） K-均值算法。根据聚类结果， ρ_k 的一种直接选择是指示函数 $\mathbf{1}(x \in k$号簇$)$。我们也可以按以下方式选择光滑单位分解：首先，将 k号簇的质心记为 a_k ，并将能够包含 k号簇中所有数据点的最小半径记为 r_k，使得这些数据点包含在 $\mathbb{B}_{r_k}($ $a_k)$ 中。然后，我们可以构造一个开覆盖$\{S_k = \mathbb{B}_{r_k+\varepsilon}(a_k)^\circ\}_{k\in[K]}$ ，其中 ε 是一个较小的正数，使得 $\{S_k\}_{k\in[K]}$ 以高概率覆盖 μ^* 的未知支集。给定该开覆盖，对于每个 $k \in [K]$，我们将一个局部分割函数定义为 $\widetilde{\rho}_k(x) = ((r_k+\varepsilon)^2 - \|x-a_k\|^2)^\gamma \cdot \mathbf{1}(x \in S_k)$，其中 $\gamma > 1$是一个调节参数。由此得到的 $\{\rho_k\}_{k\in[K]}$ 构成了 $\mathcal{M}$ 的单位分解，其中对于 $k \in [K]$有 $\rho_k = \widetilde{\rho}_k / \big(\sum_{k'=1}^K \widetilde{\rho}_{k'}\big)$ 。注意，当 γ较大时， $\widetilde{\rho}_k$ 会对远离质心 a_k 的点赋予较低权重。

惩罚项的选择：我们选择平滑性正则化分布 $\widetilde{\nu}_{k,Q_k}$ ，将其设为如第3节所述的 $(Q_k)_\#\widehat{\mu}_n^k$ 的高斯核平滑版本。这 $\widetilde{\nu}_{k,Q_k}$ 与变分自编码器中的常用高斯编码器相关，因此我们可以利用变分自编码器中的重参数化技巧来优化所需的目标函数。对于先验分布 ν，我们可以考虑简单的分布 ν_0 ，例如标准高斯分布 $\mathcal{N}(0, I_d)$，这也是传统生成建模方法中的常见做法。此外，为确保所学习流形的光滑性，我们可以考虑下文第5节备注1中描述的数据驱动先验，从而确保 ν_k 和 $\widetilde{\nu}_{k,\widehat{Q}_k}$ 具有匹配尾部。对于差异度量$\mathcal{D}(\cdot,\cdot)$，为避免对抗训练不稳定，我们可以：(1) 选择 $\mathcal{D}(\cdot,\cdot)$ 为由正定核 k刻画的（平方）最大均值差异，例如逆多二次（IMQ）核$k(x,y) = C_{im}/(C_{im} + \|x-y\|_2^2)$ 和 RBF 核 $k(x,y) = \exp(-\|x-y\|^2/C)$；(2) 当内在（潜在）维度为 $O(1)$ 阶时，可以选择 $\mathcal{D}(\cdot,\cdot)$ 作为通过“POT”软件包计算的1-瓦瑟斯坦距离（Flamary 等人，2021），该软件包使用网络单纯形算法（Bonneel等，2011）计算两个离散测度之间的最优传输映射。

解码器和编码器的构建：解码器 $\mathbf{G}$ 和编码器 $\mathbf{Q}$ 可以通过神经网络实现。然而，对于较大的 K，如果将每个解码器/编码器都视为独立的神经网络，我们就需要训练大量参数。为了解决这一问题，我们在解码器集合 $\mathbf{G}$ 和编码器集合 $\mathbf{Q}$内部采用参数共享。具体来说，对于编码器集合，我们将编码器 $\{Q_k\}_{k\in[K]}$ 之间的最后一层（输出层）设为独立参数，而将其他层设为共享参数。此外，由于解码器—编码器结构旨在重构数据，我们希望解码器 G_k 接近编码器 Q_k的逆。为此，对于解码器集合，我们反过来将解码器 $\{G_k\}_{k\in[K]}$ 之间的第一层（输入层）设为独立参数，并将其他层设为共享参数。我们采用这种参数共享方案的依据是：在编码器中，第一层（卷积层）专注于“低层”特征提取，而其他层则提取“高层”特征。因此，所述参数共享方案使网络能够利用数据不同簇之间共有的“低层”信息，从而提高模型训练效率。

目标函数的优化：回顾一下，$\widetilde{\nu}_{k,\ Q_k} = \widetilde{Q}_k(\widehat{\mu}_n^k)$，其中 $\widetilde{Q}_k$ 是随机扰动编码器 $\widetilde{Q}_k(X) = Q_k(X) + \sqrt{h}\cdot\mathcal{N}(0,\ I_d)$，而 $\widehat{\mu}_n^k$ 是重加权经验测度 $\widehat{\mu}_n^k = \frac{\widehat{\mu}_n \cdot \rho_k}{\widehat{p}_k}$。给定用于定义惩罚项和重构代价的差异度量~~$\mathcal{D}(\cdot,\ \cdot)$~~ 和代价函数 $c(\cdot,\ \cdot)$，我们可以

将目标函数改写为

$$\sum_{k=1}^{K}\Big\{\widehat{p}_k\int c\big(x,G_k(Q_k(x))\big)\,\mathrm{d}\widehat{\mu}_n^k+\lambda_k\cdot\mathcal{D}\big(\nu_k,(\widetilde{Q}_k)_{\#}\widehat{\mu}_n^k\big)\Big\}.$$

为近似该目标函数的梯度，我们通过引入辅助随机变量 $u\in$ Unif(0,1)，从测度 $\widehat{\mu}_n^k$ 中采样。具体而言，若 $u<\rho_k(X_i)$，我们就将数据 X_i 纳入经验测度 $\widehat{\mu}_n^k$；否则，将该数据从 $\widehat{\mu}_n^k$中排除。此外，可利用来自 ν_k 和 $(Q_k)_{\#}\widehat{\mu}_n^m$的有限样本估计惩罚项$\widetilde{}$。

算法1： MLDMAE 的实现算法

输入：正则化系数 $\{\lambda_k\}_{k\in[K]}$、单位分解 $\{\rho_k\}_{k\in[K]}$、差异度量 $\mathcal{D}(\cdot,\cdot)$ 和代价函数 $c(\cdot,\cdot)$、先验分布 $\{\nu_k\}_{k\in[K]}$、潜在（内在）维数 d；

数据：$X^{(n)}=\{X_1,X_2\cdots,X_n\}$；

重复

从 $X^{(n)}$中抽取一个小批量数据集 $\mathscr{D}$；

对于 $k\leftarrow 1\ to\ K$，**执行**

初始化一个空数据集 $\widetilde{\mathscr{D}}_k$；

对于 $X\in\mathscr{D}$，**执行**

从 Unif(0,1)生成随机变量 u；

如果 $u\le\rho_k(X)$ **则**

将 X 添加到数据集 $\widetilde{\mathscr{D}}_k$；

从先验分布 ν_k生成数据集 $\mathscr{L}_k$；

从 $\mathcal{N}(Q_{\phi,k}(X)\ h\,I_d)$ 为均匀选取的 X 生成数据集 $\mathscr{E}_k$，取值范围为 $\widetilde{\mathscr{D}}_k$.

首先使用一阶方法（例如 Adam，Kingma 和 Ba（2014年））迭代一步更新 ϕ 和 θ，其中目标函数：

$$\sum_{k=1}^{K}\Big\{\frac{\widehat{p}_k}{|\widetilde{\mathscr{D}}_k|}\sum_{X\in\widetilde{\mathscr{D}}_k}c\big(X,G_{\theta,k}(Q_{\phi,k}(X))\big)+\lambda_k\cdot\mathcal{D}\Big(\frac{1}{|\mathscr{L}_k|}\sum_{z\in\mathscr{L}_k}\delta_z,\quad\frac{1}{|\mathscr{E}_k|}\sum_{z\in\mathscr{E}_k}\delta_z\Big)\Big\},$$

其中 δ_z 是点 z. 上的狄拉克测度

直到 (ϕ,θ) 收敛；

基于以上讨论，我们现在可以构建算法1所述的算法来实现 MLDMAE 估计，其中我们用 $\mathbf{G}_\theta=\{G_{\theta,1},G_{\theta,2},\cdots,G_{\theta,M}\}$ 和 $\mathbf{Q}_\phi=\{Q_{\phi,1},Q_{\phi,2},\cdots,Q_{\phi,M}\}$ 分别表示由 θ 和 ϕ 参数化的解码器和编码器。

5 理论分析

本节中，我们推导 MLDMAE 估计量的有限样本误差。首先，我们陈述关于目标分布 μ^* 以及用于定义估计量 (4) 的逼近族的以下假设。

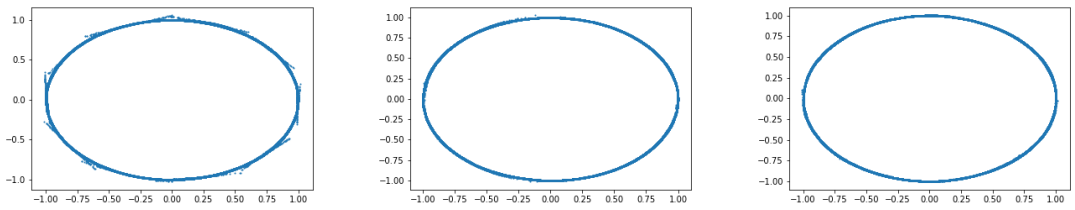
假设 A（目标分布）： 流形 $\mathcal{M}$ 上的目标分布 μ^* 满足：(1)$\mathcal{M}\subset\cup_{k\in[K]}S_k$；(2) 对任意 $k\in[K]$，存在 $G_k^*\in C_L^\beta(\mathbb{R}^d;\mathbb{R}^D)$ 和 $Q_k^*\in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d)$，使得对任意 $x\in\mathcal{M}\cap S_k$，$x=G_k^*(Q_k^*(x))$ 成立；(3) 对任意 $m\in[K]$，令 $p_k=\mathbb{E}_{\mu^*}[\rho_k]$，则 $p_k>0$ 和 $\nu_k^*=(Q_k^*)\#(\frac{\mu^*\cdot\rho_k}{p_k})\in C_L^\alpha(\mathbb{R}^d)$；令 $\Omega_k=Q_k^*(\mathcal{M}\cap S_k)$，存在函数 $g_k:\mathbb{R}^+\to\mathbb{R}^+$，使得对任意 $r>0$ 和 $z\in\Omega_k$，存在 $z'\in\Omega_k$，满足 $z\in B_r(z')$ 且 $\nu_k^*(z')\ge g_k(r)$。

假设 B（逼近族）：逼近族 $\mathcal{G}$ 满足：(1) $(\mathbf{G}^*,\ \mathbf{Q}^*,\ \mathbf{v}^*)\in\mathcal{G}$，其中 $\mathbf{G}^*=(G_1^*,\ G_2^*,\ \cdots,\ G_K^*)$、$\mathbf{Q}^*=(Q_1^*,\ Q_2^*,\ \cdots,\ Q_K^*)$ 和 $\mathbf{v}^*=(\nu_1^*,\ \nu_2^*,\ \cdots,\ \nu_K^*)$；(2) 对任意

$(\mathbf{G},\ \mathbf{Q},\ \mathbf{v}) \in \mathcal{G}$，对于任意 $k \in [K]$，均有 $G_k \in C_L^\beta(\mathbb{R}^d;\ \mathbb{R}^D)$ 和 $Q_k \in C_L^\beta(\mathbb{R}^D;\ \mathbb{R}^d)$。

示例（流形支撑分布）：对于任意定义在嵌入于 $\mathbb{R}^D$ 中的 β-光滑、d-维无边界紧子流形上的 α-光滑分布 μ^*，且该分布具有正密度，我们可以找到合适的开覆盖 $\{S_k\}_{k\in[K]}$ 和单位分解 $\{\rho_k\}_{k\in[K]}$，使假设 A 成立。此外，由逼近族 $\mathcal{G} = \{(\mathrm{G},\mathrm{Q},\mathrm{v}) : \forall k \in [K], G_k \in C_L^\beta(\mathbb{R}^d;\ \mathbb{R}^D)\ Q_k \in C_L^\beta(\mathbb{R}^D;\ \mathbb{R}^d)\ \nu_k = (V_{k\#}\nu_0) \cdot \rho_k(G_k(z))\ \mathbb{E}_{\nu_0}[\rho_k(G_k(V_k(z)))], V_k \in C_L^{\alpha+1}(\mathbb{B}_1^d;\ \mathbb{B}_1^d)\}$ 所诱导的（混合）生成模型类，其中 ν_0 是 $\mathbb{B}_1^d$ 上任意固定的 α-光滑分布，且其密度值有正下界（例如均匀分布），足以对流形支撑分布 μ^* 建模（即逼近族 $\mathcal{G}$ 满足假设 B）。特别地，当 $\alpha + 1 = \beta$ 时，我们可以将复合映射 $G_k \circ V_k$ 视为 β-光滑编码器，以避免估计先验分布；也就是说，我们可以使用逼近族 $\mathcal{G} = \{(\mathrm{G},\mathrm{Q},\mathrm{v}) :\ \forall k \in [K], G_k \in C_L^\beta(\mathbb{R}^d;\ \mathbb{R}^D)\ Q_k \in C_L^\beta(\mathbb{R}^D;\ \mathbb{R}^d)\ \nu_k = \frac{\nu_0 \cdot \rho_k(G_k(z))}{\mathbb{E}_{\nu_0}[\rho_k(G_k(z))]}\}$。更多细节见附录 B。

备注*1*。逼近族的选择表明，对于学习由流形支撑的分布，与其将先验分布 ν_k 设为某种固定的简单分布 ν_0，不如按权重 $\rho_k(G_k(\cdot))$对其重新缩放，使所得分布在收敛后具有与 $Q_{k\#}(\rho_k\mu^*/\mathbb{E}_{\mu^*}[\rho_k])$匹配的尾部。然而，在 *MLDMAE* 中引入这样的先验族可能会因函数 ρ_k高度不规则而导致训练不稳定。为解决这一问题，我们可以考虑如下“数据驱动”先验：首先，将 ν_k 固定为某种简单的固定分布 ν_0，例如均匀分布或截断正态分布；然后运行 *MLDMAE* 算法，得到编码器/解码器对 $\{(\widehat{G}_k^{[1]}, \widehat{Q}_k^{[1]})\}_{k\in[K]}$的估计量。接着，将先验分布 ν_k 固定为按 $\rho_k(G_k^{[1]}(\cdot))$缩放后的 ν_0，并以 $\{(\widehat{G}_k^{[1]}, \widehat{Q}_k^{[1]})\}_{k\in[K]}$ 为初始化运行 *MLDMAE* 算法，以得到编码器/解码器对 $\{(\widehat{G}_k^{[2]}, \widehat{Q}_k^{[2]})\}_{k\in[K]}$的估计量。上述步骤可以继续进行：将先验分布 ν_k 固定为 $\nu_0 \cdot \rho_k(G_k^{[l]}(\cdot))$，并得到估计量 $\{(\widehat{G}_k^{[l+1]}, \widehat{Q}_k^{[l+1]})\}_{k\in[K]}$；当验证误差不再改善时停止。使用这种数据驱动先验，可以显著提升 *MLDMAE* 在不同分割函数支撑集交汇处的性能，示意见图 *2*。



(a) MLDMAE：截断正态先验 (b) MLDMAE：数据驱动先验（更新一次 (c) MLDMAE：数据驱动先验更新）（更新两次）

图2：当目标测度为球面上的均匀分布时，MLDMAE在采用不同先验分布时的表现。图(a)绘制了先验分布为截断正态分布时MLDMAE估计量生成的样本。图(b)和图(c)分别绘制了采用备注1中所述数据驱动先验分布、更新一次和两次后MLDMAE估计量生成的样本，其中 ν_0 与图(a)中一样为截断正态分布。可以看出，采用简单的截断正态先验时，生成的图形在不同分割函数的交界处往往不够平滑。而按照备注1中所述策略更新一次先验分布，就能显著改善表现。

示例（具有聚类结构的分布）：另一个例子是由生成模型混合 $\mu^* = \sum_{k=1}^K p_k \cdot (G_k^*)_\# \nu_k^*$ 诱导的分布，其中各生成模型的支撑集互不相交。在这种情况下，μ^* 的支撑流形是不连通流形，我们可以简单地

选取 $\{S_k\}_{k\in[K]}$ 为任意一组互不相交的集合，使其能够覆盖生成模型的每个支撑集（即，对于 $k \in [K]$, $\text{supp}((G_k^*)_{\#}\nu_k^*) \subset S_k$ ），并令 ρ_k 为指示函数 $\mathbf{1}(x \in S_k)$。作出这样的选择后，如果对于每个 $k \in [K]$, ν_k^* 都是具有紧支撑的 α-光滑函数，且 G_k^* 是具有 β-光滑逆的 β-光滑函数，则假设 A 成立。

定理 1。 固定 $\alpha \geq 0$, $\beta \geq 1$ 以及从属于开覆盖$\{S_k\}_{k\in[K]}$的单位分解 $\{\rho_k\}_{k\in[K]}$ 。假设目标分布 μ^* 满足假设 *A*，且逼近族 $\mathcal{G}$满足假设 *B*。如果我们选择差异度量 $\mathcal{D}(\cdot,\cdot)$ 为1-*Wasserstein*距离 W_1，并选择代价函数 $c(\cdot,\cdot)$ 为平方 ℓ_2 损失，那么存在一组正则化系数$\{\lambda_k\}_{k\in[K]}$，使得当 n足够大时，

1. 如果 $\alpha = 0$，那么通过选择重加权经验测度 $\widetilde{\nu}_{k,\,Q_k} = \frac{1}{\widehat{p}_k n}\sum_{i=1}^n \delta_{Q(X_i)}\rho_k(X_i)$ 作为插件 $\widehat{p}_k = \frac{1}{n}\sum_{i=1}^n \rho_k(X_i)$ ，所得估计量 $\widehat{\mu}$ 以至少 $1-n^{-1}$的概率满足

$$W_1(\widehat{\mu}, \mu^*) \leq C\, n^{-\frac{1}{d}} \vee \frac{\log n}{\sqrt{n}}. \tag{5}$$

2. 如果 $\alpha > 0, \beta > 1$，则存在一个光滑度正则化经验测度 $\widetilde{\nu}_{k,\,Q_k}$ 作为插件，使得由此得到的估计量 $\widehat{\mu}$ 以至少 $1 - n^{-1}$ 的概率满足

$$W_1(\widehat{\mu}, \mu^*) \leq C\, n^{-\frac{(\alpha\wedge(\beta-1))+1}{2(\alpha\wedge(\beta-1))+d}} \vee \frac{\log n}{\sqrt{n}}. \tag{6}$$

定理 1 的证明中采用的光滑度正则化经验测度 $\widetilde{\nu}_{k,\,Q_k}$ 可以基于小波截断，也可以基于测度 $Q_{k\#}(\frac{\mu^*\cdot\rho_k}{p_k})$ 的核密度估计量。根据 Liang（2020）和 Tang & Yang (2022) 得出的极小~~极大~~下界，当 $\alpha \leq \beta - 1$ 时，定理 1 中的收敛速率相对于 1-瓦瑟斯坦距离而言是极小极大最优的。如果假设 A 对 $K = 1$ 成立，则陈述 (5) 可以为 LDMAE 估计器提供理论保证。环境维度D 没有出现在所推导速率的指数中；因此，定理 1 表明 MLDMAE 能适应低维子流形结构，因为该界不受“维数灾难”的影响。此外，MLDMAE 估计量可以利用目标测度的光滑性，通过在惩罚项中对经验测度进行正则化，进一步提高估计精度。

6 模拟

本节展示将 MLDMAE 方法应用于常见流形——二维螺旋和环面时的一些可视化结果。精确的数据生成分布见附录 A。惩罚项 $\mathcal{D}(\cdot,\cdot)$ 采用通过“POT”软件包计算的 1-瓦瑟斯坦距离（Flamary 等人，2021），代价函数 $c(\cdot,\cdot)$ 为平方 ℓ_2 损失。作为基准，我们还考虑了：(1) 统计学文献中常用于密度估计的经典核密度估计量（KDE），其使用高斯核；(2) 使用与 MLDMAE 相同类型的代价函数和差异度量的 LDMAE估计器。不同方法学习得到的生成器所生成的样本见图 3。估计分布与真实分布之间的 1-瓦瑟斯坦距离见表 1。可以看出，采用多个编码器和解码器能够比采用单对编码器和解码器取得显著更好的性能。具体而言，即使增加编码器/解码器族的复杂度，LDMAE 仍无法捕捉这些统计学中标准流形的正确形状。另一方面，MLDMAE 允许使用多个编码器/解码器，因此只需相对简单的编码器/解码器结构和更少的训练参数，就能正确学习流形结构。此外，在这两个示例中，MLDMAE 的表现均优于经典 KDE。

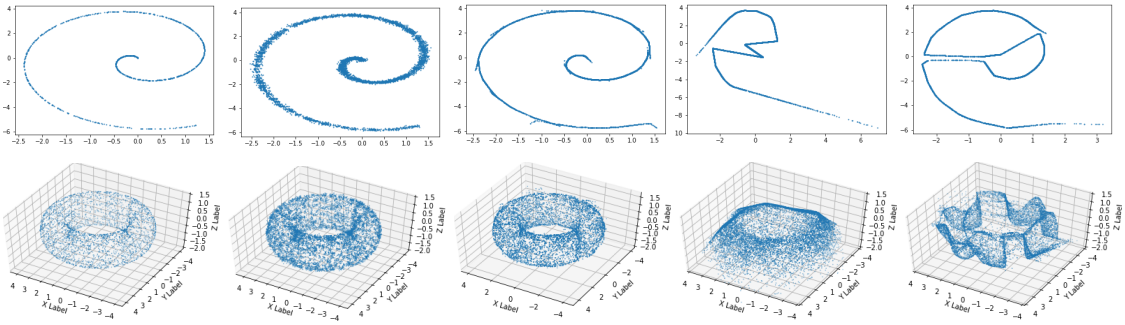


图 3：该图展示了目标测度位于螺旋和环面上时，MLDMAE、LDMAE 和 KDE 的表现。对于螺旋（上排）和环面（下排）的示例，我们分别观察到 $n = 1000$ 和 $n = 3000$ 个训练点。第一列绘制训练样本。第二列绘制由经典核密度估计量生成的样本，这相当于向原始训练样本添加高斯噪声。第三列绘制由我们提出的 MLDMAE 估计量生成的样本（螺旋为 $K = 10$，环面为 $K = 15$）；编码器和解码器由单隐藏层神经网络参数化，隐藏层大小为 128；单位分解选用第 4 节所述的光滑单位分解，其中 $\gamma = 10$；先验分布按备注1中的描述选取，其中 ν_0 为截断正态分布。第四列和第五列分别绘制由 LDMAE 估计量生成的样本，其编码器—解码器对分别由单隐藏层和双隐藏层神经网络参数化。

		KDE	MLDMAE	LDMAE（1近邻）	LDMAE（2近邻）
螺旋	W_1 距离	0.2119	0.1936	0.6315	0.4666
	训练参数数量	/	4492	1027	34051
环面	W_1 距离	0.2837	0.2335	0.9101	0.6193
	训练参数数量	/	10529	1412	34565

表 1：该表给出了螺旋和环面示例中目标测度与不同方法的分布估计器之间的 1-瓦瑟斯坦距离。我们还提供了不同方法的训练参数数量。

7 真实数据应用

本节使用三个真实世界数据集对所提出的 MLDMAE 方法进行实证评估：MNIST 手写数字（LeCun 等人，1995）、Fashion-MNIST（Xiao 等人，2017）和 CelebA 64×64（Krizhevsky 等人，2009）。作为比较，我们考虑 LDMAE 方法和变分自编码器（VAE）（Kingma 和 Welling，2013；Rezende 等人，2014；Kingma 等人，2016），这两者都是常用的基于自编码器的生成建模方法。在所有报告的实验中，我们将重构代价设为平方 ℓ_2 损失，并将先验分布 ν_k 固定为标准高斯分布 $\mathcal{N}(0, I_d)$。单位分解采用第 4 节所述的指示函数。编码器和解码器采用卷积深度神经网络建模，并按第 4 节所述进行参数共享；更多细节见附录 A。我们考虑 LDMAE 和 MLDMAE 的惩罚项中使用的两种差异度量 $\mathcal{D}(\cdot, \cdot)$：一种是通过“POT”软件包计算的 1-瓦瑟斯坦距离，另一种是采用逆多二次（IMQ）卷积核 $k(x, y) = 2d/(2d + \|x - y\|_2^2)$ 的最大均值差异（MMD）。正如 Tolstikhin 等人（2019）所述，逆多二次核的尾部要厚得多，

传统的 RBF 核 $k(x,\ y) = \exp(-\|x-y\|^2/C)$，因此能够为离群点提供更有意义的梯度。MLDMAE 的簇数 K 经过选择，使自由参数的数量约为共享参数数量的 $\frac{1}{5} \sim \frac{1}{2}$，结果表明 $K = 5$ 适用于所有示例。另一个重要因素是潜在维度，它在真实数据集中并不明确。潜在维度过小会导致较大的重建误差，进而产生噪声较大的生成样本。相反，潜在维度过大会导致编码分布奇异，从而造成数值不稳定性。我们对 Fashion-MNIST 使用 $d = 4$，对 MNIST 手写数字使用 $d = 8$，对 CelebA 使用 $d = 64$；这些设置似乎效果尚可，而且 MLDMAE 的训练稳定，并且对初始化具有鲁棒性。

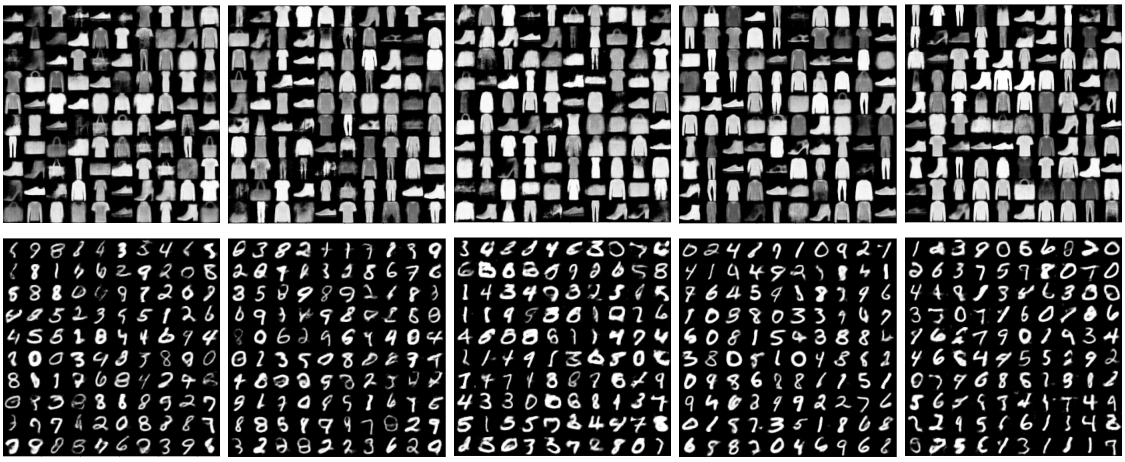
为了定量评估 MLDMAE 估计量，对于 MNIST 手写数字数据集和 Fashion-MNIST 数据集，我们考虑两类评估指标：一类是测试对数似然（Test LL），其做法是将高斯 Parzen 窗拟合到生成样本上，并报告在测试样本上评估得到的对数似然，这一方法由 Goodfellow 等人（2014）采用；另一类是测试样本与生成样本之间的 1-瓦瑟斯坦距离（W_1）。生成样本如图 4 所示，Test LL 和 W_1 距离列于表 2。可以看到，对于 Fashion-MNIST 数据集，采用 W_1 或 MMD 惩罚项的 MLDMAE 明显优于变分自编码器和 LDMAE。对于 MNIST 手写数字数据集，在 W_1 指标上，采用 MMD 惩罚项的 MLDMAE 优于采用 Wasserstein 惩罚项的 MLDMAE；这可能是因为 MNIST 手写数字数据集的潜在维度相对较大（$d = 8$），因此需要非常大的批量大小才能准确估计 Wasserstein 惩罚项，这会减少梯度更新的次数。此外，图 5 展示了 MNIST 手写数字数据集的簇数 K 增加时，Test LL 和 W_1 距离的变化趋势。可以看到，当 $M \geq 10$时，两项指标的变化趋势都趋于平稳，这与 MNIST 手写数字数据集包含 10 个数字（簇）这一事实一致。此外，当簇数在 $[1, 10]$, 范围内增加 K时，两项指标都有明显改善，而簇数 K 每增加 1，训练参数总数仅增加 7%。

表 2：MNIST 手写数字和 Fashion-MNIST 数据集：不同方法的测试对数似然和 W_1 距离；“+最大均值差异”和“$+W_1$”表示惩罚项中差异度量 $\mathcal{D}(\cdot,\cdot)$的选项。

	Fashion-MNIST		MNIST	
	测试对数似然 ↑	W_1 ↓	测试对数似然 ↑	W_1 ↓
VAE	533	77.7	393	63.2
LDMAE+最大均值差异	549	75.5	381	63.0
LDMAE+W_1	562	74.6	400	68.1
MLDMAE+最大均值差异	557	66.0	443	**56.6**
MLDMAE+W_1	**562**	**65.8**	**456**	63.1

表 3：CelebA 数据集：不同方法的弗雷歇起始距离和核起始距离。LDMAE 和 MLD-MAE 中的惩罚项选用最大均值差异惩罚项。

	弗雷歇起始距离 ↓	核起始距离 ↓
VAE	63.0	0.063
LDMAE+最大均值差异	55.5	0.057
LDMAE+SW_1	62.8	0.064
MLDMAE+最大均值差异	**51.1**	**0.051**
MLDMAE+SW_1	52.0	0.052



VAE LDMAE+最大均值差异 LDMAE+W_1 MLDMAE+最大均值差异 MLDMAE+W_1

图 4：Fashion-MNIST（上排）和 MNIST 手写数字数据集（下排）中不同方法生成的样本。第一列对应变分自编码器估计量；第二列和第三列分别对应使用最大均值差异和 W_1 惩罚项的 LMDAE 估计量；第四列和第五列分别对应使用最大均值差异和 W_1 惩罚项的 MLMDAE 估计量。

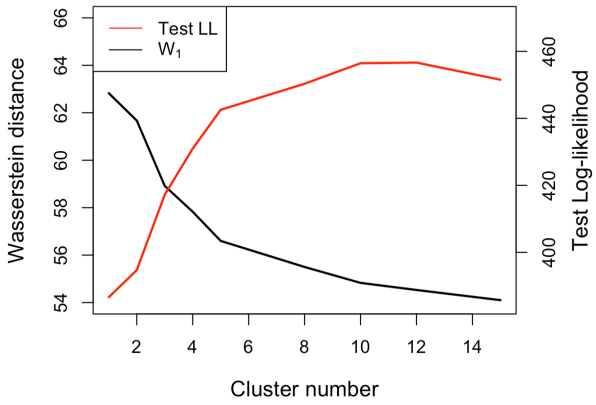


图 5：针对 MNIST 手写数字数据集，采用最大均值差异惩罚项且簇数不同的 MLDMAE 的负测试对数似然（红色）和 W_1 距离（黑色）。 K

对于 CelebA 数据集，由于环境维度 $D = 64 \times 64 \times 3$ 极大，因此我们不采用在环境空间中评估的测试对数似然或 W_1 距离，而是考虑两种常用于彩色图像数据的指标：FID（Heusel等人，2017年）和 KID（Bińkowski等人，2018年）。在这两种指标中，原始高维数据被输入一个经 ImageNet 预训练的 Inception 网络，以获得 2048 维的 Inception（特征）表示；FID 和 KID 分别是生成样本与测试样本的 Inception 表示之间的弗雷歇距离和平方 MMD。此外，如前所述， W_1 惩罚项不适用于较大的内在维度。为避免维数灾难，我们考虑所谓的切片瓦瑟斯坦距离：$SW_1(\mu, \nu) := \mathbb{E}_{\theta \sim \mathrm{Unif}(\mathbb{S}_1^{d-1})}[W_1(\mathrm{Proj}_{\theta}\#\mu, \mathrm{Proj}_{\theta}\#\nu)]$，其中 Proj_θ 表示投影到方向 θ 的投影函数，而 $\mathrm{Unif}(\mathbb{S}_1^{d-1})$表示 $\mathbb{S}_1^{d-1}$上的均匀分布。对 $\mathrm{Unif}(\mathbb{S}_1^{d-1})$取期望可通过蒙特卡洛方法估计。切片瓦瑟斯坦距离将高维概率密度切分为一组一维边缘分布，并通过 Wasserstein距离比较这些边缘分布。它具有与 Wasserstein距离相似的定性性质，但计算起来容易得多。生成样本以及 FID、KID 见图 6 和表 3。

我们可以看到，MLDMAE估计量在所有评估指标上都取得了最佳表现。采用最大均值差异惩罚项的MLDMAE 略优于采用 SW_1 惩罚项的 MLDMAE，但计算时间也更长（使用 NVIDIA A100-SXM4-40GB GPU 时，采用最大均值差异的 MLDMAE 每个 epoch 需要 150 秒，而采用 $+SW_1$ 的 MLDMAE 每个 epoch 需要 120 秒）。



（a）变分自编码器　（b）LDMAE+最大均值差异　（c）LDMAE+SW_1　（d）MLDMAE+最大均值差异　（e）MLDMAE+SW_1

图6：CelebA 数据集上不同方法生成的样本（第一列：变分自编码器；第二列和第三列：分别为采用最大均值差异和 SW_1 惩罚项的 LDMAE；第四列和第五列：分别为采用最大均值差异和SW_1 惩罚项的 MLDMAE）。

8 结论

在本研究中，我们提出了一种新方法——潜分布匹配自编码器混合（MLD-MAE），以改进传统的基于自编码器的生成建模方法，从而学习流形支撑分布。我们从理论上证明，所提出的估计量能够以极小极大最优收敛率学习流形支撑分布。此外，我们通过实验表明，采用多组编码器/解码器对后，MLDMAE得出的估计量能够显著提高目标分布估计精度。在理论分析中，我们考虑惩罚项取为Wasserstein距离的情形。至于其他一些对抗损失（例如实验中考察的MMD距离）的理论分析，则留待未来研究。

参考文献

Arjovsky, M.、Chintala, S. 与 Bottou, L.（2017年）。Wasserstein生成对抗网络。载于国际机器学习大会。PMLR。

Berenfeld, C., Rosa, P. & Rousseau, J.（2022年）。在未知流形附近估计密度：一种贝叶斯非参数方法。*arXiv* 预印本 *arXiv:2205.15717* 。

Bińkowski, M., Sutherland, D. J., Arbel, M. & Gretton, A.（2018年）。揭开 MMD 生成对抗网络的奥秘。*arXiv* 预印本 *arXiv:1801.01401* 。

Bonneel, N., Van De Panne, M., Paris, S. & Heidrich, W.（2011）。使用拉格朗日质量输运进行位移插值。载于*2011*年 *SIGGRAPH Asia* 大会论文集。

Bouzebda, S. & Didi, S.（2017年）。平稳遍历离散时间过程的多元小波密度与回归估计量：渐近结果。《统计学通讯：理论与方法》**46**，1367–1406。

Brock, A., Donahue, J. & Simonyan, K.（2018年）。用于高保真自然图像合成的大规模生成对抗网络训练。

Caffarelli, L. A.（1996年）。凸势函数映射的边界正则性——II。《数学年刊》144，453–496。

Chen, Y., Gao, Q. & Wang, X.（2022年）。推断型Wasserstein生成对抗网络。《皇家统计学会会刊：B辑（统计方法学）》84，83–113。

Divol, V.（2022年）。流形上的测度估计：一种最优传输方法。《概率论及相关领域》。

Eldering, J.（2013年）。《正规双曲不变流形：非紧情形》。Paris：Atlantis出版社。

Evans, L. C.（2010a）。偏微分方程，第19卷。美国数学学会。

Evans, L. C.（2010年b）。偏微分方程。美国罗得岛州普罗维登斯：美国数学学会。

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N. 等（2021年）。POT：Python 最优传输。机器学习研究期刊 **22**，1–8。

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. 和 Bengio, Y.（2014年）。生成对抗网络。

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S.（2017年）。通过双时间尺度更新规则训练的生成对抗网络收敛至局部纳什均衡。Jupp, P. E. & Mardia, K. V.（2009年）。方向统计学。约翰·威立父子公司。Khayatkhoei, M., Singh, M. K. & Elgammal, A.（2018年）。生成对抗网络的非连通流形学习。载于神经信息处理系统进展，S. Bengio、H. Wallach、H. Larochelle、K. Grauman、N. Cesa-Bianchi & R. Garnett 编，第31卷。Curran Associates, Inc. Kingma, D. P. & Ba, J.（2014年）。Adam：一种随机优化方法。*arXiv*预印本 *arXiv:1412.6980* 。

Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I. & Welling, M.（2016年）。使用逆自回归流改进变分推断。载于神经信息处理系统进展，D. Lee、M. Sugiyama、U. Luxburg、I. Guyon & R. Garnett 编，第29卷。Curran Associates, Inc. Kingma, D. P. & Welling, M.（2013年）。自编码变分贝叶斯。Kolouri, S., Pope, P. E., Martin, C. E. & Rohde, G. K.（2018年）。切片沃瑟斯坦自编码器：一种极其简单的生成模型。*arXiv*预印本 *arXiv:1804.01947* 。Krizhevsky, A., Hinton, G. 等（2009年）。从微型图像中学习多层特征。Lan, Z., Reich, B. J. & Bandyopadhyay, D.（2021年）。用于正定矩阵的空间贝叶斯半参数混合模型及其在弥散张量成像中的应用。加拿大统计学杂志 **49**, 129–149。LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Muller, U. A., Sackinger, E., Simard, P. 等（1995年）。分类学习算法：手写数字识别方面的比较。神经网络：统计力学视角**261**，2。Li, C.-L., Chang, W.-C., Cheng, Y., Yang, Y. & Póczos, B.（2017年）。最大均值差异生成对抗网络：深入理解矩匹配网络。神经信息处理系统进展 **30**。Liang, T.（2020年）。生成对抗网络学习分布的效果如何。Lin, L., Lazar, D., Sarpabayeva, B. & Dunson, D. B.（2020年）。流形上的鲁棒优化与推断。*arXiv*预印本 *arXiv:2006.06843* 。Lin, L., Rao, V. & Dunson, D.（2017年）。斯蒂费尔流形上的贝叶斯非参数推断。《统计学报》，535–553。Ling, Y., An, Y., Liu, M., Hasan, S. A., Fan, Y. & Hu, X.（2017年）。将额外知识整合到生物医学自然语言处理任务的词嵌入模型中。载于*2017*年国际神经网络联合会议（*IJCNN*）。IEEE。Luo, L., Yang, Z., Cao, M., Wang, L., Zhang, Y. & Lin, H.（2020年）。一种从生物医学文献中进行生物医学实体与关系抽取的基于神经网络的联合学习方法。《生物医学信息学杂志》**103**，103384。Mardia, K.（1999年）。方向统计学与形状分析。《应用统计学杂志》**26**，949–957。Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. & Kavukcuoglu, K.（2016年）。WaveNet：一种用于原始音频的生成模型。Rezende, D. J., Mohamed, S. & Wierstra, D.（2014年）。深度生成模型中的随机反向传播与近似推断。Silverman, B. W.（2018年）。用于统计学与数据分析的密度估计。Routledge。Singh, S., Uppal, A., Li, B., Li, C.-L., Zaheer, M. & Poczos, B.（2018年）。对抗损失下的非参数密度估计。载于神经信息处理系统进展，S. Bengio、H. Wallach、H. Larochelle、K. Grauman、N. Cesa-Bianchi & R. Garnett 编，第31卷。Curran Associates, Inc.

Tang, R. & Yang, Y. (2022). 对抗损失下未知子流形上分布估计的极小极大速率。*arXiv*预印本 *arXiv:2202.09030* 。Terradot, L., Durnell, N., Li, M., Li, M., Ory, J., Labigne, A., Legrain, P., Colland, F. & Waksman, G. (2004年)。幽门螺杆菌蛋白质相互作用图谱中蛋白质复合物的生化表征：复合物形成策略及IV型分泌系统中新型相互作用的证据。《分子与细胞蛋白质组学》**3**，809–819。Tolstikhin, I., Bousquet, O., Gelly, S. & Schoelkopf, B. (2019)。Wasserstein自编码器。Tsybakov, A. B. (2009)。非参数估计导论。纽约，纽约州：施普林格纽约。Villani, C. (2009)。最优传输：旧与新。柏林，海德堡：施普林格柏林海德堡。Wainwright, M. J. (2019)。高维统计：一种非渐近视角。剑桥统计与概率数学丛书。剑桥大学出版社。Xiao, H., Rasul, K. & Vollgraf, R. (2017)。Fashion-MNIST：一个用于机器学习算法基准测试的新型图像数据集。*arXiv*预印本 *arXiv:1708.07747* 。Xu, Q., Huang, G., Yuan, Y., Guo, C., Sun, Y., Wu, F. & Weinberger, K. (2018)。生成对抗网络评估指标的实证研究。*arXiv*预印本 *arXiv:1806.07755* 。You, Z.-H., Lei, Y.-K., Gui, J., Huang, D.-S. & Zhou, X. (2010)。利用流形嵌入，根据高通量实验数据评估并预测蛋白质相互作用。《生物信息学》**26**，2744–2751。Zhang, R., Ogden, R. T., Picard, M. & Srivastava, A. (2022年)。形状空间上的非参数k样本检验及其在线粒体形状分析中的应用。《皇家统计学会期刊：*C*辑（应用统计）》**71**，51–69。Zhao, S., Song, J. & Ermon, S. (2019)。InfoVAE：平衡变分自编码器中的学习与推断。《*AAAI*人工智能会议论文集》**33**，5885–5892。

附录

符号说明：我们采用本文中的符号，并进一步引入以下符号~用于技术证明。我们用 $\mathrm{N}(\mathcal{F}, d, \epsilon)$ 表示函数空间~$\mathcal{F}$ 关于伪度量 d的 ϵ-覆盖数。我们用 $\mathbb{B}_r(x)$ 表示以 x 为中心、半径为r 的闭球，其度量为 ℓ_2 距离；特别地，在不存在歧义时，我们用 $\mathbb{B}_r^d$ 表示 $\mathbb{B}_r(\mathbf{0}_d)$。我们记 $\mathbb{S}_1^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$。对于函数 $f : \Omega \to \mathbb{R}^d$，我们用 $\mathrm{J}f(x)$ 表示 f 在 Ω 中 $x \in$ 处的雅可比矩阵。对于函数 $f: \mathbb{R}^d \to \mathbb{R}$，我们用 $f^{(a)}$ 表示其混合偏导数 $\partial^{|a|} f/\partial x_1^{a_1} \cdots \partial x_d^{a_d}$。我们将半径为 $r > 0$、定义在 Ω 上的 α-光滑 Hölder（函数）类（参见 Evans (2010b)）定义为 $C_r^\alpha(\Omega) := \big\{f: \Omega \to \mathbb{R} \,\big|\, \|f\|_{C^\alpha(\Omega)} = \sum_{|a|\le\lfloor\alpha\rfloor} \max_{x\in\Omega} |f^{(a)}(x)| + \sum_{|a|=\lfloor\alpha\rfloor} \max_{x,y\in\Omega,\, x\neq y} \big|f^{(a)}(x) - f^{(a)}(y)\big| / \|x-y\|^{\tilde{\alpha}-\lfloor\alpha\rfloor} \le r\big\}$。类似地，我们用 $C_r^\alpha(\Omega; \mathbb{R}^D) = \big\{f = (f_1, \ldots, f_D) : \Omega \to \mathbb{R}^D \,\big|\, \forall j \in [D],\, f_j \in C_r^\alpha(\Omega)\big\}$ 表示其对应的向量值函数空间。对于一个 $f \in C_r^\alpha(\Omega; \mathbb{R}^D)$ 和一个多重指标 $a \in \mathbb{N}_0^d$，我们将 $f^{(a)}$ 记为 D 维向量，其第 j个分量是 f_j 的混合偏导数 $[f_j]^{(a)}$，其中 $j \in [D]$。全文中，C, c, C_0, c_0, C_1, c_1,C_2, c_2,... 通常用于表示正常数，其取值可能逐行变化，但与其他所有量均无关。

A 剩余的实现细节

A.1 模拟

螺旋中的训练点 $\mathcal{D}$按以下步骤生成：(1) 生成 $\phi_0 \sim \mathcal{N}(0, 1)$；(2) 设置$\phi = 3\pi\phi_0$；(3) 通过 $X = [\frac{\cos(\phi+2)\cdot\phi}{\pi}, {}^2\sin(\phi+2\cdot\phi)]$ 生成数据点 X。环面~~中的训练点 $\mathcal{D}$按~~以下步骤生成：(1) 生成 ϕ_0, $\phi_1 \sim \mathcal{N}(0, 1)$；(2) 设置 $\phi = 2\pi\phi_0$和 $\theta = 2\pi\phi_1$；(3) 通过 $X = [(3 + \cos(\theta))\cos(\phi), (3 + \cos(\theta))\sin(\phi), \sin(\theta)]$生成数据点 X。MLDMAE 中的簇数M 对于螺旋数据集为 $M = 10$，对于环面数据集为 $M = 15$。单位分解由 $\rho_k = \widetilde{\rho}_k / \big(\sum_{k'=1}^K \widetilde{\rho}_{k'}\big)$ 给出，其中 $\widetilde{\rho}_k = (r_k^2 - \|x - a_k\|^2)^{10} \cdot \mathbf{1}(x \in S_k)$, $\{a_k\}_{k\in[K]}$ 为 K-均值算法返回的中心，$r_k = \sup\{\|x - a_k\| : x \in \mathcal{D};\ \forall k_1 \in [K], \|x - a_k\| \le \|x - a_{k_1}\|\}$和 $S_k = \mathbb{B}_{r_k}(a_k)$。

A.2 真实数据应用

我们在 MNIST 手写数字、Fashion-MNIST 和 CelebA 上训练的模型规格见表 4、5 和 6。“共享”是指编码器之间或解码器之间的参数共享。所有模型均使用 Adam优化进行优化，学习率为 0.001, $\beta_1 = 0.9$, 以及 $\beta_2 = 0.999$。所有数据集的单位分解均选为指示函数 $\rho_k(x) = \mathbf{1}(x \in$ 簇 $m)$，其中 $k \in [K]$。用于复现实验的代码可在 https://github.com/rtang1997/MLDMAE 获取。

操作	卷积核	步幅	特征图	激活函数	是否共享?
解码器 $G_k(z): k \in [K], z \in \mathcal{N}(0, I_d)$			8		
全连接			$3 \times 3 \times 128$	ReLU	No
转置卷积	3×3	2×2	$7 \times 7 \times 64$	ReLU	Yes
转置卷积	3×3	2×2	$14 \times 14 \times 32$	ReLU	Yes
转置卷积	3×3	2×2	$28 \times 28 \times 1$	ReLU	Yes
编码器 $Q_k(x): k \in [K]$			$28 \times 28 \times 1$		
卷积	3×3	2×2	$26 \times 26 \times 3$	带泄漏的修正线性单元	Yes
卷积	3×3	2×2	$12 \times 12 \times 32$	带泄漏的修正线性单元	Yes
卷积	3×3	2×2	$5 \times 5 \times 64$	带泄漏的修正线性单元	Yes
卷积	3×3	2×2	$2 \times 2 \times 128$	带泄漏的修正线性单元	Yes
全连接			8		No
MLDMAE 的簇数 K	5				
批量大小	MLDMAE 的 256 ，WAE 和变分自编码器为 128				
训练轮数	50				
Leaky ReLU 斜率	0.1				
正则化系数（λ_k）	最大均值差异惩罚项为 100， W_1 惩罚项为 10				
MLDMAE 的带宽（h）	0.01				
训练样本数量	60k				

表 4：MNIST 手写数字数据集的编码器/解码器网络架构和超参数。

操作	卷积核	步幅	特征图	激活函数	是否共享?
解码器 $G_k(z)$: $k \in [K]$， $z \in \mathcal{N}(0$， $I_d)$			4		
全连接			$3 \times 3 \times 128$	ReLU	No
转置卷积	3×3	2×2	$7 \times 7 \times 64$	ReLU	Yes
转置卷积	3×3	2×2	$14 \times 14 \times 32$	ReLU	Yes
转置卷积	3×3	2×2	$28 \times 28 \times 1$	ReLU	Yes
编码器 $Q_k(x): k \in [K]$			$28 \times 28 \times 1$		
卷积	3×3	2×2	$26 \times 26 \times 3$	带泄漏的修正线性单元	Yes
卷积	3×3	2×2	$12 \times 12 \times 32$	带泄漏的修正线性单元	Yes
卷积	3×3	2×2	$5 \times 5 \times 64$	带泄漏的修正线性单元	Yes
卷积	3×3	2×2	$2 \times 2 \times 128$	带泄漏的修正线性单元	Yes
全连接			4		No
MLDMAE 的簇数 K	5				
批量大小	MLDMAE 为 256 ，WAE 和变分自编码器为 128				
训练轮数	50				
Leaky ReLU 斜率	0.1				
正则化系数 (λ_k)	最大均值差异惩罚项为 100， W_1 惩罚项为 10				
MLDMAE 的带宽（h）	0.01				
训练样本数量	60k				

表5：Fashion-MNIST 数据集的编码器/解码器架构和超参数。

操作	卷积核	步幅	特征图	激活函数	是否共享?
解码器 $G_k(z)$：$k \in [K], z \in \mathcal{N}(0, I_d)$			64		
全连接			$8 \times 8 \times 1024$	ReLU	No
转置卷积	5×5	2×2	$16 \times 16 \times 512$	ReLU	Yes
批量归一化					
转置卷积	5×5	2×2	$32 \times 32 \times 256$	ReLU	Yes
批量归一化					
转置卷积	5×5	2×2	$64 \times 64 \times 128$	ReLU	Yes
批量归一化					
转置卷积	3×3	1×1	$64 \times 64 \times 3$	Tanh	Yes
编码器 $Q_k(x)$：$k \in [K]$			$64 \times 64 \times 3$		
卷积	5×5	2×2	$32 \times 32 \times 128$	ReLU	Yes
批量归一化					
卷积	5×5	2×2	$16 \times 16 \times 256$	ReLU	Yes
批量归一化					
卷积	5×5	2×2	$8 \times 8 \times 512$	ReLU	Yes
批量归一化					
卷积	5×5	2×2	$4 \times 4 \times 1023$	ReLU	Yes
批量归一化					
全连接			64		No
MLDMAE 的簇数 K	5				
批量大小	MLDMAE 为 256 ，WAE 和变分自编码器为 128				
训练轮数	50				
正则化系数（λ_k）	100				
SWAE 和 MLDMAE 的带宽（h）	0.01				
训练样本数量	180k				

表6：CelebA 数据集的编码器/解码器架构和超参数。

B 子流形上的分布生成建模

环境空间 $\mathbb{R}^D$ 中的子流形可视为一种非线性“子空间”。借鉴 Tang & Yang (2022) 中的定义，我们将 $\mathbb{R}^D$ 上无边界的 d 维光滑紧子流形上的光滑分布族定义为集合 $\mathcal{P}^* = \mathcal{P}^*(d,\ D,\ \alpha,\ \beta,\ L^*)$，其中 $d \leq D$、 $\beta > 1$ 和 $\alpha \in (0,\ \beta - 1]$ 由满足以下条件的所有概率测度 $\mu \in \mathcal{P}(\mathbb{R}^D)$ 构成：

1. μ 是嵌入于$\mathbb{R}^D$中的紧致 d维子流形 $\mathcal{M}$ 上的 α-光滑分布。2. 相对于 $\mathcal{M}$ 的体积测度的密度 μ 在$\mathcal{M}$上有一致的下界 $1/L^*$ 。

3. $\mathcal{M}$ 可由图册 $\mathscr{A} = \{(U_\lambda, \phi_\lambda)\}_{\lambda \in \Lambda}$ 覆盖于 $\mathcal{M}$ 上，满足：a) 图册 $\mathscr{A}$中的每个坐标图(U, ϕ)均满足 $\|\phi^{-1}\|_{C^\beta}(\phi(U)) \leq L^*$ 且$\|\mu \circ \phi^{-1}\|_{C^\alpha}(\phi(U)) \leq L^*$；b) 对任意 $z \in \phi(U)$， $\phi^{-1}(z)$的雅可比矩阵满秩，且其所有奇异值的绝对值均有下界 $1/L^*$ 。此外，对任意$x \in \mathcal{M}$，存在一个 $\lambda \in \Lambda$，使得 U_λ，且 $\phi_\lambda(U_\lambda)$分别覆盖 $\mathbb{B}_{1/L^*}(x) \cap \mathcal{M}$ 和 $\mathbb{B}_{1/L^*}(\phi_\lambda(x))$。

我们有以下引理，用于描述能够对子流形支撑分布进行建模的生成模型类的混合。

引理 *1.* 考虑 $\mathscr{O}_K = \{S_k = B^\circ_{r_k}(a_k)\}_{m=1}^K$，对于 $k \in [K]$，选择 $\rho_k(x) = \widetilde{\rho}_k(x)/\sum_{k\in[K]}\widetilde{\rho}_k(x)$，其中$\widetilde{\rho}_k(x) = (r_k^2 - \|x-a_k\|^2)^\gamma \cdot \mathbf{1}(x \in S_k)$，对于 $\gamma \geq \alpha+1$。存在一个常数 r^*，它仅取决于 $(L^*, \beta, \alpha, d, D)$，使得对于任意 $\mu^* \in \mathcal{P}^*(d, D, \alpha, \beta, L^*)$，若满足：*(1)* $supp(\mu^*) \subset \cup_{k\in[K]} S_k$；*(2)* 对任意$k \in [K]$、$r_k \leq r^*$ 和 $a_k \in \mathcal{M}$；*(3)* 存在某些正常数 L_1^*，使得 $\min_{k\in[K]} r_k \geq L_1^*$ 且 $\inf_{x\in\mathcal{M}} \sum_{k\in[K]} \widetilde{\rho}_k(x) \geq L_1^*$。则：

1. 存在一些只依赖于 $(L^*, L_1^*, \beta, \alpha, d, D, \gamma)$ 的通用常数 (L, c)，使得对于任意 $k \in [K]$，假设 *A* 对 μ^* 成立，其上界为 L，且函数为 $g_k(r) = c(r^\gamma \wedge 1)$；

2. 考虑 $\nu_0 \in \mathcal{P}(\mathbb{B}_1^d)$，其密度为 α-光滑且有正的下界，以及逼近族

(a) $\mathcal{G}_1 = \{(G, Q, v)$: $\forall k \in [K], G_k \in C_L^\beta(\mathbb{R}^d; \mathbb{R}^D)$ $Q_k \in C_L^\beta(\mathbb{R}^D; \mathbb{R}^d)$ $\nu_k \in \mathcal{P}(\mathbb{B}_1^d)$，其中 $\nu_k \in C_L^\alpha(\mathbb{B}_1^d)\}$；*(b)*。 $\mathcal{G}_2 = \{ G\ Q\ v \quad \forall k \in [K]\ G_k \in C_L^\beta\ \mathbb{R}^d\ \mathbb{R}^D \quad Q_k \in C_L^\beta\ \mathbb{R}^D\ \mathbb{R}^d \quad v_k = \frac{V_{m\#\nu_0}\quad G_k\ z}{\mathbb{E}_{\nu_0}[\rho_k\ G_k\ V_k\ z\ ]}\ V_k \in C_L^{\alpha+1}\ \mathbb{B}_1^d\ \mathbb{B}_1^d$

(,,): ,(;)(;)(),(;); ;

(c). $\mathcal{G}_3 = \{(G, Q, v) : \forall k \in [K], G_k \in C_L^\beta(\mathbb{R}^d; \mathbb{R}^D), Q_k \in C_L^\beta(\mathbb{R}^D; \mathbb{R}^d), v_k = \frac{\nu_0 \cdot \rho_k(G_k(z))}{\mathbb{E}_{\nu_0}[\rho_k(G_k(z))]}\}$

因此，当 L 足够大时，假设 *B* 对 $\mathcal{G}_1$ 或 $\mathcal{G}_2$ 成立，也就是说，逼近族 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 都足以对 $\mathcal{P}^*(d, D, \alpha, \beta, L^*)$ 内的分布进行建模。此外，如果 $\alpha = \beta - 1$，则假设 *B* 对 $\mathcal{G}_3$ 成立。

下一个引理表明，可以根据一小部分数据找到满足引理 1 条件的 $\mathscr{O}_K$。

引理 2. 考虑任意一个支撑集为 $\mathcal{M}$ 的 $\mu^* \in \mathcal{P}^*(d, D, \alpha, \beta, L^*)$，固定任意正常数 r^*，并令 $n_1 \leq n$ 为正整数。则对于任意正常数 c，存在仅依赖于 $(d, D, \beta, L^*, r^*, c)$ 的常数 C, c_1，使得当 $C \leq n_1 \leq n$ 时，令 I_1 为 $[n]$ 的任意子集，且满足 $|I_1| = n_1$，则以大于 $1 - n_1^{-c}$ 的概率有

(1). $\mathcal{M} \subset \bigcup_{i\in I_1} \mathbb{B}_{c_1(\frac{\log n_1}{n_1})^{\frac{1}{d}}}(x_i)$；*(2).* 存在一个常数 K，它仅取决于(d, D, β, L^*, r^*)，以及一个子集 $\{a_k\}_{k=1}^K \subset \{X_i\}_{i\in I_1}$，使得

(a). $\bigcup_{i\in I_1} \mathbb{B}_{c_1(\frac{\log n_1}{n_1})^{\frac{1}{d}}}(x_i) \subset \bigcup_{k=1}^K \mathbb{B}_{r^*}(a_k)$.

(b). $\inf_{x\in\mathcal{M}} \sum_{k\in[K]} \widetilde{\rho}_k(x) > (\frac{r^*}{\sqrt{2}})^{2\gamma}$, *where* $\widetilde{\rho}_k(x) = ((r^*)^2 - \|x - a_k\|^2)^\gamma \cdot \mathbf{1}(x \in \mathbb{B}_{r^*}(a_k))$.

C 定理 5.1 的证明

我们首先考虑 $\alpha > 0$ 和 $\beta > 1$ 的情形。为简化符号说明，记 $\tilde{\alpha} = \alpha \wedge (\beta - 1)$。我们考虑两种光滑度正则化经验测度 $\widetilde{\nu}_{k,\ Q_k}$：一种基于核密度估计量，另一种基于小波估计量。

核密度估计量： 定义

$$\widetilde{\nu}_{k,Q_k}(y) = \frac{1}{n\widehat{p}_k h^d} \sum_{i=1}^n \widetilde{k}\big(\frac{y - Q_k(X_i)}{h}\big)\rho_k(X_i), \quad \widehat{p}_k = \frac{1}{n}\sum_{i=1}^n \rho_k(X_i), \tag{7}$$

对于 $h = n^{-1/(2\tilde{\alpha}+d)}$ 和 $\widetilde{k}$： $\mathbb{R}^d \to \mathbb{R}$ 满足

1.$\widetilde{k}(\cdot)$ 在 $\mathbb{R}^d$ 上是 $\lceil\tilde{\alpha}\rceil \vee \lceil\frac{d}{2}\rceil$ 光滑的，且其支撑集包含于 $[-1, 1]^d$；2. $\int_{\mathbb{R}^d} \widetilde{k}(z)\,\mathrm{d}z = 1$，并且对于任意满足 $1 \leq |j| \leq \lfloor\alpha\rfloor + 1$ 的 $j \in \mathbb{N}_0^d$， $\int \widetilde{k}(z) \cdot z^j\,\mathrm{d}z = 0$;

3. 对于任意 $z \in \mathbb{R}^d$， $\widetilde{k}(z) = \widetilde{k}(-z)$。

小波估计量： 定义 $\widetilde{\nu}_{k,\ Q_k}(y)$ 为

$$\widetilde{\nu}_{k,Q_k}(y) = \frac{1}{\widehat{p}_k}\Big(\sum_{m\in\mathbb{S}} \widetilde{a}_m^{Q_k}\phi_m(y) + \sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}} \widetilde{\theta}_{ljm}^{Q_k}\psi_{ljm}(y)\Big), \tag{8}$$

with

$$\widehat{p}_k = \frac{1}{n}\sum_{i=1}^{n}\rho_k(X_i);$$
$$\mathbb{S} = \{m \in \mathbb{Z}^d \,|\, \mathrm{supp}(\phi_m) \cap [-L,L]^d \neq \emptyset\};$$
$$\mathbb{S}_{lj} = \{m \in \mathbb{Z}^d \,|\, \mathrm{supp}(\psi_{ljm}) \cap [-L,L]^d \neq \emptyset\};$$

$$\widetilde{a}_m^{Q_k} = \frac{1}{n}\sum_{i=1}^{n}\phi_m(Q_k(X_i))\rho_k(X_i);$$
$$\widetilde{\theta}_{ljm}^{Q_k} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ljm}(Q_k(X_i))\rho_k(X_i),$$

其中 $2^{dJ} \asymp n^{\frac{d}{2\alpha+d}}$和 $\{\phi_m, \psi_{ljm}\colon\ l = 1,\ \cdots,\ 2^d-1,\ j \in \mathbb{N},\ m \in \mathbb{Z}^d\}$ 是定义在 $\mathbb{R}^d$ 上的 Besov 空间的正交小波基，定义为 $\phi_m(y) = \phi(y-m)$ 和 $\psi_{ljm}(y) = 2^{\frac{jd}{2}}\psi_l(2^j y - m)$，并且对于任意 $1 \le l \le 2^d-1$ (Bouzebda & Didi, 2017)， $\phi(\cdot)$ 和 $\psi_l(\cdot)$均为紧支撑，且具有有界的 $\lceil \alpha \vee (\frac{d}{2}-\alpha)\rceil$ 阶导数。

我们将证明这两种 $\widetilde{\nu}_{k,\ Q_k}$ 选择都能得到所需结果。根据 μ^*的假设 A，对于任意 $k \in [K]$，存在 $G_k^* \in C_L^\beta(\mathbb{R}^d;\mathbb{R}^D)$ 和 $Q_k^* \in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d)$，使得.....并且对于任意 $x \in \mathcal{M}\cap S_k$， $G_k^*(Q_k^*(x)) = x$。由于$\mathbf{G}$、$\mathbf{Q}$ 和 $\widehat{\mathbf{v}}$ 在训练目标上的最优性，我们可以得到

$$\begin{aligned}&\sum_{k=1}^{K}\Big\{\frac{1}{n}\sum_{i=1}^{n}\|X_i - \widehat{G}_k(\widehat{Q}_k(X_i))\|^2\rho_k(X_i) + \lambda_k\cdot\sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(z)\widehat{\nu}_k(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z\Big)\Big\}\\ &\le \sum_{k=1}^{K}\lambda_k\cdot\sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(z)\nu_k^*(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,Q_k^*}(z)\,\mathrm{d}z\Big),\end{aligned} \tag{9}$$

其中，回顾 $\nu_k^* = (Q_k^*)_\#(\frac{\mu^*\rho_k}{p_k})$。于是我们有以下引理。

引理 3. 对于任意固定的 c_1 和 c_2，定义

$$\widetilde{\mathcal{Q}}_k = \{Q \in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d) : \textit{the density } \nu_{k,Q}^* \textit{ of } Q_\#\big(\frac{\mu^*\cdot\rho_k}{p_k}\big) \textit{ exists and } \nu_{k,Q}^* \in C_{c_2}^{\tilde{\alpha}}(\mathbb{R}^d)\}.$$

于是存在一个常数 c_3，使得以大于 $1-\frac{1}{n^2}$ 的概率，对于任意 $k \in [K]$，都有 —

$$\sup_{Q\in\widetilde{\mathcal{Q}}_k}\sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(z)\nu_{k,Q}^*(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,Q}(z)\,\mathrm{d}z\Big) \le c_4\Big(n^{-\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} + \frac{\log n}{\sqrt{n}}\Big),$$

其中 $\widetilde{\nu}_{k,\ Q}$ 可以是(7) 中的核密度估计量，也可以是(8)中的小波估计量。

因此我们 对于任意 $k \in [K]$，选择 $\lambda_k = \lambda = \big(n^{-\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} + \text{对数}\sqrt{n}^{\,n}\big)^{-1}\cdot n^{-\frac{2\beta}{d}-1}$，则根据第二个陈述 of

引理 3 和 $Q_k^* \in \widetilde{\mathcal{Q}}_k$，以大于 $1 - M\,n^{-2}$ 的概率成立，

$$\sum_{k=1}^{K}\Big\{\frac{1}{n}\sum_{i=1}^{n}\|X_i - \widehat{G}_k(\widehat{Q}_k(X_i))\|^2\rho_k(X_i) + \lambda\cdot\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(z)\widehat{\nu}_k(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z\Big)\Big\}$$
$$\le C\,\lambda\cdot\big(n^{-\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} + \frac{\log n}{\sqrt{n}}\big).$$

因此，以大于 $1 - M\,n^{-2}$ 的概率，对于任意 $k \in [K]$,

$$\frac{1}{n}\sum_{i=1}^{n}\|X_i - \widehat{G}_k(\widehat{Q}_k(X_i))\|^2\rho_k(X_i) \le C\,n^{-\frac{2\beta}{d}-1}, \tag{10}$$

and

$$\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(z)\widehat{\nu}_k(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z\Big) \le C\left(n^{-\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} + \frac{\log n}{\sqrt{n}}\right). \tag{11}$$

接下来，我们使用以下引理来界定总体重构误差。

引理 4. 对于估计量 $\widehat{\mathbf{G}}$ 和 $\widehat{\mathbf{Q}}$，存在正常数 N、 c_1、 c_2 和 c_3，使得当$n \ge N$时，对于任意 $k \in [K]$,

(1). 以大于 $1-c_1\,n^{-3}$ 的概率，有 $\mathbb{E}[\|X-\widehat{G}_k(\widehat{Q}_k(X))\|_2\cdot\rho_k(X)] \le c_2\,n^{-\frac{\beta}{d}}\vee$对数$\sqrt{n}^{n}$; *(2).* 如果$-\tilde{\alpha} \ge 0$, 那么以大于 $1 - c_1\,n^{-3}$ 的概率，密度 $\nu^*_{k,\widehat{Q}_k}$ 关于 $(\widehat{Q}_k)_{\#}\binom{\mu^*\cdot\rho_k}{p_k}$ 存在且属于 $C^{\tilde{\alpha}}_{c_3}(\mathbb{R}^d)$。

根据引理 3 和引理 4 的第二个结论，存在常数 c， c_1 ，使得以大于 $1 - c\,n^{-2}$ 的概率有

$$\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(z)\nu^*_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z\Big) \le c_1\left(n^{-\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} + \frac{\log n}{\sqrt{n}}\right). \tag{12}$$

So c结合公式 (10)、(11) 和 (12)，以大于 $1-\frac{1}{n}$ 的概率有

$$
\begin{aligned}
W_1(\widehat{\mu},\mu^*) &= \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)} \Big(\sum_{k=1}^K \int p_k \cdot f(X)\,\mathrm{d}\Big(\frac{\mu^*\cdot\rho_k}{p_k}\Big) - \sum_{k=1}^K \int \widehat{p}_k \cdot f(X)\,\mathrm{d}(\widehat{G}_k)_{\#}\widehat{\nu}_k \Big) \\
&\overset{(i)}{\le} C\sqrt{\frac{\log n}{n}} + \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)} \Big(\sum_{k=1}^K \int \widehat{p}_k \cdot f(X)\,\mathrm{d}\Big(\frac{\mu^*\cdot\rho_k}{p_k}\Big) - \sum_{k=1}^K \int \widehat{p}_k \cdot f(X)\,\mathrm{d}(\widehat{G}_k)_{\#}\widehat{\nu}_k \Big) \\
&\le C\sqrt{\frac{\log n}{n}} + \sum_{k=1}^K \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)} \Big(\int \widehat{p}_k \cdot f(X)\,\mathrm{d}\Big(\frac{\mu^*\cdot\rho_k}{p_k}\Big) - \int \widehat{p}_k \cdot f(x)\,\mathrm{d}(\widehat{G}_k)_{\#}\widehat{\nu}_k \Big) \\
&\le C\sqrt{\frac{\log n}{n}} + \sum_{k=1}^K \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)} \Big(\int \frac{\widehat{p}_k}{p_k}\cdot f(X)\rho_k(X)\mathrm{d}\mu^* - \int \frac{\widehat{p}_k}{p_k}\cdot f(\widehat{G}_k(\widehat{Q}_k(X)))\rho_k(X)\mathrm{d}\mu^* \\
&\qquad + \int \frac{\widehat{p}_k}{p_k}\cdot f(\widehat{G}_k(\widehat{Q}_k(X)))\rho_k(X)\mathrm{d}\mu^* - \int \widehat{p}_k \cdot f(x)\,\mathrm{d}(\widehat{G}_k)_{\#}\widehat{\nu}_k \Big) \\
&\le C\sqrt{\frac{\log n}{n}} + 2\sum_{k=1}^K \mathbb{E}_{\mu^*}\Big[\|X-\widehat{G}_k(\widehat{Q}_k(X))\|\cdot\rho_k(X)\Big] \\
&\qquad + \sum_{k=1}^K \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)} \Big(\int f(\widehat{G}_k(z))\nu^*_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(\widehat{G}_k(z))\,\widehat{\nu}_k(z)\,\mathrm{d}z \Big) \\
&\le C_1\, n^{-\frac{\beta}{d}} \vee \frac{\log n}{\sqrt{n}} + \sum_{k=1}^K \Big[\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)} \Big(\int f(\widehat{G}_k(z))\nu^*_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(\widehat{G}_k(z))\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z \Big) \\
&\qquad + \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)} \Big(\int f(\widehat{G}_k(z))\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(\widehat{G}_k(z))\,\widehat{\nu}_k(z)\,\mathrm{d}z \Big) \Big] \\
&\overset{(ii)}{\le} C_1\, n^{-\frac{\beta}{d}} \vee \frac{\log n}{\sqrt{n}} + C_2 \sum_{k=1}^K \Big[\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)} \Big(\int f(z)\nu^*_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z \Big) \\
&\qquad + \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)} \Big(\int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(z)\,\widehat{\nu}_k(z)\,\mathrm{d}z \Big) \Big] \\
&\le C_2\, n^{-\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} \vee \frac{\log n}{\sqrt{n}},
\end{aligned}
\tag{13}
$$

其中，(*i*) 使用伯恩斯坦不等式得出 $\big|\widehat{p}_k - p_k\big| \le C\sqrt{\log n / n}$ 以至少$1-n^{-3}$的概率成立，而 (*ii*) 使用了以下事实：$\beta \ge 1$。

接下来考虑情形 $\alpha = 0$。定义 $\widetilde{\nu}_{k,Q} = \frac{1}{\widehat{p}_k n}\sum_{i=1}^n \delta_{Q(X_i)}\rho_k(X_i)$。对于任意 $f \in \mathrm{Lip}_1(\mathbb{R}^d)$，有

$$
\int f(z)\,\mathrm{d}\widetilde{\nu}_{k,Q} = \frac{1}{\widehat{p}_k n}\sum_{i=1}^n f(Q(X_i))\rho_k(X_i).
$$

接下来给出以下引理。

引理 5. 存在一个常数 c，使得以下结论以大于 $1-\frac{1}{n^2}$ 的概率成立：

$$
\sup_{Q\in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d)} \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)} \Big(\frac{1}{p_k}\int f(Q(x))\rho_k(x)\,\mathrm{d}\mu^* - \int f(z)\widetilde{\nu}_{k,Q}(z)\,\mathrm{d}z \Big) \le c\Big(\frac{\log n}{\sqrt{n}} + n^{-\frac{1}{d}} \Big).
$$

然后根据引理 5 和公式 (9)，选择 $\lambda_k = \lambda = \big(n^{-\frac{1}{d}} + \sqrt{\frac{\log n}{n}}\big)^{-1}\cdot n^{-\frac{2\beta}{d}-1}$，我们可得陈述 (10) 成立，且

$$
\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)} \Big(\int f(z)\widehat{\nu}_k(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z \Big) \le C\Big(n^{-\frac{1}{d}} + \frac{\log n}{\sqrt{n}} \Big).
$$

然后根据引理 4，我们有

$$\mathbb{E}[\|X-\widehat{G}_k(\widehat{Q}_k(X))\|_2\rho_k(X)] \le c_2 n^{-\frac{\beta}{d}} \vee \frac{\log n}{\sqrt{n}}.$$

因此，结合引理 5 和公式 (13)，我们有

$$\begin{aligned}
W_1(\widehat{\mu},\mu^*) &\le C\sqrt{\frac{\log n}{n}} + 2\sum_{k=1}^{K}\mathbb{E}_{\mu^*}\Big[\|X-\widehat{G}_k(\widehat{Q}_k(X))\|\cdot\rho_k(X)\Big] \\
&\qquad + \sum_{k=1}^{K}\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)}\Big(\int f(\widehat{G}_k(z))\nu^*_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(\widehat{G}_k(z))\,\widehat{\nu}_k(z)\,\mathrm{d}z\Big) \\
&\le C_1\, n^{-\frac{\beta}{d}} \vee \frac{\log n}{\sqrt{n}} + C_2\sum_{k=1}^{K}\Bigg[\sup_{f\in \mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(z)\nu^*_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z\Big) \\
&\qquad + \sup_{f\in \mathrm{Lip}_1(\mathbb{R}^D)}\Big(\int f(z)\widetilde{\nu}_{k,\widehat{Q}_k}(z)\,\mathrm{d}z - \int f(z)\,\widehat{\nu}_k(z)\,\mathrm{d}z\Big)\Bigg] \\
&\le C_2\, n^{-\frac{1}{d}} \vee \frac{\log n}{\sqrt{n}}.
\end{aligned}$$

C.1 引理 3 的证明：核密度估计量

任取 $k\in[K]$。由于 $\widetilde{\mathcal{Q}}_k \subseteq C_L^{\beta}(\mathbb{R}^D;\mathbb{R}^d)$，对于任意 $Q\in\widetilde{\mathcal{Q}}_k$，$\nu^*_{k,Q}\in C_L^{\tilde{\alpha}}(\mathbb{R}^d)$ 和 $\mathrm{supp}(\nu^*_{k,Q})\in[-L,L]^d$，都有，其中 $\nu^*_{k,Q}$ 是 $\frac{\mu^*\cdot\rho_k}{p_k}$ 经映射 Q 得到的推前测度的密度。回忆

$$\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y) = \frac{1}{nh^d}\sum_{i=1}^{n}\widetilde{k}\Big(\frac{y-Q(X_i)}{h}\Big)\rho_k(X_i).$$

由于 $\nu^*_{k,Q}$ 和 $\widetilde{\nu}_{k,Q}$ 都是紧支撑的，因此存在常数 C，使得对任意 $Q\in\mathcal{Q}_k$，

$$\sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(y)\nu^*_{k,Q}(y)\,\mathrm{d}y - \int f(y)\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y\Big) \le C\sup_{f\in\mathbf{C}_1^1(\mathbb{R}^d)}\Big(\int f(y)\nu^*_{k,Q}(y)\,\mathrm{d}y - \int f(y)\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y\Big),$$

其中 $\mathbf{C}_1^1(\mathbb{R}^d)=\big\{f\colon\ \mathbb{R}^d\to\mathbb{R}\,\big|\,\sup_{z\in\mathbb{R}^d}\sum_{j\in\mathbb{N}_0^d,\ |j|\le 1}|f^{(j)}(z)|\le 1\big\}$。接下来考虑 $f\in\mathbf{C}_1^1(\mathbb{R}^d)$，可得

$$\begin{aligned}
&\int f(y)\nu^*_{k,Q}(y)\,\mathrm{d}y - \int f(y)\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y \\
&= \frac{1}{\widehat{p}_k}\Big(\int p_k\cdot f(y)\nu^*_{k,Q}(y)\,\mathrm{d}y - \int \widehat{p}_k\cdot f(y)\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y\Big) + \int f(y)\nu^*_{k,Q}(y)\,\mathrm{d}y\cdot\big(1-\frac{p_k}{\widehat{p}_k}\big) \\
&\le \underbrace{\frac{1}{\widehat{p}_k}\Big|\int f(y)\cdot p_k\cdot\nu^*_{k,Q}(y)\,\mathrm{d}y - \int f(y)\cdot\mathbb{E}_{X^{(n)}}\big[\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y)\big]\,\mathrm{d}y\Big|}_{(A)} \\
&\quad + \underbrace{\frac{1}{\widehat{p}_k}\Big|\int f(y)\cdot\mathbb{E}_{X^{(n)}}\big[\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y)\big]\,\mathrm{d}y - \int f(y)\cdot\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y\Big|}_{(B)} + \underbrace{\Big|1-\frac{p_k}{\widehat{p}_k}\Big|}_{(C)}.
\end{aligned} \tag{14}$$

首先，对于项（C），根据伯恩斯坦不等式，以至少 $1-n^{-3}$ 的概率，有 $|p_k-\widehat{p}_k|\le C\sqrt{\text{对数}_n\, n}$；随后根据 $p_k>0$，当 n 足够大时，有 $\big|1-\frac{p_k}{\widehat{p}_k}\big|\le C\sqrt{\text{对数}_n\, n}$。为了界定项（$A$），我们采用与 Divol（2022）中引理 4.3 的证明类似的策略。回顾 $\nu^*_{k,Q}=Q_{\#}[\frac{\mu^*\cdot\rho_k}{p_k}]$，我们可以

写成

$$\begin{aligned}\int f(y)\cdot\mathbb{E}_{X^{(n)}}\left[\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y)\right]\mathrm{d}y &= \int f(y)\cdot\frac{1}{h^d}\cdot\mathbb{E}_{\mu^*}\left[\widetilde{k}(\frac{y-Q(X)}{h})\cdot\rho_k(X)\right]\mathrm{d}y\\ &= \int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot p_k\cdot\nu^*_{k,Q}(z)\,\mathrm{d}z\,\mathrm{d}y\end{aligned}$$

记 $\upsilon(\cdot)=p_k\cdot\nu^*_{k,\ Q}(\cdot)$，我们可以得到

$$\begin{aligned}&\Big|\int f(y)\cdot p_k\cdot\nu^*_{k,Q}(y)\,\mathrm{d}y-\int f(y)\cdot\mathbb{E}_{X^{(n)}}\left[\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y)\right]\mathrm{d}y\Big|\\ &=\Big|\int f(y)\cdot\upsilon(y)\,\mathrm{d}y-\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot\upsilon(z)\,\mathrm{d}z\,\mathrm{d}y\Big|\\ &=\Big|\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot(\upsilon(z)-\upsilon(y))\,\mathrm{d}z\,\mathrm{d}y\Big|.\end{aligned}$$

当 $\lfloor\tilde{\alpha}\rfloor$ 为偶数时，记为 $s=\lfloor\tilde{\alpha}\rfloor$；当 $\lfloor\tilde{\alpha}\rfloor$ 为奇数时，记为 $s=\lfloor\tilde{\alpha}\rfloor-1$。然后利用泰勒定理，我们可以分解

$$\upsilon(z)-\upsilon(y)=\sum_{\substack{j\in\mathbb{N}_0^d\\1\le|j|<s}}\frac{\upsilon^{(j)}(y)}{j!}\cdot(z-y)^j+\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\upsilon^{(j)}(y+t(z-y))\,\mathrm{d}t\cdot(z-y)^j.$$

利用以下事实：对于任意满足 $1\le|j|\le\lfloor\tilde{\alpha}\rfloor$ 的 $j\in\mathbb{N}_0^d$，$\int_{\mathbb{R}^d}\widetilde{k}(z)\cdot z^j\,\mathrm{d}z=0$ 且 $\widetilde{k}(\cdot)=\widetilde{k}(-\cdot)$，我们可以得到

$$\begin{aligned}&\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|<s}}\frac{\upsilon^{(j)}(y)}{j!}\cdot(z-y)^j\,\mathrm{d}z\,\mathrm{d}y\\ &=\int\int f(y)\cdot\widetilde{k}(z)\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|<s}}\frac{\upsilon^{(j)}(y)}{j!}\cdot z^j\cdot h^{|j|}\,\mathrm{d}z\,\mathrm{d}y=0,\end{aligned}$$

and

$$\begin{aligned}&\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\upsilon^{(j)}(y)\cdot(z-y)^j\,\mathrm{d}t\,\mathrm{d}z\,\mathrm{d}y\\ &=\int\int f(y)\cdot\widetilde{k}(z)\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\upsilon^{(j)}(y)\cdot z^j\cdot h^s\,\mathrm{d}t\,\mathrm{d}z\,\mathrm{d}y=0.\end{aligned}$$

因此，我们有

$$\begin{aligned}&\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot(\upsilon(z)-\upsilon(y))\,\mathrm{d}z\,\mathrm{d}y\\ &=\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(y+t(z-y))-\upsilon^{(j)}(y)\big)\cdot(z-y)^j\,\mathrm{d}t\,\mathrm{d}z\,\mathrm{d}y\\ &\overset{(i)}{=}\int\int f(y)\cdot\frac{1}{t^dh^d}\cdot\widetilde{k}(\frac{x-y}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(x)-\upsilon^{(j)}(y)\big)\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

~ ~其中，(i) 使用变量替换 $x = y + t(z-y)$ 和 $k(\cdot) = k(-\cdot)$。然后交换变量 x和 y，利用 s 是偶数且 k 是偶函数这一事实，我们有

$$\begin{aligned}
&\int\int f(y)\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{x-y}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(x)-\upsilon^{(j)}(y)\big)\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\\
&=\int\int f(x)\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{y-x}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(y)-\upsilon^{(j)}(x)\big)\cdot\big(\frac{y-x}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\\
&=-\int\int f(x)\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{x-y}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(x)-\upsilon^{(j)}(y)\big)\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

因此，当 $\lfloor\tilde{\alpha}\rfloor$ 为偶数时，我们可以得到

$$\begin{aligned}
&\Big|\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot(\upsilon(z)-\upsilon(y))\,\mathrm{d}z\,\mathrm{d}y\Big|\\
&=\Big|\frac{1}{2}\int\int(f(y)-f(x))\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{x-y}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(x)-\upsilon^{(j)}(y)\big)\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\Big|\\
&=\Big|\frac{1}{2}\int\int(f(y)-f(y+wth))\cdot\widetilde{k}(w)\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(y+wth)-\upsilon^{(j)}(y)\big)\cdot\big(wh\big)^j\,\mathrm{d}t\,\mathrm{d}w\,\mathrm{d}y\Big|\\
&\overset{(ii)}{\lesssim} h^{1+\tilde{\alpha}}\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_{[-L-h,L+h]^d}\int_{[-1,1]^d}\|w\|^{1+\tilde{\alpha}}\cdot\widetilde{k}(w)\cdot\int_0^1(1-t)^{s-1}\,\mathrm{d}t\,\mathrm{d}w\,\mathrm{d}y\\
&\lesssim n^{-\frac{1+\tilde{\alpha}}{2\tilde{\alpha}+d}},
\end{aligned}$$

其中，(ii) 使用 $\upsilon(\cdot)=p_k\cdot\nu_{k,\,Q}^{*}(\cdot)$ 的支撑集包含于 $[-L,L]^d$，而 $\widetilde{k}(\cdot)$ 的支撑集包含于$[-1,1]^d$。另一方面，当 $\lfloor\tilde{\alpha}\rfloor$ 为奇数时，回想一下 $s=\lfloor\tilde{\alpha}\rfloor-1$ 为偶数，我们有

$$\begin{aligned}
&\Big|\int\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}(\frac{y-z}{h})\cdot(\upsilon(z)-\upsilon(y))\,\mathrm{d}z\,\mathrm{d}y\Big|\\
&=\Big|\frac{1}{2}\int\int(f(x)-f(y))\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{x-y}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\big(\upsilon^{(j)}(x)-\upsilon^{(j)}(y)\big)\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\Big|\\
&\le\Big|\frac{1}{2}\int\int(f(x)-f(y))\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{x-y}{th})\\
&\qquad\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\Big(\upsilon^{(j)}(x)-\upsilon^{(j)}(y)-\sum_{\substack{q\in\mathbb{N}_0^d\\|q|=1}}\upsilon^{(j+q)}(y)\cdot(x-y)^q\Big)\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\Big|\\
&+\Big|\frac{1}{2}\int\int(f(x)-f(y))\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{x-y}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\sum_{\substack{q\in\mathbb{N}_0^d\\|q|=1}}\upsilon^{(j+q)}(y)\cdot(x-y)^q\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\Big|\\
&\le C\,h^{1+\tilde{\alpha}}\\
&\quad+\underbrace{\Big|\frac{1}{2}\int\int(f(x)-f(y))\cdot\frac{1}{t^d h^d}\cdot\widetilde{k}(\frac{x-y}{th})\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}\int_0^1(1-t)^{s-1}\sum_{\substack{q\in\mathbb{N}_0^d\\|q|=1}}\upsilon^{(j+q)}(y)\cdot(x-y)^q\cdot\big(\frac{x-y}{t}\big)^j\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\Big|}_{(D)}.
\end{aligned}$$

那么对于项 (D)，利用泰勒定理和 $f \in \mathbf{C}_1^1(\mathbb{R}^d)$，我们可以写成

$$f(x) - f(y) = \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=1}} \int_0^1 f^{(j)}(y + t(x-y))\,\mathrm{d}t \cdot (x-y)^j. \tag{15}$$

利用以下事实：对于任意满足 $1 \le |j| \le \lfloor \tilde{\alpha} \rfloor + 1$ 的 $j \in \mathbb{N}_0^d$，$\int_{\mathbb{R}^d} \widetilde{k}(z) \cdot z^j \,\mathrm{d}z = 0$，我们可以得到
So u

$$\begin{aligned}
&\int\int \sum_{\substack{l \in \mathbb{N}_0^d \\ |l|=1}} f^{(l)}(y) \cdot (x-y)^l \cdot \frac{1}{t^d h^d} \cdot \widetilde{k}\big(\frac{x-y}{th}\big) \cdot \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=s}} \frac{s}{j!} \int_0^1 (1-t)^{s-1} \sum_{\substack{q \in \mathbb{N}_0^d \\ |q|=1}} \upsilon^{(j+q)}(y) \cdot (x-y)^q \cdot \big(\frac{x-y}{t}\big)^j \,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y \\
&= \sum_{\substack{l \in \mathbb{N}_0^d \\ |l|=1}} \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=s}} \sum_{\substack{q \in \mathbb{N}_0^d \\ |q|=1}} \int_0^1 \frac{1}{t^d h^d} \cdot \frac{s}{j!} \cdot \int\int f^{(l)}(y) \cdot (x-y)^l \cdot \widetilde{k}\big(\frac{x-y}{th}\big)(1-t)^{s-1} \upsilon^{(j+q)}(y) \cdot (x-y)^q \cdot \big(\frac{x-y}{t}\big)^j \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t \\
&= 0.
\end{aligned}$$

因此，

$$\begin{aligned}
(D) = \Big| &\frac{1}{2} \int\int \sum_{\substack{l \in \mathbb{N}_0^d \\ |l|=1}} \int_0^1 \big(f^{(l)}(y + \omega(x-y)) - f^{(l)}(y)\big)\,\mathrm{d}\omega \cdot (x-y)^l \\
&\cdot \frac{1}{t^d h^d} \cdot \widetilde{k}\big(\frac{x-y}{th}\big) \cdot \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=s}} \frac{s}{j!} \int_0^1 (1-t)^{s-1} \sum_{\substack{q \in \mathbb{N}_0^d \\ |q|=1}} \upsilon^{(j+q)}(y) \cdot (x-y)^q \cdot \big(\frac{x-y}{t}\big)^j \,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y \Big| \\
\overset{(iii)}{=} \Big| &\frac{1}{2} \int\int \int_0^1 \int_0^1 \sum_{\substack{l \in \mathbb{N}_0^d \\ |l|=1}} \big(f^{(l)}(\theta) - f^{(l)}(y)\big) \cdot \big(\frac{\theta - y}{\omega}\big)^l \\
&\cdot \frac{1}{\omega^d t^d h^d} \cdot \widetilde{k}\big(\frac{\theta-y}{\omega th}\big) \cdot \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=s}} \frac{s}{j!}(1-t)^{s-1} \sum_{\substack{q \in \mathbb{N}_0^d \\ |q|=1}} \upsilon^{(j+q)}(y) \cdot \big(\frac{\theta-y}{\omega}\big)^q \cdot \big(\frac{\theta-y}{\omega t}\big)^j \,\mathrm{d}\omega\,\mathrm{d}t\,\mathrm{d}\theta\,\mathrm{d}y \Big|,
\end{aligned}$$

其中，(iii) 使用变量替换 $y + \omega(x-y) = \theta$。通过交换变量 θ 和 y，我们得到

$$\begin{aligned}
&\int\int \int_0^1 \int_0^1 \sum_{\substack{l \in \mathbb{N}_0^d \\ |l|=1}} \big(f^{(l)}(\theta) - f^{(l)}(y)\big) \cdot \big(\frac{\theta-y}{\omega}\big)^l \\
&\quad \cdot \frac{1}{\omega^d t^d h^d} \cdot \widetilde{k}\big(\frac{\theta-y}{\omega th}\big) \cdot \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=s}} \frac{s}{j!}(1-t)^{s-1} \sum_{\substack{q \in \mathbb{N}_0^d \\ |q|=1}} \upsilon^{(j+q)}(y) \cdot \big(\frac{\theta-y}{\omega}\big)^q \cdot \big(\frac{\theta-y}{\omega t}\big)^j \,\mathrm{d}\omega\,\mathrm{d}t\,\mathrm{d}\theta\,\mathrm{d}y \\
&= \int\int \int_0^1 \int_0^1 \sum_{\substack{l \in \mathbb{N}_0^d \\ |l|=1}} \big(f^{(l)}(y) - f^{(l)}(\theta)\big) \cdot \big(\frac{y-\theta}{\omega}\big)^l \\
&\quad \cdot \frac{1}{\omega^d t^d h^d} \cdot \widetilde{k}\big(\frac{y-\theta}{\omega th}\big) \cdot \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=s}} \frac{s}{j!}(1-t)^{s-1} \sum_{\substack{q \in \mathbb{N}_0^d \\ |q|=1}} \upsilon^{(j+q)}(\theta) \cdot \big(\frac{y-\theta}{\omega}\big)^q \cdot \big(\frac{y-\theta}{\omega t}\big)^j \,\mathrm{d}\omega\,\mathrm{d}t\,\mathrm{d}\theta\,\mathrm{d}y \\
&= -\int\int \int_0^1 \int_0^1 \sum_{\substack{l \in \mathbb{N}_0^d \\ |l|=1}} \big(f^{(l)}(\theta) - f^{(l)}(y)\big) \cdot \big(\frac{\theta-y}{\omega}\big)^l \\
&\quad \cdot \frac{1}{\omega^d t^d h^d} \cdot \widetilde{k}\big(\frac{\theta-y}{\omega th}\big) \cdot \sum_{\substack{j \in \mathbb{N}_0^d \\ |j|=s}} \frac{s}{j!}(1-t)^{s-1} \sum_{\substack{q \in \mathbb{N}_0^d \\ |q|=1}} \upsilon^{(j+q)}(\theta) \cdot \big(\frac{\theta-y}{\omega}\big)^q \cdot \big(\frac{\theta-y}{\omega t}\big)^j \,\mathrm{d}\omega\,\mathrm{d}t\,\mathrm{d}\theta\,\mathrm{d}y,
\end{aligned}$$

由此可得

$$\begin{aligned}
&\Big|\frac{1}{2}\int\int\int_0^1\int_0^1 \sum_{\substack{l\in\mathbb{N}_0^d\\|l|=1}} \big(f^{(l)}(\theta)-f^{(l)}(y)\big)\cdot\big(\frac{\theta-y}{\omega}\big)^l \\
&\quad\cdot\frac{1}{\omega^d t^d h^d}\cdot\widetilde{k}\big(\frac{\theta-y}{\omega th}\big)\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}(1-t)^{s-1}\sum_{\substack{q\in\mathbb{N}_0^d\\|q|=1}}\upsilon^{(j+q)}(y)\cdot\big(\frac{\theta-y}{\omega}\big)^q\cdot\big(\frac{\theta-y}{\omega t}\big)^j\,\mathrm{d}\omega\,\mathrm{d}t\,\mathrm{d}\theta\,\mathrm{d}y\Big|, \\
=&\Big|\frac{1}{4}\int\int\int_0^1\int_0^1 \sum_{\substack{l\in\mathbb{N}_0^d\\|l|=1}} \big(f^{(l)}(\theta)-f^{(l)}(y)\big)\cdot\big(\frac{\theta-y}{\omega}\big)^l \\
&\quad\cdot\frac{1}{\omega^d t^d h^d}\cdot\widetilde{k}\big(\frac{\theta-y}{\omega th}\big)\cdot\sum_{\substack{j\in\mathbb{N}_0^d\\|j|=s}}\frac{s}{j!}(1-t)^{s-1}\sum_{\substack{q\in\mathbb{N}_0^d\\|q|=1}}\big(\upsilon^{(j+q)}(y)-\upsilon^{(j+q)}(\theta)\big)\cdot\big(\frac{\theta-y}{\omega}\big)^q\cdot\big(\frac{\theta-y}{\omega t}\big)^j\,\mathrm{d}\omega\,\mathrm{d}t\,\mathrm{d}\theta\,\mathrm{d}y\Big| \\
\lesssim& h^{1+\tilde{\alpha}},
\end{aligned}$$

其中 最后一个不等式利用了这样一个事实： $s=\lfloor\tilde{\alpha}\rfloor-1$ 和 υ 是 $\tilde{\alpha}$ 光滑的。于是我们可以得到 t

$$(A)\lesssim n^{-\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}}.$$

剩下的任务是对项（B）进行界定，它等于

$$\begin{aligned}
(B)&=\Big|\int f(y)\cdot\mathbb{E}_{X^{(n)}}\big[\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y)\big]\,\mathrm{d}y-\int f(y)\cdot\widehat{p}_k\cdot\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y\Big| \\
&=\Big|\frac{1}{n}\sum_{i=1}^n\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}\big(\frac{y-Q(X_i)}{h}\big)\cdot\rho_k(X_i)\,\mathrm{d}y-\mathbb{E}_{\mu^*}\Big[\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}\big(\frac{y-Q(X)}{h}\big)\cdot\rho_k(X)\,\mathrm{d}y\Big]\Big|.
\end{aligned}$$

通过标准的对称化，我们可以得到

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\substack{f\in\mathbf{C}_1^1(\mathbb{R}^d)\\Q\in\widetilde{\mathcal{Q}}_k}}\Big|\frac{1}{n}\sum_{i=1}^n\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}\big(\frac{y-Q(X_i)}{h}\big)\cdot\rho_k(X_i)\,\mathrm{d}y-\mathbb{E}\Big[\int f(y)\cdot\frac{1}{h^d}\cdot k\big(\frac{y-Q(X)}{h}\big)\cdot\rho_k(X)\,\mathrm{d}y\Big]\Big|\Big] \\
&\le 2\mathbb{E}\Big[\sup_{\substack{f\in\mathbf{C}_1^1(\mathbb{R}^d)\\Q\in\widetilde{\mathcal{Q}}_k}}\Big|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\int f(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}\big(\frac{y-Q(X_i)}{h}\big)\cdot\rho_k(X_i)\,\mathrm{d}y\Big|\Big],
\end{aligned}$$

其中 $\{\varepsilon_i\}_{i=1}^n$是来自拉德马赫分布的 n独立同分布副本，即 $P(\varepsilon_i=1)=P(\varepsilon_i=-1)=0.5$。定义函数集

$$\mathcal{F}=\Big\{f:\mathcal{M}\to\mathbb{R}\,|\,f(x)=\int g(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}\big(\frac{y-Q(x)}{h}\big)\cdot\rho_k(x)\,\mathrm{d}y;\ g\in\mathbf{C}_1^1(\mathbb{R}^d);\ Q\in\widetilde{\mathcal{Q}}_k\Big\}.$$

由于存在常数 L，使得 $\mathrm{supp}(Q)=[-L,L]^d$，我们可以先考虑函数集

$$\mathcal{F}_1=\Big\{f:[-L,L]^d\to\mathbb{R}\,|\,f(z)=\int g(y)\cdot\frac{1}{h^d}\cdot\widetilde{k}\big(\frac{y-z}{h}\big)\,\mathrm{d}y;\ g\in\mathbf{C}_1^1(\mathbb{R}^d)\Big\}.$$

对于任意的 $f\in\mathcal{F}_1$和 $j\in\mathbb{N}_0^d$，若满足$1\le|j|\le\lfloor\frac{d}{2}\rfloor$，则有

$$\begin{aligned}
f^{(j)}(z)&=\int g(y)\frac{1}{h^{d+|j|}}\widetilde{k}^{(j)}\big(\frac{z-y}{h}\big)\,\mathrm{d}y \\
&=\frac{1}{h^{|j|}}\int g(z-ht)\widetilde{k}^{(j)}(t)\,\mathrm{d}t.
\end{aligned}$$

由于 $|j| \geq 1$，存在满足 $|j_1| = 1$ 的 $j_1 \in \mathbb{N}_0^d$，使得 $j - j_1$ 中的每个元素均非负。因此有

$$|f^{(j)}(z)| = \Big|\frac{1}{h^{|j|-1}} \int g^{(j_1)}(z - ht)\widetilde{k}^{(j-j_1)}(t)\,\mathrm{d}t\Big| \lesssim \frac{1}{h^{|j|-1}},$$

其中，最后一个不等式使用了 $g \in \mathbf{C}_1^1(\mathbb{R}^d)$。此外，当 $d \geq 2$ 时，对于任意满足 $|j| = \lfloor \frac{d}{2} \rfloor$ 且 $z_1, z_2 \in \mathbb{R}^d$ 的 $j \in \mathbb{N}_0^d$，有

$$\begin{aligned}|f^{(j)}(z_1) - f^{(j)}(z_2)| &= \Big|\frac{1}{h^{\lfloor \frac{d}{2} \rfloor - 1}} \int g^{(j_1)}(z_1 - ht)\widetilde{k}^{(j-j_1)}(t)\,\mathrm{d}t - \frac{1}{h^{\lfloor \frac{d}{2} \rfloor - 1}} \int g^{(j_1)}(z_2 - ht)\widetilde{k}^{(j-j_1)}(t)\,\mathrm{d}t\Big| \\ &= \Big|\frac{1}{h^{\lfloor \frac{d}{2} \rfloor - 1}} \int \frac{1}{h^d} g^{(j_1)}(y)\big(\widetilde{k}^{(j-j_1)}(\frac{z_1 - y}{h}) - \widetilde{k}^{(j-j_1)}(\frac{z_2 - y}{h})\big)\,\mathrm{d}y\Big|\end{aligned}$$

若 $\|z_1 - z_2\| \leq h$，则

$$\begin{aligned}|f^{(j)}(z_1) - f^{(j)}(z_2)| &= \Big|\frac{1}{h^{\lfloor \frac{d}{2} \rfloor - 1}} \int \frac{1}{h^d} g^{(j_1)}(y)\big(\widetilde{k}^{(j-j_1)}(\frac{z_1 - y}{h}) - \widetilde{k}^{(j-j_1)}(\frac{z_2 - y}{h})\big)\,\mathrm{d}y\Big| \\ &\leq C\,\frac{1}{h^{\lfloor \frac{d}{2} \rfloor}}\|z_1 - z_2\| \leq C\,\frac{1}{h^{\frac{d}{2}-1}}\|z_1 - z_2\|^{\frac{d}{2} - \lfloor \frac{d}{2} \rfloor};\end{aligned}$$

若 $\|z_1 - z_2\| \geq h$，则

$$|f^{(j)}(z_1) - f^{(j)}(z_2)| \leq |f^{(j)}(z_1)| + |f^{(j)}(z_2)| \leq C\,\frac{1}{h^{\lfloor \frac{d}{2} \rfloor - 1}} \leq C\,\frac{1}{h^{\frac{d}{2}-1}}\|z_1 - z_2\|^{\frac{d}{2} - \lfloor \frac{d}{2} \rfloor}.$$

因此，当 $d \geq 2$ 时，存在一个常数 C，使得对于任意 $f \in \mathcal{F}_1$，都有

$$C\,h^{\frac{d}{2}-1} f \in C_1^{\frac{d}{2}}([-L, L]^d).$$

类似地，当 $d = 1$ 时，对于任意 $z_1, z_2 \in \mathbb{R}^d$，都有

$$|f(z_1) - f(z_2)| = \Big|\int (g(z_1 - ht) - g(z_2 - ht)) \cdot \widetilde{k}\big(t\big)\,\mathrm{d}y\Big| \lesssim \|z_1 - z_2\|.$$

因此，我们可以得出

$$\mathcal{F}_1 \in C^{\frac{d}{2}}_{L_1\,h^{-(\frac{d}{2}-1)_+}}([-L, L]^d),$$

对于某个常数 L_1，其中 $(a)_+ = \max(a,\ 0)$。因此，对于任意 $\epsilon > 0$，我们都可以找到一个集合 $N_\epsilon^f \subseteq \mathcal{F}_1$，使得 $\log |N_\epsilon^f| \lesssim \epsilon^{-(\frac{d}{2}-1)_+}$，并且对于任意 $f \in \mathcal{F}_1$，存在 $\widetilde{f} \in N_\epsilon^f$，使得

$$\sup_{y \in [-L, L]^d} |f(y) - \widetilde{f}(y)| \leq \epsilon.$$

接下来，为了推导 $\mathcal{F}$ 的 ϵ-覆盖数，我们引入以下引理。

引理6。（*Tang 与 Yang (2022)* 的引理 *12*）设 $\mathcal{X}_G = \big\{x \in \mathbb{R}^D :\ \ x = G(z)\ z \in \mathbb{B}_1^d\big\}$ 是由利普希茨连续映射 $G :\ \ \mathbb{R}^d \to \mathbb{R}^D$ 导出的 d 维子流形，则对于任意 $\widetilde{\gamma} > 0$，有

$$\log \mathbf{N}\big(C_1^{\widetilde{\gamma}}(\mathbb{R}^D),\, \|\cdot\|_{L^\infty(\mathcal{X}_G)},\, \epsilon\big) \leq C\,\epsilon^{-\frac{d}{\widetilde{\gamma}}}, \quad \forall \epsilon > 0,$$

其中，$\mathbf{N}(\mathcal{F},\, d,\, \epsilon)$ 表示函数空间 $\mathcal{F}$ 相对于伪度量 d 的 ϵ-覆盖数，而 $\|f\|_{L^\infty(\mathcal{X}_G)} = \sup_{x \in \mathcal{X}_G} |f(x)|$ 表示约束在集合 $\mathcal{X}_G$ 上的函数上确界范数。

那么 由于 $\Omega_k = Q_k^*(\mathcal{M} \cap S_k)$ 是紧支撑的，且 $\mathcal{M} \cap S_k = G_k^*(\Omega_k)$，对于任意 $\epsilon > 0$，我们可以

集合 N_ϵ^Q，使得 $\sqrt{\log |N_\epsilon^Q| \lesssim (\frac{1}{\epsilon})^{\frac{d}{2\beta}}}$，并且对于任意 $Q \in \widetilde{\mathcal{Q}}_k$，存在 $\widetilde{Q} \in N_\epsilon^Q$，使得

找到一个函数 at

$$\sup_{x \in \mathcal{M} \cap S_k} \|\widetilde{Q}(x) - Q(x)\| \leq \epsilon.$$

那么，对于任意的 $Q \in \widetilde{\mathcal{Q}}_k$ 和 $f \in \mathcal{F}_1$，存在 $\widetilde{Q} \in N_\epsilon^Q$、$\widetilde{f} \in N_\epsilon^f$ 以及常数 c，使得

$$\begin{aligned}
&\left| f(Q(x))\rho_k(x) - \widetilde{f}(\widetilde{Q}(x))\rho_k(x) \right| \\
&\leq \left| f(Q(x))\rho_k(x) - \widetilde{f}(Q(x))\rho_k(x) \right| + \left| \widetilde{f}(Q(x))\rho_k(x) - \widetilde{f}(\widetilde{Q}(x))\rho_k(x) \right| \\
&\leq c\epsilon.
\end{aligned}$$

因此我们可以得到

$$\log \mathbf{N}(\mathcal{F}, \epsilon, \|\cdot\|_\infty) \lesssim \left(\frac{1}{\epsilon}\right)^{\frac{d}{\beta}} + \left(\frac{h^{-(\frac{d}{2}-1)_+}}{\epsilon}\right)^2.$$

取 $\delta = \left(\frac{1}{n}\right)^{\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} \vee \left(\frac{1}{n}\right)^{\frac{\beta}{d}}$，我们可以得到

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \int_\delta^1 \left[\left(\frac{1}{\epsilon}\right)^{\frac{d}{2\beta}} + \frac{h^{1-\frac{d}{2}} \vee 1}{\epsilon} \right] \mathrm{d}\epsilon \\
&\lesssim \frac{\log n}{\sqrt{n}} + \left(\frac{1}{n}\right)^{\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} + \left(\frac{1}{n}\right)^{\frac{\beta}{d}}.
\end{aligned}$$

根据 Dudley 的熵积分界（例如，参见 Wainwright（2019年）的定理 5.22），有 at

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{f \in \mathbf{C}_1^1(\mathbb{R}^d) \\ Q \in \widetilde{\mathcal{Q}}_k}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int f(y) \cdot \frac{1}{h^d} \cdot \widetilde{k}\left(\frac{y - Q(X_i)}{h}\right) \cdot \rho_k(X_i)\, \mathrm{d}y \right| \right] \\
&\lesssim \frac{\log n}{\sqrt{n}} + \left(\frac{1}{n}\right)^{\frac{\tilde{\alpha}+1}{2\tilde{\alpha}+d}} + \left(\frac{1}{n}\right)^{\frac{\beta}{d}}.
\end{aligned}$$

接下来应用 Talagrand 集中不等式（例如，参见 Wainwright（2019）的定理3.27）以及 $\tilde{\alpha} + 1 \leq \beta$ 这一事实。

C.2 引理 3 的证明：小波估计量

任取一个 $k \in [K]$。由于 $\widetilde{\mathcal{Q}}_k \subseteq C_L^\beta(\mathbb{R}^D; \mathbb{R}^d)$，对于任意 $Q \in \widetilde{\mathcal{Q}}_k$，都有 $\mathrm{supp}(\nu_{k,Q}^*) \subseteq [-L, L]^d$，其中 $\nu_{k,Q}^*$ 是 $\frac{\mu^* \cdot \rho_k}{p_k}$ 经映射 Q 推前所得测度的密度。此外，根据支撑集包含于 $[-L, L]^d$ 和 $0 < p_k = \mathbb{E}_{\mu^*}[\rho_k] \leq 1$ 中的 $\nu_{k,Q}^* \in C_L^\alpha(\mathbb{R}^d)$，我们可以将 $p_k \cdot \nu_{k,Q}^*(y)$ 写为

$$p_k \cdot \nu_{k,Q}^*(y) = \sum_{m \in \mathbb{S}} a_m^Q \phi_m(y) + \sum_{l=1}^{2^d-1} \sum_{j=0}^{+\infty} \sum_{m \in \mathbb{S}_{lj}} \theta_{ljm}^Q \psi_{ljm}(y),$$

其中 $\{\phi_m, \psi_{ljm} : \ l = 1, \cdots, 2^d - 1, j \in \mathbb{N}, m \in \mathbb{Z}^d\}$ 是 Besov 空间上的正交小波基，其定义在$\mathbb{R}^d$ 上为 $\phi_m(y) = \phi(y - m)$ 和 $\psi_{ljm}(y) = 2^{\frac{jd}{2}} \psi_l(2^j y - m)$,一并且对于任意$1 \leq l \leq 2^d - 1$，都有 $\phi(\cdot)$ 和 $\psi_l(\cdot)$；这些函数均为紧支撑，且具有有界的 β 阶导数（Bouzebda & Didi, 2017）。那么，存在一个常数 C，使得 $|\theta_{ljm}^Q| \leq C(2^{-dj})^{\frac{\alpha}{d}+\frac{1}{2}}$且 $a_m^Q \leq C$。回顾

$$\widehat{p}_k \cdot \widetilde{\nu}_{k,Q}(y) = \sum_{m \in \mathbb{S}} \widetilde{a}_m^Q \phi_m(y) + \sum_{l=1}^{2^d-1} \sum_{j=0}^{J} \sum_{m \in \mathbb{S}_{lj}} \widetilde{\theta}_{ljm}^Q \psi_{ljm}(y),$$

with

$$\widetilde{a}_m^Q = \frac{1}{n}\sum_{i=1}^{n}\phi_m(Q(X_i))\rho_k(X_i);$$

$$\widetilde{\theta}_{ljm}^Q = \frac{1}{n}\sum_{i=1}^{n}\psi_{ljm}(Q(X_i))\rho_k(X_i).$$

我们有

$$\mathbb{E}[\widetilde{a}_m^Q] = p_k\cdot\int\phi_m(y)\nu_{k,Q}^*(y)dy = a_m^Q,\quad \mathbb{E}[\widetilde{\theta}_{ljm}^Q] = p_k\cdot\int\psi_{ljm}(y)\nu_{k,Q}^*(y)dy = \theta_{ljm}^Q.$$

此外，由于 $\phi(\cdot)$ 和 $\psi_l(\cdot)$ 均为紧支撑，我们可以得出，存在一个常数C，使得对于任意 $1\le l\le 2^d-1$ 和 $j\in\mathbb{N}$，都有 $|\mathbb{S}_{lj}|\le C2^{dj}$ 和 $|\mathbb{S}|\le C$。由于 $\nu_{k,\,Q}^*$ 和 $\widetilde{\nu}_{k,\,Q}$ 均为紧支撑。因此，存在一个常数 C，使得对于任意 $Q\in\mathcal{Q}_k$,

$$\sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)}\Big(\int f(y)\nu_{k,Q}^*(y)\,\mathrm{d}y-\int f(y)\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y\Big)\le C\sup_{f\in C_1^1(\mathbb{R}^d)}\Big(\int f(y)\nu_{k,Q}^*(y)\,\mathrm{d}y-\int f(y)\widetilde{\nu}_{k,Q}(y)\,\mathrm{d}y\Big).$$

接下来考虑 $f\in C_1^1(\mathbb{R}^d)$，同理，我们可以改写为

$$f(y)=\sum_{m\in\mathbb{Z}^d}b_m\phi_m(y)+\sum_{l=1}^{2^d-1}\sum_{j=0}^{+\infty}\sum_{m\in\mathbb{Z}^d}f_{ljm}\psi_{ljm}(y)$$

其中 $|f_{ljm}|\le C_1(2^{-dj})^{\frac{1}{d}+\frac{1}{2}}$和 $|b_m|\le C_1$。因此我们可以得到

$$\begin{aligned}
&\int f(y)\nu_{k,Q}^*(y)dy-\int f(y)\widetilde{\nu}_{k,Q}(y)dy\\
&=\frac{1}{\widehat{p}_k}\Big(\int f(y)p_k\cdot\nu_{k,Q}^*(y)dy-\int\widehat{p}_k\cdot f(y)\widetilde{\nu}_{k,Q}(y)dy\Big)+\int f(y)\nu_{k,Q}^*(y)dy\cdot\big(1-\frac{p_k}{\widehat{p}_k}\big)\\
&=\frac{1}{\widehat{p}_k}\int f(y)\left(\sum_{m\in\mathbb{S}}(\widetilde{a}_m^Q-\mathbb{E}\widetilde{a}_m^Q)\phi_m(y)+\sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}}(\widetilde{\theta}_{ljm}^Q-\mathbb{E}\widetilde{\theta}_{ljm}^Q)\psi_{ljm}(y)\right)dy\\
&+\frac{1}{\widehat{p}_k}\int f(y)\left(\sum_{l=1}^{2^d-1}\sum_{j=J}^{\infty}\sum_{m\in\mathbb{S}_{lj}}\theta_{ljm}^Q\psi_{ljm}(y)\right)dy+\int f(y)\nu_{k,Q}^*(y)dy\cdot\big(1-\frac{p_k}{\widehat{p}_k}\big)\\
&\le\frac{1}{\widehat{p}_k}\underbrace{\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}}f_{ljm}\psi_{ljm}\left(Q(X_i)\right)\rho_k(X_i)-\mathbb{E}\Big[\sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}}f_{ljm}\psi_{ljm}\left(Q(X)\right)\rho_k(X)\Big]\right|}_{(A)}\\
&+\frac{1}{\widehat{p}_k}\underbrace{\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{m\in\mathbb{S}}b_m\phi_m(Q(X_i))\rho_k(X_i)-\mathbb{E}\Big[\sum_{m\in\mathbb{S}}b_m\phi_m(Q(X))\rho_k(X)\Big]\right|}_{(B)}+\frac{1}{\widehat{p}_k}\underbrace{\sum_{l=1}^{2^d-1}\sum_{j=J}^{+\infty}\sum_{m\in\mathbb{S}_{lj}}f_{ljm}\theta_{ljm}^Q}_{(C)}+\underbrace{\big|1-\frac{p_k}{\widehat{p}_k}\big|}_{(D)}.
\end{aligned}\tag{16}$$

首先，对于项（D），根据伯恩斯坦不等式，以至少 $1-n^{-3}$ 的概率有 $|p_k-\widehat{p}_k|\le C\sqrt{\log_n n}$成立；随后根据$p_k>0$，当 n 足够大时，我们有 $\big|1-\frac{p_k}{\widehat{p}_k}\big|\le C\sqrt{\log_n n}$。此外，对于项（$C$），由于$|f_{ljm}|\lesssim(2^{-dj})^{\frac{1}{d}+\frac{1}{2}}$、$|\theta_{ljm}^Q|\lesssim(2^{-dj})^{\frac{\alpha}{d}+\frac{1}{2}}$以及 $2^{dJ}\asymp n^{\frac{d}{2\alpha+d}}$,，我们可以得到

$$\sum_{j=J+1}^{+\infty}\sum_{l=1}^{2^d-1}\sum_{m\in\mathbb{S}_{lj}}f_{ljm}\theta_{ljm}^Q\lesssim n^{-\frac{\alpha+1}{2\alpha+d}}.$$

然后，对于项（A），通过标准对称化，我们可以得到

$$\mathbb{E}\left[\sup_{\substack{f\in C_1^1(\mathbb{R}^d)\\ Q\in\widetilde{\mathcal{Q}}_k}}\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}} f_{ljm}\psi_{ljm}\left(Q(X_i)\right)\rho_k(X_i)-\mathbb{E}\left[\sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}} f_{ljm}\psi_{ljm}\left(Q(X)\right)\rho_k(X)\right]\right|\right]$$

$$\leq 2\mathbb{E}\left[\sup_{\substack{f\in C_1^1(\mathbb{R}^d)\\ Q\in\widetilde{\mathcal{Q}}_k}}\left|\frac{1}{n}\sum_{l=1}^{2^d-1}\sum_{i=1}^{n}\varepsilon_i\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}} f_{ljm}\psi_{ljm}\left(Q(X_i)\right)\rho_k(X_i)\right|\right],$$

其中 $\{\varepsilon_i\}_{i=1}^n$ 是来自拉德马赫分布的 n 独立同分布副本，即 $P(\varepsilon_i=1)=P(\varepsilon_i=-1)=0.5$。定义函数集

$$\mathcal{F}=\left\{f:\mathcal{M}\to\mathbb{R}:f(z)=\sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}} f_{ljm}\psi_{ljm}(Q(x));|f_{ljm}|\leq(2^{-dj})^{\frac{1}{d}+\frac{1}{2}};Q\in\widetilde{\mathcal{Q}}_k\right\}.$$

首先考虑函数集

$$\mathcal{F}_1=\left\{f:[-L_3,L_3]^d\to\mathbb{R},f(y)=\sum_{l=1}^{2^d-1}\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}} f_{ljm}\psi_{ljm}(y),|f_{ljm}|\leq(2^{-dj})^{\frac{1}{d}+\frac{1}{2}}\right\}.$$

若 $\frac{1}{d}\leq\frac{1}{2}$，则存在常数 c，使得对任意 $f\in\mathcal{F}_1$，均有

$$c(2^{dJ})^{\frac{1}{d}-\frac{1}{2}}f\in C_1^{\frac{d}{2}}([-L,L]^d).$$

因此，对任意 $\epsilon>0$，我们都可以找到集合 $N_\epsilon^f\subseteq\mathcal{F}_1$，使得 $\sqrt{\log|N_\epsilon^f|}\lesssim(2^{dJ})^{(\frac{1}{2}-\frac{1}{d})_+}\epsilon$，并且对任意 $f\in\mathcal{F}_1$，存在 $\widetilde{f}\in N_\epsilon^f$ 使得

$$\sup_{y\in[-L,L]^d}|f(y)-\widetilde{f}(y)|\leq\epsilon.$$

由于 $\Omega_k=Q_k^*(\mathcal{M}\cap S_k)$ 是紧支撑的，且 $\mathcal{M}\cap S_k=G_k^*(\Omega_k)$，由引理6可知，对任意 $\epsilon>0$，我们都可以找到函数集 N_ϵ^Q，使得 $\sqrt{\log|N_\epsilon^Q|}\lesssim(\frac{1}{\epsilon})^{\frac{d}{2\beta}}$，并且对任意 $Q\in\widetilde{\mathcal{Q}}_k$，存在 $\widetilde{Q}\in N_\epsilon^Q$ 使得

$$\sup_{x\in\mathcal{M}\cap S_k}\|\widetilde{Q}(x)-Q(x)\|\leq\epsilon.$$

那么，对于任意的 $Q\in\widetilde{\mathcal{Q}}_k$ 和 $f\in\mathcal{F}_1$，存在 $\widetilde{Q}\in N_\epsilon^Q$、$\widetilde{f}\in N_\epsilon^f$ 以及一个常数 c，使得

$$\begin{aligned}&\left|f(Q(x))\rho_k(x)-\widetilde{f}(\widetilde{Q}(x))\rho_k(x)\right|\\&\leq\left|f(Q(x))\rho_k(x)-\widetilde{f}(Q(x))\rho_k(x)\right|+\left|\widetilde{f}(Q(x))\rho_k(x)-\widetilde{f}(\widetilde{Q}(x))\rho_k(x)\right|\\&\leq c\epsilon.\end{aligned}$$

因此，我们可以得到

$$\log\mathbf{N}(\mathcal{F},\epsilon,\|\cdot\|_\infty)\lesssim\left(\frac{1}{\epsilon}\right)^{\frac{d}{\beta}}+\left(\frac{2^{dJ(\frac{1}{2}-\frac{1}{d})_+}}{\epsilon}\right)^2.$$

选取 $\delta=(\frac{1}{n})^{\frac{\alpha+1}{2\alpha+d}}\vee(\frac{1}{n})^{\frac{\beta}{d}}$，我们可以得到

$$\begin{aligned}&\frac{1}{\sqrt{n}}\int_\delta^1\left[\left(\frac{\log n}{\epsilon}\right)^{\frac{d}{2\beta}}+\frac{(2^{dJ(\frac{1}{2}-\frac{1}{d})}\log n)\vee 1}{\epsilon}\right]\mathrm{d}\epsilon\\&\lesssim\frac{\log n}{\sqrt{n}}+(\frac{1}{n})^{\frac{\alpha+1}{2\alpha+d}}+(\frac{1}{n})^{\frac{\beta}{d}}.\end{aligned}$$

根据 Dudley 熵积分界，有

$$\mathbb{E}\sup_{\substack{f\in C_1^1(\mathbb{R}^d)\\ Q\in\widetilde{\mathcal{Q}}_k}}\left|\frac{1}{n}\sum_{l=1}^{2^d-1}\sum_{i=1}^{n}\varepsilon_i\sum_{j=0}^{J}\sum_{m\in\mathbb{S}_{lj}}f_{ljm}\psi_{ljm}\left(Q(X_i)\right)\rho_k(X_i)\right|$$
$$\lesssim\frac{\log n}{\sqrt{n}}+\left(\frac{1}{n}\right)^{\frac{\alpha+1}{2\alpha+d}}+\left(\frac{1}{n}\right)^{\frac{\beta}{d}}.$$

同理，我们可以得到

$$\mathbb{E}\sup_{\substack{f\in C_1^1(\mathbb{R}^d)\\ Q\in\widetilde{\mathcal{Q}}_k}}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\sum_{m\in\mathbb{S}}b_m\phi_m\left(Q(X_i)\right)\rho_k(X_i)\right|$$
$$\lesssim n^{-\frac{\beta}{d}}+\frac{\log n}{\sqrt{n}}.$$

接下来利用 Talagrand 集中不等式（例如，参见 Wainwright（2019）的定理3.27）以及 $\alpha+1\le\beta$ 这一事实，即可得出该结论。

C.3 引理 4 的证明

在以下分析中，我们固定任意一个 $k\in[K]$。由于

$$\frac{1}{n}\sum_{i=1}^{n}\|X_i-\widehat{G}_k(\widehat{Q}_k(X_i))\|_2^2\cdot\rho_k(X_i)\le n^{-\frac{2\beta}{d}-1},$$

我们可以得到

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\|X_i-\widehat{G}_k(\widehat{Q}_k(X_i))\|_2\rho_k(X_i)&\le\frac{1}{n}\sum_{i=1}^{n}\|X_i-\widehat{G}_k(\widehat{Q}_k(X_i))\|_2\sqrt{\rho_k(X_i)}\\&\le\sqrt{\frac{1}{n}\sum_{i=1}^{n}\|X_i-\widehat{G}_k(\widehat{Q}_k(X_i))\|_2^2\cdot\rho_k(X_i)}\\&\le n^{-\frac{\beta}{d}-\frac{1}{2}}.\end{aligned}$$

定义

$$\mathcal{F}_2=\{f=G\circ Q:G\in C_L^\beta(\mathbb{R}^d;\mathbb{R}^D),Q\in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d)\}.$$

因此，我们有 $\widehat{G}\circ\widehat{Q}\in\mathcal{F}_2$。此外，当 $\sup_{x\in\mathcal{M}\cap S_k}\|f_1(x)-f_2(x)\|_2\le\epsilon$ 时，有

$$\begin{aligned}&\sup_{x\in\mathcal{M}}\left|\|x-f_1(x)\|_2\rho_k(x)-\|x-f_2(x)\|_2\rho_k(x)\right|\\&\le\sup_{x\in\mathcal{M}\cap S_k}\|f_2(x)-f_1(x)\|_2\\&\le\epsilon.\end{aligned}$$

因此，考虑函数类 $\mathcal{F}_2=\{\|x-f(x)\|_2\rho_k(x)\ f\in\mathcal{F}_2\}$，根据引理6，有 $\log N(\mathcal{F}_2,\|\cdot\|_{\mathcal{M}},\epsilon)\le\log N(\mathcal{F}_2,\|\cdot\|_{\mathcal{M}},\epsilon)\lesssim(\frac{1}{\epsilon})^{\frac{d}{\beta}}$。根据 Dudley 熵积分界（例如参见 Wainwright（2019）），我们可以得到

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}_2}\left|\frac{1}{n}\sum_{i=1}^{n}\|X_i-f(X_i)\|_2\rho_k(X_i)-\mathbb{E}_{\mu^*}\left[\|X-f(X)\|_2\rho_k(X)\right]\right|\right]\le C\,n^{-\frac{\beta}{d}}\vee\frac{\log n}{\sqrt{n}}.$$

随后，根据 Talagrand 集中不等式（例如参见 Wainwright（2019）），我们可以得到存在常数 c_2，使得以概率 $1-n^{-3}$ 有

$$\mathbb{E}_{\mu^*}\left[\|X-\widehat{G}_k\circ\widehat{Q}_k(X)\|_2\cdot\rho_k(X)\right]\le c_2\left(n^{-\frac{\beta}{d}}\vee\frac{\log n}{\sqrt{n}}\right).$$

对于第二个陈述，我们先固定一个足够小的正常数 $r>0$，稍后再确定其取值。然后，对于任意 $z\in\Omega_k=Q^*(\mathcal{M}\cap S_k)$，存在 $\sigma(z)\in\Omega_k$，使得 $z\in\mathbb{B}_r(\sigma(z))$ 和 $\nu_k^*(\sigma(z))\ge g(r)>0$。令 $\mathcal{A}_z=\{\sigma(z):\ z\in\Omega_k\}$ 和 $f_k=G_k\circ Q_k\circ G_k^*$，我们采用以下引理，它为所有 $z\in\mathcal{A}_z$ 给出 $\|G_k^*(z)-f_k(z)\|_2$ 的一个上界。

引理 7. 以至少 $1-cn^{-3}$ 的概率，对于所有 $z\in\mathcal{A}_z$，

$$\sum_{\substack{j\in\mathbb{N}_0^d\\ |j|\le\lfloor\beta\rfloor}}\frac{1}{j!}\,\|\widehat{f}_k^{(j)}(\widetilde{z})-G_k^{*,(j)}(\widetilde{z})\|_2\,(\delta_n)^{|j|}\le C\cdot\big(g(r)\big)^{-\frac{2\beta}{d}-1}\cdot\Big(\frac{\log n}{n}\Big)^{\frac{\beta}{d}},\tag{17}$$

其中 $\delta_n=b_1\cdot(g(r))^{-\frac{2}{d}}\cdot\big(\frac{\log n}{n}\big)^{\frac{1}{d}}$，这里的常数 b_1 与 n 和 r 无关。

因此，根据引理 7，我们可以得到

$$\begin{aligned}&\sup_{z\in\mathcal{A}_z}\|G_k^*(z)-\widehat{f}_k(z)\|_2\le C\cdot\big(g(r)\big)^{-\frac{2\beta}{d}-1}\cdot\Big(\frac{\log n}{n}\Big)^{\frac{\beta}{d}}\\&\sup_{z\in\mathcal{A}_z}\|\mathbf{J}_{G_k^*}(z)-\mathbf{J}_{\widehat{f}_k}(z)\|_F\le C_1\cdot\big(g(r)\big)^{-\frac{2\beta-2}{d}-1}\cdot\Big(\frac{\log n}{n}\Big)^{\frac{\beta-1}{d}}.\end{aligned}$$

此外，由于 G_k^* 和 $f=G_k\circ Q_k\circ G_k^*$ 是 β-Hölder 光滑的，且满足 $\beta\ge1+\tilde{\alpha}>1$，我们有

$$\begin{aligned}&\sup_{z\in\Omega_k}\|G_k^*(z)-\widehat{f}_k(z)\|_2\le C\cdot\big(g(r)\big)^{-\frac{2\beta}{d}-1}\cdot\Big(\frac{\log n}{n}\Big)^{\frac{\beta}{d}}+C_2\,r\\&\sup_{z\in\Omega_k}\|\mathbf{J}_{G_k^*}(z)-\mathbf{J}_{\widehat{f}_k}(z)\|_F\le C_1\cdot\big(g(r)\big)^{-\frac{2\beta-2}{d}-1}\cdot\Big(\frac{\log n}{n}\Big)^{\frac{\beta-1}{d}}+C_3\,r.\end{aligned}\tag{18}$$

由于对任意 $z\in\Omega_k$，都有 $z=Q_k^*(G_k^*(z))$，因此得到 $I_d=\mathrm{J}^{Q_k^*}(G_k^*(z))\,\mathrm{J}^{G_k^*}(z)$。由于 Q_k^* 是 L-Lipschitz 的，$\mathrm{J}^{Q_k^*}(G_k^*(z))$ 的算子范数有界，这意味着对某个正常数 $c>0$ 以及 $z\in\Omega_k$，$\det(\mathrm{J}_{G_k^*}^T(z)\,\mathrm{J}^{G_k^*}(z))\ge c$。此外，由于 G_k^* 是 β-Hölder 光滑的，且满足 $\beta>1$，因此存在正常数 ϵ，使得对于 Ω_k的 ϵ-扩张 $\Omega_{k,\epsilon}=\{y\in\mathbb{B}_\epsilon(z):\ z\in\Omega_k\}$，对任意 $z\in\Omega_{k,\epsilon}$，都有 $\det(\mathrm{J}_{G_k^*}^T(z)\,\mathrm{J}^{G_k^*}(z))\ge\frac{c}{2}$。因此，式 (18) 中的第二个表达式以及 $\widehat{f}_k$的 β-Hölder 光滑性意味着，对于所有充分小的 ϵ, r 和充分大的 n，$\inf_{z\in\Omega_{k,\epsilon}}\det(\mathrm{J}_{\widehat{f}_k}^T(z)\,\mathrm{J}^{\widehat{f}_k}(z))\ge\frac{c}{4}$。现在令 $l_k=Q_k\circ G_k^*$，并利用恒等式 $f_k=G_k\circ l_k$，

$$\begin{aligned}\mathbf{J}_{\widehat{f}_k}^T(z)\,\mathbf{J}_{\widehat{f}_k}(z)&=\Big(\mathbf{J}_{\widehat{G}_k}(\widehat{l}_k(z))\,\mathbf{J}_{\widehat{l}_k}(z)\Big)^T\Big(\mathbf{J}_{\widehat{G}_k}(\widehat{l}_k(z))\,\mathbf{J}_{\widehat{l}_k}(z)\Big)\\&=\mathbf{J}_{\widehat{l}_k}^T(z)\,\mathbf{J}_{\widehat{G}_k}^T(\widehat{l}_k(z))\,\mathbf{J}_{\widehat{G}_k}(\widehat{l}_k(z))\,\mathbf{J}_{\widehat{l}_k}(z),\end{aligned}$$

通过取行列式，我们进一步得到（注意，$\mathbf{J}^{\widehat{l}_k}(z)$ 是一个方阵）

$$\det{}^2\Big(\mathbf{J}_{\widehat{l}_k}(z)\Big)\cdot\det\Big(\mathbf{J}_{\widehat{G}_k}^T(\widehat{l}_k(z))\,\mathbf{J}_{\widehat{G}_k}(\widehat{l}_k(z))\Big)\ge\frac{c}{4}.$$

由于 G_k 和 Q_k 都是 L-Lipschitz 的，我们可以进一步推得，对于所有$z\in\Omega_{k,\epsilon}$，$0<c_1\le\det(\mathbf{J}^{\widehat{l}_k}(z))\le c_2$。

我们声称，当 ϵ、r 足够小且 n 足够大时，l_k 在 $\Omega_{k,\epsilon}$ 上全局可逆。否则，假设存在 $\Omega_{k,\epsilon}$ 中的两个不同元素 z_0 和 z_1，使得 $l_k(z_0)=l_k(z_1)$。由于 $0<c_1\le$

$\det(\mathbf{J}_{\widehat{l}_k}(z)) \le c_2$ 意味着 l_k 局部可逆，即存在一个与 ϵ 无关的常数 $b_0 > 0$，使得 $\|z_0 - z_1\| \ge b_0$。根据 $\Omega_{k,\epsilon}$ 的定义以及 G_k 和 l_k的利普希茨连续性，存在$\widehat{\Omega}_k$ 中的 $\bar{z}_0$ 和 $\bar{z}_1$，使得（当 ϵ足够小时）

$$\|\bar{z}_0 - \bar{z}_1\| \ge \frac{1}{2} b_0, \quad \|\widehat{l}_k(\bar{z}_0) - \widehat{l}_k(\bar{z}_1)\|_2 \le C\epsilon \quad \text{and} \quad \|\widehat{f}_k(\bar{z}_0) - \widehat{f}_k(\bar{z}_1)\| \le C\epsilon. \tag{19}$$

上面的第三个显示式与（18）中的第一个显示式相结合可推出 $\|G_k^*(\bar{z}_0) - G_k^*(\bar{z}_1)\|_2 \le C_1(\epsilon + r)$。另一方面，由上面的第一个显示式以及 Q_k^* 的利普希茨连续性，我们有

$$\frac{1}{2} b_0 \le \|\bar{z}_0 - \bar{z}_1\| = \|Q_k^*(G_k^*(\bar{z}_0) - Q_k^*(G_k^*(\bar{z}_1))\|_2 \le C\|G_k^*(\bar{z}_0) - G_k^*(\bar{z}_1)\|_2 \le CC_1(\epsilon + r),$$

即 当 ϵ、r 取得足够小时，这将产生矛盾

令 $\widehat{l}_k^{-1} : \widehat{l}_k(\Omega_{k,\epsilon/2}) \to \Omega_{k,\epsilon/2}$ 为 $\widehat{l}_k$ 在 $\Omega_{k,\epsilon/2}$ 上的逆函数。利用霍尔德空间上的反函数定理（例如参见（Eldering, 2013）的附录 A），我们可以得出，对于某个充分大的常数 C_0，$\widehat{l}_k^{-1} \in C^{\beta}_{C_0}(\widehat{l}_k(\Omega_{k,\epsilon/2}); \mathbb{R}^d)$。因此，我们可以将 $\nu^*_{k,\widehat{Q}_k} = [\widehat{Q}_k]\#(\frac{\rho_k \mu^*}{p_k})$ 的密度函数表达式写成

$$\nu^*_{k,\widehat{Q}_k}(y) = \nu_k^*(\widehat{l}_k^{-1}(y)) \cdot \left(\det\left(\mathbf{J}^T_{\widehat{l}_k^{-1}}(x)\mathbf{J}_{\widehat{l}_k^{-1}}(x)\right)\right)^{\frac{1}{2}} \cdot \mathbf{1}\left(y \in \widehat{l}_k(\Omega_k)\right)$$

通过对 $y = \widehat{l}_k(z)$ 应用关于 $z \sim \nu_k^*$ 的变量替换。此外，由于 $\nu_k^* \in C_L^{\alpha}(\mathbb{R}^d)$，结合 $\widehat{l}_k^{-1} \in C^{\beta}_{C_0}(\widehat{l}_k(\Omega_{k,\epsilon/2}); \mathbb{R}^d)$，可知对于某个常数 C_1，有 $\nu^*_{\widehat{Q}} \in C^{\tilde{\alpha}}_{C_1}(\mathbb{R}^d)$（回忆 $\tilde{\alpha} = \alpha \wedge (\beta - 1)$）。

C.4 引理 7 的证明

证明沿用 Tang & Yang (2022) 的分析。令 $h_n = \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$ 和 $\mathcal{N}_{h_n} \subset \mathcal{A}_z$ 为 $\mathcal{A}_z$ 在 ℓ_2 距离下的一个最小 h_n-覆盖集，其基数满足 $|\mathcal{N}_{h_n}| \le C\ n^{\text{对数}}\ n$。对于任意 $\widetilde{z} \in \mathcal{N}_{h_n}$，定义 $\delta_n = b\left(\frac{\log n}{n}\right)^{\frac{1}{d}}$。

我们声称，只需证明：当 b 足够大时，以至少 $1 - n^{-3}$ 的概率，对于任意 $\widetilde{z} \in \mathcal{N}_{h_n}$，

$$\sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \|\widehat{f}_k^{(j)}(\widetilde{z}) - G_k^{*,(j)}(\widetilde{z})\|_2 \, (\delta_n)^{|j|} \le C \left(\frac{\log n}{n}\right)^{\frac{\beta}{d}}. \tag{20}$$

事实上，若此不等式成立，则可以运用基于 h_n-覆盖集的标准逼近论证。具体而言，对于任意 $z \in \mathcal{A}_z$，存在 $\widetilde{z} \in \mathcal{N}_{h_n}$ 使得 $\|z - \widetilde{z}\|_2 \le h_n = (\frac{\log n}{n})^{\frac{1}{d}}$，我们可以通过对 $G_k^*(z) - f_k(z)$ 进行泰勒展开得到

$$\begin{aligned}
\|G_k^*(z) - \widehat{f}_k(z)\|_2 &\le C \sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \|\widehat{f}_k^{(j)}(\widetilde{z}) - G_k^{*,(j)}(\widetilde{z})\|_2 \left(\frac{\log n}{n}\right)^{\frac{|j|}{d}} + C\left(\frac{\log n}{n}\right)^{\frac{\beta}{d}} \\
&\le C \left(\frac{\log n}{n}\right)^{\frac{\beta}{d}}.
\end{aligned}$$

现在来证明不等式 (20)。回顾一下，

$$\frac{1}{n} \sum_{i=1}^{n} \|X_i - \widehat{G}_k(\widehat{Q}_k(X_i))\|^2 \rho_k(X_i) \le C\, n^{-\frac{2\beta}{d} - 1}.$$

特别地，对于固定的 $\widetilde{z} \in \mathcal{N}_{h_n}$，将求和限制在 $\mathbb{B}_{\delta_n}(\widetilde{z})$ 中的那些 $Q^*(X_i)$ 上，我们进一步得到 $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$（回顾一下，$f_k = G_k \circ Q_k \circ G_k^*$）

$$\frac{1}{n}\sum_{i=1}^{n} \|G_k^*(Q_k^*(X_i)) - \widehat{f}_k(Q_k^*(X_i))\|^2 \cdot \rho_k(X_i) \cdot \mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X_i)) \leq C\, n^{-\frac{2\beta}{d}-1}.$$

$\widehat{\ }$对前述表达式中的 $G_k^*(z) - f_k(z)$ 在 $\widetilde{z}$ 处进行泰勒展开，并利用这样一个事实：对于某个充分大的常数 C_0，有 $G_k^* - \widehat{f}_k \in C_{C_0}^{\beta}(\mathbb{B}_1^d;\ \mathbb{R}^D)$，经过一些代数运算后，我们可以得到以下局部基本不等式

$$\begin{aligned} &U_n(\widetilde{z},\, \widehat{f}_k) := \\ &\frac{1}{n}\sum_{i=1}^{n} \Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\leq\lfloor\beta\rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right) (Q_k^*(X_i) - \widetilde{z})^j \Big\|_2^2 \cdot \mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X_i)) \cdot \rho_k(X_i) \\ &\leq c\Bigg((\delta_n)^{2\beta} + (\delta_n)^{\beta} \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\leq\lfloor\beta\rfloor}} \frac{1}{j!} \left\|G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right\|_2 (\delta_n)^{|j|} \Bigg) \cdot \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X_i)) \cdot \rho_k(X_i). \end{aligned} \tag{21}$$

(21) 右侧的第二个因子可以通过简单的并集界论证以及有界函数的伯恩斯坦不等式进行界定，如下所示。首先，我们可以对期望进行界定

$$\begin{aligned} &\mathbb{E}_{\mu^*}\left[\mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X)\right] \\ &\overset{(i)}{=} p_k \int_{\mathbb{B}_{\delta_n}(\widetilde{z})} \nu_k^*(z)\,\mathrm{d}z \leq C\, p_k\, \delta_n^d \\ &\leq C\, b^d\, \frac{\log n}{n}, \end{aligned} \tag{22}$$

其中，步骤 (i) 由以下事实得出：$\nu_k^* = (Q_k^*)\#(\frac{\mu^*\cdot\rho_k}{p_k})$。由于随机变量 $\mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X))\cdot\rho_k(X)$ 一致有界于 1，且不等式 (22) 和 $\rho_k \leq 1$ 意味着其方差有界于 $C_1\, b^d \log_n{}^n$，因此我们可以对所有 $\widetilde{z} \in \mathcal{N}_{h_n}$（其中 $|\mathcal{N}_{h_n}| \leq C\ {}^n\log n$）应用伯恩斯坦不等式和简单的并集界论证，从而以至少 $1 - n^{-c}$ 的概率得到，

$$\sup_{\widetilde{z}\in\mathcal{N}_{h_n}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X_i)) \cdot \rho_k(X_i) - \mathbb{E}_{\mu^*}\left[\mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X)\right] \right| \leq C_2 b^{\frac{d}{2}} \cdot \frac{\log n}{n}, \tag{23}$$

结合式 (22) 可得

$$\sup_{\widetilde{z}\in\mathcal{N}_{h_n}} \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X_i)) \right] \leq C\, b^d \cdot \frac{\log n}{n}. \tag{24}$$

$\widehat{\ }$为了分析局部基本不等式 (21) 左侧的量 $U_n(\widetilde{z},\ f_k)$，我们将借助以下引理。

引理 8。 </style}以至少 的概率，以下不等式对任意 -光滑函数($\mathbb{B}_1^d$; R^D) </style}以及 $\widetilde{z} \in \mathcal{N}_{h_{\widetilde{n}}}$ 成立，

$$\left|U_n(\widetilde{z},\, f) - \mathbb{E}_{\mu^*}[U_n(\widetilde{z},\, f)]\right| \leq C\, b^{\frac{d}{2}} \cdot \frac{\log n}{n} \cdot \left\{ \left(\frac{\log n}{n}\right)^{\frac{2\beta}{d}} + \Big[\sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\leq\lfloor\beta\rfloor}} \frac{1}{j!} \|f^{(j)}(\widetilde{z}) - G_k^{*,(j)}(\widetilde{z})\|_2 (\delta_n)^{|j|} \Big]^2 \right\},$$

其中，期望是针对 $\{X_i\}_{i=1}^n$ 中的随机性求取的。

在应用此引理之前，请注意，对于任意 $\widetilde{z} \in \mathcal{N}_{h_n}$，我们可以将期望 $\mathbb{E}_{\mu^*}[U_n(\widetilde{z},\ f_k)]$，（其中 f 已代入 f_k）界定为

$$
\begin{aligned}
&\mathbb{E}_{\mu^*}[U_n(\widetilde{z},\, \widehat{f}_k\,)] \\
&= \mathbb{E}_{\mu^*}\Bigg[\Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\left(G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right)(Q_k^*(X)-\widetilde{z})^j\Big\|_2^2 \cdot \mathbf{1}_{\mathbb{B}_{\delta_n}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X)\Bigg] \\
&\overset{(i)}{\ge} \inf_{z\in\mathbb{B}_{\delta_n}(\widetilde{z})} \nu_k^*(z) \int_{z\in\mathbb{B}_{\delta_n}(\widetilde{z})} \Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\left(G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right)(z-\widetilde{z})^j\Big\|_2^2 \mathrm{d}z \\
&\overset{(ii)}{=} \delta_n^d \inf_{z\in\mathbb{B}_{\delta_n}(\widetilde{z})} \nu_k^*(z) \int_{\mathbb{B}_1^d} \Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\delta_n^j\left(G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right)z^j\Big\|_2^2 \mathrm{d}z \\
&\ge C\, b^d \cdot \frac{\log n}{n} \cdot g(r) \cdot \int_{\mathbb{B}_1^d} \Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\delta_n^j\left(G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right)z^j\Big\|_2^2 \mathrm{d}z,
\end{aligned}
\tag{25}
$$

其中，步骤 (i) 利用了给定 $X \sim \frac{\mu^*\cdot\rho_k}{p_k}$ 时 $Q_k^*(X)$ 服从 ν_k^* 这一事实；步骤 (ii) 是通过对 $\frac{z-\widetilde{z}}{\delta_n} \to z$ 应用变量替换 $\sim$ 得出的；最后一步则利用了对于 $\widetilde{z} \in \mathcal{A}_z$，$\nu_k^*(\widetilde{z}) \ge g(r)$ 这一事实，以及 ν_k^* 的光滑性。现在，利用以下事实：对于任意 d 元多项式 $\mathcal{S}(y) = \sum_{j\in\mathbb{N}_0^d,\,|j|\le k} a_j y^j, y \in \mathbb{R}^d$，存在某个正常数 $C(d,k)$，它仅取决于 (d,k)，使得

$$
\int_{\mathbb{B}_1^d} \mathcal{S}^2(y)\,\mathrm{d}y \ge C(d,k) \sum_{j\in\mathbb{N}_0^d,\,|j|\le k} a_j^2,
$$

我们可以得到

$$
\int_{\mathbb{B}_1^d} \Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\delta_n^j\left(G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right)z^j\Big\|_2^2 \mathrm{d}z \ge c\Bigg(\sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\left\|\widehat{f}_k^{(j)}(\widetilde{z}) - G_k^{*,(j)}(\widetilde{z})\right\|_2 (\delta_n)^{|j|}\Bigg)^2. \tag{26}
$$

最后，结合公式 (21)、(22)、(25)、(26) 和引理 8，可得：以至少 $1 - cn^{-3}$ 的概率，对于任意 $\widetilde{z} \in \mathcal{N}_{h_n}$,

$$
\begin{aligned}
&b^d \cdot \frac{\log n}{n} \cdot g(r) \cdot \Bigg(\sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\left\|\widehat{f}_k^{(j)}(\widetilde{z}) - G_k^{*,(j)}(\widetilde{z})\right\|_2 (\delta_n)^{|j|}\Bigg)^2 \\
&\le C b^{\frac{d}{2}} \cdot \frac{\log n}{n} \cdot \Bigg(\sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\left\|\widehat{f}_k^{(j)}(\widetilde{z}) - G_k^{*,(j)}(\widetilde{z})\right\|_2 (\delta_n)^{|j|}\Bigg)^2 + C b^{\frac{d}{2}} \cdot \Big(\frac{\log n}{n}\Big)^{\frac{2\beta}{d}} \cdot \frac{\log n}{n} \\
&\quad + C b^d \cdot \frac{\log n}{n} \cdot \Bigg((\delta_n)^{2\beta} + (\delta_n)^{\beta} \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!}\left\|G_k^{*,(j)}(\widetilde{z}) - \widehat{f}_k^{(j)}(\widetilde{z})\right\|_2 (\delta_n)^{|j|}\Bigg).
\end{aligned}
$$

因此，通过选择足够大的 $b = b_1(g(r))^{-\frac{2}{d}}\ b_1$，并利用 $\delta_n = b\big(\frac{\log n}{n}\big)^{\frac{1}{d}}$ 的定义，即可得到所述不等式 (20)。

C.5 引理 8 的证明

该证明可由 Tang & Yang (2022) 中引理 18 的证明推出，为了完整起见，我们在此给出证明。由于 $f \in C_L^{\beta}(\mathbb{B}_1^d;\, R^D)$，对于任意满足 $|j| \le \lfloor\beta\rfloor$ 的 $z \in \mathbb{B}_1^d$ 和 $j \in \mathbb{N}_0^d$，均有 $\|f^{(j)}(z)\|_2 \le \sqrt{D}L = \overline{C}_0$。

~对于任意固定的 $\widetilde{z} \in \mathcal{N}_{h_{\widetilde{n}}}$ 和 $\delta > 0$，令

$$\bar{\mathcal{T}}(\widetilde{\delta}) = \Big\{ T = \{T_j\}_{j \in \mathbb{N}_0^d,\, |j| \le \lfloor \beta \rfloor} \in [-C_0,\, C_0]^{D \times \binom{d+\lfloor \beta \rfloor - 1}{d}} : \sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left\| T_j - G_k^{*,(j)}(\widetilde{z}) \right\|_2 (\delta_{\widetilde{z}})^{|j|} \le \widetilde{\delta} \Big\}.$$

~我们还定义了以下以 $T \in \mathcal{T}(\delta)$ 为指标的经验过程上确界，

$$\begin{aligned} & Z_n(\widetilde{\delta}) = \\ & \sup_{T \in \bar{\mathcal{T}}(\widetilde{\delta})} \Bigg| \mathbb{E}_{\mu^*} \Big[\Big\| \sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - T_j \right) (Q_k^*(X) - \widetilde{z})^j \Big\|_2^2 \cdot \mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X) \Big] \\ & - \frac{1}{n} \sum_{i=1}^n \Big[\Big\| \sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - T_j \right) (Q_k^*(X_i) - \widetilde{z})^j \Big\|_2^2 \cdot \mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X) \Big] \Bigg|, \end{aligned}$$

以及 $R_n(\widetilde{\delta}) = \mathbb{E}_{\mu^*}\left[Z_n(\widetilde{\delta}) \right]$。我们将先证明固定半径 $\widetilde{\delta} > 0$ 下的集中不等式，然后使用剥离技巧使半径可以是随机的，从而得到所需结果。

为了应用 Talagrand 集中不等式（例如，参见 Wainwright（2019）中的定理3.27）~ ~ ~来界定固定 $\delta > 0$ 下的差值 $|Z_n(\delta) - R_n(\delta)|$，我们注意到，上述第二个经验和中的每个加性分量的二阶矩均一致有界于

$$\begin{aligned} & \mathbb{E}_{\mu^*} \Bigg[\sup_{T \in \bar{\mathcal{T}}(\widetilde{\delta})} \Big(\Big\| \sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - T_j \right) (Q_k^*(X) - \widetilde{z})^j \Big\|_2^4 \cdot \mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X) \Big) \Bigg] \\ & \le \sup_{\substack{z \in \mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z}) \\ T \in \bar{\mathcal{T}}(\widetilde{\delta})}} \Big\| \sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - T_j \right) (z - \widetilde{z})^j \Big\|_2^4 \cdot \mathbb{E}_{\mu^*} \left[\mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X) \right] \\ & \le C \sup_{T \in \bar{\mathcal{T}}(\widetilde{\delta})} \Bigg(\sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left\| T_j - G_k^{*,(j)}(\widetilde{z}) \right\|_2 (\delta_{\widetilde{z}})^{|j|} \Bigg)^4 \cdot b_2^d \cdot \frac{\log n}{n} \\ & \le C\, b_2^d\, \widetilde{\delta}^4 \cdot \frac{\log n}{n}, \end{aligned}$$

这里我们使用不等式（22）来界定 $\mathbb{E}_{\mu^*}\left[\mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X) \right]$。此外，每个加性分量几乎必然有界于

$$\begin{aligned} & \sup_{z \in \mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})} \Big\| \sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - T_j \right) (Q_k^*(X) - \widetilde{z})^j \Big\|_2^2 \\ & \le C \Bigg(\sum_{\substack{j \in \mathbb{N}_0^d \\ |j| \le \lfloor \beta \rfloor}} \frac{1}{j!} \left\| T_j - G^{*,(j)}(\widetilde{z}) \right\|_2 (\delta_{\widetilde{z}})^{|j|} \Bigg)^2 \le C\, \widetilde{\delta}^2. \end{aligned}$$

基于这两个界，我们可以应用 Talagrand 集中不等式，得到对于任意$s \ge 0$,

$$\mathbb{P}\big(Z_n(\widetilde{\delta}) \ge R_n(\widetilde{\delta}) + s^2 \big) \le 2 \exp\left(- \frac{c\, n s^4}{s^2\, \widetilde{\delta}^2 + b_2^d\, \widetilde{\delta}^4 \cdot \frac{\log n}{n}} \right). \tag{27}$$

~接下来，我们通过对称化技巧和链式方法来界定期望 $R_n(\delta)$。根据一个标准的

对称化后，我们可以得到

$$R_n(\widetilde{\delta}) \le \frac{2}{\sqrt{n}}\,\mathbb{E}\Bigg[\sup_{T\in\bar{\mathcal{T}}(\widetilde{\delta})} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\Bigg[\Big\|\sum_{\substack{j\in\mathbb{N}_0^d\\ |j|\le\lfloor\beta\rfloor}}\frac{1}{j!}\left(G_k^{*,(j)}(\widetilde{z})-T_j\right)\left(Q_k^*(X)-\widetilde{z}\right)^j\Big\|_2^2\cdot\mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X_i))\cdot\rho_k(X_i)\Bigg]\right|\Bigg],$$

其中，$\{\varepsilon_i\}_{i=1}^n$ 是来自拉德马赫分布的 n 独立同分布副本，即 $\mathbb{P}(\varepsilon_i=1)=\mathbb{P}(\varepsilon_i=-1)=0.5$。由于给定 $\{X_i\}_{i=1}^n$后，上确界中的随机过程是一个具有内在度量的次高斯过程

$$\begin{aligned}
&d_n^2(T,\widetilde{T})\\
&=\frac{1}{n}\sum_{i=1}^{n}\Bigg(\Big\|\sum_{\substack{j\in\mathbb{N}_0^d\\ |j|\le\lfloor\beta\rfloor}}\frac{1}{j!}\left(G_k^{*,(j)}(\widetilde{z})-T_j\right)\left(Q_k^*(X_i)-\widetilde{z}\right)^j\Big\|_2^2\\
&\qquad\qquad-\Big\|\sum_{\substack{j\in\mathbb{N}_0^d\\ |j|\le\lfloor\beta\rfloor}}\frac{1}{j!}\left(G_k^{*,(j)}(\widetilde{z})-\widetilde{T}_j\right)\left(Q_k^*(X_i)-\widetilde{z}\right)^j\Big\|_2^2\Bigg)^2\cdot\mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X_i)))\cdot\rho_k(X_i)\\
&\le C\,\widetilde{\delta}^4\,\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X_i)))\cdot\rho_k(X_i),
\end{aligned}$$

~ – ~ – ~对于任意 $T,T\in\mathcal{T}(\delta)$，其中最后一步用到了 $\mathcal{T}(\delta)$ 的定义。结合上述结果与不等式 (22)，可得

$$\mathbb{E}_{\mu^*}\Big[\sup_{T,\widetilde{T}\in\bar{\mathcal{T}}(\delta)}d_n^2(T,\widetilde{T})\Big]\le C\,b_2^d\,\widetilde{\delta}^4\cdot\frac{\log n}{n}\quad\text{and}\quad d_n(T,\widetilde{T})\le C\widetilde{\delta}\sum_{\substack{j\in\mathbb{N}_0^d\\ |j|\le\lfloor\beta\rfloor}}\frac{1}{j!}\left\|T_j-\widetilde{T}_j\right\|_2\delta_{\widetilde{z}}^{|j|}.$$

最后，令 $\mathcal{K}_n(\delta)=\sup_{T,\widetilde{T}\in\bar{\mathcal{T}}(\delta)}d_n^2(T,\widetilde{T})$，通过应用杜德利不等式进行标准链式方法，我们可以得到

$$\begin{aligned}
R_n(\widetilde{\delta}) &\le C\,\frac{1}{\sqrt{n}}\,\mathbb{E}_{\mu^*}\Big[\int_0^{\mathcal{K}_n(\widetilde{\delta})}\sqrt{\log\frac{\widetilde{\delta}}{u}}\,\mathrm{d}u\Big]\\
&= C\,\frac{1}{\sqrt{n}}\,\mathbb{E}_{\mu^*}\Big[\mathcal{K}_n(\widetilde{\delta})\cdot\int_0^1\sqrt{\log\frac{\widetilde{\delta}}{u\cdot\mathcal{K}_n(\widetilde{\delta})}}\,\mathrm{d}u\Big]\\
&= C\,\frac{1}{\sqrt{n}}\,\mathbb{E}_{\mu^*}\Big[\mathcal{K}_n(\widetilde{\delta})\cdot\mathbf{1}(\mathcal{K}_n(\widetilde{\delta})\le b_2^{\frac{d}{2}}\widetilde{\delta}^2\sqrt{\tfrac{\log n}{n}})\int_0^1\sqrt{\log\frac{\widetilde{\delta}}{u\cdot\mathcal{K}_n(\widetilde{\delta})}}\,\mathrm{d}u\Big]\\
&+ C\,\frac{1}{\sqrt{n}}\,\mathbb{E}_{\mu^*}\Big[\mathcal{K}_n(\widetilde{\delta})\cdot\mathbf{1}(\mathcal{K}_n(\widetilde{\delta})> b_2^{\frac{d}{2}}\widetilde{\delta}^2\sqrt{\tfrac{\log n}{n}})\int_0^1\sqrt{\log\frac{\widetilde{\delta}}{u\cdot\mathcal{K}_n(\widetilde{\delta})}}\,\mathrm{d}u\Big]\\
&\le C_1\,b_2^{\frac{d}{2}}\cdot\frac{\log(n/\widetilde{\delta})}{n}\cdot\widetilde{\delta}^2,
\end{aligned}\tag{28}$$

其中我们利用了以下事实：相对于度量 d_n，$\bar{\mathcal{T}}(\widetilde{\delta})$ 的 u-覆盖熵至多为 $C_2\log\frac{\widetilde{\delta}}{u}$，对于 $u\in(0,1)$ 均成立，其中 C_2 依赖于 (d,D)（对 D 的依赖至多为多项式级）。结合不等式 (27)，我们得到对所有 $t\ge1$,

$$\mathbb{P}\Big(Z_n(\widetilde{\delta})\ge C\,t^2\,b_2^{\frac{d}{2}}\cdot\frac{\log(n/\widetilde{\delta})}{n}\cdot\widetilde{\delta}^2\Big)\le 2\exp\Big(-c\,t^2\,\log(n/\widetilde{\delta})\Big).\tag{29}$$

最后，我们应用剥离技巧，将上述关于 $Z_n(\delta)$ 的高概率界推广到随机半径 $\widetilde{\delta} = \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!} \left\|\widehat{f}_k^{(j)} - G^{*,(j)}_{\cdot,k}(\widetilde{z})\right\|_2 (\delta_{\widetilde{z}})^{|j|}$。具体而言，我们首先设定基本层级 $\bar{\delta} = \left(\frac{\log n}{n}\right)^{\frac{\beta}{d}}$，并对于 $s = 1, \cdots, S$，其中 $S \le C \log \frac{1}{\bar{\delta}}$，定义集合

$$\widetilde{\mathcal{T}}_0 = \Big\{T = \{T_j\}_{j\in\mathbb{N}_0^d, |j|\le\lfloor\beta\rfloor} \in [-C_0, C_0]^{D\times\binom{d+\lfloor\beta\rfloor-1}{d}} : \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!} \|T_j - G_k^{*,(j)}(\widetilde{z})\|_2 (\delta_{\widetilde{z}})^{|j|} \le \bar{\delta}\Big\};$$

$$\widetilde{\mathcal{T}}_s = \Big\{T = \{T_j\}_{j\in\mathbb{N}_0^d, |j|\le\lfloor\beta\rfloor} \in [-C_0, C_0]^{D\times\binom{d+\lfloor\beta\rfloor-1}{d}} : 2^{s-1}\bar{\delta} \le \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!} \|T_j - G_k^{*,(j)}(\widetilde{z})\|_2 (\delta_{\widetilde{z}})^{|j|} \le 2^s\bar{\delta}\Big\}.$$

将不等式 (29) 应用于 $\delta = 2^s\bar{\delta}$，其中 $s \in [S]$ 取足够大的常数 $t > 0$，由于 $C_1 \le -\log(2^s\bar{\delta}) \le C_2 \log n$，我们得到

$$\mathbb{P}\left(Z_n(\bar{\delta}) \ge C\, b_2^{\frac{d}{2}} \frac{\log n}{n} \bar{\delta}^2\right) + \sum_{s=1}^{S} \mathbb{P}\left(Z_n(2^s\bar{\delta}) \ge C\, b_2^{\frac{d}{2}} \frac{\log n}{n} 4^s\bar{\delta}^2\right) \le n^{-(c+1)}.$$

注意，对于任意 $T \in \widetilde{\mathcal{T}}_s$ 和任意 $s \in \{0\} \cup [S]$，事件 $Z_n(2^s\bar{\delta}) \le C\, b_2^{\frac{d}{2}} \log n\, 4^s\bar{\delta}^2$ 意味着

$$\begin{aligned}
&\left| \mathbb{E}_{\mu^*}\left[\Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - T_j\right)(Q_k^*(X) - \widetilde{z})^j \Big\|_2^2 \cdot \mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X)\right] \right. \\
&\quad \left. - \frac{1}{n}\sum_{i=1}^{n}\left[\Big\| \sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!} \left(G_k^{*,(j)}(\widetilde{z}) - T_j\right)(Q_k^*(X_i) - \widetilde{z})^j \Big\|_2^2 \cdot \mathbf{1}_{\mathbb{B}_{\delta_{\widetilde{z}}}(\widetilde{z})}(Q_k^*(X)) \cdot \rho_k(X)\right] \right| \\
&\le c_1\, b_2^{\frac{d}{2}} \frac{\log n}{n} \left\{\bar{\delta}^2 + \Big(\sum_{\substack{j\in\mathbb{N}_0^d \\ |j|\le\lfloor\beta\rfloor}} \frac{1}{j!} \|T_j - G_k^{*,(j)}(\widetilde{z})\|_2 (\delta_{\widetilde{z}})^{|j|} \Big)^2 \right\}.
\end{aligned}$$

最后，由于对于任意 $f \in C_L^\beta(\mathbb{B}_1^d; R^D)$，$T_f := \{T_{f,j} = f^{(j)}\}_{j\in\mathbb{N}_0^d, |j|\le\lfloor\beta\rfloor}$ 必属于某个 $\widetilde{\mathcal{T}}_s$，所述结论可由前面两个公式以及对 $\widetilde{z} \in \mathcal{N}_{h_n}$ 的简单并集界推出，其中 $|\mathcal{N}_{h_n}| \le C^n \log n \le C\, n$。

C.6 引理 5 的证明

首先，根据伯恩斯坦不等式，以至少 $1 - n^{-3}$ 的概率有 $|p_k - \widehat{p}_k| \le C\sqrt{\log n}$；然后根据 $p_k > 0$，对于足够大的 n，有 $\left|1 - \frac{p_k}{\widehat{p}_k}\right| \le C\sqrt{\log n}$。因此

$$\begin{aligned}
&\sup_{Q\in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d)} \sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)} \Big(\frac{1}{p_k}\int f(Q(x))\rho_k(x)\,\mathrm{d}\mu^* - \int f(z)\,\widetilde{\nu}_{k,Q}(z)\,\mathrm{d}z\Big) \\
&= \sup_{Q\in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d)} \sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)} \Big(\frac{1}{p_k}\int f(Q(x))\rho_k(x)\,\mathrm{d}\mu^* - \frac{1}{\widehat{p}_k n}\sum_{i=1}^{n} f(Q(X_i))\rho_k(X_i)\Big) \\
&\le C\sqrt{\frac{\log n}{n}} + \frac{1}{p_k} \sup_{Q\in C_L^\beta(\mathbb{R}^D;\mathbb{R}^d)} \sup_{f\in\mathrm{Lip}_1(\mathbb{R}^d)} \Big(\int f(Q(x))\rho_k(x)\,\mathrm{d}\mu^* - \frac{1}{n}\sum_{i=1}^{n} f(Q(X_i))\rho_k(X_i)\Big) \\
&\le C\sqrt{\frac{\log n}{n}} + C \sup_{f\in\mathrm{Lip}_1(\mathbb{R}^D)} \Big(\int f(x)\rho_k(x)\,\mathrm{d}\mu^* - \frac{1}{n}\sum_{i=1}^{n} f(X_i)\rho_k(X_i)\Big),
\end{aligned}$$

其中最后一个不等式是由假设 $\beta \geq 1$ 得出的。然后考虑 $f, f' \in \text{Lip}_1(\mathbb{R}^D)$ 的伪度量

$$d_n(f,f') = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\big(f(X_i)\rho_k(X_i) - f'(X_i)\rho_k(X_i)\big)^2} \leq \sup_{x\in S_k\cap\mathcal{M}} |f(x)-f'(x)| = \sup_{x\in G^*(Q^*(S_k\cap\mathcal{M}))} |f(x)-f'(x)|.$$

由引理6可得 $\log \mathbf{N}(\text{Lip}_1(\mathbb{R}^D), \|\cdot\|_{L^\infty(S_k\cap\mathcal{M})}, \epsilon) \leq C\,\epsilon^{-d}$。取 $\delta = (\frac{1}{n})^{\frac{1}{d}}$，可得

$$\frac{1}{\sqrt{n}}\int_{\delta}^{1}\sqrt{\log \mathbf{N}(\text{Lip}_1(\mathbb{R}^D), \|\cdot\|_{L^\infty(S_k\cap\mathcal{M})}, \epsilon)}\,\mathrm{d}\epsilon \leq C\,n^{-\frac{1}{d}} \vee \frac{\log n}{\sqrt{n}}.$$

因此，与引理 3 的证明中的分析类似，利用 Dudley 熵积分界和 Talagrand 集中不等式，可以得到所需结果。

D 技术细节的证明

D.1 引理 1 的证明

考虑任意一个 $\mu^* \in \mathcal{P}^*(d, D, \alpha, \beta, L^*)$。记 $\mathcal{M} = \text{supp}(\mu^*)$，首先考虑以下引理。

引理 9 (Tang & Yang (2022) 的引理 17)。存在正常数 (τ_1, L_1)，使得对于任意 $x_0 \in \mathcal{M}$，定义 $Q_{x_0}: \mathbb{R}^D \to \mathbb{R}^d$ 为 $Q_{x_0}(x) = W_{x_0}^T(x - x_0)$，其中 $W_{x_0} \in \mathbb{R}^{D\times d}$ 是 $\mathcal{M}$ 在 x_0 处切空间的一组任意正交归一基，则存在集合 U_{x_0}，满足 $\mathbb{B}_{\tau_1}(x_0) \cap \mathcal{M} \subset U_{x_0} \subset \mathcal{M}$，以及函数 $G_{x_0} \in C^{\beta}_{L_1}(\mathbb{R}^d; \mathbb{R}^D)$，使得

(1). $G_{x_0}(\mathbb{B}_1^d) = \widetilde{U}_{x_0}$ *and for any* $z \in \mathbb{B}_1^d$, $Q_{x_0}(G_{x_0}(z)) = z$;
(2). $\mu^* \circ G_{x_0}|_{\mathbb{B}_1^d} \in C^{\alpha}_{L_1}(\mathbb{B}_1^d)$ *and for any* $z \in \partial\mathbb{B}_1^d$, $\|G_{x_0}(z) - x_0\| \geq \tau_1$.

根据 Sobolev 延拓定理，存在常数 L_2，使得对于任意 $x_0 \in \mathcal{M}$，均存在 $Q_{x_0} \in C^{\beta}_{L_2}(\mathbb{R}^D)$，使得 $Q_{x_0}|_{\widetilde{U}_{x_0}} = Q_{x_0}|_{\overline{\widetilde{U}_{x_0}}}$。由于 $J_{G_{x_0}}(z)^T J_{G_{x_0}}(z)$ 在 $z \in \mathbb{B}_1^d$ 上的特征值一致有下界，且 G_{x_0} 是 β-光滑的，具有 $\beta > 1$ 和一致有界的 Hölder 范数，因此存在一个足够小的正常数 r_0，使得对于任意 $x_0 \in \mathcal{M}$，

$$\sup_{v\in\mathbb{S}_1^{d-1}=\{v\in\mathbb{R}^d:\|v\|=1\}} \frac{\sup_{z,z'\in\mathbb{B}^d_{r_0}} \|(J_{G_{x_0}}(z) - J_{G_{x_0}}(z'))v\|}{\|J_{G_{x_0}}(0)v\|} \leq \frac{1}{3}. \tag{30}$$

接着，我们选取一个足够小的正常数 $r^* \leq \tau_1/2$，使得对于任意 $x_0 \in \mathcal{M}$，$Q_{x_0}(\mathbb{B}_{r^*}(x_0)\cap\mathcal{M}) \subset \mathbb{B}^d_{r_0}$。对于任意的 $k \in [K]$，考虑 $x_0 = a_k \in \mathcal{M}$。定义 $G_k = G_{x_0}$ 和 $Q_k = Q_{x_0}$。于是我们有 $S_k \cap \mathcal{M} \subset \mathbb{B}_{\tau_1}(x_0) \cap \mathcal{M} \subset U_{x_0}$，且对于任意 $x \in \mathcal{M}\cap S_k$，$x = G_k(Q_k(x))$。此外，令 $p_k = \mathbb{E}_{\mu^*}[\rho_k(X)]$，。由于 μ^* 的密度一致有下界，且 $r_k \geq L^*$，我们有 p_k 也一致有下界。此外，我们可以将 $\nu_k = (Q_k)_{\#}(\frac{\mu^*\rho_k}{p_k})$ 写作

$$\nu_k(z) = \begin{cases} \frac{\mu^*(G_k(z))\cdot\rho_k(G_k(z))\sqrt{\det(J_{G_k}(z)^T J_{G_k}(z))}}{\int_{Q_k(\mathcal{M}\cap S_k)} \mu^*(G_k(z))\cdot\rho_k(G_k(z))\sqrt{\det(J_{G_k}(z)^T J_{G_k}(z))}\,\mathrm{d}z} & z \in \mathbb{B}_1^d; \\ 0 & o.w. \end{cases}$$

那么，根据 $\widetilde{\rho}_k(\cdot)$ 的 α-光滑性以及 $\sum_{k\in[K]}\widetilde{\rho}_k(\cdot)$ 的一致下界性，对于某个常数 L，我们有 $\nu_k(z)|_{\mathbb{B}_1^d} \in C^{\alpha}_L(\mathbb{B}_1^d)$。另一方面，根据引理 9 的第二个结论以及 G_m 的利普希茨连续性，存在正常数 ϵ，使得 $Q_k(\mathcal{M}\cap S_k) \subset \mathbb{B}^d_{1-\epsilon}$。结合

当 $z \notin Q_k(\mathcal{M} \cap S_k)$ 时 $\nu_k(z) = 0$ 这一事实，我们可以得到 $\nu_k(z) \in C^{\alpha}_{L^*}(\mathbb{R}^d)$。此外，对于任意$z \in \Omega_k = Q_k(\mathcal{M} \cap S_k)$ 以及任意 $r > 0$，我们将证明存在 $z' \in \mathbb{B}_r(z)$，使得 $\nu_k(z') \geq c(r^\gamma \wedge 1)$。首先，若 $\|G_k(z) - x_0\| \leq r^*/2$，则有

$$\nu_k(z) \geq c_1 \, \rho_k(G_k(z)) \geq c_1 \, \frac{(\frac{3r_k^2}{4})^\gamma}{M(r^*)^{2\gamma}}.$$

另一方面，若 $\|G_k(z) - G_k(0)\| = \|G_k(z) - x_0\| \geq r^*/2$，记 $z = av$ 为 $v = z/\|z\|$ 和$a = \|z\|$。则有 $a \leq r_0$ 和

$$r^*/2 \leq \|G_k(z) - G_k(0)\| \leq c_2 \, \|z\| = c_2 \, |a|.$$

若 $r \geq a$，则 $z = av \in \mathbb{B}_r(0)$ 和 $\nu_k(0) \geq c$，其中某个正常数为 c。若 $r < a$，取 $z' = (a - r)v$，则 $z' \in \mathbb{B}_r(z)$ 且

$$\begin{aligned}
\|G_k(z') - G_k(0)\| &= \sup_{l \in \mathbb{S}_1^{d-1}} \Big(l^T G_k(z') - l^T G_k(0) \Big) \\
&= \sup_{l \in \mathbb{S}_1^{d-1}} \Big(l^T G_k(z) - l^T G_k(0) + l^T G_k(z') - l^T G_k(z) \Big) \\
&\overset{(i)}{=} \sup_{l \in \mathbb{S}_1^{d-1}} \Big(l^T J_{G_k}(z_l) av - l^T J_{G_k}(z_l') rv \Big) \\
&= \sup_{l \in \mathbb{S}_1^{d-1}} \Big(l^T J_{G_k}(z_l)(a - r)v + l^T (J_{G_k}(z_l) - J_{G_k}(z_l')) rv \Big) \\
&\overset{(ii)}{\leq} \sup_{l \in \mathbb{S}_1^{d-1}} \Big(l^T J_{G_k}(z_l)(a - r)v + \frac{r}{2a} \|G_k(z) - G_k(0)\| \Big) \\
&= \frac{a - r/2}{a} \|G_k(z) - G_k(0)\| \\
&\leq r_k - \frac{r r_k}{2 r_0},
\end{aligned}$$

其中，(i) 使用中值定理，而 (ii) 使用等式 (30) 和泰勒定理，得到

$$\begin{aligned}
\frac{r}{2a} \|G_k(z) - G_k(0)\| &= \frac{r}{2} \Big\| \int_0^1 J_{G_k}(tz) \, \mathrm{d}t \cdot v \Big\| \\
&\geq \frac{r}{2} \|J_{G_k}(0) \cdot v\| - \frac{r}{2} \Big\| \int_0^1 J_{G_k}(0) - J_{G_k}(tz) \, \mathrm{d}t \cdot v \Big\| \\
&\geq \frac{3r}{2} \sup_{z, z' \in \mathbb{B}_{r_0}^d} \|(J_{G_{x_0}}(z) - J_{G_{x_0}}(z'))v\| - \frac{r}{2} \int_0^1 \|(J_{G_k}(0) - J_{G_k}(tz))v\| \, \mathrm{d}t \\
&\geq r \sup_{z, z' \in \mathbb{B}_{r_0}^d} \|(J_{G_{x_0}}(z) - J_{G_{x_0}}(z'))v\| \\
&\geq \sup_{l \in \mathbb{S}_1^{d-1}} l^T (J_{G_k}(z_l) - J_{G_k}(z_l')) rv.
\end{aligned}$$

因此有

$$\widetilde{\rho}_k(G_k(z')) = (r_k^2 - \|G_k(z') - x_0\|^2)^\gamma \cdot \geq (\frac{r r_k^2}{2 r_0})^\gamma \geq (\frac{r_k^2}{2 r_0})^\gamma r^\gamma \geq (\frac{(L_1^*)^2}{2 r_0})^\gamma r^\gamma. \tag{31}$$

因此，存在常数 c 使得

$$\nu_k(z') \geq c \, r^\gamma.$$

因此，我们有：对于 μ^*，在 $G_k^* = G_k$、$Q_k^* = Q_k$ 和 $\nu_k^* = \nu_k$、$k \in [K]$ 的条件下，假设 A 成立。Fo r

对于第二个陈述，注意到根据 $G_k^* = G_k$ 的 β-光滑性、 $Q_k^* = Q_k$ 以及$\nu_k^* = \nu_k$ 的 α-光滑性，假设 B 对逼近族 $\mathcal{G} = \mathcal{G}_1$ 显然成立。此外，考虑

$$\overline{\nu}_k(z) = \frac{\mu^*(G_k(z)) \cdot \sqrt{\det(J_{G_k}(z)^T J_{G_k}(z))}}{\int_{\mathbb{B}_1^d} \mu^*(G_k(z)) \cdot \sqrt{\det(J_{G_k}(z)^T J_{G_k}(z))}\,\mathrm{d}z}, \quad z \in \mathbb{B}_1^d.$$

那么我们有 $\nu_k(z) = \frac{\nu_k(z)\cdot\rho_k(G_k(z))\,\mathbb{E}_{\nu_k}[\rho_k(G_k(z))]}{}$， $\nu_k(z) \in C_L^\alpha(\mathbb{B}_1^d)$ 且$\inf_{z\in\mathbb{B}_1^d}\nu_k(z) \geq L_3 > 0$。因此，存在一个 $(\alpha+1)$-光滑可逆函数 $V_k\colon \mathbb{B}_1^d \to \mathbb{B}_1^d$(例如，参见 (Caffarelli, 1996; Villani, 2009))，使得 $\nu_0 = V_{m\#}\nu_k$ 且 $\nu_k = V_m^{-1}{}_\#\nu_0$。因此， $\mathcal{G}_2$ 足以对 μ^*进行建模。

对于族 $\mathcal{G}_3$，令 V_k 为 $\overline{V_k}|_{\mathbb{B}^d_{1-\epsilon/2}}$ 到 $\mathbb{R}^d$的 $(\alpha+1)$-光滑延拓。注意， V_k 具有 $(\alpha+1)$-光滑逆，且 $V_k(\mathbb{B}^d_{1-\epsilon/2}) \subset \mathbb{B}^d_{1-\epsilon_1}$ ，其中 ϵ_1为某个正常数。我们可以将 V_k^{-1} 视为 $V_k^{-1}|_{V_k(\mathbb{B}^d_{1-\epsilon/2})}$ 到 $\mathbb{R}^d$的一个α-光滑延拓。然后我们可以根据 $Q_k(\mathcal{M}\cap S_k) \subset \mathbb{B}^d_{1-\epsilon}$定义 $G_k' = G_k \circ V_k^{-1}$ 和 $Q_k' = V_k \circ Q_k$；对于任意 $x \in \mathcal{M}\cap S_k$，我们有 $G_k'(Q_k'(x)) = x$。此外，令

$$\nu_k'(z) = (Q_k')_\# \frac{\mu^*\rho_k}{p_k} = \begin{cases} \frac{\nu_0(z)\cdot\rho_k(G_k\circ V_k^{-1}(z))}{\int_{\mathbb{B}_1^d}\nu_0(z)\cdot\rho_k(G_k\circ V_k^{-1}(z))\,\mathrm{d}z}, & z \in V_k(\mathbb{B}^d_{1-\epsilon/2}), \\ 0, & o.w. \end{cases}$$

利用 $V_k^{-1}|_{V_k(\mathbb{B}^d_{1-\epsilon/2})}$具有 $(\alpha+1)$-光滑性且 Hölder 范数有界，以及当 $z \notin V_k(\mathbb{B}^d_{1-\epsilon})$ 时 $\nu_k'(z) = 0$ 这一事实，对于某个常数 L，我们有 $\nu_k' \in C_L^\alpha(\mathbb{R}^d)$。此外，回顾一下，对于任意 $z_0 \in Q_k(\mathcal{M}\cap S_k)$ 和 $r > 0$，存在 $z_0' \in Q_k(\mathcal{M}\cap S_k)$，使得 $z_0 \in \mathbb{B}_r(z_0')$ 且 $\widetilde{\rho}_k(G_k(z_0')) \geq c_1(r^\gamma \wedge 1)$。注意，由于 V_k 的利普希茨连续性，存在一个常数 $L_4 \geq 1$ ，使得

$$\|V_k(z_0) - V_k(z_0')\| \leq L_4\|z_0 - z_0'\|.$$

因此，对于任意 $z \in V_k(Q_k(\mathcal{M}\cap S_k))$ 和 $r > 0$，存在 $z' \in V_k(Q_k(\mathcal{M}\cap S_k)) \cap \mathbb{B}_r(z)$，使得$\widetilde{\rho}_k(G_k \circ V_k^{-1}(z')) \geq \frac{c_1}{L_4^\gamma}(r^\gamma \wedge 1)$，且 $\nu_k'(z) \geq c(r^\gamma \wedge 1)$。因此，当 $\alpha = \beta - 1$ 时，假设 A 在 $G_k^* = G_k'$、 $Q_k^* = Q_k'$ 和 $\nu_k^* = \nu_k'$ 下成立，而假设 B 在 $\mathcal{G} = \mathcal{G}_3$ 下成立。

D.2 引理 2 的证明

令 $\mathcal{N}_\epsilon$ 为 $\mathcal{M}$的最小 ϵ-覆盖集，其中 ϵ 是一个稍后选取的数。由引理 9 和 $\mathcal{M}$的紧致性，我们有 $|\mathcal{N}_\epsilon| \leq C_1(\frac{1}{\epsilon})^d$，其中 C_1是一个仅依赖于 $(d,\ D,\ \beta,\ L^*)$ 的正常数。然后，若 $\epsilon \leq \tau_1$，由引理 9 可知，对于任意 $x_0 \in \mathcal{M}$

$$\mathcal{P}_{\mu^*}(\mathbb{B}_\epsilon(x_0)) = \int_{Q_{x_0}(\mathbb{B}_\epsilon(x_0))} \mu^*(G_{x_0}(z))\sqrt{\det(J_{G_{x_0}}(z)^T J_{G_{x_0}}(z))}\,\mathrm{d}z \geq C_2\,\epsilon^d.$$

然后，由伯恩斯坦不等式和一个简单的并集界论证可知，以至少$1 - n_1^{-c}$的概率，对于任意 $x_0 \in \mathcal{N}_\epsilon$,

$$\Big|\frac{1}{n_1}\sum_{i\in I_1}\mathbf{1}(\|X_i - x_0\| \leq \epsilon) - \mathcal{P}_{\mu^*}(\mathbb{B}_\epsilon(x_0))\Big| \leq \frac{1}{3n_1}\log(\delta) + \sqrt{\frac{2C_2\epsilon^d\log(\delta)}{n_1}}, \quad \delta = 2C_1 n_1^c(\frac{1}{\epsilon})^d.$$

因此，存在常数 $C_3,\ C$，使得当 $n_1 \geq C$ 时，通过选择 $\epsilon = C_3(\frac{\log n_1}{n_1})^{\frac{1}{d}}$，我们可以以至少 $1 - n_1^{-c}$ 的概率保证，对于任意 $x_0 \in \mathcal{N}_\epsilon$,

$$\Big|\frac{1}{n_1}\sum_{i\in I_1}\mathbf{1}(\|X_i - x_0\| \leq \epsilon) - \mathcal{P}_{\mu^*}(\mathbb{B}_\epsilon(x_0))\Big| \leq \frac{C_2}{2}\epsilon^d \leq \frac{1}{2}\mathcal{P}_{\mu^*}(\mathbb{B}_\epsilon(x_0)).$$

因此，对于任意 $x_0 \in \mathcal{N}_\epsilon$，存在 $i \in I_1$，使得 $\|X_i - x_0\| \leq \epsilon$。进而可得，对于任意$x \in \mathcal{M}$，存在 $i \in I_1$，使得 $\|X_i - x_0\| \leq 2\epsilon$。第一个结论的证明至此完成。对于第二个结论，当 n_1 足够大时，我们有 $\epsilon = C_3(\frac{\log n_1}{n_1})^{\frac{1}{d}} \leq \frac{r^*}{16}$。令 $\widetilde{\mathcal{N}}_{r^*/4}$ 表示 $\bigcup_{i\in I_1} \mathbb{B}_{2\epsilon}(X_i)$ 的最小$r^*/4$-覆盖集。那么，$|\mathcal{N}_{r^*/4}|$ 由$\mathcal{M}$的最小 $r^*/8$-覆盖数控制。对于任意 $x_0 \in |\mathcal{N}_{r^*/4}|$，存在一个指标 $i \in I_1$，使得 $\mathbb{B}_{2\epsilon}(X_i) \cap \mathbb{B}_{r^*/4}(x_0) \neq \emptyset$。令 I_2 为所有满足条件的指标 i（其中 $x_0 \in |\mathcal{N}_{r^*/4}|$）构成的集合。那么，对于任意 $x \in \bigcup_{i\in I_1} \mathbb{B}_{2\epsilon}(X_i)$，存在 $i \in I_2$，使得

$$\|x - X_i\| \leq r^*/4 + r^*/4 + 2\epsilon \leq \frac{5r^*}{8}.$$

因此，令 $M = |I_2|$ 和 $\{a_k\}_{k=1}^K = \{X_i\}_{i\in I_2}$，可得

$$\bigcup_{i\in I_1} \mathbb{B}_{2\epsilon}(X_i) \subset \bigcup_{k\in[K]} \mathbb{B}_{r^*}(a_k),$$

and

$$\inf_{x\in\mathcal{M}} \sum_{k\in[K]} \widetilde{\rho}_k(x) \geq ((r^*)^2 - (\frac{5r^*}{8})^2)^\gamma > (\frac{(r^*)^2}{2})^\gamma.$$

证明完毕。