arXiv:2601.10133v1 [math.ST] 15 Jan 2026

无界各向同性噪声下的曲率驱动流形拟合

RUOWEI LI 和 ZHIGANG YAO

Abstract. 流形拟合旨在从高维数据中重建低维流形，其理论框架由 Fefferman 等人 [12, 13] 建立。本文研究如何根据观测值 $Y = X + \xi$ 恢复紧致C^3 子流形 $\mathcal{M} \subset \mathbb{R}^D$，该子流形的维数为 $d < D$，且正可达半径为 τ；其中，X 在 $\mathcal{M}$ 上服从均匀分布，$\xi \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_D)$ 表示各向同性高斯噪声。为将管状邻域 Γ 中的任意点 z 投影到 $\mathcal{M}$ 上，我们通过归一化局部核构造了基于样本的估计器 $F: \Gamma \to \mathbb{R}^D$，并采用理论推导得到的带宽 $r = c_D\sigma$。在样本量为$O(\sigma^{-3d-5})$ 时，我们以高概率建立了如下的一致渐近展开:

$$F(z) = \pi(z) + \frac{d}{2} H_{\pi(z)}\sigma^2 + O(\sigma^3), \qquad z \in \Gamma,$$

其中，$\pi(z)$ 是 z 在 $\mathcal{M}$ 上的投影，$H_{\pi(z)}$ 是 $\mathcal{M}$ 在 $\pi(z)$ 处的平均曲率向量。所得流形 $F(\Gamma)$ 的可达半径下界为 $c\tau$（$c > 0$），并且与 $\mathcal{M}$ 的豪斯多夫距离达到最先进水平，为 $O(\sigma^2)$。数值实验验证了重建误差的二次衰减，并展示了估计器 F 的计算效率。我们的工作提出了一种曲率驱动框架，可实现流形去噪与重建，并达到二阶精度。

1. 引言

现实世界的数据通常表现为高维观测，但其内蕴自由度很少。这与流形假设相符，该假设指出，此类数据在 $\mathbb{R}^D$ 中靠近一个维数为 $d < D$的流形 $\mathcal{M}$。为恢复这种潜在结构，流形学习方法会构造低维嵌入，使欧几里得距离近似于测地线距离，例如等距映射 [29], 、局部线性嵌入 [28, 9],和拉普拉斯特征映射 [2]。实际数据往往会受到测量或传输噪声的污染。流形拟合旨在从含噪样本中重建光滑$\mathcal{M}$，同时保留其内蕴几何和拓扑。应用包括单细胞数据去噪 [32], 、改进 RNA 结构模型 [31], 、恢复冷冻电子显微镜和纳米光子光探测中的数据 [22], ，以及风向分析 [8]。它还为流形感知生成模型奠定基础，例如生成对抗网络 [19] 和基于 CycleGAN 的流形拟合 [33]。

1.1. 几何直觉。 在黎曼几何中，切空间为光滑子流形提供一阶线性近似。要精确描述其局部嵌入，需要二阶几何信息，这些信息由平均曲率向量编码（图 1.1）。平均曲率向量定义为第二基本形式的归一化迹，或者等价地，定义为主曲率的平均值 κ_i（见第 2.2 节）。平均曲率向量是刻画子流形向法空间弯曲的基本二阶外蕴量。例如，在球面上，它指向内侧；而在马鞍面上，它为零。

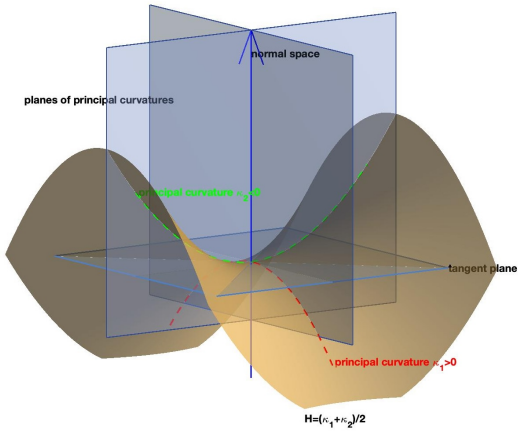


FIGURE 1.1. 马鞍面的平均曲率 H

考虑加性高斯噪声模型 $Y = X + \xi$，其中随机向量 X 在 d维流形 $\mathcal{M} \subset \mathbb{R}^D$，$D > d$ 上服从均匀分布，而 $\xi \sim \mathcal{N}(0, \sigma^2 I_D)$表示各向同性高斯噪声。如图 1.2 所示，当流形 $\mathcal{M}$上的样本 x受到各向同性环境噪声（蓝色虚线圆）的扰动时，噪声样本的密度（可直观地视为覆盖频率）蕴含着底层流形的几何信息。直观而言，平均曲率向量支配着噪声的分布，使得内法线方向的密度高于外法线方向（红色虚线）。这一现象也在脊理论框架下有所讨论 [18]。沿主曲率方向，切空间内密度函数的对称性使偏差可以忽略。因此，曲率会影响含噪点 y的密度，即 $p_\sigma(y)$。定理 3.3 和定理 3.5 给出了精确的数学刻画，这使我们的工作区别于 [18]。这一几何洞见指导了我们构建曲率驱动的流形拟合方法。

1.2. 当前流形拟合方法。我们介绍当前流形拟合方法中的两种基本技术 [18, 25, 11, 34, 14]。第一种技术（图1.3(a)）将局部主成分分析应用于

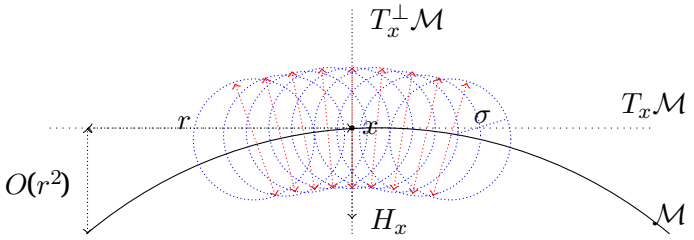


FIGURE 1.2. 该图展示了受各向同性噪声扰动的样本。底层子流形上的点 $x \in \mathcal{M}$ 受到各向同性噪声的偏移 σ （蓝色虚线圆），噪声密度的主要偏差出现在法向方向上（红色虚线）。在 x处， $T_x\mathcal{M}$、 $T_x^{\perp}\mathcal{M}$ 和H_x 分别表示切空间、法空间和平均曲率向量。在尺度 r下， $\mathcal{M}$沿法向方向偏离 $T_x\mathcal{M}$，偏差量级为 $O(r^2)$。

（PCA） [30] 对以观测样本 y为中心、半径为 r 的欧氏球内的样本进行分析。它通过最小化该邻域内所有样本到子空间的距离平方和，确定 d维子空间，$\widehat{\ }$并以此得到估计切空间 $T_y\mathcal{M}$ ，作为投影 $\pi(y)$处真实切空间 $T_{\pi(y)}\mathcal{M}$的近似。这种方法会引入两种内在误差：

$$d(\mathcal{M}, \widehat{T_y\mathcal{M}}) = d(\mathcal{M}, T_{\pi(y)}\mathcal{M}) + d(T_{\pi(y)}\mathcal{M}, \widehat{T_y\mathcal{M}}).$$

首先，用切空间对流形进行局部线性近似会产生量级为 $O(r^2)$ 的曲率诱导的法向偏差，如图 1.2 所示。因此，如何有依据地选取临界尺度r 或 k-近邻（k-NN）的数量 k 至关重要。其次，估计得到的切空间 $T_y\mathcal{M}$位于带噪点 y 处，天然不同于潜在投影点 $\pi(y)$处的切空间 $T_{\pi(y)}\mathcal{M}$ 。这些偏差会传递到后续步骤，例如法空间 [34]和密度函数 [33]的估计。第二种技术（图 1.3(b)）通过对球形或柱形邻域内的样本进行局部平均来降低抽样变异性。该技术通常通过均值漂移[15] 或核密度估计（KDE） [27]实现。带宽 r_i的选择同样至关重要：尺度过小无法抑制方差，尺度过大则会引入几何偏差。如何以理论为依据精确选择带宽，仍然是一项挑战。

接下来，我们详细介绍现有的流形拟合方法。基于德洛内三角剖分的方法假设样本无噪声 [5, 3], ，而基于反卷积的估计量在限制性噪声模型下达到极小极大速率 [16, 17]。随后， [18]提出了基于核密度估计的脊线估计方法。

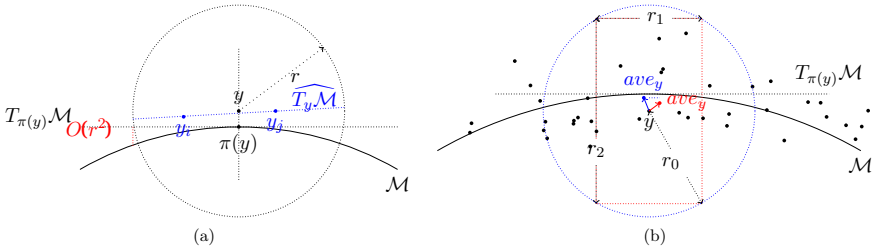


FIGURE 1.3. (a) 局部主成分分析：对于观测点 y，其在 $\mathcal{M}$ 上的投影为 $\pi(y)$，切空间为 $T_{\pi(y)}\mathcal{M}$，而 $\widehat{\ }$ 估计切空间 $T_y\mathcal{M}$由邻近点 y_i, y_j构造；(b) 局部平均：观测样本（黑点）在以 y为中心的球形（蓝色）或圆柱形（红色）邻域内的平均点为ave_y。

该方法 [25] 使用核密度估计和局部主成分分析，根据无噪声样本估计一个近似平方距离函数。后续工作 [11]提出了一种对环境噪声具有鲁棒性的方法 σ。该方法通过局部主成分分析，逐步从含噪样本中估计局部法空间，从而构造近似距离函数。流形估计量定义为该函数的零点集；以高概率，其Hausdorff误差为 $O(\sigma)$，且可达半径有下界（见定义2.1）。

[34]中的方法采用局部样本均值作为底层流形的参照，并通过对样本上的法空间进行局部加权平均来估计法空间。尽管其理论误差仍为$O(\sigma)$，但该技术显著降低了计算复杂度。

在 [14], 中，点细化通过对初始估计值周围超圆柱内的点取平均来降低采样偏差。将其整合到[11] 的框架中，可得到一个Hausdorff误差为 $O(\sigma^2)$ 且可达半径有下界的估计量。然而，该细化方法假设可以获取潜在流形上的无噪声样本，这与噪声观测模型相矛盾。

文献 [33] 提出了一种样本复杂度为 $O(\sigma^{-(d+3)})$ 的两阶段估计器。球形局部均值确定投影方向，进而定义柱形域，并在其中通过第二次局部均值获得流形估计。该估计量以高概率达到 $O(\sigma^2\log(1/\sigma))$ 的豪斯多夫距离，并具有下界受限的可达半径，但尚未实现一致估计。

本文证明，在加性高斯噪声模型下，采用理论推导得到的带宽 $r = c_D\sigma$ 进行局部径向均值漂移，能够有效地将含噪样本投影到底层流形上（图1.4）。该估计量达到最先进误差 $\frac{d}{2}H_{\pi(z)}\sigma^2$，这是由平均曲率引起的二阶偏差。因此，我们的框架所做的不仅是去噪，更是对数据生成几何进行统计反演。

1.3. 主要结果。 我们假设 $\mathcal{M}$是 $\mathbb{R}^D$中的一个紧致、 C^3-光滑 d维子流形，且环境空间 $\mathbb{R}^D$ 配备欧几里得度量 g_E 和 L^2-范数 $\|\cdot\|$。设 $\mathcal{M}$ 具有有限体积 $\mathrm{Vol}(\mathcal{M}) = \int_{\mathcal{M}} d\mu$，且可达半径为正 $\tau > 0$。考虑一个可表示为如下形式的随机向量 $Y \in \mathbb{R}^D$

$$Y = X + \xi, \tag{1.1}$$

其中， $X \in \mathbb{R}^D$ 是一个未观测的随机向量，服从相对于 $\mathcal{M}$上诱导体积测度 $d\mu$ 的均匀分布 ω ，而$\xi \sim \phi_\sigma$ 表示环境空间中的 D维各向同性高斯噪声，即

$$\phi_\sigma(\xi) = (\frac{1}{2\pi\sigma^2})^{\frac{D}{2}} e^{-\frac{\|\xi\|^2}{2\sigma^2}}.$$

Y的密度可视为 ω 与 ϕ_σ的卷积，

$$p_\sigma(y) = \int_{\mathcal{M}} \phi_\sigma(y-x)\omega(x)\, d\mu(x). \tag{1.2}$$

假设观测样本 $\{y_i\}_{i=1}^N \subset \mathbb{R}^D$ 是 Y的 N 个独立同分布实现。令 $\varphi_r : \mathbb{R}^D \to (0,\infty)$为一个光滑、紧支撑的旋转不变核，其带宽为 $r > 0$，例如截断高斯核

$$\varphi_r(z) = \phi_r(z)\chi\big(\tfrac{\|z\|}{\sqrt{2}r}\big) = (2\pi r^2)^{-\frac{D}{2}} e^{\frac{-\|z\|^2}{2r^2}}\, \chi\big(\tfrac{\|z\|}{\sqrt{2}r}\big), \tag{1.3}$$

其中 $\chi:\ [0,\infty) \to [0,1]$ 是满足以下条件的光滑径向函数

$$\chi(t) = 1 \ \text{ for } 0 \le t \le \rho_0 < 1, \qquad \chi(t) = 0 \ \text{ for } t \ge 1.$$

于是 $\varphi_r \in C_0^\infty(B_D(0,\sqrt{2r}))$。令

$$r = c_D\sigma, \quad c_D := \Big(\frac{\int_{B_D(0,1)} \varphi_1(\sqrt{2}u)\, du}{2\int_{B_D(0,1)} (u^{(1)})^2 \varphi_1(\sqrt{2}u)\, du}\Big)^{\frac{1}{2}},$$

然后我们可以构造局部均值漂移估计量 $F\colon \mathbb{R}^D \to \mathbb{R}^D$ 如下

$$F(z) := \frac{\sum_{j=1}^N \varphi_r(y_j - z) y_j}{\sum_{j=1}^N \varphi_r(y_j - z)}. \tag{1.4}$$

由于高斯分布可以近似认为在几个标准差 $C\sigma$范围内趋于消失，因此考虑 $\mathcal{M}$ 的管状邻域是合理的，表示为

$$\Gamma = \{(x,v) \in T^\perp\mathcal{M} : \|v\| \le C\sigma\}.$$

我们假设噪声带宽 σ 相对于流形 $\mathcal{M}$ 的可达半径 τ 足够小。具体而言，假设 $\sigma \le \sigma_0 := \frac{\tau-\varepsilon}{C+\sqrt{2}c_D}$ 对任意固定的 $\varepsilon \in (0,\ \tau)$ 均成立。

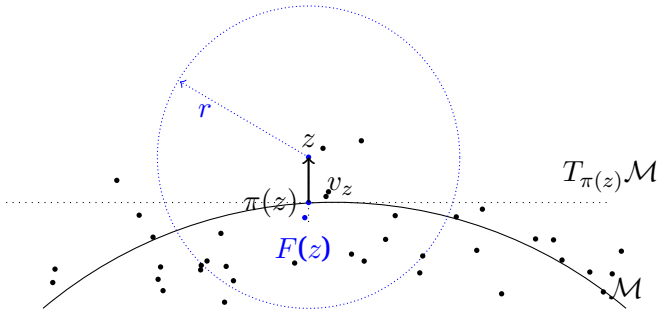


FIGURE 1.4. 构造估计量的示意图。底层子流形 $\mathcal{M}$，在点 z 处的估计量 F 由距离 r 以内的含噪样本（黑点）加权平均得到。令 $\pi(z)$表示 z 在 $\mathcal{M}$上的投影，并定义向量 $v_z = z - \pi(z)$。

定理 1.1. 若样本量 $N = O(\sigma^{-3d-5})$，则估计量 $F(z)$ 满足

$$F(z) = \pi(z) + \tfrac{d}{2} H_{\pi(z)}\sigma^2 + O(\sigma^3), \quad \textit{uniformly for all } z \in \Gamma,$$

以至少 $1-C_1\exp(-C_2\sigma^{-c})$ 的概率，其中 c、C_1、$C_2 > 0$为某些常数，且 $H_{\pi(z)}$ 是投影点 $\pi(z) \in \mathcal{M}$处的平均曲率向量。$\widehat{}$估计得到的 d维流形 $\mathcal{M} = F(\Gamma)$ 与 $\mathcal{M}$ 的豪斯多夫距离为$\frac{d}{2}H_{\pi(z)}\sigma^2$，且其可达半径下界为 $c'\tau$；当 $c' > 0$ 时，这一结论以高概率成立。

注 1.2. (1) 法向方向上的主导偏差为 $O(\sigma^2)$，与平均曲率向量成正比，而切向误差的阶为$O(\sigma^3)$。这反映了平均曲率导致的二次法向偏离；另一方面，由于对称性，切向误差会抵消到更高阶。

(2) 在平均曲率为零处，误差降至 $O(\sigma^3)$；因此，这一结论适用于任意极小子流形（其平均曲率处处为零，例如 Calabi–Yau 流形）。这一结果为极小曲面的几何重要性确立了一个新颖的统计学对应结论。此外，它还改进了 Fef- ferman 等人在 [14]中得到的 $O(\sigma^2)$ 豪斯多夫距离界。

我们的分析还将该估计量置于若干更广泛的理论框架之中。

图上的随机游走。 估计量 F 可视为在由观测点构建的局部图上进行的一步随机游走，其中马尔可夫转移概率由核 φ_r的归一化给出。该过程增强底层几何结构并抑制噪声引起的畸变，因此 $F(z)$的主导项恢复了投影 $\pi(z)$。

扩散过程。 各向同性高斯噪声模型对应于 $\mathbb{R}^D$ 上的一个扩散过程，其热核为 $k_t(x,\ y) = (4\pi t)^{-D/2}e^{-\|x-y\|^2/\ (4t)}$ 和 $\sigma^2 = 2t$。因此，式 (1.2) 中的含噪密度 p_σ 是将内蕴密度 ω 在 $\mathcal{M}$ 上扩散时间 $t = \sigma^2/2$后得到的。与试图逆转这一扩散过程的去噪方案[20] 不同，这种做法通常在数学上并不一致，因为图拉普拉斯算子为正，演化本质上仍是扩散过程。相比之下，我们的估计量直接处理扩散后的分布。它通过增强底层几何结构来抑制噪声，将投影到 $\mathcal{M}$ 上的结果恢复为主导项，而二阶偏差则由 $\mathcal{M}$的平均曲率向量明确决定。

离散平均曲率流。定理 1.1 得出

$$\lim_{\sigma\to 0^+}\frac{F(\pi(z))-\pi(z)}{\sigma^2} = \tfrac{d}{2}H_{\pi(z)}.$$

左侧的算子 $F-id$ 与基于截断归一化权重 φ_r构造的归一化图拉普拉斯算子相同。其极限行为表明，将归一化图拉普拉斯算子作用于坐标函数时，会收敛到流形上的平均曲率向量场。这一结果与扩散映射理论 [7]有所交叠。估计量 F 作用于 $\mathcal{M}$时，相当于在时间尺度 σ^2下进行离散平均曲率流，从而改善数据的几何结构和拓扑结构。在含噪情形下，我们的分析表明，$F-id$ 的主导项能够恢复潜在流形，而偏差项则源于二阶外蕴曲率。

1.4. 主要贡献。我们的方法从根本上不同于以往的流形拟合技术，因为它避开了困扰多阶段估计量的逐级误差传播。基于局部主成分分析 $\widehat{\ }$的传统方法，首先在含噪点 y处构造切空间估计 $T_y\mathcal{M}$ 。这种线性近似必然会产生阶为 $O(r^2)$的法向偏差，其中 r 为局部尺度参数。这种偏差会传播到后续步骤，例如法空间估计 [34], 、构造近似距离函数 [25, 11], ，以及将流形识别为零水平集或脊线 [18]。每个阶段都会放大初始几何失真，导致累积误差难以严格刻画和控制。

相比之下，我们利用 p_σ 在 $\mathcal{M}$上的指数映射显式表示，对数据生成过程进行统计反演。这种直接分析揭示，局部均值漂移中的最先进误差 $O(\sigma^2)$ 是一种精确的几何信号：它与底层流形在投影点 $\pi(z)$处的平均曲率向量 $H_{\pi(z)}$ 成正比。因此，作为外蕴几何核心量的平均曲率，正是从受扰数据的概率结构中显现出来的。

除了这一理论洞见之外，与现有的流形拟合框架相比，我们的方法还具有若干实际优势：

有理论依据的带宽选择。 局部平均半径 $r = c_D\sigma$ 由渐近分析（定理 3.1、定理 3.5）推导得出，其中常数 c_D 明确取决于环境维数 D 和所选的核。这为选择临界尺度参数提供了严谨且有原则的依据。许多现有方法缺乏此类指导；其表现取决于邻域大小的启发式选择（ k-NN 中的 r 或 k ），而这往往需要计算开销较大的交叉验证或调参。我们推导出的带宽不仅确保在所述理论保证下实现最优的偏差-方差权衡，也能转化为计算高效且可复现的流程。

所需参数极少。 该构造只需知道噪声带宽 σ，通常可以从数据中估计得到，或由模型给定。关键的是，它不要求预先知道流形的内在维数 d 。这在实践中很有优势，因为 d 在应用中通常未知，而对其进行估计本身就是一个独立且非平凡的问题。依赖局部主成分分析或 k-NN 图的方法，在构造切空间时本质上需要正确选择 d ，因此容易受到维数设定错误的影响。

显式且非迭代的投影。 估计量 $F(z)$ 通过单次显式的均值漂移步骤逼近流形上的投影。这与迭代方案（例如子空间约束均值漂移或沿密度脊进行梯度上升）形成对比，因为这类方案的收敛性质和不动点可能缺乏清晰的几何解释。它也不同于隐式表示方法，后者将估计流形定义为某个学习函数的零点集。这类隐式表示掩盖了输入点 z 与其投影之间的直接对应关系，使分析和实际应用都更加复杂。

1.5. 文章结构。 本文结构如下。第 2 节介绍几何和概率预备知识。第3节分析总体层面的核估计量 $F(z)$ 的渐近性质。第4节为样本估计量 $F(z)$ 收敛到其总体对应量 μ_z 建立高概率定量界。第5节对输出流形进行几何分析。第6节通过数值实验展示该方法的有效性。最后，第7节汇总所有证明。

2. Preliminaries

本节介绍本文中使用的记号，并回顾全文所用的几何与概率事实。

2.1. 记号。固定整数 d, D，并考虑维数为 $d < D$ 的紧致 C^3 子流形 $\mathcal{M} \subset \mathbb{R}^D$。环境空间 $\mathbb{R}^D$ 配备欧几里得度量 g_E 和 L^2-范数 $\|\cdot\|$。$\mathcal{M}$ 上的诱导黎曼度量记为 $g := g_E|_{\mathcal{M}}$。我们用 Vol($\mathcal{M}$) 表示 $\mathcal{M}$ 关于诱导体积测度 $d\mu$ 的体积，并假设 Vol($\mathcal{M}$) $< \infty$。对于 $m \in \mathbb{N}$、$z \in \mathbb{R}^m$ 和 $r > 0$，我们用 $B_m(z, r)$ 表示 $\mathbb{R}^m$ 中以 z 为中心、半径为 r 的开欧几里得球。

全文中，x 表示潜在流形上的一点 $\mathcal{M}$，y 表示与分布相关的一点 p_σ，z 表示环境空间中的任意一点。一般常数 $C, C_1, C_2, \ldots$ 可能因行而异，除非另有说明，它们仅依赖于流形 $\mathcal{M}$、可达半径 τ，有时也依赖于 ε的选择。特别地，这些常数在任意固定的管状邻域 $T(\tau - \varepsilon)$内均为一致的。

2.2. 子流形的几何量。我们引入若干用于刻画嵌入 $\mathcal{M} \subset \mathbb{R}^D$的几何量。我们用 ∇ 表示 $\mathbb{R}^D$ 中的环境导数（即 g_E的列维-奇维塔联络），用 $\nabla^{\mathcal{M}}$ 表示$(\mathcal{M}, g)$上的列维-奇维塔联络。$x \in \mathcal{M}$ 处的切空间和法空间分别记为 $T_x\mathcal{M}$ 和 $T_x^\perp\mathcal{M}$。切丛和法丛分别记为 $T\mathcal{M}$和 $T^\perp\mathcal{M}$。

$x \in \mathcal{M}$ 处的第二基本形式为 $\mathrm{II}_x\colon T_x\mathcal{M} \times T_x\mathcal{M} \to T_x^\perp\mathcal{M}$，定义在 $\mathcal{M}$ 上的向量场 X, Y 之间，定义如下

$$\mathrm{II}_x(X, Y) = (\nabla_X Y)^\perp,$$

其中 $(\cdot)^\perp$ 表示投影到法空间上的正交投影。这个双线性形式度量了 $\mathcal{M}$ 的外蕴曲率。对于张量 T，我们用 $\|T\|_{\mathrm{op}}$ 表示其算子范数；例如，

$$\|\mathrm{II}_x\|_{\mathrm{op}} := \sup_{\|u\|=\|v\|=1} \|\mathrm{II}_x(u, v)\|, \qquad u, v \in T_x\mathcal{M}.$$

对于每个法向方向 $n \in T_x^\perp\mathcal{M}$，与之对应的形状算子 $S_n : T_x\mathcal{M} \to T_x\mathcal{M}$ 定义如下

$$S_n(X) = -(\nabla_X n)^\top,$$

其中 $(\cdot)^\top$ 表示投影到 $T_x\mathcal{M}$ 上。第二基本形式和形状算子之间满足恒等式

$$\langle S_n(X), Y\rangle = \langle \Pi(X, Y), n\rangle.$$

S_n 的特征值是方向 n上的主曲率，相应的特征向量是主方向。

平均曲率向量 $H_x \in T_x^\perp\mathcal{M}$ 提供了一个衡量整体弯曲程度的指标:

$$H_x = \frac{1}{d}\operatorname{Tr}(\mathrm{II}_x) = \frac{1}{d}\sum_{i=1}^{d} \mathrm{II}_x(e_i, e_i), \tag{2.1}$$

其中， $\{e_i\}$ 是 $T_x\mathcal{M}$的一组标准正交基。特别地， $\langle H,\ n\rangle$ 给出了方向 n 上的平均曲率。在平均曲率流下，流形会按照与其平均曲率成比例的方式演化。极小子流形的平均曲率为零。

全局控制这些曲率量的可达半径 τ of $\mathcal{M}$。

定义2.1 （可达半径）。 设 A是 $\mathbb{R}^D$的一个闭子集。A的可达半径记为 reach(A)，是满足以下条件的最大数值：距离 A小于 τ 的任意点，在 A中都有唯一最近点 τ 。

我们始终假设 $\mathcal{M}$ 具有正可达半径 $\tau > 0$。直观而言，较大的可达半径意味着流形在局部接近其切空间，这一点可由以下引理 [10]说明。

引理 2.2 （Federer拟合半径条件）。 设 $\mathcal{M}$ 是 $\mathbb{R}^D$ 中的一个嵌入子流形，其可达半径为 $\tau > 0$。那么

$$\tau^{-1} = \sup\left\{\frac{2d(b, T_a\mathcal{M})}{\|a-b\|^2} | a, b \in \mathcal{M},\ a \neq b\right\}.$$

正可达半径意味着第二基本形式具有一致上界；结合 $\mathcal{M}$ 的正则性与紧致性，可得曲率张量及其协变导数的一致上界。

引理 2.3 ([26], 命题 6.1)。 若 $\mathcal{M} \subset \mathbb{R}^D$ 具有正可达半径 $\tau > 0$，则

$$\|\mathrm{II}\|_{op} \leq \tau^{-1}. \tag{2.2}$$

特别地，当 $\mathcal{M}$是紧致的且 C^3时，所有曲率张量和第二基本形式都具有有界的协变导数。

这意味着所有主曲率的绝对值均不超过 τ^{-1} ，从而确保流形不会弯曲得过于剧烈。

2.3. **指数映射**。 在 $x \in \mathcal{M}$ 处的指数映射 记为 $\exp_x : T_x\mathcal{M} \to \mathcal{M}$。

$x \in \mathcal{M}$ 处的单射半径 记为 inj(x)，它是满足 $\exp_x : B_d(0, r) \to \mathcal{M}$ 为微分同胚的最大半径 $r > 0$ 。 $\mathcal{M}$ 的全局单射半径定义为 $\mathrm{inj}_{\mathcal{M}} = \inf_{x\in\mathcal{M}} \mathrm{inj}(x)$。若可达半径为正 $\tau > 0$，则单射半径存在下界。

引理 2.4 ([1] 推论 1.4)。 设 $\mathcal{M} \subset \mathbb{R}^D$ 是一个具有正可达半径 $\tau > 0$的 C^2 流形，则

$$inj_{\mathcal{M}} \geq \pi\tau.$$

对于 $r <$inj(x)，指数映射引入了一个指数坐标图 $\left(\exp_x(B_d(0,\ r)),\ \exp_x^{-1}\right)$，它为 x 周围的局部坐标系提供了以下标准展开式。

引理 2.5。对于 $x \in \mathcal{M}$ 和 $u \in T_x\mathcal{M}$，当 $\|u\|$ 足够小时，指数映射满足

$$exp_x(u) = x + u + \frac{1}{2}\Pi_x(u,u) + O(\|u\|^3).$$

其中 Π_x： $T_xM \times T_xM \to T_x^{\perp}M$ 是 x处的第二基本形式。

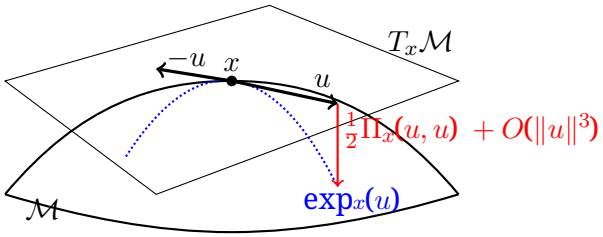


FIGURE 2.1. 指数映射及其泰勒展开。

引理 2.6。对于 $x \in \mathcal{M}$，设 $\{u_i\}_{i=1}^d$ 是 $T_x\mathcal{M}$的一组标准正交基，并用 $g_{ij}(u)$ 表示 x周围指数坐标中的度量分量。那么当 $\|u\| \to 0$时，

$$\sqrt{detg_{ij}(u)} = 1 + O(\|u\|^2).$$

2.4. 豪斯多夫距离与管状邻域。点 a 与非空集合 A 之间的距离表示为 $d(a, A) = \inf_{a' \in A} \|a - a'\|$。我们采用豪斯多夫距离来度量两个集合之间的距离。

定义 2.7（豪斯多夫距离）。设 A_1 和 A_2 是 $\mathbb{R}^D$的两个非空子集。我们将它们的豪斯多夫距离 $\text{dist}_H(A_1, A_2)$ 定义为

$$\text{dist}_H(A_1, A_2) = \max\{\sup_{a \in A_1} \inf_{b \in A_2} \|a - b\|, \sup_{b \in A_2} \inf_{a \in A_1} \|a - b\|\}.$$

对于具有正可达半径 $\tau > 0$的 $\mathcal{M}$，定义其法丛中的开 τ-管状邻域为

$$\mathcal{T}(\tau) := \{(x, v) \in T^{\perp}\mathcal{M} : \|v\| < \tau\}.$$

我们将 $(x, v) \in \mathcal{T}(\tau)$ 与点 $x + v \in \mathbb{R}^D$等同。对于 $0 < \varepsilon < \tau$，我们还引入闭管状邻域

$$T(\varepsilon) := \{(x, v) \in T^{\perp}\mathcal{M} : \|v\| \leq \varepsilon\}.$$

由于 $\mathcal{M}$ 是紧致的且 $\tau > 0$，集合 $T(\varepsilon)$ 是紧致的。

对于非空闭集 $A \subset \mathbb{R}^D$，注意到 A 的最近点是唯一的，我们将相应的投影点记为 $\pi_A(z) := \text{argmin}_{a \in A} \|z - a\|$。默认情况下，只要 $d(z, \mathcal{M}) < \tau$，我们就将到 $\mathcal{M}$ 上的最近点投影记作 $\pi(z)$。

正可达假设保证投影 $\pi\colon \mathcal{T}(\tau) \to \mathcal{M}$定义良好。由于 $\mathcal{M}$ 是 C^3，管状邻域定理蕴含 π 在 $\mathcal{T}(\tau)$ 上是 C^2。特别地，对于每个 $0 < \varepsilon < \tau$ 和 $m = 0, 1$，都存在常数 $B_{\pi,\varepsilon,m} > 0$ ，使得

$$\sup_{y \in T(\varepsilon)} \|\nabla^m \pi(y)\| \le B_{\pi,\varepsilon,m}.$$

因此，法向量 $v_y := y - \pi(y)$ 属于 C^2 y，其导数∇v_y 在 $T(\tau - \varepsilon)$ 上一致有界。

我们将反复使用一个适用于管状邻域上张量场的简单 Lipschitz 型估计。

引理 2.8。 设 $\Gamma \subset T(\tau)$ 为紧致集，且 S 是定义在 Γ 的一个开邻域上的 C^{m+1} 张量场。那么 $\nabla^m S$ 在 Γ 上是 *Lipschitz* 的：存在一个常数 $C > 0$, 仅依赖于 m 和 $\nabla^m S$ 的上界，使得对所有 x、$y \in \Gamma$,

$$\left\|\nabla^m S(y) - \nabla^m S(x)\right\|_{\mathrm{op}} \le C\|y - x\|.$$

特别地，取 $\Gamma = T(\tau - \varepsilon)$ 和 $x = \pi(y)$ ，我们得到

$$\left\|\nabla^m S(y) - \nabla^m S(\pi(y))\right\|_{\mathrm{op}} \le C\|y - \pi(y)\| = C\|v_y\|, \quad y \in T(\tau - \varepsilon).$$

2.5. 概率预备知识。 我们首先概述一下：概率测度记作 $\mathbb{P}$ （表示一般的分布），支撑在 $\mathcal{M}$ 上的分布记作 $\mathbb{P}_{\mathcal{M}}$ ，噪声水平为 $\sigma > 0$ 的含噪分布记作 p_σ 。期望记作 $\mathbb{E}(\cdot)$。

回顾一下，对随机变量 X 的通用 Chernoff 界可通过对 e^{tX} 应用 Markov 不等式得到。

引理 2.9 (Chernoff 界 [6])。 若 X 是随机变量，则对每个$t > 0$

$$\mathbb{P}(X \ge a) = \mathbb{P}\left(e^{tX} \ge e^{ta}\right) \le \frac{\mathbb{E}\left(e^{tX}\right)}{e^{ta}}.$$

请注意，Hoeffding 不等式给出了独立随机变量的平均值偏离其期望值的概率上界。

引理 2.10 (Hoeffding 不等式 [21])。 设 $X_1, \ldots, X_n$ 是相互独立的实值随机变量，并且对每个 i 都满足 $a_i \le X_i \le b_i$ 。令 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$。则对每个 $t > 0$,

$$\mathbb{P}\big(\bar{X}_n - \mathbb{E}(\bar{X}_n) \ge t\big) \le \exp\Big(-\frac{2n^2 t^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big).$$

特别地，如果对于所有 i， $b_i - a_i \le L$ ，那么

$$\mathbb{P}\big(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| \ge t\big) \le 2\exp\Big(-\frac{2nt^2}{L^2}\Big).$$

因此，我们证明以下引理，以控制独立随机向量的有界加权平均与其期望之间的偏差。

引理 2.11. 设 $\{Y_i\}_{i=1}^n$ 是 $\mathbb{R}^D$ 中的独立同分布随机向量，并设 $W:\mathbb{R}^D\to[0,\infty)$ 是一个可测权重函数。定义

$$S_n := \frac{1}{n}\sum_{i=1}^n W(Y_i)Y_i, \qquad B_n := \frac{1}{n}\sum_{i=1}^n W(Y_i),$$

并令 $s := \mathbb{E}(W(Y)Y)$, $b := \mathbb{E}(W(Y)) > 0$, $\widehat{\mu}_n := S_n/B_n$ 和 $\widehat{\mu}_w = s/b$。假设

(1) $\{Y_i : W(Y_i) > 0\} \subset B_D(z, r)$ 对所有 i 以及某些 $z\in\mathbb{R}^D$ 和 $r>0$ 均成立；*(2)* $0 \le W(Y_i) \le M$ 对所有 i 均成立；*(3)* 存在 $K>0$，使得对每个坐标 $j=1$、……、D 以及每个 i，

$$\left|(W(Y_i)Y_i)^{(j)} - \mathbb{E}((W(Y)Y)^{(j)})\right| \le K.$$

那么，对于任意 $\varepsilon>0$，我们得到

(2.3)

$$\mathbb{P}\big(\|\widehat{\mu}_n-\widehat{\mu}_w\|>\varepsilon\big) \le 2\exp\Big(-\frac{nb^2}{2M^2}\Big)+2D\exp\Big(-\frac{nb^2\varepsilon^2}{32DK^2}\Big)+2\exp\Big(-\frac{nb^4\varepsilon^2}{8\|s\|^2M^2}\Big).$$

3. Normalized local kernel estimator

本节分析总体层面局部核估计量 (1.4) 的渐近性质。我们推导总体密度 p_σ 以及相应得分函数 $\nabla\log p_\sigma(z)$ 的展开式，它们编码了底层流形的曲率信息。最后，我们证明总体级估计量得到的估计流形与底层流形足够接近。

回顾一下，对于任意 $r>0$，令 $\varphi_r:\mathbb{R}^D\to(0,\infty)$ 为一个具有紧致支集的光滑旋转不变核，其带宽为 r，例如截断高斯核

$$\varphi_r(z) = \phi_r(z)\chi\big(\tfrac{\|z\|}{\sqrt{2}r}\big) = (2\pi r^2)^{-\frac{D}{2}} e^{\frac{-\|z\|^2}{2r^2}}\,\chi\big(\tfrac{\|z\|}{\sqrt{2}r}\big), \tag{3.1}$$

其中 $\chi:[0,\infty)\to[0,1]$ 是满足以下条件的光滑截断函数

$$\chi(t)=1 \ \text{ for } 0\le t\le\rho_0<1, \qquad \chi(t)=0 \ \text{ for } t\ge 1.$$

然后 $\varphi_r\in C_0^\infty(B_D(0,\sqrt{2}r))$。根据式 (1.1) 中定义的观测样本 $\{y_i\}_{i=1}^N$，我们可以构造局部核估计量 $F:\mathbb{R}^D\to\mathbb{R}^D$，其形式为

$$F(z) := \frac{\sum_{j=1}^N \varphi_r(y_j-z)y_j}{\sum_{j=1}^N \varphi_r(y_j-z)}. \tag{3.2}$$

3.1. 总体级估计量的渐近分析。随着样本量增加，样本级归一化估计量收敛到其总体级估计量

$$\mu_z := \frac{\mathbb{E}(\varphi_r(Y-z)Y)}{\mathbb{E}(\varphi_r(Y-z))} = \frac{\int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, y\, p_\sigma(y)\, dy}{\int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, p_\sigma(y)\, dy}$$

以高概率成立，详见第4节。

首先，我们考察其期望关于截断半径 r 的渐近行为。

定理 3.1. 对于光滑密度 p_σ： $\mathbb{R}^D \to \mathbb{R}$，估计量 F： $\mathbb{R}^D \to \mathbb{R}^D$ （定义见 (3.2) ）满足如下渐近关系：

$$\mu_z = \frac{\int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, y\, p_\sigma(y)\, dy}{\int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, p_\sigma(y)\, dy} = z + r^2\, \frac{2B\, \nabla p_\sigma(z)}{A\, p_\sigma(z)} + O(r^4), \tag{3.3}$$

其中 $A = \int_{B_D(0,1)} \varphi_1(\sqrt{2}u)\, du$ 和 $B = \int_{B_D(0,1)} (u^{(1)})^2 \varphi_1(\sqrt{2}u)\, du$ 是依赖于维数 D 的常数。

注3.2。 根据噪声的带宽 σ ，取 $r = c_D\sigma = (\frac{A}{2B})^{\frac{1}{2}}\sigma$，我们得到

$$\mu_z = z + \sigma^2\, \nabla \log p_\sigma(z) + O(\sigma^4).$$

主导误差项由得分函数 $\nabla \log p_\sigma(z)$ 决定。

证明的关键在于对密度函数作局部泰勒展开，并结合局部核和积分区域的径向不变性。详情见第7节。

3.2. 流形上噪声样本的密度。 回顾一下， $\mathcal{M} \subset \mathbb{R}^D$ 是一个紧致的、 C^3-光滑的 d维子流形，并赋有诱导体积测度 $d\mu$。其体积为 $\mathrm{Vol}(\mathcal{M}) = \int_{\mathcal{M}} d\mu < \infty$。我们考虑 $\mathcal{M}$上的均匀概率分布，其密度为

$$\omega(x) = \frac{1}{\mathrm{Vol}(\mathcal{M})},\ x \in \mathcal{M}.$$

添加各向同性高斯噪声 $\phi_\sigma \sim \mathcal{N}(0,\ \sigma^2 I_D)$ 后，会在 $\mathbb{R}^{\mathbb{D}}$ 上得到一个概率分布，其定义为卷积

$$p_\sigma(y) = \int_{\mathcal{M}} \phi_\sigma(y-x)\omega(x)d\mu(x) = \frac{1}{\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{\frac{D}{2}}} \int_{\mathcal{M}} e^{-\frac{\|y-x\|^2}{2\sigma^2}} d\mu(x).$$

$\mathcal{M}$ 的半径为 $\tau - \varepsilon > 0$ 的闭管状邻域为

$$T(\tau-\varepsilon) = \{(x,v) \in T^\perp\mathcal{M} \mid \|v\| \le \tau - \varepsilon\}.$$

对于点 $y \in T(\tau)$，令 $\pi(y)$ 表示其在 $\mathcal{M}$ 上的投影，$v_y = y - \pi(y)$ 表示相应的法向位移。下述定理给出了 $p_\sigma(y)$ 关于 $\|v_y\|$和 σ的展开。

定理 3.3。设 $\mathcal{M} \subset \mathbb{R}^D$ 是一个紧致的 C^3 维数为 d、可达半径为 $\tau > 0$的子流形。对于任意 $\varepsilon \in (0, \tau)$，存在常数 $\sigma_0 > 0$，使得对于所有 $y \in T(\tau - \varepsilon)$ 和 $0 < \sigma \leq \sigma_0$,

$$p_\sigma(y) = \frac{1}{\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{\frac{D-d}{2}}} e^{-\frac{\|v_y\|^2}{2\sigma^2}} \frac{1}{\sqrt{\det A_y}} \left(1 + O(\sigma\|v_y\| + \sigma^2)\right), \tag{3.4}$$

其中 $A_y = I_d - \langle v_y, \Pi_{\pi(y)} \rangle$ 和 $\Pi_{\pi(y)}$ 是 $\mathcal{M}$在投影点 $\pi(y) \in \mathcal{M}$处的第二基本形式。$O(\cdot)$ 项关于 $y \in T(\tau - \varepsilon)$ 和$\sigma \in (0, \sigma_0]$一致。

证明基于指数映射和指数坐标中体积元在 $\pi(y)$附近的泰勒展开（引理 2.5 和 2.6）。含噪密度可分解为法向方向上的高斯项 $\exp\Big(-\frac{\|v_y\|^2}{2\sigma^2}\Big)$和依赖曲率的修正项 $(\det A_y)^{-1/2}$，余项阶为 $O(\sigma\|v_y\| + \sigma^2)$。因此， p_σ 的局部行为由法向距离 $\|v_y\|$ 和投影点处的第二基本形式 $\Pi_{\pi(y)}$ 共同决定。

为了分析导数，将定理 3.3 改写为对数密度的形式，并给出余项及其导数的一致界，较为方便。

推论 3.3.1。假设 $\mathcal{M}$ 是一个紧致的、正可达半径为 $\tau > 0$ 的 C^3 子流形。对于任意 $\varepsilon \in (0, \tau)$，存在常数 $\sigma_0 > 0$ 和 $C_{\varepsilon,m} > 0 (m = 0, 1)$，使得对所有 $y \in T($ $\tau - \varepsilon)$ 和 $0 < \sigma \leq \sigma_0$,

$$\log p_\sigma(y) = \log\big(\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{\frac{D-d}{2}}\big)^{-1} - \frac{\|v_y\|^2}{2\sigma^2} - \frac{1}{2}\log\det A_y + R(y, \sigma), \tag{3.5}$$

其中 R 满足

$$|R(y, \sigma)| \leq C_{\varepsilon,0}\big(\sigma\|v_y\| + \sigma^2\big).$$

并且 R 关于 y 的环境导数满足

$$\|\nabla_y R(y, \sigma)\|_{op} \leq C_{\varepsilon,1}\sigma.$$

注意，对数密度可展开为余项为 $O(\sigma\|v_y\|+\sigma^2)$ 的形式，而关于 y 的微分仅能保持 $O(\sigma)$ 的阶。

3.3. 得分展开。接下来，我们要推导得分函数的展开式

$$s_\sigma(y) = \nabla \log p_\sigma(y), \qquad y \in T(\tau - \varepsilon).$$

我们首先记录 $\pi(y)$ 和 v_y 关于环境坐标 y 的导数。令 $\nabla_\omega(\cdot)$ 表示沿方向 $\omega \in R^D$ 的欧几里得方向导数，并令 $\omega^\top$ 表示将 ω 正交投影到 $T_{\pi(y)}M$ 上。

引理 3.4。设 $\mathcal{M} \subset \mathbb{R}^D$ 是一个紧致的、维数为 d且可达半径为 $\tau > 0$的 C^3 子流形。对于任意 $y \in \mathcal{T}(\tau)$， $\omega \in \mathbb{R}^D$， $\pi(y)$ 的导数以及v_y 满足

$$\nabla_\omega \pi(y) = A_y^{-1}\omega^\top, \qquad \nabla_\omega v_y = \omega - A_y^{-1}\omega^\top.$$

证明利用第二基本形式对 A_y 的刻画。结合引理 3.4 与推论 3.3.1 中的对数密度展开，可得得分 $\nabla \log p_\sigma(y)$ 的如下表达式。

定理 3.5。设 $\mathcal{M} \subset \mathbb{R}^D$ 是一个具有 C^3、维数为 d 且可达半径为 $\tau > 0$ 的紧致子流形。对于任意 $\varepsilon \in (0,\ \tau)$，存在常数 $\sigma_0 > 0$，使得对于所有 $y \in T(\tau - \varepsilon)$ 以及 $0 < \sigma \leq \sigma_0$,

$$s_\sigma(y) := \nabla \log p_\sigma(y) = -\frac{v_y}{\sigma^2} + \frac{d}{2} H_{\pi(y)} + O\big(\|v_y\| + \sigma\|v_y\| + \sigma^2\big), \tag{3.6}$$

其中，$H_{\pi(y)} = \frac{1}{d}\operatorname{Tr}\big(\Pi_{\pi(y)}\big)$ 表示 $\pi(y) \in \mathcal{M}$ 处的平均曲率。$O(\cdot)$ 项关于 $y \in T(\tau - \varepsilon)$ 和 $\sigma \in (0,\ \sigma_0]$ 是一致的。

注 3.6。(1) 首项 $-v_y/\sigma^2$ 由高斯噪声引起，并沿法向方向指向流形。项 $\frac{d}{2}H_{\pi(y)}$ 则体现了平均曲率的贡献，该贡献源自 p_σ 中 $\langle v, \Pi\rangle$ 的行列式。余项 $O(\|v_y\| + \|v_y\|\sigma + \sigma^2)$ 在 y 趋近于 $\mathcal{M}$ 且 σ 趋于零时消失。

(2) 对于管状邻域内的点，在噪声足够小时，$s_\sigma(y)$ 的切向分量相较于其法向分量可以忽略。这表明，在 $T(\tau - \varepsilon)$ 中，得分场本质上是一个法向量场。

(3) 由 (3.6) 可知，满足 $v_y = \sigma^2 \frac{d}{2} H_{\pi(y)}$ 的点 y 对应于沿法向方向的密度局部极大值。这为脊理论 [18] 中的脊集提供了精确的几何刻画。

3.4. **总体层面的输出流形**。接下来，结合定理 3.1 和定理 3.5 中得分 $\nabla \log p_\sigma(z)$ 的展开，我们可以证明，总体级估计量将输出一个与底层流形 $\mathcal{M}$ 的 $O(\sigma^2)$距离为.....的流形。回顾一下，两个非空子集 $A_1,\ A_2 \subset \mathbb{R}^D$ 之间的豪斯多夫距离定义为

$$\operatorname{dist}_H(A_1, A_2) = \max\{\sup_{a\in A_1}\inf_{b\in A_2}\|a-b\|,\ \sup_{b\in A_2}\inf_{a\in A_1}\|a-b\|\}.$$

定理 3.7。设 $\mathcal{M} \subset \mathbb{R}^D$ 是一个紧致的、维数为 d 且可达半径为 $\tau > 0$ 的子流形 C^3。对于任意 $\varepsilon \in (0, \tau)$，存在常数 $\sigma_0 > 0$，使得对于 (3.3) 中定义的总体映射 $\mu\colon T(\tau-\varepsilon) \to \mathbb{R}^D$ 以及所有 $0 < \sigma \leq \sigma_0$,

$$\mu_z = \pi(z) + \frac{d}{2} H_{\pi(z)}\sigma^2 + O(\sigma^2), \tag{3.7}$$

其中，$H_{\pi(z)}$ 表示 $\pi(z) \in \mathcal{M}$处的平均曲率。此外，输出流形

$$\mu(T(\tau-\varepsilon)) = \{\mu_z :\ z \in T(\tau-\varepsilon)\}$$

满足

$$\operatorname{dist}_H(\mu(T(\tau-\varepsilon)), \mathcal{M}) \leq C\sigma^2.$$

注 3.8。 尽管原始流形 $\mathcal{M}$仅为 C^3，总体级估计量 $z \mapsto \mu_z$ 继承了核 φ_r 的无限光滑性，且像 $\mu(T(\tau-\varepsilon))$ 是一个 C^∞ 流形。

4. Sample-level approximation and sample complexity

本节将给出定量界，确保样本级估计量 $F(z)$ 以高概率收敛到其总体对应量 μ_z。利用有界随机变量的集中不等式，我们确定样本量尺度，使随机偏差相较于总体级分析中得到的确定性偏差 $O(\sigma^2)$ 可以忽略不计。

回顾一下， $\mathcal{M}$ 是一个紧致的、 C^3-光滑的 d 维子流形，位于 $\mathbb{R}^D$ 中，具有有限体积 $\mathrm{Vol}(\mathcal{M})$ 和正可达半径 τ。基于式 (1.1) 中定义的高斯噪声样本 $\mathcal{Y} = \{y_i\}_{i=1}^N$，我们构造一个局部核估计量

$$F(z) := \frac{\sum_{j=1}^N \varphi_r(y_j - z) y_j}{\sum_{j=1}^N \varphi_r(y_j - z)},\ r := c_D \sigma,$$

其中 $\varphi_r \in C_c^\infty(B_D(0,\ \sqrt{2}r))$。令 $\mathcal{Y}_z = \{y_i\}_{i=1}^n (n \le N)$ 表示在 $F(z)$ 中起作用的样本，即落在局部球 $B_D(z,\ \sqrt{2}r)$ 内的样本。并将其总体级估计量记为

$$\mu_z := \frac{\mathbb{E}(\varphi_r(Y-z)Y)}{\mathbb{E}(\varphi_r(Y-z))}.$$

我们始终在 $\mathcal{M}$ as 的闭管状邻域内进行研究

$$\Gamma = \{(x,v) \in T^\perp \mathcal{M} : \|v\| \le C\sigma\}.$$

假设噪声带宽 σ 相对于 $\mathcal{M}$ 的可达半径 τ 足够小。具体而言，设 $\sigma \le \sigma_0 := \frac{\tau-\varepsilon}{C+\sqrt{2}c_D}$ 对任意固定的 $\varepsilon \in (0,\tau)$，这确保了 $B_D(z,\sqrt{2}r) \subset T(\tau-\varepsilon)$，从而能够在第3节中进行一致估计。

4.1. 局部样本复杂度。 由于估计量 F 中的权重由截断核所界定，我们可以应用引理 2.11 得到以下定理。该定理表明，局部样本量 $n \ge O(\sigma^{-2d-5})$ 以高概率控制样本级估计量与总体级估计量之间的偏差。

定理 4.1。 对于任意 $z \in \Gamma$，设 $\mathcal{Y}_z = \{y_i\}_{i=1}^n$ 为在局部球 $B_D(z,\sqrt{2}r)$ 中观测到的含噪点，且满足 $r = c_D\sigma$ 和 $0 < \sigma \le \sigma_0$。那么对于 $n \ge O(\sigma^{-2d-5})$，有

$$\mathbb{P}(\|F(z) - \mu_z\| \le C_3\sigma^3) \ge 1 - C_1 \exp(-C_2\sigma^{-1}),$$

对于某些常数 C_1、 C_2 和 $C_3 > 0$。

4.2. 总样本复杂度。接下来，通过分析密度函数p_σ ，我们估计观测值 $Y \sim p_\sigma$ （如式 (1.2) 所定义）落入局部球 $B_D(z, \sqrt{2}r)$ 的概率，并进一步推导所需的总样本量 N。

引理 4.2. 对于任意满足 $B_D(z,\sqrt{2}r) \subset T(\tau-\varepsilon)$ 且满足 $r = c_D\sigma$和 $0 < \sigma \le \sigma_0$ 的 $z \in \Gamma$，可得

$$\mathbb{P}(Y \in B_D(z, \sqrt{2}r)) = c\, r^d (1 + O(r)), \tag{4.1}$$

其中， Y 是 (1.1) 中定义的观测值。

推论 4.2.1. 设 n 为落在 $B_D(z,\sqrt{2}r)$ 中的观测点 Y 的数量，其中这些观测点在 (1.1) 中定义，且 $z \in \overline{\Gamma}$。假设 $r = c_D\sigma$ 和 $0 < \sigma \le \sigma_0$。如果总样本量满足 $N \ge O(\sigma^{-5-3d})$，则

$$\mathbb{P}(n \ge C_1\sigma^{-5-2d}) \ge 1 - 2\exp(-C_2\sigma^{-5-2d})$$

对于某些常数 C_1、 $C_2 > 0$。

5. Sample-level estimator and output manifold

对于位于紧致 C^3 子流形 $\mathcal{M} \subset \mathbb{R}^D$ 上的数据点，在按（1.1）中定义的方式受到高斯噪声 $\mathcal{Y} = \{y_i\}_{i=1}^N$ 影响时，本节将建立（1.4）中定义的样本级估计器 F 的渐近性质。考虑管状邻域

$$\Gamma = \{(x, v) \in T^\perp\mathcal{M} : \|v\| \le C\sigma\}.$$

假设 $\sigma \le \sigma_0 := \frac{\tau-\varepsilon}{C+\sqrt{2}c_D}$对任意固定的 $\varepsilon \in (0, \tau)$ 和 $r = c_D\sigma$均成立，这保证了 $B_D(z, \sqrt{2}r) \subset T(\tau-\varepsilon)$。我们证明，当样本量为 $N \ge O(\sigma^{-3d-5})$ 时，估计器 F 能够有效地对 Γ中的点去噪，并将其映射到底层流形 $\mathcal{M}$附近。此外，我们还可以刻画输出流形 $F(\Gamma)$的几何性质。

5.1. 样本级渐近一致性。以下定理给出了样本量为$N \ge O(\sigma^{-3d-5})$时，估计器 $F(z)$ 在 Γ 上的逐点估计。

定理 5.1。 若样本量满足 $N \ge O(\sigma^{-3d-5})$，则估计量 $F(z)$ 对任意 $z \in \Gamma$ （定义见(1.4) ）均满足

$$F(z) = \pi(z) + \tfrac{d}{2} H_{\pi(z)}\sigma^2 + O(\sigma^3)$$

以至少 $1 - C_1\exp(-C_2\sigma^{-1})$ 的概率，其中 C_1、 $C_2 > 0$为某些常数，且此处 $H_{\pi(z)}$ 是 $\pi(z) \in \mathcal{M}$处的平均曲率。

备注 5.2。 上述定理表明，估计精度本质上受流形几何结构的支配。特别是，如果 $H_{\pi(z)}$ 在 $\pi(z) \in \mathcal{M}$ 处的平均曲率为零，则估计误差可降低至 $O(\sigma^3)$。

推论 5.2.1。 在定理 *5.1* 的假设下，设 $\mathcal{Y}^o = \{y_i\}_{i=1}^n := \mathcal{Y} \cap \Gamma$ 是 Γ 中的所有观测数据点 $\mathcal{Y}$。那么，在 (1.4) 中定义的估计量 $F(z)$ 满足

$$F(z) = \pi(z) + \tfrac{d}{2} H_{\pi(z)} \sigma^2 + O(\sigma^3) \qquad uniformly\ for\ all\ z \in \mathcal{Y}^o,$$

以至少 $1 - C_1' \exp(-C_2' \sigma^{-c'})$ 的概率，其中 c'、C_1'、C_2' >0为某些常数。

推论 5.2.2。 在定理 *5.1* 的假设下，我们得到

$$\lim_{\sigma \to 0^+} \frac{F(\pi(z)) - \pi(z)}{\sigma^2} = \tfrac{d}{2} H_{\pi(z)}.$$

注 5.3。 (1) 左侧的算子 $F - id$ 可视为一个归一化图拉普拉斯算子，它由带宽为 $r = c_D \sigma$的截断核权重构造而成。极限行为表明，将归一化图拉普拉斯算子作用于坐标函数时，它会收敛到流形上的平均曲率向量场。这一结果与扩散映射理论有所重叠 [7]。

(2) 估计量 F 作用于 $\mathcal{M}$，相当于在 σ^2的时间尺度上执行离散平均曲率流，从而内在地增强数据的几何和拓扑结构。

5.2. **输出流形**。 现在我们将证明，光滑流形的小扰动能够保持其几何正则性，即扰动后流形的可达半径仍接近原流形的可达半径。

引理 5.4。 设 $\mathcal{M} \subset \mathbb{R}^n$ 是一个紧致的 C^3 d 维子流形，其可达半径为 $\tau > 0$, 且 $f:\ \mathcal{M} \to \mathbb{R}^n$ 是一个 C^1 法向量场

$$f(x) = \frac{d}{2} H_x \sigma^2 + O(\sigma^3),$$

其中 H_x 是 $\mathcal{M}$ 在 x处的平均曲率向量。则对于扰动后的流形

$$\widehat{\mathcal{M}} = \{x + f(x) \mid x \in \mathcal{M}\},$$

存在一个仅依赖于 $\mathcal{M}$ 的常数 $c > 0$ ，使得当 σ 充分小时，

$$reach(\widehat{\mathcal{M}}) \geq \tau(1 - c\sigma^2) > 0.$$

接下来，我们可以证明，基于观测点构造的估计量 F 能够输出一个估计流形 $F(\Gamma)$，该流形满足正则性条件，且与原流形足够接近。

定理 5.5。 设 $F:\ \Gamma \to \mathbb{R}^D$ 是如 (1.4) 中定义的估计量。若样本量满足 $N \geq O(\sigma^{-3d-5})$，则 $F(\Gamma)$ 是一个具有正可达半径的光滑 d 维流形，并且满足

$$\mathrm{dist}_H(F(\Gamma), \mathcal{M}) \leq C' \sigma^2$$

以至少 $1 - C_1' \exp(-C_2' \sigma^{-c'})$ 的概率，其中 C'、C_1'、C_2'、$c' > 0$ 为常数。

证明对 Γ 使用标准的有限覆盖论证，以高概率说明估计流形 $F(\Gamma)$ 的几何性质。

6. Numerical experiments

本节评估离散估计器 F 在模拟数据上进行流形拟合的性能，实验设置与 [25, 33]类似。数据生成和重建流形的评估过程如下。

首先，从 d维子流形 $\mathcal{M} \subset \mathbb{R}^D$ 上按照其诱导体积测度均匀生成一个独立同分布样本$\{x_i\}_{i=1}^N$。随后添加独立高斯噪声，得到观测数据$\{y_i\}_{i=1}^N \subset \mathbb{R}^D$，其形式为

$$y_i = x_i + \xi_i, \qquad \xi_i \sim \mathcal{N}(0, \sigma^2 I_D).$$

随后，我们在 $\mathcal{M}$的一个固定管状邻域内独立生成一组测试点$\{z_i\}_{i=1}^{N_0}$，使得 $\sigma/2 \le d(z_i,\ \mathcal{M}) \le 2\sigma$（这些点可能包含观测样本点，并用于评估算法）。估计器 F 对这些测试点的估计通过以下算法1进行。

算法1yl25：流形拟合估计器。

输入： 测试点 $\{z_i\}_{i=1}^{N_0}$，观测样本 $\{y_i\}_{i=1}^{N}$，环境维数D，噪声带宽 σ。

输出： 流形 $\mathcal{M}$ 上的估计点 $\{F(z_i)\}_{i=1}^{N_0}$。

1. 计算截断半径 $r = {}^{D\gamma}_{(D/2,\ 1)}\ 2\gamma(D/2+1,\ 1)\sigma$. 2. 对每个 $z \in \{z_i\}_{i=1}^{N_0}$: – 确定指标集 $I_z = \{j : \|y_j - z\|^2$ $d(z,\ y_j) \le \sqrt{2}\, r\}$。– 计算所有 $j \in I_z$ 的高斯权重 $w_j = e^{\quad 2r^2}$ 。– 获得估计点 $F(z) = \frac{\sum_{j\in I_z} w_j y_j}{\sum_{j\in I_z} w_j}$。

算法1需要将噪声带宽 σ 作为输入。在第 6.1–6.3 节中，我们使用模型中的真实带宽，以验证定理 5.1 预测的渐近行为。实际应用中，可以根据含噪观测 [23, 24] 估计 σ 。算法1的计算复杂度为时间复杂度 $O(N_0 N D)$、空间复杂度 $O((N_0 + N)D)$。性能评估基于上确界逼近误差，定义为 $\max_j d(F(z_j)\ \mathcal{M})$，用于估计与真实流形 $\mathcal{M}$ 之间的豪斯多夫距离，同时也评估计算所需的 CPU 时间。

6.1. 渐近分析。下文中，我们采用豪斯多夫距离作为指标，验证定理 5.1 中针对噪声带宽 σ、样本量 N 和平均曲率 H建立的渐近关系。

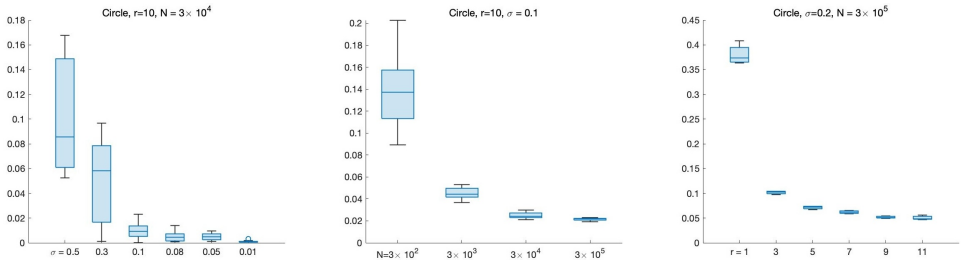


FIGURE 6.1. yl25 关于噪声带宽 σ、样本量 N 和平均曲率 H 的渐近性能。

对于半径为 r 的圆，我们固定 $N = 3 \times 10^4$、$N_0 = 100$和 $r = 10$，同时改变噪声水平 $\sigma \in \{0.5, 0.3, 0.1, 0.08, 0.05, 0.01\}$。对于每个 σ，我们随机生成 50 组不同的数据集 $\{z_i\}_{i=1}^{N_0}$，并对每组数据应用算法 yl25。重建流形与底层流形之间的豪斯多夫距离显示在图 6.1 左图中。结果表明，随着 σ 减小，豪斯多夫距离呈二次衰减，这与定理 5.1 所述的渐近行为一致。接下来，在固定 $\sigma = 0.1$ 和 $r = 10$的情况下，我们评估了 yl25 在不同样本量 $N \in \{3\times10^2, 3\times10^3, 3\times10^4, 3\times10^5\}$下的性能。相应的豪斯多夫距离显示在图 6.1 中图中。随着 N 增加，豪斯多夫距离显著减小，这可归因于对局部几何结构的估计更加准确。最后，在设定 $N = 3 \times 10^5$ 和 $\sigma = 0.2$ 后，我们考察了豪斯多夫距离随平均曲率
$\kappa = r^{-1} \in \{1^{-1}, 3^{-1}, 5^{-1}, 7^{-1}, 9^{-1}, 11^{-1}\}$的变化情况。图 6.1 右图所示结果表明，随着曲率减小，豪斯多夫距离略有下降。

6.2. 与其他流形拟合方法的比较。对于半径为 r 的圆，我们设置 $N = 3000$、$N_0 = 300$、$r = 5$和 $\sigma = 0.1$。各估计量的参数选择如下：对于 ysl23 [33], ，我们设置 $r_0 = r_1 = 5\sigma/\lg(N)$ 和 $r_2 = 10\sigma\sqrt{\log(1/\sigma)/\lg(N)}$；对于 yx19 [34], cf18[11], 和 km17 [25], ，则按照其论文中的方法将参数选为 $r = 2\sqrt{\sigma}$ 。这表明，这些方法中的局部邻域参数通常需要通过经验调优，而我们的方法则依据理论原则来选择截断半径。

每种方法均进行了 10 次独立试验，其性能汇总于图 6.2。从豪斯多夫距离来看，yl25 和 yx19 的表现略优于 ysl23 和 cf18，而这四种方法

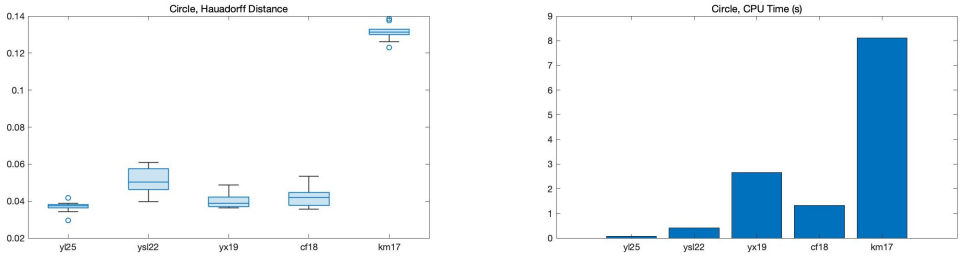


FIGURE 6.2. 使用 yl25、ysl23、yx19、cf18 和 km17 进行圆拟合的豪斯多夫距离和 CPU 时间（$N = 3000$、 $r = 5$、 $\sigma = 0.1$）。

均显著优于 km17。在计算效率方面，yl25 也展现出优势，其运行时间明显短于另外四种方法。

6.3. 其他流形的拟合。 我们展示了在其他非平凡流形上的拟合性能。对于环面，我们设置 $N = 2.5 \times 10^4$、 $N_0 = 100$，并进行 10 次独立运行。图 6.3 左侧和中间的面板将观测点云与我们的方法 yl25 得到的拟合点进行比较，右侧面板则展示了豪斯多夫距离随噪声水平降低而单调下降的情况。

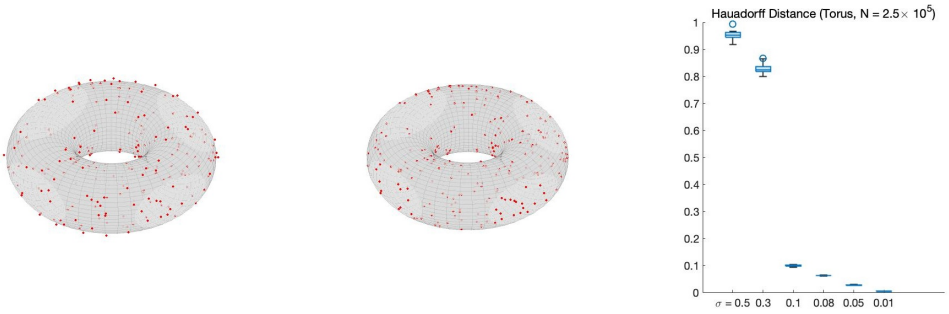


FIGURE 6.3. yl25 在环面上的拟合性能（N =2.5×10^4）。

对于 Calabi–Yau 流形，也观察到了类似的结果和趋势——这类紧致的 Ricci 平坦 K¨ahler 流形在理论物理中尤为重要 [4]。我们以费马四次曲面为具体示例，它可化简为局部坐标下的规范化方程 $x^4 + y^4 = 1, x, y \in \mathbb{C}$。通过降维和适当的投影，可在三维空间中可视化所得的四维曲面。图 6.4 展示了其投影形态。

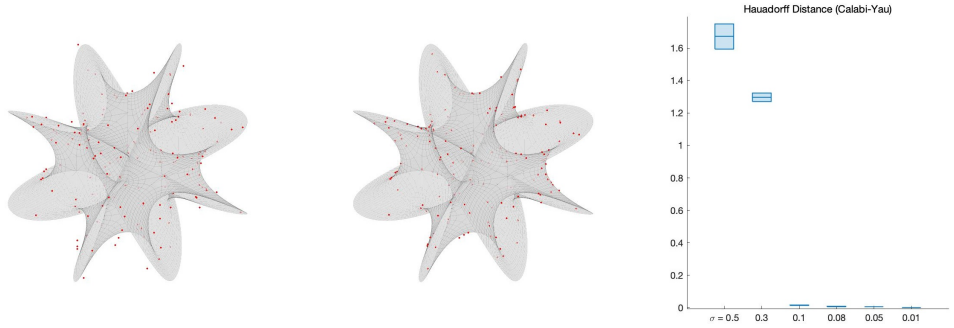


FIGURE 6.4.yl25 在 Calabi–Yau 流形（$N = 3 \times 10^5$）上的拟合性能。

数据点均匀采样，并生成了 3×10^5 个样本$\mathbb{R}^4$，其中添加了高斯噪声。如图 6.4 左图所示，初始含噪数据远离 Calabi–Yau 流形，而算法 yl25 的输出则重建得明显更接近该流形（中图）。这表明该方法对于复杂几何结构的有效性。此外，实验也证实，随着噪声水平 σ 降低，豪斯多夫距离随之减小。

数值实验表明，我们的算法为选择均值漂移估计器中的局部平均半径提供了理论指导，从而显著提升了拟合精度和计算效率。然而，理论样本量要求$N = O(\sigma^{-3d-5})$会随底层流形的维数 d 大幅增加，这给高维潜在流形带来了挑战。

7. 结论

本文提出了一种从带噪高维样本中拟合低维流形的高效方法。在高斯噪声加性模型下，我们提出了一种基于核加权平均的局部均值漂移估计器，其带宽 $r = c_D\sigma$ 由理论推导得到。该估计器将管状邻域中的点向底层流形投影，以高概率达到 $O(\sigma^2)$ 阶的一致豪斯多夫误差。与依赖切空间近似的早期方法相比，该方法具有更好的性能。

理论分析表明，主导偏差与平均曲率向量成正比；平均曲率向量是一种二阶外在几何量，而切向误差则属于更高阶 $O(\sigma^3)$。这一几何洞见不仅细化了误差刻画，还揭示了其与离散平均曲率流之间的联系，将流形学习与几何分析衔接起来。在实际应用中，该方法只需噪声水平 σ，无需迭代

方案或隐式表示，同时提供清晰的输入—输出对应关系和计算效率。

在圆、环面和 Calabi–Yau 流形上的数值实验验证了理论结果：豪斯多夫距离随 σ 呈二次衰减。该估计量在准确性和运行时间方面均优于现有方法，而所提出的带宽选择 $r = c_D\sigma$ 无需启发式调参，并提升了可复现性。然而，样本复杂度$N = O(\sigma^{-3d-5})$ 随内在维数 d 快速增长，这给维数极高的潜在流形带来了挑战。

未来的工作可以考虑更一般的分布模型，包括底层流形上的非均匀采样以及一般的无界噪声。另一个值得探索的方向是研究点云模型上的离散平均曲率流，并分析其收敛性质。

总之，本研究为流形拟合提供了一个具有理论依据、几何可解释性且实际有效的框架，其理论保证与实证表现高度吻合。本研究推动了统计流形估计领域的发展，并为几何数据分析的进一步研究奠定了基础。

8. Proofs

8.1. **引理的证明**。

引理 *2.5* 的证明。 回忆一下，$\exp_x(u) = \gamma_u(1)$，其中 $\gamma_u : [0,1] \to \mathcal{M}$ 是满足初始条件 $\gamma_u(0) = x, \gamma_u'(0) = u$的 $\mathcal{M}$ 中的测地线。考虑 $\exp_x$ 在 $T_x\mathcal{M}$原点处的泰勒展开，可写为

$$\exp_x(u) = x + I_x(u) + \frac{1}{2}Q_x(u) + \dots,$$

其中 I_x, Q_x 分别是 $T_x\mathcal{M}$中的线性型和二次型。因此

$$\gamma_u(t) = x + I_x(u)t + \frac{1}{2}Q_x(u)t^2 + O(t^3).$$

因此， $I_x(u) = \gamma_u'(0) = u$ 且 $Q_x(u) = \gamma_u''(0)$。由于 γ 是一条测地线，我们有$\nabla^{\mathcal{M}}_{\gamma'(0)}\gamma'(0) = 0$，这意味着 $\gamma_u''(0) = \Pi_x(u,u)$。证明完毕。□

引理 *2.6* 的证明。 由于 $g_{ij}|_x = \delta_{ij}$ 可知， $D\exp_x$ 是 $T_x\mathcal{M}$原点处的恒等映射，这意味着 $\frac{\partial}{\partial u_i} = d(\exp_x)_0(u_i) = u_i$. 指数坐标中的测地线参数化为

$$c(t) = t(u^1, u^2, \dots, u^d).$$

将其代入测地线方程，可得对任意 k,

$$0 = \ddot{c^k}(t) + \dot{c^i}(t)\dot{c^j}(t)\Gamma^k_{ij}(\exp(c(t))) = u^iu^j\Gamma^k_{ij}(\exp(c(t))).$$

令 $t = 0$，我们有如下二次型

$$u^iu^j\Gamma^k_{ij}(x) = 0$$

对于任意 $u^i u^j$ 和对称的 Γ_{ij}^k。因此，对于所有 $i,\ j,\ k$，均有 $\Gamma_{ij}^k(\mathbf{0}) = 0$，这意味着 $\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij} = 0$ 在原点处成立。因此 $\partial_k g_{ij}|_x = 0$ 可由

$$\partial_k g_{ij}|_x = \frac{1}{2}(\partial_k g_{ij} + \partial_j g_{ik} - \partial_i g_{jk} + \partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ik})|_x = 0.$$

因此，$\sqrt{\det g_{ij}(x)} = 1$ 及其一阶导数均为零。在 x 处进行泰勒展开，得到

$$\sqrt{\det g_{ij}(u)} = 1 + O(\|u\|^2).$$

证明至此完成。 □

引理 *2.11* 的证明。设

$$\widehat{\mu}_n - \widehat{\mu}_w = \frac{S_n}{B_n} - \frac{s}{b} = \frac{S_n - s}{B_n} - \frac{s}{b}\cdot\frac{B_n - b}{B_n}.$$

固定事件

$$\mathcal{E} := \{|B_n - b| \le b/2\}.$$

在 $\mathcal{E}$ 上，我们有 $B_n \ge b/2$，因此得到初等界

$$\|\widehat{\mu}_n - \widehat{\mu}_w\| \le \frac{2\|S_n - s\|}{b} + \frac{2\|s\|}{b^2}\,|B_n - b|. \tag{8.1}$$

要在 $\mathcal{E}$ 上强制满足 $\|\widehat{\mu}_n - \widehat{\mu}_w\| \le \varepsilon$，只需要求以下两个确定性不等式

$$\|S_n - s\| \le \frac{b\varepsilon}{4}, \qquad |B_n - b| \le \frac{b^2\varepsilon}{4\|s\|}.$$

因此，根据并集界，

(8.2)

$$\mathbb{P}\big(\|\widehat{\mu}_n - \widehat{\mu}_w\| > \varepsilon\big) \le \mathbb{P}(\mathcal{E}^c) + \mathbb{P}\Big(\|S_n - s\| > \frac{b\varepsilon}{4}\Big) + \mathbb{P}\Big(|B_n - b| > \frac{b^2\varepsilon}{4\|s\|}\Big).$$

我们使用霍夫丁不等式（引理 2.10）分别对右侧的三项进行界定。

（i）对 $\mathbb{P}(\mathcal{E}^c) = \mathbb{P}(|B_n - b| > b/2)$ **的界。**由于 $0 \le W \le M$，对标量平均值 $B_n = \frac{1}{n}\sum W_i$ 应用霍夫丁不等式，其取值范围长度为 M，得到

$$\mathbb{P}(|B_n - b| > b/2) \ \le\ 2\exp\Big(-\frac{nb^2}{2M^2}\Big). \tag{8.3}$$

（ii）对 $\mathbb{P}\big(\|S_n - s\| > b\varepsilon/4\big)$ **的界。**根据假设，每个坐标 $(W(Y)Y)^{(j)}$ 几乎必然位于以其均值为中心、长度至多为 $2K$的固定区间内。因此，对于每个坐标 j，根据霍夫丁不等式可得

$$\mathbb{P}\Big(\big|(S_n - s)^{(j)}\big| > t\Big) \le 2\exp\Big(-\frac{2nt^2}{(2K)^2}\Big) = 2\exp\Big(-\frac{nt^2}{2K^2}\Big).$$

对 D 个坐标应用并集界，并利用 $\|v\| \le \sqrt{D}\|v\|_\infty := \sqrt{D}\max_j |v^{(j)}|$，我们得到，对于任意 $\tau > 0$,

$$\mathbb{P}\big(\|S_n - s\| > \tau\big) \le 2D \exp\Big(-\frac{n\tau^2}{2DK^2}\Big).$$

代入 $\tau = b\varepsilon/4$ 可得

$$\mathbb{P}\Big(\|S_n - s\| > \frac{b\varepsilon}{4}\Big) \le 2D \exp\Big(-\frac{nb^2\varepsilon^2}{32DK^2}\Big). \tag{8.4}$$

(iii) $\mathbb{P}\big(|B_n - b| > b^2\varepsilon/(4\|s\|)\big)$的界。再次对 B_n应用霍夫丁不等式，并取阈值$b^2\varepsilon/(4\|s\|)$，可得

$$\mathbb{P}\Big(|B_n - b| > \frac{b^2\varepsilon}{4\|s\|}\Big) \le 2\exp\Big(-\frac{2ns^2}{M^2}\Big) = 2\exp\Big(-\frac{nb^4\varepsilon^2}{8\|s\|^2M^2}\Big). \tag{8.5}$$

将三个界 (8.3)、(8.4)、(8.5) 与 (8.2) 结合，我们得到显式的有限样本尾界

$$\mathbb{P}\big(\|\widehat{\mu}_n - \widehat{\mu}_w\| > \varepsilon\big) \le 2\exp\Big(-\frac{nb^2}{2M^2}\Big) + 2D\exp\Big(-\frac{nb^2\varepsilon^2}{32DK^2}\Big) + 2\exp\Big(-\frac{nb^4\varepsilon^2}{8\|s\|^2M^2}\Big).$$

证明至此完成。 □

8.2. **第3节的证明**。

定理 *3.1* 的证明。 我们将 $p_\sigma(y)$ 在 z 处展开为泰勒级数:

$$\begin{aligned} p_\sigma(y) &= p_\sigma(z) + \sum_{k=1}^{D} (y^{(k)} - z^{(k)})\, \partial_k p_\sigma(z) \\ &\quad + \frac{1}{2}\sum_{k,l=1}^{D} (y^{(k)} - z^{(k)})(y^{(l)} - z^{(l)})\, \partial^2_{kl} p_\sigma(z) + O(\|y - z\|^3). \end{aligned}$$

由于核 $\varphi_r(y-z)$ 和积分区域 $B_D(z,\sqrt{2}r)$ 都是径向对称的，因此所有关于 $(y-z)$ 的奇函数项在积分后均为零。于是,

$$\begin{aligned} &\int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, p_\sigma(y)\, dy \\ &= \int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\Big(p_\sigma(z) + \tfrac{1}{2}|y^{(1)} - z^{(1)}|^2\, \Delta p_\sigma(z) + O(\|y-z\|^4)\Big) dy. \end{aligned} \tag{8.6}$$

令 $y = z + \sqrt{2r}\,u$，并利用 $dy = (\sqrt{2r})^D du$，可得

$$\begin{aligned}
&2^{-\frac{D}{2}} \int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, p_\sigma(y)\, dy \\
&= \int_{B_D(0,1)} \varphi_1(\sqrt{2}u) \Big(p_\sigma(z) + r^2 (u^{(1)})^2 \Delta p_\sigma(z) + O(r^4) \Big) du \\
&= A p_\sigma(z) + B r^2 \Delta p_\sigma(z) + O(r^4),
\end{aligned}$$

其中

$$A = \int_{B_D(0,1)} \varphi_1(\sqrt{2}u)\, du, \qquad B = \int_{B_D(0,1)} (u^{(1)})^2\, \varphi_1(\sqrt{2}u)\, du.$$

类似地，由于 $\Delta(y\, p_\sigma(y)) = 2\nabla p_\sigma(y) + y\, \Delta p_\sigma(y)$，我们有

$$\begin{aligned}
&2^{-\frac{D}{2}} \int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, y\, p_\sigma(y)\, dy \\
&= A z\, p_\sigma(z) + B r^2 \big(2\nabla p_\sigma(z) + z\, \Delta p_\sigma(z) \big) + O(r^4).
\end{aligned} \tag{8.7}$$

于是，由 (8.6) 和 (8.7) 得

$$\begin{aligned}
\mu_z &= \frac{A z p_\sigma(z) + B r^2 (2\nabla p_\sigma(z) + z \Delta p_\sigma(z)) + O(r^4)}{A p_\sigma(z) + B r^2 \Delta p_\sigma(z) + O(r^4)} \\
&= z + r^2\, \frac{2B\, \nabla p_\sigma(z)}{A\, p_\sigma(z)} + O(r^4),
\end{aligned}$$

至此证明完毕。 □

定理 3.3 的证明。 任取 $\varepsilon \in (0, \tau)$ 和 $y \in T(\tau - \varepsilon)$。我们假设$0 < \sigma < \sigma_0$ 对某个 $\sigma_0 > 0$成立。令 $x_0 = \pi(y) \in \mathcal{M}$, $v := y - x_0 \in T_{x_0}^{\perp}\mathcal{M}$，使得 $\|v\| \le \tau - \varepsilon$。设

$$C_\sigma := \big(\mathrm{Vol}(\mathcal{M}) (2\pi\sigma^2)^{D/2} \big)^{-1}, \quad p_\sigma(y) = C_\sigma \int_{\mathcal{M}} \exp\Big(-\tfrac{1}{2\sigma^2} \|y - x\|^2 \Big)\, d\mu(x).$$

定义对称端omorphism $\langle v,\ \Pi_{x_0} \rangle$ 在 $T_{x_0}\mathcal{M}$ 上如下：$\langle \langle v,\ \Pi_{x_0} \rangle u,\ w \rangle = \langle v,\ \Pi_{x_0}(u,\ w) \rangle$，并令 $A := I - \langle v,\ \Pi_{x_0} \rangle$.。利用 $\|\Pi_{x_0}\|_{\mathrm{op}} \le \tau^{-1}$（引理 2.3）和 $\|v\| \le \tau - \varepsilon$，则

$$\lambda_{\min}(A) \ \ge\ 1 - \|v\|/\tau \ \ge\ \varepsilon/\tau \ =:\ c_\varepsilon > 0.$$

因此，A 在 $T(\tau - \varepsilon)$ 上一致正定，且具有谱隙 c_ε。

在 x_0 处的指数坐标中，我们将 $x(u)$ 写作：$= \exp_{x_0}(u)$，其中 $u \in T_{x_0}\mathcal{M} \simeq \mathbb{R}^d$。根据引理 2.4，存在 $r_0 > 0$，使得 $u \mapsto x(u)$ 是从 $B_d(0,\ r_0)$ 到 $E_0 := \exp_{x_0}(B_d(0,\ r_0))$ 的微分同胚。我们在 $B_d(0,\ r_0)$ 上有如下展开，

$$\sqrt{\det g(u)} = 1 + Q_g(u), \qquad |Q_g(u)| \le C_g \|u\|^2, \quad \text{(Lemma 2.6)}$$

$$x(u) = x_0 + u + \tfrac{1}{2}\, \Pi_{x_0}(u,\ u) + Q_e(u) \qquad \|Q_e(u)\| \le C_e \|u\|^3, \quad \text{（引理 2.5）}$$

其中常数为 C_g, C_e ，仅取决于 $\mathcal{M}$。根据 $x(u)$ 的展开，我们得到

$$\|y - x(u)\|^2 = \|v\|^2 + u^\top A u + R(u), \qquad R(u) = O\big(\|v\|\,\|u\|^3 + \|u\|^4\big),$$

其常数仅依赖于 $\mathcal{M}$, ε。对于 $r = r_0$ 并令

$$E := \exp_{x_0}\big(B_d(0,r)\big) \subset \mathcal{M}, \qquad P_\sigma(y) = C_\sigma\,(I_1 + I_2),$$

其中

$$I_1 := \int_E e^{-\|y-x\|^2/(2\sigma^2)}\, d\mu(x), \qquad I_2 := \int_{\mathcal{M}\setminus E} e^{-\|y-x\|^2/(2\sigma^2)}\, d\mu(x).$$

使用指数坐标 $x = x(u)$ 以及上述展开式，我们有

$$I_1 = \int_{\|u\|\le r} \exp\Big(-\frac{1}{2\sigma^2}\big(\|v\|^2 + u^\top A u + R(u)\big)\Big)\big(1 + Q_g(u)\big)\, du.$$

作变量代换 $z = u/\sigma$，使得 $du = \sigma^d dz$。利用 $R(u) = O(\|v\|\|u\|^3 + \|u\|^4)$、$Q_g(u) = O(\|u\|^2)$ 的界以及指数函数的泰勒展开，我们得到

$$\begin{aligned} I_1 &= e^{\frac{-\|v\|^2}{2\sigma^2}}\sigma^d \int_{\|z\|\le r/\sigma} e^{-\frac{1}{2}z^\top A z}\Big(1 - \frac{R(\sigma z)}{2\sigma^2} + O\Big(\frac{R(\sigma z)^2}{\sigma^4}\Big)\Big)\big(1 + Q_g(\sigma z)\big)\, dz \\ &= e^{\frac{-\|v\|^2}{2\sigma^2}}\sigma^d \int_{\|z\|\le r/\sigma} e^{-\frac{1}{2}z^\top A z}\Big(1 + \sigma\|v\|\phi_1(y,z,\sigma) + \sigma^2\phi_2(y,z,\sigma)\Big)\, dz, \end{aligned}$$

其中 ϕ_1, ϕ_2 是 $(y,\ z,\ \sigma)$ 的光滑函数，并满足如下形式的多项式界

$$|\phi_i(y,z,\sigma)| \le C\|z\|^m$$

对某个与 (y,σ) 无关的 m 。利用 $\mathcal{M}$ 是一个紧致 C^3 子流形以及管状邻域定理，所有曲率量和投影 $\pi(\cdot)$ 及其导数在 $T(\tau - \varepsilon)$ 上均一致有界。因此，对于 $|\alpha| \le 1$，存在一个整数 m_0 和常数 $C_\alpha > 0$ ，使得

$$\big|\partial_y^\alpha \phi_i(y,z,\sigma)\big| \ \le\ C_\alpha\,\|z\|^{m_0}.$$

由于 $\lambda_{\min}(A) \ge c_\varepsilon$，高斯尾部呈指数级小：

$$\int_{\|z\|>r/\sigma} e^{-\frac{1}{2}z^\top A z} dz \ \le\ e^{-\frac{c_\varepsilon r^2}{4\sigma^2}} \int_{\mathbb{R}^d} e^{-\frac{1}{4}z^\top A z} dz \ \le Ce^{-c/\sigma^2}.$$

因此，我们可以将积分范围扩展到整个 $\mathbb{R}^d$ ，误差呈指数级小，并写作

$$I_1 = e^{-\|v\|^2/(2\sigma^2)}\sigma^d\Big(J_0(y) + \sigma\|v\|J_1(y,\sigma) + \sigma^2 J_2(y,\sigma)\Big),$$

其中

$$J_0(y) := \int_{\mathbb{R}^d} e^{-\frac{1}{2}z^\top A z}\, dz = \frac{(2\pi)^{d/2}}{\sqrt{\det A}},$$

and

$$J_i(y,\sigma) := \int_{\mathbb{R}^d} e^{-\frac{1}{2}z^\top A z}\, \phi_i(y,z,\sigma)\, dz, \quad i=1,2.$$

利用 $\phi_i(y,\ z,\ \sigma)$ 的有界性，可得积分中的高斯矩

$$\left|\partial_y^\alpha \left(e^{-\frac{1}{2}z^\top A_y z}\phi_i(y,z,\sigma)\right)\right| \ \le\ C'_\alpha\, e^{-\frac{c_\varepsilon}{2}\|z\|^2}\, \|z\|^{m_0},$$

因此，由控制收敛可得

$$\sup_{(y,\sigma)\in T(\tau-\varepsilon)\times(0,\sigma_0]} \left|\partial_y^\alpha J_i(y,\sigma)\right| \ \le\ C''_\alpha,$$

对于某个常数 $C''_\alpha < \infty$。特别地，J_i 关于 y光滑，且所有y导数在 $T(\tau-\varepsilon)\times(0,\ \sigma_0]$上一致有界。

对于 I_2，在边界 $\|u\|=r$上，由 $\lambda_{\min}(A) \ge c_\varepsilon$ 可得

$$\|y-x(u)\|^2 = \|v\|^2 + u^\top A u + R(u) \ \ge\ \|v\|^2 + \delta_\varepsilon, \qquad \delta_\varepsilon := \tfrac{c_\varepsilon}{2}\, r^2 = \tfrac{\varepsilon}{2\tau}\, r^2.$$

于是，集合 $\{\|y-x\|^2 < \|v\|^2+\delta_\varepsilon\}$包含于 E中。因此，对于 $x\in\mathcal{M}\setminus E$，$\|y-x\|^2 \ge \|v\|^2+\delta_\varepsilon$，且

$$I_2 \le\ \mathrm{Vol}(\mathcal{M})\, \exp\Big(-\frac{\|v\|^2+\delta_\varepsilon}{2\sigma^2}\Big) \ =\ e^{-\frac{\|v\|^2}{2\sigma^2}}\, O\big(e^{-\delta_\varepsilon/(2\sigma^2)}\big),$$

它比 σ的任意次幂都指数级更小，因此通过调整 J_2，可以将项 I_2 吸收到 $\sigma^2 J_2$ 贡献中。

最后，将 I_1和 I_2合并。令 $F_i(y,\sigma) := J_i(y,\sigma)/J_0(y)$，并利用 $C_\sigma\sigma^d J_0(y)\ = (\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{\frac{D-d}{2}})^{-1}\det(A)^{-1/2}$，可得

$$p_\sigma(y) = \frac{e^{-\|v\|^2/(2\sigma^2)}}{\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{\frac{D-d}{2}}}\, \frac{1}{\sqrt{\det A}}\Big(1+\sigma\|v\|F_1(y,\sigma)+\sigma^2 F_2(y,\sigma)\Big),$$

对所有 $y\in T(\tau-\varepsilon)$ 和 $0<\sigma\le\sigma_0$均成立。由于 $\lambda_{\min}(A) \ge c_\varepsilon$，$J_0(y)$有正下界，且关于 y 光滑，因此 F_1, F_2 及其所有导数均一致有界。□

推论 *3.3.1* 的证明。固定任意 $\varepsilon\in(0,\tau)$，并考虑集合 $K := T(\tau-\varepsilon)\times(0,\sigma_0]$。根据定理 3.3，我们有

$$p_\sigma(y) = C_\sigma\, \frac{e^{-\frac{\|v_y\|^2}{2\sigma^2}}}{\sqrt{\det A_y}}\,\big(1+R_P(y,\sigma)\big),\ C_\sigma := \big(\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{D/2}\big)^{-1}, \tag{8.8}$$

其中

$$R_P(y,\sigma) = \sigma\|v_y\|\, G_1(y,\sigma) + \sigma^2 G_2(y,\sigma). \tag{8.9}$$

此处 $G_1,\ G_2$ 是定义在 K 上关于 $(y,\ \sigma)$ 的光滑函数，源自定理 3.3 证明中光滑核的高斯积分。特别地，存在常数 $C_m>0$ 使得对于 $m=1,\ 2$，

$$\sup_{(y,\sigma)\in K} \left\|\nabla_y^m G_i(y,\sigma)\right\| \le C_m, \qquad i=1,2. \tag{8.10}$$

根据管状邻域定理，投影 $\pi\colon T(\tau-\varepsilon) \to \mathcal{M}$ 以及法向分量 $v_y = y - \pi(y)$ 在 y 中是 C^2，因此它们的导数在 $T(\tau-\varepsilon)$ 上一致有界。也就是说，对于 $m = 1,\ 2$，存在 $B_m > 0$使得

$$\sup_{y\in T(\tau-\varepsilon)} \big\|\nabla_y^m v_y\big\| \le B_m. \tag{8.11}$$

$\|v_y\|$也同样成立，因为它由v_y 经过光滑算子得到。

现在利用 (8.10) 和 (8.11) 对 $\nabla_y^m R_P$ 进行界定。对于 $m = 0$，我们有

$$|R_P(y,\sigma)| \le \sigma\|v_y\|\big|G_1(y,\sigma)\big| + \sigma^2\big|G_2(y,\sigma)\big| \le C_0\big(\sigma\|v_y\| + \sigma^2\big),$$

对于某个常数 C_0。对于 $m = 1$，由于每个导数 $\nabla_y R_P$ 都是形如以下各项的有限和

$$\sigma\,\nabla_y^{k_1}\|v_y\|\,\nabla_y^{k_2} G_1(y,\sigma) \quad \text{or} \quad \sigma^2\,\nabla_y^k G_2(y,\sigma), \qquad k_1 + k_2 \le 1,\ k \le 1,$$

存在仅依赖于 ε 和 $\mathcal{M}$ 的常数 $C_{\varepsilon,1} > 0$，使得

$$\big\|\nabla_y R_P(y,\sigma)\big\| \le C_{\varepsilon,1}\big(\sigma\|v_y\| + \sigma^2\big). \tag{8.12}$$

对 (8.8) 取对数，得到

$$\log p_\sigma(y) = \log C_\sigma - \frac{\|v_y\|^2}{2\sigma^2} - \frac{1}{2}\log\det A_y + \log\big(1 + R_P(y,\sigma)\big).$$

Let

$$R_0(y,\sigma) := \log\big(1 + R_P(y,\sigma)\big).$$

如有必要，缩小 σ_0 后，我们可以假设

$$|R_P(y,\sigma)| \le \frac{1}{2}, \qquad (y,\sigma) \in K.$$

于是，$1+R_P$ 有正下界。对于 $m = 0$，由简单的不等式可得

$$|R_0(y,\sigma)| = \big|\log(1 + R_P(y,\sigma))\big| \le 2\,|R_P(y,\sigma)| \le 2C_{\varepsilon,0}\big(\sigma\|v_y\| + \sigma^2\big).$$

对于 $m = 1$，导数 $\nabla_y R_0(y,\ \sigma)$是有限项之和，每一项具有如下形式

$$c_\ell \partial_y^l R_P(y,\sigma),$$

其中 c_ℓ 是涉及对数函数各阶导数的光滑函数。

由 (8.12) 可得

$$\big\|\nabla_y R_0(y,\sigma)\big\| \le C'_{\varepsilon,1}\big(\sigma\|v_y\| + \sigma^2\big),$$

对于某些常数$C'_{\varepsilon,\ 1}$，它们仅依赖于 ε 和 $\mathcal{M}$。 □

引理 *3.4* 的证明。对于任意 $y \in \mathcal{T}(\tau)$，记 $x := \pi(y) \in \mathcal{M}$ 和 $v := v_y = y - x \in T_x^\perp\mathcal{M}$。令 $S_v\colon T_x\mathcal{M} \to T_x\mathcal{M}$ 表示与法向量 v关联的形状算子，其定义为

$$\langle S_v(u), w\rangle = \langle \Pi_x(u,w), v\rangle, \qquad S_v(u) = -(\nabla_u v)^\top,$$

对于 $u, w \in T_x\mathcal{M}$。那么我们有 $A_y = I_d - \langle v_y, \Pi_x\rangle = I_d - S_v$在 $T_x\mathcal{M}$ 上。

令 $\omega \in \mathbb{R}^D$ 为任意值，并选取一条光滑曲线 $\gamma\colon (-\delta,\delta) \to T(\tau)$，使得 $\gamma(0) = y$ 且 $\gamma'(0) = \omega$。令

$$\pi(t) := \pi(\gamma(t)) \in \mathcal{M}, \qquad v(t) := v_{\gamma(t)} = \gamma(t) - \pi(t) \in T^{\perp}_{\pi(t)}\mathcal{M}.$$

那么，对于沿着 $\pi(t)$ 的任意切向量场 $Y(t) \in T_{\pi(t)}\mathcal{M}$，我们有

$$\langle v(t), Y(t)\rangle = 0, \qquad \forall t.$$

对该恒等式关于 t 求导，并在 $t = 0$ 处取值，可得

$$0 = \frac{d}{dt}\langle v(t), Y(t)\rangle\big|_{t=0} = \langle v'(0), Y(0)\rangle + \langle v(0), \nabla_{\pi'(0)} Y\rangle, \tag{8.13}$$

其中 $v'(0) = \nabla_\omega v$，且 $\pi'(0) = \nabla_\omega \pi(y) \in T_{\pi(y)}\mathcal{M}$。根据形状算子的定义（必要时对 v 作局部延拓），我们有

$$\langle v(0), \nabla_{\pi'(0)} Y\rangle = -\langle (\nabla_{\pi'(0)} v)^\top, Y(0)\rangle = -\langle S_v(\pi'(0)), Y(0)\rangle.$$

将其代入 (8.13)，得到

$$\langle v'(0), Y(0)\rangle = -\langle S_v(\pi'(0)), Y(0)\rangle,$$

这意味着 $\nabla_\omega v$ 的切向分量满足

$$(\nabla_\omega v)^\top = -S_v(\nabla_\omega \pi(y)). \tag{8.14}$$

另一方面，根据定义 $v(t) = \gamma(t) - \pi(t)$，

$$\nabla_\omega v = \nabla_\omega(\gamma - \pi) = \omega - \nabla_\omega \pi(y). \tag{8.15}$$

取 (8.15) 的切向分量，并将其与 (8.14) 结合，可得

$$\omega^\top - \nabla_\omega \pi(y) = -S_v(\nabla_\omega \pi(y)).$$

因此，

$$\omega^\top = (I_d - S_v)\big(\nabla_\omega \pi(y)\big). \tag{8.16}$$

根据引理 2.3 和到达半径假设，我们有 $\|S_v\|_{\mathrm{op}} \le \|\Pi_x\|_{\mathrm{op}}\|v\| \le \tau^{-1}\|v\| < 1$ for $y \in \mathcal{T}(\tau)$，因此 $A_y = I - S_v$ 在 $T_x\mathcal{M}$ 上可逆。因此

$$\nabla_\omega \pi(y) = A_y^{-1}\omega^\top.$$

最后，将其代回 (8.15)，得到

$$\nabla_\omega v_y = \omega - A_y^{-1}\omega^\top.$$

□

定理 *3.5* 的证明。对于任意 $\varepsilon \in (0,\tau)$，令 $\sigma_0 > 0$ 如推论 3.3.1 中所述。对于 $y \in T(\tau - \varepsilon)$ 和 $0 < \sigma \le \sigma_0$，推论 3.3.1 给出展开式

$$\log p_\sigma(y) = C_\sigma - \frac{\|v_y\|^2}{2\sigma^2} - \frac{1}{2}\log\det A_y + R(y,\sigma), \tag{8.17}$$

其中 $C_\sigma = \log\big(\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{(D-d)/2}\big)^{-1}$ 且余项 R 满足

$$\|\nabla_y^m R(y,\sigma)\|_{\mathrm{op}} \le C_{\varepsilon,m}\big(\sigma\|v_y\| + \sigma^2\big), \quad m = 0, 1.$$

对 (8.17) 取环境梯度可得

$$\nabla \log p_\sigma(y) = -\frac{1}{2\sigma^2}\nabla\|v_y\|^2 - \frac{1}{2}\nabla \log\det A_y + \nabla R(y,\sigma). \tag{8.18}$$

利用引理 3.4，对于任意 $\omega \in \mathbb{R}^D$，有

$$\nabla_\omega v_y = \omega - A_y^{-1}\omega^\top.$$

由于 $v_y \in T^\perp_{\pi(y)}\mathcal{M}$ 和 $A_y^{-1}\omega^\top \in T_{\pi(y)}\mathcal{M}$，我们得到

$$D_\omega\|v_y\|^2 = 2\langle v_y, \nabla_\omega v_y\rangle = 2\langle v_y, \omega\rangle.$$

根据梯度的定义，这意味着

$$\nabla\|v_y\|^2 = 2v_y.$$

因此，(8.18) 中的第一项恰好为

$$-\frac{1}{2\sigma^2}\nabla\|v_y\|^2 = -\frac{v_y}{\sigma^2}.$$

定义

$$G(y) := -\frac{1}{2}\log\det A_y, \qquad y \in T(\tau-\varepsilon).$$

对于 $x \in \mathcal{M}$，我们有 $v_x = 0$ 和 $A_x = I_d$，因此 $\det A_x = 1$ 且 $G(x) = 0$。令 $u \in T_x^\perp\mathcal{M}$，满足 $\|u\| = 1$，并考虑曲线

$$\gamma(t) := x + tu, \qquad |t| \ll 1.$$

当 $|t|$ 足够小时，$\gamma(t) \in T(\tau-\varepsilon)$，$\pi(\gamma(t)) = x$ 且 $v_{\gamma(t)} = tu$。令 S_u 为在 x 处定义、与 u 相关联的形状算子。由形状算子关于法向量参数的线性性可得

$$A_{\gamma(t)} = I_d - \langle tu, \Pi_x\rangle = I_d - t\,S_u,$$

因此

$$G(\gamma(t)) = -\frac{1}{2}\log\det\big(I_d - tS_u\big).$$

在 $t = 0$ 处求导可得

$$\frac{d}{dt}G(\gamma(t))\Big|_{t=0} = -\frac{1}{2}\operatorname{Tr}(-S_u) = \frac{1}{2}\operatorname{Tr}(S_u).$$

对于平均曲率 $H_x = \frac{1}{d}\operatorname{Tr}(\Pi_x)$，我们有 $\operatorname{Tr}(S_u) = d\langle H_x, u\rangle$。因此

$$\frac{d}{dt}G(\gamma(t))\Big|_{t=0} = \frac{d}{2}\langle H_x, u\rangle,$$

这表明

$$\nabla G(x) = \frac{d}{2}H_x \in T_x^\perp\mathcal{M}.$$

现在令 $y \in T(\tau-\varepsilon)$。根据引理 2.8，我们有

$$\|\nabla G(y) - \nabla G(\pi(y))\| \le C\|v_y\|$$

对于某个常数 $C > 0$，它仅取决于 $\mathcal{M}$ 和 ε。因此

$$-\frac{1}{2}\nabla\log\det A_y = \nabla G(y) = \frac{d}{2}H_{\pi(y)} + O(\|v_y\|).$$

根据 A_y 的正则性， $\nabla F(y)$ 在 $T(\tau-\varepsilon)$ 上一致有界。由推论 3.3.1

$$\|\nabla R(y,\sigma)\| \le C_{\varepsilon,1}\big(\sigma\|v_y\| + \sigma^2\big).$$

将上述梯度代入 (8.18)，得到

$$\nabla \log p_\sigma(y) = -\frac{v_y}{\sigma^2} + \frac{d}{2} H_{\pi(y)} + O(\|v_y\| + \sigma\|v_y\| + \sigma^2),$$

其中， $O(\cdot)$ 项关于 $y \in T(\tau-\varepsilon)$ 和 $0<\sigma\le\sigma_0$ 是一致的。 □

定理 *3.7* 的证明。 对任意固定的 $\varepsilon\in(0,\tau)$，令 $\sigma_0>0$ 如定理 3.5 中所述。对于 $0<\sigma\le\sigma_0$，由定理 3.1（取 $r^2 = \frac{A}{2B}\sigma^2 = c_D^2\sigma^2$），我们可以估计总体映射 μ_z 在每个 $z\in T(\tau-\varepsilon)$ 处的值：

$$\begin{aligned}
\mu_z &= \frac{\mathbb{E}[\varphi_r(Y-z)Y]}{\mathbb{E}[\varphi_r(Y-z)]} = \frac{\int_{B_D(z,\sqrt{2}r)} \varphi_\sigma(y-z)\, y\, p_\sigma(y)\, dy}{\int_{B_D(z,\sqrt{2}r)} \varphi_\sigma(y-z)\, p_\sigma(y)\, dy} \\
&= z + \sigma^2 \nabla \log p_\sigma(z) + O(\sigma^4) \\
&= z - v_z + \sigma^2 \frac{d}{2} H_{\pi(z)} + \sigma^2 O(\|v_z\| + \sigma\|v_z\| + \sigma^2) \\
&= \pi(z) + \sigma^2 \frac{d}{2} H_{\pi(z)} + O(\sigma^2).
\end{aligned}$$

因此，对于某个常数 C， $\mu(T(\tau-\varepsilon))$ 包含于以 $\mathcal{M}$ 为中心、半径为 $C\sigma^2$ 的管状邻域中，也就是说， $\mu(T(\tau-\varepsilon)) \subset T(C\sigma^2)$，并且

$$\sup_{z\in T(\tau-\varepsilon)} d(\mu_z, \mathcal{M}) \le C\sigma^2.$$

相反，对于每个 $x\in\mathcal{M}$，存在 $z = x \in T(\tau-\varepsilon)$ 使得$\mu_z = x + O(\sigma^2)$ $v\in T_x^\perp\mathcal{M}$，这意味着

$$\sup_{x\in\mathcal{M}} d(x, \mu(T(\tau-\varepsilon))) \le C\sigma^2.$$

那么， $\mu(T(\tau-\varepsilon))$ 是满足以下条件的光滑子流形

$$\mathrm{dist}_H(\mu(T(\tau)), \mathcal{M}) \le C\sigma^2.$$

证明至此完成。 □

8.3. **第4节的证明。**

定理 *4.1* 的证明。 对于任意 $z\in\Gamma$，设 $\sigma\in(0,\sigma_0]$。由引理 2.11，令 $W(y) = \varphi_r(y-z)$，则 $M = O(r^{-D}) = O(\sigma^{-D})$ 且 $K := 2rM = O(\sigma^{1-D})$。采用定理 3.1 的论证，我们得到

$$\begin{aligned}
b &= \int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, p_\sigma(y)\, dy \\
&= \int_{B_D(0,1)} \varphi_1(\sqrt{2}u) \left(p_\sigma(z) + r^2 (u^{(1)})^2 \Delta p_\sigma(z) + O(r^4) \right) 2^{\frac{D}{2}} du \\
&= C p_\sigma(z) + O(r^2),
\end{aligned}$$

其中 $C = 2^{\frac{D}{2}} \int_{B_D(0,\,1)} \varphi_1(\sqrt{2}u)\, du$。类似地，

$$s = \int_{B_D(z,\sqrt{2}r)} \varphi_r(y-z)\, y\, p_\sigma(y)\, dy = Cz\, p_\sigma(z) + O(r^2). \tag{8.19}$$

回顾定理 3.3，我们有

$$p_\sigma(y) = \frac{1}{\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{\frac{D-d}{2}}} e^{-\frac{\|v_y\|^2}{2\sigma^2}} \frac{1}{\sqrt{\det A_y}} \left(1 + O(\sigma\|v_y\| + \sigma^2)\right).$$

由于 $r = O(\sigma)$ 和 $d(z,\ \mathcal{M}) \le O(\sigma)$，我们得到 $b = O(\sigma^{-D+d})$ 和 $s \le O(\sigma^{1-D+d})$。

根据引理 2.11，令 $\varepsilon = c\,\sigma^3$ 并选取 $n = C_0\,\sigma^{-2d-5}$。采用这一选择后，(2.3) 中指数项的指数为

$$-\frac{nb^2}{2M^2} = -\frac{O(\sigma^{-2d-5})O(\sigma^{-2D+2d)})}{O(\sigma^{-2D})} = -O(\sigma^{-5}),$$

$$-\frac{nb^2\varepsilon^2}{32DK^2} = -\frac{O(\sigma^{-2d-5})O(\sigma^{-2D+2d)})O(\sigma^6)}{O(\sigma^{2-2D})} = -O(\sigma^{-1}),$$

$$-\frac{nb^4\varepsilon^2}{8\|s\|^2M^2} \le -\frac{O(\sigma^{-2d-5})O(\sigma^{-4D+4d)})O(\sigma^6)}{O(\sigma^{2-2D+2d})O(\sigma^{-2D})} = -O(\sigma^{-1}).$$

因此，存在常数 C_1, C_2，使得

$$\mathbb{P}\big(\|\widehat{\mu}_n - \widehat{\mu}_w\| > \varepsilon\big) \le C_1 \exp(-C_2\sigma^{-1}),$$

其中 $\widehat{\mu}_n = \frac{\sum_{i=1}^n W_{(\sum_{i=1}^n W\ y_i\ )()}}{} = F(z)$ 且 $\widehat{\mu}_w = \frac{\mathbb{E}[W_{(\quad Y\ Y]\,)()}}{\mathbb{E}[W\ Y}] = \mu_z$，由此得到所述界

$$\mathbb{P}\big(\|F(z) - \mu_z\| \le c\,\sigma^3\big) \ge 1 - C_1 \exp(-C_2\sigma^{-1}),$$

已给出 $n \ge C_0\,\sigma^{-2d-5}$。证明至此完成。 □

引理 *4.2* 的证明。回忆定理 3.3 中对任意 $y \in T(\tau-\varepsilon)$ 的局部展开

$$p_\sigma(y) = \frac{1}{\mathrm{Vol}(\mathcal{M})(2\pi\sigma^2)^{\frac{D-d}{2}}} e^{-\frac{\|v_y\|^2}{2\sigma^2}} \frac{1}{\sqrt{\det A_y}} \left(1 + O(\sigma\|v_y\| + \sigma^2)\right). \tag{8.20}$$

对于 $z \in \Gamma \subset T(\tau-\varepsilon)$，在其投影 $\pi(z) \in \mathcal{M}$ 周围，引入坐标 $(u,\ n)$，其中 $u \in \mathbb{R}^d$ 是 $\mathcal{M}$ 上的局部切向坐标，$n \in \mathbb{R}^{D-d}$ 表示法向坐标。相应的体积元满足 $dy = J(u,\ n)\, du\, dn$。于是

$$\begin{aligned} \mathbb{P}(Y \in B_D(z,\sqrt{2}r)) &= \int_{B_D(z,\sqrt{2}r)} p_\sigma(y)\, dy \\ &= \int_{\|u-u_z\|\le r'} \int_{\|n-n_z\|\le r''(u)} p_\sigma(u,n)\, J(u,n)\, dn\, du \end{aligned} \tag{8.21}$$

其中 $(u_z,\ n_z)$ 是 z 的坐标，而 r'，$r''(u)$ 是与 r 同阶的量，差异至多为几何常数。

根据（8.20），对 n 的主要依赖呈高斯形式：

$$(2\pi\sigma^2)^{-(D-d)/2}\, e^{-\|n\|^2/(2\sigma^2)}.$$

由于 $r = O(\sigma)$，沿法向方向在 $\|n - n_z\| \le c\sigma$ 上的积分捕获了总高斯质量的固定比例。因此，

$$c_0\, e^{-\|n_z\|^2/(2\sigma^2)} \le \int_{\|n-n_z\|\le c\sigma} (2\pi\sigma^2)^{-(D-d)/2} e^{-\|n\|^2/(2\sigma^2)}\, dv. \tag{8.22}$$

将 (8.22) 代入 (8.21)，并利用 $J(u,\ n) = 1 + O(r)$ 且 $\frac{1}{\sqrt{\det A_{\pi(y)}}} = \frac{1}{\sqrt{\det A_{\pi(z)}}} + O(r)$，可得

$$\mathbb{P}(Y \in B_D(z, \sqrt{2}r)) \ge \frac{c_0 e^{-\|n_z\|^2/(2\sigma^2)}}{\mathrm{Vol}(\mathcal{M})\sqrt{\det A_{\pi(z)}}} \Big(1 + O(r)\Big) \int_{\|u-u_z\|\le r'} du,$$

其中 c_0 是 (8.22) 中的常数。切向积分对应于半径为 $\simeq r$ 的球的 d 维欧几里得体积，即 $\omega_d r^d$。

汇总这些估计，可得

$$\mathbb{P}(Y \in B_D(z, \sqrt{2}r)) \ge \frac{\omega_d\, c_0 e^{-\|n_z\|^2/(2\sigma^2)}}{\mathrm{Vol}(\mathcal{M})\sqrt{\det A_{\pi(z)}}}\, r^d\, (1 + O(r)). \tag{8.23}$$

证明至此完成。 □

推论 *4.2.1* 的证明。点的数量 n可视为一个样本量为 N、概率参数为 $p \ge c\sigma^d$的二项随机变量。于是，根据 Chernoff 界（引理 2.9），

$$\mathbb{P}(n \ge C_1\sigma^{-5-2d}) \ge 1 - 2\exp(-C_3\sigma^{-5-2d}).$$

当 σ 足够小时，该概率将接近 1。 □

8.4. **第 2 节的证明。**

定理 *5.1* 的证明。根据定理 3.7 中的论证，我们可以在每个 $z \in \Gamma$ 处估计总体映射 μ_z，即

$$\begin{aligned}\mu_z = \frac{\mathbb{E}(\varphi_r(Y-z)Y)}{\mathbb{E}(\varphi_r(Y-z))} &= \pi(z) + \frac{d}{2}H_{\pi(z)}\sigma^2 + \sigma^2 O(\|v_z\| + \sigma\|v_z\| + \sigma^2) \\ &= \pi(z) + \frac{d}{2}H_{\pi(z)}\sigma^2 + O(\sigma^3).\end{aligned}$$

根据定理 4.1 和推论 4.2.1，对于样本量 $N \ge O(\sigma^{-3d-5})$，我们证明 $F(z)$ 与 μ_z之间的等价性质，即

$$\mathbb{P}(\|F(z) - \mu_z\| \le c\sigma^3) \ge 1 - C_1 \exp\left(-C_2\sigma^{-1}\right).$$

因此，

$$F(z) = \pi(z) + \frac{d}{2}H_{\pi(z)}\sigma^2 + O(\sigma^3).$$

证明至此完成。 □

推论 *5.2.1* 的证明。 定理 5.1 确保对于每个固定的 $z \in \mathcal{Y}^o$，事件

$$E_z := \{F(z) = \pi(z) + \tfrac{d}{2}H_{\pi(z)}\sigma^2 + O(\sigma^3)\}$$

以至少 $1 - C_1\exp(-C_2\sigma^{-1})$ 的概率成立。因此，联合事件

$$E := \bigcap_{z\in\mathcal{Y}^o} E_z$$

以至少.....的概率成立

$$P(E) = 1 - P\left(\bigcup_{z\in\mathcal{Y}^o} E_z^c\right) \geq 1 - \sum_{z\in\mathcal{Y}^o} P(E_z^c) \geq 1 - nC_1e^{-C_2\sigma^{-1}}.$$

注意到

$$1 - nC_1e^{-C_2\sigma^{-1}} \geq 1 - C_1'\exp(-C_2'\sigma^{-c'})$$

将多项式因子 $n \leq N = O(\sigma^{-3d-5})$ 吸收到常数中后c'， C_1'， C_2'。 □

引理 *5.4* 的证明。 回顾关于可达半径的 Federer 判据（引理 2.2），不失一般性，我们可以假设

$$0 < \tau^{-1} = \sup\left\{\frac{2\cdot d(b, T_a\mathcal{M})}{\|a-b\|^2} \;\middle|\; a, b \in \mathcal{M}, a \neq b\right\}.$$

令 $p, q \in \mathcal{M}$其中 $p = x + f(x)$， $q = y + f(y)$，对 $x, y \in \mathcal{M}$。定义 $\Delta = y - x$，并将其分解为 $\Delta = \Delta^\top + \Delta^\perp$，其中 $\Delta^\top \in T_x\mathcal{M}$ 且$\Delta^\perp \in T_x^\perp\mathcal{M}$。注意，$\|\Delta^\perp\| = d(y, T_x\mathcal{M})$。

的切空间 $\widehat{\mathcal{M}}$在 p处由下式给出

$$T_p\widehat{\mathcal{M}} = \{x + f(x) + v + Df_x(v) \mid v \in T_x\mathcal{M}\}.$$

现在，我们估计 q 到 $T_p\mathcal{M}$的距离：

$$d(q, T_p\widehat{\mathcal{M}}) = \min_{v\in T_x\mathcal{M}} \|\Delta + f(y) - f(x) - v - Df_x(v)\|.$$

取 $v = \Delta^\top$，得到

$$d(q, T_p\widehat{\mathcal{M}}) \leq \|\Delta^\perp + f(y) - f(x) - Df_x(\Delta^\top)\|.$$

对 $f \in C^1$使用泰勒展开，得到

$$f(y) - f(x) = Df_x(\Delta) + O(\|\Delta\|^2) = Df_x(\Delta^\top) + Df_x(\Delta^\perp) + O(\|\Delta\|^2).$$

代入距离表达式

$$d(q, T_p\widehat{\mathcal{M}}) \leq \|\Delta^\perp + Df_x(\Delta^\perp) + O(\|\Delta\|^2)\|.$$

由于 $f(x) = \frac{d}{2}H_x\sigma^2 + O(\sigma^3)$ 且 $\mathcal{M}$ 是紧致的，我们在 $\mathcal{M}$ 上有一致成立的 $\|Df_x\| = O(\sigma^2)$。因此 $\|Df_x(\Delta^\perp)\| \leq O(\sigma^2)\cdot\|\Delta^\perp\|$。

将 Federer 判据（引理 2.2）应用于 $\mathcal{M}$，可得 $C^{-1}\|\Delta^\perp\| \leq\|\Delta\|^2 \leq C\|\Delta^\perp\|$，其中某个 $C > 0$。

综合这些估计，我们得到

$$d(q, T_p\widehat{\mathcal{M}}) \le \|\Delta^\perp\|(1 + C_1\sigma^2 + C_2),$$

对于常数 C_1, $C_2 > 0$ 其取决于 $\mathcal{M}$。

现在考虑 p 与 q 之间的距离:

$$\|p - q\| = \|\Delta + f(y) - f(x)\| = \|\Delta\|(1 + O(\sigma^2)),$$

因为 $\|f(y) - f(x)\| = O(\sigma^2\|\Delta\|)$。

将 Federer 判据应用于 $\mathcal{M}$，我们得到

$$\frac{\|p - q\|^2}{2 \cdot d(q, T_p\widehat{\mathcal{M}})} \ge \frac{\|\Delta\|^2(1 + O(\sigma^2))}{2\|\Delta^\perp\|(1 + O(1))} = \frac{\|\Delta\|^2}{2\|\Delta^\perp\|} \cdot \frac{1 + O(\sigma^2)}{1 + O(1)}.$$

由于当 σ 足够小时，$\frac{\|\Delta\|^2}{2\|\Delta^\perp\|} \ge \tau$ 和 $\frac{1+O(\sigma^2)}{1+O(1)} \ge 1 - C\sigma^2$，我们得出结论

$$\frac{\|p - q\|^2}{2 \cdot d(q, T_p\widehat{\mathcal{M}})} \ge \tau(1 - C\sigma^2).$$

对所有 $p \ne q \in \mathcal{M}$ 取下确界，即完成证明。 □

定理 *5.5* 的证明。由于 Γ 是$\mathcal{M}$的 D维管状邻域，半径为 $C\sigma$，因此它可以被有限个欧几里得球覆盖

$$\Gamma \subset \bigcup_{j=1}^{M} B_D(z_j, \rho),$$

其中 $\rho = c_0\sigma^3$，某个常数为 $c_0 > 0$，且覆盖数满足$M \le C_{\mathrm{cov}}\rho^{-D} = O(\sigma^{-3D})$。

根据推论 5.2.1，对每个中心 $z_j \in \Gamma$，都有局部展开

$$F(z_j) = \pi(z_j) + \tfrac{d}{2}H_{\pi(z_j)}\sigma^2 + O(\sigma^3)$$

其概率至少为 $1 - C_1\exp(-C_2\sigma^{-1})$，其中 $\pi(z_j)$表示z_j 在 $\mathcal{M}$ 上的投影，而 $H_{\pi(z_j)}$ 是 $\pi(z_j)$处的平均曲率向量。对所有中心应用并集界，并注意到 $M \le O(\sigma^{-3D})$仅随 σ^{-1}呈多项式增长，我们得到修正后的常数 C_1', C_2', $c' > 0$，使得

$$\mathbb{P}\Big(\forall j \le M,\ \|F(z_j) - \pi(z_j) - \tfrac{d}{2}H_{\pi(z_j)}\sigma^2\| \le C\sigma^3\Big) \ge 1 - C_1'\exp(-C_2'\sigma^{-c'}).$$

接下来，我们将这一一致界从中心 $\{z_j\}$ 推广到所有 $z \in \Gamma$。由于 F 在紧致集 Γ上是 C^∞，因此 F 是利普希茨连续的，其常数为 L_F。同样，投影映射 π 和平均曲率场 $H_{\pi(\cdot)}$在 Γ上均为 C^1，这意味着它们分别具有利普希茨连续性，常数为 L_π, L_H。因此，对于任意满足 $\|z - z'\| \le \rho$ 的 z, $z' \in \Gamma$，我们有

$$\|F(z) - F(z')\| = O(\rho), \quad \|\pi(z) - \pi(z')\| = O(\rho), \quad \|H_{\pi(z)} - H_{\pi(z')}\| = O(\rho).$$

将这些结果与中心处的展开式相结合，可得

$$F(z) = \pi(z) + \tfrac{d}{2}H_{\pi(z)}\sigma^2 + O(\sigma^3), \quad \text{uniformly for all } z \in \Gamma,$$

其概率至少为 $1 - C_1' \exp(-C_2'\sigma^{-c'})$。因此，像 $F(\Gamma)$ 以高概率是 $\mathbb{R}^D$ 中的光滑 d维子流形。由于$\|F(z) - \pi(z)\| \leq C\sigma^2$ 一致成立，我们得到

$$\sup_{z\in\Gamma} d(F(z), \mathcal{M}) \leq C\sigma^2.$$

相反，对于每个 $x \in \mathcal{M}$，存在 $z = x + O(\sigma)v_x \in T_x^{\perp}\mathcal{M} \subset \Gamma$，使得 $F(z) = x + O(\sigma^2)v_x$，这意味着

$$\sup_{x\in\mathcal{M}} d(x, F(\Gamma)) \leq C\sigma^2.$$

结合这两个界可得

$$\mathrm{dist}_H(F(\Gamma), \mathcal{M}) \leq C\sigma^2$$

其概率至少为 $1 - C_1'\exp(-C_2'\sigma^{-c'})$。此外，根据引理 5.4 的论证， $F(\Gamma)$ 的可达半径满足

$$\tau_F \geq \tau(1 - C\sigma^2) > 0$$

其中某个常数为 $C > 0$。

致谢。 我们感谢 Junhao Chen 就定理 3.3 和 3.5 的证明提供的宝贵讨论。新加坡国立大学的 Z.Y. 获新加坡教育部 Tier 2 项目 A-8001562-00-00 和 Tier 1 项目（A-8000987-00-00 和 A-8002931-00-00）资助；R.L. 是获 Tier 2 项目 A-8001562-00-00 资助的研究员。

□

References

[1] Stephanie Alexander 和 Richard Bishop. 曲率有上界空间中子空间的高斯方程与单射半径. *Geometriae Dedicata* 117 65 84, , : – .[2] 2005
Mikhail Belkin 和 Partha Niyogi. 用于降维和数据表示的拉普拉斯特征映射. *Neural computation*, 15(6):1373–1396, 2003年. [3] Jean-Daniel Boissonnat、Leonidas J Guibas 和 Steve Y Oudot. 使用见证复形在任意维数中进行流形重建. *Discrete & Computational Geometry* 42 1 37 70, 2009 , ( ): – . [4]
Eugenio Calabi. 关于典范类消失的 k¨ahler 流形. 载于 *Algebraic geometry and topology. A symposium in honor of S. Lefschetz*, 第 12 卷，页码 78–89，2015年. [5] Siu-Wing Cheng、Tamal K Dey 和 Edgar A Ramos. 根据点样本进行流形重建. 载于 *SODA*, 第 5 卷，页码 1018–1027，2005. [6] Herman Chernoff. 基于观测值之和的假设检验的渐近效率度量. *The Annals of Mathematical Statistics* 493 507, , 页码 – .[7] 1952
Ronald R. Coifman 和 St´ephane Lafon. 扩散映射. , ( ): – . *Applied and Computational Harmonic Analysis* 21 1 5 30, 2006

[8] Chinh Dang、Ammar Safaie、Mantha Phanikumar 和 Hayder Radha. 使用流形近似估计风速和风向. 载于 *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 页码 328–329, 2015年.
[9] David L. Donoho 和 Carrie Grimes. Hessian特征映射：用于高维数据的局部线性嵌入技术. *Proceedings of the National Academy of Sciences* 100 10 5591 5596, 2003 , ( ): – . [10]

Herbert Federer. 曲率测度. *Transactions of the American Mathematical Society* 93 3 418 491, 1959 , ( ): – .[11]

Charles Fefferman、Sergei Ivanov、Yaroslav Kurylev、Matti Lassas 和 Hariharan Narayanan. 将假定的流形拟合到噪声数据. 载于 *Conference On Learning Theory* 688 720 2018 , 页码 – .

PMLR, .[12] Charles Fefferman、Sergei Ivanov、Yaroslav Kurylev、Matti Lassas 和 Hariharan Narayanan. 流形的重建与插值。I：几何Whitney问题. *Foundations of Computational Mathematics*, 20(5):1035–1133, 2020年.[13] Charles Fefferman、Sergei Ivanov、Matti Lassas、Jinpeng Lu 和 Hariharan Narayanan. 流形的重建与插值 II：具有部分距离数据的黎曼流形反问题. *arXiv:2111.14528*, 2021年. [14] Charles Fefferman、Sergei Ivanov、Matti Lassas 和 Hariharan Narayanan. 将大可达半径流形拟合到噪声数据. *Journal of Topology and Analysis*, 17(02):315–396, 2025年. [15] Keinosuke Fukunaga 和 Larry D. Hostetler. 密度函数梯度的估计及其在模式识别中的应用. *IEEE Transactions on Information Theory* 21 1 32 40, 1975 , ( ): – .[16]

Christopher Genovese、Marco Perone-Pacifico、Isabella Verdinelli 和 Larry Wasserman. 极小极大流形估计. *Journal of Machine Learning Research*, 13 43 1263 1291, 2012 ( ): – .[17] Christopher R Genovese、Marco Perone-Pacifico、Isabella Verdinelli 和 Larry Wasserman. Hausdorff损失下的流形估计与奇异反卷积. *The Annals of Statistics*, 40(2):941–963, 2012年.[18] Christopher R Genovese、Marco Perone-Pacifico、Isabella Verdinelli 和 Larry Wasserman. 非参数脊线估计. *The Annals of Statistics*, 42(4):1511–1545, 2014年.[19] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio. 生成对抗网络. *Advances in neural information processing systems* 27 2014 , , . [20]

Matthias Hein 和 Markus Maier. 流形去噪. *Advances in neural information processing systems* 19 2006 , , . [21]

Wassily Hoeffding. 有界随机变量之和的概率不等式. *Journal of the American statistical association*, 58(301):13–30, 1963 年 .[22] Inki Kim 、 Renato Juliano Martins 、 Jaehyuck Jang 、 Trevon Badloe、Samira Khadir、Ho-Youl Jung、Hyeongdo Kim、Jongun Kim、Patrice Genevet 和 Junsuk Rho. 用于光探测与测距技术的纳米光子学. *Nature nanotechnology*, 16 5 508( ): –524, 2021 年. [23] Zhenglei Li、Weigang Pan、Shuxin Wang、Xiangmeng Tang 和 Hao Hu. 未知噪声环境中基于流形的点云去噪网络. *Infrared Physics & Technology* 132 104735, 2023 , : . [24] Ce Liu、William T Freeman 、Richard Szeliski 和 Sing Bing Kang. 从单幅图像估计噪声. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* 1 901 908, 2006 , : – .[25] Kitty Mohammed 和 Hariharan Narayanan. 使用核密度估计和局部主成分分析进行流形学习. *arXiv:1709.03615*, 2017年.

[26] Partha Niyogi、Stephen Smale 和 Shmuel Weinberger。如何从随机样本中高置信度地求出子流形的同调。*Geometry* 39 419 441, 2008 *Discrete & Computational* , : – 。[27] Emanuel Parzen。概率密度函数和众数的估计。 *The Annals of Mathematical Statistics* 33 3 1065 1076, 1962 , ( ): – 。也称为 Parzen 窗口法。[28] Sam T Roweis 和 Lawrence K Saul。通过局部线性嵌入进行非线性降维。 *science*, 290(5500):2323–2326, 2000。[29] Joshua B Tenenbaum、Vin de Silva 和 John C Langford。用于非线性降维的全局几何框架。 *science*, 290(5500):2319–2323, 2000。[30] Hemant Tyagi、Elif Vural 和 Pascal Frossard。黎曼流形光滑嵌入的切空间估计。 *Information and Inference: A Journal of the IMA* 2 1 69 114, 2013 , ( ): – 。通过局部主成分分析估计切空间；分析采样宽度和曲率效应。[31] 吴梦浩和 ZHIGANG YAO。一种用于聚类和多尺度 RNA 校正的基于主子流形的方法。 *ArXiv*, 2503.20513, 2025年。[32] ZHIGANG YAO、李冰洁、陆宇坤和丘成桐。通过流形拟合进行单细胞分析：RNA 聚类及更多领域的框架。 *Proceedings of the National Academy of Sciences* 121 37 2400002121, 2024 , ( ):e 。[33] ZHIGANG YAO、苏嘉吉、李冰洁和丘成桐。流形拟合。 *preprint arXiv:2304.07680* 2023 *arXiv* , 。[34] ZHIGANG YAO 和夏雨晴。无界噪声下的流形拟合。 *Machine Learning Research* 26 45 1 55, 2025 *Journal of* , ( ): – 。

Zhigang Yao: Department of Statistics and Data Science, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 117546; Center of Mathematical Sciences and Applications, Harvard University, 20 GardenStreet, Cambridge USA 02138

Email address: zhigang.yao@nus.edu.sg; zhigang.yao@cmsa.fas.harvard.edu

Ruowei Li: Department of Statistics and Data Science, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 117546*Email address*: ruoweili@nus.edu.sg