arXiv:2601.06715v1 [math.ST] 10 Jan 2026

重尾目标分布的扩散模型：得分估计与采样保证

Yifeng Yu* Lu Yu†

2026年1月13日

摘要

基于得分的扩散模型已成为生成建模的强大框架，其中得分估计是核心统计瓶颈。现有关于得分估计的保证主要聚焦于轻尾目标分布，或依赖于紧支撑等限制性假设，而这些假设在实践中的重尾数据下往往不成立。在本文中，我们研究目标分布为重尾分布、属于光滑度参数为 $\beta > 0$ 的索np列夫类时的常规（高斯分布）基于得分的扩散模型。我们同时考虑由尾参数 γ 标示的指数尾和多项式尾部衰减。利用核密度估计，我们推导出得分估计的精确极小极大速率，揭示出一种定性二分：在指数尾条件下，该速率与轻尾情形一致，至多相差多对数因子；而在多项式尾部条件下，该速率显式依赖于 γ。我们进一步为相应的连续反向动力学给出采样保证。在全变差意义下，生成分布在指数尾条件下以极小极大最优速率 $n^{-\beta/(2\beta+d)}$ 收敛（至多相差对数因子），而在多项式尾部条件下则以依赖于 γ 的速率收敛。后者的采样速率是否为极小极大最优仍是一个开放问题。这些结果刻画了重尾目标分布下得分估计及由此产生的采样精度的统计极限，将扩散理论拓展到了轻尾设定之外。

目录

1 引言 2 1.1 更多相关工作 .4

2 预备知识 **5**

2.1 基于得分的扩散模型 .5 2.2 建模设定与正则性条件 .6 2.3 基于核的得分估计 .7

3 主要结果：收敛分析 **8**

3.1 多项式衰减 .8 3.2 指数衰减 .9

4 主要结果：得分估计的极小极大下界 **11**

4.1 多项式尾部区域的极小极大下界 .11 4.2 指数尾部情形的极小极大下界 .12 4.3 最不利分布的构造 .12

*清华大学数学科学系 yyf22@mails.tsinghua.edu.cn

†香港城市大学数据科学系 lu.yu@cityu.edu.hk

5 证明概要 **12**
5.1 多项式衰减区间下采样误差的证明概要（定理 4） . 135.1.1 定理 1 的证明概要 .145.1.2 定理 3 的证明概要 .165.2 指数衰减情形下采样误差的证明概要（定理 8） 16

6 讨论 17

7 致谢 **17**

A 附录 **21**
A.1 第 2.3 节的证明 .21A.2 第3.1节的证明 .21A.2.1 定理 1 的证明 .22A.2.2 定理 3 的证明 . 25A.2.3 定理 4 的证明 .27A.3 第 3.2 节的证明 .28A.3.1 定理 5 的证明 .28A.3.2 定理 6 的证明 . 29A.3.3 定理 7 的证明 .29A.3.4 定理 8 的证明 .29A.4 上界辅助引理的证明 .30A.4.1 多项式衰减引理的证明 .30A.4.2 指数衰减引理的证明 .34A.5 多项式衰减目标的替代低噪声策略 .36
A.6 第 4 节的证明 .45A.6.1 定理 9 的证明 .45A.6.2 定理 10 的证明 .50

1 引言

扩散模型已成为现代生成建模中的核心工具，在图像生成 [DN21、HJA20、RDN+22、SME20]，自然语言处理 [PVG+21]和计算生物学 [AA22、XYS+22]方面表现出色。一种常用的表述形式是基于分数的生成模型（SGM），它通过随机微分方程实现 [SSDK+20]。概括来说，SGM 包含一个前向过程，逐步将从目标分布中抽取的数据扰动为噪声；还包含一个反向过程，旨在从噪声中重建目标分布。反向动力学取决于未知的分数函数，该函数指定了每个噪声水平下去噪的局部方向。因此，得分估计决定了反向扩散过程的准确性，通常被视为 SGM 中的核心统计瓶颈。

近年来，大量研究探讨了在扩散模型框架下，得分函数能够被多准确地学习。早期工作 [BMR20] 在 ℓ_2-范数下建立了得分估计保证，但相应的界取决于某些未知假设类的 Rademacher 复杂度。随后， [GPPX24] 表明，采用表达能力足够强的神经网络架构可大幅降低样本复杂度。近期研究进一步推动了使用神经网络进行得分估计。具体而言， [CL24]表明，对于次高斯目标分布，得分函数可以在局部

由参数具有一致界的神经网络近似，从而克服某些高维挑战。 [HRX24] 借助神经切核方法中的工具，推导出使用神经网络进行得分估计的明确样本复杂度保证。与这些可学习性结果互为补充的是， [CKMM25] 将去噪扩散概率模型 [HJA20]得分估计归约为经典的分布学习任务，例如 PAC 密度估计，并利用这一归约证明了得分估计的计算下界，其中包括针对一般高斯混合模型得分估计的密码学困难性。

其他研究探讨了在目标分布具有额外结构假设时的得分估计。例如， [GKL24] 和 [CKS24] 表明，当数据服从高斯混合分布时，可以利用分段低次多项式近似高效地学习分数函数。另一条研究路线考察了由低维结构支撑的数据。在这一方向上， [CHZW23] 表明，当数据位于未知线性子空间中时，神经网络能够准确近似得分函数；而 [ADR24] 则通过精心设计的神经网络，为由有界低维流形支撑的数据建立了准确的得分函数近似。最近， [YP25]和 [YMP25] 在宽松流形假设下研究了去噪和隐式分数匹配，并推导出非渐近近似与泛化界，其速率由内在维数决定。

另一项研究工作针对分数生成模型建立了极小极大收敛速率，刻画了采样误差随样本量和环境维数变化时所能达到的最佳水平。然而，现有分析依赖于目标分布的强正则性假设，例如次高斯尾 [SAL25, ZYLL24, CL25, DKXZ24], 、对密度的显式下界 [OAS23, FGL25]，或紧支撑 [SAL25, DKXZ24, TY24]。这些条件限制性很强，无法适用于实践中遇到的许多重尾分布，例如金融领域 [BT03, Har13]、信号与图像处理领域 [ATB03, BTS05], ，以及涉及测量误差且尾部质量不可忽略的各种科学领域 [AFT98]。因此，当前的极小极大理论并不适用于重尾目标分布，也留下了关于分数生成模型在这一情形下的行为与极限的基本问题。

在这项工作中，我们聚焦于目标分布具有重尾特征时，传统基于得分的扩散模型中的得分估计问题，以弥补这一空白。由于分数函数支配着反向扩散动力学，理解其统计复杂度对于扩散模型的收敛分析至关重要。受此启发，我们研究以下几个基本问题：

1) 在采用高斯加噪的标准基于得分的扩散流程下，当底层数据分布具有重尾特征时，得分估计的极小极大最优速率是多少？

2) 这些得分估计速率如何转化为生成分布与目标分布之间采样误差的界？

我们对这两个问题都给出了明确答案。我们建立了重尾目标分布下得分估计的首个尖锐极小极大保证。具体而言，我们考虑光滑度参数为 $\beta > 0$ 的索ობ列夫类中的密度，并分析两种具有代表性的尾部情形：(i) 指数尾部衰减，其中包括次高斯情形；(ii) 多项式尾部衰减，对应尾部更重的分布。两种情形均由尾参数 γ 刻画。对于每种情形，我们构造了基于核的得分估计量，并证明它们达到了相应的极小极大最优速率。我们还分析了相关的连续反向扩散过程，并为所得采样分布推导出收敛保证。特别是在指数衰减情形下，我们证明生成的分布达到了全变差意义下的极小极大最优采样速率。我们的主要贡献概括如下。

- 我们分析了经典的高斯前向扩散，并建立了多项式尾部情形下得分估计的近乎极小极大最优速率。具体而言，当目标密度具有多项式衰减且尾指数为 γ 时，我们证明基于核的估计量可达到速率 $n^{-\frac{\gamma+1}{d+\gamma+1}} t^{-1-\frac{d}{2}\frac{\gamma+1}{d+\gamma+1}}$（定理 1），从而将分数生成模型（SGM）的极小极大得分估计理论拓展到轻尾情形之外。

- 我们进一步研究由学习得到的得分函数驱动的相关连续反向扩散，并据我们所知，首次推导出全变差采样误差界 $\mathcal{O}\left(n^{-\frac{2\beta}{4\beta\ d+\gamma+1\ (\gamma+1)\,(\,)+d(d+2(\gamma+1))}}\right)$ 在多项式尾部区域（定理 4）。

- 对于指数衰减目标，我们在整个流程的两个阶段均获得了极小极大最优保证：高斯前向过程下的得分估计达到最优速率 $n^{-1}t^{-1-d/2}$（定理 5），而由学习得到的得分函数驱动的连续时间反向扩散在全变差意义下达到最优速率 $n^{-\beta/\ (2\beta+d)}$（定理 8），与次高斯目标情形下的已知结果相符。

综合而言，这些结果为标准基于得分的扩散模型在广泛的重尾目标类别下的得分估计及其产生的采样误差提供了统一的统计学视角。它们还揭示了指数衰减与多项式尾部衰减之间的鲜明差异。对于指数衰减目标，传统的高斯扩散框架在统计上具有稳健性，其速率与次高斯情形下的速率一致。相比之下，对于多项式衰减目标，在相同高斯前向过程下，极小极大得分估计速率明确依赖于尾指数，这种依赖尾指数的惩罚还会传递到采样误差界中，表明尾部越重，统计难度从根本上越高。

1.1 更多相关工作

扩散模型的收敛分析。 受基于得分函数的扩散模型在实证上的成功所启发，越来越多的研究开始考察其收敛性质。现有的大多数分析都假设可以获得准确的得分估计值，并着重研究由此产生的采样动力学的收敛行为。这些工作通常依赖于目标分布的有界二阶矩假设（例如，[CEAD25、CCL+22、YY25、GNZ25、CDS25、CLL23]），或施加紧支撑条件（例如，[DB22、LLT23、ADR24]）。尽管这些研究为基于扩散的采样器的稳定性、离散化误差和收敛行为提供了宝贵见解，但它们将分数函数视为已知，因此绕过了从有限数据中学习得分函数这一根本性的统计挑战。

超越高斯噪声的扩散模型。 另一类研究探讨了这样一类分数生成模型：它们在前向过程中以非高斯噪声取代高斯扰动，例如广义高斯分布 [DSL21], 伽马分布 [NRW21]、t-分布 [PPX+24] 或 α-稳定 Lévy 过程 [YPKL23，SSD25]。这些方法有着共同的动机：增强模型在数据分布不均衡或具有重尾的情形下的鲁棒性，并拓展扩散模型在高斯框架之外的适用范围。然而，这一方向的理论发展仍然有限。与高斯情形不同，非高斯扰动通常无法导出闭式转移密度或易于处理的反向随机微分方程，这使得分析由此产生的动力学或刻画采样误差变得困难。因此，这些研究大多没有给出明确的收敛保证，也没有研究估计相应得分函数的统计难度。

记号。令 $\mathbb{R}^d$ 为 d维欧氏空间。 $L^1(\mathbb{R}^d)$ 表示 $\mathbb{R}^d$ 上可积函数构成的空间。对于 $p \geqslant 1$，定义 L^p范数为 $\|f\|_{L^p} := \left(\int_{\mathbb{R}^d} |f(x)|^p \,\mathrm{d}x\right)^{1/p}$。给定可测权重 $q\colon \mathbb{R}^d \to [0,\infty)$，定义加权范数 $\|f\|_{L^p(q)} := \left(\int_{\mathbb{R}^d} |f(x)|^p q(x) \,\mathrm{d}x\right)^{1/p}$。对于任意随机向量 X，令 $\mathcal{L}(X)$ 表示其分布律，并定义 $\|X\|_{\mathbb{L}_2} := \sqrt{\mathbb{E}[\|X\|^2]}$，其中 $\|\cdot\|$ 表示欧氏范数。我们用 $\mathrm{polylog}(n)$ 简写 $(\log n)^C$，其中 $C > 0$为某个常数。对于$f, g\colon \mathbb{R}^d \to \mathbb{R}$，定义卷积 $(f * g)(x) = \int_{\mathbb{R}^d} f(x-y)g(y)\mathrm{d}y$。对于函数 $f\colon \mathbb{R}^d \to \mathbb{R}$，定义其傅里叶变换为 $\mathcal{F}[f](\omega) = \int_{\mathbb{R}^d} f(x)\, e^{-i\langle \omega, x\rangle} \mathrm{d}x$，其中 $\langle \omega, x\rangle$ 是 $\mathbb{R}^d$中的标准内积。对于多重指标 $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$和 $\omega = (\omega_1, \ldots, \omega_d) \in \mathbb{R}^d$，定义 $\omega^\alpha := \prod_{i=1}^d \omega_i^{\alpha_i}$.$\mathrm{TV}(p, q)$ 和 $\mathrm{KL}(p\|q)$ 分别表示概率测度 p 与 q之间的全变差距离和 Kullback–Leibler散度。

2 预备知识

在本节 中，我们介绍本文所研究的 SGMs，并陈述主要假设 s.

2.1 基于得分函数的扩散模型

框架。 我们考虑如下前向过程

$$\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + g(X_t, t)\mathrm{d}B_t\,,$$

其中，初始点 $X_0 \sim p_0$ 服从目标分布， B_t 表示标准d–维布朗运动。这里，漂移项 $f\colon \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$以及函数 $g\colon \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^{d\times d}$ 均为扩散参数。为确保随机微分方程（SDE）（2.1）定义良好，需要对二元组（f， g）施加一定条件。该二元组的常见选择取决于建模目标；有关详细概述，请参阅 [TZ24]。为便于说明，本文采用最简单的选择，令f（X_t， t） $= 0$且 g（X_t，t） $= I$。在这一选择下，前向过程简化为标准布朗运动

$$\mathrm{d}X_t = \mathrm{d}B_t\,. \tag{2.1}$$

解（X_t）$_{t\in[0,T]}$ 在给定 X_0 时具有闭式条件分布，即

$$X_t = X_0 + \sqrt{t}Z, \quad Z \sim \mathcal{N}(0, I_d)\,.$$

因此，在足够长的时间跨度 T 内，向原始数据 $X_0 \sim p_0$ 添加高斯噪声 B_t 后， X_T 的边缘分布 p_T 对 X_0的依赖会变得很弱；当 p_0 为轻尾分布（例如次高斯分布）时，该分布近似服从高斯分布。相反，当p_0 为重尾分布时，高斯卷积虽能平滑分布，却无法消除重尾行为。在这种情况下，X_T 仅在高概率区域内接近高斯分布，而尾部仍存在偏离。

随后，扩散模型通过反转随机微分方程（2.1）生成新数据，这将得到以下反向随机微分方程

$$\mathrm{d}X_t^{\leftarrow} = \nabla \log p_{T-t}(X_t^{\leftarrow})\mathrm{d}t + \mathrm{d}W_t\,, \tag{2.2}$$

其中 $X_0^{\leftarrow} \sim p_T$，而术语 $\nabla \log p_t$称为 分数函数，对应于 p_t。此外，W_t 表示另一个与 B_t相互独立的标准布朗运动。在温和条件下，以 $X_0^{\leftarrow} \sim p_T$初始化的反向过程 $\{X_t^{\leftarrow}\}_{0\leqslant t\leqslant T}$与前向过程 $\{X_{T-t}\}_{0\leqslant t\leqslant T}$ [And82、CCGL23]具有相同的分布。因此，从 $X_t^{\leftarrow}$到 $t = 0$运行反向扩散过程 T 将生成一个来自目标数据分布 p_0的样本。注意，密度 p_T 为

未知时，可以使用高斯分布 $\hat{p}_T = \mathcal{N}(0, TI_d)$[1]。因此，我们推导出如下定义的反向扩散过程：

$$\mathrm{d}Y_t = \nabla \log p_{T-t}(Y_t)\mathrm{d}t + \mathrm{d}W_t, \quad Y_0 \sim \hat{p}_T\,.$$

分数匹配。处理（2.2）时面临的另一项挑战是，分数函数 $\nabla \log p_t$ 未知，因为分布 p_t 并未显式给出。实践中，通常通过学习一个替代函数，从数据中近似得分函数 s_θ，其依据是分数匹配目标

$$\underset{\theta\in\Theta}{\text{minimize}} \quad \mathbb{E}[\|s_\theta(t, X_t) - \nabla \log p_t(X_t)\|^2]\,,$$

其中 $\{s_\theta : \theta \in \Theta\}$ 是一个表达能力足够强的函数类。这项优化可通过基于神经网络的参数估计器，或核密度分数估计器等非参数方法来实现。尽管神经网络在实际应用中占据主导地位，但由于核方法具有易于处理的统计性质，理论分析中通常采用核方法。本文采取后一种思路，并提出一种基于核的分数函数估计器。

将学习得到的得分估计量 $\hat{s}$ 代入反向过程（2.2），可得以下实用的连续时间反向随机微分方程：

$$\mathrm{d}\hat{Y}_t = \hat{s}_{T-t}(\hat{Y}_t)\mathrm{d}t + \mathrm{d}W_t\,, \quad \hat{Y}_0 \sim \hat{p}_T \tag{2.3}$$

我们的分析聚焦于得分估计，并对生成样本 Y_t 的分布与目标分布 p_0. 之间的全变差距离给出界。该连续逆向随机微分方程的离散化不在本文讨论范围内。实践中，该随机微分方程通过数值格式实现，而相关的离散化偏差已针对 Euler–Maruyama 方法（例如 [CLL23，GNZ25]）、指数积分器（例如 [HO10，ZC22]）、随机中点格式 [GCC24，LJ24，YY25], 以及高阶 Runge-Kutta方法 [HHL25b，HHL25a，WCW24] 进行了研究。

2.2 建模设定与正则性条件

为了在重尾情形下对得分估计误差和采样误差推导严格的界，我们对目标分布p_0的尾部衰减和正则性施加结构性假设。具体而言，我们考虑两种尾部行为，即指数衰减和多项式衰减，并重点研究属于索об列夫光滑类的目标密度，详见下文。

假设 1 (指数尾部衰减**)**。 密度 p_0 满足以下衰减条件，其中常数为 c_1、$C_1 > 0$:

$$p_0(x) \leqslant C_1 \exp(-c_1\|x\|^\gamma), \quad \gamma \in (0, \infty).$$

参数 γ 决定目标分布的尾部行为。当 $\gamma = 2$时， p_0 为高斯分布，呈现轻尾特征且具有所有阶的有限阶矩。 $\gamma = 1$ 对应于拉普拉斯型分布，其仍具有指数衰减，但尾部比高斯分布更重。当 $\gamma \in$ 属于 (0,1) 时，衰减速度较慢，因而分布具有更重的尾部，高阶矩可能不再存在。

假设 2 （多项式尾部衰减）。 目标分布 p_0 满足以下指数衰减条件：存在一个常数 $C_2 > 0$ 使得

$$p_0(x) \leqslant \frac{C_2}{(1 + \|x\|^2)^{(1+\gamma+d)/2}}, \quad \gamma \in (0, \infty).$$

[1]我们注意到，即使 p_0 具有重尾特性，在较大的 T 下使用 $\hat{p}_T = \mathcal{N}(0,\ TI_d)$ 进行初始化，在主体区域内仍然是准确的。远端尾部中任何剩余的差异都可以视为终端初始化误差。

多项式尾部衰减条件涵盖了实践中遇到的广泛重尾分布。典型示例包括多元 Student's t 分布、帕累托族，以及含有重尾成分的有限混合分布。这类分布可能具有无限方差，并且可能不存在高阶有限阶矩，因此这一情形比指数尾部情形的范围更广、挑战也更大。

除尾部行为外，我们还要求 p_0 满足温和的光滑性条件，以控制核密度估计量的逼近误差。

假设 3 (索伯列夫正则性)。 真实分布 p_0 属于索伯列夫类 $\mathcal{P}_{\mathcal{S}}(\beta, L)$，其光滑度参数为 $\beta > 0$ *[Tsy09]*。具体而言，对于 β， $L \in \mathbb{R}_+$，

$$\mathcal{P}_{\mathcal{S}}(\beta, L) = \Big\{ p \in L^1(\mathbb{R}^d) \;\Big|\; p \geqslant 0,\ \int_{\mathbb{R}^d} p(x)\,\mathrm{d}x = 1, \\ \forall \alpha \in \mathbb{N}_0^d \ with \ \sum_{i=1}^d \alpha_i = \beta,\ \int_{\mathbb{R}^d} |\omega^\alpha|^2 \, |\mathcal{F}[p](\omega)|^2 \,\mathrm{d}\omega \leqslant (2\pi)^d L^2 \Big\}.$$

该假设也用于扩散模型得分估计的既有极小极大理论 [ZYLL24]。

2.3 基于核函数的得分估计

为了得到得分函数 $s_t(x) = \nabla \log\ p_t(x)$ 的统计上可处理估计量，我们采用基于核密度估计（KDE）的非参数构造。给定来自数据分布 p_0 的 n 个独立同分布样本 $\{X^{(i)}\}_{i=1}^n$，我们首先使用核函数 φ_t 估计边缘密度 $p_t = \varphi_t * p_0$，然后构造 $\nabla \log\ p_t$ 的代入估计量。为避免低密度区域出现数值不稳定性，我们还对密度估计施加阈值。

高斯核密度估计量和得分估计量。我们将 p_t 的高斯核密度估计量及其梯度定义为

$$\hat{p}_t(x) = \frac{1}{n}\sum_{i=1}^n \varphi_t\Big(x - X^{(i)}\Big), \qquad \nabla \hat{p}_t(x) = \frac{1}{nt}\sum_{i=1}^n \Big(X^{(i)} - x\Big)\varphi_t\Big(x - X^{(i)}\Big),$$

其中， $\varphi_t(x) = \frac{1}{(2\pi t)^{d/2}} \exp\big(-\|x\|^2/2t\big)$ 是高斯转移密度。与标准核密度估计不同，这里的带宽不是调节参数；相反，扩散时间 t 自然地充当带宽。

接下来，我们定义阈值化得分估计量

$$\hat{s}_t(x) := \frac{\nabla \hat{p}_t(x)}{\hat{p}_t(x)} \mathbb{1}_{\{\hat{p}_t(x) \geqslant \rho_n\}}(x), \qquad \rho_n := \frac{\log n}{n(2\pi t)^{d/2}},$$

从而使分　　母在计算该比值的区域内有一个严格正的下界　　.

无偏性。 给定数据集后，$\hat{}p_t$ 和 $\nabla\hat{}p_t$ 分别是 p_t 和 ∇p_t的无偏估计量。特别地，

$$\mathbb{E}[\hat{p}_t(x)] = \int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x-y\|^2}{2t}\right) p_0(y)\,\mathrm{d}y = (\varphi_t * p_0)(x) = p_t(x)$$

and

$$\mathbb{E}[\nabla \hat{p}_t(x)] = \int_{\mathbb{R}^d} \nabla_x \varphi_t(x-y)\, p_0(y)\, \mathrm{d}y = \nabla_x \int_{\mathbb{R}^d} \varphi_t(x-y)\, p_0(y)\, \mathrm{d}y = \nabla p_t(x).$$

这种无偏性是我们误差分析的核心，并将在我们所考虑的重尾情形中反复使用。

3 主要结果：收敛分析

本节在上一节所述的标准扩散框架下，假设目标密度 p_0 具有重尾，为所提出的得分估计量及其后续采样过程建立理论保证。我们首先讨论多项式衰减尾部的情形，然后转而讨论多项式指数衰减尾部的情形。我们在第5 节提供证明概要，概述主要证明的整体结构，突出关键技术难点，并阐明我们的设定和论证与先前工作的不同之处。完整证明见附录。

3.1 多项式衰减

在本部分，我们首先分析得分估计误差，然后在假设目标密度 p_0 具有多项式衰减的重尾的情况下，推导由所提出的得分估计量驱动的逆向随机微分方程（2.3）所生成分布的采样误差界。

现在，我们陈述关于扩散时间 $t \in (0,\ T]$ 处得分估计的主要统计保证。以下定理给出了所提出的估计器 $\hat{s}_t$ 的均方误差界。

定理 1。 假设 p_0 满足假设 2和 3。对于任意 $t \in (0, T]$，所提出的估计器 $\hat{s}_t$ 满足

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\, \mathrm{d}x\right] \lesssim \mathrm{polylog}(n) n^{-\frac{\gamma+1}{d+\gamma+1}} t^{-1-\frac{d}{2}\frac{\gamma+1}{d+\gamma+1}}.$$

定理 1 中的均方误差界会随着 t 的减小而变差，这反映出在极小噪声水平下，得分估计在统计上变得病态。然而，对于下游采样保证而言，相关量并不是固定 t处的点态误差，而是其沿扩散轨迹随时间的累积。因此，我们引入下界 $t_0 \in (0, T)$，并在以下定理中界定 $[t_0, T]$ 上的积分误差。

定理 2。 假设 p_0 满足假设 2 和 3。那么，对于任意 $0 < t_0 < T$，

$$\int_{t_0}^{T} \mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\, \mathrm{d}x\right] \mathrm{d}t \lesssim \mathrm{polylog}(n) n^{-\frac{\gamma+1}{d+\gamma+1}} t_0^{-\frac{d(\gamma+1)}{2(d+\gamma+1)}}.$$

作为对比，现有关于分数生成模型中得分估计的分析通常聚焦于轻尾情形，例如次高斯或紧支集假设。我们的结果通过尾指数 γ，在重尾与轻尾之间进行插值。随着 γ 增大，多项式尾部条件的限制性降低，我们界中的尾部依赖惩罚也随之消失。特别是，

$$\lim_{\gamma \to \infty} \frac{\gamma+1}{d+\gamma+1} = 1.$$

因此，定理 2 中的样本量依赖性趋近于 n^{-1} （相差至多对数因子），这与次高斯假设下得分估计的已知速率一致，相关结果见 p_0 [ZYLL24]。

事实上，定理 2 控制了在扩散时间上积分的得分估计误差 $t \in [t_0, T]$。对 t_0 的依赖不可避免：随着 $t \to 0$，边缘密度 p_t 在尾部变得越来越不规则，而代入估计量 $\hat{s}_t$ 会因低密度区域中的样本稀少而出现方差膨胀。这表明，应将 $t_0 > 0$ 视为正则化参数，并在下游采样保证中避免小噪声区域 $t < t_0$。这样做会引入额外偏差，因为反向时间动力学不再一直演化回 p_0，而是演化至平滑分布 $p_{t_0} = p_0 * \varphi_{t_0}$。下一个定理将量化这种提前停止偏差的全变差距离。

定理 3. 在满足假设 *2*和 *3*且 *Sobolev* 光滑性为 $\beta \in (0, 2]$, 的条件下，定义 $p_{t_0} = p_0 * \varphi_{t_0}$。那么，提前停止偏差的全变差距离上界为

$$\mathrm{TV}(p_0, p_{t_0}) \lesssim t_0^{\frac{\beta(\gamma+1)}{d+2(\gamma+1)+2\beta}}.$$

定理 3 表明，在时刻 t_0 提前停止会产生可控的逼近误差 TV(p_0，p_{t_0})；当 $t_0 \to 0$ 时，该误差会以由 Sobolev 光滑性 β 和尾指数 γ决定的速率趋于零。结合定理 2，这便形成了标准的偏差-方差权衡：选择更小的 t_0 会降低提前停止偏差，但会通过因子 $t_0^{-\frac{d}{2}(\gamma+1)/(d+\gamma+1)}$ 增大得分估计误差。这种权衡促使我们选择 t_0，以平衡最终采样界中的两项，如下所示。

定理 4. 假设 p_0 满足假设 *2* 和*3*，其 *Sobolev* 光滑度参数为 $\beta \in (0, 2]$。令

$$t_0 = n^{-\frac{2(d+2(\gamma+1))}{4\beta(d+\gamma+1)+d(d+2(\gamma+1))}} \quad \text{and} \quad T = n^{\frac{4\beta(\gamma+1)}{4\beta(d+\gamma+1)+d(d+2(\gamma+1))}}.$$

那么，存在一个得分估计量 $(\hat{s}_t)^{t\in[0,T]}$，使得从相应的反向时间扩散过程 Y_{T-t_0} 中采样 (2.3) 满足

$$\mathbb{E}\left[\mathrm{TV}\left(p_0, \mathcal{L}(\hat{Y}_{T-t_0})\right)\right] \lesssim \mathrm{polylog}(n)\, n^{-\frac{2\beta(\gamma+1)}{4\beta(d+\gamma+1)+d(d+2(\gamma+1))}}.$$

随着 $\gamma \to \infty$对应尾部逐渐变轻，我们界中依赖尾部的退化也随之消失。在此极限下，全变差速率简化为 $n^{-\beta/(2\beta+d)}$，与 [ZYLL24] 中研究的次高斯情形以及非参数密度估计的经典极小极大速率一致 [Tsy09]。

替代低噪声策略。为求完整，我们也在附录A.5 中探讨了一种替代方法：在扩散时间 t较小的低噪声区域，使用解耦核分数估计器，从而避免提前停止。此变体需要略强的假设，且速率慢于定理 4，因此我们将其作为补充视角，而非主要保证。

3.2 指数衰减

本节研究目标分布 p_0 具有指数衰减尾部时，扩散模型中得分估计量的收敛性以及由此产生的采样误差。主要

区别在于 p_0的尾部行为不同于多项式尾部情形。与多项式尾部区域的分析相类似，我们首先在指数尾部假设下建立得分估计量的逐点均方误差界。

定理 5. 假设假设 *1* 成立。对于任意 $t \in (0, T]$，所提出的估计器 $\hat{s}_t$ 满足：

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x) \mathrm{d}x\right] \leqslant \mathrm{polylog}(n) n^{-1} t^{-1-d/2}.$$

与定理 1 相比，最显著的区别在于对样本量 n 的依赖。多项式尾部区域的速率为 $n^{-\frac{\gamma+1}{d+\gamma+1}}$，会随着尾部变得更重（$\gamma$ 更小）而明确下降；而无论 γ 取何值，指数衰减情形都能以对数因子为代价恢复近参数化速率 n^{-1}。这表明，标准扩散模型流程在统计上对广泛的指数尾部行为具有稳健性。下面的定理给出了累积分数估计误差的上界。

定理 6. 假设 p_0 满足假设 *1* 和 *3*，对于任意 $T > 0$ 以及任意 $0 < t_0 < T$，有

$$\int_{t_0}^{T} \mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\, \mathrm{d}x\right] \mathrm{d}t \lesssim \mathrm{polylog}(n) n^{-1} t_0^{-d/2}.$$

在假设 1下，目标分布 p_0 具有拉伸指数尾，其中次高斯情形是其一个特例，同时参数 γ允许尾部呈现更一般的衰减形态。定理6 表明，这种额外的一般性并不会降低统计速率，因为在 $[t_0, T]$ 上的累积分数误差仍按 $\mathrm{polylog}(n) n^{-1}$ 的量级增长，且对小噪声的依赖性为显式形式 $t_0^{-d/2}$。特别是，在次高斯情形下，该上界恢复了相同的速率 [ZYLL24，推论 3。7]。

定理 6中的界随着 t_0 趋于零而变差，这反映了低噪声区域的病态性。因此，我们在时间 t_0 停止反向动力学，并以 $p_{t_0} = p_0 * \varphi_{t_0}$为目标，这会产生提前停止偏差 $\mathrm{TV}(p_0, p_{t_0})$。下一个定理将控制指数尾下的这一偏差。

定理 7. 假设 p_0 满足假设 *1* 和假设 *3*，其中 $\beta \in (0, 2]$。那么，提前停止偏差满足

$$\mathrm{TV}(p_0, p_{t_0}) \lesssim t_0^{\beta/2} \cdot (\log(1/t_0))^{\frac{d}{2\gamma}}.$$

特别地，忽略多对数因子后，该速率由 $t_0^{\beta/2}$ 主导。

基于这一结果以及定理 6中的得分估计保证，我们现在给出目标分布 p_0 与反向时间扩散动力学（2.3）生成的样本分布之间的全变差距离界。

定理 8。 假设假设条件 *1* 和 *3* 成立。令 $t_0 = n^{-\frac{2}{2\beta+d}}$ 和 $T = n^{\frac{2\beta}{2\beta+d}}$。那么，存在一个得分估计量 $(\hat{s}_t)^{t\in[0,T]}$，使得对应反向时间扩散过程 $\hat{Y}_{T-t_0}$ 的样本满足 (2.3)

$$\mathbb{E}\big[\mathrm{TV}\big(p_0, \mathcal{L}(\hat{Y}_{T-t_0})\big)\big] \lesssim \mathrm{polylog}(n)\, n^{-\frac{\beta}{2\beta+d}}.$$

对 t_0 和 T 的选择与先前针对次高斯目标的分析一致，并得到相同的收敛速率 $n^{-\beta/(2\beta+d)}$ [ZYLL24，定理 3。8]。允许 p_0 具有拉伸指数尾，并不会使收敛速率相比次高斯或紧支撑情形变差。这表明，传统扩散模型在整个指数衰减区域内都具有鲁棒性。

比较我们对多项式衰减情形和指数衰减情形的分析，可以看出两者在性质上存在明显的二分。在多项式情形下，尾指数 γ 通过因子 $n^{-(\gamma+1)/(d+\gamma+1)}$显式出现在定理 2中。相比之下，在指数衰减情形下，定理 $\mathcal{O}(n^{-1})$中的收敛速率提升为 6（仅相差对数因子）；在相应的采样保证（定理8）中，γ 对主导项的影响实际上消失了。直观上，这源于大部分概率质量所在区域的几何特征。对于多项式衰减尾部，为了捕获大部分质量，感兴趣区域必须以多项式 速率随 n扩大。因此，必须覆盖的定义域会迅速增大，而可用样本则稀疏地分布在这个大定义域上。对于指数衰减尾部，包含几乎全部质量的区域仅随对数 增长， n因此其体积增长得慢得多。在这种情况下，估计误差的尺度为 n^{-1}（仅相差多对数因子），而尾指数 γ 仅通过这些对数项产生影响，这解释了它为何看似从主收敛速率中消失。

4 主要结果：得分估计的极小极大下界

为补充我们的上界结果，并证明所提出的得分估计量在重尾情形下的速率最优性，我们为得分估计误差建立了相匹配的极小极大下界。主要技术挑战在于构造一个假设族（候选数据分布），使其在统计上难以区分，同时诱导出彼此充分分离的得分函数。下面，我们首先给出多项式衰减和指数衰减两种情形下的极小极大下界。随后，我们将介绍驱动该构造并作为主要技术要素的最不利族。由于两种尾部情形的证明遵循相同的总体策略，且多项式衰减情形在技术上更为复杂，我们将重点讨论该情形。完整证明见附录 A.6。

4.1 多项式尾部区域下的极小极大下界

在本节中，我们将证明定理 1中得到的均方误差界几乎达到了极小极大最优。

定理 9。 令 $\mathcal{H}_\gamma$ 表示满足假设2 且具有$_{2(d+2\gamma+2)}$尾指数 γ的重尾分布类。对于任意足够大的 n 以及满足 $n^{-\frac{1}{d(d+2(\gamma+1))+2\beta(d+\gamma+1)}} \lesssim t \lesssim 1$的所有扩散时间 t， 极小极大风险满足

$$\inf_{\hat{s}_t} \sup_{p_0 \in \mathcal{H}_\gamma \cap \mathcal{P}_S(\beta,L)} \mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\mathrm{d}x\right] \gtrsim n^{-\frac{\gamma+1}{d+\gamma+1}-\delta} t^{-1-\frac{\gamma+1}{d+\gamma+1}+\delta},$$

在对数因子范围内，适用于任意小的 $\delta > 0$。

该下界与定理1 中的速率仅相差对数因子，从而确立了其极小极大最优性。特别是，对尾指数的依赖 γ 是内在的，无法在整个函数类上统一改进。

4.2 指数尾部情形下的极小极大下界

在此 在这一部分，我们证明定理5中的均方误差率几乎达到了极小极大最优 l.

定理 10. 令 $\mathcal{H}'_\gamma$ 表示满足假设 *1* 且2尾指数为 γ 的重尾分布类。对于所有足够大的 n 以及所有扩散时间 $t \gtrsim n^{-\,d+2\beta}$，极小极大风险满足

$$\inf_{\hat{s}_t} \sup_{p_0 \in \mathcal{H}'_\gamma \cap \mathcal{P}_S(\beta, L)} \mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x) \mathrm{d}x\right] \gtrsim n^{-1}\, t^{-1-d/2}.$$

该下界与定理 5 的收敛速率相符，至多相差对数因子，因此确立了极小极大最优性。

4.3 最不利分布的构造

下界的证明采用 Fano 方法，并将其应用于精心设计的目标分布有限子集。与经典非参数下界不同，后者通常在固定的紧域（如 $[-1,1]^d$）上对密度进行扰动；而在我们的设定中，为体现多项式衰减，必须将扰动置于尾部区域。

尾部中的扰动区域。 我们在包含于 $[R, 2R]^d$ 的超立方网格上引入局部扰动，其中 $R > 0$ 是一个较大的半径，将根据样本量 n 和时间 t 来选取。这个选择有两个重要原因：

- 在尾部区域，基准密度很小（数量级为 $\|x\|^{-(d+\gamma+1)}$），这使假设之间的 KL 散度保持较小，因此难以区分。
- 与此同时，我们需要相应的得分函数保持足够大的差异。可实现的差异由衰减的基准密度与扰动幅度之间的权衡决定，而扰动幅度取决于尾指数 γ。

扰动假设。 设 q_0 是一个形式为 $q_0(x) \propto (1+\|x\|^2)^{-(d+\gamma+1)/2}$ 的参考重尾密度。固定间距为 h 的网格 $\{x_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}} \subset [R, 2R]^d$，其中 $\mathcal{I}$ 表示相应的索引集。对于二进制向量 $b = (b_{\mathbf{i}})_{\mathbf{i}\in\mathcal{I}} \in \{0,1\}^{|\mathcal{I}|}$，定义扰动后的密度

$$q_b(x) = q_0(x) + \epsilon \sum_{\mathbf{i}\in\mathcal{I}} b_{\mathbf{i}}\, \omega(x - x_{\mathbf{i}}),$$

其中 ω 是一个光滑的隆起函数，其支撑尺度为 h。随后，我们选取一个有限子集 $\mathcal{B} \subseteq \{0,1\}^{|\mathcal{I}|}$，并考虑假设族 $\{q_b\}_{b\in\mathcal{B}}$。

参数耦合。 下界源于对扰动幅度 ϵ、尾部位置 R 和网格尺度 h 的精细平衡。当 $n \to \infty$ 时，为了利用较小的基准密度，扰动必须放置在更靠近尾部的位置（因此 $R \to \infty$），而 ϵ 则必须缩小，以满足信息论约束。具体而言，我们选择

$$R \asymp n^{\frac{1}{d+\gamma+1}}\, t^{-\frac{d}{2(d+\gamma+1)}},$$

这使样本量与多项式尾部所诱导的几何结构相耦合。

5 证明概要

本节概述定理4和8的高层次证明概要。

5.1 多项式衰减区间中的采样误差证明概要（定理 4）

该证明将我们的得分估计和提前停止结果统一起来，从而推导出最终的采样保证。我们利用三角不等式，将数据分布 p_0 与ˆ生成分布 $\mathcal{L}(Y_{T-t_0})$ 之间的全变差误差分解。首先，

$$\mathrm{TV}\big(p_0, \mathcal{L}(\hat{Y}_{T-t_0})\big) \leqslant \mathrm{TV}(p_0, p_{t_0}) + \mathrm{TV}\big(p_{t_0}, \mathcal{L}(\hat{Y}_{T-t_0})\big),$$

˜它将提前停止偏差与反向过程误差区分开来。接下来，令 Y 表示使用相同的学习得分函数、但从 p_T 初始化的反向过程。那么，

$$\mathrm{TV}\big(p_{t_0}, \mathcal{L}(\hat{Y}_{T-t_0})\big) \leqslant \mathrm{TV}\big(p_{t_0}, \mathcal{L}(\tilde{Y}_{T-t_0})\big) + \mathrm{TV}\big(\mathcal{L}(\tilde{Y}_{T-t_0}), \mathcal{L}(\hat{Y}_{T-t_0})\big).$$

第一项对应于得分估计误差。对于第二项，反向核下的全变差收缩给出

$$\mathrm{TV}\big(\mathcal{L}(\tilde{Y}_{T-t_0}), \mathcal{L}(\hat{Y}_{T-t_0})\big) \leqslant \mathrm{TV}\big(p_T, \mathcal{N}(0, TI_d)\big),$$

我们将其解释为初始化误差。结合这些界，可得如下分解。

$$\mathrm{TV}(p_0, \mathcal{L}(\hat{Y}_{T-t_0})) \leqslant \underbrace{\mathrm{TV}(p_0, p_{t_0})}_{\text{Early Stopping Bias}} + \underbrace{\mathrm{TV}(p_{t_0}, \mathcal{L}(\tilde{Y}_{T-t_0}))}_{\text{Score Matching Error}} + \underbrace{\mathrm{TV}(p_T, \mathcal{N}(0, TI_d))}_{\text{Initialization Error}}.$$

分数匹配误差（定理 1 和 2）。根据 Pinsker 不等式和 Girsanov 定理，采样误差受累积分数估计误差的平方根控制。

$$\mathrm{TV}\big(\mathcal{L}(\tilde{Y}_{T-t_0}), p_{t_0}\big) \lesssim \left(\int_{t_0}^{T} \mathbb{E}\Big[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\,\mathrm{d}x\Big]\mathrm{d}t\right)^{1/2}.$$

将此结果与定理 2中的重尾速率结合，可得

$$\mathrm{TV}\big(\mathcal{L}(\tilde{Y}_{T-t_0}), p_{t_0}\big) \lesssim \mathrm{polylog}(n)\, n^{-\frac{\gamma+1}{2(d+\gamma+1)}}\, t_0^{-\frac{d(\gamma+1)}{4(d+\gamma+1)}}.$$

该项随着 $t_0 \to 0$发散，因此倾向于选择较大的截断值 t_0 ，以避开小噪声区域。

提前停止偏差（定理 3）。为避开 $t = 0$附近条件数不佳的小噪声区域，我们在 t_0 处停止反向过程，并以平滑分布 $p_{t_0} = p_0 * \varphi_{t_0}$ 作为目标，而不是 p_0。定理3 表明，这会引入全变差偏差

$$\mathrm{TV}(p_0, p_{t_0}) \lesssim t_0^{\frac{\beta(\gamma+1)}{d+2(\gamma+1)+2\beta}},$$

该项以由光滑度 β 和尾指数 γ决定的速率随 $t_0 \to 0$ 消失。因此，该项倾向于选择较小的 t_0。

初始化误差（引理 19）。潜在分布 $p_T = p_0 * \varphi_T$与高斯先验 $\mathcal{N}(0, TI_d)$ 之间的差异由引理 19控制，从而得到

$$\mathrm{TV}(p_T, \varphi_T) \lesssim T^{-1/2}.$$

因此，只要选择足够大的 T，就可以使初始化误差可以忽略不计。

最终的极小极大速率通过平衡提前停止偏差与得分匹配误差之间的权衡来实现。令这两个速率相等，可得最优停止时间 t_0 以及所述收敛速率。

下面，我们给出定理 1 和3 的证明概要，解释重尾情形下出现的主要技术挑战，并重点说明我们的证明技术与现有针对轻尾情形的分析方法有何不同。

5.1.1 定理1的证明概要

我们的主要方法是通过阈值化论证来分解误差。我们将定义域按照 (i) 真实密度的大小以及 (ii) 对估计量进行的阈值化处理进行划分，以控制得分估计量 $\hat{s}_t(x)$ 的加权平均积分平方误差（平均积分平方误差）。

首先，我们根据真实密度 p_t 进行分解。我们将其划分为 $\mathbb{R}^d$ 高密度主体区域

$$\mathcal{G}_1 := \{x \in \mathbb{R}^d : p_t(x) > c_\alpha \rho_n\}$$

和低密度尾部区域

$$\mathcal{G}_2 := \{x \in \mathbb{R}^d : p_t(x) \leqslant c_\alpha \rho_n\},$$

其中， ρ_n 是给定的截断阈值， $c_\alpha > 0$ 是一个待指定的常数。其次，我们利用估计量中内置的阈值化机制：按照构造，当估计密度低于该阈值时，即当 $\hat{p}_t(x) < \rho_n$ 时，$\hat{s}_t(x)$ 被设为零。

这两种分解通过 p_t 与 $\hat{p}_t$ 之间的关系相互联系起来，而该关系由一致性集中界控制。

引理 11。 对于任意固定的 $\alpha > 0$，存在一个常数 $C_\alpha > 0$，其仅取决于 p_0 和维度 d，使得以至少 $1 - n^{-\alpha}$ 的概率，

$$\left|\hat{p}_t(x) - p_t(x)\right| < C_\alpha \left(\frac{\log n}{n(2\pi t)^{d/2}} + \sqrt{\frac{p_t(x) \log n}{n(2\pi t)^{d/2}}} \right), \qquad \textit{for all } x \in \mathbb{R}^d,$$

其中 C_α =最大值$\{\sqrt{8\alpha}, 16\alpha/3\}$。

令 A_α 表示引理 11成立的高概率事件。结合该事件与主体区/尾部区划分，我们将误差分解为

$$\begin{aligned}
\mathbb{E}\Big[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\,\mathrm{d}x\Big] = &\underbrace{\int_{\mathcal{G}_1} \mathbb{E}[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}_{A_\alpha}]\, p_t(x)\,\mathrm{d}x}_{\mathrm{I}} \\
&+ \underbrace{\int_{\mathcal{G}_2} \mathbb{E}[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}_{A_\alpha}]\, p_t(x)\,\mathrm{d}x}_{\mathrm{II}} \\
&+ \underbrace{\int_{\mathbb{R}^d} \mathbb{E}[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}_{A_\alpha^c}]\, p_t(x)\,\mathrm{d}x}_{\mathrm{III}}.
\end{aligned}$$

在项 I 和 II 中，我们进一步根据 $\hat{p}_t(x) \geqslant \rho_n$ （活跃区域，即估计得分函数的区域）或 $\hat{p}_t(x) < \rho_n$ （截断区域，即将得分函数设为零的区域）进行划分。

步骤 1：分母的下界。 主要技术难点来自于不稳定的分母

$$\hat{s}_t(x) = \frac{\nabla \hat{p}_t(x)}{\hat{p}_t(x)} \mathbb{1}_{\{\hat{p}_t(x) \geqslant \rho_n\}}.$$

我们首先推导不同区域内 $\hat{p}_t(x)$ 的下界：在交集 $\mathcal{G}_1 \cap A_\alpha$上，引理 11表明 $\hat{p}_t(x) \asymp p_t(x)$，特别地，$\hat{p}_t(x) \geqslant \frac{1}{2} p_t(x)$。在其余活跃区域中，我们使用显式阈值 $\hat{p}_t(x) \geqslant \rho_n$。

步骤 2：化归为逐点均方误差界。一旦对分母给出上界控制，得分误差便可化归为 $\hat{p}_t$ 和 $\nabla \hat{p}_t$ 的误差的线性组合。这些

误差可由以下逐点均方误差估计进行界定，而这些估计的关键在于其随真实密度$p_t(x)$进行缩放。

命题 12 ($\hat{}p_t$ **的均方误差)。** $\hat{}p_t(x)$ 的均方误差满足

$$\mathrm{MSE}(\hat{p}_t(x)) \;\leqslant\; \frac{p_t(x)}{n(2\pi t)^{d/2}}\,.$$

命题 13 ($\nabla\hat{}p_t$ **的均方误差)。** $\nabla\hat{}p_t(x)$ 的均方误差满足

$$\mathrm{MSE}(\nabla\hat{p}_t(x)) \;\leqslant\; \frac{p_t(x)}{nt(2\pi t)^{d/2}}\,.$$

步骤 3：利用主体区域与尾部引理进行积分。 将这些均方误差界代入项 I 和 II 后，会得到关于 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 的积分，其权重涉及 $p_t(x)$ 和 $s_t(x)$。接下来的两个引理对这些积分进行控制，并量化重尾假设下主体区域和尾部区域的几何性质。

引理 14 (亦即引理16)。 在假设 *2* 下，我们有

$$|\mathcal{G}_1| \lesssim \big(C_\gamma(t)\rho_n\big)^{-\frac{d}{1+\gamma+d}},$$
$$\int_{\mathcal{G}_1} \|s_t(x)\|^2\,\mathrm{d}x \lesssim \big(C_\gamma(t)\big)^{-\frac{1}{m}}\, m\, t^{-1}\rho_n^{-\frac{1}{m}-\frac{d}{(1+\gamma+d)m'}}.$$

引理 15 (亦即引理17)。 在假设 *2* 下，有

$$\int_{\mathcal{G}_2} p_t(x)\,\mathrm{d}x \lesssim \big(C_\gamma(t)\big)^{\frac{d}{d+\gamma+1}}\rho_n^{\frac{\gamma+1}{d+\gamma+1}},$$
$$\int_{\mathcal{G}_2} \|s_t(x)\|^2 p_t(x)\,\mathrm{d}x \lesssim \varepsilon^{-1}t^{-1}\big(C_\gamma(t)\big)^{\frac{d(1-\varepsilon)}{d+\gamma+1}}\rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{d+\gamma+1}}.$$

这里 m，$m'>1$ 满足 $1/m+1/m'=1$，$C_\gamma(t)$ 表示 γ 的一个函数，t 和 d，其显式形式见附录。结合项 I–III 的界，并通过选择 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 来最优平衡各项贡献，可得 ρ_n，从而得到定理 1 所述的收敛速率。

注释 1（重尾情形中的统计瓶颈）。我们注意到，定理 $n^{-\frac{\gamma+1}{d+\gamma+1}}$ 中的速率 1 并非宽松分析造成的假象。它反映了重尾的两种内在几何效应，这些效应从根本上改变了相较于传统次高斯情形的偏差—方差权衡：

- **主体区域中的多项式体积扩张（$\mathcal{G}_1$）**。在轻尾情形下，高密度区域 $\mathcal{G}_1$（估计在此处较为稳定）的体积通常仅随 n 对数增长。而在重尾情形下，引理 14表明，$|\mathcal{G}_1|$ 随 $\rho_n^{-d/(d+\gamma+1)}$ 呈多项式增长。这种快速扩张会稀释有效样本密度，并增大主体区域中的积分方差。

- **尾部中的概率质量很大（$\mathcal{G}_2$）**。$\mathcal{G}_2$ 中的截断误差由低于阈值 ρ_n 的概率质量控制。对于次高斯尾，该概率质量呈指数衰减，因此可以进行激进的截断，同时将误差降至可忽略的程度。相比之下，引理15表明，多项式尾部在 $\mathcal{G}_2$中仍保留不可忽略的概率质量，其尺度关系为 $\rho_n^{(\gamma+1)/(d+\gamma+1)}$，这迫使我们更保守地选择 ρ_n，并在主体区域的方差与尾部偏差之间形成紧张的权衡。

注释 2（低噪声区域中的得分估计困难）。我们的分析还阐明了为什么朴素的核估计量会在 $t \to 0$ 时变得次优。在无界设置下，指数衰减目标和多项式衰减目标都会产生尾部区域，其中数据点极为稀疏。随着 t 减小，有效核尺度 $\sqrt{t}$ 变得小于这些区域中样本之间的典型距离，因此估计量几乎无法获得数据，其局部方差也会急剧增大。在紧支撑条件下不会出现这一现象，此时主要挑战是边界偏差，而不是大范围的空缺区域。

关键区别在于这些稀疏区域承载了多少概率质量。对于指数型衰减，远尾部的概率质量下降得极快，因此高方差区域对 L^2 误差的贡献始终可以忽略不计，收敛速率也可以达到紧支撑情形的水平（例如参见 [DKXZ24]）。相比之下，对于多项式衰减， p_0 的尾部下降得慢得多，因此这些稀疏区域仍然包含相当可观的概率质量。密度p_t 下降得不够快，无法抵消方差爆炸的影响，最终这些区域会主导全局估计误差，并使收敛速率变差。

5.1.2 定理3的证明概要

证明首先将全变差距离拆分为来自半径为 R 的中心区域和互补尾部区域的贡献。在中心区域中，我们通过 H¨older 不等式，利用 L^2-误差来界定 L^1-误差，这会引入一个与体积平方根成正比的因子，即 $R^{d/2}$。随后，在频域中利用普朗歇尔定理控制 L^2 平滑误差 $\|p_0 - p_{t_0}\|_{L^2}$：利用 p_0 的索伯列夫正则性 β，我们得到 $t_0^{\beta/2}$ 阶的衰减。在尾部区域中，误差由重尾假设控制；该假设意味着半径为 R 的球外部的概率质量以 $R^{-(\gamma+1)}$ 阶多项式衰减。综合这些估计可得如下形式的上界

$$R^{d/2} t_0^{\beta/2} \;+\; R^{-(\gamma+1)}.$$

所述速率可通过选择 R 来平衡中心区域中的平滑误差与尾部的截断误差。

5.2 指数衰减情形下采样误差的证明概要（定理 8）

证明步骤与多项式衰减区间中的相同，但在指数尾条件下，得分估计界所依赖的几何控制要强得多。因此，得分匹配项不再受到多项式重尾惩罚的影响。

我们从全变差误差的相同分解开始

$$\mathrm{TV}(p_0, \mathcal{L}(\hat{Y}_{T-t_0})) \leqslant \underbrace{\mathrm{TV}(p_0, p_{t_0})}_{\text{Early Stopping Bias}} + \underbrace{\mathrm{TV}(p_{t_0}, \mathcal{L}(\tilde{Y}_{T-t_0}))}_{\text{Score Matching Error}} + \underbrace{\mathrm{TV}(p_T, \mathcal{N}(0, TI_d))}_{\text{Initialization Error}}.$$

得分匹配误差（定理 5 和 6）。与之前一样，Pinsker 不等式和 Girsanov 定理将采样误差控制在累积得分均方误差的平方根范围内。关键区别在于，在指数衰减情形下，相关几何量的表现要好得多：有效支撑体积 $|\mathcal{G}_1|$仅以对数速度增长（引理 21），尾部质量则以超多项式速度衰减（引理22）。将这些界代入得分估计的均方误差界，可得更快的累积速率

$$\mathrm{TV}\big(\mathcal{L}(\tilde{Y}_{T-t_0}), p_{t_0}\big) \;\lesssim\; \mathrm{polylog}(n)\, n^{-1/2}\, t_0^{-d/4}.$$

特别是，对 n的依赖近乎参数化，这体现了扩散流程在指数尾分布下的稳健性。

提前停止偏差（定理7）。在时间 t_0 停止反向动力学，同样会将p_0 替换为平滑目标分布 $p_{t_0} = p_0 * \varphi_{t_0}$。在指数衰减情形下，截断项可以忽略，偏差完全由索伯列夫正则性决定：

$$\mathrm{TV}(p_0, p_{t_0}) \lesssim t_0^{\beta/2}.$$

因此，与多项式情形一样，这一项倾向于令 t_0 取较小值。
初始化误差（引理19）。与高斯先验之间的初始化失配 $p_T = p_0 * \varphi_T$ 满足与之前相同的界：

$$\mathrm{TV}(p_T, \varphi_T) \lesssim T^{-1/2},$$

并且通过令 T 足够大，可以使其变得可以忽略。

通过平衡提前停止偏差 $t_0^{\beta/2}$ 与得分匹配误差$n^{-1/2}t_0^{-d/4}$，可得到最终收敛速率。由此得到最优截断阈值

$$t_0 \asymp n^{-\frac{2}{2\beta+d}},$$

以及相应的收敛速率

$$\mathrm{TV}\big(p_0, \mathcal{L}(\hat{Y}_{T-t_0})\big) \lesssim \mathrm{polylog}(n)\, n^{-\frac{\beta}{2\beta+d}}.$$

6 讨论

在本文中，我们刻画了重尾目标分布的标准得分生成模型框架中，得分估计的基本极小极大界以及由此产生的采样误差。我们表明，当目标密度呈指数衰减时，标准高斯加噪流程对尾部行为具有统计稳健性。相比之下，对于多项式衰减目标，这种稳健性会失效，而如何设计有原则的补救方法仍是一个开放问题。

一种可能的方向是采用重尾噪声（例如稳定型或 Lévy 型扰动 [YPKL23，SSD25]）来修改前向过程。然而，这将超出高斯扩散族的范畴，并引入截然不同且更具挑战性的分析问题，因此，将极小极大理论扩展到这一领域并非易事，也超出了本文的研究范围。另一种有前景的方法是考虑经过调节的或数据依赖噪声调度，以避免在主导统计误差的区域放置过多质量；对于这类补救方法的系统研究，我们将留待未来工作。

此外，在多项式衰减区间内，我们证明了得分估计的极小极大最优性，并为由学习得到的得分驱动的连续时间反向扩散的采样误差提供了保证，但目前尚不清楚该采样速率是否为极小极大最优。要弥合这一差距，要么需要针对采样问题建立匹配的极小极大下界，要么需要进行更精细的分析，以更准确地刻画得分估计误差如何通过反向扩散动力学传播。这个问题将留待未来工作。

7 致谢

本研究得到香港城市大学启动基金和香港研究资助局资助项目 21306325 的支持。

参考文献

[AA22] Namrata Anand 和 Tudor Achim，利用等变去噪扩散概率模型进行蛋白质结构与序列生成，arXiv 预印本 arXiv:2205.15019（2022年）。[ADR24] Iskander Azangulov、George Deligiannidis 和 Judith Rousseau，流形假设下高维扩散模型的收敛性，arXiv 预印本 arXiv:2409.18804（2024年）。[AFT98] Robert Adler、Raya Feldman 和 Murad Taqqu，《重尾实用指南：统计技术与应用》，Springer Science & Business Media，1998年。[And82] Brian DO Anderson，反向时间扩散方程模型，《随机过程及其应用》12（1982年），第3期，313–326页。[ATB03] Alin Achim、Panagiotis Tsakalides 和 Anastasios Bezerianos，基于重尾建模的贝叶斯小波收缩合成孔径雷达图像去噪，《IEEE 地球科学与遥感汇刊》41（2003年），第8期，1773–1784页。[BMR20] Adam Block、Youssef Mroueh 和 Alexander Rakhlin，利用去噪自编码器和朗之万采样进行生成建模，arXiv 预印本 arXiv:2002.00107（2020年）。[BT03] Brendan O Bradley 和 Murad S Taqqu，金融风险与重尾，《金融重尾分布手册》，爱思唯尔，2003年，第35–103页。[BTS05] Alexia Briassouli、Panagiotis Tsakalides 和 Athanasios Stouraitis，重尾中的隐藏消息：基于α稳定模型的DCT域水印检测，《IEEE多媒体汇刊》7（2005年），第4期，700–715页。[CCGL23] Patrick Cattiaux、Giovanni Conforti、Ivan Gentil 和 Christian L´eonard，有限熵条件下扩散过程的时间反演，《Annales de l'Institut Henri Poincar´e (B) Probabilit´es et Statistiques》，第59卷，Institut Henri Poincar´e，2023年，第1844–1881页。[CCL+22] Sitan Chen、Sinho Chewi、Jerry Li、Yuanzhi Li、Adil Salim 和 Anru R Zhang，采样如同学习得分函数一样容易：具有最小数据假设的扩散模型理论，arXiv 预印本 arXiv:2209.11215（2022年）。[CDS25] Giovanni Conforti、Alain Durmus 和 Marta Gentiloni Silveri，最小数据假设下基于得分的扩散模型的KL收敛保证，《SIAM数据科学数学期刊》7（2025年），第1期，86–109页。[CEAD25] Paula Cordero-Encinar、O Deniz Akyildiz 和 Andrew B Duncan，扩散退火朗之万蒙特卡洛用于生成式建模的非渐近分析，arXiv 预印本 arXiv:2502.09306（2025年）。[CHZW23] Minshuo Chen、Kaixuan Huang、Tuo Zhao 和 Mengdi Wang，低维数据上扩散模型的得分函数近似、估计与分布恢复，国际机器学习会议，PMLR，2023年，第4672–4712页。[CKMM25] Sinho Chewi、Alkis Kalavasis、Anay Mehrotra 和 Omar Montasser，DDPM得分匹配与分布学习，arXiv 预印本 arXiv:2504.05161（2025年）。[CKS24] Sitan Chen、Vasilis Kontonis 和 Kulin Shah，利用高效得分匹配学习一般高斯混合，arXiv 预印本 arXiv:2404.18893（2024年）。[CL24] Frank Cole 和 Yulong Lu，基于得分的生成模型在学习一类次高斯概率分布时突破维数灾难，arXiv 预印本 arXiv:2402.08082（2024年）。[CL25] Changxiao Cai 和 Gen Li，扩散模型概率流常微分方程的极小极大最优性，arXiv 预印本 arXiv:2503.09583（2025年）。[CLL23] Hongrui Chen、Holden Lee 和 Jianfeng Lu，基于得分的生成式建模的改进分析：最小光滑性假设下的易用界，国际机器学习会议，PMLR，2023年，第4735–4763页。

[DB22] Valentin De Bortoli, 《流形假设下去噪扩散模型的收敛性》, arXiv预印本 arXiv:2208.05314（2022年）。[DKXZ24] Zehao Dou、Subhodh Kotekal、Zhehao Xu 和 Harrison H Zhou, 《从最优得分匹配到最优采样》, arXiv预印本 arXiv:2409.07032（2024年）。[DN21] Prafulla Dhariwal 和 Alexander Nichol, 《扩散模型在图像合成方面击败生成对抗网络》, 神经信息处理系统进展 34（2021年）, 8780–8794。[DSL21] Jacob Deasy、Nikola Simidjievski 和 Pietro Li`o, 《重尾去噪得分匹配》, arXiv预印本 arXiv:2112.09788（2021年）。[FGL25] Jianqing Fan、Yihong Gu 和 Ximing Li, 《使用ReLU神经网络和扩散模型对可因子分解密度进行最优估计》, arXiv预印本 arXiv:2510.03994（2025年）。[GCC24] Shivam Gupta、Linda Cai 和 Sitan Chen, 《使用随机化中点进行更快速的基于扩散的采样：串行与并行》, arXiv预印本 arXiv:2406.00924（2024年）。[GKL24] Khashayar Gatmiry、Jonathan Kelner 和 Holden Lee, 《使用扩散模型学习高斯混合模型》, arXiv预印本 arXiv:2404.18869（2024年）。[GNZ25] Xuefeng Gao、Hoang M Nguyen 和 Lingjiong Zhu, 《一类基于得分的生成模型的Wasserstein收敛保证》, 机器学习研究期刊 26（2025年）, 第43号, 1–54。[GPPX24] Shivam Gupta、Aditya Parulekar、Eric Price 和 Zhiyang Xun, 《改进的扩散模型训练样本复杂度界》, 神经信息处理系统进展 37（2024年）, 40976–41012。[Har13] Andrew C Harvey, 《波动率与重尾的动态模型：及其在金融和经济时间序列中的应用》, 第52卷, 剑桥大学出版社, 2013年。[HHL25a] Daniel Zhengyu Huang、Jiaoyang Huang 和 Zhengjiang Lin, 《基于得分的生成模型中概率流 ODE 的收敛分析》, IEEE信息论汇刊（2025年）。[HHL25b] , 《扩散概率模型中高阶 ODE 求解器的快速收敛》, arXiv预印本 arXiv:2506.13061（2025年）。[HJA20] Jonathan Ho、Ajay Jain 和 Pieter Abbeel, 《去噪扩散概率模型》, 神经信息处理系统进展 33（2020年）, 6840–6851。[HO10] Marlis Hochbruck 和 Alexander Ostermann, 《指数积分器》, Acta Numerica 19（2010年）, 209–286。[HRX24] Yinbin Han、Meisam Razaviyayn 和 Renyuan Xu, 《扩散模型中基于神经网络的得分估计：优化与泛化》, arXiv预印本 arXiv:2401.15604（2024年）。[LJ24] Gen Li 和 Yuchen Jiao, 《扩散概率模型的改进收敛速率》, arXiv预印本 arXiv:2410.13738（2024年）。[LLT23] Holden Lee、Jianfeng Lu 和 Yixin Tan, 《一般数据分布下基于得分的生成建模的收敛性》, 国际算法学习理论会议, PMLR, 2023年, 第946–985页。[NRW21] Eliya Nachmani、Robin San Roman 和 Lior Wolf, 《去噪扩散伽马模型》, arXiv预印本 arXiv:2110.05948（2021年）。[OAS23] Kazusato Oko、Shunta Akiyama 和 Taiji Suzuki, 《扩散模型是极小极大最优的分布估计器》, 国际机器学习会议, PMLR, 2023年, 第26517–26582页。[PPX+24] Kushagra Pandey、Jaideep Pathak、Yilun Xu、Stephan Mandt、Michael Pritchard、Arash Vahdat 和 Morteza Mardani, 《重尾扩散模型》, arXiv预印本 arXiv:2410.14171（2024年）。

[PVG+21] Vadim Popov、Ivan Vovk、Vladimir Gogoryan、Tasnima Sadekova 和 Mikhail Kudinov，《Grad-TTS：一种用于文本转语音的扩散概率模型》，国际机器学习会议，PMLR，2021年，第8599–8608页。[RDN+22] Aditya Ramesh、Prafulla Dhariwal、Alex Nichol、Casey Chu 和 Mark Chen，《使用 CLIP潜变量的分层文本条件图像生成》，arXiv预印本 arXiv:2204.06125 1（2022年），第2号，3。[SAL25] Arthur St´ephanovitch、Eddie Aamari 和 Cl´ement Levrard，《基于得分的生成模型的泛化界：一种合成证明》，arXiv预印本 arXiv:2507.04794（2025年）。[SME20] Jiaming Song、Chenlin Meng 和 Stefano Ermon，《去噪扩散隐式模型》，arXiv预印本 arXiv:2010.02502（2020年）。[SSD25] Dario Shariatian、Umut Simsekli 和 Alain Oliviero Durmus，《基于重尾扩散与去噪 Lévy 的概率模型》，第十三届学习表征国际会议，2025年。[SSDK+20] Yang Song、Jascha Sohl-Dickstein、Diederik P Kingma、Abhishek Kumar、Stefano Ermon 和 Ben Poole，《通过随机微分方程进行基于得分的生成建模》，arXiv预印本 arXiv:2011.13456（2020年）。[Tsy09] Alexandre B. Tsybakov，《非参数估计导论》，第11卷，施普林格，2009年。[TY24] Rong Tang 和 Yun Yang，《扩散模型对流形结构的自适应性》，国际人工智能与统计会议，PMLR，2024年，第1648–1656页。[TZ24] Wenpin Tang 和 Hanyang Zhao，《通过随机微分方程理解基于得分的扩散模型——技术教程》，arXiv预印本 arXiv:2402.07487（2024年）。[WCW24] Yuchen Wu、Yuxin Chen 和 Yuting Wei，《随机 Runge-Kutta 方法：可证明的扩散模型加速》，arXiv预印本 arXiv:2410.04760（2024年）。[XYS+22] Minkai Xu、Lantao Yu、Yang Song、Chence Shi、Stefano Ermon 和 Jian Tang，《GeoDiff：用于分子构象生成的几何扩散模型》，arXiv预印本 arXiv:2203.02923（2022年）。[YMP25] Konstantin Yakovlev、Anna Markovich 和 Nikita Puchkin，《隐式分数匹配遇上去噪分数匹配：改进的收敛速率与对数密度 Hessian 估计》，arXiv预印本 arXiv:2512.24378（2025年）。[YP25] Konstantin Yakovlev 和 Nikita Puchkin，《宽松流形假设下去噪分数匹配的泛化误差界》，arXiv预印本 arXiv:2502.13662（2025年）。[YPKL23] Eun Bi Yoon、Keehun Park、Sungwoong Kim 和 Sungbin Lim，《基于得分的生成模型与 L´évy 过程》，神经信息处理系统进展 36（2023年），40694–40707。[YY25] Yifeng Yu 和 Lu Yu，《推进基于得分模型的 Wasserstein 收敛分析：来自离散化与二阶加速的见解》，arXiv预印本 arXiv:2502.04849（2025年）。[ZC22] Qinsheng Zhang 和 Yongxin Chen，《使用指数积分器快速采样扩散模型》，arXiv预印本 arXiv:2204.13902（2022年）。[ZYLL24] Kaihong Zhang、Caitlyn H. Yin、Feng Liang 和 Jingbo Liu，《基于得分的扩散模型的极小极大最优性：超越密度下界假设》，机器学习研究论文集 235（2024年），60134–60178。

A 附录

补充符号。 我们用 $\mathbb{B}_r(x) := \{y \in \mathbb{R}^d : \|y - x\| \leqslant r\}$ 表示以 x 为中心、半径为 r 的欧几里得球，并简记 $\mathbb{B}_r$ 为 $\mathbb{B}_r(0)$。

A.1 第 2.3 节的证明2.3

证明。 [命题 12] 的证明 根据均方误差的定义以及 $\hat{}p_t$，的构造，我们有

$$\mathrm{MSE}(\hat{p}_t(x)) \leqslant \mathrm{Var}(\hat{p}_t(x)) \leqslant \frac{1}{n}\mathrm{Var}(\varphi_t(X^{(i)} - x)) .$$

设 $Z_0 \sim p_0$ 是 $\{X^{(i)}\}_{i=1}^n$. 的一个独立副本。于是我们有

$$\begin{aligned}
\frac{1}{n}\mathrm{Var}(\varphi_t(X^{(i)} - x)) &\leqslant \frac{1}{n}\mathbb{E}[\varphi_t^2(Z_0 - x)] \\
&= \frac{1}{n}\int_{\mathbb{R}^d} \frac{1}{(2\pi t)^d} \exp\left(-\frac{\|x-y\|_2^2}{t}\right) p_0(y)\,\mathrm{d}y \\
&\leqslant \frac{1}{n(2\pi t)^{d/2}}\int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x-y\|_2^2}{2t}\right) p_0(y)\,\mathrm{d}y \\
&= \frac{p_t(x)}{n(2\pi t)^{d/2}} .
\end{aligned}$$

将此结果与前面的公式结合，即得所需结论。 □

证明。 [命题 13] 的证明：令 $Z_0 \sim p_0$ 为 $\{X^{(i)}\}_{i=1}^n$. 的一个独立副本。由于 $\nabla\hat{}p_t(x)$ 是无偏的，我们有

$$\begin{aligned}
\mathrm{MSE}(\nabla\hat{p}_t(x)) &= \mathrm{Var}(\nabla\hat{p}_t(x)) \\
&= \frac{1}{nt^2}\mathrm{Var}((X^{(i)} - x)\varphi_t(X^{(i)} - x)) \\
&\leqslant \frac{1}{nt^2}\mathbb{E}\big[(Z_0 - x)^2\varphi_t^2(Z_0 - x)\big] \\
&= \frac{1}{nt^2}\int_{\mathbb{R}^d} \frac{(y-x)^2}{(2\pi t)^d} \exp\left(-\frac{\|x-y\|_2^2}{t}\right) p_0(y)\mathrm{d}y \\
&\leqslant \frac{1}{nt(2\pi t)^{d/2}}\int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x-y\|_2^2}{2t}\right) p_0(y)\mathrm{d}y \\
&= \frac{p_t(x)}{nt(2\pi t)^{d/2}}
\end{aligned}$$

如所 desired。 □

A.2 第 3.1 节的证明

定义事件

$$A_\alpha = \left\{ x \in \mathbb{R}^d : |\hat{p}_t(x) - p_t(x)| < C_\alpha\left(\frac{\log n}{n(2\pi t)^{d/2}} + \sqrt{\frac{p_t(x)\log n}{n(2\pi t)^{d/2}}}\right)\right\} .$$

根据引理11，可得 $\mathbb{P}(A_\alpha^c) < n^{-\alpha}$。我们进一步将空间划分为主体区域$\mathcal{G}_1$ 和尾部区域 $\mathcal{G}_2$ ，通过

$$\mathcal{G}_1 := \{x : p_t(x) > c_\alpha \rho_n\}, \qquad \mathcal{G}_2 := \{x : p_t(x) \leqslant c_\alpha \rho_n\},$$

其中，选择 c_α 使其满足 $c_\alpha > 2$ 和 $c_\alpha > 2C_\alpha(1+\sqrt{c_\alpha})$。那么，对于任意 $x \in \mathcal{G}_1 \cap A_\alpha$,

$$\begin{aligned}\hat{p}_t(x) &> p_t(x) - C_\alpha\big(\rho_n + \sqrt{p_t(x)\rho_n}\big) \\ &> p_t(x) - C_\alpha\left(\frac{p_t(x)}{c_\alpha} + \frac{p_t(x)}{\sqrt{c_\alpha}}\right) > \Big(1 - C_\alpha(c_\alpha^{-1} + c_\alpha^{-1/2})\Big)p_t(x) > \frac{1}{2}p_t(x),\end{aligned}$$

因此，$\hat{p}_t(x) > \frac{1}{2}p_t(x) > \frac{c_\alpha}{2}\rho_n > \rho_n$。这表明，在主体区域 $\mathcal{G}_1$ 中，密度估计远离零有界（在事件 A_α 上），而 $\mathcal{G}_2$ 则隔离出了需要进行阈值化以确保稳定性的低密度区域。

A.2.1 定理 1 的证明1

证明。 证明的关键在于将积分区域划分为高密度区域 $\mathcal{G}_1$（其中核估计量能够良好地集中）和低密度尾部区域 $\mathcal{G}_2$（其中估计量会被截断）。我们将 L^2 得分匹配误差分为三个部分

$$\begin{aligned}&\mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\,\mathrm{d}x\right] \\ &= \int_{\mathbb{R}^d} \mathbb{E}\big[\|\hat{s}_t(x) - s_t(x)\|^2\big] p_t(x)\,\mathrm{d}x \\ &= \underbrace{\int_{\mathcal{G}_1} \mathbb{E}\big[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}\{\omega : x \in A_\alpha(\omega)\}\big] p_t(x)\,\mathrm{d}x}_{\mathrm{I}} + \underbrace{\int_{\mathcal{G}_2} \mathbb{E}\big[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}\{\omega : x \in A_\alpha(\omega)\}\big] p_t(x)\,\mathrm{d}x}_{\mathrm{II}} \\ &\quad + \underbrace{\int_{\mathbb{R}^d} \mathbb{E}\big[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}\{\omega : x \in A_\alpha^c(\omega)\}\big] p_t(x)\,\mathrm{d}x}_{\mathrm{III}}.\end{aligned}$$

步骤 1。 注意，当 $x \in \mathcal{G}_1 \cap A_\alpha$ 时，$\hat{p}_t(x) > \max\{\rho_n, \frac{1}{2}p_t(x)\}$，于是

$$\begin{aligned}\mathrm{I} &\leqslant \int_{\mathcal{G}_1} \mathbb{E}[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}\{\hat{p}_t(x) > \max(\rho_n, \frac{1}{2}p_t(x))\}] p_t(x)\,\mathrm{d}x \qquad (A.1)\\ &\leqslant \int_{\mathcal{G}_1} 2\mathbb{E}\left[\frac{\|\nabla\hat{p}_t(x) - \nabla p_t(x)\|^2 + \|s_t(x)\|^2|\hat{p}_t(x) - p_t(x)|^2}{\hat{p}_t(x)^2} \mathbb{1}\{\hat{p}_t(x) > \frac{1}{2}p_t(x)\}\right] p_t(x)\,\mathrm{d}x \\ &\leqslant 8\int_{\mathcal{G}_1} \mathbb{E}\big[\|\nabla\hat{p}_t(x) - \nabla p_t(x)\|^2 + \|s_t(x)\|^2|\hat{p}_t(x) - p_t(x)|^2\big]\frac{1}{p_t(x)}\,\mathrm{d}x \\ &\leqslant 8\int_{\mathcal{G}_1} \big(\mathrm{MSE}(\nabla\hat{p}_t(x)) + \mathrm{MSE}(\hat{p}_t(x)) \cdot \|s_t(x)\|^2\big) \cdot \frac{1}{p_t(x)}\,\mathrm{d}x\end{aligned}$$

根据命题12 和13，我们有

$$\mathrm{I} \leqslant 8|\mathcal{G}_1| \cdot \frac{1}{nt(2\pi t)^{d/2}} + 8\int_{\mathcal{G}_1} \|s_t(x)\|^2\,\mathrm{d}x \cdot \frac{1}{n(2\pi t)^{d/2}} \qquad (A.2)$$

引理 16。 在假设2下，我们有

$$|\mathcal{G}_1| \lesssim (C_\gamma(t)\rho_n)^{-\frac{d}{1+\gamma+d}},$$

and

$$\int_{\mathcal{G}_1} \|s_t(x)\|^2 \mathrm{d}x \lesssim C_\gamma(t)^{-\frac{1}{m}} m t^{-1} \rho_n^{-\frac{1}{m}-\frac{d}{(1+\gamma+d)m'}},$$

其中

$$C_\gamma(t) = \begin{cases} C_1(\gamma,d) + 2^{1+\gamma+d}C_0, & 0 < t \leqslant 1 \\ C_2(\gamma,d,t) + 2^{1+\gamma+d}C_0, & t > 1 \end{cases}.$$

根据 $m = \frac{1}{\varepsilon}$ 引理 16，我们得到

$$\int_{\mathcal{G}_1} \|s_t(x)\|^2 \mathrm{d}x \lesssim C_\gamma(t)^{\varepsilon} \varepsilon^{-1} t^{-1} \rho_n^{-\varepsilon-\frac{d}{1+\gamma+d}}.$$

将其代回不等式（A.2）可得

$$\begin{aligned} \mathrm{I} &\lesssim (C_\gamma(t)\rho_n)^{-\frac{d}{1+\gamma+d}} \cdot \frac{1}{nt(2\pi t)^{d/2}} + (C_\gamma(t)^{\varepsilon}\varepsilon^{-1}t^{-1}\rho_n^{-\varepsilon-\frac{d}{1+\gamma+d}}) \cdot \frac{1}{n(2\pi t)^{d/2}} \\ &= C_\gamma(t)^{-\frac{d}{1+\gamma+d}} (t\log n)^{-1} \rho_n^{\frac{1+\gamma}{1+\gamma+d}} + (C_\gamma(t)^{\varepsilon}\varepsilon^{-1}\rho_n^{-\varepsilon})(t\log n)^{-1}\rho_n^{\frac{1+\gamma}{1+\gamma+d}} \\ &= \left(C_\gamma(t)^{-\frac{d}{1+\gamma+d}} + C_\gamma(t)^{\varepsilon}\varepsilon^{-1}\rho_n^{-\varepsilon}\right)(t\log n)^{-1}\rho_n^{\frac{1+\gamma}{1+\gamma+d}}. \end{aligned}$$

步骤 2。 回顾 $\mathcal{G}_2$ 和 $\hat{}s_t(x)$的定义，我们有

$$\mathrm{II} \leqslant \int_{\mathcal{G}_2} \mathbb{E}\big[\|s_t(x)\|^2 \mathbb{1}\{\hat{p}_t(x) < \rho_n\}\big] p_t(x)\,\mathrm{d}x + \int_{\mathcal{G}_2} \mathbb{E}\big[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}\{\hat{p}_t(x) \geqslant \rho_n\}\big] p_t(x)\,\mathrm{d}x.$$

引理 17。 在假设 2下，有

$$\int_{\mathcal{G}_2} p_t(x)\,\mathrm{d}x \lesssim C_\gamma(t)^{\frac{d}{d+\gamma+1}} \rho_n^{\frac{\gamma+1}{d+\gamma+1}},$$

and

$$\int_{\mathcal{G}_2} \|s_t(x)\|^2 p_t(x)\,\mathrm{d}x \lesssim \varepsilon^{-1}t^{-1}C_\gamma(t)^{\frac{d(1-\varepsilon)}{d+\gamma+1}} \rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{d+\gamma+1}}.$$

应用引理 17，II 的第一项可界定为

$$\int_{\mathcal{G}_2} \mathbb{E}\big[\|s_t(x)\|^2 \mathbb{1}\{\hat{p}_t(x) < \rho_n\}\big] p_t(x)\,\mathrm{d}x \leqslant \int_{\mathcal{G}_2} \|s_t(x)\|^2 p_t(x)\,\mathrm{d}x \lesssim \varepsilon^{-1}t^{-1}C_\gamma(t)^{\frac{d(1-\varepsilon)}{d+\gamma+1}} \rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{(d+\gamma+1)}}.$$

对于 II 的第二项，根据命题 12 和命题 13，并采用与式（A.1）中类似的方法，可得

$$\int_{\mathcal{G}_2} \mathbb{E}\big[\|\hat{s}_t(x) - s_t(x)\|^2 \mathbb{1}\{\hat{p}_t(x) \geqslant \rho_n\}\big] p_t(x)\,\mathrm{d}x$$

$$
\begin{aligned}
&\leqslant \int_{\mathcal{G}_2} 2\mathbb{E}\left[\frac{\|\nabla\hat{p}_t(x)-\nabla p_t(x)\|^2+\|s_t(x)\|^2|\hat{p}_t(x)-p_t(x)|^2}{\hat{p}_t(x)^2}\mathbb{1}\{\hat{p}_t(x)\geqslant\rho_n\}\right]p_t(x)\,\mathrm{d}x\\
&\leqslant 2\rho_n^{-2}\int_{\mathcal{G}_2}\big(\mathrm{MSE}(\nabla\hat{p}_t(x))+\mathrm{MSE}(\hat{p}_t(x))\cdot\|s_t(x)\|^2\big)p_t(x)\,\mathrm{d}x\\
&\leqslant 2\rho_n^{-2}\left(\int_{\mathcal{G}_2}\frac{p_t(x)}{nt(2\pi t)^{d/2}}p_t(x)\,\mathrm{d}x+\int_{\mathcal{G}_2}\frac{p_t(x)}{n(2\pi t)^{d/2}}\cdot\|s_t(x)\|^2p_t(x)\,\mathrm{d}x\right)\\
&\leqslant 2\rho_n^{-1}\left(\frac{c_\alpha}{nt(2\pi t)^{d/2}}\int_{\mathcal{G}_2}p_t(x)\,\mathrm{d}x+\frac{c_\alpha}{n(2\pi t)^{d/2}}\int_{\mathcal{G}_2}\|s_t(x)\|^2p_t(x)\,\mathrm{d}x\right).
\end{aligned}
$$

再次应用引理 17，即可推出

$$
\begin{aligned}
&\int_{\mathcal{G}_2}\mathbb{E}\big[\|\hat{s}_t(x)-s_t(x)\|^2\mathbb{1}\{\hat{p}_t(x)\geqslant\rho_n\}\big]p_t(x)\,\mathrm{d}x\\
&\lesssim 2\rho_n^{-1}\left(\frac{c_\alpha}{nt(2\pi t)^{d/2}}C_\gamma(t)^{\frac{d}{d+\gamma+1}}\rho_n^{\frac{\gamma+1}{d+\gamma+1}}+\frac{c_\alpha}{n(2\pi t)^{d/2}}\varepsilon^{-1}t^{-1}C_\gamma(t)^{\frac{d(1-\varepsilon)}{d+\gamma+1}}\rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{d+\gamma+1}}\right)\\
&=2c_\alpha(t\log n)^{-1}\left(C_\gamma(t)^{\frac{d}{d+\gamma+1}}\rho_n^{\frac{\gamma+1}{d+\gamma+1}}+\varepsilon^{-1}C_\gamma(t)^{\frac{d(1-\varepsilon)}{d+\gamma+1}}\rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{d+\gamma+1}}\right)
\end{aligned}
$$

步骤 3。 根据霍尔德不等式，我们得到

$$
\begin{aligned}
\mathrm{III}&\leqslant\left(\int_{\mathbb{R}^d}\mathbb{E}\big[\|\hat{s}_t(x)-s_t(x)\|^4\big]p_t(x)\,\mathrm{d}x\right)^{1/2}\left(\int_{\mathbb{R}^d}\mathbb{E}[\mathbb{1}\{x\in A_\alpha^c\}]p_t(x)\,\mathrm{d}x\right)^{1/2}\\
&\lesssim n^{-\alpha/2}\left(\int_{\mathbb{R}^d}\|s_t(x)\|^4p_t(x)\,\mathrm{d}x+\int_{\mathbb{R}^d}\mathbb{E}\big[\|\hat{s}_t(x)\|^4\big]p_t(x)\,\mathrm{d}x\right)^{1/2}.
\end{aligned}
$$

引理18（引理 F.16，见 [ZYLL24]）。设 X 和 G 是相互独立的 d维随机向量，满足 $X\sim p_0$，$G\sim\mathcal{N}(0,I_d)$。并且 $Y=X+\sqrt{t}G$。则，

$$
\mathbb{E}[\|\nabla\log p_Y(Y)\|^m]\leqslant t^{-\frac{m}{2}}d^{\frac{m}{2}}(m-1)!!\,.
$$

根据引理18，第一项可由以下式子界定

$$
\int_{\mathbb{R}^d}\|s_t(x)\|^4p_t(x)\,\mathrm{d}x\leqslant 3t^{-2}d^2\,,
$$

第二项可由以下式子界定

$$
\begin{aligned}
\int_{\mathbb{R}^d}\mathbb{E}\|\hat{s}_t(x)\|^4p_t(x)\,\mathrm{d}x&=\int_{\mathbb{R}^d}\mathbb{E}\big[\|\hat{s}_t(x)\|^4\mathbb{1}\{\hat{p}_t(x)>\rho_n\}\big]p_t(x)\,\mathrm{d}x\\
&\leqslant\rho_n^{-4}\int_{\mathbb{R}^d}\mathbb{E}\big[\|\nabla\hat{p}_t(x)\|^4\big]p_t(x)\,\mathrm{d}x\,.
\end{aligned}
$$

回顾 $\nabla p_t(x)$的定义，可知

$$
\begin{aligned}
\mathbb{E}\big[\|\nabla\hat{p}_t(x)\|^4\big]&=\frac{1}{(nt)^4}\mathbb{E}_{X^{(i)}\sim p_0}\left[\left\|\sum_{i=1}^n(X^{(i)}-x)\varphi_t(X^{(i)}-x)\right\|^4\right]\\
&\leqslant\frac{1}{t^4}\mathbb{E}_{X\sim p_0}\big[\|(X-x)\varphi_t(X-x)\|^4\big]\\
&\leqslant\frac{1}{t^4(2\pi)^{d/2}}\sup_{r>0}r^2\exp(-2r/t)
\end{aligned}
$$

$$\leqslant \frac{1}{t^2 e^2 (2\pi)^{d/2}}\,.$$

因此，

$$\mathrm{III} \lesssim n^{-\alpha/2}\Big(3t^{-2}d^2 + \rho_n^{-4}[t^2e^2(2\pi)^{d/2}]^{-1}\Big)^{1/2}\,.$$

我们可以选择足够大的 α，以覆盖 t 和 n 的量级。

结合步骤 1–3，可得

$$\begin{aligned}\mathbb{E}\bigg[\int_{\mathbb{R}^d}\|\hat{s}_t(x)-s_t(x)\|^2 p_t(x)\,\mathrm{d}x\bigg] \lesssim{}& \Big(C_\gamma(t)^{-\frac{d}{1+\gamma+d}} + C_\gamma(t)^{\varepsilon}\varepsilon^{-1}\rho_n^{-\varepsilon}\Big)(t\log n)^{-1}\rho_n^{\frac{1+\gamma}{1+\gamma+d}}\\ &+ 2c_\alpha(t\log n)^{-1}\bigg(C_\gamma(t)^{\frac{d}{d+\gamma+1}}\rho_n^{\frac{\gamma+1}{d+\gamma+1}} + \varepsilon^{-1}C_\gamma(t)^{\frac{d(1-\varepsilon)}{d+\gamma+1}}\rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{d+\gamma+1}}\bigg)\\ &+ \varepsilon^{-1}t^{-1}C_\gamma(t)^{\frac{d(1-\varepsilon)}{d+\gamma+1}}\rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{(d+\gamma+1)}}\end{aligned}$$

如所需。 □

A.2.2 定理证明3

证明。 首先，引入截断半径 $R>0$，可将 p_0 与 p_{t_0} 之间的全变差距离分解为中心部分和尾部部分。由詹森不等式可得

$$\begin{aligned}\mathrm{TV}(p_0,p_{t_0}) &= \frac{1}{2}\int_{\mathbb{R}^d}|p_0(x)-p_{t_0}(x)|\,\mathrm{d}x\\ &= \frac{1}{2}\int_{\|x\|\leqslant R}|p_0(x)-p_{t_0}(x)|\,\mathrm{d}x + \frac{1}{2}\int_{\|x\|>R}|p_0(x)-p_{t_0}(x)|\,\mathrm{d}x\\ &\leqslant \frac{1}{2}\sqrt{\mathrm{Vol}(\mathbb{B}_R)}\sqrt{\int_{\mathbb{R}^d}|p_0(x)-p_{t_0}(x)|^2\,\mathrm{d}x} + \frac{1}{2}\int_{\|x\|>R}(p_0(x)+p_{t_0}(x))\,\mathrm{d}x\\ &\lesssim R^{d/2}\|p_0-p_{t_0}\|_{L^2} + \int_{\|x\|>R}p_0(x)\,\mathrm{d}x + \int_{\|x\|>R}p_{t_0}(x)\,\mathrm{d}x\,.\end{aligned} \tag{A.3}$$

对于第二项，我们有

$$\int_{\|x\|>R}p_0(x)\,\mathrm{d}x \leqslant C_0\int_R^\infty r^{-(d+\gamma+1)}r^{d-1}\,\mathrm{d}r = C_0\int_R^\infty r^{-\gamma-2}\,\mathrm{d}r \lesssim C_0R^{-(\gamma+1)}\,.$$

对于第二个尾部项，回忆一下 $X_{t_0}=X_0+\sqrt{t_0}Z$，其中 $X_0\sim p_0$ 和 $Z\sim\mathcal{N}(0,\ I_d)$。当 R 足够大（相对于 $\sqrt{t_0}$）时，X_0 的重尾会占主导地位，超过高斯噪声。具体而言，

$$\begin{aligned}\mathbb{P}(\|X_{t_0}\|>R) &\leqslant \mathbb{P}(\|X_0\|>R/2) + \mathbb{P}(\|\sqrt{t_0}Z\|>R/2)\\ &\leqslant C_0(R/2)^{-(\gamma+1)} + \exp\bigg(-\frac{R^2}{8t_0}\bigg)\,.\end{aligned}$$

由于当 R 较大时，多项式衰减占主导，因此我们有：

$$\text{Tail Error} \lesssim R^{-(\gamma+1)}\,. \tag{A.4}$$

接下来这一步由普朗歇尔定理得出，

$$\begin{aligned}\|p_0 - p_{t_0}\|_{L^2}^2 &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |\mathcal{F}[p_0](\omega) - \mathcal{F}[p_{t_0}](\omega)|^2 \,\mathrm{d}\omega \\ &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |\mathcal{F}[p_0](\omega)|^2 |1 - \phi_{t_0}(\omega)|^2 \,\mathrm{d}\omega\,,\end{aligned}$$

其中 $\phi_{t_0}(\omega) = \exp(-\frac{t_0\|\omega\|^2}{2})$。我们在阈值$K = t_0^{-1/2}$ 处将积分分为低频和高频部分。

注意到 $|1 - \phi_{t_0}(\omega)| \leqslant 2$，因此可得

$$\begin{aligned}\int_{\|\omega\|\geqslant t_0^{-1/2}} |\mathcal{F}[p_0]|^2 |1 - \phi_{t_0}|^2 \,\mathrm{d}\omega &\leqslant 4\int_{\|\omega\|\geqslant t_0^{-1/2}} |\mathcal{F}[p_0]|^2 \frac{\|\omega\|^{2\beta}}{\|\omega\|^{2\beta}} \,\mathrm{d}\omega \\ &\leqslant 4(t_0^{-1/2})^{-2\beta}\int_{\mathbb{R}^d} \|\omega\|^{2\beta} |\mathcal{F}[p_0]|^2 \,\mathrm{d}\omega \\ &\lesssim t_0^{\beta} L^2\,.\end{aligned}$$

由于 $|1 - e^{-x}|^2 \leqslant x^2$ 对于 $x \geqslant 0$，因此有$|1 - \phi_{t_0}(\omega)|^2 \leqslant (\frac{t_0\|\omega\|^2}{2})^2 = \frac{t_0^2\|\omega\|^4}{4}$。于是，

$$\begin{aligned}\int_{\|\omega\|< t_0^{-1/2}} |\mathcal{F}[p_0]|^2 |1 - \phi_{t_0}|^2 \,\mathrm{d}\omega &\lesssim t_0^2 \int_{\|\omega\|< t_0^{-1/2}} |\mathcal{F}[p_0]|^2 \|\omega\|^4 \,\mathrm{d}\omega \\ &= t_0^2 \int_{\|\omega\|< t_0^{-1/2}} |\mathcal{F}[p_0]|^2 \|\omega\|^{2\beta} \|\omega\|^{4-2\beta} \,\mathrm{d}\omega\,.\end{aligned}$$

当 $\beta \leqslant 2$时，有 $4 - 2\beta \geqslant 0$。因此， $\|\omega\|^{4-2\beta} \leqslant (t_0^{-1/2})^{4-2\beta} = t_0^{\beta-2}$ ，一个 d

$$\int_{\|\omega\|< t_0^{-1/2}} |\mathcal{F}[p_0]|^2 |1 - \phi_{t_0}|^2 \,\mathrm{d}\omega \lesssim t_0^2 \cdot t_0^{\beta-2} \int_{\mathbb{R}^d} \|\omega\|^{2\beta} |\mathcal{F}[p_0]|^2 \lesssim t_0^{\beta} L^2\,.$$

结合高频和低频部分，我们得到

$$\|p_0 - p_{t_0}\|_{L^2} \lesssim \sqrt{t_0^{\beta}} = t_0^{\beta/2}\,. \tag{A.5}$$

步骤 3：最优半径选择。 将（A.4）和（A.5）代回显示式（A.3）可得

$$\mathrm{TV}(p_0, p_{t_0}) \lesssim R^{d/2} t_0^{\beta/2} + R^{-(\gamma+1)}\,.$$

为优化这个界，我们通过选择 R 使两项贡献达到平衡，从而满足

$$R^{\frac{d}{2}+\gamma+1} \asymp t_0^{-\beta/2}\,,$$

由此得到如下选择：

$$R \asymp t_0^{-\frac{\beta}{d+2(\gamma+1)}}\,.$$

将这个最优的 R 代回前一个公式，得到

$$\mathrm{TV}(p_0, p_{t_0}) \lesssim \left(t_0^{-\frac{\beta}{d+2(\gamma+1)}}\right)^{-(\gamma+1)} = t_0^{\frac{\beta(\gamma+1)}{d+2(\gamma+1)}}$$

如所需。 □

A.2.3 定理4的证明

证明。 回想一下，反向过程 $X^{\leftarrow}$ 由 Y_t 近似。令 $(Y_t)_{t\in[0,T]}$ 为将 $Y_0 \sim \mathcal{N}(0, TI_d)$ 替换为 $\tilde{Y}_0 \sim p_T$ 后得到的 $(Y_t)_{t\in[0,T]}$，也就是说，$(\tilde{Y}_t)_{t\in[0,T]}$ 满足：

$$\mathrm{d}\tilde{Y}_t = \hat{s}_{T-t}(\tilde{Y}_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad \tilde{Y}_t \sim p_T\,.$$

根据三角不等式，有

$$\begin{aligned}&\mathbb{E}\Big[\mathrm{TV}(\mathcal{L}(X_0), \mathcal{L}(\hat{Y}_{T-t_0}))\Big]\\ &\leqslant \mathrm{TV}(\mathcal{L}(X_0), \mathcal{L}(X^{\leftarrow}_{T-t_0})) + \mathbb{E}[\mathrm{TV}(\mathcal{L}(X^{\leftarrow}_{T-t_0}), \mathcal{L}(\tilde{Y}_{T-t_0}))] + \mathbb{E}[\mathrm{TV}(\mathcal{L}(\tilde{Y}_{T-t_0}), \mathcal{L}(\hat{Y}_{T-t_0}))]\,.\end{aligned}$$

根据定理 3，$\mathrm{TV}(\mathcal{L}(X_0)\ \mathcal{L}(X^{\leftarrow}_{T-t_0})) \lesssim n^{-\frac{2}{\beta(\gamma+)}\frac{1}{4\beta(d+\gamma+1)+d(d+2(\gamma+1))}}$。对于第二项，根据 Pinsker 不等式和数据处理不等式，我们得到

$$\mathrm{TV}(\mathcal{L}(X^{\leftarrow}_{T-t_0}), \mathcal{L}(\tilde{Y}_{T-t_0})) \lesssim \sqrt{\mathrm{KL}(\mathcal{L}(X^{\leftarrow}_{T-t_0}) \| \mathcal{L}(\tilde{Y}_{T-t_0}))} \leqslant \sqrt{\mathrm{KL}(\mathbb{P}_{X^{\leftarrow}} \| \mathbb{P}_{\tilde{Y}})}$$

其中，$\mathbb{P}_{X^{\leftarrow}}$ 和 $\mathbb{P}_{\tilde{Y}}$ 是 $(X^{\leftarrow}_t)_{t\in[0,T-t_0]}$ 和 $(Y_t)_{t\in[0,T-t_0]}$ 的路径测度。于是，我们有

$$\begin{aligned}\mathrm{KL}(\mathbb{P}_{X^{\leftarrow}} \| \mathbb{P}_{\tilde{Y}}) &\leqslant \frac{1}{2}\int_0^{T-t_0}\int_{\mathbb{R}^d} \|\hat{s}_{T-t}(x) - s_{T-t}(x)\|^2\,\mathrm{d}x\,\mathrm{d}t\\ &= \frac{1}{2}\int_{t_0}^{T}\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2\,\mathrm{d}x\,\mathrm{d}t\end{aligned}$$

于是 $\mathbb{E}[\mathrm{KL}(\mathbb{P}_{X^{\leftarrow}} \| \mathbb{P}_{\tilde{Y}})]$ 可由定理 2给出上界，并根据 Jensen 不等式，我们有

$$\mathbb{E}[\mathrm{TV}(\mathcal{L}(X^{\leftarrow}_{T-t_0}), \mathcal{L}(\tilde{Y}_{T-t_0}))] \lesssim \mathbb{E}\left[\sqrt{\mathrm{KL}(\mathbb{P}_{X^{\leftarrow}} \| \mathbb{P}_{\tilde{Y}})}\right] \leqslant \sqrt{\mathbb{E}[\mathrm{KL}(\mathbb{P}_{X^{\leftarrow}} \| \mathbb{P}_{\tilde{Y}})]}\,.$$

由此可得

$$\mathbb{E}[\mathrm{TV}(\mathcal{L}(X^{\leftarrow}_{T-t_0}), \mathcal{L}(\tilde{Y}_{T-t_0}))] \lesssim \mathrm{polylog}(n) n^{-\frac{\gamma+1}{2(d+\gamma+1)}} t_0^{-\frac{d(\gamma+1)}{4(d+\gamma+1)}} \asymp \mathrm{polylog}(n) n^{-\frac{2\beta(\gamma+1)}{4\beta(d+\gamma+1)+d(d+2(\gamma+1))}}\,.$$

由于 Y_{T-t_0} 与 Y_{T-t_0} 唯一的区别在于它们的初始分布，我们有

$$\mathbb{E}[\mathrm{TV}(\mathcal{L}(\tilde{Y}_{T-t_0}), \mathcal{L}(\hat{Y}_{T-t_0}))] \leqslant \mathrm{TV}(\mathcal{L}(X_T), \mathcal{N}(0, TI_d)) \lesssim \frac{1}{\sqrt{T}}\,.$$

$4\beta(\gamma+1)$ 最后一个不等式由引理 19 得出。取 $T = n^{4\beta(d+\gamma+1)+d(d+2(\gamma+1))}$，可得

$$\mathbb{E}[\mathrm{TV}(\mathcal{L}(\tilde{Y}_{T-t_0}), \mathcal{L}(\hat{Y}_{T-t_0}))] \lesssim n^{-\frac{2\beta(\gamma+1)}{4\beta(d+\gamma+1)+d(d+2(\gamma+1))}}\,.$$

证明至此完成。 □

引理 19。 设 p_0 是定义在 $\mathbb{R}^d$ 上的概率密度，且具有有限的一阶矩 $M_1 := \int_{\mathbb{R}^d} \|z\|\, p_0(z)\mathrm{d}z < \infty$. 令 φ_T 表示 $\mathcal{N}(0, TI_d)$ 的密度，并定义 $p_T := p_0 * \varphi_T$。则有

$$\mathrm{TV}(p_T, \varphi_T) \leqslant \frac{M_1}{2\sqrt{T}}.$$

A.3 第3.2节的证明

命题 12、13 和引理11均与尾部衰减的行为无关，因此无需修改即可继续成立。对于得分估计的全局收敛定理，我们逐步重新审视证明过程：只有引理 16、 17和 23 需要修改，之后即可立即得到所需结果。

引理 20 （引理23的修改版）。 在假设 *1*下，有

$$p_t(x) \leqslant \mathcal{C}_\gamma(t) \exp\left(-c_\gamma(t) \min(\|x\|^2, \|x\|^\gamma)\right).$$

其中 $\mathcal{C}_\gamma(t) := \frac{1}{(2\pi t)^{d/2}} + C_1$ 且 $c_\gamma(t) := \text{最小值}\left(\frac{1}{8t}, \frac{c_1}{2^\gamma}\right)$。

A.3.1 定理5的证明

证明。 证明过程遵循与定理 1 的证明相同的基于阈值的分解策略。我们将加权积分均方误差分解为来自高密度主体区域 $\mathcal{G}_1$和低密度尾部区域 $\mathcal{G}_2$的贡献。

回顾式（A.1）和（A.2）中的分解。在主体区域 $\mathcal{G}_1$上，误差由方差项主导，其阶为 $|\mathcal{G}_1| \cdot \frac{1}{n(2\pi t)^{d/2}}$。在假设 1 （指数衰减）下，下面这个对引理16~~的改~~编引理 $\mathcal{G}_1$ 表明，其大小仅按对数增长。

引理 21 (引理的修改 16)。 在假设 *1* 下，对于阈值 $\rho_n = \log n \frac{1}{n(2\pi t)^{d/2}}$，高密度区域 $\mathcal{G}_1 = \{x : p_t(x) \geqslant c_\alpha \rho_n\}$ 的体积满足

$$|\mathcal{G}_1| \lesssim (\log(1/\rho_n))^{d/\gamma}$$

and

$$\int_{\mathcal{G}_1} \|s_t(x)\|^2 \mathrm{d}x \lesssim \mathrm{polylog}(nt) \cdot \varepsilon t^{-1} \rho_n^{-\varepsilon}.$$

因此，主体区域上的积分误差满足

$$\mathrm{I} \lesssim \mathrm{polylog}(n) \frac{1}{nt^{1+d/2}}.$$

对于尾部区域 $\mathcal{G}_2$，误差由尾部的概率质量控制。

引理 22 (引理的修改 17)。 在假设 *1* 下，对于阈值 $\rho_n = \log n \frac{1}{n(2\pi t)^{d/2}}$，尾部区域 $\mathcal{G}_2 = \{x : p_t(x) < c_\alpha \rho_n\}$ 的概率质量满足

$$\int_{\mathcal{G}_2} p_t(x) \, \mathrm{d}x \lesssim \rho_n (\log(1/\rho_n))^{d/k}$$

and

$$\int_{\mathcal{G}_2} \|s_t(x)\|^2 p_t(x) \, \mathrm{d}x \lesssim \varepsilon^{-1} t^{-1} \rho_n^{1-\varepsilon}.$$

由于指数衰减，这一质量相对于阈值 ρ_n 超多项式速率衰减，使得偏差和截断误差相较于主体区域中的方差项可以忽略不计。

综合这些部分，总均方误差由主体区域的有效体积缩放后的参数方差速率主导

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x) \mathrm{d}x\right] \lesssim \mathrm{polylog}(n) n^{-1} t^{-(1+d/2)}.$$

□

A.3.2 定理6的证明

证明。 我们对扩散时间区间 $[t_0, T]$内定理5中得出的逐点均方误差界进行积分

$$\begin{aligned}\int_{t_0}^{T} \mathbb{E}\Big[\|\hat{s}_t - s_t\|^2_{L^2(p_t)}\Big]\mathrm{d}t &\lesssim \int_{t_0}^{T} \mathrm{polylog}(n) n^{-1} t^{-(1+d/2)}\mathrm{d}t \\ &= \mathrm{polylog}(n) n^{-1}\left[\frac{t^{-d/2}}{-d/2}\right]\Bigg|_{t_0}^{T} \\ &\lesssim \mathrm{polylog}(n) n^{-1} t_0^{-d/2}.\end{aligned}$$

这就得到了所需结果。 □

A.3.3 定理7的证明

证明。 证明过程与定理 3 的证明类似，使用截断半径 R 来分解全变差。

$$\mathrm{TV}(p_0, p_{t_0}) \leqslant \frac{1}{2} R^{d/2} \|p_0 - p_{t_0}\|_{L^2} + \frac{1}{2}\int_{\|x\|>R} (p_0(x) + p_{t_0}(x))\mathrm{d}x.$$

对于中心区域中的平滑误差，索博列夫假设给出了与多项式情形相同的界（通过普朗歇尔定理推导）

$$\|p_0 - p_{t_0}\|_{L^2} \lesssim t_0^{\beta/2}.$$

对于尾部区域，在假设 1 下，质量呈指数衰减。具体而言，当 R 足够大时，

$$\int_{\|x\|>R} p_0(x)\mathrm{d}x \leqslant C\exp(-cR^{\gamma}),$$

并且卷积p_{t_0} 满足类似的尾部界。我们选择半径 R ，使尾部误差与平滑误差相当（相差至多对数因子），即 $\exp(-cR^{\gamma}) \approx t_0^{\beta/2}$ 。这意味着：

$$R \asymp (\log(1/t_0))^{1/\gamma}.$$

将该半径代回分解中，体积因子会贡献一个多对数项 $R^{d/2} \asymp (\log(1/t_0))^{d/2\gamma}$。因此，总误差由平滑项主导

$$\mathrm{TV}(p_0, p_{t_0}) \lesssim t_0^{\beta/2}(\log(1/t_0))^{\frac{d}{2\gamma}}.$$

□

A.3.4 定理 8 的证明8

证明。 证明过程沿用定理 4 中完全相同的论证，将全变差误差分解为三个部分——提前停止偏差、分数匹配误差和初始化误差。

$$\mathbb{E}[\mathrm{TV}(p_0, \mathcal{L}(\hat{Y}_{T-t_0}))] \leq \underbrace{\mathrm{TV}(p_0, p_{t_0})}_{\mathrm{I}} + \underbrace{\mathbb{E}[\mathrm{TV}(p_{t_0}, \mathcal{L}(\tilde{Y}_{T-t_0}))]}_{\mathrm{II}} + \underbrace{\mathrm{TV}(p_T, \mathcal{N}(0, TI_d))}_{\mathrm{III}}.$$

对于提前停止偏差 I，定理 7 证明，在指数尾条件下，误差由索博列夫平滑项 $t_0^{\beta/2}$ 主导。对于分数匹配误差 II，应用 Pinsker 不等式以及定理 6 中的累积均方误差界，可得收敛速率为多对数(n) $n^{-1/2}t_0^{-d/4}$。对于初始化误差 III，引理 19 保证其以 $T^{-1/2}$的速率衰减，当 T足够大时，该误差趋于零。平衡偏差 $t_0^{\beta/2}$ 与得分误差 $n^{-1/2}t_0^{-d/4}$，即可得到最优停止时间 $t_0 \asymp n^{-\frac{2}{2\beta+d}}$以及文中所述的极小极大收敛速率。 □

A.4 上界辅助引理的证明

A.4.1 多项式衰减引理的证明

证明。 [引理的证明 11] 注意，$\hat{}p_t$ 是一个无偏估计量，命题 12 给出了 MSE($\hat{}p_t$) 的上界，

$$\mathrm{MSE}(\hat{p}_t(x)) \leqslant \frac{p_t(x)}{n(2\pi t)^{d/2}}.$$

此外，还有

$$\max_{1\leqslant i\leqslant n}\frac{1}{n}\Big(\varphi_t(X^{(i)}-x)-\mathbb{E}[\varphi_t(X^{(i)}-x)]\Big) \leqslant \frac{2}{n}\|\varphi_t(x)\|_\infty = \frac{2}{n(2\pi t)^{d/2}}.$$

根据伯恩斯坦不等式，有

$$\mathbb{P}(|\hat{p}_t(x)-p_t(x)| \geqslant \varepsilon) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2p_t(x)/(n(2\pi t)^{d/2})+4\varepsilon/(3n(2\pi t)^{d/2})}\right).$$

于是，以至少 $1-\delta$ 的概率，有

$$|\hat{p}_t(x)-p_t(x)| < \max\left\{\sqrt{\frac{4p_t(x)\log(2/\delta)}{n(2\pi t)^{d/2}}}, \frac{8\log(2/\delta)}{3n(2\pi t)^{d/2}}\right\}.$$

令 $\delta = n^{-\alpha}$，我们得到

$$\mathbb{P}\left(|\hat{p}_t(x)-p_t(x)| < \max\left\{\sqrt{\frac{4p_t(x)\log(2n)\alpha}{n(2\pi t)^{d/2}}}, \frac{8\log(2n)\alpha}{3n(2\pi t)^{d/2}}\right\}\right) \geqslant 1-n^{-\alpha}$$

如所需。 □

为证明引理 16 和 17，我们需要以下引理

引理 23。 在假设 *2*下，当 $\|x\|>1$时，有

$$p_t(x) \leqslant C_\gamma(t)(1+\|x\|^2)^{-(1+d+\gamma)/2}$$

其中

$$C_\gamma(t) = \begin{cases} C_1(\gamma,d)+2^{1+\gamma+d}C_0, & 0<t\leqslant 1 \\ C_2(\gamma,d,t)+2^{1+\gamma+d}C_0, & t>1 \end{cases}.$$

证明。 [引理的证明 23] 设 $R = \|x\|/2$。我们将密度分解为 p_t

$$p_t(x) = p_0 * \varphi_t(x) = \underbrace{\int_{\|y\|\leqslant R} p_0(y)\varphi_t(x-y)\,\mathrm{d}y}_{\mathrm{I}} + \underbrace{\int_{\|y\|>R} p_0(y)\varphi_t(x-y)\,\mathrm{d}y}_{\mathrm{II}}$$

右侧第一项可按如下方式进行界定

$$\begin{aligned}
\mathrm{I} &\leqslant \int_{\|y\|\leqslant R} p_0(y) \sup_{\|y\|\leqslant R} \varphi_t(x-y)\,\mathrm{d}y \\
&= \int_{\|y\|\leqslant R} p_0(y) \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x\|^2}{8t}\right)\mathrm{d}y \\
&\leqslant \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x\|^2}{8t}\right).
\end{aligned}$$

记

$$F(t,r) := \frac{1}{(2\pi t)^{d/2}} \exp(-r/8t)(1+r)^{(1+\gamma+d)/2}.$$

固定 $r \geqslant 1$ 时，$F(t,r)$ 的最大点为 $t(r) = r/(4d)$，因此可得

$$\begin{aligned}
\sup_{0<t\leqslant 1, r\geqslant 1} F(t,r) &= \max\left\{ \sup_{r\leqslant 4d} \left(\frac{2d}{\pi}\right)^{d/2} e^{-d/2}(1+r)^{\frac{1+\gamma+d}{2}} r^{-d/2}, \sup_{r>4d} (2\pi)^{-d/2} e^{-r/8}(1+r)^{\frac{1+\gamma+d}{2}} \right\} \\
&\leqslant \max\left\{ \left(\frac{2d^2}{\pi e(1+\gamma)}\right)^{d/2} \left(\frac{1+\gamma+d}{d}\right)^{\frac{1+\gamma+d}{2}}, (2\pi)^{-d/2} e^{-\frac{3+4\gamma+4d}{8}} [4(1+\gamma+d)]^{\frac{1+\gamma+d}{2}} \right\} \\
&:= C_1(\gamma, d)
\end{aligned}$$

其中，$C_1(\gamma,d)$ 与 t 无关。固定 $t > 1$ 时，$F(t,r)$ 的最大点为 $r(t) = 4t(1+\gamma+d) - 1$，于是

$$\begin{aligned}
\sup_{r\geqslant 1} F(t,r) &= \frac{t^{(1+\gamma)/2}}{(2\pi)^{d/2}} \exp\left(-\frac{4t(1+\gamma+d)-1}{8t}\right) [4(1+\gamma+d)]^{(1+\gamma+d)/2} \\
&\leqslant \frac{t^{(1+\gamma)/2}}{(2\pi)^{d/2}} e^{-\frac{1+\gamma+d}{4}} [4(1+\gamma+d)]^{\frac{1+\gamma+d}{2}} \\
&:= C_2(\gamma, d, t)
\end{aligned}$$

因此，当 $0 < t \leqslant 1$ 时，$\mathrm{I} \leqslant C_1(\gamma,d)(1+\|x\|^2)^{-(1+\gamma+d)/2}$；当 $t > 1$ 时，$\mathrm{I} \leqslant C_2(\gamma,d,t)(1+\|x\|^2)^{-(1+\gamma+d)/2}$。第二项

$$|\mathrm{II}| \leqslant \int_{\|y\|>R} \frac{C_0}{(1+\|y\|^2)^{(1+\gamma+d)/2}} \cdot \varphi_t(x-y)\,\mathrm{d}y \leqslant \frac{2^{1+\gamma+d} C_0}{(1+\|x\|^2)^{(1+\gamma+d)/2}}.$$

将这两项合并，我们得到

$$p_t(x) \leqslant \frac{C_\gamma(t)}{(1+\|x\|^2)^{(1+\gamma+d)/2}},$$

其中

$$C_\gamma(t)=\begin{cases}C_1(\gamma,d)+2^{1+\gamma+d}C_0, & 0<t\leqslant 1\\ C_2(\gamma,d,t)+2^{1+\gamma+d}C_0, & t>1\end{cases}.$$

□

证明。 [引理的证明 16] 由引理 23，我们有

$$p_t(x)\leqslant\frac{C_\gamma(t)}{(1+\|x\|^2)^{(1+\gamma+d)/2}}.$$

由于 $\rho_n=\overset{n}{\cancel{\text{对数}n\ 2\pi t}}()^{d/2}$，因此 $p_t(x)>c_\alpha\rho_n$ 推出

$$\|x\|\leqslant(C_\gamma(t)/(c_\alpha\rho_n))^{\frac{1}{1+\gamma+d}}\lesssim(C_\gamma(t)\rho_n)^{-\frac{1}{1+\gamma+d}}.$$

因此，我们得到

$$|\mathcal{G}_1|\leqslant\left|\left\{x:\|x\|\leqslant(C_\gamma(t)\rho_n)^{-\frac{1}{1+\gamma+d}}\right\}\right|\lesssim(C_\gamma(t)\rho_n)^{-\frac{d}{1+\gamma+d}}.$$

根据霍尔德不等式，有

$$\begin{aligned}\int_{\mathcal{G}_1}\|s_t(x)\|^2\mathrm{d}x&=\int_{G_1}\|s_t(x)\|^2\frac{1}{p_t(x)}p_t(x)\,\mathrm{d}x\\&\leqslant\left(\int_{\mathcal{G}_1}\|s_t(x)\|^{2m}p_t(x)\,\mathrm{d}x\right)^{1/m}\left(\int_{\mathcal{G}_1}p_t(x)^{-m'}p_t(x)\,\mathrm{d}x\right)^{1/m'},\end{aligned}$$

其中 m，$m'>1$ 且满足 $\frac{1}{m}+\frac{1}{m'}=1$。由引理 18，有

$$\left(\int_{\mathcal{G}_1}\|s_t(x)\|^{2m}p_t(x)\,\mathrm{d}x\right)^{1/m}\lesssim mt^{-1},$$

利用 $|\mathcal{G}_1|$的上界，我们有

$$\begin{aligned}\left(\int_{\mathcal{G}_1}p_t(x)^{-m'}p_t(x)\mathrm{d}x\right)^{1/m'}&=\left(\int_{\mathcal{G}_1}p_t(x)^{1-m'}\mathrm{d}x\right)^{1/m'}\\&\leqslant[|\mathcal{G}_1|(c_\alpha\rho_n)^{1-m'}]^{1/m'}\\&\leqslant C_\gamma(t)^{-\frac{1}{m}}\rho_n^{-\frac{1}{m}-\frac{d}{(1+\gamma+d)m'}}.\end{aligned}$$

合并这两项，可得

$$\int_{\mathcal{G}_1}\|s_t(x)\|^2\mathrm{d}x\lesssim C_\gamma(t)^{-\frac{1}{m}}mt^{-1}\rho_n^{-\frac{1}{m}-\frac{d}{(1+\gamma+d)m'}}.$$

□

证明。 [引理 17] 的证明。注意，

$$\int_{\mathcal{G}_2}p_t(x)\,\mathrm{d}x=\mathbb{P}_{X\sim p_t}(p_t(X)\leqslant c_\alpha\rho_n).$$

根据 H¨older 不等式和引理 18，可得

$$\begin{aligned}\int_{\mathcal{G}_2}\|s_t(x)\|^2 p_t(x)\,\mathrm{d}x &= \mathbb{E}_{X\sim p_t}\left[\|s_t(X)\|^2\mathbb{1}\{p_t(x)\leqslant c_\alpha\rho_n\}\right]\\ &\leqslant \mathbb{E}_{X\sim p_t}[\|s_t(X)\|^{2m}]^{\frac{1}{m}}\mathbb{E}[\mathbb{1}\{p_t(X)\leqslant c_\alpha\rho_n\}]^{1-\frac{1}{m}}\\ &\lesssim mt^{-1}\mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n)^{1-\frac{1}{m}}\,.\end{aligned}$$

令 $R_n = \left(\frac{C_\gamma(t)}{c_\alpha\rho_n}\right)^{\frac{1}{d+\gamma+1}}$，根据引理 23，可得

$$\{x:\|x\|\geqslant R_n\}\subset\{x:p_t(x)\leqslant c_\alpha\rho_n\}\,.$$

然后，我们可以将该概率分解为

$$\begin{aligned}\mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n) &= \mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n,\|X\|<R_n)+\mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n,\|X\|\geqslant R_n)\\ &\leqslant \mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n,\|X\|<R_n)+\mathbb{P}(\|X\|\geqslant R_n)\\ &= \int_{\|x\|<R_n} p_t(x)\mathbb{1}\{p_t(x)\leqslant c_\alpha\rho_n\}\,\mathrm{d}x+\int_{\|x\|\geqslant R_n}\frac{C_\gamma(t)}{(1+\|x\|^2)^{(d+\gamma+1)/2}}\mathrm{d}x\\ &\leqslant c_\alpha\rho_n\mathrm{Vol}(\mathbb{B}_{R_n})+C_\gamma(t)\cdot\frac{2\pi^{d/2}}{\Gamma(d/2)}\int_{R_n}^{\infty}\rho^{-\gamma-2}\mathrm{d}\rho \qquad (A.6)\\ &\leqslant c_\alpha\rho_n\cdot\frac{\pi^{d/2}}{\Gamma(d/2+1)}R_n^d+C_\gamma(t)\cdot\frac{2\pi^{d/2}}{\Gamma(d/2)}R_n^{-\gamma-1}\\ &= c_\alpha^{\frac{\gamma+1}{d+\gamma+1}}C_\gamma(t)^{\frac{d}{d+\gamma+1}}\left(\frac{\pi^{d/2}}{\Gamma(d/2+1)}+\frac{2\pi^{d/2}}{\Gamma(d/2)}\right)\cdot\rho_n^{\frac{\gamma+1}{d+\gamma+1}}\,.\end{aligned}$$

因此，我们得到

$$\int_{\mathcal{G}_2}p_t(x)\,\mathrm{d}x\lesssim C_\gamma(t)^{\frac{d}{d+\gamma+1}}\rho_n^{\frac{\gamma+1}{d+\gamma+1}}\,,$$

and

$$\int_{\mathcal{G}_2}\|s_t(x)\|^2 p_t(x)\,\mathrm{d}x\lesssim\varepsilon^{-1}t^{-1}C_\gamma(t)^{\frac{d(1-\varepsilon)}{d+\gamma+1}}\rho_n^{\frac{(\gamma+1)(1-\varepsilon)}{(d+\gamma+1)}}$$

如所需。 □

证明。 [引理 19] 的证明我们从全变差距离的定义开始

$$\mathrm{TV}(p_T,\varphi_T)=\frac{1}{2}\int_{\mathbb{R}^d}|p_T(x)-\varphi_T(x)|\mathrm{d}x.$$

将 p_T 的卷积表示代入全变差距离，并应用三角不等式，得到

$$\begin{aligned}|p_T(x)-\varphi_T(x)| &= \left|\int_{\mathbb{R}^d}p_0(z)\varphi_T(x-z)\mathrm{d}z-\varphi_T(x)\right|\\ &= \left|\int_{\mathbb{R}^d}p_0(z)\varphi_T(x-z)\mathrm{d}z-\int_{\mathbb{R}^d}p_0(z)\varphi_T(x)\mathrm{d}z\right|\end{aligned}$$

$$= \left| \int_{\mathbb{R}^d} p_0(z)(\varphi_T(x-z) - \varphi_T(x))\mathrm{d}z \right|$$
$$\leqslant \int_{\mathbb{R}^d} p_0(z)|\varphi_T(x-z) - \varphi_T(x)|\mathrm{d}z.$$

关于 x 对等式两边在 $\mathbb{R}^d$ 上积分，然后交换积分次序，得到

$$\int_{\mathbb{R}^d} |p_T(x) - \varphi_T(x)|\mathrm{d}x \leqslant \int_{\mathbb{R}^d} p_0(z)\left(\int_{\mathbb{R}^d} |\varphi_T(x-z) - \varphi_T(x)|\mathrm{d}x\right)\mathrm{d}z.$$

根据全变差距离的定义，有

$$\int_{\mathbb{R}^d} |\varphi_T(x-z) - \varphi_T(x)|\mathrm{d}x = 2\,\mathrm{TV}(\varphi_T(\cdot - z)\|\varphi_T).$$

将其代入前一个不等式，得到

$$\int_{\mathbb{R}^d} |p_T(x) - \varphi_T(x)|\mathrm{d}x \leqslant 2\int_{\mathbb{R}^d} p_0(z)\,\mathrm{TV}(\varphi_T(\cdot - z), \varphi_T)\mathrm{d}z.$$

两边除以因子 2，得到

$$\mathrm{TV}(p_T, \varphi_T) \leqslant \int_{\mathbb{R}^d} p_0(z)\,\mathrm{TV}(\varphi_T(\cdot - z), \varphi_T)\mathrm{d}z.$$

根据 Pinsker 不等式，有

$$\mathrm{TV}(\varphi_T(\cdot - z), \varphi_T) \leqslant \sqrt{\frac{1}{2}\,\mathrm{KL}(\mathcal{N}(z, TI_d), \mathcal{N}(0, TI_d))} = \frac{\|z\|}{2\sqrt{T}}.$$

因此，

$$\begin{aligned}\mathrm{TV}(p_T, \varphi_T) &\leqslant \int_{\mathbb{R}^d} p_0(z) \cdot \frac{\|z\|}{2\sqrt{T}}\mathrm{d}z \\ &= \frac{1}{2\sqrt{T}} \int_{\mathbb{R}^d} \|z\| p_0(z)\mathrm{d}z \\ &= \frac{M_1}{2\sqrt{T}}.\end{aligned}$$

由于根据假设 $M_1 < \infty$，证明完成。 □

A.4.2 指数衰减引理的证明

证明。 [引理 20] 的证明

$$p_t(x) = \int_{\mathbb{R}^d} p_0(y)\varphi_t(x-y)\,\mathrm{d}y.$$

为了分析较大 $\|x\|$ 时的衰减行为，我们将积分区域 $\mathbb{R}^d$ 划分为两个区域。令 $R = \|x\|/2$。我们定义区域 $D_1 = \{y : \|x-y\| \geqslant R\}$ 以及 $D_2 = \{y : \|x-y\| < R\}$。

我们首先考虑在第一区域 D_1 上的积分。在该区域中， x 与y 之间的距离满足 $\|x-y\| \geqslant \|x\|/2$。因此，高斯核 $\varphi_t(x-y)$ 会迅速衰减。我们可以推导出该核函数的上界

$$\varphi_t(x-y) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x-y\|^2}{2t}\right) \leqslant \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x\|^2}{8t}\right).$$

由于 $\int_{D_1} p_0(y)\mathrm{d}y \leqslant 1$，因此我们有

$$I_1 = \int_{D_1} p_0(y)\varphi_t(x-y)\,\mathrm{d}y \leqslant \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x\|^2}{8t}\right).$$

接下来，我们转向第二区域 D_2。在该区域中，噪声较小，满足 $\|x-y\| < \|x\|/2$。根据反三角不等式， y 的大小有如下下界

$$\|y\| = \|x-(x-y)\| \geqslant \|x\| - \|x-y\| > \frac{\|x\|}{2}.$$

利用初始密度上的假设 1 p_0，对于任意 $y \in D_2$，我们有

$$p_0(y) \leqslant C_1 \exp(-c_1\|y\|^\gamma) \leqslant C_1 \exp\left(-c_1\frac{\|x\|^\gamma}{2^\gamma}\right).$$

与前一步类似，我们将 p_0 在积分区域外取上确界，以此界定该区域的贡献。由于 $\int_{D_2} \varphi_t(x-y)\mathrm{d}y \leqslant 1$，因此有

$$I_2 = \int_{D_2} p_0(y)\varphi_t(x-y)\,\mathrm{d}y \leqslant C_1 \exp\left(-2^{-\gamma}c_1\|x\|^\gamma\right).$$

结合这两个界，总密度满足：

$$p_t(x) = I_1 + I_2 \leqslant \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x\|^2}{8t}\right) + C_1 \exp\left(-2^{-\gamma}c_1\|x\|^\gamma\right).$$

当 $\|x\|$ 足够大时，衰减较慢的项将主导总和。若 $\gamma < 2$，则 $\exp(-2^{-\gamma}c_1\|x\|^\gamma)$ 项占主导。若 $\gamma \geqslant 2$，则高斯分布项占主导，或以相近的速率衰减。因此，我们得出 $p_t(x) \leqslant \mathcal{C}_\gamma(t)\exp\left(-c_\gamma(t)\min(\|x\|^2, \|x\|^\gamma)\right)$，其中 $\mathcal{C}_\gamma(t) = \frac{1}{(2\pi t)^{d/2}} + C_1$，且 $c_\gamma(t) = \min\left(\frac{1}{8t}, \frac{c_1}{2^\gamma}\right)$。 □

Proof. [引理 21] 的证明。条件 $\mathcal{C}_\gamma(t)\exp(-c_\gamma(t)\|x\|^\gamma) \geqslant c_\alpha\rho_n$ 意味着

$$\|x\|^\gamma \leqslant \frac{1}{c_\gamma(t)} \log\left(\mathcal{C}_\gamma(t)/c_\alpha\rho_n\right).$$

因此，半径为 $R_n \asymp (\log(1/\rho_n))^{1/\gamma}$。体积按 $R_n^d \asymp (\log(1/\rho_n))^{d/\gamma}$ 的比例增长。与 n 中的多项式因子相比，这种对数增长可以忽略不计。因此，我们得到

$$\begin{aligned}\int_{\mathcal{G}_1} \|s_t(x)\|^2\mathrm{d}x &\lesssim mt^{-1}\left(\int_{\mathcal{G}_1} p_t(x)^{1-m'}\mathrm{d}x\right)^{1/m'} \\ &\leqslant mt^{-1}\left[|\mathcal{G}_1|(c_\alpha\rho_n)^{1-m'}\right]^{1/m'}\end{aligned}$$

$$\leqslant mt^{-1}\left[(\log(\mathcal{C}_\gamma(t)/c_\alpha\rho_n)/c_\gamma(t))^{d/\gamma}(c_\alpha\rho_n)^{1-m'}\right]^{1/m'}$$
$$\lesssim \mathrm{polylog}(nt)\cdot mt^{-1}c_\gamma(t)^{d/(\gamma m')}\rho_n^{-1/m}$$

正如所需。 □

证明。 [引理 22] 令 $k=\min(\gamma,2)$。当 $\|x\|$充分大时，尾部的行为如 $\exp(-c_\gamma(t)\|x\|^k$。于是，尾部概率有如下界

$$\begin{aligned}\mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n) &\leqslant \mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n, \|X\|<R_n)+\mathbb{P}(\|X\|\geqslant R_n)\\ &=\int_{\|x\|<R_n} p_t(x)\mathbb{1}\{p_t(x)\leqslant c_\alpha\rho_n\}\,\mathrm{d}x+\int_{\|x\|\geqslant R_n}\mathcal{C}_\gamma(t)\exp(-c_\gamma(t)\|x\|^k)\,\mathrm{d}x\\ &\leqslant c_\alpha\rho_n\mathrm{Vol}(\mathbb{B}_{R_n})+\mathcal{C}_\gamma(t)\frac{2\pi^{d/2}}{\Gamma(d/2)}\int_{R_n}^\infty r^{d-1}\exp(-c_\gamma(t)r^k)\,\mathrm{d}r\\ &\leqslant c_\alpha\rho_n\frac{\pi^{d/2}}{\Gamma(d/2+1)}R_n^d+\mathcal{C}_\gamma(t)\frac{2\pi^{d/2}}{\Gamma(d/2)}\cdot\frac{R_n^{d-k}}{kc_\gamma(t)}\exp(-c_\gamma(t)R_n^k)\,.\end{aligned}$$

利用标准的尾部渐近

$$\int_R^\infty r^{d-1}e^{-cr^k}\mathrm{d}r\sim\frac{R^{d-k}}{ck}e^{-cR^k}\qquad\text{as } R\to\infty$$

我们得到

$$\begin{aligned}\mathbb{P}(p_t(X)\leqslant c_\alpha\rho_n) &= c_\alpha\rho_nC_1R_n^d+\mathcal{C}_\gamma(t)\exp(-c_\gamma(t)R_n^k)\cdot C_2R_n^{d-k}\\ &\leqslant C'\rho_n(R_n^d+R_n^{d-k})\\ &\lesssim \rho_nR_n^d\\ &\asymp \rho_n(\log(1/\rho_n))^{d/k}\,.\end{aligned}$$

其中，最后一步使用了由尾部衰减推导出的关系

$$\exp(-c_\gamma(t)R_n^k)\asymp\rho_n$$

由此可得

$$R_n\asymp(\log(1/\rho_n))^{1/k}\,.$$

证明至此完成。

□

A.5 多项式衰减目标的替代低噪声策略

如第 5.1.1 节所述，在低噪声区域（扩散时间较小 t）中，基于内在扩散尺度 $\sqrt{t}$ 的朴素估计量会因样本稀疏性而在重尾区域遭遇方差爆炸。为缓解这一问题而不采用提前停止，我们引入一种解耦核分数估计器。本节中，我们假设 p_0 满足假设 3，其中 $\beta>2$；此外，为了实现逐点误差控制，我们进一步对 ¨施加 H $p_0$ölder 连续性条件。

假设 4 (H¨older 正则性)。目标密度 p_0 有界，且属于 $H\ddot{o}lder$ 类 $\Sigma(\beta, L)$，其中 $\beta > 2$。令 $\ell = \lfloor \beta \rfloor$ 表示 β 的整数部分。假设 p_0 是 ℓ 次连续可微，且其 ℓ 阶偏导数是 $H\ddot{o}lder$ 连续的

$$|D^\alpha p_0(x) - D^\alpha p_0(y)| \leqslant L\|x - y\|^{\beta-\ell}, \quad \forall x, y \in \mathbb{R}^d, |\alpha| = \ell,$$

其中， D^α 表示与多重指标 α对应的混合偏导数， $L > 0$ 是一个与 x 和 y无关的通用常数。

我们注意到，对于所有 t，H¨older 常数 L均保持不变，因为

$$\begin{aligned}
|D^\alpha p_t(x) - D^\alpha p_t(y)| &= |D^\alpha (p_0 * \varphi_t)(x) - D^\alpha (p_0 * \varphi_t)(y)| \\
&= \int_{\mathbb{R}^d} [(D^\alpha p_0)(x - z) - (D^\alpha p_0)(y - z)]\varphi_t(z)\mathrm{d}z \\
&\leqslant \int_{\mathbb{R}^d} L\|x - y\|^{\beta-\ell}\varphi_t(z)\mathrm{d}z \\
&= L\|x - y\|^{\beta-\ell}.
\end{aligned}$$

得分估计量构造。令 $K: \mathbb{R}^d \to \mathbb{R}$ 为一个阶数为$\ell = \lfloor \beta \rfloor$、支撑集为 $[-1, 1]^d$的紧支撑核函数，满足对于所有满足 $1 \leqslant |\alpha| \leqslant \ell$的多重指标$\alpha$，有 $\int_{\mathbb{R}^d} K(u)\mathrm{d}u = 1$, $\int_{\mathbb{R}^d} u^\alpha K(u)\,\mathrm{d}u = 0$，且 K， $\nabla K \in L^2(\mathbb{R}^d)$。

p_0 的核密度估计量（带宽为 $h > 0$ ）定义为

$$\hat{f}_h(y) = \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{y - X_i}{h}\right), \quad \text{where } X_i \overset{\text{i.i.d.}}{\sim} p_0.$$

密度估计量通过与高斯核 φ_t卷积来定义

$$\hat{p}_t(x) := (\hat{f}_h * \varphi_t)(x) = \int_{\mathbb{R}^d} \hat{f}_h(y)\varphi_t(x - y)\,\mathrm{d}y.$$

我们使用统一阈值 $\rho_n = C_\rho h^\beta$对得分估计量进行稳定化，其中 $C_\rho > 0$ 取足够大的值。具体而言，

$$\hat{s}_t(x) := \frac{\nabla \hat{p}_t(x)}{\hat{p}_t(x)}\mathbb{1}_{\{\hat{p}_t(x) \geqslant \rho_n\}}.$$

我们首先利用傅里叶分析建立全局平均积分平方误差界，然后推导所提出的得分估计量的逐点误差界。

命题 24 （全局平均积分平方误差界）。 假设 *3* 成立。设核函数 K满足傅里叶域近似性质

$$|\mathcal{F}[K](v) - 1| \leqslant C_K \min(\|v\|^\beta, 1), \qquad \textit{for small } \|v\|,$$

对于某个 C_K，其取决于 K 和 d。那么，对于任意 $t > 0$，解耦估计量满足

$$\begin{aligned}
\|\nabla \hat{p}_t - \nabla p_t\|_{\mathbb{L}_2}^2 &\leqslant (C_K^2 + 1)L^2 h^{2(\beta-1)} + \frac{\|\nabla K\|_{L^2}^2}{nh^{d+2}} \\
\|\hat{p}_t - p_t\|_{\mathbb{L}_2}^2 &\leqslant (C_K^2 + 1)L^2 h^{2\beta} + \frac{\|K\|_{L^2}^2}{nh^d},
\end{aligned}$$

特别是，这些界限关于 t. 是一致的

证明。 我们采用傅里叶变换约定下的普朗歇尔定理 n

$$\mathcal{F}[g](\omega) = \int_{\mathbb{R}^d} g(x) e^{-i\omega\cdot x} \mathrm{d}x$$

。随后，我们将平均积分平方误差分解为积分平方偏差和积分方差。我们重点讨论 $\nabla\hat{} p_t$。密度情形可用相同的论证得出。

我们先讨论积分平方偏差。注意，

$$\mathbb{E}[\nabla \hat{p}_t] = \nabla(p_0 * K_h) * \varphi_t \,,$$

因此，点态偏差可写为

$$\mathrm{Bias}(x) = \nabla((p_0 * K_h) - p_0) * \varphi_t(x) \,.$$

取傅里叶变换可得

$$\mathcal{F}[\mathrm{Bias}](\omega) = i\omega(\mathcal{F}[K](h\omega) - 1)\mathcal{F}[p_0](\omega) e^{-t\|\omega\|^2/2}.$$

由普朗歇尔定理，偏差的平方 $\mathbb{L}_2$范数因此为

$$\|\mathrm{Bias}(\nabla \hat{p}_t)\|_{\mathbb{L}_2}^2 = \|\mathrm{Bias}(\nabla \hat{p}_t)\|_{L^2}^2 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \|\omega\|^2 |\mathcal{F}[K](h\omega) - 1|^2 |\mathcal{F}[p_0](\omega)|^2 e^{-t\|\omega\|^2} \,\mathrm{d}\omega.$$

利用界 $e^{-t\|\omega\|^2} \leqslant 1$ 以及核的逼近性质

$$|\mathcal{F}[K](z) - 1| \leqslant C_K \|z\|^\beta \qquad \text{when } \|z\| < 1$$

以及该界

$$|\mathcal{F}[K](z) - 1| \leqslant 2 \qquad \text{when } \|z\| \geqslant 1,$$

我们将积分拆分为区域 $\|h\omega\| < 1$ 和 $\|h\omega\| \geqslant 1$上的积分。由此得到

$$\begin{aligned}
\|\mathrm{Bias}(\nabla \hat{p}_t)\|_{\mathbb{L}_2}^2 &\leqslant \frac{1}{(2\pi)^d} \left(C_K^2 \int_{\|h\omega\|<1} \|\omega\|^2 \|h\omega\|^{2\beta} |\mathcal{F}[p_0](\omega)|^2 \,\mathrm{d}\omega + \int_{\|h\omega\|\geqslant 1} \|\omega\|^2 \cdot 4 |\mathcal{F}[p_0](\omega)|^2 \,\mathrm{d}\omega \right) \\
&\leqslant \frac{h^{2(\beta-1)}}{(2\pi)^d} \left(C_K^2 \int_{\|h\omega\|<1} \|\omega\|^{2\beta} |\mathcal{F}[p_0](\omega)|^2 \,\mathrm{d}\omega + \int_{\|h\omega\|\geqslant 1} \|\omega\|^{2\beta} |\mathcal{F}[p_0](\omega)|^2 \,\mathrm{d}\omega \right) \\
&\leqslant \frac{(C_K^2 + 1) L^2 h^{2(\beta-1)}}{(2\pi)^d} \int_{\mathbb{R}^d} \|\omega\|^{2\beta} |\mathcal{F}[p_0](\omega)|^2 \,\mathrm{d}\omega \\
&\leqslant (C_K^2 + 1) L^2 h^{2(\beta-1)},
\end{aligned}$$

其中，最后一步由以下事实得出： $\int_{\mathbb{R}^d} \|\omega\|^{2\beta} |\mathcal{F}[p_0](\omega)|^2 \mathrm{d}\omega \leqslant (2\pi)^d L^2$。

接下来，我们分析积分方差。定义 $\Psi_{h,\ t} = \nabla K_h * \varphi_t$，则 $\nabla\hat{} p_t$ 的积分方差有界于 $\frac{1}{n}\|\Psi_{h,\ t}\|_{\mathbb{L}_2}^2$。根据Young卷积不等式和 $\neg\|\varphi_t\|_{L^1} = 1$，有

$$\|\Psi_{h,t}\|_{\mathbb{L}_2} = \|\Psi_{h,t}\|_{L^2} \leqslant \|\nabla K_h\|_{L^2} \|\varphi_t\|_{L^1} = \|\nabla K_h\|_{L^2}.$$

根据核梯度的尺度关系，可得

$$\nabla K_h(x) = h^{-(d+1)} \nabla K(x/h) \,,$$

因此

$$\|\nabla K_h\|_{L^2}^2 = \int_{\mathbb{R}^d} h^{-2(d+1)} |\nabla K(x/h)|^2 \mathrm{d}x = h^{-(d+2)} \|\nabla K\|_{L^2}^2.$$

代入该恒等式可得积分方差项 $\frac{\|\nabla K\|_{L^2}^2}{nh^{d+2}}$。

□

命题 25 （一致点态偏差）。 在假设 4下，对于任意 $t > 0$ 和 $x \in \mathbb{R}^d$，解耦估计量的点态偏差满足：

$$|\mathbb{E}[\hat{p}_t(x)] - p_t(x)| \leqslant C_{B,0} h^{\beta},$$
$$\|\mathbb{E}[\nabla \hat{p}_t(x)] - \nabla p_t(x)\| \leqslant C_{B,1} h^{\beta-1},$$

其中 $C_{B,0} = L \cdot \frac{(2d)^{\ell}}{\ell!} \int_{\mathbb{R}^d} \|u\|^{\beta} |K(u)| \mathrm{d}u$ 以及 $C_{B,1} = L \cdot \frac{(2d)^{\ell-1}}{(\ell-1)!} \int_{\mathbb{R}^d} \|u\|^{\beta-1} \|\nabla K(u)\| \mathrm{d}u$，其仅取决于 L、d、ℓ 和核函数。

证明。 我们首先研究密度估计量 $\hat{p}_t$ 的偏差。该偏差可表示为

$$\mathrm{Bias}(\hat{p}_t)(x) = ((p_0 * K_h) - p_0) * \varphi_t(x).$$

由于 φ_t 是概率密度，与 φ_t 进行卷积是一个 L^{∞}-压缩。因此，

$$\sup_x |\mathrm{Bias}(\hat{p}_t)(x)| \leqslant \sup_y |(p_0 * K_h)(y) - p_0(y)|.$$

对于任意固定的 y，利用带积分余项的多元泰勒公式，将 $p_0(y - hu)$ 在 y 附近展开，可得

$$\begin{aligned} p_0(y - hu) = p_0(y) &+ \sum_{1 \leqslant |\alpha| \leqslant \ell} \frac{(-h)^{|\alpha|} u^{\alpha}}{\alpha!} D^{\alpha} p_0(y) \\ &+ \frac{(-h)^{\ell}}{\ell!} \sum_{|\alpha| = \ell} \alpha! \int_0^1 (1-s)^{\ell-1} [D^{\alpha} p_0(y - shu) - D^{\alpha} p_0(y)] \mathrm{d}s \, u^{\alpha}. \end{aligned}$$

多项式项在积分后消失，因为核函数具有所需的消失矩。根据假设4，余项满足

$$|R_{\ell}(y, -hu)| \leqslant L \cdot \frac{(2d)^{\ell}}{\ell!} \|hu\|^{\beta}.$$

因此，

$$|(p_0 * K_h)(y) - p_0(y)| \leqslant \int_{\mathbb{R}^d} |K(u)| \cdot L \cdot \frac{(2d)^{\ell}}{\ell!} h^{\beta} \|u\|^{\beta} \mathrm{d}u =: C_{B,0} h^{\beta}.$$

Next, 我们分析梯度估计量 $\nabla \hat{p}_t$的偏差。它可以写为 $\nabla \hat{p}_t$:

$$\mathrm{Bias}(\nabla \hat{p}_t)(x) = (\nabla (p_0 * K_h) - \nabla p_0) * \varphi_t(x).$$

利用恒等式 $\nabla (p_0 * K_h) = (\nabla p_0) * K_h$，我们可以用 ∇p_0 替代 p_0，重复相同的泰勒展开论证。由于当 $\beta > 1$ 时， $\nabla p_0 \in \Sigma(\beta - 1, L)$，因此可得与之前完全相同的界，只需将 β 替换为 $\beta - 1$，并将 K 替换为 ∇K。

□

命题 26（一致逐点方差）。对于任意 $t>0$ 和 $x\in\mathbb{R}^d$，解耦估计量的逐点方差由局部密度 $p_t^*(x)$ = 上确界$_{\|y-x\|\leqslant h}\, p_t(y)$ 所界定，即

$$\mathrm{Var}(\hat{p}_t(x)) \leqslant \frac{\|K\|_{L^2}^2\, p_t^*(x)}{nh^d}\,, \quad \mathrm{Var}(\nabla\hat{p}_t(x)) \leqslant \frac{\|\nabla K\|_{L^2}^2\, p_t^*(x)}{nh^{d+2}}\,,$$

其中这些界与 t 无关。

证明。我们首先界定密度估计量 $\hat{p}_t$ 的方差。注意，

$$\hat{p}_t(x) = \frac{1}{n}\sum_{i=1}^n Z_i, \qquad Z_i = (K_h * \varphi_t)(x - X_i)\,.$$

由于 Z_i 是独立同分布的，我们有

$$\mathrm{Var}(\hat{p}_t(x)) \leqslant \frac{1}{n}\mathbb{E}[Z_1^2]\,.$$

根据詹森不等式，对概率密度 φ 的卷积应用该不等式可得

$$(K_h * \varphi_t)^2(z) \leqslant (K_h^2 * \varphi_t)(z).$$

这意味着

$$\mathbb{E}[Z_1^2] \leqslant (K_h^2 * p_t)(x)\,.$$

若 K 的支撑集是紧集，则 $K_h^2(u) = h^{-2d}K^2(u/h)$ 的支撑集包含于 $[-h,h]^d$，通过变量替换可得

$$(K_h^2 * p_t)(x) = \frac{1}{h^d}\int_{\mathbb{R}^d} K^2(u)p_t(x-hu)\mathrm{d}u \leqslant \frac{p_t^*(x)}{h^d}\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u = \frac{\|K\|_{L^2}^2 p_t^*(x)}{h^d}.$$

其中 $p_t^*(x) := \sup_{\|u\|\leqslant 1} p_t(x-hu)$。除以 n 后，即可得到 $\hat{p}_t(x)$ 的方差上界。

接下来，我们对梯度估计量 $\nabla\hat{p}_t$的方差进行界定。注意，

$$\nabla\hat{p}_t(x) = \frac{1}{n}\sum_{i=1}^n W_i, \qquad W_i = (\nabla K_h * \varphi_t)(x - X_i)\,.$$

利用缩放关系

$$\nabla K_h(u) = h^{-(d+1)}\nabla K(u/h)\,,$$

我们看到，与 K_h相比，梯度会额外引入一个因子 h^{-1}。应用与密度情形相同的詹森不等式和紧支撑论证，可得到类似的上界，其中 $\|K\|_{L^2}^2$ 替换为 $\|\nabla K\|_{L^2}^2$，h^d 替换为 h^{d+2}.

□

引理 27（高概率集中）。设 $\hat{p}_t(x) = (f_h * \varphi_t)(x)$，其带宽为 $h \asymp n^{-1/(2\beta+d)}$，阈值为 $\rho_n = C_\rho h^\beta$，其中 $C_\rho > \max(2C_{B,0}, L\cdot\frac{(2d)^\ell)!}{2^\ell})$。对于任意 $t>0$ 和 ~~$x\in$~~ $\mathbb{R}^d$，以至少 $1-n^{-\alpha}$ 的概率：

$$|\hat{p}_t(x) - p_t(x)| \leqslant C_{B,0}h^\beta + C_\alpha\left(\sqrt{p_t^*(x)\frac{\log n}{nh^d}} + \frac{\log n}{nh^d}\right),$$

其中 C_α = 最大值$\left(\sqrt{8\alpha\|K\|_{L^2}^2}, \frac{8\alpha\|K\|_\infty}{3}\right)$。在集合 $G_1 = \{x : p_t(x) \geqslant 2\rho_n\}$ 上，对于充分大的 n，我们进一步有$\hat{p}_t(x) \geqslant \frac{1}{2}p_t(x)$。

证明。 我们首先对 $Z_i = (K_h * \varphi_t)(x - X_i)$ 应用伯恩斯坦不等式，其中 $|Z_i| \leqslant h^{-d}\|K\|_\infty =: M$，然后 $\mathrm{Var}(Z_i) \leqslant \frac{\|K\|_{L^2}^2 p_t^*(x)}{h^d}$。令 $\lambda = \max\left(\sqrt{4\alpha\sigma^2 \log(2n)\frac{4\alpha M}{3n}\log(\frac{2n}{3n})}\right)$。这确保了

$$\mathbb{P}(|\hat{p}_t - \mathbb{E}\hat{p}_t| > \lambda) \leqslant n^{-\alpha}.$$

总误差分解为偏差和随机误差

$$|\hat{p}_t - p_t| \leqslant |\mathbb{E}\hat{p}_t - p_t| + |\hat{p}_t - \mathbb{E}\hat{p}_t| \leqslant C_{B,0}h^\beta + C_\alpha\left(\sqrt{p_t^*(x)\frac{\log n}{nh^d}} + \frac{\log n}{nh^d}\right),$$

其中 $C_\alpha = \max\left(\sqrt{8\alpha\|K\|_{L^2}^2}, \frac{8\alpha\|K\|_\infty}{3}\right)$。由 p_t 的 H¨older 正则性可得

$$|p_t(x + hu) - p_t(x)| \leqslant L \cdot \frac{(2d)^\ell}{\ell!}\|hu\|^\beta .$$

因此，对于 $x \in G_1$，有

$$p_t^*(x) \leqslant p_t(x) + L \cdot \frac{(2d)^\ell}{\ell!}h^\beta \leqslant 2p_t(x) .$$

因此，随机项满足

$$\sqrt{p_t^*(x)\frac{\log n}{nh^d}} \leqslant \sqrt{2p_t(x) \cdot \frac{\log n}{C_\rho^2}\rho_n^2} = \sqrt{\frac{h^\beta \log n}{C_\rho}}p_t(x).$$

我们还有

$$\frac{\log n}{nh^d} \leqslant \frac{h^\beta \log n}{C_\rho}\rho_n.$$

因此，

$$|\hat{p}_t - p_t| \leqslant \frac{1}{2}\rho_n + C_\alpha\left(\sqrt{\frac{h^\beta \log n}{C_\rho}} + \frac{h^\beta \log n}{2C_\rho}\right)p_t(x) \leqslant \frac{1}{2}p_t(x)$$

当 n 足够大时，意味着 $\hat{p}_t(x) \geqslant \frac{1}{2}p_t(x)$。 □

假设5 (分数正则性**)**。 初始分布 p_0 满足 $\mathbb{E}_{p_0}[\|s_0(x)\|^{2k}] < \infty$，其中某个 $k \geqslant 1$。

引理28 (分数矩的单调性**)**。 假设初始分数函数 $s_0 = \nabla$对数 p_0满足 $\mathbb{E}_{p_0}[\|s_0(x)\|^{2k}] < \infty$，其中某个 $k \geqslant 1$。那么，对于任意 $t > 0$，扩散分数 s_t 的矩随 t非递增

$$\mathbb{E}_{p_t}[\|s_t(x)\|^{2k}] \leqslant \mathbb{E}_{p_0}[\|s_0(x)\|^{2k}].$$

证明。 我们利用将扩散分数解释为条件期望的方法（通常称为 Tweedie公式）。分数函数可以表示为

$$s_t(x) = \nabla \log p_t(x) = \frac{\nabla p_t(x)}{p_t(x)} = \frac{\int_{\mathbb{R}^d} \nabla p_0(y)\varphi_t(x - y)\,\mathrm{d}y}{p_t(x)}.$$

利用恒等式 $\nabla p_0(y) = s_0(y)p_0(y)$，我们得到

$$s_t(x) = \int_{\mathbb{R}^d} s_0(y)\frac{p_0(y)\varphi_t(x-y)}{p_t(x)}\,\mathrm{d}y.$$

注意到 $w(y|x) := \frac{p_0(y)\varphi_t(x-y)}{p_t(x)}$ 恰好对应后验密度 $p(y|x)$（即给定含噪观测 x时，干净数据 y 的条件分布）。因此， $s_t(x)$ 是初始得分函数的后验期望

$$s_t(x) = \mathbb{E}_{Y|X=x}[s_0(Y)].$$

考虑凸函数 $\Phi(u) = \|u\|^{2k}$ （当 $k \geqslant 1$时）。根据条件期望的詹森不等式，有

$$\|s_t(x)\|^{2k} = \left\|\mathbb{E}_{Y|X=x}[s_0(Y)]\right\|^{2k} \leqslant \mathbb{E}_{Y|X=x}\left[\|s_0(Y)\|^{2k}\right].$$

现在，我们针对边缘密度 $p_t(x)$ 计算全局期望

$$\begin{aligned}\mathbb{E}_{p_t}[\|s_t(x)\|^{2k}] &= \int_{\mathbb{R}^d} \|s_t(x)\|^{2k} p_t(x)\,\mathrm{d}x \\ &\leqslant \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} \|s_0(y)\|^{2k}\frac{p_0(y)\varphi_t(x-y)}{p_t(x)}\,\mathrm{d}y\right)p_t(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \|s_0(y)\|^{2k} p_0(y)\varphi_t(x-y)\,\mathrm{d}y\,\mathrm{d}x.\end{aligned}$$

根据富比尼定理，有

$$\mathbb{E}_{p_t}[\|s_t(x)\|^{2k}] \leqslant \int_{\mathbb{R}^d} \|s_0(y)\|^{2k} p_0(y)\left(\int_{\mathbb{R}^d}\varphi_t(x-y)\,\mathrm{d}x\right)\mathrm{d}y.$$

由于 φ_t 是概率密度，且 $\int \varphi_t(x-y)\,\mathrm{d}x = 1$，因此

$$\mathbb{E}_{p_t}[\|s_t(x)\|^{2k}] \leqslant \int_{\mathbb{R}^d} \|s_0(y)\|^{2k} p_0(y)\,\mathrm{d}y = \mathbb{E}_{p_0}[\|s_0(x)\|^{2k}]$$

如所需。

□

现在，我们已准备好为所提出的得分估计量建立均方误差。

定理 29 (ˆ 的均方误差s_t)。 假设 2、3 以及 $\beta > 2$、4 和 5 均成立。令 $h \asymp n^{-\frac{1}{2\beta+d}}$并令 $\rho_n = C_\rho h^\beta \asymp n^{-\frac{\beta}{2\beta+d}}$。对于任意 $t > 0$，得分估计误差满足

$$\mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\,\mathrm{d}x\right] \lesssim n^{-\frac{\beta-2}{2\beta+d}} + n^{-\frac{\beta(\gamma+1)}{k'(2\beta+d)(d+\gamma+1)}},$$

其中 $\frac{1}{k} + \frac{1}{k'} = 1$。

证明。 我们根据密度估计量 ˆp_t 的表现，将加权平均积分平方误差 $\mathcal{R} = \mathbb{E}\left[\int_{\mathbb{R}^d} \|\hat{s}_t(x) - s_t(x)\|^2 p_t(x)\,\mathrm{d}x\right]$ 分解为两个不同的部分，以分析得分估计误差：

$$\mathcal{R} = E_{\text{trunc}} + E_{\text{est}},$$

其中 $E_{\text{trunc}} = \mathbb{E}\int_{\{\hat{p}_t<\rho_n\}} \|s_t\|^2 p_t \,\mathrm{d}x$，且 $E_{\text{est}} = \mathbb{E}\int_{\{\hat{p}_t\geqslant\rho_n\}} \left\| \frac{\nabla \hat{p}_t}{\hat{p}_t} - s_t \right\|^2 p_t \,\mathrm{d}x$。部分 E_{trunc}对应于 $\hat{p}_t$ 过小、无法定义有效得分估计的情形。部分E_{est} 则对应于 $\hat{p}_t$ 足够大、可以直接计算得分估计的情形。

我们引入两个辅助区域，以细化误差分析。定义理论高密度区域 $G_1 = \{x : \ p_t(x) \geqslant 2\rho_n\}$ 以及尾部区域 $G_2 = \{x : \ p_t(x) < 2\rho_n\}$。回顾引理 27 中的高概率事件 A_α。在该事件上，对于所有 $x \in G_1$，一致界 $|\hat{p}_t(x) - p_t(x)| \leqslant \frac{1}{2}p_t(x)$均成立。补事件 A_α^c 的概率为 $\mathbb{P}(A_\alpha^c) \leqslant n^{-\alpha}$，当 n 较大时该概率可忽略不计。

现在我们分析 E_{trunc}。为界定这一截断误差，我们将积分拆分到辅助区域 G_1 和 G_2上。该拆分将密度的主要质量与其尾部区分开来，从而简化界的推导：

$$E_{\text{trunc}} = \mathbb{E}\left[\int_{G_2\cap\{\hat{p}_t<\rho_n\}} \|s_t\|^2 p_t \,\mathrm{d}x\right] + \mathbb{E}\left[\int_{G_1\cap\{\hat{p}_t<\rho_n\}} \|s_t\|^2 p_t \,\mathrm{d}x\right].$$

我们首先考虑尾部区域的贡献。我们将这部分称为尾部截断误差。拆分后的第一项由 $\|s_t\|^2 p_t$ 在 G_2上的总质量界定。我们对 $\|s_t\|^2$应用具有共轭指数 k 和 k'的 Hölder 不等式，其中 $\frac{1}{k} + \frac{1}{k'} = 1$。假设 5 确保 $\mathbb{E}_{p_t}[\|s_t\|^{2k}] < \infty$，因此该不等式给出：

$$\int_{G_2} \|s_t\|^2 p_t \,\mathrm{d}x \leqslant \left(\int_{\mathbb{R}^d} \|s_t\|^{2k} p_t(x) \,\mathrm{d}x\right)^{1/k} \left(\int_{G_2} p_t(x) \,\mathrm{d}x\right)^{1/k'} \lesssim (\mathbb{P}(p_t(X) < 2\rho_n))^{1/k'}.$$

结合按照式（A.6）中的相同论证，尾部概率 $\mathbb{P}(p_t(X) < 2\rho_n)$ 的尺度关系为 $\rho_n^{\frac{1}{d+\gamma+1}}$。代入缩放关系 $\rho_n \asymp h^\beta$ 后，该项进一步满足尺度关系 $h^{\frac{\beta}{k'}(\gamma+1)(d+\gamma+1)}$。

接下来考虑高密度区域的贡献。在高概率事件A_α上，任意 $x \in G_1$ 都满足 $\hat{p}_t(x) \geqslant \frac{1}{2}p_t(x)$。由于对于 $x \in G_1$，$p_t(x) \geqslant 2\rho_n$，因此我们有 $\hat{p}_t(x) \geqslant \rho_n$。因此，集合 $G_1 \cap \{\hat{p}_t < \rho_n\}$ 在 A_α 上为空。该项的唯一贡献来自 A_α^c，其为 $O(n^{-\alpha})$，可以忽略。我们可以将这个可忽略项吸收到最终界中的常数因子里。

现在我们分析 E_{est}，即估计误差，其中 $\hat{p}_t \geqslant \rho_n$。下界 $\hat{p}_t \geqslant \rho_n$ 确保得分估计中的分母有下界。这给出了 $\hat{p}_t^{-1} \leqslant \rho_n^{-1}$，这是后续分析的一个关键界。利用三角不等式，我们将得分误差分解为两个部分。

$$\left\| \frac{\nabla \hat{p}_t}{\hat{p}_t} - \frac{\nabla p_t}{p_t} \right\|^2 p_t \lesssim T_1 + T_2,$$

其中 $T_1 = \frac{\|\nabla \hat{p}_t - \nabla p_t\|^2}{\hat{p}_t^2} p_t$ 且 $T_2 = \frac{|\hat{p}_t - p_t|^2}{\hat{p}_t^2} \|s_t\|^2 p_t$。

我们首先推导加权因子 $p_t(x)\,\hat{p}_t^2(x)$ 在所有满足 $\hat{p}_t(x) \geqslant \rho_n$ 的 x 上的一致界。这个一致界简化了 T_1 和 T_2 的分析。对于 $x \in G_1$，事件 A_α 给出 $\hat{p}_t(x) \geqslant \frac{1}{2}p_t(x)$。由此可得 $p_t(x)\,\hat{p}_t^2(x) \leqslant p_t^4(x)$。由于 $p_t(x)$ 对于 $x \in G_1$ 满足 $\geqslant 2\rho_n$，我们进一步得到 $p_t(x)\,\hat{p}_t^2(x) \leqslant \rho_n^2$。对于 $x \in G_2$，条件 $\hat{p}_t(x) \geqslant \rho_n$ 和 $p_t(x) < 2\rho_n$ 直接推出 $p_t(x)\,\hat{p}_t^2(x) < 2\rho_n^2 = \rho_n^2$。因此，在估计量的活跃区域内，加权因子 $p_t(x)\,\hat{p}_t^2(x)$ 一致地满足 $O(\rho_n^{-1})$。

我们首先分析梯度分量 T_1。我们使用加权因子的一致界，并代入命题24中的均方积分误差界。这一代入将梯度误差与密度估计的已知结果联系起来：

$$\mathbb{E}\left[\int_{\{x:\hat{p}_t(x)\geqslant\rho_n\}} T_1\,\mathrm{d}x\right] \lesssim \frac{1}{\rho_n}\|\nabla\hat{p}_t-\nabla p_t\|_{\mathbb{L}_2}^2 \leqslant \frac{1}{\rho_n}\left[(C_K^2+1)L^2h^{2(\beta-1)}+\frac{\|\nabla K\|_{L^2}^2}{nh^{d+2}}\right].$$

缩放关系 $\rho_n \asymp h^\beta$ 告诉我们，括号中的主导项为 $(C_K^2+1)L^2h^{2(\beta-1)}$。代入这一缩放关系后，得到 $T_1 \lesssim h^{-\beta}\cdot h^{2\beta-2}=h^{\beta-2}$。我们的假设 $\beta>2$ 确保该项收敛。

接下来，我们分析密度分量 T_2。我们再次应用 H¨older 不等式，将得分矩与密度误差项分离。这种分离利用了假设5 和引理28中得分矩的有界性，

$$\begin{aligned}\mathbb{E}\left[\int_{\{x:\,\hat{p}_t(x)\geqslant\rho_n\}} T_2\,\mathrm{d}x\right] &= \mathbb{E}\left[\int_{\{x:\,\hat{p}_t(x)\geqslant\rho_n\}}\left(\frac{|\hat{p}_t(x)-p_t|^2}{\hat{p}_t^2}\right)\|s_t\|^2 p_t\,\mathrm{d}x\right]\\ &\leqslant \left(\mathbb{E}_{p_t}[\|s_t\|^{2k}]\right)^{1/k}\left(\mathbb{E}\int_{\{x:\,\hat{p}_t(x)\geqslant\rho_n\}}\left(\frac{|\hat{p}_t-p_t|^2}{\hat{p}_t^2}\right)^{k'} p_t\,\mathrm{d}x\right)^{1/k'}.\end{aligned}$$

第一项由常数 $M_{2k}^{1/k}$ 控制。对于第二项，我们使用 $p_t^*(x)=\sup^{u\in[-h,h]^d} p_t(x+hu)$ 的上界。p_t 的 H¨older 正则性给出 $p_t^*(x)\leqslant p_t(x)+\frac{L}{(2d\,\ell)!}h^\beta$。因此，

$$\begin{aligned}&\mathbb{E}\int_{\{x:\,\hat{p}_t(x)\geqslant\rho_n\}}\left(\frac{|\hat{p}_t-p_t|^2}{\hat{p}_t^2}\right)^{k'} p_t\,\mathrm{d}x\\ &\leqslant \mathbb{E}\int_{\{x:\,\hat{p}_t(x)\geqslant\rho_n\}}\frac{1}{\hat{p}_t^{2k'-1}}|\hat{p}_t-p_t|^{2k'}\mathrm{d}x\\ &\leqslant \mathbb{E}\int_{\{x:\,\hat{p}_t(x)\geqslant\rho_n\}}\frac{1}{\hat{p}_t^{2k'-1}}\left(C_{B,0}^2h^{2\beta}+\frac{\|K\|_{L^2}^2p_t^*(x)}{nh^d}\right)^{k'-1}|\hat{p}_t-p_t|^2\mathrm{d}x\\ &\leqslant \left[\frac{1}{\rho_n^{2k'-1}}\left(C_{B,0}^2h^{2\beta}+\frac{\|K\|_{L^2}^2L(2d)^\ell h^\beta}{nh^d}\right)^{k'-1}+\frac{1}{\rho_n^{k'}}\left(\frac{\|K\|_{L^2}^2}{nh^d}\right)^{k'-1}\right]\mathbb{E}\int_{\{x:\,\hat{p}_t(x)\geqslant\rho_n\}}|\hat{p}_t-p_t|^2\mathrm{d}x\\ &\leqslant \left[\frac{1}{\rho_n^{2k'-1}}\left(C_{B,0}^2h^{2\beta}+\frac{\|K\|_{L^2}^2L(2d)^\ell h^\beta}{nh^d}\right)^{k'-1}+\frac{1}{\rho_n^{k'}}\left(\frac{\|K\|_{L^2}^2}{nh^d}\right)^{k'-1}\right]\left[(C_K^2+1)L^2h^{2\beta}+\frac{\|K\|_{L^2}^2}{nh^d}\right]\\ &\asymp h^\beta.\end{aligned}$$

结合 E_{trunc} 和 E_{est}的结果，总得分估计误差由两项主导：一是梯度估计误差，其尺度关系为 $h^{\beta-2}$；二是尾部截断误差，其尺度关系为 $h^{\beta(\gamma+1)k'(d+\gamma+1)}$：

$$\mathcal{R}\lesssim h^{\beta-2}+h^{\frac{\beta(\gamma+1)}{k'(d+\gamma+1)}}.$$

代入带宽缩放 $h\asymp n^{-\frac{1}{2\beta+d}}$，即可得到所述误差率：

$$\mathcal{R}\lesssim n^{-\frac{\beta-2}{2\beta+d}}+n^{-\frac{\beta(\gamma+1)}{k'(2\beta+d)(d+\gamma+1)}}.$$

□

备注。 我们通过比较控制这两种过程的特征时间尺度，说明另一种策略并非最优。

- **提前停止时间**（t_0）。在定理4中，采样误差的上界形式为

$$\mathrm{Err}(t) \lesssim \underbrace{\mathrm{Bias}(t)}_{\downarrow \text{ as } t\downarrow 0} + \underbrace{\mathrm{ScoreErr}(t)}_{\uparrow \text{ as } t\downarrow 0},$$

其中，提前停止偏差随时间递减：$\mathrm{Bias}(t) \asymp t^{\frac{\beta}{d+2(\gamma+1)}}$，而得分估计项则随时间递增：$\mathrm{ScoreErr}(t) \asymp t^{-\frac{d}{4(\gamma+1)(d+\gamma+1)}}$。我们将 t_0 定义为该上界的最小化点，即随着 t 减小所带来的边际偏差降低，恰好被得分估计引入的额外方差抵消之处。

- **切换时间**（$t_{\text{切换}}$）。对于另一种（混合）策略，解耦得分估计器的速率受定理29中两个分量中较慢者的限制。定义

$$\kappa_{\mathrm{grad}} := \frac{\beta-2}{2\beta+d}, \qquad \kappa_{\mathrm{tail}} := \frac{\beta(\gamma+1)}{k'(2\beta+d)(d+\gamma+1)}, \qquad \kappa_{\mathrm{dec}} := \min\{\kappa_{\mathrm{grad}}, \kappa_{\mathrm{tail}}\},$$

因此， $n^{-\kappa_{\mathrm{dec}}}$ 是解耦估计量所能达到的最佳逐点均方误差率（相差常数因子）。

最优混合规则在每个噪声水平 t下，选择逐点均方误差较小的估计量。因此，切换时间 t_{switch} 由两个均方误差界相当时的交叉点确定。令解耦速率 $n^{-\kappa_{\mathrm{dec}}}$ 等于标准估计器速率 $n^{-\kappa_1} t^{-(1+d\kappa_1/2)}$，其中 $\kappa_1 := \frac{\gamma+1}{d+\gamma+1}$，可得

$$t_{\mathrm{switch}} \asymp n^{-\frac{2(\kappa_1-\kappa_{\mathrm{dec}})}{2+d\kappa_1}}.$$

由于解耦估计量需要 $\beta > 2$，且通常满足 $\kappa_{\mathrm{dec}} < \kappa_1$，因此交叉点出现在远低于极小极大最优停止尺度的噪声水平处，也就是说，对于较大的 n，有 $t_{\mathrm{switch}} \ll t_0$。特别地，要达到 t_0，此时已经处于解耦估计量的逐点均方误差大于标准估计器的区域。因此，使用解耦估计量继续深入深度小噪声区域 $t \in [0, t_0]$，只会累积额外的方差，而几乎无法进一步降低偏差，因而无法改善总体速率。这说明为何在 t_0 处提前停止仍是极小极大最优的。

A.6 节证明4

在这部分中，我们给出得分估计的极小极大下界证明。

A.6.1 定理9的证明

选取一个固定的 $\psi \in C_c^\infty([-1/2,1/2]^d)$，满足 $\int_{[-1/2,1/2]^d} \psi = 0$ 和 $\nabla\psi(0) = v_0 \neq 0$，$L_\psi = \sup_{x\in[-1/2,1/2]^d} \|\nabla^2\psi\|$。令

$$h := \Lambda\sqrt{t}, \quad \Lambda \geqslant \frac{4L_\psi C_d}{\|\nabla\psi(0)\|},$$

并基于 ψ 和 h构造扰动，方法如下：

$$\omega(x) := \psi(x/h)\,.$$

那么， ω 的支撑集为 $[-h/2, h/2]^d$ 和 $\int_{[-h/2,h/2]^d} \omega = 0$。取

$$q_0 = C_\gamma(1 + \|x\|^2)^{-(d+\gamma+1)/2}, \quad q_b = q_0 + \epsilon \sum_{\mathbf{i}\in\mathcal{I}} b_{\mathbf{i}}\omega(x - x_{\mathbf{i}})\,.$$

那么， q_0， $q_b \in H_\gamma \cap \mathcal{P}_{\mathcal{S}}(L,\ \beta)$。令

$$f_0 = \varphi_t * q_0, \quad f_b = \varphi_t * q_b\,.$$

得分分离度。注意

$$\begin{aligned}
&\int_{\mathbb{R}^d} \|s_b(x) - s_{b'}(x)\|^2 f_0(x)\mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left\| \frac{\nabla f_b}{f_b} - \frac{\nabla f_{b'}}{f_{b'}} \right\|^2 f_0(x)\mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left\| \frac{\nabla f_b \cdot (f_{b'} - f_b) + f_b \cdot (\nabla f_b - \nabla f_{b'})}{f_b(x) f_{b'}(x)} \right\|^2 f_0(x)\mathrm{d}x\,.
\end{aligned}$$

记$\Delta f_{b,\ b'} = f_b - f_{b'}$，$\Delta\nabla f_{b,\ b'} = \nabla f_b - \nabla f_{b'}$，则有

$$\begin{aligned}
\nabla f_b \cdot \Delta f_{b',b} &= \nabla f_b \cdot \epsilon \sum_{\mathbf{i}\in\mathcal{I}} (b'_{\mathbf{i}} - b_{\mathbf{i}})(\varphi_t * \omega)(x - x_{\mathbf{i}}) \\
&= \epsilon \nabla f_0 \cdot \sum_{\mathbf{i}\in\mathcal{I}} (b'_{\mathbf{i}} - b_{\mathbf{i}})(\varphi_t * \omega)(x - x_{\mathbf{i}}) \\
&\quad + \epsilon^2 \left[\sum_{\mathbf{i}\in\mathcal{I}} b_{\mathbf{i}}(\varphi_t * \nabla\omega)(x - x_{\mathbf{i}})\right] \left[\sum_{\mathbf{i}\in\mathcal{I}} (b'_{\mathbf{i}} - b_{\mathbf{i}})(\varphi_t * \omega)(x - x_{\mathbf{i}})\right] \\
f_b \cdot \Delta\nabla f_{b,b'} &= \epsilon f_b \cdot \sum_{\mathbf{i}\in\mathcal{I}} (b_{\mathbf{i}} - b'_{\mathbf{i}})(\varphi_t * \nabla\omega)(x - x_{\mathbf{i}}) \\
&= \epsilon f_0 \cdot \sum_{\mathbf{i}\in\mathcal{I}} (b_{\mathbf{i}} - b'_{\mathbf{i}})(\varphi_t * \nabla\omega)(x - x_{\mathbf{i}}) \\
&\quad + \epsilon^2 \left[\sum_{\mathbf{i}\in\mathcal{I}} b_{\mathbf{i}}(\varphi_t * \omega)(x - x_{\mathbf{i}})\right] \left[\sum_{\mathbf{i}\in\mathcal{I}} (b_{\mathbf{i}} - b'_{\mathbf{i}})(\varphi_t * \nabla\omega)(x - x_{\mathbf{i}})\right]
\end{aligned}$$

将其代回前面的公式可得

$$\begin{aligned}
&\int_{\mathbb{R}^d} \|s_b(x) - s_{b'}(x)\|^2 f_0(x)\mathrm{d}x \\
&\geqslant \epsilon^2 \int_{[R,2R]^d} \left\| \frac{\sum_{\mathbf{i}\in\mathcal{I}} (b'_{\mathbf{i}} - b_{\mathbf{i}})[\nabla f_0 \cdot (\varphi_t * \omega)(x - x_{\mathbf{i}}) - f_0 \cdot (\varphi_t * \nabla\omega)(x - x_{\mathbf{i}})]}{f_0^2} \right\|^2 f_0(x)\mathrm{d}x \\
&= \epsilon^2 \sum_{\mathbf{i}\in\mathcal{I}} \int_{D_{\mathbf{i}}} \frac{1}{f_0(x)} \left\| (b_{\mathbf{i}} - b'_{\mathbf{i}}) \left[\frac{\nabla f_0}{f_0} \cdot (\varphi_t * \omega)(x - x_{\mathbf{i}}) - (\varphi_t * \nabla\omega)(x - x_{\mathbf{i}}) \right] + \xi_{\mathbf{i}}(x) \right\|^2 \mathrm{d}x\,,
\end{aligned}$$

其中

$$\xi_{\mathbf{i}}(x)=\sum_{\mathbf{j}\neq\mathbf{i}\in\mathcal{I}}(b_{\mathbf{j}}-b'_{\mathbf{j}})\left[\frac{\nabla f_0}{f_0}\cdot(\varphi_t*\omega)(x-x_{\mathbf{j}})-(\varphi_t*\nabla\omega)(x-x_{\mathbf{j}})\right].$$

注意，当 x 远离 $[-h/2,h/2]^d$时，我们可以推导出 $\varphi_t*\omega$的一个上界

$$\begin{aligned}|(\varphi_t*\omega)(x)|&=\left|\int_{\mathbb{R}^d}\varphi_t(x-y)\omega(y)\mathrm{d}y\right|\\&\leqslant\int_{[-h/2,h/2]^d}\varphi_t(x-y)|\omega(y)|\mathrm{d}y\\&\leqslant\sup_{y\in[-h/2,h/2]^d}\varphi_t(x-y)\cdot\|\omega\|_\infty\\&\leqslant\sup_{y\in[-h/2,h/2]^d}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\|x-y\|_\infty^2}{2t}\right)\cdot\|\omega\|_\infty.\end{aligned}$$

类似地，可以对 $\varphi_t*\nabla\omega$ 进行界定；于是当 $x\in x_{\mathbf{i}}+[-r,r]^d$时，其中 $r=\delta h$，且 δ 是一个与 t和 ϵ 无关的小常数，

$$\begin{aligned}\|\xi_{\mathbf{i}}(x)\|&\leqslant\sum_{\mathbf{j}\neq\mathbf{i}\in\mathcal{I}}\left\|\frac{\nabla f_0}{f_0}\cdot(\varphi_t*\omega)(x-x_{\mathbf{j}})\right\|+\|(\varphi_t*\nabla\omega)(x-x_{\mathbf{j}})\|\\&\leqslant\left(\frac{1}{R}\|\omega\|_\infty+\|\nabla\omega\|_\infty\right)\cdot\sum_{k=1}^{R}\sum_{\|x_{\mathbf{j}}-x_{\mathbf{i}}\|_\infty=kh}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{[(k-1/2)h-r]^2}{2t}\right)\\&=\left(\frac{1}{R}\|\omega\|_\infty+\|\nabla\omega\|_\infty\right)\cdot\sum_{k=1}^{R}[(2k+1)^d-(2k-1)^d]\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{[(k-1/2)h-r]^2}{2t}\right)\\&\leqslant C(\delta,d)\left(\frac{1}{R}\|\omega\|_\infty+\|\nabla\omega\|_\infty\right)\cdot t^{d/2}h^{-d}\\&=C(\delta,d)\left(\frac{1}{R}\|\psi\|_\infty+\frac{1}{\Lambda\sqrt{t}}\|\nabla\psi\|_\infty\right)\cdot\Lambda^{-d}.\end{aligned}$$

我们还注意到

$$\begin{aligned}&\left\|\frac{\nabla f_0}{f_0}\cdot(\varphi_t*\omega)(x-x_{\mathbf{i}})-(\varphi_t*\nabla\omega)(x-x_{\mathbf{i}})\right\|\\\geqslant&\|(\varphi_t*\nabla\omega)(x-x_{\mathbf{i}})\|-\left\|\frac{\nabla f_0}{f_0}(\varphi_t*\omega)(x-x_{\mathbf{i}})\right\|\end{aligned}$$

我们分别处理此公式右侧的两项。令 $\delta<\|v_0\|/(4L_\psi)$，由此可得

$$\|\nabla\omega(u)-\nabla\omega(0)\|\leqslant\sup_{x\in[-h/2,h/2]^d}\|\nabla^2\omega(x)\|\cdot\|u\|\leqslant L_\psi r/h^2<\frac{\|v_0\|}{4h},$$

那么，对于 $\|\nabla\omega(u)\|>\frac{3}{4}\|\nabla\omega(0)\|$ $x\in x_{\mathbf{i}}+[-r,r]^d$。记

$$g_t(u):=(\varphi_t*\nabla\omega)(u)=\int_{\mathbb{R}^d}\varphi_t(u-y)\nabla\omega(y)\mathrm{d}y.$$

有

$$\begin{aligned}\|g_t(u)-\nabla\omega(u)\| &= \left\|\int_{\mathbb{R}^d}\varphi_t(u-y)(\nabla\omega(y)-\nabla\omega(u))\mathrm{d}y\right\| \\ &\leqslant \int_{\mathbb{R}^d}\varphi_t(u-y)\frac{L_\psi}{h^2}\|u-y\|\mathrm{d}y \\ &= \frac{L_\psi\sqrt{t}}{h^2}\mathbb{E}[\|Z\|] \\ &\leqslant \frac{L_\psi\sqrt{dt}}{h^2}\,,\end{aligned}$$

其中 $Z\sim\mathcal{N}(0,\ I_d)$，最后一个不等式由詹森不等式得出。因此，我们有

$$\|g_t(u)\|\geqslant\|\nabla\omega(u)\|-\|g_t(u)-\nabla\omega(u)\|\geqslant\frac{3\|v_0\|}{4h}-\frac{L_\psi\sqrt{d}}{\Lambda h}\geqslant\frac{\|v_0\|}{2h}$$

and

$$\left\|\frac{\nabla f_0}{f_0}\cdot(\varphi_t*\omega)(x-x_{\mathbf{i}})\right\|\leqslant\frac{1}{R}\|\omega\|_\infty=\frac{1}{R}\|\psi\|_\infty\,.$$

令 $R>4h\|\psi\|_\infty/\|v_0\|$，则可得

$$\left\|\frac{\nabla f_0}{f_0}\cdot(\varphi_t*\omega)(x-x_{\mathbf{i}})-(\varphi_t*\nabla\omega)(x-x_{\mathbf{i}})\right\|\geqslant\frac{\|v_0\|}{4h}\,.$$

汇集各部分可得

$$\begin{aligned}&\int_{\mathbb{R}^d}\|s_b(x)-s_{b'}(x)\|^2f_0(x)\mathrm{d}x \\ &\geqslant\epsilon^2\sum_{\mathbf{i}\in\mathcal{I}}R^{d+\gamma+1}\int_{x_{\mathbf{i}}+[-r,r]^d}\left\|(b_{\mathbf{i}}-b'_{\mathbf{i}})\left[\frac{\nabla f_0}{f_0}\cdot(\varphi_t*\omega)(x-x_{\mathbf{i}})-(\varphi_t*\nabla\omega)(x-x_{\mathbf{i}})\right]+\xi_{\mathbf{i}}(x)\right\|^2\mathrm{d}x \\ &\geqslant\epsilon^2\sum_{\mathbf{i}\in\mathcal{I}}R^{d+\gamma+1}(2r)^d\mathbb{1}\{b_{\mathbf{i}}\neq b'_{\mathbf{i}}\}\left(\frac{\|v_0\|}{4\Lambda\sqrt{t}}-C(r,d)\left(\frac{1}{R}\|\psi\|_\infty+\frac{1}{\Lambda\sqrt{t}}\|\nabla\psi\|_\infty\right)\cdot\Lambda^{-d}\right)^2 \\ &\gtrsim\epsilon^2\sum_{\mathbf{i}\in\mathcal{I}}R^{d+\gamma+1}t^{d/2-1}\mathbb{1}\{b_{\mathbf{i}}\neq b'_{\mathbf{i}}\}\,.\end{aligned}$$

因此，我们有

$$\int_{\mathbb{R}^d}\|s_b(x)-s_{b'}(x)\|^2f_0(x)\mathrm{d}x\gtrsim\epsilon^2R^{d+\gamma+1}t^{d/2-1}d_H(b,b')\,.$$

其中 $d_H(b,\ b')=\sum_{\mathbf{i}\in\mathcal{I}}\mathbf{1}\{b_{\mathbf{i}}\neq b'_{\mathbf{i}}\}$ 是汉明距离。由吉尔伯特-瓦尔沙莫夫界可知，存在一个大小呈指数级增长的打包集 $\mathcal{B}'\subset\mathcal{B}$，其最小汉明距离与维度成线性关系，即 $|\mathcal{B}'|\geqslant\exp(c_0(R/h)^d)$ 且 $\min_{b\neq b'\in\mathcal{B}'}d_H(b,\ b')\geqslant c_0(R/h)^d$，其中某个普适常数为 c_0。因此，我们得到

$$\|s_b-s_{b'}\|^2_{L^2(f_0)}\gtrsim\epsilon^2R^{2d+\gamma+1}t^{-1}\,.$$

界定 KL 半径。

对于每个 $b \in \mathcal{B}$, f_b 与 f_0 之间的 KL散度满足

$$\begin{aligned}
\mathrm{KL}(f_b \| f_0) &\leqslant \chi^2(f_b \| f_0) \\
&= \int_{\mathbb{R}^d} \frac{(f_b(x) - f_0(x))^2}{f_0(x)} \mathrm{d}x \\
&= \sum_{\mathbf{i} \in \mathcal{I}} \mathbf{1}\{b_{\mathbf{i}} = 1\} \int_{D_{\mathbf{i}}} \frac{\left[\epsilon \sum_{\mathbf{j} \in \mathcal{I}} (\varphi_t * \omega)(x - x_{\mathbf{j}})\right]^2}{f_0(x)} \mathrm{d}x \\
&\leqslant \sum_{\mathbf{i} \in \mathcal{I}} \epsilon^2 R^{d+\gamma+1} \int_{D_{\mathbf{i}}} \left[(\varphi_t * \omega)(x - x_{\mathbf{i}}) + \sum_{\mathbf{j} \neq \mathbf{i} \in \mathcal{I}} (\varphi_t * \omega)(x - x_{\mathbf{j}}) \right]^2 \mathrm{d}x \,.
\end{aligned}$$

同样，有

$$|(\varphi_t * \omega)(x - x_{\mathbf{i}})| \leqslant \|\omega\|_\infty = \|\psi\|_\infty$$

and

$$\begin{aligned}
\left| \sum_{\mathbf{j} \neq \mathbf{i} \in \mathcal{I}} (\varphi_t * \omega)(x - x_{\mathbf{j}}) \right| &\leqslant \sum_{k=1}^{R} \sum_{\|x_{\mathbf{j}} - x_{\mathbf{i}}\|_\infty = kh} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{|(k-1)h|^2}{2t} \right) \\
&\leqslant \sum_{k=1}^{R} [(2k+1)^d - (2k-1)^d] \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{|(k-1)h|^2}{2t} \right) \\
&\leqslant \frac{2d \cdot 3^{d-1}}{\sqrt{2\pi}} \left[1 + 2^{\frac{3d}{2}-2} \Gamma\left(\frac{d}{2}\right) t^{d/2} h^{-d} \right] \\
&\leqslant \frac{2d \cdot 3^{d-1}}{\sqrt{2\pi}} \left[1 + 2^{\frac{3d}{2}-2} \Gamma\left(\frac{d}{2}\right) \Lambda^{-d} \right] .
\end{aligned}$$

因此,

$$\mathrm{KL}(f_b \| f_0) \lesssim \epsilon^2 R^{2d+\gamma+1}$$

Fano方法。 由于观测值是独立同分布的，我们得到 $\max_{b \in \mathcal{B}} \mathrm{KL}(f_b^{\otimes n} \| f_0^{\otimes n}) \lesssim n\epsilon^2 R^{2d+\gamma+1} t^{-d/2}$。令 $R = \epsilon^{-\rho}$，通过选择 $\rho = \frac{1-\delta}{d+\gamma+1}$ 和 $\epsilon = cn^{-1}t^{-d/2}$，对于足够小的常数 c，我们得到

$$\max_{b \in \mathcal{B}} \mathrm{KL}(f_b^{\otimes n} \| f_0^{\otimes n}) \lesssim n\epsilon^2 R^{2d+\gamma+1} \leqslant cR^d t^{-d/2} \leqslant c \log(|\mathcal{B}'|).$$

应用Fano不等式可得

$$\begin{aligned}
\inf_{\hat{s}} \sup_{b \in \mathcal{B}'} \mathbb{E}_{f_b} \|\hat{s} - s_b\|_{L^2(f_0)}^2 &\gtrsim \min_{b \neq b' \in \mathcal{B}'} \|s_b - s_b'\|_{L^2(f_0)}^2 \gtrsim \epsilon^2 \epsilon^{-\frac{(2d+\gamma+1)(1-\delta)}{d+\gamma+1}} t^{-1} \\
&= \epsilon^{\frac{(1+\gamma)(1-\delta)}{1+\gamma+d}} t^{-1} \\
&= n^{-\frac{(1+\gamma)(1-\delta)}{1+\gamma+d}} t^{-1-\frac{(1+\gamma)(1-\delta)}{1+\gamma+d} \cdot \frac{d}{2}} .
\end{aligned}$$

注意,

$$\|\hat{s} - s_b\|_{L^2(f_b)}^2 - \|\hat{s} - s_b\|_{L^2(f_0)}^2$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^d} \|\hat{s}(x) - s_b(x)\|^2 (f_b(x) - f_0(x))\,\mathrm{d}x \\
&\leqslant \int_{[R,2R]^d} \|\hat{s}(x) - s_b(x)\|^2 \epsilon \|\omega\|_\infty \,\mathrm{d}x \\
&\asymp \epsilon^\delta \int_{[R,2R]^d} \|\hat{s}(x) - s_b(x)\|^2 f_0(x)\,\mathrm{d}x \\
&\asymp n^{-\delta} \|\hat{s} - s_b\|^2_{L^2(f_0)}\,.
\end{aligned}
$$

由此可得,

$$
\inf_{\hat{s}} \sup_{b \in \mathcal{B}'} \mathbb{E}_{f_b} \|\hat{s} - s_b\|^2_{L^2(f_b)} \gtrsim n^{-\frac{\gamma+1}{d+\gamma+1} - \delta'} t^{-1-\frac{\gamma+1}{d+\gamma+1}+\delta'}
$$

对于任意小的 δ'。

为确保扰动密度 q_b 构成有效的最坏情形构造，我们必须验证三个约束：重尾衰减、索伯列夫正则性，以及局部扰动的几何一致性。回顾我们的参数选择： $h \asymp t^{1/2}$、 $\epsilon \asymp n^{-1}t^{-d/2}$和 $R \asymp \epsilon^{-\frac{1}{d+\gamma+1}}$。

重尾性（$q_b \in \mathcal{H}_\gamma$）： 扰动密度必须保持非负，并满足多项式尾部衰减。这要求扰动被基准密度所控制:

$$
\epsilon \|\omega\|_\infty \lesssim \inf_{x \in [R,2R]^d} q_0(x) \asymp R^{-(d+\gamma+1)}.
$$

[1]代入 $R \asymp \epsilon^{-\frac{1}{d+\gamma+1}}$后，上界变为 ϵ，这对t不施加额外约束。

索伯列夫正则性（$q_b \in \mathcal{P}_\mathcal{S}(\beta,\ L)$）： 我们假设密度属于阶数为$\beta$的索伯列夫球，因此要求 $\|q_b - q_0\|_{H^\beta} \lesssim L$。扰动 $\Delta(x) = \epsilon \sum_{\mathbf{i}} b_{\mathbf{i}} \omega(x - x_{\mathbf{i}})$由约$N \asymp (R/h)^d$ 个互不相交的小波组成。其索伯列夫半范数的平方按以下方式缩放

$$
\|D^\beta \Delta\|^2_{L^2} \asymp \sum_{\mathbf{i} \in \mathcal{I}} \int |D^\beta(\epsilon \omega(x - x_{\mathbf{i}}))|^2 \mathrm{d}x \asymp \left(\frac{R}{h}\right)^d \cdot \epsilon^2 h^{-2\beta} \cdot h^d = R^d \epsilon^2 h^{-2\beta}.
$$

条件 $\|D^\beta \Delta\|_{L^2} \lesssim 1$ 意味着 $R^{d/2}\epsilon t^{-\beta/2} \lesssim 1$ （因为 $h \asymp t^{1/2}$）。代入 $\epsilon \asymp n^{-1}t^{-d/2}$和 $R \asymp (nt^{d/2})^{\frac{1}{d+\gamma+1}}$，可得关于 t的下界

$$
R^{d/2}(n^{-1}t^{-d/2})t^{-\beta/2} \lesssim 1\,,
$$

这意味着

$$
(nt^{d/2})^{\frac{d}{2(d+\gamma+1)}} n^{-1} t^{-(d+\beta)/2} \lesssim 1.
$$

[2]整理各项可得 $t \gtrsim n^{(\cdots)/(d+\gamma+1)}$。

几何一致性 ($h \lesssim R$)： 小波支撑尺度 h 必须小于尾部位置 R。这要求 $t^{1/2} \lesssim (nt^{d/2})^{\frac{1}{d+\gamma+1}}$。整理后得到 $t^{\gamma+1} \lesssim n^2$；在 $n \to \infty$ 且 $t \to 0$的相关范围内，这显然成立。

A.6.2 定理10的证明

指数尾部情形下最不利假设的构造总体上遵循与多项式情形相同的方式。然而，一个关键差异使得分析大为

简单：**在指数衰减情形下，最优速率由主体区域而非尾部决定**。

在多项式尾部情形中，极小极大速率由扩展尾部[1]区域中的样本稀缺所决定，这迫使我们必须在较大的半径 $R_n \asymp n^{\,d+\gamma+1}$ 处放置扰动。对于指数衰减目标，有效支撑集仅按对数速度增长，因此只需在一个固定紧集（或增长极其缓慢的区域）上引入扰动，即可获得参数化速率 n^{-1} （扩散时间依赖性除外）。

假设构造。 令 $q_0(x) \propto \exp(-\|x\|^\gamma)$ 为满足假设 1 的基准密度。与多项式构造不同，我们选择将扰动网格 $\{x_{\mathbf{i}}\}_{\mathbf{i}\in\mathcal{I}}$ 置于单位立方体 $[0,1]^d$ (或任意 q_0-质量有正下界的区域）中。定义

$$q_b(x) = q_0(x) + \epsilon \sum_{\mathbf{i}\in\mathcal{I}} b_{\mathbf{i}}\omega(x - x_{\mathbf{i}}),$$

其中，小波 ω 与定理 9 的证明中相同，网格间距设为$h = \Lambda\sqrt{t}$。令

$$f_0 = \varphi_t * q_0, \quad f_b = \varphi_t * q_b\,.$$

尺度关系。 由于扰动的支撑位于一个有界区域内，且 $q_0(x) \asymp 1$，因此在得分分离度和 KL 散度界中都会出现的权重 $1/f_0(x)$ 不会引入任何额外的多项式因子（与多项式尾部情形相反）。关键的尺度关系变为

1. **得分函数分离度**。 沿用多项式情形中的方法，并利用扰动区域上的 $f_0(x) \asymp 1$ ，我们得到

$$\|s_b - s_{b'}\|^2_{L^2(f_0)} \gtrsim \epsilon^2\, t^{d/2-1}\, d_H(b,b') \gtrsim \epsilon^2\, t^{-1}.$$

   这里，来自支撑体积的因子 $t^{d/2}$ 与网格密度 h^{-d} 相消，主导项仅剩梯度尺度 $\|\nabla\omega\|^2 \propto h^{-2} \asymp t^{-1}$。

2. **KL 散度**。 同样，KL 界不再需要支付尾部惩罚。利用 a χ^2 型控制，并注意到网格点的数量为 $M \asymp (1/h)^d \asymp t^{-d/2}$，我们有

$$\mathrm{KL}(f_b\|f_0) \lesssim \sum_{\mathbf{i}\in\mathcal{I}} \int \frac{(\epsilon\varphi_t * \omega)^2}{f_0} \asymp M\epsilon^2 t^d \asymp \epsilon^2 t^{d/2}.$$

   这里我们利用了单个网格单元的贡献为 t^d 量级这一事实。

Fano论证。令 $\mathcal{B}$ 为一个满足 $\log|\mathcal{B}| \asymp M \asymp t^{-d/2}$ 的打包集。只要满足以下条件，Fano不等式便适用

$$\max_{b\in\mathcal{B}} \mathrm{KL}(f_b^{\otimes n}\|f_0^{\otimes n}) \lesssim \log|\mathcal{B}|.$$

利用上述界，这可简化为

$$n\,\epsilon^2\, t^{d/2} \lesssim t^{-d/2}\,,$$

这意味着

$$\epsilon^2 \lesssim \frac{1}{n}\, t^{-d}.$$

为匹配我们上界中的时间依赖性，我们在局部（固定-t）情形下进行分析，并选择

$$\epsilon \asymp n^{-1/2}\, t^{-d/4},$$

这确保了 $n\epsilon^2 t^{d/2} \lesssim 1$ ，并且足以用于局部 Fano 构造。将此选择代入分离界可得

$$\text{Minimax risk} \gtrsim \epsilon^2\, t^{-1} \asymp n^{-1}\, t^{-(1+d/2)}.$$

约束验证。 接下来需要检查该构造定义了有效假设。

- **正性**。 我们要求在扰动区域上 $\epsilon\|\omega\|_\infty \lesssim \inf q_0$ 。由于定义域有界，$\inf q_0 \asymp 1$，当 n 足够大时，该条件对 $\epsilon \asymp n^{-1/2}t^{-d/4}$ 成立。
- **索伯列夫正则性**。 我们施加 $\|D^\beta \Delta\|_{L^2} \lesssim 1$。结合 $\epsilon \asymp n^{-1/2}t^{-d/4}$ 和 $h \asymp t^{1/2}$，可得

  $$n^{-1/2}t^{-d/4} \cdot t^{-\beta/2} \lesssim 1\,,$$

  这等价于

  $$t \gtrsim n^{-\frac{2}{d+2\beta}}\,.$$

  这与定理 8 中允许的时间范围一致。

结论。 由此得到的下界为 $n^{-1}t^{-(1+d/2)}$，除对数因子外，与定理5中的上界相匹配。这表明，在指数尾条件下，统计难度主要由参数尺度 n^{-1} 以及扩散核的病态性 $t^{-(1+d/2)}$决定，而不会因尾部稀疏性而额外增加惩罚。