利用扩散过程实现高精度、无维数依赖的采样

Khashayar Gatmiry SitanChen AdilSalim 加州大学伯克利分校
哈佛大学

arXiv:2601.10708v1 [cs.LG] 15 Jan 2026

摘要

扩散模型在从丰富的多模态分布中采样方面取得了显著的实证成果。其推理过程依赖于对某个微分方程进行数值求解。该微分方程无法以闭式求解，而通过离散化求解通常需要许多小步迭代，才能生成高质量样本。更准确地说，先前研究表明，扩散模型离散化方法的迭代复杂度关于环境维数和精度倒数呈多项式增长 $1/\varepsilon$。在本研究中，我们提出一种新的扩散模型求解器，它巧妙地结合了低次数近似与配点法 [LSV18]，，并证明其迭代复杂度呈多对数级增长，关于 $1/\varepsilon$，从而首次为一种仅使用对数据分布得分函数（的近似）访问的基于扩散的采样器提供了“高精度”保证。此外，我们的界不显式依赖环境维数；更准确地说，维数仅通过目标分布支撑集的有效半径影响求解器的复杂度。

目录

1 引言 1
1.1 我们的贡献 .11.2 相关工作 .21.3 路线图 .3

2 预备知识与我们的算法 4
2.1 扩散模型 .42.2 一种截然不同的离散化观点 .52.3 配点法 .52.4 我们的算法 .6

3 证明概述 7
3.1 沿真实反向过程的低次数近似 .73.2 沿算法的低次数近似 .83.3 采样器的收敛性 .10

4 展望
11A 更多相关工作 16B 符号说明 16C 导数计算 17C.1 引理 3.1 的证明 .18C.2 引理 C.1 的证明. 19C.3 引理 3.2 的证明 .21C.4 引理 3.3 的证明 .22**D 引理 3.4 的证明 23E 定理 3.5 的证明 24F 命题 3.6 的证明 30**F.1 预备知识.30**G 定理 3.7 的证明 34H 推论 3.9 的证明 37**H.1 欠阻尼朗之万蒙特卡罗 . 37H.2 完成证明 .38**I U**

PP对……运动的估计 **P概率** **y 流 ODE 38**

1 引言

扩散模型 [SDWMG15，SE19，HJA20，DN21，SSDK+21，SDME21，VKK21] 是图像生成及其他模态领域的主流范式。它们通过数值模拟由根据训练数据学习得到的漂移项所驱动的某个微分方程的反向过程，从而从高维分布中采样。反向过程旨在撤销某种噪声过程；通过足够精确地模拟反向过程，我们便能从纯噪声中生成该分布的新样本。

这种方法在实践中的成功，近年来推动了大量理论研究，旨在理解扩散模型能够轻松从丰富的多模态分布中采样的机制。这些研究借鉴了对数凹采样领域的丰富文献中的工具，并得出了一个惊人的结论：只要能够沿噪声过程获取足够准确的分布得分函数估计，扩散模型就能在高维空间中高效地从任意分布中采样，包括高度非对数凹的分布（有关这方面文献的概述，请参见第 1.2 节）。

最早的此类结果表明，对于任何二阶矩有界的光滑分布，只要分数估计 ε 足够准确，就能在 $(d\,1/\varepsilon)$ 次迭代内，将全变差距离上的采样误差控制在 [LLT23、CCL+23、CLL23、BDBDD24、CDS23] 所示的范围内。这些结果研究的是在离散时间模拟反向过程的采样器。离散化会引入偏差，因此必须将离散化步长设得足够小，即为 $\mathrm{poly}(\varepsilon, 1/d)$，以确保采样器的轨迹与连续时间反向过程的轨迹足够接近。

尽管近期研究显著改进了这些离散化界（见第1.2节），但其中仍有一个方面特别缺乏理解，那就是对 ε的依赖。在对数凹采样文献中，沿这一维度已形成清晰的分类：(1)“低精度”方法，如朗之万蒙特卡罗，其迭代复杂度随 $1/\varepsilon$呈多项式增长；(2)“高精度”方法会校正离散化偏差，例如通过 Metropolis 调整进行校正，其迭代复杂度关于 呈多对数增长： $1/\varepsilon$。

对于扩散模型，第二类保证仍鲜有研究。因此，我们提出以下问题：

扩散建模是否存在迭代复杂度为 O(多对数$1/\varepsilon$) 的采样器？

1.1 我们的贡献

在此，我们对此给出肯定回答。我们将聚焦于以下这类目标分布 分布 q:

假设 1 (有界加噪声)。 令 $R, \sigma > 0$。存在一个分布 q_{pre}，其支撑集位于半径为 R、以原点为中心的球内，且该球位于 $\mathbb{R}^d$中，使得 $q = q_{\mathsf{pre}} \star \mathcal{N}(0, \sigma^2 I)$。

换言之，我们考虑从一个分布中分布 q 进行采样，该分布是一个紧支撑分布与高斯噪声的卷积。我们的界将随R/σ呈多项式增长。满足假设1的一个自然例子是各向同性高斯混合；更一般地，正如 [CCL+23]所指出的，满足假设1 的分布自然地刻画了扩散模型在实践中试图通过提前停止进行采样的对象。σ 的正性也使得我们能够证明在全变差距离等强基于散度的度量下的收敛保证。

在本文中，我们考虑通过近似模拟反向过程，从分布 q 进行采样。我们首先证明一个关键的结构性质：沿反向过程的得分函数的高阶时间导数有界，这意味着得分函数可以在每个点处用时间的低次多项式近似。该表述较为技术性，因此我们将其留至补充材料。

利用这一结果，我们展示如何调整 [LSV18] 的一种专用常微分方程求解器，以模拟反向过程中的步进，并论证我们的模拟在所有时间上都与

反向过程的轨迹保持接近，见算法2。因此，我们证明算法2 能够高效地从分布 q中采样，从而得到以下主要保证：

定理 1.1（非正式表述，参见推论 3.9）。 设 ε，R， $\sigma > 0$。假设分布 q 满足参数为 $R1$ 和， σ的假设 1。若可获得利普希茨得分估计，且其估计误差有界于 L^2并具有次指数尾部，则存在一种基于扩散的算法（参见算法2），可输出来自分布 ε的样本，使其与 q 的全变差（TV）距离在 $\tilde{O}((R\cdot \mathrm{polylog}\ 1/\varepsilon)$次迭代内达到。

据我们所知，这是首个仅假设能够获得足够精确的得分函数，便实现对 $1/\varepsilon$ 的多对数依赖（而非多项式依赖）的基于扩散的采样器。尽管与先前工作相比，我们的结果对得分误差分布提出了更强的假设，但值得指出的是，即使假设得分估计完全准确，先前工作的分析仍会对 $1/\varepsilon$ 产生多项式依赖。

我们结果的另一项优势是，算法的迭代复杂度与环境维数 d 无关，而取决于假设 1 中的半径 R。虽然一般而言 R 可能随 d 增长，但高斯混合情形提供了一个自然的场景，其中 R 可能远小于 d。例如，在分布学习理论中，最具挑战性的情形恰恰是分量中心之间的距离为 $\Theta(\sqrt{\log k})$ [HL18、KSS18、DKS18、LL22]。这对应于 $R = \Theta(\sqrt{\log k})$ 且 $\sigma = 1$ 的情形；对此，我们的采样器的迭代复杂度仅随 k 和 $1/\varepsilon$ 呈多对数级增长，而现有的基于扩散的方法在此例中会随 d 和 $1/\varepsilon$ 呈多项式级增长（关于一项近期在高斯混合情形下实现维数无关速率的结果 [LCW25]，参见第 1.2 节）。

1.2 相关工作

我们在附录 A中回顾了近期关于扩散模型收敛界的大量研究，此处着重介绍与我们的结果最相关的研究方向。

对 ε 的依赖。 首先，实践中高阶求解器在扩散采样中的普及 [LZB+22、LZB+25、ZBR+23]与此相呼应，许多研究借鉴高阶数值求解器的思想，试图改善基于扩散的采样对 ε 的依赖，例如 [LHE+24、WCW24、LHE+24]，这些工作在相对最小的假设下，实现了固定的多项式加速，改善了对 ε 的依赖。

Huang 等人 [HHL25a, HHL25b] 和 Li 等人 [LZWC25] 使用高阶求解器实现了任意多项式加速，即速率按 $O(d^{1+O(1/K)}/\varepsilon^{1/K})$ 缩放，其中常数 K 可以任意大。但我们注意到，尽管这似乎表明可以取 $K = \tilde{\Theta}(\log(1/\varepsilon))$，从而实现与我们工作类似的多对数级缩放，但这些研究的分析实际上假设迭代次数至少随 K 指数增长（例如参见 [LZWC25]中的讨论），因此它们获得的加速至多是多项式级，而非像我们工作中那样的指数级。此外，这些研究的假设与我们的假设无法直接比较。例如， [HHL25a, HHL25b] 假设得分估计至少具有二阶光滑性； [LZWC25] 不要求得分估计具有任何光滑性，但假设其雅可比矩阵接近真实得分的雅可比矩阵。相比之下，我们需要假设得分估计具有一阶光滑性（这有其合理性，因为真实得分是利普希茨的），并且得分误差为次指数分布。最后，我们指出，这三项研究与我们的工作都假设数据分布具有紧支撑。

据我们所知，唯一真正实现指数级改进的 ε 依赖关系的先前工作是 [HZD+24]。不过，该工作采用的访问模型比本领域通常研究的更强，因为除了得分函数之外，它还假设能够近似查询任意点之间的对数密度比，以实现 Metropolis-Hastings 滤波器。我们还注意到独立的

以及同期的 [Wai25] 工作将反向过程分解为从一系列强对数凹后验中采样，并指出，利用现成的强对数凹采样高精度保证，可以使 ε 基于扩散的采样对 [HZD+24]呈对数依赖。不过，这些现成的算法不仅需要知道得分函数，还需要知道数据分布与噪声卷积后的对数密度，这与 [HZD+24]类似。

无维数依赖的界。关于我们对 R/σ 而非 d 的依赖，已有一系列相关工作[LY24a、LHC25、HWC24、PAD24] 研究从内在维数较低的分布中进行采样。他们表明，对于任意内在维数较低的分布——其内在维数通过支撑集的覆盖数 k 来量化——去噪扩散概率模型的迭代复杂度为 $O(k^4/\varepsilon^2)$。尽管这为基于扩散的采样器提供了另一个速率能够适应分布底层几何结构的例子，但它与我们的结果无法直接比较：支撑集可能是高维的，但半径有界，反之亦然。另一方面， [LCW25] 考虑了各向同性高斯混合这一特殊情形，并采用与本文完全不同的技术证明，扩散模型可以在 $T = \tilde{O}(\mathrm{polylog}(k)/\varepsilon)$ 次迭代内从这些分布中采样，前提是高斯分布的均值范数至多为 T^C，其中 C 为任意大的绝对常数。我们指出，该结果与我们的结果无法直接比较。我们的结果对 $1/\varepsilon$ 具有多对数依赖，并且保证适用于远超各向同性高斯混合这一特殊情形的广泛分布类别。另一方面，对于高斯混合， [LCW25] 的结果在其他参数上的标度更优：我们对混合中任意中心的最大范数具有多项式依赖，而他们对半径的依赖可以任意小。

扩散过程的配置法。 最后，我们注意到，配点法（见第 2.3 节）已在扩散过程的背景下得到研究，但主要是作为一种 并行化 采样器步骤的方法 [AHL+23，GCC24，以及 CRYR24]，，但并未使用低次多项式近似。这些工作实现了多对数级的 轮复杂度，但所需的 总工作量 仍然是目标精度的多项式级。相比之下，我们工作的关键技术要素是在时间上建立得分函数的低次可逼近性，这使我们能够证明，通过 [LSV18] 在概率流常微分方程的一个较短时间窗口上使用低次多项式实现的配点法，以 指数 速率收敛，即使在顺序设置下也是如此。

高精度对数凹采样。 关于高精度采样器的文献传统上以对数凹分布，以及更一般地满足函数不等式的分布为中心，内容极其丰富，无法在此一一详述；有关详细综述，请参阅 [Che23] 的第7章。我们在附录A 中简要概述。我们的技术并未借鉴这部分文献，而是基于 [LSV18]，的工作；该工作提出了我们所采用的、通过低次近似实现配点法的一般框架。受与逻辑回归相关的采样问题启发，他们将该方法用于为形如 分布 $q \propto e^{-f}$ 的概率密度构造高精度采样器，其中 $f(x) = \sum_i \phi_i(\langle a_i x\rangle) + \lambda\|x\|^2$。

1.3 路线图

在第 2 节中，我们介绍技术预备知识并描述我们的采样器。本节还将说明该采样器如何实现定理 1.1 所保证的高精度。在第 3 节中，我们概述主要结构性结果的证明，该结果表明反向过程的漂移项可由时间的低次多项式良好近似；同时也介绍完成定理 1.1 证明的主要步骤。完整的证明细节见附录。

2 预备知识与我们的算法

符号说明。令 γ^d 表示 d 维标准高斯分布。给定区间 $I \subseteq \mathbb{R}^d_{\geq 0}$，令 $\mathcal{C}(I, \mathbb{R}^d)$ 表示从 $I \to \mathbb{R}^d$ 到连续映射的空间。对于定义在 $\mathbb{R}^d$ 上的任意分布 p，$L^2(p)$ 表示 $\mathbb{R}^d$ 中平方可积随机向量的空间。回顾全变差距离 $TV(P, Q) = \frac{1}{2}\int_{\mathbb{R}^d} |p - q| dx$（其中 p、q 为概率密度）以及二阶瓦瑟斯坦距离 $W_2^2(P, Q) = \inf_{\pi \in \Pi(P,Q)} \int \|x - y\|^2 d\pi(x, y)$ 的定义，其中 $\Pi(P, Q)$ 是耦合的集合。最后，对于任意向量 $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$，$\|x\|_\infty = \max_i |x_i|$；若 x 是随机向量，则 $\|x\|_{p,\infty} = \max_i (\mathbb{E}|x_i|^p)^{1/p}$。

2.1 扩散模型

这里我们回顾扩散模型的基础知识；关于更详细的综述，请参阅 [CMFW24] 和 [NBZA24]。

扩散模型由两个主要部分构成：前向过程和反向过程。前向过程是由如下形式的随机微分方程驱动的噪声过程

$$dx_t = -x_t\,dt + \sqrt{2}\,dB_t, \qquad x_0 \sim q_{\text{pre}},$$

其中 $(B_t)_{t\geq 0}$ 是 $\mathbb{R}^d$ 上的标准布朗运动。给定 x_0，时间 t 处的过程服从 $e^{-t}x_0 + \sqrt{1-e^{-2t}}g$，其中 $g \sim \gamma^d$；将该边缘分布记为 $\overrightarrow{p}_t$，其中 $\overrightarrow{p}_0 = q_{\text{pre}}$。

我们将前向过程运行至时间 T。定义

$$\sigma_t = \sqrt{1 - e^{-2(T-t)}}. \tag{1}$$

该反向过程 旨在消除这一噪声过程，即将 $\overrightarrow{p}_T$ 变换为 $\overrightarrow{p}_0$。为方便符号说明，我们将 $\overrightarrow{p}_{T-t}$ 记为 q_t，使得 $q_T = q_{\text{pre}}$。这种反向过程的一种形式由概率流常微分方程给出，其定义如下：

$$dy_t = (y_t + \nabla \ln q_t(y_t))\,dt, \tag{2}$$

其中，$\nabla \ln q_t$ 称为得分函数。其关键性质是：若 $y_0 \sim q_0$，则 $y_T \sim q_T = q_{\text{pre}}$。实际应用中，使用得分估计值 $s_t \approx \nabla \ln q_t$ 代替实际的得分函数来运行该过程；在理论文献中，通常假设二者在 $L_2(q_t)$ 意义下足够接近。在本文中，我们将采用一个稍强的假设：

假设 2（次指数得分误差）。存在某个参数 $\varepsilon_{\text{err}} > 0$，使得得分估计值 s_t：$\mathbb{R}^d \to \mathbb{R}^d$ 产生的误差的次指数范数至多为 ε_{err}，对所有 $0 \leq t \leq T$ 均成立。也就是说，对所有 $y \geq 0$，$\mathbb{P}_t[\|s_t - \nabla \ln q_t\| \geq z] \leq 2\exp(-z/\varepsilon_{\text{err}})$。

我们将能否将这一条件放宽为更标准的 L^2 精度得分估计留作开放问题，但值得注意的是，即使在完美得分估计的假设下，此前也尚不清楚如何实现高精度保证。

此外，不是在 $y_0 \sim q_0$ 处初始化常微分方程，而是将其初始化于 $y_0 \sim \gamma^d$，其中该点对应某个噪声分布，该分布满足 $\pi \approx q_0$，且易于从中采样。例如，在标准 OU 过程的例子中，我们可以取 $\pi = \gamma^d$，因为前向过程会以指数级速度收敛到 γ^d。由于容易从 π 中采样，因此可以使用概率流常微分方程从 q 中近似生成新的样本。

最后，我们进一步假设得分估计是利普希茨连续的。

假设 3（L 利普希茨得分估计值）。s_t is $\tilde{L}$-利普希茨：$\forall x, y \in \mathbb{R}^d$，$\|s_t(x) - s_t(y)\| \leq \tilde{L}\|x - y\|$。

2.2 对离散化截然不同的看法

传统上，要在离散时间中模拟连续时间常微分方程，会使用欧拉–丸山离散化等数值方法。这类离散化方案只能通过在一个时间窗口内使用固定梯度，近似模拟连续时间常微分方程。例如，要在时间窗口 $[0,h]$,内用一个梯度步近似常微分方程 $x_t = f_t(x_t)dt$，欧拉–丸山方法使用起始时刻 0 的梯度，得到更新式 $\hat{x}_h = hf_0(\hat{x}_0) + x_0$，其中 $\hat{x}_h$ 表示离散化后的过程。这里，h 是算法的步长。不幸的是，这类离散化方法在步长大小上存在一个根本性的、依赖维度的限制。

为了说明这一点，考虑对所有时间 $t \geq 0$ 都成立的向量场 $f_t = x$。在这种情况下，离散化步为 $\hat{x}_h = h(1)_{i=1}^d + x_0 = ((1+h))_{i=1}^d$；而常微分方程 $x_t = f_t(x_t)dt$ 的解，在初始条件 $x_0 = (1)_{i=1}^d$ 下为 $x_t = (e^t)_{i=1}^d$。这意味着 $\|x_h - \hat{x}_h\| = \|(e^h - (1+h))_{i=1}^d\| = \Theta(h^2\sqrt{d})$。因此，为了使偏差 $\|\hat{x}_h - \overline{x}_h\|$ 保持在 $O(1)$ 的范围内，步长 h 必须为 $O(1/d^{1/4})$。这是因为，沿着常微分方程解 $x(t)$ 前进时，向量场 $f_t(x_t)$ 可能会与其初始值大相径庭，因此使用固定向量 $f_0(x_0)$ 作为速度，就无法前进那么远。

然而，如果我们事先知道$f_t(x_t)$沿着路径 $(x_t)_{[0,h]}$的函数形式，或许就能利用这一点，通过采取更大的步长来模拟常微分方程。一种方法是利用f 的高阶光滑性$_t$ 并运行高阶数值求解器；正如第1.2节所述，许多先前研究采用了这一方法 [LWCC23、WCW24、LHE+24、HHL25a、HHL25b、LZWC25], 但据我们所知，这些方法都无法在 ε中实现指数级加速。

相反，我们选择了另一种策略。假设我们知道 $f_t(x_t)$属于一个以 $\{\phi_j(t)\}_{j=1}^D$为基的低维子空间，也就是说，对于某些系数 $(c_i)_{i=1}^D \in \mathbb{R}^D$，有 $f_t(x_t) = \sum_{i=1}^D c_i\phi_i$；那么，如果能够估计这些系数，我们就可以预测较大时间 t 下的路径 x_t 。在这项工作中，我们证明了一个或许令人意外的结果：在对目标分布仅施加最小条件的情况下，概率流常微分方程路径上得分函数的每个坐标 $f_t(x_t)$，都可以用 t 的低次多项式很好地近似。随后，我们证明，可以结合配点法与我们的得分估计来估计每个坐标对应的这些系数。值得注意的是，这使我们能够采取更长的离散步长来跟踪常微分方程路径，而且不显式依赖维度。我们强调，尽管配点法（即皮卡迭代）已用于先前的扩散模型理论研究 [GCC24, CRYR24, SBE+24]，但我们方法的关键创新在于利用低次多项式逼近。

2.3 配点法

配点法是一种通过不动点迭代来近似求解常微分方程的数值格式。这里考虑一个一般的初值问题：$dx_t = f_t(x_t)dt$，且对所有 $t \in [0,H]$，$x_0 = v$。该问题可写成积分表示 $x_t = v + \int_0^t f_s(x_s)ds$，可将其视为方程 $x = \mathcal{T}(x)$的不动点解，其中算子 $\mathcal{T}\colon \mathcal{C}([0,H],\mathbb{R}^d) \to \mathcal{C}([0,H],\mathbb{R}^d)$定义为 $\mathcal{T}(x)_t \triangleq v + \int_0^t f_s(x_s)ds$；若存在 $k \in \mathbb{N}$使得 $\mathcal{T}^k$ 是 L-利普希茨的，且 $L < 1$，则该算子有唯一不动点。 [LSV18] 中的作者提出了如下求解该不动点方程的方法。假设解的时间导数的每个坐标，即 $t \mapsto f_t(x_t)$，都可以用低次多项式很好地近似。令 $\phi_1, \ldots, \phi_D\colon [0,H] \to \mathbb{R}$ 为适当选取的多项式基，通过多项式插值，我们可以找到节点 $c_1, \ldots, c_D$，使得对所有 $i, j \in [D]$，均有 $\phi_j(c_i) = \mathbb{1}[i = j]$，因此特别地，

$$\frac{dx_t}{dt} \approx \sum_{j=1}^{D} f_{c_j}(x_{c_j})\phi_j(t).$$

像之前一样将其写成积分形式，便得到近似不动点方程：

$$x \approx \mathcal{T}_\phi(x), \qquad \mathcal{T}_\phi(x)_t \triangleq v + \int_0^t \sum_{j=1}^{D} f_{c_j}(x_{c_j})\phi_j(s)\,\mathrm{d}s\,. \tag{3}$$

这表明了一种自然的算法：在不动点迭代过程中，无需维护整个连续时间解$x:[0,H]\to\mathbb{R}^d$，只需维护这些值 $(x_{c_j})_{j\in[D]}$，并根据公式（3）更新它们；这相当于一次矩阵向量乘法。该算法总结如下，见算法 1 。

算法 1：PICARD($(f_t)_{t\in[0,H]}$, v, $(c_j)_{j=1}^{D}$, N) 输入：向量场 $f_t:\mathbb{R}^d\to\mathbb{R}^d$ 对于 $t\in[0,H]$，初值 v，迭代次数 N 输出：初值问题的近似解 $\hat{x}_H$ 1 定义 $A_\phi\in\mathbb{R}^{D\times D}$，使其满足 $(A_\phi)_{i,j}=\int_0^{c_j}\phi_i(s)\mathrm{d}s$ 2 定义 $X^{(0)}\in\mathbb{R}^{d\times D}$ 为 $v1_D^\top$，其中 1_D 为全 1 向量。3 对于 $t=0,\dots,N-1$，执行 4 定义 $F_c(X^{(t)})\in\mathbb{R}^{d\times D}$，使其满足 $F_c(X^{(t)})_{:,j}=f_{c_j}(X^{(t)}_{:,j})$ 5 $X^{(t+1)}\leftarrow v1_D^\top+F_c(X^{(t)})A_\phi$6 结束 7 返回 $v+\int_0^H\sum_{i=1}^{D}F_{c_i}(X^{(N)}_{:,i})\phi_i(s)\,\mathrm{d}s$

// (3)

[LSV18] 给出了在适当选择多项式基和节点的情况下，配点法能够收敛到常微分方程的足够精确解的条件。我们需要将其保证调整到我们的设定中，以考虑分数估计误差。完整的保证将在第 3 节给出。我们对多项式基要求满足以下条件 ϕ:

定义 2.1。 我们称 ϕ 是 γ_ϕ-有界的，如果对于所有基元素 $\phi_j:\ [0,H]\to\mathbb{R}$ 和 $0\le x\le H$, $\sum_{j=1}^{D}\left|\int_0^H\phi_j(s)\,\mathrm{d}s\right|\le\gamma_\phi t$.

~在本文中，具体而言，我们将取 $\phi_j(x)=\phi_j(2x/H-1)$，其中

$$\widetilde{\phi}_j(x)=\frac{\sqrt{1-c_j^2}\cos(D\cos^{-1}x)}{D(x-c_j)},$$

其中 $c_j=\cos(\frac{2j-1}{2D}\pi)$ 是切比雪夫多项式的 j 次根，该多项式的次数为 D。正如~[LSV18 的证明，引理 2。9]中所示，$\widetilde{\phi}_j$ 是一个次数为 j 的多项式，满足 $\widetilde{\phi}_j(c_i)=\mathbb{1}[i=j]$；此外，[LSV18]指出，根据 [LV17，引理 91]，

$$\sum_{j=1}^{D}\left|\int_{-1}^{1}\widetilde{\phi}_j(x)\mathrm{d}x\right|\le 2000\,.$$

~经过变量替换，将尺度从 $\widetilde{\phi}_j$ 调整为 ϕ_j后，我们得出结论：该基 (ϕ_j) 是 $\gamma_\phi=O(H)$有界的。

2.4 我们的算法

现在，我们可以将前几小节中的各项要素结合起来，对我们的采样器进行描述。该算法只是将反向过程划分为一系列较小的时间窗口，每个时间窗口的长度为

算法 2： COLLOCATIONDIFFUSION

输入： 得分估计 s_t 满足假设 2 和 3，目标误差 $0 < \varepsilon < 1$，去噪调度 $0 < t_1 < \cdots < t_n = T$，皮卡迭代深度 N

输出： 从满足 $W_2(\hat{q}, q) \le \varepsilon$ 的分布 $\hat{}$分布 q 中采样

$\hat{x}_0 \sim \gamma^d$

对于 $k =$ 分布 q, ..., $n-1$, n **执行**

  $\hat{x}_{t_k} = \mathrm{PICARD}((s_t)_{t\in[t_{k-1},t_k]}, \hat{x}_{t_{k-1}}, N)$

结束

返回 $\hat{x}_{t_n}$

该长度为 $\tilde{O}(\sigma^2 R_2)$，并通过 PICARD 配点法成功求解每个时间窗口中的相应初值问题。由于 PICARD 调用并不能得到精确解，误差会有所累积，但这些误差可以变得极小，因为 PICARD 能够以极高的精度求解相关常微分方程。

3 证明概述

本节介绍证明的主要组成部分，具体证明留待附录给出。回顾一下，我们方法的关键在于证明一个结构性结果，表明沿反向过程的得分函数的时间导数可以得到控制。

3.1 沿真实反向过程的低次数近似

我们的出发点是，将反向过程中点 y_t 处得分函数的时间导数，与前向过程中生成 y_t 的干净样本的后验分布的高阶矩联系起来。

我们从以下推导开始。

引理 3.1。假设 $y_t = X_t + \sigma_t \xi$，其中 $X_t = e^{-(T-t)}\bar{X}$，且 $\bar{X} \sim q_{\mathrm{pre}}$，并且 $\xi \sim \gamma^d$。我们使用符号说明 $\mathbb{E}^{t,y_t}$ 表示给定 $y_t = X_t + \sigma_t \xi$ 时 X_t 的条件期望，并用 q^{t,y_t} 表示该后验分布的密度。定义后验均值为 $\mu_t(y) \triangleq \mathbb{E}^{t,y} X_t$。

概率流常微分方程的向量场的时间导数可计算如下：

$$
\begin{aligned}
\partial_t(y_t + \nabla \log q_t(y_t)) = &\left(y_t + \frac{\mathbb{E}^{t,y_t} X_t - y_t}{\sigma_t^2}\right) \\
&+ \frac{1}{\sigma_t^2}\mathbb{E}^{t,y_t} X_t + \frac{1-\sigma_t^2}{2\sigma_t^6}\mathbb{E}^{t,y_t}\left(\langle y_t, X_t\rangle + \|X_t\|^2\right)\left(X_t - X_t'\right) + \frac{4}{\sigma_t^4}\mathbb{E}^{t,y_t}\left(\langle y_t, X_t\rangle - \|X\|^2\right)\left(X_t - X_t'\right) \\
&+ \frac{1}{\sigma_t^2}\left(\frac{1}{\sigma_t^2} - 1\right) y_t - \frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t} X_t + \frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t}\left(2\left\langle y_t, \frac{X_t''}{\sigma_t^2}\right\rangle - \langle X_t, y_t\rangle - \frac{1}{\sigma_t^2}\left\langle X_t, X_t''\right\rangle\right)(X_t - X_t').
\end{aligned}
$$

这里，X_t, X_t', X_t'' 表示从后验分布中独立同分布抽取的样本，其期望算子为 $\mathbb{E}^{t,y_t}$。

多次迭代这一计算后，我们得到以下关键计算结果，用于表示概率流常微分方程向量场的高阶导数：

L引理 3.2。 反向常微分方程向量场的第 k 阶导数可以展开为 s

$$\partial_t^k\left(y_t+\nabla \ln q_t(y_t)\right)$$
$$=\sum_{r_1,r_2,r_3,r_4}\ \sum_{\mathbf{i}\in[k]^{r_3},\mathbf{k}\in[k]^{r_4},\mathbf{l}\in[k]^{r_4}} \frac{\left(1-\sigma_t^2\right)^{r_1}}{\sigma_t^{r_2}}\mathbb{E}^{t,y_t}\prod_{j=1}^{r_3}\left\langle y_t, X^{(\mathbf{i}_j)}\right\rangle\prod_{j=1}^{r_4}\left\langle X^{(\mathbf{k}_j)}, X^{(\mathbf{l}_j)}\right\rangle\left(a_{\mathbf{i},\mathbf{k},\mathbf{l}}X+b_{\mathbf{i},\mathbf{k},\mathbf{l}}y_t\right)$$

其中 $r_1, r_2, r_3 \le k$ 和 $r_2 \le 4k$，以及系数 $(a_{\mathbf{i}、\mathbf{k}、\mathbf{l}}、b_{\mathbf{i}、\mathbf{k}、\mathbf{l}})$ 满足 $\sum_{\mathbf{i}\in[k]^{r_3}、\mathbf{k}\in[k]^{r_4}、\mathbf{l}\in[k]^{r_4}}\left|a_{\mathbf{i}、\mathbf{k}、\mathbf{l}}\right|+\left|b_{\mathbf{i}、\mathbf{k}、\mathbf{l}}\right| \le (74k)^k$。这里的随机变量 $X^{(i)}$ 彼此独立，并根据后验分布 q^t、^{y}进行分布，而 i、k、l 是任意的索引元组。

这一结果可以从与 Tweedie 公式相同的角度来理解，该公式将 X 的后验均值与得分函数联系起来。文献中关于 Tweedie 公式的高阶推广 [MSLE21]通常将后验分布的矩与得分函数相对于 空间变量的导数联系起来，而在引理 3.2 中，我们考虑的是相对于 时间变量的导数。

我们方法的一个重要特点是，对得分函数高阶导数的多重积分公式进行界定，且该界与环境维数 d无关；这对于获得仅取决于分布有效半径R 的迭代复杂度界至关重要。更具体地说，有效半径的界使我们能够对如下形式的期望进行界定 $\mathbb{E}[\prod^{r_4}{}_{j=1}\ \langle X^{(\mathbf{k}_j)}, X^{(\mathbf{l}_j)}\rangle\ ]$，并以与维数无关的量界定如下形式的期望 $\mathbb{E}[\prod^{r_3}{}_{j=1}\ \langle y, X^{(\mathbf{i}_j)}\rangle\ ]$，且后者以高概率成立。由此，我们可以通过应用柯西-施瓦茨不等式，推导出引理 3.2 左侧的界（见引理 C.4）。利用这些公式，我们得到了首个关键技术要素：对得分函数高阶时间导数的无维数界：

引理 3.3。 若分布 q 满足假设 *1*，$(y_t)_t$是 (2)的一个解，且 $y_0 \sim q_T$，则概率流常微分方程中向量场的第 k^{th} 阶导数有如下界

$$\left\|\partial_t^k\left(y_t+\nabla \log q_t(y_t)\right)\right\|_{p,\infty} \lesssim R\left(\frac{74k}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t}+(kp)^{1/2}\right)^{2k}$$

注意，这些界随导数的阶数呈单指数增长，并随 $(R/\sigma_t)^k$ 变化。如果这些导数界逐点成立，那么我们就可以直接利用泰勒余项定理，说明真实分数可由低次多项式很好地近似。但归根结底，引理 3.3仅针对真实反向过程所诱导的分布给出了高概率陈述。我们需要处理偏离真实反向过程的情况：一方面，采样器的初始化来自高斯分布，而非正确的边缘分布；另一方面，在后续每个应用配点法的窗口中，由于前面窗口累积的离散化误差，初始化也会略微偏离正确的边缘分布。接下来，我们将说明如何处理这些偏差。

3.2 沿算法进行低次数近似

为了论证实际采样器轨迹上的低次数近似性质，我们需要证明引理3.2 中的后验分布对条件的扰动具有鲁棒性。

为此，我们的关键工具是以下耦合引理：

引理 3.4 (耦合)。 设 $\tilde{y}$ 是一个随机变量，使得 $\|y-\tilde{y}\| \le \delta$ 对于 $\delta \le \frac{1}{6(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})}$，那么，给定 $T-t \ge 1$，在事件 $\mathcal{A}_1$ （方程 (13)中定义的）下，我们有

$$\mathrm{TV}(q^{t,y}, q^{t,\tilde{y}}) \le 8\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}). \tag{4}$$

特别地，等式 (4) 以至少 $1-\varepsilon_1$ 的概率成立。

结合引理 3.3 和3.4，我们得到一个界，刻画了从一个与真实边缘分布略有偏差的分布出发时，沿反向过程的得分函数高阶时间导数：

定理 3.5。 设 $\delta \in (0,1)$。若 q 满足假设 *1*，且 $(\tilde{y}_t)_t$通过从一个与 q 在 W δ中相差 2 的分布初始化反向过程0 并在连续时间下运行概率流常微分方程得到。那么，对于 δ 足够小的情况（其严格刻画见引理 *E.6*），以至少 $1-\delta_2$的概率（关于初始化的随机性），概率流常微分方程沿轨迹的向量场的 k^{th} 阶导数在 $\tilde{y}_t$ 处有界

$$\left\|\partial_t^k\left(\tilde{y}_t + \nabla \log q_t(\tilde{y}_t)\right)\right\|_\infty \lesssim R\left(\frac{k}{\sigma_t^2}\right)^k \left(\frac{R}{\sigma_t}\right)^{2k} + (k\log(74kd/\delta_2))^{2k}\,.$$

该定理为我们的方法提供了理论依据。由于这是一个关于从真实采样器正确的中间点初始化的过程的迭代项的高概率界，我们现在可以应用前述泰勒截断论证，沿采样器本身的轨迹而非理想化反向过程的轨迹，得到所需的低次数近似。我们将结合 [LSV18] 中的技术，证明以下 PICARD 的收敛界（算法 1），并回顾第 2.3节中的符号说明：

命题 3.6 （非正式表述，见附录）。 假设PICARD 使用一个 γ_ϕ-有界多项式基 ϕ运行，其节点为 (c_j)。对于任意 $h \leq \frac{1}{2\tilde{}L\gamma_\phi}$，对于真实的概率流常微分方程 (y_t)以及任意 $x \in \mathcal{C}([t_0, t_0+h],\ \mathbb{R}^d)$,

$$\left\|T_\phi^{\circ m}(y) - y\right\|_{[t_0,t_0+h]} \leq 2\Big(\varepsilon_{\mathrm{ld}} + \max_{1\leq j\leq D}\left\|s_{c_j}(y(c_j)) - y(c_j) - \nabla \log q_{c_j}(y(c_j))\right\|\Big)\big(1+\gamma_\phi\big)h \tag{5}$$

$$\|T_\phi^{\circ m}(x) - T_\phi^{\circ m}(y)\|_{[t_0,t_0+h]} \leq \frac{1}{2^m}\|x-y\|_{[t_0,t_0+h]}\,, \tag{6}$$

其中，$\|\cdot\|_{[a,b]}$ 表示区间 $[a,\ b]$, 上的上确界范数，而 $\varepsilon_{\mathrm{ld}}$ 表示沿概率流常微分方程对得分函数进行逼近时产生的逼近误差（按上确界范数计），所用的多项式属于 ϕ的张成空间。

上文的式（5）表明，真实概率流常微分方程的轨迹是精确皮卡迭代的精确不动点，而对于基于多项式插值的近似皮卡迭代 *PICARD* （算法1）而言，该轨迹在局部是一个近似不动点。式（6）表明，任何局部接近这一真实轨迹的轨迹都会以指数速率向真实轨迹收敛，直至误差达到式（5）中的界所给出的水平。

一个重要的注意事项是，求解器能够运行的时间只能与基的有界性γ_ϕ 成反比（回忆定义 2.1），而该有界性可能是常数阶；同时，运行时间也只能与概率流常微分方程向量场的利普希茨连续性成反比。这就是我们无法直接使用配置法一次性求解概率流常微分方程的原因。尽管如此，关键在于，正如上面的第二个不等式所示，PICARD 产生的误差会随着迭代次数增加而呈指数递减，但还需额外考虑由低次数近似和分数估计误差带来的误差。随后，我们将串联多次应用上述界，得到最终的保证。

最后，我们指出，尽管该命题的核心思想源自 [LSV18], 但我们的界的一项创新之处在于，我们考虑了分数估计误差，并证明由此产生的误差不会在皮卡迭代中呈指数级累积。这对于确保我们的采样器足够接近概率流常微分方程的真实轨迹至关重要。

3.3 采样器的收敛性

现在，我们已经具备证明主要结果所需的全部要素。

我们首先说明如何获得一个W_2-接近分布 q的采样器。与我们的总变差收敛保证相比，这一保证有两个令人称道的特点：1）对分数估计误差的要求宽松得多；2）只需运行算法 CollocationDiffusion，也就是说，只需使用常微分方程求解器模拟真实的反向过程，无需进行任何后处理 [LSV18] 即可实现离散步。

定理 *3.7.* 设 $\varepsilon_{err} \le O(\varepsilon/ln(d/R))$, $\sigma > 0$ 且 $R \ge 1$。假设分布 q 满足参数为 R, σ 的假设 *1*。给定满足假设 *2* 和 *3* 的得分估计 s_t，以至少 $1 - O(\varepsilon(R/\sigma)^2)$ 的概率，算法 *2* 输出来自某个分布 $\hat{q}$ 的样本，使得 $W_2(q, \hat{q}) \le \tilde{O}(\varepsilon \cdot log^2(1/\varepsilon))$ *in* $\tilde{O}((R/\sigma)^2 \cdot log(1/\varepsilon))$ 轮数，并且每轮进行 $\tilde{O}(log(1/\varepsilon))$ 次皮卡迭代，总计调用得分估计预言机 $\tilde{O}((R/\sigma)^2 \cdot log^2(1/\varepsilon))$ 次。

注 3.8。上文中，$\tilde{O}(\cdot)$ 隐藏了 $\log d$、$\log R$、$\log \tilde{L}$、$\log\log(1/\varepsilon)$ 等因子。此外，假设 $R \ge 1$ 仅为简化界。完整定理请参见附录 *G*。

这里我们非正式地概述该结果的证明，证明细节留待附录给出。其思路是在所有时间点紧密追踪真实的概率流常微分方程轨迹，并依次在较小的时间窗口上运行 PICARD。

假设在反向过程的模拟中，截至某个时间 t，采样器的迭代结果服从某个分布，且该分布与真实边缘分布之间的W_2 距离较小。接下来，考虑再模拟 h 个时间步以逼近分布 q_{T-t-h}。只需运行 CollocationDiffusion即可实现这一点，前提是 h 足够小；由于初始分布与真实边缘分布之间的W_2距离很小，泰勒截断论证结合向量场高阶导数的界，保证我们可以安全地应用命题 3.6，使任何由W_2 差异造成的误差都能以指数速度趋于零，最终仅留下固定量的分数估计误差和多项式逼近误差。

请注意，W_2 差异所导致的误差指数收缩，对我们的结果至关重要。每次运行皮ICARD时，我们都可以借此“重置”偏离真实反向过程曲线的程度。这让人想起近期另一种分析概率流常微分方程的方法，该方法由 [CCL+24]提出。在那项工作中，作者尝试在较短的时间窗口内模拟常微分方程，并在每个窗口中使用朴素的 Wasserstein 耦合论证。在他们的设定中，Wasserstein 耦合会产生一些误差，他们希望这些误差在整个采样器运行过程中线性累积，而不是指数累积。为此，他们在每个新时间窗口开始时向采样器注入少量噪声，从而有效地“重启”Wasserstein 耦合。相比之下，我们不通过注入噪声来重启耦合，而是使用皮卡迭代来抑制不断累积的 Wasserstein 误差。

提升至全变差接近性。我们可以运用上述来自 [CCL+24] 的思路，推导出主定理的另一项推论。具体而言，利用欠阻尼朗之万动力学的正则化性质，我们在 Wasserstein距离下的采样保证（定理 3.7）可通过标准论证转化为全变差距离下的保证 [GCC24、CCL+24]，前提是进一步假设真实的原始得分函数 $\nabla \ln$ 分布 q 是利普希茨的。

推论 *3.9*。令 ε、R、$\sigma > 0$。假设 q 是一个满足假设 *1*（参数为 R、σ）的分布，并且 $\nabla ln\, q$ 是 L-利普希茨的。给定满足假设 *2* 的得分估计 s_t，其中 $\varepsilon_{sc} \le \tilde{O}(\frac{\varepsilon^{2/3}}{L^{1/2} d^{1/6} R^{2/3}})$，算法 *3* 输出来~~自分布~~ $\hat{q}$ 的样本，并满足 $TV(\hat{q}, q) \le \varepsilon$ *in* $\tilde{O}\left((R/\sigma)^2\right)$ 次迭代。

在高精度情形下，从 Wasserstein 距离的接近性出发，却能如此容易地转换为全变差距离的接近性，这似乎有些反直觉，因为无需进行任何形式的 Metropolis 调整，就必须运行欠阻尼朗之万蒙特卡洛。关键思路是简单地 缩短运行欠阻尼朗之万的时间，从而使离散化带来的偏差不会过大；这一思路源于 [GCC24]。我们在附录 H 中对这一论证进行了形式化。

4 展望

在这项工作中，我们提出了一种基于扩散的采样器，可从一大类分布中进行采样，其中高斯混合是一个特例。该采样器的迭代复杂度随 $1/\varepsilon$ （“高精度”）呈多对数级增长，且不显式依赖维度，而是取决于分布的有效半径。关键技术步骤是建立概率流常微分方程向量场高阶导数的界，进而将其近似为低次多项式。这个近似结果使我们能够利用配点法的一种变体——一种用于数值求解常微分方程的不动点迭代方法，由 [LSV18] 提出。如果在足够小的时间窗口内运行，该方法会以指数速度收敛到真实的常微分方程解。

我们的工作存在一些局限，也由此引出了值得未来研究的有趣问题。首先，我们能否以某种方式放宽对底层分布q的假设？我们不必坚持认为它是分布q_{pre} 的加噪版本，且该分布的支撑集是紧致的；也可以仅要求q_{pre} 具有有界矩。我们还可以要求经过若干步后与q 达到全变差或 KL 散度收敛，且步数仅随 $1/\sigma$呈对数级增长，这与低精度区域中已知的结果相符 [CLL23、BDD23]。最后，为了在得分访问条件下实现采样的高精度保证，得分误差具有次指数尾部这一假设是否真的必要？

致谢

我们感谢 Sinho Chewi、Holden Lee、Yin Tat Lee 和 Jianfeng Lu 就离散化界提出的宝贵建议。SC 获得 NSF CAREER 奖项 CCF-2441635 的支持。

参考文献

[AC24] Jason M Altschuler 和 Sinho Chewi。通过算法预热实现更快的高精度对数凹采样。*ACM* 汇刊，71(3):1–55，2024年。[AHL$^+$23] Nima Anari、Yizhi Huang、Tianyu Liu、Thuy-Duong Vuong、Brian Xu 和 Katherine Yu。通过连续游走实现并行离散采样。载于第*55*届*ACM*计算理论年度研讨会论文集，第103–116页，2023年。[BDBDD24] Joe Benton、Valentin De Bortoli、Arnaud Doucet 和 George Deligiannidis。通过随机局部化获得扩散模型的近似 d 阶线性收敛界。载于第十二届国际学习表征会议，2024年。[BDD23] Joe Benton、George Deligiannidis 和 Arnaud Doucet。流匹配方法的误差界。*arXiv* 预印本 *arXiv:2305.16860*，2023年。[BMR22] Adam Block、Youssef Mroueh 和 Alexander Rakhlin。结合去噪自编码器和朗之万采样的生成建模。*arXiv* 预印本 *2002.00107*，2022年。

[CCL+23] 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简单：具有最小数据假设的扩散模型理论。载于国际学习表征会议，2023年。[CCL+24] Sitan Chen、Sinho Chewi、Holden Lee、Yuanzhi Li、Jianfeng Lu 和 Adil Salim。概率流常微分方程已被证明具有快速性。神经信息处理系统进展，36，2024年。[CDD23] Sitan Chen、Giannis Daras 和 Alex Dimakis。超越线性扩散过程的复原—退化：对 DDIM 类采样器的非渐近分析。载于国际机器学习会议，第4462–4484页。PMLR，2023年。[CDS23] Giovanni Conforti、Alain Durmus 和 Marta Gentiloni Silveri。无需提前停止的得分扩散模型：有限 Fisher 信息即为所需。arXiv 预印本 arXiv:2308.12240，2023年。[CDWY20] Yuansi Chen、Raaz Dwivedi、Martin J Wainwright 和 Bin Yu。Metropolized哈密顿蒙特卡罗的快速混合：多步梯度的优势。机器学习研究杂志，21(92):1–72，2020年。[Che23] Sinho Chewi。对数凹采样。书稿见 https://chewisinho. github. io，2023年。[CLA+21] Sinho Chewi、Chen Lu、Kwangjun Ahn、Xiang Cheng、Thibaut Le Gouic 和 Philippe Rigollet。Metropolis 调整朗之万算法的最优维度依赖性。载于学习理论会议，第1260–1300页。PMLR，2021年。[CLL23] Hongrui Chen、Holden Lee 和 Jianfeng Lu。得分函数生成建模的改进分析：最小光滑性假设下的易用界。载于国际机器学习会议，第4735–4763页。PMLR，2023年。[CMFW24] Minshuo Chen、Song Mei、Jianqing Fan 和 Mengdi Wang。扩散模型概览：应用、引导生成、统计速率与优化。arXiv 预印本 arXiv:2404.07771，2024年。[CRYR24] Haoxuan Chen、Yinuo Ren、Lexing Ying 和 Grant Rotskoff。通过并行采样加速扩散模型：亚线性时间复杂度的推理。神经信息处理系统进展，37:133661–133709，2024年。[DB22] Valentin De Bortoli。流形假设下去噪扩散模型的收敛性。机器学习研究汇刊，2022年。[DBTHD21] Valentin De Bortoli、James Thornton、Jeremy Heng 和 Arnaud Doucet。扩散 Schrödinger 桥及其在基于得分函数的生成建模中的应用。载于 M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang 和 J. Wortman Vaughan 编，神经信息处理系统进展，第34卷，第17695–17709页。Curran Associates, Inc.，2021年。[DCWY19] Raaz Dwivedi、Yuansi Chen、Martin J Wainwright 和 Bin Yu。对数凹采样：Metropolis-Hastings 算法速度很快。机器学习研究杂志，20(183):1–42，2019年。

[DKS18] Ilias Diakonikolas、Daniel M Kane 和 Alistair Stewart。列表可解码稳健均值估计与球形高斯混合学习。载于《第50届ACM SIGACT计算理论年会论文集》，第1047–1060页，2018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。载于 M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang 和 J. Wortman Vaughan 编，《神经信息处理系统进展》第34卷，第8780–8794页。Curran Associates, Inc.，2021年。[GCC24] Shivam Gupta、Linda Cai 和 Sitan Chen。采用随机中点的更快基于扩散的采样：顺序式与并行式。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。arXiv预印本 arXiv:2404.18869，2024年。[GPPX23] Shivam Gupta、Aditya Parulekar、Eric Price 和 Zhiyang Xun。扩散模型的样本高效训练。arXiv预印本 arXiv:2311.13745，2023年。[GW12] Arnaud Guillin 和 Feng-Yu Wang。退化的福克-普朗克方程：Bismut公式、梯度估计与哈纳克不等式。《微分方程杂志》253(1):20–40，2012年。[HHL25a] Daniel Zhengyu Huang、Jiaoyang Huang 和 Zhengjiang Lin。基于分数的生成模型中概率流常微分方程的收敛分析。《IEEE信息论汇刊》，2025年。[HHL25b] Daniel Zhengyu Huang、Jiaoyang Huang 和 Zhengjiang Lin。扩散概率模型中高阶常微分方程求解器的快速收敛。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》33:6840–6851，2020年。[HL18] Samuel B Hopkins 和 Jerry Li。混合模型、稳健性与平方和证明。载于《第50届ACM SIGACT计算理论年会论文集》，第1021–1034页，2018年。[HWC24] Zhihan Huang、Yuting Wei 和 Yuxin Chen。去噪扩散概率模型能够自适应于未知低维性，并达到最优性。arXiv预印本 arXiv:2410.18784，2024年。[HZD+24] Xunpeng Huang、Difan Zou、Hanze Dong、Zhang Zhang、Yian Ma 和 Tong Zhang。反向转移核：加速扩散推理的灵活框架。《神经信息处理系统进展》37:95515–95578，2024年。[JL24] Yuchen Jiao 和 Gen Li。扩散模型的依赖于实例的收敛理论。arXiv预印本 arXiv:2410.13738，2024年。[JZL25] Yuchen Jiao、Yuchen Zhou 和 Gen Li。一般分布下去噪扩散概率模型的最优收敛分析。arXiv预印本 arXiv:2510.27562，2025年。[KSS18] Pravesh K Kothari、Jacob Steinhardt 和 David Steurer。通过平方和实现稳健矩估计与改进的聚类。载于《第50届ACM SIGACT计算理论年会论文集》，第1035–1046页，2018年。

[LCW25] Gen Li、Changxiao Cai 和 Yuting Wei。近似高斯混合的扩散模型无维度收敛。arXiv预印本 arXiv:2504.05300，2025年。[LHC25] Jiadong Liang、Zhihan Huang 和 Yuxin Chen。扩散模型的低维适应：全变差收敛。arXiv预印本 arXiv:2501.12982，2025年。[LHE+24] Gen Li、Yu Huang、Timofey Efimov、Yuting Wei、Yuejie Chi 和 Yuxin Chen。可证明地加速基于得分的扩散模型的收敛。arXiv预印本 arXiv:2403.03852，2024年。[LJ24] Gen Li 和 Yuchen Jiao。改进扩散概率模型的收敛速率。载于《第十三届国际学习表征会议》，2024年。[LL22] Allen Liu 和 Jerry Li。多项式时间内以近乎最优的分离度聚类混合物。载于《第54届ACM SIGACT计算理论年会论文集》，第1248–1261页，2022年。[LLT22] Holden Lee、Jianfeng Lu 和 Yixin Tan。具有多项式复杂度的基于得分的生成建模收敛。《神经信息处理系统进展》35:22870–22882，2022年。[LLT23] Holden Lee、Jianfeng Lu 和 Yixin Tan。一般数据分布的基于得分的生成建模收敛。载于《国际算法学习理论会议》，第946–985页。PMLR，2023年。[LST20] Yin Tat Lee、Ruoqi Shen 和 Kevin Tian。对数光滑梯度集中与 Metropolized哈密顿蒙特卡罗 更紧的运行时间。载于《学习理论会议》，第2565–2597页。PMLR，2020年。[LSV18] Yin Tat Lee、Zhao Song 和 Santosh S Vempala。常微分方程的算法理论与从良态对数凹概率密度中采样。arXiv预印本 arXiv:1812.06243，2018年。[LV17] Yin Tat Lee 和 Santosh S Vempala。多面体中的测地线行走。载于《第49届ACM SIGACT计算理论年会论文集》，第927–940页，2017年。[LWCC23] Gen Li、Yuting Wei、Yuxin Chen 和 Yuejie Chi。迈向扩散生成模型更快的非渐近收敛。arXiv预印本 arXiv:2306.09251，2023年。[LWCC24] Gen Li、Yuting Wei、Yuejie Chi 和 Yuxin Chen。扩散模型概率流常微分方程的精确收敛理论。arXiv预印本 arXiv:2408.02320，2024年。[LWYL22] Xingchao Liu、Lemeng Wu、Mao Ye 和 Qiang Liu。让我们搭建桥梁：理解并扩展扩散生成模型。arXiv预印本 arXiv:2208.14699，2022年。[LY24a] Gen Li 和 Yuling Yan。适应基于得分的扩散模型中的未知低维结构。arXiv预印本 arXiv:2405.14861，2024年。[LY24b] Gen Li 和 Yuling Yan。在最小假设下扩散概率模型的 O (d/t) 收敛理论。arXiv预印本 arXiv:2409.18959，2024年。[LZB+22] Cheng Lu、Yuhao Zhou、Fan Bao、Jianfei Chen、Chongxuan Li 和 Jun Zhu。Dpm-solver：用于扩散概率模型采样的快速常微分方程求解器，可在约10步内完成采样。《神经信息处理系统进展》35:5775–5787，2022年。

[LZB+25] Cheng Lu、Yuhao Zhou、Fan Bao、Jianfei Chen、Chongxuan Li 和 Jun Zhu。Dpm-solver++：用于扩散概率模型引导采样的快速求解器。《机器智能研究》，第 1–22 页，2025年。[LZWC25] Gen Li、Yuchen Zhou、Yuting Wei 和 Yuxin Chen。通过高阶近似加速扩散模型。arXiv 预印本 arXiv:2506.24042，2025年。[MSLE21] Chenlin Meng、Yang Song、Wenzhe Li 和 Stefano Ermon。通过去噪估计数据分布的高阶梯度。《神经信息处理系统进展》，34：25359–25369，2021年。[NBZA24] Preetum Nakkiran、Arwen Bradley、Hattie Zhou 和 Madhu Advani。逐步扩散：基础教程。arXiv 预印本 arXiv:2406.08929，2024年。[PAD24] Peter Potaptchik、Iskander Azangulov 和 George Deligiannidis。流形假设下扩散模型的线性收敛。arXiv 预印本 arXiv:2410.09046，2024年。[Pid22] Jakiw Pidstrigach。基于分数的生成模型能够检测流形。载于 S. Koyejo、S. Mo- hamed、A. Agarwal、D. Belgrave、K. Cho 和 A. Oh 编，《神经信息处理系统进展》，第 35 卷，第 35852–35865 页。Curran Associates, Inc.，2022年。[SBE+24] Andy Shih、Suneel Belkhale、Stefano Ermon、Dorsa Sadigh 和 Nima Anari。扩散模型的并行采样。《神经信息处理系统进展》，36，2024年。[SDME21] Yang Song、Conor Durkan、Iain Murray 和 Stefano Ermon。基于得分的扩散模型的最大似然训练。载于 M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang 和 J. Wortman Vaughan 编，《神经信息处理系统进展》，第 34 卷，第 1415–1428 页。Curran Associates, Inc.，2021年。[SDWMG15] Jascha Sohl-Dickstein、Eric Weiss、Niru Maheswaranathan 和 Surya Ganguli。利用非平衡热力学进行深度无监督学习。载于 Francis Bach 和 David Blei 编，《第 32 届国际机器学习会议论文集》，《机器学习研究论文集》第 37 卷，第 2256–2265 页，法国里尔，2015年7月。PMLR。[SE19] Yang Song 和 Stefano Ermon。通过估计数据分布的梯度进行生成建模。载于 H. Wallach、H. Larochelle、A. Beygelzimer、F. d'Alché-Buc、E. Fox 和 R. Garnett 编，《神经信息处理系统进展》，第 32 卷。Curran Associates, Inc.，2019年。[SSDK+21] 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。载于《国际学习表征会议》，2021年。[VKK21] Arash Vahdat、Karsten Kreis 和 Jan Kautz。在潜空间中进行基于得分的生成建模。载于 M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang 和 J. Wortman Vaughan 编，《神经信息处理系统进展》，第 34 卷，第 11287–11302 页。Curran Associates, Inc.，2021年。[Wai25] Martin J Wainwright。无需扩散过程的基于得分采样：一种简单且模块化方案的指导。arXiv 预印本 arXiv:2512.24152，2025年。[WCW24] Yuchen Wu、Yuxin Chen 和 Yuting Wei。随机 Runge-Kutta 方法：扩散模型的可证明加速。arXiv 预印本 arXiv:2410.04760，2024年。

[WSC22] Keru Wu、Scott Schmidler 和 Yuansi Chen。对数凹采样中 Metropolis 调整朗之万算法的极小极大混合时间。机器学习研究杂志，23(270)：1–63，2022年。[WY22] Andre Wibisono 和 Kaylee Y. Yang。非精确朗之万算法的 KL 散度收敛及其在基于得分的生成模型中的应用。*arXiv* 预印本 *2211.01512*，2022年。[ZBR+23] Wenliang Zhao、Lujia Bai、Yongming Rao、Jie Zhou 和 Jiwen Lu。UniPC：用于扩散模型快速采样的统一预测校正框架。神经信息处理系统进展，36：49842–49869，2023年。[ZHH+25] Matthew S Zhang、Stephen Huan、Jerry Huang、Nicholas M Boffi、Sitan Chen 和 Sinho Chewi。即使对于去噪扩散概率模型，亚线性迭代也可能足够。*arXiv* 预印本 *arXiv:2511.04844*，2025年。

进一步的相关工作

关于扩散过程离散化界的其他工作。近年来，大量研究给出了扩散模型的一般收敛保证 [DBTHD21, BMR22, DB22, LLT22, LWYL22, Pid22, WY22, CCL+23, CDD23, LLT23, LWCC23, BDD23, CCL+24, BDBDD24, CLL23, GPPX23, LWCC24, LY24b, CDS23, GCC24]。其中一系列工作 [CCL+23, LLT23, CLL23, BDBDD24, CDS23, LY24b] 分析了去噪扩散概率模型（DDPM），即概率流常微分方程的随机类比，并证明了 $\tilde{O}(d)$ 的迭代复杂度界。近期，[BDBDD24, CDS23] 在仅假设数据分布具有有界二阶矩的条件下建立了这些界；而最近，[LY24b]进一步将其推广至具有有界一阶矩的分布，并使其对 $1/\varepsilon$ 的依赖仅为线性。另一类工作 [CCL+24, CDD23, LWCC23, LHE+24, GCC24, LWCC24, LJ24, JL24] 研究了概率流常微分方程，并证明了类似的速率。一个显著区别是，在光滑性假设下，已知可以实现对 d 的次线性依赖 [CCL+24, GCC24, LJ24, JL24]。最近，[ZHH+25, JZL25]表明，仅使用去噪扩散概率模型也可以实现对 d 的次线性缩放。然而，就目前而言，所有这些工作都仅适用于“低精度”情形，也就是说，这些工作已知可达到的最佳 ε 仍为 $\mathrm{poly}(1/\varepsilon)$，即使在光滑性假设下也是如此。

对数凹分布的高精度采样。高精度对数凹采样研究中的一个核心思想是：在精度较低的采样器每一步之后，都追加一个 Metropolis-Hastings 滤波器步骤。将这一做法与朗之万蒙特卡罗或哈密顿蒙特卡罗相结合，分别产生了广受欢迎的 MALA 和 MHMC 方法。一系列研究 [DCWY19、CDWY20、LST20、CLA+21、WSC22], 最终迎来了近期的突破 [AC24], , 确立了 MALA 用于对数凹分布采样的迭代复杂度为 $\tilde{O}(\sqrt{d}\,\mathrm{poly}(1/\varepsilon))$。

B 符号说明

我们将使用定义在 $\mathbb{R}^d$ 上的数据分布 分布 q 。在全文中，$\bar{X}$ 表示从分布 q中采样得到的随机向量，而X_t表示$X_t = \sqrt{1-\sigma_t^2}\,\bar{X}$的缩放。同样，我们用 $\bar{x}$和x_t表示这些随机向量的实现。我们用 ξ 表示从 γ^d中采样得到的随机向量。给定t、y，我们用分布 $q^{t,y}$表示在 $\sqrt{1-\sigma_t^2}\,\bar{X}+\sigma_t\xi = y$条件下$X_t$的后验分布。我们用 $\mathbb{E}^{t,y}$表示关于分布 $q^{t,y}$的期望，并记作 $\mu_t(y) \triangleq \mathbb{E}^{t,y}X_t$。

我们将使用以下噪声调度

$$\sigma_t \triangleq \sqrt{1-e^{-2(T-t)}},.$$

在全文中，我们令 $q_t \triangleq$概率分布$(\sqrt{1-\sigma_t^2}\,\bar{X}+\sigma_t\xi)$，从而 $q_{\text{pre}}=q_T$。[1]

给定 t，令 $F^*{}_t:\mathbb{R}^d\to\mathbb{R}^d$ 表示映射

$$F_t^*(y)=y+\nabla\ln q_t(y). \tag{7}$$

若上下文已明确指出 t，我们将省略下标。注意

$$\nabla\ln q_t(y)=\frac{\mathbb{E}^{t,y}X_t-y}{\sigma_t^2}=\frac{\mu_t(y)-y}{\sigma_t^2}. \tag{8}$$

我们用 $F_t(y)$ 表示得分函数的估计值。若上下文已明确，我们会省略 F_t 和 $F^*{}_t$ 中的时间索引 t。此外，我们考虑以下过程：

- (y_t)：由真实的概率流常微分方程定义的过程

$$\mathrm{d}y_t=F_t^*(y_t)\,\mathrm{d}t\,,\qquad y_0\sim q_0$$

- $(\hat{y}_t)$：由我们的 PICARD 算法给出的过程1 对于 $t\in[t_i,\ t_{i+1}]$，对于所有 $1\le i\le n$，a s

$$\hat{y}_t\triangleq x_{t_i}+\int_0^{t-t_i}\sum_{i=1}^{D}F_{c_i}(X_{:,i})\phi_i(s)\,\mathrm{d}s,$$

其中 $X_{:,\,i}$ 定义于算法 1中。

- $(\tilde{y}_t)$：在区间 $[t_i,\ t_{i+1}]$内，该过程由从算法迭代点 $\tilde{y}(t_i)$ 出发的真实概率流常微分方程给出，对所有 $1\le i\le n-1$均成立，如 2 中所定义。具体而言，$\tilde{y}(0)=y_{t_i}$，并且对于所有 $t\in[t_i,\ t_{i+1}]$:

$$\frac{d}{dt}\tilde{y}(t)=F_t^*(\tilde{y}(t)).$$

在证明过程中，对于固定的 $1\le i\le n-1$，我们将区间 $[t_i,\ t_{i+1}]$记为 $[t_0,t_0+h]$。对于任意曲线 $z(t)$：$t\in[t_0,t_0+h]\to\mathbb{R}^d$，在计算中我们还使用以下上确界范数：

$$\|z\|_{[t_0,t_0+h]}=\sup_{t_0\le t\le t_0+h}\|z(t)\|.$$

C 导数计算

为了证明得分函数 $F^*{}_t(y_t)$是关于时间 t 的低次函数，我们的方法是界定 $F^*{}_t(y_t)$关于时间的高阶导数。有了这样的界，我们便能利用下文回顾的泰勒余项定理，证明$F^*{}_t(y_t)$与其 k 阶泰勒近似足够接近。

[1]请注意，先前研究中的一些惯例是对时间进行参数化，使得分布 $q_{\ }=$分布 q_0，，但我们不采用这种做法，因为我们的选择能让证明中的符号说明更加简洁。

事实 C.1. 对于函数 $g:[t_0,t_0+h]\to\mathbb{R}$，定义次数为 k 的低次近似 $P_{\le k}g(s)$ 如下：

$$P_{\le k}g(s)=\sum_{r=0}^{k-1}\frac{d^r g(s)}{ds^r}\Big|_{s=t_0}\frac{(s-t_0)^r}{r!}.$$

根据泰勒余项定理，我们有

$$\left\|g-P_{\le k}g\right\|_{[t_0,t_0+h]}\le\frac{\max_{s\in[t_0,t_0+h]}\left|\frac{d^k g(s)}{ds^k}\right|}{k!}h^k.$$

有鉴于此，我们现在开始证明正文中所述的得分函数高阶导数的与维数无关的界（引理 3.1-3.3）。

引理 3.2中的界主要基于引理 C.1和 3.1。引理 C.1包含两部分，分别在下文的引理 C.2、C.3 中给出，用于计算后验均值 $\mu_t(y)$关于时间（t）和空间（y）变量的导数。在引理 3.3 中，我们将这些要素结合起来，得到得分函数高阶导数关于时间的界。

C.1 引理3.1的证明

为方便起见，我们在此重述引理 3.1：

引理 3.1. 设 $y_t=X_t+\sigma_t\xi$，其中 $X_t=e^{-(T-t)}\bar{X}$，且对于 $\bar{X}\sim q_{\text{pre}}$和 $\xi\sim\gamma^d$。我们用符号$\mathbb{E}^{t,\,y_t}$表示给定 $y_t=X_t+\sigma_t\xi$时 X_t 的条件期望，并用 $q^{t,\,y_t}$表示该后验分布的密度。定义后验均值 $\mu_t(y)\triangleq\mathbb{E}^{t,\,y}X_t$。

概率流常微分方程的向量场的时间导数可计算为

$$\begin{aligned}\partial_t(y_t+\nabla\log q_t(y_t))=&\left(y_t+\frac{\mathbb{E}^{t,y_t}X_t-y_t}{\sigma_t^2}\right)\\&+\frac{1}{\sigma_t^2}\mathbb{E}^{t,y_t}X_t+\frac{1-\sigma_t^2}{2\sigma_t^6}\mathbb{E}^{t,y_t}\left(\langle y_t,X_t\rangle+\|X_t\|^2\right)\left(X_t-X_t'\right)+\frac{4}{\sigma_t^4}\mathbb{E}^{t,y_t}\left(\langle y_t,X_t\rangle-\|X\|^2\right)\left(X_t-X_t'\right)\\&+\frac{1}{\sigma_t^2}\left(\frac{1}{\sigma_t^2}-1\right)y_t-\frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t}X_t+\frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t}\left(2\left\langle y_t,\frac{X_t''}{\sigma_t^2}\right\rangle-\langle X_t,y_t\rangle-\frac{1}{\sigma_t^2}\left\langle X_t,X_t''\right\rangle\right)(X_t-X_t').\end{aligned}$$

这里，X_t，X_t'，X_t'' 表示从后验分布中独立同分布抽取的样本，其期望算子为 $\mathbb{E}^{t,\,y_t}$。

为完成此推导，我们需要以下计算，其证明见下文第 C.2 节。

引理 C.1。 按照引理 *3.1*中的符号说明，对于任意 $v\in\mathbb{R}^d$，

$$\begin{aligned}\partial_t\mu_t(y)&=\mu_t(y)+\frac{1-\sigma_t^2}{2\sigma_t^4}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle+\|X_t\|^2\right)\left(X_t-X_t'\right)+\frac{4}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle-\|X_t\|^2\right)\left(X_t-X_t'\right),\\D(\mu_t(y))[v]&=\frac{1}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y-X_t,v\rangle\right)(X_t-X_t'),\end{aligned}$$

其中，右侧的期望是针对给定 y 时从后验分布中抽取的独立同分布副本 X，X'而言的。

证明　　引理 3.1 由引理 C.1得出，即结合引理 C.2和 C.3：

证明。 我们可以写成

$$\partial_t(y_t+\nabla\ln q_t(y_t))=y_t+\nabla\ln q_t(y_t)+\partial_t\nabla\ln q_t(y_t)$$
$$=y_t+\nabla\ln q_t(y_t)+\partial_t\nabla\ln q_t(y)\Big|_{y=y_t}+\nabla^2 q_t(y_t)(y_t+\nabla\ln q_t(y_t)).$$

现在利用引理 C.3 以及 $v=y_t+\nabla\ln$ 分布 $q_t(y_t)$，以及（8），可得

$$\nabla^2 q_t(y_t)(y_t+\nabla\ln q_t(y_t))=\frac{1}{\sigma_t^2}\nabla\left(\frac{\mathbb{E}^{t,y}X_t-y}{\sigma_t^2}\right)\Big|_{y=y_t}[y_t+\nabla\ln q_t(y_t)]$$
$$=\frac{1}{\sigma_t^2}\left(-y_t+\frac{y_t-\mathbb{E}^{t,y_t}X_t}{\sigma_t^2}+\mathbb{E}^{t,y_t}\frac{1}{\sigma_t^2}\left(\langle y_t-X_t,y_t+\nabla\ln q_t(y_t)\rangle-\|y_t\|^2+\frac{\|y_t\|^2}{\sigma_t^2}\right)(X_t-X_t')\right)$$
$$=\frac{1}{\sigma_t^2}(\frac{1}{\sigma_t^2}-1)y_t-\frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t}X_t+\frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t}\left(2\left\langle y_t,\frac{X_t''}{\sigma_t^2}\right\rangle-\langle X_t,y_t\rangle-\frac{1}{\sigma_t^2}\langle X_t,X_t''\rangle\right)(X_t-X_t').$$

结合引理 C.3，我们得到

$$\partial_t(y_t+\nabla\ln q_t(y_t))=\partial_t y_t+\partial_t(\nabla\ln q_t(y))\Big|_{y=y_t}+\nabla^2 q_t(y_t)(y_t+\nabla\ln q_t(y_t))$$
$$=\partial_t y_t+\frac{1}{\sigma_t^2}\partial_t(\mathbb{E}^{t,y}X_t)\Big|_{y=y_t}+\nabla^2 q_t(y_t)(y_t+\nabla\ln q_t(y_t))$$
$$=\left(y_t+\frac{\mathbb{E}^{t,y_t}X_t-y_t}{\sigma_t^2}\right)$$
$$+\frac{1}{\sigma_t^2}\mathbb{E}^{t,y}X_t+\frac{1-\sigma_t^2}{2\sigma_t^6}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle+\|X_t\|^2\right)(X_t-X_t')+\frac{4}{\sigma_t^4}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle-\|X_t\|^2\right)(X_t-X_t')$$
$$+\frac{1}{\sigma_t^2}(\frac{1}{\sigma_t^2}-1)y_t-\frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t}X_t+\frac{1}{\sigma_t^4}\mathbb{E}^{t,y_t}\left(2\left\langle y_t,\frac{X_t''}{\sigma_t^2}\right\rangle-\langle X_t,y_t\rangle-\frac{1}{\sigma_t^2}\langle X_t,X_t''\rangle\right)(X_t-X_t'). \quad (9)$$

其中 最后一项是对分母中的 σ_t 求导所得。 □

C.2 引理的证明C.1

下面我们证明引理 C.1 的两个部分，分别计算后验均值 $\mu_t(y)=\mathbb{E}^{t,y}X_t$关于时间和空间的导数。

引理 C.2。 给定任意时间 t 和 $y\in\mathbb{R}^d$，后验分布 $q^{t,y}$ 的期望的时间导数可计算为

$$\partial_t\mathbb{E}^{t,y}X_t=\mathbb{E}^{t,y}X_t+\frac{1-\sigma_t^2}{2\sigma_t^4}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle+\|X_t\|^2\right)(X_t-X_t')+\frac{4}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle-\|X_t\|^2\right)(X_t-X_t'),$$

其中 X,X'是从后验分布 $q^{t,y}$ 中独立抽取的样本。

证明。有

$$\mathbb{E}^{t,y}X_t=\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}x_t q_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}$$

其中随机变量为 $\xi\sim\gamma^d$、$X_t=\sqrt{1-\sigma_t^2}\bar{X}_t$ 和 $\bar{X}\sim$分布 q_{pre}，其中 $y=X_t+\sigma_t\xi$，我们分别用 $\bar{x}$ 和 $x_t=\sqrt{1-\sigma_t^2}\bar{x}$ 表示 $\bar{X}$ 和 X_t 的取值。于是

$$\begin{aligned}\partial_t\mathbb{E}^{t,y}X_t=&\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}x_tq_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}\\&+4\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}x_t\frac{\langle y-x_t,x_t\rangle}{2\sigma_t^2}q_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}-4\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}x_tq_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}\frac{\langle y-x_t,x_t\rangle}{2\sigma_t^2}q_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}\\&+\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}x_t\frac{\|y-x_t\|^2}{2\sigma_t^4}(1-\sigma_t^2)q_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}-\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}x_tq_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}\frac{\|y-x_t\|^2}{2\sigma_t^4}(1-\sigma_t^2)q_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}},\end{aligned}$$

其中，第一行是对不在指数中的x_t 求 ∂_t 导数得到的（注意，$x_t=\sqrt{1-\sigma_t^2}\bar{x}=e^{-(T-t)}\bar{x}$是$t$的函数）；第二行是对指数中的$x_t$求导数得到的；最后一行是对指数中的 $1/\sigma^2{}_t$ 求导数得到的。注意，在这些情况下，先对分子再对分母求导，会分别在第二行和第三行中产生x_t与〈$y-x_t$、x_t〉 与 $\|y-x_t\|^2$之间的协方差项:

$$\begin{aligned}\partial_t\mathbb{E}^{t,y}X_t&=\mathbb{E}^{t,y}X_t+\text{Cov}\left(\frac{(1-\sigma_t^2)\|y-X_t\|^2}{2\sigma_t^4},X_t\right)+4\,\text{Cov}\left(\frac{\langle y-X_t,X_t\rangle}{\sigma_t^2},X_t\right)\\&=\mathbb{E}^{t,y}X_t+\frac{1-\sigma_t^2}{2\sigma_t^4}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle+\|X_t\|^2\right)\left(X_t-\mathbb{E}^{t,y}X_t\right)+\frac{4}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle-\|X_t\|^2\right)\left(X_t-\mathbb{E}^{t,y}X_t\right)\\&=\mathbb{E}^{t,y}X_t+\frac{1-\sigma_t^2}{2\sigma_t^4}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle+\|X_t\|^2\right)\left(X_t-X_t'\right)+\frac{4}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y,X_t\rangle-\|X_t\|^2\right)\left(X_t-X_t'\right),\end{aligned}$$

其中，$X_{\prime t}$是X_t的一个独立副本。 □

Lemma C.3. 后验均值关于 y 的方向导数为

$$D(\mathbb{E}^{t,y}X_t)[v]=\frac{1}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y-X_t,v\rangle\right)(X_t-X_t')=\frac{1}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y-X_t,v\rangle+\frac{\|y\|^2}{\sigma_t^2}-\|y\|^2\right)(X_t-X_t').$$

证明。我们有

$$\begin{aligned}D\left(\mathbb{E}^{t,y}X_t\right)[v]=&-\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}\left\langle\frac{y-x_t}{\sigma_t^2},v\right\rangle x_tq_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}+\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}x_tq_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}\frac{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}\left\langle\frac{y-x_t}{\sigma_t^2},v\right\rangle q_{\text{pre}}(\bar{x})d\bar{x}}{\int e^{-\frac{\|y-x_t\|^2}{2\sigma_t^2}}q_{\text{pre}}(\bar{x})d\bar{x}}\\&=\text{Cov}\left(\left\langle\frac{y-X_t}{\sigma_t^2},v\right\rangle,X_t\right)\\&=\mathbb{E}^{t,y}\frac{1}{\sigma_t^2}\left(\langle y-X_t,v\rangle+\frac{\|y\|^2}{\sigma_t^2}-\|y\|^2\right)(X_t-\mathbb{E}^{t,y}X_t)\\&=\frac{1}{\sigma_t^2}\mathbb{E}^{t,y}\left(\langle y-X_t,v\rangle+\frac{\|y\|^2}{\sigma_t^2}-\|y\|^2\right)(X_t-X_t'),\end{aligned}$$

其中我们利用了这样一个事实：在计算协方差时，y是常数，因此加上 $\|y\|^2_{\sigma_t^2} - \|y\|^2$不会改变协方差的值。 □

C.3 引理证明3.2

本节证明引理 3.2 中关于得分函数高阶导数的界，为方便起见，在此重述：

引理 3.2。 反向常微分方程向量场的第 k 阶导数可展开为

$$\begin{aligned}&\partial_t^k(y_t + \nabla \ln q_t(y_t))\\ &= \sum_{r_1,r_2,r_3,r_4} \sum_{\mathbf{i}\in[k]^{r_3},\mathbf{k}\in[k]^{r_4},\mathbf{l}\in[k]^{r_4}} \frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \mathbb{E}^{t,y_t} \prod_{j=1}^{r_3} \left\langle y_t, X^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle X^{(\mathbf{k}_j)}, X^{(\mathbf{l}_j)} \right\rangle (a_{\mathbf{i},\mathbf{k},\mathbf{l}} X + b_{\mathbf{i},\mathbf{k},\mathbf{l}} y_t)\end{aligned}$$

其中 $r_1, r_2, r_3 \le k$ 且 $r_2 \le 4k$，系数 $(a_{\mathbf{i},\mathbf{k},\mathbf{l}}, b_{\mathbf{i},\mathbf{k},\mathbf{l}})$ 满足 $\sum_{\mathbf{i}\in[k]^{r_3},\mathbf{k}\in[k]^{r_4},\mathbf{l}\in[k]^{r_4}} |a_{\mathbf{i},\mathbf{k},\mathbf{l}}| + |b_{\mathbf{i},\mathbf{k},\mathbf{l}}| \le (74k)^k$。这里，随机变量 $X^{(i)}$ 独立地服从后验分布 $q^{t,y}$，而 i、k、l 是任意的索引元组。

证明。 为了控制高阶导数 $\partial_k t(y_t + \nabla \ln$ 分布 $q_t(y_t))$，引理 3.1 促使我们分析对以下形式的项求导后会发生什么

$$\frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \mathbb{E}^{t,y_t} \prod_{j=1}^{r_3} \left\langle y_t, X_t^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle X_t^{(\mathbf{k}_j)}, X_t^{(\mathbf{l}_j)} \right\rangle X_t, \tag{10}$$

$$\frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \mathbb{E}^{t,y_t} \prod_{j=1}^{r_3} \left\langle y_t, X_t^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle X_t^{(\mathbf{k}_j)}, X_t^{(\mathbf{l}_j)} \right\rangle y_t, \tag{11}$$

对于某些整数 $r_1, r_2, r_3, r_4, r \le k$；这类项在取 k 阶导数后自然会出现。特别地，我们有

$$\begin{aligned}\partial_t \mathbb{E}^{t,y_t} X_t &= \partial_t \mathbb{E}^{t,y} X_t \Big|_{y=y_t} + D(\mathbb{E}^{t,y} X_t)[y_t + \nabla \ln q_t(y_t)]\\ &= \mathbb{E}^{t,y} X_t + \frac{1-\sigma_t^2}{2\sigma_t^4} \mathbb{E}^{t,y} \left(\langle y, X_t\rangle + \|X_t\|^2\right)(X_t - X_t') + \frac{4}{\sigma_t^2} \mathbb{E}^{t,y} \left(\langle y, X_t\rangle - \|X_t\|^2\right)(X_t - X_t')\\ &+ \frac{1}{\sigma_t^2}\left(\frac{1}{\sigma_t^2} - 1\right) y_t - \frac{1}{\sigma_t^4} \mathbb{E}^{t,y_t} X_t + \frac{1}{\sigma_t^4} \mathbb{E}^{t,y_t} \left(2\left\langle y_t, \frac{X_t''}{\sigma_t^2}\right\rangle - \langle X_t, y_t\rangle - \frac{1}{\sigma_t^2}\left\langle X_t, X_t''\right\rangle\right)(X_t - X_t'),\end{aligned}$$

and

$$\partial_t(y_t) = y_t + \nabla \ln q_t(y_t) = y_t + \frac{\mathbb{E}^{t,y_t} X_t - y_t}{\sigma_t^2}.$$

这里的关键是跟踪公式（11）中出现的X_t和y_t的总数。我们用 $\kappa = 2(r_3 + r_4)$表示这个数。首先要注意，每进行一步求导，我们要么对公式（11）中的某个y_t求导，要么对某个X_t求导。无论哪种情况，X_t和y_t的总数最多增加 2，而分母中 $\sigma^2{}_t$ 的幂次，即r_2，，最多增加 4。因此，$\kappa \le 2k$。对 σ_t 求导：

$$\partial_t \frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} = -r_2 \frac{(1-\sigma_t^2)^{r_1+1}}{\sigma_t^{r_2+2}} + \frac{-2r_1(1-\sigma_t^2)^{r_1-1}}{\sigma_t^{r_2-1}}, \tag{12}$$

这会使 r_1 最多增加 1，使 r_2 最多增加 2。因此，$r_1 \le k$ 且 $r_2 \le 4k$。另一方面，对 y_t 或 X_t 求导最多产生 $34 = 1+2+4\times2\times2+2\times2+2+1+4\times2$ 个新项（将系数为 4 的项按四次计数），而根据公式（12）对 σ_t 求导最多产生 $r_2 + 2r_1 \le 6k$ 个新项，这意味着总共对 k 次求导后，最多产生 $(2\times34+6)k = 74k$ 个新项。因此，总体而言，对 k 次求导后最多产生 $(74k)^k$ 个项。

□

C.4 引理证明3.3

我们正式陈述并证明关于导数的上界 $\partial_{kt}(y_t + \nabla \ln$分布 $q_t(y_t))$，见引理 3.3，为方便起见，在此重述如下：

引理 3.3。 若分布 q 满足假设 *1*，$(y_t)_t$是方程 (2)的解，且 $y_0 \sim$分布 q_T，则沿概率流常微分方程的向量场的第 k^{th} 阶导数有如下界

$$\left\|\partial_t^k (y_t + \nabla \log q_t(y_t))\right\|_{p,\infty} \lesssim R\left(\frac{74k}{\sigma_t^2}\right)^k \left(\frac{R}{\sigma_t} + (kp)^{1/2}\right)^{2k}$$

证明。 证明可直接由引理 3.2结合下文的引理 C.4 得出。 □

在引理C.4 中，我们对引理 p-范数的各项进行界定，这些项出现在引理 3.2中。

引理 C.4。我们有

$$\left\|\frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \prod_{j=1}^{r_3} \left\langle y_t, X_t^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle X_t^{(\mathbf{k}_j)}, X_t^{(\mathbf{l}_j)} \right\rangle X_t \right\|_{p,\infty} \vee \left\|\frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \prod_{j=1}^{r_3} \left\langle y, X_t^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle X_t^{(\mathbf{k}_j)}, X_t^{(\mathbf{l}_j)} \right\rangle y_t \right\|_{p,\infty}$$
$$\le R\left(\frac{1}{\sigma_t^2}\right)^k \left(\frac{R}{\sigma_t} + (kp)^{1/2}\right)^{2k},$$

其中，$\|v\|_{p,\infty}$ 表示向量 v 的 p 阶矩的无穷范数。

证明。 首先，由于q_0 的支撑集半径有界于R，我们有

$$\left\|\frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \prod_{j=1}^{r_3} \left\langle y_t, X_t^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle X_t^{(\mathbf{k}_j)}, X_t^{(\mathbf{l}_j)} \right\rangle X_t \right\|_{p,\infty} \le \frac{1}{\sigma_t^{r_2}} \left\|\prod_{j=1}^{r_3} \left\langle y_t, X_t^{(\mathbf{i}_j)} \right\rangle\right\|_p R^{2r_4+1}.$$

现在利用 [GKL24], 中引理 4.4 的推导，可得

$$\left\|\prod_{j=1}^{r_3} \left\langle y_t, X_t^{(\mathbf{i}_j)} \right\rangle\right\|_p \le R^{2r_3}\left(1 + \frac{(r_3 p)^{1/2}\sigma_t}{R}\right)^{r_3}.$$

将此代入上式并使用 $R \ge 1$:

$$\begin{aligned}
\text{LHS} \le \frac{R^{2r_3+2r_4+1}}{\sigma_t^{r_2}}\left(1 + \frac{(r_3 p)^{1/2}\sigma_t}{R}\right)^{r_3} &\le R\left(\frac{R}{\sigma_t^2}\right)^{2k}\left(1 + \frac{(kp)^{1/2}\sigma_t}{R}\right)^{k} \\
&\le R\left(\frac{1}{\sigma_t}\right)^{2k}\left(\frac{R}{\sigma_t}\right)^{2k}\left(1 + \frac{(kp)^{1/2}\sigma_t}{R}\right)^{2k} \\
&\le R\frac{1}{\sigma_t^{2k}}\left(\frac{R}{\sigma_t} + (kp)^{1/2}\right)^{2k}
\end{aligned}$$

对于另一项同理，使用柯西-施瓦茨不等式

$$\begin{aligned}\left\|\frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}}\prod_{j=1}^{r_3}\left\langle y, X_t^{(\mathbf{i}_j)}\right\rangle\prod_{j=1}^{r_4}\left\langle X_t^{(\mathbf{k}_j)}, X_t^{(\mathbf{l}_j)}\right\rangle y\right\|_{p,\infty} &\le \frac{1}{\sigma_t^{r_2}}\left\|\prod_{j=1}^{r_3}\left\langle y, X_t^{(\mathbf{i}_j)}\right\rangle\right\|_{2p} R^{2r_4}\|y\|_{2,\infty}\\ &\le \frac{1}{\sigma_t^{r_2-1}}R^{2r_3}\left(1+\frac{(2r_3p)^{1/2}\sigma_t}{R}\right)^{r_3}R^{2r_4}\\ &\le \frac{1}{\sigma_t^{2k}}\left(\frac{R}{\sigma_t}+(kp)^{1/2}\right)^{2k}.\end{aligned}$$

□

D 引理 3.4 的证明3.4

我们通过将引理 3.3 中的不等式从 y 转换到附近的 $\tilde{y}$ 来证明定理 *3.5*$_t$，方法是对 y 条件下 $\tilde{y}$的分布与 $\tilde{y}$ 之间进行耦合。注意，引理 3.3 中的论证并不能直接应用于 y，因为它不像 $\tilde{y}$ 那样遵循真实的反向过程。

首先，回顾一下 $\bar{X} \sim$分布 q_{pre}以及 $\xi \sim \gamma^d$和 $y_t = X_t + \sigma_t\xi, X_t = \sqrt{1-\sigma_t^2}\bar{X}$。现在，对于每个 $\varepsilon_1 > 0$，考虑事件

$$\mathcal{A}_1 := \left\{\|y_t - X_t\| \le \sqrt{d} + \sqrt{\ln(1/\varepsilon_1)}\right\}, \tag{13}$$

并注意到 $\mathbb{P}(\mathcal{A}_1) \ge 1-\varepsilon_1$，因为 $\sigma_t \le 1$。

引理 3.4 (耦合)。 设 $\tilde{y}$ 是一个随机变量，使得 $\|y - \tilde{y}\| \le \delta$ 对于 $\delta \le \frac{1}{6(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})}$，那么，给定 $T - t \ge 1$，在方程 中定义的事件 (13)下，我们有

$$\mathrm{TV}(q^{t,y}, q^{t,\tilde{y}}) \le 8\delta(\sqrt{d} + \sqrt{\ln(1/\varepsilon_1)}). \tag{4}$$

特别地，式 (4) 成立的概率至少为 $1-\varepsilon_1$。

证明。 由于 $T - t \ge 1$，我们有

$$\sigma_t^2 = 1 - e^{-2(T-t)} \ge 1 - e^{-2} \ge 0.5.$$

令 $p(x、y)$表示 $(X_t、y_t)$的联合分布，其中$y_t = X_t + \xi$、$X_t = \sqrt{1-\sigma_t^2}\bar{X}$、$\bar{X} \sim$分布 q_{pre}，以及 $\xi \sim N(0, \sigma^2{}_tI)$。注意，分布 $q^{t,y} = p(\cdot\,|y)$。我们有

$$\begin{aligned}\left|\ln\left(\frac{p(y\mid x)}{p(\tilde{y}\mid x)}\right)\right| &= \frac{1}{2\sigma_t^2}\left|\|y-x\|^2 - \|\tilde{y}-x\|^2\right|\\ &\le \left|\|y-x\|^2 - \|\tilde{y}-x\|^2\right|\\ &\le 2\|y-\tilde{y}\|\,\|y-x\| + \|y-\tilde{y}\|^2\\ &\le 2\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}) + \delta^2\\ &\le 3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}),\end{aligned}$$

并且根据对 δ 的假设界，我们有 $2\delta(\sqrt{d} + \sqrt{\ln(1/\varepsilon_1)}) + \delta^2 \le 0.5$，这意味着

$$(1-3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}))p(y\mid x) \le p(\tilde{y}\mid x) \le p(y\mid x)(1+3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})).$$

两边同乘 $p(x)$，得到

$$(1-3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}))p(y\mid x)p(x)\le p(\tilde{y}\mid x)p(x)\le p(y\mid x)p(x)(1+3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})).$$

但根据贝叶斯法则，很容易看出这意味着

$$(1-3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}))^2 q^{t,y}(x)\le q^{t,\tilde{y}}(x)\le q^{t,y}(x)(1+3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}))^2.$$

鉴于 $(1+3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}))^2\le 1+8\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})$，根据对 δ 的假设界限，我们得出

$$\mathrm{TV}(q^{t,y},q^{t,\tilde{y}})\le 8\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}).$$ □

E 定理证明 3.5

这里我们证明，对于曲线 $\tilde{y}$ 而非 y，其向量场的导数也满足类似的界，其中y 和 $\tilde{y}$ 定义见第 B节。本节中，我们始终假设 $\tilde{y}_t$ 和 y_t 在窗口 δ 中相距不超过 $[t_0,t_0+h]$:

$$\|y_t-\tilde{y}_t\|_{[t_0,t_0+h]}\le\delta. \tag{14}$$

稍后在定理 G.1的证明中，我们将通过归纳法证明可以假设等式（14）成立。为使表达清晰，在不致引起歧义时，我们将省略变量的 t 索引。首先，我们重述定理 3.5，并给出基于本节其余内容的简短证明。随后，我们将详细证明这些中间引理。

定理 3.5. 设 $\delta\in(0,1)$。若分布 q 满足假设 *1*，且 $(\tilde{y}_t)_t$ 由以下方式给出：从一个与分布 q 在 W_2 距离上相差不超过 δ的分布进行初始化0，并在连续时间下运行概率流常微分方程。则对于 δ 足够小的情形（其严格刻画见引理 *E.6*），以至少 $1-\delta_2$的概率（关于初始化的随机性），概率流常微分方程中向量场沿轨迹的第 k^{th} 阶导数在 $\tilde{y}_t$ 处有界，界为

$$\left\|\partial_t^k(\tilde{y}_t+\nabla\log q_t(\tilde{y}_t))\right\|_\infty\lesssim R\left(\frac{k}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t}\right)^{2k}+(k\log(74kd/\delta_2))^{2k}.$$

证明。为简洁起见，本节余下部分省略 y_t、X_t 中的时间下标 t，并记为 $y=X+\sigma_t\xi$，其中 $\tilde{X}\sim$ qpre，$X=\sqrt{1-\sigma_t^2}\tilde{X}$ 且 $\xi\sim\gamma^d$。令 $p(.|y)$ 表示在给定 y 的条件下 X 的分布，并令 $\tilde{X}$ 为条件分布 $p(.|\tilde{y})$ 的一个样本。对于指标 i，令 $X^{(i)}$ 和 $\tilde{X}^{(i)}$ 分别为 X 和 $\tilde{X}$ 的独立同分布样本。现在应用引理 E.6，并取 $\delta_1=\frac{\delta_2}{60k)^k}$，得到

$$\mathbb{P}\left(\left|\mathbb{E}_{\{\tilde{X}^{(i)}\}}\frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}}\prod_{j=1}^{r_3}\left\langle\tilde{y},\tilde{X}^{(i_j)}\right\rangle\prod_{j=1}^{r_4}\left\langle\tilde{X}^{(k_j)},\tilde{X}^{(l_j)}\right\rangle X\right|_\infty\ge eR\left(\frac{1}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t}\right)^{2k}+e(k\log(74kd/\delta_2))^k\right)\le\delta_2.$$

对所有项应用并集界，可得其概率至少为 $1-\delta_2$

$$\left\|\partial_t^k(\tilde{y}_t+\nabla\log q_t(\tilde{y}_t))\right\|_\infty\le e(R+\delta+\sigma_t+1)\left(\frac{kh}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t}\right)^{2k}+e(k\log(74kd/\delta_2))^{2k}.$$ □

接下来，我们通过对项 $\left\|\prod_{j=1}^{r_3}\left\langle y, X^{(\mathbf{i}_j)}\right\rangle\right\|_p$ 求得一个悲观界，从而为所有由索引 $\mathbf{i}$ 构成的有序元组 $\mathbb{E}\left|\prod_{j=1}^{r_3}\left\langle \tilde{y}, X^{(\mathbf{i}_j)}\right\rangle\right|$ 推导出一个悲观界，其中每个 $X^{(i)}$ 都表示从后验分布 $q^{t,y}$ 中抽取的一个独立样本。

引理 E.1. 对于所有索引元组$\mathbf{i}$，我们有

$$\mathbb{E}\left|\prod_{j=1}^{r_3}\left\langle \tilde{y}, X^{(\mathbf{i}_j)}\right\rangle\right| \le \left(R^2\left(1+\frac{(r_3 p)^{1/2}\sigma_t}{R}\right)+\delta R\right)^{r_3}$$

证明。 根据柯西-施瓦茨不等式，

$$\begin{aligned}
\left\|\prod_{j=1}^{r_3}\left\langle \tilde{y}, X^{(\mathbf{i}_j)}\right\rangle\right\|_p &\le \sum_{S\subseteq[k]}^{r_3}\left\|\prod_{j\in S}\left\langle y, X^{(\mathbf{i}_j)}\right\rangle\right\|_p\left\|\prod_{j\notin S}\left\langle \tilde{y}-y, X^{(\mathbf{i}_j)}\right\rangle\right\|_p \\
&\le \sum_{\ell=1}^{r_3}\binom{r_3}{\ell}\left(R^2\left(1+\frac{(r_3 p)^{1/2}\sigma_t}{R}\right)\right)^{\ell}(\delta R)^{r_3-\ell} \\
&\le \left(R^2\left(1+\frac{(r_3 p)^{1/2}\sigma_t}{R}\right)+\delta R\right)^{r_3}. \qquad \square
\end{aligned}$$

引理 E.2. 对于 $\nu \triangleq \left(3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})\right)^{1/p}$，对于所有索引元组 $\mathbf{i}$，我们有

$$\left\|1\left\{\mathcal{A}_1\right\}\prod_{j=1}^{r_3}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right\|_p \le \left(R^2\left(1+\frac{(2r_3 p)^{1/2}\sigma_t}{R}\right)+\delta R+\nu R\left(\sqrt{2pd}+R+\delta\right)\right)^{r_3}. \tag{15}$$

证明。首先，请注意我们有以下界限：

$$\begin{aligned}
\left\|\prod_{j=1}^{r_3}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right\|_p^p &= \mathbb{E}\prod_{j=1}^{r_3}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle^p \\
&\le \mathbb{E}\prod_{j=1}^{r_3}\|\tilde{y}\|^p\left\|\tilde{X}^{(\mathbf{i}_j)}\right\|^p \\
&\le \mathbb{E}\prod_{j=1}^{r_3}\left(\|y-\tilde{y}\|+\|\xi\|+\left\|X^{\mathbf{i}_j}\right\|\right)^p\left\|\tilde{X}^{(\mathbf{i}_j)}\right\|^p \\
&\le \mathbb{E}\prod_{j=1}^{r_3}(\delta+\|\xi\|+R)^p R^p \\
&\le \mathbb{E}\sum_{\ell=1}^{r_3 p}\binom{r_3 p}{\ell}\|\xi\|^{\ell}(R+\delta)^{r_3 p-\ell}R^{r_3 p} \\
&\le \sum_{\ell=1}^{r_3 p}\binom{r_3 p}{\ell}\left(\sqrt{\ell d}\right)^{\ell}(R+\delta)^{r_3 p-\ell}R^{r_3 p} \\
&= \left(R\left(\sqrt{pd}+R+\delta\right)\right)^{r_3 p}, \qquad (16)
\end{aligned}$$

这意味着

$$\left\|\prod_{j=1}^{r_3}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right\|_p \le \left(R\left(\sqrt{pd}+R+\delta\right)\right)^{r_3}.$$

因此，应用引理 3.4 并结合 $\varepsilon_1=\delta$，在事件 $\mathcal{A}_1$ 上，我们有

$$TV(q^{t,y}, q^{t,\tilde{y}}) \le 3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}),$$

这意味着，对于所有 m,

$$TV(X^{(m)}, \tilde{X}^{(m)}) \le 3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}).$$

因此，在给定 y 的条件下，$X^{(m)}$ 与 $\tilde{X}^{(m)}$ 之间的最优耦合满足

$$\mathbb{P}\left(X^{(m)}=\tilde{X}^{(m)}\right) \ge 1-3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}) \tag{17}$$

现在有了这个耦合，对于子集 $\mathcal{U}\subset[r_3]$,，令 $\mathcal{A}_2=\mathcal{A}_2(\mathcal{U})$ 表示以下事件：对于所有$j\in\mathcal{U}$，$X^{(\mathbf{i}_j)}=\tilde{X}^{(\mathbf{i}_j)}$。那么我们可以写为

$$\begin{aligned}
&\mathbb{E}\left(1\left\{\mathcal{A}_1\cap\mathcal{A}_2(\mathcal{U})\right\}\prod_{j=1}^{r_3}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right)^p \\
&=\mathbb{E}\left(1\left\{\mathcal{A}_1\cap\mathcal{A}_2(\mathcal{U})\right\}\prod_{j\in\mathcal{U}}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right)^p\left(1\left\{\mathcal{A}_1\cap\mathcal{A}_2(\mathcal{U})\right\}\prod_{j\notin\mathcal{U}}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right)^p. \\
&\le\sqrt{\mathbb{E}\left(1\left\{\mathcal{A}_1\cap\mathcal{A}_2(\mathcal{U})\right\}\prod_{j\in\mathcal{U}}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right)^{2p}}\sqrt{\mathbb{E}\left(1\left\{\mathcal{A}_1\cap\mathcal{A}_2(\mathcal{U})\right\}\prod_{j\notin\mathcal{U}}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right)^{2p}}
\end{aligned} \tag{18}$$

但另一方面，由方程（17）以及耦合的独立性可知，我们有

$$\mathbb{P}(\mathcal{A}_2(\mathcal{U})|\mathcal{A}_1) \le \left(3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})\right)^{r_3-|\mathcal{U}|}.$$

因此，与（16）中的推导类似

$$\begin{aligned}
&\mathbb{E}\left(1\left\{\mathcal{A}_1\cap\mathcal{A}_2(\mathcal{U})\right\}\prod_{j\notin\mathcal{U}}\left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)}\right\rangle\right)^p \\
&\le \mathbb{E}1\left\{\mathcal{A}_1\right\}\sum_{\ell=1}^{(r_3-|\mathcal{U}|)p}\binom{(r_3-|\mathcal{U}|)p}{\ell}\|\xi\|^\ell(R+\delta)^{(r_3-|\mathcal{U}|)p-\ell}R^{(r_3-|\mathcal{U}|)p}\mathbb{P}_{|y}(\mathcal{A}_2(\mathcal{U})|\mathcal{A}_1) \\
&\le \mathbb{P}_{|y}(\mathcal{A}_2(\mathcal{U})|\mathcal{A}_1)\sum_{\ell=1}^{(r_3-|\mathcal{U}|)p}\binom{(r_3-|\mathcal{U}|)p}{\ell}\|\xi\|^\ell(R+\delta)^{(r_3-|\mathcal{U}|)p-\ell}R^{(r_3-|\mathcal{U}|)p} \\
&\le \left(3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})\right)^{|\mathcal{U}|}\left(R\left(\sqrt{pd}+R+\delta\right)\right)^{(r_3-|\mathcal{U}|)p} \\
&= \nu^{p(r_3-|\mathcal{U}|)}\left(R\left(\sqrt{pd}+R+\delta\right)\right)^{(r_3-|\mathcal{U}|)p},
\end{aligned}$$

其中 $\nu \triangleq \left(3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})\right)^{1/p}$。这意味着

$$\left\| 1\left\{\mathcal{A}_1 \cap \mathcal{A}_2(\mathcal{U})\right\} \prod_{j\notin\mathcal{U}}^{r_3} \left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \right\rangle \right\|_p \le \left(3\delta(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})\right)^{1/p} \left(R\left(\sqrt{pd}+R+\delta\right)\right)^{r_3-|\mathcal{U}|}. \tag{19}$$

对于第一项，我们可以利用耦合，将 $\tilde{X}^{(\mathbf{i}_j)}$ 替换为 $X^{(\mathbf{i}_j)}$，然后使用引理 E.1：

$$\begin{aligned}
\mathbb{E}\left(1\left\{\mathcal{A}_1 \cap \mathcal{A}_2(\mathcal{U})\right\} \prod_{j\in\mathcal{U}} \left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \right\rangle \right)^p &= \mathbb{E}\left(1\left\{\mathcal{A}_1 \cap \mathcal{A}_2(\mathcal{U})\right\} \prod_{j\in\mathcal{U}} \left\langle \tilde{y}, X^{(\mathbf{i}_j)} \right\rangle \right)^p \\
&\le \mathbb{E}\left(\prod_{j\in\mathcal{U}} \left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \right\rangle \right)^p \\
&\le \left(R^2\left(1+\frac{(r_3 p)^{1/2}\sigma_t}{R}\right)+\delta R \right)^{p|\mathcal{U}|}.
\end{aligned} \tag{20}$$

结合公式（20）和（19），并代入公式（18）：

$$\begin{aligned}
&\mathbb{E}\left(1\left\{\mathcal{A}_1\right\} \prod_{j=1}^{r_3} \left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \right\rangle \right)^p \\
&= \sum_{\mathcal{U}\subseteq[r_3]} \mathbb{E}\left(1\left\{\mathcal{A}_1 \cap \mathcal{A}_2(\mathcal{U})\right\} \prod_{j=1}^{r_3} \left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \right\rangle \right)^p \\
&\le \sum_{\mathcal{U}\subseteq[r_3]} \left(R^2\left(1+\frac{(2r_3 p)^{1/2}\sigma_t}{R}\right)+\delta R \right)^{p|\mathcal{U}|} \left(\nu R\left(\sqrt{2pd}+R+\delta\right)\right)^{(r_3-|\mathcal{U}|)p} \\
&\le \left(R^2\left(1+\frac{(2r_3 p)^{1/2}\sigma_t}{R}\right)+\delta R+\nu R\left(\sqrt{2pd}+R+\delta\right) \right)^{r_3 p}.
\end{aligned}$$

□

引理 E.3（耦合变量的副本界）。 假设我们有

$$\delta \le R \wedge \frac{1}{\left(\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}\right)(\sqrt{2pd}/R+1)^p}.$$

那么，对于第B节中定义的 $\tilde{y}$，令 $\tilde{X}^{(i)}$ 为来自后验分布 $q^{t,\tilde{y}}$ 的独立样本。那么，对于满足 $r_2 \le 2(r_3+r_4) \le 2k$ 的整数 r_2, r_3, r_4，在此前定义的事件 $\mathcal{A}_1$ 下，对于所有索引元组 $\mathbf{i},\mathbf{k},\mathbf{l}$，

$$\begin{aligned}
\left\| 1\left\{\mathcal{A}_1\right\} \frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \prod_{j=1}^{r_3} \left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle \tilde{X}^{(\mathbf{k}_j)}, \tilde{X}^{(\mathbf{l}_j)} \right\rangle \tilde{X} \right\|_{p,\infty} &\lesssim \frac{R}{\sigma_t^{2k}} \left(\frac{R}{\sigma_t} + (kp)^{1/2} \right)^{2k}, \\
\left\| 1\left\{\mathcal{A}_1\right\} \frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \prod_{j=1}^{r_3} \left\langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \right\rangle \prod_{j=1}^{r_4} \left\langle \tilde{X}^{(\mathbf{k}_j)}, \tilde{X}^{(\mathbf{l}_j)} \right\rangle \tilde{y} \right\|_{p,\infty} &\lesssim \frac{R+1}{\sigma_t^{2k}} \left(\frac{R}{\sigma_t} + (kp)^{1/2} \right)^{2k},
\end{aligned}$$

其中，$\|v\|_{p,\infty}$ 表示向量 v 的 p 阶矩的无穷范数。

证明。注意，给定关于 δ 的条件，则引理（15）中的界E.2 里，第二项被第一项支配，这意味着

$$\left\| 1\{\mathcal{A}_1\} \prod_{j=1}^{r_3} \langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \rangle \right\|_p \lesssim \left(R^2 \left(1 + \frac{(r_3 p)^{1/2} \sigma_t}{R} \right) + \delta R \right)^{r_3} \lesssim \left(R^2 \left(1 + \frac{(r_3 p)^{1/2} \sigma_t}{R} \right) \right)^{r_3}.$$

因此 ，由$r_2 \le 2(r_3 + r_4)$，并利用分布 q_0 的支撑集半径有界，为 R ：

$$\begin{aligned}
&\left\| 1\{\mathcal{A}_1\} \frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \prod_{j=1}^{r_3} \langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \rangle \prod_{j=1}^{r_4} \langle \tilde{X}^{(\mathbf{k}_j)}, \tilde{X}^{(\mathbf{l}_j)} \rangle \tilde{X} \right\|_{p,\infty} \\
&\le \frac{1}{\sigma_t^{r_2}} \left\| 1\{\mathcal{A}_1\} \prod_{j=1}^{r_3} \langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \rangle \right\|_p R^{2r_4+1} \\
&\lesssim R \left(\frac{1}{\sigma_t^2} \right)^{r_3+r_4} \left(\frac{R^2}{\sigma_t^2} \right)^{r_4} \left(\frac{R^2}{\sigma_t^2} + \frac{R}{\sigma_t} (r_3 p)^{1/2} \right)^{r_3} \\
&\le \frac{R}{\sigma_t^{2k}} \left(\frac{R}{\sigma_t} \right)^{2r_4} \left(\left(\frac{R}{\sigma_t} + (r_3 p)^{1/2} \right)^2 \right)^{r_3} \\
&\le \frac{R}{\sigma_t^{2k}} \left(\frac{R}{\sigma_t} + (kp)^{1/2} \right)^{2k}.
\end{aligned}$$

其中我们利用了$r_3 + r_4 \le k$这一事实。对另一项也类似，利用柯西-施瓦茨不等式

$$\begin{aligned}
&\left\| 1\{\mathcal{A}_1\} \frac{(1-\sigma_t^2)^{r_1}}{\sigma_t^{r_2}} \prod_{j=1}^{r_3} \langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \rangle \prod_{j=1}^{r_4} \langle \tilde{X}^{(\mathbf{k}_j)}, \tilde{X}^{(\mathbf{l}_j)} \rangle y \right\|_{p,\infty} \\
&\le \frac{1}{\sigma_t^{r_2}} \left\| \prod_{j=1}^{r_3} \langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \rangle \right\|_{2p} R^{2r_4} \left(\|\tilde{y} - y\|_{2,\infty} + \|y\|_{2,\infty} \right) \\
&\le \frac{1}{\sigma_t^{r_2}} \left\| \prod_{j=1}^{r_3} \langle \tilde{y}, \tilde{X}^{(\mathbf{i}_j)} \rangle \right\|_{2p} R^{2r_4} (R + \sigma_t + \delta) \\
&\lesssim \frac{R+1}{\sigma_t^{2k}} \left(\frac{R}{\sigma_t} + (kp)^{1/2} \right)^{2k}.
\end{aligned}$$

□

引理 E.4 （高阶导数的高概率控制——固定时间）。 在引理*E.3* 的设定下，令

$$g(t) = \tilde{y}_t + \nabla \log q_t(\tilde{y}_t),$$

其中， $\tilde{}y(t)$ 由公式(7)定义。于是，对于常数 c_1 和任意坐标 i,

$$\mathbb{P}\left(\left| \partial_t^k g_i(t) \right| \ge c_1 (60k)^k (R+1) \left(\frac{1}{\sigma_t^2} \right)^k \left(\frac{R}{\sigma_t} + \sqrt{\ln(1/\varepsilon_1)} \right)^{2k} \right) \le \varepsilon_1.$$

证明。根据引理E.3 和3.2，存在常数c_1 使得

$$\begin{aligned}\left\|\partial_t^k g_i(t)\right\|_p &= \left\|\partial_t^k(y_{t,i} + \partial_i \log q_t(y_t))\right\|_p \\ &\le (c_1/e)(60k)^k(R+1)\left(\frac{1}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t} + (kp)^{1/2}\right)^{2k}.\end{aligned}$$

此外，对于固定的$j \le d$，令 $\mathcal{A}_3 \subseteq \mathcal{A}_1$ 为 $\mathcal{A}_1$ 中的事件，其中

$$\left|\partial_t^k g_i(t)\right| \ge c_1(60k)^k(R+1)\left(\frac{1}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t} + \sqrt{\ln(1/\varepsilon_1)}\right)^{2k}.$$

对 y 取期望

$$\begin{aligned}&\mathbb{E}_y\mathbb{E}_{\{\tilde{X}^{(i)}\}}\left|\partial_t^k g_i(t)\right|^p \\ &\ge \mathbb{P}(\mathcal{A}_3)\left(c_1(60k)^k(R+1)\left(\frac{1}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t} + \sqrt{\ln(1/\varepsilon_1)}\right)^{2k}\right)^p.\end{aligned}$$

现在令 $p = \ln(1/\varepsilon_1)/k$，得到

$$\mathbb{P}(\mathcal{A}_3) \le \left(\frac{\frac{R}{\sigma_t} + (kp)^{1/2}}{e\frac{R}{\sigma_t} + e\sqrt{\ln(1/\varepsilon_1)}}\right)^{2kp} = \left(\frac{1}{e}\right)^{2\ln(1/\varepsilon_1)} \le \varepsilon_1. \qquad \square$$

引理 E.5. 在与引理 *E.4*相同的设定下，对于

$$\delta \le R \wedge \frac{1}{\left(\sqrt{d} + \sqrt{\ln(1/\varepsilon_1)}\right)\left(\sqrt{2\ln(1/\varepsilon)d}/(R\sqrt{k}) + 1\right)^{\ln(1/\varepsilon_1)/k}},$$

对于某个常数 c_1，有

$$\mathbb{P}\left(\sup_{\tilde{y}:\|\tilde{y}-y\|\le\delta}\left|\partial_t^k g_i(t)\Big|_{s=0}\right| \ge c_1(60k)^k(R+1)\left(\frac{1}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t} + \sqrt{\ln(1/\varepsilon_1)}\right)^{2k}\right) \le \varepsilon_1.$$

证明。 注意， $\tilde{y}$ 可以是邻域 $\|y - \tilde{y}\| \le \delta$中$y$ 的任意映射。因此，我们取 $\tilde{y}$ 为

$$\tilde{y}(y) \triangleq \operatorname{argmax}_{\tilde{y}:\|\tilde{y}-y\|\le\delta}\left|\partial_t^k g_i(t)\Big|_{t=0}\right|.$$

因此，结论由引理 E.4 得出。 $\square$

引理 E.6。 设两条曲线，即真实曲线和算法曲线之间的距离，上界为

$$\tilde{\delta} \triangleq R \wedge \frac{1}{\left(\sqrt{d} + \sqrt{\ln(1/\varepsilon_1)}\right)\left(\sqrt{2\ln(1/\varepsilon_1)d}/(R\sqrt{k}) + 1\right)^{\ln(1/\varepsilon_1)/k}}, \tag{21}$$

即 $\sup_{t\in[t_0,t_0+h]}\|y(t) - \tilde{y}(t)\| \le \delta$，其中 $\tilde{y}(t)$ 由引理 *E.5* 定义，而 y 的定义与 $y(0) = y$ 类似。于是

$$\mathbb{P}\left(\sup_{t\in[t_0,t_0+h]}\left|\partial_t^k g_i(t)\Big|_{t=0}\right| \ge C_1\right) \le 2\varepsilon_1.$$

其中

$$C_1 \triangleq c_1(60k)^k(R+1)\left(\frac{1}{\sigma_t^2}\right)^k\left(\frac{R}{\sigma_t}+\sqrt{\ln\left(\frac{R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}}{\varepsilon_1}\right)}\right)^{2k}.$$

证明。 假设我们取一个覆盖 $\mathcal{C}$区间 $[t_0, t_0+h]$，其精度为 δ 。然后，利用引理 E.5 并应用并集界，特别是由于 2δ 满足该引理的条件，我们得到（利用$h \le 1$）

$$\mathbb{P}\left(\sup_{\tilde{y}:\|\tilde{y}-y\|\le 2\tilde{\delta}, y\in\mathcal{C}}\left|\partial_t^k g_i(t)\Big|_{t=0}\right| \ge C_1\right) \le \varepsilon_1. \tag{22}$$

另一方面，对于每个 $s \in [t_0, t_0+h]$，，若我们假设 $\bar{s} \in \mathcal{C}$ 是覆盖中距离s最近的元素，那么根据引理 I.3 以及覆盖精度的选取：

$$\|y(\bar{t})-y(t)\| \le \tilde{\delta}.$$

结合这一点、曲线之间的距离以及三角不等式，我们得到

$$\|\tilde{y}(t)-y(\bar{t})\| \le 2\tilde{\delta}.$$

因此，我们证明了曲线 $\tilde{y}$ 在时间区间 $[t_0, t_0+h]$ 上的每个点都接近覆盖 $\mathcal{C}$中的某个点。结合方程（22），可知其成立的概率至少为 $1-2\varepsilon_1$

$$\mathbb{P}\left(\sup_{t\in[t_0,t_0+h]}\left|\partial_t^k g_i(t)\Big|_{t=0}\right| \ge C_1\right) \le 2\varepsilon_1. \qquad \square$$

F 命题的证明3.6

本节利用我们建立的框架，通过皮卡迭代对概率流常微分方程上的得分函数进行低次数近似。

F.1 预备知识

这里我们回顾结合多项式逼近的皮卡迭代的基本设置，该设置由 [LSV18]提出。

定义 F.1。 给定向量场 $F(x): \mathbb{R}^d \to \mathbb{R}^d$，受皮卡迭代启发并遵循 *[LSV18]*,定义算子 $T(x)$，它作用于曲线 $x(t): [t_0, t_0+h] \to \mathbb{R}^d$ 如下

$$T(x)(t) = x(t_0) + \int_{t_0}^{t_0+h} F(x(s))ds.$$

此外，假设我们有一组光滑一维函数的基 $\{\phi_j\}_{j=1}^D$，以及满足 $1 \le \forall i, j \le D,\ \phi_j(c_i) = 0\ if\ i \ne j\ and\ \phi_i(c_i) = 1$ 的点 $\{c_j\}_{j=1}^D$。那么，定义与基 $\{\phi_j\}_{j=1}^D$ 对应的 T 算子近似 T_ϕ 为

$$T_\phi(x)(t) = \int_{t_0}^{t_0+h}\sum_{j=1}^{D} F(x(c_j))\phi_j(s)ds. \tag{23}$$

特别地，我们选取 ϕ_j，使其构成次数小于 D 的一元多项式的基；我们将它们选作点 cj 对应的拉格朗日乘子多项式。这样，式（23）中的积分便可以用闭式求出。

定义 F.2. 我们称基 $\phi = \{\phi_j\}_{j=1}^{D}$ 有界，当且仅当

$$\sum_j \left| \int_{t_0}^{t_0+h} \phi_j(s)\, ds \right| \le \gamma_\phi h.$$

接下来，我们推广 [LSV18] 的方法，并证明 $T\phi$ 关于上确界范数的一个重要利普希茨性质 $\|\cdot\|$，当向量场 F 接近一个利普希茨向量场时。

引理 F.3. 假设基 ϕ 有界，且向量场 F_s 是 $\tilde{L}$ 利普希茨的。那么，对于任意 x、$y \in C([t_0, t_0+h], \mathbb{R}^d)$ 且 $x(t_0) = y(t_0)$，

$$\begin{aligned} \|T_\phi(x) - T_\phi(y)\|_{[t_0,t_0+h]} &\le \tilde{L}\|x-y\|_{[t_0,t_0+h]}\gamma_\phi h, \\ \|T_\phi^{\circ\ell}(x) - T_\phi^{\circ\ell}(y)\|_{[t_0,t_0+h]} &\le \left(\tilde{L}\gamma_\phi h\right)^\ell \|x-y\|_{[t_0,t_0+h]}. \end{aligned} \tag{24}$$

证明。F 对于每个 $1 \le s \le h$，利用我们对得分函数的估计所具有的 $\tilde{L}$ 利普希茨性质 ，

$$\begin{aligned} \|F_t(x(t)) - F_t(y(t))\| &\le \tilde{L}\|x(t) - y(t)\| \\ &\le \tilde{L}\|x-y\|_{[t_0,t_0+h]}. \end{aligned}$$

因此

$$\begin{aligned} \|T_\phi(x) - T_\phi(y)\|_{[t_0,t_0+h]} &= \sup_{t_0 \le h' \le t_0+h} \left\| \int_{t_0}^{t_0+h'} \sum_{j=1}^{D} \left(F_{c_j}(x(c_j)) - F_{c_j}(y(c_j))\right)\phi_j(s) ds \right\| \\ &\le \sup_{1 \le j \le D} \left\| F_{c_j}(x(c_j)) - F_{c_j}(y(c_j)) \right\|_\infty \sum_j \left| \int_{t_0}^{t_0+h} \phi_j(s) ds \right| \\ &\le \tilde{L}\|x-y\|_{[t_0,t_0+h]}\gamma_\phi h. \end{aligned}$$

（24）中的第二行24源于应用（ℓ 次。 □

定义 F.4。 （低次向量场）令 y_t 为常微分方程 $\dot{y}_t = F^*t(y_t)$ 的解。如果向量场 F^* 沿y是低次的，则称其接受以下低次近似：

$$\left\| F^* \circ y - P_{\le D}(F^* \circ y) \right\|_{[t_0,t_0+h]} \le \varepsilon_{\mathrm{ld}},$$

其中，曲线 $P_{\le D}(F^* \circ y)$ 是 $F^* \circ y$ 的近似，其坐标是时间变量 $t \in [t_0, t_0+h]$ 中的次数至多为 D 的多项式。

引理 F.5。 对于 $\dot{y}_t = F^*t(y_t)$，已知 F^* 是根据定义 *F.4* 的低次向量场，我们有

$$\left\| T_\phi^{\circ m} y - y \right\| \le \frac{\left(\tilde{L}\gamma_\phi h\right)^m - 1}{\left(\tilde{L}\gamma_\phi h\right) - 1} \left(\varepsilon_{\mathrm{ld}} + \max_{j=1}^{D} \left\| F(y(c_j)) - F^*(y(c_j)) \right\|\right)\left(1 + \gamma_\phi\right)h.$$

证明。 我们可以写作 $T_\phi(y) = y^{t_0} + S_\phi(F \circ y)$，其中对于任意曲线 $z \in \mathcal{C}([t_0, t_0 + h], \mathbb{R}^d)$ 定义

$$S_\phi(z)(.) = \sum_{j=1}^{D} z(c_j) \int_{t_0}^{\cdot} \phi_j(s) ds.$$

注意，由于 $P_{\leq D}(F^* \circ y)$ 的次数至多为 D，且 ϕ_j 是切比雪夫点 $(c_i)^D{}_{j=1}$处的拉格朗日乘子多项式，因此

$$P_{\leq D}(F^* \circ y) = \sum_{j=1}^{D} P_{\leq D}(F^* \circ y)(c_j) \phi_j,$$

根据 $S_\phi(z)$ 的定义，这意味着

$$\int_{t_0}^{\cdot} P_{\leq D}(F^* \circ y)(s) ds = S_\phi\left(P_{\leq D}(F^* \circ y)\right).$$

现在将其与引理 F.3 结合，并使用低次数定义 F.4

$$\begin{aligned}
&\left\| S_\phi(F \circ y) - \int_{t_0}^{\cdot} P_{\leq D}(F^* \circ y)(s) ds \right\|_{[t_0, t_0+h]} \\
&= \left\| S_\phi(F \circ y) - S_\phi\left(P_{\leq D}(F^* \circ y)\right) \right\|_{[t_0, t_0+h]} \\
&= \left\| \sum_{j=1}^{D} \left(F \circ y \Big|_{c_j} - P_{\leq D}(F^* \circ y) \Big|_{c_j} \right) \left(y_{t_0} + \int_{t_0}^{\cdot} \phi_j(s) ds \right) \right\|_{[t_0, t_0+h]} \\
&\leq (\max_{j=1}^{D} \left\| F^*(y(c_j)) - P_{\leq D}\left(F^*(y(c_j))\right) \right\| + \left\| F(y(c_j)) - F^*(y(c_j)) \right\|) \sum_{j=1}^{D} \left\| y_{t_0} + \int_{t_0}^{\cdot} \phi_j(s) ds \right\|_{[t_0, t_0+h]} \\
&\leq (\varepsilon_{\mathsf{ld}} + \max_{j=1}^{D} \left\| F(y(c_j)) - F^*(y(c_j)) \right\|) \sum_{j=1}^{D} \left\| y_{t_0} + \int_{t_0}^{\cdot} \phi_j(s) ds \right\|_{[t_0, t_0+h]} \\
&\leq (\varepsilon_{\mathsf{ld}} + \max_{j=1}^{D} \left\| F(y(c_j)) - F^*(y(c_j)) \right\|) \gamma_\phi h. \qquad (25)
\end{aligned}$$

但根据$\varepsilon_{\mathsf{ld}}$的定义，

$$\left\| P_{\leq D}(F^* \circ y) - F^* \circ y \right\|_{[t_0, t_0+h]} \leq \varepsilon_{\mathsf{ld}},$$

由恒等式 $y - y_{t_0} = \int_{t_0}^{\cdot} F \circ y(s) ds$ 可得

$$\begin{aligned}
\left\| \int_{t_0}^{\cdot} P_{\leq D}(F^* \circ y)(s) ds - (y - y_{t_0}) \right\|_{[t_0, t_0+h]} &= \left\| \int_{t_0}^{\cdot} P_{\leq D}(F^* \circ y)(s) ds - \int_{t_0}^{\cdot} F^* \circ y(s) ds \right\|_{[t_0, t_0+h]} \\
&\leq \varepsilon_{\mathsf{ld}} h \qquad (26)
\end{aligned}$$

结合式（25）和式（26）

$$\begin{aligned}\left\|T_\phi(y)-y\right\|_{[t_0,t_0+h]} &\le \left\|T_\phi(y)-y_{y_{t_0}}-\int_{t_0}^{\cdot} P_{\le D}\left(F^*\circ y\right)ds\right\|_{[t_0,t_0+h]}\\ &\quad+\left\|\int_{t_0}^{\cdot} P_{\le D}\left(F^*\circ y\right)ds-\left(y-y_{t_0}\right)\right\|_{[t_0,t_0+h]}\\ &\le \left(\varepsilon_{\mathsf{ld}}+\max_{j=1}^{D}\left\|F(y(c_j))-F^*(y(c_j))\right\|\right)\left(1+\gamma_\phi\right)h.\end{aligned}$$

现在应用引理 F.3 次 -1 次

$$\begin{aligned}\left\|T_\phi^{\circ i}(y)-T_\phi^{\circ(i-1)}(y)\right\| &\le \left(\tilde{L}\gamma_\phi h\right)^{i-1}\left\|T_\phi(y)-y\right\|_{[0,h]}\\ &\le \left(\tilde{L}\gamma_\phi h\right)^{i-1}\left(\varepsilon_{\mathsf{ld}}+\max_{j=1}^{D}\left\|F(y(c_j))-F^*(y(c_j))\right\|\right)\left(1+\gamma_\phi\right)h \qquad (27)\end{aligned}$$

对 27) 求和，其中 $i=1,\dots,m$:

$$\begin{aligned}\left\|T_\phi^{\circ m}(y)-y\right\| &\le \sum_{i=1}^{m}\left\|T_\phi^{\circ i}y-T_\phi^{\circ(i-1)}y\right\|\\ &\le \frac{\left(\tilde{L}\gamma_\phi h\right)^m-1}{\left(\tilde{L}\gamma_\phi h\right)-1}\left(\varepsilon_{\mathsf{ld}}+\max_{j=1}^{D}\left\|F(y(c_j))-F^*(y(c_j))\right\|\right)\left(1+\gamma_\phi\right)h.\end{aligned}$$

□

推论 F.6 (近似皮卡迭代的影响)。 对于精确的概率流常微分方程 $\dot{y}_t=F^*{}_t(y_t)$，给定 F^* 是根据定义 *F.4* 定义的低次向量场，且给定 $\tilde{L}\gamma_\phi h\le\frac{1}{2}$，对于任意曲线 $x\in\mathcal{C}([t_0,t_0+h],\ \mathbb{R}^d)$有

$$\begin{aligned}&\left\|T_\phi^{\circ m}(y)-y\right\|_{[t_0,t_0+h]}\le 2\left(\varepsilon_{\mathsf{ld}}+\max_{j=1}^{D}\left\|F(y(c_j))-F^*(y(c_j))\right\|\right)\left(1+\gamma_\phi\right)h\\ &\left\|T_\phi^{\circ m}(x)-T_\phi^{\circ m}(y)\right\|_{[t_0,t_0+h]}\le\frac{1}{2^m}\left\|x-y\right\|_{[t_0,t_0+h]}.\end{aligned}$$

证明。 这可直接由引理F.3和 F.5得出。 □

引理 F.7。 回顾第 *B节中 F 和 F的定义。对于初始点 $\left\|\bar{y}_{t_0}-y_{t_0}\right\|\le\varepsilon_p$，定义 y_s，$\tilde{y}_s$，以及 $\bar{y}_s$ 对于 $t_0\le s\le t_0+h$ 如下

$$\begin{aligned}\dot{y}_s &= F_s^*(y_s),\\ \dot{\tilde{y}}_s &= F_s^*(\tilde{y}_s),\\ \bar{y}(s) &= \bar{y}_{t_0}+(s-t_0)F_{t_0}(\bar{y}_{t_0}).\end{aligned}$$

那么，在曲线 $\tilde{y}$ 满足低次数假设 *F.4*、时间区间 $[t_0,t_0+h]$ 内，并选取步长 $h=O\left(\frac{1}{1+R_2}\right)$ 且假设 $\tilde{L}\le\frac{1}{2}$ 和 $\sigma_t\ge 1$的条件下，

$$\begin{aligned}\left\|y-T_\phi^{\circ m}(\bar{y})\right\|_{[t_0,t_0+h]} &\le 2\varepsilon_p(1+2(1+\gamma_\phi)h)+\left(\varepsilon_{\mathsf{ld}}+\max_{j=1}^{D}\left\|F(y(c_j))-F^*(y(c_j))\right\|\right)\left(1+\gamma_\phi\right)h\\ &\quad+\frac{h}{2^m}\left(\tilde{L}\varepsilon_p+\varepsilon_{\mathsf{err}}+\sqrt{d}R(R^2+\ln(1/\varepsilon_1))+(2+4R^2)\varepsilon_p\right).\end{aligned}$$

证明。首先，由于我们有假设 $T - t_0 - h \geq 1$，因此 $L_s \leq (2+4R^2), \forall s \in [t_0, t_0+h]$。因此，由于我们知道切比雪夫基的 $\gamma_\phi = O(1)$，假设 $h \leq O\left(\frac{1}{1+R_2}\right)$ 满足推论 3.6 关于 h 的前提条件。另一方面，由引理 I.2，并再次利用条件 $h = O\left(\frac{1}{1+R_2}\right)$，我们得到 $\|y - \tilde{y}\|_{[t_0,t_0+h]} \leq 2\varepsilon_p$。现在将其与推论 3.6 结合：

$$\begin{aligned}
\left\|y - T_\phi^{\circ m}(\bar{y})\right\|_{[t_0,t_0+h]} &\leq \|y-\tilde{y}\|_{[t_0,t_0+h]} + \left\|\tilde{y} - T_\phi^{\circ m}(\tilde{y})\right\|_{[t_0,t_0+h]} + \left\|T_\phi^{\circ m}(\bar{y}) - T_\phi^{\circ m}(\tilde{y})\right\|_{[t_0,t_0+h]} \\
&\leq 2\varepsilon_p + 2(\varepsilon_{\mathsf{ld}} + \max_{j=1}^{D}\left\|F(\tilde{y}(c_j)) - F^*(\tilde{y}(c_j))\right\|)\left(1+\gamma_\phi\right)h + \frac{1}{2^m}\|\bar{y}-\tilde{y}\|_{[t_0,t_0+h]} \\
&\leq 2\varepsilon_p + 2(\varepsilon_{\mathsf{ld}} + 2\|y-\tilde{y}\|_{[t_0,t_0+h]} + \max_{j=1}^{D}\left\|F(y(c_j)) - F^*(y(c_j))\right\|)\left(1+\gamma_\phi\right)h \\
&\quad + \frac{1}{2^m}\|\bar{y}-\tilde{y}\|_{[t_0,t_0+h]} \\
&\leq 2\varepsilon_p(1+2(1+\gamma_\phi)h) + 2(\varepsilon_{\mathsf{ld}} + \max_{j=1}^{D}\left\|F(y(c_j)) - F^*(y(c_j))\right\|)\left(1+\gamma_\phi\right)h \\
&\quad + \frac{1}{2^m}\|\bar{y}-\tilde{y}\|_{[t_0,t_0+h]} \qquad (28)
\end{aligned}$$

但根据引理 I.3，以至少 $1-\varepsilon_1$ 的概率，对所有 $t_0 \leq s \leq t_0+h$

$$\left\|F_s^*(\tilde{y}_s)\right\| \leq 6\left(R + \sqrt{d} + \sqrt{\ln(1/\varepsilon_1)}\right). \qquad (29)$$

因此，利用引理 I.1 和I.2，

$$\begin{aligned}
\frac{d}{ds}\|\bar{y}_s - \tilde{y}_s\| &\leq \frac{\langle F_{t_0}(\tilde{y}_{t_0}) - F_s^*(\tilde{y}_s), \bar{y}_t - y_t\rangle}{\|\bar{y}_t - y_t\|} \\
&\leq \left\|F_{t_0}(\tilde{y}_{t_0}) - F_t^*(\tilde{y}_t)\right\| \\
&\leq \left\|F_{t_0}(\tilde{y}_{t_0}) - F_{t_0}(y_{t_0})\right\| + \left\|F_{t_0}(y_{t_0}) - F_{t_0}^*(y_{t_0})\right\| \\
&\quad + \left\|F_{t_0}^*(y_{t_0}) - F_t^*(y_t)\right\| + \left\|F_t^*(y_t) - F_t^*(\tilde{y}_t)\right\| \\
&\leq \tilde{L}\varepsilon_p + \varepsilon_{\mathsf{err}} + \sqrt{d}R(R^2 + \ln(1/\varepsilon_1)) + (2+4R^2)e^{\frac{5R^2}{\sigma_t^2}h}\varepsilon_p.
\end{aligned}$$

这意味着

$$\|\bar{y} - \tilde{y}\|_{[t_0,t_0+h]} \leq h\left(\left\|F_{t_0}(\tilde{y}_{t_0}) - F_{t_0}^*(\tilde{y}_{t_0})\right\| + 12\left(R + \sqrt{d} + \sqrt{\ln(1/\varepsilon_1)}\right)\right).$$

将其代回式（28）

$$\begin{aligned}
\left\|y - T_\phi^{\circ m}(\bar{y})\right\| &\leq 2\varepsilon_p(1+2(1+\gamma_\phi)h) + (\varepsilon_{\mathsf{ld}} + \max_{j=1}^{D}\left\|F(y(c_j)) - F^*(y(c_j))\right\|)\left(1+\gamma_\phi\right)h \\
&\quad + \frac{h}{2^m}\left(\tilde{L}\varepsilon_p + \varepsilon_{\mathsf{err}} + \sqrt{d}R(R^2 + \ln(1/\varepsilon_1)) + (2+4R^2)e^{\frac{5R^2}{\sigma_t^2}h}\varepsilon_p\right).
\end{aligned}$$

利用假设 $h = O\left(\frac{1}{1+R_2}\right)$ 即可完成证明。 □

G 定理 3.7 的证明

本节将所有内容汇总起来，给出我们的主要端到端结果，即如何从目标分布中采样，使其与某个在全变差意义下 $\tilde{O}(\varepsilon_{\mathsf{err}}R^2)$-接近、且在 W_2 意义下与真实分布 q $\tilde{O}(\varepsilon_{\mathsf{err}})$-接近的分布。

定理 *G.1*（定理 *3.7* 的形式化版本）。对于满足界 $\varepsilon err \le \varepsilon_1 / ln(d/R)$ 和 $\varepsilon err \le \big((R/\sigma)\wedge(1/\sqrt{d})\big)/(ln(\sigma/R)+ln(d))^2$ 的误差参数 $\varepsilon_1, \varepsilon err > 0$，给定步长 $h = \gamma k(1+(R/\sigma)^2)(ln(1/\varepsilon_1)+ln(d))$，其中 γ 为足够小的常数，且 $k \triangleq ln(1/\varepsilon err)$ 是多项式逼近的次数，以至少

$$1-O\Big(\varepsilon_1 T(R/\sigma)^2\ln(1/\varepsilon_{\mathrm{err}})(\ln(1/\varepsilon)+\ln(d))\Big),$$

对于 $T = \ln(R/\sigma)+\ln(d)+\ln(1/\varepsilon_{\mathrm{err}})$，算法 *2*输出一个样本，其分布与目标测度 q 的 *2*-瓦瑟斯坦距离不超过误差

$$O(Tk\varepsilon_{\mathrm{err}}\log(1/\varepsilon_1)) = O\Big((\ln(R/\sigma)+\ln(d)+\ln(1/\varepsilon_{\mathrm{err}}))\ln(1/\varepsilon_{\mathrm{err}})\ln(k/\varepsilon_1)\varepsilon_{\mathrm{err}}\Big),$$

in

$$O\Big(T(R/\sigma)^2\ln(1/\varepsilon_{\mathrm{err}})(\ln(1/\varepsilon_1)+\ln(d))\Big)$$

轮数和 $m = \ln(R/\sigma)+\ln(d)+\ln\ln(1/\varepsilon_1)+\ln(1/\varepsilon_{\mathrm{err}})+\ln(\tilde{L})$ 每轮的皮卡迭代次数。

证明。 我们可以先缩放分布，从缩放后的分布中采样，然后再缩放回去。注意，这个缩放过程不会改变比值 R/σ。因此，不失一般性，我们假设$\sigma_0 = \sigma = 1$，因为我们的界只依赖于量 R/σ。我们将前向过程运行至时间

$$T := \Theta(\ln(R)+\ln(d)+\ln(1/\varepsilon_{\mathrm{err}})). \tag{30}$$

现在我们通过归纳证明，对于所有 i，都有 $\left\|y - T^{\circ m}_{\phi}(\bar{y})\right\|_{[t^{(i)},\, t^{(i)}+h]} = O((t^{(i)}+h)k\varepsilon_{\mathrm{err}}\log(k/\varepsilon_1))$。首先，注意到我们可以将目标分布 q_{pre} 与 γ^d 之间的 Wasserstein距离界定为

$$W_2(q_{\mathrm{pre}},\gamma^d) \le W_2(q_{\mathrm{pre}},\delta_{\{0\}}) + W_2(\delta_{\{0\}},\gamma^d),$$

其中 $\delta_{\{0\}}$ 是位于原点的点质量。但

$$W_2(q_{\mathrm{pre}},\delta_{\{0\}}) = \mathbb{E}_{q_{\mathrm{base}}}\|Y\|^2 \le R,$$
$$W_2(\gamma^d,\delta_{\{0\}}) = \sqrt{d}.$$

因此

$$W_2(q_{\mathrm{pre}},\gamma^d) \le R+\sqrt{d}.$$

现在根据 Wass erstein 收缩性质以及奥恩斯坦–乌伦贝克过程，并根据 T 在（30）中的选取，可得

$$W_2(q_T,\gamma^d) \le \left(R+\sqrt{d}\right)e^{-T} = O(\varepsilon_{\mathrm{err}}).$$

为了证明区间 $[t^{(i)}$ 的归纳步骤，其中 $t^{(i)}+h] = [t_0, t_0+h],$，根据归纳假设，我们知道对于所有之前的区间 $(t^{(j)},\ t^{(j)}+h)$，当 $j < i$ 时：

$$\left\|y - T^{\circ m}_{\phi}(\bar{y})\right\|_{[t^{(j)},t^{(j)}+h]} \le O((t^{(j)}+h)k\varepsilon_{\mathrm{err}}\log(k/\varepsilon_1)),$$

根据算法迭代项的定义，即 $\hat{y} = T^{\circ m}_{\phi}(\bar{y})$ 意味着 s

$$\|y-\hat{y}\|_{[0,t_0]} \le O(t_0 k\varepsilon_{\mathrm{err}}\log(k/\varepsilon_1)).$$

利用引理 I.2，进一步可知，在给定 h 的选择下，

$$\|\tilde{y} - y\|_{[t_0, t_0+h]} = O(t_0 k \varepsilon_{\text{err}} \log(k/\varepsilon_1)).$$

因此，我们可以在引理 $\varepsilon_p = O(t_0 k \varepsilon_{\text{err}} \log(k/\varepsilon_1))$ 中使用 F.7。

现在，给定假设 $\varepsilon_{\text{err}} \le (1/\ln(d/R))\varepsilon_1$ 以及 $\varepsilon_{\text{err}} \le \left(R \wedge (1/\sqrt{d})\right)/(\ln(1/\varepsilon_1) + \ln(d))^2$，并且由于 $\sigma_t \ge \sigma_0 \ge 1$ 容易检查出 $\delta = O(t_0 k \varepsilon_{\text{err}} \log(k/\varepsilon_1))$ 满足引理 E.6 的假设。因此，我们可以应用引理 E.6，其中

$$k = \ln(1/\varepsilon_{\text{err}}),$$

利用以下事实：$\sigma_t \ge \frac{1}{2}$ 以及 $R \ge 1$ 存在常数 c_2, c_0，使得对于区间 $[t t_0 + h]$：t，th

$$\mathbb{P}\left(\sup_{s \in [t_0, t_0+h]} \left| \partial_s^k g_i^{(\tilde{y}(s))}(s) \right|_{s=0} \right| \ge \left(c_2 k^2 R^2 (\ln(1/\varepsilon_1) + \ln(d)) \right)^k \Bigg) \le 2\varepsilon_1.$$

此外，将其与事实 C.1 结合，并选取步长

$$h = \frac{\gamma}{k(1+R^2)(\ln(1/\varepsilon_1) + \ln(d))}$$

对于足够小的常数 γ，我们以至少 $1 - 2\varepsilon_1$ 的概率得到

$$\left\| g - P_{\le k} g \right\|_{[t_0, t_0+h]} \le \left(h c_2 e k R^2 (\ln(1/\varepsilon_1) + \ln(d)) \right)^k \le \varepsilon_{\text{err}}. \tag{31}$$

因此，现在我们可以使用引理 F.7 并结合

$$m := \ln(R) + \ln(d) + \ln\ln(1/\varepsilon_1) + \ln(1/\varepsilon_{\text{err}}) + \ln(\tilde{L})$$

次皮卡迭代；利用切比雪夫基 $\gamma_\phi = O(1)$ 的事实 $(\phi_0, \ldots, \phi_k)$ 对于第 2.3 节中定义的多项式，令 $\varepsilon_{\text{ld}} = \varepsilon_{\text{err}}$ 根据方程 (31) 和假设 F.4，再利用假设 2，可得

$$\begin{aligned}
\left\| y - T_\phi^{\circ m}(\bar{y}) \right\|_{[t_0, t_0+h]} &= O\left(\varepsilon_p + h \left(\varepsilon_{\text{err}} + \sum_{j=1}^{k} \left\| F(y(c_j)) - F^*(y(c_j)) \right\| \right) \right) \\
&\quad + \frac{h}{2^m} O\left(\tilde{L} \varepsilon_p + \varepsilon_{\text{err}} + \sqrt{d} R (R^2 + \ln(1/\varepsilon_1)) + (2 + 4R^2)\varepsilon_p \right) \\
&\le O(\varepsilon_p + h k \varepsilon_{\text{err}} \log(k/\varepsilon_1)) \\
&\quad + \frac{h}{2^m} O\left(\tilde{L} \varepsilon_p + \varepsilon_{\text{err}} + \sqrt{d} R (R^2 + \ln(1/\varepsilon_1)) + (2 + 4R^2)\varepsilon_p \right)., \qquad (32)
\end{aligned}$$

最后一行由m的选取可得。

现在利用$h = O(\frac{1}{1+R_2})$这一事实，我们有

$$\frac{h}{2^m}(2 + 4R^2)\varepsilon_p = O(\frac{1}{2^m} T k \varepsilon_{\text{err}} \log(1/\varepsilon_1)) = O(\frac{T/h}{2^m} h k \varepsilon_{\text{err}} \log(1/\varepsilon_1)) = O(h k \varepsilon_{\text{err}} \log(k/\varepsilon_1)),$$

其中我们利用了这一事实：

$$T/h = O\left((1+R^2) \ln(1/\varepsilon_{\text{err}})^2 (\ln(1/\varepsilon_1) + \ln(d)) \right) = O(2^m).$$

类似地，不难验证，取m 时，我们有

$$\frac{h}{2^m}\left(\tilde{L}\varepsilon_p + \varepsilon_{\text{err}} + \sqrt{d}R(R^2 + \ln(1/\varepsilon_1))\right) = O(hk\varepsilon_{\text{err}}\log(k/\varepsilon_1))$$

因此，我们可以对32进行上界估计：

$$\mathbb{E}\left\|y - T_\phi^{\circ m}(\bar{y})\right\|_{[t_0,t_0+h]} \le O(\varepsilon_p) + O(hk\varepsilon_{\text{err}}\log(k/\varepsilon_1)) = O((t_0+h)k\varepsilon_{\text{err}}\log(k/\varepsilon_1)).$$

这就证明了归纳步骤。因此，总体而言，我们证明了

$$\|y - \hat{y}\|_{[0,T]} = O(Tk\varepsilon_{\text{err}}\log(1/\varepsilon_1)) = O\Big((\ln(R) + \ln(d) + \ln(1/\varepsilon_{\text{err}}))\ln(1/\varepsilon_{\text{err}})\log(k/\varepsilon_1)\varepsilon_{\text{err}}\Big),$$

经过 $O\Big(TR^2\ln(1/\varepsilon_{\text{err}})(\ln(1/\varepsilon_1)+\ln(d))\Big)$轮后，其中每一轮我们都应用皮卡迭代 $m = \ln(R)+\ln(d)+\ln\ln(1/\varepsilon_1)+\ln(1/\varepsilon_{\text{err}})+\ln(\tilde{L})$ 次。注意，该 Wasserstein 保证成立的概率至少为 $1-O(\varepsilon_1 T/h) = 1-O\Big(\varepsilon_1 T R^2\ln(1/\varepsilon_{\text{err}})(\ln(1/\varepsilon)+\ln(d))\Big)$，这是应用并集界之后得到的。

□

H 推论3.9 的证明

本节将说明如何利用欠阻尼朗之万蒙特卡罗，将我们的采样器所达到的 Wasserstein 界提升为全变差界，从而证明推论3.9。相关伪代码见算法 3，我们对算法 2 稍作修改，在算法2结束时，运行一小段时间的欠阻尼朗之万蒙特卡罗算法（见附录 H.1）。这一步通常被扩散模型文献称为校正步骤。由此得到的算法 3 满足推论 3.9中所述的全变差保证。

算法 3： CORRECTEDCOLLOCATIONDIFFUSION($(s_t), \varepsilon$)

输入： 得分估计 s_t 满足假设 2，目标误差 $0 < \varepsilon < 1$

输出： 从满足 $\text{TV}(\hat{q}, q) \le \varepsilon$ 的分布 $\hat{q}$ 中采样

1 $\eta \leftarrow \Theta(\min(\varepsilon^{5/3}/(L^{1/4}d^{1/2}), \varepsilon\sigma^2/R^2))$，其中 $L \le \text{poly}(d/\eta)$ 是利普希茨常数q的上界。

2 $x \leftarrow$COLLOCATIONDIFFUSION($(s_t), \varepsilon$)

3 运行欠阻尼朗之万蒙特卡罗（参见附录 H.1），摩擦参数为 $\Theta(\sqrt{L})$ 且步长 $h = \varepsilon^{2/3}/(d^{1/3}M(\eta)^{1/3}L^{1/2})$，对于 $M(\eta)$步，其中 $M(\eta)$表示步数用于在上一步运行 COLLOCATIONDIFFUSION。设所得样本为 x'。

4 返回x'

H.1 欠阻尼朗之万蒙特卡罗

本节为求完整，简要回顾欠阻尼朗之万蒙特卡罗；在我们的算法中，它仅用于最后阶段，将 Wasserstein 接近性转换为全变差接近性。

给定分布 q 的对数密度估计值 s 以及摩擦参数 γ，步长为 h 且分数估计值为 s 的欠阻尼朗之万蒙特卡罗算法定义为

$$\begin{aligned} \mathrm{d}x_t &= v_t\,\mathrm{d}t \\ \mathrm{d}v_t &= (s(x_{\lfloor t/h\rfloor h}) - \gamma v_t)\,\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t\,, \end{aligned}$$

其中，B_t 是标准布朗运动。

我们将使用 [CCL+24]的以下结果，它是 [GW12]短时间正则化的一个推论：

定理 H.1 （[CCL+24]中的定理 A.5，重新表述）。 设 $q \propto e^{-H}$ 是定义在 $\mathbb{R}^d$ 上的一个分布，其中 ∇H 是 L-利普希茨的。设 p 是定义在 $\mathbb{R}^d$上的任意分布。假设 s： $\mathbb{R}^d \to \mathbb{R}^d$满足 $\|s - \nabla H\|^2 L_2(q) \le \varepsilon^2$ sc。设 $T \lesssim 1/\sqrt{L}$。

如果 p_N 表示从 p 初始化，使用摩擦参数为 $\Theta(\sqrt{L})$、步长为 h 的欠阻尼朗之万蒙特卡罗运行 T/h 步所得到的分布，那么

$$\mathrm{TV}(p_N, q) \lesssim \frac{W_2(p,q)}{L^{1/4}T^{3/2}} + \frac{\varepsilon_{\mathrm{sc}} T^{1/2}}{L^{1/4}} + L^{3/4}T^{1/2}d^{1/2}h\,. \tag{33}$$

H.2 完成证明

先验地 看来，方程（33）中的第三项会迫使我们选择 $h = O(d^{-1/2}\varepsilon)$，这意味着迭代复杂度为 $d^{1/2}/\varepsilon$。这里我们采用 [GCC24]的一个思路：由于在应用定理 H.1时，$W_2(p,q)$ 可以取得相当小的值，因此我们可以令 T 较小，从而得到更优的迭代复杂度界 T/h。

推论 3.9 的证明。为调节 η，令 $\mathrm{M}(\eta) = (\mathrm{R}/\sigma)^2 \cdot \mathrm{polylog}(1/\varepsilon, \mathrm{d}, \mathrm{R}, \tilde{\mathrm{L}})$ 表示算法 2 达到定理 3.7 中 Wasserstein 误差 η 所需的迭代次数。在定理 H.1 中，取 $\mathrm{h} = \varepsilon^{2/3}/(\mathrm{d}^{1/3}\mathrm{M}(\eta)^{1/3}\mathrm{L}^{1/2})$，$\mathrm{T} = \mathrm{M}(\eta)\mathrm{h}$，则可得：从算法 2 给出的分布出发，若运行欠阻尼朗之万蒙特卡罗 $\mathrm{M}(\eta)$ 次，将得到一个分布 p_N，使得

$$\mathrm{TV}(p_N, q) \lesssim \frac{\eta L^{1/4}d^{1/2}}{M(\eta)^{2/3}\varepsilon^{2/3}} + \frac{\varepsilon_{\mathrm{sc}} M(\eta)^{1/3}\varepsilon^{1/3}}{L^{1/2}d^{1/6}} + \varepsilon\,, \tag{34}$$

在算法 2 成功达到 Wasserstein 误差 η这一事件的条件下。后者发生的概率为 $1 - \eta(R/\sigma)^2$，因此，算法 3 的全变差误差为式（34）加上 $\eta(R/\sigma)^2$。

如果取 $\eta = \min(\varepsilon^{5/3}/(L^{1/4}d^{1/2}), \varepsilon\sigma^2/R^2,,$ 那么 $\eta(R/\sigma)^2 \le \varepsilon$，$M(\eta) \le \tilde{O}(R/\sigma)^2 \cdot \mathrm{polylog}(1/\varepsilon, d, \tilde{L}, L)$，且右侧的第一项有界于 ε。根据假设 $\varepsilon_{\mathrm{sc}} \le \tilde{O}(\frac{\varepsilon^{2/3} L^{1/2} R^{1/6}}{(R/\sigma)^{2/3}} dR$，右侧的第二项有界~~于 ε。将~~上文中的 ε 替换为 $c\varepsilon$，其中 c 是充分小的常数，即可得到所声称的界。

□

I 概率流常微分方程运动的上界估计

在本节中，我们证明得分函数的一些有用性质。首先，我们对得分函数 $\nabla \log$ 分布 q_t 的光滑性进行界定。

引理 I.1。 （真实分数算子范数的上界）对于 $t \ge 1$，我们有

$$\begin{aligned}\|\nabla(y_t + \nabla \log q_t(y_t))\| &\le e^{-(T-t)}\left(2 + 4R^2\right),\\ \left\|\nabla^2 \log q_t(y_t)\right\|_{\mathrm{op}} &\le 2 + 4e^{-(T-t)}R^2,\end{aligned}$$

并且对于 $2(T-t) < 1$,

$$\|\nabla(y_t + \nabla \log q_t(y_t))\| \le \frac{5R^2}{(T-t)^2}.$$

证明。 不难验证，log分布 q_t 的海森矩阵可以写成如下形式：

$$-\nabla^2 \log q_t(y) = -\frac{1}{\sigma_t^4}\mathrm{Cov}(q^{t,y}) + \frac{1}{\sigma_t^2} I.$$

但利用分布 q 的支撑集半径 R 假设，我们可以将协方差界定为

$$\mathrm{Cov}(q^{t,y}) \le \frac{1-\sigma_t^2}{\sigma_t^4} R^2 I.$$

因此

$$-\nabla\left(y_t + \nabla \log q_t(y_t)\right) = \frac{1-\sigma_t^2}{\sigma_t^2} I - \frac{1-\sigma_t^2}{\sigma_t^4} R^2 I. \tag{35}$$

现在，对于 $t \ge 1$，我们有 $\sigma^2{}_t = (1-e^{-2(T-t)}) \ge 1 - 1/e^2 \ge 0.5$。因此，对于 $t \ge 1$,

$$\left\|\nabla\left(y_t + \nabla \log q_t(y_t)\right)\right\| \le (1-\sigma_t^2)\left(2+4R^2\right) = e^{-2(T-t)}\left(2+4R^2\right).$$

同理

$$\left\|\nabla^2 \log q_t(y_t)\right\| \le 2 + (1-\sigma_t^2)\left(4R^2\right) = 2 + 4e^{-2(T-t)}R^2.$$

另一方面，对于 $s = 2(T-t) < 1$，利用 $e^{-s} \le 1 - s + \frac{s^2}{2} \le 1 - \frac{s}{2}$ 我们有 $\sigma^2{}_t = 1 - e^{-(T-t)} \ge T - t$。然后，根据假设 $R \ge 1$ 和方程（35），我们有

$$\left\|\nabla\left(y_t + \nabla \log q_t(y_t)\right)\right\| \le \left(\frac{2}{T-t} + \frac{4R^2}{(T-t)^2}\right) \le \frac{5R^2}{(T-t)^2}. \qquad \square$$

接着，我们可以利用这个光滑性界，控制两个按照相同的概率流常微分方程演化的过程随时间推移而发散的程度：

引理 I.2. （从相近点出发的真实常微分方程解之间的距离）对于按照概率流常微分方程演化的 $y_t, \tilde{y}_t$，即

$$\frac{d}{dt} y_t = y_t + \nabla \log q_t(y_t)$$
$$\frac{d}{dt} \tilde{y}_t = \tilde{y}_t + \nabla \log q_t(\tilde{y}_t),$$

其初始条件满足 $\left\|y_{t_0} - \tilde{y}_{t_0}\right\| \le \varepsilon$，那么对于满足 $T - (t_0 + s) \ge 1$ 的时间窗口 s，我们有

$$\left\|y_{t_0+s} - \tilde{y}_{t_0+s}\right\| \le e^{\frac{5R^2}{\sigma_t^2}s}\varepsilon$$

证明。 注意，对于 $t_0 \le t \le t_0 + s$，我们有 $t = T - t \ge 1$ 根据我们的假设，因此由引理 I.1：

$$\begin{aligned}
\frac{d}{dt}\left\|y_t - \tilde{y}_t\right\| &= \frac{\left\langle y_t - \tilde{y}_t, \nabla \log q_t(y_t) - \nabla \log q_t(\tilde{y}_t)\right\rangle}{\left\|y_t - \tilde{y}_t\right\|} \\
&= \frac{\int_{r=0}^1 (y_t - \tilde{y}_t)^\top \nabla\left(y_t + q_t(y_t + r(\tilde{y}_t - y_t))\right)(y_t - \tilde{y}_t)\, dr}{\left\|y_t - \tilde{y}_t\right\|} \\
&\le e^{-t}(2 + 4R^2)\frac{\left\|y_t - \tilde{y}_t\right\|^2}{\left\|y_t - \tilde{y}_t\right\|} \\
&= e^{-t}(2 + 4R^2)\left\|y_t - \tilde{y}_t\right\|,
\end{aligned}$$

这意味着

$$\frac{d}{dt}\ln(\|y_t-\tilde{y}_t\|)\le e^{-t}(2+4R^2). \tag{36}$$

对方程（36）从 $t=t_0$ 到 $t=t_0+s$ 进行积分，我们得到第一部分所需的结果。类似地，对于 $T-t_0<1$的情形：

$$\frac{d}{dt}\ln(\|y_t-\tilde{y}_t\|)\le\frac{5R^2}{t^2}, \tag{37}$$

这意味着（使用不等式 $\sigma^2{}_t=1-e^{-t}\le t$，其中 $t=T-(t_0+s)$）

$$\left\|y_{t_0+s}-\tilde{y}_{t_0+s}\right\|\le e^{\frac{5R^2}{T-(t_0+s)}s}\varepsilon\le e^{\frac{5R^2}{\sigma_t^2}s}\varepsilon. \qquad \square$$

引理 I.3。（概率流常微分方程的分布保证）沿着反向常微分方程

$$\dot{y}_t=F_t^*(y_t) \tag{38}$$

对于 $F^*t(y)=y+\nabla\log q_t(y)$，我们以概率 $1-\varepsilon_1$ 得到，

$$\begin{aligned}
&\forall t\in[t_0,t_0+h],\|y_t\|\le e\left(R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}\right),\\
&\forall t\in[t_0,t_0+h],\left\|F_t^*(y_t)\right\|\le 6\left(R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}\right),\\
&\forall s_1,s_2,\left\|y_{s_1}-y_{s_2}\right\|\le 6(s_2-s_1)\left(R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}\right)\\
&\left\|F_{t_0}^*(y_{t_0})-F_{t_0+h}^*(y_{t_0+h})\right\|\lesssim\sqrt{d}R(R^2+\ln(1/\varepsilon)).
\end{aligned}$$

证明。回忆一下，我们可以写成 $y_t=X_t+\sigma_t\xi$ 其中 $F^*{}_t(y)=y_t+\frac{\mathbb{E}^{t,y}X-y_{t_0}}{\sigma_t^2}$。现在对于时间 t_0，使用 $t_0\ge 0.5$，其中 $t_0\ge 1$，正如我们在引理 I.1 中所证明的，我们有

$$\left\|F_{t_0}^*(y_{t_0})\right\|=\mathbb{E}\left\|y_{t_0}+\frac{\mathbb{E}^{t,y_{t_0}}X-y_{t_0}}{\sigma_{t_0}^2}\right\| \tag{39}$$

$$\le\left(1+\frac{1}{\sigma_{t_0}^2}\right)\left\|y_{t_0}\right\|+\frac{1}{\sigma_{t_0}^2}R \tag{40}$$

$$\le\frac{3}{2}\left\|y_{t_0}\right\|+\frac{R}{2} \tag{41}$$

现在由于 $\|\xi\|$ 是次高斯的，我们以至少 $1-\varepsilon_1$ 的概率得到，

$$\left\|y_{t_0}\right\|\le R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}.$$

另一方面，根据定义（38）

$$\frac{d}{ds}\|y_s\|\le\left\|F_s^*(y_s)\right\|\le\frac{3}{2}\|y_s\|+\frac{R}{2},$$

因此截至时间s，我们得到

$$\|y_s\|\le(R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)})e^{s\left(R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}\right)},$$

由$h \leq \frac{1}{R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}}$这一事实可知，

$$\|y_s\| \leq e\left(R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}\right).$$

第二部分由（41）得出。对于第三部分，我们有

$$\begin{aligned}\left\|y_{s_1}-y_{s_2}\right\| &= \left\|\int_{s_1}^{s_2} F_s^*(y_s)ds\right\| \\ &\leq \int_{s_1}^{s_2} \left\|F_s^*(y_s)\right\| ds \\ &\leq 6(s_2-s_1)\left(R+\sqrt{d}+\sqrt{\ln(1/\varepsilon_1)}\right).\end{aligned}$$

对于最后一部分，利用引理 3.3：

$$\left(\mathbb{E}\left\|\partial_t F_t^*(y_t)\right\|^p\right)^{1/p} \lesssim \sqrt{d}R(R+\sqrt{p})^2.$$

从t_0 积分到t_0+h：

$$\left(\mathbb{E}\left\|F_{t_0}^*(y_{t_0})-F_{t_0+h}^*(y_{t_0+h})\right\|^p\right)^{1/p} \lesssim h\sqrt{d}R(R+\sqrt{p})^2,$$

这意味着以至少 ε_1 的概率，我们有

$$\left\|F_{t_0}^*(y_{t_0})-F_{t_0+h}^*(y_{t_0+h})\right\| \lesssim \sqrt{d}R(R+\sqrt{\ln(1/\varepsilon)})^2.$$

证明完毕。 □