arXiv:2303.03384v1 [cs.LG] 6 Mar 2023

超越线性扩散的恢复—退化：DDIM 型采样器的非渐近分析

Sitan Chen* Giannis Daras† Alexandros G. Dimakis‡加州大学
伯克利分校 德克萨斯大学奥斯汀分校 德克萨斯大学奥斯汀分校

2023年3月7日

摘要

我们为扩散生成建模中使用的确定性采样器建立了一个非渐近分析框架。近期有若干研究利用吉尔萨诺夫定理和插值论证的链式法则变体等工具分析随机采样器。然而，将这些技术用于确定性采样器时，得到的界毫无意义。我们通过揭示概率流常微分方程中的一步演化可分解为两个步骤，为确定性采样提供了一种新的操作性解释：1）恢复步骤，在某个无限接近的过去时刻，对条件对数似然执行梯度上升；2）退化步骤，利用指向当前迭代点的噪声运行前向过程。这一视角使我们能够将去噪扩散隐式模型推广到一般的非线性正向过程。随后，我们在数据分布满足温和条件的情况下，为这类采样器首次建立了多项式收敛界。

1 引言

扩散模型 [SDWMG15，HJA20，SE19] 已成为生成建模的一种强大框架。其核心组成部分之一是在不同尺度上扰动样本，逐渐将数据塑造成噪声。扰动过程也称为前向过程，可由其定义的中间分布 $\{q_t\}_{t\in[0,T]}$完整描述。扩散模型通过近似中间分布的分数函数，即对数似然的梯度，来学习逆转前向过程 q_t。

得分函数学习完成后，可以通过运行与前向过程相关联的反向随机微分方程（SDE）来生成样本 [And82，SSDK+20]。然而在实践中，只能对 SDE 进行适当的离散化；又由于采样过程具有递归性质，先前步骤的离散化误差可能不断累积，导致采样漂移偏离真实反向过程。其他误差来源包括估计分数时的近似误差 [SHG+22，HJA20，ND21] 以及初始分布。近期研究已探讨如何控制反向 SDE 中误差的传播，包括 [BMR22，DBTHD21，DB22，LWYL22，LLT22a，Pid22，LLT22b，CLL22，CCL+22]。

第二类采样方法是确定性采样器。如[SSDK+20]所述，可以通过确定性常微分方程过程推导出这类采样器，即概率流常微分方程。该方程满足相同的福克-普朗克方程（因此具有相同的边缘分布 $\{q_t\}$），与反向 SDE 相同。

*电子邮箱：sitanc@berkeley.edu
†电子邮箱：giannisdara@utexas.edu
‡电子邮箱：dimakis@austin.utexas.edu

另一项工作 DDIM [SME21], 通过考虑一种产生相同训练目标、但具有不同反向过程的非马尔可夫扩散过程，推导出确定性采样器。这两种表述经重新参数化后 [SSDK+20，KAAL22]. DDIM 采样器可以解释为反复执行两个步骤的组合：恢复步骤，在时间 t恢复当前迭代点的某种粗略最终重建；退化步骤则将这一粗略估计扰动至时间 $t+h$。这一解释可以扩展到一般的线性扰动过程 [ZTC22、DDT+22、BBC+22、ZTC22]。

确定性采样器相比随机采样器具有诸多优势。尽管后者通常在图像生成方面达到最先进水平，但需要大量函数评估，因此在许多应用中并不实用。近期研究 [KAAL22]显著缩小了确定性采样器与随机采样器在样本质量上的差距。确定性采样器通常快得多[SME21，ND21] 而且还有助于计算似然 [HJA20，SSDK+20]。此外，加速扩散模型最成功的技术之一——渐进式蒸馏 [SH22], 需要确定性采样器。确定性采样器还使人们能够探索训练后网络的语义潜在空间[KJU22]。

尽管确定性采样器意义重大，目前对其理论的理解仍然有限。具体而言，与随机采样器不同，目前尚无关于其非渐近收敛行为的分析。要获得此类分析颇具挑战，因为吉尔萨诺夫定理——在实现反向 SDE 时用于界定误差传播的主要工具——以及相关技术，对于确定性采样器都会给出无效界（见第 5 节）。

我们的贡献主要有两点。首先，我们为反向常微分方程提出了一种新的操作性解释，将 DDIM采样推广到任意非线性前向过程。

定理1.1 （非正式表述，见第 3 节）。 记 h 为离散化概率流常微分方程时所采用的无穷小步长。令 $\ell \in \mathbb{N}$ 为满足 $\ell \to \infty$ 且 $\ell h \to 0$的参数。对于任意前向过程，将概率流常微分方程运行 h 的时间，等价于执行以下两个步骤：1）通过对条件对数似然进行梯度上升，将当前迭代点恢复到过去 ℓh 个时间步；2）通过沿当前迭代点所指方向的噪声模拟前向过程，使其退化 $(\ell-1)h$ 个步骤。

随后，我们通过非渐近证明补充这一新的渐近结果，证明基于该操作性解释的采样器会收敛到真实过程。这为扩散模型中的随机采样器近期非渐近分析 [CCL+22、LLT22b、CLL22]提供了确定性采样的对应结果：

定理 1.2 （非正式表述，参见定理 4.1）。 在对数据分布的光滑性作出温和假设的条件下（特别是，该分布可以是任意非对数凹的），定理 *1.1* 所导出的确定性采样器生成的样本，其相对于数据分布的 KL 散度很小，前提是 ℓh 和 ℓ^{-1} 相对于维度及其他问题特定参数呈多项式级小。

作为推论，我们的技术表明，概率流常微分方程的欧拉离散化也满足相同的界；据我们所知，这为该采样器提供了首个非渐近分析。

内容提要。 第 2 节简要回顾扩散生成建模中的正向过程和反向过程。第 3 节介绍我们对一般非线性扩散的概率流常微分方程所作的新操作性解释。第 4 节正式陈述我们的主要非渐近保证——定理 4.1，并概述证明，将技术细节留至附录。

在附录A 中，我们说明了第3节中出现的某个学习率参数的选择依据。在附录B 中，我们为定理 4.1的证明提供预备计算。在附录 C中，

我们给出了由具有相似漂移的常微分方程驱动的两个过程之间距离的一个一般界，其中一个过程是离散时间过程的插值。最后，在附录D 中，我们将这一一般界应用于我们的设定，界定概率流常微分方程与我们的采样器之间的漂移差异，并证明定理4.1。

2 预备知识

在本文中，我们考虑由如下形式的随机微分方程（SDE）驱动的一般前向过程：

$$\mathrm{d}x_t = f_t(x_t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t, \qquad x_0 \sim q\,, \tag{1}$$

其中（W_t）是 $\mathbb{R}^d$中的标准布朗运动。令 q_t 表示 x_t的分布，即 $q_0 = q$.

假设我们将前向过程运行至终止时间 $T > 0$。在对扩散施加的温和条件下（例如参见 [And82、F85、CCGL22]），本文所考虑的过程均满足这些条件，因此存在一个由随机微分方程给出的适当反向过程，使得其在时间 t的边缘分布为 q_{T-t}。为方便起见，我们通常将 q_{T-t}称为$q_t^{\leftarrow}$。

事实上，具有这一性质的随机微分方程构成一个完整的族。对于任意 $\lambda \geq 0$，考虑过程（$x^{\leftarrow,\lambda}_t$）$_{0\leq t\leq T}$，其形式为

$$\mathrm{d}x^{\leftarrow,\lambda}_t = -\Big\{f_{T-t}(x^{\leftarrow,\lambda}_t) - \frac{1+\lambda^2}{2}g(T-t)^2\nabla \ln q_t^{\leftarrow}(x^{\leftarrow,\lambda}_t)\Big\}\,\mathrm{d}t + \lambda g(T-t)\mathrm{d}W_t, \qquad x_0^{\leftarrow,\lambda} \sim q_0^{\leftarrow}\,. \tag{2}$$

通过检验福克-普朗克方程，可以看出，$x^{\leftarrow,\lambda}_t$ 的边缘分布确实由$q_t^{\leftarrow}$给出。

该族中的一个值得注意的过程对应于 $\lambda = 0$的情形。这是一个确定性过程，记为（$x_t^{\leftarrow}$）$_{0\leq t\leq T}$，由概率流常微分方程 [SSDK+20]驱动。

$$\mathrm{d}x_t^{\leftarrow} = -\Big\{f_{T-t}(x_t^{\leftarrow}) - \frac{1}{2}g(T-t)^2\nabla \ln q_t^{\leftarrow}(x_t^{\leftarrow})\Big\}\,\mathrm{d}t\,, \tag{3}$$

其中 $x_0^{\leftarrow} \sim q_0^{\leftarrow}$。

在扩散模型文献中，前向过程有两种常见选择：方差爆炸（VE）随机微分方程 [SSDK+20、SE19、SE20]，其对应于 $f_t(x_t) = 0$, $g(t) = \sqrt{\mathrm{d}\sigma_t^2/\mathrm{d}t}$，其中 σ_t^2 为某个递增函数；以及方差保持（VP）随机微分方程 [HJA20]，其对应于 $f_t(x_t) = -\frac{1}{2}\beta_t x_t$, $g(t) = \sqrt{\beta_t}$，其中 β_t 为某个方差调度。这两种选择被用于最先进的扩散模型 [DN21、KSS+22]，并构成了 DALL·E2 [RDN+22]、Imagen [SCS+22] 和 Stable Diffusion [RBL+22] 等系统的基础。

3 Ope概率流常微分方程的有理解释

3.1 预备知识：线性随机微分方程与 DDIM

我们首先回顾与方差爆炸（VE） [SSDK+20] 随机微分方程相关的概率流常微分方程，如何被解释为去噪扩散隐式模型（DDIM） [SME21]。为简化说明，我们专门讨论 $\sigma_t^2 = t$的情形，它对应于前向过程

$$\mathrm{d}x_t = \mathrm{d}W_t, \qquad x_0 \sim q\,.$$

根据（3），相应的概率流常微分方程为：

$$\mathrm{d}x_t^{\leftarrow} = \frac{1}{2}\nabla \ln q_t^{\leftarrow}(x_t^{\leftarrow})\,\mathrm{d}t, \qquad x_0^{\leftarrow} \sim q_T, \tag{4}$$

因此，$x_t^{\leftarrow}$的边缘分布为 $q_t^{\leftarrow}$，对任意 $0 \le t \le T$皆如此。DDIM 的视角为（4）提供了一种有趣的操作性解释。取一个无穷小步长 h，并考虑以下过程：给定 $x_t^{\leftarrow}$，生成 $x_{t+h}^{\leftarrow}$。我们首先估计前向过程的起点 x_0。注意到

$$x_t^{\leftarrow} = x_{T-t} = x_0 + \varepsilon\sqrt{T-t} \tag{5}$$

对于 $\varepsilon \sim \mathcal{N}(0,\ \mathrm{Id})$，根据 Tweedie 公式 [Efr11]，在给定 x_0的条件下，$x_t^{\leftarrow}$的后验分布的均值，即 $\mathbb{E}[x_0|x_t^{\leftarrow}]$，恰好为：

$$z \triangleq \mathbb{E}[x_0|x_t^{\leftarrow}] = x_t^{\leftarrow} + (T-t)\,\nabla \ln q_t^{\leftarrow}(x_t^{\leftarrow})\,.$$

从 z 出发，沿着前向过程从时间 0 退化到时间 $T-t$，最终会得到 $z + \gamma\sqrt{T-t}$，其中使用了某个高斯噪声 $\gamma \sim \mathcal{N}(0, \mathrm{Id})$。

以下是 DDIM 背后的核心思想：假设我们不再将 γ 作为以下方程的解：

$$x_t^{\leftarrow} = z + \gamma\sqrt{T-t}\,,$$

也就是说，假设我们将 γ 视为将 z 退化为 $x_t^{\leftarrow}$所需的“模拟噪声”，而不是新采样的高斯噪声。

现在设想运行前向过程，将 z从时间 0 退化到时间 $T-(t+h)$，但使用该模拟噪声 $\gamma = \frac{x_t^{\leftarrow} - z}{\sqrt{T-t}}$ 代替高斯噪声。事实证明，由此得到的向量（我们将其定义为 $x_{t+h}^{\leftarrow}$）近似等于从 $x_t^{\leftarrow}$出发运行时长为 h 的概率流常微分方程所得到的结果！

事实上，以这种方式对 z 进行退化后得到的结果是

$$\begin{aligned} x_{t+h}^{\leftarrow} &\triangleq z + \sqrt{T-(t+h)}\cdot\frac{x_t^{\leftarrow} - z}{\sqrt{T-t}} \\ &= x_t^{\leftarrow} + (T-t)\cdot\left(1 - \sqrt{1 - \frac{h}{T-t}}\right)\cdot\nabla \ln q_{T-t}(x_t^{\leftarrow})\,. \end{aligned}$$

注意，当 $h \to 0$时，迭代点 $x_{t+h}^{\leftarrow}$趋近于 $x_t^{\leftarrow} + \frac{h}{2}\nabla \ln q_{T-t}(x_t^{\leftarrow})$。因此，如上所述的解释确实恢复了概率流常微分方程（4），正如所声称的那样。上述结论可以相当容易地推广到任意线性扩散 [DDT+22、BBC+22]。

3.2 一般扩散

现在考虑这样的情形：前向过程由任意的、可能是非线性的扩散给出，如第 2 节中的公式（1）所示，因此相应的概率流常微分方程由公式（3）给出。不幸的是，一旦我们脱离线性设定，上一节中的操作性解释就会失效。关键问题在于，当我们构造前向过程起点的估计 z时，不再存在关于 $x_t^{\leftarrow}$条件后验均值的简单表达式。

恢复算子。为绕过这一问题，我们的第一个洞见是：我们不再为前向过程起点推导估计，而是为过去 ℓh 个时间单位处的过程推导估计，即在前向过程的 $T - t - \ell h$ 时刻。在上一节中，我们隐含地取了 $\ell = (T-t)/h$，但现在 ℓ 是一个需要调节的参数。关键在于，选择

ℓ 使得 $\ell h \to 0$ 从而我们可以围绕 $T-t$进行线性化。在与（5）类比的情况下，我们得到近似关系

$$\begin{aligned} x_t^{\leftarrow} = x_{T-t} &\approx x_{T-t-\ell h} + \ell h\, f_{T-t-\ell h}(x_{T-t-\ell h}) + g(T-t-\ell h)\sqrt{\ell h}\cdot\varepsilon \\ &\approx x_{T-t-\ell h} + \ell h\, f_{T-t}(x_t^{\leftarrow}) + g(T-t)\sqrt{\ell h}\cdot\varepsilon \end{aligned}$$

对于 $\varepsilon \sim \mathcal{N}(0, \mathrm{Id})$，近似误差至多为加性误差 $o(h)$。重新整理后可知， $x_{T-t-\ell h}$只是 $x_t^{\leftarrow} - \ell h f_{T-t}(x_t^{\leftarrow})$ 加上方差为 $\ell h g(T-t)^2$的高斯噪声。因此，再次利用 Tweedie 公式，我们发现给定$x_t^{\leftarrow}$ 时， $x_{T-t-\ell h}$ 的后验分布的均值近似为

$$z \triangleq x_t^{\leftarrow} - \ell h \left\{ f_{T-t}(x_t^{\leftarrow}) - g(T-t)^2 \nabla \ln q_t^{\leftarrow}(x_t^{\leftarrow}) \right\}. \tag{6}$$

借用 [BBC+22]中的术语，我们将从 $x_t^{\leftarrow}$到 z 的映射称为恢复算子。形式上，对于 $t > s > 0$，定义恢复算子 $R_{t\to s}(\cdot)$ 如下:

$$R_{t\to s}(x) \triangleq x - (t-s) f_t(x) + (t-s) g(t)^2 \nabla \ln q_t(x) \tag{7}$$

使得 $z = R_{T-t\to T-t-\ell h}(x_t^{\leftarrow})$。

作为梯度上升的恢复算子。这里我们简要指出，换个角度理解恢复算子也很有启发性：它可以看作一次梯度上升。

形式上，给定时刻 $0 < t < s$，考虑最大化条件对数似然 $\ln q_s^{\leftarrow}(\cdot \mid x_t^{\leftarrow})$。这等价于最大化

$$\ell_{x_t^{\leftarrow}}(x) \triangleq \ln q_t^{\leftarrow}(x_t^{\leftarrow} \mid x_s^{\leftarrow} = x) + \ln q_s^{\leftarrow}(x). \tag{8}$$

对于 s ，它比 t稍大一个无穷小量，$x_t^{\leftarrow}$在给定$x_s^{\leftarrow} = x$ 的条件下，其分布为高斯分布，均值和协方差近似为 $x + f_{T-s}(x)$（$t-s$）以及 $g(T-t)^2(t-s)$ Id。因此，我们可以计算（8）的梯度，得到

$$\nabla \ell_{x_t^{\leftarrow}}(x) \approx \frac{1}{g(T-t)^2(t-s)} \left(\mathrm{Id} + (t-s)\,\nabla f_{T-s}(x)\right) \cdot \left(x_t^{\leftarrow} - x - f_{T-s}(x)\,(t-s)\right) + \nabla \ln q_s^{\leftarrow}(x)\,. \tag{9}$$

现在考虑从$x_t^{\leftarrow}$出发，采用学习率 η 进行一次梯度步，得到 $x_t^{\leftarrow} + \eta \nabla \ell_{x_t^{\leftarrow}}(x_t^{\leftarrow})$ 。在附录A中，我们证明，在 q 服从高斯分布且前向过程为 Ornstein-Uhlenbeck 过程这一特殊情形下，仅通过一步梯度上升来最大化条件对数似然时，学习率的正确选择为

$$\eta \triangleq 2g(T-t)^2 \cdot (t-s)\,. \tag{10}$$

在这种情况下，注意到

$$\begin{aligned} x_t^{\leftarrow} + \eta \nabla \ell_{x_t^{\leftarrow}}(x_t^{\leftarrow}) &\approx x_t^{\leftarrow} - (t-s) f_{T-t}(x_t^{\leftarrow}) - (t-s)^2 (\nabla f_{T-s}(x_t^{\leftarrow})) f_{T-s}(x_t^{\leftarrow}) \\ &\qquad + (t-s) g(T-t)^2 \nabla \ln q_s^{\leftarrow}(x_t^{\leftarrow}) \\ &\approx x_t^{\leftarrow} - (t-s) f_{T-t}(x_t^{\leftarrow}) + (t-s) g(T-t)^2 \nabla \ln q_t^{\leftarrow}(x_t^{\leftarrow}), \end{aligned} \tag{11}$$

其中，在第二步中，我们舍去了二阶项 $(t-s)^2(\nabla f_{T-s}(x_t^{\leftarrow}))f_{T-s}(x_t^{\leftarrow})$，并将 $(t-s)g(T-t)^2\nabla \ln q_s^{\leftarrow}(x_t^{\leftarrow})$ 近似为一阶项 $(t-s)g(T-t)^2\nabla \ln\, q_t^{\leftarrow}(x_t^{\leftarrow})$。现在注意，对于 $s = t - \ell h$，(11) 的更新规则与 (6) 的更新规则相同。

退化算子。推导的其余部分与上一节的思路类似。给定噪声向量 $\gamma \in \mathbb{R}^d$，定义退化算子 $D^{\gamma}_{s,t}(\cdot)$ 如下：

$$D^{\gamma}_{s\to t}(x) \triangleq x + f_s(x)(t-s) + g(s)\sqrt{t-s}\cdot\gamma\,. \tag{12}$$

该算子只是从时间 s 开始，以 γ 作为噪声，对前向过程进行时长为 $t-s$ 的 Euler-Maruyama 离散化。

从 z 开始，沿着前向过程将其从 $T-t-\ell h$ 时刻退化至 $T-t$ 时刻，最终会得到 $D^{\gamma}_{T-t-\ell h\to T-t}(z) = z + \ell h\, f_{T-t-\ell h}(z) + g(T-t-\ell h)\sqrt{\ell h}\cdot\gamma$，其中 $\gamma\sim\mathcal{N}$ 是某个高斯噪声，服从 $(0, \mathrm{Id})$ 分布。与前面一样，我们改为令 γ 表示将 z 退化为 $x_t^{\leftarrow}$ 所需的模拟噪声；在这种情况下，它由下式的解给出

$$x_t^{\leftarrow} = z + \ell h\, f_{T-t-\ell h}(z) + g(T-t-\ell h)\sqrt{\ell h}\cdot\gamma\,.$$

为了得到反向过程的下一个迭代点 $x_{t+h}^{\leftarrow}$，我们使用 γ 将 z 从时间 $T-t-\ell h$ 退化至 $T-t-h$。结果如下

$$\begin{aligned}
x_{t+h}^{\leftarrow} &= z + (\ell-1)h\, f_{T-t-\ell h}(z) + g(T-t-\ell h)\sqrt{(\ell-1)h}\cdot\frac{x_t^{\leftarrow} - z - \ell h\, f_{T-t-\ell h}(z)}{g(T-t-\ell h)\sqrt{\ell h}} \\
&\approx z + (\ell-1)h f_{T-t}(x_t^{\leftarrow}) + \sqrt{1-1/\ell}\cdot\ell h g(T-t)^2\nabla\ln q_{T-t}(x_t^{\leftarrow}) \qquad (13)\\
&= x_t^{\leftarrow} - h f_{T-t}(x_t^{\leftarrow}) + \ell h\cdot\left(1-\sqrt{1-1/\ell}\right)\cdot g(T-t)^2\nabla\ln q_{T-t}(x_t^{\leftarrow})\,. \qquad (14)
\end{aligned}$$

其中在第二步中，我们将 $f_{T-t-\ell h}(z)$ 近似为 $f_{T-t}(x_t^{\leftarrow})$，并舍去了 $o(\ell h)$ 项，因为我们假设 $\ell h\to 0$。最后，由于 $\ell\to\infty$，右侧收敛至 $x_t^{\leftarrow} - h\{f_{T-t}(x_t^{\leftarrow}) - \frac{1}{2}g(T-t)^2\nabla\ln q_{T-t}(x_t^{\leftarrow})\}$，这就得到了概率流常微分方程的欧拉离散化。注意，只有这最后一步需要取 $\ell\to\infty$。最后，取 $h\to 0$ 时，上述结果便得到概率流常微分方程（3）。

我们在第 4 节开头对该操作性解释所提示的采样算法进行了更为正式的描述，并在该节中给出了这一采样器的非渐近分析的主要结果。

3.3 对其他采样器的扩展

我们为将 DDIM 扩展到非线性正向过程而提出的操作性解释，可以相对直接地用于描述更一般的采样器。例如，在公式（2）中，我们定义了一个更一般的反向过程族，其中每个过程在时间 t 都具有正确的边缘分布。这些过程可以通过类似的操作性解释轻松描述。

具体而言，考虑以下过程 $(\widetilde{x}^{\lambda}_{kh})_{k\in\{0,\dots,T/h\}}$。给定迭代点 $\widetilde{x}^{\lambda}_{kh}$，前一个迭代点 $\widetilde{x}^{\lambda}_{(k-1)h}$ 定义如下：

$$\widetilde{x}^{\lambda}_{(k-1)h} = D^{\gamma'}_{(k-\ell)h\to(k-1)h}(z), \quad \text{for } z \triangleq R_{kh\to(k-\ell)h}(\widetilde{x}^{\lambda}_{kh})\,, \tag{15}$$

$$\gamma' = \sqrt{1-\frac{\lambda^2}{\ell-1}}\,\gamma + \frac{1}{\sqrt{\ell-1}}\lambda\nu, \tag{16}$$

$$\gamma : D^{\gamma}_{(k-\ell)h\to kh}(z) = \widetilde{x}^{\lambda}_{kh}, \qquad \nu\sim\mathcal{N}(0,\mathrm{Id})\,.$$

注意，我们使用与之前相同的恢复算子得到 z，然后用它来估计噪声 γ。该框架的关键变化在于，现在从 z 到 $x_{t+h}^{\leftarrow}$ 进行加噪时，

我们不再仅使用估计噪声，而是使用估计噪声 γ 与新噪声 ν 的线性组合。

参数 λ 在此控制更新规则与确定性采样器的接近程度。显然，当 $\lambda = 0$ 时，我们得到与之前相同的完全确定性采样器。当 $\lambda = 1$ 时，采样器则变为 [SSDK+20] 的反向 SDE 采样器。系数经过专门选择，使得如果 γ 实际上是从 $\mathcal{N}(0, \mathrm{Id})$ 中抽取的样本，而非模拟噪声，那么 γ' 同样会是从$\mathcal{N}(0, \mathrm{Id})$ 中抽取的样本。

注意，

$$\begin{aligned}\widetilde{x}^{\lambda}_{(k-1)h} &= z + (\ell-1)h\, f_{(k-\ell)h}(z) + g((k-\ell)h)\sqrt{(\ell-1)h}\cdot\gamma' \\ &\approx z + (\ell-1)h\, f_{kh}(\widetilde{x}^{\lambda}_{kh}) + g(kh)\sqrt{(\ell-1)h}\cdot\gamma' \end{aligned} \tag{17}$$

其中在第二步中，我们用 $f_{(k-\ell)h}(z)$ 和 $g((k-\ell)h)$ 分别近似为 $f_{kh}(\widetilde{x}^{\lambda}_{kh})$ 和 $g(kh)$，并舍去了$o(h)$ 项。根据 (16) 以及 (20) 中估计噪声 γ 的定义，我们有

$$\gamma' = \sqrt{1-\frac{\lambda^2}{\ell-1}}\cdot\frac{\widetilde{x}^{\lambda}_{kh} - z - \ell h\, f_{(k-\ell)h}(z)}{g((k-\ell)h)\sqrt{\ell h}} + \frac{1}{\sqrt{\ell-1}}\lambda\nu \approx \sqrt{1-\frac{\lambda^2}{\ell-1}}\cdot\frac{\widetilde{x}^{\lambda}_{kh} - z - \ell h\, f_{kh}(z)}{g(kh)\sqrt{\ell h}} + \frac{1}{\sqrt{\ell-1}}\lambda\nu\,,$$

其中我们用 $f_{(k-\ell)h}(z)$ 和 $g((k-\ell)h)$ 分别近似为 $f_{kh}(z)$，并用 $g(kh)$ 近似 $g(kh)$。将其代入 (17)，并回顾 (15) 中 z 的定义，我们有

$$\begin{aligned}\widetilde{x}^{\lambda}_{(k-1)h} &\approx \widetilde{x}^{\lambda}_{kh} - h\, f_{kh}(\widetilde{x}^{\lambda}_{kh}) + \ell h\, g(kh)^2 \nabla \ln q_{kh}(\widetilde{x}^{\lambda}_{kh}) \\ &\qquad - \sqrt{1-\frac{1}{\ell}}\cdot\sqrt{1-\frac{\lambda^2}{\ell-1}}\cdot \ell h\, g(kh)^2\nabla\ln q_{kh}(\widetilde{x}^{\lambda}_{kh}) + \lambda\sqrt{h}\, g(kh)^2\,\nu \\ &\overset{(\ell\to\infty)}{=} \widetilde{x}^{\lambda}_{kh} - h\,\{f_{kh}(\widetilde{x}^{\lambda}) - \frac{1+\lambda^2}{2} g(kh)^2\nabla\ln q_{kh}(\widetilde{x}^{\lambda})\} + \lambda\sqrt{h}\, g(kh)^2\nu\,,\end{aligned}$$

并且在最后一步中，我们利用了

$$\lim_{\ell\to\infty} \ell\Big(1-\sqrt{1-\frac{1}{\ell}}\cdot\sqrt{1-\frac{\lambda^2}{\ell-1}}\Big) = \frac{1+\lambda^2}{2}\,.$$

4 离散化分析

下文中，我们针对由式（14）刻画的 DDIM 类采样器给出非渐近收敛分析。在第 4.1 节中，我们正式定义所讨论的采样器。随后在第 4.2节中，我们陈述主要结果；据我们所知，这些结果构成了首个针对扩散模型确定性采样的收敛分析。在第 4.3 节中，我们给出证明概述，并将大部分技术细节留待附录 B、C 和 D。

4.1 DDIM 类采样器

受第 3.2 节讨论的启发，我们的分析将聚焦于如下按时间反向定义的过程 $(\widetilde{x}_{kh})_{k\in\{0,\dots,T/h\}}$。迭代点 $\widetilde{x}_T$ 从 $q_0^{\leftarrow}$ 中采样。给定迭代点 $\widetilde{x}_{kh}$，前一迭代点 $\widetilde{x}_{(k-1)h}$ 定义如下：

$$\widetilde{x}_{(k-1)h} = D^{\gamma}_{(k-\ell)h\to(k-1)h}(z), \quad \text{for } z \triangleq R_{kh\to(k-\ell)h}(\widetilde{x}_{kh}) \text{ and } \gamma:\ D^{\gamma}_{(k-\ell)\to kh}(z) = \widetilde{x}_{kh}\,, \tag{18}$$

其中 R 和 D 分别在（7）和（12）中定义。由于 z 是恢复当前迭代点 $\widetilde{x}_{kh}$所得的结果，因此有

$$z = \widetilde{x}_{kh} - \ell h\, f_{kh}(\widetilde{x}_{kh}) + \ell h\, g(kh)^2\nabla\ln q_{kh}(\widetilde{x}_{kh}). \tag{19}$$

下一个迭代点 $\widetilde{x}_{(k-1)h}$ 是通过对 z 进行时长为 $(\ell-1)h$ 的退化得到的，其中噪声向量取为模拟噪声 γ。更准确地说，γ 是可用于对 z 进行时长为 ℓh 的退化、从而得到 $\widetilde{x}_{kh}$ 的噪声向量。由于 γ 是 $D^{\gamma}{}_{(k-\ell)\,h\to kh}(z) = \widetilde{x}_{kh}$ 的解，因此等价的表述方式是通过

$$\gamma = \frac{\widetilde{x}_{kh} - z - \ell h\, f_{(k-\ell)h}(z)}{g((k-\ell)h)\sqrt{\ell h}}. \tag{20}$$

注意，(18) 在 $k < \ell$ 时没有定义；在这种情况下，我们按照 Euler-Maruyama 离散化进行更新：

$$\widetilde{x}_{(k-1)h} = \widetilde{x}_{kh} - h(f_{kh}(\widetilde{x}_{kh}) - \frac{1}{2}g(kh)^2\nabla \ln q_{kh}(\widetilde{x}_{kh})) \quad \text{if } k < \ell\,.$$

后文中，记

$$\widetilde{x}^{\leftarrow}_{kh} \triangleq \widetilde{x}_{T-kh}$$

为。

4.2 结果陈述

我们对前向过程 (x_t) 和数据分布作出以下温和假设：

假设 1. 对所有 $t \geq 0$，参数 $L_{f;\mathrm{t}}$、L_g、$L_{f;\mathrm{x}}$、$L_{\mathsf{sc},\,t}$、R、$g_{\max}$、β、$M \geq$1、$c > 0$ 满足以下条件：

1. $f_t(x)$ 关于 t 是 $L_{f;\mathrm{t}}$-利普希茨的，关于 x 是 $L_{f;\mathrm{x}}$-利普希茨的。2. $g^2(t)$ 关于 t是 L_g-利普希茨的。3. $\|f_t(0)\| \leq R$。4. $g(t) \leq g_{\max}$。5. $\nabla \ln q_t^{\leftarrow}(x)$ 关于 x 是 $L_{\mathsf{sc},\,t}$-利普希茨的，并且满足

$$\|\nabla \ln \frac{q_t^{\leftarrow}}{q_s^{\leftarrow}}(x)\| \leq \beta|t-s|^c(1 + \|x\| + \|\nabla q_t^{\leftarrow}(x)\|)$$

对所有 $s \geq 0$ 均成立。记 $\sup_{t\geq 0} L_{\mathsf{sc},\,t}$ 为 $L_{\mathsf{sc},\,*}$。

6. $\nabla f_t(x)$ 和 $\nabla^2 \ln q_t^{\leftarrow}$ 在算子范数下是 L_{high}-利普希茨的。

注释1. 我们注意到，假设 1 的前四部分以及第 6 部分的前半部分都相当温和，任何合理选择的前向过程都满足这些条件。例如，对于奥恩斯坦-乌伦贝克过程 $\mathrm{d}x_t = -x_t\,\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$，我们可以取 $L_{f;\mathrm{t}} = 0, L_{f;\mathrm{x}} = 1, L_g = 0, R =$0,$g_{\max} = \sqrt{2}$，并且对于所有 x，$\nabla f_t(x) = -\mathrm{Id}$，因此显然在算子范数下是利普希茨的。第 5 部分确保得分函数 $\nabla \ln q_t^{\leftarrow}$ 在空间或时间发生扰动时变化不大。前者是基于分数的生成建模中有关得分函数离散化界的文献里的标准假设 [BMR22, CCL+22, LLT22a, LLT22b, CLL22]，；后者则适用于合理选择的前向过程。例如，对于奥恩斯坦-乌伦贝克过程，我们可以取 $c = 1/2$ 和$\beta = \Theta(L_{\mathsf{sc},*}\sqrt{d})$（参见例如 [LLT22a] 中的引理 C.12）。

假设1 与先前针对基于得分的生成模型所做分析中的假设，主要区别在于第6 部分的后半部分：我们假设高阶光滑性满足 $q_t^{\leftarrow}$。正如我们将在第C节看到的，这对我们的分析至关重要，因为在计算费舍尔信息的时间导数时会自然出现其三阶导数。

$\ln q_t^{\leftarrow}$，正如第*4.3*节所述。在该节中我们讨论过，只有在常微分方程设定下才需要计算此类时间导数，这也解释了为何先前对随机采样器的分析无需作出这样的假设。

在这些条件下，我们证明，我们的离散化过程能够以预设误差逼近真实反向过程 ε，前提是 ℓ 和 $(\ell h)^{-1}$ 大于某些关于以下参数呈多项式有界的量： $1/\varepsilon$ 以及假设 1中的所有参数：

定理 4.1。 令 $\varepsilon > 0$。令 $\widetilde{p}$ 表示过程 $(\widetilde{x}_{kh}^{\leftarrow})$ 在时刻 T的分布。假设 *1* 成立，并定义

$$\Lambda \triangleq \exp\left(\int_0^T (L_{f;\mathrm{x}}^2 + g_{\max}^2 L_{\mathsf{sc},t})\,\mathrm{d}t\right) \qquad \textit{and} \qquad \Lambda' \triangleq \exp\left(\int_0^T (L_{f;\mathrm{x}}^2 + g_{\max}^2 L_{\mathsf{sc},\lfloor t/h \rfloor h})\,\mathrm{d}t\right). \quad (21)$$

那么，存在量 $\mathfrak{C}_1$ 和 $\mathfrak{C}_2$，它们关于 $L_{f;\mathrm{t}}$、$L_{f;\mathrm{x}}$、L_g、R、$g_{\max}$、β、$L_{\mathsf{sc}*}$ L_{high} Λ Λ' d $\mathbb{E}\|x_0^{\leftarrow}\|^2$ $1/\varepsilon$ $\widetilde{p}\|q \le \varepsilon$ $\ell \ge \mathfrak{C}_1$ $\ell h \le \mathfrak{C}_2^{-1}$ 、、、、、、、和呈多项式有界，并且满足 KL ( ), 前提是 和 。

注释2。我们简要说明上述定理中出现的量 Λ、Λ'。通常我们认为 $L_{f;\mathrm{x}}$和 $g_{\max}$是常数量级，因此 Λ和 Λ'随 $\exp(\int_0^T L_{\mathsf{sc},t}\,\mathrm{d}t)$ 和 $\exp(\int_0^T L_{\mathsf{sc},\lfloor t/h \rfloor h}\,\mathrm{d}t)$ 呈多项式增长。虽然这随 T 呈指数增长，但 Ornstein-Uhlenbeck 等合理前向过程的指数收敛意味着，我们应认为 T 随 d/ε 呈对数增长。尽管直觉上可能会认为 Λ、Λ' 随 $L_{\mathsf{sc}、*}$呈指数增长，但我们在附录C中的例 1 证明，这些量实际上随 d 及其他参数（如 $L_{\mathsf{sc}、*}$）呈多项式增长，例如数据分布为高斯分布时。总而言之，这表明我们的非渐近保证在假设 1 中所有相关参数上具有多项式复杂度。

在实际应用中，过程 $(\widetilde{x}_{kh}^{\leftarrow})$ 会初始化为前向过程的平稳测度 q^*（经过适当的重新缩放后），而非初始化为 $q_0^{\leftarrow}$。正如[LLT22a、CCL+22和LLT22b], 所指出的，在替代初始化下，该过程最终迭代点与在初始化 $\widetilde{x}_0^{\leftarrow} \sim q^*$下的最终迭代点之间的 KL 散度，不超过这两个过程初始迭代点之间的 KL 散度。然而，由于 q^*的平稳性，后者的 KL 散度等价于前向过程的平稳测度与前向过程在时刻 T 的分布之间的 KL 散度。当 T 较大时，该 KL 散度通常呈指数级小，例如前向过程为 Ornstein-Uhlenbeck 过程时。通过 Pinsker不等式将 KL 散度转化为全变差，并应用三角不等式，可得在这种替代初始化下， $\widetilde{x}_T^{\leftarrow}$与数据分布 q 之间的全变差不超过定理 4.1 中的误差界，加上 q_T 与平稳分布之间的距离。形式化地，我们得到以下结果：

推论 4.2。设 $\varepsilon > 0$。设 $f_t(x) = -x$ 和 $g(t) = \sqrt{2}$，使得式 (1) 中的前向过程对应于标准的 Ornstein-Uhlenbeck 过程。定义过程 $(\overline{x}_{kh})$ 为按与式 (18) 相同的更新规则得到的过程，但其中 $\overline{x}_T$从 $\mathcal{N}(0, \mathrm{Id})$ 而非 $q_0^{\leftarrow}$中采样。令 p 表示 $\overline{x}_0$的分布。假设 $\nabla^2 \ln q_t^{\leftarrow}$关于算子范数是 L_{high}-利普希茨的，并定义

$$\Lambda \triangleq \exp\left(\int_0^T L_{\mathsf{sc},t}\,\mathrm{d}t\right) \qquad \textit{and} \qquad \Lambda' \triangleq \exp\left(\int_0^T L_{\mathsf{sc},\lfloor t/h \rfloor h}\,\mathrm{d}t\right).$$

那么存在一些量 $\mathfrak{C}_1$ 和 $\mathfrak{C}_2$，它们关于 d, $L_{\mathsf{sc},\ *}$, L_{high}, Λ, Λ'和$1/\varepsilon$ 呈多项式有界，使得

$$\mathsf{TV}(p, q) \le \varepsilon + \sqrt{\mathsf{KL}\left(q \| \mathcal{N}(0, \mathrm{Id})\right)}\,\exp(-T)$$

由 $\ell \ge \mathfrak{C}_1$ 和 $\ell h \le \mathfrak{C}_2^{-1}$给出。

4.3 证明概述

我们的离散化分析采用插值式论证，类似于对数凹采样文献中使用的方法 [VW19、CEL+21、WY22] 以及近期一些关于基于分数的生成建模的分析 [LLT22a、LLT22b、CLL22]。这里我们介绍这一论证的设定，并重点说明分析常微分方程而非随机微分方程时所体现出的关键技术差异。

我们先考虑一个一般情形，给定如下两个随机过程 $(y_t)^{t\in[0,T]}$ 和 $(y'_t)^{t\in[0,T]}$ 。过程 (y_t) 由任意常微分方程给出

$$\mathrm{d}y_t = \mu_t(y_t)\,\mathrm{d}t\,. \tag{22}$$

最终我们将取 μ_t 为 $-f_{T-t}+\frac{1}{2}g(T-t)^2\nabla \ln q_{T-t}$，使得 (22) 是与 (1) 中前向过程相关联的概率流常微分方程。过程 $(y'_t$ 的构造方法是先对 (y_t) 作离散时间近似，例如通过以下更新规则

$$y'_{(k+1)h} = y'_{kh} + h\cdot \mu'_{kh}(y_{kh})$$

对所有整数 $k=0,\ 1,\ \ldots\ldots,\ T/h$。最终我们将取 μ'_{kh} 为 $-f_{T-kh}+\frac{1}{2}g(T-kh)^2\nabla \ln q_{T-kh}$，再加上来自 (13) 中近似以及取 $\ell\to\infty$ 所产生的误差项（这些误差项的显式形式见附录 B）。

然后，为了对所有实数值 $t\in[0,T]$ 获得 y'_t，我们考虑这些迭代点的线性插值：如果 $k=\lfloor t/h\rfloor$，那么我们定义 $y_t = y_{kh}+(t-kh)\mu'_{kh}(y_{kh})$。我们将其写为

$$\mathrm{d}y'_t = \mu'_{kh}(y'_{kh})\,\mathrm{d}t\,.$$

假设这两个过程都以相同的分布初始化，即 $y_0,\ y'_0\sim\pi$ 对于某个概率测度 π 定义在 $\mathbb{R}^d$上，那么我们希望根据 t控制 y_t 上的边缘分布与 y'_t 上的边缘分布之间的统计距离。将这些分布分别记为 π_t 和π'_t，我们证明下面这个一般性界，它是我们工作的技术核心。首先，我们对这两个过程作出如下假设。当我们将这些过程具体化为（$x_t^{\leftarrow}$）和（$\widetilde{x}_t^{\leftarrow}$）时，这些假设将由假设 1推出：

假设 2。对于所有 $0\le t\le T$，存在参数 L_t、L'_t、$M\ge 1$ 和 $\zeta_t>0$，使得：

1. $\nabla\ln\pi_t$和 μ_t是 L_t-利普希茨。2. $\nabla\mu_t$ 在算子范数下是 M-利普希茨。
3. μ'_t 是 L'_t-利普希茨。4. $\mathbb{E}[\|\mu_t(y'_t) - \mu'_{kh}(y'_{kh})\|^2]\le\zeta_t^2$。5. 对所有 $0\le t\le T,\quad h\le 1/2L'_t$。

我们简要地结合最终应用来解释这些假设：我们的应用是界定离散化过程的误差。在该应用中，条件 1 和 2 适用于真实连续过程。前者直接源于我们对真实过程的边缘分布具有（二阶）光滑性的标准假设。后者直接源于我们对三阶光滑性的假设；这一假设强于分析反向 随机微分方程 所需的条件，但很可能是分析反向常微分方程所必需的。

条件3 和4 是我们最终将为离散化过程（见第D节）确立的性质。粗略而言，它们规定离散化概率流 ODE 中的漂移项是利普希茨连续的，并且平均而言与真实 ODE 的漂移项接近。

最后，条件5 仅对应于对离散化过程步长的约束。

为方便起见，我们还将定义以下量

$$L \triangleq \max_t L_t, \quad L' \triangleq \max_t L'_t, \quad \zeta^2 \triangleq \int_0^T \zeta_t^2 \,\mathrm{d}t, \quad \Lambda \triangleq \exp\Big(\int_0^T L_t \,\mathrm{d}t\Big), \quad \Lambda' \triangleq \exp\Big(\int_0^T L'_t \,\mathrm{d}t\Big). \tag{23}$$

本节的主要结果是给出 π'_T 与 π_T之间 KL 散度的一个界：

定理 4.3。

$$\mathsf{KL}\left(\pi'_T \| \pi_T\right) \lesssim \Lambda^{O(1)} L'^{1/2} \zeta^2 + (\Lambda^{O(1)} + \Lambda'^{O(1)})(L_0'^{1/2} d^{1/2} + M d T^{1/2})\, \zeta T^{1/2}\,.$$

证明这一结果的关键是，在整个 $t \in [0,T]$上对 KL（$\pi'_t \| \pi_t$）的时间导数给出一个一致的界，然后通过积分即可得到 KL（$\pi'_t \| \pi_t$）的界。

利用两个过程的福克-普朗克方程给出的 π'_t、π_t密度的时间导数，可以显式地计算该时间导数：

$$\begin{aligned} \partial_t \pi_t &= -\mathrm{div}(\pi_t \cdot \mu_t), \\ \partial_t \pi'_t &= -\mathrm{div}(\pi'_t \cdot \widehat{\mu}_{t,kh})\,, \end{aligned}$$

对于 $\widehat{\mu}_{t,\ kh}(x) \quad \mathbb{E}[\mu'_{kh}(y'_{kh}) \quad | \ y'_t = x]$。其中 $\widehat{\mu}_{t,\ kh}$ 是给定位置时刻 kh、在未来时刻 t的漂移的条件期望。计算（见引理 C.4）可得

$$\partial_t \mathsf{KL}\left(\pi'_t \| \pi_t\right) = \int \pi'_t \langle \nabla \ln \frac{\pi'_t}{\pi_t}, \widehat{\mu}_{t,kh} - \mu_t \rangle\,. \tag{24}$$

到目前为止，这些都是标准步骤。然而，此处我们的分析不同于插值方法的通常应用。实际上，如果驱动 y_t 和y'_t 的常微分方程是配备了额外布朗运动项的随机微分方程，那么式（24）将额外包含一个负项，该项为费舍尔信息的某个倍数，作用于 π'_t 和 π_t之间。用公式表示，这意味着我们将不再使用式（24），而是有

$$\partial_t \mathsf{KL}\left(\pi'_t \| \pi_t\right) = \int \pi'_t \langle \nabla \ln \frac{\pi'_t}{\pi_t}, \widehat{\mu}_{t,kh} - \mu_t \rangle - C \int \pi'_t \|\nabla \ln \frac{\pi'_t}{\pi_t}\|^2, \tag{25}$$

其中某个 $C > 0$ 取决于布朗运动的量。式（25）中费舍尔信息项的优势在于，我们可以应用杨氏不等式，方便地将上述量上界为以下量的某个倍数：

$$\int \pi'_t \, \|\widehat{\mu}_{t,kh} - \mu_t\|^2, \tag{26}$$

从而完全避免处理 $\nabla \ln \frac{\pi'_t}{\pi_t}$。粗略地说，量（26）对应于离散过程在时刻 kh 的漂移与连续过程在时刻 t的漂移之间的期望平方差。如果前一过程在时刻 kh 和 t之间的变化不太大，并且漂移 μ'_{kh} 和 μ_t 在平均意义下足够接近，则该量很小。我们在第D 节中验证了概率流常微分方程满足这两个条件。

在常微分方程设定下，情况更加棘手。为处理（24），我们改为应用柯西-施瓦茨不等式，得到

$$\partial_t \mathsf{KL}\left(\pi'_t \| \pi_t\right) \le \left(\int \pi'_t \|\nabla \ln \frac{\pi'_t}{\pi_t}\|^2\right)^{1/2} \cdot \left(\int \pi'_t \|\widehat{\mu}_{t,kh} - \mu_t\|\right)^{1/2},$$

此后，主要的技术难点在于确保右侧的第一项（同样对应于 π'_t 与 π_t之间的费舍尔信息）不会随 t 而爆炸。在引理 C.5和 C.7中，我们展示了如何将该量的时间导数多项式地界定在若干问题特定参数（例如维度以及 μ_t的光滑性）之内。综上，我们得到以下界限。技术细节留待补充材料给出，下面简要概述如何控制这些量的时间导数：

引理 4.4 （参见引理 C.5 和 C.7）。 对于所有 $0 \le t \le T$,

$$\mathbb{E}_{\pi'_t}[\|\nabla \ln \pi'_t\|^2] \lesssim \Lambda'^{O(1)}(L'_0 d + M^2 d^2 t)$$
$$\mathbb{E}_{\pi'_t}[\|\nabla \ln \pi_t\|^2] \lesssim \Lambda^{O(1)}(L'_0 d + M^2 d^2 t + L'\zeta^2)$$

证明概要。 计算 $\partial_t \int \pi'_t \|\nabla \ln \pi'_t\|^2$时，出现的一项是 $\partial_t \ln \pi'_t$。利用 π'_t的福克-普朗克方程，我们可以推导出 $\partial_t \ln \pi'_t$ 的表达式（见命题 C.6）。结合分部积分计算可得

$$\begin{aligned}\partial_t \int \pi'_t \|\nabla \ln \pi'_t\|^2 &= -2 \int \pi'_t \big(\langle \nabla \operatorname{div} \widehat{\mu}_{t,kh}, \nabla \ln \pi'_t \rangle + (\nabla \ln \pi'_t)^\top (\nabla \widehat{\mu}_{t,kh})(\nabla \ln \pi'_t)\big) \\ &\lesssim \sup_x \|\nabla \widehat{\mu}_{t,kh}(x)\|_{\mathsf{op}} \int \pi'_t \|\nabla \ln \pi'_t\|^2 + \int \pi'_t \|\nabla \operatorname{div} \widehat{\mu}_{t,kh}\|^2,\end{aligned}$$

其中最后一步用到了杨氏不等式。 μ_t 的利普希茨连续性使我们能够对 $\sup\|\nabla \widehat{\mu}_{t,\ kh}\|_{\mathsf{op}}$进行界定，而 μ_t 的高阶光滑性则使我们能够对 $\|\nabla \operatorname{div} \widehat{\mu}_{t,\ kh}\|$进行界定。 □

这是分析中唯一出现三阶导数的部分，也是用到假设 1 的第1部分（对应于假设 2 的第2部分）的地方。上述论证中有一点较为微妙：如何从真实漂移 μ_t的光滑性，推导出$\widehat{\mu}_{t,\ kh}$（一个形式复杂的条件期望）的光滑性。为联系二者，我们利用这样一个事实：当步长 h足够小时，离散时间过程是可逆的（引理 C.2），因此$\widehat{\mu}_{t,\ kh}$ 可以表示为 μ'_{kh} 与某个确定性函数的复合。

总而言之，引理所推出的费舍尔信息界 4.4 使我们能够像在随机微分方程情形中一样，将控制 ∂_tKL 散度 $(\pi'_t \| \pi_t)$ 转化为控制式（ 26 ）所刻画的漂移差异；我们将在附录 D 中完成这一步。

5 相关工作

近年来，扩散模型取得了巨大进展，并已超越其他深度生成模型，例如生成对抗网络（GAN）[DN21、SSDK+20、DDDD22、KSS+22]。其应用涵盖蛋白质生成 [AA22、TYT+22、SDH+22、CSJ+22]、医学影像 [JAD+21、AJD+22]、三维数据 [PJBM22] 等诸多领域；全面综述请参见 [YZS+22] 。

随机采样器的非渐近分析， [BMR22，DBTHD21，DB22，LWYL22，LLT22a，Pid22，LLT22b，CLL22，CCL+22] 借鉴了对数凹采样领域丰富文献中的工具（近期综述见 [Che22] ），从而为扩散模型提供了收敛保证。这些研究聚焦于前向过程为 Ornstein-Uhlenbeck 过程、反向过程由一个随机微分方程描述的情形。值得注意的是，[CCL+22、LLT22b、CLL22] 近期的研究表明，在对数据分布 q 作出温和假设（例如光滑且二阶矩有界）的条件下，对反向 SDE 进行适当离散化，并运行多项式数量级的步数，所生成的样本在统计距离上接近数据分布。

在我们的工作之前，对于任何前向过程对应的概率流常微分方程，都没有已知的非渐近界。此前针对扩散模型的非渐近分析并不充分，因为它们要么依赖吉尔萨诺夫定理 [CCL+22] ，要么依赖基于链式法则的变体[CLL22] ，采用 [VW19], 的插值论证；正如我们下面简要解释的那样，这两种方法在确定性采样器情形下都只能得到空洞的界。

非正式地说，吉尔萨诺夫定理不仅可以用于界定算法最终迭代点的分布之间的距离，甚至还可以界定两个过程的轨迹分布之间的距离——请注意，根据数据处理不等式，后者距离的上界也是前者距离的上界。反向过程的每一步都具有随机性，因此即使后者距离也很小。然而，如果没有随机性，该距离将是无穷大。

基于链式法则的论证 [CLL22] 建立了一个与吉尔萨诺夫定理类似的界；特别是，当算法与理想化过程以相同的分布初始化时，这两种论证给出的界是相同的。

最后，我们指出， [LLT22a] 使用了不依赖链式法则的插值论证，但其分析与对数采样文献中关于朗之万蒙特卡洛的现有分析类似，[VW19、CEL+21]利用了算法与理想化过程之间 KL 散度的时间导数表达式中出现的某个费舍尔信息项。正如我们在第 4.3 节所解释的，对于常微分方程，这个费舍尔信息项并不会出现。

6 结论

在本文中，我们从操作层面对概率流常微分方程进行了诠释，将其视为一个由两个步骤迭代构成的过程：通过梯度上升进行恢复，并朝当前迭代点方向进行退化。这一视角也扩展到了包含布朗运动分量的反向过程。我们的操作性诠释与 [BBC+22 中提出的采样器高度一致，DDT+22] 并将去噪扩散隐式模型 [SME21] 的框架推广到了一般的非线性正向过程。

我们工作的主要技术贡献，是对源自该框架的确定性采样器进行了非渐近分析。尽管先前的工作 [CCL+22、LLT22b、CLL22] 在底层反向过程为随机微分方程时，对扩散模型进行了非渐近分析，但据我们所知，我们的分析是常微分方程设定下的首例同类分析。我们的证明基于插值论证，但与该方法以往应用的关键区别在于，采样器的确定性要求我们控制算法与真实反向过程之间费舍尔信息的时间导数。

局限性与未来方向。最显而易见的改进方向，是进一步优化对维度、得分函数的利普希茨性等各种参数的定量依赖关系。直观而言，相较于使用随机微分方程，概率流常微分方程中没有布朗运动，理应带来更好的维度依赖性；但在本文中，我们只能为确定性采样器确立一个关于 d的多项式迭代复杂度。此外，为方便起见，本文忽略了得分估计误差的问题。理论上，可以像 [WY22] 中那样，使用测度变换类论证，在得分估计误差具有次高斯尾部时获得保证；但若要处理像 L_2 中那样仅有一个得分估计误差 [CCL+22，LLT22b，CLL22]。

我们还将一个问题留作开放问题：为了获得概率流常微分方程的非渐近保证，我们关于高阶光滑性的假设是否确有必要。

除了这些技术改进之外，我们还提出了一些值得探索的实证方向。首先，我们的离散化过程引入了许多新的超参数，可以尝试对其进行

调优，以在实践中提升性能。即使对于线性扩散，探索调优 ℓ 的效果也很有意思；在 DDIM 中，目前它被设为 $(T-t)/h$。此外，探索恢复过程中的参数（如学习率和梯度上升的步数），以及使用动量或高阶优化方法，如何帮助构建更好的采样器，也很有意义。我们预计，不同的恢复过程可以还原其他离散化框架，例如 Heun 方法这类二阶方法。从实证角度看，我们预计，优化学习率和步数能够得到计算开销更低、样本质量更高的确定性采样器。

致谢。SC 感谢 Sinho Chewi、Holden Lee、Yuanzhi Li、Jianfeng Lu 和 Adil Salim 围绕确定性基于分数的生成建模和插值方法开展了许多富有启发性的讨论。作者还要感谢 Sinho Chewi、Holden Lee 和 Adil Salim 对本研究早期版本提出的宝贵意见。

本研究得到美国国家科学基金会资助项目 CCF 1763702、AF 1901292、CNS 2148141、TripodsCCF 1934932、IFML CCF 2019844、德克萨斯高级计算中心（TACC）的支持，并获得西部数据、WNCG IAP、德克萨斯大学奥斯汀分校机器学习实验室（MLL）、思科及 Archie Straiton 捐赠教职奖学金提供的研究资助。SC 得到了美国国家科学基金会奖项 2103300 的支持。GD 获得了奥纳西斯奖学金、博多萨基奖学金和莱文蒂斯奖学金的支持。

参考文献

[AA22] Namrata Anand 和 Tudor Achim。使用等变去噪扩散概率模型生成蛋白质结构和序列。arXiv 预印本 arXiv:2205.15019，2022年。[AJD+22] Marius Arvinte、Ajil Jalal、Giannis Daras、Eric Price、Alex Dimakis 和 Jonathan I Tamir。使用基于分数的模型进行磁共振成像重建的单次适配。载于国际磁共振医学学会年会，2022年。[And82] Brian DO Anderson。反向时间扩散方程模型。《随机过程及其应用》，12(3):313–326，1982年。[BBC+22] Arpit Bansal、Eitan Borgnia、Hong-Min Chu、Jie S Li、Hamid Kazemi、Furong Huang、Micah Goldblum、Jonas Geiping 和 Tom Goldstein。冷扩散：无需噪声即可逆转任意图像变换。arXiv 预印本 arXiv:2208.09392，2022年。[BMR22] Adam Block、Youssef Mroueh 和 Alexander Rakhlin。使用去噪自编码器和朗之万采样进行生成建模。arXiv 电子预印本，2022年。[CCGL22] 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。2022年9月。[CCL+22] Sitan Chen、Sinho Chewi、Jerry Li、Yuanzhi Li、Adil Salim 和 Anru R Zhang。采样如同学习分数一样简单：具有最小数据假设的扩散模型理论。arXiv 预印本 arXiv:2209.11215，2022年。[CEL+21] Sinho Chewi、Murat A. Erdogdu、Mufan B. Li、Ruoqi Shen 和 Matthew Zhang。从 Poincar´e 到对数索博列夫条件下的朗之万蒙特卡洛分析。arXiv 电子预印本，2021年。

[Che22] Sinho Chewi。对数凹采样。2022年。书稿草案见 https://chewisinho.github.io/。[CLL22] Hongrui Chen、Holden Lee 和 Jianfeng Lu。改进基于分数的生成建模分析：在最小光滑性假设下给出易用的界。arXiv 预印本 arXiv:2211.01916，2022年。[CSJ+22] Gabriele Corso、Hannes St¨ark、Bowen Jing、Regina Barzilay 和 Tommi Jaakkola。DiffDock：用于分子对接的扩散步骤、扭转与转动。arXiv 预印本 arXiv:2210.01776，2022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。扩散薛定谔桥及其在基于分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。[DDDD22] Giannis Daras、Yuval Dagan、Alexandros G Dimakis 和 Constantinos Daskalakis。分数引导的中间层优化：逆问题的快速朗之万混合。arXiv 预印本 arXiv:2206.09104，2022年。[DDT+22] Giannis Daras、Mauricio Delbracio、Hossein Talebi、Alexandros G Dimakis 和 Pey-man Milanfar。软扩散：针对一般性扰动的分数匹配。arXiv 预印本 arXiv:2209.05442，2022年。[DN21] Prafulla Dhariwal 和 Alexander Nichol。扩散模型在图像合成方面胜过生成对抗网络。《神经信息处理系统进展》，34:8780–8794，2021年。[Efr11] Bradley Efron。Tweedie 公式与选择偏差。《美国统计学会杂志》，106(496):1602–1614，2011年。.[F85] Hans F¨ollmer。用熵方法研究扩散过程的时间反演。载于《随机微分系统》（马赛-吕米尼，1984），《控制信息科学讲义》第69卷，第156–163页。施普林格，柏林，1985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。[JAD+21] Ajil Jalal、Marius Arvinte、Giannis Daras、Eric Price、Alexandros G Dimakis 和 Jon Tamir。利用深度生成先验实现鲁棒压缩感知磁共振成像。《神经信息处理系统进展》，34:14938–14954，2021年。[KAAL22] Tero Karras、Miika Aittala、Timo Aila 和 Samuli Laine。阐明基于扩散的生成模型的设计空间。arXiv 预印本 arXiv:2206.00364，2022年。[KJU22] Mingi Kwon、Jaeseok Jeong 和 Youngjung Uh。扩散模型早已拥有语义潜在空间。arXiv 预印本 arXiv:2210.10960，2022年。

[KSS+22] Dongjun Kim、Seungjae Shin、Kyungwoo Song、Wanmo Kang 和 Il-Chul Moon。《软截断：用于高精度分数估计的基于分数的扩散模型通用训练技术》。发表于国际机器学习会议，页码 11201–11228。PMLR，2022年。[LLT22a] Holden Lee、Jianfeng Lu 和 Yixin Tan。《基于分数的生成建模的多项式复杂度收敛性》。arXiv 电子印本，2022年。[LLT22b] Holden Lee、Jianfeng Lu 和 Yixin Tan。《一般数据分布下基于分数的生成建模的收敛性》。arXiv 预印本 arXiv:2209.12381，2022年。[LWYL22] Xingchao Liu、Lemeng Wu、Mao Ye 和 Qiang Liu。《让我们搭建桥梁：理解并拓展扩散生成模型》。arXiv 预印本 arXiv:2208.14699，2022年。[ND21] Alexander Quinn Nichol 和 Prafulla Dhariwal。《改进的去噪扩散概率模型》。发表于国际机器学习会议，页码 8162–8171。PMLR，2021年。[Pid22] Jakiw Pidstrigach。《基于分数的生成模型能够检测流形》。arXiv 电子印本，2022年。[PJBM22] Ben Poole、Ajay Jain、Jonathan T. Barron 和 Ben Mildenhall。《Dreamfusion：使用二维扩散实现文本到三维生成》。arXiv，2022年。[RBL+22] Robin Rombach、Andreas Blattmann、Dominik Lorenz、Patrick Esser 和 Bj¨orn Ommer。《使用潜在扩散模型进行高分辨率图像合成》。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，页码 10684–10695，2022年。[RDN+22] Aditya Ramesh、Prafulla Dhariwal、Alex Nichol、Casey Chu 和 Mark Chen。《利用 CLIP潜变量进行分层文本条件图像生成》。arXiv 预印本 arXiv:2204.06125，2022年。[SCS+22] Chitwan Saharia、William Chan、Saurabh Saxena、Lala Li、Jay Whang、Emily Denton、Seyed Kamyar Seyed Ghasemipour、Burcu Karagol Ayan、S Sara Mahdavi、Rapha Gontijo Lopes 等。《具备深层语言理解能力的逼真文本到图像扩散模型》。arXiv 预印本 arXiv:2205.11487，2022年。[SDH+22] Arne Schneuing、Yuanqi Du、Charles Harris、Arian Jamasb、Ilia Igashov、Weitao Du、Tom Blundell、Pietro Li´o、Carla Gomes、Max Welling 等。《使用等变扩散模型进行基于结构的药物设计》。arXiv 预印本 arXiv:2210.13695，2022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》。发表于国际机器学习会议，页码 2256–2265。PMLR，2015年。[SE19] Yang Song 和 Stefano Ermon。《通过估计数据分布的梯度进行生成建模》。神经信息处理系统进展，第32卷，2019年。[SE20] Yang Song 和 Stefano Ermon。《改进基于分数的生成模型训练技术》。神经信息处理系统进展，第33卷，页码 12438–12448，2020年。

[SH22] Tim Salimans 和 Jonathan Ho。《用于快速采样扩散模型的渐进式蒸馏》。arXiv 预印本 arXiv:2202.00512，2022 年。[SHG+22] Vikash Sehwag 、 Caner Hazirbas 、 Albert Gordo 、 Firat Ozgenel 和 Cristian Canton。《利用扩散模型从低密度区域生成高保真数据》。发表于 IEEE/CVF计算机视觉与模式识别会议论文集，页码 11492–11501，2022年。[SME21] Jiaming Song、Chenlin Meng 和 Stefano Ermon。《去噪扩散隐式模型》。发表于国际学习表征会议，2021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分数的生成建模》。arXiv 预印本 arXiv:2011.13456，2020年。[TYT+22] Brian L Trippe、Jason Yim、Doug Tischer、Tamara Broderick、David Baker、Regina Barzilay 和 Tommi Jaakkola。《用于基序支架构建问题的三维蛋白质骨架扩散概率建模》。arXiv 预印本 arXiv:2206.04119，2022年。[VW19] Santosh Vempala 和 Andre Wibisono。《未校正朗之万算法的快速收敛：等周性已足够》。发表于神经信息处理系统进展 32，页码 8094–8106。Curran Associates, Inc.，2019年。[WY22] Andre Wibisono 和 Kaylee Yingxi Yang。《非精确朗之万算法在 KL 散度下的收敛性及其在基于分数的生成模型中的应用》。arXiv 预印本 arXiv:2211.01512，2022年。[YZS+22] Ling Yang、Zhilong Zhang、Yang Song、Shenda Hong、Runsheng Xu、Yue Zhao、Yingxia Shao、Wentao Zhang、Bin Cui 和 Ming-Hsuan Yang。《扩散模型：方法与应用综述》。arXiv 预印本 arXiv:2209.00796，2022年。[ZTC22] Qinsheng Zhang、Molei Tao 和 Yongxin Chen。《广义去噪扩散隐式模型》。arXiv 预印本 arXiv:2206.05564，2022年。

A 调整学习率

本节通过考虑数据分布 q 为各向同性高斯分布且前向过程为 Ornstein-Uhlenbeck 过程的特殊情形，来说明在将恢复算子解释为梯度上升时选择学习率（10）的理由。

首先，回顾损失函数 $\ell_{x_t^{\leftarrow}}$在（8）中的定义。一般而言，从 $x_t^{\leftarrow}$出发，使用学习率 η 进行一步梯度上升，得到迭代点

$$x_t^{\leftarrow} + \eta \nabla \ell_{x_t^{\leftarrow}}(x_t^{\leftarrow}) = x_t^{\leftarrow} + \eta\Big(-\frac{1}{g(T-t)^2}\big(\mathrm{Id} + (t-s)\nabla f_{T-s}(x_t^{\leftarrow})\big) f_{T-s}(x_t^{\leftarrow}) + \nabla \ln q_s^{\leftarrow}(x_t^{\leftarrow})\Big)\,. \quad (27)$$

现在假设 $q \sim \mathcal{N}(0, \sigma^2\,\mathrm{Id})$，并且

$$f_t(x) = -\alpha x \qquad \text{and} \qquad g(t) = \beta\sqrt{2}\,.$$

那么， $q_t^{\leftarrow}$由 $\mathcal{N}(0, (e^{-2\alpha t}\sigma^2 + \frac{\beta^2}{\alpha}(1 - e^{-2\alpha t}))\,\mathrm{Id})$ 给出，条件对数似然 $\ln\, q_s^{\leftarrow}(x \mid x_t^{\leftarrow})$ 是关于 x 的二次函数，因此在使 $\nabla \ell_{x_t^{\leftarrow}}$ 为零的 x 处取得最大值。

在这种情况下，(9) 可简化为

$$\nabla \ell_{x_t^{\leftarrow}} \approx \frac{1}{2\beta^2(t-s)}(1+\alpha(s-t))(x_t^{\leftarrow} - x(1+\alpha(s-t))) - \frac{1}{\mathbb{V}[q_s^{\leftarrow}]}\, x\,,$$

其中， $\mathbb{V}[q_s^{\leftarrow}]$ 表示 $q_s^{\leftarrow}$的方差。令右侧为零并求解 x 可得

$$x = \frac{1+\alpha(s-t)}{\mathbb{V}[q_{T-s}] + (1+\alpha(s-t))^2}\, x_t^{\leftarrow} = \Big(1 + \big(\alpha - \frac{2\beta^2}{\mathbb{V}[q_{T-s}]}\big)\cdot(t-s) + O(|t-s|^2)\Big)\, x_t^{\leftarrow} \tag{28}$$

是 $\nabla \ell_{x_t^{\leftarrow}}$的一个（近似）驻点。

经过一次梯度步后的下一个迭代点（27）可简化为

$$x_t^{\leftarrow} + \eta \nabla \ell_{x_t^{\leftarrow}}(x_t^{\leftarrow}) \approx x_t^{\leftarrow} - \frac{\eta}{2\beta^2(t-s)}(1+\alpha(s-t))\cdot\alpha(s-t)\cdot x_t^{\leftarrow} - \frac{\eta}{e^{-2\alpha s}\sigma^2 + \frac{\beta^2}{\alpha}(1-e^{-2s})}\, x_t^{\leftarrow}.$$

最后，我们注意到，取

$$\eta \triangleq 2\beta^2(t-s)\,, \tag{29}$$

上式可简化为

$$x_t^{\leftarrow} - (1+\alpha(s-t))\cdot\alpha(s-t)\cdot x_t^{\leftarrow} - \frac{2\beta^2(t-s)}{e^{-2\alpha s}\sigma^2 + \frac{\beta^2}{\alpha}(1-e^{-2s})}\, x_t^{\leftarrow} = \Big(1 + \big(\alpha - \frac{2\beta^2}{\mathbb{V}[q_{T-s}]}\big)\cdot(t-s) + O(|t-s|^2)\Big)\, x_t^{\leftarrow}\,,$$

该结果与（28）在二阶项意义下相符。因此，当数据服从高斯分布且前向过程为 Ornstein-Uhlenbeck 过程时，如 $t-s \to 0$ 所示，为确保在一阶近似下，单次梯度步将我们从 $x_t^{\leftarrow}$带到条件对数似然 $\ln q_s^{\leftarrow}(x \mid x_t^{\leftarrow}$的最大值点， η 的正确取值由（29）给出；这与正文中的（10）相对应，正如所声称的那样。

B 证明预备知识

设 $h > 0$ 和 $\ell \in \mathbb{N}$ 为离散化参数。定义

$$\delta_\ell \triangleq 1 - \sqrt{1 - 1/\ell} = \frac{1}{2\ell} + O(1/\ell^2)$$

and

$$\xi_\ell = \delta_\ell - \frac{1}{2\ell} = O(1/\ell^2)$$

回顾方程（$\widetilde{x}_{kh}$）$_{k\in\{0,\dots,T/h\}}$ 中过程的定义：18。此处，我们将式（18）中的更新规则改写如下，以便清楚地看出它与 Euler-Maruyama 离散化的相似之处：

$$\begin{aligned}
\widetilde{x}_{(k-1)h} &= z + (\ell-1)h\, f_{(k-\ell)h}(z) + g((k-\ell)h)\sqrt{(\ell-1)h}\cdot\gamma \\
&= z + (\ell-1)h\, f_{(k-\ell)h}(z) + g((k-\ell)h)\sqrt{(\ell-1)h}\cdot\frac{\widetilde{x}_{kh} - z - \ell h\, f_{(k-\ell)h}(z)}{g((k-\ell)h)\sqrt{\ell h}} \\
&= (1-\delta_\ell)\widetilde{x}_{kh} + \delta_\ell z + (\ell\delta_\ell - 1)h f_{(k-\ell)h}(z) \\
&= \widetilde{x}_{kh} + (\ell\xi_\ell - 1/2)\, h f_{(k-\ell)h}(z) \\
&\qquad - \delta_\ell\big(\ell h\, f_{kh}(\widetilde{x}_{kh}) - \ell h\, g(kh)^2 \nabla \ln q_{kh}(\widetilde{x}_{kh})\big).
\end{aligned}$$

注意，$\delta_\ell \eta_k = hg(kh)^2\,(1/2 + \ell\xi_\ell)$，因此我们可以进一步将其改写为

$$
\begin{aligned}
&= \widetilde{x}_{kh} + (\ell\xi_\ell - 1/2)\, h f_{(k-\ell)h}(z) \\
&\quad - h(1/2 + \ell\xi_\ell)\big(f_{kh}(\widetilde{x}_{kh}) - g(kh)^2 \nabla \ln q_{kh}(\widetilde{x}_{kh})\big) \\
&= \widetilde{x}_{kh} - h\left\{f_{kh}(\widetilde{x}_{kh}) - \frac{1}{2} g(kh)^2 \nabla \ln q_{kh}(\widetilde{x}_{kh})\right\} + v_{kh}^{(1)}(\widetilde{x}_{kh}) + \cdots + v_{kh}^{(3)}(\widetilde{x}_{kh}),
\end{aligned}
$$

其中超额项为

$$
\begin{aligned}
v_{kh}^{(1)}(\widetilde{x}_{kh}) &\triangleq \ell\xi_\ell h f_{(k-\ell)h}(z) \cdot \mathbb{1}[k \geq \ell] \\
v_{kh}^{(2)}(\widetilde{x}_{kh}) &\triangleq \frac{h}{2}(f_{(k-\ell)h}(z) - f_{kh}(\widetilde{x}_{kh})) \cdot \mathbb{1}[k \geq \ell] \\
v_{kh}^{(3)}(\widetilde{x}_{kh}) &\triangleq h\ell\xi_\ell\big(-f_{kh}(\widetilde{x}_{kh}) + g(kh)^2 \nabla \ln q_{kh}(\widetilde{x}_{kh})\big)\mathbb{1}[k \geq \ell] \cdot .
\end{aligned}
$$

注意，随着 $\ell \to \infty$ 和 $h\ell \to 0$，超额项趋于零，过程 $(\widetilde{x}_{kh})$ 收敛到由 Euler-Maruyama 离散化给出的过程。

在后续章节中，我们通过插值论证对这一点进行定量分析。令 $(\widetilde{x}_t)_{0 \leq t \leq T}$ 表示离散过程 $(\widetilde{x}_{kh})_{h=0,\ \ldots,\ T/h}$ 的线性插值，并令 $(\widetilde{x}_t^{\leftarrow})$ 表示时间反演过程 $\widetilde{x}_t^{\leftarrow}, \widetilde{x}_{T-t}$。具体而言，对于任意 $kh \leq t < (k+1)h$，

$$
\begin{aligned}
\mathrm{d}\widetilde{x}_t^{\leftarrow} = -\Big\{ & f_{T-kh}(\widetilde{x}_{kh}^{\leftarrow}) - \frac{1}{2} g(T-kh)^2 \nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow}) \\
& - \frac{1}{h}\left(v_{T-kh}^{(1)}(\widetilde{x}_{kh}^{\leftarrow}) + \cdots + v_{T-kh}^{(3)}(\widetilde{x}_{kh}^{\leftarrow})\right)\Big\}\,\mathrm{d}t. \qquad (30)
\end{aligned}
$$

我们注意到，即使没有上述超额项（此时上述过程就是概率流常微分方程的 Euler-Maruyama 离散化），也没有已有研究给出非渐近分析，证明这种离散化以多项式速度收敛到连续时间概率流常微分方程。接下来的分析既能控制超额项，也能建立这样的非渐近分析。

C 插值论证

本节给出一般性界，说明由离散化常微分方程驱动的分布与由连续时间常微分方程驱动的分布之间的 KL 散度如何随时间变化。在本节中，我们始终考虑由以下常微分方程给出的两个随机过程 $(y_t)_{t\in[0,T]}$ 和 $(y'_t)_{t\in[0,T]}$，它们定义在 $\mathbb{R}^d$ 上：

$$
\mathrm{d}y_t = \mu_t(y_t)\,\mathrm{d}t \qquad (31)
$$

$$
\mathrm{d}y'_t = \mu'_{kh}(y'_{kh})\,\mathrm{d}t, \quad k = \lfloor t/h \rfloor, \qquad (32)
$$

其中，对于某个定义在 $\mathbb{R}^d$ 上的概率测度 π，有 $y_0, y'_0 \sim \pi$。过程 (y'_t) 等价于对一个离散时间过程进行线性插值：该过程通过以下更新，从第 k 个迭代点 y'_{kh} 转移到第 $(k+1)$ 个迭代点 $y'_{(k+1)h}$：

$$
y'_{(k+1)h} = y'_{kh} + h\,\mu'_{kh}(y'_{kh})\,.
$$

我们令 π_t, π'_t 分别表示 y_t, y'_t 的分布。当我们最终应用本节得到的估计时，将令 (y'_t) 表示对概率流常微分方程进行离散化所得的过程，并令 (y_t) 表示连续时间下真实的概率流常微分方程。

本节中的界在满足假设 2的条件下成立，为方便起见，现重述如下：

假设 2. 对所有 $0 \le t \le T$，存在参数 L_t、L'_t、$M \ge 1$ 和 $\zeta_t > 0$，使得：

1. $\nabla \ln \pi_t$ 和 μ_t 均为 L_t-利普希茨。
2. $\nabla \mu_t$ 在算子范数下为 M-利普希茨。
3. μ'_t 为 L'_t-利普希茨。
4. $\mathbb{E}[\|\mu_t(y_t) - \mu'_{kh}(y'_{kh})\|^2] \le \zeta_t^2$。
5. 对所有 $0 \le t \le T$，$h \le 1/2L'_t$。

为方便起见，我们还回顾一下在（23）中定义的量：

$$L \triangleq \max_t L_t, \quad L' \triangleq \max_t L'_t, \quad \Lambda \triangleq \exp\Big(\int_0^T L_t \,\mathrm{d}t\Big), \quad \Lambda' \triangleq \exp\Big(\int_0^T L'_t \,\mathrm{d}t\Big), \quad \zeta^2 \triangleq \int_0^T \zeta_t^2 \,\mathrm{d}t$$

并重述关于 π'_T 和 π_T之间 KL 散度的主要界 ：

定理 4.3。

$$\mathsf{KL}\left(\pi'_T \| \pi_T\right) \lesssim \Lambda^{O(1)} L'^{1/2} \zeta^2 + (\Lambda^{O(1)} + \Lambda'^{O(1)})(L_0'^{1/2} d^{1/2} + M d T^{1/2})\, \zeta T^{1/2}\,.$$

例 1。这里我们给出一个简单的例子，说明当 (y_t) 对应于反转从高斯分布出发的 Ornstein-Uhlenbeck 过程的概率流常微分方程时，Λ' 随 d 和 L' 呈多项式增长，而非指数增长。

定义 $\pi_t^{\rightarrow}$：对于 $0 \le t \le T$，将从 $\mathcal{N}(0, \frac{1}{L}\mathsf{Id})$ 出发运行 Ornstein-Uhlenbeck 过程 t 时间后得到的边缘分布记为 $\pi_t^{\rightarrow}$，其中 L 足够大，并考虑相应的反向常微分方程

$$\mathrm{d}y_t = (y_t + \nabla \ln \pi_t(y_t))\,\mathrm{d}t,$$

其中 $\pi_t, \pi_{T-t}^{\rightarrow}$ 表示 $(y_t)_{t\in[0,T]}$ 的边缘分布。具体而言， π_t 由 $\mathcal{N}(0, \frac{1}{L_t}\mathsf{Id})$ 给出，对于$L_t = (e^{-2(T-t)}/L + 1 - e^{-2(T-t)})^{-1}$。注意，

$$\Lambda' = \exp\Big(\int_0^T L_t \,\mathrm{d}t\Big) = \exp\Big(\frac{1}{2}\ln(1 + (e^{2T} - 1)L)\Big).$$

由于 $\mathsf{KL}\left(\mathcal{N}(0, \frac{1}{L}\mathsf{Id}) \| \mathcal{N}(0, \mathsf{Id})\right) = \frac{d}{2}(\ln\, L - 1 + \frac{1}{\ell}) \lesssim d \ln\, L$，我们必须让前向过程运行 $T \approx \frac{1}{2}\ln(d \ln\, L)$ 时间，才能使 $\pi_T^{\rightarrow}$ 接近 $\mathcal{N}(0, \mathsf{Id})$。在这种情况下， $\Lambda' \lesssim \sqrt{dL}$ ~~$\ln L$~~。

我们首先推导 (π'_t) 和 (π_t) 所对应的福克-普朗克方程。

命题 C.1. (π'_t) 和 (π_t) 的分布满足

$$\begin{aligned} \partial_t \pi_t &= -\mathrm{div}(\pi_t \cdot \mu_t) \\ \partial_t \pi'_t &= -\mathrm{div}(\pi'_t \cdot \widehat{\mu}_{t,kh}), \end{aligned}$$

其中

$$\widehat{\mu}_{t,kh}(x) \triangleq \mathbb{E}[\mu'_{kh}(y'_{kh}) \mid y'_t = x].$$

当上下文能够明确看出 k 时，为简化记号，我们将 $_{kh}$,$\widehat{\mu}_t$记为 $\widehat{\mu}_t$ 。

证明。(π_t) 的福克-普朗克方程为

$$\partial_t \pi_t = -\mathrm{div}(\pi_t \cdot \mu_t).$$

对于插值过程 (π'_t)，在给定时间 kh 的条件下，$(\pi'_t)^{kh \le t < (k+1)h}$ 的福克–普朗克方程记作 $(\pi'_{t|kh})^{kh \le t < (k+1)h}$，其形式为

$$\partial_t \pi'_{t|kh}(x) = -\mathrm{div}_x(\pi'_{t|kh}(x) \cdot \mu'_{kh}(y'_{kh})). \tag{33}$$

若 Π'_{kh} 表示 $\sigma(y'_t \mid 0 \le t \le kh$ 上的概率测度，那么对 (33) 两边关于 Π'_{kh}积分，可得

$$\begin{aligned}
\partial_t \pi'_t(x) &= -\int \mathrm{div}_x(\pi'_t(x \mid \xi) \cdot \mu'_{kh}(y'_{kh}))\, \Pi'_{kh}(\mathrm{d}\xi) \\
&= -\mathrm{div}_x \int \pi'_t(x \mid \xi) \cdot \mu'_{kh}(y'_{kh})\, \Pi'_{kh}(\mathrm{d}\xi) \\
&= -\mathrm{div}_x\big(\pi'_t(x) \int \mu'_{kh}(y'_{kh})\, \Pi'_{kh|t}(\mathrm{d}\xi \mid y'_t = x)\big) \\
&= -\mathrm{div}_x(\pi'_t(x) \cdot \mathbb{E}[\mu'_{kh}(y'_{kh}) \mid y'_t = x]) \\
&= -\mathrm{div}_x(\pi'_t(x) \cdot \widehat{\mu}_{t,kh}(x)).
\end{aligned}$$

□

事实证明，由于我们假设步长 h 足够小（见假设2的条件 5 ），条件期望 $\widehat{\mu}_{t,kh}$ 具有简单的形式。对于任意 k，常微分方程 $\mathrm{d}y'_t = \mu'_{kh}(y'_{kh})\,\mathrm{d}t$ 定义了一个映射 $F_{kh\to t}$：$\mathbb{R}^d \to \mathbb{R}^d$，对任意 $kh \le t \le (k+1)h$ 均有

$$F_{kh\to t}(z) = z + (t - kh)\mu'_{kh}(z)$$

因此，从时间 kh 的 z 出发，运行常微分方程直至时间 t，最终得到 $F_{kh\to t}(z)$。当 h 足够小时，$F_{kh\to t}$ 是可逆的：

引理 C.2. 设 $h \le 1/2L'$。那么，对于任意 z、$z' \in \mathbb{R}^d$，

$$\frac{1}{2}\|z - z'\| \le \|F_{kh\to t}(z) - F_{kh\to t}(z')\| \le \frac{3}{2}\|z - z'\|. \tag{34}$$

特别地，$F_{kh\to t}$ 存在唯一的 2-利普希茨逆映射 $F^{-1}_{kh\to t}$： $\mathbb{R}^d \to \mathbb{R}^d$，因此

$$\widehat{\mu}_{t,kh}(x) = \mu'_{kh}(F^{-1}_{kh\to t}(x)). \tag{35}$$

此外，$\widehat{\mu}_{t,\ kh}$ 是 $O(L'_t)$-利普希茨的。

今后，当 k、h、t 可从上下文中明确得知时，我们将把逆映射 $F^{-1}_{kh\to t}$ 简称为 F^{-1}。

证明。对于第一个界，注意到

$$\|F_{kh\to t}(z) - F_{kh\to t}(z')\| \ge \|z - z'\| - (t - kh)\|\mu'_{kh}(z) - \mu'_{kh}(z')\| \ge (1 - h \cdot L'_{kh})\,\|z - z'\|,$$

因此，(34) 的证明类似以对事实理的第一部分，回顾一下，$'$ $\mathbb{R}^d$ 上的双利普希茨函数是双射，因此 $F_{kh\to t}$ 有唯一的逆函数 $F^{-1}_{kh\to t}$。为说明后一个函数是 2-利普希茨的，对于任意 z_0、z'_0，我们可以在 (34) 的下界中取 $z = F^{-1}_{kh\to t}(z_0)$ 和 $z' = F^{-1}_{kh\to t}(z'_0)$，从而得出 $\frac{1}{2}\|F^{-1}_{kh\to t}(z_0) - F^{-1}_{kh\to t}(z'_0)\| \le \|z_0 - z'_0\|$，如所需。由此可得式 (35)，因为给定 $y'_t = x$ 时，y'_{kh} 的分布是集中在 $F^{-1}_{kh\to t}(x)$ 处的点质量。

唯一仍需验证的部分是 $\widehat{\mu}_{t,kh}$ 的利普希茨连续性。这是因为 $\widehat{\mu}_{t,kh}$ 是一个 L'_t-利普希茨函数与一个 2-利普希茨函数的复合。 □

我们还将使用 μ_t （条件 2，假设2）的三阶光滑性的以下简单推论：

引理 C.3. 对所有 x、$x' \in \mathbb{R}^d$，有

$$\sup\|\nabla \operatorname{div} \mu_t\| \le Md \qquad and \qquad \sup\|\nabla \operatorname{div} \widehat{\mu}_{t,kh}\| \le 2Md$$

证明。第一个界直接由

$$|\operatorname{div} \mu_t(x) - \operatorname{div} \mu_t(x')| = |\operatorname{Tr} \nabla \mu_t(x) - \operatorname{Tr} \nabla \mu_t(x')| \le \|\nabla \mu_t(x) - \nabla \mu_t(x')\|_{\mathsf{tr}} \le Md\|x - x'\|.$$

对于第二个界，注意到

$$|\operatorname{div} \widehat{\mu}_t(x) - \operatorname{div} \widehat{\mu}_t(x')| \le \|\nabla \mu'_{kh}(F^{-1}(x)) - \nabla \mu'_{kh}(F^{-1}(x'))\|_{\mathsf{op}} \le dM\|F^{-1}(x) - F^{-1}(x')\| \le 2Md$$

如所述。 □

我们现在 可以计算 π'_t 与 π 之间 KL 散度的时间导数 $_t$.

引理 C.4。

$$\partial_t \mathsf{KL}\left(\pi'_t \| \pi_t\right) \le \zeta_t \left(\int \pi'_t \|\nabla \ln \pi'_t - \nabla \ln \pi_t\|^2\right)^{1/2}$$

证明。我们可以计算

$$\begin{aligned}
\partial_t \mathsf{KL}\left(\pi'_t \| \pi_t\right) &= \int (\partial_t \pi'_t) \ln \frac{\pi'_t}{\pi_t} + \int \pi'_t \, \partial_t \ln \frac{\pi'_t}{\pi_t} = \int (\partial_t \pi'_t) \ln \frac{\pi'_t}{\pi_t} + \int \pi'_t \, \frac{\partial_t (\pi'_t/\pi_t)}{\pi'_t/\pi_t} \\
&= \int (\partial_t \pi'_t) \ln \frac{\pi'_t}{\pi_t} + \int \pi_t \cdot \frac{\pi_t \partial_t \pi'_t - \pi'_t \partial_t \pi_t}{{\pi_t}^2} \\
&= \int (\partial_t \pi'_t) \ln \frac{\pi'_t}{\pi_t} - \int \frac{\pi'_t}{\pi_t} \partial_t \pi_t \\
&= -\int \operatorname{div}(\pi'_t \cdot \widehat{\mu}_{t,kh}) \, \ln \frac{\pi'_t}{\pi_t} + \int \frac{\pi'_t}{\pi_t} \operatorname{div}(\pi_t \cdot \mu_t) \\
&= \int \pi'_t \, \langle \widehat{\mu}_{t,kh}, \nabla \ln \frac{\pi'_t}{\pi_t} \rangle - \int \pi_t \, \langle \nabla \frac{\pi'_t}{\pi_t}, \mu_t \rangle \\
&= \int \pi'_t \, \langle \nabla \ln \frac{\pi'_t}{\pi_t}, \widehat{\mu}_{t,kh} - \mu_t \rangle.
\end{aligned}$$

由柯西-施瓦茨不等式，便可得到该引理：

$$\int \pi'_t \|\widehat{\mu}_{t,kh} - \mu_t\|^2 = \mathbb{E}_{\pi'_t}[\|\mu'_{kh}(F^{-1}(y'_t)) - \mu_t(y'_t)\|^2] = \mathbb{E}_{\pi'_t}[\|\mu'_{kh}(y'_{kh}) - \mu_t(y'_t)\|^2] \le \zeta_t^2. \qquad \square$$

我们需要控制引理 $\int \pi'_t \|\nabla \ln \pi'_t - \nabla \ln \pi_t\|^2$ 中的费舍尔信息 C.4。为此，我们将在下文的引理 $\int \pi'_t \|\nabla \ln \pi'_t\|^2$ 和 $\int \pi'_t \|\nabla \ln \pi_t\|^2$ 中界定 C.5 和 C.7 的时间导数，并应用三角不等式。

引理 C.5。

$$\partial_t \int \pi'_t \|\nabla \ln \pi'_t\|^2 \lesssim L'_t \int \pi'_t \|\nabla \ln \pi'_t\|^2 + M^2 d^2$$

特别地，根据 Gr ¨onwall 不等式，对任意 $0 \le t \le T$，有

$$\int \pi'_t \|\nabla \ln \pi'_t\|^2 \lesssim \Lambda'^{O(1)} (L'_0 d + M^2 d^2 t) \tag{36}$$

证明。我们有

$$\begin{aligned}\partial_t \int \pi_t' \|\nabla \ln \pi_t'\|^2 &= -\int \operatorname{div}(\pi_t' \cdot \widehat{\mu}_t)\|\nabla \ln \pi_t'\|^2 + \int \pi_t' \partial_t \|\nabla \ln \pi_t'\|^2 \\ &= 2\int \pi_t'\big(\langle \widehat{\mu}_t, (\nabla^2 \ln \pi_t')\nabla \ln \pi_t'\rangle + \langle \partial_t \nabla \ln \pi_t', \nabla \ln \pi_t'\rangle\big) \\ &= 2\int \pi_t'\big(\langle \widehat{\mu}_t, (\nabla^2 \ln \pi_t')\nabla \ln \pi_t'\rangle + \langle \nabla(-\operatorname{div} \widehat{\mu}_t - \langle \nabla \ln \pi_t', \widehat{\mu}_t\rangle), \nabla \ln \pi_t'\rangle\big), \quad (37)\end{aligned}$$

其中最后一步使用了下方命题 C.6 的第一部分。注意到，我们可以将括号中的后一项（见（37））写成

$$\langle -\nabla \operatorname{div} \widehat{\mu}_t - (\nabla^2 \ln \pi_t')\widehat{\mu}_t - (\nabla \widehat{\mu}_t)\nabla \ln \pi_t', \nabla \ln \pi_t'\rangle.$$

在这三项中，第二项恰好与（37）中的第一项相消。综合以上各项可得

$$\begin{aligned}\partial_t \int \pi_t' \|\nabla \ln \pi_t'\|^2 &= -2\int \pi_t'\big(\langle \nabla \operatorname{div} \widehat{\mu}_t, \nabla \ln \pi_t'\rangle + (\nabla \ln \pi_t')^\top (\nabla \widehat{\mu}_t)(\nabla \ln \pi_t')\big) \\ &\lesssim \sup \|\nabla \widehat{\mu}_t\|_{\mathsf{op}} \int \pi_t' \|\nabla \ln \pi_t'\|^2 + \int \pi_t' \|\nabla \operatorname{div} \widehat{\mu}_t\|^2,\end{aligned}$$

其中最后一步使用了杨氏不等式。引理的第一部分由引理 C.2和 C.3推出。对于第二部分，Gr¨on wall 不等式给出

$$\mathbb{E}_{\pi_t'}[\|\nabla \ln \pi_t'\|^2] \le \Lambda'^{O(1)}\Big(\int \pi \|\nabla \ln \pi\|^2 + M^2 d^2 t\Big).$$

最后，我们注意到

$$\int \pi \|\nabla \ln \pi\|^2 = -\int \pi \Delta \ln \pi \le L_0' d \quad (38)$$

由分部积分以及假设 2 的条件 1。 □

我们指出，当边缘分布 $h \to 0$和 $\{\pi'_{T-t}\}_{t\in[0,T]}$分别通过从球对称高斯分布出发运行奥恩斯坦-乌伦贝克过程得到时，引理 C.5是紧的。

在上述证明中，我们需要进行以下计算：

命题 C.6。

$$\begin{aligned}\partial_t \ln \pi_t' &= -\operatorname{div} \widehat{\mu}_{t,kh} - \langle \nabla \ln \pi_t', \widehat{\mu}_{t,kh}\rangle \\ \partial_t \ln \pi_t &= -\operatorname{div} \mu_t - \langle \nabla \ln \pi_t, \mu_t\rangle.\end{aligned}$$

接下来，我们进行与引理 C.5 类似的计算，以界定 $\mathbb{E}_{\pi_t'}[\|\nabla \ln \pi_t\|^2]$的时间导数：

引理 C.7。

$$\partial_t \int \pi_t' \|\nabla \ln \pi_t\|^2 \lesssim L_t \int \pi_t' \|\nabla \ln \pi_t\|^2 + M^2 d^2 + L_t \zeta_t^2.$$

特别地，根据 Gr ¨onwall 不等式，对于任意 $0 \le t \le T$，我们有

$$\mathbb{E}_{\pi_t'}[\|\nabla \ln \pi_t\|^2] \lesssim \Lambda^{O(1)}(L_0' d + M^2 d^2 t + L'\zeta^2) \quad (39)$$

证明。我们有

$$\begin{aligned}\partial_t \int \pi_t' \|\nabla \ln \pi_t\|^2 &= -\int \operatorname{div}(\pi_t' \cdot \widehat{\mu}_t)\|\nabla \ln \pi_t\|^2 + \int \pi_t' \partial_t \|\nabla \ln \pi_t\|^2 \\ &= 2\int \pi_t'\big(\langle \widehat{\mu}_t, (\nabla^2 \ln \pi_t)\nabla \ln \pi_t\rangle + \langle \partial_t \nabla \ln \pi_t, \nabla \ln \pi_t\rangle\big) \\ &= 2\int \pi_t'\big(\langle \widehat{\mu}_t, (\nabla^2 \ln \pi_t)\nabla \ln \pi_t\rangle + \langle \nabla(-\operatorname{div} \mu_t - \langle \nabla \ln \pi_t, \mu_t\rangle), \nabla \ln \pi_t\rangle\big), \quad (40)\end{aligned}$$

其中在最后一步中，我们使用了命题 C.6 的第二部分。注意，括号中（40）的后一项可以写成

$$\langle -\nabla \operatorname{div} \mu_t - (\nabla^2 \ln \pi_t)\mu_t - (\nabla \mu_t)\nabla \ln \pi_t, \nabla \ln \pi_t\rangle.$$

这三项中，第二项几乎与（40）中的第一项相消。综合以上各项，可得不等式

$$\begin{aligned}\partial_t \int \pi_t' \|\nabla \ln \pi_t\|^2 = -2\int \pi_t'\big(&\langle \nabla \operatorname{div} \mu_t, \nabla \ln \pi_t\rangle + (\nabla \ln \pi_t)^\top (\nabla \mu_t)(\nabla \ln \pi_t) \\ &+ (\mu_t - \widehat{\mu}_t)^\top (\nabla^2 \ln \pi_t)\nabla \ln \pi_t\big) \\ \lesssim \sup\|\nabla \mu_t\|_{\mathsf{op}} &\int \pi_t' \|\nabla \ln \pi_t\|^2 + \int \pi_t' \|\nabla \operatorname{div} \mu_t\|^2 \\ &+ 2\sup\|\nabla^2 \ln \pi_t\|_{\mathsf{op}} \Big(\int \pi_t' \|\nabla \ln \pi_t\|^2\Big)^{1/2} \Big(\int \pi_t' \|\mu_t - \widehat{\mu}_t\|^2\Big)^{1/2} \\ \lesssim L_t &\int \pi_t' \|\nabla \ln \pi_t\|^2 + \int \pi_t' \|\nabla \operatorname{div} \mu_t\|^2 + L_t \int \pi_t' \|\mu_t - \widehat{\mu}_t\|^2\end{aligned}$$

其中在倒数第二步和最后一步中，我们使用了杨氏不等式；在最后一步中，我们使用了假设 2 的条件 1。引理的第一部分由引理 C.2 以及假设 2的条件4推出。引理的第二部分由 Gr¨onwall 不等式和（38）推出。 □

现在，我们可以结合引理 C.4、C.5 和 C.7 来证明定理 4.3:

定理 *4.3* 的证明。 根据三角不等式以及方程（36）和（39），

$$\Big(\int \pi_t' \|\nabla \ln \pi_t' - \nabla \ln \pi_t\|^2\Big)^{1/2} \lesssim (\Lambda^{O(1)} + \Lambda'^{O(1)})(L_0'^{1/2} d^{1/2} + M d t^{1/2}) + \Lambda^{O(1)} L'^{1/2} \zeta_t,$$

因此，对 $t \in [0, T]$ 积分引理 C.4 中的界，可得

$$\mathsf{KL}\left(\pi_T' \| \pi_T\right) \lesssim (\Lambda^{O(1)} + \Lambda'^{O(1)})(L_0'^{1/2} d^{1/2} + M d T^{1/2}) \int_0^T \zeta_t \, \mathrm{d}t + \Lambda^{O(1)} L'^{1/2} \zeta^2 \,.$$

最后，我们通过柯西-施瓦茨不等式对 $\int_0^T \zeta_t \, \mathrm{d}t \le \zeta T^{1/2}$ 进行界定。 □

最后，我们给出一个范数界，它将在后续内容中发挥作用：

引理 C.8. 对任意 $0 \le t \le T$ 和任意 $c > 0$,

$$\partial_t \, \mathbb{E}\|y_t'\|^2 \le \mathbb{E}\|\mu_{kh}'\|^2 + \mathbb{E}\|y_t'\|^2 \,.$$

证明. 回忆一下， $y_t' = y_{kh}' + (t - kh)\, \mu_{kh}'(y_{kh}')$，因此

$$\mathbb{E}\|y_t'\|^2 = \mathbb{E}\|y_{kh}'\|^2 + (t - kh)^2 \, \mathbb{E}\|\mu_{kh}'(y_{kh}')\|^2 + 2(t - kh)\, \mathbb{E}\langle y_{kh}', \mu_{kh}'(y_{kh}')\rangle \,.$$

对 t 求导，得到

$$\partial_t \mathbb{E}\|y_t'\|^2 = 2(t - kh)\, \mathbb{E}\|\mu_{kh}'(y_{kh}')\|^2 + 2\, \mathbb{E}\langle y_{kh}', \mu_{kh}'(y_{kh}')\rangle = 2\, \mathbb{E}\langle y_t', \mu_{kh}'(y_{kh}')\rangle \,,$$

因此，根据杨氏不等式即可得到该引理。 □

D 漂移差异的界定

我们希望应用定理 4.3 ，其中 (y_t) 和 (y'_t) 分别由 $(x_t^{\leftarrow})$ 和 $(\widetilde{x}_t^{\leftarrow})$ 给出，并定义于方程（3）和（30）中。对于这些过程，方程 (μ_{kh}) 和 (μ'_t) 以及（31）和（32）中的漂移分别为

$$\begin{aligned}\mu_t(x) &\triangleq -f_{T-t}(x) + \frac{1}{2}g(T-t)^2\nabla \ln q_t^{\leftarrow}(x)\\ \mu'_{kh}(x) &\triangleq -f_{T-kh}(x) + \frac{1}{2}g(T-kh)^2\nabla \ln q_{kh}^{\leftarrow}(x) - \frac{1}{h}(v_{T-kh}^{(1)}(x) + \cdots + v_{T-kh}^{(3)}(x))\,,\end{aligned} \tag{41}$$

且两个过程均以分布 $\pi = q_T$ 作为初始分布。一般而言，前一个过程的边缘分布 (π_t) 为 $(q_t^{\leftarrow})$。我们将后一个过程的边缘分布 (π'_t) 记为 (p_t)。

D.1 漂移的光滑性

现在我们验证假设 2 的前三个部分。

引理 D.1。 假设 *1* 的第 *2* 部分成立，其中

$$L_t \triangleq \Theta(L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},t})\,.$$

证明。根据假设 1 的第 5 部分， $\nabla \ln q_t^{\leftarrow}$ 是 $L_{\mathsf{sc},\,t}$-利普希茨的。由于 μ_t 是一个 $L_{f;\mathsf{x}}$-利普希茨函数与一个 $\frac{1}{2}g_{\max}^2 L_{\mathsf{sc},\,t}$-利普希茨函数之和，因此结论成立。 □

引理 D.2. 假设 *2*的第 *2* 部分成立，且

$$M \triangleq (1 + g_{\max}^2/2)L_{\mathsf{high}} = \Theta(g_{\max}^2 L_{\mathsf{high}})\,.$$

证明。 由假设 1 的第 6部分可知， $\nabla\mu_t$ 是一个 L_{high}-利普希茨函数与一个 $g_{\max}^2 L_{\mathsf{high}}/2$-利普希茨函数之和。 □

引理 D.3. 对所有整数 $\ell \le k \le T/h$，恢复算子 $R_{kh\to(k-\ell)h}$ 是 $O(1)$-利普希茨的。证明。对于任意 x、x'，有

$$\begin{aligned}\|R_{kh\to(k-\ell)h}(x) - R_{kh\to(k-\ell)h}(x')\| &\le \|x-x'\| + \ell h\,\|f_{kh}(x) - f_{kh}(x')\|\\ &\quad + \ell h\, g(kh)^2\,\|\nabla \ln q_{kh}(x) - \nabla \ln q_{kh}(x')\|\\ &\le (1 + \ell h L_{f;\mathsf{x}} + \ell h g_{\max}^2 L_{\mathsf{sc},kh})\,\|x-x'\| \lesssim \|x-x'\|.\end{aligned}$$ □

引理 D.4. $\frac{1}{h}(v_1 + \cdots + v_3)$ 是 $O(L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},kh})$-利普希茨的。

证明。由引理 D.3， $f_{(k-\ell)h}(z) = f_{(k-\ell)h}(R_{kh\to(k-\ell h)}(\widetilde{x}_{kh}^{\leftarrow})$ 是一个 $L_{f;\mathsf{x}}$-利普希茨函数与一个关于 $\widetilde{x}_{kh}^{\leftarrow}$ 的 $O(1)$-利普希茨函数的复合，因此 $\frac{1}{h}v_1$ 是 $O(\ell\xi_\ell L_{f;\mathsf{x}})$-利普希茨的。同理， $\frac{1}{h}v_2$ 是一个 $O(L_{f;\mathsf{x}})$-利普希茨函数与一个关于 $\widetilde{x}_{kh}^{\leftarrow}$ 的 $L_{f;\mathsf{x}}/2$-利普希茨函数之差，因此它是 $O(L_{f;\mathsf{x}})$-利普希茨的。最后，$\frac{1}{h}v_3$ 是一个 $\ell\xi_\ell L_{f;\mathsf{x}} \ll L_{f;\mathsf{x}}$-利普希茨函数与一个 $g_{\max}^2 L_{\mathsf{sc},kh}$-利普希茨函数之和，因此它是 $(L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},kh}$-利普希茨的。 □

引理 D.5. μ'_{kh} 如 (41) 中定义的 $O(L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},kh})$-利普希茨。特别地，假设 2 的第 3部分成立，其中

$$L'_t \triangleq \Theta(L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},kh})$$

对所有 $kh \le t < (k+1)h$ 均成立。

证明。注意到 $f_{T-kh}(\cdot) - \frac{1}{2}g(T-kh)^2\nabla \ln q_{kh}^{\leftarrow}(\cdot)$ 是 $O(L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},\,kh})$-利普希茨的，因此由引理D.4可得结论。 □

D.2 漂移之间的距离

我们的离散化分析主要致力于验证假设2的第4部分。为方便起见，我们将 $v^{(1)}_{T-kh}(z)\ldots,v^{(3)}_{T-kh}(z)$ 记作 $v_1,\ldots,v_3$。此后，假设

$$h \ll \min((RL_{f;\mathsf{x}}\ell)^{-1},(g^2_{\max}L_{\mathsf{sc},*})^{-1}) \tag{42}$$

对于任意 $kh \le t \le (k+1)h$，我们有

$$\begin{aligned}\mathbb{E}\|\mu_t(\widetilde{x}^{\leftarrow}_t)-\mu'_{kh}(\widetilde{x}^{\leftarrow}_{kh})\|^2 \lesssim{}& \mathbb{E}\|f_{T-t}(\widetilde{x}^{\leftarrow}_t)-f_{T-kh}(\widetilde{x}^{\leftarrow}_{kh})\|^2\\ &+\mathbb{E}\|g(T-t)^2\nabla\ln q^{\leftarrow}_t(\widetilde{x}^{\leftarrow}_t)-g(T-kh)^2\nabla\ln q^{\leftarrow}_{kh}(\widetilde{x}^{\leftarrow}_{kh})\|^2\\ &+\frac{1}{h^2}(\mathbb{E}\|v_1\|^2+\cdots\mathbb{E}\|v_3\|^2)\,.\end{aligned} \tag{43}$$

我们首先对超额项 $v_1,\ldots,v_3$ 进行界定。我们关注 $T-kh \ge \ell h$ 的情形，因为否则根据定义有 $v_1=v_2=v_3=0$。

引理 D.6。

$$\frac{1}{h^2}\mathbb{E}[\|v_1\|^2+\cdots+\|v_3\|^2] \lesssim \varepsilon_1 \max_{k'\in\{0,1,\ldots,T/h\}}\mathbb{E}\|\nabla\ln q^{\leftarrow}_{k'h}(\widetilde{x}^{\leftarrow}_{k'h})\|^2+\varepsilon_2$$

for

$$\begin{aligned}\varepsilon_1 &\triangleq \exp(O(L^2_{f;\mathsf{x}}T))(\ell^{-2}+\ell^2h^2L^2_{f;\mathsf{x}})\,g^4_{\max}\\ \varepsilon_2 &\triangleq \exp(O(L^2_{f;\mathsf{x}}T))(\ell^{-2}+\ell^2h^2L^2_{f;\mathsf{x}})(\mathbb{E}\|\widetilde{x}^{\leftarrow}_0\|^2+R^2+\ell^2h^2L^2_{f;\mathsf{t}})\,.\end{aligned}$$

证明。回顾

$$v_1=\ell\xi_\ell h f_{T-(k-\ell)h}(z)\,,$$

因此我们有

$$\begin{aligned}\mathbb{E}\|v_1\|^2 &= \ell^2\xi^2_\ell h^2\,\mathbb{E}\|f_{T-(k-\ell)h}(z)\|^2\\ &\lesssim \ell^{-2}h^2(\mathbb{E}\|f_{T-(k-\ell)h}(\widetilde{x}^{\leftarrow}_{kh})\|^2+L^2_{f;\mathsf{x}}\,\mathbb{E}\|z-\widetilde{x}^{\leftarrow}_{kh}\|^2)\\ &\lesssim \ell^{-2}h^2(L^2_{f;\mathsf{x}}\,\mathbb{E}\|\widetilde{x}^{\leftarrow}_{kh}\|^2+R^2+L^2_{f;\mathsf{x}}\,\mathbb{E}\|z-\widetilde{x}^{\leftarrow}_{kh}\|^2)\,.\end{aligned} \tag{44}$$

回顾一下，

$$v_2=\frac{h}{2}(f_{T-(k-\ell)h}(z)-f_{T-kh}(\widetilde{x}^{\leftarrow}_{kh}))\,,$$

因此我们有

$$\begin{aligned}\mathbb{E}\|v_2\|^2 &= \frac{h^2}{4}\,\mathbb{E}\|f_{T-(k-\ell)h}(z)-f_{T-kh}(\widetilde{x}^{\leftarrow}_{kh})\|^2\\ &\lesssim h^2(\ell^2h^2L^2_{f;\mathsf{t}}+L^2_{f;\mathsf{x}}\,\|z-\widetilde{x}^{\leftarrow}_{kh}\|^2)\,.\end{aligned} \tag{45}$$

回顾一下，

$$v_3=h\ell\xi_\ell\big(-f_{T-kh}(\widetilde{x}^{\leftarrow}_{kh})+g(T-kh)^2\nabla\ln q^{\leftarrow}_{(k+\ell)h}(\widetilde{x}^{\leftarrow}_{kh})\big)\,,$$

因此我们有

$$\begin{aligned}\mathbb{E}\|v_3\|^2 &= h^2\ell^2\xi^2_\ell\,\mathbb{E}\|-f_{T-kh}(\widetilde{x}^{\leftarrow}_{kh})+g(T-kh)^2\nabla\ln q^{\leftarrow}_{kh}(\widetilde{x}^{\leftarrow}_{kh})\|^2\\ &\lesssim \ell^{-2}h^2(R^2+L^2_{f;\mathsf{x}}\,\mathbb{E}\|\widetilde{x}^{\leftarrow}_{kh}\|^2+g^4_{\max}\,\mathbb{E}\|\nabla\ln q^{\leftarrow}_{kh}(\widetilde{x}^{\leftarrow}_{kh})\|^2)\end{aligned} \tag{46}$$

结合方程（44）、（45）和（46），我们得到

$$\frac{1}{h^2}\mathbb{E}[\|v_1\|^2+\cdots+\|v_3\|^2] \lesssim (\ell^2h^2L_{f;\mathsf{t}}^2+\ell^{-2}R^2)+\ell^{-2}g_{\max}^4\,\mathbb{E}\|\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + \ell^{-2}L_{f;\mathsf{x}}^2\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2+L_{f;\mathsf{x}}^2\,\mathbb{E}\|z-\widetilde{x}_{kh}^{\leftarrow}\|^2\,. \tag{47}$$

回顾式（19）可知

$$z=\widetilde{x}_{kh}-\ell h\left(f_{kh}(\widetilde{x}_{kh})-g(T-kh)^2\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh})\right),$$

so

$$\begin{aligned}\|z-\widetilde{x}_{kh}^{\leftarrow}\|^2 &\lesssim \ell^2h^2(1+\ell^2h^2L_{f;\mathsf{x}}^2)\,\|f_{T-kh}(\widetilde{x}_{kh}^{\leftarrow})\|^2+\ell^2h^2g_{\max}^4\,\|\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2\\ &\lesssim \ell^2h^2\,(L_{f;\mathsf{x}}^2\,\|\widetilde{x}_{kh}^{\leftarrow}\|^2+R^2)+\ell^2h^2g_{\max}^4\,\|\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2,\end{aligned} \tag{48}$$

其中，第二步使用了式（42）。将其代入式（47），并利用下文的引理 C.8 对 $\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2$进行界定，即可得到所需的界。 □

引理 D.7. 对任意整数 $0\le k\le T/h$ 以及任意 $kh\le t<(k+1)h$,

$$\mathbb{E}\|\mu_t(\widetilde{x}_t^{\leftarrow})-\mu'_{kh}(\widetilde{x}_{kh}^{\leftarrow})\|^2\lesssim\varepsilon_1'\max_{k'\in\{0,1,\ldots,T/h\}}\mathbb{E}\|\nabla\ln q_{k'h}^{\leftarrow}(\widetilde{x}_{k'h}^{\leftarrow})\|^2+\varepsilon_2'$$

for

$$\varepsilon_1'\triangleq\varepsilon_1+h^2L_g^2+g_{\max}^4(h^2L_{f;\mathsf{x}}^2+h^2g_{\max}^4L_{\mathsf{sc},*}^2+g_{\max}^4\beta^2h^{2c})\cdot\exp(O(L_{f;\mathsf{x}}^2T)) \tag{49}$$

$$\begin{aligned}\varepsilon_2'\triangleq{}&\varepsilon_2+g_{\max}^4\beta^2h^{2c}+(\mathbb{E}\|\widetilde{x}_0^{\leftarrow}\|^2+R^2+\ell^2h^2L_{f;\mathsf{t}}^2)\\ &\times(h^2L_{f;\mathsf{x}}^2+h^2g_{\max}^4L_{\mathsf{sc},*}+g_{\max}^4\beta^2h^{2c})\cdot\exp(O(L_{f;\mathsf{x}}^2T))\,.\end{aligned} \tag{50}$$

特别地，对于任意 $\delta>0$，若

$$\begin{aligned}&\ell\gtrsim\delta^{-1/2}(g_{\max}^2+R+\ell hL_{f;\mathsf{t}}+\mathbb{E}\|x_0^{\leftarrow}\|^2)\cdot\exp(O(L_{f;\mathsf{x}}^2T))\\ &h\lesssim\min\big\{\mathrm{poly}(L_g,L_{f;\mathsf{t}},R,g_{\max},L_{\mathsf{sc},*},L_{f;\mathsf{x}},\mathbb{E}\|x_0^{\leftarrow}\|^2)^{-1}\,\ell^{-1}\delta^{1/2},(\delta/(g_{\max}^4\beta^2))^{1/2c}\big\}\cdot\exp(O(L_{f;\mathsf{x}}^2T))\,,\end{aligned} \tag{51}$$

则 $\varepsilon_1',\varepsilon_2'\le\delta$。

证明。 利用43式（43）右侧第一项关于 f 在时间和空间上的利普希茨连续性，可以对其进行界定：

$$\mathbb{E}\|f_{T-kh}(\widetilde{x}_{kh}^{\leftarrow})-f_{T-t}(\widetilde{x}_t^{\leftarrow})\|^2\lesssim L_{f;\mathsf{x}}^2\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}-\widetilde{x}_t^{\leftarrow}\|^2+h^2L_{f;\mathsf{t}}^2\,.$$

对于 (43) 右侧的第二项，我们可以利用 g^2 和分数的利普希茨连续性：

$$\begin{aligned}&\mathbb{E}\|g(T-t)^2\,\nabla\ln q_t^{\leftarrow}(\widetilde{x}_t^{\leftarrow})-g(T-kh)^2\,\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2\\ &\quad\lesssim h^2L_g^2\,\mathbb{E}\|\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2+g(T-t)^4\,\mathbb{E}\|\nabla\ln\frac{q_{kh}^{\leftarrow}}{q_t^{\leftarrow}}(\widetilde{x}_{kh}^{\leftarrow})\|^2+g(T-t)^4\,L_{\mathsf{sc},t}^2\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}-\widetilde{x}_t^{\leftarrow}\|^2\\ &\quad\lesssim(h^2L_g^2+g(T-t)^4\beta^2h^{2c})\,\mathbb{E}\|\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2\\ &\qquad+g(T-t)^4\beta^2h^{2c}\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2+g(T-t)^4\beta^2h^{2c}+g(T-t)^4L_{\mathsf{sc},t}^2\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}-\widetilde{x}_t^{\leftarrow}\|^2\\ &\quad\lesssim(h^2L_g^2+g_{\max}^4\beta^2h^{2c})\,\mathbb{E}\|\nabla\ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2\\ &\qquad+g_{\max}^4\beta^2h^{2c}\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2+g_{\max}^4\beta^2h^{2c}+g_{\max}^4L_{\mathsf{sc},t}^2\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}-\widetilde{x}_t^{\leftarrow}\|^2\,.\end{aligned}$$

将上述界代入 (43)，可得

$$\begin{aligned}
&\mathbb{E}\|\mu_t(\widetilde{x}_t^{\leftarrow}) - \mu'_{kh}(\widetilde{x}_{kh}^{\leftarrow})\|^2 \\
&\quad\lesssim (L_{f;\mathrm{x}}^2 + g_{\max}^4 L_{\mathsf{sc},t}^2)\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow} - \widetilde{x}_t^{\leftarrow}\|^2 + (h^2 L_g^2 + g_{\max}^4 \beta^2 h^{2c})\,\mathbb{E}\|\nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 \\
&\qquad + g_{\max}^4 \beta^2 h^{2c}\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2 + g_{\max}^4 \beta^2 h^{2c} + \frac{1}{h^2}\,\mathbb{E}[\|v_1\|^2 + \cdots + \|v_3\|^2].
\end{aligned}$$

应用 $\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow} - \widetilde{x}_t^{\leftarrow}\|^2$ 和 $\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2$ 引理 D.8 和 D.9 中的界，并注意到$L_{f;\mathrm{x}}^2 + g_{\max}^4 L_{\mathsf{sc},\,t}^2 \ll 1/h^2$ 由（42），可知该引理可由引理 D.6 以及方程（ ε_1'， ε_2' （49）、（50）中 的定义推出。注意，在引理陈述中对 ℓ， h 所作的假设界中，我们用 $\mathbb{E}\|x_0^{\leftarrow}\|^2$ 替换了 $\mathbb{E}\|\widetilde{x}_0^{\leftarrow}\|^2$；这是因为这两个量相同。 □

D.3 运动与范数界

引理 D.8。对于任意整数 $0 < k \le T/h$ 以及任意 $kh \le t < (k+1)h$,

$$\begin{aligned}
\mathbb{E}\|\widetilde{x}_t^{\leftarrow} - \widetilde{x}_{kh}^{\leftarrow}\|^2 \lesssim h^2 \cdot \exp(O(L_{f;\mathrm{x}}^2 T))\Big(&\mathbb{E}\|\widetilde{x}_0\|^2 + R^2 + \ell^2 h^2 L_{f;\mathrm{t}}^2 \\
&+ g_{\max}^4 \max_{k\in\{0,1,\ldots,T/h\}} \mathbb{E}\|\nabla \ln q_t^{\leftarrow}(\widetilde{x}_t^{\leftarrow})\|^2\Big) + \mathbb{E}[\|v_1\|^2 + \cdots + \|v_3\|^2]\,.
\end{aligned}$$

证明。根据插值过程的定义,

$$\widetilde{x}_t^{\leftarrow} = \widetilde{x}_{kh}^{\leftarrow} - (t-kh)\,\{f_{T-kh}(\widetilde{x}_{kh}^{\leftarrow}) - \frac{1}{2} g(T-kh)^2 \nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow}) + \frac{1}{h}(v_1 + \cdots + v_3)\},$$

so

$$\mathbb{E}\|\widetilde{x}_t^{\leftarrow} - \widetilde{x}_{kh}^{\leftarrow}\|^2 \lesssim h^2\,\mathbb{E}\|f_{T-kh}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + h^2 g_{\max}^4\,\mathbb{E}\|\nabla \ln\, q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + \mathbb{E}[\|v_1\|^2 + \cdots + \|v_3\|^2]\,.$$

Th使用假设1 的第1 部分以及引理D.9 即可得到，证明

$$\mathbb{E}\|f_{T-kh}(\widetilde{x}_{kh}^{\leftarrow})\|^2 \lesssim \exp(O(L_{f;\mathrm{x}}^2 T))\Big(\mathbb{E}\|\widetilde{x}_0\|^2 + R^2 + \ell^2 h^2 L_{f;\mathrm{t}}^2 + g_{\max}^4 \max_{k\in\{0,1,\ldots,T/h\}} \mathbb{E}\|\nabla \ln q_t^{\leftarrow}(\widetilde{x}_t^{\leftarrow})\|^2\Big)\,,$$

这里用到了 $\exp(O(L_{f;\mathrm{x}}^2 T)) \cdot L_{f;\mathrm{x}}^2 = \exp(O(L_{f;\mathrm{x}}^2 T))$. □

引理 D.9。对于所有 $0 \le t \le T$,

$$\mathbb{E}\|\widetilde{x}_t^{\leftarrow}\|^2 \lesssim \exp(O(L_{f;\mathrm{x}}^2 T))\left(\mathbb{E}\|\widetilde{x}_0^{\leftarrow}\|^2 + R^2 + \ell^2 h^2 L_{f;\mathrm{t}}^2 + g_{\max}^4 \max_{k\in\{0,1,\ldots,T/h\}} \mathbb{E}\|\nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2\right).$$

证明。根据引理C.8,

$$\begin{aligned}
\partial_t\,\mathbb{E}\|\widetilde{x}_t^{\leftarrow}\|^2 &\lesssim \mathbb{E}\|\widetilde{x}_t^{\leftarrow}\|^2 + \mathbb{E}\|f_{T-kh}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + g_{\max}^4\,\mathbb{E}\|\nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + \frac{1}{h^2}\,\mathbb{E}[\|v_1\|^2 + \cdots + \|v_3\|^2] \\
&\lesssim \mathbb{E}\|\widetilde{x}_t^{\leftarrow}\|^2 + L_{f;\mathrm{x}}^2\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2 + g_{\max}^4\,\mathbb{E}\|\nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + \mathbb{E}\|z - \widetilde{x}_{kh}^{\leftarrow}\|^2 + R^2 + \ell^2 h^2 L_{f;\mathrm{t}}^2 \\
&\lesssim \mathbb{E}\|\widetilde{x}_t^{\leftarrow}\|^2 + L_{f;\mathrm{x}}^2\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2 + g_{\max}^4\,\mathbb{E}\|\nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + R^2 + \ell^2 h^2 L_{f;\mathrm{t}}^2\,,
\end{aligned}$$

其中，第二步用到了（47）和 z 在（19）中的定义，第三步用到了（48）以及由（ $\ell h \ll 1$ （42）可知。应用 Gr¨onwall 不等式于反向过程中的时间区间 $t \in [kh,(k+1)\,h]$，可得

$$\begin{aligned}
\mathbb{E}\|\widetilde{x}_t^{\leftarrow}\|^2 &\lesssim \exp(O(h)) \cdot \left((1 + hL_{f;\mathrm{x}}^2)\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2 + h(g_{\max}^4\,\mathbb{E}\|\nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + R^2 + \ell^2 h^2 L_{f;\mathrm{t}}^2)\right) \\
&\lesssim \exp(cL_{f;\mathrm{x}}^2 h)\,\mathbb{E}\|\widetilde{x}_{kh}^{\leftarrow}\|^2 + h\,\exp(O(h)) \cdot (g_{\max}^4\,\mathbb{E}\|\nabla \ln q_{kh}^{\leftarrow}(\widetilde{x}_{kh}^{\leftarrow})\|^2 + R^2 + \ell^2 h^2 L_{f;\mathrm{t}}^2)
\end{aligned}$$

对所有 $t \in [kh,(k+1)h]$ ，某个绝对常数 $c > 0$ 均成立。特别地，该界对$t = (k+1)h$成立。将此迭代 T/h 次，即可得到所需的界。 □

回忆一下 Λ、Λ' 在（21）中的定义。

引理 D.10. 对所有整数 $0 \le k \le T/h$,

$$\mathbb{E}\|\nabla \ln q^{\leftarrow}_{kh}(\widetilde{x}^{\leftarrow}_{kh})\|^2 \lesssim \Lambda^{O(1)}\Big((L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},*})d + g_{\max}^4 L_{\mathsf{high}}^2 d^2 T \\ + L_{\mathsf{sc},*} T \max_{t\in[0,T]} \mathbb{E}\|\mu_t(\widetilde{x}^{\leftarrow}_t) - \mu'_{\lfloor t/h\rfloor h}(\widetilde{x}^{\leftarrow}_{\lfloor t/h\rfloor h})\|^2\Big)$$

证明。 证明由引理 D.1、D.2、D.5以及引理 C.7 中的界可得， ζ_t ,
$\mathbb{E}\|\mu_t\ \widetilde{x}^{\leftarrow}_t\ - \mu'_{kh}\ \widetilde{x}^{\leftarrow}_{kh}\ \|^2$ $\zeta^2 = \int_0^T \zeta_t^2\ \ t \le T$ $_t\,\zeta_t^2$ Λ Λ'
( ) ( ) 且 d max。注意，在 和 的定义中，被积函数中有一个 $L_{f;\mathsf{x}}^2$ 项，尽管其定义中只有一个 $L_{f;\mathsf{x}}$ 项L_t ，见引理 D.1。这样放宽条件是为了吸收上述分析其他部分中出现的 exp($O(L_{f;\mathsf{x}}^2 T)$) 项。 □

D.4 综合起来

定理 *4.1*的证明。 令 $\delta > 0$ 为稍后选取的一个小参数，并假设 h， ℓ 满足 (51)。于是，对引理 D.7 中的界在 $0 \le t \le T$ 上积分，并应用引理 D.10，可得

$$\begin{aligned}\zeta^2 &\triangleq \int_0^T \mathbb{E}\|\mu_t(\widetilde{x}^{\leftarrow}_t) - \mu'_{\lfloor t/h\rfloor h}(\widetilde{x}^{\leftarrow}_{\lfloor t/h\rfloor h})\|^2\,\mathrm{d}t \\ &\lesssim \delta T + \delta\Lambda^{O(1)}\Big((L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},*})dT + (1+g_{\max}^2)^2 L_{\mathsf{high}}^2 d^2 T^2 \\ &\quad + L_{\mathsf{sc},*} T \int_0^T \mathbb{E}\|\mu_t(\widetilde{x}^{\leftarrow}_t) - \mu'_{\lfloor t/h\rfloor h}(\widetilde{x}^{\leftarrow}_{\lfloor t/h\rfloor h})\|^2\,\mathrm{d}t\Big).\end{aligned}$$

前提是

$$\delta \le \frac{1}{2}\Lambda^{-O(1)} L_{\mathsf{sc},*}^{-1} T^{-1},$$

我们可以重新整理，从而得出

$$\zeta^2 \lesssim \delta\Lambda^{O(1)}\left((L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},*})dT + g_{\max}^4 L_{\mathsf{high}}^2 d^2 T^2\right).$$

根据定理4.3,

$$\mathsf{KL}\left(\pi'_T \| \pi_T\right) \lesssim (\Lambda^{O(1)} + \Lambda'^{O(1)})(L_0'^{1/2} d^{1/2} + M d T^{1/2})\,\zeta T^{1/2} + \Lambda^{O(1)} L'^{1/2}\zeta^2$$

我们将取 δ 足够小，使得 $\zeta^2 \le 1$；在这种情况下，通过将 L'_0 上界估计为 L'，上式至多为

$$\begin{aligned}&\lesssim (\Lambda^{O(1)} + \Lambda'^{O(1)})(L'^{1/2} d^{1/2} + M d T^{1/2})\,\zeta T^{1/2} \\ &\lesssim (\Lambda^{O(1)} + \Lambda'^{O(1)})\big((L_{f;\mathsf{x}}^{1/2} + g_{\max} L_{\mathsf{sc},*}^{1/2})d^{1/2} + g_{\max}^2 L_{\mathsf{high}} d T^{1/2}\big) \\ &\qquad \times \big((L_{f;\mathsf{x}}^{1/2} + g_{\max} L_{\mathsf{sc},*}^{1/2})d^{1/2} T^{1/2} + g_{\max}^2 L_{\mathsf{high}} d T\big)\delta^{1/2} T^{1/2} \\ &\lesssim (\Lambda^{O(1)} + \Lambda'^{O(1)})\big((L_{f;\mathsf{x}} + g_{\max}^2 L_{\mathsf{sc},*})dT + g_{\max}^4 L_{\mathsf{high}}^2 d^2 T^2\big)\delta^{1/2} T^{1/2}\end{aligned}$$

我们取 δ ，使得上式至多为目标精度 ε。根据（51），取 h， ℓ 并满足定理陈述中的界限，即可实现这一点。 □