arXiv:2302.08049v1 [math.ST] 16 Feb 2023

欠阻尼朗之万蒙特卡洛的改进离散化分析

Matthew Zhang* Sinho Chewi† Mufan (Bill) Li‡
Krishnakumar Balasubramanian§ Murat A.Erdogdu¶

2023年2月17日

摘要

欠阻尼朗之万蒙特卡洛（ULMC）是一种算法，它利用粒子在势阱中运动时的动量，从未归一化密度中进行采样。我们针对两个核心问题，对 ULMC 进行了新的分析：(1) *Can we obtain improved sampling guarantees beyond strong log-concavity?* (2) *Can we achieve acceleration for sampling?*

对于问题 (1)，此前关于 ULMC 的结果仅在满足对数索伯列夫不等式和严格的海森光滑性条件时才成立。在此，我们通过去除海森光滑性条件，并考虑满足 Poincar´e 不等式的分布，放宽了这些假设。我们的分析实现了当前最优的维数依赖性，并且足够灵活，可以处理弱光滑势函数。作为推论，我们还首次在强对数凹条件下，为不满足海森光滑性的 ULMC 获得了 KL 散度保证；这一结果基于关于欠阻尼朗之万扩散过程中对数 Sobolev 常数的一项新结论。

对于（2），Cao、Lu 和 Wang（2020年）近期取得的突破，通过偏微分方程方法，首次在连续时间下为采样建立了加速结果。我们的离散化分析将他们的结果转化为算法保证，其条件数依赖性确实优于此前关于欠阻尼朗之万蒙特卡罗的研究，不过，离散时间下能否实现完全加速仍有待解决。

（1）和（2）都需要 Rényi 离散化界，这比通常采用的 Wasserstein耦合论证更具挑战性。我们通过一种基于吉尔萨诺夫定理、灵活且易于扩展至更一般情形的离散化分析来解决这一问题。

1 引言

从高维分布中采样 $\pi \propto \exp(-U)$ 于 $\mathbb{R}^d$，当归一化常数未知且仅给定势函数 U 时，在包括经济学、物理学和科学计算在内的多个应用领域中，这一问题的重要性日益凸显 [JP10；Von11；KPB20]。这一问题的近期进展得益于它与优化领域之间的紧密联系，这一联系始于 [JKO98]的开创性工作；参见 [Che23] 了解相关介绍。

鉴于基于动量的算法在优化领域取得了成功 [Nes83], 研究用于采样的基于动量的算法也就顺理成章了。人们希望这类方法能够改善收敛估计对条件数等关键问题参数的依赖 κ、维度 d和误差容限 ϵ。其中一种方法是欠阻尼朗之万蒙特卡罗（ULMC），它是对欠阻尼朗之万扩散（ULD）的一种离散化：

$$\begin{aligned} \mathrm{d}x_t &= v_t \,\mathrm{d}t\,, \\ \mathrm{d}v_t &= -\gamma v_t \,\mathrm{d}t - \nabla U(x_t)\,\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t\,, \end{aligned} \tag{ULD}$$

*多伦多大学计算机科学系和Vector研究所， matthew.zhang@mail.utoronto.ca
†麻省理工学院数学系， schewi@mit.edu
‡多伦多大学统计科学系和Vector研究所， mufan.li@mail.utoronto.ca§加州大学戴维斯分校统计系， kbala@ucdavis.edu
¶多伦多大学计算机科学系和Vector研究所， erdogdu@cs.toronto.edu

其中， $\{B_t\}_{t\geq 0}$ 是标准的 d维布朗运动。ULD 的平稳分布为$\mu(x,v) \propto \exp(-U(x) - \|v\|^2/2)$，特别地， μ 的 x边缘分布就是所需的目标分布 π。因此，通过在离散化时采用较小的步长并进行大量迭代，ULMC 将从 π中生成一个近似样本。

我们还注意到，在 $\gamma = 0$这一极限情形下，欠阻尼朗之万蒙特卡罗与哈密顿蒙特卡洛算法非常相似；已知该算法在某些有限的设定中能够实现加速，并具有更小的离散化误差[Vis21；AGS22；BM22；WW22]。

尽管目前尚无关于欠阻尼朗之万蒙特卡罗的分析能够在采样方面实现加速（即对条件数 κ呈平方根依赖），但已知欠阻尼朗之万蒙特卡罗能够改善对其他参数（如维数 d 和误差容限 ϵ [Che+18a；Che+18b；DR20], ）的依赖，至少在 Wasserstein 度量下的保证是如此。然而，与针对较简单的（过阻尼）朗之万蒙特卡洛算法的广泛研究相比，现有的欠阻尼朗之万蒙特卡罗分析并不容易扩展到 KL 散度和 Ré´nyi 散度等更强的性能度量。因此，欠阻尼朗之万蒙特卡罗结果的适用范围受到限制；参见第1.1节的讨论。

鉴于这些不足，在本研究中，我们提出以下两个问题：

1. 我们能否通过 ULMC 获得超越强对数凹情形的采样保证?
2. 我们能否通过 ULMC 获得采样的加速收敛保证?

1.1 我们的贡献

我们通过为 ULMC 提出一种新的吉尔萨诺夫离散化界来回答上述两个问题。我们的界在强 Rényi 散度度量下成立，并适用于一般假设（尤其是不要求目标分布具有强对数凹性 π，且允许势函数弱光滑）。因此，它为 ULMC 带来了以下新的最先进结果：

- 我们获得了一个 ϵ^2-保证，其 KL散度的迭代复杂度为 $\widetilde{\mathcal{O}}(\kappa^{3/2}d^{1/2}\epsilon^{-1})$，适用于强对数凹且对数光滑的分布，并消除了 [Ma+21]的 Hessian 的 Lipschitz 假设；其中， κ 是该分布的条件数。

- 在对数 Sobolev 不等式（对数 Sobolev 不等式）和 L-光滑势函数的条件下，我们在全变差距离下获得了 ϵ 保证，迭代复杂度为 $\widetilde{\mathcal{O}}(C^{3/2}$ 对数 Sobolev 不等式$L^{3/2}d^{1/2}\epsilon^{-1})$，且同样无需假设海森矩阵满足 Lipschitz 条件。对于这类分布，就维数依赖性而言，这是目前最先进的保证。
- 我们获得了 ϵ^2阶 Rényi 散度度量下的保证，适用于任意阶数，且在 R´中 $[1,2$，迭代复杂度为 $\widetilde{\mathcal{O}}(C^{3/2}$ Poincaré 不等式 $L^{3/2}\,d^2\epsilon^{-1})$；其条件是满足 Poincaré 不等式，且势函数 ´满足 L-光滑；在对数凹条件下，复杂度可改善为 $\widetilde{\mathcal{O}}(C$ Poincaré 不等式$Ld^2\epsilon^{-1})$。这是已知的 ULMC 在这些条件下的首个保证，并且显著改进了 LMC 在相同条件下的相应结果 [Che+21]。
- 在 Poincar´e 情形下，我们还考虑弱光滑势函数（即梯度为 H¨older 连续且系数为 $s \in (0,1]$），这更真实地反映了满足 Poincar´e 不等式的分布所具有的微妙光滑性。

现在，我们结合现有文献，更详细地讨论我们的结果。

较弱假设下的保证。 先前的工作， [Che+18b；DR20；GT20], 要求目标分布具有强对数凹性。而对于基于等周假设开展的研究，据我们所知，只有 [Ma+21]进一步对势函数作出了限制性较强的利普希茨海森条件假设。相比之下，我们对 U的海森矩阵不作此类假设，并在对数索伯列夫不等式（LSI）或更弱的 Poincar´e 不等式（PI）假设下得到结果；已知在后一种情形下，采样分析颇具挑战性[Che+21]。

如上所述，对于满足 LSI 和光滑性假设的分布，我们的采样结果在维数依赖性方面达到了当前最优水平（$d^{1/2}$）；相比之下，此前最优结果对 d [Che+21呈线性依赖；Che+22]。此外，在 Poincar´e 情形下，我们还可以考虑弱光滑势函数，而此前在 ULMC 的背景下尚未研究过这类势函数。

更强度量下的保证。实现这些结果的关键，是我们在 R´enyi 散度度量下进行的离散化分析。事实上，ULD 在 LSI 或 PI 条件下的连续时间收敛结果，分别适用于 KL 或 R´enyi 散度度量；而要将这些保证转换为 ULMC算法的结果，就必须研究 R´enyi 度量下的离散化。这是主要的技术挑战，因为我们不再能够依赖文献中标准的 Wasserstein耦合论证 [Che+18b；DR20]。两个值得注意的例外是：R´enyi 离散化论证 [GT20], ，其维数依赖性并非最优；以及 εKL 散度论证 [Ma+21]，后者要求严格的光滑性假设。

在本文中，我们基于一种新的轨迹上的对数Sobolev不等式（将在下文进一步讨论），首次在不具备Hessian光滑性的情况下，为通过ULMC从强对数凹且对数光滑分布中进行采样提供KL散度保证。

迈向采样加速。 我们的工作还受到 [CLW20], 的突破性成果的启发，该成果首次在连续时间内为ULD实现了加速收敛保证。我们的离散化界使我们能够将这一结果转化为算法保证，并确实改善了ULMC对条件数 κ[1] 的依赖，而以往结果的依赖至少为 $\kappa^{3/2}$；在对数凹情形下，我们的依赖关系关于 κ 为线性。尽管这仍未达到证明采样完全加速的目标（即将其改善至 $\kappa^{1/2}$），但我们的结果进一步表明，有望通过ULMC实现加速。

ULD轨迹上的新对数Sobolev不等式。 最后，为了证明强对数凹情形下的KL散度保证，我们在ULD上建立了一种新的对数Sobolev不等式（命题 14），该结果本身也具有独立的意义。尽管对于过阻尼朗之万扩散，此类结果此前已有研究，但据我们所知，对于欠阻尼版本而言，这是一个新的结果。

1.2 更多相关工作

朗之万蒙特卡洛。 对于标准的 LMC 算法， $\mathcal{W}_2$ 中针对强对数凹测度的非渐近速率估计最早由 [Dal17a], 给出。关于 KL 散度的保证是在对数索伯列夫不等式下由 KL 散度由 [VW19], 提出，该工作建立了一个颇具吸引力的连续时间框架，用于在泛函不等式条件下分析 LMC。尽管困难重重，´该结果仍由 [GT20推广至 Rényi 散度；EHZ22]。与此同时，也有大量文献研究了在耗散性等尾部增长条件下的 KL 散度KL 收敛 [RRT17；EMS18；EH21；Mou+22]，而这些条件通常蕴含泛函不等式。

与当前工作最相关的是， [Che+21] 将 [VW19] 的连续时间方法推广到了 Rényi 散度，并进一步运用吉尔萨诺夫定理提出了一种新颖的离散化分析，该方法也适用于弱光滑势函数。本文基于 [Che+21] 中引入的 Girsanov技术来研究欠阻尼朗之万蒙特卡罗。

欠阻尼朗之万扩散。 ULMC 是欠阻尼朗之万扩散（ULD）的一种离散化。最早由 [Kol34] 和 [H¨or67] 在关于次椭圆性的开创性研究中提出，人们很快认识到，要定量地证明其收敛至平稳分布颇具技术难度，更不用说捕捉任何加速现象了。 [Vil02；Vil09] 的开创性工作发展了次强制性方法，在泛函不等式条件下首次给出了收敛保证；另见 [H´er06；DMS09；DMS15；RS18]。关于 ULD 定性和定量收敛结果的全面讨论，也可参阅 [Ber+22] 及其中的参考文献。

如前所述， [CLW20] 最近的突破在连续时间下实现了 χ_2-散度的加速收敛，条件是目标分布 π 为对数凹分布。这项工作基于 $\mathcal{H}^{-1}$ [Alb+19]所采用的对偶索 Sobole夫空间方法。然而，由于该方法依赖于 L^2 空间的对偶性及其与 Poincaré 不等式´的联系，因此难以推广到 L^p 空间或其他泛函不等式。

其他离散化方法。 此后，针对这一问题已提出许多其他离散化方案[SL19；Li+19；HBE20；FLO21；Mon21；FRS22；JLS23], 不过，迄今为止的所有分析都仅限于 $\mathcal{W}_2$ 距离，且未能在条件数方面实现加速 κ。

[1]在庞加莱不等式的情形下，条件数为 $\kappa := C_{\mathsf{PI}}L$，这与强对数凹情形下的定义一致。

1.3 文章结构

本文其余部分安排如下：第 2 节回顾所需的定义和假设。第 3 节陈述主要结果，并简要概述其证明。第 4 节通过一些示例阐述定理的若干推论。第 5 节简要概述主要结果的证明，最后在第 6 节讨论未来的研究方向。

2 背景

2.1 记号

下文中，我们将使用 $\|\cdot\|$ 表示向量的 2-范数。一般而言，我们只处理在 $\mathbb{R}^d$ 上具有密度的测度；为方便起见，略微滥用记号，将测度与其密度混同。记号 $a = \mathcal{O}(b)$ 表示存在一个绝对常数 $C > 0$，使得 $a \le Cb$；$\widetilde{\mathcal{O}}(\cdot)$ 表示隐藏了对数因子。类似地，如果存在常数 c、$C > 0$ 使得 $cb \le a \le Cb$，我们记作 $a = \Theta(b)$；$\widetilde{\Theta}(\cdot)$ 表示隐藏了对数因子。平稳测度（在位置坐标上）为 $\pi \propto \exp(-U)$，而 U 将称为势函数。我们将使用 $L^2(\pi)$ 表示测试函数 f，其中 $\mathbb{E}_\pi f^2 < \infty$；使用 $\mathcal{H}^1(\pi)$ 表示弱可微的 $L^2(\pi)$ 函数，其中 $\partial_{x_i} f \in L^2(\pi)$。最后，记号 $\lesssim$、$\gtrsim$、$\asymp$ 分别表示 $\le$、$\ge$、$=$，均忽略绝对常数。后续章节将引入更多记号。

2.2 定义与假设

在本小节中，我们将定义相关过程、散度和等周不等式。首先，我们通过以下随机微分方程（SDE）定义 ULMC 算法：

$$\begin{aligned} \mathrm{d}x_t &= v_t \,\mathrm{d}t\,, \\ \mathrm{d}v_t &= -\gamma v_t \,\mathrm{d}t + \nabla U(x_{kh}) \,\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t\,, \end{aligned} \tag{ULMC}$$

其中 $t \in [kh, (k+1)h)$ 对于某个步长 $h > 0$。注意，这种 ULMC 表述可以通过闭式形式积分（见附录 A）。

接下来，我们定义两个概率分布 μ 和 π 在 $\mathbb{R}^d$上的几种距离度量。我们将全变差距离定义为

$$\|\mu - \pi\|_{\mathsf{TV}} \coloneqq \sup |\mu(A) - \pi(A)|\,, \tag{2.1}$$

其中，上确界 取遍博雷尔可测集 $A \subset \mathbb{R}^d$。我们进一步定义 KL 散度 a s

$$\mathsf{KL}(\mu \,\|\, \pi) \coloneqq \int \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \log \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \,\mathrm{d}\pi\,, \tag{2.2}$$

并且，如果 μ 不关于 π绝对连续，则 KL 散度$(\mu \,\|\, \pi) \coloneqq +\infty$。最后，我们定义阶数为 $q > 1$ 的 R´enyi 散度如下：

$$\mathcal{R}_q(\mu \,\|\, \pi) \coloneqq \frac{1}{q-1} \log \int \left|\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right|^q \mathrm{d}\pi\,,$$

类似地，如果 $\mu \not\ll \pi$，则 $\mathcal{R}_q(\mu \,\|\, \pi) \coloneqq +\infty$。对于任意阶数，R´enyi 散度都是 KL 散度的上界，即 KL 散度$(\mu \,\|\, \pi) \le \mathcal{R}_q(\mu \,\|\, \pi)$，且 $\mathcal{R}_q$ 关于 q单调。特别地，当 $q = 2$时，还可得到χ_2 散度，即 $\chi_2(\mu \,\|\, \pi) = \exp(\mathcal{R}_2(\mu \,\|\, \pi)) - 1$。

我们的主要结果是在以下光滑性条件下给出的。

定义1（光滑性）。如果势函数 U 可微，且 ∇U 满足s-H ¨older 连续性条件，则称其为（L, s）-弱光滑的：

$$\|\nabla U(x) - \nabla U(y)\| \le L \,\|x - y\|^s\,, \tag{2.3}$$

对所有 x、$y \in \mathbb{R}^d$，以及某些 $L \ge 0$、$s \in (0, 1]$。在 $s = 1$ 的特殊情形下，我们称该势函数是 L-光滑的，或者称 ∇U 是 L-Lipschitz 的。

We c进行三条分析路线。第一条假设势函数具有强凸性，即：

定义 2（强　　凸性）。若对某个 $m \geq 0$，对所有 $x, y \in \mathbb{R}^d$，势函数 U 都是 m-强凸的　　：

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq \frac{m}{2} \|x - y\|^2 .$$

在上述 $m = 0$ 情形下，我们称 U 是凸的。若势函数 U 是（强）凸的，则称分布 $\pi \propto \exp(-U)$ 是（强）对数凹的。

第二个更一般的假设是对数索伯列夫不等式。

定义 3（对数索伯列夫不等式）。若测度 π 对参数为 $C_{\mathsf{LSI}} > 0$ 的对数索伯列夫不等式（LSI）成立，则对所有 $g \in \mathcal{H}^1(\pi)$：

$$\mathsf{ent}_\pi(g^2) \leq 2C_{\mathsf{LSI}} \, \mathbb{E}_\pi[\|\nabla g\|^2] , \tag{LSI}$$

其中 $\mathsf{ent}_\pi(g^2) := \mathbb{E}_\pi[g^2 \log(g^2 / \mathbb{E}_\pi[g^2])]$。

一个 m-强凸势函数已知满足常数为LSI的（ m^{-1} [BGL14]）。更一般地，我们可以考虑以下较弱的等周不等式，它对应于（LSI）的线性化。

定义 4（Poincar´e 不等式）。测度 π 满足参数为 $C_{\mathsf{PI}} > 0$的 Poincar ´e 不等式，当对所有 $g \in \mathcal{H}^1(\pi)$：

$$\mathsf{var}_\pi(g) \leq C_{\mathsf{PI}} \, \mathbb{E}_\pi[\|\nabla g\|^2] , \tag{PI}$$

其中 $\mathsf{var}_\pi(g) = \mathbb{E}_\pi[|g - \mathbb{E}_\pi[g]|^2]$。

条件（LSI）和（PI）是马尔可夫扩散理论以及采样中对平稳分布所作的标准假设 [BGL14；VW19；Che+21；Che23]。已知一大类目标分布都满足这些条件，例如对数凹分布或某些混合分布 [Che21；CCN21]。

我们将一个 m-强对数凹目标分布的条件数定义为：其势函数为（L，s）-弱光滑时，条件数为 κ，L/m。如果目标分布不满足强凸性，而仅满足（LSI）（分别为（PI）），则条件数分别为 κ，$C_{\mathsf{LSI}}L$（分别为 κ，$C_{\mathsf{PI}}L$）。

最后，我们整理了几个温和的假设，以简化下文界的计算；这些假设也曾出现在以往的研究中。特别请参见 [Che+21，附录 A]。

假设 1。位置坐标的范数期望可由某个常数定量界定，$\mathbb{E}_\pi[\|\cdot\|] \leq \mathfrak{m} = \widetilde{\mathcal{O}}(d)$[2]，其中某个常数为 $\mathfrak{m} < \infty$。此外，我们假设 $\nabla U(0) = 0$ (不失一般性)，并且 $U(0) - \min U = \mathcal{O}(d)$。

3 主要定理

下文中，我们始终令动量的初始分布 ρ_0 等于平稳分布 $\rho \propto \exp(-\|\cdot\|^2/2)$。那么，在假设1 下，我们可以为位置找到一个初始分布 π_0，它是中心高斯分布，其方差在附录D中给出，使得 π_0 相对于KL 散度 $\mathcal{R}_q$具有某个适当有界的初始散度（例如 π）。最后，我们通过从分布 $\mu_0(x, v) = \pi_0(x) \times \rho_0(v)$中采样来初始化欠阻尼朗之万蒙特卡罗，即 x 和 v 相互独立。

3.1 KL 散度和全变差距离中的收敛

为了陈述我们关于 ULMC 在 KL 散度 和 全变差距离方面的结果，我们借助 [Ma+21] 中关于连续时间的以下结果。该结果基于熵型次椭圆性论证，并对坐标进行时间变换（证明见附录 B.1）。

[2]例如，当 $U(x) = \|x\|^\alpha$ 对 $1 \leq \alpha \leq 2$成立时，这一结论成立。

引理 5 (改编自 [Ma+21，命题 1])。 定义李雅普诺夫函数a l

$$\mathcal{F}(\mu' \,\|\, \mu) \triangleq \mathsf{KL}(\mu' \,\|\, \mu) + \mathbb{E}_{\mu'}\left[\left\|\mathfrak{M}^{1/2} \nabla \log \frac{\mu'}{\mu}\right\|^2\right], \quad \textit{where} \quad \mathfrak{M} = \begin{bmatrix} \frac{1}{4L} & \frac{1}{\sqrt{2L}} \\ \frac{1}{\sqrt{2L}} & 4 \end{bmatrix} \otimes I_d \,. \tag{3.1}$$

对于目标分布 π ，其满足 L-光滑性，且满足(对数 Sobolev 不等式) ，参数为 C_{LSI}，令 $\gamma = 2\sqrt{2L}$。则概率分布 μ_t 满足欠阻尼朗之万扩散的

$$\partial_t \mathcal{F}(\mu_t \,\|\, \mu) \le -\frac{1}{10 C_{\mathsf{LSI}} \sqrt{2L}} \mathcal{F}(\mu_t \,\|\, \mu) \,.$$

现在我们将更精确地陈述主要结果。首先，在强对数凹性和光滑性条件下，我们得到以下 KL散度保证。

定理6(KL 散度下的收敛：强对数凹情形)。设势函数 U 为 m-强凸且 L-光滑，并且满足假设*1*。那么，对于

$$h = \widetilde{\Theta}\Big(\frac{\epsilon m^{1/2}}{L d^{1/2}}\Big) \quad \textit{and} \quad \gamma \asymp \sqrt{L},$$

对于$\hat{}\mu_{Nh}$，以下结论成立：以中心高斯分布为初始分布（方差见附录*D*）的欠阻尼朗之万蒙特卡罗第 N次迭代的概率分布：

$$\mathsf{KL}(\hat{\mu}_{Nh} \,\|\, \mu) \le \epsilon^2 \qquad \textit{after} \qquad N = \widetilde{\Theta}\Big(\frac{\kappa^{3/2} d^{1/2}}{\epsilon}\Big) \quad \textit{iterations}.$$

这里，我们说明将KL 散度的误差容限设为 ϵ^2的原因。根据 Pinsker 和 Talagrand 的输运不等式，我们知道 KL 散度的量级为全变差距离[2]， $\mathcal{W}_2^2$。因此，这使得我们能够公平地比较以KL 散度、全变差距离和 $\mathcal{W}_2$为衡量标准的收敛保证。将强凸性假设弱化为（对数 Sobolev 不等式），我们得到一个关于全变差距离的结果。

定理 7 (在 全变差距离 下的收敛（满足 对数 Sobolev 不等式）)。 设势函数满足 L-光滑性，满足常数为C_{LSI}的假设(对数 Sobolev 不等式)，并满足假设 *1*。那么，对于

$$h = \widetilde{\Theta}\Big(\frac{\epsilon}{C_{\mathsf{LSI}}^{1/2} L d^{1/2}}\Big), \quad \textit{and} \quad \gamma \asymp \sqrt{L},$$

以下结论对 $\hat{}\mu_{Nh}$成立，其中 N表示以中心高斯分布（其方差见附录 *D* 中指定）为初始分布的欠阻尼朗之万蒙特卡罗的第 N 次迭代的概率分布：

$$\|\hat{\mu}_{Nh} - \mu\|_{\mathsf{TV}} \le \epsilon \qquad \textit{after} \qquad N = \widetilde{\Theta}\Big(\frac{C_{\mathsf{LSI}}^{3/2} L^{3/2} d^{1/2}}{\epsilon}\Big) \quad \textit{iterations}.$$

3.2 在 $\mathcal{R}_q$ 中的收敛及改善对 κ 的依赖性

为了陈述我们在 $\mathcal{R}_q$ 中的收敛结果，我们还采用了以下技术性假设，该假设源自[CLW20]。

假设 2. $\mathcal{H}^1(\mu) \hookrightarrow L^2(\mu)$ 是一个紧嵌入。其次，假设 U 二阶连续可微，并且对于所有 $x \in \mathbb{R}^d$，有

$$\|\nabla^2 U(x)\| \le \mathfrak{L}\,(1 + \|\nabla U(x)\|)\,.$$

注。 *[Hoo81，定理 3。1]* 表明，如果势函数具有超线性尾部增长，即 $U(x) \propto \|x\|^\alpha$ 对 $\alpha > 1$ 及充分大的 $\|x\|$成立，那么该假设的第一部分总是满足的。如果尾部严格线性，我们则可以构造一个任意接近的、具有超线性尾部的近似；因此，对于本文所考虑的所有目标分布，该条件通常都成立。正如 *[CLW20]*, 中也指出的，上述假设纯粹出于技术原因而提出，很可能并非必要条件。

当梯度满足相同常数的 L-光滑性时，假设的第二部分成立。在凸情形或 $\nabla^2 U$ 有下界的情形中，常数 $\mathfrak{L}$ 不会出现在界中。因此，对于这种情况下的弱光滑势函数，我们可以用二次可微势函数进行近似，从而得到速率估计。

鉴于上述讨论，我们强调，这一额外假设在很大程度上并不妨碍我们结果的适用性。在该假设下，[CLW20] 在连续时间下证明了关于（欠阻尼朗之万扩散）的如下保证。

引理 8 （快速收敛于 L^2；改编自 [CLW20，定理 1]）。 在假设 *2*下，如果 π还满足 （PI），且常数为 C_{PI}，则对于 μ_t 在 μ_0处初始化的 ULD 的概率分布，以下结论成立，其中 $C_0 > 0$ 是绝对常数：

$$\chi_2(\mu_t \,\|\, \mu) \le C_0 \exp\bigl(-\mathfrak{q}(\gamma)\, t\bigr)\, \chi_2(\mu_0 \,\|\, \mu)\,,$$

指数中的系数为

$$\mathfrak{q}(\gamma) := \frac{C_{\mathsf{PI}}^{-1}\gamma}{C_0\,(C_{\mathsf{PI}}^{-1} + R^2 + \gamma^2)}, \tag{3.2}$$

且常数 R 为

$$R = \begin{cases} 0 & \text{if } U \text{ convex}, \\ \sqrt{K} & \text{if } \inf_{x\in\mathbb{R}^d} \nabla^2 U(x) \succeq -K I_d\,, \\ \mathfrak{L}\sqrt{d} & \text{if } \|\nabla^2 U(x)\|_{\mathrm{op}} \le \mathfrak{L}\,(1 + \|\nabla U(x)\|) \text{ for all } x \in \mathbb{R}^d\,. \end{cases}$$

注。 在强对数凹情形下，引理 *8* 实际上给出了阶为 $\sqrt{m}$ 的更快衰减速度，*5*其条件数依赖性为 $m/\sqrt{L}$。

我们的最终结果利用了上述 ULD 的加速收敛保证，并首次建立了 ULMC 在 d维实数空间 R´ényi 散度下的界，同时改善了条件数依赖性。

定理 9 （在 $\mathcal{R}_q$ 下的收敛（PI）。令势函数 （L， s）-弱光滑，满足 （PI），常数为 C_{PI}，并满足假设*1*。此外，还满足附加技术条件假设 *2*。则对于 $\xi \in$ （0，1）

$$h = \tilde{\Theta}\Bigl(\frac{\gamma^{1/(2s)}\epsilon^{1/s}\xi^{1/s}\mathfrak{q}(\gamma)^{1/(2s)}}{L^{1/s}d^{1/2}\,(L\vee d)^{1/(2s)}}\Bigr),$$

以下结论适用于$\hat{}\mu_{Nh}$：从中心高斯分布初始化的欠阻尼朗之万蒙特卡罗第 N次迭代的概率分布（其方差见附录 *D*），其中 $q = 2-\xi \in [1,2)$，且 $\mathfrak{q}$ 由(3.2)定义：

$$\mathcal{R}_q(\hat{\mu}_{Nh} \,\|\, \mu) \le \epsilon^2 \qquad \textit{after} \qquad N = \tilde{\Theta}\Bigl(\frac{L^{1/s}\,d^{1/2}\,(L\vee d)^{1+1/(2s)}}{\gamma^{1/(2s)}\,\epsilon^{1/s}\,\xi^{1/s}\,\mathfrak{q}(\gamma)^{1+1/(2s)}}\Bigr) \quad \textit{iterations}\,.$$

注。最优选择是取 $\gamma \asymp \sqrt{C^{-1}}$ Poincaré 不等式+ R^2。若势函数 U 是凸的，则设 $\gamma \asymp \mathfrak{q}(1/\sqrt{C_{\mathsf{PI}}}) \asymp 1/\sqrt{C_{\mathsf{PI}}}$，据 *[CLW20]*，这是已知的最优选择。因此，在凸且光滑的情形下，迭代复杂度对条件数的依赖性为 κ，优于 κ^2中所见的依赖性 *[Che+21]*。对维度 d 和误差容限 ϵ的依赖性也有所改善。

4 示例

示例 10。我们考虑势函数 $U(x) = \sqrt{1+\|x\|^2}$，它满足常数为 $\mathcal{O}(d)$ [Bob03] 的（PI），且是 (1, 1)-光滑的。假设满足假设 2 的紧嵌入条件，定理 9 表明，在针对 γ 优化后，对于 $\mathcal{R}_{2-\xi}$ 中的 ϵ^2-保证，其~复杂度为 $\mathcal{O}(d^3\xi^{-1}\epsilon^{-1})$；这是因为此处的势函数是对数凹的。在这种情况下，其维数依赖性与采用拒绝采样的近端采样器 [Che+22 推论 8]，的维数依赖性相同，为 $\widetilde{\mathcal{O}}(d^3)$；这优于 [Che+21 定理 8]，，后者对于相同的保证只能达到$\widetilde{\mathcal{O}}(d^4\epsilon^{-2})$。不过，值得注意的是，后两项工作对任意阶的 Rényi 散度都能得到这些结果，并不限于阶数 $q = 2-\xi < 2$；而目前使用我们的欠阻尼朗之万蒙特卡罗结果还无法达到这一点。

例11。考虑一个 m-强对数凹且 L-对数光滑的分布。贝叶斯回归中存在这类非平凡示例（参见例如 [Dal17b，第 6] 节）；我们将考察第一个示例，其中$\pi(x) \propto \exp(-\|x-\boldsymbol{a}\|^2/2)+\exp(-\|x+\boldsymbol{a}\|^2/2)$，其中某个 $\boldsymbol{a}\in\mathbb{R}^d$ 满足：$\|\boldsymbol{a}\|=1/3$。这里，我们的定理6给出的复杂度为 $N=\widetilde{\mathcal{O}}(d^{1/2}\epsilon^{-1})$，从而获得关于 KL 散度的 ϵ^2-保证。相比之下，海森矩阵为$\nabla^2 U(x) = I_d - 4\boldsymbol{a}\boldsymbol{a}^\top \exp(2x^\mathsf{T}\boldsymbol{a})/(1+\exp(2x^\mathsf{T}\boldsymbol{a}))^2,$，其具有 $L_\mathsf{H}\asymp d$，其中 L_H 是海森矩阵在弗罗贝尼乌斯范数下的 Lipschitz 常数。因此，[Ma+21 定理 1] 将其表述为 $N=\widetilde{\mathcal{O}}(d^{1/2}L_H m^{-2}\epsilon^{-1})$；在此情形下，要获得相同的 ϵ^2-精度保证，需要 $N=\widetilde{\mathcal{O}}(d^{3/2}\epsilon^{-1})$。其维数依赖性更差。最后，我们可以与 [GT20 定理 28], 所得到的离散化界进行比较；将其与我们的连续时间结果结合（采用与定理 6 相同的证明技巧），可得到 $N=\widetilde{O}(d^{1/2}\epsilon^{-2})$ 次迭代。其关于 ϵ 的阶数并非最优，但维数依赖性相同。

例12。我们可以分析满足参数为C_LSI的对数索伯列夫不等式的 L-光滑分布。一种情况是考虑强凸势函数的任意有界扰动。在这种情况下，令 $U_{\boldsymbol{a}}$ 为例11中目标分布的势函数。接着，考虑修正势函数为 $U_{\boldsymbol{a}}+f$的目标分布，其中 $\sup_x|f(x)|\vee\|\nabla f(x)\|\vee\|\nabla^2 f(x)\|_{\mathsf{op}}\le\mathfrak{B}$，对某个 $\mathfrak{B}<\infty$成立，并令 $\nabla^2 f$为 $\mathcal{O}(d)$-Frobenius Lipschitz。我们可以利用 Holley–Stroock 引理 [HS87]，对该势函数的对数 Sobolev 常数进行界定。设这个新势函数的条件数为 κ。我们以$N=\widetilde{\mathcal{O}}(\kappa^{3/2}d^{1/2}\epsilon^{-1})$达到 ϵ-精度的全变差距离。相比之下，先前的界 [Ma+21，定理 1] 给出 $N=\tilde{\mathcal{O}}(\kappa^2 d^{3/2}\epsilon^{-1})$，才能达到相同的全变差距离保证，其维度依赖更差。不过请注意，[Ma+21，定理 1] 中的保证是针对 KL 散度的，这比全变差距离更强。最后，我们注意到 [GT20] 要求强对数凹性，因此无法在这种情形下提供保证。

例 13。考虑一个 $(1,s)$-弱对数光滑目标分布，它是对数凹的，并满足带有 $C_\mathsf{PI}=\mathcal{O}(d)$ 的 Poincar´e 不等式。因此，定理 9 给出 $N=\tilde{\mathcal{O}}(d^{2+1/s}\xi^{-1/s}\epsilon^{-1/s})$，从而为 $\mathcal{R}_{2-\xi}$ 提供 ϵ^2-保证。[Che+21 中的定理 7] 则为相同的保证给出 $N=\tilde{\mathcal{O}}(d^{3+2/s}\epsilon^{-2/s})$，在两个参数上都更差。另一方面，考虑势函数为 $U(x)=\|x\|^\alpha$ 的分布这一特定情形，它具有 $C_\mathsf{PI}=\mathcal{O}(d^{2/\alpha-1})$ [Bob03]，，且是对数凸的、$(1,\alpha-1)$-弱对数光滑的。因此，定理 9 给出 $N=\tilde{\mathcal{O}}(d^{\alpha/(\alpha-1)}\xi^{-1/(\alpha-1)}\epsilon^{-1/(\alpha-1)})$，为 $\mathcal{R}_{2-\xi}$ 散度中的 ϵ^2-精度提供保证。与 [Che+21 的例 9], 中得到的速率估计相比，这一结果差了 d 倍，因为他们利用了比泛函不等式更强的一类不等式，可在 (PI) 与 (LSI) 之间进行插值，而我们的分析无法捕捉这一改进。尽管如此，我们的收敛保证在 ϵ 依赖性方面仍然更好。

5 证明概要

5.1 连续时间结果

对于满足 Poincaré 不等式和对数 Sobolev 不等式的结果，我们利用 ´ 中陈述的现有结果 [CLW20；Ma+21]，并在引理 5 和 8 中加以呈现。这些结果使我们能够用指数衰减量来界定 $\chi^2(\mu_t\,\|\,\mu)$、$\mathsf{KL}(\mu_t\,\|\,\mu)$。

在强凸性的附加假设下，我们可以在另一组坐标（ϕ，ψ）$\mathcal{M}(x,\ v)\triangleq(x,\ x+\frac{2}{\gamma}v)$ 下得到压缩（见附录 B）。因此，我们可以在这些替代坐标中考察连续时间迭代序列的分布以及目标分布 $\mu_t^{\mathcal{M}}$，$\mu^{\mathcal{M}}$ 分别。由此，我们得到以下命题。

命题 14 (沿轨迹的对数 Sobolev 不等式)。假设 U 是 m-强凸且 L- 光滑。令 $\mu_t^{\mathcal{M}}$表示连续时间欠阻尼朗之万扩散的概率分布，其中 $\gamma=c\sqrt{L}$对应于 $c\ge\sqrt{2}$ 在 $(\phi,\ \psi)$ 坐标中。假设初始分布 μ_0 满足 (对数 Sobolev 不等式) 常数（在变换后的坐标中）为 $C_\mathsf{LSI}(\mu_0^{\mathcal{M}})$，则 $\{\mu_t^{\mathcal{M}}\}_{t\ge0}$ 满足 (对数 Sobolev 不等式)，且其常数可以一致地上界为

$$C_\mathsf{LSI}(\mu_t^{\mathcal{M}})\le\exp\Big(-m\sqrt{\frac{2}{L}}\,t\Big)\,C_\mathsf{LSI}(\mu_0^{\mathcal{M}})+\frac{2}{m}\,.$$

这一命题的证明的核心思路是分析坐标（ϕ，ψ）下欠阻尼朗之万扩散的离散化（欠阻尼朗之万蒙特卡罗）。注意，这可以写成以下形式

对于某个矩阵 $\overline{\Sigma} \in \mathbb{R}^{2d\times 2d}$ 和函数 $\bar{F}\colon\ \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$,

$$(\phi_{(k+1)h}, \psi_{(k+1)h}) \overset{d}{=} \bar{F}(\phi_{kh}, \psi_{kh}) + \mathcal{N}(0, \overline{\Sigma})\,.$$

¯这是由一个确定性函数 F 和一个高斯分布相加构成的：前者给出从（ϕ, ψ）出发的 ULMC 下一次迭代的均值，后者给出所得迭代的方差。具体而言，我们证明，对于坐标（$\phi(x,v)$ $\psi(x,v)$） $\triangleq (x, x + \frac{2}{\gamma}v)$，可以在 F 意义下找到几乎必然的严格¯压缩，即

$$\|\bar{F}\|_{\mathrm{Lip}} \le 1 - \frac{m}{\sqrt{2L}}\, h + \mathcal{O}(Lh^2)\,,$$

其中，按记号滥用， $\bar{F}\colon\ \mathbb{R}^{2d} \to \mathbb{R}^{2d}$；函数 $g\colon\ \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ 的半范数 $\|g\|_{\mathrm{Lip}}$ 指该函数的 Lipschitz 常数。

¯由于 F 在 h足够小时是压缩映射，每次推前都会将对数 Sobolev 常数乘以一个因子[VW19，引理 19]。与此同时，高斯卷积至多会使对数 Sobolev 常数增加一个加性常数 [Cha04，推论 3。1]。因此，每次迭代的对数 Sobolev 常数构成一个（截断的）几何级数，故可由无穷级数上界控制。顺带一提，这也可用于对 ULMC 迭代序列的对数 Sobolev 常数进行上界估计。对取适当的极限，同时保持不变，即可得到命题中所述的界。因此，考虑 KL 散度的分解，简单应用柯西–施瓦茨不等式可得

$$\begin{aligned}
\mathsf{KL}(\hat{\mu}_t^{\mathcal{M}} \,\|\, \mu^{\mathcal{M}}) &= \int \log \frac{\hat{\mu}_t^{\mathcal{M}}}{\mu^{\mathcal{M}}}\, \mathrm{d}\hat{\mu}_t^{\mathcal{M}} = \mathsf{KL}(\hat{\mu}_t^{\mathcal{M}} \,\|\, \mu_t^{\mathcal{M}}) + \int \log \frac{\mu_t^{\mathcal{M}}}{\mu^{\mathcal{M}}}\, \mathrm{d}\hat{\mu}_t^{\mathcal{M}} \\
&\le \mathsf{KL}(\hat{\mu}_t^{\mathcal{M}} \,\|\, \mu_t^{\mathcal{M}}) + \mathsf{KL}(\mu_t^{\mathcal{M}} \,\|\, \mu^{\mathcal{M}}) + \sqrt{\chi^2(\hat{\mu}_t^{\mathcal{M}} \,\|\, \mu_t^{\mathcal{M}}) \times \mathsf{var}_{\mu_t^{\mathcal{M}}}\Big(\log \frac{\mu_t^{\mathcal{M}}}{\mu^{\mathcal{M}}}\Big)}\,.
\end{aligned}$$

对于 $\mu_t^{\mathcal{M}}$的对数索伯列夫不等式蕴含 Poincar´e 不等式，这使我们能够利用 Fisher 信息 $\mathsf{FI}(\mu_t^{\mathcal{M}} \,\|\, \mu^{\mathcal{M}}) = \mathbb{E}_{\mu_t^{\mathcal{M}}} \|\nabla \log(\mu_t^{\mathcal{M}}/\mu^{\mathcal{M}})\|^2$ 来界定方差项。该项可通过 [Ma+21] 中用于得到我们的 全变差距离 界的相同熵型次椭圆性论证加以界定，而其余两项则分别通过离散化分析和再次运用熵型次椭圆性论证来处理。

5.2 离散化分析

我们用于控制离散化误差的主要结果如下所示。

命题 15。 令 $(\hat{}\mu_t)^{t\ge 0}$ 表示 (欠阻尼朗之万蒙特卡罗) 的概率分布，并令 $(\mu_t)^{t\ge 0}$ 表示连续时间欠阻尼朗之万扩散 (欠阻尼朗之万扩散) 的概率分布，二者均以某个 μ_0 初始化。假设势函数 U 是(L, s)-弱光滑的。若步长 h 满足

$$h = \widetilde{\mathcal{O}}_s\Big(\frac{\gamma^{1/(2s)}\, \epsilon^{1/s}}{L^{1/s}\, T^{1/(2s)}\, (d + \mathcal{R}_2(\mu_0 \,\|\, \mu^{(a)}))^{1/2}}\Big)\,, \tag{5.1}$$

¯其中记号 $\mathcal{O}_s$ 隐藏了依赖于 s 的常数，以及包括 $\log N$在内的多重对数因子，并且$\mu^{(a)}$ 是修正目标分布（详见附录 *C.3* ），则

$$\mathcal{R}_q(\hat{\mu}_T \,\|\, \mu_T) \le \epsilon^2\,.$$

注。 关于 h的条件仅通过对数因子依赖于 N。其次，该结论是在一般假设下证明的，并且可以与 $\mathcal{R}_q$ 中的连续时间结果结合，适用于任意设定，例如 *[Che+21]*中讨论的对数 Sobolev 不等式或 Lata la–Oleszkiewicz 不等式。

下面我们概述该结果的证明。与[Che+21]的工作类似，我们首先运用数据处理不等式，从而可以用迭代序列的时间边缘分布之间的 Rényi 散度上界，来界定路径测度之间的 Rényi 散度´´

$$\mathcal{R}_q(\hat{\mu}_T \,\|\, \mu_T) \le \mathcal{R}_q(P_T \,\|\, Q_T)\,,$$

其中 P_T, Q_T分别表示路径空间上的欠阻尼朗之万蒙特卡罗和（欠阻尼朗之万扩散）的概率测度 $C([0,T],\mathbb{R}^{2d})$。随后，我们运用吉尔萨诺夫定理，它使我们能够通过两个过程的漂移项之差，对路径散度进行精确界定：

$$\mathcal{R}_{2q}(P_T \,\|\, Q_T) \lesssim \log \mathbb{E} \exp\Big(\frac{4q^2}{\gamma} \int_0^T \|\nabla U(x_t) - \nabla U(x_{\lfloor t/h \rfloor h})\|^2 \,\mathrm{d}t\Big) .$$

还需对期望中的项进行界定。为此，我们通过对以下事件进行条件化来实现：$\sup_{t\in[0,T]} \|x_t - x_{\lfloor t/h \rfloor h}\|^2$ 被一个随 $h \to 0$趋于零的量所界定；我们必须证明该事件以足够高的概率发生。为证明这一点，我们首先进行单步分析，即对上述量在 $T \le h$的情形下进行界定。与朗之万蒙特卡洛相比，该分析的主要改进在于，随机微分方程（欠阻尼朗之万扩散）与（欠阻尼朗之万蒙特卡罗）在位置坐标上完全一致，而漂移项之间的差异仅体现在动量上。对动量积分后，位置坐标上的误差阶更高（主导项为 $\mathcal{O}(dh^2)$），相比之下，在 $\mathcal{O}(dh)$ 中观察到的是 [Che+21，引理 24]）。

将该分析从单步扩展到整个时间区间的技术，与 [Che+21]中所采用的方法十分接近。具体而言，我们得到 $\|x_t\|$ 对 $\|x_{kh}\|$ 在区间 $t \in [kh,(k+1)h$上的依赖关系。由于势函数仅满足 Poincar´e 不等式，对后者进行控制相当复杂，因为这等价于证明迭代序列具有次高斯尾界，而目标分布本身在位置坐标上并非次高斯分布。通过与辅助势函数进行比较，我们可以证明，对于所选择的初始化，迭代序列在直到 N 的所有迭代次数内均保持次高斯性（尽管相应常数会不断增大）。最后，这使我们能够重新得到上述命题中的离散化结果。

6 结论

本文在多个情形下，为欠阻尼朗之万蒙特卡洛算法提供了当前最先进的收敛保证。特别是，我们的离散化分析（命题 15）具有普适性，可推广到任意阶的 R´enyi 散度，并适用于势函数满足各种条件（如 Latała–Oleszkiewicz 条件、弱光滑性等）的情形。因此，我们的结果是全面理解 ULMC 算法的重要一步。然而，当前连续时间分析技术的局限性使我们无法获得更强的迭代复杂度结果。更具体地说，我们尚不清楚如何分析阶数大于 2 的 R´enyi 散度，也不清楚当势函数满足对数索伯列夫不等式时，是否可能出现超收缩衰减。其次，我们基于吉尔萨诺夫变换的离散化方法在条件数方面目前仍非最优（这一点已在 [Che+21] 中指出），因此离散化后无法得到预期的 $\sqrt{\kappa}$ 依赖关系。若要改进这一结果，就需要改进证明技巧。我们相信，本文提出的结果和技巧将激发未来的相关研究。

致谢

感谢 Jason M. Altschuler、Alain Durmus 和 Aram-Alexandre Pooladian 与我们进行了富有帮助的交流。KB 得到了美国国家科学基金会资助（DMS-2053918）的支持。SC 得到了美国国家科学基金会 TRIPODS 计划（项目编号 DMS- 2022448）的支持。MAE 得到了加拿大自然科学与工程研究委员会资助 [2019-06167],康诺特新研究者奖、CIFAR 人工智能讲席计划和 CIFAR 人工智能催化剂资助的支持。ML 得到了安大略省研究生奖学金和 Vector研究所的支持。

参考文献

[AGS22] S. Apers、S. Gribling 和 D. Szil´agyi.《用于高效高斯采样的哈密顿蒙特卡罗：长步长与随机步长》。载于：arXiv 预印本 arXiv:2209.12771（2022）。[Alb+19] D. Albritton 等。《动力学福克–普朗克方程的变分方法》。载于：arXiv 预印本 arXiv:1902.04037（2019）。[Ber+22] E. Bernard 等。《利用舒尔补实现次强制性》。载于：Annales Henri Lebesgue 5（2022），第 523–557 页。[BGL14] D. Bakry、I. Gentil 和 M. Ledoux.《马尔可夫扩散算子的分析与几何》。第 348 卷。数学科学基础丛书 [数学科学基本原理]。Springer，Cham，2014，第 xx+552 页。[BLM13] S. Boucheron、G. Lugosi 和 P. Massart.《集中不等式：独立性的非渐近理论》，Michel Ledoux 作序。Oxford University Press，Oxford，2013，第 x+481 页。[BM22] N. Bou-Rabee 和 M. Marsden.《采用分层蒙特卡罗时间积分的未经校正的哈密顿蒙特卡罗马尔可夫链蒙特卡罗》。载于：arXiv 预印本 arXiv:2211.11003（2022）。[Bob03] S. G. Bobkov.《某些球对称概率测度的谱间隙与集中性》。载于：泛函分析的几何方面。第 1807 卷。数学讲义。Springer，Berlin，2003，第 37–43 页。[CCN21] H.-B. Chen、S. Chewi 和 J. Niles-Weed.《混合分布的无维数依赖对数 Sobolev 不等式》。载于：Journal of Functional Analysis 281.11（2021），第 109236 页。[CD76] J. Chandra 和 P. W. Davis.《格罗恩沃尔不等式的线性推广》。载于：Proceedings of the American Mathematical Society 60.1（1976），第 157–160 页。[Cha04] D. Chafai.《熵、凸性与泛函不等式：关于 Φ-熵与 Φ-Sobolev 不等式》。载于：J. Math. Kyoto Univ. 44.2（2004），第 325–363 页。[Che+18a] X. Cheng 等。《非凸情形下朗之万动力学的尖锐收敛速率》。载于：arXiv 预印本 arXiv:1805.01648（2018）。[Che+18b] X. Cheng 等。《欠阻尼朗之万马尔可夫链蒙特卡罗：非渐近分析》。载于：学习理论会议。PMLR。2018，第 300–323 页。[Che+21] S. Chewi 等。《从 Poincar´e到对数 Sobolev 的朗之万蒙特卡罗分析》。载于：arXiv 预印本 arXiv:2112.12662（2021）。[Che+22] Y. Chen 等。《采样近端算法的改进分析》。载于：第三十五届学习理论会议论文集。P.-L. Loh 和 M. Raginsky 编。第 178 卷。机器学习研究论文集。PMLR，2022年7月，第 2984–3014 页。[Che21] Y. Chen.《KLS 猜想中等周系数的近乎常数下界》。载于：Geom. Funct. Anal. 31.1（2021），第 34–61 页。[Che23] S. Chewi.《对数凹采样》。书稿可见于 https://chewisinho.github.io/。2023年。[CLW20] Y. Cao、J. Lu 和 L. Wang.《欠阻尼朗之万动力学的显式 L^2-收敛速率估计》。载于：arXiv 电子预印本，arXiv:1908.04746（2020）。[Dal17a] A. S. Dalalyan.《采样与优化之间更深入且更强的类比：朗之万蒙特卡罗与梯度下降》。载于：2017年学习理论会议论文集。S. Kale 和 O. Shamir 编。第 65 卷。机器学习研究论文集。PMLR，2017，第 678–689 页。[Dal17b] A. S. Dalalyan.《从光滑对数凹密度进行近似采样的理论保证》。载于：Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79.3（2017），第 651–676 页。[DMS09] J. Dolbeault、C. Mouhot 和 C. Schmeiser.《具有线性松弛项的动力学方程的次强制性》。载于：Comptes Rendus Mathematique 347.9-10（2009），第 511–516 页。

[DMS15] J. Dolbeault、C. Mouhot 和 C. Schmeiser。《保持质量的线性动理学方程的次强制性》。载于：《美国数学会汇刊》367.6（2015年），第3807–3828页。[DR20] A. S. Dalalyan 和 L. Riou-Durand。《使用动理学朗之万扩散从对数凹密度中采样》。载于：Bernoulli 26.3（2020年），第1956–1988页。[EH21] M. A. Erdogdu 和 R. Hosseinzadeh。《关于朗之万蒙特卡洛的收敛性：尾部增长与光滑性之间的相互作用》。载于：《第三十四届学习理论会议论文集》。M. Belkin 和 S. Kpotufe 编。第134卷。《机器学习研究论文集》。PMLR，2021年，第1776–1822页。[EHZ22] M. A. Erdogdu、R. Hosseinzadeh 和 S. Zhang。《朗之万蒙特卡洛在卡方散度和R´ényi散度下的收敛性》。载于：《第25届人工智能与统计国际会议论文集》。G. Camps-Valls、F. J. R. Ruiz 和 I. Valera 编。第151卷。《机器学习研究论文集》。PMLR，2022年3月，第8151–8175页。[EMS18] M. A. Erdogdu、L. Mackey 和 O. Shamir。《使用离散化扩散进行全局非凸优化》。载于：《第32届神经信息处理系统国际会议论文集》。2018年，第9694–9703页。[FLO21] J. Foster、T. Lyons 和 H. Oberhauser。《欠阻尼朗之万马尔可夫链蒙特卡洛的移位常微分方程方法》。载于：arXiv预印本 arXiv:2101.03446（2021年）。[FRS22] J. Foster、G. d. Reis 和 C. Strange。《满足交换条件的随机微分方程的高阶分裂方法》。载于：arXiv预印本 arXiv:2210.17543（2022年）。[GT20] A. Ganesh 和 K. Talwar。《通过离散化朗之万马尔可夫链蒙特卡洛的R´ényi散度分析实现更快的差分隐私采样器》。载于：《神经信息处理系统进展》。H. Larochelle 等编。第33卷。Curran Associates, Inc.，2020年，第7222–7233页。[HBE20] Y. He、K. Balasubramanian 和 M. A. Erdogdu。《随机化中点采样方法的遍历性、偏差与渐近正态性》。载于：《神经信息处理系统进展》。H. Larochelle 等编。第33卷。Curran Associates, Inc.，2020年，第7366–7376页。[H´er06] F. H´erau。《线性非齐次弛豫玻尔兹曼方程的次强制性与指数时间衰减》。载于：《渐近分析》46.3-4（2006年），第349–359页。[Hoo81] J. G. Hooton。《有限测度空间上的紧Sobolev嵌入》。载于：《数学分析与应用杂志》83.2（1981年），第570–581页。[H¨or67] L. H¨ormander。《次椭圆二阶微分方程》。载于：《数学学报》119（1967年），第147–171页。[HS87] R. Holley 和 D. Stroock。《对数Sobolev不等式与随机Ising模型》。载于：《统计物理学杂志》46.5-6（1987年），第1159–1194页。[JKO98] R. Jordan、D. Kinderlehrer 和 F. Otto。《福克–普朗克方程的变分形式》。载于：《SIAM数学分析杂志》29.1（1998年），第1–17页。[JLS23] T. Johnston、I. Lytras 和 S. Sabanis。《无需梯度利普希茨连续性的动理学朗之万马尔可夫链蒙特卡洛采样——强凸情形》。载于：arXiv预印本 arXiv:2301.08039（2023年）。[JP10] M. Johannes 和 N. Polson。《连续时间金融计量经济学中的马尔可夫链蒙特卡洛方法》。载于：《金融计量经济学手册：应用》。爱思唯尔，2010年，第1–72页。[Kol34] A. Kolmogorov。《随机运动（布朗运动理论）》。载于：《数学年刊》（1934年），第116–117页。[KPB20] I. Kobyzev、S. J. Prince 和 M. A. Brubaker。《归一化流：当前方法的介绍与综述》。载于：《IEEE模式分析与机器智能汇刊》43.11（2020年），第3964–3979页。[Li+19] X. Li 等。《随机Runge–Kutta方法加速朗之万蒙特卡洛及其他方法》。载于：《神经信息处理系统进展》第32卷（2019年）。

[Ma+21] Y.-A. Ma 等。《是否存在梯度型马尔可夫链蒙特卡洛的Nesterov 加速类似物？》载于：Bernoulli 27.3（2021年），第1942–1992页。[Mir17] I. Mironov。《Rényi差分隐私》。载于：《2017年IEEE第30届计算机安全基础研讨会（CSF）》。IEEE，2017年，第263–275页。[Mon21] P. Monmarché。《采用标准分裂格式的高维马尔可夫链蒙特卡洛与欠阻尼朗之万扩散》。载于：Electronic Journal of Statistics 15.2（2021年），第4117–4166页。[Mou+22] W. Mou 等。《朗之万扩散离散化的改进界：无需凸性的近最优速率》。载于：Bernoulli 28.3（2022年），第1577–1601页。[Nes83] Y. E. Nesterov。《一种求解凸规划问题且具有收敛速率$O(\frac{1}{k^2})$的方法》。载于：Doklady Akademii Nauk。第269卷，第3期。俄罗斯科学院，1983年，第543–547页。[Oks13] B. Oksendal。《随机微分方程：介绍及其应用》。Springer Science & Business Media，2013年。[RRT17] M. Raginsky、A. Rakhlin 和 M. Telgarsky。《通过随机梯度朗之万动力学进行非凸学习：非渐近分析》。载于：《2017年学习理论会议论文集》。S. Kale 和 O. Shamir 编。第65卷。《机器学习研究论文集》。PMLR，2017年，第1674–1703页。[RS18] J. Roussel 和 G. Stoltz。《朗之万动力学的谱方法及相关误差估计》。载于：ESAIM：数学建模与数值分析 52.3（2018年），第1051–1083页。[SL19] R. Shen 和 Y. T. Lee。《对数凹采样的随机化中点方法》。载于：《神经信息处理系统进展》。H. Wallach 等编。第32卷。Curran Associates, Inc.，2019年。[Vil02] C. Villani。《玻尔兹曼方程的流体力学极限》。载于：Astérisque，SMF 282（2002年），第365–405页。[Vil09] C. Villani。《次强制性》。载于：《美国数学会回忆录》202.950（2009年），第iv+141页。[Vis21] N. K. Vishnoi。《用于采样的哈密顿蒙特卡罗方法介绍》。载于：arXiv预印本 arXiv:2108.12107（2021年）。[Von11] U. Von Toussaint。《物理学中的贝叶斯推断》。载于：《现代物理评论》83.3（2011年），第943页。[VW19] S. Vempala 和 A. Wibisono。《未校正朗之万算法的快速收敛：等周

-

足够”。载于：神经信息处理系统进展。H. Wallach 等编，第 32 卷。Curran Associates, Inc., 2019 年。[WW22] J.-K. Wang 和 A. Wibisono。《通过切比雪夫积分时间加速哈密顿蒙特卡罗》。载于：arXiv 预印本 arXiv:2207.02189（2022年）。

A Ex欠阻尼朗之万扩散的显式形式 n

回想一下，我们按照随机微分方程（x_t， v_t）显式地演化时间 $t \in [kh, (k+1)h)$，这里为方便起见再次列出：欠阻尼朗之万蒙特卡罗。

$$\mathrm{d}x_t \triangleq v_t \,\mathrm{d}t\,, \tag{A.1}$$

$$\mathrm{d}v_t \triangleq -\gamma v_t + \nabla U(x_{kh})\,\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t\,. \tag{A.2}$$

因此　　由于我们在非线性项中固定位置 x_{kh}，因此可以得到显式解　　n

$$x_{(k+1)h} = x_{kh} + \gamma^{-1}\left(1-\exp(-\gamma h)\right) v_{kh} - \gamma^{-1}\left(h - \gamma^{-1}\left(1-\exp(-\gamma h)\right)\right)\nabla U(x_{kh}) + W_k^x\,, \tag{A.3}$$

$$v_{(k+1)h} = \exp(-\gamma h)\, v_{kh} - \gamma^{-1}\left(1-\exp(-\gamma h)\right)\nabla U(x_{kh}) + W_k^v\,, \tag{A.4}$$

其中 $(W_k^x, W_k^v)_{k\in\mathbb{N}}$ 是由变量对组成的独立序列，每对变量服从联合分布

$$\begin{bmatrix} W_k^x \\ W_k^v \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \frac{2}{\gamma}\left(h - \frac{2}{\gamma}\left(1-\exp(-\gamma h)\right) + \frac{1}{2\gamma}\left(1-\exp(-2\gamma h)\right)\right) & * \\ \frac{1}{\gamma}\left(1-2\exp(-\gamma h)+\exp(-2\gamma h)\right) & 1-\exp(-2\gamma h) \end{bmatrix}\right),$$

其中 $*$ 与左下角的元素相同。

B 连续时间结果

B.1 熵型次椭圆性

引理 5 的证明基于对 [Ma+21]中李雅普诺夫函数衰减论证的改编 [Vil09]（基于熵型次椭圆性，参见 [DR20, 引理1]），并结合时间变换论证。为完整起见，我们在下面给出详细过程。

引理 5 的证明。首先注意到，变量 x_t， v_t 在满足 $\gamma = 2\sqrt{2L}$ 条件时，按照（欠阻尼朗之万蒙特卡罗）可变换为（$\tilde{x}_t$， $\tilde{v}_t$） = （$x_{t\sqrt{\xi}}$， $\frac{1}{\sqrt{\xi}}\, v_{t\sqrt{\xi}}$），其满足以下过程：

$$\mathrm{d}\tilde{x}_t = \xi\tilde{v}_t\,\mathrm{d}t\,,$$

$$\mathrm{d}\tilde{v}_t = -\xi\tilde{\gamma}\tilde{v}_t\,\mathrm{d}t - \nabla U(\tilde{x}_t)\,\mathrm{d}t + \sqrt{2\tilde{\gamma}}\,\mathrm{d}B_t\,,$$

其中 $\tilde{\gamma} = 2$， $\xi = 2L$，是满足 [Ma+21命题 1]的参数。根据该命题，我们知道，由下式给出的李雅普诺夫泛函

$$\tilde{\mathcal{F}}(\tilde{\mu}' \parallel \tilde{\mu}) = \mathsf{KL}(\tilde{\mu}' \parallel \tilde{\mu}) + \mathbb{E}_{\tilde{\mu}'}\left[\left\|\mathfrak{N}^{1/2}\nabla \log\frac{\tilde{\mu}'}{\tilde{\mu}}\right\|^2\right], \qquad \text{where } \mathfrak{N} = \frac{1}{L}\begin{bmatrix} 1/4 & 1/2 \\ 1/2 & 2 \end{bmatrix} \otimes I_d\,,$$

以 $\partial_t\tilde{\mathcal{F}}(\tilde{\mu}_t \parallel \tilde{\mu}) \le -\frac{1}{10C_{\mathsf{LSI}}}\tilde{\mathcal{F}}(\tilde{\mu}_t \parallel \tilde{\mu})$ 的速率衰减。这里，对数 Sobolev 不等式常数在我们的坐标变换下保持不变，但此时 $\tilde{\mu}_t$ 表示 $(\tilde{x}_t, \tilde{v}_t)$ 的联合概率分布，而平稳测度具有 $\tilde{\mu}(\tilde{x}, \tilde{v}) \propto \pi(\tilde{x}) \times \exp(-\xi\|\tilde{v}\|^2/2)$ 的形式。将变量变换逆转即可立即得到定理的结论，其中动量的梯度按 $\xi^{1/2}$ 放大，而时间按 $\xi^{1/2}$ 缩短。

□

B.2 欠阻尼朗之万蒙特卡罗的压缩

本节中，我们证明 ULMC 的压缩结果，并据此推导欠阻尼朗之万扩散轨迹上的对数 Sobolev 不等式。以 (x, v) 为初始状态的 ULMC，其下一次迭代的均值为

$$F(x,v) \coloneqq \Big(x + \frac{1-\exp(-\gamma h)}{\gamma}\, v - \frac{h - \gamma^{-1}\left(1-\exp(-\gamma h)\right)}{\gamma}\nabla U(x),$$
$$\exp(-\gamma h)\, v - \frac{1-\exp(-\gamma h)}{\gamma}\nabla U(x)\Big)\,.$$

我们将使用以下坐标变换

$$(\phi, \psi) \coloneqq \mathcal{M}(x, v) \coloneqq \left(x, x + \frac{2}{\gamma} v\right).$$

在这些新坐标中，从(ϕ, ψ)出发的欠阻尼朗之万蒙特卡罗下一次迭代的均值为 $\bar{F}(\phi, \psi)$，其中 $\bar{F} = \mathcal{M} \circ F \circ \mathcal{M}^{-1}$。由于 $\mathcal{M}^{-1}(\phi, \psi) = (\phi, \frac{\gamma}{2}(\psi - \phi))$，我们可以明确写出

$$\begin{aligned}\bar{F}(\phi, \psi) = \Big(&\phi + \frac{1 - \exp(-\gamma h)}{2}(\psi - \phi) - \frac{h - \gamma^{-1}(1 - \exp(-\gamma h))}{\gamma} \nabla U(\phi), \\ &\phi + \frac{1 + \exp(-\gamma h)}{2}(\psi - \phi) - \frac{h + \gamma^{-1}(1 - \exp(-\gamma h))}{\gamma} \nabla U(\phi)\Big).\end{aligned}$$

引理 16. 考虑上述定义的映射 $\bar{F}\colon\ \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$。假设 $m I_d \preceq \nabla^2 U \preceq L I_d$。那么，对于 $h \lesssim 1$ 和 $\gamma = c\sqrt{L}$，存在某个 $c \geq \sqrt{2}$，使得 $\bar{F}$ 是一个压缩映射，其参数为

$$\|\bar{F}\|_{\mathrm{Lip}} \leq 1 - \frac{m}{\sqrt{2L}} h + O(Lh^2).$$

证明. 我们计算偏导数

$$\begin{aligned}\partial_\phi \bar{F}(\phi, \psi)_\phi &= \frac{1 + \exp(-\gamma h)}{2} I_d - \frac{h - \gamma^{-1}(1 - \exp(-\gamma h))}{\gamma} \nabla^2 U(\phi), \\ \partial_\phi \bar{F}(\phi, \psi)_\psi &= \frac{1 - \exp(-\gamma h)}{2} I_d - \frac{h + \gamma^{-1}(1 - \exp(-\gamma h))}{\gamma} \nabla^2 U(\phi), \\ \partial_\psi \bar{F}(\phi, \psi)_\phi &= \frac{1 - \exp(-\gamma h)}{2} I_d, \\ \partial_\psi \bar{F}(\phi, \psi)_\psi &= \frac{1 + \exp(-\gamma h)}{2} I_d.\end{aligned}$$

令 $a \coloneqq \exp(-\gamma h)$，且 $b \coloneqq \frac{2}{\gamma}(h + \gamma^{-1}(1 - \exp(-\gamma h)))$。由于

$$\frac{h - \gamma^{-1}(1 - \exp(-\gamma h))}{\gamma} = O(h^2),$$

我们有

$$\|\nabla \bar{F}(\phi, \psi)\|_{\mathrm{op}} \leq \frac{1}{2} \Big\| \underbrace{\begin{bmatrix} (1+a) I_d & (1-a) I_d - b \nabla^2 U(\phi) \\ (1-a) I_d & (1+a) I_d \end{bmatrix}}_{=:A} \Big\|_{\mathrm{op}} + O(Lh^2).$$

于是,

$$AA^\intercal = \begin{bmatrix} (1+a)^2 I_d + ((1-a) I_d - b \nabla^2 U(\phi))^2 & * \\ 2(1-a^2) I_d - (1+a) b \nabla^2 U(\phi) & \{(1-a)^2 + (1+a)^2\} I_d \end{bmatrix},$$

其中右上角的元素由对称性确定。由于 $1 - a = \Theta(\gamma h)$ 和 $b = O(h/\gamma)$，可将其简化如下:

$$\begin{aligned}&\Big\| AA^\intercal - 2 \underbrace{\begin{bmatrix} (1+a^2) I_d & (1-a^2) I_d - b \nabla^2 U(\phi) \\ (1-a^2) I_d - b \nabla^2 U(\phi) & (1+a^2) I_d \end{bmatrix}}_{=:B} \Big\|_{\mathrm{op}} \\ &\quad \leq O\Big(\frac{L^2 h^2}{\gamma^2} + Lh^2\Big).\end{aligned}$$

可以验证，矩阵 B 的特征值为 $1+a^2 \pm (1-a^2-b\lambda)$，其中 λ 遍历 $\nabla^2 U(\phi)$ 的特征值。因此，我们可以对其进行界定

$$\|B\|_{\mathrm{op}} \le \max\{2a^2 + Lb, 2 - bm\}\,.$$

我们注意到

$$\begin{aligned} 2a^2 + Lb &= 2\exp(-2\gamma h) + \frac{2L\left(h + \gamma^{-1}\left(1-\exp(-\gamma h)\right)\right)}{\gamma} \\ &= 2\left\{1 - 2\gamma h + \frac{2Lh}{\gamma} + O(\gamma^2 h^2 + Lh^2)\right\}. \end{aligned}$$

为了使其严格小于 2，我们必须取 $\gamma > \sqrt{L}$。对于 $c \ge \sqrt{2}$，我们选择 $\gamma = c\sqrt{L}$，此时

$$\begin{aligned} \|B\|_{\mathrm{op}} &\le 2\max\left\{1 - c\sqrt{L}\,h,\ 1 - m\sqrt{\frac{2}{L}}\,h\right\} + O(Lh^2) \\ &= 2\left(1 - m\sqrt{\frac{2}{L}}\,h\right) + O(Lh^2)\,. \end{aligned}$$

由此可得

$$\|AA^{\mathsf{T}}\|_{\mathrm{op}} \le 4\left(1 - m\sqrt{\frac{2}{L}}\,h\right) + O(Lh^2)$$

因此

$$\|\nabla \bar{F}(\phi,\psi)\|_{\mathrm{op}} \le \sqrt{1 - m\sqrt{\frac{2}{L}}\,h} + O(Lh^2) \le 1 - \frac{m}{\sqrt{2L}}\,h + O(Lh^2)\,.$$

□

ULMC 迭代为

$$(x_{(k+1)h}, v_{(k+1)h}) \overset{d}{=} F(x_{kh}, v_{kh}) + \mathcal{N}(0,\Sigma)\,,$$

其中，Σ 是 LMC 更新中高斯随机向量的协方差。在新坐标下，这次迭代可写为

$$(\phi_{(k+1)h}, \psi_{(k+1)h}) \overset{d}{=} \bar{F}(\phi_{kh}, \psi_{kh}) + \mathcal{N}(0, \mathcal{M}\Sigma\mathcal{M}^{\mathsf{T}})\,.$$

记作 $\mathcal{M}\Sigma\mathcal{M}^{\mathsf{T}} = \bar{\Sigma} \otimes I_d$，我们可以计算

$$\begin{aligned} \bar{\Sigma}_{1,1} &= \frac{2h}{\gamma} - \frac{3}{\gamma^2} + \frac{4\exp(-\gamma h)}{\gamma^2} - \frac{\exp(-2\gamma h)}{\gamma^2} = O(\gamma h^3)\,, \\ \bar{\Sigma}_{1,2} &= \frac{2h}{\gamma} - \frac{1}{\gamma^2} + \frac{\exp(-2\gamma h)}{\gamma^2} = O(h^2)\,, \\ \bar{\Sigma}_{2,2} &= \frac{2h}{\gamma} + \frac{5}{\gamma^2} - \frac{8\exp(-\gamma h)}{\gamma^2} + \frac{3\exp(-2\gamma h)}{\gamma^2} = \frac{4h}{\gamma^2} + O(h^2)\,. \end{aligned}$$

由此可得

$$\|\bar{\Sigma}\|_{\mathrm{op}} \le \frac{4h}{\gamma} + O(h^2)\,.$$

因此，$C_{\mathsf{LSI}}(\mathcal{N}(0, \mathcal{M}\Sigma\mathcal{M}^{\mathsf{T}})) \le \frac{4h}{\gamma^2} + O(h^2)$。

命题 17。设 $\hat{\mu}_t^{\mathcal{M}} :=$ 概率分布(ϕ_t, ψ_t)。那么，对于所有 $\varepsilon > 0$，只要 $h > 0$ 足够小（取决于 ε），就有

$$C_{\mathsf{LSI}}(\hat{\mu}_{Nh}^{\mathcal{M}}) \le \Big(1 - \big(m\sqrt{\frac{2}{L}} - \varepsilon\big)\, h\Big)^N C_{\mathsf{LSI}}(\hat{\mu}_0^{\mathcal{M}}) + \frac{4}{2m - \varepsilon\sqrt{2L}} + O\big(\frac{h\sqrt{L}}{m}\big)\,.$$

证明。对数 Sobolev 不等式常数的演化满足

$$\begin{aligned} C_{\mathsf{LSI}}(\hat{\mu}_{(k+1)h}^{\mathcal{M}}) &\le \|\bar{F}\|_{\mathrm{op}}^2\, C_{\mathsf{LSI}}(\hat{\mu}_{kh}^{\mathcal{M}}) + C_{\mathsf{LSI}}\big(\mathcal{N}(0, \mathcal{M}\Sigma\mathcal{M}^{\mathsf{T}})\big) \\ &\le \Big(1 - m\sqrt{\frac{2}{L}}\, h + O(Lh^2)\Big)\, C_{\mathsf{LSI}}(\hat{\mu}_{kh}^{\mathcal{M}}) + \frac{4h}{\gamma} + O(h^2)\,. \end{aligned}$$

当 h 足够小时，我们有

$$C_{\mathsf{LSI}}(\hat{\mu}_{(k+1)h}^{\mathcal{M}}) \le \Big(1 - \big(m\sqrt{\frac{2}{L}} - \varepsilon\big)\, h\Big)\, C_{\mathsf{LSI}}(\hat{\mu}_{kh}^{\mathcal{M}}) + \frac{4h}{\gamma} + O(h^2)\,.$$

迭代可得，

$$C_{\mathsf{LSI}}(\hat{\mu}_{Nh}^{\mathcal{M}}) \le \Big(1 - \big(m\sqrt{\frac{2}{L}} - \varepsilon\big)\, h\Big)^N C_{\mathsf{LSI}}(\hat{\mu}_0^{\mathcal{M}}) + \frac{4}{2m - \varepsilon\sqrt{2L}} + O\big(\frac{h\sqrt{L}}{m}\big)\,.$$

证明完毕。 □

推论 18. 现在令 $\mu_t^{\mathcal{M}}$表示连续时间欠阻尼朗之万扩散的概率分布，其中$\gamma = c\sqrt{L}$ 对应于(ϕ, ψ)坐标中的 $c \ge \sqrt{2}$。那么，

$$C_{\mathsf{LSI}}(\mu_t^{\mathcal{M}}) \le \exp\Big(-m\sqrt{\frac{2}{L}}\, t\Big)\, C_{\mathsf{LSI}}(\mu_0^{\mathcal{M}}) + \frac{2}{m}\,.$$

证明。在前述命题中，令 $h \searrow 0$，同时令 $Nh \to t$，然后令 $\varepsilon \searrow 0$。 □

C 离散化分析

我们考虑 [Ma+21] 中使用的离散化，其微分形式如下：

$$\begin{aligned} d\hat{x}_t &= \hat{v}_t\, \mathrm{d}t\,, \\ d\hat{v}_t &= -\gamma \hat{v}_t\, \mathrm{d}t - \nabla U(\hat{x}_{kh})\, \mathrm{d}t + \sqrt{2\gamma}\, \mathrm{d}B_t\,, \end{aligned}$$

并将变量 $\hat{w}_t$ 定义为元组 $(\hat{x}_t, \hat{v}_t)$，其中 $t \in [kh, (k+1)h]$。

C.1 技术引理

定理 19（吉尔萨诺夫定理，改编自 [Oks13，定理 8.6.8]）。设 $(x_t)_{t\ge0}$、$(b_t^P)_{t\ge0}$、$(b_t^Q)_{t\ge0}$ 是适应于同一滤过的随机过程，且 $\sigma \in \mathbb{R}^{d\times d}$ 是任意常数矩阵（可以是退化矩阵）。令 P_T 和 Q_T 为路径空间 $C([0,T]; \mathbb{R}^d)$ 上的概率测度，使得 $(w_t)_{t\ge0}$ 按照以下方式演化：

$$\begin{aligned} \mathrm{d}w_t &= b_t^P\, \mathrm{d}t + \sigma\, \mathrm{d}B_t^P \qquad \textit{under } P_T\,, \\ \mathrm{d}w_t &= b_t^Q\, \mathrm{d}t + \sigma\, \mathrm{d}B_t^Q \qquad \textit{under } Q_T\,, \end{aligned}$$

其中，B^P是一个 P_T-布朗运动，而 B^Q 是一个 Q_T-布朗运动。此外，假设存在一个过程 $(u_t)_{t\ge0}$，使得

$$\sigma\, u_t = b_t^P - b_t^Q\,,$$

and

$$\mathbb{E}^{Q_T}\exp\Big(2q^2\int_0^T\|u_s\|^2\,\mathrm{d}s\Big)<\infty\,,$$

因此，若我们将 $\sigma^\dagger$ 定义为 σ的Moore–Penrose 伪逆，则根据前述假设，有 $u_t=\sigma^\dagger\ (b_t^P-b_t^Q)$。接下来，

$$\frac{\mathrm{d}P_T}{\mathrm{d}Q_T}=\exp\Big(\int_0^T\langle\sigma^\dagger\,(b_t^P-b_t^Q),\mathrm{d}B_t^Q\rangle-\frac{1}{2}\int_0^T\|\sigma^\dagger\,(b_t^P-b_t^Q)\|^2\,\mathrm{d}t\Big)\,.$$

事实上，我们只需要以下推论。

推论 20. 对于任意事件 $\mathcal{E}$ 和 $q\geq 1$,

$$\mathbb{E}^{Q_T}\big[(\frac{\mathrm{d}P_T}{\mathrm{d}Q_T})^q{}_{\mathbb{I}}\mathcal{E}\big]\leq\sqrt{\mathbb{E}\Big[\exp\Big(2q^2\int_0^T\|\sigma^\dagger\,(b_t^P-b_t^Q)\|^2\,\mathrm{d}t\Big)\,{}_{\mathbb{I}}\mathcal{E}\Big]}\,.$$

证明。利用柯西–施瓦茨不等式，再利用伊藤引理，我们得到

$$\begin{aligned}\mathbb{E}^{Q_T}\big[(\frac{\mathrm{d}P_T}{\mathrm{d}Q_T})^q{}_{\mathbb{I}}\mathcal{E}\big]&=\mathbb{E}^{Q_T}\Big[\exp\Big(q\int_0^T\langle\sigma^\dagger\,(b_t^P-b_t^Q),\mathrm{d}B_t^Q\rangle-\frac{q}{2}\int_0^T\|\sigma^\dagger\,(b_t^P-b_t^Q)\|^2\,\mathrm{d}t\Big)\,{}_{\mathbb{I}}\mathcal{E}\Big]\\&\leq\sqrt{\mathbb{E}^{Q_T}\Big[\exp\Big((2q^2-q)\int_0^T\|\sigma^\dagger\,(b_t^P-b_t^Q)\|^2\,\mathrm{d}t\Big)\,{}_{\mathbb{I}}\mathcal{E}\Big]}\\&\quad\times\underbrace{\sqrt{\mathbb{E}^{Q_T}\Big[\exp\Big(2q\int_0^T\langle\sigma^\dagger\,(b_t^P-b_t^Q),\mathrm{d}B_t^Q\rangle-2q^2\int_0^T\|\sigma^\dagger\,(b_t^P-b_t^Q)\|^2\,\mathrm{d}t\Big)\,{}_{\mathbb{I}}\mathcal{E}\Big]}}_{=1}\\&\leq\sqrt{\mathbb{E}^{Q_T}\Big[\exp\Big(2q^2\int_0^T\|\sigma^\dagger\,(b_t^P-b_t^Q)\|^2\,\mathrm{d}t\Big)\,{}_{\mathbb{I}}\mathcal{E}\Big]}\,.\end{aligned}$$

这里，我们利用了以下事实： $t\mapsto\exp(\int_0^t\langle u_\tau,\mathrm{d}B_\tau\rangle-\frac{1}{2}\int_0^t\|u_\tau\|^2\mathrm{d}\tau)$ 是一个局部鞅。 □

对于过程 (x_t,v_t)，我们可以确定以下结果:

$$\sigma=\begin{bmatrix}0&0\\0&\sqrt{2\gamma}\,I_d\end{bmatrix},\qquad b_t^P=\begin{bmatrix}v_t\\-\gamma v_t-\nabla U(x_t)\end{bmatrix},\qquad b_t^Q=\begin{bmatrix}v_t\\-\gamma v_t-\nabla U(x_{\lfloor t/h\rfloor h})\end{bmatrix}.$$

在此情况下， $\|\sigma^\dagger(b_t^P-b_t^Q)\|\equiv\frac{1}{\sqrt{2\gamma}}\|\nabla U(x_{\lfloor t/h\rfloor h})-\nabla U(x_t)\|$。

我们还采用以下引理，证明从略，出自 [Che+21]。

引理 21 （测度变换，引自[Che+21，引理 21]）。 设 μ， ν 为概率测度，且 E 为任意事件。则，

$$\mu(E)\leq\nu(E)+\sqrt{\chi_2(\mu\,\|\,\nu)\,\nu(E)}\,.$$

特别地，若 μ和 ν 是定义在 $\mathbb{R}^d$ 上的概率测度，并且

$$\nu\{\|\cdot\|\geq R_0+\eta\}\leq C\exp(-c\eta^2)\qquad \textit{for all } \eta\geq 0\,,$$

其中 $C\geq 1$，则

$$\mu\Big\{\|\cdot\|\geq R_0+\sqrt{\frac{1}{c}\,\mathcal{R}_2(\mu\,\|\,\nu)}+\eta\Big\}\leq 2C\exp\big(-\frac{c\eta^2}{2}\big)\qquad \textit{for all } \eta\geq 0\,.$$

引理 22. 设 $(B_t)_{t\ge 0}$ 是 $\mathbb{R}^d$ 中的标准布朗运动。那么，若 $\lambda \ge 0$ 且 $h \le 1/(4 \quad \lambda)$，

$$\mathbb{E}\exp\bigl(\lambda \sup_{t\in[0,h]} \|B_t\|^2\bigr) \le \exp(6dh\lambda)\,.$$

特别地，对所有 $\eta \ge 0$，

$$\mathbb{P}\bigl\{ \sup_{t\in[0,h]} \|B_t\| \ge \eta\bigr\} \le 3\exp\bigl(-\frac{\eta^2}{6dh}\bigr)\,.$$

引理 23 ([GT20，引理 14])。设 $Y>0$ 是一个随机变量。假设对所有 $0<\delta<1/2$，存在一个事件 $\mathcal{E}_\delta$，其概率至少为 $1-\delta$，使得

$$\mathbb{E}[Y^2 \mid \mathcal{E}_\delta] \le \frac{v}{\delta^\xi}$$

对于某个 $\xi<1$。那么，$\mathbb{E}\,Y \le 4\sqrt{v}$。

引理 24（矩阵 Gr¨onwall 不等式）。设 $x:\ \mathbb{R}_+ \to \mathbb{R}^d$，且 $c\in\mathbb{R}^d$、$A\in\mathbb{R}^{d\times d}$，其中 A 的元素均为非负数。假设以下不等式按分量成立：

$$x(t) \le c + \int_0^t Ax(s)\,\mathrm{d}s\,, \qquad \textit{for all } t\ge 0\,. \tag{C.1}$$

那么，以下不等式成立，其中 $I_d\in\mathbb{R}^{d\times d}$ 是 d 维单位矩阵：

$$x(t) \le (AA^\dagger e^{At} - AA^\dagger + I_d)\,c\,, \tag{C.2}$$

其中，$A^\dagger$ 是 A 的 Moore–Penrose 伪逆（当 A 可逆时，这等价于标准逆）。

证明。这是 [CD76 的主定理]的一个特例。 □

C.2 欠阻尼朗之万蒙特卡罗的移动界

接下来，我们将证明连续时间朗之万扩散的移动界。以下引理是关于高斯向量范数集中 [的一个标准事实，参见 BLM13 的定理 5。5]。

引理 25（范数集中）。以下集中性成立：对于所有 $\eta\ge 0$，

$$\rho(\|\cdot\| \ge \sqrt{d}+\eta) \le \exp\Bigl(-\frac{\eta^2}{2}\Bigr)\,.$$

注意，由于动量变量v中的布朗运动分量，$\|v_t - v_0\|$ 的大小为 $\mathcal{O}(\sqrt{dt})$；这与过阻尼Langevin扩散增量的大小具有相同的阶。然而，如果我们仅考虑 x坐标上的增量，则可得到以下界。

引理 26。设 $(x_t,v_t)_{t\ge 0}$ 表示从 (x_0,v_0)出发的连续时间欠阻尼朗之万扩散，并假设势函数的梯度 ∇U 满足 $\nabla U(0) = 0$ 且为 H ¨older 连续（满足 (2.3)）。此外，假设 $h \lesssim L^{-1/2}\wedge\gamma^{-1}$ 且 $0\le\lambda\lesssim \frac{1}{\gamma^s d^s h^{3s}}$。那么，

$$\log\mathbb{E}\exp\bigl(\lambda \sup_{t\in[0,h]} \|x_t - x_0\|^{2s}\bigr) \lesssim \bigl(L^{2s}h^{4s}\,(1+\|x_0\|^{2s^2}) + h^{2s}\,\|v_0\|^{2s} + \gamma^s d^s h^{3s}\bigr)\,\lambda\,.$$

证明。对于插值时刻，我们将使用 Gr¨onwall 矩阵不等式（引理 24），以及以下关于 x 的方程:

$$\begin{aligned}
\|x_t - x_0\| &\le \left\| \int_0^t v_\tau \,\mathrm{d}\tau \right\| \le h\|v_0\| + \left\| \int_0^t (v_\tau - v_0)\,\mathrm{d}\tau \right\| \\
&\le h\|v_0\| + \left\| \int_0^t \int_0^\tau \gamma v_{\tau'} \,\mathrm{d}\tau' \,\mathrm{d}\tau \right\| + \left\| \int_0^t \int_0^\tau \nabla U(x_{\tau'}) \,\mathrm{d}\tau' \,\mathrm{d}\tau \right\| \\
&\quad + \left\| \int_0^t \int_0^\tau \sqrt{2\gamma}\,\mathrm{d}B_{\tau'} \,\mathrm{d}\tau \right\| \\
&\le h\|v_0\| + \gamma h \left(h\|v_0\| + \int_0^t \|v_\tau - v_0\| \,\mathrm{d}\tau \right) + Lh^2 \\
&\quad + Lh \left(h\|x_0\|^s + \int_0^t \|x_\tau - x_0\| \,\mathrm{d}\tau \right) + \sqrt{2\gamma}\, h \sup_{t\in[0,h]} \|B_t\| .
\end{aligned}$$

这里我们使用 ∇U 的 H¨older 性质以及 $\|x\|^s \le 1 + \|x\|$。v 的情形同理:

$$\begin{aligned}
\|v_t - v_0\| &\le \left\| \int_0^t \gamma v_\tau \,\mathrm{d}\tau \right\| + \left\| \int_0^t \nabla U(x_\tau) \,\mathrm{d}\tau \right\| + \left\| \int_0^t \sqrt{2\gamma}\,\mathrm{d}B_\tau \right\| \\
&\le \gamma \left(h\|v_0\| + \int_0^t \|v_\tau - v_0\| \,\mathrm{d}\tau \right) + Lh + L \left(h\|x_0\|^s + \int_0^t \|x_\tau - x_0\| \,\mathrm{d}\tau \right) \\
&\quad + \sqrt{2\gamma} \sup_{t\in[0,h]} \|B_t\| .
\end{aligned}$$

因此，我们可以使用 Gr¨onwall 不等式的矩阵形式（引理 24）。应用该引理时，设 $c = c_1 + c_2$ 满足 c_1, c_2。首先，对于 c_1:

$$A = \begin{bmatrix} Lh & \gamma h \\ L & \gamma \end{bmatrix}, \qquad c_1 = \begin{bmatrix} Lh^2\|x_0\|^s + \gamma h^2\|v_0\| + Lh^2 + \sqrt{2\gamma}\, h \sup_{t\in[0,h]}\|B_t\| \\ Lh\|x_0\|^s + \gamma h\|v_0\| + Lh + \sqrt{2\gamma}\sup_{t\in[0,h]}\|B_t\| \end{bmatrix}.$$

注意到 c_1 属于 A 的像空间，因此 $AA^\dagger c_1 = c_1$；类似地，注意到 $\exp(At)\, c_1$ 属于 A 的像空间（利用矩阵指数的幂级数表示），于是对于第一个分量，我们得到:

$$\begin{aligned}
&\sup_{t\in[0,h]} \|x_0 - x_t\| \\
&\quad \le h \exp\big((Lh+\gamma)\,h\big) \left(\gamma h\|v_0\| + Lh\|x_0\|^s + Lh + \sqrt{2\gamma} \sup_{t\in[0,h]} \|B_t\| \right) + c_2 \text{ term} \\
&\quad \le 2h \left(\gamma h\|v_0\| + Lh\|x_0\|^s + Lh + \sqrt{2\gamma} \sup_{t\in[0,h]} \|B_t\| \right) + c_2 \text{ term} ,
\end{aligned}$$

其中在第二行中，我们取 $h \lesssim \frac{1}{\sqrt{L}+\gamma}$。现在，取

$$c_2 = \begin{bmatrix} h\|v_0\| \\ 0 \end{bmatrix},$$

我们得到如下结果（其中 $\mathbf{v}^{(1)}$ 表示向量 $\mathbf{v}$ 的第一个分量）:

$$\left(\left(AA^\dagger \left(e^{Ah} - I_{2d} \right) + I_{2d} \right) c_2 \right)_{(1)} = \frac{Lhe^{(Lh+\gamma)\,h} + \gamma}{Lh+\gamma}\, h\|v_0\|.$$

最后，对于 $h \lesssim \frac{1}{\sqrt{L}+\gamma}$，其可由 $2h\|v_0\|$ 上界估计。利用引理 22 并将其代入该表达式，即可完成证明。

□

C.3 迭代序列的次高斯性

类似于 [Che+21]，我们引入一个修正势函数，以证明 ULMC 迭代序列的次高斯性。首先，我们考虑 x-坐标上的一个修正分布，其参数为 a，（β，S）其中S，$\beta \geq 0$:

$$\pi^{(a)} \propto \exp(-U^{(a)}), \qquad U^{(a)}(x) \triangleq U(x) + \frac{\beta}{2}(\|x\| - S)_+^2 . \tag{C.3}$$

修正势函数满足以下性质。

引理27 （修正势函数的性质，[Che+21，引理 23]）。设 $\pi^{(a)}$ 和 $U^{(a)}$ 如（C.3）中所定义。假设 $\nabla U(0) = 0$ 且 ∇U 满足（2.3）。则以下断言成立。

1.（次高斯　尾界）假设选择 S 使得 $\pi(B(0,S)) \geq 1/2$。则，对所有 $\eta \geq 0$,

$$\pi^{(a)}\{\|\cdot\| \geq S + \eta\} \leq 2\exp\Big(-\frac{\beta\eta^2}{2}\Big) .$$

2.（梯度增长）梯度 $\nabla U^{(a)}$ 满足

$$\|\nabla U^{(a)}(x)\| \leq L + (\beta + L)\|x\| .$$

那么，令 $\{(x_t^{(a)}, v_t^{(a)})\}_{t\geq 0}$ 为势函数 $U^{(a)}$ 和 $\mu^{(a)} := \pi^{(a)} \otimes \rho$ 对应的欠阻尼朗之万扩散的解，可得以下引理:

引理 28。假设 $h \lesssim (\beta + L)^{-1/2} \wedge \gamma^{-1} \wedge d^{-1/2}$，且 $\beta \leq 1$。那么，对于所有 $\delta \in (0,1)$，以至少 $1-\delta$ 的概率,

$$\sup_{t \leq Nh} \|x_t^{(a)}\| - S \lesssim (\beta + L) S h^2 + \sqrt{\frac{1}{\beta}\mathcal{R}_2(\mu_0^{(a)} \,\|\, \mu^{(a)})} + \sqrt{\frac{1}{\beta}\log\frac{16N}{\delta}} .$$

证明。我们可以结合测度变换引理（引理 21）以及引理 25、27 中的次高斯尾界，得出以至少 $1-\delta$ 的概率，以下事件同时成立:

$$\begin{aligned}
\max_{k\leq N} \|x_{kh}^{(a)}\| &\leq S + \sqrt{\frac{2}{\beta}\mathcal{R}_2(\mu_0^{(a)} \,\|\, \mu^{(a)})} + \sqrt{\frac{4}{\beta}\log\frac{8N}{\delta}} \\
\max_{k\leq N} \|v_{kh}^{(a)}\| &\leq \sqrt{d} + \sqrt{2\,\mathcal{R}_2(\mu_0^{(a)} \,\|\, \mu^{(a)})} + \sqrt{4\log\frac{4N}{\delta}} .
\end{aligned}$$

这里我们结合并集界以及 $t \mapsto \mathcal{R}_2(\mu_t^{(a)} \,\|\, \mu^{(a)})$ 关于 t 的单调性。

对于插值时刻，我们将使用 Gr¨onwall 矩阵不等式，以及针对 x 的以下不等式:

$$\begin{aligned}
\|x_{kh}^{(a)} - x_{kh+t}^{(a)}\| &\leq h\|v_{kh}^{(a)}\| + \Big\|\int_0^t\int_0^\tau \gamma v_{kh+\tau'}^{(a)}\,\mathrm{d}\tau'\,\mathrm{d}\tau\Big\| + \Big\|\int_0^t\int_0^\tau \nabla U^{(a)}(x_{kh+\tau'}^{(a)})\,\mathrm{d}\tau'\,\mathrm{d}\tau\Big\| \\
&\quad + \Big\|\int_0^t\int_0^\tau \sqrt{2\gamma}\,\mathrm{d}B_{kh+\tau'}\,\mathrm{d}\tau\Big\| \\
&\leq h\|v_{kh}^{(a)}\| + \gamma h\Big(h\|v_{kh}^{(a)}\| + \int_0^t \|v_{kh+\tau}^{(a)} - v_{kh}^{(a)}\|\,\mathrm{d}\tau\Big) + Lh^2 \\
&\quad + (\beta + L)h\Big(h\|x_{kh}^{(a)}\| + \int_0^t \|x_{kh+\tau}^{(a)} - x_{kh}^{(a)}\|\,\mathrm{d}\tau\Big) \\
&\quad + \sqrt{2\gamma}\,h \sup_{\tau\in[0,h]} \|B_{kh+\tau} - B_{kh}\| .
\end{aligned}$$

同样，

$$\begin{aligned}\|v_{kh}^{(a)}-v_{kh+t}^{(a)}\| &\le \left\|\int_0^t \gamma v_{kh+\tau}^{(a)}\,\mathrm{d}\tau\right\| + \left\|\int_0^t \nabla U^{(a)}(x_{kh+\tau}^{(a)})\,\mathrm{d}\tau\right\| + \left\|\int_0^t \sqrt{2\gamma}\,\mathrm{d}B_{kh+\tau}\right\| \\ &\le \gamma\left(h\|v_{kh}^{(a)}\| + \int_0^t \|v_{kh+\tau}^{(a)}-v_{kh}^{(a)}\|\,\mathrm{d}\tau\right) + Lh \\ &\quad + (\beta+L)\left(h\|x_{kh}^{(a)}\| + \int_0^t \|x_{kh+\tau}^{(a)}-x_{kh}^{(a)}\|\,\mathrm{d}\tau\right) \\ &\quad + \sqrt{2\gamma}\sup_{\tau\in[0,h]}\|B_{kh+\tau}-B_{kh}\|.\end{aligned}$$

因此，我们可以类似于引理 ¨24 中的做法，应用矩阵 Gronwall 不等式，令$c=c_1+c_2$ 表示以下矩阵：

$$\begin{aligned}A &= \begin{bmatrix}(\beta+L)\,h & \gamma h\\ (\beta+L) & \gamma\end{bmatrix},\\ c_1 &= \begin{bmatrix}(\beta+L)\,h^2\|x_{kh}^{(a)}\| + \gamma h^2\|v_{kh}^{(a)}\| + Lh^2 + \sqrt{2\gamma}\,h\sup_{t\in[0,h]}\|B_{kh+t}-B_{kh}\|\\ (\beta+L)\,h\|x_{kh}^{(a)}\| + \gamma h\|v_{kh}^{(a)}\| + Lh + \sqrt{2\gamma}\sup_{t\in[0,h]}\|B_{kh+t}-B_{kh}\|\end{bmatrix},\\ c_2 &= \begin{bmatrix}h\|v_{kh}^{(a)}\|\\ 0\end{bmatrix}.\end{aligned}$$

注意， c_1 同样属于 A 的像空间，因此 $(AA^\dagger - I_2)\,c = 0$。最后，计算矩阵指数后，我们得到

$$\begin{aligned}\sup_{t\le h}\|x_{kh}^{(a)}-x_{kh+t}^{(a)}\| &\le h\exp\big((\beta+L)\,h^2+\gamma h\big)\Big((\beta+L)\,h\|x_{kh}^{(a)}\| + \gamma h\|v_{kh}^{(a)}\| + Lh \\ &\qquad + \sqrt{2\gamma}\sup_{t\le h}\|B_{kh+t}-B_{kh}\|\Big) \\ &\quad + h\,\frac{(\beta+L)\exp\big((\beta+L)\,h^2+\gamma h\big)h+\gamma}{(\beta+L)\,h+\gamma}\,\|v_{kh}^{(a)}\| \\ &\le 2h\left((\beta+L)\,h\|x_{kh}^{(a)}\| + \|v_{kh}^{(a)}\| + Lh + \sqrt{2\gamma}\sup_{t\le h}\|B_{kh+t}-B_{kh}\|\right),\end{aligned}$$

其中在第二行中，我们取 $h \lesssim \quad {}^{1}\,{}_{\overline{(\beta+L)^{1/2}}} \wedge \frac{1}{\gamma}$。注意，这与引理26中的计算完全类似。

随后，我们可以应用并集界，得到对于任意 S_1、 S_2,

$$\begin{aligned}&\mathbb{P}\Big\{\sup_{t\in[0,Nh]}\|x_t^{(a)}\|\ge\eta\Big\}\\ &\quad\le \mathbb{P}\Big\{\max_{k=0,1,\dots N-1}\|x_{kh}^{(a)}\|\ge S_1\Big\} + \mathbb{P}\Big\{\max_{k=0,1,\dots N-1}\|v_{kh}^{(a)}\|\ge S_2\Big\}\\ &\qquad + \sum_{k=0}^{N-1}\mathbb{P}\Big\{\sup_{t\in[0,h]}\|x_{kh+t}^{(a)}-x_{kh}^{(a)}\|\ge\eta-S_1\Big\}\\ &\quad\le \mathbb{P}\Big\{\max_{k=0,1,\dots N-1}\|x_{kh}^{(a)}\|\ge S_1\Big\} + \mathbb{P}\Big\{\max_{k=0,1,\dots N-1}\|v_{kh}^{(a)}\|\ge S_2\Big\}\\ &\qquad + \sum_{k=0}^{N-1}\mathbb{P}\Big\{\sup_{t\in[0,h]}\sqrt{2\gamma}\,\|B_{kh+t}-B_{kh}\|\ge\frac{\eta-S_1}{2h}-(\beta+L)\,S_1h-S_2-Lh\Big\}.\end{aligned}$$

随后，分别取 $S_1 = S + \sqrt{\frac{2}{\beta}\,\mathcal{R}_2(\mu_0^{(a)}\,\|\,\mu^{(a)})} + \sqrt{\frac{4}{\beta}}$对数 $\frac{8N}{\delta}$, $S_2 = \sqrt{d} + \sqrt{2\,\mathcal{R}_2(\mu_0^{(a)}\,\|\,\mu^{(a)})} +$ _____

$\sqrt{\text{4对数}\frac{4N}{\delta}}$，我们利用布朗运动尾界（引理 22）得到概率为 $1-2\delta$：

$$\sup_{t\le Nh}\|x_t^{(a)}\|-S_1\lesssim(\beta+L)\,S_1h^2+S_2h+Lh^2+\sqrt{\gamma dh^3\log\frac{3N}{\delta}}\,.$$

若假设 $\beta\le 1$ 和 $h\lesssim\frac{1}{\sqrt{d}}$，则我们可以进一步简化此界，得到

$$\sup_{t\le Nh}\|x_t^{(a)}\|-S\lesssim(\beta+L)\,Sh^2+\sqrt{\frac{1}{\beta}\mathcal{R}_2(\mu_0^{(a)}\,\|\,\mu^{(a)})}+\sqrt{\frac{1}{\beta}\log\frac{8N}{\delta}}\,.$$

证明到此结束。 □

为了将这种次高斯性转移到原始的欠阻尼朗之万过程，我们考虑这两个过程之间卡方散度的如下界。

命题 29. 设 Q_T、$Q_T^{(a)}$ 分别表示在相同的初始化 μ_0下，原始扩散过程和修正扩散过程在路径空间上的概率分布。那么，若 $\beta\lesssim\frac{\gamma}{T}\wedge L$ 且 $h\lesssim(\beta+L)^{-1/2}\wedge\gamma^{-1}\wedge d^{-1/2}$，则

$$\mathcal{R}_2(Q_T\,\|\,Q_T^{(a)})\lesssim\frac{\beta^2L^2S^2Th^4}{\gamma}+\frac{\beta T}{\gamma}\left(\mathcal{R}_2(\mu_0\,\|\,\mu^{(a)})+\log N\right).$$

证明。对引理 28中的事件进行条件化，记该事件为 $\mathcal{E}_\delta$，其中 $\delta\le 1/2$，然后利用吉尔萨诺夫定理（推论20），可得（取充分小的 h，使诺维科夫条件成立）

$$\begin{aligned}\log\mathbb{E}\Big[\Big(\frac{\mathrm{d}Q_T}{\mathrm{d}Q_T^{(a)}}\Big)^4{}_{\mathbb{I}\mathcal{E}_\delta}\Big]&\le\frac{1}{2}\log\mathbb{E}\Big[\exp\Big(\frac{16}{\gamma}\int_0^T\|\nabla U(x_t^{(a)})-\nabla U^{(a)}(x_t^{(a)})\|^2\,\mathrm{d}t\Big)_{\mathbb{I}\mathcal{E}_\delta}\Big]\\&\le\frac{1}{2}\log\mathbb{E}\Big[\exp\Big(\frac{16\beta^2}{\gamma}\int_0^T\big(\|x_t^{(a)}\|-S\big)_+^2\,\mathrm{d}t\Big)_{\mathbb{I}\mathcal{E}_\delta}\Big]\\&\lesssim\frac{\beta^2T}{\gamma}\Big\{(\beta+L)^2\,S^2h^4+\frac{1}{\beta}\mathcal{R}_2(\mu_0\,\|\,\mu^{(a)})+\frac{1}{\beta}\log\frac{16N}{\delta}\Big\}.\end{aligned}$$

若取 $\beta\lesssim\gamma/T$ 并且 $L\ge\beta$，我们可以利用引理 23 得到

$$\mathcal{R}_2(Q_T\,\|\,Q_T^{(a)})=\log\mathbb{E}\Big[\Big(\frac{\mathrm{d}Q_T}{\mathrm{d}Q_T^{(a)}}\Big)^2\Big]\lesssim\frac{\beta^2L^2S^2Th^4}{\gamma}+\frac{\beta T}{\gamma}\left(\mathcal{R}_2(\mu_0\,\|\,\mu^{(a)})+\log N\right).$$

证明至此结束。 □

命题 30. 考虑从 μ_0 初始化的连续时间扩散过程 $(x_t,v_t)_{t\ge0}$。对于 $h\lesssim(\beta+L)^{-1/2}\wedge\gamma^{-1}\wedge d^{-1/2}$，$S\asymp\mathfrak{m}$ 和 $\beta\asymp\frac{\gamma}{T}$，当 $\delta\in(0,1/2)$ 时，以下结论以概率 $1-\delta$ 成立：

$$\max_{k=0,\dots,N-1}\|x_{kh}\|\lesssim\mathfrak{m}+\sqrt{\frac{T}{\gamma}\left(\mathcal{R}_2(\mu_0\,\|\,\mu^{(a)})+\log\frac{N}{\delta}\right)},$$

$$\max_{k=0,\dots,N-1}\|v_{kh}\|\lesssim\sqrt{d}+\sqrt{\mathcal{R}_2(\mu_0\,\|\,\mu^{(a)})+\log\frac{N}{\delta}}\,.$$

证明。 回顾引理 28 的证明可知，以概率 $1-\delta$，

$$\max_{k=0,\dots N-1}\|x_{kh}^{(a)}\|\lesssim S+\sqrt{\frac{1}{\beta}\mathcal{R}_2(\mu_0\,\|\,\mu^{(a)})}+\sqrt{\frac{1}{\beta}\log\frac{8N}{\delta}}\,.$$

特别地，这立即意味着以下结论成立：对于 $\eta\ge0$，

$$\mathbb{P}\Big(\max_{k=0,\dots N-1}\|x_{kh}^{(a)}\|\gtrsim S+\sqrt{\frac{1}{\beta}\mathcal{R}_2(\mu_0\,\|\,\mu^{(a)})}+\sqrt{\frac{1}{\beta}\log\frac{8N}{\delta}}+\eta\Big)\lesssim N\exp(-c\beta\eta^2)\,,$$

其中 $c > 0$ 为一个通用常数。

然后，利用测度变换（引理 21）以及命题29中的界，取$S \asymp \mathfrak{m}$，我们以概率 $1-\delta$

$$\begin{aligned}\max_{k=0,\dots N-1} \|x_{kh}\| &\lesssim S + \sqrt{\frac{1}{\beta}\mathcal{R}_2(\mu_0 \parallel \mu^{(a)})} + \sqrt{\frac{1}{\beta}\mathcal{R}_2(Q_T \parallel Q_T^{(a)})} + \sqrt{\frac{1}{\beta}\log\frac{N}{\delta}} \\ &\lesssim \mathfrak{m} + \sqrt{\frac{1}{\beta}\left(\mathcal{R}_2(\mu_0 \parallel \mu^{(a)}) + \log\frac{N}{\delta}\right) + \frac{\beta L^2 T h^4 \mathfrak{m}^2}{\gamma}}\,.\end{aligned}$$

我们选择 $\beta \asymp \gamma/T$，使得

$$\max_{k=0,\dots N-1} \|x_{kh}\| \lesssim \mathfrak{m} + \sqrt{\frac{T}{\gamma}\left(\mathcal{R}_2(\mu_0 \parallel \mu^{(a)}) + \log\frac{N}{\delta}\right)}\,.$$

最后，结合并集界来控制 $\|v_{kh}\|$ 引理 25中的量，我们便得到该命题。 □

C.4 完成离散化证明

我们按照 [Che+21]的证明进行。

证明。 [命题 15] 的证明。令 $\{x_t\}_{t\geq 0}$遵循连续时间过程。令 P_T， Q_T分别表示与插值过程和连续时间扩散相对应的路径空间上的测度，二者均以 $\mu_0 = \pi_0 \otimes \mathcal{N}(0, I_d$为初始值。然后，定义

$$G_t \triangleq \frac{1}{\sqrt{2\gamma}}\int_0^t \langle \nabla U(x_\tau) - \nabla U(x_{\lfloor \tau/h \rfloor h}), \mathrm{d}B_\tau\rangle - \frac{1}{4\gamma}\int_0^t \|\nabla U(x_\tau) - \nabla U(x_{\lfloor \tau/h\rfloor h})\|^2 \,\mathrm{d}\tau.$$

根据吉尔萨诺夫定理（定理 19），我们立即利用伊藤公式得到

$$\begin{aligned}\mathbb{E}_{Q_T}\left[\left(\frac{\mathrm{d}P_T}{\mathrm{d}Q_T}\right)^q\right] - 1 &= \mathbb{E}\exp(qG_T) - 1 \\ &= \frac{q\,(q-1)}{4\gamma}\,\mathbb{E}\int_0^T \exp(qG_t)\,\|\nabla U(x_t) - \nabla U(x_{\lfloor t/h\rfloor h})\|^2\,\mathrm{d}t \\ &\leq \frac{q^2}{4\gamma}\int_0^T \sqrt{\mathbb{E}[\exp(2qG_t)]\,\mathbb{E}[\|\nabla U(x_t) - \nabla U(x_{\lfloor t/h\rfloor h})\|^4]}\,\mathrm{d}t\,.\end{aligned}$$

分别对这些项进行界定，首先利用推论20 和（2.3）可得

$$\begin{aligned}\mathbb{E}\exp(2qG_t) &\leq \sqrt{\mathbb{E}\exp\Big(\frac{4q^2}{\gamma}\int_0^t \|\nabla U(x_r) - \nabla U(x_{\lfloor r/h\rfloor h})\|^2\,\mathrm{d}r\Big)} \\ &\leq \sqrt{\mathbb{E}\exp\Big(\frac{4L^2q^2}{\gamma}\int_0^t \|x_r - x_{\lfloor r/h\rfloor h}\|^{2s}\,\mathrm{d}r\Big)}\,.\end{aligned}$$

现在让我们在事件

$$\mathcal{E}_{\delta,kh} \triangleq \left\{\max_{j=0,1,\dots,k-1}\|x_{kh}\| \leq R_\delta^x,\ \max_{j=0,1,\dots,k-1}\|v_{kh}\| \leq R_\delta^v\right\}.$$

根据命题 30，在选择时我们可以得到 $\mathbb{P}(\mathcal{E}_{\delta,kh}) \geq 1-\delta$

$$\begin{aligned}R_\delta^x &\lesssim \mathfrak{m} + \sqrt{\frac{\gamma}{T}\left(\mathcal{R}_2(\mu_0 \parallel \mu^{(a)}) + \log\frac{N}{\delta}\right)}\,, \\ R_\delta^v &\lesssim \sqrt{d} + \sqrt{\mathcal{R}_2(\mu_0 \parallel \mu^{(a)}) + \log\frac{N}{\delta}}\,.\end{aligned}$$

接下来，我们通过一些细致的步骤来界定所需的量。

单步误差。首先考虑单个区间上的误差 $[0,h]$。若假设步长满足 $h \lesssim (\gamma^{1-s}/(L^2 d^s q^2))^{1/(1+3s)}$，引理26 表明

$$\begin{aligned}\log \mathbb{E}\exp\Big(\frac{8L^2q^2}{\gamma}\int_0^h \|x_t - x_0\|^2 \,\mathrm{d}t\Big) &\le \log \mathbb{E}\exp\Big(\frac{8L^2hq^2}{\gamma}\sup_{t\in[0,h]}\|x_t - x_0\|^2\Big)\\ &\lesssim \frac{L^{2+2s}h^{1+4s}q^2}{\gamma}\left(1+\|x_0\|^{2s^2}\right) + \frac{L^2h^{1+2s}q^2}{\gamma}\|v_0\|^{2s}\\ &\quad + \frac{L^2 d^s h^{1+3s}q^2}{\gamma^{1-s}}.\end{aligned}$$

迭代。令 $\{\mathcal{F}_t\}_{t\ge 0}$ 表示滤过，则写作 $H_t = \int_0^t \|x_r - x_{\lfloor r/h\rfloor h}\|^2 \,\mathrm{d}r$，我们可以对 $\mathcal{F}_{(N-1)h}$ 取条件并迭代单步界。

$$\begin{aligned}&\log \mathbb{E}\Big[\exp\Big(\frac{8L^2q^2}{\gamma}H_{Nh}\Big)\,\mathbb{I}^{\mathcal{E}_{\delta,Nh}}\Big]\\ &\le \log \mathbb{E}\Big[\exp\Big(\frac{8L^2q^2}{\gamma}H_{(N-1)h}\\ &\qquad + \mathcal{O}\Big(\frac{L^{2+2s}h^{1+4s}q^2}{\gamma}\left(1+\|x_{(N-1)h}\|^{2s^2}\right)\\ &\qquad\quad + \frac{L^2h^{1+2s}q^2}{\gamma}\|v_{(N-1)h}\|^{2s} + \frac{L^2d^sh^{1+3s}q^2}{\gamma^{1-s}}\Big)\Big)\,\mathbb{I}^{\mathcal{E}_{\delta,Nh}}\Big]\\ &\le \log \mathbb{E}\Big[\exp\Big(\frac{8L^2q^2}{\gamma}H_{(N-1)h}\Big)\,\mathbb{I}^{\mathcal{E}_{\delta,(N-1)h}}\Big]\\ &\quad + \mathcal{O}\Big(\frac{L^{2+2s}h^{1+4s}q^2}{\gamma}(R_\delta^x)^{2s^2} + \frac{L^2h^{1+2s}q^2}{\gamma}(R_\delta^v)^{2s} + \frac{L^2d^sh^{1+3s}q^2}{\gamma^{1-s}}\Big).\end{aligned}$$

现在我们作出一些额外的简化假设，以得到更易于解释的界：假设 $\gamma/T \le 1$和 $h \lesssim \frac{1}{L}(1\wedge \frac{d^{1/2}}{\mathfrak{m}^s})$。在这些假设下，

$$\begin{aligned}&\log \mathbb{E}\Big[\exp\Big(\frac{8L^2q^2}{\gamma}H_{Nh}\Big)\,\mathbb{I}^{\mathcal{E}_{\delta,Nh}}\Big]\\ &\le \log \mathbb{E}\Big[\exp\Big(\frac{8L^2q^2}{\gamma}H_{(N-1)h}\Big)\,\mathbb{I}^{\mathcal{E}_{\delta,(N-1)h}}\Big]\\ &\quad + \mathcal{O}\Big(\frac{L^2h^{1+2s}q^2}{\gamma}\left(d + \mathcal{R}_2(\mu_0\,\|\,\mu^{(a)}) + \log\frac{N}{\delta}\right)^s\Big).\end{aligned}$$

完成这一迭代后可得

$$\log \mathbb{E}\Big[\exp\Big(\frac{8L^2q^2}{\gamma}H_{Nh}\Big)\,\mathbb{I}^{\mathcal{E}_{\delta,Nh}}\Big] \lesssim \frac{L^2Th^{2s}q^2}{\gamma}\left(d + \mathcal{R}_2(\mu_0\,\|\,\mu^{(a)}) + \log\frac{N}{\delta}\right)^s.$$

最后，当23时，应用引理

$$h \lesssim_s \frac{\gamma^{1/(2s)}}{L^{1/s}T^{1/(2s)}q^{1/s}} \tag{C.4}$$

（其中 $\lesssim_s$ 隐藏着一个依赖于 s的常数），我们得到

$$\log \mathbb{E}\Big[\exp\Big(\frac{4L^2q^2}{\gamma}H_{Nh}\Big)\Big] \lesssim 1 + \frac{L^2Th^{2s}q^2}{\gamma}\left(d + \mathcal{R}_2(\mu_0\,\|\,\mu^{(a)}) + \log N\right)^s.$$

接下来需要选择合适的步长 h ，使整个量小于 1。具体而言，只需选择

$$h \lesssim \widetilde{\mathcal{O}}_s\Big(\frac{\gamma^{1/(2s)}}{L^{1/s}T^{1/(2s)}q^{1/s}\left(d+\mathcal{R}_2(\mu_0 \parallel \mu^{(a)})\right)^{1/2}}\Big)\,. \tag{C.5}$$

第二项。接下来需要对原表达式中的另一项进行界定。由引理26可得

$$\mathbb{E}[\exp(\lambda \left\|x_{kh+t}-x_{kh}\right\|^{2s}) \mid w_{kh}] \lesssim 1\,,$$

只要将 λ 选得足够小，即

$$\lambda \asymp \frac{1}{\gamma^s d^s h^{3s}} \wedge \frac{1}{L^{2s}h^{4s}\left(1+\left\|x_{kh}\right\|\right)^{2s^2}} \wedge \frac{1}{h^{2s}\left\|v_{kh}\right\|^{2s}}\,.$$

这立即给出一个尾界：对于 $\eta \geq 0$,

$$\mathbb{P}\{\left\|x_{kh+t}-x_{kh}\right\|^{4s} \geq \eta \mid w_{kh}\} \lesssim \exp(-\lambda\sqrt{\eta})\,.$$

积分可得

$$\begin{aligned}\sqrt{\mathbb{E}[\|\nabla U(x_t)-\nabla U(x_{kh})\|^4]} &\leq L^2\sqrt{\mathbb{E}[\|x_t-x_{kh}\|^{4s}]} \lesssim L^2\sqrt{\mathbb{E}\,\frac{1}{\lambda^2}} \\ &\lesssim L^2\gamma^s d^s h^{3s} + L^{2+2s}h^{4s}\sqrt{1+\mathbb{E}[\|x_{kh}\|^{4s^2}]} + L^2h^{2s}\sqrt{\mathbb{E}[\|v_{kh}\|^{4s}]}\,.\end{aligned}$$

我们可以通过对先前的尾界（命题30）进行积分来估计期望 :

$$\begin{aligned}\sqrt{\mathbb{E}[\|\nabla U(x_t)-\nabla U(x_{kh})\|^4]} &\lesssim L^2\gamma^s d^s h^{3s} + L^{2+2s}h^{4s}\left(\mathfrak{m}+\frac{T}{\gamma}\left(\mathcal{R}_2(\mu_0 \parallel \mu^{(a)})+\log N\right)\right)^{2s^2} \\ &\quad + L^2h^{2s}\left(d+\mathcal{R}_2(\mu_0 \parallel \mu^{(a)})+\log N\right)^s \\ &\leq \widetilde{\mathcal{O}}\Big(L^2h^{2s}\left(d+\mathcal{R}_2(\mu_0 \parallel \mu^{(a)})\right)^s\Big)\,,\end{aligned}$$

前提是 $h \leq \widetilde{\mathcal{O}}(\frac{1}{L}(\frac{d^{1/2}}{\mathfrak{m}^s} \wedge \mathcal{R}_2(\frac{\gamma\mathcal{R}_2}{T})^{s/2}))$, 其中 $\mathcal{R}_2 = \mathcal{R}_2(\mu_0 \parallel \mu^{(a)})$。在我们的应用中，这个条件并不起主导作用，可以忽略。

合并各项界。 最后，我们可以合并这些步骤，得到：只要满足C.5)，步长条件成立，

$$\mathbb{E}_{Q_T}\Big[\Big(\frac{\mathrm{d}P_T}{\mathrm{d}Q_T}\Big)^q\Big]-1 \leq \widetilde{\mathcal{O}}\Big(\frac{Tq^2}{\gamma}L^2h^{2s}\left(d+\mathcal{R}_2(\mu_0 \parallel \mu^{(a)})\right)^s\Big)\,.$$

最后，满足以下步长条件即可将 R´enyi 散度限制在 ϵ^2以内:

$$h \lesssim \widetilde{\mathcal{O}}_s\Big(\frac{\gamma^{1/(2s)}\epsilon^{1/s}}{L^{1/s}T^{1/(2s)}q^{1/s}\left(d+\mathcal{R}_2(\mu_0 \parallel \mu^{(a)})\right)^{1/2}}\Big)\,.$$

证明至此完成。 □

D 主要结果的证明

首先，我们汇总一些关于可行初始化的结果，这些结果来自 [Che+21]。回顾一下， $\pi^{(a)}$ 是附录 C.3 中引入的修正分布。令

$$\pi_0 = \mathcal{N}(0, \varsigma I_d),$$

其中 $\varsigma = (2L+\beta)^{-1}$ 是高斯分布的方差， $\beta \asymp 1/T$ 是修正势函数中出现的参数。 T 的取值取决于所采用的假设；我们在主要假设下汇总了以下条件：

$$T = \begin{cases} \widetilde{\Theta}\left(\frac{L+d}{\mathfrak{q}(\gamma)}\right) & \pi \text{ satisfies (PI)} \\ \widetilde{\Theta}(\sqrt{L}C_{\mathsf{LSI}}) & \pi \text{ satisfies (LSI), or is strongly log-concave,} \end{cases}$$

其中 $\mathfrak{q}(\gamma)$ 在 (3.2) 中定义。

引理 31（改编自 [Che+21，附录 A]）。假设 π 满足（PI）和 H¨older 连续性条件（2.3），并且 $\nabla U(0) = 0$, $U(0) - \min U \lesssim d$。那么对于$\pi_0 = \mathcal{N}(0, (2L+\beta)^{-1} I_d)$，以下两个性质成立，其中 β是修正势函数中出现的参数：

$$\mathcal{R}_q(\pi_0 \parallel \pi) \le \widetilde{\mathcal{O}}(\beta + L + d)\,,$$
$$\mathcal{R}_q(\pi_0 \parallel \pi^{(a)}) \le \widetilde{\mathcal{O}}(\beta + L + d)\,.$$

证明。应用 [Che+21的引理 30] 或 [Che+21的引理 31]。 □

根据我们的分析，取 $\beta \lesssim L$，并且如果还满足 $L \lesssim d$，那么可以合理地预期 $\mathcal{R}_q(\pi_0 \parallel \pi \quad \mathcal{R}_q\ \pi_0 \parallel \pi^{(a)} \quad \le \widetilde{\mathcal{O}}\ d \qquad \mu_0 = \pi_0 \otimes \rho \qquad \mathcal{R}_q\ \mu_0 \parallel \mu \ = \mathcal{R}_q\ \pi_0 \parallel \pi \qquad \mathcal{R}_q\ \mu_0 \parallel \mu^{(a)} =) ( ) ( )$。令，使得 $( ) ( )$，类似地 $( ) ( )$。$\mathcal{R}_q\ \pi_0 \parallel \pi^{(a)}$

以 下引理给出了初始化时 Fisher 信息值的一个上界。

引理 32. 在前一个引理的条件下，初始化 $\mu_0 = \pi_0 \otimes \rho$ 也满足 $\mathsf{FI}(\mu_0 \parallel \mu) \lesssim Ld + L^{1-s} d^s$。

证明. 注意到由于 $\nabla U(0) = 0$, $\|\nabla \log \pi(x)\|^2 = \|\nabla U(x)\|^2 \le L^2 \|x\|^{2s}$。其次， π_0满足 $\mathbb{E}_{x\sim\pi_0}[\|x\|^2] \lesssim d/L$。因此，

$$\begin{aligned} \mathsf{FI}(\mu_0 \parallel \mu) = \mathbb{E}_{\pi_0}\left[\left\|\nabla \log \frac{\pi_0}{\pi}\right\|^2\right] &\le \mathbb{E}_{x\sim\pi_0}[\|\nabla U(x) - (2L+\beta)\, x\|^2] \\ &\lesssim L^2\, \mathbb{E}_{x\sim\mu_0}[\|x\|^2 + \|x\|^{2s}] \lesssim Ld + L^{1-s} d^s\,, \end{aligned}$$

其中最后一步用到了詹森不等式。 □

D.1 Poincar´e 不等式

证明。 [定理的证明9] 引理8 中的连续时间结果表明

$$T \gtrsim \frac{1}{\mathfrak{q}(\gamma)} \log \frac{\chi_2(\mu_0 \parallel \mu)}{\varepsilon^2} \implies \chi_2(\mu_T \parallel \mu) \le \epsilon^2\,.$$

˜注意到存在一个可行的初始化，使得 $\log \chi_2(\mu_0 \parallel \mu) \le \mathcal{O}(L+d)$，那么当我们选择 $T = \widetilde{\mathcal{O}}($ ^{1}时，该条件得到满足 $\overline{\mathfrak{q}(\gamma)}\ (L + d + \log\ \frac{1}{\epsilon}))$。这也表明，对于 $\epsilon \lesssim 1$，有 $\mathcal{R}_2(\mu_T \parallel \mu) = \log(1 + \chi_2(\mu_T \parallel \mu)) \lesssim \epsilon^2$。
注意以下关于 R´enyi 散度 [的分解（弱三角不等式），参见例如Mir17，命题 11]:

$$\mathcal{R}_q(P_1 \parallel P_2) \le \frac{q - 1/\mathfrak{c}}{q-1}\, \mathcal{R}_{\mathfrak{c}q}(P_1 \parallel P_3) + \mathcal{R}_{\mathfrak{d}(q-1/\mathfrak{c})}(P_3 \parallel P_2),$$

对于任意有效的 H¨older 共轭对 $\mathfrak{c}, \mathfrak{d}$，即$\frac{1}{\mathfrak{c}} + \frac{1}{\mathfrak{d}} = 1$, $\mathfrak{c}, \mathfrak{d} > 1$，以及任意三个概率分布$P_1, P_2, P_3$。

在我们的情形中，令 $q = 2 - \xi$ 和 $\mathfrak{d}(q - 1/\mathfrak{c}) = 2$，解出 $\mathfrak{c}$、 $\mathfrak{d}$ 后，可得 $\xi \le 1/2$ 如下：

$$\mathcal{R}_{2-\xi}(P_1 \parallel P_2) \le 2\mathcal{R}_{2/\xi}(P_1 \parallel P_3) + \mathcal{R}_2(P_3 \parallel P_2)\,.$$

因此，令 $P_1 = \hat{}\mu_{Nh}$，$P_2 = \mu$，$P_3 = \mu_{Nh}$，并将这一结果与命题15的离散化界结合起来，便得到

$$\mathcal{R}_{2-\xi}(\hat{\mu}_{Nh} \,\|\, \mu) \lesssim \mathcal{R}_{2/\xi}(\hat{\mu}_{Nh} \,\|\, \mu_{Nh}) + \mathcal{R}_2(\mu_{Nh} \,\|\, \mu) \lesssim \epsilon^2 \,,$$

只要

$$h = \widetilde{\Theta}\Big(\frac{\gamma^{1/(2s)} \epsilon^{1/s} \xi^{1/s} \mathfrak{q}(\gamma)^{1/(2s)}}{L^{1/s} d^{1/2} \, (L \vee d)^{1/(2s)}}\Big) \,,$$
$$N = \widetilde{\Theta}\Big(\frac{L^{1/s} d^{1/2} \, (L \vee d)^{1+1/(2s)}}{\gamma^{1/(2s)} \epsilon^{1/s} \xi^{1/s} \mathfrak{q}(\gamma)^{1+1/(2s)}}\Big) \,.$$

证明至此完成。 □

D.2 对数索伯列夫不等式

D.2.1 KL 散度

证明。 [定理 6] 的证明。我们在扭曲坐标（ϕ，ψ）中给出以下定理，这些坐标曾用于引理 14。考虑 KL 散度 的分解，并使用柯西–施瓦茨不等式：

$$\begin{aligned}
\mathsf{KL}(\hat{\mu}_T^{\mathcal{M}} \,\|\, \mu^{\mathcal{M}}) &= \int \log \frac{\hat{\mu}_T^{\mathcal{M}}}{\mu^{\mathcal{M}}} \, \mathrm{d}\hat{\mu}_T^{\mathcal{M}} \\
&= \mathsf{KL}(\hat{\mu}_T^{\mathcal{M}} \,\|\, \mu_T^{\mathcal{M}}) + \int \log \frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}} \, \mathrm{d}\hat{\mu}_T^{\mathcal{M}} \\
&= \mathsf{KL}(\hat{\mu}_T^{\mathcal{M}} \,\|\, \mu_T^{\mathcal{M}}) + \mathsf{KL}(\mu_T^{\mathcal{M}} \,\|\, \mu^{\mathcal{M}}) + \int \log \frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}} \, \mathrm{d}(\hat{\mu}_T^{\mathcal{M}} - \mu_T^{\mathcal{M}}) \\
&\le \mathsf{KL}(\hat{\mu}_T^{\mathcal{M}} \,\|\, \mu_T^{\mathcal{M}}) + \mathsf{KL}(\mu_T^{\mathcal{M}} \,\|\, \mu^{\mathcal{M}}) + \sqrt{\chi^2(\hat{\mu}_T^{\mathcal{M}} \,\|\, \mu_T^{\mathcal{M}}) \times \mathsf{var}_{\mu_T^{\mathcal{M}}}\Big(\log \frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}}\Big)} \,.
\end{aligned}$$

利用引理 14中的迭代序列的对数索伯列夫不等式，我们得到（根据对数索伯列夫不等式蕴含具有相同常数的 Poincar´不等式这一推论）

$$\mathsf{var}_{\mu_T^{\mathcal{M}}}\Big(\log \frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}}\Big) \le C_{\mathsf{LSI}}(\mu_T^{\mathcal{M}}) \, \mathbb{E}_{\mu_T^{\mathcal{M}}}\Big[\Big\|\nabla \log \frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}}\Big\|^2\Big] \,,$$

whe^re 我们将对数函数 $\frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}}$ 代入式（Poincaré 不等式）中的函数。这里，$C_{\mathsf{LSI}}(\mu_T^{\mathcal{M}}) \lesssim 1/m$ 对所有 $t \ge 0$。

由于 $\mu^{\mathcal{M}} = \mathcal{M}_{\#}\mu$，因此 $\mu^{\mathcal{M}}(\phi, \psi) \propto \mu(\mathcal{M}^{-1}(\phi, \psi))$。因此，

$$\nabla \log \mu^{\mathcal{M}} = (\mathcal{M}^{-1})^{\mathsf{T}} \nabla \log \mu \circ \mathcal{M}^{-1} \,,$$

对 $\nabla \log \mu_T^{\mathcal{M}}$ 也同理。由此得到表达式

$$\mathbb{E}_{\mu_T^{\mathcal{M}}}\Big[\Big\|\nabla \log \frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}}\Big\|^2\Big] = \mathbb{E}_{\mu_T}\Big[\Big\|(\mathcal{M}^{-1})^{\mathsf{T}} \nabla \log \frac{\mu_T}{\mu}\Big\|^2\Big] \,.$$

此外，有

$$\mathcal{M}^{-1} \, (\mathcal{M}^{-1})^{\mathsf{T}} = \begin{bmatrix} 1 & -\gamma/2 \\ -\gamma/2 & \gamma^2/2 \end{bmatrix} .$$

对于 $c_0 > 0$ 和 $\mathfrak{M}$ 在附录 B.1中定义，我们有

$$L \, \mathfrak{M} - c_0 \, \mathcal{M}^{-1} \, (\mathcal{M}^{-1})^{\mathsf{T}} = \begin{bmatrix} 1/4 - c_0 & \sqrt{L} \, (1/\sqrt{2} + c_0 \sqrt{2}) \\ \sqrt{L} \, (1/\sqrt{2} + c_0 \sqrt{2}) & L \, (4 - c_0) \end{bmatrix} .$$

行列式为 $L\,((\frac{1}{4}-c_0)\,(4-c_0)\;-(\,\frac{1}{\sqrt{2}}+c_0\sqrt{2})^2)\;>0$，当 $c_0>0$ 足够小时成立。这表明$\mathcal{M}^{-1}\,(\mathcal{M}^{-1})^{\mathsf{T}}\preceq c_0^{-1}L\,\mathfrak{M}$，因此

$$\mathbb{E}_{\mu_T^{\mathcal{M}}}\Big[\Big\|\nabla\log\frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}}\Big\|^2\Big]\lesssim L\,\mathsf{FI}_{\mathfrak{M}}(\mu_T\parallel\mu)\,.$$

这里我们定义

$$\mathsf{FI}_{\mathfrak{M}}(\mu'\parallel\mu)\triangleq\mathbb{E}_{\mu'}\big[\|\mathfrak{M}^{1/2}\,\nabla\log\frac{\mu'}{\mu}\|^2\big]$$

通过引理 5 对 Fisher 信息的衰减使我们能够设定

$$T\gtrsim C_{\mathsf{LSI}}\sqrt{L}\log\Big(\frac{\kappa}{\epsilon^2}\,\big(\mathsf{KL}(\mu_0\parallel\mu)+\mathsf{FI}_{\mathfrak{M}}(\mu_0\parallel\mu)\big)\Big)\implies\mathsf{var}_{\mu_T^{\mathcal{M}}}\Big(\log\frac{\mu_T^{\mathcal{M}}}{\mu^{\mathcal{M}}}\Big)\lesssim\epsilon^2\,.$$

相同的 T 选择还确保了 $\mathsf{KL}(\mu_T^{\mathcal{M}}\parallel\mu^{\mathcal{M}})\;\leq\varepsilon^2$。根据我们的初始化（引理 32），利用以下事实可以直接估计

$$\mathsf{FI}_{\mathfrak{M}}(\mu_0\parallel\mu)\lesssim\frac{1}{L}\,\mathsf{FI}(\pi_0\parallel\pi)\lesssim d\,,$$

以及 $\mathsf{KL}(\mu_0\parallel\mu)\;\lesssim d\log\;\kappa$，因此我们对 T的条件是（取 $C_{\mathsf{LSI}}\leq m^{-1}$）

$$T\geq\widetilde{\mathcal{O}}\Big(\frac{\sqrt{L}}{m}\log\frac{\kappa d}{\epsilon^2}\Big)\,.$$

还要注意，这要求 $\gamma\asymp\sqrt{L}$。对于其余的 $\chi^2(\hat{}\mu_T\parallel\mu_T)$ 和 $\mathsf{KL}(\hat{}\mu_T\parallel\mu_T)$ 项，我们应用命题 15，其中 $T=Nh$ 取指定值，目标精度为 ϵ，并取 $q=2$ 和 $s=1$，由此得到

$$h=\widetilde{\Theta}\Big(\frac{\epsilon m^{1/2}}{Ld^{1/2}}\Big)\,,$$

with

$$N=\widetilde{\Theta}\Big(\frac{\kappa^{3/2}d^{1/2}}{\epsilon}\Big)$$

（使用 $N=T/h$）。 □

D.2.2 全变差距离

证明。 [定理 7] 的证明。首先注意，全变差距离是一个恰当的度量，因此满足三角不等式。接下来，应用两次 Pinsker 不等式，

$$\begin{aligned}\|\hat{\mu}_{Nh}-\mu\|_{\mathsf{TV}}&\leq\|\hat{\mu}_{Nh}-\mu_{Nh}\|_{\mathsf{TV}}+\|\mu_{Nh}-\mu\|_{\mathsf{TV}}\\&\lesssim\sqrt{\mathsf{KL}(\hat{\mu}_{Nh}\parallel\mu_{Nh})}+\sqrt{\mathsf{KL}(\mu_{Nh}\parallel\mu)}\,.\end{aligned}$$

这些项可以分别进行界定。与前一个定理的证明类似，利用引理 5，只需取

$$T\geq\widetilde{\mathcal{O}}\Big(C_{\mathsf{LSI}}\sqrt{L}\log\frac{d}{\epsilon^2}\Big)\,,$$

对于另一项，只需对任意值的 q，$\gamma\asymp\sqrt{L}$ 结合对 T 的要求使用命题 15，即可得到:

$$h=\widetilde{\Theta}\Big(\frac{\epsilon}{C_{\mathsf{LSI}}^{1/2}Ld^{1/2}}\Big)\,,$$

with

$$N=\widetilde{\Theta}\Big(\frac{C_{\mathsf{LSI}}^{3/2}L^{3/2}d^{1/2}}{\epsilon}\Big)\,,$$

（使用 $N=T/h$）。 □