arXiv:2303.00570v1 [math.ST] 1 Mar 2023

重尾采样中离散化 Itô扩散的均方分析

Ye He* Tyler Farghly† Krishnakumar Balasubramanian‡ Murat A. Erdogdu§

2023年3月2日

摘要

我们通过对与加权 Poincar´e 不等式相关的一类自然 Itô扩散进行离散化，分析从一类重尾分布中进行采样的复杂度。基于均方分析，我们确立了获得一个其分布在 Wasserstein-2 度量下 ϵ 接近目标分布的样本所需的迭代复杂度。在本文中，我们将均方分析推至极限，即始终只要求目标密度具有有限方差，这是进行均方分析的最低要求。为获得显式估计，我们在各种假设下计算了与重尾目标相关的某些矩的上界。对于只能获得未归一化目标密度的函数评估这一情形，我们还通过使用高斯平滑技术估计梯度，给出了类似的迭代复杂度结果。我们基于多元 t-分布给出了说明性示例。

1 引言

从给定目标密度中采样 $\pi: \mathbb{R}^d \to \mathbb{R}$ 这一问题广泛存在于统计学、机器学习、运筹学和应用数学的各类问题中。马尔可夫链蒙特卡罗（MCMC）算法是一类常用的采样算法（Robert and Casella，1999年；Andrieu 等，2003；Hairer 等，2006；Brooks 等，2011；Meyn and Tweedie，2012年；Leimkuhler and Matthews，2016；Douc 等，2018年）；该领域一种广泛采用的方法是对以目标密度为平稳密度的 Itô 扩散进行离散化。过阻尼朗之万扩散是一种常用的扩散过程，

$$dX_t = \nabla \log \pi(X_t)dt + \sqrt{2}dB_t, \tag{1}$$

其中 B_t 是一个 d 维布朗运动。例如，未校正朗之万算法（Rossky 等，1978年）、Metropolis 校正朗之万算法（Roberts and Tweedie，1996年；Roberts and Rosenthal，1998年）以及近端采样器（Titsias and Papaspiliopoulos，2018年；Lee 等，2021；Vono 等，2022年）都是由式（1）的不同离散化得到的。在轻尾假设下，即当密度 π 的尾部呈指数级快速衰减时，式（1）中的扩散过程 X_t 会以指数速度收敛至以 π 为平稳密度，这促使人们将式（1）的离散化用于实际采样算法。过去十年间，各种离散化方法的非渐近迭代复杂度已得到充分研究，从而较为全面地阐明了轻尾密度的采样问题。

受稳健统计学（Kotz and Nadarajah，2004年；Jarner and Roberts，2007年；Kamatani，2018年）、多重比较程序（Genz 等，2004年；Genz and Bretz，2009）、贝叶斯统计学（Gelman 等，2008年；Ghosh 等，2018年）以及统计机器学习（Balcan and Zhang，2017；Nguyen 等，2019；S¸im¸sekli 等，2020；Diakonikolas 等，2020）的应用所启发，本文关注从重尾密度中采样，例如尾部呈多项式衰减的密度。

*数学系，加利福尼亚大学戴维斯分校。 leohe@ucdavis.edu
†统计学系，牛津大学，英国。 farghly@stats.ox.ac.uk
‡统计学系，加利福尼亚大学戴维斯分校。 kbala@ucdavis.edu§计算机科学系和统计科学系，多伦多大学。 erdogdu@cs.toronto.edu

当目标密度 π 为重尾时，方程（1）的解在若干感兴趣的度量下不会以指数速度收敛到其平稳密度。事实上，Roberts and Tweedie（1996年）的定理 2.4 表明，若 $|\nabla \log \pi(x)| \to 0$ as $|x| \to \infty$，则方程（1）的解不是指数遍历的。另一方面，文献中的标准结果（例如 Wang（2006年）；Bakry 等人（2014））表明，方程（1）的解在 χ^2 度量下快速指数收敛到其平衡密度，等价于密度 π 满足 Poincar´e 不等式，而这又要求 π 具有指数衰减尾部。此外，当 π 具有多项式衰减尾部时，收敛速度只有次指数或多项式级别（Wang，2006年，第 4 章）。因此，由方程（1）中的 Langevin 扩散离散化得到的算法仅适用于从轻尾、指数衰减的密度中采样；对于重尾密度的采样，这类算法则效率较低。

因此，我们的重尾采样方法基于对某些自然的 Itô 扩散过程进行离散化，这些扩散过程出现在以下加权 Poincaré´ 不等式的背景下（Blanchet 等人，2009；Bobkov 和 Ledoux，2009）。这类不等式可视为将 Brascamp-Lieb 不等式（针对对数凹密度类建立）推广到一类重尾密度。

定理 1 (加权 Poincaré´不等式；Bobkov 和 Ledoux 中的定理 2.3 (2009))。设目标密度具有如下形式 $\pi_\beta \propto V^{-\beta}$ 其中 $\beta > d$ 且 $V \in \mathcal{C}^2(\mathbb{R}^d)$ 为正、凸函数，且 $(\nabla^2 V)^{-1}(x)$ 对所有 $x \in \mathbb{R}^d$均有定义。对于任意光滑且 π_β-可积函数 g 定义在 $\mathbb{R}^d$ 上，并且 $G = Vg$,

$$(\beta+1)Var_{\pi_\beta}(g) \leq \int_{\mathbb{R}^d} \frac{\langle (\nabla^2 V)^{-1}\nabla G, \nabla G\rangle}{V} d\pi_\beta + \frac{d}{\beta - d}\left(\int_{\mathbb{R}^d} g d\pi_\beta\right). \tag{2}$$

满足定理 1中的条件、因而满足（2）的重尾密度的一个典型例子是多元 t-分布。具体而言，我们考虑以下 Itô 扩散过程

$$dX_t = -(\beta-1)\nabla V(X_t)dt + \sqrt{2V(X_t)}dB_t, \tag{3}$$

其中 $(B_t)_{t\geq 0}$ 是 $\mathbb{R}^d$上的标准布朗运动。（3）中的 Itô扩散以指数速度收敛到目标 π_β，在 χ^2-散度意义下，只要它满足加权 Poincar´e 不等式及其他温和的假设；详情见命题1 。因此，我们研究用于重尾密度采样的（3）的 Euler-Maruyama 离散化的预言机复杂度。我们的证明基于均方分析技术；这是一种分析随机微分方程数值离散化的常用技术，例如可参见Milstein 和 Tretyakov（2004年）的综述。本文的结果将均方分析推向了极限；我们考虑的重尾密度无一例外都只需具有有限方差，而这正是使用该技术的最低要求。

1.1 我们的贡献

在这项工作中，我们作出以下贡献：

- 在定理2中，我们给出了（3）的 Euler-Maruyama 离散化为获得一个样本所需迭代次数的上界，该样本与目标密度之间的 Wasserstein-2 距离不超过 ϵ。所得上界以目标密度的若干（一阶和二阶）矩π表示。我们的证明技术基于均方分析；我们表明，对于多元 t-分布，只要密度具有有限方差（这是进行均方分析的必要条件），我们的分析就不是空泛的。

- 虽然定理2 假设可以获取未归一化目标密度函数的精确梯度（称为一阶设置），但在定理3中，我们基于高斯平滑技术分析了根据函数评估估计梯度的情形（即零阶情形）。

- 我们给出了几个说明性示例，以突出一阶情形与零阶情形下结果的差异。具体而言，在第 5 节中，我们证明，对于自由度较小的多元 t-分布（因此属于真正的重尾情形），梯度估计误差由离散化误差主导。而对于自由度较大的情形（因此属于相对适度的重尾情形），离散化误差与梯度估计误差的阶相当。因此，零阶算法通过使用小批量梯度估计器，达到了与一阶算法相同的迭代复杂度。

1.2 相关工作

过去十年间，人们广泛分析了式（1）的不同离散化方法的非渐近迭代复杂度。例如，Dalalyan（2017）；Durmus 和 Moulines（2017）；Dalalyan 和 Karagulyan（2019）；Durmus 等（2019）；Lee 等（2020）；Shen 和 Lee（2019）；He 等（2020）；Chen 等（2020）；Durmus 等（2019）；Dalalyan 等（2019）；Li 和 Erdogdu（2020）；Chen 等（2020）；Chewi 等（2021a）及其参考文献，研究了在各种轻尾假设下未校正朗之万算法（ULA）的性质。特别是，Vempala 和 Wibisono（2019）；Erdogdu 和 Hosseinzadeh（2021）；Chewi 等（2021a）分析了 ULA 在适用于轻尾密度的各种泛函不等式下的表现。此外，Balasubramanian 等（2022）近期的工作分析了目标密度仅具有 H¨older 连续性时（取平均的）ULA 的表现，尽管采用的是较弱的 Fisher信息度量。

一些研究工作，例如 Dwivedi 等人（2019）；Chewi 等人（2021b）；Wu 等人（2022），分析了轻尾情形下的 Metropolis校正朗之万算法（MALA）。Lee 等（2021）；Chen 等人（2022）在各种轻尾假设下分析了近端采样算法。Dalalyan 和 Riou-Durand（2020）；Bou-Rabee 等人（2020）；Chen 等人（2020）；Ma 等人（2021）；Monmarch´e（2021）；Cao 等人（2021）；Wang 和 Wibisono（2022）；Chen 和 Vempala（2022）等研究分析了广泛使用的哈密顿蒙特卡罗算法以及欠阻尼朗之万扩散的离散化的迭代复杂度。对于轻尾情形下实践中使用的其他马尔可夫链蒙特卡罗算法的非渐近分析，我们也推荐感兴趣的读者参阅 Lu 和 Wang（2022）；Ding 和 Li（2021）。

在重尾采样的背景下，Kamatani（2018）在渐近情形中研究了经过适当修正的 Metropolis 随机游走的缩放极限。Johnson 和 Geyer（2012）针对 Metropolis随机游走算法提出了一种变量变换方法。该方法基于某些可逆变换将重尾密度转换为轻尾密度，从而可以借助有关轻尾采样算法的丰富研究成果。Yang 等人（2022）近期也探讨了类似思路。还值得一提的是，Deligiannidis 等人（2019）、Durmus 等人（2020）以及 Bierkens 等人（2019）在重尾情形下采用变换方法，分别证明了弹跳粒子采样器和之字形采样器的渐近指数遍历性。我们还指出，Andrieu 等（2021a）和 Andrieu 等（2021b）的近期研究，在多项式重尾情形下，为分段确定性马尔可夫过程蒙特卡罗、独立Metropolis-Hastings采样器和伪边缘方法等其他采样方法建立了类似的次指数遍历性结果。S.im.sekli 等人（2020）、Huang 等人（2021）以及 Zhang 和 Zhang（2022）的研究，在连续时间情形下，为由 α-稳定过程驱动、以重尾密度为平衡分布的扩散过程建立了指数遍历性结果。然而，如何为这些扩散过程的实际离散化获得收敛结果，仍然基本上是一个悬而未决的问题。

关于重尾采样非渐近预言机复杂度分析的文献极为有限。Chandrasekaran 等人(2009)研究了使用 Metropolis 随机游走算法从 s-凹分布中采样的迭代复杂度。He 等人(2022年)研究了 ULA 在一类变换密度上的应用（即通过可逆变换将重尾密度变换为轻尾密度，这与Johnson 和 Geyer (2012年)的做法类似），并建立了非渐近预言机复杂度结果。不过，他们主要关注各向同性密度的情形。Li 等人(2019)分析了一类一般 Itô扩散过程的离散化，这类扩散过程具有重尾平衡密度。第Li 等人(2019)节提供了与该研究的详细比较。5。

近期的研究包括Hsieh 等人(2018年)；Zhang 等人(2020年)；Chewi 等人(2020年)；Ahn 和 Chewi (2021)；Jiang (2021)；Li 等人(2022年)也研究了基于镜像朗之万扩散过程离散化的采样。上述研究主要关注从约束密度中采样。连续时间收敛通常在所谓的镜像 Poincar´e 不等式下进行分析；这类不等式是对 Brascamp-Lieb 不等式沿不同于加权 Poincar´e 不等式的方向所作的推广。Li 等人Li 等人(2022年)的离散化分析基于均方分析。

如前所述，我们的工作借鉴了关于加权函数不等式的文献，这些不等式由重尾密度满足。加权 Poincaré 不等式由 Blanchet 等人(2009) 和 Bobkov 和 Ledoux (2009) 提出，并通过 Brascamp-Lieb 不等式的一种推广证明适用于 s-凹密度类。感兴趣的读者还可参阅 Cattiaux etal. (2010, 2011);Bonnefont et al. (2016); Cordero-Erausquin and Gozlan (2017); Cattiaux et al.(2019，了解各种推广

and 对以下工作的改进：Blanchet 等人（2009）和 Bobkov 和 Ledoux（2009）。

1.3 记号

本文余下部分将始终使用以下记号。

- $\langle \cdot, \cdot \rangle$ 表示欧几里得内积，$|\cdot|$ 表示欧几里得范数。
- 对于两个矩阵 A 和 B，$A \preceq B$ 表示 $B - A$ 是半正定的。任意 $d \times d$ 矩阵 A 的 2-范数记为 $\|A\|_2$。I_d 是 $d \times d$ 单位矩阵。
- Δ 表示拉普拉斯算子，∇ 表示给定函数的梯度。
- $\mathcal{C}^2(\mathbb{R}^d)$表示 $\mathbb{R}^d$ 上所有二阶连续可微实值函数的集合。$\mathcal{C}_c^2(\mathbb{R}^d)$表示 $\mathcal{C}^2(\mathbb{R}^d)$中所有具有紧支撑的函数的集合。
- 两个概率测度在 $\mathbb{R}^d$、μ 和 ν 上的 Wasserstein-2 距离为

$$W_2(\mu,\nu) := \inf_{\zeta \in C(\mu,\nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \zeta(dx,dy) \right)^{\frac{1}{2}}.$$

  其中，$C(\mu,\nu)$ 表示定义在 $\mathbb{R}^d \times \mathbb{R}^d$ 上的、其边缘分布分别为 μ 和 ν 的所有测度的集合。
- Th从概率测度 ν 到概率测度 μ 的 e χ^2 散度定义为

$$\chi^2(\nu|\mu) := \int_{\mathbb{R}^d} \left(\frac{\nu(dx)}{\mu(dx)} - 1 \right)^2 \mu(dx).$$

- 伽马函数和贝塔函数定义如下：

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt, \quad \forall\, z > 0, \quad \text{and} \quad B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1} dt, \quad \forall\, x, y > 0.$$

- 对于依赖于 d的两个正的量 $f(d)$ $g(d)$，如果存在常数 $C > 0$，使得对于所有 $d > 1$均有 $f(d) \leq C g(d)$，则我们定义 $f(d) = O(g(d))$。如果存在常数 C_1、$C_2 > 0$，使得对于所有 $d > 1$均有 $C_1 g(d) \leq f(d) \leq C_2 g(d)$，则我们定义 $f(d) = \Theta(g(d))$。我们使用 O 在 O记号中隐藏对数因子。

1.4 组织结构

在第 2 节中，我们首先在某些有利于离散化分析的假设下，建立式（3）中 Itô扩散的指数遍历性。接下来，我们给出式（3）的 Euler-Maruyama 离散化的非渐近预言机复杂度的主要结果。在第 3 节中，我们提供重尾情形下所需的矩计算，以便从第 2 节的结果中获得显式速率。在第 4 节中，我们将结果扩展到零阶情形。在第 5 节中，我们给出若干说明性示例。我们在第 6 节中进一步讨论假设的含义。证明见第 7 节以及附录 A、B 和C。

2 Itô离散化与加权Poincaré不等式

本节的目标是分析式（3）中的 Itô扩散，该扩散过程以特定类别的重尾密度作为其平稳密度。令 X_0 服从分布 ρ_0，并将 X_t 的分布记为 ρ_t，对所有 $t \geq 0$均成立。对于任意函数 $\psi \in \mathcal{C}_c^2$（$\mathbb{R}^d$），式（3）的无穷小生成元为

$$\mathcal{L}\psi = -(\beta - 1)\langle \nabla V, \nabla \psi \rangle + V \Delta \psi. \tag{4}$$

因此，与（3）对应的福克-普朗克方程为

$$\partial_t \rho_t = \nabla \cdot (\beta \rho_t \nabla V + V \nabla \rho_t) = \nabla \cdot \left(\rho_t V \nabla \log \frac{\rho_t}{\pi_\beta} \right). \tag{5}$$

由此可知，在定理 1的条件下，$\pi_\beta \propto V^{-\beta}$ 是（3）的唯一平稳密度。接下来，我们考察（3）收敛到其平稳密度的性质。为此，我们引入以下假设。

假设 **1.** 存在正常数 C_V，使得对于所有 $x \in \mathbb{R}^d$,

$$\frac{\langle (\nabla^2 V)^{-1}(x)\nabla V(x), \nabla V(x)\rangle}{V(x)} \le C_V.$$

当 V具有径向对称性时，即当 $V(x)$ 时：$= \phi(|x|)$ 对某个 $\phi \in \mathcal{C}^2(\mathbb{R}_+)$ 成立，假设 1 中的条件可简化如下。注意到

$$\nabla V(x) = \frac{\phi'(|x|)}{|x|}x, \qquad \text{and} \qquad \nabla^2 V = \left(\phi''(|x|) - \frac{\phi'(|x|)}{|x|}\right)\frac{x\otimes x}{|x|^2} + \frac{\phi'(|x|)}{|x|}I_d,$$

其中 $\otimes$ 表示外积。因此，只要 ϕ 满足以下条件即可：

$$\phi'(r) \le (\phi''(r)r) \wedge (C_V\phi(r)/r), \text{ for all } r \ge 0.$$

例如，当 ϕ 是一个 p阶多项式，且 $p \ge 2$ 并具有非负系数时，该性质对 $C_V = p$ 成立。

接下来给出定理 1的以下推论，该推论的动机来自 Bobkovand Ledoux (2009) 第 2 节中的讨论。

推论**1**。 考虑定理 *1*的情形，并进一步假设假设 *1*在 $C_V \in (0, \beta+1)$下成立，则对于任意光滑的 π_β-可积函数， ϕ 在 $\mathbb{R}^d$上，

$$Var_{\pi_\beta}(\phi) \le \left(\sqrt{\beta+1} - \sqrt{C_V}\right)^{-2} \int_{\mathbb{R}^d} \langle V(x)(\nabla^2 V)^{-1}(x)\nabla\phi(x), \nabla\phi(x)\rangle \pi_\beta(x)dx. \tag{6}$$

证明。 我们从（2）出发，假设 $\int_{\mathbb{R}^d} g d\pi_\beta = 0$。于是（2）可改写为

$$(\beta+1)\int_{\mathbb{R}^d} g(x)^2\pi_\beta(x)dx \le \int_{\mathbb{R}^d} \frac{\langle (\nabla^2 V)^{-1}(x)\nabla(gV)(x), \nabla(gV)(x)\rangle}{V(x)}\pi_\beta(x)dx.$$

现在，注意到我们有以下基本界

$$\langle A(u+v), (u+v)\rangle \le r\langle Au, u\rangle + \frac{r}{r-1}\langle Av, v\rangle, \quad u, v \in \mathbb{R}^d, r > 1,$$

对于任意正定对称矩阵 $A \in \mathbb{R}^{d\times d}$。因此，我们得到

$$\begin{aligned}(\beta+1)\int_{\mathbb{R}^d} g(x)^2\pi_\beta(x)dx \le r\int_{\mathbb{R}^d} &\frac{\langle (\nabla^2 V)^{-1}(x)g(x)\nabla V(x), g(x)\nabla V(x)\rangle}{V(x)}\pi_\beta(x)dx \\ &+ \frac{r}{r-1}\int_{\mathbb{R}^d} \frac{\langle (\nabla^2 V)^{-1}(x)V(x)\nabla g(x), V(x)\nabla g(x)\rangle}{V(x)}\pi_\beta(x)dx.\end{aligned}$$

利用假设1中的条件，我们进一步得到

$$\begin{aligned}(\beta+1)\int_{\mathbb{R}^d} g(x)^2\pi_\beta(x)dx \le rC_V &\int_{\mathbb{R}^d} g(x)^2\pi_\beta(x)dx \\ &+ \frac{r}{r-1}\int_{\mathbb{R}^d} \langle V(x)(\nabla^2 V)^{-1}(x)\nabla g(x), \nabla g(x)\rangle \pi_\beta(x)dx,\end{aligned}$$

这意味着，对于任意 $r \in (1, (\beta+1)/C_V)$,

$$\int_{\mathbb{R}^d} g(x)^2\pi_\beta(x)dx \le \frac{r}{(r-1)(\beta+1-rC_V)}\int_{\mathbb{R}^d} \langle V(x)(\nabla^2 V)^{-1}(x)\nabla g(x), \nabla g(x)\rangle \pi_\beta(x)dx.$$

取 $r := \sqrt{\frac{\beta+1}{C_V}} > 1$，可得对于所有满足 $\int g d\pi_\beta = 0$ 的 g，并且

$$\int_{\mathbb{R}^d} g(x)^2\pi_\beta(x)dx \le \left(\sqrt{\beta+1} - \sqrt{C_V}\right)^{-2} \int_{\mathbb{R}^d} \langle V(x)(\nabla^2 V)^{-1}(x)\nabla g(x), \nabla g(x)\rangle \pi_\beta(x)dx.$$

对于所有一般的 ϕ，令 $g=\phi-\int\phi d\pi_\beta$，可得

$$Var_{\pi_\beta}(\phi)\leq\left(\sqrt{\beta+1}-\sqrt{C_V}\right)^{-2}\int_{\mathbb{R}^d}\langle V(x)(\nabla^2V)^{-1}(x)\nabla\phi(x),\nabla\phi(x)\rangle\pi_\beta(x)dx.$$

□

W时 V 是强凸的，假设 1 满足以下充分条件。

假设 2. 函数 $V\colon\mathbb{R}^d\to(0,\infty)$ 二阶连续可微，且 V 满足

(1) V是 α-强凸的，即对所有 $x\in\mathbb{R}^d$，$\nabla^2V(x)\succeq\alpha I_d$。

(2) 存在一个正常数 C_V，使得对所有 $x\in\mathbb{R}^d$，

$$\frac{\langle\nabla V(x),\nabla V(x)\rangle}{V(x)}\leq\alpha C_V.$$

以下结果可由假设 2直接得到。

引理 1。 设 $\beta>d$。若假设 *2* 在 $C_V\in(0,\beta+1)$下成立，则对于任意光滑的、 π_β 可积函数 ϕ在 $\mathbb{R}^d$上，有

$$Var_{\pi_\beta}(\phi)\leq\alpha^{-1}\left(\sqrt{\beta+1}-\sqrt{C_V}\right)^{-2}\int_{\mathbb{R}^d}V(x)|\nabla\phi(x)|^2\pi_\beta(x)dx.\tag{7}$$

由（7），我们可以证明沿着（ χ^2-散度的 3）中的指数衰减。下面命题的证明是标准的，此处给出证明以求完整。

命题 1。 在引理 *1*的条件下，对于 (X_t) 如下扩散过程 (3)，其中 ρ_t 是 X_t的分布，我们有

$$\chi^2(\rho_t|\pi_\beta)\leq\exp\left(-2\alpha\left(\sqrt{\beta+1}-\sqrt{C_V}\right)^2t\right)\chi^2(\rho_0|\pi_\beta).\tag{8}$$

命题 *1*的证明。 首先，根据（ χ^2（$\rho_t|\pi$）可以利用（5）计算其导数，

$$\begin{aligned}\frac{d}{dt}\chi^2(\rho_t|\pi_\beta)&=\frac{d}{dt}\int_{\mathbb{R}^d}\left(\frac{\rho_t(x)}{\pi_\beta(x)}-1\right)^2\pi_\beta(x)dx\\&=2\int_{\mathbb{R}^d}\partial_t\rho_t(x)\left(\frac{\rho_t(x)}{\pi_\beta(x)}-1\right)dx\\&=-2\int_{\mathbb{R}^d}\left\langle\nabla\left(\frac{\rho_t}{\pi_\beta}\right)(x),\nabla\log\left(\frac{\rho_t}{\pi_\beta}\right)(x)\right\rangle V(x)\rho_t(x)dx\\&=-2\int_{\mathbb{R}^d}V(x)\left|\nabla\left(\frac{\rho_t}{\pi_\beta}\right)(x)\right|^2\pi_\beta(x)dx.\end{aligned}$$

根据（7），可得

$$\begin{aligned}\frac{d}{dt}\chi^2(\rho_t|\pi_\beta)&\leq-2\alpha\left(\sqrt{\beta+1}-\sqrt{C_V}\right)^2\mathrm{Var}_{\pi_\beta}\left(\frac{\rho_t}{\pi_\beta}\right)\\&=-2\alpha\left(\sqrt{\beta+1}-\sqrt{C_V}\right)^2\chi^2(\rho_t|\pi_\beta).\end{aligned}$$

最后，(8) 根据格朗沃尔不等式得证。 □

上述结果表明，对于满足假设 π_β 2 的这一类，(3) 中的 Itô扩散会以指数速度收敛至其平稳密度。因此，(3) 的时间离散化为从该类密度中采样提供了一种实用方法。(3) 的 Euler-Maruyama 离散化如下：

$$x_{k+1} = x_k - h(\beta-1)\nabla V(x_k) + \sqrt{2hV(x_k)}\xi_{k+1}, \tag{9}$$

其中 $h > 0$ 为步长，且 $\{\xi\}_{k=1}^{\infty}$ 是取值于 $\mathbb{R}^d$ 的一列独立同分布标准高斯随机向量。现在，我们给出关于从 π_β 采样时，式（9）迭代复杂度的主要结果。我们基于均方分析，在 W_2 度量下陈述离散化结果。特别地，我们强调，命题 1 除假设 $\beta > d$外，还要求条件 2，而下面的定理 2 则不要求。第 6节将重新审视这些条件，并提供更多见解。如何在更强的 χ^2-散度下获得收敛结果，将留待未来研究。

定理 **2.** 设 V满足参数为 $L > 0$的梯度Lipschitz条件，并满足假设 *2*，其中

$$\delta := \frac{\beta - 1 - \frac{1}{4}C_V d}{\frac{1}{4}C_V d} > 0. \tag{10}$$

设（x_k）$_{k=0}^{\infty}$ 由式 （9）生成，其中 ν_k 表示 x_k的分布，对所有 $k \geq 0$均成立。那么取步长为，

$$h < \min\left(\frac{1}{4(\beta-1)L}, \frac{2\delta}{3(1+\delta)\alpha(\beta-1)}\right),$$

马尔可夫链 (x_k) 上 Wasserstein-2 距离的衰减可以用以下方程描述：对于所有 $k \geq 1$,

$$W_2(\nu_k, \pi_\beta) \leq (1-A)^k W_2(\nu_0, \pi_\beta) + \frac{C}{A} + \frac{B}{\sqrt{A(2-A)}}. \tag{11}$$

其中 A, B 以及 C 分别见 (51)、(52) 和 (53)。

备注1（常数 δ）。现在，我们基于指数压缩性论证，说明常数 δ 的定义及其所满足条件的动机。

定义 1（指数压缩性）。设 X_t、 Y_t 是同一个随机微分方程（SDE）的两个不同解，其初始条件分别为 x、y 。如果存在常数 $\kappa > 0$，使得下式成立，则称该 SDE 是 W_2-指数压缩的：

$$W_2(L(X_t), L(Y_t)) \leq e^{-\kappa t}\ |x-y|,$$

其中， $L(X)$ 表示 X 的分布律。

一致耗散性是一致耗散性的充分条件（Gorham 等人，*2019*，定理 10）。方程 (3) 的一致耗散条件可表示为

$$-(\beta-1)\langle \nabla V(x) - \nabla V(y), x-y\rangle + \frac{1}{2}\left\|\sqrt{2V(x)}I_d - \sqrt{2V(y)}I_d\right\|_F^2 \leq -\kappa|x-y|^2,$$

或等价地表示为

$$-(\beta-1)\langle \nabla V(x) - \nabla V(y), x-y\rangle + d|\sqrt{V(x)} - \sqrt{V(y)}|^2 \leq -\kappa|x-y|^2.$$

当 V 满足假设*2*时，上述一致耗散条件的一个充分条件为

$$-\alpha(\beta-1)|x-y|^2 + \frac{d}{4}\alpha C_V|x-y|^2 \leq -\kappa|x-y|^2,$$

或等价地，

$$\alpha\left(\beta - 1 - \frac{d}{4}C_V\right) \leq \kappa.$$

充分条件与以下条件一致： $\delta > 0$ 定理 *2*中的条件，这也启发了定理*2*中的假设。

注 2 （迭代复杂度）。 根据定理 2，我们可以计算在 Wasserstein-2 距离下达到 ϵ 精度所需的迭代复杂度阶数。根据 (51)、(52)、(53)，可得

$$\frac{C}{A} = \frac{9(\delta+1)L}{\alpha\delta} d^{\frac{1}{2}} h^{\frac{1}{2}} \mathbb{E}_{\pi_\beta}\left[V(X)\right]^{\frac{1}{2}} + \frac{6(\delta+1)L}{\alpha\delta}(\beta-1)h\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}},$$
$$\frac{B}{\sqrt{A(2-A)}} \leq \frac{8(\delta+3)}{\delta} d^{\frac{1}{2}} h^{\frac{1}{2}} \mathbb{E}_{\pi_\beta}\left[V(X)\right]^{\frac{1}{2}} + \frac{8(\delta+3)}{\delta}(\beta-1)h\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}.$$

上述公式表明

$$\frac{C}{A} + \frac{B}{\sqrt{A(2-A)}} \leq \frac{9(\delta+3)}{\delta}\left(1+\frac{L}{\alpha}\right)\left(d^{\frac{1}{2}} h^{\frac{1}{2}} \mathbb{E}_{\pi_\beta}\left[V(X)\right]^{\frac{1}{2}} + (\beta-1)h\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}\right).$$

因此，当步长 h 满足以下条件时，我们得到 $\frac{C}{A} + \frac{B}{\sqrt{A(2-A)}} < \epsilon/2$

$$h < \min\left\{ \frac{\delta^2 \mathbb{E}_{\pi_\beta}\left[V(X)\right]^{-1}\epsilon^2}{81d(\delta+3)^2(1+\frac{L}{\alpha})^2}, \frac{\delta\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{-\frac{1}{2}}\epsilon}{81(\beta-1)(\delta+3)(1+\frac{L}{\alpha})} \right\}. \tag{12}$$

定义 $K_\epsilon = \log\left(2W_2(\nu_0,\pi_\beta)/\epsilon\right)$，则对所有满足 $k \geq K$ 的情形，都有 $W_2(\nu_k,\pi_\beta) < \epsilon$

$$\begin{aligned} K &= \frac{3(1+\delta)}{\alpha(\beta-1)\delta h^*} K_\epsilon \\ &\leq 273 \max\left\{ \frac{(\delta+3)^3(1+\frac{L}{\alpha})^2 d\mathbb{E}_{\pi_\beta}\left[V(X)\right]}{\alpha\delta^3(\beta-1)\epsilon^2}, \frac{(\delta+3)^2(1+\frac{L}{\alpha})\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}}{\alpha\delta^2\epsilon} \right\} K_\epsilon. \end{aligned} \tag{13}$$

回顾 δ 在 (10)中的定义。 K 的阶数取决于 δ的阶数。也就是说，我们有以下两种情况:

- 如果 $\delta = O(1)$ 且 $\beta = O(d)$，则有

$$K = \tilde{O}\left(\frac{1}{\alpha\epsilon^2}\left(1+\frac{L}{\alpha}\right)^2 \mathbb{E}_{\pi_\beta}\left[V(X)\right] + \frac{1}{\alpha\epsilon}\left(1+\frac{L}{\alpha}\right)\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}\right).$$

- 如果 $\delta = O(1/d)$ 且 $\beta = O(d)$，则有

$$K = \tilde{O}\left(\frac{d^3}{\alpha\epsilon^2}\left(1+\frac{L}{\alpha}\right)^2 \mathbb{E}_{\pi_\beta}\left[V(X)\right] + \frac{d^2}{\alpha\epsilon}\left(1+\frac{L}{\alpha}\right)\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}\right).$$

为了从备注 2 中获得更明确的迭代复杂度界，需要计算以下两个量的界： $\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]$ 和 $\mathbb{E}_{\pi_\beta}\left[V(X)\right]$。

3 矩界

本节将在定理 2 的条件下计算矩界。

3.1 示例：多元 t-分布

我们首先从各向同性情形开始。

命题 2. 设 $\pi_\beta = Z_\beta^{-1} V^{-\beta}$ 与 $\beta > d/2+1$、 $V(x) = 1+|x|^2$ 以及 $Z_\beta = \int_{\mathbb{R}^d}(1+|x|^2)^{-\beta}dx$。我们有

$$\mathbb{E}_{\pi_\beta}\left[V(X)\right] = \frac{\beta-1}{\beta-1-\frac{d}{2}} \quad and \quad \mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] = \frac{2d}{\beta-1-\frac{d}{2}}. \tag{14}$$

证明. 令 $A_d(1)$ 表示 d 维单位球面的表面积。通过标准计算可得，对于所有 $\beta > \frac{d}{2}$，

$$Z_\beta = \int_{\mathbb{R}^d} (1+|x|^2)^{-\beta} dx = \int_0^\infty (1+r^2)^{-\beta} r^{d-1} dr A_d(1) = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \int_0^\infty (1+R)^{-\beta} R^{\frac{d}{2}-1} dR$$
$$= \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \int_0^1 u^{\frac{d}{2}-1}(1-u)^{\beta-\frac{d}{2}-1} du = \frac{\pi^{\frac{d}{2}} B(\frac{d}{2}, \beta-\frac{d}{2})}{\Gamma(\frac{d}{2})},$$

其中 B 是贝塔函数。在上述计算中，第二个恒等式由转换到极坐标得出。第三个恒等式由代入 $R = r^2$ 得出，第四个恒等式由代入 $u = R/(1+R)$ 得出。因此，对于所有 $\beta > d/2+1$，我们有

$$\mathbb{E}_{\pi_\beta}[V(X)] = Z_\beta^{-1} \int_{\mathbb{R}^d} (1+|x|^2)(1+|x|^2)^{-\beta} dx = \frac{Z_{\beta-1}}{Z_\beta} = \frac{\pi^{\frac{d}{2}} B(\frac{d}{2}, \beta-1-\frac{d}{2})}{\Gamma(\frac{d}{2})} \frac{\Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}} B(\frac{d}{2}, \beta-\frac{d}{2})}$$
$$= \frac{B(\frac{d}{2}, \beta-1-\frac{d}{2})}{B(\frac{d}{2}, \beta-\frac{d}{2})} = \frac{\Gamma(\frac{d}{2})\Gamma(\beta-1-\frac{d}{2})}{\Gamma(\beta-1)} \frac{\Gamma(\beta)}{\Gamma(\frac{d}{2})\Gamma(\beta-\frac{d}{2})} = \frac{\beta-1}{\beta-1-\frac{d}{2}}.$$

其中，第四个恒等式由贝塔函数的性质得出：$B(x,y) = \Gamma(x)\Gamma(y)\Gamma(x+y)$；第五个恒等式则由 Γ 函数的性质得出：$\Gamma(1+z) = z\Gamma(z)$。对于另一个期望，我们有

$$\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] = Z_\beta^{-1} \int_{\mathbb{R}^d} |2x|^2 (1+|x|^2)^{-\beta} dx = 4Z_\beta^{-1} A_{d-1}(1) \int_0^\infty r^2 (1+r^2)^{-\beta} r^{d-1} dr$$
$$= \frac{4\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) Z_\beta} \int_0^\infty R^{\frac{d}{2}} (1+R)^{-\beta} dR = \frac{4\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) Z_\beta} \int_0^1 u^{\frac{d}{2}} (1-u)^{\beta-\frac{d}{2}-2} du$$
$$= \frac{4\pi^{\frac{d}{2}} B(\frac{d}{2}+1, \beta-\frac{d}{2}-1)}{\Gamma(\frac{d}{2})} \frac{\Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}} B(\frac{d}{2}, \beta-\frac{d}{2})} = \frac{4B(\frac{d}{2}+1, \beta-\frac{d}{2}-1)}{B(\frac{d}{2}, \beta-\frac{d}{2})}$$
$$= 4\frac{\Gamma(\frac{d}{2}+1)\Gamma(\beta-\frac{d}{2}-1)}{\Gamma(\beta)} \frac{\Gamma(\beta)}{\Gamma(\frac{d}{2})\Gamma(\beta-\frac{d}{2})} = \frac{2d}{\beta-\frac{d}{2}-1},$$

这里，我们在上述计算中应用了相同的代换以及贝塔函数和伽马函数的性质。

□

注释 3. 若 π_β 是各向同性多元 t-分布类，结合命题2中的结果，则关于维度参数 d 的两个期望的阶如下：

- 当 $\beta > \frac{d}{2}+1$ 且 $\beta - 1 - \frac{d}{2} = O(d)$ 时，我们有
$$\mathbb{E}_{\pi_\beta}[V(X)] = O(1), \qquad and \qquad \mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] = O(1).$$
- 当 $\beta > \frac{d}{2}+1$ 且 $\beta - 1 - \frac{d}{2} = O(1)$ 时，我们有
$$\mathbb{E}_{\pi_\beta}[V(X)] = O(d), \qquad and \qquad \mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] = O(d).$$

对于一类一般的非各向同性多元 t 分布，我们考虑 $\pi_\beta = Z_\beta^{-1} V^{-\beta}$ 其中 $V(x) = 1 + x^T \Sigma x$ 其中 Σ 是一个严格正定的 $d \times d$ 矩阵。在 Roth (2012年中，已证明对于任意$\beta > \frac{d}{2}$，归一化常数为

$$Z_\beta = \frac{\Gamma(\frac{\nu}{2})\pi^{\frac{d}{2}}\sqrt{\det(\Sigma)}}{\Gamma(\frac{\nu+d}{2})} = \frac{\Gamma(\beta-\frac{d}{2})\pi^{\frac{d}{2}}\sqrt{\det(\Sigma)}}{\Gamma(\beta)}.$$

因此，对于任意 $\beta > \frac{d}{2}+1$，我们有

$$\mathbb{E}_{\pi_\beta}[V(X)] = \frac{Z_{\beta-1}}{Z_\beta} = \frac{\Gamma(\beta)\Gamma(\beta-1-\frac{d}{2})}{\Gamma(\beta-1)\Gamma(\beta-\frac{d}{2})} = \frac{\beta-1}{\beta-1-\frac{d}{2}},$$

and

$$\begin{aligned}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] &= Z_\beta^{-1}\int_{\mathbb{R}^d}\langle\nabla V(x), V(x)^{-\beta}\nabla V(x)\rangle dx \\ &= -Z_\beta^{-1}\int_{\mathbb{R}^d} V(x)\nabla\cdot\left(V(x)^{-\beta}\nabla V(x)\right)dx \\ &= \beta\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] - Z_\beta^{-1}\int_{\mathbb{R}^d}\Delta V(x)V(x)^{-(\beta-1)}dx.\end{aligned}$$

上述恒等式意味着

$$\begin{aligned}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] &= (\beta-1)^{-1}Z_\beta^{-1}\int_{\mathbb{R}^d}\Delta V(x)V(x)^{-(\beta-1)}dx \\ &\le (\beta-1)^{-1}Z_\beta^{-1}\int_{\mathbb{R}^d}\mathrm{trace}(\nabla^2 V(x))V(x)^{-(\beta-1)}dx \\ &\le \frac{\mathrm{trace}(\Sigma)}{\beta-1}\mathbb{E}_{\pi_\beta}\left[V(X)\right] \\ &\le \frac{\mathrm{trace}(\Sigma)}{\beta-1-\frac{d}{2}}\end{aligned}$$

其中第二个不等式是由 $\nabla^2 V(x) = \Sigma$这一事实得出的。

注 4. 若 π_β 属于非各向同性多元 t 分布这一类，则关于维度参数 d的两个期望的阶如下，

- 当 $\beta > \frac{d}{2}+1$ 且 $\beta-1-\frac{d}{2} = O(d)$ 时，我们有

$$\mathbb{E}_{\pi_\beta}\left[V(X)\right] = O(1), \qquad and \qquad \mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] = O(d^{-1}\mathrm{trace}(\Sigma)).$$

- 当 $\beta > \frac{d}{2}+1$ 且 $\beta-1-\frac{d}{2} = O(1)$ 时，我们有

$$\mathbb{E}_{\pi_\beta}\left[V(X)\right] = O(d), \qquad and \qquad \mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] = O(\mathrm{trace}(\Sigma)).$$

3.2 具有类二次型 V 且在 bal 外部的非各向同性密度 l

本节中，我们估计一类非各向同性密度的期望，其形式为 $\pi_\beta \propto V^{-\beta}$，其中 V满足以下李雅普诺夫条件:

$$\exists\ \varepsilon, R>0 \text{ such that } \Delta V(x) - (\beta-1)\frac{|\nabla V(x)|^2}{V(x)} \le -\varepsilon \qquad \forall\ |x|\ge R. \tag{15}$$

上述李雅普诺夫条件刻画了在半径为R的球外呈“二次型”的 V 类。若假设 V 具有利普希茨梯度，那么当 β 足够大时，只要 V在 $|x|\ge R$ 处满足 PL 不等式 $|\nabla V(x)|^2 \ge a^2 V(x)$，其中 $a>0$ 满足某些条件，上述假设便成立；二次增长正是由该不等式推得的。特别地，若 V 满足参数为 L的梯度利普希茨假设，则对所有 $\beta \ge 1 + a^{-2}(dL+\varepsilon)$,

$$\Delta V(x) - (\beta-1)\frac{|\nabla V(x)|^2}{V(x)} \le dL - (\beta-1)a^2 \le -\varepsilon \qquad \forall\ |x|\ge R,$$

从而得到（15）中的李雅普诺夫条件。

命题 3. 若 $V\in\mathcal{C}^2(\mathbb{R}^d)$ 是正的， L-梯度利普希茨且满足 (15)，则有

$$\mathbb{E}_{\pi_\beta}\left[V(x)\right] \le (dL+\varepsilon)\max_{|x|\le R} V(x), \quad and \quad \mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] \le \frac{dL\,(dL+\varepsilon)}{(\beta-1)}\max_{|x|\le R} V(X). \tag{16}$$

证明. 由于 $\mathcal{L}$ 是遍历的，其平稳分布为 π_β，因此有

$$\mathbb{E}_{\pi_\beta}[V(X)] = \lim_{t\to\infty} \mathbb{E}[V(X_t)],$$

满足（X_t）$_{t\geq 0}$ 是以初始条件 $X_0 = x$为初值条件的（3）的解。我们首先给 $\mathbb{E}[V$（X_t）$]$ 取一个上界，然后令 $t \to \infty$。令（P_t）$_{t\geq 0}$ 为（3）的马尔可夫半群，则

$$\frac{d}{dt}\mathbb{E}_{\pi_\beta}[V(X_t)] = \frac{d}{dt}P_tV(x) = P_t\mathcal{L}V(x).$$

由（4），可得

$$\begin{aligned}\mathcal{L}V(x) &= V(x)\left[\Delta V(x) - (\beta-1)\frac{|\nabla V(x)|^2}{V(x)}\right]\\ &\leq V(x)\left(-\varepsilon 1_{|x|\geq R} + dL1_{|x|<R}\right)\\ &\leq -\varepsilon V(x) + (dL+\varepsilon)\max_{|x|\leq R} V(x),\end{aligned}$$

其中 fi　　第一个不等式由（15）以及 $\Delta V \leq d\left\|\nabla^2 V\right\|_2$这一事实得出。因此我们得到　　in

$$\frac{d}{dt}P_tV(x) \leq -\varepsilon P_tV(x) + (dL+\varepsilon)\max_{|x|\leq R} V(x),$$

由格朗沃尔不等式可知

$$\mathbb{E}_{\pi_\beta}[V(X_t)] = P_tV(x) \leq V(x)e^{-\varepsilon t} + \left(1-e^{-\varepsilon t}\right)(dL+\varepsilon)\max_{|x|\leq R} V(x).$$

因此，令 $t \to \infty$，可得 $\mathbb{E}_{\pi_\beta}[V(X)] \leq (dL+\varepsilon)\max_{|x|\leq R} V(x)$。对于另一个期望，我们有

$$\begin{aligned}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] &= Z_\beta^{-1}\int_{\mathbb{R}^d}\langle\nabla V(x), V(x)^{-\beta}\nabla V(x)\rangle dx\\ &= -Z_\beta^{-1}\int_{\mathbb{R}^d} V(x)\nabla\cdot\left(V(x)^{-\beta}\nabla V(x)\right)dx\\ &= \beta\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] - Z_\beta^{-1}\int_{\mathbb{R}^d}\Delta V(x)V(x)^{-(\beta-1)}dx.\end{aligned}$$

由上述恒等式可知

$$\begin{aligned}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] &= (\beta-1)^{-1}Z_\beta^{-1}\int_{\mathbb{R}^d}\Delta V(x)V(x)^{-(\beta-1)}dx\\ &\leq (\beta-1)^{-1}Z_\beta^{-1}\int_{\mathbb{R}^d}\mathrm{trace}(\nabla^2V(x))V(x)^{-(\beta-1)}dx\\ &\leq (\beta-1)^{-1}Z_\beta^{-1}dL\int_{\mathbb{R}^d}V(x)^{-(\beta-1)}dx\\ &= \frac{dL}{\beta-1}\mathbb{E}_{\pi_\beta}[V(X)]\\ &\leq \frac{dL(dL+\varepsilon)}{\beta-1}\max_{|x|\leq R}V(x).\end{aligned}$$

□

3.3 一般情形

接下来讨论一般情形，其中 $\pi_\beta = Z_\beta^{-1}V^\beta$ 且 $V \in \mathcal{C}^2(\mathbb{R}^d)$ 为正的，因此存在常数 α、$L > 0$ 和 $\alpha I_d \preceq \nabla^2V(x) \preceq LI_d$，对所有 $x \in \mathbb{R}^d$均成立。由于 V是强凸的，因此存在唯一的$x^* \in \mathbb{R}^d$，使得对所有 $x \in \mathbb{R}^d$，$V(x) \geq V(x^*) > 0$，且 $\nabla V(x^*) = 0$。不失一般性，我们假设 $x^* = 0$。

命题4。设 $\beta > \frac{d}{2}+1$。若 $V \in \mathcal{C}^2(\mathbb{R}^d)$ 为正的、 α-强凸且 L-梯度利普希茨，则对任意 $r \in (0, \beta - \frac{d}{2} - 1)$，有

$$\mathbb{E}_{\pi_\beta}[V(X)] \le \left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}} V(0) \left(\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}}, \tag{17}$$

$$\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] \le \frac{dL}{\beta-1}\left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}} V(0) \left(\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}}. \tag{18}$$

证明。对于任意 $r \in (0, \beta - \frac{d}{2} - 1)$，我们有

$$\mathbb{E}_{\pi_\beta}[V(X)] = \frac{\int_{\mathbb{R}^d} V(x)V(x)^{-\beta}dx}{Z_\beta} = \frac{Z_{\beta-1}}{Z_\beta} \le \left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}} V(0) \left(\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}}.$$

其中，最后一个不等式由引理 3 得出。对于另一个期望，我们有

$$\begin{aligned}
\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] &= Z_\beta^{-1}\int_{\mathbb{R}^d} \langle \nabla V(x), V(x)^{-\beta}\nabla V(x)\rangle dx \\
&= -Z_\beta^{-1}\int_{\mathbb{R}^d} V(x)\nabla\cdot\left(V(x)^{-\beta}\nabla V(x)\right)dx \\
&= \beta\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] - Z_\beta^{-1}\int_{\mathbb{R}^d}\Delta V(x)V(x)^{-(\beta-1)}dx.
\end{aligned}$$

上述恒等式意味着

$$\begin{aligned}
\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] &= (\beta-1)^{-1}Z_\beta^{-1}\int_{\mathbb{R}^d}\Delta V(x)V(x)^{-(\beta-1)}dx \\
&\le (\beta-1)^{-1}Z_\beta^{-1}\int_{\mathbb{R}^d}\mathrm{trace}(\nabla^2 V(x))V(x)^{-(\beta-1)}dx \\
&\le (\beta-1)^{-1}Z_\beta^{-1}dL\int_{\mathbb{R}^d}V(x)^{-(\beta-1)}dx \\
&= \frac{dL}{\beta-1}\frac{Z_{\beta-1}}{Z_\beta} \\
&\le \frac{dL}{\beta-1}\left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}} V(0)\left(\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}}.
\end{aligned}$$

其中，最后一个不等式同样由引理 3 得出。 □

注 5。 式 (17) 和 (18) 中出现了伽马函数之比。利用伽马函数的性质，可以将该比值显式写出。

- 当 d 为偶数且 $d = 2k$ 对于某个整数 k时，

$$\begin{aligned}
\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})} &= \frac{\Gamma(r)}{\Gamma(\frac{d}{2}+r)}\frac{\Gamma(\beta)}{\Gamma(\beta-\frac{d}{2})} = \frac{\Gamma(r)}{\Gamma(r)\prod_{i=1}^k(\frac{d}{2}+r-i)}\frac{\Gamma(\beta-\frac{d}{2})\prod_{i=1}^k(\beta-i)}{\Gamma(\beta-\frac{d}{2})} \\
&= \frac{\prod_{i=1}^k(\beta-i)}{\prod_{i=1}^k(\frac{d}{2}+r-i)} \le \left(\frac{\beta-\frac{d}{2}}{r}\right)^{\frac{d}{2}},
\end{aligned}$$

- 当 d 是一个奇数，且对于某个整数 k有 $d = 2k-1$ 时，

$$\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})} = \frac{\Gamma(r)}{\Gamma(\frac{d}{2}+r)}\frac{\Gamma(\beta)}{\Gamma(\beta-\frac{d}{2})} = \frac{\Gamma(r)}{\Gamma(\frac{1}{2}+r)\prod_{i=1}^{k-1}(\frac{d}{2}+r-i)}\frac{\Gamma(\beta-\frac{d}{2}+\frac{1}{2})\prod_{i=1}^{k-1}(\beta-i)}{\Gamma(\beta-\frac{d}{2})}$$

$$
\begin{aligned}
&= \frac{\prod_{i=1}^{k-1}(\beta-i)}{\prod_{i=1}^{k-1}(\frac{d}{2}+r-i)} \frac{r^{-1}\Gamma(r+1)}{\Gamma(\frac{1}{2}+r)} \frac{\Gamma(\beta-\frac{d}{2}+\frac{1}{2})}{\Gamma(\beta-\frac{d}{2})} \\
&\leq \left(\frac{\beta-\frac{d}{2}+\frac{1}{2}}{r+\frac{1}{2}}\right)^{k-1} r^{-1}(1+r)^{\frac{1}{2}} \left(\beta-\frac{d}{2}+\frac{1}{2}\right)^{\frac{1}{2}} \\
&\leq \sqrt{\frac{1+r}{r}} \left(\frac{\beta-\frac{d}{2}}{r}\right)^{\frac{d}{2}},
\end{aligned}
$$

其中第一个不等式由 Gautschi 不等式推出（Ismail 和 Muldoon，*1994*）。

注释 6. 结合定理 *4* 以及注释 *5*中的上界，我们可以得到 $\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]$和 $\mathbb{E}_{\pi_\beta}\left[V(X)\right]$的估计：对于任意 $r \in (0, \beta-\frac{d}{2}-1)$,

$$
\mathbb{E}_{\pi_\beta}\left[V(X)\right] \leq V(0)\left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}} \left(\frac{1+r}{r}\right)^{\frac{1}{2(\beta-\frac{d}{2}-r)}} \left(\frac{\beta-\frac{d}{2}}{r}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}}, \tag{19}
$$

$$
\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right] \leq \frac{V(0)dL}{\beta-1}\left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}} \left(\frac{1+r}{r}\right)^{\frac{1}{2(\beta-\frac{d}{2}-r)}} \left(\frac{\beta-\frac{d}{2}}{r}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}}. \tag{20}
$$

4 零阶 Itˆo 离散化

此前我们考虑的是函数 V 的梯度可由解析方式获得的情形，现在我们考虑只能获得函数评估的情形。该情形称为零阶情形，近期在 Dwivedi 等人（2019）；Lee 等（2021）；Roy 等人（2022年）关于采样复杂度的工作中得到了研究。在该情形下，我们通过零阶信息，即函数评估，构造梯度的近似。为简便起见，我们考虑获得精确函数评估的情形。基于高斯平滑技术（Nesterov 和 Spokoiny,2017；Roy 等人, 2022年），对于任意 $x \in \mathbb{R}^d$，我们将零阶梯度估计器 $g_{\sigma,\, m}(x)$ 定义为

$$
g_{\sigma,m}(x) := \frac{1}{m}\sum_{i=1}^{m} \frac{V(x+\sigma u_i)-V(x)}{\sigma} u_i \tag{21}
$$

其中，假设 $u_i \sim \mathcal{N}(0, I_d)$ 是独立同分布的。参数 m 称为批量大小参数。于是，用于采样 π_β 的零阶算法为

$$
x_{k+1} = x_k - h(\beta-1)g_{\sigma,m}(x_k) + \sqrt{2V(x_k)}\xi_{k+1} \tag{22}
$$

其中， $h > 0$ 是步长， $\{\xi_{k+1}\}_{k=0}^{\infty}$ 是取值于 $\mathbb{R}^d$中的一列独立同分布的标准高斯随机向量。根据 Balasubramanian 和 Ghadimi (2022年) 和 Roy 等人 (2022年)，我们回顾 $g_{\sigma,\, m}$ 的以下性质。

命题 5. *(*Roy 等人，2022年，第 8.1 节）假设 V 是 L-梯度Lipschitz。定义 $\zeta_k = g_{\sigma,\, m}(x_k) - \nabla V(x_k)$ 其中 $g_{\sigma,\, m}$ 定义见 (21) 且 $\{x_k\}_{k=0}^{\infty}$ 由 (22)生成。对任意 $k \geq 0$,

$$
\mathbb{E}\left[\left|\mathbb{E}\left[\zeta_k | x_k\right]\right|^2\right] \leq L^2\sigma^2 d, \tag{23}
$$

and

$$
\mathbb{E}\left[\left|\zeta_k - \mathbb{E}\left[\zeta_k | x_k\right]\right|^2\right] \leq \frac{\sigma^2}{2m}L^2(d+3)^3 + \frac{2(d+5)}{m}\mathbb{E}\left[|\nabla V(x_k)|^2\right]. \tag{24}
$$

定理 3. 假设 V 是参数为 $L > 0$ 的梯度Lipschitz函数，并满足假设2 ，其中 δ见 (10)。令 g_σ, m 定义见 (21) 和 $(x_k)_{k=0}^{\infty}$ 由 (22) 生成，且 $x_k \sim \nu_k$ 对所有 $k \geq 0$ 均成立。则对于时间步长

$$h < \min\left\{\frac{2\delta}{3(1+\delta)\alpha(\beta-1)}, \frac{\alpha m\delta}{24(1+\delta)(\beta-1)(d+5)L^2}, \frac{1}{4(\beta-1)L}\right\}, \tag{25}$$

沿马尔可夫链 $(x_k)_{k=0}^{\infty}$ 的 Wasserstein-2 距离衰减可由以下方程描述。对于所有 $k \geq 1$,

$$W_2(\nu_k, \pi_\beta) \leq (1-A')^k W_2(\nu_0, \pi_\beta) + \frac{C'}{A'} + \frac{B'}{\sqrt{A'(2-A')}}. \tag{26}$$

其中 A', B' 和 C' 分别由 (60)、(61) 和 (62)给出。

注 7. 利用定理 3，我们可以研究达到 ε-精度（以 Wasserstein-2 距离衡量）所需的迭代复杂度。接下来的讨论聚焦于迭代复杂度对维数和 ε的依赖关系。当 $\beta = \Theta(d)$ 且 α时， $L = \Theta(1)$，并且当 h 满足 (25)时，我们有

$$A' = O\left(\delta dh\right), \qquad \frac{C'}{A'} = O\left(\frac{(dh\mathbb{E}_{\pi_\beta}\left[V(X)\right])^{\frac{1}{2}} + dh\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}} + \sigma d^{\frac{1}{2}}}{\delta}\right),$$
$$\frac{B'}{\sqrt{A'(2-A')}} = O\left(\left(\frac{dh}{\delta} + \frac{dh^{\frac{1}{2}}}{(\delta m)^{\frac{1}{2}}}\right)\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}} + \frac{(dh)^{\frac{1}{2}}}{\delta}\mathbb{E}_{\pi_\beta}\left[V(X)\right] + \frac{\sigma d^2 h^{\frac{1}{2}}}{(\delta m)^{\frac{1}{2}}}\right).$$

为确保 $W_2(\nu_K, \pi_\beta) < \varepsilon$，我们要求以下各项都

$$(1-A')^K W_2(\nu_0, \pi_\beta), \qquad \frac{C'}{A'}, \qquad \frac{B'}{\sqrt{A'(2-A')}},$$

小于 $\varepsilon/3$。令 $\sigma = \varepsilon\delta/\sqrt{d}$，并且

$$h = O\left(\min\left\{\frac{(\varepsilon\delta)^2}{d}\mathbb{E}_{\pi_\beta}\left[V(X)\right]^{-1}, \frac{\varepsilon\delta}{d}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{-\frac{1}{2}}, \frac{\varepsilon^2\delta m}{d^2}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{-1}\right\}\right),$$

因此，我们得到迭代复杂度 K 的阶为

$$K = \tilde{O}\left(\max\left\{\frac{1}{\varepsilon^2\delta^3}\mathbb{E}_{\pi_\beta}\left[V(X)\right], \frac{1}{\varepsilon\delta^2}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}, \frac{d}{\varepsilon^2\delta^2 m}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]\right\}\right). \tag{27}$$

因此，函数评估次数为 mK。

5 说明性示例

现在，我们通过说明性示例来阐明研究结果的影响。

5.1 多元 t 分布：大自由度

我们首先考虑自由度为 $d+2$ 的各向同性多元 t 分布。我们选择$V(x) = 1+|x|^2$、 $\beta = d+1$ 和 $\pi_\beta(x) \propto V(x)^{-\beta} = (1+|x|^2)^{-(d+1)}$。采用这一选择后， V和 β使得 V满足假设 2，其中 $\alpha = 2$、 $C_V = 2$，并且 V 是具有 $L = 2$ 的 L-Lipschitz 梯度。定理 2 中的常数 δ 变为 $\delta = 1$。此外，根据命题 2， $\mathbb{E}_{\pi_\beta}[V(X)] = 2$ 且 $\mathbb{E}_{\pi_\beta}[|\nabla V(X)|^2] = 4$。

5.1.1 一阶算法

根据定理 2 和式 (13)，要达到 Wasserstein-2 距离下的 ϵ 精度，迭代复杂度的阶为 $\tilde{O}(1/\epsilon^2)$。采用相同的 V 和 β 选择后，我们检验 Li 等人（2019）中定理 1 的条件。扩散过程 (3) 是 α'-一致耗散的，其中 $\alpha' = d$；而式 (9) 给出的欧拉离散化的局部偏差阶为 $(p_1, p_2) = (1, 3/2)$ 和 $(\lambda_1, \lambda_2) = (\Theta(d^5)\ \Theta(d^4))$。推导上述常数的详细计算见附录 B。因此，根据 Li 等人（2019）中的定理 1，要达到 Wasserstein-2 距离下的 ϵ 精度，迭代复杂度的阶为 $\tilde{O}(d^3/\epsilon^2)$。因此，与 Li 等人（2019）的结果相比，我们获得了与维数无关的迭代复杂度，从而确保 Wasserstein-2 距离下达到 ϵ 精度。

5.1.2 零阶算法

根据定理3 和（27），要使 Wasserstein-2 距离达到 ε精度，迭代复杂度的阶为 $\tilde{O}\left(\left(1 \vee d/m\right)/\varepsilon^2\right)$。当 $m = 1$时，迭代复杂度 $K \sim \tilde{O}\left(d/\varepsilon^2\right)$ 和函数评估次数 mK 也具有相同的阶 $\tilde{O}\left(d/\varepsilon^2\right)$。若选择批量大小 $m = d$，则可获得与维度无关的迭代复杂度 $K \sim \tilde{O}\left(1/\varepsilon^2\right)$，但函数评估次数的阶为 $\tilde{O}\left(d/\varepsilon^2\right)$。因此，我们注意到，对于自由度较大的多元 t-分布，估计梯度的成本会影响采样复杂度。

5.2 多元 t-分布：小自由度

现在考虑自由度为 3 的各向同性多元 t-分布。我们用 π_β 表示相应的密度函数。选择 3 作为自由度仅仅是为了方便；本例的结果适用于自由度严格大于 2 的所有情形，这对应于方差有限的情况。我们取 $V(x) = 1 + |x|^2$，$\beta = (d+3)/2$ 和$\pi_\beta(x) \propto V(x)^{-\beta} = (1+|x|^2)^{-(d+3)/2}$。根据上述选择，$V$和 β，V满足假设 2，其中$\alpha = 2$，$C_V = 2$，且 V是具有 $L = 2$的 L-Lipschitz 梯度。因此，定理 2 中的常数 δ 为$\delta = 1/d$。根据命题 2，$\mathbb{E}_{\pi_\beta}[V(X)] = d+1$和 $\mathbb{E}_{\pi_\beta}[|\nabla V(X)|^2] = 4d$。

5.2.1 一阶算法

根据定理 2 和 (13)，要达到 Wasserstein-2 距离下的 ϵ 精度，迭代复杂度为 $\tilde{O}(d^4/\epsilon^2)$ 量级。在采用相同的 V 和 β 选择时，我们检验 Li 等人（ 2019 ）中定理 1 的条件。扩散过程 (3) 是 α'- 一致耗散的，且具有 $\alpha' = 1$；式 (9) 给出的欧拉离散化的局部偏差阶为 $(p_1, p_2) = (1, 3/2)$ 和 $(\lambda_1, \lambda_2) = (\Theta(d^5)\ \Theta(d^4))$。推导这些常数的详细计算见附录 B。因此，根据 Li 等人（2019）中的定理 1，要达到 Wasserstein-2 距离下的 ϵ 精度，迭代复杂度为 $\tilde{O}(d^6/\epsilon^2)$ 量级。即使在这种极端重尾情形下（即仅方差存在），为确保 Wasserstein-2 距离下的 ϵ 精度，我们仍能得到维度依赖呈多项式增长的迭代复杂度。此外，与 Li 等人（2019）的结果相比，我们的分析有助于将维度指数降低一半。

5.2.2 零阶算法

根据定理 3 和（27），要在 Wasserstein-2 距离下达到 ε 精度，迭代复杂度的阶为 $\tilde{O}\left(\max\{d^4/\varepsilon^2,\ d^{\frac{5}{2}}/\varepsilon,\ d^4/\varepsilon^2 m\}\right)$。因此，对于任意批量大小 m，迭代复杂度
$K = \tilde{O}\ d^4/\varepsilon^2$ $m = 1$ $mK = ()$。取，函数评估次数 $\tilde{O}$
的阶相同，即（d^4/ε^2）。

注 8. 第 *5.2.2* 节讨论的例子揭示了以下重要观察：增大批量大小并不能改善迭代复杂度。为了解释这一点，我们将 (9) 和 (22)都视为连续动力学 (3)的近似。对于一阶算法，近似误差仅来自 Euler-Maruyama 离散化。对于零阶算法，近似误差则同时来自 Euler-Maruyama 离散化和零阶梯度估计=。当 Euler-Maruyama 离散化带来的误差占主导时，最优批量大小始终为 1 且

零阶算法的预言机复杂度与一阶算法的迭代复杂度相同。当零阶梯度估计带来的误差占主导时，我们需要根据 d 选择较大的批量大小，从而使零阶算法的迭代复杂度与一阶算法的迭代复杂度相同，同时零阶预言机复杂度会大出 m 倍。

6 关于假设的进一步结果与补充见解

在第 2 节中，我们给出了关于 V 的充分条件，使得当 $\beta > d$ 时， $\pi_\beta \propto V^{-\beta}$ 满足带权 Poincar´e 不等式，其权重为 V。在本节中，我们通过引入以下假设，放宽第 2 节中的条件。

假设 3。函数 $V\colon \mathbb{R}^d \to (0,\infty)$ 二阶连续可微，并且 V 满足

(1) 对所有 $x \in \mathbb{R}^d$， $\nabla^2 V(x)$ 均可逆。

(2) 存在 $\gamma \in \left(0, \frac{\beta}{d+2}\right]$，使得

$$\sup_{x\in\mathbb{R}^d} \left\| V(x)^{\gamma-1} \left(\nabla^2 V_\gamma\right)^{-1}(x) \right\|_2 \le C_V(\gamma),$$

其中 $V_\gamma := V^\gamma$，且 $C_V(\gamma)$ 是依赖于 γ 的正常数。

引理 2。 在假设 3下，对于任意光滑函数 $\phi \in L^2(\pi_\beta)$，

$$Var_{\pi_\beta}(\phi) \le C_{WPI} \int_{\mathbb{R}^d} |\nabla\phi(x)|^2 V(x)\pi_\beta(x)dx, \quad with \quad C_{WPI} = C_V(\gamma)\left(\frac{\beta}{\gamma}-1\right)^{-1}. \tag{28}$$

证明。首先我们定义 为 $V_\gamma := V^\gamma$。取 $\beta' = \beta - 2\gamma$。对于 $\pi_{\beta'} \propto V^{-\beta'}$，我们可以将其写成 $\pi_{\beta'} \propto {V_\gamma}^{-a}$ 其中

$$a = \frac{\beta'}{\gamma} = \frac{\beta-2\gamma}{\gamma} \ge d,$$

其中，不等式由以下事实得出： $\gamma \in \left(0, \frac{\beta}{d+2}\right]$。因此，我们可以将定理 1 应用于 $\pi_{\beta'} \propto {V_\gamma}^{-a}$，从而对于任意光滑的 $\pi_{\beta'}$平方可积函数 g，满足 $\mathbb{E}_{\pi_{\beta'}}[g(X)] = 0$ 以及 $G = V_\gamma g$,

$$(a+1)\int_{\mathbb{R}^d} g(x)^2\pi_{\beta'}(x)dx \le \int_{\mathbb{R}^d} \frac{\langle(\nabla^2 V_\gamma)^{-1}(x)\nabla G(x), \nabla G(x)\rangle}{V_\gamma(x)}\pi_{\beta'}(x)dx. \tag{29}$$

由于 $\beta' = \beta - 2\gamma$， （29）等价于

$$(a+1)\int_{\mathbb{R}^d} \frac{|G(x)|^2}{V(x)} V(x)^{-(\beta-1)}dx \le \int_{\mathbb{R}^d} \langle(\nabla^2 V_\gamma)^{-1}(x)\nabla G(x), \nabla G(x)\rangle V(x)^{-(\beta'+\gamma)}dx. \tag{30}$$

在假设3下，我们有

$$\begin{aligned}
&\int_{\mathbb{R}^d} \langle(\nabla^2 V_\gamma)^{-1}(x)\nabla G(x), \nabla G(x)\rangle V(x)^{-(\beta'+\gamma)}dx \\
&\le C_V(\gamma)\int_{\mathbb{R}^d} |\nabla G(x)|^2 V(x)^{1-\gamma} V(x)^{-(\beta'+\gamma)}dx \\
&= C_V(\gamma)\int_{\mathbb{R}^d} |\nabla G(x)|^2 V(x)^{-(\beta-1)}dx,
\end{aligned}$$

其中，最后一个恒等式由以下事实得出： $\beta' = \beta - 2\gamma$。结合式（30），可得

$$(a+1)\int_{\mathbb{R}^d} \frac{|G(x)|^2}{V(x)} V(x)^{-(\beta-1)}dx \le C_V(\gamma)\int_{\mathbb{R}^d} |\nabla G(x)|^2 V(x)^{-(\beta-1)}dx. \tag{31}$$

由于 $G = V^{\gamma}g$, G 光滑， π_β平方可积且 $\mathbb{E}_{\pi_{\beta-\gamma}}[G(X)] = 0$。对于任意 π_β平方可积的 ϕ，令 $G = \phi - \mathbb{E}_{\pi_{\beta-\gamma}}[\phi(X)]$，于是可得

$$\int_{\mathbb{R}^d} |\phi(x) - \mathbb{E}_{\pi_{\beta-\gamma}}[\phi(X)]|^2 \pi_\beta(x)dx \leq \frac{C_V(\gamma)}{a+1} \int_{\mathbb{R}^d} |\nabla\phi(x)|^2 V(x)\pi_\beta(x)dx. \tag{32}$$

因此，对于任意光滑且平方可积的 π_β ϕ,

$$Var_{\pi_\beta}(\phi) = \inf_{c\in\mathbb{R}} \int_{\mathbb{R}^d} |\phi(x) - c|^2 \pi_\beta(x)dx \leq \frac{C_V(\gamma)}{a+1} \int_{\mathbb{R}^d} |\nabla\phi(x)|^2 V(x)\pi_\beta(x)dx,$$

这等价于（28）中 $C_{\mathrm{WPI}} = {}^{C_V}{}_{(a+1}{}^{\gamma} = C_{V)}{}_{(\gamma)} \left(\frac{\beta}{\gamma} - 1\right)^{-1}$。 □

注 9。引理2可应用于满足 t-分布的多元 V（x）$= 1 + |x|^2$类。当$\beta \in (\frac{d+2}{2}, d]$取 $\gamma = \frac{\beta}{d+2}$时，假设3成立，且

$$C_V(\gamma) = \frac{(d+2)^2}{2\beta(2\beta - d - 2)}.$$

因此，引理2 推出，自由度为 t的多元分布 $\nu \in (2, d]$ 满足加权 Poincar ´é 不等式，权重为 $1 + |x|^2$，且

$$C_{WPI} = \frac{(d+2)^2}{\nu(d+1)(d+\nu)}.$$

详细　　推导上述常数的计算过程见附录C。

作为引理 2的直接推论，我们得到以下关于 (3) 的收敛结果。

命题 6。 在假设 3下，若 (X_t) 满足 (3)，且 ρ_t 是 X_t的分布，则有

$$\chi^2(\rho_t|\pi_\beta) \leq \exp\left(-C_V(\gamma)^{-1}\left(\frac{\beta}{\gamma} - 1\right)t\right)\chi^2(\rho_0|\pi_\beta). \tag{33}$$

对于多元 t-分布的情形，命题 6 使我们能够证明 (3) 在自由度更小（因而尾部更重）时的 χ^2 散度指数收敛，这与命题 1相比更为显著。

6.1 引理 1与引理 2之间的关系

引理 2 的结果补充了引理 1的结果。它可用于研究当 π_β 且 $\beta \leq d$时的加权庞加莱不等式（WPI）。特别地，当 $\beta \leq d$ 时，若 $\pi_\beta \propto V^{-\beta}$和 V满足假设 2，且 $C_V \in (0, \frac{d+2}{d+2-\beta})$，则 V 满足假设 3。因此 π_β 满足加权庞加莱不等式。在命题 7中，我们将正式证明这一关系。

命题　　**7.** 当 $\beta \leq d$时，若假设2 在 $C_V \in (0, \frac{d+2}{d+2-\beta})$下成立，则假设3 成立　　.

证明。首先， $\nabla^2 V$ 可逆，因为 $\nabla^2 V \succeq \alpha I_d$。接下来，我们证明存在 $\gamma \in (0, \frac{\beta}{d+2}]$，使得$\left\|V(x)^{\gamma-1}(\nabla^2 V_\gamma)^{-1}(x)\right\|_2 \leq C_V(\gamma)$对所有 $x \in \mathbb{R}^d$均成立。这等价于证明存在 $\gamma \in (0, \frac{\beta}{d+2}]$，使得 $\left\|V(x)^{1-\gamma}(\nabla^2 V_\gamma)(x)\right\|_2 > 0$ 对所有 $x \in \mathbb{R}^d$均成立。根据C节中的计算，我们有

$$\nabla^2 V_\gamma(x) = \gamma V(x)^{\gamma-1}\left((\gamma-1)V(x)^{-1}\nabla V(x)^T\nabla V(x) + \nabla^2 V(x)\right).$$

因此

$$\begin{aligned} V(x)^{1-\gamma}(\nabla^2 V_\gamma)(x) &= \gamma\left(\nabla^2 V(x) - (1-\gamma)V(x)^{-1}\nabla V(x)^T\nabla V(x)\right) \\ &\succeq \alpha\gamma\left(1 - (1-\gamma)C_V\right)I_d, \end{aligned}$$

其中不等式由假设 2 得出。最后我们证明存在 $\gamma \in (0, \frac{\beta}{d+2}]$ 使得 $1-(1-\gamma)C_V > 0$。注意到

$$1-(1-\gamma)C_V > 0 \implies \gamma > 1 - \frac{1}{C_V}.$$

由于 $C_V \in \left(0, \frac{d+2}{d+2-\beta}\right)$，我们有

$$1 - \frac{1}{C_V} < \frac{\beta}{d+2}$$

因此，存在常数 $\gamma \in \left(0, \frac{\beta}{d+2}\right]$，使得 $\left\|V(x)^{1-\gamma}(\nabla^2 V_\gamma)(x)\right\|_2 > 0$ 对所有 $x \in \mathbb{R}^d$ 都成立。

□

6.2 定理 2与命题 6 之间的关系

命题 6 研究连续动力学（3），而定理 2 研究离散化（9）的收敛。定理2 中的条件可以推出命题 6中的条件。在命题6 中，我们仅假设假设 3。在定理 2中，我们假设 (i) 假设 2，(ii)$\delta = \frac{\beta-1-\frac{1}{4}C_V d}{\frac{1}{4}C_V d} > 0$，以及 (iii) V满足梯度利普希茨条件。下一个命题将说明，这三个假设合在一起可以推出假设 3。

命题8。若假设2成立，且满足 $\delta = \frac{\beta-1-\frac{1}{4}C_V d}{\frac{1}{4}C_V d} > 0$和 V是 L-梯度利普希茨连续的，则假设3 成立。

命题6的证明。在假设2 及 L-梯度利普希茨连续性假设下，我们有 V 是“本质二次型”。也就是说，假设 V在 x^*处取得全局最小值，则对所有 $x \in \mathbb{R}^d$,

$$V(x^*) + \frac{\alpha}{2}|x-x^*|^2 \le V(x) \le V(x^*) + \frac{L}{2}|x-x^*|^2.$$

因此，对所有 $x \in \mathbb{R}^d$,

$$\frac{|\nabla V(x)|^2}{V(x)} \le \frac{L^2|x-x^*|^2}{V(x^*)+\frac{\alpha}{2}|x-x^*|^2} \le \frac{2L^2}{\alpha},$$

这意味着假设2-(2) 在 $C_V = \frac{2L^2}{\alpha^2}$ 下成立。此外,

$$V(x)^{1-\gamma}(\nabla^2 V_\gamma)(x) \succeq \alpha\gamma\left(1-(1-\gamma)C_V\right) I_d = \alpha\gamma\left(1-2(1-\gamma)\frac{L^2}{\alpha^2}\right) I_d.$$

条件 $\delta = \frac{\beta-1-\frac{1}{4}C_V d}{\frac{1}{4}C_V d} > 0$等价于条件 $\beta > \frac{L^2}{2\alpha^2}d + 1$。注意，对所有 $d \ge 1$，我们有

$$\left(1 - \frac{\alpha^2}{2L^2}\right)(d+2) < \frac{L^2}{2\alpha^2}d + 1$$

因此，对任意

$$\beta > \frac{L^2}{2\alpha^2}d + 1 > \left(1 - \frac{\alpha^2}{2L^2}\right)(d+2),$$

我们可以选择 $\gamma = \frac{\beta}{d+2}$ 并得到

$$\begin{aligned} V(x)^{1-\gamma}(\nabla^2 V_\gamma)(x) &\succeq \frac{2L^2\beta}{\alpha(d+2)}\left(\frac{\alpha^2}{2L^2} + \frac{\beta}{d+2} - 1\right) I_d \\ &= \frac{2L^2\beta}{\alpha(d+2)^2}\left(\beta - \left(1 - \frac{\alpha^2}{2L^2}\right)(d+2)\right) I_d \end{aligned}$$

因此，假设3-(2) 在 $\gamma = \beta/(d+2)$ 下成立，并且

$$C_V(\gamma) = \frac{\alpha(d+2)^2}{2L^2\beta}\left(\beta - \left(1 - \frac{\alpha^2}{2L^2}\right)(d+2)\right)^{-1} > 0.$$

由于假设 3-(1) 在假设 2 下自动成立，因此证明至此完成。

□

7 主要结果的证明

7.1 定理 2 和定理 3 的证明

本节通过均方分析给出定理 2 和定理 3 的证明。我们先给出以下中间结果。

命题 9。 设 $(X_t)_{t\geq 0}$ 满足 (3)，其中 $X_t \sim \rho_t$ 对所有 $t \geq 0$均成立。如果 V 是梯度利普希茨的，参数为L，那么有

$$\begin{aligned}\mathbb{E}\left[|X_t - X_0|^2\right] \leq 4\left[(\beta-1)^2t^2\mathbb{E}\left[|\nabla V(X_0)|^2\right] + td\mathbb{E}\left[V(X_0)\right]\right] \\ \exp\left(4(\beta-1)^2L^2t^2 + d(\beta-1)L^2t^2 + 2dLt\right).\end{aligned} \tag{34}$$

命题 *9* 的证明。 根据（3），

$$\mathbb{E}[|X_t - X_0|^2] \leq 2(\beta-1)^2\mathbb{E}\left[\left|\int_0^t \nabla V(X_s)ds\right|^2\right] + 4d\mathbb{E}\left[\int_0^t V(X_s)ds\right],$$

其中

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^t \nabla V(X_s)ds\right|^2\right] &\leq 2\mathbb{E}\left[\left(\int_0^t |\nabla V(X_s) - \nabla V(X_0)|ds\right)^2\right] + 2\mathbb{E}\left[\left(\int_0^t |\nabla V(X_0)|ds\right)^2\right] \\ &\leq 2t\mathbb{E}\left[\int_0^t |\nabla V(X_s) - \nabla V(X_0)|^2 ds\right] + 2t\mathbb{E}\left[\int_0^t |\nabla V(X_0)|^2 ds\right] \\ &\leq 2L^2t\int_0^t \mathbb{E}\left[|X_s - X_0|^2\right]ds + 2t^2\mathbb{E}\left[|\nabla V(X_0)|^2\right],\end{aligned} \tag{35}$$

and

$$\begin{aligned}&\mathbb{E}\left[\int_0^t V(X_s)ds\right] \\ &\leq \mathbb{E}\left[\int_0^t V(X_0) + \langle\nabla V(X_0), X_s - X_0\rangle + \frac{L}{2}|X_s - X_0|^2 ds\right] \\ &= t\mathbb{E}\left[V(X_0)\right] + \frac{L}{2}\mathbb{E}\left[\int_0^t |X_s - X_0|^2 ds\right] - (\beta-1)\mathbb{E}\left[\int_0^t\int_0^s \langle\nabla V(X_0), \nabla V(X_u)\rangle du ds\right] \\ &\leq t\mathbb{E}\left[V(X_0)\right] + \frac{L}{2}\mathbb{E}\left[\int_0^t |X_s - X_0|^2 ds\right] - \frac{(\beta-1)t^2}{2}\mathbb{E}\left[|\nabla V(X_0)|^2\right] \\ &\quad - (\beta-1)\mathbb{E}\left[\int_0^t\int_0^s \langle\nabla V(X_0), \nabla V(X_u) - \nabla V(X_0)\rangle du ds\right] \\ &\leq t\mathbb{E}\left[V(X_0)\right] + \frac{L}{2}\mathbb{E}\left[\int_0^t |X_s - X_0|^2 ds\right] - \frac{(\beta-1)t^2}{2}\mathbb{E}\left[|\nabla V(X_0)|^2\right] \\ &\quad + \frac{(\beta-1)t^2}{2}\mathbb{E}\left[|\nabla V(X_0)|^2\right] + \frac{\beta-1}{4}\mathbb{E}\left[\int_0^t\int_0^s |\nabla V(X_u) - \nabla V(X_0)|^2 du ds\right]\end{aligned} \tag{36}$$

$$\leq t\mathbb{E}\left[V(X_0)\right] + \frac{L}{2}\mathbb{E}\left[\int_0^t |X_s - X_0|^2 ds\right] + \frac{(\beta-1)L^2}{4}\mathbb{E}\left[\int_0^t \int_0^s |X_u - X_0|^2 du ds\right]$$
$$\leq t\mathbb{E}\left[V(X_0)\right] + \left(\frac{L}{2} + \frac{(\beta-1)L^2 t}{4}\right)\mathbb{E}\left[\int_0^t |X_s - X_0|^2 ds\right]. \tag{37}$$

根据（35）和（36），可得

$$\mathbb{E}[|X_t - X_0|^2] \leq \int_0^t \left[4(\beta-1)^2 L^2 t + 2dL + d(\beta-1)L^2 t\right]\mathbb{E}\left[|X_s - X_0|^2\right] ds + 4dt\mathbb{E}\left[V(X_0)\right]$$
$$+ 4(\beta-1)^2 t^2 \mathbb{E}\left[|\nabla V(X_0)|^2\right].$$

由格朗沃尔不等式，我们因此得到

$$\mathbb{E}[|X_t - X_0|^2] \leq 4\left[(\beta-1)^2 t^2 \mathbb{E}\left[|\nabla V(X_0)|^2\right] + dt\mathbb{E}\left[V(X_0)\right]\right]$$
$$\exp\left(4(\beta-1)^2 L^2 t^2 + d(\beta-1)L^2 t^2 + 2dLt\right).$$

□

基于上述命题，我们现在证明定理2 如下。

定理 2的证明。 我们对 (9) 进行均方分析。令 $(X_t)_{t\geq 0}$ 服从 (3)，其中 $X_0 \sim \pi_\beta$。由于 π_β是 (3) 的唯一平稳分布， $X_t \sim \pi_\beta$ 对所有 $t \geq 0$均成立。根据 (9)，我们可以计算 X_h 与 x_1之间的差值，

$$X_h - x_1 = X_0 - \int_0^h (\beta-1)\nabla V(X_t)dt + \int_0^t \sqrt{2V(X_t)}dB_t - \left(x_0 - (\beta-1)hy_0 + \sqrt{2hV(x_0)}\xi_1\right)$$
$$= (X_0 - x_0) - (\beta-1)h\left(\nabla V(X_0) - \nabla V(x_0)\right) - \int_0^h (\beta-1)\left(\nabla V(X_t) - \nabla V(X_0)\right)dt$$
$$\int_0^h \left(\sqrt{2V(X_t)} - \sqrt{2V(x_0)}\right)dB_t$$
$$:= U_1 + U_2 + U_3,$$

其中

$$U_1 := (X_0 - x_0) - (\beta-1)h\left(\nabla V(X_0) - \nabla V(x_0)\right), \tag{38}$$
$$U_2 := -\int_0^h (\beta-1)\left(\nabla V(X_t) - \nabla V(X_0)\right)dt, \tag{39}$$
$$U_3 := \int_0^h \left(\sqrt{2V(X_t)} - \sqrt{2V(x_0)}\right)dB_t. \tag{40}$$

因此，根据三角不等式，

$$\mathbb{E}[|X_h - x_1|^2|]^{\frac{1}{2}} \leq \mathbb{E}[|U_1 + U_3|^2]^{\frac{1}{2}} + \mathbb{E}[|U_2|^2]^{\frac{1}{2}}.$$

由于 U_1 适应于 $\mathcal{F}_0$ 和 $\mathbb{E}[U_3|\mathcal{F}_0] = 0$，我们得到

$$\mathbb{E}[|U_1 + U_3|^2|\mathcal{F}_0] = |U_1|^2 + \mathbb{E}[|U_3|^2|\mathcal{F}_0]$$
$$= |(X_0 - x_0) - (\beta-1)h\left(\nabla V(X_0) - \nabla V(x_0)\right)|^2$$
$$+ \mathbb{E}\left[\int_0^h \left\|\sqrt{2V(X_t)}I_d - \sqrt{2V(x_0)}I_d\right\|_F^2 dt|\mathcal{F}_0\right].$$

由于 V 是 α-强凸且 L-梯度利普希茨，因此满足

$$\langle X_0 - x_0, \nabla V(X_0) - \nabla V(x_0)\rangle \geq \frac{\alpha L}{\alpha + L}|X_0 - x_0|^2 + \frac{1}{\alpha + L}|\nabla V(X_0) - \nabla V(x_0)|^2.$$

因此，当 $h \leq \frac{2}{(\beta-1)(\alpha+L)}$ 时，

$$
\begin{aligned}
&|(X_0 - x_0) - (\beta-1)h\left(\nabla V(X_0) - \nabla V(x_0)\right)|^2 \\
=&|X_0 - x_0|^2 - 2(\beta-1)h\langle X_0 - x_0, \nabla V(X_0) - \nabla V(x_0)\rangle + (\beta-1)^2h^2|\nabla V(X_0) - \nabla V(x_0)|^2 \\
\leq&\left(1 - \frac{2(\beta-1)\alpha L h}{\alpha+L}\right)|X_0 - x_0|^2 + (\beta-1)h\left((\beta-1)h - \frac{2}{\alpha+L}\right)|\nabla V(X_0) - \nabla V(x_0)|^2 \\
\leq&\left(1-(\beta-1)\alpha h\right)^2|X_0 - x_0|^2. \qquad (41)
\end{aligned}
$$

同时，对于任意 $r>0$，我们有

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^h \left\|\sqrt{2V(X_t)} - \sqrt{2V(x_0)}\right\|_F^2 dt\right] \\
=&d\mathbb{E}\left[\int_0^h |\sqrt{2V(X_t)} - \sqrt{2V(x_0)}|^2 dt\right] \\
\leq&d\left(h\left(\sqrt{2V(X_0)} - \sqrt{2V(x_0)}\right)^2 + \mathbb{E}\left[\int_0^h \left|\sqrt{2V(X_t)} - \sqrt{2V(X_0)}\right|^2 dt\right]\right) \\
&+ 2d|\sqrt{2V(X_0)} - \sqrt{2V(x_0)}|h^{\frac{1}{2}}\mathbb{E}\left[\int_0^h \left|\sqrt{2V(X_t)} - \sqrt{2V(X_0)}\right|^2 dt\right] \\
\leq&d(1+r)h\left(\sqrt{2V(X_0)} - \sqrt{2V(x_0)}\right)^2 + d(1+r^{-1})\mathbb{E}\left[\int_0^h \left|\sqrt{2V(X_t)} - \sqrt{2V(X_0)}\right|^2 dt\right].
\end{aligned}
$$

注意，根据假设2，我们有

$$
|\nabla(\sqrt{2V(x)})| = \frac{\sqrt{2}|\nabla V(x)|}{2\sqrt{V(x)}} \leq \frac{\sqrt{2\alpha C_V}}{2},
$$

对所有 $x \in \mathbb{R}^d$ 均成立。因此

$$
(\sqrt{2V(X_0)} - \sqrt{2V(x_0)})^2 \leq \frac{\alpha C_V}{2}|X_0 - x_0|^2, \qquad (42)
$$

and

$$
\int_0^h |\sqrt{2V(X_t)} - \sqrt{2V(X_0)}|^2 dt \leq \frac{\alpha C_V}{2}\int_0^h |X_t - X_0|^2 dt. \qquad (43)
$$

由（42）和（43），可得

$$
\begin{aligned}
\mathbb{E}[\int_0^h \left\|\sqrt{2V(X_t)} - \sqrt{2V(x_0)}\right\|_F^2 dt] \leq& \frac{\alpha C_V dh(1+r)}{2}\mathbb{E}[|X_0 - x_0|^2] \\
&+ \frac{\alpha C_V d(1+r^{-1})}{2}\int_0^h \mathbb{E}[|X_t - X_0|^2]dt.
\end{aligned} \qquad (44)
$$

接下来，我们将命题 9 应用于 $\mathbb{E}[|X_t - X_0|^2]$。特别地，当

$$
t \in [0,h] \quad \text{and} \quad h < \frac{1}{4(\beta-1)L},
$$

我们有

$$
\begin{aligned}
\mathbb{E}[|X_t - X_0|^2] \leq& \left(4dt\mathbb{E}\left[V(X_0)\right] + 4(\beta-1)^2t^2\mathbb{E}\left[|\nabla V(X_0)|^2\right]\right)\exp(1) \\
\leq& 12dt\mathbb{E}\left[V(X_0)\right] + 12(\beta-1)^2t^2\mathbb{E}\left[|\nabla V(X_0)|^2\right]. \qquad (45)
\end{aligned}
$$

结合（44）和（45），当 $h < \frac{1}{4(\beta-1)L}$ 时，我们有

$$\begin{aligned}
&\mathbb{E}[\int_0^h \left\|\sqrt{2V(X_t)} - \sqrt{2V(x_0)}\right\|_F^2 dt] \\
\leq &\frac{1}{2}\alpha C_V d(1+r)h\mathbb{E}[|X_0 - x_0|^2] \qquad (46)\\
&+ 6\alpha C_V d(1+r^{-1})\int_0^h \left(dt\mathbb{E}\left[V(X_0)\right] + (\beta-1)^2 t^2 \mathbb{E}\left[|\nabla V(X_0)|^2\right]\right) dt \\
= &\frac{1}{2}\alpha C_V d(1+r)h\mathbb{E}[|X_0 - x_0|^2] \qquad (47)\\
&+ 3\alpha C_V d^2(1+r^{-1})h^2\mathbb{E}\left[V(X_0)\right] + 2\alpha C_V d(\beta-1)^2(1+r^{-1})h^3\mathbb{E}\left[|\nabla V(X_0)|^2\right].
\end{aligned}$$

由（41）和（46），可得

$$\begin{aligned}
&\mathbb{E}[|U_1 + U_3|^2] \\
\leq &\left(1 - 2(\beta-1)\alpha h + (\beta-1)^2\alpha^2 h^2 + \frac{1}{2}\alpha C_V d(1+r)h\right)\mathbb{E}\left[|X_0 - x_0|^2\right] \\
&+ 3\alpha C_V d^2(1+r^{-1})h^2\mathbb{E}\left[V(X_0)\right] + 2\alpha C_V d(\beta-1)^2(1+r^{-1})h^3\mathbb{E}\left[|\nabla V(X_0)|^2\right] \\
\leq &\left(1 - 2(\beta-1)\alpha h + (\beta-1)^2\alpha^2 h^2 + \frac{1}{2}\alpha C_V d(1+r)h\right)\mathbb{E}\left[|X_0 - x_0|^2\right] \\
&+ 2\alpha C_V d(1+r^{-1})h^2\left(3d\mathbb{E}\left[V(X_0)\right] + 2(\beta-1)^2 h\mathbb{E}\left[|\nabla V(X_0)|^2\right]\right). \qquad (48)
\end{aligned}$$

由于 $C_V < \frac{4}{(\beta-1)d}$，记作 $\delta = \frac{(\beta-1) - \frac{1}{4}C_V d}{\frac{1}{4}C_V d} > 0$。我们有

$$\begin{aligned}
&1 - 2(\beta-1)\alpha h + (\beta-1)^2\alpha^2 h^2 + \frac{1}{2}\alpha C_V d(1+r)h \\
= &1 - 2(\beta-1)\alpha h + (\beta-1)^2\alpha^2 h^2 + 2(\beta-1)\alpha\frac{1+r}{1+\delta}h \\
= &\left[1 - \alpha(\beta-1)(1 - \frac{1+2r}{1+\delta})h\right]^2 + \alpha^2(\beta-1)^2 h^2 \\
&- 2\alpha(\beta-1)\frac{r}{1+\delta}h - \alpha^2(\beta-1)^2 h^2\left(\frac{\delta - 2r}{1+\delta}\right)^2.
\end{aligned}$$

通过选取 $r = \frac{\delta}{3}$，对于任意 $h \in \left(0, \frac{2\delta}{\alpha(1+\delta)(\beta-1)}\right]$，有

$$\begin{aligned}
&1 - 2(\beta-1)\alpha h + (\beta-1)^2\alpha^2 h^2 + \frac{1}{2}\alpha C_V d(1+r)h \\
\leq &\left[1 - \alpha(\beta-1)\frac{\delta}{3(1+\delta)}h\right]^2 + \alpha^2(\beta-1)^2 h\left(h - \frac{2\delta}{3(1+\delta)}\alpha^{-1}(\beta-1)^{-1}\right) \\
\leq &\left[1 - \alpha(\beta-1)\frac{\delta}{3(1+\delta)}h\right]^2.
\end{aligned}$$

取 $r = \delta/3$，式（48）可改写为

$$\begin{aligned}
\mathbb{E}[|U_1 + U_3|^2] \leq &\left(1 - \frac{\alpha(\beta-1)\delta}{3(1+\delta)}h\right)^2\mathbb{E}[|X_0 - x_0|^2] \\
&+ \frac{8\alpha(\beta-1)(3+\delta)h^2}{(1+\delta)\delta}\left(3d\mathbb{E}\left[V(X_0)\right] + 2(\beta-1)^2 h\mathbb{E}\left[|\nabla V(X_0)|^2\right]\right). \qquad (49)
\end{aligned}$$

接下来，利用（45）中的界，当 $h < \frac{1}{4(\beta-1)L}$ 时，

$$\mathbb{E}[|U_2|^2] \leq (\beta-1)^2 L^2\mathbb{E}\left[\left(\int_0^h |X_t - X_0| dt\right)^2\right]$$

$$\leq (\beta-1)^2L^2h\int_0^h \mathbb{E}\left[|X_t-X_0|^2\right]dt$$
$$\leq 6d(\beta-1)^2L^2h^3\mathbb{E}\left[V(X_0)\right]+4(\beta-1)^4L^2h^4\mathbb{E}\left[|\nabla V(X_0)|^2\right]. \tag{50}$$

结合（49）和（50），当 $h<\min\left(\frac{1}{4(\beta-1)L},\ \frac{2\delta}{3(1+\delta)\alpha(\beta-1)}\right)$ 时，

$$\mathbb{E}\left[|X_h-x_1|^2\right]^{\frac{1}{2}}\leq\left[(1-A)^2\mathbb{E}\left[|X_0-x_0|^2\right]+B^2\right]^{\frac{1}{2}}+C,$$

with

$$A=\frac{\alpha(\beta-1)\delta}{3(1+\delta)}h, \tag{51}$$
$$B=\frac{4\alpha^{\frac{1}{2}}(\beta-1)^{\frac{1}{2}}(3+\delta)^{\frac{1}{2}}h}{(1+\delta)^{\frac{1}{2}}\delta^{\frac{1}{2}}}\left(d^{\frac{1}{2}}\mathbb{E}_{\pi_\beta}\left[V(X)\right]^{\frac{1}{2}}+(\beta-1)h^{\frac{1}{2}}\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}\right), \tag{52}$$
$$C=3d^{\frac{1}{2}}(\beta-1)Lh^{\frac{3}{2}}\mathbb{E}_{\pi_\beta}\left[V(X)\right]^{\frac{1}{2}}+2(\beta-1)^2Lh^2\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}}. \tag{53}$$

上述分析对每一步均成立，因此对于所有 $k\geq 1$，我们得到

$$\mathbb{E}\left[|X_{kh}-x_k|^2\right]^{\frac{1}{2}}\leq\left[(1-A)^2\mathbb{E}\left[|X_{(k-1)h}-x_{k-1}|^2\right]+B^2\right]^{\frac{1}{2}}+C.$$

根据（Dalalyan 和 Karagulyan，2019，引理 9），其中 A，B，C 由（51）、（52）、（53）给出，对于所有 $k\geq 1$，

$$\mathbb{E}\left[|X_{kh}-x_k|^2\right]^{\frac{1}{2}}\leq(1-A)^k\mathbb{E}\left[|X_0-x_0|^2\right]^{\frac{1}{2}}+\frac{C}{A}+\frac{B}{\sqrt{A(2-A)}}.$$

取 X_0 使得 $W_2(\nu_0,\ \pi_\beta)\ =\mathbb{E}\left[|X_0-x_0|^2\right]^{\frac{1}{2}}$，可得（11）。

□

现在我们证明定理 3。

定理 *3* 的证明。 沿用定理 2证明中的相同策略和记号，我们有

$$X_h-x_1=U_1+U_2+U_3+(\beta-1)h\mathbb{E}[\zeta_0|x_0]+(\beta-1)h\left(\zeta_0-\mathbb{E}[\zeta_0|x_0]\right), \tag{54}$$

其中 U_1、U_2、U_3 为 分别定义于（38）、（39）和（40）中，且 $\zeta_0=g_{\sigma,\ m}(x_0)\ -\nabla V(x_0)$。因此，我们有

$$\begin{aligned}\mathbb{E}\left[|X_h-x_1|^2\right]^{\frac{1}{2}}&\leq\mathbb{E}\left[|U_1+U_3+(\beta-1)h\left(\zeta_0-\mathbb{E}[\zeta_0|x_0]\right)|^2\right]^{\frac{1}{2}}\\&\quad+\mathbb{E}\left[|U_2|^2\right]^{\frac{1}{2}}+(\beta-1)h\mathbb{E}\left[|\mathbb{E}[\zeta_0|x_0]|^2\right]^{\frac{1}{2}}\\&=\left\{\mathbb{E}\left[|U_1+U_3|^2\right]+(\beta-1)^2h^2\mathbb{E}\left[|\zeta_0-\mathbb{E}[\zeta_0|x_0]|^2\right]\right\}^{\frac{1}{2}}\\&\quad+\mathbb{E}\left[|U_2|^2\right]^{\frac{1}{2}}+(\beta-1)h\mathbb{E}\left[|\mathbb{E}[\zeta_0|x_0]|^2\right]^{\frac{1}{2}}.\end{aligned} \tag{55}$$

根据定理2 和命题 5的证明，当

$$h<\min\left(\frac{1}{4(\beta-1)h},\frac{2\delta}{3(1+\delta)\alpha(\beta-1)}\right),$$

我们有

$$\begin{aligned}\mathbb{E}\left[|X_h-x_1|^2\right]^{\frac{1}{2}}\leq\Bigg\{&(1-A)^2\mathbb{E}\left[|X_0-x_0|^2\right]+B^2+\frac{\sigma^2}{2m}L^2(\beta-1)^2(d+3)^3h^2\\&+\frac{2(d+5)(\beta-1)^2h^2}{m}\mathbb{E}\left[|\nabla V(x_0)|^2\right]\Bigg\}^{\frac{1}{2}}+C+L\sigma(\beta-1)d^{\frac{1}{2}}h,\end{aligned} \tag{56}$$

其中 A, B, C 定义于（51）、（52）、（53）。利用 V 是梯度 Lipschitz 连续的事实，我们有 e

$$\begin{aligned}\mathbb{E}\left[|\nabla V(x_0)|^2\right] &\leq \mathbb{E}\left[(|\nabla V(X_0)| + L|X_0 - x_0|)^2\right] \\ &\leq 2\mathbb{E}\left[|\nabla V(X_0)|^2\right] + 2L^2\mathbb{E}\left[|X_0 - x_0|^2\right]. \end{aligned} \tag{57}$$

将（57）代入（56），得到

$$\begin{aligned}\mathbb{E}\left[|X_h - x_1|^2\right]^{\frac{1}{2}} \leq &\left\{(1-A)^2\mathbb{E}\left[|X_0 - x_0|^2\right] + \frac{4(d+5)(\beta-1)^2L^2h^2}{m}\mathbb{E}\left[|X_0 - x_0|^2\right] + B^2\right. \\ &\left. + \frac{\sigma^2}{2m}L^2(\beta-1)^2(d+3)^3h^2 + \frac{4(d+5)(\beta-1)^2h^2}{m}\mathbb{E}\left[|\nabla V(X_0)|^2\right]\right\}^{\frac{1}{2}} \\ &+ C + L\sigma(\beta-1)d^{\frac{1}{2}}h. \end{aligned} \tag{58}$$

当我们选择满足以下条件的步长时

$$h < \min\left\{\frac{2(1+\delta)}{\alpha(\beta-1)\delta}, \frac{\alpha m\delta}{24(1+\delta)(\beta-1)(d+5)L^2}\right\},$$

我们有

$$(1-A)^2 + \frac{4(d+5)(\beta-1)^2L^2h^2}{m} \leq \left(1 - \frac{A}{2}\right)^2.$$

因此我们有

$$\mathbb{E}\left[|X_h - x_1|^2\right]^{\frac{1}{2}} \leq \left\{(1-A')^2\mathbb{E}\left[|X_0 - x_0|^2\right] + B'^2\right\}^{\frac{1}{2}} + C', \tag{59}$$

其中

$$A' = \frac{\alpha(\beta-1)\delta}{6(1+\delta)}h, \tag{60}$$

$$\begin{aligned}B' = &\left(\frac{4\alpha^{\frac{1}{2}}(\beta-1)^{\frac{3}{2}}(3+\delta)^{\frac{1}{2}}h^{\frac{3}{2}}}{(1+\delta)^{\frac{1}{2}}\delta^{\frac{1}{2}}} + \frac{2(\beta-1)(d+5)^{\frac{1}{2}}h}{m^{\frac{1}{2}}}\right)\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}} \\ &+ \frac{4\alpha^{\frac{1}{2}}(\beta-1)^{\frac{1}{2}}d^{\frac{1}{2}}(3+\delta)^{\frac{1}{2}}h}{(1+\delta)^{\frac{1}{2}}\delta^{\frac{1}{2}}}\,\mathbb{E}_{\pi_\beta}\left[V(X)\right]^{\frac{1}{2}} + \frac{\sigma L(\beta-1)(d+3)^{\frac{3}{2}}}{m^{\frac{1}{2}}}h, \end{aligned} \tag{61}$$

$$C' = 3L(\beta-1)d^{\frac{1}{2}}h^{\frac{3}{2}}\mathbb{E}_{\pi_\beta}\left[V(X)\right]^{\frac{1}{2}} + 2L(\beta-1)^2h^2\mathbb{E}_{\pi_\beta}\left[|\nabla V(X)|^2\right]^{\frac{1}{2}} + \sigma L(\beta-1)d^{\frac{1}{2}}h. \tag{62}$$

证明的其余部分与定理 2 的证明相同，因此得到（26）。 □

致谢

YH 的部分研究经费来自 NSF TRIPODS资助项目 CCF-1934568。TF 的研究经费来自工程与物理科学研究委员会（EP/T5178）和 DeepMind奖学金。KB 的部分研究经费来自 NSF资助项目 DMS-2053918。MAE 的研究经费来自 NSERC 资助项目 [2019-06167],康诺特新研究者奖、CIFAR人工智能主席项目和 CIFAR人工智能催化剂资助项目。YH、KB 和 MAE 曾于 2021年秋季访问西蒙斯计算理论研究所，并参与“优化与采样中的几何方法”项目；本研究的部分工作是在此期间完成的。

参考文献

K. Ahn 和 S. Chewi。通过镜像朗之万算法实现高效约束采样。 神经信息处理系统进展，34:28405–28418, 2021。 (引自第 3页。)

C. Andrieu、N. De Freitas、A. Doucet 和 M. I. Jordan。面向机器学习的马尔可夫链蒙特卡罗简介。《机器学习》，50(1):5–43，2003。（引自第1页。）C. Andrieu、P. Dobson 和 A. Q. Wang。分段确定性马尔可夫过程蒙特卡罗方法的次几何次协调性。《电子概率期刊》，26，2021a。（引自第3页。）C. Andrieu、A. Lee、S. Power 和 A. Q. Wang。通过弱 Poincar´e 不等式比较马尔可夫链及其在伪边际马尔可夫链蒙特卡罗中的应用。arXiv 预印本 arXiv:2112.05605，2021b。（引自第3页。）D. Bakry、I. Gentil 和 M. Ledoux。《马尔可夫扩散算子的分析与几何》，第103卷。施普林格，2014。（引自第2页。）K. Balasubramanian 和 S. Ghadimi。零阶非凸随机优化：处理约束、高维性和鞍点。《计算数学基础》，22(1):35–76，2022年。（引自第13页。）K. Balasubramanian、S. Chewi、M. A. Erdogdu、A. Salim 和 S. Zhang。迈向非对数凹采样理论：朗之万蒙特卡罗的一阶平稳性保证。载于《学习理论会议》，第2896–2923页。PMLR，2022年。（引自第3页。）M.-F. F. Balcan 和 H. Zhang。在 s-凹分布下样本和计算效率俱佳的学习算法。神经信息处理系统进展，30，2017。（引自第1页。）J. Bierkens、G. Roberts 和 P.-A. Zitt。之字形过程的遍历性。《应用概率年刊》，29(4):2266–2301，2019。（引自第3页。）A. Blanchet、M. Bonforte、J. Dolbeault、G. Grillo 和 J. L. V´azquez。通过熵估计研究快速扩散方程的渐近行为。《理性力学与分析档案》，191(2):347–385，2009。（引自第2、3和4页。）S. Bobkov 和 M. Ledoux。适用于柯西测度及其他凸测度的加权 Poincar´e 型不等式。《概率年刊》，37(2):403–427，2009。（引自第2、3、4和5页。）M. Bonnefont、A. Joulin 和 Y. Ma。球对称对数凹概率测度的谱隙及其他情形。《泛函分析杂志》，270(7):2456–2482，2016。（引自第3页。）N. Bou-Rabee、A. Eberle 和 R. Zimmer。哈密顿蒙特卡罗的耦合与收敛。《应用概率年刊》，30(3):1209–1250，2020年。（引自第3页。）S. Brooks、A. Gelman、G. Jones 和 X.-L. Meng。《马尔可夫链蒙特卡罗手册》。CRC出版社，2011。（引自第1页。）Y. Cao、J. Lu 和 L. Wang。欠阻尼朗之万动力学随机化算法的复杂度。《数学科学通讯》，19(7):1827–1853，2021。（引自第3页。）P. Cattiaux、N. Gozlan、A. Guillin 和 C. Roberto。重尾分布的泛函不等式及其在等周性中的应用。《电子概率期刊》，15:346–385，2010。（引自第3页。）P. Cattiaux、A. Guillin 和 L.-M. Wu。关于加权对数索博列夫不等式的一些评述。《印第安纳大学数学杂志》，第1885–1904页，2011。（引自第3页。）P. Cattiaux、A. Guillin、P. Monmarch´e 和 C. Zhang。Langevin 扩散的熵乘子法与加权对数索博列夫不等式。《泛函分析杂志》，277(11):108288，2019。（引自第3页。）K. Chandrasekaran、A. Deshpande 和 S. Vempala。s-凹函数采样：基于凸性的等周性的极限。载于《近似、随机化与组合优化：算法与技术》，第420–433页。施普林格，2009。（引自第3页。）Y. Chen、R. Dwivedi、M. J. Wainwright 和 B. Yu。Metropolis化哈密顿蒙特卡罗的快速混合：多步梯度的优势。《机器学习研究杂志》，21:92–1，2020年。（引自第3页。）

Y. Chen、S. Chewi、A. Salim 和 A. Wibisono。用于采样的近端算法的改进分析。arXiv 预印本 arXiv:2202.06386，2022年。（引见于第 3 页。）Z. Chen 和 S. S. Vempala。强对数凹分布的哈密顿蒙特卡罗最优收敛速率。Theory of Computing，18(1):1–18，2022年。（引见于第 3 页。）S. Chewi、T. Le Gouic、C. Lu、T. Maunu、P. Rigollet 和 A. Stromme。镜像朗之万扩散的指数遍历性。神经信息处理系统进展，33:19573–19585，2020年。（引见于第 3 页。）S. Chewi、M. A. Erdogdu、M. B. Li、R. Shen 和 M. Zhang。从 Poincar´e 到对数 Sobolev 的朗之万蒙特卡罗分析。arXiv 预印本 arXiv:2112.12662，2021a。（引见于第 3 页。）S. Chewi、C. Lu、K. Ahn、X. Cheng、T. Le Gouic 和 P. Rigollet。Metropolis 调整的朗之万算法的最优维数依赖性。载于学习理论会议，第 1260–1300 页。PMLR，2021b。（引见于第 3 页。）D. Cordero-Erausquin 和 N. Gozlan。对数凹分布的加权 Poincar´e 不等式的输运证明。《伯努利》，23(1):134–158，2017年。（引见于第 3 页。）A. S. Dalalyan。光滑对数凹密度近似采样的理论保证。英国皇家统计学会杂志：B辑，79(3):651–676，2017年。（引见于第 3 页。）A. S. Dalalyan 和 A. Karagulyan。针对梯度不准确的朗之万蒙特卡罗算法的易用保证。《随机过程及其应用》，129(12):5278–5311，2019年。（引见于第 3 和 23 页。）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。《伯努利》，26(3):1956–1988，2020年。（引见于第 3 页。）A. S. Dalalyan、A. Karagulyan 和 L. Riou-Durand。非强对数凹目标分布的离散化朗之万算法误差界定。arXiv 预印本 arXiv:1906.08530，2019年。（引见于第 3 页。）G. Deligiannidis、A. Bouchard-Cˆot´e 和 A. Doucet。弹跳粒子采样器的指数遍历性。统计年鉴，47(3):1268–1287，2019年。（引见于第 3 页。）I. Diakonikolas、V. Kontonis、C. Tzamos 和 N. Zarifis。在结构化分布下学习带有 Massart噪声的半空间。载于学习理论会议，第 1486–1513 页。PMLR，2020年。（引见于第 1 页。）Z. Ding 和 Q. Li。朗之万蒙特卡罗：随机坐标下降与方差缩减。《机器学习研究杂志》，22:205–1，2021年。（引见于第 3 页。）R. Douc、E. Moulines、P. Priouret 和 P. Soulier。马尔可夫链。施普林格，2018年。（引见于第 1 页。）A. Durmus 和 E. Moulines。未校正朗之万算法的非渐近收敛分析。《应用概率年刊》，27(3):1551–1587，2017年。（引见于第 3 页。）A. Durmus、S. Majewski 和 B. Miasojedow。通过凸优化分析朗之万蒙特卡罗。《机器学习研究杂志》，20(1):2666–2711，2019年。（引见于第 3 页。）A. Durmus、A. Guillin 和 P. Monmarch´e。弹跳粒子采样器的几何遍历性。《应用概率年刊》，30(5):2069–2098，2020年。（引见于第 3 页。）R. Dwivedi、Y. Chen、M. J. Wainwright 和 B. Yu。对数凹采样：梅特ropolis-黑斯廷斯算法速度很快。机器学习研究期刊，20:1–42，2019年。（引见于第 3 和 13 页。）M. A. Erdogdu 和 R. Hosseinzadeh。论朗之万蒙特卡罗的收敛：尾部增长与光滑性的相互作用。载于学习理论会议，第 1776–1822 页。PMLR，2021年。（引见于第 3 页。）

A. Gelman、A. Jakulin、M. G. Pittau 和 Y.-S. Su。用于逻辑回归及其他回归模型的弱信息默认先验分布。《应用统计年鉴》，2(4):1360–1383，2008年。（引见于第 1 页。）A. Genz 和 F. Bretz。多元正态分布和 t概率的计算，第 195 卷。施普林格科学与商业媒体，2009年。（引见于第 1 页。）A. Genz、F. Bretz 和 Y. Hochberg。多元 t 积分的近似及其在多重比较程序中的应用。载于《多重比较程序新进展》，第 24–32 页。数学统计学会，2004年。（引见于第 1 页。）J. Ghosh、Y. Li 和 R. Mitra。柯西先验分布在贝叶斯逻辑回归中的应用。《贝叶斯分析》，13(2):359–383，2018年。（引见于第 1 页。）J. Gorham、A. B. Duncan、S. J. Vollmer 和 L. Mackey。利用扩散过程衡量样本质量。《应用概率年刊》，29(5):2884–2928，2019年。（引见于第 7 页。）E. Hairer、C. Lubich 和 G. Wanner。几何数值积分：常微分方程的保结构算法，第 31 卷。施普林格科学与商业媒体，2006年。（引见于第 1 页。）Y. He、K. Balasubramanian 和 M. A. Erdogdu。关于随机中点采样方法的遍历性、偏差与渐近正态性。神经信息处理系统进展，33，2020年。（引见于第 3 页。）Y. He、K. Balasubramanian 和 M. A. Erdogdu。通过变换非校正朗之万算法进行重尾采样。arXiv 预印本 arXiv:2201.08349，2022年。（引见于第 3 页。）Y.-P. Hsieh、A. Kavis、P. Rolland 和 V. Cevher。镜像朗之万动力学。神经信息处理系统进展，31，2018年。（引见于第 3 页。）L.-J. Huang、M. B. Majka 和 J. Wang。通过稳定过程驱动的随机微分方程近似重尾分布。《伯努利》，27(3):2040–2068，2021年。（引见于第 3 页。）M. Ismail 和 M. E. Muldoon。伽马函数和 q-伽马函数的不等式与单调性性质。载于《近似与计算：纪念 Walter Gautschi 文集》，第 309–323 页。施普林格，1994年。（引见于第 13 页。）S. Jarner 和 G. Roberts。重尾蒙特卡罗马尔可夫链算法的收敛性。《斯堪的纳维亚统计学杂志》，34(4):781–815，2007年。（引见于第 1 页。）Q. Jiang。镜像朗之万蒙特卡罗：等周条件下的情形。神经信息处理系统进展，34:715–725，2021年。（引见于第 3 页。）L. T. Johnson 和 C. J. Geyer。通过变量变换在随机游走Metropolis算法中实现几何遍历性。统计年鉴，40(6):3050–3076，2012年。（引见于第 3 页。）K. Kamatani。高维重尾目标概率分布下马尔可夫链蒙特卡罗的高效策略。《伯努利》，24(4B):3711–3750，2018年。（引见于第 1 和 3 页。）S. Kotz 和 S. Nadarajah。多元 t 分布及其应用。剑桥大学出版社，2004年。（引见于第 1 页。）Y. T. Lee、R. Shen 和 K. Tian。对数光滑梯度集中与更紧的 Metropolis化哈密顿蒙特卡罗运行时间。载于学习理论会议，第 2565–2597 页，2020年。（引见于第 3 页。）Y. T. Lee、R. Shen 和 K. Tian。使用受限高斯预言机进行结构化对数凹采样。载于学习理论会议，第 2993–3050 页。PMLR，2021年。（引见于第 1、3 和 13 页。）B. Leimkuhler 和 C. Matthews。分子动力学：确定性与随机数值方法。施普林格，2016年。（引见于第 1 页。）

M. B. Li 和 M. A. Erdogdu。用于求解半定规划的黎曼朗之万算法。arXiv 预印本 arXiv:2010.11176，2020年。（引自第3页。）R. Li、M. Tao、S. S. Vempala 和 A. Wibisono。镜像朗之万算法以渐消偏差收敛。载于《国际算法学习理论会议》，第718–742页。PMLR，2022年。（引自第3页。）X. Li、Y. Wu、L. Mackey 和 M. A. Erdogdu。随机龙格–库塔加速朗之万蒙特卡罗及其他方法。载于《神经信息处理系统进展》，第32卷，2019。（引自第3、15和31页。）J. Lu 和 L. Wang。强对数凹分布的锯齿采样算法的复杂度。《统计与计算》，32(3):1–12，2022年。（引自第3页。）Y.-A. Ma、N. S. Chatterji、X. Cheng、N. Flammarion、P. L. Bartlett 和 M. I. Jordan。基于梯度的马尔可夫链蒙特卡罗是否存在类似 Nesterov 加速的方法？《伯努利》，27(3):1942–1992，2021。（引自第3页。）S. P. Meyn 和 R. L. Tweedie。马尔可夫链与随机稳定性。施普林格科学与商业媒体，2012年。（引自第1页。）G. N. Milstein 和 M. V. Tretyakov。数学物理中的随机数值方法，第456卷。施普林格，2004年。（引自第2页。）P. Monmarch´e。采用标准分裂格式的欠阻尼朗之万扩散高维马尔可夫链蒙特卡罗。《电子统计学期刊》，15(2):4117–4166，2021。（引自第3页。）Y. Nesterov 和 V. Spokoiny。凸函数的随机无梯度最小化。《计算数学基础》，17(2):527–566，2017。（引自第13页。）T. H. Nguyen、U. S¸im¸sekli 和 G. Richard。非凸优化中分数阶朗之万蒙特卡罗的非渐近分析。载于《国际机器学习会议》，第4810–4819页，2019。（引自第1页。）C. Robert 和 G. Casella。蒙特卡罗统计方法，第2卷。施普林格，1999年。（引自第1页。）G. Roberts 和 J. Rosenthal。朗之万扩散离散近似的最优缩放。《皇家统计学会杂志：B辑（统计方法学）》，60(1):255–268，1998年。（引自第1页。）G. Roberts 和 R. Tweedie。朗之万分布及其离散近似的指数收敛。《伯努利》，第341–363页，1996年。（引自第1和2页。）P. J. Rossky、J. D. Doll 和 H. L. Friedman。布朗动力学作为智能蒙特卡罗模拟。《化学物理学杂志》，69(10):4628–4633，1978年。（引自第1页。）M. Roth。关于多元 t 分布。Link¨oping University Electronic Press，2012年。（引自第9页。）A. Roy、L. Shen、K. Balasubramanian 和 S. Ghadimi。用于贝叶斯推断的朗之万扩散随机零阶离散化。《伯努利》，28(3):1810–1834，2022年。（引自第13页。）R. Shen 和 Y. T. Lee。用于对数凹采样的随机化中点法。载于《第33届神经信息处理系统国际会议论文集》，第2100–2111页，2019。（引自第3页。）U. S¸im¸sekli、L. Zhu、Y. W. Teh 和 M. Gurbuzbalaban。分数阶欠阻尼朗之万动力学：在重尾梯度噪声下通过动量重新定位随机梯度下降。载于《国际机器学习会议》，第8970–8980页，2020年。（引自第1和3页。）M. K. Titsias 和 O. Papaspiliopoulos。辅助型基于梯度的采样算法。《皇家统计学会杂志：B辑（统计方法学）》，80(4):749–767，2018年。（引自第1页。）S. Vempala 和 A. Wibisono。未校正朗之万算法的快速收敛：等周性已足够。载于《神经信息处理系统进展》，第32卷，2019。（引自第3页。）

M. Vono、D. Paulin 和 A. Doucet。使用 ADMM 型分裂实现具有无维数依赖收敛速率的高效 MCMC 采样。机器学习研究期刊，23:1–69，2022年。（引自第1页。）F.-Y. Wang。泛函不等式、马尔可夫半群与谱理论。爱思唯尔，2006。（引自第2页。）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。arXiv 预印本 arXiv:2207.02189，2022。（引自第3页。）K. Wu、S. Schmidler 和 Y. Chen。对数凹采样的 Metropolis 调整的朗之万算法的极小极大混合时间。机器学习研究期刊，23(270):1–63，2022年。（引自第3页。）J. Yang、K. Latuszy´nski 和 G. Roberts。立体投影马尔可夫链蒙特卡罗。arXiv 预印本 arXiv:2205.12112，2022。（引自第3页。）K. S. Zhang、G. Peyr´e、J. Fadili 和 M. Pereyra。镜像朗之万蒙特卡罗的 Wasserstein 控制。载于学习理论会议，第3814–3841页。PMLR，2020年。（引自第3页。）

X. Zhang 和 X. Zhang。由 α-稳定过程驱动的超临界随机微分方程的遍历性与重尾采样。arXiv 预印本 arXiv:2201.10158，2022年。（引文见第3页。）

A 第3.3节的计算

引理 3。设 $\beta > \frac{d}{2}+1$。若 $V \in C^2(\mathbb{R}^d)$ 是正的、 α-强凸且 L-梯度利普希茨，则对于任意 $r \in (0, \beta - \frac{d}{2} - 1)$，

$$\frac{Z_{\beta-1}}{Z_\beta} \le \left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}} V(0) \left(\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}}. \tag{63}$$

证明。由于 $V(x) \le V(0) + \frac{L}{2}|x|^2$，我们知道对于任意 $r \in (0, \beta - \frac{d}{2} - 1)$， $Z_{\frac{d}{2}+r}$ 有限且 $\pi_{\frac{d}{2}+r}$ 是一个概率测度。因此

$$\begin{aligned}
\frac{Z_{\beta-1}}{Z_\beta} &= \frac{\int_{\mathbb{R}^d} V(x)^{-(\beta-1)} dx}{\int_{\mathbb{R}^d} V(x)^{-\beta} dx} \\
&= \frac{Z_{\frac{d}{2}+r} \int_{\mathbb{R}^d} V(x)^{-(\beta-\frac{d}{2}-1-r)} \pi_{\frac{d}{2}+r}(x) dx}{Z_{\frac{d}{2}+r} \int_{\mathbb{R}^d} V(x)^{-(\beta-\frac{d}{2}-r)} \pi_{\frac{d}{2}+r}(x) dx} \\
&\le \frac{\left(\int_{\mathbb{R}^d} V(x)^{-(\beta-\frac{d}{2}-r)} \pi_{\frac{d}{2}+r}(x) dx\right)^{\frac{\beta-\frac{d}{2}-1-r}{\beta-\frac{d}{2}-r}}}{\int_{\mathbb{R}^d} V(x)^{-(\beta-\frac{d}{2}-r)} \pi_{\frac{d}{2}+r}(x) dx} \\
&= \left(\int_{\mathbb{R}^d} V(x)^{-(\beta-\frac{d}{2}-r)} \pi_{\frac{d}{2}+r}(x) dx\right)^{-\frac{1}{\beta-\frac{d}{2}-r}} \\
&= \left(Z_{\frac{d}{2}+r}\right)^{\frac{1}{\beta-\frac{d}{2}-r}} \left(\int_{\mathbb{R}^d} V(x)^{-\beta} dx\right)^{-\frac{1}{\beta-\frac{d}{2}-r}} \\
&\le \left(Z_{\frac{d}{2}+r}\right)^{\frac{1}{\beta-\frac{d}{2}-r}} \left(\int_{\mathbb{R}^d} (V(0) + \frac{L}{2}|x|^2)^{-\beta} dx\right)^{-\frac{1}{\beta-\frac{d}{2}-r}}.
\end{aligned}$$

对于积分 $\int_{\mathbb{R}^d}(V(0)+\frac{L}{2}|x|^2)^{-\beta}dx$，我们可以通过变换到极坐标并进行代换来计算，

$$\begin{aligned}
\int_{\mathbb{R}^d} (V(0) + \frac{L}{2}|x|^2)^{-\beta} dx &= A_{d-1}(1) \int_0^\infty (V(0) + \frac{L}{2}R^2)^{-\beta} R^{d-1} dR \\
&= \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \int_0^\infty (V(0) + V(0)R_L)^{-\beta} (\frac{2V(0)}{L})^{\frac{d}{2}-1} R_L{}^{\frac{d}{2}-1} \frac{2V(0)}{L} dR_L
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2^{\frac{d}{2}}\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})L^{\frac{d}{2}}V(0)^{\beta-\frac{d}{2}}}\int_0^{\infty}(1+R_L)^{-\beta}R_L^{\frac{d}{2}-1}dR_L \\
&= \frac{2^{\frac{d}{2}}\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})L^{\frac{d}{2}}V(0)^{\beta-\frac{d}{2}}}\int_0^{1}u^{\frac{d}{2}-1}(1-u)^{\beta-\frac{d}{2}-1}du \\
&= \frac{2^{\frac{d}{2}}\pi^{\frac{d}{2}}B(\frac{d}{2},\beta-\frac{d}{2})}{\Gamma(\frac{d}{2})L^{\frac{d}{2}}V(0)^{\beta-\frac{d}{2}}},
\end{aligned}
$$

其中，第二个恒等式是通过代换 $R_L = LR^2/(2V(0))$ 得出的，第四个恒等式是通过代换 $u = \frac{R_L}{1+R_L}$ 得出的。对于 $Z_{\frac{d}{2}+r}$，我们有

$$
\begin{aligned}
Z_{\frac{d}{2}+r} &= \int_{\mathbb{R}^d} V(x)^{-\frac{d}{2}-r}dx \\
&\le \int_{\mathbb{R}^d}\left(V(0)+\frac{\alpha}{2}|x|^2\right)^{-\frac{d}{2}-r}dx \\
&= \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}\int_0^{\infty}\left(V(0)+\frac{\alpha}{2}R^2\right)^{-\frac{d}{2}-r}R^{d-1}dR \\
&= \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}\int_0^{\infty}(V(0)+V(0)R_\alpha)^{-\frac{d}{2}-r}\left(\frac{2V(0)}{\alpha}\right)^{\frac{d}{2}-1}R_\alpha{}^{\frac{d}{2}-1}\frac{2V(0)}{\alpha}dR_\alpha \\
&= \frac{2^{\frac{d}{2}}\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\alpha^{\frac{d}{2}}V(0)^r}\int_0^{\infty}(1+R_\alpha)^{-\frac{d}{2}-r}R_\alpha{}^{\frac{d}{2}-1}dR_\alpha \\
&= \frac{2^{\frac{d}{2}}\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\alpha^{\frac{d}{2}}}\int_0^{1}u^{\frac{d}{2}-1}(1-u)^{r-1}du \\
&= \frac{2^{\frac{d}{2}}\pi^{\frac{d}{2}}B(\frac{d}{2},r)}{\Gamma(\frac{d}{2})\alpha^{\frac{d}{2}}V(0)^r}.
\end{aligned}
$$

因此，我们可以进一步得到

$$
\begin{aligned}
\frac{Z_{\beta-1}}{Z_\beta} &\le \left(\frac{2^{\frac{d}{2}}\pi^{\frac{d}{2}}B(\frac{d}{2},r)}{\Gamma(\frac{d}{2})\alpha^{\frac{d}{2}}V(0)^r}\frac{\Gamma(\frac{d}{2})L^{\frac{d}{2}}V(0)^{\beta-\frac{d}{2}}}{2^{\frac{d}{2}}\pi^{\frac{d}{2}}B(\frac{d}{2},\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}} \\
&= \left(\frac{L^{\frac{d}{2}}V(0)^{\beta-\frac{d}{2}-r}}{\alpha^{\frac{d}{2}}}\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}} \\
&= \left(\frac{L}{\alpha}\right)^{\frac{\frac{d}{2}}{\beta-\frac{d}{2}-r}}V(0)\left(\frac{\Gamma(\beta)\Gamma(r)}{\Gamma(\frac{d}{2}+r)\Gamma(\beta-\frac{d}{2})}\right)^{\frac{1}{\beta-\frac{d}{2}-r}}.
\end{aligned}
$$

□

B 第 5 节的计算 5.1和 5.2

设 $\pi_\beta(x) \propto V(x)^{-\beta} = (1+|x|^2)^{-\beta}$ 满足 $\beta > \frac{d+2}{2}$。 V的梯度和海森矩阵为

$$
\nabla V(x) = 2x, \qquad \nabla^2 V(x) = 2I_d.
$$

因此 V 是 α-强凸，且 $\alpha = 2$ 且 L-梯度利普希茨，且 $L = 2$。（3）可化简为

$$
dX_t = b(x)dt + \sigma(X_t)dB_t, \tag{64}
$$

其中 $b(x) = -2(\beta-1)x$ 以及 $\sigma(x) = \sqrt{2}(1+|x|^2)^{\frac{1}{2}}I_d$。

接下来，我们考察一致耗散条件：

$$\begin{aligned}
&\langle b(x)-b(y),x-y\rangle+\frac{1}{2}\left\|(1+|x|^2)^{\frac{1}{2}}I_d-(1+|y|^2)^{\frac{1}{2}}I_d\right\|_F^2\\
=&-2(\beta-1)|x-y|^2+d|(1+|x|^2)^{\frac{1}{2}}-(1+|y|^2)^{\frac{1}{2}}|^2\\
\leq&-2(\beta-1-\frac{d}{2})|x-y|^2,
\end{aligned}\tag{65}$$

其中，不等式成立是因为 $x\mapsto(1+|x|^2)^{\frac{1}{2}}$ 是 1-Lipschitz。因此，扩散过程 (64) 是α'-一致耗散的，且 $\alpha'=2(\beta-1-\frac{d}{2})$。特别地，当 $\alpha'=d$时， $\beta=d+1$且当 $\alpha'=1$时， $\beta=\frac{d+3}{2}$。

最后，我们考察方程 (64) 的欧拉离散化的局部偏差。我们采用 Li 等人 (2019) 中的相同记号。根据 (Li 等人, 2019，引理 29)， $p_1=1$ 和

$$\lambda_1=2\left(\mu_1(b)^2+\mu_1^F(\sigma)^2\right)\left(\pi_{1,2}(b)+\pi_{1,2}^F(\sigma)\right)(1+\mathbb{E}[|\tilde{X}_0|^2]+2\pi_{1,2}(b)\alpha'^{-1}).$$

根据 (Li 等人, 2019，引理 29)， $p_2=\frac{3}{2}$ 和

$$\lambda_2=\mu_1(b)\left(\pi_{1,2}(b)+\pi_{1,2}^F(\sigma)\right)(1+\mathbb{E}[|\tilde{X}_0|^2]+2\pi_{1,2}(b)\alpha'^{-1}),$$

with

$$\begin{aligned}
\mu_1(b)&:=\sup_{x,y\in\mathbb{R}^d,x\neq y}\frac{|b(x)-b(y)|}{|x-y|}=2(\beta-1),\\
\mu_1^F(\sigma)&:=\sup_{x,y\in\mathbb{R}^d,x\neq y}\frac{\|\sigma(x)-\sigma(y)\|_F}{|x-y|}=\sqrt{2d},\\
\pi_{1,2}(b)&:=\sup_{x\in\mathbb{R}^d}\frac{|b(x)|^2}{1+|x|^2}=4(\beta-1)^2,\\
\pi_{1,2}^F(\sigma)&:=\sup_{x\in\mathbb{R}^d}\frac{\|\sigma(x)\|_F^2}{1+|x|^2}=2d.
\end{aligned}$$

λ_1 和 λ_2 在维度参数 d 中的阶为：

$$\begin{aligned}
\lambda_1&=\Theta\left(\left((\beta-1)^2+d\right)\left((\beta-1)^2+2d\right)\left(1+(\beta-1)^2\alpha'^{-1}\right)\right),\\
\lambda_2&=\Theta\left((\beta-1)\left((\beta-1)^2+2d\right)\left(1+(\beta-1)^2\alpha'^{-1}\right)\right).
\end{aligned}$$

因此，我们有

- 当 $\beta=d+1$时， $(\lambda_1,\lambda_2)=(\Theta(d^5)\ \Theta(d^4))$,
- 当 $\beta=\frac{d+3}{2}$ 时， $(\lambda_1,\lambda_2)=(\Theta(d^5)\ \Theta(d^4))$。

C 的计算（注记 9

In t柯西类分布的例子， $V(x)=1+|x|^2$ 和 $V_\gamma:=V^\gamma$。当 $\gamma>\frac{1}{2}$ ，

$$\begin{aligned}
\nabla V_\gamma(x)&=\gamma V(x)^{\gamma-1}\nabla V(x)\\
\nabla^2V_\gamma(x)&=\gamma(\gamma-1)V(x)^{\gamma-2}\nabla V(x)^T\nabla V(x)+\gamma V(x)^{\gamma-1}\nabla^2V(x)\\
&=\gamma V(x)^{\gamma-1}\left((\gamma-1)V(x)^{-1}\nabla V(x)^T\nabla V(x)+\nabla^2V(x)\right).
\end{aligned}$$

代入 $V(x)=1+|x|^2$，得

$$\nabla V_\gamma(x)=2\gamma(1+|x|^2)^{\gamma-1}x$$

$$\begin{aligned}\nabla^2 V_\gamma(x) &= 2\gamma(1+|x|^2)^{\gamma-1}\left(I_d + 2(\gamma-1)\frac{|x|^2}{1+|x|^2}\frac{x^T x}{|x|^2}\right)\\ &= 2\gamma(1+|x|^2)^{\gamma-1}\left((I_d - \frac{x^T x}{|x|^2}) + \left(1-2(1-\gamma)\frac{|x|^2}{1+|x|^2}\right)\frac{x^T x}{|x|^2}\right),\end{aligned}$$

and

$$(\nabla^2 V_\gamma)^{-1}(x) = \frac{1}{2\gamma}(1+|x|^2)^{1-\gamma}\left((I_d - \frac{x^T x}{|x|^2}) + \frac{1+|x|^2}{1+(2\gamma-1)|x|^2}\frac{x^T x}{|x|^2}\right).$$

当 $\beta \in \left(\frac{d+2}{2}, d\right]$, $\gamma = \frac{\beta}{d+2} \in \left(\frac{1}{2}, 1\right]$时,

$$(\nabla^2 V_\gamma)^{-1}(x) \preceq \frac{1}{2\gamma(2\gamma-1)}(1+|x|^2)^{1-\gamma} I_d = \frac{(d+2)^2}{2\beta(2\beta-d-2)}(1+|x|^2)^{1-\gamma} I_d.$$

因此，假设3在 $C_V(\gamma) = \left(\frac{d+2}{2\beta}\right)^2 \left(\frac{2\beta}{2\beta-d-2}\right)$ 的条件下成立。对于柯西分布 $\pi_\beta \propto (1+|x|^2)^{-\beta} = (1+|x|^2)^{-\frac{d+\nu}{2}}$, 其中 $\beta \in (\frac{d+2}{2}, d]$, 即根据引理 2, $\nu \in (2, d]$, π_β 满足带权 Poincaŕe 不等式，其权重为 $1+|x|^2$，带权 Poincaŕe 常数为

$$C_{\mathrm{WPI}} = C_V(\gamma)\left(\frac{\beta}{\gamma} - 1\right)^{-1} = \frac{(d+2)^2}{2(d+1)\beta(2\beta-d-2)} = \frac{(d+2)^2}{\nu(d+1)(d+\nu)}.$$