arXiv:2304.04095v2 [stat.ML] 8 Jun 2023

光滑性与等周性条件下 Metropolis 调整朗之万算法混合性的简单证明

Yuansi Chen[1] 和 Khashayar Gatmiry[2]

[1]杜克大学[2]麻省理工学院

摘要

我们研究了在 $\mathbb{R}^d$上采样目标密度的 Metropolis 调整朗之万算法（MALA）的混合时间。我们假设目标密度满足 ψ_μ- 等周性，并且其 Hessian矩阵的算子范数和迹分别有界于 L 和Υ。我们的主要结果表明，从热启动出发，为达到 ϵ-全变差距离，MALA 经过 $O\left(\frac{(L\Upsilon)}{\psi_\mu^2}\frac{1}{2}\log\left(\frac{1}{\epsilon}\right)\right)$ 次迭代即可达到混合。值得注意的是，该结果适用于超越对数凹采样的情形，并且混合时间仅依赖于 Υ，而不是其上界 Ld。在 m-强对数凹和L-对数光滑采样情形下，我们的界恢复了 MALA 先前的极小极大混合界 [WSC22]。

1 引言

本文旨在阐明当前的证明技术能够说明在 $\mathbb{R}^d$ 上采样满足等周性的光滑目标密度时，Metropolis 调整朗之万算法（MALA）的混合时间。尽管 MALA 的混合时间在对数凹分布的情形下已有充分研究，但当采用略弱的假设（如等周性）时，混合性质将如何变化，仍是一个值得探讨的问题。在对未调整朗之万算法（ULA）的研究中，Vempala 和 Wibisono [VW19]曾质疑等周性假设对混合的影响。我们希望针对 MALA 提出同样的问题。具体而言，我们认为，现有的证明技术（例如截断导通率和多元分部积分）——这些技术曾出现在 [WSC22] 中——足以回答这一问题。

Freund 等人 [FMZ22] 研究了在何种条件下，朗之万算法的混合时间与维度无关。他们证明，这类算法的混合时间可以取决于目标密度的 Hessian矩阵的迹，而不是维度本身。因此，与复合采样算法结合后，可以得到一种混合时间与维度无关的算法。我们旨在探究，是否也能证明 MALA 的混合时间仅取决于 Hessian矩阵的迹，而非 $d^{\frac{1}{2}}$ 对 [WSC22]的依赖。

基于上述两点动机，我们研究 MALA 对目标密度的混合时间 μ，该目标密度满足系数为 ψ_μ的Cheeger 等周不等式，并且其Hessian矩阵的算子范数和迹分别有界于 L 和 Υ。我们给出一个简单的证明，表明从热启动开始，MALA 经过 $O\left(\frac{(L\Upsilon)^{\frac{1}{2}}}{\psi_\mu^2}\right)$次迭代即可达到 ϵ-全变差距离的目标密度。与已有工作 [WSC22], 相比，我们的结果表明，其中提出的现有证明技术（例如截断导通率和多元分部积分）足以处理具有等周性的光滑目标密度上的采样问题。当应用于从 m-强对数凹且 L-对数光滑的密度中采样时，我们的混合时间上界恢复了 MALA 先前的极小极大混合界 [WSC22]。此外，我们的混合时间界中对 Υ 的依赖使我们能够在若干场景下获得较弱的维度依赖。

2 预备知识

本节介绍 MALA 算法，并提供建立马尔可夫链混合时间所需的背景知识。

2.1 马尔可夫链基础

我们考虑从目标测度 μ 中进行采样的问题，该测度相对于 $\mathbb{R}^d$ 上的勒贝格测度具有密度。给定一个转移核为 P 的马尔可夫链：$\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^n) \to \mathbb{R}_{\geq 0}$，其中 $\mathcal{B}(\mathbb{R}^d)$ 表示 $\mathbb{R}^d$ 上的博雷尔 σ-代数，k 步转移核 P^k 递归定义为 $P^k(x, dy) := \int_{z\in\mathbb{R}^n} P^{k-1}(x, dz)P(z, dy)$。如果马尔可夫链相对于目标测度 μ 可逆，则

$$\mu(dx)P(x, dy) = \mu(dy)P(y, dx).$$

可以将该马尔可夫链与一个转移算子 $\mathcal{T}_P$ 联系起来。

$$\mathcal{T}_P(\nu)(S) := \int_{y\in\mathbb{R}^n} d\nu(y)P(y, S), \quad \forall S \in \mathcal{B}(\mathbb{R}^n).$$

换言之，当 ν 是当前状态的分布时，$\mathcal{T}_P(\nu)$ 给出下一状态的分布。而 $\mathcal{T}_P^n(\nu) := \mathcal{T}_{P^n}(\nu)$ 则是经过 n 步后状态的分布。

s-导通率。对于 $s \in (0,1)$，我们将具有平稳测度 μ 的马尔可夫链 P 的 s-导通率 Φ_s 定义如下

$$\Phi_s(P) := \inf_{S: s<\mu(S)<1-s} \frac{\int_S P(x, S^c)\mu(dx)}{\min\{\mu(S), \mu(S^c)\} - s}. \tag{1}$$

与导通率（$s = 0$ 情形）相比，s-导通率允许我们忽略分布中导通率难以界定的小部分。

惰性马尔可夫链。给定一个转移核为 P 的马尔可夫链。我们将其惰性变体P惰性定义为：以至少 $\frac{1}{2}$ 的概率停留在同一状态，定义如下。

$$P^{\text{lazy}}(x, S) := \frac{1}{2}\delta_{x\to S} + \frac{1}{2}P(x, S).$$

这里，$\delta_{x\to\cdot}$ 是位于 x 处的狄拉克测度。由于惰性变体只会使收敛速度降低一个常数因子，为了简化理论分析，本文研究惰性马尔可夫链。

总变差距离。设两个概率分布 $\mathcal{P}_1$、$\mathcal{P}_2$ 之间的总变差（TV）距离为

$$\mathrm{d}_{\mathrm{TV}}(\mathcal{P}_1, \mathcal{P}_2) := \sup_{A\in\mathcal{B}(\mathbb{R}^n)} |\mathcal{P}_1(A) - \mathcal{P}_2(A)|,$$

其中 $\mathcal{B}(\mathbb{R}^n)$是 $\mathbb{R}^n$上的博雷尔σ代数。如果 $\mathcal{P}_1$ 和 $\mathcal{P}_2$ 分别具有密度 p_1 和 p_2，我们可以写作 $\mathrm{d}_{\mathrm{TV}}(\mathcal{P}_1, \mathcal{P}_2) = \frac{1}{2}\int |p_1(x) - p_2(x)|\, dx$。

热启动。我们称初始测度 μ_0 是 M-热启动的，如果它满足

$$\sup_{S\in\mathcal{B}(\mathbb{R}^n)} \frac{\mu_0(S)}{\mu(S)} \le M.$$

混合时间。对于误差容限 $\epsilon \in (0, 1)$，具有初始分布 μ_0 和目标分布 μ 的马尔可夫链 P，其总变差距离 ϵ-混合时间定义为

$$\tau^P_{\text{mix}}(\epsilon, \mu_0, \mu) := \inf\left\{n \in \mathbb{N} \mid \mathrm{d}_{\mathrm{TV}}\left(\mathcal{T}^n_P(\mu_0), \mu\right) \le \epsilon\right\}.$$

2.2 MALA 基础知识

给定第 k 次迭代时的状态 x_k，MALA 会提出一个新状态 $y_{k+1} \sim \mathcal{N}(x_k - h\nabla f(x_k)\ 2h\mathbb{I}_d)$。随后，为了得到下一个状态，它会使用 Metropolis-Hastings校正来决定接受还是拒绝 y_{k+1}。我们用 $\mathcal{P}_x = \mathcal{N}(x - h\nabla f(x)\ 2h\mathbb{I}_d)$ 表示 MALA 在 x 时的提议核。

MALA 的一步可视为Metropol化哈密顿蒙特卡罗 (HMC) 的一步，其中包含单个蛙跳步 [Nea11]。使用HMC 中的记号，MALA 的工作方式如下：在每次迭代中，给定步长 $\eta = (2h)^{\frac{1}{2}}$ 和当前状态 $q_0 \in \mathbb{R}^d$，抽取一个独立的高斯 $p_0 \sim \mathcal{N}(0, \mathbb{I}_d)$，下一状态的提议定义为 q_η 如下

$$\begin{aligned} q_\eta &= q_0 + \eta p_0 - \frac{\eta^2}{2}\nabla f(q_0) \\ p_\eta &= p_0 - \frac{\eta}{2}\nabla f(q_0) - \frac{\eta}{2}\nabla f(q_\eta). \end{aligned} \tag{2}$$

Metropolis-Hastings校正用于确保正确的平稳测度，接受概率为

$$\mathcal{A}(q_0, p_0) := \min\left\{1, \frac{\exp\left(-f(q_\eta) - \frac{1}{2}\|p_\eta\|_2^2\right)}{\exp\left(-f(q_0) - \frac{1}{2}\|p_0\|_2^2\right)}\right\}. \tag{3}$$

参见[Nea11， WSC22]，有关 MALA与 HMC之间联系的更多细节请参阅相关文献。下文中，我们采用仅包含一个蛙跳步的 Metropolized HMC 记号，其中 $\eta = (2h)^{\frac{1}{2}}$。

2.3 关于目标密度的假设

以下是可以对目标密度施加的正则性假设。

二阶可微函数 f： $\mathbb{R}^d \to \mathbb{R}$ 是 L-光滑的，当且仅当 $\nabla^2 f_x \preceq L\mathbb{I}_d$、$\forall x \in \mathbb{R}^d$。目标密度 $\mu \propto e^{-f}$ 被称为 L-对数光滑的，是指 f 是 L-光滑的。

- 目标密度 $\mu \sim e^{-f}$ 满足系数为 ψ_μ 的切格尔等周不等式；对于 $\mathbb{R}^d$的任意划分 (S_1, S_2, S_3)，均满足以下不等式

$$\mu(S_3) \geq \psi_\mu \cdot \mathfrak{D}(S_1, S_2) \cdot \mu(S_1)\mu(S_2).$$

其中，集合 S_1、S_2 之间的距离定义为 $\mathfrak{D}(S_1, S_2) = \inf_{x\in S_1, y\in S_2}\|x - y\|_2$。

- 如果存在 $\lambda > 0$，使得目标密度 $\mu \propto e^{-f}$ 具有亚指数尾，那么它就是亚指数尾的。

$$\lim_{\|x\|_2\to\infty} e^{-\lambda\|x\|_2} e^{-f(x)} \to 0.$$

注意，如果 μ 满足 Cheeger 等周不等式，那么它也具有次指数尾。更准确地说，′的 Poincar μ 常数可由 Cheeger 等周系数通过 Cheeger 不等式 [Che15、Maz60]（另见 [Mil09]中的定理 1.1）得到，而 ′的 Poincar μ 常数则意味着由 Gromov 和 Milman 得出的次指数尾界 [GM83]（另见 [GRS15]中的定理 1.7）。

3 主要结果

定理 1.设 $\mu \propto e^{-f}$ 是定义在 $\mathbb{R}^d$上的 L-对数光滑目标密度，并具有Cheeger 等周系数 ψ_μ。设 $\Upsilon = \sup_{x\in\mathbb{R}^d}$迹$(\nabla^2 f_x)$。那么存在普适常数$c_0$和 c_1，使得对于任意误差容限 $\epsilon \in (0, 1)$，从任意 M-warm 初始测度 μ_0出发，具有步长的惰性 MALA 的混合时间

$$h = \frac{\eta^2}{2} = \frac{c_0}{\max\left\{(L\Upsilon)^{\frac{1}{2}}, L\log\left(\frac{(L\Upsilon)^{\frac{1}{4}}}{\psi_\mu}\frac{M}{\epsilon}\right)\right\}}$$

满足

$$\tau_{mix}(\epsilon, \mu_0, \mu) \leq c_1 \frac{\max\left\{(L\Upsilon)^{\frac{1}{2}}, L\log\left(\frac{(L\Upsilon)^{\frac{1}{4}}}{\psi_\mu}\frac{M}{\epsilon}\right)\right\}}{\psi_\mu^2}\log\left(\frac{M}{\epsilon}\right).$$

该定理表明，取阶为 $1\ (L\Upsilon)^{\frac{1}{2}}$ 的步长后，在忽略对数因子~~的情~~况下，混合时间的阶为 $(L\Upsilon)^{\frac{1}{2}}\psi_\mu^2$。注意，在 L-对数光滑假设下，Hessian矩阵的所有迹上的上确界 Υ 总是有上界 Ld。特别地，在 m-强对数凹且 L-对数光滑密度在 [WSC22], 中的采样设置下，我们的混合时间上界为 $d^{\frac{1}{2}}\kappa$，其中 $\kappa=\frac{L}{m}$，因为 m-强对数凹密度的等周系数为 $m^{\frac{1}{2}}$。我们的混合时间上界恢复了 MALA [WSC22] 此前的极小极大混合界，精确到常数因子。此外，这里还有一个我们的界更优的对数凹采样示例：令 Hessian矩阵满足

$$\operatorname{diag}\left(\frac{L}{d},\frac{L}{2d},\ldots,\frac{L}{2d}\right)\preceq\nabla^2 f_x\preceq\operatorname{diag}\left(L,\frac{L}{d},\ldots,\frac{L}{d}\right).$$

直接应用 [WSC22] 会导致混合时间对维度的依赖为 $d^{\frac{3}{2}}$，因为 κ 的量级为 d，并且它要求采用更小的步长。然而，定理 1中的界给出的混合时间量级为 d，因为 Υ 的量级仅为 $2L$。

请注意，在对数凹采样的情形下，定理 1中所述的热启动可通过运行欠阻尼朗之万算法若干步 [AC23]以算法方式获得。

为了证明定理 1，我们遵循通过下界估计马尔可夫链的导通率来分析混合时间的框架 [SJ89、LS93]。以下引理将混合时间分析问题归结为下界估计 s-导通率 Φ_s。

引理 1 (Lovasz 和 Simonovits [LS93])。考虑一个转移核为 P、平稳测度为 μ的可逆惰性马尔可夫链。设 μ_0是一个 M-热初始测度。设 $0<s<\frac{1}{2}$。则

$$\mathrm{d_{TV}}\left(\mathcal{T}_P^n(\mu_0),\mu\right)\leq Ms+M\left(1-\frac{\Phi_s^2}{2}\right)^n.$$

使用 s-导通率代替通常的导通率，使我们能够专注于具有良好混合行为的高概率区域。为了在高概率区域内给出导通率的下界，我们遵循 [WSC22], 中定理 3 的证明，该定理扩展了 Lov´asz [Lov99] 的一个结果（另见 [DCWY19]中的引理 2）。直观地说，该结果表明，只要在高概率区域内有 $\mathrm{d_{TV}}\left(\mathcal{T}_x,\mathcal{T}_y\right)\leq 1-\rho$，当 $\|x-y\|_2\leq\Delta$ 时，我们就有$\Phi_s\geq\frac{\rho\Delta}{\psi_\mu}$。因此，剩下的任务是界定转移重叠度，即两个相近点 $\mathcal{T}_x$ 和 $\mathcal{T}_y$ 之间的 TV距离 x， y。

我们通过以下三角不等式来界定 TV 距离，

$$\mathrm{d_{TV}}\left(\mathcal{T}_x,\mathcal{T}_y\right)\leq\mathrm{d_{TV}}\left(\mathcal{T}_x,\mathcal{P}_x\right)+\mathrm{d_{TV}}\left(\mathcal{P}_x,\mathcal{P}_y\right)+\mathrm{d_{TV}}\left(\mathcal{P}_y,\mathcal{T}_y\right).$$

为界定上述项，我们分两步进行。首先，我们界定提议分布的重叠度，即两个相近点 x、y 所对应的 $\mathcal{P}_x$ 与 $\mathcal{P}_y$ 之间的TV 距离。根据Pinsker不等式，该TV 距离的上界由KL 散度给出。由于这两个提议分布均为高斯分布，KL 散度具有闭式表达。基于这一思路，我们有以下引理

引理 2. 设 $\eta^2 L \leq 1$。对于任意 x、$y \in \mathbb{R}^d$，有

$$\mathrm{d}_{\mathrm{TV}}\left(\mathcal{P}_x, \mathcal{P}_y\right) \leq \frac{2}{\eta}\|x-y\|_2 .$$

证明从略，因为它直接由 [DCWY19]中的引理 7 得出。接下来，我们界定接受率。由于 $\mathrm{d}_{\mathrm{TV}}\left(\mathcal{T}_x, \mathcal{P}_x\right)=1-\mathbb{E}_{z \sim \mathcal{N}\left(0, \mathbb{I}_d\right)}\left[\mathcal{A}(x, z)\right]$。只需以高概率控制式（3）中的接受率。

引理 3. 假设 f 是 L-光滑的，且步长满足

$$\eta^4 \leq \frac{1}{4096 \max\left\{L^2 \log(1/\delta), L\Upsilon\right\}},$$

对于任意 $\delta \in (0,1)$，存在一个集合 $\Lambda \subset \mathbb{R}^{d \times d}$ 满足 $\mathbb{P}_{(q_0,p_0)\sim\mu}((q_0,p_0) \in \Lambda) \geq 1-\delta$，使得对于 $(q_0,p_0) \in \Lambda$ 和 (q_η,p_η) （定义见式 (2)），有

$$-f(q_\eta)-\frac{1}{2}\|p_\eta\|_2^2+f(q_0)+\frac{1}{2}\|p_0\|_2^2 \geq -\frac{1}{4}. \tag{4}$$

该引理的证明构成了本文的创新之处，并在第 4 节给出。

定理 *1* 的证明。 定理 1 的证明是在 [WSC22] 中定理 3 的证明基础上进行改写，以纳入引理 4中对接受率的新颖控制。证明的其余部分完全相同。为完整起见，我们将证明列出。

应用引理1，取

$$s=\frac{\epsilon}{2M}, n \geq \frac{2}{\Phi_s^2}\log\frac{2M}{\epsilon},$$

只需给出导通率的下界 Φ_s。令 $\delta \in (0,\ 1)$ 为稍后选定的值，并令 Λ为引理3中的集合。定义

$$\Xi:=\left\{q_0 \in \mathbb{R}^d : \mathbb{P}_{p_0 \sim \mathcal{N}\left(0, \mathbb{I}_d\right)}\left((q_0,p_0) \in \Lambda\right) \geq \frac{15}{16}\right\}.$$

然后 $\mathbb{P}_{q_0\sim\mu}(q_0 \in \Xi) \geq 1-16\delta$。应用引理 3，我们得到

$$\mathcal{A}(q_0,p_0) \geq e^{-\frac{1}{4}},\ \text{for}\ (q_0,p_0) \in \Lambda.$$

因此，我们有

$$\mathrm{d}_{\mathrm{TV}}\left(\mathcal{T}_{q_0}, \mathcal{P}_{q_0}\right)=1-\mathbb{E}_{p_0 \sim \mathcal{N}\left(0, \mathbb{I}_d\right)}[\mathcal{A}(q_0,p_0)] \leq 1-\frac{15}{16}e^{-\frac{1}{4}} \leq \frac{1}{3}, \forall q_0 \in \Xi.$$

结合引理2，可得

$$\begin{aligned}\mathrm{d}_{\mathrm{TV}}\left(\mathcal{T}_x^{\mathrm{lazy}}, \mathcal{T}_y^{\mathrm{lazy}}\right) &\leq \frac{1}{2}+\frac{1}{2}\mathrm{d}_{\mathrm{TV}}\left(\mathcal{T}_x, \mathcal{T}_y\right) \\ &\leq \frac{1}{2}+\frac{1}{2}\left(\frac{2}{3}+\frac{2\|x-y\|_2}{\eta}\right),\end{aligned} \tag{5}$$

其中 $\mathcal{T}_x^{\text{lazy}}$ 是 $\mathcal{T}_x$ 的懒惰版本。为了给出 s-导通率的下界，考虑任意满足 $\mu(S) \in (s, \frac{1}{2}]$ 的 S 可测集。定义集合

$$S_1 := \left\{ x \in S \mid \mathcal{T}_x^{\text{lazy}} < \frac{1}{16} \right\}, S_2 := \left\{ x \in S^c \mid \mathcal{T}_x^{\text{lazy}} < \frac{1}{16} \right\}, S_3 := (S_1 \cup S_2)^c.$$

Note S_1 和 S_2 是导通性不佳的集合。现在分两种情况 .

如果 $\mu(S_1) \leq \frac{1}{2}\mu(S)$ 或 $\mu(S_2) \leq \frac{1}{2}\mu(S^c)$，那么导通性良好的集合足够大:

$$\int_S \mathcal{T}_x^{\text{lazy}}(S^c)\mu(dx) \geq \frac{1}{32}\mu(S).$$

否则，对于任意 $x \in S_1 \cap \Xi$ 和 $y \in S_2 \cap \Xi$，我们有

$$\mathrm{d_{TV}}\left(\mathcal{T}_x^{\text{lazy}}, \mathcal{T}_y^{\text{lazy}}\right) \geq \left|\mathcal{T}_x^{\text{lazy}}(S) - \mathcal{T}_y^{\text{lazy}}(S)\right| \geq \frac{15}{16} - \frac{1}{16} \geq \frac{7}{8}.$$

结合等式 (5)，我们得到 $\mathfrak{D}(S_1 \cap \Xi,\ S_2 \cap \Xi) \geq \frac{\eta}{24}$。在$\mu$上的等周不等式给出

$$\mu(S_3 \cup \Xi^c) \geq \frac{\eta}{24}\psi_\mu \mu(S_1 \cap \Xi)\mu(S_2 \cap \Xi).$$

由于 $\mu(\Xi) \geq 1 - 16\delta$，在条件下去掉 Ξ 之外的部分后

$$16\delta \leq \min\left\{\frac{1}{4}s, \frac{1}{1536}\eta\psi_\mu s\right\}, \tag{6}$$

我们得到

$$\mu(S_3) \geq c\eta\psi_\mu\mu(S),$$

对于某个普适常数 $c > 0$。在这种情况下，我们有

$$\int_S \mathcal{T}_x^{\text{lazy}}(S^c)\mu(dx) \geq \frac{1}{32}\mu(S_3) \geq \frac{c}{32}\eta\psi_\mu\mu(S).$$

结合这两种情况， s-导通率的下界如下所示。

$$\Phi_s \geq \frac{1}{32}\min\left\{1, c\eta\psi_\mu\right\}.$$

δ 必须根据公式（6）进行选择，且步长 η 必须根据引理3中的第一个公式进行选择。

□

4 接受率控制

在本节中，我们证明引理 3。也就是说，我们给出步长为 η 时 MALA 接受率的下界，并且对 (q_0, p_0) 以高概率成立。

4.1 中间集中性结果

利用 Υ 和 ℓ，定义常数

$$\Upsilon_\ell := \Upsilon + 2(\ell - 1)L. \tag{7}$$

引理4。 令整数 $\ell \geq 1$。假设 e^{-f} 具有亚指数尾， f 是 L-光滑的，且上确界$x\in\mathbb{R}^d$迹($\nabla^2 f_x$) $\leq \Upsilon$，令 Υ_ℓ 由式 (7)定义，则

$$\left[\mathbb{E}_{q\sim e^{-f}} \|\nabla f(q)\|_2^{2\ell}\right]^{\frac{1}{\ell}} \leq \Upsilon_\ell.$$

引理 5. 设整数 $\ell \geq 1$ 和 $x \in \mathbb{R}^d$。假设 迹($\nabla^2 f_x$) $\leq \Upsilon$，令 Υ_ℓ 由式 (7)定义，则

$$\left[\mathbb{E}_{p\sim\mathcal{N}(0,\mathbb{I}_d)} \left(p\nabla^2 f_x p\right)^\ell\right]^{\frac{1}{\ell}} \leq \Upsilon_\ell$$

引理6。 令整数 $\ell \geq 1$， $t \geq 0$ 和 $t^2 L \leq 1$。假设 f 是 L-光滑的，且其 Hessian矩阵的迹上界为 $\sup_{x\in\mathbb{R}^d}$迹($\nabla^2 f_x$) $\leq \Upsilon$，令 Υ_ℓ 由公式 (7)定义，则

$$\left[\mathbb{E}_{(q_0,p_0)\sim\mu} \left(p_0^\top \nabla^2 f_{q_t} p_0\right)^\ell\right]^{\frac{1}{\ell}} \leq 2\Upsilon_\ell,$$

其中 $q_t = q_0 + tp_0 - \frac{t^2}{2}\nabla f(q_0)$ 如公式 (2)中所定义。

引理 7. 设整数 $\ell \geq 1$， $0 \leq t \leq \eta$ 且 $\eta^2 L \leq 1$。假设 f 是 L-光滑的，且上确界$x\in\mathbb{R}^d$迹($\nabla^2 f_x$) $\leq \Upsilon$，令 Υ_ℓ 由公式 (7)定义，则

$$\left[\mathbb{E}_{(q_0,p_0)\sim\mu} \|\nabla f(q_t) - \nabla f(q_0)\|_2^{2\ell}\right]^{\frac{1}{\ell}} \leq 4t^2 L\Upsilon_\ell, \ \ and$$
$$\left[\mathbb{E}_{(q_0,p_0)\sim\mu} \|\nabla f(q_t) - \nabla f(q_\eta)\|_2^{2\ell}\right]^{\frac{1}{\ell}} \leq 4(\eta - t)^2 L\Upsilon_\ell,$$

其中 $q_t = q_0 + tp_0 - \frac{t^2}{2}\nabla f(q_0)$ 如方程 (2).（2）中所定义。

4.2 引理 3 的证明

Give对于整数 $\ell \geq 1$，我们引入随机变量 q 在 $\mathbb{R}^d$ a 上的 ℓ 阶范数 s

$$\|\mathbf{q}\|_\ell := \left(\mathbb{E}\,\|\mathbf{q}\|_2^\ell\right)^{\frac{1}{\ell}}.$$

这个 $\|\cdot\|_\ell$ 范数满足三角不等式

$$\|\mathbf{q}_1 + \mathbf{q}_2\|_\ell \leq \|\mathbf{q}_1\|_\ell + \|\mathbf{q}_2\|_\ell.$$

引理 3 的证明。正如先前研究 [DCWY19、WSC22], 所指出的，直接应用泰勒级数展开不足以得到良好的接受率界。我们证明的新意在于谨慎地运用分部积分。

令 $\Delta_\eta(q_0,p_0) := f(q_\eta) - f(q_0) + \frac{1}{2}\|p_\eta\|_2^2 - \frac{1}{2}\|p_0\|_2^2$。为简便起见，如果上下文清楚，我们省略对 (q_0,p_0) 的依赖。我们的目标是给出 Δ_η 的高概率上界。回顾一下，步长为 η 的 MALA 提案为

$$q_\eta = q_0 + \eta p_0 - \frac{\eta^2}{2}\nabla f(q_0),$$
$$p_\eta = p_0 - \frac{\eta}{2}\nabla f(q_0) - \frac{\eta}{2}\nabla f(q_\eta).$$

展开接受率（4）中的各项，可得

$$\begin{aligned}
f(q_\eta) - f(q_0) &= \int_0^\eta \nabla f(q_t)^\top (p_0 - t\nabla f(q_0))dt, \text{ and} \\
\|p_\eta\|_2^2 - \|p_0\|_2^2 &= \left\|p_0 - \eta\nabla f(q_0) + \frac{\eta}{2}\left(\nabla f(q_0) - \nabla f(q_\eta)\right)\right\|_2^2 - \|p_0\|_2^2 \\
&= \|p_0 - \eta\nabla f(q_0)\|_2^2 + 2\left(p_0 - \eta\nabla f(q_0)\right)^\top \left(\frac{\eta}{2}\left(\nabla f(q_0) - \nabla f(q_\eta)\right)\right) \\
&\quad + \left\|\frac{\eta}{2}\left(\nabla f(q_0) - \nabla f(q_\eta)\right)\right\|_2^2 - \|p_0\|_2^2
\end{aligned}$$

根据 $\nabla f(q_0)$ 和 p_0 重新排列各项后，我们得到

$$\begin{aligned}
&\Delta_\eta(q_0,p_0) \\
&= f(q_\eta) - f(q_0) + \frac{1}{2}\|p_\eta\|_2^2 - \frac{1}{2}\|p_0\|_2^2 \\
&= \underbrace{\nabla f(q_0)^\top \left[-\int_0^\eta t(\nabla f(q_t) - \nabla f(q_\eta))dt\right] + p_0^\top \left[\int_0^\eta \left(\nabla f(q_t) - \frac{1}{2}\nabla f(q_0) - \frac{1}{2}\nabla f(q_\eta)\right)dt\right]}_{=:B_\eta(q_0,p_0)} \\
&\quad + \frac{\eta^2}{8}\|\nabla f(q_\eta) - \nabla f(q_0)\|_2^2 .
\end{aligned} \tag{8}$$

对于偶整数 $\ell \geq 1$，我们考虑 Δ_η 的 ℓ 阶矩，如下所示

$$\left[\mathbb{E}_{(q_0,p_0)\sim\mu\times\mathcal{N}(0,\mathbb{I}_d)}\Delta_\eta^\ell\right]^{\frac{1}{\ell}} \leq 2\left(\left[\mathbb{E}_{(q_0,p_0)\sim\mu\times\mathcal{N}(0,\mathbb{I}_d)}B_\eta^\ell\right]^{\frac{1}{\ell}} + \frac{\eta^2}{8}\mathbb{E}\left[\|\nabla f(q_\eta) - \nabla f(q_0)\|_2^{2\ell}\right]^{\frac{1}{\ell}}\right). \tag{9}$$

根据引理 7，上面的第二项有界。剩下需要界定 $\mathbb{E}B_\eta^\ell$。定义向量 $v_\eta(q_0,p_0) \in \mathbb{R}^{2d}$ 如下

$$v_\eta(q_0,p_0) := \begin{bmatrix} -\int_0^\eta t(\nabla f(q_t) - \nabla f(q_\eta))dt \\ \int_0^\eta \left(\nabla f(q_t) - \frac{1}{2}\nabla f(q_0) - \frac{1}{2}\nabla f(q_\eta)\right)dt \end{bmatrix}. \tag{10}$$

然后我们可以写成

$$B_\eta(q_0,p_0) = v_\eta(q_0,p_0)^\top \begin{bmatrix} \nabla f(q_0) \\ p_0 \end{bmatrix}.$$

约束 B_η 的主要策略是对 q_0 和 p_0联合应用分部积分。v_η 关于（q_0， p_0）的导数是

$$
\begin{aligned}
&\mathbf{D}v_\eta(q_0,p_0)\\
&=\begin{bmatrix}\frac{\partial v_1}{\partial q_0} & \frac{\partial v_1}{\partial p_0}\\ \frac{\partial v_2}{\partial q_0} & \frac{\partial v_2}{\partial p_0}\end{bmatrix}\\
&=\begin{bmatrix} -\int_0^\eta t\left[\left(\mathbb{I}_d-\frac{t^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_t}-\left(\mathbb{I}_d-\frac{\eta^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_\eta}\right]dt & \int_0^\eta t\left(t\nabla^2 f_{q_t}dt-\eta\nabla^2 f_{q_\eta}\right)dt\\ \int_0^\eta\left(\mathbb{I}_d-\frac{t^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_t}dt-\frac{\eta}{2}\nabla^2 f_{q_0}-\frac{\eta}{2}\left(\mathbb{I}_d-\frac{\eta^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_\eta} & \int_0^\eta t\nabla^2 f_{q_t}dt-\frac{\eta^2}{2}\nabla^2 f_{q_\eta}\end{bmatrix}
\end{aligned}
$$

我们有

$$
\begin{aligned}
\mathbb{E}B_\eta^\ell &= \mathbb{E}B_\eta^{\ell-1}v_\eta^\top\begin{bmatrix}\nabla f(q_0)\\ p_0\end{bmatrix}\\
&\overset{(i)}{=}\mathbb{E}B_\eta^{\ell-1}\underbrace{\operatorname{trace}\left(\mathbf{D}v_\eta\right)}_{A_1}\\
&\quad+(\ell-1)\mathbb{E}B_\eta^{\ell-2}\underbrace{v_\eta^\top\begin{bmatrix}\nabla^2 f_{q_0} & 0\\ 0 & \mathbb{I}_d\end{bmatrix}v_\eta}_{A_2}\\
&\quad+(\ell-1)\mathbb{E}B_\eta^{\ell-2}\underbrace{\begin{bmatrix}\nabla f(q_0)\\ p_0\end{bmatrix}^\top\mathbf{D}v_\eta v_\eta}_{A_3}
\end{aligned}
\tag{11}
$$

(i) 应用多元分部积分（我们对 $\begin{bmatrix}\nabla f(\ q_0\\ p_0\end{bmatrix}e^{-f(q_0)})-\frac{1}{2}\|p_0\|_2^2$进行积分，并且边界项消失，因为 e^{-f} 具有亚指数尾），由此得到三项。

方程中的第一项 (11)。 我们有

$$
\begin{aligned}
A_1 &= \operatorname{trace}\left(\mathbf{D}v_\eta\right)\\
&= \operatorname{trace}\left(\int_0^\eta\frac{t^3}{2}\nabla^2 f_{q_0}\nabla^2 f_{q_t}-\frac{\eta^4}{4}\nabla^2 f_{q_0}\nabla^2 f_{q_\eta}\right)\\
&\le \frac{1}{2}\eta^4 L\Upsilon.
\end{aligned}
$$

公式中的第二项 (11)。 我们有

$$
\begin{aligned}
& A_2 \\
&= v_\eta^\top \begin{bmatrix} \nabla^2 f_{q_0} & 0 \\ 0 & \mathbb{I}_d \end{bmatrix} v_\eta \\
&= \left(\int_0^\eta t(\nabla f(q_t) - \nabla f(q_\eta)) dt \right)^\top \nabla^2 f_{q_0} \left(\int_0^\eta t(\nabla f(q_t) - \nabla f(q_\eta)) dt \right) \\
&\quad + \left\| \int_0^\eta \nabla f(q_t) dt - \frac{\eta}{2} \nabla f(q_\eta) - \frac{\eta}{2} \nabla f(q_0) \right\|_2^2 \\
&\leq L \left[\int_0^\eta t \left\| \nabla f(q_t) - \nabla f(q_\eta) \right\|_2 dt \right]^2 \\
&\quad + \frac{1}{4} \left[\int_0^\eta \left\| \nabla f(q_t) - \nabla f(q_0) \right\|_2 + \left\| \nabla f(q_t) - \nabla f(q_\eta) \right\|_2 dt \right]^2 .
\end{aligned}
$$

因此，我们有

$$
\begin{aligned}
& \left\| A_2 \right\|_{\frac{\ell}{2}} \\
&\leq L \left\| \int_0^\eta t \left\| \nabla f(q_t) - \nabla f(q_\eta) \right\|_2 dt \right\|_\ell^2 + \frac{1}{4} \left\| \int_0^\eta \left\| \nabla f(q_t) - \nabla f(q_0) \right\|_2 + \left\| \nabla f(q_t) - \nabla f(q_\eta) \right\|_2 dt \right\|_\ell^2 \\
&\overset{(i)}{\leq} L \left[\frac{\eta^2}{2} \sup_{t \in [0,\eta]} \left\| \nabla f(q_t) - \nabla f(q_\eta) \right\|_\ell \right]^2 + \frac{1}{4} \left[\eta \sup_{t \in [0,\eta]} \left\| \nabla f(q_t) - \nabla f(q_\eta) \right\|_\ell + \left\| \nabla f(q_t) - \nabla f(q_0) \right\|_\ell \right]^2 \\
&\overset{(ii)}{\leq} \eta^6 L^2 \Upsilon_{\frac{\ell}{2}} + 4\eta^4 L \Upsilon_{\frac{\ell}{2}} \\
&\leq 5\eta^4 L \Upsilon_{\frac{\ell}{2}} .
\end{aligned}
$$

（i）由詹森不等式得到。（ii）由引理 7。

公式中的第三项 (11)。A_3 由以下四项组成

$$
\begin{aligned}
&A_3\\
&=\nabla f(q_0)^\top\left[-\int_0^\eta t\left[\left(\mathbb{I}_d-\frac{t^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_t}-\left(\mathbb{I}_d-\frac{\eta^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_\eta}\right]dt\right]\left[-\int_0^\eta t(\nabla f(q_t)-\nabla f(q_\eta))dt\right]\\
&\quad+\nabla f(q_0)^\top\left[\int_0^\eta t\left(t\nabla^2 f_{q_t}dt-\eta\nabla^2 f_{q_\eta}\right)dt\right]\left[\int_0^\eta\left(\nabla f(q_t)-\frac{1}{2}\nabla f(q_0)-\frac{1}{2}\nabla f(q_\eta)\right)dt\right]\\
&\quad+p_0^\top\left[\int_0^\eta\left(\mathbb{I}_d-\frac{t^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_t}dt-\frac{\eta}{2}\left(\mathbb{I}_d-\frac{\eta^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{q_\eta}-\frac{\eta}{2}\nabla^2 f_{q_0}\right]\left[-\int_0^\eta t(\nabla f(q_t)-\nabla f(q_\eta))dt\right]\\
&\quad+p_0^\top\left[\int_0^\eta t\nabla^2 f_{q_t}dt-\frac{\eta^2}{2}\nabla^2 f_{q_\eta}\right]\left[\int_0^\eta\left(\nabla f(q_t)-\frac{1}{2}\nabla f(q_0)-\frac{1}{2}\nabla f(q_\eta)\right)dt\right]\\
&\le\eta^2 L\int_0^\eta t\left\|\nabla f(q_0)\right\|_2\left\|\nabla f(q_t)-\nabla f(q_\eta)\right\|_2 dt\\
&\quad+\frac{1}{2}\eta^3 L\int_0^\eta\left\|\nabla f(q_0)\right\|_2\left(\left\|\nabla f(q_t)-\nabla f(q_0)\right\|_2+\left\|\nabla f(q_t)-\nabla f(q_\eta)\right\|_2\right)dt\\
&\quad+\int_0^\eta L^{\frac{1}{2}}\left(\left(p_0^\top\nabla^2 f_{q_t}p_0\right)^{\frac{1}{2}}+\frac{t^2}{2}L\left(p_0^\top\nabla^2 f_{q_0}p_0\right)^{\frac{1}{2}}\right)dt\left[\int_0^\eta t\left\|\nabla f(q_t)-\nabla f(q_\eta)\right\|_2 dt\right]\\
&\quad+\frac{1}{2}\eta^2 L\left[\left(p_0^\top\nabla^2 f_{q_t}p_0\right)^{\frac{1}{2}}+\left(p_0^\top\nabla^2 f_{q_\eta}p_0\right)^{\frac{1}{2}}\right]\left[\int_0^\eta\left(\left\|\nabla f(q_t)-\nabla f(q_0)\right\|_2+\left\|\nabla f(q_t)-\nabla f(q_\eta)\right\|_2\right)dt\right].
\end{aligned}
$$

The 上述方程中第一项的 $\frac{\ell}{2}$ 阶矩可按如下方式界定 s

$$
\begin{aligned}
&\mathbb{E}\left(\eta^2 L\int_0^\eta t\left\|\nabla f(q_0)\right\|_2\left\|\nabla f(q_t)-\nabla f(q_\eta)\right\|_2 dt\right)^{\frac{\ell}{2}}\\
&\le\eta^{\frac{3\ell}{2}-1}L^{\frac{\ell}{4}}\int_0^\eta\mathbb{E}\left(tL^{\frac{1}{2}}\left\|\nabla f(q_0)\right\|_2\left\|\nabla f(q_t)-\nabla f(q_\eta)\right\|_2\right)^{\frac{\ell}{2}}dt\\
&\le\frac{1}{2}\eta^{\frac{3\ell}{2}-1}L^{\frac{\ell}{4}}\int_0^\eta(t^2L)^{\frac{\ell}{2}}\mathbb{E}\left\|\nabla f(q_0)\right\|_2^\ell+\mathbb{E}\left\|\nabla f(q_t)-\nabla f(q_\eta)\right\|_2^\ell dt\\
&\overset{Lem.\ 4,\ 7}{\le}\left(\eta^5 L^{\frac{3}{2}}\Upsilon_{\frac{\ell}{2}}\right)^{\frac{\ell}{2}}.
\end{aligned}
$$

同样，应用引理 4 5 6 和 7，$\frac{\ell}{2}$ 其他三项的 阶矩分别可界定为 $\left(\eta^5 L^{\frac{3}{2}}\Upsilon_{\frac{\ell}{2}}\right)^{\frac{\ell}{2}}$，$\left(2\eta^4 L\Upsilon_{\frac{\ell}{2}}\right)^{\frac{\ell}{2}}$，$\left(2\eta^4 L\Upsilon_{\frac{\ell}{2}}\right)^{\frac{\ell}{2}}$因此，

$$
\begin{aligned}
&\left\|\!\left|A_3\right|\!\right\|_{\frac{\ell}{2}}\\
&\le\eta^5 L^{\frac{3}{2}}\Upsilon_{\frac{\ell}{2}}+\eta^5 L^{\frac{3}{2}}\Upsilon_{\frac{\ell}{2}}+2\eta^4 L\Upsilon_{\frac{\ell}{2}}+2\eta^4 L\Upsilon_{\frac{\ell}{2}}\\
&\overset{(i)}{\le}6\eta^4 L\Upsilon_{\frac{\ell}{2}}.
\end{aligned}
$$

(i) 使用了条件 $\eta^2 L\le\frac{1}{2}$。

结合 将关于 A_1、A_2 和 A_3 的界代入方程（11），并应用 H¨older 不等式,

我们得到

$$\begin{aligned}\mathbb{E}B_\eta^\ell &\le \left(\mathbb{E}B_\eta^\ell\right)^{\frac{\ell-1}{\ell}}\left(\mathbb{E}A_1^\ell\right)^{\frac{1}{\ell}} + \left(\mathbb{E}B_\eta^\ell\right)^{\frac{\ell-2}{\ell}}\left(\mathbb{E}A_2^{\frac{\ell}{2}}\right)^{\frac{2}{\ell}} + + \left(\mathbb{E}B_\eta^\ell\right)^{\frac{\ell-2}{\ell}}\left(\mathbb{E}A_3^{\frac{\ell}{2}}\right)^{\frac{2}{\ell}} \\ &\le \left(\mathbb{E}B_\eta^\ell\right)^{\frac{\ell-1}{\ell}}\left(\eta^4 L\Upsilon\right) + \left(\mathbb{E}B_\eta^\ell\right)^{\frac{\ell-2}{\ell}}\left(11\eta^4 L\Upsilon_{\frac{\ell}{2}}\right).\end{aligned}$$

对 $\mathbb{E}B_\eta^\ell$ 求解不等式，得到

$$\left[\mathbb{E}B_\eta^\ell\right]^{\frac{1}{\ell}} \le \eta^4 L\Upsilon + \left(11\eta^4 L\Upsilon_{\frac{\ell}{2}}\right)^{\frac{1}{2}}.$$

将上述界代入式（9），得到

$$\left[\mathbb{E}_{(q_0,p_0)\sim\mu\times\mathcal{N}(0,\mathbb{I}_d)}\Delta_\eta^\ell\right]^{\frac{1}{\ell}} \le 3\eta^4 L\Upsilon_\ell + 4\left(\eta^4 L\Upsilon_{\frac{\ell}{2}}\right)^{\frac{1}{2}}.$$

根据马尔可夫不等式，得到

$$\mathbb{P}_{(q_0,p_0)\sim\mu\times\mathcal{N}(0,\mathbb{I}_d)}\left(|\Delta_\eta(q_0,p_0)| \ge \alpha\right) \le \frac{\mathbb{E}\Delta_\eta^\ell}{\alpha^\ell}.$$

取 $\alpha = e^{\frac{1}{2}}\left[3\eta^4 L\Upsilon_\ell + 4\left(\eta^4 L\Upsilon_{\frac{\ell}{2}}\right)^{\frac{1}{2}}\right]$ 和 $\ell = 2\left\lceil\log\left(\frac{1}{\delta}\right)\right\rceil$，则右侧小于 $e^{-\ell/2} \le \delta$。最后，若取步长使得 $\eta^4 L\Upsilon \le \frac{1}{4096}$ 和 $\eta^4 L^2\log\left(\frac{1}{\delta}\right) \le \frac{1}{4096}$，则

$$\alpha = e^{\frac{1}{2}}\left[3\eta^4 L\Upsilon_\ell + 4\left(\eta^4 L\Upsilon_{\frac{\ell}{2}}\right)^{\frac{1}{2}}\right] \le \frac{1}{4}.$$

□

5 第 4.1 节的结果证明4.1

本节中，我们证明第 4.1 节中的中间浓缩结果。

引理的证明 *4*. 我们有

$$\begin{aligned}&\mathbb{E}_{q\sim e^{-f}}\left\|\nabla f(q)\right\|_2^{2\ell} \\ &= \mathbb{E}\left\|\nabla f(q)\right\|_2^{2(\ell-1)}\nabla f(q)^\top\nabla f(q) \\ &\overset{(i)}{=} \mathbb{E}\left\|\nabla f(q)\right\|_2^{2(\ell-1)}\operatorname{trace}\left(\nabla^2 f_q\right) + (\ell-1)\mathbb{E}\left\|\nabla f(q)\right\|_2^{2(\ell-2)}\operatorname{trace}\left(2\nabla f(q)^\top\nabla^2 f_q\nabla f(q)\right) \\ &\le \mathbb{E}\left\|\nabla f(q)\right\|_2^{2(\ell-1)}\left(\operatorname{trace}\left(\nabla^2 f_q\right) + 2(\ell-1)L\right).\end{aligned}$$

(i) 使用多元分部积分（或格林恒等式，$\nabla u = \nabla f(q)e^{-f(q)}, \nabla v = \|\nabla f(q)\|_2^{2(\ell-1)}\nabla f(q)$），由于 e^{-f} 具有次指数尾，边界项为零。由 H¨older 不等式可得

$$\mathbb{E}_{q\sim e^{-f}}\left\|\nabla f(q)\right\|_2^{2\ell} \le \left(\Upsilon + 2(\ell-1)L\right)^\ell = \Upsilon_\ell^\ell.$$

□

引理的证明5。我们有

$$\begin{aligned}
&\mathbb{E}_{p\sim\mathcal{N}(0,\mathbb{I}_d)}\left(p^\top\nabla^2 f_x p\right)^\ell \\
&= \mathbb{E}_{p\sim\mathcal{N}(0,\mathbb{I}_d)}\left(p^\top\nabla^2 f_x p\right)^{\ell-1} p^\top\nabla^2 f_x p \\
&\overset{(i)}{=} \mathbb{E}_{p\sim\mathcal{N}(0,\mathbb{I}_d)}\left(p^\top\nabla^2 f_x p\right)^{\ell-1}\operatorname{trace}\left(\nabla^2 f_x\right) + 2(\ell-1)\mathbb{E}_{p\sim\mathcal{N}(0,\mathbb{I}_d)}\left(p^\top\nabla^2 f_x p\right)^{\ell-2} p^\top\nabla^2 f_x^2 p \\
&\le \mathbb{E}_{p\sim\mathcal{N}(0,\mathbb{I}_d)}\left(p^\top\nabla^2 f_x p\right)^{\ell-1}\left(\Upsilon + 2(\ell-1)L\right).
\end{aligned}$$

(i) 应用多元分部积分（$\nabla u = pe^{-\frac{1}{2}\|p\|_2^2}, \nabla v = \left(p^\top\nabla^2 f_x p\right)^{\ell-1}\nabla^2 f_x p$），且边界项消失。由 H¨older 不等式，我们得到所需结果。 □

引理的证明 6。 情形 $t = 0$ 直接由引理 5 得出。下文中，我们假设 $t > 0$。令 (q_0, p_0) 为从 μ 中抽取的一个随机变量。定义

$$\begin{aligned}
\hat{q}_t &:= q_t = q_0 + tp_0 - \frac{t^2}{2}\nabla f(q_0) \\
\hat{p}_t &:= p_0 - \frac{t}{2}\nabla f(q_0).
\end{aligned}$$

对于熟悉 HMC 的读者，上述过程对应于蛙跳积分的半步。ˆ 定义离散前向映射 F_t，即将 (q_0, p_0) 映射到 $(\hat{q}_t, \hat{p}_t)$ 的映射

$$\begin{bmatrix}\hat{q}_t \\ \hat{p}_t\end{bmatrix} = \hat{F}_t\left(\begin{bmatrix}q_0 \\ p_0\end{bmatrix}\right).$$

其导数为

$$\begin{bmatrix}\frac{\partial\hat{q}_t}{\partial q_0} & \frac{\partial\hat{q}_t}{\partial p_0} \\ \frac{\partial\hat{p}_t}{\partial q_0} & \frac{\partial\hat{p}_t}{\partial p_0}\end{bmatrix} = \begin{bmatrix}\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0} & t\mathbb{I}_d \\ -\frac{t}{2}\nabla^2 f_{q_0} & \mathbb{I}_d\end{bmatrix}.$$

ˆ雅可比矩阵 $\mathrm{J}F_t = 1$。根据隐函数定理，由于雅可比矩阵处处不为零，ˆ F_t 可逆。我们还有其逆映射的导数

$$\begin{bmatrix}\frac{\partial q_0}{\partial\hat{q}_t} & \frac{\partial q_0}{\partial\hat{p}_t} \\ \frac{\partial p_0}{\partial\hat{q}_t} & \frac{\partial p_0}{\partial\hat{p}_t}\end{bmatrix} = \begin{bmatrix}\mathbb{I}_d & -t\mathbb{I}_d \\ \frac{t}{2}\nabla^2 f_{q_0} & \mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0}\end{bmatrix}.$$

根据上述导数，我们有

$$\begin{aligned}
\frac{\partial}{\partial\hat{p}_t}e^{-\frac{1}{2}\|p_0\|_2^2} &= -\frac{\partial p_0}{\partial\hat{p}_t}p_0 e^{-\frac{1}{2}\|p_0\|_2^2} = -\left(\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0}\right)p_0 e^{-\frac{1}{2}\|p_0\|_2^2} \\
\frac{\partial}{\partial\hat{p}_t}e^{-f(q_0)} &= -\frac{\partial q_0}{\partial\hat{p}_t}\nabla f(q_0)e^{-f(q_0)} = t\nabla f(q_0)e^{-f(q_0)}.
\end{aligned}$$

取 ℓ 阶矩，并将其写成适合进行分部积分的形式

$$\begin{aligned}
&\mathbb{E}\left(p_0^\top\nabla^2 f_{\hat{q}_t}p_0\right)^\ell \\
&= \mathbb{E}\left(p_0^\top\nabla^2 f_{\hat{q}_t}p_0\right)^{\ell-1}p_0^\top\nabla^2 f_{\hat{q}_t}\left(\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0}\right)p_0 + \mathbb{E}\left(p_0^\top\nabla^2 f_{\hat{q}_t}p_0\right)^{\ell-1}p_0^\top\nabla^2 f_{\hat{q}_t}\frac{t^2}{2}\nabla^2 f_{q_0}p_0.
\end{aligned}$$

上面的第一项为

$$
\begin{aligned}
&\mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-1} p_0^\top \nabla^2 f_{\hat{q}_t}\left(\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0}\right) p_0 \\
&= \int\int \left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-1} p_0^\top \nabla^2 f_{\hat{q}_t}\left(\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0}\right) p_0 e^{-f(q_0)} e^{-\frac{1}{2}\|p_0\|_2^2} \frac{1}{(2\pi)^{d/2}} dp_0 dq_0 \\
&\overset{(i)}{=} \int\int \left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-1} p_0^\top \nabla^2 f_{\hat{q}_t}\left(\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0}\right) p_0 e^{-f(q_0)} e^{-\frac{1}{2}\|p_0\|_2^2} \frac{1}{(2\pi)^{d/2}} d\hat{p}_t d\hat{q}_t \\
&= -\int\int \left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-1} e^{-f(q_0)} p_0^\top \nabla^2 f_{\hat{q}_t} \frac{\partial}{\partial \hat{p}_t} e^{-\frac{1}{2}\|p_0\|_2^2} \frac{1}{(2\pi)^{d/2}} d\hat{p}_t d\hat{q}_t \\
&\overset{(ii)}{=} \mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-1} \operatorname{trace}\left[\left(\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0}\right)\nabla^2 f_{\hat{q}_t} + t p_0^\top \nabla^2 f_{\hat{q}_t} \nabla f(q_0)\right] \\
&\quad + 2(\ell-1)\mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-2} p_0^\top \nabla^2 f_{\hat{q}_t}(\mathbb{I}_d - \frac{t^2}{2}\nabla^2 f_{q_0})\nabla^2 f_{\hat{q}_t} p_0 \\
&\le \mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-\frac{1}{2}} t L^{\frac{1}{2}} \|\nabla f(q_0)\|_2 + \mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-1} (\Upsilon + 2(\ell-1)L)\,.
\end{aligned}
$$

(i) 由变量代换可得，且雅可比为 1。(ii) 应用多元分部积分（$\nabla u = \frac{\partial}{\partial \hat{p}_t} e^{-\frac{1}{2}\|p_0\|_2^2}$关于 $\hat{p}_t$，∇v为其余部分）。由于 e^{-f} 具有亚指数尾，边界项趋于零。

与第二项合并，我们得到

$$
\begin{aligned}
&\mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell} \\
&\le \mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-\frac{1}{2}} \left(t L^{\frac{1}{2}} \|\nabla f(q_0)\|_2 + t^2 L \left\|\nabla^2 f_{q_0}^{\frac{1}{2}} p_0\right\|_2\right) + \mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell-1} (\Upsilon + 2(\ell-1)L)\,.
\end{aligned}
$$

根据 H¨older 不等式，我们有

$$
\begin{aligned}
&\mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell} \\
&\le \left(t L^{\frac{1}{2}}\right)^{2\ell} \mathbb{E}\,\|\nabla f(q_0)\|_2^{2\ell} + \left(t^2 L\right)^{2\ell} \mathbb{E}\left(p_0 \nabla^2 f_{q_0} p_0\right)^{\ell} + (\Upsilon + 2(\ell-1)L)^{\ell}\,.
\end{aligned}
$$

结合引理4 和引理 5，我们得出

$$
\left[\mathbb{E}\left(p_0^\top \nabla^2 f_{\hat{q}_t} p_0\right)^{\ell}\right]^{\frac{1}{\ell}} \le (1 + t L^{\frac{1}{2}})\,(\Upsilon + 2(\ell-1)L) \le 2Cl.
$$

□

引理的证明 7. 由于我们有

$$
\begin{aligned}
\|\nabla f(q_t) - \nabla f(q_0)\|_2 &= \left\|\int_0^t \nabla^2 f_{q_s}\,(p_0 - s\nabla f(q_0))\,ds\right\|_2 \\
&\le \int_0^t L^{\frac{1}{2}} \left(p_0^\top \nabla^2 f_{q_s} p_0\right)^{\frac{1}{2}} + sL\,\|\nabla f(q_0)\|_2\,ds,
\end{aligned}
$$

由此可得

$$\begin{aligned}\mathbb{E}_{(q_0,p_0)\sim\mu}\left\|\nabla f(q_t)-\nabla f(q_0)\right\|_2^{2\ell} &\le \mathbb{E}t^{2\ell-1}\int_0^t\left(L^{\frac{1}{2}}\left(p_0^\top\nabla^2 f_{q_s}p_0\right)^{\frac{1}{2}}+sL\left\|\nabla f(q_0)\right\|_2\right)^{2\ell}ds\\ &\le \mathbb{E}t^{2\ell-1}2^{\ell-1}\int_0^t\left(L^\ell\left(p_0^\top\nabla^2 f_{q_s}p_0\right)^\ell+s^{2\ell}L^{2\ell}\left\|\nabla f(q_0)\right\|_2^{2\ell}\right)ds\\ &\overset{(i)}{\le}\left(4t^2L\Upsilon_\ell\right)^\ell.\end{aligned}$$

这里（i）由引理4 和 6。 □

参考文献

[AC23] Jason M Altschuler 和 Sinho Chewi。通过算法热启动实现更快、更高精度的对数凹采样。arXiv 预印本 arXiv:2302.10249，2023 年。[Che15] Jeff Cheeger。拉普拉斯算子最小特征值的下界。载于《分析问题》，第 195–200 页。普林斯顿大学出版社，2015 年。[DCWY19] Raaz Dwivedi、Yuansi Chen、Martin J. Wainwright 和 Bin Yu。对数凹采样：Metropolis-Hastings 算法速度很快。《机器学习研究杂志》，20(183):1–42，2019 年。[FMZ22] Yoav Freund、Yi-An Ma 和 Tong Zhang。朗之万算法的收敛时间何时与维度无关？一种复合优化视角。《机器学习研究杂志》，23(214):1–32，2022 年。[GM83] Mikhael Gromov 和 Vitali D Milman。等周不等式的一个拓扑应用。《美国数学杂志》，105(4):843–854，1983 年。[GRS15] Nathael Gozlan、Cyril Roberto 和 Paul-Marie Samson。从无维度浓缩到 Poincar´é 不等式。《变分学与偏微分方程》，52(3-4):899–925，2015 年。[Lov99] L´aszl´o Lov´asz。Hit-and-run 算法的混合速度很快。《数学规划》，86(3):443–461，1999 年。[LS93] L´aszl´o Lov´asz 和 Mikl´os Simonovits。凸体中的随机游走与改进的体积算法。《随机结构与算法》，4(4):359–412，1993 年。[Maz60] Vladimir Gilelevich Maz'ya。函数空间的区域类与嵌入定理。载于《苏联科学院院刊》，第 133 卷，第 527–530 页。俄罗斯科学院，1960 年。[Mil09] Emanuel Milman。凸性在等周性、谱隙与浓缩中的作用。《数学发明》，177(1):1–43，2009 年。

[Nea11] Radford M Neal。使用哈密顿动力学的 MCMC。《马尔可夫链蒙特卡罗手册》，2(11):2，2011。[SJ89] Alistair Sinclair 和 Mark Jerrum。近似计数、均匀生成与快速混合的马尔可夫链。《信息与计算》，82(1):93–133，1989。[VW19] Santosh Vempala 和 Andre Wibisono。未调整朗之万算法的快速收敛：等周性已足够。《神经信息处理系统进展》，32，2019。[WSC22] Keru Wu、Scott Schmidler 和 Yuansi Chen。对数凹采样中 Metropolis 调整朗之万算法的极小极大混合时间。《机器学习研究杂志》，23(1):12348–12410，2022。