arXiv:2611.32322v1 [math.PR] 24 Nov 2026

共振破缺：Metropolis 调整的随机化哈密顿蒙特卡罗的平方根加速度

BY NAWAF BOU-RABEE[1,2,a]

[1]数学系，罗格斯大学 数学科学，罗格斯大学，anawaf.bourabee@rutgers.edu

[2]计算数学中心，Flatiron 研究所

固定积分时间的哈密顿蒙特卡罗可能因共振而无法充分混合，而随机化积分时间可以消除共振。我们证明，对于满足我们称为非共振的傅里叶条件的任意积分时间分布，采用速度 Verlet 和随机积分时间的 Metropolis 调整的 HMC，都能从热启动出发，在与条件数无关的转移次数内实现混合；均匀分布、指数分布和三角分布都是这类分布的例子。对于正比于 e^{-U}在 $\mathbb{R}^d$上的目标密度，其中 U 是具有 m-强凸性、 L-Lipschitz连续梯度、 L_H-Lipschitz连续海森矩阵且满足 $\kappa = L/m$的函数，要达到固定的全变差精度，所需的梯度计算次数在忽略常数和对数因子后为$\sqrt{\kappa}d + \sqrt{L_H}(d/m)^{3/4}$。这是该采样器在高斯目标分布之外首个具有条件数平方根依赖的界。证明将未调整提议的相对熵收缩转化为校正链的导通率界，并表明积分时间分布仅通过其特征函数产生影响。对于固定步长，海森正则性不可或缺：若不具备这一正则性，即使从热启动出发，该类中仍存在一些目标分布，使计算成本至少达到 $\kappa^{2/3}$的量级（忽略对数因子）。

1. 引言。 设 π 是定义在 $\mathbb{R}^d$ 上的概率测度，其密度正比于e^{-U}，其中 U是 m-强凸的，具有 L-Lipschitz 梯度和 L_H-Lipschitz 海森矩阵，且令 $\kappa = L/m$ 为条件数。我们证明，采用速度 Verlet 且使用从非共振持续时间分布（定义2.1）中抽取的随机化积分时间的 Metropolis 校正哈密顿蒙特卡罗（HMC），经过 $m^{-1/2}$次梯度计算后，可达到全变差误差$\varepsilon \in (0,1]$，其初始分布为 L^2-热启动，期望梯度计算次数为

$$\mathcal{O}\Big(\big[\sqrt{\kappa D} + \sqrt{L_H}\,(D/m)^{3/4}\big]\log\tfrac{2M}{\varepsilon}\Big), \qquad D = 1 + d + \log(M/\varepsilon), \tag{1.1}$$

，其中 $M \geq 1$ 给出了 $\|\mathrm{d}\nu_0/\mathrm{d}\pi\|_{L^2(\pi)}$ 的界$\nu_0 \ll \pi$ （定理 1 和推论 1）。在定理 2的持续时间分布条件下，我们还证明，仅有强凸性和 Lipschitz 梯度并不足以保证梯度代价达到 $\sqrt{\kappa}$的量级，即使采用热启动也是如此：对于每个在目标分布确定之前指定的步长所选取的步长（包括定理 1 中使用的步长）下，存在一个目标分布，使得开销至少为 $\kappa^{2/3}$，相差对数因子。

对于每次转移仅执行一个 Verlet 步的 Metropolis 化哈密顿蒙特卡罗，已知的混合界随 κ [13] 线性增长。更长的哈密顿轨迹可以改善这一依赖关系，但固定积分时间可能产生共振。对于高斯目标分布，考虑一个半周期能整除积分时间的模态：无论抽取的动量为何，流都会将该坐标的位置映射为其初始值或相反数，因此马尔可夫链在该坐标上永远无法混合。对积分时间进行适当随机化可以避免这一障碍。

MSC2020 学科分类。 主要分类：65C05、60J05；次要分类：65C40。关键词与短语。 哈密顿蒙特卡罗、随机化积分时间、Metropolis校正、混合时间、条件数、受限导通率。

Mitra、Srinivasan、Wang 和 Wibisono [17, 定理 1 和推论 1] 证明，采用均值为 $m^{-1/2}$ 量级的三角分布或指数分布持续时间的精确流随机化HMC，每次转移都会使相对熵按几何速度收缩，从而使预期总积分时间为$m^{-1/2}$，相差对数因子。

给定动量 $v \in \mathbb{R}^d$，哈密顿量为 $H(x,v) = U(x) + |v|^2/2$，且 $\Pi = \pi \otimes N(0, I_d)$ 是 $\mathbb{R}^{2d}$ 上的玻尔兹曼-吉布斯分布。每次转移都会抽取一个高斯动量，并从固定分布 ρ 中独立抽取一个持续时间 τ，然后以步长 h 执行 $J = \lfloor \tau/h \rfloor + B$ 个 Verlet 步，其中 B 是一个独立的公平伯努利变量。根据 H 的变化进行 Metropolis 检验，以决定接受或拒绝端点；若拒绝，马尔可夫链将停留在当前位置。将位置核记为 P。持续时间分布 ρ 按定义 2.1 的意义是非共振的：对于常数 $b_\rho > 0$ 和 $0 < \delta_\rho \le 1$，$\mathbb{E}\tau \le b_\rho/\sqrt{m}$ 和 $\mathrm{Re}\,\widehat{\rho}(\omega) \le 1 - \delta_\rho$ 对于 $\omega \ge 2\sqrt{m}$。

主要结果。定理 1 给出，在经过 $n = \mathcal{O}(\log(2M/\varepsilon))$ 次转移后，$\mathrm{TV}(\nu_0 P^n, \pi) \le \varepsilon$；其中常数仅依赖于 b_ρ、δ_ρ，因此条件数只影响每次转移的成本。当且仅当 $L_H \le L\sqrt{m}/\sqrt{D}$ 时，(1.1) 中的第二项不大于第一项。该项源自引理 6.2 中的三次Verlet能量余项 $L_H h^3 |w_n|^3/12$，其中 w_n 是半步速度：当 $|w_n|^2$ 取能量截断 $R = \mathcal{O}(D)$ 的量级时，长度为 $m^{-1/2}$ 量级的轨迹上累积的误差为 $L_H h^2 D^{3/2} m^{-1/2}$ 量级；要使其保持有界，就要求 h 为 $L_H^{-1/2} D^{-3/4} m^{1/4}$ 量级，因此每条轨迹需要 $\sqrt{L_H}(D/m)^{3/4}$ 次梯度计算。对于高斯目标分布上的该采样器，当 $h\sqrt{L} \le 1$ 时，$\sqrt{\kappa}$ 这一依赖无法进一步改善，即使采用热启动也是如此（命题 B.1）。

定理 2说明了放宽海森正则性的障碍。对于每个给定的步长$0 < h \le 1$，它构造了一个满足 $m = \kappa^{-1}$、$L = 1$ 且海森矩阵不连续的强凸势函数，以及一个密度至多为 π 的初始分布；对于该初始分布，满足（2.12）的持续时间分布至少需要数量级为 $\kappa^{2/3}$ 的期望梯度计算次数（忽略对数因子），才能达到全变差误差 $1/4$。该定理并未断言上界中的指数或L_H依赖关系是最优的。

1.1. 贡献。 定理 1 首次给出了超越高斯目标分布、针对具有随机化积分时间的 Metropolis 调整的 HMC 的平方根条件数界；此前唯一同类界是 [3]的高斯分析。

第一项贡献，即引理 6.9，将未调整链的相对熵收缩转化为调整后链的受限导通率界。相对熵收缩是 [17] 精确随机 HMC 以及[4] 未调整 OBABO 的平方根界背后的机制；Metropolis 步骤无法保持这一性质，而 [4] 通过一种需要从点初始状态获得熵估计的耦合来规避这一问题。命题6.8建立了未调整提议的收缩性；它是 [17, 引理 3], 的离散时间版本，其中 [4] 的自由输运不等式控制动量踢之间的速度熵，而海森正则性控制累积能量误差。我们将其应用于单个初始分布，即 Π 在给定 $\{X \in A, H \le R\}$条件下的条件分布；该分布相对于 Π 的相对熵接近 $\log(1/\pi(A))$。如果提议以高概率停留在 A 中，它几乎会保留全部熵，因此收缩性迫使退出概率具有一个正的下界；而能量控制使这些退出很可能被接受。该引理针对一般的 Metropolis 调整核，其提议满足此类收缩性；结合谱下界，它通过 [2]的弱庞加莱不等式给出混合性。

第二项贡献在于发现，由于 Verlet 映射保持相空间体积，积分时间分布对调整后的核的狄利克雷形式的影响仅通过其特征函数体现（引理 5.2）。因此，受限导通率仅针对三角分布参考分布进行了证明，并可推广到所有具有显式拒绝误差的非共振分布。

同一表示还给出了谱下界。这就是为什么定义 2.1 中的傅里叶条件是自然假设。

第三项贡献是定理2，其证明指出，阻碍因素是数值不稳定性，而非共振：Verlet 轨道在构造的势函数的快速坐标上扩张，由此导致的拒绝阻碍了沿其慢坐标的运动。

1.2. 定理 1 证明概要。证明分为三部分。受限导通率与混合。对于 Borel 集 $A \subseteq \mathbb{R}^d$，离开 A 的平稳流为 $Q_P(A, A^c) = \int_A P(x, A^c)\, \pi(\mathrm{d}x)$。命题 A 基于 Andrieu、Lee、Power 和 Wang 的弱庞加莱不等式 [2],，适用于满足以下条件的可逆核：其流满足 $Q_P(A, A^c) \geq \phi_P \pi(A)$，只要 $s \leq \pi(A) \leq 1/2$，且其谱下界大于 $-1 + \beta_P$。对于依赖于 M、ε 的适当阈值 s，该命题给出从 L^2-热启动出发，在 $\mathcal{O}((\beta_P \phi_P^2)^{-1} \log(2M/\varepsilon))$ 次转移后实现混合。

参考分布的受限导通率。对于$[0, 4/\sqrt{m}]$ 上采用相同舍入方式的递减三角形持续时间，命题 6.8 证明：对于具有有界能量和有限熵、且满足 $\mathbb{E}[V \mid X] = 0$的初始分布，位置提议的相对熵至多为初始联合相对熵的固定比例，再加上离散化误差。证明借鉴了[17, 引理 3]: 引理6.6 控制平方 Wasserstein 距离沿自由漂移的二阶变分，而引理 6.7 将其与动量踢步及 U 的凸性结合起来。引理 6.9 利用引理 6.2 的能量界控制接受率，并将此估计转化为调整后的核的受限导通率界。

转移与谱下界。引理 5.2 利用一次 Verlet 步所诱导的酉算子、对密度水平集的积分以及拒绝项，表示调整后的核的狄利克雷形式，因此对积分时间分布进行傅里叶比较，便可比较它们的狄利克雷形式，并得到显式的拒绝误差。我们将持续时间条件化在 $\tau \leq \tau_*$上，其中 τ_*的量级为 $m^{-1/2}$，且保留概率有正的下界，因此物理积分时间至多为 $\tau_* + h$，能量估计可控制此误差。引理 5.3 给出了与三角形参考分布的比较，且对所有可行步长均一致；在 $\mathbf{1}_A$ 处取值，并将条件核纳入完整混合分布，即得命题 C。对完整持续时间分布另行使用傅里叶界，可得命题 B的谱下界。

1.3. 与 Persistence Pays 的关系。 配套论文 [4] 针对 Metropolis 校正的 OBABO（一种带有部分动量刷新和非可逆更新的接受/拒绝均匀随机变量的动能朗之万方案），在具有不同对数因子的情况下，获得了梯度代价（1.1）的两个主导项 [19]。二者的本质区别在于可逆性。 [4] 中的采样器不可逆，因此其分析通过耦合展开：两个校正链副本通过各自未校正链的周期端点进行耦合，这需要从点初始状态出发的熵估计，以及对每个周期内拒绝次数的控制。

本文研究的采样器相对于 π可逆。这使谱方法成为可能：受限导通率和谱下界通过弱庞加莱不等式给出混合性质，而引理 5.2 中的傅里叶比较则将递减三角分布的导通率传递给每个非共振分布。熵估计仅应用于 Π 在给定一组位置和一个能量壳的条件下，并不需要从点初始状态出发控制未调整链。两篇论文都采用了自由输运熵不等式，以及速度 Verlet 的能量误差的精确表达式（引理 6.2）。

Oberdörster [20] 随后将 [4] 的熵论证与 Chen、Gatmiry 和 Jiang [8] 的能量误差分析以及针对高概率区域的稳定性结果相结合，证明了对于采用持久均匀随机变量的校正的 OBABO，在海森矩阵的 Frobenius 范数上满足海森矩阵利普希茨连续性假设时，从高概率集合中的确定性初始状态出发，其梯度复杂度为 $(\sqrt{\kappa} + L_H^{1/2} m^{-3/4})\, d^{1/4}$阶，并且该结论是一致成立的。本文保留谱范数

假设(A3) 以及热启动。其维度依赖性来自能量控制引理，其中可以替换为 [20] 中所使用的同类 Frobenius 范数估计，而不会影响对积分时间分布的处理。

1.4. 相关工作。人们早已认识到哈密顿蒙特卡罗对积分时间的敏感性。Mackenzie [16]提出时长随机化以避免共振；Neal [18,第 5.4.2 节3] 讨论了这一补救方法，以及在较短轨迹与不必要地过长的轨迹之间如何选择。Hoffman 和 Gelman [10] 提出的 No-U-Turn采样器根据 U 形转向判据在局部选择轨迹长度。Bou-Rabee、CARPENTER 和 Marsden [6] 提出的 GIST 框架通过一个通用的辅助变量构造，将随机化 HMC 与无 U 形转向采样器纳入其中，并将无 U 形转向采样器解释为一种局部自适应的随机化 HMC。WALNUTS[5] 将局部自适应扩展到轨迹上的蛙跳步长。

对于持续时间服从指数分布的精确流随机化HMC，Bou-Rabee 和 Sanz-Serna[7, 定理 39] 通过 Foster–Lyapunov 漂移条件、小化条件和 Harris 定理确立了几何遍历性。Lu 和 Wang [15] 通过变分亚强制性论证获得了显式的 L^2 收敛速率。对于具有 Lipschitz连续梯度的 m强凸势函数，当刷新率为该量级时，其物理时间下的收敛速率为 $\sqrt{m}$ 量级。这些精确流采样结果并未考虑数值积分及其梯度代价。

我们离散化的熵论证与概率测度空间中的优化有关，这方面的研究由 Wibisono [25] 发展而来。Fu 和 Wibisono [9] 证明了通过随机化积分时间并将速度重置为零，可以实现加速哈密顿下降。Wang、Srinivasan、Fu、Mitra、Wilson 和 Wibisono [24] 利用哈密顿轨迹上的加权平均实现了确定性加速。两项工作都包含具有加速复杂度保证的离散优化算法。Mitra、Srinivasan、Wang 和 Wibisono [17, 在附录 B]中将平均论证推广到相对熵，并以高斯动量刷新取代零速度重启。

对于高斯目标分布，Wang 和 Wibisono [23] 证明，采用积分时间的确定性切比雪夫调度时，精确 HMC 可以实现加速。Apers、Gribling 和 Szilágyi [3]分析了采用 Verlet 方法的 Metropolis 调整的随机化 HMC，并以梯度代价 $O(\sqrt{\kappa}\, d^{1/4})$ 达到全变差误差 ε，其中还需乘上额外的对数因子。他们的证明使用了 Hanson–Wright 集中不等式和受限导通率；他们还利用未调整哈密顿蒙特卡罗构造了热启动。他们的维数依赖比我们更优。

Lee、Shen 和 Tian [13] 通过梯度集中和受限导通率，证明每次转移采用一个蛙跳步的 Metropolis 化哈密顿蒙特卡罗可在 κd 次迭代内实现混合（相差对数因子）。他们的下界 [14, 定理 4] 指出，对于每个固定步长以及每次转移采用固定数量 j 个蛙跳步的情形，存在一个高斯分布目标，其弛豫时间至少为常数倍的 $\kappa\sqrt{d}/(j\sqrt{\log d})$，其中 $\kappa = \mathcal{O}(d^4)$。该构造中的共振模态取决于固定持续时间，因此这一界不适用于随机化时长。

1.5. 论文结构。 第 2 节给出假设、采样器、持续时间分布及主要结果。第 3 节由三个命题证明定理 1，这些命题分别在第 4、5 和 6节中得到证明：混合判据、谱下界和受限导通率界。第 7 节证明定理 2。附录 A 包含第 6.2节所用的正则化论证。附录 B 和 C 分别给出高斯下界，以及三种持续时间分布的显式傅里叶常数。

1.6. 形式化验证。 定理 1 和 2 以及命题 B.1的陈述及其所依据的定义，均已完成形式化，并在 Lean 4 中基于 Mathlib 库对其证明进行了检查；除 Lean 内核的三个标准公理外，未使用其他公理。该形式化遵循本文给出的证明，包括明确常数，并涵盖命题 A、

B、 C 和6.8 以及引理 5.1、 5.2、 5.3、 5.4、 6.1、 6.9、 7.1 和 7.2。文献中的两个结果作为假设条件参与定理 1 的证明：Talagrand 输运熵不等式（引理 6.3）以及 [4] 的插值估计（引理 6.4）。Lean源代码、将每条形式化陈述与本文对应内容进行匹配的表格，以及构建日志，均可在以下网址获取： https://github.com/nawafbourabee/breaking-resonance-lean （存档于 https://doi.org/10.5281/zenodo.24483261）。

2. 设置与主要结果。

2.1. 假设与记号。 令 $U\colon \mathbb{R}^d \to \mathbb{R}$ 为目标势函数，其中 $d \geq 1$。记$|\cdot|$为欧几里得范数，并记 $\|M\|_{\mathrm{op}} = \sup_{|z|=1} |Mz|$为其在矩阵上诱导的算子范数。我们根据需要作出以下假设。

(A1) U是 m-强凸的，其中 $m > 0$。(A2) U可微，且 ∇U是 L-Lipschitz连续的，其中 $L > 0$。(A3) U是二阶连续可微的，并且对于每个 x， $y \in \mathbb{R}^d$， $\|\nabla^2 U(x) - \nabla^2 U(y)\|_{\mathrm{op}} \leq L_H |x - y|$，其中 $L_H \geq 0$。

记 $\kappa = L/m$ 为条件数。

设 $k \geq 1$，并设 ν， μ 为 $\mathbb{R}^k$上的概率测度。我们用 $k = d$ 表示位置分布，用$k = 2d$ 表示位置和速度的联合分布。它们的相对熵为

$$\mathrm{KL}(\nu\|\mu) = \begin{cases} \displaystyle\int_{\mathbb{R}^k} f(z)\log f(z)\,\mu(\mathrm{d}z), & \nu \ll \mu, \\ +\infty, & \text{otherwise}, \end{cases}$$

其中，在绝对连续的情形下， $f = \mathrm{d}\nu/\mathrm{d}\mu$ 为 Radon–Nikodym导数，且 $0 \log 0 = 0$。它们的全变差距离为

$$\mathrm{TV}(\nu,\mu) = \sup_{A} |\nu(A) - \mu(A)|,$$

即对 Borel 集 $A \subseteq \mathbb{R}^k$取上确界。它们的 2-Wasserstein距离为

$$W_2(\nu,\mu) = \left(\inf_{\gamma \in \Gamma(\nu,\mu)} \int_{\mathbb{R}^k \times \mathbb{R}^k} |x - y|^2 \,\gamma(\mathrm{d}x, \mathrm{d}y) \right)^{1/2},$$

其中， $\Gamma(\nu,\ \mu)$ 是 $\mathbb{R}^k \times \mathbb{R}^k$ 上以 ν为第一边缘分布、以 μ为第二边缘分布的概率测度集合。

记 $Z = \int_{\mathbb{R}^d} e^{-U(x)}\mathrm{d}x$，并假设其有限，定义

$$\pi(\,\mathrm{d}x) = Z^{-1} e^{-U(x)}\,\mathrm{d}x, \qquad \gamma_d = N(0, I_d), \qquad \Pi = \pi \otimes \gamma_d. \tag{2.1}$$

2.2. *Metropolis* 调整的随机化 *HMC*。 在 (A1)下，记U的唯一极小值点为 x_*。固定步长 $h > 0$ 和持续时间分布 ρ。哈密顿量为

$$H(x,v) = U(x) - U(x_*) + \tfrac{1}{2}|v|^2. \tag{2.2}$$

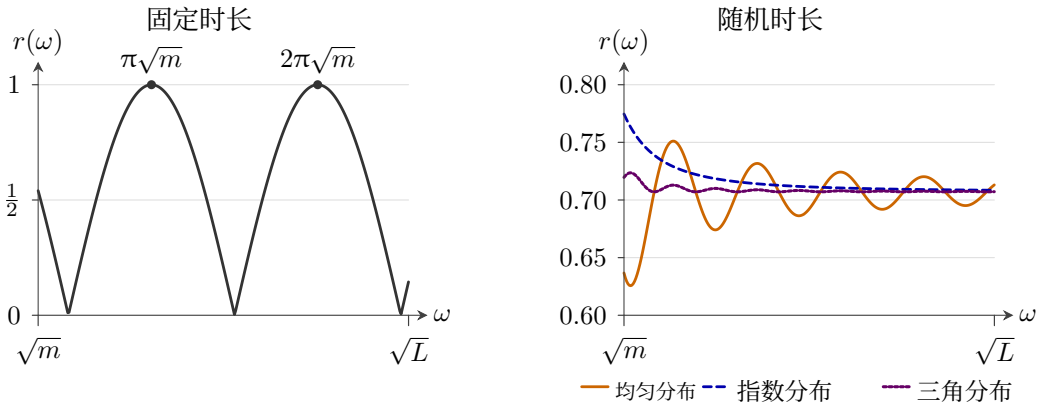


图1。精确高斯哈密顿蒙特卡洛在 $[\sqrt{m},\sqrt{L}]$、$L/m=64$ 上的均方根收缩因子 $r(\omega)$：固定时长（左）和随机时长（右）。

算法 2.1（Metropolis 调整的随机化 HMC）。给定当前的位置 X，按如下方式返回下一个位置 X^+。

1. **时长和动量随机化**。彼此独立地抽取 $V\sim\gamma_d$、$\tau\sim\rho$，以及一个公平的伯努利分布 B，并且它们均与 X 独立。设积分时间

$$J=\lfloor\tau/h\rfloor+B. \tag{2.3}$$

2. **Verlet 积分**。给定 $(x_0,v_0)=(X,V)$，执行以下更新 J 次

$$\begin{aligned} w_k &= v_k-\tfrac{h}{2}\nabla U(x_k),\\ x_{k+1} &= x_k+hw_k,\\ v_{k+1} &= w_k-\tfrac{h}{2}\nabla U(x_{k+1}). \end{aligned} \tag{2.4}$$

记 $(Q,W)=(x_J,v_J)$。

3. **Metropolis 检验**。独立地抽取 $Z_{\mathrm{acc}}\sim\mathrm{Unif}(0,1)$，并计算

$$\alpha(X,V;Q,W)=1\wedge\exp\{H(X,V)-H(Q,W)\}. \tag{2.5}$$

若 $Z_{\mathrm{acc}}\le\alpha(X,V;Q,W)$，则返回 $X^+=Q$；否则返回 $X^+=X$。

位置核。设 Φ_h 为 (2.4) 中的 Verlet 映射，并记 $\Phi_h^{\circ j}$ 为其 j 重复复合，其中 $\Phi_h^{\circ 0}$ 是 $\mathbb{R}^{2d}$ 上的恒等映射。对于 $z=(x,v)$ 和 $j\ge 0$，令 $Q_j(z)$ 为 $\Phi_h^{\circ j}(z)$ 的位置分量，并设 $\alpha_j(z)=1\wedge\exp\{H(z)-H(\Phi_h^{\circ j}(z))\}$。对于可测集 $A\subseteq\mathbb{R}^d$，位置核为

$$K_j(x,A)=\int_{\mathbb{R}^d}\left[\alpha_j(x,v)\mathbf{1}_A(Q_j(x,v))+(1-\alpha_j(x,v))\mathbf{1}_A(x)\right]\gamma_d(\mathrm{d}v),$$

$$P(x,A)=\sum_{j\ge 0}\mathbb{P}(J=j)K_j(x,A).$$

此处 K_j 使用 j Verlet 步，且 P 是算法 2.1的位置核。特别地，从 n 经过 ν_0 次迭代后，ν_0P^n的分布为 。

2.3. 非共振积分时间分布。采样器从分布 ρ中抽取持续时间 τ。使用 Verlet 步长 h 时，它将 τ 转换为整数值的积分时间 J，以 为单位进行度量，

物理积分时间为 Jh。我们首先使用精确高斯哈密顿蒙特卡洛解释选择 ρ 的原因，然后定义决定 J 的舍入规则。

对于精度矩阵为 Ω 的中心高斯目标分布，精确哈密顿蒙特卡罗会抽取一个新速度$v \sim N(0,\ I_d)$，依据哈密顿方程 $\dot{x} = v, \dot{v} = -\Omega x$ 运行，持续时间为 τ，且该时长独立于位置和速度，然后返回终点位置。该流可精确计算且保持哈密顿量不变，因此每个提议都会被接受。

精确高斯哈密顿蒙特卡罗中的共振。图1展示了时长随机化对精确高斯哈密顿蒙特卡罗中收缩的影响。对于高斯目标分布，设 $\lambda_i \in [m, L]$ 为一个精度特征值，并记 $\omega_i = \sqrt{\lambda_i}$。从不同位置出发，使用相同的初始动量和相同的持续时间 τ，生成两条精确哈密顿蒙特卡洛轨迹。若 Δ_i 是它们在模态 i上的初始位置差，则一次转移后的差为$\Delta_i^+ = \text{余弦}(\omega_i\tau)\Delta_i$。对于固定时长 s，绝对差值之比为 $|\text{余弦}(\omega_i s)|$，在共振 $\omega_i s \in \pi\mathbb{Z}$处等于一。

对于 $\Delta_i \neq 0$，令 $r(\omega_i)$ 表示 Δ_i^+ 除以 $|\Delta_i|$的均方根。由于两条轨迹具有相同的初始动量和持续时间，$\Delta_i^+ = \cos(\omega_i\tau)\Delta_i$ 仅随 $\tau \sim \rho$随机变化，因此

$$r(\omega_i)^2 = \mathbb{E}_\rho[\cos^2(\omega_i\tau)] = \frac{1 + \operatorname{Re}\widehat{\rho}(2\omega_i)}{2},$$

其中 $\widehat{\rho}(\omega) = \mathbb{E}_\rho[e^{i\omega\tau}]$是 ρ的特征函数，因此 $\operatorname{Re}\widehat{\rho}(\omega) = \mathbb{E}_\rho\cos(\omega\tau)$。当 $r(\omega_i) < 1$时，模态 i 中的位置差在均方意义下收缩，也就是说，当 $\operatorname{Re}\widehat{\rho}(2\omega_i) < 1$时。加倍后的频率 $2\omega_i = 2\sqrt{\lambda_i}$ 均位于 $[2\sqrt{m}, 2\sqrt{L}]$中。要求 $\operatorname{Re}\widehat{\rho} \le 1 - \delta_\rho$在整个半直线 $[2\sqrt{m}, \infty)$ 上成立，可使每个模态都以一个与 L无关、因而也与 κ无关的因子收缩；这就是定义 2.1 中的第二个条件。

定义 2.1（非共振持续时间分布）。设 $b_\rho > 0$ 和 $0 < \delta_\rho \le 1$。定义在 $(0, \infty)$ 上的概率分布 ρ，若满足以下条件，则称其具有常数 b_ρ, δ_ρ的非共振性质：

$$\mathbb{E}_\rho[\tau] \le \frac{b_\rho}{\sqrt{m}}, \qquad \sup_{\omega \ge 2\sqrt{m}} \operatorname{Re}\widehat{\rho}(\omega) \le 1 - \delta_\rho. \tag{2.6}$$

均方根收缩因子 $r(\omega_i)$ 在每种模态下均不超过 $\sqrt{1 - \delta_\rho/2}$。这两个条件迫使 $\mathbb{E}_\rho\tau$ 的阶为 $m^{-1/2}$：第一个条件给出 $\mathbb{E}_\rho\tau \le b_\rho/\sqrt{m}$；第二个条件在 $\omega = 2\sqrt{m}$处，结合 $1 - \text{余弦}\, x \le |x|$，给出 $\delta_\rho \le \mathbb{E}_\rho[1 - \text{余弦}(2\sqrt{m}\,\tau)] \le 2\sqrt{m}\,\mathbb{E}_\rho\tau$。

例子。以下三个分布满足定义 2.1，其中 $b_\rho = 2$ 且 $\delta_\rho = 3/4$。最后一列使用 $u = \omega/\sqrt{m} \ge 2$。

τ的分布	$\mathbb{E}_\rho\tau$	$\widehat{\rho}$的实部(ω)
$[0, 2/\sqrt{m}]$上的均匀分布	$1/\sqrt{m}$	$\sin(2u)/(2u) \le 1/4$
指数分布	$1/\sqrt{m}$	$1/(1+u^2) \le 1/5$
$[0, 4/\sqrt{m}]$上的递减三角分布	$4/(3\sqrt{m})$	$2(1 - \text{余弦}(4u))/(4u)^2 \le 1/16$

随机舍入。向下舍入会将 $[jh, (j+1)h)$ 中的每个持续时间都映射到相同的积分时间 j。对于二次势函数，j Verlet步在海森矩阵的每个特征方向上都表现为旋转；如果 j 步在某个特征方向上恰好完成一个周期，那么此区间内的每个持续时间都会使该坐标返回初始值。我们改为以相等的概率在 $\lfloor \tau/h \rfloor$ 和$\lfloor \tau/h \rfloor + 1$ 之间进行选择。额外的一步会推进相位。用傅里叶意义下的表述，这一步会将 $\lfloor \tau/h \rfloor$ 的特征函数乘以 $(1 + e^{i\theta})/2$，其模在 $2\pi\mathbb{Z}$之外严格小于一。对于间隙为 δ_ρ 的非共振持续时间分布，我们采用满足 $h^2 m \le \delta_\rho/16$ 的步长，并将由此得到的 J 分布称为非共振积分时间分布。图 2 展示了三角形持续时间分布下的舍入规则。

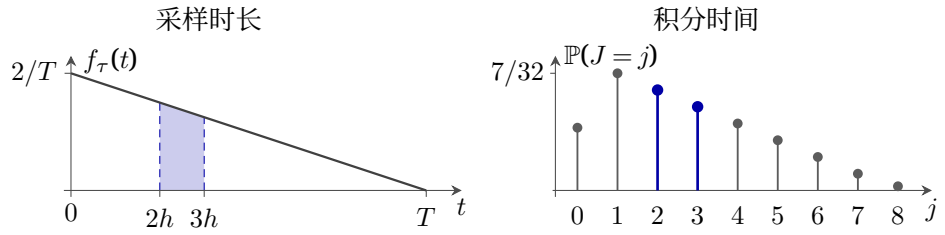


图 2. $[0,T]$ 上的三角形时长和 J 的积分时间（针对 $h=T/8$）。在阴影区间内抽取一次，即以相等概率得到 $J=2$ 或 $J=3$。

2.4. 混合与梯度复杂度。 若初始位置分布 ν_0 是一个 L^2-热启动，且其界为 $M\ge 1$，则其密度 $f_{\mathrm{init}}=\mathrm{d}\nu_0/\mathrm{d}\pi$ 满足

$$\|f_{\mathrm{init}}\|_{L^2(\pi)}\le M. \tag{2.7}$$

定理 1（热启动下的混合）。假设 $U\colon\ \mathbb{R}^d\to\mathbb{R}$ 满足 *(A1)* 至 *(A3)*，ν_0是一个 L^2-热启动，其界为 *(2.7)* 中的 M，且 ρ 是一个非共振持续时间分布，其常数为定义 *2.1* 中的 b_ρ、δ_ρ。令 $\varepsilon\in(0,1]$ 和 $D=1+d+\log(M/\varepsilon)$。存在仅依赖于 b_ρ、δ_ρ 的常数 c_1、$c_2>0$，使得若对于某个整数有 $h=4/(N\sqrt{m})$

$$N\ge c_1\left[\sqrt{\kappa D}+\sqrt{L_H}\,(D/m)^{3/4}\right] \tag{2.8}$$

以及 $n\ge c_2\log(2M/\varepsilon)$，则步长为 h 的位置核 P 满足

$$\mathrm{TV}(\nu_0P^n,\pi)\le\varepsilon. \tag{2.9}$$

c_1 和 c_2 的明确取值见（3.4）。由于 $h=4/(N\sqrt{m})$，条件（2.8）给出 $h\le 4/(c_1\sqrt{LD})$，这比 Verlet 在二次势函数上的线性稳定性条件 $h\sqrt{L}<2$ 更强；证明还要求整个轨迹上的能量误差较小，以控制 Metropolis 拒绝。转移次数与条件数无关，条件数仅影响每次转移所需的梯度计算次数。

推论 1（梯度复杂度）。在定理 *1*的假设下，设 C_{init} 为采样初始分布 ν_0 并设置采样器参数所需的梯度计算次数，设 C_∇ 为初始化及 n 次转移所需的梯度计算总次数。若 N 和 n 是定理 *1*所允许的最小整数，则

$$\mathbb{E}C_\nabla\le\mathbb{E}C_{\mathrm{init}}+1+\mathcal{O}\Big(\left[\sqrt{\kappa D}+\sqrt{L_H}\,(D/m)^{3/4}\right]\log\tfrac{2M}{\varepsilon}\Big), \tag{2.10}$$

其中隐含常数仅取决于 b_ρ 和 δ_ρ。

证明。先计算一次初始梯度。每个 Verlet 步都会重用其起点处的梯度，并在终点处计算一个新梯度。经过 Metropolis 检验后，保留返回位置处的梯度。按此方式重用已存储的梯度，

$$C_\nabla=C_{\mathrm{init}}+1+\sum_{i=1}^{n}J_i. \tag{2.11}$$

随机舍入和定义 2.1 给出

$$\mathbb{E}J_i\le\frac{\mathbb{E}_\rho\tau}{h}+\frac12=\frac{N\sqrt{m}}{4}\mathbb{E}_\rho\tau+\frac12\le\frac{b_\rho N}{4}+\frac12,$$

以及 N 和 n 的取值给出（2.10）。 □

注 2.2（凸性和输运-熵不等式已足够）。定理 1 的证明中，强凸性仅通过 Talagrand 不等式（6.9）以及证明中积分时间 $T \asymp 4/\sqrt{m}$ 的选择发挥作用。其他地方仅用到 U 的凸性：在引理6.7 中用于确定输运项的符号，在尾界（3.7）中也有应用。因此，对于任意凸 U，只要其梯度和海森矩阵满足 Lipschitz连续，且其 Gibbs 分布对某个 6.9）成立，其中 $m>0$，定理 1 仍以相同常数成立；在全文中用此 m 替代强凸常数即可。根据 Holley-Stroock 扰动引理 [11] 和 Otto-Villani 定理 [21],，这也涵盖了有界扰动后的强凸势函数，前提是扰动后仍保持凸性，且梯度和海森矩阵满足 Lipschitz连续。定理 2和命题B.1 中的下界则按原文所述使用强凸性。

2.5. 下界。 对于 $T>0$，考虑满足以下条件的持续时间分布 ρ

$$\mathbb{P}(\tau<t)\le\frac{2t}{T}\quad(t\ge0),\qquad \frac{T}{3}\le\mathbb{E}\tau<\infty. \tag{2.12}$$

第一个条件限制抽取到较短持续时间的概率，第二个条件则为均值设定了下界。在 $m=\kappa^{-1}$ 时，第 2.3 节中的均匀分布和指数分布满足条件（2.12），其中 $T=2\sqrt{\kappa}$；递减三角分布则在 $T=4\sqrt{\kappa}$ 时满足条件；这就是定理 1 中的持续时间尺度，其中 $\mathbb{E}_\rho\tau$ 的阶为 $m^{-1/2}=\sqrt{\kappa}$。

定理 2（固定步长下 Hessian 正则性的必要性）。设定

$$q_*=10^{-3},\qquad d_f=\lceil 6q_*^{-1}\log\kappa\rceil,\qquad d=d_f+2,\qquad \Lambda_\kappa=\log(e+\log\kappa).$$

对于每个 $\kappa\ge64$、$0<h\le1$ 和 $T\in[2\sqrt{\kappa},4\sqrt{\kappa}]$，都存在一个势函数 $U_h:\ \mathbb{R}^d\to\mathbb{R}$ 满足 *(A1)* 和 *(A2)*，其最优常数为 $m=\kappa^{-1}$ 和 $L=1$；并且存在一个 *Borel* 集 A_+，满足 $\pi(A_+)=1/2$，使得以下结论成立。令 $\nu=2\mathbf{1}_{A_+}\pi$，并令 P 为算法 *2.1* 在步长为 h 且持续时间分布满足 *(2.12)* 时的位置核。对于每个整数 $n\ge0$，$TV(\nu P^n,\pi)\le1/4$ 蕴含

$$\mathbb{E}N_\nabla\ge c_3^{2/3}c_4^{1/3}\frac{\kappa^{2/3}}{\Lambda_\kappa^{1/3}}, \tag{2.13}$$

其中，N_∇ 是这 n 次转移中的梯度计算次数，c_3 和 $c_4>0$ 是绝对常数。

由于 $d=d_f+2$ 仅以对数速度增长于 κ 和 $d\nu/d\pi\le2$，(2.13) 排除了对该算法的均匀 $\sqrt{\kappa}$ 梯度界乘以 d 和 $\log\kappa$ 中的任意固定多项式的可能性。

3. 定理1的证明。我们先固定证明中使用的常数和参数，然后陈述一个一般的混合判据（命题A），并建立采样器的谱下界（命题B）和受限导通率下界（命题C）。

固定一个持续时间截断 $\tau_*>0$，使得 $p_S:=\mathbb{P}(\tau\le\tau_*)\ge1-\delta_\rho/4$，并令 $T_*=\tau_*+4/\sqrt{m}$。由马尔可夫不等式可得 $\tau_*=4b_\rho/(\delta_\rho\sqrt{m})$。对于有界持续时间分布，也可以取其支撑集上端点，从而得到 $p_S=1$。

对每个整数 $N\ge16/\sqrt{\delta_\rho}$，令 $h=4/(N\sqrt{m})$。令 μ_J为 J的完整分布，w为其在 $\tau\le\tau_*$条件下的分布，并令 q 为对 $[0,4/\sqrt{m}]$上的递减三角形持续时间应用相同舍入规则所得的分布。均匀地选取 $0<c$、$\beta_P\le1$，使得对每个 $\theta\in[-\pi,\pi]$，

$$1-\operatorname{Re}\widehat{w}(\theta)\ge c\big(1-\operatorname{Re}\widehat{q}(\theta)\big),\qquad 1+\operatorname{Re}\widehat{\mu}_J(\theta)\ge\beta_P. \tag{3.1}$$

这里，帽号表示特征函数。引理 5.3 给出了每种非共振持续时间分布的显式正值选择；引理 5.4 给出了三个示例对应的选择。定义

$$\eta = \frac{c}{600}, \qquad \phi_P = \frac{p_S c}{60}. \tag{3.2}$$

第一个量界定了能量增加，第二个量是受限导通率界。

显式常数。在定理1的设定中，取截断 $\tau_* = 4b_\rho/(\delta_\rho\sqrt{m})$，使得$T_* = C_T/\sqrt{m}$ 其中 $C_T = 4(b_\rho + \delta_\rho)/\delta_\rho$，令

$$s = \frac{\varepsilon^4}{64M^4}, \qquad T = \frac{4}{\sqrt{m}}, \qquad R = 3d\log 2 + 2\log\frac{1}{\eta s}, \tag{3.3}$$

并令 $C_R = 2\log(64/\eta)$。定理1 中的常数为

$$c_1 = \max\Big\{\frac{16}{\sqrt{\delta_\rho}},\ \sqrt{\frac{48}{\eta}}\,\max\big\{\sqrt{C_R},\ \sqrt{C_T}\,C_R^{3/4}\big\}\Big\}, \qquad c_2 = \frac{6}{\beta_P\phi_P^2}. \tag{3.4}$$

由引理 5.3，c 和 β_P 都有仅依赖于 b_ρ、δ_ρ和 $p_S \ge 1 - \delta_\rho/4$ 的正下界，这是由 τ_*的选取保证的。常数 c_2 通过 ϕ_P依赖于 p_S；用这个下界替换p_S后，得到的 η、ϕ_P、c_1 和 c_2 仅依赖于 b_ρ、δ_ρ。下面的证明将通过三个推论使用这些选取，我们现在推导如下：尾界$\Pi(H > R) \le \eta s$，以及条件

$$N \ge \frac{16}{\sqrt{\delta_\rho}}, \qquad h^2(LR + L_H T_* R^{3/2}) \le \frac{\eta}{3}, \tag{3.5}$$

$$n \ge \frac{2}{\beta_P\phi_P^2}\log\frac{8M^2}{\varepsilon^3}. \tag{3.6}$$

在 (A1)条件下，玻尔兹曼-吉布斯分布满足以下尾界

$$\Pi(H > R) \le 2^{3d/2}e^{-R/2} \qquad (R \ge 0). \tag{3.7}$$

为此，将 x_* 平移至原点，并令 $u(x) = U(x) - U(x_*)$。对于 $0 < t \le 1$，由凸性可得 $u(tx) \le tu(x)$，因此 $Z(t) = \int_{\mathbb{R}^d} e^{-tu(x)}\mathrm{d}x$ 满足

$$Z(t) \le \int_{\mathbb{R}^d} e^{-u(tx)}\,\mathrm{d}x = t^{-d}Z(1).$$

因此 $\mathbb{E}_\pi e^{u(X)/2} \le 2^d$。独立动能贡献 $\mathbb{E}_{\gamma_d} e^{|V|^2/4} = 2^{d/2}$。由马尔可夫不等式可得（3.7），结合 R 如（3.3）所示，可得 $\Pi(H > R) \le \eta s$。此外$R \le C_R D$，因为 $\eta \le 1/600$。

对于 N， n 以及 $h = T/N$如定理 1中所示， $\kappa D \ge 1$给出 $N \ge c_1 \ge 16/\sqrt{\delta_\rho}$；界 $R \le C_R D$、$T_* = C_T/\sqrt{m}$ 和 $\sqrt{a} + \sqrt{b} \ge \sqrt{a+b}$给出

$$N^2 \ge \frac{48}{\eta}\Big(C_R\kappa D + L_H C_T C_R^{3/2}\frac{D^{3/2}}{m^{3/2}}\Big) \ge \frac{48}{\eta m}\big(LR + L_H T_* R^{3/2}\big),$$

这是条件（3.5）中的第二个条件，其中 $h^2 = 16/(N^2 m)$；而（3.6）由下式得出：$\log(8M^2/\varepsilon^3) \le 3\log(2M/\varepsilon)$，该式对 $M \ge 1$ 成立。本节余下部分将在（3.5）和（3.6）的条件下证明（2.9）。

对于概率测度 π 和 π-可逆核 P，记 $Pf(x) = \int_{\mathbb{R}^d} f(y)P(x, \mathrm{d}y)$，并且

$$Q_P(A, A^c) = \int_A P(x, A^c)\,\pi(\mathrm{d}x),$$

$$\mathcal{E}_P(f) = \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (f(x) - f(y))^2\, P(x, \mathrm{d}y)\pi(\mathrm{d}x) = \langle f, (I - P)f\rangle_{L^2(\pi)},$$

其中，A 是 $\mathbb{R}^d$的Borel子集，$f \in L^2(\pi)$，而 I 是恒等算子。

命题 A（由受限导通率得到的混合）。设 P 是定义在 $\mathbb{R}^d$上的马尔可夫核，且关于概率测度 π可逆。假设对于某些 $0 < s \le 1/2$ 和$0 < \phi_P$，$\beta_P \le 1$,

$$Q_P(A, A^c) \ge \phi_P \pi(A) \quad \textit{for every Borel } A \textit{ with } s \le \pi(A) \le \tfrac{1}{2}, \tag{3.8}$$

$$\|f\|^2_{L^2(\pi)} - \|Pf\|^2_{L^2(\pi)} \ge \beta_P \mathcal{E}_P(f), \qquad f \in L^2(\pi). \tag{3.9}$$

若概率测度 $\nu_0 \ll \pi$ 具有密度 f_{init}，且$\|f_{\text{init}}\|_{L^2(\pi)} \le M$，那么对于 $0 < \varepsilon \le 1$以及任意整数 $k \ge 0$,

$$\mathrm{TV}(\nu_0 P^k, \pi) \le \frac{\varepsilon}{2} + \frac{1}{2}\left(e^{-k\beta_P\phi_P^2/2}M_\infty + 2sM_\infty^2\right)^{1/2}, \qquad M_\infty = \frac{4M^2}{\varepsilon}.$$

命题A 在第4节中证明。对于该采样器，以下命题证明了（3.9）。

命题 B（谱下界）。假设*(A1)*和*(A2)*。设 ρ 是具有常数 b_ρ、δ_ρ的非共振持续时间分布。设 $N \ge 1$ 为整数，且 $h = 4/(N\sqrt{m})$。若 $h^2 m \le \delta_\rho/16$，则 P 关于 π 可逆，并且，令 β_P 满足（3.1），

$$\|f\|^2_{L^2(\pi)} - \|Pf\|^2_{L^2(\pi)} \ge \beta_P \mathcal{E}_P(f), \qquad f \in L^2(\pi).$$

命题 B 在第5.4节中得证。剩下需要验证受限导通率假设（3.8）。

命题 *C*（采样器的受限导通率）。设 U 满足*(A1)– (A3)*，且 ρ 是常数为 b_ρ、δ_ρ的非共振持续时间分布。按照*(3.1)*选取 τ_*、c，并令 $\eta = c/600$、$\phi_P = p_S c/60$。设 $0 < s \le 1/2$ 和 $R \ge 1$ 满足 $\Pi(H > R) \le \eta s$。若对于整数 $N \ge 16/\sqrt{\delta_\rho}$ 有$h = 4/(N\sqrt{m})$，且

$$h^2(LR + L_H T_* R^{3/2}) \le \eta/3,$$

then

$$Q_P(A, A^c) \ge \phi_P \pi(A) \quad \textit{for every Borel } A \subset \mathbb{R}^d \textit{ with } s \le \pi(A) \le \tfrac{1}{2}.$$

命题 C在第 6.4节中得证。

表1记录了各常数和截断的作用。其余量为表达式 $T_* = \tau_* + 4/\sqrt{m}$、$\eta = c/600$和 $\phi_P = p_S c/60$。

表1：定理/证明中的常数和截断

符号	条件	Role
c	(3.1)	傅里叶比较
β_P	(3.1)	谱下界
p_S	$\mathbb{P}(\tau \le \tau_*)$	保留时长概率
τ_*	$p_S \ge 1-\delta_\rho/4$	时长截断
R	$R \ge 1, \Pi(H>R) \le \eta s$	能量截断

对于这三种持续时间分布，表 2 给出了对每个可容许的 N均有效的选择。它们在引理 5.4中得到了证明。

表 2。三种持续时间分布的选择。

Law	c	β_P	$\sqrt{m}\,\tau_*$	p_S
三角分布	1	1	4	1
均匀分布	$1/2$	$3/4$	2	1
指数分布	$3/5$	1	8	$1-e^{-8}$

定理 1 的证明。我们应用命题 A 的受限导通率准则，其中 s 如 (3.3) 所示。假设 (A1)–(A3)、 ρ 的非共振性以及 $M \ge 1$、 $0 < \varepsilon \le 1$，给出了命题 B 和命题 C 的目标分布及持续时间分布假设条件。根据 (3.3) 中 R 的选取，截断满足 $\Pi(H>R) \le \eta s$，而 (3.5) 是命题 C 中的能量条件。它还给出 $N \ge 16/\sqrt{\delta_\rho}$，因此满足命题 B 所要求的 $h^2 m = 16/N^2 \le \delta_\rho/16$。于是两个命题都适用。引理 5.1 给出可逆性；命题 B 和命题 C 分别给出 (3.9) 和 (3.8)。这些常数满足 $0 < \phi_P$、 $\beta_P \le 1$。

由于 $0 < s \le 1/64$ 和 ν_0 是一个 L^2-热启动，其界为 M，命题 A 的所有假设条件均得到满足。取 $M_\infty = 4M^2/\varepsilon$和 $2sM_\infty^2 = \varepsilon^2/2$，可得

$$\mathrm{TV}(\nu_0 P^n, \pi) \le \frac{\varepsilon}{2} + \frac{1}{2}\left(e^{-n\beta_P \phi_P^2/2} M_\infty + \frac{\varepsilon^2}{2} \right)^{1/2}.$$

右侧至多为 ε，当

$$n \ge \frac{2}{\beta_P \phi_P^2} \log \frac{8M^2}{\varepsilon^3}, \tag{3.10}$$

这正是（3.6）所要求的。 □

4. 由受限导通率得到的混合。 我们证明命题 A。该论证适用于一般的可逆马尔可夫核。对于有界实值函数 f，记 $\mathrm{osc}(f)$ 为其在 π 下的本质上确界与本质下确界之差。

命题A的证明。根据 [2, 注 41 和引理 42], (3.8)，可得弱庞加莱不等式

$$\mathrm{Var}_\pi(f) \le \frac{2}{\phi_P^2} \mathcal{E}_P(f) + 2s\,\mathrm{osc}(f)^2 \tag{4.1}$$

对每个有界实值可测 f都成立。事实上，按照 [2, 定义 34], (3.8)中的记号，$\Phi_P^{\mathrm{W}}(s) \geq \phi_P$。所引用的结果给出了参数 r 处的弱Poincaré系数上界$2[\Phi_P^{\mathrm{W}}(r/2)]^{-2}$；取 $r = 2s$。为满足这些结果关于每个阈值处均有正剖面函数的假设，先将其应用于$(1-\delta)P(x,\mathrm{d}y) + \delta\pi(\mathrm{d}y)$，然后令 $\delta \downarrow 0$。

首先，设 f_0 是相对于 π 的概率密度，且 $0 \leq f_0 \leq M_\infty$。可逆性表明，经过 k 步后的密度为 $f_k = P^k f_0$，而马尔可夫性质给出 $0 \leq f_k \leq M_\infty$。记作$V_k = \|f_k - 1\|^2_{L^2(\pi)}$。由（3.9）得，

$$V_k - V_{k+1} = \|f_k - 1\|^2_{L^2(\pi)} - \|P(f_k - 1)\|^2_{L^2(\pi)} \geq \beta_P \mathcal{E}_P(f_k).$$

利用（4.1）以及 $\operatorname{osc}(f_k) \leq M_\infty$ 可得

$$V_{k+1} \leq (1 - \beta_P \phi_P^2/2) V_k + \beta_P \phi_P^2 s M_\infty^2.$$

由于 $V_0 = \int_{\mathbb{R}^d} f_0^2 \mathrm{d}\pi - 1 \leq M_\infty - 1 \leq M_\infty$，迭代和柯西-施瓦茨不等式给出

$$V_k \leq e^{-k\beta_P\phi_P^2/2} M_\infty + 2sM_\infty^2, \qquad \mathrm{TV}(f_0\pi P^k, \pi) \leq \tfrac{1}{2}\sqrt{V_k}. \tag{4.2}$$

对于 L^2-暖初始分布，按如下方式截断并重新归一化 f_{init}。

$$M_{\mathrm{cap}} = \frac{2M^2}{\varepsilon}, \qquad f_{\mathrm{clip}} = \min\{f_{\mathrm{init}}, M_{\mathrm{cap}}\}, \qquad Z_{\mathrm{cap}} = \int_{\mathbb{R}^d} f_{\mathrm{clip}}\,\mathrm{d}\pi, \qquad f_0^{\mathrm{cap}} = \frac{f_{\mathrm{clip}}}{Z_{\mathrm{cap}}}.$$

L^2 假设给出

$$1 - Z_{\mathrm{cap}} = \int_{\mathbb{R}^d} (f_{\mathrm{init}} - M_{\mathrm{cap}})_+ \,\mathrm{d}\pi \leq \frac{M^2}{M_{\mathrm{cap}}} = \frac{\varepsilon}{2}, \tag{4.3}$$

$$\mathrm{TV}(f_{\mathrm{init}}\pi, f_0^{\mathrm{cap}}\pi) \leq 1 - Z_{\mathrm{cap}} \leq \varepsilon/2, \tag{4.4}$$

$$0 \leq f_0^{\mathrm{cap}} \leq \frac{M_{\mathrm{cap}}}{1 - \varepsilon/2} \leq \frac{4M^2}{\varepsilon} =: M_\infty. \tag{4.5}$$

对于 (4.3)，使用 $(f_{\mathrm{init}} - M_{\mathrm{cap}})_+ \leq f_{\mathrm{init}}^2 / M_{\mathrm{cap}}$。对于 (4.4)，两个概率密度均支配公共次密度 f_{clip}，其质量为 Z_{cap}。将 (4.2) 应用于 f_0^{cap}，并结合 (4.4) 以及 P下的全变差收缩，即可得到所述界。 □

5. 傅里叶比较与谱下界。 三角分布 q 为熵论证提供参考核。条件分布 w 限制轨迹长度，从而能够控制能量误差。完整分布 μ_J 定义采样器，并用于建立其谱下界。

5.1. 可逆性。

引理 5.1（可逆性）。 假设 U 连续可微且满足 $Z < \infty$。那么，每个 K_j 都关于 π 可逆，核 K_j 的任意与状态无关的混合也同样如此。

证明。 H中的加法常数可以任意选取。将 $\mathsf{Kick}_s(x,v) = (x, v - s\nabla U(x))$ 记作踢步，将 $\mathsf{Drift}_s(x,v) = (x + sv, v)$ 记作漂移。二者都是保持勒贝格体积不变的可逆剪切变换。Verlet 更新 (2.4) 为 $\Phi_h = \mathsf{Kick}^{h/2} \circ \mathsf{Drift}_h \circ \mathsf{Kick}^{h/2}$。对于动量翻转 $F(x,v) = (x,-v)$，

$$F \circ \mathsf{Kick}_s \circ F = \mathsf{Kick}_{-s}, \qquad F \circ \mathsf{Drift}_s \circ F = \mathsf{Drift}_{-s}.$$

因此，踢步与漂移的对称顺序给出 $F\circ\Phi_h\circ F=\Phi_h^{-1}$。由此可知，对于每个整数 $j\ge 0$，$\Psi_j=F\circ\Phi_h^{\circ j}$ 都是保体积对合。

令 p_Π 为 Π 的勒贝格密度，并定义

$$a_j(z)=p_\Pi(z)\alpha_j(z)=\min\{p_\Pi(z),p_\Pi(\Phi_h^{\circ j}(z))\}=\min\{p_\Pi(z),p_\Pi(\Psi_j(z))\}.$$

最后一个等式成立，是因为动量翻转保持 H 不变。它也保持位置不变，因此 $Q_j(z)$ 是 $\Psi_j(z)$ 的位置分量。对于可测的 $A,\ A'\subseteq\mathbb{R}^d$，$\int_A K_j(x,\ A')\,\pi(\mathrm{d}x)$ 的接受部分为

$$\int_{\mathbb{R}^{2d}}\mathbf{1}_A(x)\mathbf{1}_{A'}(Q_j(z))a_j(z)\,\mathrm{d}z.$$

代换 $z\mapsto\Psi_j(z)$ 交换两个位置参数，并保持 $a_j(z)\mathrm{d}z$ 不变，因为 $a_j(\Psi_j(z))=a_j(z)$。拒绝部分支撑在对角线上。因此，$\pi(\mathrm{d}x)K_j(x,\ \mathrm{d}y)$ 是对称的。因此，混合核 P 关于 π 可逆。□

5.2. 校正核的比较。对于定义在积分时间上的分布 $w=(w_j)_{j\ge0}$，记 $P_w=\sum_j w_jK_j$，并令 $\widehat{w}(\theta)=\sum_j w_je^{ij\theta}$。对于实数 $f\in L^2(\pi)$，记 $f^\sharp(x,\ v)=f(x)$，并定义

$$\mathrm{Rej}_w(f)=\sum_j w_j\int_{\mathbb{R}^{2d}}(p_\Pi(z)-a_j(z))f^\sharp(z)^2\,\mathrm{d}z.$$

引理 5.2（调整后的核的比较）。设 w 和 q 是定义在非负整数上的分布。假设对于每个 $\theta\in[-\pi,\pi]$，都有

$$1-\mathrm{Re}\,\widehat{w}(\theta)\ge c\big(1-\mathrm{Re}\,\widehat{q}(\theta)\big),\qquad \mathrm{Re}\,\widehat{w}(\theta)\ge-1+\beta, \tag{5.1}$$

其中 $c>0$ 和 $0<\beta\le1$。仅第二个不等式便推出 $P_w\succeq(-1+\beta)I$ 在 $L^2(\pi)$ 上成立，而仅第一个不等式便推出

$$\langle f,(I-P_w)f\rangle_\pi\ge c\langle f,(I-P_q)f\rangle_\pi-\mathrm{Rej}_w(f). \tag{5.2}$$

证明。在相空间的勒贝格分布 L^2下，算子 $\mathcal{U}g=g\circ\Phi_h$ 是酉的。令 $S_w=\sum_j w_j(\mathcal{U}^j+\mathcal{U}^{-j})/2$。对于 $f\in L^2(\pi)$ 和 $u>0$，令 $g_u=f^\sharp\mathbf{1}_{\{p_\Pi\ge u\}}$。在此超水平集上，$1\le p_\Pi/u$，因此

$$\|g_u\|_{L^2(\mathrm{d}z)}^2\le u^{-1}\int_{\mathbb{R}^{2d}}|f^\sharp(z)|^2p_\Pi(z)\,\mathrm{d}z=u^{-1}\|f\|_{L^2(\pi)}^2.$$

特别地，谱定理适用于 g_u。恒等式

$$a_j(z)=\int_0^\infty\mathbf{1}_{\{p_\Pi(z)\ge u\}}\mathbf{1}_{\{p_\Pi(\Phi_h^{\circ j}(z))\ge u\}}\,\mathrm{d}u$$

将二次型中接受的部分写成对密度水平集的积分。加上拒绝部分可得

$$\langle f,P_wf\rangle_\pi=\int_0^\infty\langle g_u,S_wg_u\rangle\,\mathrm{d}u+\mathrm{Rej}_w(f). \tag{5.3}$$

事实上，对每个 j的积分都是接受贡献 $\int_{\mathbb{R}^{2d}}a_j(z)f^\sharp(z)f^\sharp(\Phi_h^{\circ j}(z))\,\mathrm{d}z$。根据 $2|ab|\le a^2+b^2$、关于 a_j的两个界以及体积保持，它绝对可积，绝对积分至多为 $\|f\|_{L^2(\pi)}^2$。此外，根据托内利定理，$\int_0^\infty\|g_u\|_{L^2(\mathrm{d}z)}^2\,\mathrm{d}u=\|f\|_{L^2(\pi)}^2$。

关于 $\mathcal{U}$ 的谱定理将 (5.1) 转化为 $I-S_w\succeq c(I-S_q)$ 和 $S_w\succeq(-1+\beta)I$。后者与 $\mathrm{Rej}_w(f)\ge0$ 共同给出了所声称的谱下界。从 $\|f\|_{L^2(\pi)}^2$ 中减去 (5.3)，并将同一恒等式与 q 一起使用，即得到 (5.2)，因为 $c\,\mathrm{Rej}_q(f)\ge0$。□

5.3. 舍入后的傅里叶界。 作为比较，使用递减三角密度$2(T-t)/T^2$在 $[0,T]$上。对于 $T=Nh$，使用算法 2.1 中的随机舍入，并采用该密度。区间 $[jh,(j+1)h)$ 的概率为 $[2(N-j)-1]/N^2$，在$J=j$和 $J=j+1$之间等分。因此，概率 $q_j=\mathbb{P}(J=j)$ 为

$$q_0=\frac{2N-1}{2N^2},\qquad q_j=\frac{2(N-j)}{N^2}\quad(1\le j<N),\qquad q_N=\frac{1}{2N^2}. \tag{5.4}$$

写出 $P_{\mathrm{ref}}=P_q$ 的核。定理 1 的证明 $T=4/\sqrt{m}$ 使用 P_{ref}，并将其与给定分布 ρ对应的采样器的核进行比较。

引理 5.3（舍入后的傅里叶界）。设 ρ 是具有常数 b_ρ、δ_ρ的非共振对象。取 $\tau_*>0$满足 $\mathbb{P}(\tau>\tau_*)\le\delta_\rho/4$，并令 w、q、μ_J如 (3.1) 中所示。假设 $h=4/(N\sqrt{m})$ 且 $h^2m\le\delta_\rho/16$。置

$$S=\sqrt{m}\,\tau_*,\quad u_0=\frac{1}{S+1},\quad k=2\lceil S+1\rceil,\quad \zeta=\frac{\delta_\rho}{4k^2},\quad c_0=\min\left\{\frac{\delta_\rho}{64},\zeta,1-\cos(\zeta/2)\right\}.$$

对于完整分布，令

$$k_\rho=2\lceil 2(b_\rho+1)\rceil,\qquad \zeta_\rho=\frac{\delta_\rho}{2k_\rho^2}.$$

于是 (3.1) 成立，其中 $c=c_0/8$ 和 $\beta_P=1-$余弦$(\zeta_\rho/2)$均为正数。

证明。令 $Y=\sqrt{m}\,\tau$。省略的质量至多为 $\delta_\rho/4$。该省略的质量对 $\mathbb{E}[1-$余弦$(uY)]$的贡献至多为 $\delta_\rho/2$。因此，在条件分布下，只要 $|u|\ge2$，就有 $\mathbb{E}[1-$余弦$(uY)]\ge\delta_\rho/2$。对于下面的条件估计，期望均取自 $\tau\le\tau_*$。令$u_0=1/(S+1)$，并令 k 为满足 $ku_0\ge2$ 的最小偶整数。对于 $|u|\ge u_0$，不等式

$$1-\cos(kx)\le k^2(1-\cos x),\qquad 1-\cos(kx)\le k^2(1+\cos x)$$

给出 $1\pm\mathbb{E}$余弦$(uY)\ge\delta_\rho/(2k^2)$。第二个不等式可由第一个将 x 平移 π 得到。

我们首先证明

$$1-\operatorname{Re}\widehat{w}(\theta)\ge c_0\min\{\theta^2/(h^2m),1\},\quad -\pi\le\theta\le\pi,$$

通过考虑三个频率范围来证明。步长条件和 $\delta_\rho\le1$ 蕴含 $h\sqrt{m}\le1/4$。对于小频率，我们使用 J的二阶矩下界。如果 $|\theta|\le u_0h\sqrt{m}$，那么 $h\sqrt{m}\,J\le Y+h\sqrt{m}\le S+1$，因此 $|\theta J|\le1$。由于 $1-$余弦$(2Y)\le2Y^2$，$u=2$处的间隙给出 $\mathbb{E}Y^2\ge\delta_\rho/4$。此外，$|h\sqrt{m}\,J-Y|\le h\sqrt{m}$蕴含 $Y^2\le2h^2mJ^2+2h^2m$。因此$h^2m\,\mathbb{E}J^2\ge\delta_\rho/8-h^2m\ge\delta_\rho/16$。现在对 $|x|\le1$使用 $1-$余弦 $x\ge x^2/4$，得到

$$1-\operatorname{Re}\widehat{w}(\theta)\ge\frac{\delta_\rho}{64}\frac{\theta^2}{h^2m},\qquad \operatorname{Re}\widehat{w}(\theta)\ge\cos1.$$

对于中间频率，非共振会产生一个大于舍入误差的间隙。如果$|\theta|\ge u_0h\sqrt{m}$ 且 $|\theta|\le\zeta=\delta_\rho/(4k^2)$，则

$$\left|\mathbb{E}\cos(\theta J)-\mathbb{E}\cos(\theta Y/(h\sqrt{m}))\right|\le|\theta|.$$

因此，$1-\operatorname{Re}\widehat{w}(\theta)\ge\zeta$以及 $1+\operatorname{Re}\widehat{w}(\theta)\ge\zeta$。

对于其余频率 $\zeta\le|\theta|\le\pi$，独立伯努利步骤在 $|\widehat{w}(\theta)|$中给出因子$|(1+e^{i\theta})/2|\le$余弦$(\zeta/2)<1$。

根据所述 c_0，这些估计给出 $1-\operatorname{Re}\widehat{w}(\theta)\ge c_0\min\{\theta^2/(h^2m),1\}$，以及 $\operatorname{Re}\widehat{w}(\theta)\ge-1+[1-$余弦$(\zeta/2)]$。由于 $Nh\sqrt{m}=4$，参考权重满足

$$1-\operatorname{Re}\widehat{q}(\theta)\le\min\{N^2\theta^2/2,2\}\le8\min\{\theta^2/(h^2m),1\}.$$

这证明了（3.1）中的第一个不等式，其中 $c = c_0/8$。

对于完整分布，令 $u_1 = 1/[2(b_\rho + 1)]$。若 $|\theta| \le u_1 h\sqrt{m}$，则

$$\operatorname{Re}\widehat{\mu}_J(\theta) \ge 1 - |\theta|\mathbb{E}J \ge 1 - u_1(b_\rho + h\sqrt{m}) \ge \tfrac{1}{2}.$$

对于 $|u| \ge u_1$，不加条件时，同样的偶整数不等式给出 $1 \pm \mathbb{E}\cos(uY) \ge \delta_\rho/k_\rho^2$。由于 $|J - Y/(h\sqrt{m})| \le 1$，对于 $u_1 h\sqrt{m} \le |\theta| \le \zeta_\rho$，我们得到 $1 + \operatorname{Re}\widehat{\mu}_J(\theta) \ge \zeta_\rho$。对于 $\zeta_\rho \le |\theta| \le \pi$，伯努利因子给出 $|\widehat{\mu}_J(\theta)| \le \cos(\zeta_\rho/2)$。这三个范围证明了所述的 β_P 取值。对于一般截断，$S = 4b_\rho/\delta_\rho$，因此常数仅取决于 b_ρ，δ_ρ。 □

5. 命题 *B* 的证明。

命题 B 的证明。由引理 5.1 可得可逆性。将引理 5.2 的谱断言直接应用于完整分布 μ_J。由此得到 $P \succeq (-1+\beta_P)I$，其中 β_P 满足 (3.1)；这里不使用能量截断。引理 5.3 保证该常数对于所有可行的 N 都一致为正。可逆马尔可夫压缩算子 P 的谱包含于 $[-1+\beta_P, 1]$。在此区间上，$1-\lambda^2 = (1-\lambda)(1+\lambda) \ge \beta_P(1-\lambda)$。由谱定理可得

$$\|f\|^2_{L^2(\pi)} - \|Pf\|^2_{L^2(\pi)} \ge \beta_P \mathcal{E}_P(f), \qquad f \in L^2(\pi). \tag{5.5}$$

□

引理5.4 （三种持续时间分布）。设 $N \ge 16$ 且 $h = 4/(N\sqrt{m})$。对于表 2 中的每种持续时间分布，所列出的 τ_*、c、β_P 满足 (3.1)，且保留概率为 p_S。

引理 5.4 的证明（包括三种分布的显式常数）见附录 C。

6. 受限导通率。我们分四步证明命题 C。首先，引理 6.1和6.2 控制沿轨迹的 Verlet 能量误差，从而界定拒绝概率。其次，对于积分时间为 $T = 4/\sqrt{m}$ 的三角形持续时间分布（5.4），命题 6.8表明，提议步骤会收缩相对熵：提议位置的分布 x_J 比初始分布 π 更接近一个固定因子。该论证是 [4], 在第 6.2 节中给出的加权熵估计的离散时间版本。第三，引理 6.9 将这一收缩转化为基于三角分布构造的核的受限导通率界 P_{ref}。第四，第 6.4 节利用第 $P5$ 节的比较引理 5.2，将该界推广到一般非共振持续时间分布的核。证明中仅使用三角分布；定理适用于满足定义 2.1 的所有分布。

归一化 $U(x_*) = 0$并写作 $g = \nabla U$。对于 $h > 0$和 $T = Nh$，应用 Verlet 更新（2.4），执行 N 步。两次踢步之间，位置自由移动：在第 n 步中，它为 $x_n + \sigma w_n$，其中 $0 \le \sigma \le h$ 是半步速度。第 w_n 节中的熵估计沿这些自由飞行段证明，然后在步端点处求值。6.2

6.1. 能量控制。 当海森矩阵满足 Lipschitz连续时，Verlet 步的能量误差可分解为修正能量的增量和一个步长三次阶的余项；修正能量的增量会沿轨迹望远镜式地累加。海森正则性正是在此处发挥作用。该恒等式及三次阶界见 [4, 引理 2.1]；证明很短，我们将其列出。

引理 6.1（速度 Verlet 的修正能量）。 假设 *(A3)*。记 $H_n = H(x_n, v_n)$，并定义修正能量 F_n 以及余项 r_n 如下

$$F_n = H_n - \frac{h^2}{8}|g(x_n)|^2, \qquad r_n = U(x_{n+1}) - U(x_n) - \tfrac{1}{2}\langle g(x_n) + g(x_{n+1}), x_{n+1} - x_n\rangle.$$

Then

$$F_{n+1} - F_n = r_n, \qquad |r_n| \le \frac{L_H h^3}{12}|w_n|^3. \tag{6.1}$$

证明。令 $a = x_{n+1} - x_n = hw_n$。代入 $v_n = w_n + (h/2)g(x_n)$ 和 $v_{n+1} = w_n - (h/2)g(x_{n+1})$，得到

$$\tfrac{1}{2}(|v_{n+1}|^2 - |v_n|^2) = -\tfrac{1}{2}\langle g(x_n) + g(x_{n+1}), a\rangle + \tfrac{h^2}{8}(|g(x_{n+1})|^2 - |g(x_n)|^2).$$

加上势能变化即可证明 $F_{n+1} - F_n = r_n$。对 $t \mapsto \langle g(x_n + ta)\ a\rangle$ 应用微积分基本定理，得到

$$r_n = \int_0^1 (\tfrac{1}{2} - t)a^{\mathsf{T}}\nabla^2 U(x_n + ta)a\, dt.$$

在积分内减去 $\nabla^2 U\ x_n + a/2$ $\int_0^1\ t -$ $1/2\ ^2 dt = 1/12$ 6 1 ()。其系数积分为零，而 ()。由此证明了（。）中的第二个断言。 □

以下能量壳层界是本研究设定所特有的。较小的 a_T 控制 F_n 的累积变化，并使轨迹保持在有界能量区域内；b_h 控制从 F_n 到 H_n 的端点修正。

引理 6.2（能量控制）。假设 *(A1)–(A3)*。设 $R \ge 1$，$h\sqrt{L} \le 1$，并且

$$a_T = L_H T h^2 R^{3/2}, \qquad b_h = h^2 L R, \qquad 2a_T + \frac{1}{2}b_h \le R. \tag{6.2}$$

对于满足 H（x_0、v_0）$\le R$ 的每个输入，所有 $0 \le n \le N$ 均满足

$$F_n \le R + 2a_T, \qquad H_n \le 2R, \qquad |H_n - H_0| \le 2a_T + b_h. \tag{6.3}$$

对于 $0 \le n < N$，还有 $|w_n|^2 \le 8R$ 和 $\sum_{i<n}|r_i| \le 2a_T$。特别地，这条轨迹的每个可能终点都以至少 $e^{-(2a_T + b_h)}$ 的概率通过 *Metropolis* 检验。

证明。界 $|g(x)|^2 \le 2LU(x)$ 可由在 $x - L^{-1}g(x)$ 处应用下降引理并利用 $U \ge 0$ 得到。因此

$$(1 - h^2L/4)H_n \le F_n \le H_n, \qquad |w_n|^2 \le 2(1 + h^2L/4)H_n \le 4F_n. \tag{6.4}$$

由于 $h^2L \le 1$，第一个界推出 $F_n \ge 3H_n/4$。中间的界由柯西-施瓦茨不等式应用于 $v_n - (h/2)g(x_n)$得到，权重为 1 和 $h\sqrt{L}/2$。若 $F_i \le 2R$ 对之前的指标成立，则 $|w_i|^2 \le 8R$ 对那些指标成立，而（6.1）中的三次界给出

$$\sum_{i<n}|r_i| \le \frac{2}{3}L_H T h^2 (2R)^{3/2} \le 2a_T.$$

根据（6.1）中的恒等式，$F_n \le F_0 + 2a_T \le R + 2a_T \le 2R$。从 $F_0 \le R$ 开始进行归纳，可证明该结论对每个指标都成立。此外，（6.2）给出

$$H_n \le \frac{R + 2a_T}{1 - h^2L/4} \le 2R,$$

因为最后一个不等式等价于 $2a_T + b_h/2 \le R$。另外 $|w_n|^2 \le 8R$。最后，

$$|H_n - H_0| \le 2a_T + \frac{h^2 L}{4}(H_n + H_0) \le 2a_T + \tfrac{3}{4}b_h \le 2a_T + b_h.$$

接受的下界直接由其定义得出。 □

现在将引理 6.2 应用于命题C证明中出现的轨迹，所用参数见表 1。参考轨迹以及在 $\tau \le \tau_*$ 条件下的轨迹，其物理积分时间均至多为 T_*。事实上， $4/\sqrt{m} \le T_*$ 且

$$Jh \le \tau + h \le \tau_* + \frac{1}{4\sqrt{m}} \le T_*,$$

由于 $N \ge 16/\sqrt{\delta_\rho}$给出 $h\sqrt{m} \le 1/4$。令 $a_* = L_H T_* h^2 R^{3/2}$ 以及 $b_h = h^2 LR$。命题 C 中的步长假设恰好是

$$a_* + b_h = h^2(LR + L_H T_* R^{3/2}) \le \eta/3. \tag{6.5}$$

由于 $R \ge 1$和 $\eta = c/600 \le 1/600$，这意味着 $h^2 L = b_h/R < 1$和 $2a_* + b_h/2 \le 2\eta/3 < R$。因此，引理6.2的假设条件成立。

令 $N_* = \lfloor T_*/h \rfloor$。将引理 6.2 应用于物理积分时间 $N_* h$。其累积误差参数 $a_{N_* h}$ 至多为 a_*，其端点误差参数为 b_h。因此，对于$H(x_0,\ v_0) \le R$，每个端点 $(Q,\ W)$ 在时间 T_* 内到达时，其能量增加至多为$b_h + 2a_* \le \eta$，因此

$$\alpha(X, V; Q, W) \ge e^{-\eta} \ge 1 - \eta. \tag{6.6}$$

6.2. *Verlet* 提案的熵。 对于位置和速度的联合分布 p，其位置边缘分布为 ν，定义

$$E_{\mathrm{joint}}(p) = \mathrm{KL}(p\|\Pi), \qquad E_{\mathrm{pos}}(\nu) = \mathrm{KL}(\nu\|\pi), \qquad E_v(p) = \mathrm{KL}(p\|\nu \otimes \gamma_d). \tag{6.7}$$

熵链式法则给出 $E_{\mathrm{joint}}(p) = E_{\mathrm{pos}}(\nu) + E_v(p)$。当 ν 和 π 具有有限的微分熵以及 U 的有限期望时，记它们的勒贝格密度为 $\nu(x)$ 和 $\pi(x)$，并将负微分熵之差定义为

$$\Delta\mathcal{L}(\nu) = \int_{\mathbb{R}^d} \nu(x)\log\nu(x)\,\mathrm{d}x - \int_{\mathbb{R}^d} \pi(x)\log\pi(x)\,\mathrm{d}x = E_{\mathrm{pos}}(\nu) - \big(\mathbb{E}_\nu U - \mathbb{E}_\pi U\big). \tag{6.8}$$

我们首先回顾 Talagrand 输运熵不等式 [22]。在加权熵论证中，Talagrand 不等式通过初始相对熵界定初始平方 Wasserstein 距离与目标分布之间的距离。

引理 6.3（Talagrand 输运熵不等式）。假设*(A1)*，并令 π 由 (2.1)给出。对于 ν 上的每个概率测度 $\mathbb{R}^d$，满足 $\mathrm{KL}(\nu\|\pi) < \infty$，

$$\frac{1}{2}W_2(\nu,\pi)^2 \le \frac{1}{m}\mathrm{KL}(\nu\|\pi). \tag{6.9}$$

[4]的自由输运熵不等式通过自由飞行两个端点的相对熵之和，对飞行过程中的速度相对熵进行界定。结合 Verlet 能量恒等式，我们可以界定 KL(x_J的分布)$\|\pi$，其中 x_J 是为 Metropolis 检验提出的位置，而 J 服从截断三角分布。

引理 6.4（[4, 命题 3.1 和推论 3.5]）。 对于步长 $h>0$，令 X，V为 $\mathbb{R}^d$中的随机向量，记

$$p_t = \mathrm{Law}(X+tV, V), \qquad \nu_t = \mathrm{Law}(X+tV),$$

并定义 $e(t) = \mathrm{KL}(p_t \,\|\, \nu_t \otimes \gamma_d)$ 为 $0 \le t \le h$。若 $e(0)+e(h)<\infty$，则

$$e(t) \le e(0)+e(h), \qquad 0 \le t \le h. \tag{6.10}$$

特别地，所有中间熵均有限。

这就是 [4, 推论 3.5]。

下面的自由输运估计用条件速度分布的一阶矩和二阶矩来界定 $\frac{1}{2}W_2(\nu_t,\pi)^2$的二阶导数。以下恒等式将这些矩转化为相对于 γ_d 的相对熵；当 $G=\Sigma^{-1}$ 时，它就化为高斯分布的最大熵性质。

引理 6.5 （相对于高斯分布的相对熵）。设 μ 是 $\mathbb{R}^d$ 上具有有限二阶矩、均值为 u 、协方差为 Σ 的概率测度，且 G 为对称正定矩阵。记 $\Sigma: G = \mathrm{tr}(\Sigma G)$ 为 *Frobenius* 内积。若 $\mathrm{KL}(\mu\|\gamma_d)<\infty$，则

$$2\,\mathrm{KL}(\mu\|\gamma_d) = |u|^2 + \mathrm{tr}\,\Sigma - \Sigma : G + \log\det G + 2\,\mathrm{KL}\big(\mu \| N(u, G^{-1})\big).$$

在所有情况下，

$$|u|^2 + \mathrm{tr}\,\Sigma - \Sigma : G + \log\det G \le 2\,\mathrm{KL}(\mu\|\gamma_d),$$

当且仅当 $\mu = N$（u、G^{-1}）时取等号。

证明。如果 $\mathrm{KL}(\mu\|\gamma_d)=\infty$，不等式显然成立且为严格不等式。否则，令 $\widetilde{\gamma}=N(u, G^{-1})$。$\widetilde{\gamma}$与 γ_d的密度比的对数是二次函数，因此在μ下可积。因此 $\mathrm{KL}(\mu\|\widetilde{\gamma})$有限，并且

$$2\,\mathrm{KL}(\mu\|\gamma_d) = 2\,\mathrm{KL}(\mu\|\widetilde{\gamma}) + \int_{\mathbb{R}^d} \big[|v|^2 - (v-u)^{\mathsf{T}} G (v-u) + \log\det G\big]\,\mu(dv).$$

计算该积分即可得到恒等式。$\mathrm{KL}(\mu\|\widetilde{\gamma})$的非负性给出了不等式及等号成立的情形。 □

令 $p=p_0$ 为位置和速度的联合分布 （X，V），其中 p_t 和 ν_t 如引理 6.4所述。对于具有密度的位置分布 ν，令 T_ν 为其到 π的 Brenier 映射。的流为

$$\mathcal{J}(p) = \int_{\mathbb{R}^{2d}} \langle v, x - T_\nu(x)\rangle\, p(dx,dv).$$

使用 E_v 和 $\Delta\mathcal{L}$，它们分别来自（6.7）和（6.8）。

引理 6.6（自由输运估计）。假设 *(A1)* 和 *(A2)*。令 p 具有紧支集，并满足 $\mathrm{KL}(p\|\Pi)<\infty$。对于 $0 \le a \le b < \infty$，

$$\mathcal{J}(p_b) - \mathcal{J}(p_a) \le \int_a^b \{2E_v(p_t) - \Delta\mathcal{L}(\nu_t)\}\,dt. \tag{6.11}$$

此外，$\mathcal{W}(t) = \frac{1}{2}W_2(\nu_t,\pi)^2$连续可微，并且 $\mathcal{W}'(t) = \mathcal{J}(p_t)$。因此，在分布意义下，

$$\mathcal{W}''(t) \le 2E_{\mathrm{joint}}(t) - 3E_{\mathrm{pos}}(t) + U_\Delta(t), \tag{6.12}$$

其中

$$E_{\mathrm{joint}}(t) = \mathrm{KL}(p_t\|\Pi), \qquad E_{\mathrm{pos}}(t) = \mathrm{KL}(\nu_t\|\pi), \qquad U_\Delta(t) = \mathbb{E}U(X+tV) - \int_{\mathbb{R}^d} U\,d\pi.$$

证明。我们首先将 $\mathcal{W}$ 的导数与流 $\mathcal{J}(p_t)$联系起来。对于位置平滑后的分布，我们随后利用条件速度协方差和高斯熵不等式来界定其二阶差分。对该界进行积分并消除平滑处理后，得到 (6.11)。假设 (A1) 和 (A2) 仅通过 π发挥作用：根据 (A1)，它具有正密度和有限二阶矩，因此与 π 的最优耦合及 Brenier 映射均存在；根据 (A2)，势函数至多按二次速度增长，因此 $\int_{\mathbb{R}^d} U\, d\pi$ 和 $\Delta\mathcal{L}(\nu_t)$ 均为有限值。这两个假设也用于引理A.1，步骤 2 至 5 均依赖于该引理。

步骤*1*：一阶变分。恒等式 $\mathcal{W}' = \mathcal{J}(p_t)$ 是沿曲线 $t \mapsto \nu_t$ [1, 的 W_2 一阶变分公式，见定理 8.4.7]；我们给出简短论证，因为我们需要在每个 t 处的导数及其连续性。有限相对熵保证 (X, V) 存在密度。自由剪切保持绝对连续性，因此每个位置分布 ν_t 都有密度。令$M_v = \mathbb{E}|V|^2$。将 $X + tV$ 与 π 作最优耦合，并在源端沿 ϵV 移动时保持其目标坐标不变。由此得到的传输成本给出 $\mathcal{W}(t+\epsilon)$ 的上界。交换 t 和 $t+\epsilon$ 的角色，可得下界

$$\epsilon\mathcal{J}(p_{t+\epsilon}) - \tfrac{1}{2}\epsilon^2 M_v \le \mathcal{W}(t+\epsilon) - \mathcal{W}(t) \le \epsilon\mathcal{J}(p_t) + \tfrac{1}{2}\epsilon^2 M_v, \qquad \epsilon > 0.$$

我们声称，当 $s \to t$ 时，$\mathcal{J}(p_s) \to \mathcal{J}(p_t)$。分布 p_s 在 W_2 中收敛到 p_t。将 ν_s 与 π 的最优耦合提升为 p_s 与 π 的耦合，同时携带速度。提升后的这些耦合是紧的，其每个极限点都是 ν_t 与 π 的最优耦合，这是由定理 5.20] 得出的 [22,。由于 ν_t 有密度，该最优耦合唯一，并由 Brenier 映射 T_t 诱导。因此，提升后的耦合弱收敛到 $(X_t, V_t, T_t(X_t))$ 的分布，而 $\mathcal{J}(p_s)$ 是在时刻 s 对提升耦合积分 $\langle v, x - y\rangle$ 所得。被积函数由 $|v|^2 + |x|^2 + |y|^2$ 控制，其积分收敛，因此它一致可积，且积分收敛到 $\mathcal{J}(p_t)$。将此应用于双侧估计中的 $t + \epsilon \to t$，可得$\mathcal{W}'(t) = \mathcal{J}(p_t)$，且导数连续。

步骤*2*：二阶变分。固定一个平滑参数 $\iota \in (0, 1]$, 将在步骤 5 中趋于零）。向 X 添加协方差为 ιI_d 的独立高斯噪声，并保持 V不变。用 p^ι 表示由此得到的位置和速度的联合分布，并在本次计算中省略上标。步骤 1 中的耦合给出了导数，但其二阶差分界仅为 M_v。为得到熵界，我们使用条件平均速度的流来表示相同的位置分布。令 n_t 为位置密度，并记

$$u_t(y) = \mathbb{E}[V \mid X + tV = y], \qquad \Sigma_t(y) = \mathrm{Cov}(V \mid X + tV = y), \tag{6.13}$$

其中 X 现在表示平滑后的位置。平滑使 n_t 为正，且空间矩场光滑。在紧时间区间上，u_t、Σ_t 及其一阶空间导数均有界，并且根据引理 A.1(a) 和 (b)，∇的对数 n_t 至多呈线性增长。

自由输运方程的零阶和一阶速度矩分别给出连续性方程和动量方程

$$\partial_t n_t + \mathrm{div}(n_t u_t) = 0, \qquad \partial_t(n_t u_t) + \mathrm{div}\big(n_t(u_t \otimes u_t + \Sigma_t)\big) = 0.$$

展开第二个方程并使用第一个方程。因此，u_t 的流的加速度为

$$a_t = \partial_t u_t + (Du_t)u_t = -\frac{\mathrm{div}(n_t\Sigma_t)}{n_t} = -\,\mathrm{div}\,\Sigma_t - \Sigma_t \nabla \log n_t. \tag{6.14}$$

特别地，a_t 关于空间是连续的，且至多呈线性增长。

定义流 $\Phi_{t,\,\sigma}$ 如下：

$$\partial_\sigma \Phi_{t,\sigma}(y) = u_{t+\sigma}\big(\Phi_{t,\sigma}(y)\big), \qquad \Phi_{t,0}(y) = y.$$

连续性方程给出 $n_{t+\sigma}(\Phi_{t,\,\sigma}(y))\ \det D_y\Phi_{t,\,\sigma}(y) = n_t(y)$，这是通过对左侧关于 σ 求导得到的。因此，$\Phi_{t,\,\sigma}$ 将 ν_t 输运至 $\nu_{t+\sigma}$。其加速度为 $a_{t+\sigma}(\Phi_{t,\,\sigma}(y))$，所以

$$\Phi_{t,\sigma}(y) = y + \sigma u_t(y) + \tfrac{1}{2}\sigma^2 a_t(y) + o_{L^2(n_t\,dy)}(\sigma^2).$$

对于余项，将加速度积分两次。差值 $a_{t+r}(\Phi_{t,\,r}(y)) - a_t(y)$ 在 $r \to 0$ 时逐点趋于零，并且当 r 较小时有界于 $C(1+|y|)$，因为 u_t 有界，且 a_t 具有线性增长。因此，该差值在 $L^2(n_t\,dy)$ 中趋于零。

令 $T_t = T_{\nu_t}$。对于 $Y \sim \nu_t$，配对 $(\Phi_{t,\sigma}(Y)\ T_t(Y))$ 将 $\nu_{t+\sigma}$ 与 π 耦合。因此，其代价给出了 $\mathcal{W}(t+\sigma)$ 的上界。对于 $\sigma = \pm\epsilon$，展开该代价可得

$$\begin{aligned}\mathcal{W}(t+\sigma) &\le \tfrac{1}{2}\int_{\mathbb{R}^d} |\Phi_{t,\sigma}(y) - T_t(y)|^2 n_t(y)\,dy \\ &= \mathcal{W}(t) + \sigma\int_{\mathbb{R}^d} \langle y - T_t(y), u_t(y)\rangle n_t(y)\,dy \\ &\quad + \tfrac{1}{2}\sigma^2 \int_{\mathbb{R}^d} \big(|u_t(y)|^2 + \langle y - T_t(y), a_t(y)\rangle\big) n_t(y)\,dy + o(\sigma^2).\end{aligned}$$

平方范数中的误差为 $o(\sigma^2)$ 由柯西-施瓦茨不等式可得。记 $D_\epsilon^2 f(t) = [f(t+\epsilon) + f(t-\epsilon) - 2f(t)]/\epsilon^2$。将两个估计相加并利用（6.14）可得

$$\limsup_{\epsilon\to 0} D_\epsilon^2 \mathcal{W}(t) \le \int_{\mathbb{R}^d} |u_t|^2 n_t\,dy - \int_{\mathbb{R}^d} \langle y - T_t(y), \operatorname{div}(n_t\Sigma_t)(y)\rangle\,dy. \tag{6.15}$$

记 $T_t = \nabla\varphi_t$，其中 φ_t 为凸函数。Brenier 势不必是二阶连续可微的，因此我们利用其分布 Hessian 进行分部积分。其奇异部分贡献了一个非正项。将 $D^2\varphi_t = G_t\,dy + D^s\varphi_t$分解。两部分均为半正定，且 G_t 是 T_t的近似导数；参见定理 11.1] [22, 之后的讨论。为对散度项进行分部积分，令 $0 \le \chi_k \uparrow 1$ 为光滑截断函数，在半径为 k的球上恒等于一，且满足 $|\nabla\chi_k| \le C/k$。于是

$$\begin{aligned}&-\int_{\mathbb{R}^d} \chi_k \langle y - T_t(y), \operatorname{div}(n_t\Sigma_t)(y)\rangle\,dy \\ &= \int_{\mathbb{R}^d} \langle y - T_t(y), n_t\Sigma_t\nabla\chi_k\rangle\,dy + \int_{\mathbb{R}^d} \chi_k n_t \Sigma_t : (I - G_t)\,dy - \int_{\mathbb{R}^d} \chi_k n_t\Sigma_t : dD^s\varphi_t.\end{aligned}$$

最后一项非正，因为 Σ_t 和奇异海森矩阵测度都是半正定的。第一项趋于零，因为 Σ_t 有界，且 $y - T_t(y) \in L^2(n_t\,dy)$。左侧收敛，因为 a_t 线性增长。舍去非正项后，也可得递增积分 $\int_{\mathbb{R}^d} \chi_k n_t\Sigma_t : G_t$ 有界，因此由单调收敛可得

$$-\int_{\mathbb{R}^d} \langle y - T_t(y), \operatorname{div}(n_t\Sigma_t)(y)\rangle\,dy \le \int_{\mathbb{R}^d} n_t\Sigma_t : (I - G_t)\,dy. \tag{6.16}$$

步骤 3：熵界。两个位置分布都有密度且具有有限二阶矩，因此 Brenier 换元公式 [22，定理 11.1 和例 11.2] 给出 $n_t(y) = \pi(T_t(y))\det G_t(y)$ 几乎处处成立。因此，$G_t(y)$ 几乎处处正定。对于这样的 y，条件分布 $\mu_y = \mathcal{L}(V \mid X + tV = y)$ 的均值为 $u_t(y)$，协方差为 $\Sigma_t(y)$；根据链式法则和引理 A.1(d)，它相对于 γ_d 的相对熵对几乎所有 y 都是有限的。取 $\mu = \mu_y$ 和 $G = G_t(y)$，由引理 6.5 可得

$$|u_t(y)|^2 + \operatorname{tr}\Sigma_t(y) - \Sigma_t(y) : G_t(y) + \log\det G_t(y) \le 2\,\mathrm{KL}(\mu_y\|\gamma_d).$$

对 $n_t(y)\,dy$ 进行积分。根据链式法则，右侧积分为 $2E_v(p_t)$；对于左侧的最后一项，我们使用

$$\int_{\mathbb{R}^d} n_t \log\det G_t\,dy = \int_{\mathbb{R}^d} n_t \log n_t\,dy - \int_{\mathbb{R}^d} \pi\log\pi\,dy = \Delta\mathcal{L}(\nu_t).$$

引理 A.1(c) 中的高斯界保证了 $\log n_t$ 的可积性，并且 $\log \pi(T_t)$ 可积，因为 T_t 将 n_t 输运至 π。因此，积分不等式为

$$\int_{\mathbb{R}^d} n_t\big(|u_t|^2 + \operatorname{tr}\Sigma_t - \Sigma_t : G_t\big)\,dy \le 2E_v(p_t) - \Delta\mathcal{L}(\nu_t). \tag{6.17}$$

结合（6.15）、（6.16）和（6.17）可得

$$\begin{aligned}\limsup_{\epsilon\to 0} D_\epsilon^2\mathcal{W}(t) &\le \int_{\mathbb{R}^d} n_t\big(|u_t|^2 + \operatorname{tr}\Sigma_t - \Sigma_t : G_t\big)\,dy \\ &\le 2E_v(p_t) - \Delta\mathcal{L}(\nu_t).\end{aligned}$$

步骤 *4*：对时间积分。现在我们从二阶差分的这个界转向积分不等式（6.11）。在相同的 t 处对 ϵ取两种符号，并应用步骤 1 中双侧估计的上界，可得 $D_\epsilon^2\mathcal{W}(t) \le M_v$。对于一个光滑的非负测试函数$\zeta$，其在 $(0,\infty)$ 上具有紧支撑，将 D_ϵ^2 从 $\mathcal{W}$ 移至 ζ 积分号下。利用上界 M_v应用反向 Fatou 引理，即得

$$\langle \mathcal{W}'', \zeta\rangle = \lim_{\epsilon\to 0}\int_{(0,\infty)} \zeta D_\epsilon^2\mathcal{W}\,dt \le \int_{(0,\infty)} \zeta\,[2E_v(p_t) - \Delta\mathcal{L}(\nu_t)]\,dt.$$

根据引理 A.1(d)，右侧在局部上可积。由于 $\mathcal{W}' = \mathcal{J}(p_t)$ 是连续的，函数

$$t \longmapsto \mathcal{J}(p_t) - \int_a^t \big[2E_v(p_r) - \Delta\mathcal{L}(\nu_r)\big]\,dr$$

是连续的，且具有非正的分布导数，因此是非递增的。比较其在 a 和 b 处的值，便得到平滑后的分布的（6.11）。

步骤 *5*：去除平滑化。我们利用附录 A 中的收敛断言取极限：（i）和（ii）给出积分的收敛，而（iii）给出端点流的收敛。这就证明了原始分布的（6.11）。最后，利用链式法则和（6.8）可得

$$2E_v(p_t) - \Delta\mathcal{L}(\nu_t) = 2\big(E_{\text{joint}}(t) - E_{\text{pos}}(t)\big) - \big(E_{\text{pos}}(t) - U_\Delta(t)\big).$$

结合 $\mathcal{W}' = \mathcal{J}$，可得（6.12）。 □

对于加权估计，假设 (A1) 和 (A2)。设 $T = Nh$，并设 p_0 为位置和速度的初始联合分布。假设其有限相对熵 E_0 =KL 散度$(p_0\|\Pi)$，支撑于 $\{H \le R\}$，且满足

$$\mathbb{E}[v_0 \mid x_0] = 0. \tag{6.18}$$

当 Π 在 $\{x_0 \in A, H \le R\}$条件化时，条件（6.18）成立。

对于 $t = nh + \sigma$、$0 \le \sigma \le h$，定义

$$\begin{aligned} E_{\text{pos}}(n,\sigma) &= \mathrm{KL}(\mathcal{L}(x_n + \sigma w_n)\|\pi), & E_{\text{joint}}(n,\sigma) &= \mathrm{KL}(\mathcal{L}(x_n + \sigma w_n, w_n)\|\Pi), \\ E_n^x &= \mathrm{KL}(\mathcal{L}(x_n)\|\pi), & U_\Delta(t) &= \mathbb{E}U(x_n + \sigma w_n) - \mathbb{E}_\pi U, \\ \mathcal{W}(t) &= \tfrac{1}{2}W_2(\mathcal{L}(x_n + \sigma w_n), \pi)^2. \end{aligned}$$

这里，E_0 表示初始联合熵，而 E_0^x 表示初始位置熵；特别地，$E_0^x \le E_0$。记 $E_{\text{pos}}(t) = E_{\text{pos}}(n, \sigma)$和 $E_{\text{joint}}(t) = E_{\text{joint}}(n, \sigma)$为 $nh < t < (n+1)h$，并令

$$I_{\text{pos}} = \int_0^T (T-t)E_{\text{pos}}(t)\,dt, \qquad I_{\text{joint}} = \int_0^T (T-t)E_{\text{joint}}(t)\,dt.$$

定义

$$\omega_n = \int_0^h (T - nh - \sigma)\, d\sigma = h(T - nh) - \tfrac{1}{2}h^2,$$
$$b_n = \mathbb{E}[F_n - H(x_0, v_0)],$$
$$Q_n = \mathbb{E}[U(x_{n+1}) - U(x_n) - h\langle g(x_n), w_n\rangle] \geq 0,$$
$$I_{\text{end}} = \sum_n \omega_n (E_n^x + E_{n+1}^x), \qquad I_{\text{mod}} = \sum_n \omega_n b_n,$$
$$Q_\Sigma = \sum_n \omega_n Q_n, \qquad \Delta_U = \mathbb{E}U(x_N) - \mathbb{E}U(x_0),$$
$$Q_{\text{int}} = \sum_n \int_0^h (T - nh - \sigma)\mathbb{E}[U(x_n + \sigma w_n) - U(x_n) - \sigma\langle g(x_n), w_n\rangle]\, d\sigma \geq 0.$$

所有关于 n 的求和都从 0 到 $N-1$。

引理 6.7（加权熵估计）。假设*(A1)* 和*(A2)*成立，且初始联合分布 p_0 具有有限相对熵 E_0，其支撑集包含于 $\{H \leq R\}$，并满足(6.18)。则

$$3I_{\text{end}} \leq \mathcal{W}(0) + \tfrac{5}{2}E_0T^2 + 5I_{\text{mod}} + 3Q_\Sigma + \frac{h^2}{4}\Delta_U - \frac{h^2}{12}\sum_n Q_n - Q_{\text{int}}. \tag{6.19}$$

证明。步骤 *1*：每段漂移上的自由输运。插值位置在每个时刻 nh处连续。在每段漂移开始时，联合分布具有紧支撑，且相对于 Π具有有限相对熵。此前的映射连续且保持勒贝格体积，并且 H在初始支撑集的像上有界。因此，引理6.6（其使用(A1) 和(A2)）适用于每段漂移。

步骤 *2*：踢步，以及对权重 $T-t$的积分。在时刻 nh （对于 $1 \leq n < N$），插值速度变化了 $-hg(x_n)$。令 T_n 为从 $\mathcal{L}(x_n)$ 到 π 的 Brenier 映射，并设

$$A_n = \mathbb{E}\langle g(x_n), x_n - T_n(x_n)\rangle = U_\Delta(nh) + C_n^U,$$

其中

$$C_n^U = \mathbb{E}\big[U(T_n(x_n)) - U(x_n) - \langle g(x_n), T_n(x_n) - x_n\rangle\big] \geq 0.$$

等式利用了 $T_n(x_n) \sim \pi$，不等式则源于 U 在 (A1)中的凸性。因此，单侧一阶变分满足

$$\mathcal{W}'(nh+) - \mathcal{W}'(nh-) = -hA_n \quad (1 \leq n < N), \qquad \mathcal{W}'(0+) = -\tfrac{h}{2}A_0.$$

最后一个恒等式使用了 (6.18) 和初始半踢步。将 (6.12) 乘以权重 $T-t$ 进行积分，并以相同权重加上 $\mathcal{W}'$ 的每次跳跃。由于权重是线性的，且在 T 处为零，分部积分后左侧仅留下端点项 $\mathcal{W}(T) - \mathcal{W}(0) - T\mathcal{W}'(0+)$。对于一般的光滑权重 $a(t)$，同样的计算还会产生一个内部项 $\int_0^T a''(t)\mathcal{W}(t)\mathrm{d}t$。代入初始半踢步和内部跳跃后，得到

$$\mathcal{W}(T) + 3I_{\text{pos}} \leq \mathcal{W}(0) + 2I_{\text{joint}} + \mathcal{Q} - h\left[\frac{T}{2}C_0^U + \sum_{n=1}^{N-1}(T - nh)C_n^U\right], \tag{6.20}$$

$$\mathcal{Q} = \int_0^T (T-t)U_\Delta(t)\, dt - h\left[\frac{T}{2}U_\Delta(0) + \sum_{n=1}^{N-1}(T - nh)U_\Delta(nh)\right].$$

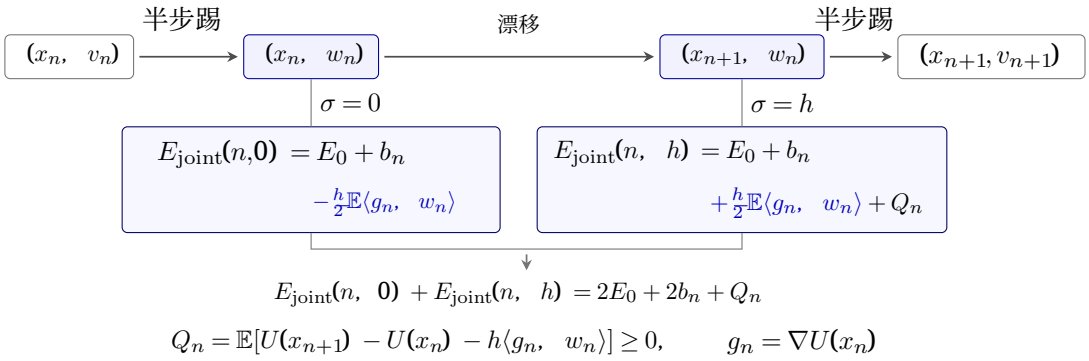


图 3。一次速度 *Verlet* 步。两个漂移端点熵中的线性项符号相反。它们相加后留下共同项 $2(E_0 + b_n)$ 以及势函数余项 Q_n。

步骤 *3*：求积误差 $\mathcal{Q}$。定义

$$\mathrm{Chord}_n(\sigma) = (1-\sigma/h)\mathbb{E}U(x_n) + (\sigma/h)\mathbb{E}U(x_{n+1}) - \mathbb{E}U(x_n + \sigma w_n) \geq 0,$$

其中，不等式再次源于 U 的凸性。在区间 $[nh,(n+1)h]$ 上，线性插值函数的两个端点系数分别为 $h(T-nh)/2 - h^2/6$ 和 $h(T-nh)/2 - h^2/3$。相邻区间的贡献在每个内部时刻 nh 相加为 $h(T-nh)$，从而留下边界修正以及线性插值函数与势函数之间的差异。

$$\mathcal{Q} = \frac{h^2}{6}\left[U_\Delta(T) - U_\Delta(0)\right] - \sum_{n=0}^{N-1}\int_0^h (T - nh - \sigma)\,\mathrm{Chord}_n(\sigma)\,d\sigma. \tag{6.21}$$

步骤 *4*：沿 *Verlet* 步进行能量核算。根据体积保持，相对于 Π 的联合相对熵的变化等于 H 的期望变化。由于 $v_n = w_n + (h/2)g(x_n)$，展开可得

$$\begin{aligned} E_{\text{joint}}(n,\sigma) = E_0 + b_n + (\sigma - h/2)\mathbb{E}\langle g(x_n), w_n\rangle \\ + \mathbb{E}[U(x_n + \sigma w_n) - U(x_n) - \sigma\langle g(x_n), w_n\rangle]. \end{aligned} \tag{6.22}$$

线性项在两个漂移端点处的线性项相加后会抵消，如图 3所示。我们得到 in

$$E_{\text{joint}}(n,0) + E_{\text{joint}}(n,h) = 2E_0 + 2b_n + Q_n, \tag{6.23}$$

$$I_{\text{joint}} - \tfrac{1}{2}E_0T^2 = I_{\text{mod}} - \frac{h^2}{12}\Delta_U + \frac{h^2}{12}\sum_n Q_n + Q_{\text{int}}. \tag{6.24}$$

对于后一个恒等式，线性项积分为

$$\int_0^h (T - nh - \sigma)(\sigma - h/2)\,d\sigma = -h^3/12, \qquad \sum_n h\mathbb{E}\langle g(x_n), w_n\rangle = \Delta_U - \sum_n Q_n.$$

步骤 *5*：转移到位置边缘分布并得出结论。在每个区间 $[nh,(n+1)h]$中，引理6.4 给出

$$E_{\text{joint}}(n,\sigma) - E_{\text{pos}}(n,\sigma) \leq E_{\text{joint}}(n,0) - E_n^x + E_{\text{joint}}(n,h) - E_{n+1}^x.$$

两边乘以 $T - nh - \sigma$ 并积分，可得

$$I_{\text{end}} \leq I_{\text{pos}} - I_{\text{joint}} + \sum_n \omega_n[E_{\text{joint}}(n,0) + E_{\text{joint}}(n,h)]. \tag{6.25}$$

将（6.25）乘以 3，并代入（6.23）和（6.20）。利用 $\mathcal{W}(T)\ge 0$，并舍去涉及 C_n^U 的非正项，可得

$$3I_{\mathrm{end}} \le \mathcal{W}(0) - I_{\mathrm{joint}} + \mathcal{Q} + 3E_0T^2 + 6I_{\mathrm{mod}} + 3Q_\Sigma.$$

代入（6.24）和（6.21），并舍去涉及 Chord_n 的非正项，即得（6.19）。

□

以下命题汇总了通过将引理 [17, Lemma 3] 证明中的加权熵论证应用于速度 Verlet 而得到的估计。

命题 *6.8*（提议分布的相对熵）。假设 *(A1)–(A3)*。设 $T = Nh \ge 2/\sqrt{m}$、$h\sqrt{L}\le 1$、$R\ge 1$，且 $a_T = \overline{L}_H T h^2 R^{3/2}$ 和 $b_h = h^2 LR$ 满足 $2a_T + b_h/2 \le R$，如式（*6.2*）所示。设初始位置和速度 (x_0, v_0) 的分布支撑于 $\{H\le R\}$，其相对于 Π 的有限相对熵为 E_0，且 $\mathbb{E}[v_0 \mid x_0] = 0$ 如式（*6.18*）所示。若 J 独立地从（*5.4*）中抽取，则提议位置 x_J 满足

$$\mathrm{KL}(\mathcal{L}(x_J)\|\pi) \le \left(\frac{5}{6} + \frac{1}{3mT^2}\right) E_0 + \frac{5}{3}a_T + \frac{49}{24}b_h. \tag{6.26}$$

证明。引理6.2 与凸性可得

$$b_n \le 2a_T, \quad I_{\mathrm{mod}} \le a_T T^2, \quad Q_n \le \tfrac{1}{2}Lh^2\mathbb{E}|w_n|^2 \le 4Lh^2R, \quad Q_\Sigma \le 2b_hT^2, \quad \Delta_U \le 2R.$$

此外，$\mathcal{W}(0) \le E_0/m$ 由输运-熵不等式（引理6.3）以及对位置边缘分布应用熵数据处理可得。将引理6.7中的估计（6.19 ）除以 $3T^2$，舍去最后两个非正项，并利用 $LT^2 \ge 4$。由此得到

$$\frac{I_{\mathrm{end}}}{T^2} \le \left(\frac{5}{6} + \frac{1}{3mT^2}\right) E_0 + \frac{5}{3}a_T + \left(2 + \frac{1}{24}\right) b_h. \tag{6.27}$$

三角分布赋予 $2\omega_n/T^2$ 以 $[nh,\ (n+1)\,h)$ 的概率。随机舍入以相同概率选取该区间上的 $J = n$ 或 $J = n+1$。因此，对于（5.4）中的概率，$I_{\mathrm{end}}/T^2 = \sum_{j=0}^{N} q_j E_j^x$。混合分布下相对熵的凸性给出$\mathrm{KL}(\mathcal{L}(x_J)\|\pi) \le I_{\mathrm{end}}/T^2$，从而证明了该结论。

□

6.3. 熵收缩的受限导通率。下一个引理将 Metropolis 调整核的提议分布的相对熵收缩转化为该核的受限导通率界。该收缩只需对某个特定的初始分布族成立；在实际应用中，只需证明它对一个包含此分布族的类别中的每个初始分布都成立。

引理 6.9（熵收缩蕴含受限导通率）。设 $\bar\pi$ 是 $\mathbb{R}^d\times\mathcal{V}$ 上的概率测度，其中 $\mathcal{V}$ 是可测空间，第一边缘分布为 π，给定第一坐标时的可测条件分布为 $\bar\pi(\cdot\mid x)$，并设 K 是具有可测提议函数 $\Upsilon:\mathbb{R}^d\times\mathcal{V}\to\mathbb{R}^d$ 和接受概率 $\alpha:\mathbb{R}^d\times\mathcal{V}\to[0,1]$ 的 *Metropolis* 校正核：从 x 出发，抽取 $V\sim\bar\pi(\cdot\mid x)$，并以概率 (x,V) 移至 $\Upsilon(x,V)$，使得

$$K(x,A) = \int_{\mathcal{V}} \big[\alpha(x,v)\mathbf{1}_A(\Upsilon(x,v)) + (1-\alpha(x,v))\mathbf{1}_A(x)\big]\,\bar\pi(\mathrm{d}v\mid x). \tag{6.28}$$

令 $s>0$ 和 $\eta\in(0,1)$，并设 $S\subseteq\mathbb{R}^d\times\mathcal{V}$ 为一个可测集，满足 $\bar\pi(S^c)\le\eta s$，在该集合上提案以至少 $1-\eta$ 的概率被接受；也就是说，对所有$(x,v)\in S$，都有 $\alpha(x,v)\ge 1-\eta$。

假设存在某个 $\theta \in [0,1)$ 和 $\delta_0 \geq 0$，使得对每个满足 $s \leq \pi(A) \leq 1/2$ 的博雷尔 $A \subseteq \mathbb{R}^d$，提议都能将条件分布 $p_A := \bar{\pi}(\cdot \mid \{x \in A\} \cap S)$ 的相对熵收缩：

$$\mathrm{KL}(\mathrm{Law}(\Upsilon(Z)) \| \pi) \leq \theta \, \mathrm{KL}(p_A \| \bar{\pi}) + \delta_0 \qquad \textit{for } Z \sim p_A. \tag{6.29}$$

那么，对每个这样的 A 以及每个 $\phi_0 \in (0, (1-\theta)/2]$，满足

$$H_2(\phi_0) + \frac{\theta\eta}{1-\eta} + \delta_0 \leq (1-\theta)\log 2, \qquad H_2(p) = -p\log p - (1-p)\log(1-p), \tag{6.30}$$

$$Q_K(A, A^c) \geq (1-\eta)^2 \, \phi_0 \, \pi(A). \tag{6.31}$$

注意，若 $\lambda := (1-\theta)\log 2 - \theta\eta/(1-\eta) - \delta_0 > 0$，则 (6.30) 对 $\phi_0 = \min\{(1-\theta)/2, \lambda/(2\log(2e/\lambda))\}$ 成立，因为 $H_2(p) \leq p\log(e/p)$ 在 $(0, 1/2]$ 上成立，且 $\log(2e/\lambda) + \log\log(2e/\lambda) \leq 2\log(2e/\lambda)$。反之，当 $\lambda \leq 0$ 时，不存在满足 (6.30) 的 $\phi_0 > 0$。

证明。固定满足 $\pi(A) \in [s, 1/2]$ 的 A，并取陈述中给出的 ϕ_0，记 $E_A = \{x \in A\} \cap S$ 和 $m_A = \bar{\pi}(E_A)$。由于 $m_A \geq \pi(A) - \bar{\pi}(S^c)$，$\bar{\pi}(S^c) \leq \eta s$ 且 $s \leq \pi(A)$，

$$m_A \geq (1-\eta)\pi(A), \qquad \mathrm{KL}(p_A \| \bar{\pi}) = \log\frac{1}{m_A} \leq \log\frac{1}{\pi(A)} - \log(1-\eta). \tag{6.32}$$

令 ν_{prop} 为提议 $\Upsilon(Z)$ 在 $Z \sim p_A$ 和 $r = \nu_{\mathrm{prop}}(A^c)$ 下的分布。记 $z = (x, v)$，将下方积分限制在 E_A（其中 $\alpha \geq 1-\eta$），并使用 $p_A = \mathbf{1}_{E_A}\bar{\pi}/m_A$，

$$\begin{aligned} Q_K(A, A^c) &= \int_{\mathbb{R}^d \times \mathcal{V}} \mathbf{1}_A(x)\,\alpha(z)\,\mathbf{1}_{A^c}(\Upsilon(z))\,\bar{\pi}(\mathrm{d}z) \geq (1-\eta)\int_{E_A} \mathbf{1}_{A^c}(\Upsilon(z))\,\bar{\pi}(\mathrm{d}z) \\ &\geq (1-\eta)\,m_A \int_{\mathbb{R}^d \times \mathcal{V}} \mathbf{1}_{A^c}(\Upsilon(z))\,p_A(\mathrm{d}z) = (1-\eta)\,m_A\,r \geq (1-\eta)^2\pi(A)\,r \end{aligned}$$

由（6.32），只需证明 $r \geq \phi_0$。

记 $H_A = \log(1/\pi(A))$。在 $x \mapsto \mathbf{1}_A(x)$ 下，分布 ν_{prop} 和 π 变为成功概率分别为 $1-r$ 和 $\pi(A)$ 的伯努利分布；而且相对熵不会增加，因此

$$\begin{aligned} \mathrm{KL}(\nu_{\mathrm{prop}} \| \pi) &\geq \mathrm{KL}\big(\mathrm{Bernoulli}(1-r) \,\big\|\, \mathrm{Bernoulli}(\pi(A))\big) \\ &\geq (1-r)\log\frac{1-r}{\pi(A)} + r\log\frac{r}{1-\pi(A)} \\ &\geq (1-r)H_A + r\log\frac{1}{1-\pi(A)} - H_2(r), \end{aligned} \tag{6.33}$$

其中 $0\log 0 = 0$。另一方面，由（6.29）、（6.32）以及 $-\log(1-\eta) \leq \eta/(1-\eta)$，

$$\mathrm{KL}(\nu_{\mathrm{prop}} \| \pi) \leq \theta H_A + \frac{\theta\eta}{1-\eta} + \delta_0.$$

假设 $r < \phi_0$。结合这两个界，并将 θH_A 移到左边，得到

$$(1-\theta-2r)H_A + r\log\frac{1}{\pi(A)(1-\pi(A))} - H_2(r) \leq \frac{\theta\eta}{1-\eta} + \delta_0, \tag{6.34}$$

其中，我们将 $rH_A + r\log(1/(1-\pi(A)))$ 写作 $r\log(1/(\pi(A(1-\pi(A)))))$，并从第一项中减去 rH_A。左侧的两个系数均非负，因为 $r < \phi_0 \leq (1-\theta)/2$，且由于 $\pi(A) \leq 1/2$，我们有 $H_A \geq \log 2$ 和 $\log(1/(\pi(A)(1-$

π（A））））$\geq 2\log 2$。将每一项替换为其下界后，式（6.34）的前两项至少为（$1-\theta-2r$）log $2+2r\log 2=$（$1-\theta$）log 2，因此

$$H_2(r)\geq(1-\theta)\log 2-\frac{\theta\eta}{1-\eta}-\delta_0\geq H_2(\phi_0)$$

由（6.30）。但是 H_2 在 $[0,1/2]$ 上单调递增，且 $r<\phi_0\leq 1/2$，因此 H_2（r）$<H_2$（ϕ_0），矛盾。因此 $r\geq\phi_0$。 □

6.4. 转移到采样器。现在我们将本节的各项要素结合起来。第6.1 节的能量估计表明，在 $\{H\leq R\}$上（其补集的质量很小），提案至少以 $1-\eta$的概率被接受；命题6.8 给出熵收缩；引理6.9 将二者转化为参考核的受限导通率P_{ref}；引理5.2 则将该界转移到非共振分布 ρ的核 P。

命题 C 的证明。步长条件给出 (6.5)，因此 $a_*+b_h\leq\eta/3$。我们对 P_{ref} 应用引理 6.9，其中 $\mathcal{V}=\mathbb{R}^d\times\{0,\dots,N\}$, $\bar\pi=\pi\otimes\gamma_d\otimes q$, $\Upsilon(x,v,j)=Q_j(x,v)$,$\alpha(x,v,j)=\alpha_j(x,v)$ 以及 $S=\{H\leq R\}$。上述能量估计给出 $\bar\pi(S^c)=\Pi(H>R)\leq\eta s$，并且在 $\{H\leq R\}$ 上有 $\alpha_j\geq 1-\eta$，满足引理的要求。每个 p_A 都支撑在 $\{H\leq R\}$ 上，具有有限相对熵，并且满足 (6.18)，因为 H 关于 v 是偶函数；因此，利用 $T=4/\sqrt{m}$ 的命题 6.8 给出

$$\mathrm{KL}(\mathrm{Law}(x_J)\|\pi)\leq\frac{41}{48}E_0+\frac{5}{3}a_*+\frac{49}{24}b_h\leq\frac{41}{48}E_0+\frac{49}{72}\eta,$$

也就是说，(6.29) 取 $\theta=41/48$和 $\delta_0=49\eta/72$。令 $\phi_0=1/[50(1-\eta)^2]$。那么 $\phi_0\leq 7/96=(1-\theta)/2$，且由于 $\eta=c/600\leq 1/600$,

$$H_2\left(\frac{1}{50(1-\eta)^2}\right)+\frac{41\eta}{48(1-\eta)}+\frac{49}{72}\eta<\frac{7}{48}\log 2,$$

因为左侧关于 η 递增，且在 $\eta=1/600$处小于 $101/1000$，而右侧大于 $101/1000$。这就是 (6.30)，因此 (6.31) 推得 $Q_{P_{\mathrm{ref}}}(A,\ A^c)\geq\pi(A)/50$，其中 $s\leq\pi(A)\leq 1/2$。

令 w 为给定 $\tau\leq\tau_*$时 J 的条件分布。其轨迹的物理积分时间至多为 T_*。对于 $s\leq\pi(A)\leq 1/2$，(6.6) 给出

$$\begin{aligned}\mathrm{Rej}_w(\mathbf{1}_A)&\leq\eta\pi(A)+\Pi(H>R)\\&\leq\eta\pi(A)+\eta s\leq 2\eta\pi(A).\end{aligned}$$

由 (3.1)，引理5.2给出

$$Q_{P_w}(A,A^c)\geq(c/50-2\eta)\pi(A)=\frac{c}{60}\pi(A),$$

因为 $\eta=c/600$。完整的核 P给出此条件核权重 $p_S=\mathbb{P}(\tau\leq\tau_*)$。剩余分量具有非负的平稳流。因此

$$Q_P(A,A^c)\geq\phi_P\pi(A),\qquad\phi_P=\frac{p_S c}{60}>0. \tag{6.35}$$

□

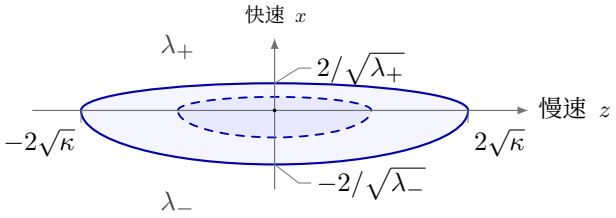


图 4。$h=1$ 和 $\kappa=256$ 的一个快速坐标与慢坐标。曲线分别为 $\psi_h(x)+z^2/(2\kappa)=1/2$ 和 2；它们相对于模态的密度分别为 $e^{-1/2}$ 和 e^{-2}。

7. 固定步长下 Hessian 正则性的必要性。 定理 2表明，对于采用固定 Verlet 步长且经 Metropolis 调整的随机化 HMC，均匀平方根梯度界不能省略 Hessian 正则性。反例中的势函数是强凸的，且具有 Lipschitz 梯度；初始密度至多是目标密度的两倍。较低的接受概率给出了下界。

证明首先针对一个快速坐标构造一组扩张的 Verlet 轨迹。对于具有 d_f 个独立副本的情形，至少一个坐标上的扩张会使得在经过足够多的 Verlet 步后难以接受。Metropolis 检验会将所有坐标作为一个整体接受或拒绝，因此快速坐标上的大幅能量增加可能使慢坐标保持在其初始半直线上。我们对获得第一次接受所需的梯度代价进行界定。

7.1. 目标分布的构造。 该示例区分快速坐标和慢坐标的作用。对于给定的步长 $0<h\le 1$，选择如下所规定的正曲率 $\lambda_->0$ 和 $\lambda_+>\lambda_-$，并令

$$\psi_h(x)=\begin{cases}\lambda_+x^2/2, & x\ge 0,\\ \lambda_-x^2/2, & x<0.\end{cases} \tag{7.1}$$

对于 d_f 个快速坐标和两个额外坐标 w，z，完整的势函数为

$$U_h(x_1,\dots,x_{d_f},w,z)=\sum_{i=1}^{d_f}\psi_h(x_i)+\frac{w^2}{2}+\frac{z^2}{2\kappa}. \tag{7.2}$$

目标位置坐标彼此独立。每个 x_i 在每条半直线上都有高斯密度，且负半轴一侧更宽。慢坐标恰好是 $z\sim N(0,\kappa)$；额外坐标 $w\sim N(0,1)$ 使 ∇U_h 的最小 Lipschitz 常数等于 1。图4展示了一个快速坐标与慢坐标。

7.2. 扩张的 *Verlet* 轨道。 现在我们给出所承诺的 $\lambda_\pm$ 的选取，并证明该定理。首先根据给定的步长定义整数周期长度：

$$p=\left\lceil\frac{2\pi}{h}\right\rceil\ge 7,\qquad p_{\mathrm{cyc}}=3p.$$

轨道在每个周期中会在正侧经过 p 步，在负侧经过 $2p$。在未扰动的角度 $\alpha=\pi/p$ 和 $\beta=\pi/(2p)$ 下，这些恰好是半周期，因此一个完整周期会返回到起始的位置对。将角度扰动参数 $0<\delta\le(80p^2)^{-1}$，其取值将在证明末尾确定。

$$\alpha=\frac{\pi}{p},\quad \beta=\frac{\pi}{2p},\quad a=\alpha+\frac{\delta}{p},\quad b=\beta-\frac{\delta}{4p}. \tag{7.3}$$

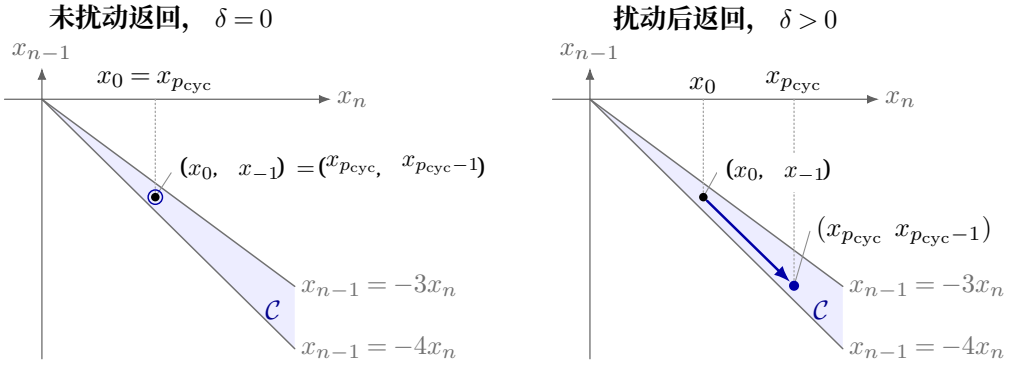


图 5。在位置对平面中经过 p个周期 $= 3p$ 步后返回。未扰动周期返回起点。扰动周期返回到锥上，且第一坐标更大。右图中的位移经过放大。

选择

$$\lambda_+ = \frac{4\sin^2(a/2)}{h^2}, \qquad \lambda_- = \frac{4\sin^2(b/2)}{h^2}, \tag{7.4}$$

并将 ψ_h 简记为 ψ，用于标量计算。函数（7.1）连续可微。 ψ_h' 的每个割线斜率都位于 $[\lambda_-, \lambda_+]$中：跨越零点的割线斜率是两侧斜率的加权平均。因此 ψ_h 是 λ_--强凸的，而 λ_+是 λ_+-光滑的。我们将使用以下界

$$\frac{1}{16} \le \lambda_+ < 1, \qquad \frac{1}{64} \le \lambda_- < \lambda_+, \qquad 2\pi \le ph < 2\pi + 1. \tag{7.5}$$

事实上， $a \ge \pi/p$和 $b \ge 3/\ （2p）$。在 $[0, \pi/2]$上，正弦不等式 $\sin t \ge 2t/\pi$给出 $\lambda_+ \ge 4/\ （ph）$ $^2 \ge 1/16$ 以及 $\lambda_- \ge 9/[\pi^2\ （ph）\ ^2] \ge 1/64$。最后的数值不等式例如可由 $\pi^2 < 10$ 和 $（2\pi + 1）\ ^2 < 54$推出。此外， $\lambda_+ \le [\ （\pi + \delta）\ /\ （ph）\ ]^2 < 1$。

令 Φ 表示对 ψ 执行的一次标量速度 Verlet 步。由于……，位置-速度坐标 $(x_n,\ \ v_n)$ 可以替换为连续位置 $(x_n,\ \ x_{n-1})$。

$$x_{n-1} = x_n - hv_n - \frac{h^2}{2}\psi'(x_n). \tag{7.6}$$

Verlet 递推为

$$x_{n+1} = 2x_n - x_{n-1} - h^2\psi'(x_n).$$

因此，这一对量按如下矩阵演化：

$$\begin{pmatrix} x_{n+1} \\ x_n \end{pmatrix} = A(\theta) \begin{pmatrix} x_n \\ x_{n-1} \end{pmatrix}, \qquad A(\theta) = \begin{pmatrix} 2\cos\theta & -1 \\ 1 & 0 \end{pmatrix}, \tag{7.7}$$

其中，当 $x_n > 0$ 时， $\theta = a$；当 $x_n < 0$ 时， $\theta = b$。

标量 Verlet 递推属于 Lagarias 和 Rains 研究的分段线性保面积映射族 [12]。他们的定理 2.4 对有理旋转数下的动力学进行了分类，其中包括存在扩张射线和发散轨道的情形。我们给出明确的参数选择，并对不变锥、扩张率和中间迭代作出定量估计，进而利用这些结果得到采样下界。图 5展示了锥中一点经过一个周期后的返回情况。

引理 7.1（一个返回锥）。设

$$\mathcal{C} = \{(x, y) : x > 0,\ -4x \le y \le -3x\}.$$

从 $\mathcal{C}$中的任意一对出发，递推(7.7) 使用 $A(a)$ 连续 p 步，随后$A(b)$ 连续 $2p$ 步。经过 $p_{\mathrm{cyc}} = 3p$ 步后，它返回 $\mathcal{C}$。此外，对于每个整数$k \geq 0$,

$$(x_{kp_{\mathrm{cyc}}}, x_{kp_{\mathrm{cyc}}-1}) \in \mathcal{C}, \qquad x_{kp_{\mathrm{cyc}}} \geq \chi^k x_0, \qquad \chi = 1 + \frac{4}{5}\delta b_0, \tag{7.8}$$

其中 $b_0 = \tan(\beta/2)$/(2余弦 β)。一个周期中的每个中间位置的绝对值，至少是该周期开始时正向位置的一半。

证明。我们首先验证中间位置的符号，然后证明锥的不变性和锥的扩张。

未扰动符号序列。对于 $0 < \theta < \pi$，定义

$$C(\theta) = \begin{pmatrix} \cot\theta - \csc\theta \\ \csc\theta - \cot\theta \end{pmatrix}.$$

由于 $A(\theta)$ =余弦 $\theta\, I$ + 正弦 $\theta\, C(\theta)$ 且 $C(\theta)^2 = -I$,

$$A(\theta)^j = \cos(j\theta) I + \sin(j\theta) C(\theta). \tag{7.9}$$

在 $\delta = 0$ 处，经过 j 个正向步后，第一坐标为

$$\frac{\sin((j+1)\alpha)x - \sin(j\alpha)y}{\sin\alpha}.$$

对于 $0 \leq j < p$，它为正，且 $A(\alpha)^p = -I$。从 $(-x, -y)$ 出发，带有 β 的类似公式在接下来的 $2p$ 次迭代中为负，并且 $A(\beta)^{2p} = -I$。所有这些位置的绝对值都至少为 x。为验证最后一个断言，使用 $-y \geq x$ 以及

$$\sin((j+1)\theta) + \sin(j\theta) \geq \sin\theta, \qquad 0 \leq j < \pi/\theta,$$

对于 $\theta = \alpha$或 β。左侧关于 j在 $[0, \pi/\theta - 1]$上是凹函数，并且在两个端点处均等于 $\sin\theta$。

符号序列的保持。 对矩阵使用诱导无穷范数。对于 $0 < \theta < \pi/2$,

$$\|C(\theta)\|_\infty = \cot(\theta/2) \leq \frac{2}{\theta}, \qquad \|C'(\theta)\|_\infty = \frac{1}{2\sin^2(\theta/2)} \leq \frac{\pi^2}{2\theta^2}.$$

由于 $a \geq \alpha = \pi/p$ 和 $b \geq 3/(2p)$，可得

$$\begin{aligned} &\|C(\theta)\|_\infty \leq 2p/3, \quad \|C'(\theta)\|_\infty \leq p^2/2 \qquad (\alpha \leq \theta \leq a), \\ &\|C(\theta)\|_\infty \leq 4p/3, \quad \|C'(\theta)\|_\infty \leq 20p^2/9 \quad (b \leq \theta \leq \beta). \end{aligned} \tag{7.10}$$

根据（7.9），$A(\theta)^j$ 的导数的范数至多为

$$j\big(1 + \|C(\theta)\|_\infty\big) + \|C'(\theta)\|_\infty.$$

乘以 δ/p（对于 $j \leq p$），或乘以 $\delta/(4p)$（对于 $j \leq 2p$），表明每个部分块的变化至多为 $2p\delta$。两个完整块为

$$\begin{aligned} \mathsf{M}_a &= A(a)^p = -\cos\delta\, I - \sin\delta\, C(a), \\ \mathsf{M}_b &= A(b)^{2p} = -\cos(\delta/2) I + \sin(\delta/2) C(b). \end{aligned}$$

具体而言， $\|\mathsf{M}_a + I\|_\infty \le p\delta$ 和 $\|\mathsf{M}_b + I\|_\infty \le p\delta$，分别由 $1 - \cos t \le t^2/2$ 和 $|\sin t| \le t$ 给出，其中$t \ge 0$。在负块期间，比较还包括前一个正块。利用 $\|A(b)^j\|_\infty \le 1 + 4p/3 \le 2p$ 可得

$$\|A(b)^j \mathsf{M}_a + A(\beta)^j\|_\infty \le (2p)(p\delta) + 2p\delta \le \frac{5}{2}p^2\delta.$$

Sinc在 $\|(x,\ y)\|_\infty \le 4x$ 上，扰动在任意迭代步对位置的改变至多为

$$10p^2\delta x \le \frac{x}{8} < x/2. \tag{7.11}$$

因此，所有预设的符号仍然严格成立，并且每个中间绝对值至少为 $x/2$。因此，下面的矩阵乘积描述了由（7.7）生成的轨迹。

返回映射的展开。令 $\mathsf{M}_{\mathrm{lin}} = C(\alpha) - \frac{1}{2}C(\beta)$。我们将证明

$$\mathsf{M}_b\mathsf{M}_a = I + \delta\mathsf{M}_{\mathrm{lin}} + \mathsf{E}_\delta, \qquad \|\mathsf{E}_\delta\|_\infty < \frac{\delta b_0}{20}.$$

我们将利用这个界来证明两个锥边界都向内移动，并且第一坐标增加。记 $\Delta_a = \mathsf{M}_a + I$ 和 $\Delta_b = \mathsf{M}_b + I$。利用 $|\sin t - t| \le t^3/6$ 对 $t \ge 0$，并结合（7.10）中的导数界可得

$$\|\Delta_a + \delta C(\alpha)\|_\infty \le \frac{\delta^2}{2} + \frac{p\delta^3}{9} + \frac{p\delta^2}{2} \le \frac{2}{3}p\delta^2,$$

$$\|\Delta_b - \tfrac{\delta}{2}C(\beta)\|_\infty \le \frac{\delta^2}{8} + \frac{p\delta^3}{36} + \frac{5}{18}p\delta^2 \le \frac{1}{3}p\delta^2.$$

为了界定乘积项，先将矩阵相乘，再取范数：

$$C(b)C(a) = \frac{1}{\sin a \sin b}\begin{pmatrix} \cos a\cos b - 1 & \cos a - \cos b \\ \cos a - \cos b & \cos a\cos b - 1 \end{pmatrix}.$$

由于 $0 < b < a \le \pi/3$ 和 $a/b \le 3$,

$$\|C(b)C(a)\|_\infty = \tan(a/2)\cot(b/2) \le \frac{2a}{\sqrt{3}\,b} < 4.$$

将 M_a 和 M_b 的显示公式代入展开 $\Delta_b\Delta_a$，因此得到

$$\|\Delta_b\Delta_a\|_\infty \le \frac{\delta^4}{16} + \frac{5}{12}p\delta^3 + 2\delta^2 \le 3\delta^2.$$

因此，返回矩阵 $I - \Delta_a - \Delta_b + \Delta_b\Delta_a$ 满足

$$\mathsf{M}_b\mathsf{M}_a = I + \delta\mathsf{M}_{\mathrm{lin}} + \mathsf{E}_\delta, \qquad \|\mathsf{E}_\delta\|_\infty \le (p+3)\delta^2 \le \frac{3}{2}p\delta^2. \tag{7.12}$$

令 $c_0 = \cos\beta$，并利用 $\alpha = 2\beta$，可得精确恒等式

$$\mathsf{M}_{\mathrm{lin}} = b_0\begin{pmatrix} -(1+c_0) & -1 \\ 1 & 1+c_0 \end{pmatrix}, \qquad b_0 = \frac{1-c_0}{2\sin\beta\, c_0} = \frac{\tan(\beta/2)}{2c_0} \ge \frac{\pi}{8p}. \tag{7.13}$$

特别是,

$$\frac{\|\mathsf{E}_\delta\|_\infty}{\delta b_0} \le \frac{12}{\pi}p^2\delta \le \frac{12}{80\pi} < \frac{1}{20}. \tag{7.14}$$

在边界 $y = -4x$ 上， $y + 4x$ 在 $I + \delta\mathsf{M}_{\mathrm{lin}}$ 作用下的一阶变化为 $\delta b_0(9 - 8c_0)x \ge \delta b_0 x$。$\mathsf{E}_\delta$ 产生的误差至多为 $20\|\mathsf{E}_\delta\|_\infty x$，因此像位于该边界的内侧。在 $y = -3x$ 上， $y + 3x$ 的一阶变化为 $\delta b_0(4 -$

$6c_0)x_\circ$ 由于 $c_0 \geq$ 余弦$(\pi/8) > \sqrt{3/2}$, 这至多为 $-(3\sqrt{3}-4)\delta b_0 x$; 误差至多为$16\|\mathsf{E}_\delta\|_\infty x < (4/5)$ $\delta b_0 x_\circ$ 由于 $3\sqrt{3}-4 > 4/5$, 该边界也向内移动。最后，在整个锥内，像的第一坐标至少为

$$\big[1+\delta b_0(2-c_0)-4\|\mathsf{E}_\delta\|_\infty\big]x \geq (1+\tfrac{4}{5}\delta b_0)x = \chi x > 0.$$

对于返回矩阵 $M = \mathsf{M}_b\mathsf{M}_a$，出射斜率是入射斜率的函数$s = y/x$，其表达式为 $(M_{21}+M_{22}s)/(M_{11}+M_{12}s)_\circ$ 其导数为 $\det M/(M_{11}+M_{12}s)^2 > 0$，因为 $\det M = 1$，且出射第一坐标为正。因此，两个边界不等式使得所有出射斜率都属于 $[-4,-3]_\circ$ 返回保持 $\mathcal{C}$不变，迭代即可证明（7.8）。 □

7.3. 能量增长与拒绝。 下面的穿越事件 E 的概率与 h 成正比。其像在一个周期内的并集的概率有一个下界，且该下界与 h 无关。令

(7.15)
$$E = \left\{(x,v) : \frac{h}{10} \leq x \leq \frac{h}{5},\ -4x \leq \cos(a)x - hv \leq -3x\right\}, \qquad E_* = \bigcup_{s=0}^{p_{\mathrm{cyc}}-1} \Phi^{\circ s}(E).$$

令 $H_f(x,\ v) = \psi(x) + v^2/2$，并令 Π_f 为位置和速度的联合分布，其密度正比于 $e^{-H_f(x,\ v)}_\circ$

LEMMA 7.2（一组膨胀轨迹）。 这些 p_{cyc} 在 (7.15) 中的像除边界外互不相交，$H_f < 2$ 在 E_*上，并且

(7.16)
$$\Pi_f(E_*) \geq q_*.$$

对于每个 $z \in E_*$ 以及每个整数 $n \geq 0$,

(7.17)
$$H_f(\Phi^{\circ n}(z)) \geq \frac{1}{20}\chi^{2\lfloor n/p_{\mathrm{cyc}}\rfloor}.$$

证明。在 E 中，速度区间的宽度为 x/h，因此其勒贝格面积为

$$|E| = \int_{h/10}^{h/5} \frac{x}{h}\,dx = \frac{3h}{200}.$$

此外， $0 \leq v \leq 1$ 这里每个像的面积都相同，因为一个 Verlet 步是保体积剪切的复合。若 E 遇到 $\Phi^{\circ j}(E)$ 对于 $0 < j < p_{\mathrm{cyc}}$，一个 E轨迹会在其第一个周期结束前与 $x_j > 0 > x_{j-1}$ 发生第二次交叉。引理7.1排除了这种交叉。由可逆性可知，任意两个像彼此不相交。

对于角度为 θ 的迭代，引入二次型

$$Q_\theta(x,y) = \frac{x^2+y^2-2\cos\theta\, xy}{h^2}.$$

它在 $A(\theta)$ 下是不变量。由 $v =$(余弦 $\theta\, x - y)/h$和 $\lambda = 4$正弦$^2(\theta/2)/h^2$，直接展开可得

(7.18)
$$Q_\theta = v^2 + \lambda(1-h^2\lambda/4)x^2, \qquad \frac{Q_\theta}{2} \leq H_f \leq \frac{Q_\theta}{2(1-h^2\lambda/4)}.$$

在 E, $Q_a \leq 1 + 1/25 = 26/25_\circ$ 在正到负穿越处，有 $xy < 0$ 且余弦$\,b >$ 余弦 $a > 0_\circ$ 因此，从 Q_a变为 Q_b会使该值增加不到两倍，因为 $x^2+y^2 \geq 2|xy|$ 且

$$\frac{x^2+y^2+2\cos b\,|xy|}{x^2+y^2+2\cos a\,|xy|} < 2.$$

在时间 p_{cyc}之前，只有这一次曲率切换。在整个 E_*期间，因此， $Q \le 52/25$。由于 $h^2\lambda < 1$, (7.18) 给出 $H_f < 2$。

Π_f 的归一化常数为

$$Z_f = \pi(\lambda_+^{-1/2} + \lambda_-^{-1/2}) \le 12\pi.$$

利用不相交性、面积保持和 $p_{\mathrm{cyc}}h \ge 6\pi$,

$$\Pi_f(E_*) \ge \frac{3p_{\mathrm{cyc}}h}{200}\frac{e^{-2}}{12\pi} \ge \frac{9}{1200}e^{-2} > 10^{-3} = q_*.$$

为了得到未来界，先从 E 开始，并考虑一个以$(x,\ y) \in \mathcal{C}$开头的周期。在其正向穿越处， $Q_a(x,\ y) \ge 10x^2/h^2$。二次型在每个曲率恒定的区块内保持不变，并在从正到负的切换处增大。因此，直到下一次正向穿越， $H_f \ge 5(x/h)^2$。在下一次正向穿越处，使用新的周期和引理7.1。由于 $x_0 \ge h/10$，这就证明了

$$H_f(\Phi^{\circ n}(z)) \ge \chi^{2\lfloor n/p_{\mathrm{cyc}}\rfloor}/20, \qquad z \in E.$$

如果 $z = \Phi^{\circ s}(z_0) \in E_*$ 且 $0 \le s < p_{\mathrm{cyc}}$，则在时刻 $n+s$ 应用相同的界，并使用 $\lfloor (n+s)/p_{\mathrm{cyc}}\rfloor \ge \lfloor n/p_{\mathrm{cyc}}\rfloor$。 □

拒绝截断。返回乘积目标分布（7.2）。由（7.5），对于 $\kappa \ge 64$，其最优曲率界为 $m = \kappa^{-1}$ 和 $L = 1$。引理 5.1 适用于该目标分布：踢与漂移剪切变换保持体积不变，且 Verlet 步在动量翻转下可逆。因此 P 对于 π是可逆的。令 $\Pi = \pi \otimes N(0,\ I_d)$，并将 H_0 记为从 Π. 抽取的点的能量。

在 Π 下， d_f 位置-速度对 $(x_i,\ v_i)$ 相互独立，因此引理 7.2 给出

$$\Pi(\text{no fast coordinate belongs to } E_*) \le (1-q_*)^{d_f} \le \kappa^{-6}. \tag{7.19}$$

每个标量位置能量都服从 Gamma$(1/2,\ 1)$ 分布。对于非对称坐标，这可通过将任一半直线上的变量作变量代换得到 $\lambda_\pm x^2/2$：在给定任一符号的条件下，得到的都是相同的 Gamma 分布。每个速度能量独立地服从相同的分布。因此H_0 服从 Gamma$(d,\ 1)$ 分布，且 $\mathbb{E}_\Pi e^{H_0/2} = 2^d$。令

$$H_{\mathrm{cut}} = 2d + 12\log\kappa.$$

由马尔可夫不等式可得

$$\Pi(H_0 > H_{\mathrm{cut}}) \le 2^d e^{-H_{\mathrm{cut}}/2} \le \kappa^{-6}. \tag{7.20}$$

令 $\mathcal{B}$ 为（7.19）和（7.20）中的两个例外相位事件的并集。那么

$$\Pi(\mathcal{B}) \le 2\kappa^{-6}. \tag{7.21}$$

定义

$$\ell_\kappa = \log\bigl(20[H_{\mathrm{cut}} + 6\log\kappa]\bigr), \qquad j_0 = p_{\mathrm{cyc}}\left\lceil \frac{\ell_\kappa}{2\log\chi}\right\rceil. \tag{7.22}$$

物理积分时间所对应的截断为 $t_0 = j_0 h$。如果一个进入的相位点位于 $\mathcal{B}$之外，则至少有一个快速坐标满足（7.17）。对于每个 $Jh \ge t_0$，其最终能量至少为 $H_{\mathrm{cut}} + 6\log\kappa$。所有其他坐标的能量均为非负，而初始总能量至多为 H_{cut}。因此，哈密顿能量的变化 ΔH 满足

$$\Delta H \ge 6\log\kappa, \qquad 1 \wedge e^{-\Delta H} \le \kappa^{-6} \quad \text{on } \mathcal{B}^c \cap \{J \ge j_0\}. \tag{7.23}$$

被拒绝的转移会使位置保持不变，而下一次尝试会重新抽取动量。因此，我们需要一个以位置为条件的接受率界。对于 $V \sim N(0,\ I_d)$，定义

$$p_{\mathrm{bad}}(x) = \mathbb{P}_V((x,V) \in \mathcal{B}), \qquad \mathcal{G} = \{x : p_{\mathrm{bad}}(x) \le \kappa^{-2}\}.$$

由（7.21）和马尔可夫不等式可得，

$$\pi(\mathcal{G}^c) \le 2\kappa^{-4}. \tag{7.24}$$

在每个固定的 $x \in \mathcal{G}$ 下，一次新的尝试接受的概率至多为

$$a_0 = \min\{1, \mathbb{P}(J < j_0) + \kappa^{-2} + \kappa^{-6}\}. \tag{7.25}$$

相关的比较是在 t_0 与持续时间尺度 T 之间进行的。由于 $J \ge \lfloor \tau/h \rfloor$，

$$\mathbb{P}(J < j_0) \le \mathbb{P}(\tau < t_0) \le \min\{1, 2t_0/T\} \le \min\{1, t_0/\sqrt{\kappa}\}.$$

每个周期所需的物理时间为 $p_{\mathrm{cyc}}h = 3ph$。界 $b_0 < 1$ 和 $\delta \le (80p^2)^{-1}$ 给出$0 < \log \chi \le 1/32$。由于 $3ph < 22$ 且 $\ell_\kappa \ge 1$，式（7.22）给出

$$t_0 + 2 \le \frac{3ph}{2}\frac{\ell_\kappa}{\log \chi} + 3ph + 2 \le \left(11 + \frac{24}{32}\right)\frac{\ell_\kappa}{\log \chi} < 12\,\frac{\ell_\kappa}{\log \chi}.$$

式（7.25）中的两个例外概率之和至多为 $2/\sqrt{\kappa}$。结合短持续时间界，可得

$$a_0 \le \min\left\{1, \frac{12\ell_\kappa}{\sqrt{\kappa}\log \chi}\right\}. \tag{7.26}$$

7.4. **混合与梯度下界。** 设 $A_+ = \{z \ge 0\}$，其中 z 是（7.2）中的慢坐标，并设

$$\nu = \pi(\cdot \mid A_+) = 2\mathbf{1}_{A_+}\pi. \tag{7.27}$$

其相对于 π 的密度的 $L^2(\pi)$ 范数为 $\sqrt{2}$，且 $\mathrm{KL}(\nu\|\pi) = \log 2$。

由（7.24）及 $d\nu/d\pi \le 2$，

$$\mathbb{P}_\nu(X_0 \notin \mathcal{G}) \le 4\kappa^{-4}. \tag{7.28}$$

在首次接受之前，位置始终精确地保持为 X_0。在 $\{X_0 \in \mathcal{G}\}$，每次转移被接受的条件概率至多为 a_0，无论此前是否遭到拒绝，因为每次尝试时，动量、持续时间、舍入变量以及用于 Metropolis 检验的均匀随机变量都是独立抽取的。

固定一个整数 $n \ge 0$。对于 $1 \le i \le n$，令 F_i 为以下事件：$X_0 \in \mathcal{G}$ 且前 $i-1$ 次转移均被拒绝。事件 F_i 发生且第 i 次转移被接受的概率至多为$a_0\mathbb{P}(F_i) \le a_0$。对 $i = 1, \ldots, n$ 求和并利用（7.28）可得

$$\mathbb{P}(\text{an acceptance occurs in the first } n \text{ transitions}) \le 4\kappa^{-4} + na_0. \tag{7.29}$$

在其补事件上，输出仍属于 A_+。由于 $\pi(A_+) = 1/2$，

$$\mathrm{TV}(\nu P^n, \pi) \ge \frac{1}{2} - 4\kappa^{-4} - na_0. \tag{7.30}$$

因此，对于 $\kappa \ge 64$，全变差精度 $1/4$ 要求

$$n \ge \frac{1}{5a_0} \ge \frac{1}{5}\max\left\{1, \frac{\sqrt{\kappa}\log \chi}{12\ell_\kappa}\right\}. \tag{7.31}$$

梯度计数。第 i 次转移需要进行 J_i 次新的梯度计算。由于 $\mathbb{E}\tau \geq T/3$ 和 $J \geq \tau/h - 1$,

$$\mathbb{E}J \geq \frac{2\sqrt{\kappa}}{3h} - 1 \geq \frac{\sqrt{\kappa}}{2h}, \qquad \kappa \geq 64. \tag{7.32}$$

因此,

$$\mathbb{E}N_\nabla \geq \sum_{i=1}^{n} \mathbb{E}J_i = n\mathbb{E}J. \tag{7.33}$$

结合（7.32）、（7.31）以及 $2\pi \leq ph < 2\pi + 1$ 可得

$$\mathbb{E}N_\nabla \geq \frac{1}{10(2\pi+1)} \max\left\{ p\sqrt{\kappa}, \frac{p\kappa \log \chi}{12\ell_\kappa} \right\}.$$

这两项分别对应一次尝试的代价，以及因拒绝而需要进行额外尝试的代价。较小的步长会增加第一项代价；下面选择的扰动则量化了较大步长下的拒绝代价。取 $\delta = （80p^2）^{-1}$。由（7.8）、（7.13）以及$\log（1+t）\geq 15t/16$ 对于 $0 \leq t \leq 1/15$,

$$\log \chi \geq \frac{3}{4}\delta b_0 \geq \frac{3\pi}{2560p^3}.$$

此外，$d \leq 6q_*^{-1}\log\kappa + 3$ 可得 $H_{\mathrm{cut}} + 6\log\kappa = 2d + 18\log\kappa \leq （13/q_*）（1+\log\kappa）$。因此

$$\ell_\kappa \leq \log(260/q_*) + \log(1 + \log\kappa) \leq 8\Lambda_\kappa,$$

其中最后一个不等式利用了 $\log（260/q_*）< 13 < 7\Lambda_\kappa$，对于 $\kappa \geq 64$。代入这些界可得

$$\mathbb{E}N_\nabla \geq \max\left\{ c_3 p\sqrt{\kappa}, \frac{c_4\kappa}{p^2\Lambda_\kappa} \right\}, \tag{7.34}$$

其中

$$c_3 = \frac{1}{10(2\pi+1)}, \qquad c_4 = \frac{\pi}{81920}\, c_3.$$

取两项的加权几何平均值，权重为 $2/3$ 和 $1/3$，可消去 p并得到（2.13）。定理 2得证。

初始分布 ν 可以显式采样。要采样一个快速坐标，以概率

$$\frac{\sqrt{\lambda_-}}{\sqrt{\lambda_+} + \sqrt{\lambda_-}},$$

然后在正符号情况下将坐标设为 $|Z|/\sqrt{\lambda_+}$，在负符号情况下设为 $-|Z|/\sqrt{\lambda_-}$，其中 Z 服从标准正态分布。抽取 w 个标准正态分布样本，并抽取 $z = \sqrt{\kappa}\,|Z_0|$ 个服从 Z_0标准正态分布的样本。这些独立抽样恰好给出（7.27）。

附录 A：光滑化与趋于极限

我们完成引理6.6证明中的正则化论证。

引理 A.1（平滑后概率律的正则性）。假设 *(A1)*和 *(A2)*。设 p 是 $\mathbb{R}^{2d}$ 上的紧支撑概率测度，且满足 KL 散度$(p\|\Pi) < \infty$。对于 $(X,V) \sim p$ 以及一个独立的 $\xi \sim \gamma_d$，令

$$p^\iota = \mathrm{Law}(X + \sqrt{\iota}\xi, V), \qquad 0 < \iota \le 1,$$
$$p_t^\iota = \mathrm{Law}(Y_t, V), \qquad Y_t = X + tV + \sqrt{\iota}\xi.$$

令 n_t 为 $\nu_t^\iota = \mathrm{Law}(Y_t)$ 的密度，并令 $u_t(y) = \mathbb{E}[V \mid Y_t = y]$ 以及 $\Sigma_t(y) = \mathrm{Cov}(V \mid Y_t = y)$。在每个有限时间区间 $[0,b]$ 上，以下结论成立。

(a) 场 u_t 和 Σ_t 在(t,y) 上光滑，并且与其一阶空间导数一起，在 $[0,b]\times\mathbb{R}^d$ 上有界。

(b) 存在某个 $C > 0$,

$$|\nabla \log n_t(y)| \le C(1 + |y|), \qquad t \in [0,b],\ y \in \mathbb{R}^d.$$

(c) 存在某些 c、$C > 0$,

$$c\exp(-C(1+|y|^2)) \le n_t(y) \le C\exp(-c|y|^2), \qquad t \in [0,b],\ y \in \mathbb{R}^d.$$

常数*(a)*、*(b)* 和*(c)*中的常数可能取决于 ι。

(d) （(6.11)）中出现的所有熵对于平滑后的概率律都是有限的，并且

$$\sup_{\substack{0<\iota\le 1\\ 0\le t\le b}} \left\{\mathrm{KL}(p_t^\iota\|\Pi) + \mathrm{KL}(\nu_t^\iota\|\pi) + E_v(p_t^\iota) + \left|2E_v(p_t^\iota) - \Delta\mathcal{L}(\nu_t^\iota)\right|\right\} < \infty.$$

证明。给定 $Y_t = y$ 时，$(X,\ V)$ 的条件分布相对于 p 的权重正比于 $\exp(-|y - X - tV|^2/(2\iota))$。对任意有界标量函数 $F(X,\ V)$ 的归一化期望求导，得到

$$\nabla_y \mathbb{E}[F \mid Y_t = y] = \iota^{-1}\,\mathrm{Cov}(F, X + tV \mid Y_t = y).$$

将此结果应用于 $F = V_i$ 和 $F = V_iV_j$。$(X,\ V)$ 有界，因此得出关于 $u_t,\ \Sigma_t$ 及其空间导数的所述界。通过在紧集上求导可知，这些矩关于 $(t,\ y)$ 光滑。得分函数为 $\nabla\log\, n_t(y) = \iota^{-1}(\mathbb{E}[X + tV \mid Y_t = y] - y)$，具有线性增长。最后，$n_t$ 是一族高斯分布的平移混合，其协方差为 ιI_d，且在紧时间区间上中心有界。因此，在这些区间上，它下界为 $c\exp(-C(1+|y|^2))$，上界为 $C\exp(-c|y|^2)$，其中 $c,\ C > 0$ 可依赖于 ι 和该区间。

记微分熵为 h。由于能量在 p 的支集上有界，有限的联合相对熵意味着 $\mathrm{h}(p)$ 有限。剪切变换的雅可比行列式为 1，而空间卷积通过凹性增加熵，因此 $\mathrm{h}(p_t^\iota) = \mathrm{h}(p^\iota) \ge \mathrm{h}(p)$。由于 Π 具有密度 e^{-H}/Z_Π，且具有Z_Π 归一化常数，对于相空间上的任意分布 q

$$\mathrm{KL}(q\|\Pi) = -\mathrm{h}(q) + \mathbb{E}_q H + \log Z_\Pi.$$

将此应用于 $q = p_t^\iota$。熵项的上界为 $-\mathrm{h}(p)$。由于 H 根据(A2) 至多具有二次增长，且平滑后的概率律对于 $0 < \iota \le 1$ 和 $0 \le t \le b$具有一致有界的二阶矩，因此能量项有界。于是 $\mathrm{KL}(p_t^\iota\|\Pi)$ 在此范围内一致有界。根据链式法则 $E_{\mathrm{joint}} = E_{\mathrm{pos}} + E_v$，两项均非负，因此 E_{pos} 和 E_v 均由联合熵控制。最后，根据（6.8），$2E_v - \Delta\mathcal{L} = 2E_v - E_{\mathrm{pos}} + U_\Delta$，且 $|U_\Delta|$ 由相同的增长性和矩界可知有界，因此该被积函数在 $[0,b]$上一致有界。

□

对每个固定的 t，联合分布 p_t^ι 及其位置边缘分布 ν_t^ι 在 W_2 下分别收敛至 p_t 和 ν_t。为从平滑后的分布的 (6.11) 推出原始分布的 (6.11)，令$\iota \to 0$ 并使用以下收敛断言。

(i) 联合微分熵和位置微分熵均收敛。凹性给出卷积下的下界，而在 W_2 收敛下的上半连续性则给出反向的极限不等式。后者可由相对熵对固定高斯分布的下半连续性以及二阶矩的收敛推出。

（ii）能量期望通过二次增长和连续性收敛，因此（6.11）右侧通过控制收敛定理收敛。

（iii）端点流 $\mathcal{J}(p_a^\iota)$ 和 $\mathcal{J}(p_b^\iota)$ 通过引理6.6证明中步骤 1 的连续性论证收敛到 $\mathcal{J}(p_a)$ 和 $\mathcal{J}(p_b)$，其中用 $\iota \to 0$ 代替 $s \to t$。
这就证明了原始分布的（6.11）。最后，$\mathcal{W}' = \mathcal{J}(p_t)$、链式法则和（6.8）给出（6.12）。

附录 B：高斯目标分布的梯度计算次数下界

命题 B.1. 设 $0 < m \le L$ 和 $U(x_1, x_2) = \frac{1}{2} m x_1^2 + \frac{1}{2} L x_2^2$。考虑位置核 P，其 *Verlet* 步长 h 满足 $h\sqrt{L} \le 1$。积分时间 J 可以服从任意分布 $0 < \mathbb{E}J < \infty$，并在每次转移时重新抽取，且独立于位置、速度和 *Metropolis* 均匀随机变量。令 $A = \{x_1 \le 0\}$ 和 $\nu_0 = 2\mathbf{1}_A \pi$。对于任意整数 $n \ge 0$，

$$\mathrm{TV}(\nu_0 P^n, \pi) \le \tfrac{1}{4} \quad \Longrightarrow \quad \mathbb{E}C_\nabla \ge n\mathbb{E}J \ge \frac{3\pi}{16 h\sqrt{m}} \ge \frac{3\pi}{16}\sqrt{\kappa}, \tag{B.1}$$

其中 C_∇ 是 (2.11)中的梯度计数。

证明。令 $a = h\sqrt{m}$, $r = \sqrt{1 - a^2/4}$ 和 $\theta = 2\arcsin(a/2)$。设 (x_1, x_2) 为转移开始时抽取的位置，(v_1, v_2) 为抽取的速度。两个坐标在 Verlet 方法下独立演化，求解第一坐标中的标量递推可得，经过 j 步后其提议值为 $Y_j/\sqrt{m}$，其中

$$Y_j = \cos(j\theta) X + \frac{\sin(j\theta)}{r} V, \qquad X = \sqrt{m}\, x_1, \quad V = v_1.$$

在 Π、X，V是相互独立的标准正态变量。设 $\vartheta_j \in [0, \pi]$ 为 $(1,\ 0)$ 与 $(\cos(j\theta),\ \sin(j\theta)/r)$ 之间的角度。旋转对称性给出 $\mathbb{P}(X \le 0,\ Y_j > 0) = \vartheta_j/(2\pi)$。由于接受只会移除这些交点，并且 $\pi(A) = 1/2$,

$$\frac{\int_A K_j(x, A^c)\, \pi(\mathrm{d}x)}{\pi(A)} \le \frac{\vartheta_j}{\pi}.$$

$(\cos t,\ \sin t/r)$ 的连续极角的导数为 $r/(r^2\cos^2 t + \sin^2 t) \le 1/r$，因此 $\vartheta_j \le j\theta/r$。此外，$\theta \le a/r$ 和 $r^2 \ge 3/4$，因为 $a \le h\sqrt{L} \le 1$。因此，对 J 取平均可得

$$\frac{\int_A P(x, A^c)\, \pi(\mathrm{d}x)}{\pi(A)} \le \frac{4}{3\pi} h\sqrt{m}\, \mathbb{E}J. \tag{B.2}$$

每个 K_j 由引理 π 保持5.1，因此 P也如此。由于 $\nu_0 \le 2\pi$ 作为测度相等，且 P为正且 π-不变，$\nu_0 P^i \le 2\pi P^i = 2\pi$ 对每个 $i \ge 0$均成立。对于从(X_i)出发的 ν_0位置链，有

$$\mathbb{P}(X_i \in A, X_{i+1} \in A^c) \le 2\int_A P(x, A^c)\, \pi(\mathrm{d}x) \le \frac{4}{3\pi} h\sqrt{m}\, \mathbb{E}J.$$

由于 $X_0 \in A$，在时间 n 之前到达 A^c 至少需要发生一次这样的穿越：

$$\{X_n \in A^c\} \subseteq \bigcup_{i=0}^{n-1} \{X_i \in A,\, X_{i+1} \in A^c\},$$

因此，利用并集界可得 $\mathbb{P}(X_n \in A^c) \le \sum_{i<n} \mathbb{P}(X_i \in A,\ X_{i+1} \in A^c) \le \frac{4}{3\pi} h\sqrt{m}\, n\mathbb{E}J$。在集合 A^c 上检验全变差距离，该集合满足 $\pi(A^c) = 1/2$ 且 $\nu_0 P^n(A^c) = \mathbb{P}(X_n \in A^c)$，可得

$$\mathrm{TV}(\nu_0 P^n, \pi) \ge \pi(A^c) - \nu_0 P^n(A^c) = \tfrac{1}{2} - \mathbb{P}(X_n \in A^c) \ge \tfrac{1}{2} - \frac{4}{3\pi} h\sqrt{m}\, n\mathbb{E}J.$$

这就证明了（B.1），因为 $\mathbb{E}C_\nabla \ge \mathbb{E}\sum_{i=1}^n J_i = n\mathbb{E}J$ 由（2.11）可得。最后，$h\sqrt{m} = (h\sqrt{L})/\sqrt{\kappa} \le 1/\sqrt{\kappa}$ 给出了最后一个界。 □

附录 C：三种持续时间分布的傅里叶界

证明 引理5.4。只需考虑 $0 < \theta < \pi$。令 $u = \theta/2$ 且 $B_N = \sin(Nu)/(N\tan u)$。对权重（5.4）求和可得

$$\mathrm{Re}\,\widehat{q}(\theta) = \frac{\sin^2(Nu)}{N^2 \sin^2 u} - \frac{\sin^2(Nu)}{N^2} = B_N^2.$$

端点处的取值由连续性可得。因此，对于三角分布 $w = q$，$c = \beta_P = 1$ 均成立。谱下界在 $\theta = \pi$ 处取到。

对于均匀持续时间，对取整区间积分可得

$$\mathrm{Re}\,\widehat{\mu}_J(\theta) = \begin{cases} B_N, & N \text{ even}, \\ \cos u\, B_N, & N \text{ odd}. \end{cases}$$

对于偶数 N，从 $1 - B_N$ 中减去 $(1 - B_N^2)/2$，得到 $(1 - B_N)^2/2$。对于奇数 N，相应的差为 $[(B_N - \text{余弦 } u)^2 + \text{正弦}^2 u]/2$。因此，对每个$N$，$c = 1/2$ 均成立。对于偶数 N，当 $\theta \to 0$ 时，该比值趋于 $1/2$，因此在这种情况下，比较常数是最优的。此外，$B_N = (u/\text{ 正切 } u)\,\mathrm{sinc}(Nu)$，其中 $\mathrm{sinc}\ x = \text{正弦 } x/x$（当 $x \ne 0$ 时），且 $\mathrm{sinc}\ 0 = 1$。不等式 $\mathrm{sinc}\ x \ge -1/4$ 对 $x \ge 0$ 成立：在 $[0, \pi]$ 上显然成立；在 $[\pi, 4]$ 上，利用正弦 $x \ge -(x - \pi)$ 及 $\pi > 3$；在 $[4, \infty)$ 上，利用正弦 $x \ge -1$。$u/$ 正切 u 和余弦 u 均属于 $[0, 1]$，由此可证 $\beta_P = 3/4$ 成立。

对于指数分布持续时间，令 $r = e^{-4/N}$ 和 $z = e^{i\theta}$。完整的特征函数及其实部为

$$\widehat{\mu}_J(\theta) = \frac{(1+z)(1-r)}{2(1-rz)}, \qquad \mathrm{Re}\,\widehat{\mu}_J(\theta) = \frac{\tanh^2(2/N)}{\tanh^2(2/N) + \tan^2 u}.$$

实部非负，且在 $\theta = \pi$ 处为零，因此 $\beta_P = 1$。我们先证明全分布比较

$$1 - \mathrm{Re}\,\widehat{\mu}_J(\theta) \ge \frac{5}{8}(1 - B_N^2). \tag{C.1}$$

令 $v = N\tan u$，$x = Nu$ 和 $y = v^2$。由于 $N\tanh(2/N) \le 2$，左侧至少为$y/(4+y)$。对于 $y \ge 20/3$，该值至少为 $5/8$，这就足够了。对于 $0 \le y \le 7$，我们有 $x \le v < \pi$，并且

$$\frac{x}{v} = \frac{\arctan(v/N)}{v/N} \ge 1 - \frac{y}{768}, \qquad \mathrm{sinc}\, x \ge \mathrm{sinc}\, v \ge f(y) := 1 - \frac{y}{6} + \frac{y^2}{120} - \frac{y^3}{5040} > 0.$$

第一个不等式用到了 $N \ge 16$，最后一个不等式则是正弦函数的交错泰勒界。因此 $B_N^2 \ge Q(y) := (1 - y/768)^2 f(y)^2$。对于多项式

$$\mathcal{P}(y) = 1 - \frac{5}{8}(4+y)\frac{1-Q(y)}{y},$$

及其在零点处的连续取值，$\mathcal{P}(y) \ge 1/250$ 在 $[0,7]$上。事实上，在每个区间 $[j, j+1]$，$j = 0, \dots, 6$上，$\mathcal{P}$的八次伯恩斯坦系数都是可以精确计算的有理数，且均大于 $1/250$；其中最小值出现在 $[3,4]$, 上，约为0.0048。由于多项式在一个区间上的下界不小于其伯恩斯坦系数的最小值，这就证明了（C.1），其中重叠范围为 $[0,7]$ 和 $[20/3, \infty)$。

接下来需要以 $\tau \le 8/\sqrt{m}$ 为条件。令 $t = e^{-8}$。根据无记忆性以及 $8/(h\sqrt{m}) = 2\overline{N}$，

$$\widehat{w}(\theta) = \widehat{\mu}_J(\theta)\frac{1 - te^{2N i\theta}}{1-t}.$$

利用 $1 - \text{余弦}(a+b) \le 2(1 - \text{余弦}\, a) + 2(1 - \text{余弦}\, b)$ 可得

$$1 - \operatorname{Re}\widehat{w}(\theta) \ge \frac{(1-2t)(1 - \operatorname{Re}\widehat{\mu}_J(\theta)) - 2t(1 - \cos(2N\theta))}{1-t}.$$

对于 $x = N\theta/2$，我们有

$$1 - B_N^2 \ge \frac{2}{7}\min\{x^2, 1\}, \qquad 1 - \cos(2N\theta) = 1 - \cos(4x) \le \min\{8x^2, 2\}.$$

对于 $[0,1]$, 上的第一个不等式，使用 sinc 函数 $x \le 1 - 19x^2/120$，得到 $1 - \text{sinc 函数}^2\, x \ge 4199x^2/14400 > 2x^2/7$。对于 $x \ge 1$，$|\text{sinc 函数 } x| \le \sin 1$ 在 1 处给出相同的界。因此，第二个表达式与 $1 - B_N^2$ 的比值至多为 28。我们得出

$$\frac{1 - \operatorname{Re}\widehat{w}(\theta)}{1 - \operatorname{Re}\widehat{q}(\theta)} \ge \frac{(1-2t)5/8 - 56t}{1-t} \ge \frac{3}{5}.$$

最后一个数值不等式由 $t < 1/2980$ 得出，并由 $e^8 \ge \sum_{j=0}^{30} 8^j/j! > 2980$ 认证。此外，$e^{-8} < 3/16$，因此这个截断满足 $\delta_\rho = 3/4$ 的省略质量要求。 □

致谢。感谢AndreWibisonofor对本文初稿提出意见，其中包括备注 2.2 中的观察。本文中的想法均为我本人所有。我使用了公开可用的大语言模型来协助计算、技术性引理的证明、数值实验、图表绘制、起草以及 Lean 形式化。

参考文献

[1] AMBROSIO, L.、GIGLI, N. 和 SAVARÉ, G.（2008）。《梯度流：度量空间与概率测度空间》，第2版。苏黎世联邦理工学院数学讲义。巴塞尔：Birkhäuser。[2] ANDRIEU, C.、LEE, A.、POWER, S. 和 WANG, A. Q.（2023）。马尔可夫链的弱 Poincaré 不等式：理论与应用。arXiv: 2312 11689 1 2312 11689 1。v。[3] https://arxiv.org/abs/。v APERS, S.、GRIBLING, S. 和 SZILÁGYI, D.（）。用于高效高斯采样的哈密顿蒙特卡罗：长步与随机步。Journal of Machine Learning Research 25 1–30。https://jmlr.org/pap ers/v25/23-1521.html[4] BOU-RABEE, N.（2026）。《Persistence Pays：Metropolis 校正的动力学 Langevin 蒙特卡罗中平方根条件数依赖与非可逆更新的接受/拒绝均匀随机变量》。arXiv: 2610 35781 2610 35781。。[5] https://arxiv.org/abs/。BOU-RABEE, N.、CARPENTER, B.、KLEPPE, T. S. 和 LIU, S.（2026）。轨道内自适应蛙跳无转向采样器。Journal of Machine Learning Research 27 1 1 452 64 27 25 –。https://jmlr.org/papers/v / - .html

[6] BOU-RABEE, N.、CARPENTER, B. 和 MARSDEN, M.（2026）。GIST：Gibbs自调节的局部自适应哈密顿蒙特卡罗。Statistics Surveys **20** 135 179 10 1214 26 156 –。https://doi.org/ 。/ -SS [7] BOU-RABEE, N. 和 SANZ-SERNA, J. M.（2017）。随机化哈密顿蒙特卡罗。Annals of Applied Probability **27** 2159 2194 10 1214 16 1255 –。https://doi.org/ 。/ -AAP [8] CHEN, Y.、GATMIRY, K. 和 JIANG, M.（2023）。何时 Metropolized哈密顿蒙特卡罗能够在理论上优于 Metropolis调整的Langevin算法？ arXiv:2304.04724。https://arxiv.org/abs/2304.047 24 [9] FU, Q. 和 WIBISONO, A.（2025）。用于优化的哈密顿下降算法：通过随机化积分时间实现加速率。在 Advances in Neural Information Processing Systems 38 93628–93685。https://arxiv.org/abs/ 2505 12553 。HOFFMAN, M. D. 和 GELMAN, A.（2014）。[10] No-U-Turn采样器：在哈密顿蒙特卡罗中自适应设置路径长度。Journal of Machine Learning Research 15 1593–1623。https://jmlr.org/paper s/v15/hoffman14a.html [11] HOLLEY, R. 和 STROOCK, D.（1987）。对数Sobolev不等式与随机Ising模型。Journal of Statistical Physics **46** 1159 1194 10 1007 01011161 –。https://doi.org/ 。/BF [12] LAGARIAS, J. C. 和 RAINS, E.（2005）。一族分段线性保面积平面映射的动力学 I：有理旋转数。Journal of Difference Equations and Applications 11 1089–1108。https://arxiv.org/abs/math/0301294 [13] LEE, Y. T.、SHEN, R. 和 TIAN, K.（2020）。对数光滑梯度集中与 Metropolized哈密顿蒙特卡罗 更紧的运行时间。在 Proceedings of the 33rd Conference on Learning Theory。Proceedings of Machine Learning Research 125 2565–2597。https://proceedings.mlr.press/v125/lee20b. html [14] LEE, Y. T.、SHEN, R. 和 TIAN, K.（2021）。适定条件分布上 Metropolized采样方法的下界。在 Advances in Neural Information Processing Systems 34 18812–18824。https://arxiv.org/abs/ 2106 05480 。LU, J. 和 WANG, L.（2022）。关于 MCMC算法中分段确定性马尔可夫过程的显式 L^2-收敛速率估计。Annals of Applied Probability 32 1333–1361。https://doi.org/10.1214/21-AAP1710 [15] MACKENZIE, P. B.（1989）。一种改进的混合蒙特卡罗方法。Physics Letters B 226 369–371。https://doi.org/ 10 1016 0370 2693 89 91212 4 。/ -（）- MITRA, S.、SRINIVASAN, V.、WANG, X. 和 WIBISONO, A.（2026）。随机化哈密顿蒙特卡罗的加速混合时间。arXiv: 2607 12902 2607 12902 。。https://arxiv.org/abs/ 。[17] NEAL, R. M.（2011）。使用哈密顿动力学的 MCMC。在 Handbook of Markov Chain Monte Carlo（S. Brooks、A. Gelman、G. L. Jones 和 X. L. Meng 编）113–162。Chapman & Hall/CRC。https: //arxiv.org/abs/ 1206 1901 。NEAL, R. M.（）。为 Metropolis接受/拒绝决策非可逆地更新均匀分布 $[0, 1]$ 值。[18] arXiv: 2001 11950 2001 11950 。。https://arxiv.org/abs/ 。OBERDÖRSTER, S.（2020）。通过 Dobrushin收缩与加速哈密顿蒙特卡罗实现高概率均匀混合。arXiv: 2611 21305 2611 21305 。。https://arxiv.org/abs/ 。[20] OTTO, F. 和 VILLANI, C.（2026）。Talagrand不等式的推广及其与对数Sobolev不等式的联系。Journal of Functional Analysis 173 361–400。https://doi.org/10.1006/jfan.1999.35 57 [21] VILLANI, C.（2000）。《最优传输：过去与现在》。Grundlehren der mathematischen Wissenschaften 338 10 1007 978 3 540 71050 9 。Springer。https://doi.org/ 。/ - - - - [22] WANG, J. K. 和 WIBISONO, A.（2009）。通过 Chebyshev积分时间加速哈密顿蒙特卡罗。在 Proceedings of the 11th International Conference on Learning Representations。https://arxiv.or g/abs/ 2207 02189 。WANG, X.、SRINIVASAN, V.、FU, Q.、MITRA, S.、WILSON, A. 和 WIBISONO, A.（2023）。通过具有确定性积分时间的哈密顿动力学加速凸优化。在 Proceedings of the 39th Conference on Learning Theory。Proceedings of Machine Learning Research 336 6692–6742。https://proceedings.mlr.press/v336/wang26c.html [23] WIBISONO, A.（2026）。测度空间中的优化即采样：作为复合优化问题的 Langevin动力学。在 Proceedings of the 31st Conference on Learning Theory。Proceedings of Machine Learning Research 75。https://proceedings.mlr.press/v75/wibisono18a.html [24]