arXiv:2601.09019v1 [stat.ML] 13 Jan 2026

通过单次耦合分析未校正哈密顿蒙特卡罗的尾部敏感 KL 与 Rényi 收敛

Nawaf Bou-Rabee* Siddharth Mitra† Andre Wibisono‡

2026年1月

摘要

哈密顿蒙特卡罗（HMC）算法是高维场景中应用最广泛的采样方法之一，但对于量化相对密度失配的散度（如 Kullback–Leibler（KL）散度和 Rényi 散度），其收敛性质仍缺乏深入理解。这些散度自然决定了 Metropolis 校正马尔可夫链的接受概率和热启动要求。在本文中，我们构建了一个框架，将未校正哈密顿蒙特卡罗（uHMC）的 Wasserstein 收敛保证提升为尾部敏感的 KL 和 Rényi 散度收敛保证。我们的方法基于单次耦合，并利用它建立 uHMC 转移核的正则化性质。该正则化性质使得 Wasserstein–2 混合时间界和渐近偏差界可以提升为 KL 散度界；类似地，Orlicz–Wasserstein 界也可以提升为 Rényi 散度界。这与 Bou-Rabee 和 Eberle（2023）先前的工作相呼应，他们通过核平滑将 Wasserstein–1 界提升为全变差距离界。因此，我们的结果为相对密度失配提供了定量控制，阐明了离散化偏差在强散度中的作用，并为未校正采样以及为 Metropolis 校正马尔可夫链生成热启动提供了有依据的保证。

目录

1 引言 2 预备知识 7 3 单次耦合 12 4 主要结果与示例 16 5 获得 KL 散度保证的关键步骤 25 6 获得 R´ényi 散度保证的关键步骤 26 7 讨论 28 A 哈密顿蒙特卡罗背景知识 31 B 概率核的正则化性质 32 C 通过单次耦合进行正则化 34 D 用于混合时间的单次耦合映射 35

*罗格斯大学数学科学系、Flatiron 研究所计算数学中心。nawaf.bourabee@rutgers.edu

†耶鲁大学计算机科学系。siddharth.mitra@yale.edu

‡耶鲁大学计算机科学系。andre.wibisono@yale.edu

E 用于渐近偏差的单次耦合映射 35F KL 散度下的混合时间 41G KL 散度下的渐近偏差 43H R´ényi 散度下的混合时间 44I R´ényi 散度下的渐近偏差 48J 信息收缩 54K KL 散度和 R´ényi 散度的凸性 54L 通过 uLA 进行交叉正则化 55M 辅助引理 58参考文献 59

1 引言

设 ν 是定义在 $\mathbb{R}^d$ 上、密度与 e^{-f}成正比的绝对连续概率测度。从 ν 中采样是许多应用中的一项基础算法任务，包括计算统计学和机器学习。当无法从 ν 中进行精确采样时，一种标准方法是马尔可夫链蒙特卡罗（MCMC），它构造一个合适的马尔可夫链 $(X_k)_{k\geq 0}$，并利用其迭代生成来自 ν [RC04，AG07，Liu08，DMPS18]的近似样本。

现代马尔可夫链蒙特卡罗算法设计中的一种常见范式，是从相对于 ν 遍历的连续时间随机过程出发，再对其动力学进行离散化；例如参见 [RT96，MSH02，DE24]。此类离散化方法通常会引入偏差：当由此得到的离散时间马尔可夫链具有不变分布 $\tilde{\nu}$时，该分布通常不同于 ν。

消除这种离散化偏差的一种经典方法，是加入 Metropolis–Hastings校正 [MRR+53，Has70]，从而得到一个具有不变分布的校正马尔可夫链 ν。然而，在本文中，我们关注未经校正的方案。作出这一选择有多方面的动机。

首先，在某些应用中，目标分布 ν 本身仅能近似定义，因此，Metropolis校正针对的是近似模型的不变分布，而非理想化的目标分布。例如，相互作用粒子近似中的非线性测度 [Szn91、DM04]；无限维空间上测度的有限维近似（如贝叶斯逆问题和条件扩散） [Stu10、HSVW05]；以及似然通过数值离散化的常微分方程或偏微分方程隐式定义的统计模型 [MM21]。在这些情形下，Metropolis校正无法消除偏差的主要来源，也可能无法有效解决这些问题，因此，未经校正的马尔可夫链蒙特卡罗方法自然成为研究对象。

其次，即使原则上可以进行精确的Metropolis校正，理论分析和经验结果都表明，校正马尔可夫链的性能可能对初始分布的选择十分敏感。如果链从目标分布的尾部初始化，提议通常是局部的，接受概率也可能极低，尤其是在高维情形下。因此，该链可能需要大量迭代才能到达 ν的典型集，从而导致初始收敛缓慢，实际性能下降。

为了获得校正马尔可夫链有意义的收敛保证，并避免实践中接受概率退化，因此必须控制该链分配给目标分布 ν 下罕见区域的概率质量。这通常通过热启动条件来实现，该条件要求初始分布与 ν 足够接近，可用具有尾部敏感性的散度（如 Kullback–Leibler（KL）散度或 R´enyi 散度）来衡量。这类散度

能够控制相对密度失配，并防止初始化在 ν 极小的区域内分配过多概率质量。这类热启动假设是校正型马尔可夫链蒙特卡罗方法理论分析中的标准假设 [DCWY19，CG23，AC24a，SWW24，SWW25，KV25a，KV25b]。

相比之下，仅用全变差距离等较弱度量表述的热启动条件，并不能控制尾部的相对密度。全变差距离通过概率质量的绝对差异，以加法方式控制偏差；而 KL 散度和 R´enyi 散度则通过密度比，以乘法方式控制偏差。因此，某个分布在全变差意义下可能与 ν 十分接近，却仍会在 ν下呈指数级低概率的区域赋予不可忽略的质量。这种失配可能导致接受概率趋于零，并在校正链中形成瓶颈（见图 1）。

直接为校正马尔可夫链获得热启动通常很困难，因为这往往需要事先了解 ν 的典型集。因此，未校正马尔可夫链自然可以发挥辅助作用：当其在尾部敏感散度下具有定量收敛保证时，便可用于高效且严谨地为校正链生成热启动初始化。

从这一角度看，未校正马尔可夫链具有双重作用：在目标本身只是近似分布的情形下，它们可以作为自然的采样方案；同时，它们也是为后续校正算法生成热启动的有效机制。这种两阶段策略——先从冷启动运行未校正链以进入典型集，再切换到校正链进行精确采样——已在理论和实践中证明行之有效[DE24，AC24a，AGS24，STG21，GHS15，SMP19]。

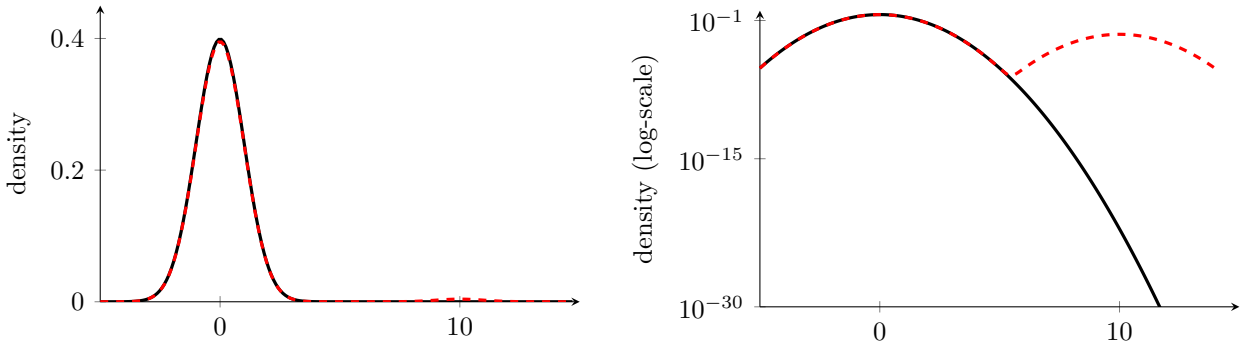
这些考虑自然引出了关于未校正马尔可夫链的两个定量问题，我们将通过尾部敏感的 KL 散度和 R´enyi 散度来研究。第一，X_k的分布以多快的速度收敛到其不变分布 $\tilde{\nu}$？第二，$\tilde{\nu}$ 与预期目标 ν有多接近？前者关乎混合时间保证，后者关乎渐近偏差。二者共同决定了达到平稳性所需的计算成本，以及该链逼近目标分布的精度。

在这项工作中，我们假设 f 可微，且其梯度可以计算。在此背景下，我们为未校正哈密顿蒙特卡罗（uHMC）建立了 KL 散度和 R´enyi 散度下的混合时间界与渐近偏差界。以这些尾部敏感散度表述的保证，尤其适用于为 Metropolis 校正马尔可夫链提供良好的初始状态，因为它们能在引入 Metropolis 校正之前，对相对密度失配进行定量控制。

1.1 主要结果

哈密顿蒙特卡罗算法通过近似模拟与目标密度相关的哈密顿动力学来生成提议 $\nu \propto e^{-f}$。在连续时间下，这些动力学能够精确保持ν，并由此得到精确的哈密顿蒙特卡罗（eHMC）。然而在实践中，必须对哈密顿流进行离散化，从而得到未校正的哈密顿蒙特卡罗方案，其不变分布通常与ν不同。本文始终研究采用速度 *Verlet*（或 leapfrog）积分器 [BRSS18、HHIL06]实现的未校正哈密顿蒙特卡罗。由于该积分器简单、具有时间可逆性和辛结构，因此是哈密顿蒙特卡罗中使用的标准离散化方法。我们将由此得到的马尔可夫链记为 uHMC-v。uHMC-v 的每次转移包括：(i) 高斯速度刷新，随后进行 (ii) 若干个确定性的 Verlet 步，以近似哈密顿流。

我们的主要结果为 uHMC-v 在 Kullback–Leibler（KL散度）和 R´enyi 散度下建立了定量的收敛保证。这些保证具有两种互补形式：混合



(a) 线性–y 图　　(b) 半对数–y 图

图 1：分布 $\pi = \mathcal{N}(0,\ 1)$ 绘制为 [　] 和 $\mu = 0.99\mathcal{N}(0,\ 1) + 0.01\mathcal{N}(10,\ 1)$，如(a)线性–$y$ 轴和(b)半对数–y 轴所示。对于这些分布，$\mathrm{TV}(\mu,\ \pi) \le 0.01$，$\mathrm{KL}(\mu \,\|\, \pi) \ge 0.4$，以及 $\mathrm{R}_2(\mu \,\|\, \pi) \ge 90$。线性–$y$ 尺度显示出较大的重叠（较小的全变差距离），而半对数–y 尺度突出了尾部的不一致性（较大的 KL 散度和 R´ényi–2 散度）。

时间界，用于控制收敛到 uHMC-v 的不变分布；渐近偏差界，则用于量化该不变分布与目标分布之间的差异 ν。相关结果将在第4.1 节和4.2节中介绍。

通过一次性耦合进行正则化。 我们分析所依据的关键技术要素，是 uHMC-v转移核在尾部敏感散度下具有正则化 （或平滑）性质。该性质通过单次耦合建立，它使我们能够将初始分布之间的 Wasserstein 型距离，与经过一次马尔可夫转移后的更强散度联系起来。

为将这一概念形式化，设$\mathbf{P}$ 是 $\mathbb{R}^d$ 上的马尔可夫核， W 表示与给定代价函数相关的 Wasserstein 距离，D 表示某种概率散度或距离。若存在常数 $C > 0$ ，使得

$$\mathrm{D}(\mu\mathbf{P} \,\|\, \pi\mathbf{P}) \ \le\ C\, W(\mu, \pi), \tag{1}$$

对所有概率测度 $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$，均有此性质。

不等式（1）体现了随机更新 $\mathbf{P}$能够平滑奇异或对齐不佳的初始分布这一思想。例如，若（1）在 KL 散度或 R´ényi 散度下成立，那么将 $\mathbf{P}$ 作用于两个不同的狄拉克测度 δ_x 和 δ_y 后，散度将变为有限，而更新前的散度则为无穷大。D的变分表示还能推出其他结论。例如，当（1）在全变差距离下成立时，核$\mathbf{P}$ 会将有界（可能不连续的）测试函数映射为 Lipschitz 函数；参见第B节。

为什么 uHMC-v 的正则化并非易事？ uHMC-v 中的正则化效应源于每次转移开始时的高斯速度刷新。然而，要证明（1）在技术上颇为棘手，因为注入的随机性随后会通过多个确定性的 Verlet 步传播。与在每个离散化步骤中显式注入高斯噪声的格式不同——例如动能朗之万动力学的分裂格式 [Mon21、命题 3 和 22] ，或指数欧拉积分器 [ACZ25、引理 4。2] —uHMC-v 中的随机性必须在经过非线性映射的复合传递后进行分析。

交叉正则化与渐近偏差。要获得渐近偏差界，仅比较 uHMC-v 的两条轨迹是不够的。相反，我们必须比较 uHMC-v 的单步与 eHMC 的单步。这就引出了一个交叉正则化问题，我们需要界定如下形式的散度

$$\mathsf{D}(\mu\mathbf{P} \,\|\, \pi\mathbf{Q}),$$

其中$\mathbf{P}$ 表示 uHMC-v 核，$\mathbf{Q}$ 表示 eHMC 核。

这种比较不可避免，因为 KL 散度和 R´ényi 散度既不满足三角不等式，也不具对称性。所需的界通过一种新颖的单次耦合构造（图 14and19 2(b)）得以建立，参见引理 14 和 19。据我们所知，此类交叉正则化分析此前尚未出现在 uHMC-v 的研究中。

相比之下，在全变差距离下进行分析时——它具有对称性且满足三角不等式——交叉比较任务可简化为对相同初始点的比较；参见 [BRE23，图 1(b) 和定理 7]。针对动能朗之万动力学，已有研究在 KL 散度下获得了比较离散动力学与精确动力学的相关正则化界 [ACZ25，引理 4。2] 此外，针对未校正朗之万算法，也有研究在 KL 散度和 R´ényi 散度下获得了相关结果 [AC24c，引理 4.7 和 C。2], 这些结果通常借助吉尔萨诺夫定理。相比之下，我们的方法依赖耦合论证，并能处理 Verlet 积分器离散流更复杂的结构。

端到端保证与模块化。 通过结合正则化界和交叉正则化界，以及现有的 uHMC-v 的 Wasserstein–2 和 Orlicz–Wasserstein 压缩性估计，我们得到了 KL 散度和 R´ényi 散度的端到端混合时间界和渐近偏差界（第 4.3.1节）。

我们的框架的一个重要特点是其模块化。Wasserstein 收缩不必由 uHMC-v 本身驱动：只要某个离散时间马尔可夫核满足适当的 Wasserstein 压缩性假设，就可以使用，但要求链的最后一步是一次 uHMC-v 更新。我们通过将结果应用于 [BRM25] 中引入的 uHMC 分层积分器（第 4.3.2节），展示了这种灵活性。

最终的梯度复杂度取决于驱动 Wasserstein 收缩的核的选择；参见推论 2 和 5。我们的框架还允许使用不同的核来实现 Wasserstein 收缩和正则化。在第 7节，我们概述了如何将未校正朗之万动力学用作正则化核，并讨论了相关的权衡。

贡献概述。我们的主要贡献如下：

- 我们为 uHMC-v 建立了 KL 散度下的混合时间保证（定理 1），并验证了强对数凹目标所需的 Wasserstein–2 假设。
- 我们推导了 uHMC 在 KL 散度下的渐近偏差界（定理 2），该结果适用于链的最后一步为 uHMC-v 的情形，并将其应用于 uHMC-v 和分层积分器 uHMC-s [BRM25]。
- 基于 Orlicz–Wasserstein压缩性，我们为 uHMC-v 建立了 R´ényi 散度下的混合时间保证（定理 3）。
- 我们推导了 uHMC 在 R´ényi 散度下的渐近偏差界（定理 4），同样假设最后一步为 uHMC-v 更新。
- 我们运用 KL 混合时间界，研究强对数凹目标下 uHMC-v 的信息收缩（第 4.4节）。

1.2 尾部敏感的收敛保证的影响

Kullback–Leibler（KL）散度（定义4）和 R´enyi 散度（定义 5）是表述收敛保证的自然度量，因为它们能够控制其他多种常用统计距离。具体而言，通过 Pinsker不等式，它们可以给出全变差距离的界；在适当的运输不等式下，可以给出 Wasserstein 距离的界；还可以给出 χ^2 散度的界。这些关系可由以下不等式链概括：

$$2\mathsf{TV}^2(\mu,\pi) \le \mathsf{KL}(\mu \,\|\, \pi) \le \log\bigl(1+\chi^2(\mu \,\|\, \pi)\bigr) \le \mathsf{R}_q(\mu \,\|\, \pi),$$

对所有概率测度 μ，π 以及所有 $2 \le q < \infty$均成立。除了这些含义之外，KL散度和 R´enyi 散度中的收敛保证还带来若干进一步的影响：

- 校正采样器的热启动生成。 对尾部敏感散度的控制提供了相对密度失配的定量界，并确保未经校正的马尔可夫链输出的分布适合作为 Metropolis校正采样的初始分布；例如参见 [DCWY19、CG23、AC24a、SWW24、SWW25、KV25a、KV25b]。

- 马尔可夫链中的依赖性。 KL散度收敛控制链上的信息收缩，并给出时间依赖性的定量界，从而刻画获得近似独立样本的速率；参见 [LMW25] 和第 4.4节。

- 差分隐私。 R´enyi 散度在差分隐私迭代算法的分析中发挥核心作用，其中隐私保证通过马尔可夫核的复合与正则化性质获得 [Mir17、FMTT18、GT20、CYS21、ABT24、BFMG25]。

1.3 Th哈密顿蒙特卡罗的理论背景与相关工作 lo

为基于哈密顿动力学的算法（例如精确哈密顿蒙特卡洛（eHMC））建立严格的收敛保证 [DKPR87，Nea11], 未校正哈密顿蒙特卡罗（uHMC）、Metropolis 校正的 HMC 和无转向采样器（NUTS） [HG14] 一直是一个活跃的研究领域。尽管如此，与朗之万动力学及其离散时间对应方法相比，人们对基于哈密顿量的采样方法的理论认识仍明显不足，尤其是在以强尾部敏感散度衡量收敛时。

对于过阻尼朗之万动力学的离散化方法，包括未校正朗之万算法（uLA） [RT96，RR98] 以及近端采样器 [LST21], ，在相对温和的假设下，已有关于 Kullback–Leibler 散度、Rényi 散度乃至更广泛的散度族（如 α-散度）的收敛保证´；例如，参见 ΦCha [；另见04]例如 [VW23、CEL$^+$24、MW25、CCSW22]以及综述 [Che24]。这些结果对相对密度失配进行了细致控制，并在理解基于朗之万动力学的采样器的热启动要求和瞬态行为方面发挥了核心作用。

相比之下，现有关于基于哈密顿量的方法的理论保证较为有限。eHMC 的收敛结果主要是在 Wasserstein 距离下建立的 [CV19，MS21]。对于 uHMC，在 Wasserstein 距离和全变差距离下均有收敛保证 [BREZ20，BRE23，BRS23]。对于 Metropolis 校正的 HMC，已有研究分析了其在全变差距离 [CG23] 以及在 χ^2 散度 [CDWY20]下的收敛性。关于 NUTS 的结果目前仅限于高斯目标分布下的全变差距离 [BRO24]。

通过分析修正熵泛函的衰减， [CDMS23]得到了 uHMC 在 KL 散度下的渐近偏差界。在强对数凹目标分布的情形下，该

研究得到的渐近 KL 偏差阶为 $O(h^4d^4)$，其中 h 表示 Verlet 步长， d 表示维度。相比之下，在可比的假设下，我们的结果得到的偏差阶为 $O(h^4d^3)$（参见第 4.3.1节和表 1）。据我们所知，此前没有研究为 uHMC——甚至 eHMC——在 R´ényi 散度下的收敛提供保证。

鉴于基于哈密顿量的采样器在实践中应用广泛，这一理论缺口尤为引人注目，尤其是在高维贝叶斯推断中。哈密顿蒙特卡洛及其变体构成了现代概率编程系统的计算核心，例如 Stan [CGH+17] 和 PyMC [PHF10], 在这些系统中，众所周知，算法性能对初始化和尾部行为高度敏感。

本研究通过构建一个针对尾部特性的未校正哈密顿蒙特卡罗收敛分析系统框架，弥补了这一缺口。 我们利用单次耦合和核正则化，将现有的 Wasserstein 保证提升至 KL 散度和 R´ényi 散度。我们的结果对相对密度失配提供了定量控制，阐明了离散化偏差在强散度中的作用，并为未校正采样及 Metropolis 校正的 HMC 热启动提供了严格保证。

2 预备知识

令 $\mathcal{P}(\mathbb{R}^d)$ 表示 $\mathbb{R}^d$上的概率分布集合。当概率分布$\rho \in \mathcal{P}(\mathbb{R}^d)$相对于 $\mathbb{R}^d$上的勒贝格测度绝对连续时，我们将 ρ 与其密度（相对于勒贝格测度）等同，并且也用 ρ 表示该密度： $\mathbb{R}^d \to [0, \infty)$ 。本文中，我们假设所有密度均满足使相关陈述和结果定义良好所需的正则性条件。

我们用 $\delta_x \in \mathcal{P}(\mathbb{R}^d)$ 表示点 $x \in \mathbb{R}^d$ 处的狄拉克分布，并用 $\Gamma(\mu,\ \pi)$ 表示两个概率分布 μ, $\pi \in \mathcal{P}(\mathbb{R}^d)$ 之间的耦合集合。定义在 $\mathbb{R}^d$ 上的标准高斯分布 $\mathcal{N}(0,\ I)$ 记为 γ_d。随机变量 X 的分布记为 $\mathcal{L}(X)$。对于任意 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 和整数 $k \geq 1$，我们定义其 k 阶矩

$$m_k(\mu) \coloneqq \mathop{\mathbb{E}}_{X\sim\mu}\left[\|X\|^k\right].$$

2.1 概率距离与散度

我们首先回顾两个概率分布之间 Wasserstein–p 距离的定义。我们关于 KL 散度的结果（定理 1 和2）将 Wasserstein–2 保证提升为 KL 散度保证。

定义 1 (Wasserstein–p 距离)。 设 $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$ ，并设 $1 \leq p < \infty$。 μ 与 π 之间的 *Wasserstein–p*距离定义为

$$W_p(\mu,\pi) \coloneqq \left(\inf_{\gamma\in\Gamma(\mu,\pi)} \mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[\|x-y\|^p\right]\right)^{\frac{1}{p}}.$$

R´ényi 散度结果（定理 3 和 4）通过将 Orlicz-Wasserstein 保证提升为 R´ényi 散度中的界，进一步强化了这些保证。为了形式化这些概念，我们首先回顾随机变量 [Ver18，第 2.7 章中的 Orlicz 范数定义。1], 然后用它来定义概率测度之间的 Orlicz-Wasserstein 距离（定义3）。

定义 2（Orlicz 范数）。设 $\psi:\ \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ 为一个递增凸函数，满足$\psi(0)=0$ 且 $\psi(x) \to \infty$ 当 $x \to \infty$ 时。随机变量 $X \sim \mu$ 关于 *Orlicz* 函数 ψ 的 *Orlicz* 范数定义为

$$\|X\|_\psi := \inf\left\{\lambda > 0 : \mathop{\mathbb{E}}_{X\sim\mu}\left[\psi\left(\frac{\|X\|}{\lambda}\right)\right] \leq 1\right\}.$$

注 1. 不同的 *Orlicz* 函数 ψ 刻画了随机变量 X的不同尾部行为。例如，如果 $\psi(x) = e^x - 1$，则 $\|X\|_\psi$ 有限等价于 X 具有次指数尾部；而如果 $\psi(x) = e^{x^2} - 1$，则 $\|X\|_\psi$ 有限对应于次高斯尾部行为。关于这些等价关系的详细讨论，参见 *[Ver18第 2.5 章和第 2 章 7]* 。

注 2. 本文中，我们固定采用 *Orlicz* 函数

$$\psi(x) := e^{x^2} - 1,$$

使其对应的 *Orlicz* 范数能够控制次高斯尾部。

实值随机变量 $X \sim \mu$ 的矩母函数（m.g.f.）是函数$M_X\colon \mathbb{R} \to \mathbb{R}_{>0}$，定义为 $M_X(t) := \mathbb{E}_{X\sim\mu}[e^{tX}]$，只要期望有限即可。以下引理将在我们的分析中发挥作用：在 Orlicz 范数有限的假设下，该引理给出了 $\|X\|^2$ 的矩母函数的界。

Le引理 1. 假设 $\|X\|_\psi \leq K$，其中 $\psi(x) = e^{x^2} - 1$。那么对于 $c \in [0, K^{-2}]$，

$$\mathbb{E}\left[e^{c\|X\|^2}\right] \leq 2^{cK^2}.$$

证明。 根据 Orlicz 范数的定义以及假设 $\|X\|_\psi \leq K$，

$$\mathbb{E}\left[e^{\|X\|^2/K^2}\right] \leq 2.$$

对于 $c \in [0, K^{-2}]$，我们记

$$\mathbb{E}\left[e^{c\|X\|^2}\right] = \mathbb{E}\left[\left(e^{\|X\|^2/K^2}\right)^{cK^2}\right].$$

由于 $cK^2 \in [0,1]$，且函数 $t \mapsto t^{cK^2}$ 在 $(0,\ \infty)$ 上是凹函数，根据詹森不等式可得

$$\mathbb{E}\left[e^{c\|X\|^2}\right] \leq \left(\mathbb{E}\left[e^{\|X\|^2/K^2}\right]\right)^{cK^2} \leq 2^{cK^2}.$$

□

定义 3 (Orlicz-Wasserstein 距离)。 令 $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$。 μ 与 π 之间的 *Orlicz–Wasserstein* 距离定义为

$$W_\psi(\mu,\pi) := \inf_{\substack{\gamma\in\Gamma(\mu,\pi)\\(X,Y)\sim\gamma}} \|X-Y\|_\psi = \inf_{\gamma\in\Gamma(\mu,\pi)} \inf\left\{\lambda > 0 : \mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\|x-y\|^2}{\lambda^2}}\right] \leq 2\right\}.$$

若 $W_\psi(\mu,\ \pi) \le c$（其中某个常数为 $c > 0$），这意味着存在一个耦合 $\gamma^* \in \Gamma(\mu,\ \pi)$ 以及某个 $\lambda^* \le c$，使得

$$\mathop{\mathbb{E}}_{(x,y)\sim\gamma^*}\left[e^{\frac{\|x-y\|^2}{\lambda^{*2}}}\right] \le 2\,.$$

因此，正如 Wasserstein–2 距离通过最优耦合控制平方差的期望，Orlicz-Wasserstein 距离控制平方差的指数矩。

近年来，Orlicz-Wasserstein 距离已用于采样，尤其用于分析 R´enyi 散度 [AC24a、AC24b、AC24c], ；并且注意到，对于任意有限的 p，Orlicz-Wasserstein 距离都给出 Wasserstein– p [AC24a距离的上界，注 3。4], 也就是说，对于任意 $\mu,\ \pi \in \mathcal{P}(\mathbb{R}^d)$ 以及 $p \ge 1$

$$\frac{1}{\sqrt{2p}} W_p(\mu,\pi) \le W_\psi(\mu,\pi)\,.$$

现在我们定义概率分布之间的 KL 散度和 R´enyi 散度。

定义 4（KL 散度）。概率分布 $\mu,\pi \in \mathcal{P}(\mathbb{R}^d)$ 且满足 $\mu \ll \pi$ 时，其 *KL* 散度定义为

$$\mathsf{KL}(\mu \,\|\, \pi) := \mathop{\mathbb{E}}_{x\sim\mu}\left[\log\frac{\mu(x)}{\pi(x)}\right].$$

若 $\mu \not\ll \pi$，则 $\mathsf{KL}(\mu \,\|\, \pi) := \infty$。

定义 5（R´enyi–q 散度）。设 q 是一个整数 $1 < q < \infty$。概率分布 $\mu,\pi \in \mathcal{P}$之间，阶数为q 的 *R´enyi* 散度在满足 $\mu \ll \pi$ 时定义为（$\mathbb{R}^d$）

$$\mathsf{R}_q(\mu \,\|\, \pi) = \frac{1}{q-1}\log \mathop{\mathbb{E}}_{x\sim\pi}\left[\left(\frac{\mu(x)}{\pi(x)}\right)^q\right].$$

若 $\mu \not\ll \pi$，则 $\mathsf{R}_q(\mu \,\|\, \pi) := \infty$。

阶数为 $q = 1$ 的 R´enyi 散度按极限意义定义，与 KL 散度（定义 4）一致，即 $\mathsf{R}_1(\mu \,\|\, \pi) := \lim_{q\to 1}\mathsf{R}_q(\mu \,\|\, \pi) = \mathsf{KL}(\mu \,\|\, \pi)$。有关 KL 散度和 R´enyi 散度的更多性质和背景，可参见 [VEH14、PW24]。R´enyi 散度关于其阶数单调递增 [VEH14，定理 3]：对于任意分布 μ、$\pi \in \mathcal{P}(\mathbb{R}^d)$ 和 $1 \le q \le q' < \infty$,

$$\mathsf{R}_q(\mu \,\|\, \pi) \le \mathsf{R}_{q'}(\mu \,\|\, \pi)\,, \tag{2}$$

并且还满足以下弱三角不等式 [Mir17，命题 11][1], 即，对于任意分布 μ，π，$\rho \in \mathcal{P}(\mathbb{R}^d)$ 和 $q \ge 2$,

$$\mathsf{R}_q(\mu \,\|\, \pi) \le \frac{3}{2}\mathsf{R}_{2q}(\mu \,\|\, \rho) + \mathsf{R}_{2q-1}(\rho \,\|\, \pi)\,. \tag{3}$$

R´对于足够大的阶数，即使是绝对连续测度，[VW23，例 2]也表明

$$\mathsf{R}_q(\mathcal{N}(0,\sigma^2 I) \,\|\, \mathcal{N}(0,\lambda^2 I)) = \infty$$

如果 $\sigma^2 > \lambda^2$ 且 $q \ge \frac{\sigma^2}{\sigma^2-\lambda^2}$，则反映出高斯尾部衰减率不匹配，尽管测度绝对连续，仍会导致散度。

[1]我们介绍 并非一个简化的上界；更精确的表述见 [Mir17，命题 1 1].

2.2 哈密顿蒙特卡洛

我们简要回顾哈密顿蒙特卡罗（HMC）框架并确定记号。如下文所述，HMC 的目标概率分布 ν 定义在 $\mathbb{R}^d$ 上，其密度正比于 e^{-f}，其中$f\colon\ \mathbb{R}^d \to \mathbb{R}$ 是一个连续可微的势函数。该方法通过引入辅助速度变量扩充状态空间，并在由此得到的相空间中运用哈密顿动力学，从而定义一个作用于 $\mathbb{R}^d$上的马尔可夫转移核；这一方法最初由 [DKPR87]提出，并在 [Nea11]中以教学方式进行了阐述。

我们首先介绍位置–速度（相）空间 $\mathbb{R}^{2d}$。在全文中，我们只考虑最简单的单位质量矩阵情形。这一限制并不失一般性：任何常数正定质量矩阵都可以通过对速度（或动量）进行线性变量变换吸收，也称为转换到质量加权坐标，从而得到一个质量为单位的等价哈密顿系统。由于我们的论证只依赖于哈密顿流的结构性质，而不依赖于相空间的具体参数化，为了清晰起见，我们采用单位质量的设定。

据此，我们将哈密顿量 H 定义为：$\mathbb{R}^{2d} \to \mathbb{R}$

$$H(x,v) \coloneqq f(x) + \tfrac{1}{2}\|v\|^2.$$

相应的哈密顿动力学（x_t，v_t）$\in \mathbb{R}^{2d}$ 由以下常微分方程组给出

$$\frac{dx_t}{dt} = \nabla_v H(x_t, v_t) = v_t, \qquad \frac{dv_t}{dt} = -\nabla_x H(x_t, v_t) = -\nabla f(x_t). \tag{4}$$

对于初始条件（x_0，v_0）$=$（x，v），我们用（x_t（x，v）v_t（x，v）表示时刻$t \geq 0$时（4）的解。相应的哈密顿流映射为

$$\mathsf{HF}_t\colon \mathbb{R}^{2d} \to \mathbb{R}^{2d}, \qquad \mathsf{HF}_t(x,v) \coloneqq (x_t(x,v), v_t(x,v)).$$

我们进一步定义位置投影

$$q_t \coloneqq \Pi_1 \circ \mathsf{HF}_t,$$

使得 $q_t(x,\ v) = x_t(x,\ v)$ 对所有 $(x,\ v) \in \mathbb{R}^{2d}$ 均成立。

精确哈密顿蒙特卡罗（eHMC）算法 [BREZ20，BRS23]构造了一个马尔可夫链（X_k）$_{k\geq 0}$，其状态空间为 $\mathbb{R}^d$，具体如下。给定初始分布 $X_0 \sim \rho$，更新规则为

$$X_{k+1} = q_T(X_k, \xi_k), \tag{5}$$

其中 $(\xi_k)_{k\geq 0}$ 是从标准高斯分布 γ_d 中独立抽取的样本，$T > 0$ 是固定的积分时间。我们将相应的转移核记为

$$\mathbf{P}_T\colon \mathcal{P}(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d),$$

因此，如果 $X_k \sim \rho_k$，那么 $X_{k+1} \sim \rho_{k+1} = \rho_k \mathbf{P}_T$，进而 $\rho_k = \rho \mathbf{P}_T^k$。尽管这一点无法直接从上述构造得出，但目标分布 ν 是 eHMC 的不变分布，即 $\nu \mathbf{P}_T = \nu$。这是因为哈密顿流保持哈密顿量不变，因而也保持密度 e^{-H} 不变；此外，速度刷新步骤保持辅助动量分布不变。此外，eHMC 关于 ν可逆；这一额外性质依赖于.....的时间可逆性

算法 1： 速度 Verlet 积分器 ($\mathsf{HF}_{t_1,\ t_2}$)

输入： 积分时间 $t_1 > 0$，步长 $t_2 > 0$，其中 $t_2 \mid t_1$，初始条件 $(x,\ v) \in \mathbb{R}^{2d}$。设定 $x_0 := x,\ v_0 := v$，以及 $N := t_1/t_2$。**对于** $j = 0$ 到 $N-1$ **执行** $x_{j+1} = x_j + t_2 v_j - \frac{t_2^2}{2}\nabla f(x_j)$ $v_{j+1} = v_j - \frac{t_2}{2}\left(\nabla f(x_j) + \nabla f(x_{j+1})\right)$ **结束循环** **输出：** $(x_N,\ v_N)$

哈密顿流，以及速度刷新步骤会从辅助动量分布中独立地重新采样速度这一事实。简要回顾请参见附录A。

eHMC 更新（5）要求精确求解哈密顿动力学（*4*），这通常难以实现。用数值积分器替代精确流会引入离散化误差；若不进行 Metropolis校正，由此得到的方法便属于未经调整的哈密顿蒙特卡罗 (uHMC) [BRS23、BRE23]。uHMC 一词涵盖了任何基于离散化的哈密顿流近似。本文重点讨论速度Verlet（蛙跳）积分器 [BRSS18、HHIL06], ，由此得到本文研究的 uHMC-v 算法。关于其他离散化方案（例如分层蒙特卡罗积分器 [BRM25],）的扩展将在第4.3.2节讨论。

为定义 uHMC-v，我们引入一种基于速度 Verlet 积分器的哈密顿流离散时间近似。令 $\tilde{\mathsf{HF}}_{t_1,t_2}\colon \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ 表示与速度 Verlet 方案对应的数值流映射，其总积分时间为 $t_1 > 0$，步长为 $t_2 > 0$，其中 t_2整除 t_1。对于任意初始条件 $(x, v) \in \mathbb{R}^{2d}$，值 $\tilde{\mathsf{HF}}_{t_1,t_2}(x, v)$ 由算法 1 定义。

在关于 f 的标准正则性假设下，速度 Verlet 积分器是哈密顿动力学的二阶精确辛近似，并且 $\tilde{\mathsf{HF}}_{t_1,t_2}(x, v)$ 在固定 $\tilde{\mathsf{HF}}_{t_1}(x, v)$ 的情况下，当 $t_2 \to 0$ 时收敛于精确流；例如参见 t_1HHIL06、BRSS18]。

我们定义相应的位置映射

$$\tilde{q}_{t_1,t_2} := \Pi_1 \circ \tilde{\mathsf{HF}}_{t_1,t_2},$$

使得 $\tilde{q}_{t_1,t_2}(x, v)$ 等于经过 t_1/t_2 次积分步骤后数值轨迹的位置分量。

从初始分布 $X_0 \sim \rho$ 出发，uHMC-v 算法通过以下更新生成马尔可夫链 $(X_k)_{k\geq 0}$ on $\mathbb{R}^d$

$$X_{k+1} = \tilde{q}_{T,h}(X_k, \xi_k), \tag{6}$$

其中 $(\xi_k)_{k\geq 0}$ 是从标准高斯分布 γ_d 中独立抽取的样本，$T > 0$ 是积分时间，$h > 0$ 是步长。此更新定义了一个马尔可夫转移核

$$\tilde{\mathbf{P}}_{T,h}\colon \mathcal{P}(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d),$$

因此，X_k 的分布律为 $\rho\tilde{\mathbf{P}}_{T,h}^k$。

在对 f、T和 h满足适当的正则性与稳定性假设时，核 $\mathbf{P}_{T,h}$ 存在一个不变分布，记作 $\tilde{\nu}_h$。一般而言，$\tilde{\nu}_h$ 取决于 h，且不同于目标分布 ν。此外，在附加假设下，$\tilde{\nu}_h$ 当 ν 时收敛至 $h \to 0$；例如，参见 [BRS23、BRE23]。

在本文中，我们始终假设，只要速度 Verlet 步长 $h > 0$，它就整除积分时间 $T > 0$，即 $T/h \in \mathbb{N}$。我们还指出，在形式极限 $h = 0$,˜下，数值流 $\mathrm{HF}_{T,\ h}$ 与精确哈密顿流 HF_T一致，而 uHMC-v 则退化为 eHMC。

注 3 （可逆性）。 eHMC 核$\mathbf{P}_T$关于目标分布ν可逆。相较之下，当不变分布 $\tilde{\nu}_h$ 存在时，uHMC 核$\mathbf{P}_{T,\ h}$ 通常关于该分布不可逆。缺乏可逆性是数值积分器离散化误差所致；即使采用速度 *Verlet* 等辛格式，且不进行 *Metropolis*校正，这一现象仍然存在。

3 单次耦合

单次 耦合 是两个马尔可夫链 [RR02、MS10、BRE23]的一步转移核之间的一种耦合。这类耦合通常旨在最大化两条链在单次转移后合并的概率。传统上，这类耦合用于同一马尔可夫链的两个副本，并用来控制其混合行为。

在本研究中，我们以两种概念上不同的方式运用单次耦合。首先，我们在两个 uHMC-v 副本之间构造单次耦合，以建立正则化界；这些界刻画单次转移如何平滑或收缩不同初始条件之间的差异，并直接导出混合时间保证。其次，我们在 uHMC-v 与 eHMC 之间构造单次耦合，以获得交叉正则化界；这些界比较离散化马尔可夫链与其连续时间对应过程的演化，并帮助我们控制离散化引入的渐近偏差。

具体而言，KL 正则化界和交叉正则化界分别在引理12和14 中得到证明；相应的 R´enyi 正则化界和交叉正则化界则在引理17 和 19中得到证明。所有这些结果均通过适当构造的单次耦合加以证明。

文献 [BRE23] 研究了 uHMC-v 的单次耦合，以将 Wasserstein–1 收敛保证提升为全变差距离保证。在该文中，速度 Verlet 方案被刻画为一个离散时间变分问题的解。在适当条件下，该变分问题被证明是强凸的，这意味着速度 Verlet 更新是变分目标函数的唯一极小化者 [BRE23，推论 14]。这一唯一性性质对分析 uHMC-v 的单次耦合具有重要意义：它意味着存在唯一的单次耦合映射，能够耦合马尔可夫链两个副本的速度，使这两个链在一步后合并；参见 [BRE23，图 1]。

我们在此采用一种密切相关的策略，以获得 uHMC-v 在 Kullback–Leibler 和 R´enyi 散度下的保证。用于推导混合时间界的单次耦合映射（图2(a)）与 [BRE23]中引入的映射一致。为控制渐近偏差，我们考虑单次耦合构造的一种变体（图2(b)），它将离散化核 uHMC-v 与连续时间核 eHMC 耦合。虽然 [BRE23]关注从相同初始状态出发的耦合，但我们的分析将这一视角扩展到从不同点初始化的离散化动力学与精确动力学之间的耦合。

类似的耦合论证见 Marsden 的博士论文第4章[Mar25], ，其中使用单次耦合映射来为采用随机化积分器实现的校正后的哈密顿蒙特卡罗的受限导通率剖面建立下界。本分析以受限

导通率剖面框架，出自 [CDWY20], ，该框架通过将单步重叠与导通率下界联系起来，为校正后的哈密顿蒙特卡罗建立混合时间界。

备注 4 （核正则化与Harnack不等式之间的关系）。 关于马尔可夫核在*KL*散度和*R´enyi*散度中的正则化性质、*Wang*型*Harnack*不等式以及反向泛函不等式之间，已有成熟的等价性结果 *[Wan97、 BGL01、ATW06、 BGL15、 BE21、AC24b]*；另请参见附录 *B*。这些等价关系从函数不等式的角度刻画正则化界，而我们通过单次耦合建立了这些界。

现在我们介绍势函数 f 的正则性假设（回顾 $\nu \propto e^{-f}$），在这些假设下可以得到单步耦合映射的估计。

假设 1。 势函数 $f\colon\ \mathbb{R}^d \to \mathbb{R}$ 满足以下条件

(a) 函数 f 在 0 处取得全局最小值，且 $f(0) = 0$。*(b)* 函数 $f \in \mathcal{C}^2(\mathbb{R}^d)$ 满足 $\|\nabla^2 f(x)\|_{\mathsf{op}} \le L$，对所有 $x \in \mathbb{R}^d$ 均成立。*(c)* 函数 $f \in \mathcal{C}^3(\mathbb{R}^d)$ 满足 $\|\nabla^3 f(x)\|_{\mathsf{op}} \le M$，对所有 $x \in \mathbb{R}^d$ 均成立。*(d)* 函数 $f \in \mathcal{C}^4(\mathbb{R}^d)$ 满足 $\|\nabla^4 f(x)\|_{\mathsf{op}} \le N$，对所有 $x \in \mathbb{R}^d$ 均成立。

这里，对于 $k \ge 2$, $\nabla^k f(x)$ 被视为定义在 $(\mathbb{R}^d)^k$ 上的对称 k 线性形式，而 $\|\cdot\|_{\mathsf{op}}$ 表示由 $\mathbb{R}^d$ 上的欧几里得范数诱导的相应算子范数，即

$$\|T\|_{\mathsf{op}} \coloneqq \sup\big\{|T(u_1,\dots,u_k)| : \|u_1\| = \cdots = \|u_k\| = 1\big\}.$$

为分析混合时间（第 3.1 节）而引入的单次耦合映射，使我们能够通过将问题转化为界定标准高斯分布与扰动高斯分布之间的散度，来控制单次转移后的散度。该扰动由单次耦合映射诱导，具体如下引理所述。引理2的证明见第 C 节。

˜引理 *2.* 回顾一下，$P_{T,h}$ 表示更新由 *(6)* 给出的 *uHMC-v* 转移核。设$x, y \in \mathbb{R}^d$为任意两点，$D \in \{KL, R_q\}$ 为 *KL* 散度（定义 *4*）或 *Rényi* 散度（定义 *5*），且 $\varphi_{x,y} : \mathbb{R}^d \to \mathbb{R}^d$ 为满足 *(7)* 的映射。则对于满足假设 *1(b)* 的势函数，任意满足 $LT^2 \le \frac{2}{5}\pi^2$ 的积分时间 $T > 0$，以及步长 $h \ge 0$，

$$\mathsf{D}(\delta_x \tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y \tilde{\mathbf{P}}_{T,h}) \le \mathsf{D}((\varphi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d)\,.$$

注 5。 我们假设步长 $h > 0$ 整除积分时间 T，即 $T = mh$ 其中某个整数为 $m \ge 1$。结合约束条件 $LT^2 \le \frac{2}{5}\pi^2$，可得

$$Lh^2 = L\Big(\frac{T}{m}\Big)^2 \le LT^2 \le \frac{2}{5}\pi^2 < 4,$$

因此 $h\sqrt{L} < 2$，这就是速度 *Verlet* 积分器的稳定性条件 *[BRSS18]*。

为了得到渐近偏差的界，我们证明 uHMC-v 的交叉正则化界。为分析渐近偏差（第 3.2 节）而引入的单次耦合映射，使我们能够用标准高斯分布与扰动高斯分布之间的散度来界定一步之后的散度。与混合时间分析一样，扰动由单次耦合映射本身诱导，具体形式见下述引理。引理 3 的证明见C 节。

˜引理 *3.* 回顾一下，$P_{T,h}$ 和 P_T 分别表示 *uHMC-v* 和 *eHMC* 的转移核，其更新分别由式 *(6)* 和式 *(5)* 给出。设 $x, y \in \mathbb{R}^d$ 为任意两点，$D \in \{KL, R_q\}$ 表示 *KL* 散度（定义 *4*）或 *R´enyi* 散度（定义 *5*），且 $\Phi_{x,y}: \mathbb{R}^d \to \mathbb{R}^d$ 是满足式 *(10)* 的映射。则对于满足假设 *1(b)* 的势函数，任意满足$LT^2 \leq \frac{2}{5}\pi^2$的积分时间 $T > 0$，以及步长 $h \geq 0$，

$$\mathsf{D}(\delta_x \tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y \mathbf{P}_T) \leq \mathsf{D}((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d)\,.$$

引理2 和3 表明，从狄拉克初始化出发的正则化保证和交叉正则化保证，都可以归结为界定标准高斯分布与扰动高斯分布之间的散度。在这两种情况下，扰动均由相应的单次耦合映射所诱导。本节其余部分将介绍这些耦合映射，并建立它们的正则性界，进而利用这些界控制扰动高斯分布与标准高斯分布之间的散度。

3.1 正则化的单次耦合映射

对于任意 x，$y \in \mathbb{R}^d$，令 $\varphi_{x,\,y}\colon \mathbb{R}^d \to \mathbb{R}^d$ 表示满足以下条件的映射

$$\tilde{q}_{T,h}(x,v) = \tilde{q}_{T,h}\big(y, \varphi_{x,y}(v)\big), \qquad \forall\, v \in \mathbb{R}^d. \tag{7}$$

也就是说，$\varphi_{x,\,y}$ 将位置 x 处的初始速度 v 映射为位置 y 处的速度，使得两条离散化轨迹经过时间 T后的位置重合。

图2(a) 展示了这一单次耦合映射，该映射由 [BRE23] 提出，用于获得 uHMC-v 在全变差距离下的混合时间保证。对于任意 x，$y \in \mathbb{R}^d$，只要 $\varphi_{x,\,y}$满足（7），且 f是 L–光滑的，并且 $LT^2 \leq \frac{2}{5}\pi^2$ [BRE23，推论 14]。

我们有如下关于映射及其雅可比矩阵的二阶正则性估计。引理 4和 5 的证明见第D节。

引理 4. 假设条件 *1(a)-(b)* 成立。进一步假设 $h \geq 0$且 $L(T^2 + Th) \leq \frac{1}{12}$。那么对于任意 x，y，$v \in \mathbb{R}^d$

$$\|\varphi_{x,y}(v) - v\| \leq \frac{3}{2T}\|x - y\|. \tag{8}$$

引理 5. 假设条件 *1(a)-(c)* 成立。进一步假设 $h \geq 0$且 $L(T^2 + Th) \leq \frac{1}{12}$。那么对于任意 x，y，$v \in \mathbb{R}^d$

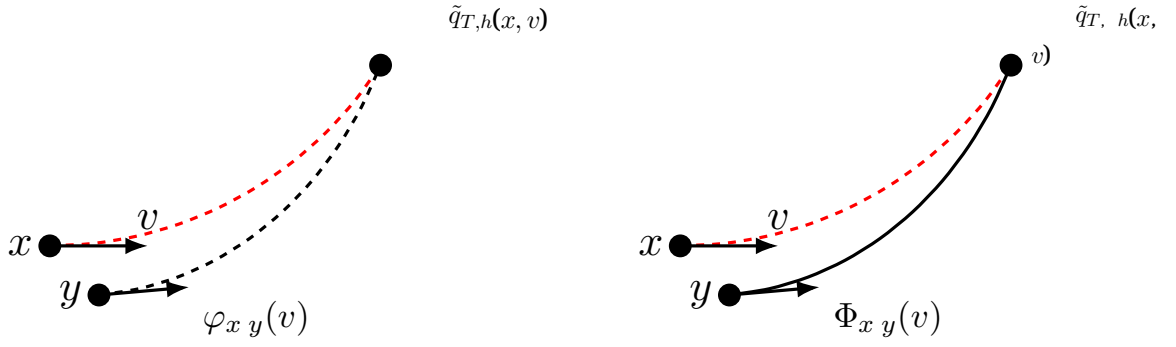
$$\|\nabla\varphi_{x,y}(v) - I\|_{\mathsf{op}} \leq \min\left\{\frac{2}{9}, \frac{11}{2}MT^2\|x - y\|\right\}. \tag{9}$$

注 6. 引理 *4*和 *5*是对 *[BRE23]* 中引理 *25* 和 *26* 的重述，采用了我们的记号和假设。为完整起见，我们在此列出。

3.2 交叉正则化的一次映射

对于任意 $x, y \in \mathbb{R}^d$，令 $\Phi_{x,y}\colon \mathbb{R}^d \to \mathbb{R}^d$ 表示满足以下条件的映射

$$\tilde{q}_{T,h}(x,v) = q_T\big(y, \Phi_{x,y}(v)\big), \qquad \forall\, v \in \mathbb{R}^d. \tag{10}$$



(a) $\tilde{q}_{T,h}(x,v) = \tilde{q}_{T,h}(y,\varphi_{x\ y}(v))$ (b) $\tilde{q}_{T,h}(x,v) = q_T(y,\Phi_{x\ y}(v))$

图 2：单次耦合。(a) 为了获得 uHMC-v 的正则化界，初始速度按如下方式耦合：$\tilde{q}_{T,\,h}(x,\ v) = \tilde{q}_{T,\,h}(y,\ \varphi_{x,\,y}(v))$。(b) 为了获得 uHMC-v 的交叉正则化界，初始速度按如下方式耦合：$\tilde{q}_{T,\,h}(x,\ v) = q_T(y,\ \Phi_{x,\,y}(v))$。

也就是说，一次映射 $\Phi_{x,y}$ 将离散动力学中位置 x处的初始速度 v 映射为精确哈密顿流中位置 y 处的速度，使得两条轨迹在经过时间 T后位置重合。

该映射见图 2(b)。对于任意 x，$y \in \mathbb{R}^d$，在假设 f是 L–光滑且 $LT^2 \le \frac{2}{5}\pi^2$的条件下，满足（10）的映射 $\Phi_{x,\,y}$ 的存在性已在第 21 节的引理 E中确立。如第E节所述，可以将该映射视为由 [BRE23]中研究的 uHMC-v 单次耦合映射复合而成，并据此进行分析。

我们有以下关于映射及其雅可比矩阵的二阶正则性估计，这些估计有助于获得 KL 散度中的渐近偏差保证（定理 2）。引理 6 和 7在第 E 节中证明。

引理 6。 假设条件 *1(a)-(c)* 成立。进一步假设 $h \ge 0$且 $L(T^2+Th) \le \frac{1}{12}$。那么，对于任意 x，y，$v \in \mathbb{R}^d$

$$\|\Phi_{x,y}(v) - v\| \le \frac{3}{2T}\|x-y\| + 2h^2(LT^{-1}\|x\| + L\|v\| + MT^{-1}\|x\|^2 + MT\|v\|^2)\,.$$

引理 7。 假设条件 *1(a)-(d)* 成立。进一步假设 $h \ge 0$且 $L(T^2+Th) \le \frac{1}{12}$。那么，对于任意 x，y，$v \in \mathbb{R}^d$

$$\|\nabla\Phi_{x,y}(v) - I\|_{\mathsf{op}} \le \min\left\{\frac{15}{18}, \frac{11}{2}MT^2\|x-y\| + \frac{22}{9}h^2Q(x,v)\right\}$$

其中

$$Q(x,v) := L + M\|x\| + MT\|v\| + (M^2T^2+N)\|x\|^2 + (M^2T^2+N)T^2\|v\|^2\,.$$

引理 6 和 7 通过分析二阶梯形近似得到，因此呈现出 h^2–对步长的依赖，以及对位置 x和速度 v的二次依赖。对 v 的二次依赖给 R´enyi 散度界带来了困难，因为要控制扰动高斯分布与标准高斯分布之间的 R´enyi 散度，需要对形如 $\mathbb{E}_{v\sim\gamma_d}[e^{\lambda\|v\|^4}]$的矩给出界，而对于所有 $\lambda > 0$，这些矩都会发散。

为了获得 R´enyi 散度中的有限交叉正则化界（引理 18），我们因此基于一阶近似，为 $\Phi_{x,y}$ 推导较弱的逐点估计和雅可比估计，这些估计

并未利用速度 Verlet 积分器的二阶精度。所得界见引理8 和9，证明见附录 E.1。

引理8。 假设条件 *1(a)-(b)* 成立。进一步假设 $h \geq 0$ 并且 $L(T^2 + Th) \leq \frac{1}{12}$。那么，对于任意 $x, y,\ v \in \mathbb{R}^d$

$$\|\Phi_{x,y}(v) - v\| \leq \frac{3}{2T}\|x - y\| + \frac{7}{5}h\left(\frac{1}{5T}\|v\| + \frac{7}{36}L\|x\|\right).$$

引理9。 假设条件 *1(a)-(c)* 成立。进一步假设 $h \geq 0$ 并且 $L(T^2 + Th) \leq \frac{1}{12}$。那么，对于任意 $x, y,\ v \in \mathbb{R}^d$

$$\|\nabla\Phi_{x,y}(v) - I\|_{\mathsf{op}} \leq \min\left\{\frac{15}{18}, \frac{11}{2}MT^2\|x - y\| + \frac{22}{135}h\left(\frac{2}{T} + \frac{16}{5}MT\|x\| + 20MT^2\|v\|\right)\right\}.$$

为简化本文余下部分的记号，我们引入以下常数，它们汇总了引理8和 9中的逐点估计和雅可比估计所涉及的系数。

定义6。 定义以下常数，它们对应于引理 *8* 和 *9*中的逐点估计和雅可比估计的系数：

$$\mathsf{p}_{xy} \coloneqq \frac{3}{2T}\ , \qquad \mathsf{p}_v \coloneqq \frac{7}{25T}\ , \qquad \mathsf{p}_x \coloneqq \frac{49}{180}L\ ,$$
$$\mathsf{j}_{xy} \coloneqq \frac{11}{2}MT^2\ , \qquad \mathsf{j}_c \coloneqq \frac{44}{135T}\ , \qquad \mathsf{j}_v \coloneqq \frac{440}{135}MT^2\ , \qquad \mathsf{j}_x \coloneqq \frac{352}{675}MT\ .$$

采用这一记号，引理8 可写为

$$\|\Phi_{x,y}(v) - v\| \leq \mathsf{p}_{xy}\|x - y\| + h\,\mathsf{p}_v\|v\| + h\,\mathsf{p}_x\|x\|,$$

而引理 9可写为

$$\|\nabla\Phi_{x,y}(v) - I\|_{\mathsf{op}} \leq \min\left\{\frac{15}{18},\ \mathsf{j}_{xy}\|x - y\| + h\,\mathsf{j}_c + h\,\mathsf{j}_v\|v\| + h\,\mathsf{j}_x\|x\|\right\}.$$

4 主要结果与示例

本节给出我们的主要结果，确立 uHMC 在 KL 散度和 R´ényi 散度下的混合时间和渐近偏差界。随后，我们在第4.3节将这些结果应用于强对数凹目标分布，并讨论相关示例；此外，我们还在第4.4节介绍一项关于信息收缩的应用。

4.1 KL 散度保证

我们首先陈述 KL 散度下的混合时间保证和渐近偏差保证。回顾一下，我们的证明策略是将 Wasserstein–2 保证提升为 KL 散度保证；因此，要获得 uHMC-v 在 KL 散度下的˜混合时间保证，我们要求 uHMC-v 核 $\mathbf{P}_{T,h}$ 满足：

˜**假设 2.** 设 $\mathbf{Q}$ 为一个离散时间马尔可夫转移核，其平稳分布为 $\tilde{\nu}_{\tilde{\mathbf{Q}}}$，且存在常数 $c_1 \geq 1$、$c_2 > 0$，使得对于任意整数 $k \geq 1$ 和初始概率分布 μ

$$W_2(\mu\tilde{\mathbf{Q}}^k, \tilde{\nu}_{\tilde{\mathbf{Q}}}) \leq c_1 e^{-c_2 k}\, W_2(\mu, \tilde{\nu}_{\tilde{\mathbf{Q}}}).$$

我们关于 uHMC-v 的 KL 散度混合时间的主要结果如下定理所述。我们在第 1节 5.1中详细说明证明定理 1所需的主要步骤，并在第 F.4节证明定理。

定理 1。 令 $\mu \in \mathcal{P}(\mathbb{R}^d)$。假设假设条件 *1(a)-(c)* 成立。假设对于 $\tilde{\mathbf{Q}} = \tilde{\mathbf{P}}_{T,h}$，假设条件 *2*成立，其不变分布为 $\tilde{\nu}_h$。假设 $h \geq 0$ 且 $L(T^2 + Th) \leq \frac{1}{12}$。那么，对于任意整数 $k \geq 0$ —

$$\mathsf{KL}(\mu \tilde{\mathbf{P}}_{T,h}^{k+1} \,\|\, \tilde{\nu}_h) \leq c_1^2 e^{-2c_2 k} \left(\frac{9}{4T^2} + 20dM^2T^4 \right) W_2^2(\mu, \tilde{\nu}_h) \,.$$

回顾一下， $\tilde{\nu}_h$ 是 uHMC-v 的不变分布，而定理1表明，沿 uHMC-v 的分布以指数速度收敛至 $\tilde{\nu}_h$ 。在命题 2 中，我们验证了当 ν 为强对数凹时，uHMC-v 满足假设条件 1 ，该命题位于第 4.3.1 节。定理 1 与命题 1共同推出推论 1，说明了 uHMC-v 达到所需精度、使其与平稳分布之间的 KL 散度处于目标范围内所需的迭代次数。

˜KL 散度下的渐近偏差界（定理 2）适用于一般的核 $\mathbf{Q}$ ，只要˜马尔可夫链的最后一步使用 uHMC-v 核 $\mathbf{P}_{T,\,h}$；这些保证要求核˜$\mathbf{Q}$ 满足假设条件 2 和3。

˜**假设 3.** 设$\mathbf{Q}$ 为一个离散时间马尔可夫转移核，其平稳分布为 $\tilde{\nu}_{\tilde{\mathbf{Q}}}$，且满足 $W_2(\tilde{\nu}_{\tilde{\mathbf{Q}}}, \nu) \leq \Delta_h$，其中 $\Delta_h \geq 0$，并且当 $h \to 0$ 时 $\Delta_h \to 0$。

描述 uHMC 的 KL 散度渐近偏差的主要结果如下。我们将在第 2 节 5.2 概述定理 2 的证明，并在第 G.3节证明该定理。

定理 2. 设 $\mu \in \mathcal{P}(\mathbb{R}^d)$，并假设假设 *1(a)-(d)* 成立。进一步假设 $h \geq 0$，$L(T^2 + Th) \leq \frac{1}{12}$，且对于马尔可夫核—˜$\mathbf{Q}$，假设 *2*和 *3*成立，其不变分布为 $\tilde{\nu}_{\tilde{\mathbf{Q}}}$。那么，对于任意整数 $k \geq 0$

$$\mathsf{KL}(\mu \tilde{\mathbf{Q}}^k \tilde{\mathbf{P}}_{T,h} \,\|\, \nu) \leq 2c_1^2 e^{-2c_2 k} \Big(\frac{9}{4T^2} + \frac{363}{2} dM^2T^4 \Big) W_2^2(\mu, \tilde{\nu}_{\tilde{\mathbf{Q}}}) + \mathsf{bias}$$

其中

$$\mathsf{bias} \leq 2\Delta_h^2 \Big(\frac{9}{4T^2} + \frac{363}{2} dM^2T^4 \Big) + 4h^4 \Bigg[45dL^2 + \Big(4\frac{L^2}{T^2} + 45dM^2 \Big) m_2(\mu \tilde{\mathbf{Q}}^k) + (4L^2 + 45dM^2T^2)d$$
$$+ \Big(4\frac{M^2}{T^2} + 45d(M^2T^2 + N)^2 \Big) m_4(\mu \tilde{\mathbf{Q}}^k) + \big(4M^2T^2 + 45d(M^2T^4 + NT^2)^2 \big) d(d+2) \Bigg] \,.$$

我们在第2 节和3 节验证 uHMC-v 的假设，并在第4.3.1节验证采用随机积分方案实现的 uHMC [BRM25] 在第4.3.2节中的假设。因此，我们得到了各条链生成一个样本所需梯度评估总次数的显式界 X ，该样本的分布满足$\mathsf{KL}(\mathcal{L}(X) \,\|\, \nu) \leq \epsilon$；见推论2 和5。我们发现，Verlet 积分格式对目标精度的依赖更有利 ϵ，而分层方案 [BRM25] 对维度的依赖更优 d ，尤其是在 $\epsilon < \frac{1}{d}$ 的情形下。 –

4.2 Rényi 散度保证

我们关于 R´ényi 散度的结果，是通过将 Orlicz-Wasserstein 保证提升为 R´ényi˜散度得到的。uHMC-v 在 R´ényi 散度下的混合时间（定理3) 要求$\mathbf{P}_{T,\,h}$ 满足：

˜**假设 4.** 设$\mathbf{Q}$ 是一个离散时间马尔可夫转移核，其平稳分布为 $\tilde{\nu}_{\tilde{\mathbf{Q}}}$，且存在常数 $c_1 \geq 1$、$c_2 > 0$，使得对于任意整数 $k \geq 1$ 和初始概率分布 μ

$$W_\psi(\mu\tilde{\mathbf{Q}}^k, \tilde{\nu}_{\tilde{\mathbf{Q}}}) \leq c_1 e^{-c_2 k}\, W_\psi(\mu, \tilde{\nu}_{\tilde{\mathbf{Q}}})\,.$$

下述定理描述了 uHMC-v 在 R´ényi 散度下的混合时间。我们在第 3 节概述定理 3 的证明，并在第 H.4 节证明定理 3。

定理 3. 设 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 成立，并假设假设 *1(a)-(c)* 成立。进一步假设 $h \geq 0$、$L(T^2 + Th) \leq \frac{1}{12}$，且假设 *4* 对 ˜$\tilde{\mathbf{Q}} = \tilde{\mathbf{P}}_{T,h}$ 成立，其不变分布为 $\tilde{\nu}_h$。对于任意 $1 < q < \infty$，定义

$$\delta_1 := (q-1)\left(8dMT^2 + \frac{3\sqrt{d}}{2T}\right),\ \delta_2 := \frac{9q(q-1)}{8T^2}\,.$$

那么对于任意整数

$$k \geq \frac{1}{c_2}\log\left(\frac{c_1 W_\psi(\mu, \tilde{\nu}_h)}{4\sqrt{\log 2}}\left(\delta_1 + \sqrt{\delta_1^2 + 16(\log 2)\delta_2}\right)\right)$$

可得

$$\mathsf{R}_q(\mu\tilde{\mathbf{P}}_{T,h}^{k+1} \,\|\, \tilde{\nu}_h) \leq \frac{\delta_1 + \delta_2}{q-1}\sqrt{\log 2}\, c_1^2 e^{-c_2 k}\, \max\{1, W_\psi^2(\mu, \tilde{\nu}_h)\}\,.$$

回顾一下，$\tilde{\nu}_h$ 是 uHMC-v 的平稳分布，而定理3 表明，uHMC-v 的分布会以指数速度收敛到$\tilde{\nu}_h$，前提是迭代次数超过某个阈值。该阈值较为宽松，仅对参数呈对数依赖。我们在命题3中验证，对于强对数凹目标分布，uHMC-v 满足假设4。在此情形下，推论3 给出了 uHMC-v 所需的步数，使其输出的样本在 R´ényi 散度意义下达到相对于平稳分布的期望精度。

˜R´ényi 散度中的渐近偏差保证适用于一般核$\mathbf{Q}$，只要马尔可夫链的最后˜一步采用 uHMC-v 转移核$\mathbf{P}_{T,\,h}$；这些保证要求核$\mathbf{Q}$ 满足假设4和5。

˜**假设5**。设$\mathbf{Q}$ 为一个离散时间马尔可夫转移核，其平稳分布为 $\tilde{\nu}_{\tilde{\mathbf{Q}}}$，并且 $W_\psi(\tilde{\nu}_{\tilde{\mathbf{Q}}}, \nu) \leq \Delta_h$，其中 $\Delta_h \geq 0$，且当 $h \to 0$时 $\Delta_h \to 0$。

描述 uHMC 在 R´ényi 散度下渐近偏差的主要结果如下定理所示。我们在第 4 节概述定理 6.2 的证明，并在第 4 节证明定理 I.3。

定理 4。设 $\mu \in \mathcal{P}(\mathbb{R}^d)$，假设 $X \sim \nu$ 满足 $\|X\|_\psi \leq K_\nu$，且假设 *1(a)-(c)* 成立。另请回顾定义 *6*中定义的常数。进一步假设 *4* 和 *5* 成立˜，对于马尔可夫核 $\mathbf{Q}$，其平稳分布为 $\tilde{\nu}_{\tilde{\mathbf{Q}}}$。设 $1 < q < \infty$ 并定义

$$s := 2q - 1, \quad u^2 := 1 + \frac{1}{4s}, \quad \delta_1 := s\left(8dMT^2 + \frac{3\sqrt{d}}{2T}\right), \quad \delta_2 := \frac{9qs}{4T^2}\,.$$

假设

$$\Delta_h \le \min\left\{ \frac{\sqrt{2\log 2}}{4s(\frac{\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_{xy}+\mathsf{p}_{xy}u)}{\sqrt{2}} + h\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x+\mathsf{p}_x u))}, \frac{1}{\sqrt{8s(\mathsf{p}_{xy}^2(1+3su^2)+2h^2\mathsf{p}_x^2(1+3su^2))}} \right\},$$

$$h \le \min\left\{ \frac{u-1}{\mathsf{p}_v}, \frac{\sqrt{\log 2}}{2\sqrt{2}s\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x+\mathsf{p}_x u)K_\nu}, \frac{1}{4K_\nu \mathsf{p}_x\sqrt{(1+3su^2)s}} \right\},$$

并且 $L(T^2+Th) \le \frac{1}{12}$。那么对于任意整数

$$k \ge \frac{1}{c_2}\log\left(\frac{c_1 W_\psi(\mu,\tilde{\nu}_{\tilde{\mathbf{Q}}})}{4\sqrt{\log 2}}\left(\delta_1+\sqrt{\delta_1^2+16(\log 2)\delta_2}\right)\right)$$

有

$$\mathsf{R}_q(\mu\tilde{\mathbf{Q}}^k\tilde{\mathbf{P}}_{T,h}\,\|\,\nu) \le \frac{3(\delta_1+\delta_2)}{2(2q-1)}\sqrt{\log 2}\,c_1^2 e^{-c_2 k}\,\max\{1, W_\psi^2(\mu,\tilde{\nu}_{\tilde{\mathbf{Q}}})\} + \mathsf{bias}$$

其中

$$\begin{aligned}\mathsf{bias} = {} & dh(6\mathsf{j}_c + 6\sqrt{d}\,\mathsf{j}_v + 108sdh\mathsf{j}_v^2 + h\mathsf{p}_v^2 + 2\mathsf{p}_v) + 2\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_{xy}+\mathsf{p}_{xy}u)\Delta_h \\ & + \mathsf{p}_{xy}^2(1+3su^2)\Delta_h^2 + 2h\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x+\mathsf{p}_x u)(\Delta_h + K_\nu + \sqrt{\Delta_h^2+K_\nu^2}\,) \\ & + 2h^2\mathsf{p}_x^2(1+3su^2)(\Delta_h^2+K_\nu^2)\,.\end{aligned}$$

定理4 中关于 Δ_h 的上界意味着，在固定积分方案（并验证其满足假设5 ）后，可得到步长 h 的上界。我们在命题4中验证了 uHMC-v 对强对数凹目标分布满足假设5。验证假设4和5后，推论 4 给出了 uHMC-v 的一阶预言机复杂度，使其输出的样本与目标分布之间的 Rényi 散度达到 ϵ 精度。

注 7 (Rényi 散度与 KL 复杂度)。 在 *Rényi* 散度意义下达到给定精度所需的梯度评估总次数，其界（推论 *4*）不如 *KL* 散度意义下相应的界（推论 *2*和*5*）理想。这一差距源于我们的分析本身：在 *Rényi* 散度设定下，单次耦合构造只能利用速度 *Verlet* 积分器的一阶精度，而 *KL* 分析则能够利用其二阶精度；参见第 *3.2* 节。

4.3 示例

我们考虑目标分布 $\nu \propto e^{-f}$ 满足 α-强对数凹（α-SLC），这意味着势函数 $f \in \mathcal{C}^2(\mathbb{R}^d)$ 对某个 α 是 $\alpha > 0$-强凸的，即 $\nabla^2 f(x) \succeq \alpha I$ 对所有 $x \in \mathbb{R}^d$均成立。现在，我们针对这一情形，讨论采用两种不同离散化方案的定理：速度 Verlet 方案（uHMC-v）和分层积分方案 [BRM25] ，后者是一种随机积分器。

4.3.1 采用速度 Verlet 积分方案的强对数凹目标

在目标分布为 α-SLC 的条件下，我们验证了 ˜uHMC-v 核 $\mathbf{P}_{T,h}$ 满足假设 2、3、4、5 ，其平稳分布为 $\tilde{\nu}_h$。

KL 散度保证的 2 和 3 项假设。 uHMC-v 的 Wasserstein–p 距离混合时间保证见于 [BRS23, 定理 18], 该结果验证了假设 2。我们在下文重述该结果。

命题 1（来自 [BRS23] 的定理 18，改编版）。设 μ，$\pi \in \mathcal{P}(\mathbb{R}^d)$，$\nu$ 为 α-强对数凹函数，并假设条件 *1*（*a*）-（*b*）成立。进一步假设 $h \geq 0$ 且 $LT^2 \leq \frac{1}{20}$。那么，对于任意 $p \in [1, \infty)$ 以及整数 $k \geq 0$

$$W_p(\mu \tilde{\mathbf{P}}_{T,h}^k, \pi \tilde{\mathbf{P}}_{T,h}^k) \leq \Big(1 - \frac{\alpha T^2}{10}\Big)^k W_p(\mu, \pi)\,.$$

因此，当 $\tilde{\nu}_h \tilde{\mathbf{P}}_{T,h} = \tilde{\nu}_h$ 时，条件 2 在 $c_1 = 1$ 和 $c_2 = \frac{\alpha T^2}{10}$ 下成立。

为验证条件 3，我们先使用 Wasserstein–2 距离的三角不等式，由此可得

$$\begin{aligned} W_2(\tilde{\nu}_h, \nu) &\leq W_2(\nu, \nu \tilde{\mathbf{P}}_{T,h}) + W_2(\nu \tilde{\mathbf{P}}_{T,h}, \tilde{\nu}_h) \\ &\leq W_2(\nu, \nu \tilde{\mathbf{P}}_{T,h}) + \Big(1 - \frac{\alpha T^2}{10}\Big) W_2(\nu, \tilde{\nu}_h) \end{aligned}$$

其中第二个不等式由命题1得出。因此可得

$$\frac{\alpha T^2}{10} W_2(\tilde{\nu}_h, \nu) \leq W_2(\nu, \nu \tilde{\mathbf{P}}_{T,h})\,. \tag{11}$$

为对此进行界定，[BRE23，引理 23]（采用简化后的常数）指出，在条件 1（a）-（c）以及 $L(T^2 + Th) \leq \frac{1}{6}$ 下，对于任意 x，$v \in \mathbb{R}^d$

$$\max_{s \leq T} \|q_s(x,v) - \tilde{q}_{s,h}(x,v)\| \leq h^2 \Big(\frac{1}{5} L\|x\| + \frac{9}{10} LT\|v\| + \frac{1}{120} M\|x\|^2 + \frac{3}{10} MT^2 \|v\|^2 \Big)\,. \tag{12}$$

因此，设 $X_0 \sim \nu$ 且 $\xi \sim \gamma_d$，并定义 $X_1 := q_T(X_0, \xi)$ $\tilde{X}_1 := \tilde{q}_{T,h}(X_0, \xi)$。则（12）推出

$$\|X_1 - \tilde{X}_1\| \leq h^2 \Big(\frac{1}{5} L\|X_0\| + \frac{9}{10} LT\|\xi\| + \frac{1}{120} M\|X_0\|^2 + \frac{3}{10} MT^2 \|\xi\|^2 \Big)\,.$$

将两边平方，并对 a_1，a_2，a_3，$a_4 \in \mathbb{R}$ 应用 $(a_1 + a_2 + a_3 + a_4)^2 \leq 4(a_1^2 + a_2^2 + a_3^2 + a_4^2)$，得到

$$\|X_1 - \tilde{X}_1\|^2 \leq 4h^4 \Big(\frac{1}{25} L^2 \|X_0\|^2 + \frac{81}{100} L^2 T^2 \|\xi\|^2 + \frac{1}{120^2} M^2 \|X_0\|^4 + \frac{9}{100} M^2 T^4 \|\xi\|^4 \Big)\,.$$

对 $X_0 \sim \nu$ 和 $\xi \sim \gamma_d$ 取期望，并根据定义 1，得到

$$W_2^2(\nu, \nu \tilde{\mathbf{P}}_{T,h}) \leq 4h^4 \Big(\frac{1}{25} L^2 \mathop{\mathbb{E}}_{\nu} \big[\|X_0\|^2\big] + \frac{81}{100} L^2 T^2 \mathop{\mathbb{E}}_{\gamma_d} \big[\|\xi\|^2\big] + \frac{1}{120^2} M^2 \mathop{\mathbb{E}}_{\nu} \big[\|X_0\|^4\big] + \frac{9}{100} M^2 T^4 \mathop{\mathbb{E}}_{\gamma_d} \big[\|\xi\|^4\big] \Big)\,.$$

利用 $\mathbb{E}_{\gamma_d}\big[\|\xi\|^2\big] = d$ 和 $\mathbb{E}_{\gamma_d}\big[\|\xi\|^4\big] = d(d+2)$，并结合（11），得到以下命题。

命题 2。 设 ν 为 α-强对数凹函数，并假设 *1 (a)-(c)* 成立。进一步假设 $h \geq 0$ 且 $L(T^2 + Th) \leq \frac{1}{20}$。则

$$W_2(\tilde{\nu}_h, \nu) \leq h^2 \frac{20}{\alpha T^2} \Big(\frac{1}{25} L^2 \mathop{\mathbb{E}}_{\nu} \big[\|X_0\|^2\big] + \frac{81}{100} dL^2 T^2 + \frac{1}{120^2} M^2 \mathop{\mathbb{E}}_{\nu} \big[\|X_0\|^4\big] + \frac{9}{100} d(d+2) M^2 T^4 \big] \Big)^{\frac{1}{2}}\,.$$

这验证了 uHMC-v 的假设 3 。因此，由于命题 1 和 2 分别验证了 uHMC-v 的假设2和 3 ，我们可以从定理1和 2中得到以下推论。

推论 1 （uHMC-v 的 KL 散度混合时间）。 在定理 *1* 和命题 *1* 成立的条件下，uHMC-v从 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 初始化后，输出样本 X_k ，使得KL（$\mathcal{L}(X_k) \,\|\, \tilde{\nu}_h$） $\leq \epsilon$ ，其中

$$k \geq 1 + \frac{5}{\alpha T^2} \log\left(\frac{\left(\frac{9}{4T^2} + 20dM^2T^4\right) W_2^2(\mu, \tilde{\nu}_h)}{\epsilon} \right).$$

我们还给出了 uHMC-v 为获得目标分布 ν 的 ϵ–精度近似所需的一阶梯度复杂度，其误差度量为 KL 散度。假设 $W_2(\mu, \tilde{\nu}_h)$ 关于 d 为多项式关系，其中 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 为初始分布，我们具体说明了该复杂度对精度ϵ 和维度 d 的依赖关系。

推论 2 （KL 散度保证下的预言机复杂度）。 在定理 *2*以及命题 *1*和 *2*成立的条件下，uHMC-v 输出样本 X，使得KL（$\mathcal{L}(X) \,\|\, \nu$） $\leq \epsilon$，需要 $O\left(d^{3/4}\epsilon^{-1/4}\log(d/\epsilon)\right)$ 次一阶预言机调用。

R´ényi 散度保证的假设 4 和 5 。为了验证假设4，注意到 [BRS23，引理 19]，（证明 [BRS23，定理 18] （即命题1）的关键引理）可以扩展，用于建立 uHMC-v 在 Orlicz-Wasserstein 距离下的混合时间保证。为此， [BRS23，引理 19] 指出，对于任意 x， y， $v \in \mathbb{R}^d$

$$\|\tilde{q}_{T,h}(x, v) - \tilde{q}_{T,h}(y, v)\| \leq \left(1 - \frac{\alpha T^2}{10}\right)\|x - y\|.$$

现在假设 $X_0 \sim \mu$， $Y_0 \sim \pi$，且 $\xi \sim \gamma_d$ 并定义 $X_1 := \tilde{q}_{T,h}(X_0, \xi)$， $Y_1 := \tilde{q}_{T,h}(Y_0, \xi)$。根据 [BRS23，引理 19]，我们知道

$$\|X_1 - Y_1\| \leq \left(1 - \frac{\alpha T^2}{10}\right)\|X_0 - Y_0\|.$$

根据定义 2，注意，对于两个随机变量 Z， Z'，若 $\|Z\| \leq \rho\|Z'\|$ 对于某个 $\rho > 0$，则 $\|Z\|_\psi \leq \rho\|Z'\|_\psi$。因此我们得到

$$\|X_1 - Y_1\|_\psi \leq \left(1 - \frac{\alpha T^2}{10}\right)\|X_0 - Y_0\|_\psi.$$

对 X_0、 Y_0 的初始耦合取下确界，并根据定义 3，可得

$$W_\psi(\mu\tilde{\mathbf{P}}_{T,h}, \pi\tilde{\mathbf{P}}_{T,h}) \leq \left(1 - \frac{\alpha T^2}{10}\right)W_\psi(\mu, \pi).$$

将此过程迭代多步，可得以下命题。

命题 3。 设 μ， $\pi \in \mathcal{P}(\mathbb{R}^d)$， ν 是 α-强对数凹的，并假设假设 *1* (*a*) - (*b*) 成立。进一步假设 $h \geq 0$ 并且 $LT^2 \leq \frac{1}{20}$。那么对于任意整数$k \geq 0$

$$W_\psi(\mu\tilde{\mathbf{P}}_{T,h}^k, \pi\tilde{\mathbf{P}}_{T,h}^k) \leq \left(1 - \frac{\alpha T^2}{10}\right)^k W_\psi(\mu, \pi).$$

接下来需要验证假设 5。这与命题 2的证明概要类似。我们首先结合三角不等式以及命题3，得到

$$\frac{\alpha T^2}{10} W_\psi(\tilde{\nu}_h, \nu) \le W_\psi(\nu, \nu\tilde{\mathbf{P}}_{T,h}) \,. \tag{13}$$

˜令 $X_0 \sim \nu$， $\xi \sim \boldsymbol{\gamma}_d$，并定义 $X_1 := q_T(X_0, \xi)$ $X_1 := \tilde{q}_{T,\,h}(X_0, \xi)$。则由引理 27可得

$$\|X_1 - \tilde{X}_1\| \le h \left(\frac{6}{25}\|\xi\| + \frac{7}{30} LT \|X_0\| \right) .$$

将两边除以 $\lambda > 0$ （稍后确定）并平方，得到

$$\frac{\|X_1 - \tilde{X}_1\|^2}{\lambda^2} \le \frac{2h^2}{\lambda^2} \left(\frac{36}{625}\|\xi\|^2 + \frac{49}{900} L^2 T^2 \|X_0\|^2 \right)$$

因此，

$$\exp\left(\frac{\|X_1 - \tilde{X}_1\|^2}{\lambda^2} \right) \le \exp\left(\frac{2h^2}{\lambda^2} \left(\frac{36}{625}\|\xi\|^2 + \frac{49}{900} L^2 T^2 \|X_0\|^2 \right) \right) .$$

对 $X_0 \sim \nu$ 和 $\xi \sim \boldsymbol{\gamma}_d$ 两边取期望，得到

$$\mathbb{E}\left[\exp\left(\frac{\|X_1 - \tilde{X}_1\|^2}{\lambda^2} \right) \right] \le \mathbb{E}\left[\exp\left(\frac{h^2}{\lambda^2} \frac{72}{625}\|\xi\|^2 \right) \right] \mathbb{E}\left[\exp\left(\frac{h^2}{\lambda^2} \frac{98}{900} L^2 T^2 \|X_0\|^2 \right) \right]$$

利用性质 高斯分布和引理 1，可进一步将其上界估计为 y

$$\mathbb{E}\left[\exp\left(\frac{\|X_1 - \tilde{X}_1\|^2}{\lambda^2} \right) \right] \le \left(1 - 2\frac{h^2}{\lambda^2}\frac{72}{625} \right)^{-\frac{d}{2}} 2^{\frac{h^2}{\lambda^2}\frac{98}{900} L^2 T^2} .$$

只要 $\frac{h^2}{\lambda^2}\frac{72}{625} \le \frac{1}{2}$和 $\frac{h^2}{\lambda^2}\frac{98}{900} L^2 T^2 \le \frac{1}{K_\nu^2}$。为确保这一点，并使右侧的每一项都小于 $\sqrt{2}$，我们选择满足以下条件的 λ

$$\lambda \ge h \, \max\left\{ \sqrt{d}, \frac{7}{15} LTK_\nu \right\} .$$

因此，在该 条件下，右侧为 ≤ 2。因此，根据定义 3 可得 at

$$W_\psi(\nu, \nu\tilde{\mathbf{P}}_{T,h}) \le h \, \max\left\{ \sqrt{d}, \frac{7}{15} LTK_\nu \right\} ,$$

将其与式（13）结合，即得以下命题。

命题 4. 设 ν 为 α-强对数凹，假设 $X \sim \nu$ 满足 $\|X\|_\psi \le K_\nu$，并假设条件 *1(a)-(b)* 成立。进一步假设 $h \ge 0$ 且 $L(T^2 + Th) \le \frac{1}{20}$。则

$$W_\psi(\tilde{\nu}_h, \nu) \le h \frac{10}{\alpha T^2} \max\left\{ \sqrt{d}, \frac{7}{15} LTK_\nu \right\} .$$

命题3和 4验证了 uHMC-v 的假设4 和5，由此得到以下推论。

推论 3 (uHMC-v 在 R´ényi 散度下的混合时间)。 在满足以下条件、使定理 *3* 和命题 *3* 成立时，uHMC-v 从 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 初始化后，输出一个样本 X_k，使得$\mathsf{R}_q(\mathcal{L}(X_k) \,\|\, \tilde{\nu}_h) \le \epsilon$ 对于

$$k \ge 1 + \frac{10}{\alpha T^2} \log\left(\frac{2\left(\delta_1 + 2\sqrt{\log 2}\max\left\{\delta_2, \sqrt{\delta_2}\right\}\right)\max\left\{1, W_\psi^2(\mu, \tilde{\nu}_h)\right\}}{\epsilon} \right).$$

现在给出 uHMC-v 的一阶梯度复杂度，以获得目标分布 ν 在 R´ényi 散度下的 ϵ 精度近似。假设 $W_\psi(\mu,\ \tilde{\nu}_h)$ 关于 d 呈多项式增长，其中 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 是初始分布，并说明复杂度对精度ϵ 和维度 d 的依赖关系。

推论 4 (保证 R´ényi 散度的预言机复杂度)。 在满足定理 *4* 和命题 *3* 及 *4* 的条件下，uHMC-v 输出一个样本 X，使得 $\mathsf{R}_q(\mathcal{L}(X) \,\|\, \nu) \le \epsilon$，且需要 $O\left(d^{3/2}\epsilon^{-1}\log(d/\epsilon)\right)$ 次一阶预言机调用。

4.3.2 强对数凹目标分布与分层积分方案

我们验证给定于 [BRM25]中的 uHMC 方案满足假设 2、3、4。我们将该方案称为˜uHMC-s（分层方案），将转移核记为$\mathbf{Q}_{T,\ h}$，并将该马尔可夫链的平稳分布记为 $\tilde{\nu}_{\tilde{\mathbf{Q}}_{T,\ h}}$。

假设 2 和 4 适用于 uHMC-s，这一点可由 [BRM25，引理 5]得出。BRM25 中的几乎必然压缩性 [BRM25，引理 5] 立即可得任意 Wasserstein–p 距离（当 p有限时）以及 Orlicz-Wasserstein 距离的混合时间保证（类似于命题3的证明概要）。BRM25] 中的定理 6 明确针对 Wasserstein–2 距离表述，但我们在下文给出由 [BRM25，引理 5]推出的更一般定理。

命题 5 (由 [BRM25]中的引理 5 推出)。 设 $\mu,\ \pi \in \mathcal{P}(\mathbb{R}^d),\ \ \nu$ 是 α-强对数凹的，并假设假设 *1(a)-(b)* 成立。进一步假设 $h \ge 0$ 且 $LT^2 \le \frac{1}{8}$。那么对于 $W \in \{W_p, W_\psi\}$ 任意 $p \in [1, \infty)$，以及任意整数 $k \ge 0$

$$W(\mu\tilde{\mathbf{Q}}_{T,h}^k, \pi\tilde{\mathbf{Q}}_{T,h}^k) \le \left(1 - \frac{\alpha T^2}{6}\right)^k W(\mu, \pi)\,.$$

假设3 由 [BRM25，定理 10], 可得，如下所述。

命题 6(BRM 中的定理 10 [BRM25])。 设 ν 为 α-强对数凹，且假设*1(a)-(b)*成立。进一步假设 $h \ge 0$ 且 $LT^2 \le \frac{1}{8}$。则

$$W_2(\tilde{\nu}_{\tilde{\mathbf{Q}}_{T,h}}, \nu) \le h^{\frac{3}{2}} 142\sqrt{d}\frac{6}{\alpha T^2}\left(\frac{L}{\alpha}\right)^{\frac{1}{2}} L^{\frac{1}{4}}\,.$$

验证 uHMC-s 的假设5 的一种方法是将 [BRM25引理 7]推广到 ψ- Orlicz 精度；本文不探讨这个问题。命题5 和6与定理2 共同给出以下一阶复杂度保证：对于目标分布的 KL 散度精度，uHMC-s 的保证。正如预期，与 Verlet 积分格式的推论2 相比，我们可以看到维度依赖性有所改善，但对误差 ϵ的依赖性更差，正如 [BRM25] 对 Wasserstein–2 距离所述。我们展示精度 ϵ 和维度 d 的依赖性，并假设 $W_2(\mu,\ \tilde{\nu}_h)$ 关于 d 呈多项式关系，其中 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 是初始分布。

推论5 (分层方案的 KL 散度保证之预言机复杂度**)**。 在满足定理*2* 和命题*5* 及*6* 的条件下，uHMC-s *(*最后一步采用uHMC-v*)* 输出一个样本 X，使得$\mathsf{KL}(\mathcal{L}(X)\,\|\,\nu) \le \epsilon$，需要 $O\left(\max\left\{d^{2/3}\epsilon^{-1/3}, d^{3/4}\epsilon^{-1/4}\right\}\log(d/\epsilon)\right)$ 次一阶预言机调用。

4.4 信息收缩的应用

本文研究的混合时间和渐近偏差保证，描述的是马尔可夫链上边缘分布与平稳分布或目标分布之间的散度；然而，它们并不能直接量化马尔可夫链上的两个迭代何时（近似）独立，这需要分析两个迭代的联合分布，并考察其何时接近乘积分布。因此，两个迭代之间的互信息刻画了它们的依赖性；这是一种强独立性概念，因为它意味着迭代之间的协方差很小 [LMW25，引理 9]，并且它们相对于测试函数相互独立。量化马尔可夫链生成的样本之间的依赖性，有助于将这些样本用于下游应用；这也促使 [LMW25] 研究互信息泛函沿朗之万动力学及其离散化方法（如未校正朗之万算法）的收缩 [RT96、VW23] 以及近端采样器 [LST21、CCSW22]。利用 [LMW25附录 I], 中讨论的从混合时间到信息收缩的一般化简，我们运用定理 1 得到 uHMC-v 信息收缩的界。

两个随机变量之间的互信息是其联合分布的函数，定义如下。

定义 7。 给定定义在 $\mathbb{R}^d$ 上的两个随机变量 X, Y，满足 $(X,Y) \sim \rho^{XY} \in \mathcal{P}(\mathbb{R}^{2d})$，互信息泛函 互信息 : $\mathcal{P}(\mathbb{R}^{2d}) \to \mathbb{R}$ 定义为

$$\mathsf{MI}(X;Y) \equiv \mathsf{MI}(\rho^{XY}) \coloneqq \mathsf{KL}\left(\rho^{XY} \,\|\, \rho^X \otimes \rho^Y\right).$$

设 $X_0 \sim \mu$ 其中 μ 任意，并设 $X_k \sim \mu\, \tilde{\mathbf{P}}_{T,h}^k$，那么我们希望研究 互信息$(X_0; X_k)$。注意，当 $\mu = \tilde{\nu}_h$ 时，这将告诉我们何时能沿 uHMC-v 马尔可夫链获得近似独立同分布样本。现在我们重述 [LMW25 引理 13], ，该引理描述了如何将一般马尔可夫转移核$\tilde{\mathrm{P}}\,\mathbf{P}$.

~命题 7 (引理 13，出自 [LMW25])。 设 $\mathbf{P}$ 是一个平稳分布为$\tilde{\nu}$的马尔可夫链。设 $X_0 \sim \mu \in \mathcal{P}(\mathbb{R}^d)$ 为初始迭代点，并通过 $X_k \sim \mu\, \tilde{\mathbf{P}}^k$ 定义后续迭代点，其中$k \ge 1$为整数。给定误差阈值 $\epsilon \ge 0$, 假设存在 $\tau \in \mathbb{N}$ 使得 KL散度$(\delta_x\, \tilde{\mathbf{P}}^\tau \,\|\, \tilde{\nu}) \le \epsilon$ 对所有 $x \in \mathbb{R}^d$均成立。那么，我们有

$$\mathsf{MI}(X_0; X_\tau) \le \epsilon\,.$$

根据定理1，特别是推论1 满足命题7关于从狄拉克初始化获得混合保证的要求，因此我们得到以下推论，描述 uHMC-v 的信息压缩速率。

推论 6. 假设定理*1*和命题*1*所需的条件成立。给定$X_0 \sim \mu \in \mathcal{P}(\mathbb{R}^d)$，定义 $X_k \sim \mu\, \tilde{\mathbf{P}}_{T,h}^k$，其中整数为 $k \ge 1$。则以下结论成立

$$\mathsf{MI}(X_0; X_k) \le e^{-\frac{\alpha T^2}{5}(k-1)}\left(\frac{9}{4T^2} + 20dM^2T^4\right)\mathop{\mathbb{E}}_{(X,Y)\sim\mu\otimes\tilde{\nu}_h}\left[\|X-Y\|^2\right].$$

我们在第 6 节 J中证明推论。注意，如果马尔可夫链初始化于 $X_0 \sim \delta_x$ ，对于某个 $x \in \mathbb{R}^d$，那么 $X_0 = x$ 与任意后续迭代之间的互信息为 0。

5 获得 KL 散度保证的关键步骤

5.1 混合时间

˜在本小节中，我们将马尔可夫转移核固定为 uHMC-v，即 $\mathbf{P}_{T,\ h}$。我们从引理 2 开始，该引理表明，对于任意两个点 $x,\ y \in \mathbb{R}^d$

$$\mathsf{KL}(\delta_x \tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y \tilde{\mathbf{P}}_{T,h}) \le \mathsf{KL}((\varphi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d)\,,$$

其中请回忆，$\varphi_{x,\ y}$ 是满足（7）的映射。接下来，我们在引理 10 中界定标准高斯分布与扰动标准高斯分布之间的 KL 散度。我们在 F.1节中证明引理 10 。

引理 10. 设 $\psi : \mathbb{R}^d \to \mathbb{R}^d$ 是可微双射。假设存在常数 $0 \le \mathfrak{m}_1 < \infty$ 和 $0 \le \mathfrak{m}_2 < 1$，使得对于所有 $v \in \mathbb{R}^d$, $\|\psi(v) - v\| \le \mathfrak{m}_1$ 和 $\|\nabla\psi(v) - I\|_{\mathsf{op}} \le \mathfrak{m}_2$。则

$$\mathsf{KL}(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \frac{1}{2}\mathfrak{m}_1^2 + \frac{d\,\mathfrak{m}_2^2}{2(1-\mathfrak{m}_2)}\,.$$

结合引理 2 和引理 10 以及 $\psi = \varphi_{x,\ y}$ (7)，可得 uHMC-v 从狄拉克初始化出发时的 KL 散度正则化界，见引理 11。我们在 F.2节中证明引理 11。

引理 11. 设 $x, y \in \mathbb{R}^d$ 并假设条件 *1(a)-(c)* 成立。此外，假设 $h \ge 0$ 且 $L(T^2 + Th) \le \frac{1}{12}$。则

$$\mathsf{KL}(\delta_x \tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y \tilde{\mathbf{P}}_{T,h}) \le \left(\frac{9}{4T^2} + 20dM^2T^4\right)\|x - y\|^2.$$

现在，我们可以利用引理 11 将引理 29 应用于从狄拉克初始化推广到一般初始化的情形˜，其中 $\mathbf{P} = \mathbf{Q} = \mathbf{P}_{T,h}$，由此为 uHMC-v 在 KL 散度下给出正则化保证。我们在引理 12 中陈述该结果，并在第 F.3 节证明。

引理 12. 设 $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$ 并假设条件 *1(a)-(c)* 成立。此外，假设 $h \ge 0$ 且 $L(T^2 + Th) \le \frac{1}{12}$。则

$$\mathsf{KL}(\mu \tilde{\mathbf{P}}_{T,h} \,\|\, \pi \tilde{\mathbf{P}}_{T,h}) \le \left(\frac{9}{4T^2} + 20dM^2T^4\right) W_2^2(\mu, \pi)\,.$$

˜ ˜结合引理 12与条件 2 对于 $\mathbf{Q} = \mathbf{P}_{T,h}$ 可得定理 1。我们在第 1 节证明定理 F.4。

5.2 渐近偏差

我们首先给出引理 3，它表明对于任意两个点 $x,\ y \in \mathbb{R}^d$

$$\mathsf{KL}(\delta_x \tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y \mathbf{P}_T) \le \mathsf{KL}((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d)\,,$$

其中回顾一下，$\Phi_{x,\ y}$ 是满足（10）的映射。我们在以下引理中界定KL散度($(\Phi_{x,\ y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d$)。我们在第13节G.1中证明引理 13。

引理 13。假设条件*1(a)-(d)*成立。进一步假设 $h \geq 0$ 且 $L(T^2+Th) \leq \frac{1}{12}$。那么对于任意 $x,\ y \in \mathbb{R}^d$

$$\mathsf{KL}((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \leq \Big(\frac{9}{4T^2}+\frac{363}{2}dM^2T^4\Big)\|x-y\|^2+ 4h^4\Bigg[45dL^2+\Big(4\frac{L^2}{T^2}+45dM^2\Big)\|x\|^2+\Big(4\frac{M^2}{T^2}+45d(M^2T^2+N)^2\Big)\|x\|^4 +(4L^2+45dM^2T^2)d+\big(4M^2T^2+45d(M^2T^4+NT^2)^2\big)d(d+2)\Bigg].$$

现在，我们可以得到 uHMC-v 的 KL 散度交叉正则化界。我们在14 引理G.2中证明。

引理 14. 设 μ 是一个分布，使得 $\mathbb{E}_\mu[\|x\|^2]=m_2(\mu)<\infty$ 且 $\mathbb{E}_\mu[\|x\|^4]=m_4(\mu)<\infty$，并且 π 是一个分布，使得 $W_2(\mu,\pi)<\infty$。假设假设条件*1(a)-(d)*成立。进一步假设 $h \geq 0$ 且 $L(T^2+Th) \leq \frac{1}{12}$。则

$$\mathsf{KL}(\mu\tilde{\mathbf{P}}_{T,h} \,\|\, \pi\mathbf{P}_T) \leq \Big(\frac{9}{4T^2}+\frac{363}{2}dM^2T^4\Big)W_2^2(\mu,\pi)+ 4h^4\Bigg[45dL^2+\Big(4\frac{L^2}{T^2}+45dM^2\Big)m_2(\mu)+\Big(4\frac{M^2}{T^2}+45d(M^2T^2+N)^2\Big)m_4(\mu) +(4L^2+45dM^2T^2)d+\big(4M^2T^2+45d(M^2T^4+NT^2)^2\big)d(d+2)\Bigg].$$

引理 14与 假设条件2 和3 共同推出定理2，我们将在第G.3节证明该定理 .

6 获得 R´ényi 散度保证的关键步骤

6.1 混合时间

˜在本小节中，我们固定转移核为 uHMC-v，即 $\mathbf{P}_{T,\,h}$。引理 2 表明，对于任意两个点 $x,\ y \in \mathbb{R}^d$

$$\mathsf{R}_q(\delta_x\tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y\tilde{\mathbf{P}}_{T,h}) \leq \mathsf{R}_q((\varphi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d)\,,$$

其中回顾一下，$\varphi_{x,\,y}$ 是满足（7）的映射。接下来，我们在引理 15中界定扰动标准高斯分布与标准高斯分布之间的 R´ényi 散度。我们在 H.1 节证明引理 15。

引理 15。设 $\psi:\ \mathbb{R}^d \to \mathbb{R}^d$ 是一个可微双射。假设存在常数 $0 \leq \mathfrak{m}_1 < \infty$ 和 $0 \leq \mathfrak{m}_2 < 1$，使得对于所有 $v \in \mathbb{R}^d$, $\|\psi(v)-v\| \leq \mathfrak{m}_1$ 以及 $\|\nabla\psi(v)-I\|_{\mathsf{op}} \leq \mathfrak{m}_2$。那么对于任意 $1<q<\infty$

$$\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \leq \frac{d\mathfrak{m}_2}{1-\mathfrak{m}_2}+\sqrt{d}\,\mathfrak{m}_1+\frac{q\mathfrak{m}_1^2}{2}\,.$$

结合引理2 和15 以及$\psi=\varphi_{x,\,y}$ (7) 可得 uHMC-v 在 R´ényi 散度下从狄拉克初始分布出发的正则化界，见引理16。我们在第16节H.2中证明引理16。

引理 16. 设 x， $y \in \mathbb{R}^d$ 并假设条件*1(a)-(c)* 成立。此外，假设 $h \geq 0$ 且 $L(T^2+Th) \leq \frac{1}{12}$。那么，对于任意 $1 < q < \infty$

$$\mathsf{R}_q(\delta_x \tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y \tilde{\mathbf{P}}_{T,h}) \leq \left(8dMT^2 + \frac{3\sqrt{d}}{2T}\right)\|x-y\| + \frac{9q}{8T^2}\|x-y\|^2 .$$

现在，我们可以利用引理16 和引理30，将引理 = **从狄拉克初始化推广到一般初始化** =。T其中h我们在引理17 中陈述这一结果，并在第H.3节证明。

引理 17。设 μ， $\pi \in \mathcal{P}(\mathbb{R}^d)$，并假设假设条件*1(a)-(c)* 成立。进一步假设 $h \geq 0$ 且 $L(T^2+Th) \leq \frac{1}{12}$。那么，对于任意 $1 < q < \infty$

$$\mathsf{R}_q(\mu \tilde{\mathbf{P}}_{T,h} \,\|\, \pi \tilde{\mathbf{P}}_{T,h}) \leq \inf_{\gamma \in \Gamma(\mu,\pi)} \frac{1}{q-1} \log \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \left[e^{(q-1)\left[\left(8dMT^2 + \frac{3\sqrt{d}}{2T}\right)\|x-y\| + \frac{9q}{8T^2}\|x-y\|^2\right]} \right] .$$

关于Q的假设条件 = $\mathbf{P}_{T,\,h}$ 结合引理17 可得定理3，我们将在第H.4节证明。

6.2 渐近偏差

我们先给出引理3，它表明对于任意两个点 x， $y \in \mathbb{R}^d$

$$\mathsf{R}_q(\delta_x \tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y \mathbf{P}_T) \leq \mathsf{R}_q((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) ,$$

其中回顾一下，$\Phi_{x,\,y}$ 是一个满足（10）的映射。我们在以下引理中给出$\mathsf{R}_q((\Phi_{x,\,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d)$ 的界。我们在第18I.1 节证明引理。

引理18。 假设条件 *1(a)-(c)* 成立，并回顾定义 *6*中定义的常数。此外，进一步假设

$$0 \leq h \leq \frac{25T}{7}\left[\left(1 + \frac{1}{4(q-1)}\right)^{\frac{1}{2}} - 1\right]$$

并且 $L(T^2+Th) \leq \frac{1}{12}$。那么对于任意的 x、$y \in \mathbb{R}^d$ 和 $1 < q < \infty$

$$\begin{aligned}\mathsf{R}_q((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \leq{}& \mathsf{p}_{xy}^2\|x-y\|^2(1+3(q-1)(h\mathsf{p}_v+1)^2) + \sqrt{d}\,\|x-y\|(6\sqrt{d}\,\mathsf{j}_{xy} + \mathsf{p}_{xy}(h\mathsf{p}_v+1)) \\ &+ h^2\mathsf{p}_x^2\|x\|^2(1+3(q-1)(h\mathsf{p}_v+1)^2) + h\sqrt{d}\,\|x\|(6\sqrt{d}\,\mathsf{j}_x + \mathsf{p}_x(h\mathsf{p}_v+1)) \\ &+ dh(6\mathsf{j}_c + 6\sqrt{d}\,\mathsf{j}_v + 108(q-1)dh\mathsf{j}_v^2 + h\mathsf{p}_v^2 + 2\mathsf{p}_v) .\end{aligned}$$

我们在第 19 节证明引理I.2。

引理 19。 设 μ， $\pi \in \mathcal{P}(\mathbb{R}^d)$，假设条件 *1(a)-(c)* 成立，并回顾定义*6*中定义的常数。此外，进一步假设

$$0 \leq h \leq \frac{25T}{7}\left[\left(1 + \frac{1}{4(q-1)}\right)^{\frac{1}{2}} - 1\right]$$

并且 $L(T^2+Th) \leq \frac{1}{12}$。那么对于任意 $1 < q < \infty$

$$\mathsf{R}_q(\mu \tilde{\mathbf{P}}_{T,h} \,\|\, \pi \mathbf{P}_T) \leq \inf_{\gamma \in \Gamma(\mu,\pi)} \frac{1}{q-1} \log \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \left[e^{(q-1)(\alpha_0 + \alpha_1\|x-y\| + \alpha_2\|x-y\|^2 + \alpha_3\|x\| + \alpha_4\|x\|^2)} \right]$$

其中

$$\begin{aligned}
\alpha_0 &\coloneqq dh(6\mathsf{j}_c + 6\sqrt{d}\,\mathsf{j}_v + 108(q-1)dh\mathsf{j}_v^2 + h\mathsf{p}_v^2 + 2\mathsf{p}_v)\\
\alpha_1 &\coloneqq \sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_{xy} + \mathsf{p}_{xy}(h\mathsf{p}_v + 1))\\
\alpha_2 &\coloneqq \mathsf{p}_{xy}^2(1 + 3(q-1)(h\mathsf{p}_v + 1)^2)\\
\alpha_3 &\coloneqq h\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x + \mathsf{p}_x(h\mathsf{p}_v + 1))\\
\alpha_4 &\coloneqq h^2\mathsf{p}_x^2(1 + 3(q-1)(h\mathsf{p}_v + 1)^2)\,.
\end{aligned}$$

引理 19 结合假设 4 和 5 可得定理 4，证明见第 I.3 .

7 讨论

本文研究了 uHMC 在 KL 散度和 R´ényi 散度下的混合时间与渐近偏差。为得到这些结果，我们利用单次耦合推导出 uHMC-v 在 KL 和 R´ényi 散度下的正则化与交叉正则化结果，从而能够将 Wasserstein 保证提升至 KL 和 R´ényi 散度。

我们现在讨论理论保证及其推论，并将其与其他采样方法的结果进行比较，尤其是未校正朗之万算法（uLA） [RT96、RR98、VW23]，并概述未来的研究方向。

我们总结了不同采样方法所需的梯度预言机复杂度，以输出一个样本X，使得 $\mathsf{D}(\mathcal{L}(X) \,\|\, \nu) \le \epsilon$，见表 1；这里考虑强对数凹目标分布$\nu$、精度 $\epsilon > 0$和散度 D；回顾一下，我们在第 4.3 节验证了强对数凹情形下的所有假设。表1 中的第 1–6 行给出了固定核方案的保证，其中在整个采样过程中使用同一个马尔可夫核；第 7–10 行对应于复合核方案，这体现了渐近偏差界的模块化：第一个马尔可夫核驱动 Wasserstein 收缩（例如第 7 行中的 uHMC-s），第二个马尔可夫核则通过交叉正则化将保证提升至强散度（例如第 7 行中的 uHMC-v）。

据我们所知，唯一一项专门研究 uHMC-v 在 KL 散度下保证的已有工作是 [CDMS23] （表 1，第 3 行），该工作利用次椭圆性技术，在目标分布满足对数 Sobolev 不等式假设的条件下，给出了 KL 散度下的渐近偏差保证。相比之下，对于强对数凹目标分布这一常见设定，我们在推论 2 （表 1，第 2 行）中获得了更低的梯度复杂度。此外，渐近偏差界通常并不意味着混合时间保证；据我们所知，定理1 是首个关于 uHMC-v 在 KL 散度下混合时间的保证，并且具有多种应用，例如研究马尔可夫链中互信息的压缩（见第 4.4节）。除了我们在 R´ényi 散度下给出的混合时间和渐近偏差界（第 4.2节）之外，据我们所知，eHMC 或 uHMC 尚无超越 KL 散度的保证。如第 1节所述，我们在 R´ényi 散度下的保证使 uHMC-v 能够为校正马尔可夫链生成热启动（表 1，第 4 行）。正如第 1.1节所强调的，由于速度 Verlet 积分器具有复杂的结构，分析该积分器颇具挑战性；值得注意的是，我们关于 uHMC-v 在 KL 散度下的梯度复杂度（表 1，第 2 行）与已有关于 uHMC-v 在全变差距离下的最佳梯度复杂度（表 1，第 1 行）相当。表1第 2 行还表明，与 uLA 在 KL 散度下的梯度复杂度（表 1，第 5 行）相比，uHMC-v 在维度依赖性和精度依赖性方面均更优；事实上，uHMC-v 具有最佳的精度

	算法	梯度复杂度	散度 (D)	参考文献
	固定核方案			
1	uHMC-v	$d^{3/4}/\epsilon^{1/4}$	TV^2	[BRE23]
2	uHMC-v	$d^{3/4}/\epsilon^{1/4}$	KL	推论 2
3	uHMC-v	$d/\epsilon^{1/4}$	KL	[CDMS23]
4	uHMC-v	$d^{3/2}/\epsilon$	R_q	推论 4
5	uLA	d/ϵ	KL	[VW23]
6	uLA	d/ϵ	R_q	[CEL+24]
	复合核方案			
7	uHMC-s + uHMC-v	$\max\{d^{2/3}/\epsilon^{1/3}, d^{3/4}/\epsilon^{1/4}\}$	KL	推论 5
8	uHMC-v +uLA	$d^{5/8}/\epsilon^{3/8}$	KL	章节 L.1
9	uHMC-s + uLA	$d^{1/2}/\epsilon^{1/2}$	KL	章节 L.1
10	uHMC-v +uLA	$d^{3/4}/\epsilon^{3/4}$	R_q	第 L.2 节

表 1：输出一个样本 X，使得 $\mathsf{D}(\mathcal{L}(X) \,\|\, \nu) \le \epsilon$，其中$\nu$ 是强对数凹目标分布，D 表示统计距离或散度。为便于展示，“梯度复杂度”列省略了对数因子，仅表示其对维度 d 和精度 ϵ的依赖性。第 1–6 行对应于在整个采样过程中始终使用固定核的马尔可夫链蒙特卡洛方法；第 7–10 行体现了我们框架的模块化特性，对应于由第一个马尔可夫核驱动 Wasserstein 收缩、由第二个核进行交叉正则化的方法，例如，第 7 行对应于由 uHMC-s 驱动的 Wasserstein 收缩，以及来自 uHMC-v 的交叉正则化。在第 1 行中，我们给出平方全变差距离的保证，以便将其与 KL 散度和 R´ényi 散度置于可比尺度上——Pinsker不等式和 (2) 意味着，对于$q > 1$、 μ、 $\pi \in \mathcal{P}(\mathbb{R}^d)$，$2\mathsf{TV}^2(\mu,\pi) \le \mathsf{KL}(\mu \,\|\, \pi) \le \mathsf{R}_q(\mu \,\|\, \pi)$。

表 1 所列所有方法之间的依赖关系1，其中可以利用速度 Verlet 积分器的二阶精度（表 1，第 1 至 3 行）。

我们证明策略的一个关键方面，是将获得混合时间和渐近偏差界所需的 Wasserstein 假设，与核的正则化和交叉正则化性质解耦。这种解耦是我们模块化框架的核心，使得我们能够为复合核方案建立渐近偏差界：其中 Wasserstein 收缩由一个核驱动，而负责将保证提升至 KL散度或 R´ényi 散度的最终交叉正则化步骤则由另一个核实现。这种模块化在以下情形中特别有价值：对于给定核而言，建立其在 KL散度或 R´ényi 散度下的交叉正则化性质较为困难，但仍可推导出 Wasserstein 混合和渐近偏差界。据我们所知，此前唯一一种能够将 Wasserstein 保证扩展到更强散度下渐近偏差界的方法，是 [AC24c] 中研究的局部误差框架。然而，正如作者在 [AC24c 的附录 A 中指出的，2], 由于其框架中的 Wasserstein–∞ 假设过于严格，该方法目前仅限于 KL 散度，无法在 R

é´nyi 散度下给出渐近偏差界。表1中的第 7–10 行对应于使用我们框架推导出的复杂度保证，其中两个不同的核分别负责 Wasserstein 收缩和交叉正则化——第 7–9 行涉及 KL 散度，第 10 行涉及 R´ényi 散度。第7行对应于由 uHMC-s [BRM25] 控制 Wasserstein 收缩，并由 uHMC-v 提供交叉正则化的情形；我们将在第 4.3.2 节讨论这一点。我们在附录L中概述了由 uLA 核提供交叉正则化时的梯度复杂度，这对应于表1中的最后三行。通过 uLA 实现交叉正则化所得到的保证表明，与始终使用 uLA 相比，使用 uHMC方法驱动 Wasserstein 收缩能够带来显著改进（表 1，第 5 行）；同时，uLA 的交叉正则化性质相比 uHMC-v 的交叉正则化表现出更温和的维度依赖性。因此，当维度依赖性比精度更为重要，或希望获得 R´ényi 散度下的保证时，uLA 颇具吸引力。未来的工作可以研究考虑其他马尔可夫核时能够获得的梯度复杂度保证，并进一步考察该框架如何与强对数凹性、光滑性等其他问题参数相互作用；本文对此不作重点讨论。

展望未来，研究能否为 uHMC 在 Φ-散度 [Cha04], 方面获得混合时间和渐近偏差界，以及能否获得 Wasserstein–p保证，从而将 uHMC 的适用范围扩展到强对数凹性之外，也很值得关注。需要指出的是，为方便起见，我们将假设 2 和4 表述为指数快速收缩，不过，我们的证明技巧也适用于收缩速率较弱的 Wasserstein 收缩，只是该速率会影响最终的 KL 或 R´ényi 收敛。

致谢。 SM 和 AW 获得了美国国家科学基金会项目 CCF-2403391 和 CCF-2443097 的资助。

哈密顿蒙特卡罗背景介绍

我们讨论精确哈密顿蒙特卡洛的若干关键性质 [DKPR87、Nea11]，用于从目标概率分布 $\nu \propto e^{-f}$（其支撑集为 $\mathbb{R}^d$）进行采样。全文中，我们采用第 2.2 节中的记号。

- 哈密顿量守恒。设 $(x, v) \in \mathbb{R}^{2d}$ 为某个初始条件。那么对于任意 $t \geq 0$,$H(\mathsf{HF}_t(x, v)) = H(x, v)$, 因此哈密顿量 H 沿着哈密顿动力学 (4) 保持守恒。为此，记 $(x_t, v_t) \equiv (x_t(x, v)\ v_t(x, v)) = \mathsf{HF}_t(x, v)$。然后，利用链式法则和哈密顿动力学 (4) 的定义，可得

$$\frac{d}{dt} H(x_t, v_t) = \langle \nabla f(x_t), v_t \rangle + \langle v_t, -\nabla f(x_t) \rangle = 0\,,$$

  因此，哈密顿量沿着哈密顿动力学（4）保持守恒。

- **辛性**。定义如下 $2d \times 2d$ 矩阵

$$J := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

  若对于所有 $(x,\ v) \in \mathbb{R}^{2d}$ 和 $t \geq 0$，流 HF_t 都满足以下条件，则称其为辛的：

$$(\nabla \mathsf{HF}_t(x, v))^\top J\, \nabla \mathsf{HF}_t(x, v) = J. \tag{14}$$

  固定 $(x, v) \in \mathbb{R}^{2d}$，并记 $M_t := (\nabla \mathsf{HF}_t(x, v))^\top J\, \nabla \mathsf{HF}_t(x, v)$。注意到 $M_0 = J$。可以验证 $M_t = 0$，因此对于所有 $t \geq 0$, $M_t = M_0 = J$，均有此结论。于是，辛性条件 (14) 对所有 $(x, v) \in \mathbb{R}^{2d}$ 和 $t \geq 0$ 均成立。

  辛结构的一个重要推论是，哈密顿流保持体积，即对所有 $(x, v) \in \mathbb{R}^{2d}$ 和 $t \geq 0$，$\det(\nabla \mathsf{HF}_t(x, v)) = 1$。为此，固定 $(x, v) \in \mathbb{R}^{2d}$，并对 (14) 两边取行列式，得到 $(\det(\nabla \mathsf{HF}_t(x, v)))^2 = 1$。因此，$\det(\nabla \mathsf{HF}_t(x, v)) = \pm 1$。由于 $t \mapsto \det(\nabla \mathsf{HF}_t(x, v))$ 连续，且 $\det(\nabla \mathsf{HF}_0(x, v)) = 1$，所以对所有 $t \geq 0$，$\det(\nabla \mathsf{HF}_t(x, v)) = 1$。

- **时间反演**。设沿哈密顿动力学方程（4）演化的解在时间 $t \geq 0$从初始状态（$x,\ v$）出发时为（$x_t(x,\ v)\ v_t(x,\ v)$）。那么，哈密顿动力学满足以下时间反演性质：对于 $t \geq 0$,

$$\mathsf{HF}_t(x_t(x, v), -v_t(x, v)) = (x, -v)\,.$$

- **平稳分布**。我们希望证明，目标分布 ν 是 eHMC 核 $\mathbf{P}_T$ 的平稳分布，即 $\nu \mathbf{P}_T = \nu$。证明将用到前面提到的哈密顿动力学性质。

  令 $\nu_{\mathsf{P}}(x,\ v) \propto e^{-H(x,v)} \in \mathcal{P}(\mathbb{R}^{2d})$ 表示定义在相空间 $\mathbb{R}^{2d}$ 上的联合目标分布，因此，$\nu_{\mathsf{P}} = \nu \otimes \gamma_d$。证明 ν_{P} 在相空间上经过一步精确哈密顿蒙特卡洛后保持不变，即可说明目标分布 ν 是位置空间上 eHMC 核$\mathbf{P}_T$ 的平稳分布。回顾一下，精确哈密顿蒙特卡洛由以下步骤组成：(i) 从 γ_d 中抽取新的速度（独立于位置）；(ii) 沿哈密顿流演化。步骤 (i) 不影响位置边缘分布，并从 γ_d 中抽取速度，该分布是速度边缘分布

ν_{P}的，因此 ν_{P} 在步骤 (i) 下保持不变。故只需检验 ν_{P} 在哈密顿流下保持不变，即 $(\mathsf{HF}_t)_{\#}\nu_{\mathsf{P}} = \nu_{\mathsf{P}}$。

令 $(x, v) \in \mathbb{R}^{2d}$ 并定义 $(x_t, v_t) \equiv (x_t(x, v)\ v_t(x, v)) = \mathsf{HF}_t(x, v)$。令 $\tilde{\nu}_{\mathsf{P}} := (\mathsf{HF}_t)_{\#}\nu_{\mathsf{P}}$。那么，根据换元公式，我们有

$$\nu_{\mathsf{P}}(x, v) = \tilde{\nu}_{\mathsf{P}}(\mathsf{HF}_t(x, v)) \left| \det(\nabla \mathsf{HF}_t(x, v)) \right| = \tilde{\nu}_{\mathsf{P}}(\mathsf{HF}_t(x, v)),$$

最后一个等号成立是因为哈密顿流保持体积。因此可得

$$\tilde{\nu}_{\mathsf{P}}(\mathsf{HF}_t(x, v)) = \nu_{\mathsf{P}}(x, v) = Z^{-1} e^{-H(x,v)} = Z^{-1} e^{-H(\mathsf{HF}_t(x,v))}$$

最后一个等号成立是因为哈密顿动力学过程中哈密顿量守恒，而 Z 是某个归一化常数。因此，我们得出结论 $\tilde{\nu}_{\mathsf{P}} = \nu_{\mathsf{P}}$。

采用速度 Verlet 积分器的未校正的 HMC 也具有辛性和时间反演性质；然而，它并不守恒哈密顿量，因此会导致有偏的平稳分布。

B 概率核的正则化性质

设 $\mathbf{P}$ 是一个作用于分布 $\mu \in \mathcal{P}(\mathbb{R}^d)$ 的单步概率核，其作用方式为

$$(\mu\mathbf{P})(x) = \int \mu(z)\mathbf{P}(z, x)\, dz$$

并作用于实值函数或观测量 $f\colon\ \mathbb{R}^d \to \mathbb{R}$ 通过

$$(\mathbf{P}f)(x) = \int f(z)\mathbf{P}(x, z)\, dz\,.$$

这样，转移核既作用于分布，也作用于函数，分别输出分布和函数。

在本节中，我们使用 f，g 表示实值函数，使用 μ，π 表示概率分布。我们还将分布 μ 下函数 f 的期望记作

$$\mu(f) = \int f(x)\mu(x)\, dx\,.$$

根据上述定义和 Fubini 定理，

$$(\mu\mathbf{P})(f) = \int f(z)\Big(\int \mu(x)\mathbf{P}(x, z)\, dx\Big)\, dz = \int \mu(x)\Big(\int f(z)\mathbf{P}(x, z)\, dz\Big)\, dx = \int \mu(x)(\mathbf{P}f)(x)\, dx = \mu(\mathbf{P}f)\,,$$

因此，对于转移核 $\mathbf{P}$、概率分布 μ 和函数 f，

$$(\mu\mathbf{P})(f) = \mu(\mathbf{P}f)\,. \tag{15}$$

B.1 从 Wasserstein–1 到全变差

在本节中，我们研究1，其中D = 全变差，并且 $W = W_1$。 与 π 之间的全变差（*TV*）距离定义为

$$\mathsf{TV}(\mu,\pi) = \frac{1}{2}\int |\mu(x) - \pi(x)|\, dx = \frac{1}{2}\sup_{\|f\|_\infty \le 1} \Big| \int f(x)(\mu(x) - \pi(x))\, dx \Big|$$

其中 $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$。令

$$\|f\|_{\mathsf{Lip}} = \sup_{x\neq y} \frac{|f(x) - f(y)|}{\|x - y\|}\,.$$

现在假设 1) 对 D =全变差 和 $W = W_1$成立，即

$$\mathsf{TV}(\mu\mathbf{P}, \pi\mathbf{P}) \le C\, W_1(\mu,\pi) \tag{16}$$

对核 $\mathbf{P}$ 以及任意分布 μ,π成立。我们在下面的引理中证明正则化 (16) 的一个推论。

引理 20。 假设 (16) 对马尔可夫核 $\mathbf{P}$ 以及任意两个分布 $\mu,\pi \in \mathcal{P}(\mathbb{R}^d)$成立。令 f 为满足 $\|f\|_\infty < \infty$的实值函数。则

$$\|\mathbf{P}f\|_{\mathsf{Lip}} \le 2C\, \|f\|_\infty\,. \tag{17}$$

证明。 根据全变差距离的定义，

$$\begin{aligned}
\mathsf{TV}(\mu\mathbf{P}, \pi\mathbf{P}) &= \frac{1}{2}\sup_{\|g\|_\infty\le 1} \Big| \int g(x)((\mu\mathbf{P})(x) - (\pi\mathbf{P})(x))\, dx \Big| \\
&= \frac{1}{2}\sup_{\|g\|_\infty\le 1} \big|\mu(\mathbf{P}g) - \pi(\mathbf{P}g)\big| \\
&\ge \frac{1}{2}\big|\mu(\mathbf{P}g) - \pi(\mathbf{P}g)\big|
\end{aligned}$$

其中第二个等号由（15）成立，因此最终结论为

$$\big|\mu(\mathbf{P}g) - \pi(\mathbf{P}g)\big| \le 2\mathsf{TV}(\mu\mathbf{P}, \pi\mathbf{P})\,,$$

对于任意有界函数 g 满足 $\|g\|_\infty \le 1$。利用（16），我们得到

$$\big|\mu(\mathbf{P}g) - \pi(\mathbf{P}g)\big| \le 2C\, W_1(\mu,\pi)$$

对于任意 $\mu,\ \pi \in \mathcal{P}(\mathbb{R}^d)$。令 $x,\ y \in \mathbb{R}^d$为满足 $x \neq y$的任意对象。取 $\mu = \delta_x$和 $\pi = \delta_y$，可推出 $W_1(\delta_x,\ \delta_y) = \|x - y\|$，因此，

$$\big|(\mathbf{P}g)(x) - (\mathbf{P}g)(y)\big| \le 2C\, \|x - y\|\,.$$

因此，对于任意满足 $\|g\|_\infty < 1$ 的 g，我们得出

$$\|\mathbf{P}g\|_{\mathsf{Lip}} \le 2C\,.$$

对于满足 $\|f\|_\infty < \infty$的函数 f，上述论证适用于 $g = \frac{f}{\|f\|_\infty}$，因此我们可以得出，对于所有有界函数 f,

$$\|\mathbf{P}f\|_{\mathsf{Lip}} \le 2C\, \|f\|_\infty\,.$$

证明至此完成。 □

因此我们可以看出，在条件（16）下，马尔可夫核会将有界函数正则化为 Lipschitz 函数。反过来同样成立，即我们可以由（16）推出（17），但这里不作具体说明。

B.2 Wasserstein–2 到 KL 散度

考虑 (1)，其中 D = KL散度和 $W = W_2^2$，即假设存在常数 $C > 0$ 使得 t

$$\mathsf{KL}(\mu\mathbf{P} \,\|\, \pi\mathbf{P}) \le C\, W_2^2(\mu, \pi) \tag{18}$$

对所有 μ，$\pi \in \mathcal{P}(\mathbb{R}^d)$ 均成立。KL 散度的 Donsker–Varadhan变分表示指出，对所有 μ, $\pi \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathsf{KL}(\pi_1 \,\|\, \pi_2) = \sup_{f:\mathbb{R}^d\to\mathbb{R}} \left(\mathop{\mathbb{E}}_{\pi_1}[f] - \log \mathop{\mathbb{E}}_{\pi_2}[e^f] \right), \tag{19}$$

其中，上确界取遍所有可测函数。因此，将 (19) 应用于 (18)，并结合 (15)，可得对所有 f,

$$\mu(\mathbf{P}f) - \log \pi(\mathbf{P}e^f) \le C\, W_2^2(\mu, \pi)\,.$$

特别地，对函数 g 取 $\mu = \delta_x$、$\pi = \delta_y$ 和 $f = \log g$，可得

$$(\mathbf{P}\log g)(x) \le \log(\mathbf{P}g)(y) + C\,\|x - y\|^2\,. \tag{20}$$

对于 $x = y$，(20) 退化为凹函数 log 的詹森不等式。对于 $x \neq y$，(20) 是一个对数Harnack不等式 [Wan97]，它用观测量刻画一步熵平滑 (18)；关于正则化 (18) 与 Harnack不等式 (20) 之间的对偶性，以及 R´enyi 散度的类似结论，详见 [AC24b] 。

C 正则化与单次耦合

引理的证明 *2*。 根据 uHMC 转移核的定义，我们有如下结论

$$\begin{aligned}
\mathsf{D}(\delta_x\tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y\tilde{\mathbf{P}}_{T,h}) &\overset{(6)}{=} \mathsf{D}(\mathcal{L}(\tilde{q}_{T,h}(x,\xi)) \,\|\, \mathcal{L}(\tilde{q}_{T,h}(y,\xi))) \\
&\overset{(7)}{=} \mathsf{D}(\mathcal{L}(\tilde{q}_{T,h}(y,\varphi_{x,y}(\xi))) \,\|\, \mathcal{L}(\tilde{q}_{T,h}(y,\xi))) \\
&\le \mathsf{D}(\mathcal{L}(\tilde{\mathsf{HF}}_{T,h}(y,\varphi_{x,y}(\xi))) \,\|\, \mathcal{L}(\tilde{\mathsf{HF}}_{T,h}(y,\xi))) \\
&= \mathsf{D}(\mathcal{L}(y,\varphi_{x,y}(\xi)) \,\|\, \mathcal{L}(y,\xi)) \\
&= \mathsf{D}(\mathcal{L}(\varphi_{x,y}(\xi)) \,\|\, \mathcal{L}(\xi)) \\
&= \mathsf{D}((\varphi_{x,y})_{\#}\gamma_d \,\|\, \gamma_d)
\end{aligned}$$

其中，不等式成立是因为投影不会增大散度，第三个等号成立是˜因为$\mathsf{HF}_{T,\,h}$ 是一个双射映射。陈述中的假设保证映射（7）的存在性和唯一性，这归因于 [BRE23，推论 14]。

□

引理的证明 *3*。 根据 uHMC 和 eHMC 转移核的定义，我们有如下结果

$$\begin{aligned}
\mathsf{D}(\delta_x\tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y\mathbf{P}_T) &\overset{(5),(6)}{=} \mathsf{D}(\mathcal{L}(\tilde{q}_{T,h}(x,\xi)) \,\|\, \mathcal{L}(q_T(y,\xi))) \\
&\overset{(10)}{=} \mathsf{D}(\mathcal{L}(q_T(y,\Phi_{x,y}(\xi))) \,\|\, \mathcal{L}(q_T(y,\xi))) \\
&\le \mathsf{D}(\mathcal{L}(\mathsf{HF}_T(y,\Phi_{x,y}(\xi))) \,\|\, \mathcal{L}(\mathsf{HF}_T(y,\xi))) \\
&= \mathsf{D}(\mathcal{L}(y,\Phi_{x,y}(\xi)) \,\|\, \mathcal{L}(y,\xi)) \\
&= \mathsf{D}(\mathcal{L}(\Phi_{x,y}(\xi)) \,\|\, \mathcal{L}(\xi)) \\
&= \mathsf{D}((\Phi_{x,y})_{\#}\gamma_d \,\|\, \gamma_d)
\end{aligned}$$

其中，不等式成立是因为投影不会增加散度，第三个等号成立是因为 HF_T 是一个双射映射。陈述中的假设确保该映射的存在性和唯一性（10），这由引理 21得出。

□

D 混合时间的一次耦合映射

引理的证明 *4*。引理 4 中陈述的假设满足 [BRE23的假设，引理 25]因此 [BRE23的引理 25] 推出（8）。 □

引理的证明 *5*。引理 5 中陈述的假设满足 [BRE23的假设，引理 26]。因此，$\frac{11}{2}MT^2\|x-y\|$（9）中的上界由 [BRE23的引理 26]得出。还需得到（全局）上界 $_{2\ 9}$，如（9）中所述。由于引理 5 的假设蕴含 [BRE23的引理 26], 的全局分析指出： [BRE23的引理 26] 指出

$$\|\nabla\varphi_{x,y}(v) - I\|_{\mathsf{op}} \leq \frac{(12/5)LT^2}{1-(6/5)LT^2}.$$

利用 $LT^2 \leq \frac{1}{12}$，可得

$$\|\nabla\varphi_{x,y}(v) - I\|_{\mathsf{op}} \leq \frac{2}{9}.$$

□

E 渐近偏差的一次耦合映射

回顾用于获得渐近偏差界的一次映射（10）。对于任意 $x,\ y \in \mathbb{R}^d$，它是一个映射 $\Phi_{x,\ y}$：$\mathbb{R}^d \to \mathbb{R}^d$ 使得对于任意 $v \in \mathbb{R}^d$

$$\tilde{q}_{T,h}(x,v) = q_T(y, \Phi_{x,y}(v))\,.$$

现在说明如何通过组合用于界定混合时间的一次映射（7）与另一个映射 ϕ_x来得到这样的映射，其定义如下。对于任意 $x \in \mathbb{R}^d$，令 ϕ_x： $\mathbb{R}^d \to \mathbb{R}^d$ 为一个映射，使得对于任意 $v \in \mathbb{R}^d$

$$\tilde{q}_{T,h}(x,v) = q_T(x, \phi_x(v))\,. \tag{21}$$

在势函数满足 ϕ_x（21）的条件下， L-光滑性以及 $LT^2 \leq \frac{2}{5}\pi^2$，根据 [BRE23，推论 14], 存在满足（21）的

给定（7）和（21），注意以下几点

$$\tilde{q}_{T,h}(x,v) \overset{(21)}{=} q_T(x,\phi_x(v)) \overset{(7)}{=} q_T(y, \varphi_{x,y}(\phi_x(v))) = q_T(y,\Phi_{x,y}(v)) \tag{22}$$

因此，满足（10）的映射 $\Phi_{x,\ y}$ 是 $\Phi_{x,\ y} = \varphi_{x,\ y} \circ \phi_x$。注意，在等式序列中使用（7）时，我们将其用于 $h = 0$。

引理 21. 对于任意 $x,\ y \in \mathbb{R}^d$，存在映射 $\Phi_{x,\ y}$： $\mathbb{R}^d \to \mathbb{R}^d$ 满足 (10) 当势函数满足假设 *1(b)* 且 $LT^2 \leq \frac{2}{5}\pi^2$ 时。 –

证明. 由 (22) 可知，对于任意 x， $y \in \mathbb{R}^d$，满足以下条件的映射 $\Phi_{x,\ y}$

$$\tilde{q}_{T,h}(x,v) = q_T(y, \Phi_{x,y}(v))$$

是 $\Phi_{x,\ y} = \varphi_{x,\ y} \circ \phi_x$ 其中 $\varphi_{x,\ y}$ 满足（7），且 ϕ_x 满足（21）。在 $\varphi_{x,\ y}$ 和 ϕ_x存在的条件是 f为 L–光滑，且 $LT^2 \leq \frac{2}{5}\pi^2$根据 [BRE23，推论 14]，由此可推出存在映射 $\Phi_{x,\ y}$ 在相同条件下满足（10）。 □

在推导映射 $\Phi_{x,\ y}$的正则性界之前，请注意，映射 $\varphi_{x,\ y}$ 的逐点正则性估计和雅可比正则性估计分别见引理 4 和 5 。接下来，我们陈述满足（ ϕ_x （21）的映射 [BRE23]中的类似界。

引理 22。 假设条件 *1(a)-(c)* 成立。进一步假设 $h \geq 0$且 $L\left(T^2 + Th\right) \leq \frac{1}{6}$。那么，对任意 x，– $v \in \mathbb{R}^d$

$$\|\phi_x(v) - v\| \leq 2h^2(LT^{-1}\|x\| + L\|v\| + MT^{-1}\|x\|^2 + MT\|v\|^2)\,.$$

引理 23。 假设条件 *1(a)-(d)* 成立。进一步假设 $h \geq 0$且 $L\left(T^2 + Th\right) \leq \frac{1}{6}$。那么，对任意 x，– $v \in \mathbb{R}^d$

$$\|\nabla\phi_x(v) - I\|_{\mathsf{op}} \leq \min\left\{\frac{1}{2}, 2h^2(L + M\|x\| + MT\|v\| + (M^2T^2 + N)\|x\|^2 + (M^2T^4 + NT^2)\|v\|^2)\right\}\,.$$

引理22 和23分别对应于 [BRE23]中的引理27和引理28，但常数更为简化。现在，我们可以得到映射 $\Phi_{x,y}$ 满足（10）时的逐点估计和雅可比矩阵估计。

引理 *6* 的证明。任取 $v \in \mathbb{R}^d$。注意到

$$\|\Phi_{x,y}(v) - v\| = \|\varphi_{x,y}(\phi_x(v)) - \phi_x(v) + \phi_x(v) - v\| \leq \|\varphi_{x,y}(\phi_x(v)) - \phi_x(v)\| + \|\phi_x(v) - v\|\,.$$

结论现在由引理 4 和22 得出。 □

引理 *7* 的证明。任取 $v \in \mathbb{R}^d$。我们知道

$$\nabla\Phi_{x,y}(v) = \underbrace{\nabla\varphi_{x,y}(\phi_x(v))}_{A} \cdot \underbrace{\nabla\phi_x(v)}_{B}\,.$$

因此

$$\begin{aligned}
\|\nabla\Phi_{x,y}(v) - I\|_{\mathsf{op}} &= \|AB - I\|_{\mathsf{op}} \\
&= \|A(B-I) + (A-I)\|_{\mathsf{op}} \\
&\leq \|A\|_{\mathsf{op}}\|B-I\|_{\mathsf{op}} + \|A-I\|_{\mathsf{op}} \\
&\leq (1 + \|A-I\|_{\mathsf{op}})\|B-I\|_{\mathsf{op}} + \|A-I\|_{\mathsf{op}} \\
&= \|A-I\|_{\mathsf{op}} + \|B-I\|_{\mathsf{op}} + \|A-I\|_{\mathsf{op}}\|B-I\|_{\mathsf{op}}\,.
\end{aligned}$$

现在我们使用引理 5 和 23 ，因为它们分别给出了 $\|A-I\|_{\mathsf{op}}$ 和 $\|B-I\|_{\mathsf{op}}$ 的界。定义

$$Q(x,v) \coloneqq L + M\|x\| + MT\|v\| + (M^2T^2 + N)\|x\|^2 + (M^2T^4 + NT^2)\|v\|^2\,.$$

于是我们有 $\|\nabla\Phi_{x,\ y}(v) - I\|_{\mathsf{op}} \leq$

$$\min\left\{\frac{2}{9}, \frac{11}{2}MT^2\|x-y\|\right\} + \min\left\{\frac{1}{2}, 2h^2Q(x,v)\right\} + \min\left\{\frac{1}{9}, \frac{4}{9}h^2Q(x,v), \frac{11}{4}MT^2\|x-y\|\right\}.$$

因此，我们有

$$\|\nabla\Phi_{x,y}(v) - I\|_{\mathsf{op}} \leq \min\left\{\frac{15}{18}, \frac{11}{2}MT^2\|x-y\| + \frac{22}{9}h^2Q(x,v)\right\}.$$

□

E.1 一阶正则性

回顾式（22），满足式（10）的映射 $\Phi_{x,y}$ 如何由分别满足式（7）和式（21）的映射 $\varphi_{x,y}$ 与 ϕ_x 复合而成。因此，引理 22 和引理 23 中关于 v 的二次依赖会传递到引理 6 和引理 7 中针对 $\Phi_{x,y}$ 的正则性估计上。这使得我们得到了 uHMC 在 KL 散度下渐近偏差的尖锐界（定理 2 ）。然而，将证明推广到 R´enyi 散度颇具挑战，因为要得到引理 13 的对应结果，需要计算 $\mathbb{E}_{v\sim\gamma_d}[e^{\lambda\|v\|^4}]$，而该量对任意 $\lambda > 0$ 都会发散。

引理 22 和引理 23 中关于 v 的二次依赖源于二阶梯形近似 [BRE23，引理 22]；因此，为了在映射 ϕ_x 的正则性估计中减弱对 v 的依赖，我们考虑一阶梯形近似，并在引理 24 和引理 25 中给出相应的估计。随后，我们将这些结果推广到引理 8 和引理 9 中的映射 $\Phi_{x,y}$ 。由此，我们得到了定理 4 中 uHMC 在 R´enyi 散度下的渐近偏差界。证明引理 24、25、8 和 9 所需的 Verlet 格式一阶误差界见 E.2 节。

引理 24. 假设条件 *1(a)-(b)* 成立。进一步假设 $h \geq 0$ 并且 $L(T^2+Th) \leq \frac{1}{6}$。那么对于任意 $x,\ v \in \mathbb{R}^d$

$$\|\phi_x(v) - v\| \leq \frac{7}{5}h\left[\frac{1}{5}T^{-1}\|v\| + \frac{7}{36}L\|x\|\right].$$

证明。 根据 [BRE23，引理 27], 可得

$$\|\phi_x(v) - v\| \leq \frac{7}{6T}\max_{s\leq T}\|\tilde{q}_{s,h}(x,v) - q_s(x,v)\|.$$

因此，我们的引理27 推出了所述界。 □

引理 25. 假设条件 *1(a)-(c)* 成立。进一步假设 $h \geq 0$ 并且 $L(T^2+Th) \leq \frac{1}{6}$。那么对于任意 $x,\ v \in \mathbb{R}^d$

$$\|\nabla\phi_x(v) - I\|_{\mathsf{op}} \leq \min\left\{\frac{1}{2}, \frac{2}{15}h\left(2T^{-1} + \frac{16}{5}MT\|x\| + 20MT^2\|v\|\right)\right\}.$$

证明。 全局界 1 2 的推导方式与 [BRE23，引理 28]相同。要得到局部界， [BRE23，引理 28]指出

$$\|\nabla\phi_x(v) - I\|_{\mathsf{op}} \leq \frac{10}{9T}\|\nabla_2\tilde{q}_{T,h}(x,v) - \nabla_2 q_T(x,v)\|_{\mathsf{op}} + \frac{14}{15}MT^3\|\phi_x(v) - v\|.$$

代入引理 24和28 中的结果并化简，可得

$$\|\nabla\phi_x(v) - I\|_{\mathsf{op}} \le \frac{2}{15}h\left(2T^{-1} + \frac{16}{5}MT\|x\| + 20MT^2\|v\|\right).$$

□

现在我们可以证明引理8 和9。

引理*8*的证明。证明过程与引理6的证明类似。任取 $v \in \mathbb{R}^d$。注意到

$$\|\Phi_{x\ y}{}^{(v)} - v\| = \|\varphi_{x,y}(\phi_x(v)) - \phi_x(v) + \phi_x(v) - v\| \le \|\varphi_{x,y}(\phi_x(v)) - \phi_x(v)\| + \|\phi_x(v) - v\|.$$

结论现在由引理4和24推出。

□

引理*9*的证明。证明过程与引理7的证明类似。任取 $v \in \mathbb{R}^d$。我们知道

$$\|\nabla\Phi_{x,y}(v) - I\|_{\mathsf{op}} \le \|A - I\|_{\mathsf{op}} + \|B - I\|_{\mathsf{op}} + \|A - I\|_{\mathsf{op}}\|B - I\|_{\mathsf{op}}$$

其中 $A := \nabla\varphi_{x,\ y}(\phi_x(v)),\ B := \nabla\phi_x(v).$

现在我们使用引理 5和 25，因为它们分别给出了 $\|A-I\|_{\mathsf{op}}$和 $\|B-I\|_{\mathsf{op}}$的界。定义

$$Q(x,v) := 2T^{-1} + \frac{16}{5}MT\|x\| + 20MT^2\|v\|.$$

于是有 $\|\nabla\Phi_{x,\ y}(v) - I\|_{\mathsf{op}} \le$

$$\min\left\{\frac{2}{9}, \frac{11}{2}MT^2\|x-y\|\right\} + \min\left\{\frac{1}{2}, \frac{2}{15}h\,Q(x,v)\right\} + \min\left\{\frac{1}{9}, \frac{4}{135}h\,Q(x,v), \frac{11}{4}MT^2\|x-y\|\right\}.$$

因此有

$$\|\nabla\Phi_{x,y}(v) - I\|_{\mathsf{op}} \le \min\left\{\frac{15}{18}, \frac{11}{2}MT^2\|x-y\| + \frac{22}{135}h\,Q(x,v)\right\}.$$

□

E.2 Verlet 格式的一阶误差界

˜回顾第 2.2 节，任意满足 t_1 可整除 t_2 的情形下，$\tilde{q}_{t_1,\ t_2}$ 定义为 $\tilde{q}_{t_1,\ t_2} := \Pi_1 \circ \mathrm{HF}_{t_1,\ t_2\circ}$ ˜ 类似地，令 $\tilde{v}_{t_1,\ t_2} := \Pi_2 \circ \mathrm{HF}_{t_1,\ t_2\circ}$ 当 $t_2 = 0$ 时，我们将 $\tilde{v}_{t_1,\ 0}$ 记作 $v_{t_1\circ}$ 进一步定义

$$\lfloor t\rfloor_h := \max\{s \in h\mathbb{Z} \ :\ s \le t\} \quad \text{and} \quad \lceil t\rceil_h := \min\{s \in h\mathbb{Z} \ :\ s \ge t\} \quad \text{for } h > 0.$$

以下引理是 [BRE23 引理 22] 的一阶类比结果。

引理 26。 固定 $T > 0$，令 $u:\ [0,T] \to \mathbb{R}$ 为满足以下条件的可微函数：最大值$_{s\in[0,T]}\,|u'(s)| \le B_1$。那么，对于任意满足 h 整除 T的 $h > 0$，有

$$\left\|\int_0^T (T-s)u(s)\,ds - \frac{1}{2}\int_0^T (T-s)[u(\lfloor s\rfloor_h) + u(\lceil s\rceil_h)]\,ds\right\| \le \frac{B_1 hT^2}{2}.$$

证明。令 $t_k = kh$ 以及 $t_{k+1} = (k+1)h$，并考虑区间 $[t_k, t_{k+1}]$。根据 [BRE23，引理 22],，该区间上的误差为

$$\epsilon_k = \left\| \int_{t_k}^{t_{k+1}} \left[\frac{s^2}{2} - Ts - \alpha_k \right] u'(s)\, ds \right\|$$

其中

$$\alpha_k = \frac{1}{4}(t_{k+1}^2 - 2Tt_{k+1} - 2Tt_k + t_k^2)\,.$$

于是我们有

$$\epsilon_k \le B_1 \int_{t_k}^{t_{k+1}} \left| \frac{s^2}{2} - Ts - \alpha_k \right| ds\,.$$

我们计划用积分的最大值乘以宽度 h来界定该积分。令

$$F(s) = \frac{s^2}{2} - Ts - \alpha_k$$

并注意到 F在 $[t_k, t_{k+1}]$上是凸且递减的，因为当 $s \in [t_k, t_{k+1}]$时，$F'(s) = s - T \le 0$。因此，F在区间 $[t_k, t_{k+1}]$ 上的最大值将在其某个端点处取得。我们看到

$$F(t_k) = \frac{h}{2}\left(T - \frac{(t_k + t_{k+1})}{2} \right)$$

and

$$F(t_{k+1}) = -\frac{h}{2}\left(T - \frac{(t_k + t_{k+1})}{2} \right).$$

因此

$$\epsilon_k \le B_1 h \left[\frac{h}{2}\left(T - \frac{(t_k + t_{k+1})}{2} \right) \right] \le \frac{B_1 h^2 T}{2}\,.$$

为得到总误差，我们乘以 $\frac{T}{h}$，从而得到

$$\left\| \int_0^T (T-s)u(s)\, ds - \frac{1}{2}\int_0^T (T-s)[u(\lfloor s \rfloor_h) + u(\lceil s \rceil_h)]\, ds \right\| \le \frac{B_1 h T^2}{2}\,.$$

□

以下引理是 [BRE23，引理 23]的对应版本。

引理 27。假设*1 (a)-(b)*成立。此外，假设 $h \ge 0$ 并且 $L(T^2 + Th) \le \frac{1}{6}$。对于任意 x，$v \in \mathbb{R}^d$ –

$$\max_{s \le T} \|q_s(x,v) - \tilde{q}_{s,h}(x,v)\| \le h\left(\frac{6}{25}\|v\| + \frac{7}{30} LT\|x\| \right).$$

证明。我们首先利用 [BRE23，引理 20]，结合 $h = 0$ 并在假设 $LT^2 \le \frac{1}{6}$ 下得到 [BRE23，（57）]

$$\max_{s \le T} \|v_s(x,v)\| \le \frac{7}{6} LT\|x\| + \frac{6}{5}\|v\|\,. \tag{23}$$

考虑简写记号 q_t 和 $\tilde{q}_t$ 分别表示 $q_t(x,\ v)$ 以及 $\tilde{q}_{t,\ h}(x,\ v)$。根据 [BRE23，引理 23],可得

$$
\begin{aligned}
q_T - \tilde{q}_T \ &= \ \mathrm{I} + \mathrm{II} + \mathrm{III} \ \text{ where} \\
\mathrm{I} \ &:= \ \frac{1}{2}\int_0^T (T-s)[\nabla f(\tilde{q}_{\lfloor s\rfloor_h}) - \nabla f(q_{\lfloor s\rfloor_h}) + \nabla f(\tilde{q}_{\lceil s\rceil_h}) - \nabla f(q_{\lceil s\rceil_h})]\,ds \\
&\quad - \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[\nabla f(\tilde{q}_{\lceil s\rceil_h}) - \nabla f(q_{\lceil s\rceil_h}) - (\nabla f(\tilde{q}_{\lfloor s\rfloor_h}) - \nabla f(q_{\lfloor s\rfloor_h}))]\,ds \\
\mathrm{II} \ &:= \ -\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[\nabla f(q_{\lceil s\rceil_h}) - \nabla f(q_{\lfloor s\rfloor_h})]\,ds \\
\mathrm{III} \ &:= \ -\int_0^T (T-s)\nabla f(q_s)\,ds + \frac{1}{2}\int_0^T (T-s)[\nabla f(q_{\lfloor s\rfloor_h}) + \nabla f(q_{\lceil s\rceil_h})\,ds\,.
\end{aligned}
$$

为了对 $\|$进行界定，我们有 [BRE23中的（60）式]

$$
|\mathrm{I}| \le LT^2 \max_{s\le T}\|q_s - \tilde{q}_s\| \le \ \frac{1}{6}\max_{s\le T}\|q_s - \tilde{q}_s\|
$$

而 $|\mathrm{II}|$ 在 [BRE23中的（61）式]给出了其上界

$$
|\mathrm{II}| \le Lh^2\left(\frac{7}{72}\|x\| + \frac{3}{5}T\|v\|\right).
$$

接下来，我们利用引理 26 及 $u(s) = -\nabla f(q_s(x,v))$，其中

$$
|u'(s)| \le L\|v_s\| \overset{(23)}{\le} \frac{7}{6}L^2T\|x\| + \frac{6}{5}L\|v\| =: B_1\,.
$$

因此，再利用 $LT^2 \le \frac{1}{6}$，我们得到

$$
|\mathrm{III}| \le \frac{h}{12}\left[\frac{6}{5}\|v\| + \frac{7}{6}LT\|x\|\right].
$$

因此

$$
\max_{s\le T}\|q_s(x,v) - \tilde{q}_{s,h}(x,v)\| \le \frac{6}{5}\left[|\mathrm{II}| + |\mathrm{III}|\right]
$$

化简后得到

$$
\max_{s\le T}\|q_s(x,v) - \tilde{q}_{s,h}(x,v)\| \le \frac{6}{5}h\left[\left(LhT\frac{3}{5} + \frac{1}{10}\right)\|v\| + \left(Lh\frac{7}{72} + LT\frac{7}{72}\right)\|x\|\right].
$$

对 $h \le T$ 取上界并使用 $LTh \le \frac{1}{6}$，化简后可得

$$
\max_{s\le T}\|q_s(x,v) - \tilde{q}_{s,h}(x,v)\| \le h\left(\frac{6}{25}\|v\| + \frac{7}{30}LT\|x\|\right).
$$

□

下面的引理是 [BRE23，引理 24]的对应版本。

引理 28。 假设条件 *1(a)-(c)* 成立。进一步假设 $h \geq 0$ 且 $L(T^2+Th) \leq \frac{1}{6}$。对于任意 $x,\ v \in \mathbb{R}^d$，

$$\max_{s\leq T} \|\nabla_2 q_s(x,v) - \nabla_2 \tilde{q}_{s,h}(x,v)\|_{\mathsf{op}} \leq \frac{6}{25} h \left(1 + \frac{7}{5} MT^2\|x\| + \frac{216}{25} MT^3 \|v\|\right).$$

证明。将 $q_t(x,v)$ 简记为 q_t，将 $\tilde{q}_{t,h}(x,v)$ 简记为 $\tilde{q}_t$。根据 [BRE23]，可将 $\nabla_2 q_T - \nabla_2 \tilde{q}_T = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}$ 写出，其中 I、II、III、IV 的定义见 [BRE23, (64)]。

BRE [23] 方程 (65)、(66) 和 (67) 给出了 $\|\mathrm{I}\|_{\mathsf{op}}$、$\|\mathrm{II}\|_{\mathsf{op}}$ 和 $\|\mathrm{III}\|_{\mathsf{op}}$ 的界，分别如下

$$\begin{aligned}
\|\mathrm{I}\|_{\mathsf{op}} &\leq \frac{1}{6} \max_{s\leq T} \|\nabla_2 q_s - \nabla_2 \tilde{q}_s\|_{\mathsf{op}} \\
\|\mathrm{II}\|_{\mathsf{op}} &\leq \frac{6}{5} MT^3 \max_{s\leq T} \|q_s - \tilde{q}_s\| \\
\|\mathrm{III}\|_{\mathsf{op}} &\leq \frac{3}{5} h^2 LT + h^2 M \left(\frac{7}{60} T\|x\| + \frac{18}{25} T^2 \|v\|\right).
\end{aligned}$$

为求 $\|\mathrm{IV}\|_{\mathsf{op}}$ 的界，我们应用引理 26，其中 $u(s) = -\nabla^2 f(q_s)\nabla_2 q_s$ 满足

$$\|u'(s)\|_{\mathsf{op}} \leq \frac{6}{5}\left(M\left(\frac{7}{36}\|x\| + \frac{6}{5}T\|v\|\right) + L\right) =: B_1 .$$

因此

$$\|\mathrm{IV}\|_{\mathsf{op}} \leq \frac{3hT^2}{5}\left(M\left(\frac{7}{36}\|x\| + \frac{6}{5}T\|v\|\right) + L\right).$$

因此我们得到

$$\max_{s\leq T} \|\nabla_2 q_s(x,v) - \nabla_2 \tilde{q}_{s,h}(x,v)\|_{\mathsf{op}} \leq \frac{6}{5}\left[\|\mathrm{II}\|_{\mathsf{op}} + \|\mathrm{III}\|_{\mathsf{op}} + \|\mathrm{IV}\|_{\mathsf{op}}\right].$$

利用引理 27 对 $\|\mathrm{II}\|_{\mathsf{op}}$ 求界并化简，得到

$$\begin{aligned}
\max_{s\leq T} \|\nabla_2 q_s(x,v) - \nabla_2 \tilde{q}_{s,h}(x,v)\|_{\mathsf{op}} \leq \frac{6}{25} hT \Bigg[& 3hL + 3LT + \|x\|\Big(\frac{7}{10} LMT^2(h+T) + \frac{7}{12} hM + \frac{7}{12} TM\Big) \\
& + \|v\|\Big(\frac{36}{5} MT^2\big(\frac{3}{5} LhT + \frac{1}{10}\big) + \frac{18}{5} hMT + \frac{18}{5} MT^2\Big)\Bigg].
\end{aligned}$$

对 $h \leq T$ 取上界，并使用 $L(T^2+Th) \leq \frac{1}{6}$，我们得到

$$\max_{s\leq T} \|\nabla_2 q_s(x,v) - \nabla_2 \tilde{q}_{s,h}(x,v)\|_{\mathsf{op}} \leq \frac{6}{25} h \left(1 + \frac{7}{5} MT^2\|x\| + \frac{216}{25} MT^3 \|v\|\right).$$

□

F KL 散度下的混合时间

F.1 引理 10 的证明

证明。 根据变量替换公式，对于任意 $v \in \mathbb{R}^d$

$$((\psi^{-1})_{\#}\boldsymbol{\gamma}_d)(v) = \boldsymbol{\gamma}_d(\psi(v)) \det \nabla\psi(v)$$

其中 $\det \nabla\psi(v) > 0$ 因为 $\|\nabla\psi(v) - I\|_{\mathsf{op}} < 1$。利用 KL 散度在同时进行双射变换时的不变性 [PW24，推论 2。18], 我们对 $\mathsf{KL}(\boldsymbol{\gamma}_d \| (\psi^{-1})_{\#}\boldsymbol{\gamma}_d)$ 进行界定，而非对 $\mathsf{KL}(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d)$。我们有

$$\begin{aligned}
\mathsf{KL}(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) &= \mathsf{KL}(\boldsymbol{\gamma}_d \,\|\, (\psi^{-1})_{\#}\boldsymbol{\gamma}_d) \\
&= \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\Big[\log \frac{\boldsymbol{\gamma}_d(v)}{((\psi^{-1})_{\#}\boldsymbol{\gamma}_d)(v)}\Big] \\
&= \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\Big[\log \frac{\boldsymbol{\gamma}_d(v)}{\boldsymbol{\gamma}_d(\psi(v))|\det \nabla\psi(v)|}\Big] \\
&= \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\Big[-\frac{1}{2}\|v\|^2 + \frac{1}{2}\|\psi(v)\|^2 - \log|\det \nabla\psi(v)|\Big] \\
&= \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\Big[\frac{1}{2}\|\psi(v) - v\|^2 + \langle \psi(v) - v, v\rangle - \log|\det \nabla\psi(v)|\Big] \\
&= \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\Big[\frac{1}{2}\|\psi(v) - v\|^2 + \mathsf{Tr}(\nabla\psi(v) - I) - \log|\det \nabla\psi(v)|\Big],
\end{aligned}$$

其中最后一步使用了分部积分。由于 $\|\nabla\psi(v) - I\|_{\mathsf{op}} \le \mathfrak{m}_2 < 1$ 对所有 $v \in \mathbb{R}^d$，[Rum18、定理 11] 推出

$$\mathsf{Tr}(\nabla\psi(v) - I) - \log|\det \nabla\psi(v)| \le \frac{\|\nabla\psi(v) - I\|_F^2}{2(1 - \|\nabla\psi(v) - I\|_{\mathsf{op}})}\,.$$

利用 $\|\cdot\|_F^2 \le d\|\cdot\|_{\mathsf{op}}^2$ 和界 $\mathfrak{m}_1$ 与 $\mathfrak{m}_2$，可得

$$\mathsf{KL}(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \frac{1}{2}\mathfrak{m}_1^2 + \frac{d\,\mathfrak{m}_2^2}{2(1-\mathfrak{m}_2)}\,.$$

□

F.2 引理 11 的证明

证明。 应用引理 2 和 10 以及 $\psi = \varphi_{x,\ y}$ (7) 和界 $\mathfrak{m}_1$ 与 $\mathfrak{m}_2$ （分别由引理 4和 5 给出），可得

$$\mathsf{KL}(\delta_x\tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y\tilde{\mathbf{P}}_{T,h}) \le \frac{9}{4T^2}\|x-y\|^2 + \frac{1089}{56}dM^2T^4\|x-y\|^2.$$

对 $\frac{1089}{56} < 20$ 取上界即可完成证明。 □

F.3 引理 12 的证明

证明。 结合引理 11 和 29，并取 γ 为最优的 W_2–耦合，使其耦合 μ 与π 即可完成证明。

□

F.4 定理 1的证明

证明。 当 $k = 0$时，利用引理 12 以及 $\pi = \tilde{\nu}_h$即可得到该结果。对于 $k \ge 1$，注意到

$$\mathsf{KL}(\mu\tilde{\mathbf{P}}_{T,h}^{k+1} \,\|\, \tilde{\nu}_h) = \mathsf{KL}((\mu\tilde{\mathbf{P}}_{T,h}^{k})\tilde{\mathbf{P}}_{T,h} \,\|\, \tilde{\nu}_h) \le \left(\frac{9}{4T^2} + 20dM^2T^4\right) W_2^2(\mu\tilde{\mathbf{P}}_{T,h}^{k}, \tilde{\nu}_h)$$

其中不等式由引理12得出。应用假设2 即可完成证明 f. □

G KL散度中的渐近偏差

G.1 引理证明13

证明。 固定任意点 x、 $y \in \mathbb{R}^d$。按照第 F.1节中引理 10的证明所采用的变量替换和化简，我们得到

$$\mathsf{KL}((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) = \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\frac{1}{2}\|\Phi_{x,y}(v)-v\|^2 + \mathsf{Tr}(\nabla\Phi_{x,y}(v)-I) - \log|\det\nabla\Phi_{x,y}(v)|\right].$$

根据引理 7， $\|\nabla\Phi_{x,\ y}(v)-I\|_{\mathsf{op}} < 1$ 对于所有 $v\in\mathbb{R}^d$，因此 [Rum18、定理 1。1] 表明

$$\begin{aligned}\mathsf{Tr}(\nabla\Phi_{x,y}(v)-I) - \log|\det\nabla\Phi_{x,y}(v)| &\le \frac{\|\nabla\Phi_{x,y}(v)-I\|_F^2}{2(1-\|\nabla\Phi_{x,y}(v)-I\|_{\mathsf{op}})}\\ &\le \frac{d\|\nabla\Phi_{x,y}(v)-I\|_{\mathsf{op}}^2}{2(1-\|\nabla\Phi_{x,y}(v)-I\|_{\mathsf{op}})}\\ &\le 3d\|\nabla\Phi_{x,y}(v)-I\|_{\mathsf{op}}^2\end{aligned}$$

其中，在第二个不等式中，我们利用了 $\|\cdot\|_F^2 \le d\|\cdot\|_{\mathsf{op}}^2$；在最后一个不等式中，我们利用了引理 7。因此，我们有

$$\mathsf{KL}((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\frac{1}{2}\|\Phi_{x,y}(v)-v\|^2 + 3d\|\nabla\Phi_{x,y}(v)-I\|_{\mathsf{op}}^2\right].$$

引理 6 指出

$$\|\Phi_{x,y}(v)-v\| \le \frac{3}{2T}\|x-y\| + 2h^2R(x,v)\,,$$

其中

$$R(x,v) \coloneqq LT^{-1}\|x\| + L\|v\| + MT^{-1}\|x\|^2 + MT\|v\|^2\,,$$

以及引理 7指出

$$\|\nabla\Phi_{x,y}(v)-I\|_{\mathsf{op}} \le \frac{11}{2}MT^2\|x-y\| + \frac{22}{9}h^2Q(x,v)\,,$$

其中

$$Q(x,v) \coloneqq L + M\|x\| + MT\|v\| + (M^2T^2+N)\|x\|^2 + (M^2T^4+NT^2)\|v\|^2\,.$$

因此，我们有

$$\mathsf{KL}((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d\|\boldsymbol{\gamma}_d) \le \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\frac{1}{2}\left(\frac{3}{2T}\|x-y\| + 2h^2R(x,v)\right)^2 + 3d\left(\frac{11}{2}MT^2\|x-y\| + \frac{22}{9}h^2Q(x,v)\right)^2\right].$$

利用 $(a_1+a_2)^2 \le 2a_1^2+2a_2^2$ 对于 a_1， $a_2\in\mathbb{R}$ 可得

$$\mathsf{KL}((\Phi_{x,y})_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \left(\frac{9}{4T^2} + \frac{363}{2}dM^2T^4\right)\|x-y\|^2 + 4h^4\mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[R^2(x,v) + d\frac{242}{27}Q^2(x,v)\right].$$

分别将 $(a_1+a_2+\cdots+a_k)^2 \le k(a_1^2+a_2^2+\cdots+a_k^2)$ 应用于 a_1、.....、 $a_k\in\mathbb{R}$，其中 $k=4$，并在$R^2(x,\ v)$ 和 $Q^2(x,\ v)$ 上取 5，然后化简常数，可得

$$\begin{aligned}\mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[R^2(x,v) + d\frac{242}{27}Q^2(x,v)\right] &\le 45dL^2 + \left(4\frac{L^2}{T^2} + 45dM^2\right)\|x\|^2 + \left(4\frac{M^2}{T^2} + 45d(M^2T^2+N)^2\right)\|x\|^4\\ &\quad + \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[(4L^2+45dM^2T^2)\|v\|^2 + \left(4M^2T^2 + 45d(M^2T^4+NT^2)^2\right)\|v\|^4\right].\end{aligned}$$

利用 $\mathbb{E}_{\boldsymbol{\gamma}_d}[\|v\|^2] = d$ 和 $\mathbb{E}_{\boldsymbol{\gamma}_d}[\|v\|^4] = d(d+2)$，即可得到所需的界。

□

G.2 引理的证明14

˜证明。我们从引理 29 出发，其中 $\mathbf{P} = \mathbf{P}_{T,\ h}$ 且 $\mathbf{Q} = \mathbf{P}_T$。取 γ 为最优的W_2耦合，使其连接 μ与 π，由引理3和 13，结论即得。

□

G.3 定理的证明 2

˜ ˜证明。令 $\mathbf{P}$ 和 $\mathbf{P}$ 分别简记为 $\mathbf{P}_{T,h}$ 和 $\mathbf{P}_T$。

For $k = 0$，引理 14 给出了 KL散度$(\mu\mathbf{P} \parallel \nu)$ 的上界，因此定理 2 成立。
现在假设 $k \geq 1$。由于 KL散度$(\mu\, \tilde{}\mathbf{Q}^k\, \tilde{}\mathbf{P} \parallel \nu) =$ KL散度$((\mu\, \tilde{}\mathbf{Q}^k)\, \tilde{}\mathbf{P} \parallel \nu\mathbf{P})$，我们可以应用引理 14 得到

$$\begin{aligned}\mathsf{KL}(\mu\tilde{\mathbf{Q}}^k\tilde{\mathbf{P}} \parallel \nu) \leq \Big(\frac{9}{4T^2}+\frac{363}{2}dM^2T^4\Big)W_2^2(\mu\tilde{\mathbf{Q}}^k,\nu)+ \\ 4h^4\Bigg[45dL^2+\Big(4\frac{L^2}{T^2}+45dM^2\Big)m_2(\mu\tilde{\mathbf{Q}}^k)+\Big(4\frac{M^2}{T^2}+45d(M^2T^2+N)^2\Big)m_4(\mu\tilde{\mathbf{Q}}^k) \\ +(4L^2+45dM^2T^2)d+\big(4M^2T^2+45d(M^2T^4+NT^2)^2\big)d(d+2)\Bigg].\end{aligned}$$

利用三角不等式和 $(a_1+a_2)^2 \leq 2a_1^2+2a_2^2$，对于 $a_1,\ a_2 \in \mathbb{R}$，我们有

$$\begin{aligned}W_2^2(\mu\tilde{\mathbf{Q}}^k,\nu) &\leq 2W_2^2(\mu\tilde{\mathbf{Q}}^k,\tilde{\nu}_{\tilde{\mathbf{Q}}})+2W_2^2(\tilde{\nu}_{\tilde{\mathbf{Q}}},\nu) \\ &\leq 2c_1^2e^{-2c_2k}\,W_2^2(\mu,\tilde{\nu}_{\tilde{\mathbf{Q}}})+2\Delta_h^2\end{aligned}$$

其中，最后一个不等式由假设 2 和3得出。因此，我们得到

$$\mathsf{KL}(\mu\tilde{\mathbf{Q}}^k\tilde{\mathbf{P}} \parallel \nu) \leq 2c_1^2e^{-2c_2k}\Big(\frac{9}{4T^2}+\frac{363}{2}dM^2T^4\Big)W_2^2(\mu,\tilde{\nu}_{\tilde{\mathbf{Q}}})+\mathsf{bias}$$

其中

$$\begin{aligned}\mathsf{bias} \leq 2\Delta_h^2\Big(\frac{9}{4T^2}+\frac{363}{2}dM^2T^4\Big)+4h^4\Bigg[45dL^2+\Big(4\frac{L^2}{T^2}+45dM^2\Big)m_2(\mu\tilde{\mathbf{Q}}^k) \\ +\Big(4\frac{M^2}{T^2}+45d(M^2T^2+N)^2\Big)m_4(\mu\tilde{\mathbf{Q}}^k)+(4L^2+45dM^2T^2)d+\big(4M^2T^2+45d(M^2T^4+NT^2)^2\big)d(d+2)\Bigg].\end{aligned}$$

□

H R´ényi散度的混合时间

H.1 引理 15 的证明

证明。回想一下，$\boldsymbol{\gamma}_d$ 和 $\psi_{\#}\boldsymbol{\gamma}_d$ 都具有相对于勒贝格测度的密度，分别记为 $\boldsymbol{\gamma}_d(v)$ 和 $(\psi_{\#}\boldsymbol{\gamma}_d)(v)$。由于 $\|\nabla\psi(v) - I\|_{\mathsf{op}} < 1$，对于所有 $v \in \mathbb{R}^d$，我们都有 $\det\nabla\psi(v) > 0$，由变量替换公式可得

$$(\psi_{\#}\boldsymbol{\gamma}_d)(\psi(v)) = \frac{\boldsymbol{\gamma}_d(v)}{\det\nabla\psi(v)}, \qquad v \in \mathbb{R}^d. \tag{24}$$

根据 R´ényi 散度（定义5），

$$\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) = \frac{1}{q-1}\log \mathop{\mathbb{E}}_{V\sim\boldsymbol{\gamma}_d}\left[\left(\frac{(\psi_{\#}\boldsymbol{\gamma}_d)(V)}{\boldsymbol{\gamma}_d(V)}\right)^q\right].$$

重新写出被积函数，

$$\left(\frac{(\psi_{\#}\boldsymbol{\gamma}_d)(v)}{\boldsymbol{\gamma}_d(v)}\right)^q \boldsymbol{\gamma}_d(v) = \left(\frac{(\psi_{\#}\boldsymbol{\gamma}_d)(v)}{\boldsymbol{\gamma}_d(v)}\right)^{q-1} (\psi_{\#}\boldsymbol{\gamma}_d)(v),$$

我们可以等价地将 R´ényi 散度表示为

$$\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) = \frac{1}{q-1}\log \mathop{\mathbb{E}}_{U\sim\psi_{\#}\boldsymbol{\gamma}_d}\left[\left(\frac{(\psi_{\#}\boldsymbol{\gamma}_d)(U)}{\boldsymbol{\gamma}_d(U)}\right)^{q-1}\right].$$

现在我们使用推前测度的定义性质：对于任意可测函数$F\colon\ \mathbb{R}^d \to \mathbb{R}$,

$$\mathop{\mathbb{E}}_{U\sim\psi_{\#}\boldsymbol{\gamma}_d}[F(U)] = \mathop{\mathbb{E}}_{V\sim\boldsymbol{\gamma}_d}[F(\psi(V))].$$

将此恒等式应用于

$$F(u) = \left(\frac{(\psi_{\#}\boldsymbol{\gamma}_d)(u)}{\boldsymbol{\gamma}_d(u)}\right)^{q-1},$$

我们得到

$$\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) = \frac{1}{q-1}\log \mathop{\mathbb{E}}_{V\sim\boldsymbol{\gamma}_d}\left[\left(\frac{(\psi_{\#}\boldsymbol{\gamma}_d)(\psi(V))}{\boldsymbol{\gamma}_d(\psi(V))}\right)^{q-1}\right].$$

代入变量替换恒等式（24）后，即可得到一个完全以$V\sim\boldsymbol{\gamma}_d$表示的表达式，我们将在证明的余下部分对其进行界定。特别地，

$$\begin{aligned}\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) &= \frac{1}{q-1}\log \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\left(\frac{\boldsymbol{\gamma}_d(v)|\det\nabla\psi(v)|^{-1}}{\boldsymbol{\gamma}_d(\psi(v))}\right)^{q-1}\right]\\ &= \frac{1}{q-1}\log \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\exp\left(\frac{q-1}{2}(\|\psi(v)\|^2-\|v\|^2)\right)|\det\nabla\psi(v)|^{-(q-1)}\right]\end{aligned}$$

我们还有以下界

$$\begin{aligned}\|\psi(v)\|^2-\|v\|^2 &= 2\langle v,\psi(v)-v\rangle + \|\psi(v)-v\|^2\\ &\le 2\|v\|\|\psi(v)-v\| + \mathfrak{m}_1^2\\ &\le 2\mathfrak{m}_1\|v\| + \mathfrak{m}_1^2\end{aligned}$$

and

$$|\det\nabla\psi(v)| \ge (1-\mathfrak{m}_2)^d.$$

因此，我们得到

$$\begin{aligned}\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) &= \frac{1}{q-1}\log \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\exp\left(\frac{q-1}{2}(\|\psi(v)\|^2-\|v\|^2)\right)|\det\nabla\psi(v)|^{-(q-1)}\right]\\ &\le \frac{1}{q-1}\log \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\exp\left(\frac{q-1}{2}(2\mathfrak{m}_1\|v\|+\mathfrak{m}_1^2)\right)(1-\mathfrak{m}_2)^{-d(q-1)}\right]\\ &= \frac{1}{q-1}\left[-d(q-1)\log(1-\mathfrak{m}_2) + \frac{q-1}{2}\mathfrak{m}_1^2 + \log \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[e^{(q-1)\mathfrak{m}_1\|v\|}\right]\right].\end{aligned}$$

应用引理31，并结合 $c=(q-1)\mathfrak{m}_1$，可得

$$\begin{aligned}\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) &\le d\log\frac{1}{1-\mathfrak{m}_2}+\frac{\mathfrak{m}_1^2}{2}+\sqrt{d}\,\mathfrak{m}_1+\frac{(q-1)\mathfrak{m}_1^2}{2}\\ &= d\log\frac{1}{1-\mathfrak{m}_2}+\sqrt{d}\,\mathfrak{m}_1+\frac{q\mathfrak{m}_1^2}{2}\end{aligned}$$

使用 $\log\frac{1}{1-x}\le\frac{x}{1-x}$ for $x\in[0,1)$，我们得到

$$\mathsf{R}_q(\psi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \frac{d\mathfrak{m}_2}{1-\mathfrak{m}_2}+\sqrt{d}\,\mathfrak{m}_1+\frac{q\mathfrak{m}_1^2}{2}\,.$$

□

H.2 引理的证明 16

证明。 应用引理 2 和 15 以及 $\psi=\varphi_{x,\,y}$ (7) 和界 $\mathfrak{m}_1$ 和 $\mathfrak{m}_2$，分别由引理 4和 5 给出，可得

$$\mathsf{R}_q(\delta_x\tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y\tilde{\mathbf{P}}_{T,h}) \le \frac{99}{14}dMT^2\|x-y\|+\frac{3\sqrt{d}}{2T}\|x-y\|+\frac{9q}{8T^2}\|x-y\|^2$$

化简常数后，我们得到一个上界:

$$\mathsf{R}_q(\delta_x\tilde{\mathbf{P}}_{T,h} \,\|\, \delta_y\tilde{\mathbf{P}}_{T,h}) \le \left(8dMT^2+\frac{3\sqrt{d}}{2T}\right)\|x-y\|+\frac{9q}{8T^2}\|x-y\|^2\,.$$

□

H.3 引理的证明 17

证明。 直接应用引理30以及引理16即可证明该结论。 □

H.4 定理3的证明

证明。 令 $\mathbf{P}$ 和 $\tilde{\nu}$ 分别表示$\mathbf{P}_{T,\,h}$ 和 $\tilde{\nu}_h$。我们首先应用引理 17 得到

$$\mathsf{R}_q(\mu\tilde{\mathbf{P}}^{k+1} \,\|\, \tilde{\nu}) \le \inf_{\gamma\in\Gamma(\mu\tilde{\mathbf{P}}^k,\tilde{\nu})}\frac{1}{q-1}\log\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{(q-1)\left[\left(8dMT^2+\frac{3\sqrt{d}}{2T}\right)\|x-y\|+\frac{9q}{8T^2}\|x-y\|^2\right]}\right].$$

令 $\delta_1\coloneqq(q-1)\left(8dMT^2+\frac{3\sqrt{d}}{2T}\right)$ 和 $\delta_2\coloneqq\frac{9q}{(q-1)8T^2}$。然后将下确界移入对数中，可得

$$\mathsf{R}_q(\mu\tilde{\mathbf{P}}^{k+1} \,\|\, \tilde{\nu}) \le \frac{1}{q-1}\log\inf_{\gamma\in\Gamma(\mu\tilde{\mathbf{P}}^k,\tilde{\nu})}\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\delta_1\|x-y\|+\delta_2\|x-y\|^2}\right].$$

对于任意 $\beta>0$，都有 $\delta_1\|x-y\|\le\frac{\beta^2\delta_1^2}{2}+\frac{\|x-y\|^2}{2\beta^2}$。我们稍后将选择 β（与 x,y 无关）。应用这一点，我们得到

$$\begin{aligned}\mathsf{R}_q(\mu\tilde{\mathbf{P}}^{k+1} \,\|\, \tilde{\nu}) &\le \frac{1}{q-1}\log\inf_{\gamma\in\Gamma(\mu\tilde{\mathbf{P}}^k,\tilde{\nu})}\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\beta^2\delta_1^2}{2}+\frac{\|x-y\|^2}{2\beta^2}+\delta_2\|x-y\|^2}\right]\\ &= \frac{1}{q-1}\log\inf_{\gamma\in\Gamma(\mu\tilde{\mathbf{P}}^k,\tilde{\nu})}\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\beta^2\delta_1^2}{2}+(\delta_2+\frac{1}{2\beta^2})\|x-y\|^2}\right].\end{aligned}$$

将指数项和对数项拆开，我们得到

$$(q-1)\mathsf{R}_q(\mu\tilde{\mathbf{P}}^{k+1} \,\|\, \tilde{\nu}) \le \frac{\beta^2\delta_1^2}{2} + \log \inf_{\gamma\in\Gamma(\mu\tilde{\mathbf{P}}^k,\tilde{\nu})} \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \left[e^{(\delta_2+\frac{1}{2\beta^2})\|x-y\|^2}\right].$$

~ ~在假设 4 对 $\mathbf{Q} = \mathbf{P}_{T,\ h}$成立的条件下，我们知道

$$W_\psi(\mu\tilde{\mathbf{P}}^k, \tilde{\nu}) \le c_1 e^{-c_2 k}\, W_\psi(\mu, \tilde{\nu})\,.$$

根据定义 3，这意味着存在 $\gamma^* \in \Gamma(\mu\tilde{\mathbf{P}}^k,\ \tilde{\nu})$ 以及某个 $\lambda^* \le c_1 e^{-c_2 k}\, W_\psi(\mu,\ \tilde{\nu})$，使得

$$\mathop{\mathbb{E}}_{(x,y)\sim\gamma^*} \left[e^{\frac{\|x-y\|^2}{\lambda^{*2}}}\right] \le 2\,.$$

此外，利用引理 1，我们知道，对于任意 $a \in [0, \frac{1}{\lambda^{*2}}]$ —

$$\mathop{\mathbb{E}}_{(x,y)\sim\gamma^*} \left[e^{a\|x-y\|^2}\right] \le 2^{a\lambda^{*2}}\,.$$

因此，只要 $\delta_2 + \frac{1}{2\beta^2} \le \frac{1}{\lambda^{*2}}$，我们就可以写成

$$\begin{aligned}
(q-1)\mathsf{R}_q(\mu\tilde{\mathbf{P}}^{k+1} \,\|\, \tilde{\nu}) &\le \frac{\beta^2\delta_1^2}{2} + \log \mathop{\mathbb{E}}_{(x,y)\sim\gamma^*} \left[e^{(\delta_2+\frac{1}{2\beta^2})\|x-y\|^2}\right] \\
&\le \frac{\beta^2\delta_1^2}{2} + \log 2^{(\delta_2+\frac{1}{2\beta^2})\lambda^{*2}} \\
&= \frac{\beta^2\delta_1^2}{2} + \left(\delta_2 + \frac{1}{2\beta^2}\right)\lambda^{*2}\log 2 \\
&\le \delta_2(\log 2)c_1^2 e^{-2c_2k}\, W_\psi^2(\mu,\tilde{\nu}) + \frac{\beta^2\delta_1^2}{2} + c_1^2 e^{-2c_2k}\, W_\psi^2(\mu,\tilde{\nu})\frac{\log 2}{2\beta^2}\,.
\end{aligned}$$

为了最小化上界，我们取

$$\beta^2 = \frac{c_1 e^{-c_2 k}\, W_\psi(\mu,\tilde{\nu})\sqrt{\log 2}}{\delta_1}\,.$$

代入后，我们的上界为

$$(q-1)\mathsf{R}_q(\mu\tilde{\mathbf{P}}^{k+1} \,\|\, \tilde{\nu}) \le \delta_2(\log 2)c_1^2 e^{-2c_2k}\, W_\psi^2(\mu,\tilde{\nu}) + \delta_1\sqrt{\log 2}\, c_1 e^{-c_2k}\, W_\psi(\mu,\tilde{\nu})\,.$$

经过基本化简可得上界

$$(q-1)\mathsf{R}_q(\mu\tilde{\mathbf{P}}^{k+1} \,\|\, \tilde{\nu}) \le (\delta_1+\delta_2)\sqrt{\log 2}\, c_1^2 e^{-c_2k}\, \max\{1, W_\psi^2(\mu,\tilde{\nu})\}\,.$$

然而，这一结论成立的前提是

$$\delta_2 + \frac{\delta_1 e^{c_2k}}{2c_1 W_\psi(\mu,\tilde{\nu})\sqrt{\log 2}} \le \frac{e^{2c_2k}}{c_1^2 W_\psi^2(\mu,\tilde{\nu})}$$

其中我们代入了 β^2 的最优值，并且还利用了 $\lambda^* \le c_1 e^{-c_2 k}\, W_\psi(\mu,\ \tilde{\nu})$。只要.....成立，

$$k \ge \frac{1}{c_2}\log\left(\frac{c_1 W_\psi(\mu,\tilde{\nu})}{4\sqrt{\log 2}}\left(\delta_1 + \sqrt{\delta_1^2 + 16(\log 2)\delta_2}\right)\right)\,.$$

□

I R´enyi 散度中的渐近偏差

I.1 引理18 的证明

证明。 我们用 Φ 作为 $\Phi_{x、y}$的简写。我们首先进行变量替换，这与引理 15的证明类似，即有

$$\mathsf{R}_q(\Phi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) = \frac{1}{q-1}\log \mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[\exp\left(\frac{q-1}{2}(\|\Phi(v)\|^2 - \|v\|^2)\right)|\det\nabla\Phi(v)|^{-(q-1)}\right].$$

引理8 指出

$$\|\Phi(v)-v\| \le \mathsf{p}_{xy}\|x-y\| + h\mathsf{p}_v\|v\| + h\mathsf{p}_x\|x\|$$

其中 $\mathsf{p}^{xy} := \frac{3}{2T}$、$\mathsf{p}_v := \frac{7}{25T}$ 以及 $\mathsf{p}_x := \frac{49}{180}L$。引理 9 指出

$$\|\nabla\Phi(v) - I\|_{\mathsf{op}} \le \min\left\{\frac{15}{18}, \mathsf{j}_{xy}\|x-y\| + h\mathsf{j}_c + h\mathsf{j}_v\|v\| + h\mathsf{j}_x\|x\|\right\}$$

其中 $\mathsf{j}^{xy} := \frac{11}{2}MT^2$、$\mathsf{j}_c := \frac{44}{135T}$、$\mathsf{j}_v := \frac{440}{135}MT^2$ 和 $\mathsf{j}_x := \frac{352}{675}MT$。

我们有

$$\begin{aligned}
\|\Phi(v)\|^2 - \|v\|^2 &\le 2\|v\|\|\Phi(v)-v\| + \|\Phi(v)-v\|^2 \\
&\le 2\|v\|(\mathsf{p}_{xy}\|x-y\| + h\mathsf{p}_v\|v\| + h\mathsf{p}_x\|x\|) + (\mathsf{p}_{xy}\|x-y\| + h\mathsf{p}_v\|v\| + h\mathsf{p}_x\|x\|)^2 \\
&= h\mathsf{p}_v\|v\|^2(h\mathsf{p}_v+2) + 2(h\mathsf{p}_v+1)\|v\|(\mathsf{p}_{xy}\|x-y\| + h\mathsf{p}_x\|x\|) + (\mathsf{p}_{xy}\|x-y\| + h\mathsf{p}_x\|x\|)^2
\end{aligned}$$

为了界定雅可比矩阵，注意到对于任意满足 $\|A-I\|_{\mathsf{op}} < 1$ 的矩阵 A，$|\det A| \ge (1-\|A-I\|_{\mathsf{op}})^d$。因此

$$\begin{aligned}
|\det\nabla\Phi(v)|^{-(q-1)} &\le (1-\|\nabla\Phi(v)-I\|_{\mathsf{op}})^{-(q-1)d} \\
&\le \exp\left((q-1)d\frac{\|\nabla\Phi(v)-I\|_{\mathsf{op}}}{1-\|\nabla\Phi(v)-I\|_{\mathsf{op}}}\right) \\
&\le \exp\left(6(q-1)d\|\nabla\Phi(v)-I\|_{\mathsf{op}}\right) \\
&\le \exp\left(6(q-1)d(\mathsf{j}_{xy}\|x-y\| + h\mathsf{j}_c + h\mathsf{j}_v\|v\| + h\mathsf{j}_x\|x\|)\right)
\end{aligned}$$

综合以上各项可得

$$\exp\left(\frac{q-1}{2}(\|\Phi(v)\|^2 - \|v\|^2)\right)|\det\nabla\Phi(v)|^{-(q-1)} \le \exp\left((q-1)(\mathsf{c}_0 + \mathsf{c}_1\|v\| + \mathsf{c}_2\|v\|^2)\right)$$

其中

$$\begin{aligned}
\mathsf{c}_0 &:= \frac{(\mathsf{p}_{xy}\|x-y\| + h\mathsf{p}_x\|x\|)^2}{2} + 6d\mathsf{j}_{xy}\|x-y\| + 6dh\mathsf{j}_c + 6dh\mathsf{j}_x\|x\| \\
\mathsf{c}_1 &:= (h\mathsf{p}_v+1)(\mathsf{p}_{xy}\|x-y\| + h\mathsf{p}_x\|x\|) + 6dh\mathsf{j}_v \\
\mathsf{c}_2 &:= \frac{h\mathsf{p}_v(h\mathsf{p}_v+2)}{2}.
\end{aligned}$$

令 $s := q-1$。因此我们有

$$\begin{aligned}
\mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[e^{\left(s(\mathsf{c}_0+\mathsf{c}_1\|v\|+\mathsf{c}_2\|v\|^2)\right)}\right] &= e^{s\mathsf{c}_0}\mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[e^{\left(s(\mathsf{c}_1\|v\|+\mathsf{c}_2\|v\|^2)\right)}\right] \\
&\le e^{s\mathsf{c}_0}\left(\mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[e^{2s\mathsf{c}_1\|v\|}\right]\right)^{\frac{1}{2}}\left(\mathop{\mathbb{E}}_{v\sim\boldsymbol{\gamma}_d}\left[e^{2s\mathsf{c}_2\|v\|^2}\right]\right)^{\frac{1}{2}} \\
&\le e^{s\mathsf{c}_0}e^{s\mathsf{c}_1\sqrt{d}+s^2\mathsf{c}_1^2}(1-4s\mathsf{c}_2)^{-\frac{d}{4}}
\end{aligned}$$

其中最后一个不等式由引理 31 以及高斯分布的性质得出，并且在任意情况下都成立

$$s\mathsf{c}_2 < \frac{1}{4}\,.$$

取对数并除以 s，得到

$$\mathsf{R}_q(\Phi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \mathsf{c}_0 + \sqrt{d}\,\mathsf{c}_1 + s\mathsf{c}_1^2 - \frac{d}{4s}\log(1 - 4s\mathsf{c}_2)\,.$$

对 $x \in [0, \frac{1}{2}]$使用 $-\log(1-x) \le \frac{x}{1-x} \le 2x$，得到

$$\mathsf{R}_q(\Phi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \mathsf{c}_0 + \sqrt{d}\,\mathsf{c}_1 + s\mathsf{c}_1^2 + 2d\mathsf{c}_2$$

只要满足以下条件，该结论就成立

$$s\mathsf{c}_2 \le \frac{1}{8}\,.$$

当.....时，条件 $s\mathsf{c}_2 \le \frac{1}{8}$ 成立

$$0 < h \le \frac{1}{\mathsf{p}_v}\left(\left(1 + \frac{1}{4s}\right)^{\frac{1}{2}} - 1\right) = \frac{25T}{7}\left(\left(1 + \frac{1}{4s}\right)^{\frac{1}{2}} - 1\right).$$

为简化上界

$$\mathsf{R}_q(\Phi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le \mathsf{c}_0 + \sqrt{d}\,\mathsf{c}_1 + s\mathsf{c}_1^2 + 2d\mathsf{c}_2$$

注意

$$\begin{aligned}
\mathsf{c}_0 &\le \mathsf{p}_{xy}^2\|x-y\|^2 + 6d\mathsf{j}_{xy}\|x-y\| + h^2\mathsf{p}_x^2\|x\|^2 + 6dh\mathsf{j}_x\|x\| + 6dh\mathsf{j}_c \\
\sqrt{d}\,\mathsf{c}_1 &= \sqrt{d}(h\mathsf{p}_v+1)\mathsf{p}_{xy}\|x-y\| + \sqrt{d}(h\mathsf{p}_v+1)h\mathsf{p}_x\|x\| + 6d^{3/2}h\mathsf{j}_v \\
s\mathsf{c}_1^2 &\le 3s(h\mathsf{p}_v+1)^2\mathsf{p}_{xy}^2\|x-y\|^2 + 3s(h\mathsf{p}_v+1)^2h^2\mathsf{p}_x^2\|x\|^2 + 108sd^2h^2\mathsf{j}_v^2 \\
2d\mathsf{c}_2 &\le dh^2\mathsf{p}_v^2 + 2dh\mathsf{p}_v\,.
\end{aligned}$$

因此，将同类项合并后可得

$$\begin{aligned}
\mathsf{R}_q(\Phi_{\#}\boldsymbol{\gamma}_d \,\|\, \boldsymbol{\gamma}_d) \le\ & \mathsf{p}_{xy}^2\|x-y\|^2(1 + 3s(h\mathsf{p}_v+1)^2) + \sqrt{d}\,\|x-y\|(6\sqrt{d}\,\mathsf{j}_{xy} + \mathsf{p}_{xy}(h\mathsf{p}_v+1)) \\
& + h^2\mathsf{p}_x^2\|x\|^2(1 + 3s(h\mathsf{p}_v+1)^2) + h\sqrt{d}\,\|x\|(6\sqrt{d}\,\mathsf{j}_x + \mathsf{p}_x(h\mathsf{p}_v+1)) \\
& + dh(6\mathsf{j}_c + 6\sqrt{d}\,\mathsf{j}_v + 108sdh\mathsf{j}_v^2 + h\mathsf{p}_v^2 + 2\mathsf{p}_v)\,.
\end{aligned}$$

□

I.2 引理19 的证明

˜证明。 通过应用引理 30，其中 $\mathbf{P} = \mathbf{P}_{T,\ h}$ 以及 $\mathbf{Q} = \mathbf{P}_T$ 再结合引理 18，可得

$$\mathsf{R}_q(\mu\tilde{\mathbf{P}}_{T,h} \,\|\, \pi\mathbf{P}_T) \le \inf_{\gamma\in\Gamma(\mu,\pi)} \frac{1}{q-1}\log \mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{(q-1)(\alpha_0+\alpha_1\|x-y\|+\alpha_2\|x-y\|^2+\alpha_3\|x\|+\alpha_4\|x\|^2)}\right]$$

其中 α_0、α_2、α_3、α_4 定义为引理 18所得结果的系数。

□

I.3 定理 4 的证明

我们先陈述以下推论，其证明与定理 3 的证明非常相似，因此省略。该推论有助于证明定理 4。

推论 7。考虑与定理 *3* 相同的条件，并假设 *4* 成立。对于一个~一般的马尔可夫核 $\mathbf{Q}$，其平稳分布为 $\tilde{\nu}_{\tilde{\mathbf{Q}}}$。进一步定义 δ_1 和 δ_2，如定理 *3*中所述。那么对于任意整数

$$k \geq \frac{1}{c_2} \log \left(\frac{c_1 W_\psi(\mu, \tilde{\nu}_{\tilde{\mathbf{Q}}})}{4\sqrt{\log 2}} \left(\delta_1 + \sqrt{\delta_1^2 + 16(\log 2)\delta_2} \right) \right)$$

有

$$\mathsf{R}_q(\mu \tilde{\mathbf{Q}}^k \mathbf{P}_{T,h} \parallel \tilde{\nu}_{\tilde{\mathbf{Q}}} \tilde{\mathbf{P}}_{T,h}) \leq \frac{\delta_1 + \delta_2}{q-1} \sqrt{\log 2}\, c_1^2 e^{-c_2 k} \max\{1, W_\psi^2(\mu, \tilde{\nu}_{\tilde{\mathbf{Q}}})\}\,.$$

~ ~定理 *4* 的证明。令$\mathbf{P}$ 表示 $\mathbf{P}_{T,\,h}$。应用 R´enyi 散度的弱三角不等式（3）以及关于阶数（2）的单调性，可得

$$\mathsf{R}_q(\mu \tilde{\mathbf{Q}}^k \tilde{\mathbf{P}} \parallel \nu) \leq \frac{3}{2} \mathsf{R}_{2q}(\mu \tilde{\mathbf{Q}}^k \tilde{\mathbf{P}} \parallel \tilde{\nu}_{\tilde{\mathbf{Q}}} \tilde{\mathbf{P}}) + \mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}} \tilde{\mathbf{P}} \parallel \nu)\,.$$

第一项可直接由推论 7 进行界定，从而得到定理 4陈述中的第一项。为界定第二项，引理 19 表明

$$\mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}} \tilde{\mathbf{P}} \parallel \nu) \leq \inf_{\gamma \in \Gamma(\tilde{\nu}_{\tilde{\mathbf{Q}}}, \nu)} \frac{1}{2q-1} \log \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \left[e^{(2q-1)(\alpha_0 + \alpha_1 \|x-y\| + \alpha_2 \|x-y\|^2 + \alpha_3 \|x\| + \alpha_4 \|x\|^2)} \right]$$

其中

$$\begin{aligned}
\alpha_0 &\coloneqq dh(6\mathsf{j}_c + 6\sqrt{d}\,\mathsf{j}_v + 108(2q-1)dh\mathsf{j}_v^2 + h\mathsf{p}_v^2 + 2\mathsf{p}_v) \\
\alpha_1 &\coloneqq \sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_{xy} + \mathsf{p}_{xy}(h\mathsf{p}_v + 1)) \\
\alpha_2 &\coloneqq \mathsf{p}_{xy}^2(1 + 3(2q-1)(h\mathsf{p}_v + 1)^2) \\
\alpha_3 &\coloneqq h\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x + \mathsf{p}_x(h\mathsf{p}_v + 1)) \\
\alpha_4 &\coloneqq h^2\mathsf{p}_x^2(1 + 3(2q-1)(h\mathsf{p}_v + 1)^2)\,.
\end{aligned} \tag{25}$$

将指数中的常数项分离出来，我们得到

$$(2q-1)\mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}} \tilde{\mathbf{P}} \parallel \nu) \leq (2q-1)\alpha_0 + \log \inf_{\gamma \in \Gamma(\tilde{\nu}_{\tilde{\mathbf{Q}}}, \nu)} \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \left[e^{(2q-1)(\alpha_1 \|x-y\| + \alpha_2 \|x-y\|^2 + \alpha_3 \|x\| + \alpha_4 \|x\|^2)} \right].$$

对项 $\alpha_1\|x-y\|$ 和 $\alpha_3\|x\|$ 应用杨氏不等式，我们得到

$$\alpha_1 \|x-y\| \leq \frac{\beta_1^2 \alpha_1^2}{2} + \frac{\|x-y\|^2}{2\beta_1^2}$$

对于任意稍后将选定的 $\beta_1 > 0$，以及

$$\alpha_3 \|x\| \leq \frac{\beta_3^2 \alpha_3^2}{2} + \frac{\|x\|^2}{2\beta_3^2}$$

对于任意稍后也将选定的 $\beta_3 > 0$ 。定义如下

$$\eta_0 \coloneqq \frac{\beta_1^2\alpha_1^2}{2} + \frac{\beta_3^2\alpha_3^2}{2}$$
$$\eta_1 \coloneqq \alpha_2 + \frac{1}{2\beta_1^2}$$
$$\eta_2 \coloneqq \alpha_4 + \frac{1}{2\beta_3^2}$$

从而得到

$$(2q-1)\mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}}\tilde{\mathbf{P}} \parallel \nu) \le (2q-1)(\alpha_0+\eta_0) + \log \inf_{\gamma\in\Gamma(\tilde{\nu}_{\tilde{\mathbf{Q}}},\nu)} \mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{(2q-1)(\eta_1\|x-y\|^2+\eta_2\|x\|^2)}\right]. \quad (26)$$

利用 $\eta_2\|x\|^2 \le 2\eta_2\|x-y\|^2 + 2\eta_2\|y\|^2$，我们得到

$$\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{(2q-1)(\eta_1\|x-y\|^2+\eta_2\|x\|^2)}\right] \le \mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{(2q-1)(\eta_1+2\eta_2)\|x-y\|^2} e^{2(2q-1)\eta_2\|y\|^2)}\right]. \quad (27)$$

利用 H¨older 不等式，我们得到

$$\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{(2q-1)(\eta_1+2\eta_2)\|x-y\|^2} e^{2(2q-1)\eta_2\|y\|^2)}\right] \le \underbrace{\left(\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{2(2q-1)(\eta_1+2\eta_2)\|x-y\|^2}\right]\right)^{1/2}}_{\mathrm{I}} \underbrace{\left(\mathop{\mathbb{E}}_{y\sim\nu}\left[e^{4(2q-1)\eta_2\|y\|^2)}\right]\right)^{1/2}}_{\mathrm{II}}. \quad (28)$$

为处理 I，注意到假设5 告诉我们

$$W_\psi(\tilde{\nu}_{\tilde{\mathbf{Q}}}, \nu) \le \Delta_h .$$

根据定义 3，这意味着存在一个耦合 $\gamma^* \in \Gamma(\tilde{\nu}_{\tilde{\mathbf{Q}}},\ \nu)$ 以及某个 $\lambda^* \le \Delta_h$ 使得

$$\mathop{\mathbb{E}}_{(x,y)\sim\gamma^*}\left[e^{\frac{\|x-y\|^2}{\lambda^{*2}}}\right] \le 2 .$$

此外，利用引理 1，我们知道对于任意 $a \in [0, \frac{1}{\lambda^{*2}}]$ —

$$\mathop{\mathbb{E}}_{(x,y)\sim\gamma^*}\left[e^{a\|x-y\|^2}\right] \le 2^{a\lambda^{*2}} .$$

因此，只要

$$2(2q-1)(\eta_1+2\eta_2) \le \frac{1}{\lambda^{*2}} \quad (29)$$

我们就得到

$$\mathrm{I} \le 2^{(2q-1)(\eta_1+2\eta_2)\lambda^{*2}} . \quad (30)$$

我们在29证明结束时验证（

为处理 II, 回想一下 ν 具有有限的 ψ–范数，即对于 $X \sim \nu$， $\|X\|_\psi \le K_\nu$。因此，利用引理 1，只要

$$4(2q-1)\eta_2 \le \frac{1}{K_\nu^2} , \quad (31)$$

我们可以得出结论

$$\text{Ⅱ} \le 2^{2(2q-1)\eta_2 K_\nu^2}\,. \tag{32}$$

我们将在证明末尾验证（31）。

因此，(26)、(27)、(28)、(30)和(32)给出

$$(2q-1)\mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}}\tilde{\mathbf{P}} \,\|\, \nu) \le (2q-1)(\alpha_0+\eta_0) + (2q-1)(\eta_1+2\eta_2)\lambda^{*2}\log 2 + 2(2q-1)\eta_2 K_\nu^2 \log 2\,.$$

将两边除以 $2q-1$，得到

$$\mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}}\tilde{\mathbf{P}} \,\|\, \nu) \le \alpha_0+\eta_0 + (\eta_1+2\eta_2)\lambda^{*2}\log 2 + 2\eta_2 K_\nu^2 \log 2\,.$$

β_1 和 β_3 的最优值分别为 β_1^* 和 β_3^*，由下式给出

$$\beta_1^{*2} = \frac{\lambda^*\sqrt{\log 2}}{\alpha_1}\ ,\ \beta_3^{*2} = \frac{\sqrt{2(\lambda^{*2}+K_\nu^2)\log 2}}{\alpha_3}\,.$$

在这些最优选择下

$$\eta_0 = \frac{\alpha_1\lambda^*\sqrt{\log 2}}{2} + \frac{\alpha_3\sqrt{2(\lambda^{*2}+K_\nu^2)\log 2}}{2}$$

and

$$\eta_1+2\eta_2 = \alpha_2 + \frac{\alpha_1}{2\lambda^*\sqrt{\log 2}} + 2\alpha_4 + \frac{\alpha_3}{\sqrt{2(\lambda^{*2}+K_\nu^2)\log 2}}$$

and

$$\eta_2 = \alpha_4 + \frac{\alpha_3}{2\sqrt{2(\lambda^{*2}+K_\nu^2)\log 2}}\,.$$

因此，化简常数并利用 λ^*，$K_\nu \le \sqrt{\lambda^{*2}+K_\nu^2}$，得到

$$\mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}}\tilde{\mathbf{P}} \,\|\, \nu) \le \alpha_0 + 2\alpha_1\lambda^* + \alpha_2\lambda^{*2} + 2\alpha_3\Big(\lambda^* + K_\nu + \sqrt{\lambda^{*2}+K_\nu^2}\Big) + 2\alpha_4(\lambda^{*2}+K_\nu^2)\,.$$

对 $\lambda^* \le \Delta_h$ 取上界可得

$$\mathsf{R}_{2q}(\tilde{\nu}_{\tilde{\mathbf{Q}}}\tilde{\mathbf{P}} \,\|\, \nu) \le \alpha_0 + 2\alpha_1\Delta_h + \alpha_2\Delta_h^2 + 2\alpha_3\Big(\Delta_h + K_\nu + \sqrt{\Delta_h^2+K_\nu^2}\Big) + 2\alpha_4(\Delta_h^2+K_\nu^2)\,.$$

为进一步简化，注意到引理19中的步长约束可写为

$$h\mathsf{p}_v + 1 \le \left(1 + \frac{1}{4(2q-1)}\right)^{\frac{1}{2}} =: u$$

并进一步定义 $s := 2q-1$。这给出了以下界

$$\begin{aligned}
\alpha_1 &\le \sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_{xy} + \mathsf{p}_{xy}u)\\
\alpha_2 &\le \mathsf{p}_{xy}^2(1+3su^2)\\
\alpha_3 &\le h\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x + \mathsf{p}_x u)\\
\alpha_4 &\le h^2\mathsf{p}_x^2(1+3su^2)\,.
\end{aligned} \tag{33}$$

接下来需要检查满足这些约束所需的条件。

为确保 (29)，**即** $2(2q-1)(\eta_1+2\eta_2) \le \frac{1}{\lambda^{*2}}$ 要满足这一判据，我们要求

$$2(2q-1)\left(\alpha_2+2\alpha_4+\frac{\alpha_1}{2\lambda^*\sqrt{\log 2}}+\frac{\alpha_3}{\sqrt{2(\lambda^{*2}+K_\nu^2)\log 2}}\right)\le\frac{1}{\lambda^{*2}}$$

利用 $\lambda^* \le \sqrt{\lambda^{*2}+K_\nu^2}$ 以及 $\lambda^* \le \Delta_h$，一个充分条件是

$$2(2q-1)\left[\alpha_2\Delta_h^2+2\alpha_4\Delta_h^2+\frac{\alpha_1}{2\sqrt{\log 2}}\Delta_h+\frac{\alpha_3}{\sqrt{2\log 2}}\Delta_h\right]\le 1$$

这可以改写为

$$(\alpha_2+2\alpha_4)\Delta_h^2+\frac{\Delta_h}{\sqrt{2\log 2}}\Big(\frac{\alpha_1}{\sqrt{2}}+\alpha_3\Big)-\frac{1}{2(2q-1)}\le 0\,.$$

这是关于 Δ_h 的二次式，并利用引理 32 可得

$$\Delta_h \le \min\left\{\frac{\sqrt{2\log 2}}{4(2q-1)(\frac{\alpha_1}{\sqrt{2}}+\alpha_3)},\frac{1}{\sqrt{8(2q-1)(\alpha_2+2\alpha_4)}}\right\}.$$

要使其成立，一个充分条件是代入式（33）中的上界。代入这些值并利用 $s:=2q-1$ 可得

$$\Delta_h \le \min\left\{\frac{\sqrt{2\log 2}}{4s(\frac{\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_{xy}+\mathsf{p}_{xy}u)}{\sqrt{2}}+h\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x+\mathsf{p}_x u))},\frac{1}{\sqrt{8s(\mathsf{p}_{xy}^2(1+3su^2)+2h^2\mathsf{p}_x^2(1+3su^2))}}\right\}.$$

确保 （31）**成立，即** $4\,(2q-1)\,\eta_2 \le \frac{1}{K_\nu^2}$ 为确保满足此准则，我们要求

$$4(2q-1)\left(\alpha_4+\frac{\alpha_3}{2\sqrt{2(\lambda^{*2}+K_\nu^2)\log 2}}\right)\le\frac{1}{K_\nu^2}$$

这与以下条件相同

$$\frac{2\alpha_3(2q-1)K_\nu^2}{\sqrt{2\log 2}(1-4\alpha_4(2q-1)K_\nu^2)}\le\sqrt{\lambda^{*2}+K_\nu^2}$$

这由.....得以保证

$$\frac{2\alpha_3(2q-1)K_\nu^2}{\sqrt{2\log 2}(1-4\alpha_4(2q-1)K_\nu^2)}\le K_\nu\,.$$

重新整理可得

$$4\sqrt{\log 2}\alpha_4(2q-1)K_\nu^2+\sqrt{2}(2q-1)\alpha_3K_\nu-\sqrt{\log 2}\le 0\,.$$

要使其成立，一个充分条件是将（ α_3、 α_4 ）中的上界代入33。代入这些值并使用 $s:=2q-1$ 可得

$$4\sqrt{\log 2}\,h^2\mathsf{p}_x^2(1+3su^2)sK_\nu^2+\sqrt{2}sh\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x+\mathsf{p}_x u)K_\nu-\sqrt{\log 2}\le 0$$

这是关于 h 的二次函数。引理 32于是给出

$$h\le\min\left\{\frac{\sqrt{\log 2}}{2\sqrt{2}s\sqrt{d}\,(6\sqrt{d}\,\mathsf{j}_x+\mathsf{p}_x u)K_\nu},\frac{1}{4K_\nu\mathsf{p}_x\sqrt{(1+3su^2)s}}\right\}.$$

□

备注 8 (证明策略)。 定理 *4* 的证明也可以通过应用引理 *19*，从一开始就对 $\mathsf{R}_q(\mu\,\tilde{\mathbf{Q}}^k\,\tilde{\mathbf{P}}\ \|\ \nu)$ 进行界定来完成。相反，我们应用 *R´enyi* 散度的弱三角不等式，将混合时间保证所带来的贡献显式表示出来；这一保证通过推论 *7* 得出。这使得定理 *4* 中的指数衰减项与定理 *3*中的指数压缩项十分相似。

J 信息收缩

推论 *6* 的证明。设 (X_0, X_k) 的联合分布为 $(X_0, X_k) \sim \rho_{0,k}$，其边缘分布为 $X_0 \sim \rho_0 = \mu$ 和 $X_k \sim \rho_k = \mu\,\tilde{\mathbf{P}}^k_{T,h}$。根据定义 7，可得

$$\begin{aligned}
\mathsf{MI}(X_0; X_k) &= \mathsf{KL}(\rho_{0,k} \,\|\, \rho_0 \otimes \rho_k) \\
&= \mathop{\mathbb{E}}_{x\sim\rho_0}\left[\mathsf{KL}(\rho_{k|0=x} \,\|\, \rho_k)\right] \\
&= \mathop{\mathbb{E}}_{x\sim\rho_0}\left[\mathsf{KL}(\rho_{k|0=x} \,\|\, \tilde{\nu}_h)\right] - \mathsf{KL}(\rho_k \,\|\, \tilde{\nu}_h) \\
&\le \mathop{\mathbb{E}}_{x\sim\rho_0}\left[\mathsf{KL}(\rho_{k|0=x} \,\|\, \tilde{\nu}_h)\right] \\
&= \mathop{\mathbb{E}}_{x\sim\rho_0}\left[\mathsf{KL}(\delta_x\tilde{\mathbf{P}}^k_{T,h} \,\|\, \tilde{\nu}_h)\right] \\
&\le e^{-\frac{\alpha T^2}{5}(k-1)}\Big(\frac{9}{4T^2} + 20dM^2T^4\Big) \mathop{\mathbb{E}}_{x\sim\rho_0}\mathop{\mathbb{E}}_{y\sim\tilde{\nu}_h}\|x-y\|^2 \\
&= e^{-\frac{\alpha T^2}{5}(k-1)}\Big(\frac{9}{4T^2} + 20dM^2T^4\Big) \mathop{\mathbb{E}}_{(X,Y)\sim\rho_0\otimes\tilde{\nu}_h}\left[\|X-Y\|^2\right]
\end{aligned}$$

其中，第二个等号由 KL 散度的链式法则得到，第三个等号可通过展开两边验证，第一个不等式由 KL 散度的非负性得到，第二个不等式则由定理 2 和命题 1 得到。

□

K KL散度与R´enyi散度的凸性

以下引理有助于将正则化和交叉正则化保证从狄拉克初始化推广到一般初始化。设定 $\mathbf{P} = \mathbf{Q}$ 有助于获得正则化保证（引理 12 和 17），进而得到混合时间结果；而 $\mathbf{P} \neq \mathbf{Q}$ 用于推导交叉正则化保证（引理 14 和 19），进而得到渐近偏差界。以下关于 $\mathbf{P} = \mathbf{Q}$ 的情形的引理见 [AC24b，定理 3。7]。

引理 29。 令 μ、$\pi \in \mathcal{P}(\mathbb{R}^d)$。则对于任意两个马尔可夫核 $\mathbf{P}$ 和 $\mathbf{Q}$

$$\mathsf{KL}(\mu\mathbf{P} \,\|\, \pi\mathbf{Q}) = \mathsf{KL}(\mathop{\mathbb{E}}_{x\sim\mu}[\delta_x\mathbf{P}] \,\|\, \mathop{\mathbb{E}}_{y\sim\pi}[\delta_y\mathbf{Q}]) \le \inf_{\gamma\in\Gamma(\mu,\pi)} \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \mathsf{KL}(\delta_x\mathbf{P} \,\|\, \delta_y\mathbf{Q}).$$

证明。 根据定义，等式成立，其中 $\mu\mathbf{P}$ 和 $\pi\mathbf{Q}$。为证明不等式，任取一个耦合 $\gamma \in \Gamma(\mu, \pi)$。由于 KL 散度具有联合凸性，不等式对 γ 成立。由于该不等式对任意耦合均成立，对所有耦合取下确界即可得出结论。 □

引理 30。 设 $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$。则对于 $q > 1$ 以及任意两个马尔可夫核 $\mathbf{P}$ 和 $\mathbf{Q}$

$$\mathsf{R}_q(\mu\mathbf{P} \,\|\, \pi\mathbf{Q}) \le \inf_{\gamma \in \Gamma(\mu,\pi)} \frac{1}{q-1} \log \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \left[e^{(q-1)\mathsf{R}_q(\delta_x\mathbf{P} \| \delta_y\mathbf{Q})} \right] .$$

证明。 对于分布 $\rho, \nu \in \mathcal{P}(\mathbb{R}^d)$，定义

$$F_q(\rho \,\|\, \nu) := \mathop{\mathbb{E}}_{\nu} \left[\left(\frac{\rho}{\nu} \right)^q \right] = \int \rho(z)^q \nu(z)^{1-q} \, dz \,,$$

因此，R´enyi 散度（定义 5）为 $\mathsf{R}_q(\rho \,\|\, \nu) = \frac{1}{q-1} \log F_q(\rho \,\|\, \nu)$。注意，泛函$F_q(\rho \,\|\, \nu)$ 关于（ρ, ν）联合凸，因为函数 $g(u, v) = u^q v^{1-q}$ 对于 $u, v \ge 0$, $q > 1$ 是联合凸的。

取一个耦合 $\gamma \in \Gamma(\mu, \pi)$。由于 F_q 是联合凸的，我们有

$$F_q(\mu\mathbf{P} \,\|\, \pi\mathbf{Q}) \le \mathop{\mathbb{E}}_{(x,y)\sim\gamma} [F_q(\delta_x\mathbf{P} \,\|\, \delta_q\mathbf{Q})] \,.$$

根据 F_q 的定义，我们得到

$$\exp((q-1)\mathsf{R}_q(\mu\mathbf{P} \,\|\, \pi\mathbf{Q})) \le \mathop{\mathbb{E}}_{(x,y)\sim\gamma} [\exp((q-1)\mathsf{R}_q(\delta_x\mathbf{P} \,\|\, \delta_y\mathbf{Q}))] \,.$$

两边取对数并除以 $q-1$，得到

$$\mathsf{R}_q(\mu\mathbf{P} \,\|\, \pi\mathbf{Q}) \le \frac{1}{q-1} \log \mathop{\mathbb{E}}_{(x,y)\sim\gamma} \left[e^{(q-1)\mathsf{R}_q(\delta_x\mathbf{P} \| \delta_y\mathbf{Q})} \right] .$$

由于该不等式对任意耦合 $\gamma \in \Gamma(\mu, \pi)$ 都成立，因此它对耦合的下确界也成立，从而得到该引理。

□

通过uLA进行L 交叉正则化

未校正朗之万算法（uLA） [RT96、RR98、Dal17、DM19、VW23] 用于从目标分布 $\nu \propto e^{-f}$ 中 $\mathbb{R}^d$采样，从 $X_0 \sim \mu$开始，其定义如下更新规则

$$X_{k+1} = X_k - \eta \nabla f(X_k) + \sqrt{2\eta}\, \xi_k$$

其中， $k \ge 0$ 是整数， $\eta > 0$ 是步长，并且对所有 $k \ge 0$都有 $\xi_k \sim \boldsymbol{\gamma}_d$ 。当 $\eta \to 0$时，我们得到由下式给出的朗之万动力学

$$dX_t = -\nabla f(X_t)\, dt + \sqrt{2}\, dW_t$$

其中， W_t 是标准布朗运动。朗之万动力学以目标分布 ν 作为˜其平稳分布，我们将uLA的平稳分布记为 $\tilde{\nu}_\eta$。令$\mathbf{P}_\eta$ 表示uLA执行一步的转移核，并令$\mathbf{P}_\eta$ 表示朗之万
˜ ˜动力学运行时间 η的转移核。回顾一下，$\mathbf{P}_{T,\ h}$ 表示uHMC-v转移核，$\mathbf{Q}_{T,\ h}$ 表示uHMC-s转移核，而eHMC的核为$\mathbf{P}_T$。

我们的目标是获得两步马尔可夫链的渐近偏差界，其中 Wasserstein 收缩由 uHMC 驱动，交叉正则化由 uLA 提供（见第 7节），并且

研究相应的梯度复杂度，以输出一个样本 X，使得 $\mathsf{D}(\mathcal{L}(X)\,\|\,\nu) \le \epsilon$，对于强对数凹目标分布 ν、精度 $\epsilon > 0$ 以及散度 $\mathsf{D} \in \{\mathsf{KL}, \mathsf{R}_q\}$。由于我们希望研究最终梯度复杂度对维度 d 和精度 ϵ的依赖关系，因此我们简化后续分析，仅纳入相关参数的依赖关系，即维度 d、uLA步长 η和uHMC步长 h。我们还使用 $\gtrsim$, $\lesssim$ 来隐藏绝对常数。

L.1 KL 散度中的渐近偏差

在 [AC24c 的引理 4 中，给出了 uLA 的一个交叉正则化界。7]他们证明，在假设 1(b) 成立且 $\eta \lesssim \frac{1}{L}$ 时，对所有 x，以下结论均成立：$y \in \mathbb{R}^d$

$$\mathsf{KL}(\delta_x \tilde{\mathbf{P}}_\eta \,\|\, \delta_y \mathbf{P}_\eta) \lesssim \frac{\|x-y\|^2}{\eta} + \eta^3 \|\nabla f(x)\|^2 + \eta^2 d\,.$$

对最优 Wasserstein–2耦合下、分别来自两个分布 x 和 y 的（ μ， $\pi \in \mathcal{P}$（$\mathbb{R}^d$）取期望，并应用引理 29 可得

$$\mathsf{KL}(\mu \tilde{\mathbf{P}}_\eta \,\|\, \pi \mathbf{P}_\eta) \lesssim \frac{W_2^2(\mu,\pi)}{\eta} + \eta^3 \mathop{\mathbb{E}}_{\mu}[\|\nabla f(x)\|^2] + \eta^2 d\,, \tag{34}$$

从而得到 uLA 对应的引理 14 。与定理 2 的证明类似，接下来我们将结合 Wasserstein–2 混合和 Wasserstein–2 偏差保证。我们分别讨论 uHMC-v 和 uHMC-s，并假设目标分布具有强对数凹性。

由 uHMC-v 驱动的 Wasserstein 收缩性。取 $\mu = \rho \tilde{\mathbf{P}}_{T,h}^k$ 以及 $\pi = \nu$ (对于初始分布 ρ 和目标分布 ν) 代入 (34)，再结合三角不等式可得

$$\mathsf{KL}(\rho \tilde{\mathbf{P}}_{T,h}^k \tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \frac{W_2^2(\rho \tilde{\mathbf{P}}_{T,h}^k, \tilde{\nu}_h)}{\eta} + \frac{W_2^2(\tilde{\nu}_h, \nu)}{\eta} + \eta^3 \mathop{\mathbb{E}}_{\rho \tilde{\mathbf{P}}_{T,h}^k}[\|\nabla f(x)\|^2] + \eta^2 d\,.$$

命题1和2，结合$\mathbb{E}_{\rho \tilde{\mathbf{P}}_{T,h}^k}[\|\nabla f(x)\|^2] = O(d)$（参见 [AC24c，引理 C.3]），得到以下结果

$$\mathsf{KL}(\rho \tilde{\mathbf{P}}_{T,h}^k \tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \frac{e^{-k} W_2^2(\rho, \tilde{\nu}_h)}{\eta} + \frac{h^4 d^2}{\eta} + \eta^3 d + \eta^2 d\,.$$

为使右侧达到所需精度 ϵ，取 $h \asymp \epsilon^{3/8} d^{-5/8}$ 和 $\eta \asymp \sqrt{\epsilon/d}$，结合 $k = O(\log(d/\epsilon))$ 次迭代，可得二阶预言机复杂度为$O(d^{5/8}\epsilon^{-3/8}\log(d/\epsilon))$. 结果见表1，第8行。

由 uHMC-s 驱动的 Wasserstein 收缩性。我们首先取 $\mu = \rho \tilde{\mathbf{P}}_{T,h}^k$ 以及 $\pi = \nu$（对于初始分布 ρ 和目标分布 ν），在（34）中，结合三角不等式可得

$$\mathsf{KL}(\rho \tilde{\mathbf{Q}}_{T,h}^k \tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \frac{W_2^2(\rho \tilde{\mathbf{Q}}_{T,h}^k, \tilde{\nu}_{\tilde{\mathbf{Q}}_{T,h}})}{\eta} + \frac{W_2^2(\tilde{\nu}_{\tilde{\mathbf{Q}}_{T,h}}, \nu)}{\eta} + \eta^3 \mathop{\mathbb{E}}_{\rho \tilde{\mathbf{Q}}_{T,h}^k}[\|\nabla f(x)\|^2] + \eta^2 d\,.$$

命题5和6以及 [AC24c，引理 C。3]可得

$$\mathsf{KL}(\rho\tilde{\mathbf{P}}_{T,h}^{k}\tilde{\mathbf{P}}_{\eta}\,\|\,\nu)\lesssim\frac{e^{-k}W_{2}^{2}(\rho,\tilde{\nu}_{\tilde{\mathbf{Q}}_{T,h}})}{\eta}+\frac{h^{3}d}{\eta}+\eta^{3}d+\eta^{2}d\,.$$

给定所需精度 ϵ，取 $h\asymp\sqrt{\epsilon/d}$，$\eta\asymp\overline{\sqrt{}}\epsilon/d$，当迭代 $k=O(\log(d/\epsilon))$ 次时，预言机复杂度为 $O(d^{1/2}\epsilon^{-1/2}\log(d/\epsilon))$。这对应于表 1第 9 行。

L.2 R´ényi 散度中的渐近偏差

AC24b 中描述了 uLA 在 R´ényi 散度下的一种正则化性质，见 [AC24b，(3.14)]，，其中证明了对所有 x，$y\in\mathbb{R}^{d}$

$$\mathsf{R}_{q}(\delta_{x}\tilde{\mathbf{P}}_{\eta}\,\|\,\delta_{y}\tilde{\mathbf{P}}_{\eta})\lesssim\frac{\|x-y\|^{2}}{\eta}\,.$$

应用引理 30 可得，对任意 μ，$\pi\in\mathcal{P}(\mathbb{R}^{d})$，

$$\mathsf{R}_{q}(\mu\tilde{\mathbf{P}}_{\eta}\,\|\,\pi\tilde{\mathbf{P}}_{\eta})\lesssim\inf_{\gamma\in\Gamma(\mu,\pi)}\log\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\|x-y\|^{2}}{\eta}}\right].$$

将此应用于 $\mu=\rho\,\tilde{}\mathbf{P}_{T,\,h}^{k}$（其中 ρ 是某个初始分布）以及 $\pi=\tilde{\nu}_{h}$（其中 $\tilde{\nu}_{h}$ 是 uHMC-v 的平稳分布），可得

$$\mathsf{R}_{q}(\rho\tilde{\mathbf{P}}_{T,h}^{k}\tilde{\mathbf{P}}_{\eta}\,\|\,\tilde{\nu}_{h}\tilde{\mathbf{P}}_{\eta})\lesssim\inf_{\gamma\in\Gamma(\rho\tilde{\mathbf{P}}_{T,h}^{k},\tilde{\nu}_{h})}\log\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\|x-y\|^{2}}{\eta}}\right]=\log\inf_{\gamma\in\Gamma(\rho\tilde{\mathbf{P}}_{T,h}^{k},\tilde{\nu}_{h})}\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\|x-y\|^{2}}{\eta}}\right].$$

在 uHMC-v 满足 Orlicz-Wasserstein 混合假设（命题 3）的条件下，并按照定理 3证明中的类似计算可得

$$\mathsf{R}_{q}(\rho\tilde{\mathbf{P}}_{T,h}^{k}\tilde{\mathbf{P}}_{\eta}\,\|\,\tilde{\nu}_{h}\tilde{\mathbf{P}}_{\eta})\lesssim\frac{e^{-k}\,W_{\psi}(\rho,\tilde{\nu}_{h})}{\eta}\tag{35}$$

只要 $k\gtrsim$ 对数 $^{W_{\psi}(\rho,\,\tilde{\nu}_{h})\,\eta}$。这给出了推论7 的对应结果，适用于 uLA 交叉正则化，并特化为由 uHMC-v 驱动的 Wasserstein 收缩。

现在我们要证明定理4 在我们设定下的对应结果，并首先陈述 uLA 在 R´Rényi 散度 [AC24c，引理 C。2], 其中他们证明，在假设1(b) 成立且对于 $\eta\lesssim\frac{1}{L}$，对所有 x，$y\in\mathbb{R}^{d}$，

$$\mathsf{R}_{q}(\delta_{x}\tilde{\mathbf{P}}_{\eta}\,\|\,\delta_{y}\mathbf{P}_{\eta})\lesssim\frac{\|x-y\|^{2}}{\eta}+\eta^{3}\|\nabla f(x)\|^{2}+\eta^{2}d\,.\tag{36}$$

应用 Rényi 散度的弱三角不等式（3），我们有

$$\mathsf{R}_{q}(\rho\tilde{\mathbf{P}}_{T,h}^{k}\tilde{\mathbf{P}}_{\eta}\,\|\,\nu)\lesssim\mathsf{R}_{2q}(\rho\tilde{\mathbf{P}}_{T,h}^{k}\tilde{\mathbf{P}}_{\eta}\,\|\,\tilde{\nu}_{h}\tilde{\mathbf{P}}_{\eta})+\mathsf{R}_{2q}(\tilde{\nu}_{h}\tilde{\mathbf{P}}_{\eta}\,\|\,\nu\mathbf{P}_{\eta})\,.$$

右侧第一项由（35）控制，而对于第二项，我们将从（36）出发。为此，结合引理 30和（36）可得

$$\mathsf{R}_{2q}(\tilde{\nu}_{h}\tilde{\mathbf{P}}_{\eta}\,\|\,\nu)\lesssim\inf_{\gamma\in\Gamma(\tilde{\nu}_{h},\nu)}\log\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\|x-y\|^{2}}{\eta}+\eta^{3}\|\nabla f(x)\|^{2}+\eta^{2}d}\right]$$

假设1（a）–（b）使我们可以将其界定为

$$\mathsf{R}_{2q}(\tilde{\nu}_h\tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \inf_{\gamma\in\Gamma(\tilde{\nu}_h,\nu)} \log \mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\frac{\|x-y\|^2}{\eta}+\eta^3\|x\|^2+\eta^2 d}\right]$$

可进一步简化为

$$\mathsf{R}_{2q}(\tilde{\nu}_h\tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \eta^2 d + \log \inf_{\gamma\in\Gamma(\tilde{\nu}_h,\nu)} \mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{\|x-y\|^2(\frac{1}{\eta}+2\eta^3)+2\|y\|^2\eta^3}\right].$$

利用 H¨older 不等式，我们有

$$\mathsf{R}_{2q}(\tilde{\nu}_h\tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \eta^2 d + \log \inf_{\gamma\in\Gamma(\tilde{\nu}_h,\nu)} \underbrace{\left(\mathop{\mathbb{E}}_{(x,y)\sim\gamma}\left[e^{2\|x-y\|^2(\frac{1}{\eta}+2\eta^3)}\right]\right)^{1/2}}_{\mathrm{I}} \underbrace{\left(\mathop{\mathbb{E}}_{y\sim\nu}\left[e^{4\|y\|^2\eta^3}\right]\right)^{1/2}}_{\mathrm{II}}.$$

按照定理 4 证明中处理 I 和 II的类似步骤，在步长满足适当假设的条件下，并利用 uHMC-v 的奥尔利奇-瓦瑟斯坦偏差保证（命题 4），我们得到以下结果

$$\mathsf{R}_{2q}(\tilde{\nu}_h\tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \eta^2 d + \Big(\frac{1}{\eta}+\eta^3\Big)h^2 d + \eta^3 K_\nu^2 .$$

考虑 $K_\nu = O(\sqrt{d})$，因此我们得到

$$\mathsf{R}_{2q}(\tilde{\nu}_h\tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \eta^2 d + \Big(\frac{1}{\eta}+\eta^3\Big)h^2 d + \eta^3 d .$$

由此可得整体渐近偏差保证为

$$\mathsf{R}_q(\rho\tilde{\mathbf{P}}_{T,h}^k\tilde{\mathbf{P}}_\eta \,\|\, \nu) \lesssim \frac{e^{-k}\, W_\psi(\rho,\tilde{\nu}_h)}{\eta} + \eta^2 d + \Big(\frac{1}{\eta}+\eta^3\Big)h^2 d + \eta^3 d .$$

为使总误差不超过精度 ϵ，我们设 $\eta \asymp (\epsilon/d)^{1/2}$ 和 $h \asymp (\epsilon/d)^{3/4}$，从而得到整体梯度复杂度 $O(d^{3/4}\epsilon^{-3/4}\log(d/\epsilon))$，如表 1第 10 行所示。

M 辅助引理

引理 31. 令 $V \sim \boldsymbol{\gamma}_d$，并设 $m_d := \mathbb{E}\,\|V\|$。那么对于任意 $c \geq 0$,

$$\mathbb{E}\left[e^{c\|V\|}\right] \leq \exp\left(c\, m_d + \frac{c^2}{2}\right).$$

特别地，由于 $m_d \leq \sqrt{d}$,

$$\mathbb{E}\left[e^{c\|V\|}\right] \leq \exp\left(c\sqrt{d} + \frac{c^2}{2}\right).$$

证明。函数 $f(v) = \|v\|$ 是 1-Lipschitz 的。对于标准高斯向量 $V \sim \boldsymbol{\gamma}_d$，高斯集中不等式 [Led01] 指出，对于任意 $c \in \mathbb{R}$,

$$\mathbb{E}\exp\big(c(f(V) - \mathbb{E}\, f(V))\big) \leq \exp\left(\frac{c^2}{2}\right).$$

两边乘以 $\exp(c\,\mathbb{E}\,f(V))$，并利用 $f(V) = \|V\|$，得到

$$\mathbb{E}\, e^{c\|V\|} \le \exp\left(c\,\mathbb{E}\,\|V\| + \frac{c^2}{2}\right).$$

最后，由詹森不等式可得 $\mathbb{E}\,\|V\| \le (\mathbb{E}\,\|V\|^2)^{1/2} = \sqrt{d}$。 □

引理 32. 设 $f(x) = ax^2 + bx + c$ 是定义在 $x \in \mathbb{R}_{\ge 0}$ 上的二次函数，其中 $a > 0, b > 0, c < 0$。那么，对于 $f(x) \le 0$，只需满足

$$x \le \min\left\{\frac{-c}{2b}, \sqrt{\frac{-c}{4a}}\right\}.$$

证明。假设

$$x \le \min\left\{\frac{-c}{2b}, \sqrt{\frac{-c}{4a}}\right\}.$$

Then $bx \le -\frac{c}{2}$且 $ax^2 \le -\frac{c}{4}$。因此 $f(x) = ax^2 + bx + c \le \frac{c}{4}$，这与 $c < 0$ 一致，因为 ≤ 0 . □

参考文献

[ABT24] Jason M Altschuler、Jinho Bok 和 Kunal Talwar。凸优化中带噪随机梯度下降的隐私性。《SIAM 计算杂志》，53(4):969– 1001，2024年。[AC24a] Jason M Altschuler 和 Sinho Chewi。通过算法热启动实现更快的高精度对数凹采样。《ACM 杂志》，71(3):1–55，2024年。[AC24b] Jason M Altschuler 和 Sinho Chewi。移位组合 I：Harnack 不等式与反向输运不等式。《IEEE 信息论汇刊》，2024年。[AC24c] Jason M Altschuler 和 Sinho Chewi。移位组合 III：KL 散度的局部误差框架。《arXiv 预印本 arXiv:2412.17997》，2024年。[ACZ25] Jason M Altschuler、Sinho Chewi 和 Matthew S Zhang。移位组合 IV：欠阻尼 Langevin 与具有部分加速的数值离散化方法。《arXiv 预印本 arXiv:2506.23062》，2025年。[AG07] Søren Asmussen 和 Peter W. Glynn。《随机模拟：算法与分析》。Springer，纽约，2007年。[AGS24] Simon Apers、Sander Gribling 和 D´aniel Szil´agyi。高效高斯采样的哈密顿蒙特卡罗：长步长与随机步长。《Journal of Machine Learning Research》，25(348):1–30，2024年。[ATW06] Marc Arnaudon、Anton Thalmaier 和 Feng-Yu Wang。曲率下界无界的流形上的 Harnack 不等式与热核估计。《数学科学通报》，130(3):223–233，2006年。

[BE21] Fabrice Baudoin 和 Nathaniel Eldredge。马尔可夫核的输运不等式及其应用。Electronic Journal of Probability，26:1–39，2021年。[BFMG25] Mario Bravo、Juan P Flores-Mella 和 Crist´obal Guzm´an。连续模下投影朗之万算法的混合时间与隐私分析。arXiv 预印本 arXiv:2501.04134，2025年。[BGL01] Sergey G Bobkov、Ivan Gentil 和 Michel Ledoux。哈密顿–雅可比方程的超收缩性。《纯粹与应用数学杂志》，80(7):669–696，2001年。[BGL15] Dominique Bakry、Ivan Gentil 和 Michel Ledoux。关于 Harnack 不等式与最优输运。比萨高等师范学院科学分部年刊，14:705–727，2015年。[BRE23] Nawaf Bou-Rabee 和 Andreas Eberle。未校正哈密顿蒙特卡罗的混合时间保证。《伯努利》，29(1):75–104，2023年。[BREZ20] Nawaf Bou-Rabee、Andreas Eberle 和 Raphael Zimmer。哈密顿蒙特卡罗的耦合与收敛。《应用概率年刊》，30(3):1209–1250，2020年。[BRM25] Nawaf Bou-Rabee 和 Milo Marsden。采用分层蒙特卡罗时间积分的未校正哈密顿马尔可夫链蒙特卡罗。《应用概率年刊》，35(1):360–392，2025年。[BRO24] Nawaf Bou-Rabee 和 Stefan Oberd¨orster。无转向采样器的混合与高斯集中几何。arXiv 预印本 arXiv:2410.06978，2024年。[BRS23] Nawaf Bou-Rabee 和 Katharina Schuh。平均场模型中未校正哈密顿蒙特卡罗的收敛。Electronic Journal of Probability，28:1–40，2023年。[BRSS18] Nawaf Bou-Rabee 和 Jes´us Maria Sanz-Serna。几何积分器与哈密顿蒙特卡罗方法。Acta Numerica，27:113–206，2018年。[CCSW22] Yongxin Chen、Sinho Chewi、Adil Salim 和 Andre Wibisono。采样近端算法的改进分析。载于《学习理论会议》，第2984–3014页。PMLR，2022年。[CDMS23] Evan Camrud、Alain Durmus、Pierre Monmarch´e 和 Gabriel Stoltz。未校正广义哈密顿蒙特卡罗的二阶定量界。arXiv 预印本 arXiv:2306.09513，2023年。[CDWY20] Yuansi Chen、Raaz Dwivedi、Martin J Wainwright 和 Bin Yu。梅特罗波利斯化哈密顿蒙特卡罗的快速混合：多步梯度的优势。Journal of Machine Learning Research，21(92):1–72，2020年。[CEL+24] Sinho Chewi、Murat A Erdogdu、Mufan Li、Ruoqi Shen 和 Matthew S Zhang。从庞加莱不等式到对数索博列夫不等式的朗之万蒙特卡罗分析。计算数学基础，页码1–51，2024年。[CG23] Yuansi Chen 和 Khashayar Gatmiry。梅特罗波利斯化哈密顿蒙特卡罗何时能够被证明优于梅特罗波利斯校正朗之万算法？arXiv 预印本 arXiv:2304.04724，2023年。

[CGH+17] Bob Carpenter、Andrew Gelman、Matthew D Hoffman、Daniel Lee、Ben Goodrich、Michael Betancourt、Marcus Brubaker、Jiqiang Guo、Peter Li 和 Allen Riddell。Stan：一种概率编程语言。《统计软件杂志》，76:1–32，2017年。[Cha04] Djalil Chafa¨ı。《熵、凸性与函数不等式：关于 Φ-熵与Φ-Sobolev不等式》。《京都大学数学杂志》，2004年。[Che24] Sinho Chewi。《对数凹采样》，2024年。草稿见：https://chewisinho.github.io。[CV19] Zongchen Chen 和 Santosh Vempala。强对数凹分布的哈密顿蒙特卡罗最优收敛率。载于 APPROX-RANDOM，2019。[CYS21] Rishav Chourasia、Jiayuan Ye 和 Reza Shokri。朗之万扩散与噪声梯度下降的差分隐私动力学。《神经信息处理系统进展》，34:14771–14781，2021年。[Dal17] Arnak Dalalyan。采样与优化之间更进一步且更强的类比：朗之万蒙特卡罗与梯度下降。载于学习理论会议，第678–689页。PMLR，2017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。Journal of Machine Learning Research，20(183):1–42，2019。[DE24] Alain Durmus 和 Andreas Eberle。高维非精确马尔可夫链蒙特卡罗方法的渐近偏差。《应用概率年刊》，34(4):3435–3468，2024年。[DKPR87] Simon Duane、Anthony D Kennedy、Brian J Pendleton 和 Duncan Roweth。混合蒙特卡罗。《物理快报B》，195(2):216–222，1987。[DM04] Pierre Del Moral。《费曼–卡茨公式：具有应用的谱系与相互作用粒子系统》。《概率及其应用》。Springer，2004年。[DM19] Alain Durmus 和 ´Eric Moulines。通过未校正朗之万算法进行高维贝叶斯推断。《伯努利》，25(4A)，2019。[DMPS18] Randal Douc、 ´Eric Moulines、Pierre Priouret 和 Philippe Soulier。《马尔可夫链》。《施普林格运筹学与金融工程丛书》。Springer，Cham，2018年。[FMTT18] Vitaly Feldman、Ilya Mironov、Kunal Talwar 和 Abhradeep Thakurta。通过迭代实现隐私放大。载于2018年IEEE第59届计算机科学基础年会（FOCS），第521–532页。IEEE，2018年。[GHS15] Andrew Golightly、Daniel A Henderson 和 Chris Sherlock。用于随机动力学模型精确推断的延迟接受粒子马尔可夫链蒙特卡罗。《统计与计算》，25(5):1039–1055，2015年。

[GT20] Arun Ganesh 和 Kunal Talwar。通过对离散化朗之万马尔可夫链蒙特卡罗进行 Rényi 散度分析，实现更快速的差分隐私采样器。《神经信息处理系统进展》，33:7222–7233，2020年。[Has70] W. Keith Hastings。使用马尔可夫链的蒙特卡罗采样方法及其应用。Biometrika，57(1):97–109，1970年。[HG14] Matthew D Hoffman 和 Andrew Gelman。无转向采样器：在哈密顿蒙特卡罗中自适应设定路径长度。机器学习研究杂志，15(1):1593–1623，2014年。[HHIL06] Ernst Hairer、Marlis Hochbruck、Arieh Iserles 和 Christian Lubich。几何数值积分。Oberwolfach报告，3(1):805–882，2006年。[HSVW05] Martin Hairer、Andrew M. Stuart、Jochen Voss 和 Petter Wiberg。路径采样中产生的随机偏微分方程分析。第一部分：高斯情形。Communications in Mathematical Sciences，3(4):587–603，2005年。[KV25a] Yunbum Kook 和 Santosh S Vempala。高维冷启动下更快速的对数凹采样。arXiv预印本 arXiv:2505.01937，2025年。[KV25b] Yunbum Kook 和 Santosh S Vempala。通过算法扩散对对数凹函数进行采样与积分。载于第57届ACM计算理论研讨会年会论文集，第924–932页，2025年。[Led01] Michel Ledoux。《测度集中现象》，第89卷《数学调查与专著》。美国数学学会，2001年。[Liu08] Jun S. Liu。《科学计算中的蒙特卡罗策略》。Springer，纽约，第2版，2008年。[LMW25] Jiaming Liang、Siddharth Mitra 和 Andre Wibisono。通过 Φ-互信息的压缩刻画朗之万动力学及其算法中样本的依赖关系。载于Nika Haghtalab和Ankur Moitra（编），第三十八届学习理论会议论文集，《机器学习研究论文集》。PMLR，2025年。[LST21] Yin Tat Lee、Ruoqi Shen 和 Kevin Tian。利用受限高斯预言机进行结构化对数凹采样。载于学习理论会议，第2993–3050页。PMLR，2021年。[Mar25] Milo Steven Marsden。《哈密顿蒙特卡罗的随机化积分器与自适应步长选择》。博士论文，斯坦福大学，2025年6月。[Mir17] Ilya Mironov。Rényi差分隐私。载于2017年IEEE第30届计算机安全基础研讨会（CSF），第263–275页。IEEE，2017年。[MM21] Takeru Matsuda 和 Yuto Miyatake。带离散化误差量化的常微分方程模型估计。SIAM/ASA不确定性量化杂志，9(1):302–331，2021年。

[Mon21] Pierre Monmarch´e. 欠阻尼朗之万扩散的标准分裂格式高维马尔可夫链蒙特卡罗。《电子统计学杂志》，15(2):4117–4166，2021年。[MRR+53] Nicholas Metropolis、Arianna W Rosenbluth、Marshall N Rosenbluth、Augusta H Teller 和 Edward Teller。高速计算机进行状态方程计算。《化学物理学杂志》，21(6):1087–1092，1953。[MS10] Neal Madras 和 Deniz Sezer。马尔可夫链收敛的定量界：Wasserstein 距离和全变差距离。《伯努利》，16(1):882–908，2010。[MS21] Oren Mangoubi 和 Aaron Smith。强对数凹分布上的哈密顿蒙特卡罗混合：连续动力学。《应用概率年刊》，31(5):2019–2045，2021年。[MSH02] Jonathan C. Mattingly、Andrew M. Stuart 和 Desmond J. Higham。随机微分方程及其近似的遍历性：局部利普希茨向量场与退化噪声。《随机过程及其应用》，101(2):185–232，2002。[MW25] Siddharth Mitra 和 Andre Wibisono。未校正朗之万算法与近端采样器中 Φ-散度的快速收敛。载于第36届国际算法学习理论会议，2025年。[Nea11] Radford M Neal。利用哈密顿动力学进行马尔可夫链蒙特卡罗。《马尔可夫链蒙特卡罗手册》，2(11):2，2011。[PHF10] Anand Patil、David Huard 和 Christopher J Fonnesbeck。PyMC：用于 Python 的贝叶斯随机建模。《统计软件杂志》，35:1–81，2010。[PW24] Yury Polyanskiy 和 Yihong Wu。《信息论：从编码到学习》。剑桥大学出版社，2024年。草稿见：http://www.stat.yale.edu/~yw562/teaching/itbook-export.pdf。[RC04] Christian P. Robert 和 George Casella。《蒙特卡罗统计方法》。施普林格统计学丛书。Springer，纽约，第2版，2004年。[RR98] Gareth O Roberts 和 Jeffrey S Rosenthal。朗之万扩散离散近似的最优尺度。《英国皇家统计学会杂志：B辑（统计方法学）》，60(1):255–268，1998。[RR02] Gareth O Roberts 和 Jeffrey S Rosenthal。特定随机递归序列的单次耦合。《随机过程及其应用》，99(2):195–208，2002。[RT96] Gareth O. Roberts 和 Richard L. Tweedie。朗之万分布及其离散近似的指数收敛。《伯努利》，1996。[Rum18] Siegfried M Rump。扰动单位矩阵行列式的估计。《线性代数及其应用》，558:101–107，2018年。[SMP19] Georgia K Stuart、Susan E Minkoff 和 Felipe Pereira。一种用于地震反演与不确定性量化的两阶段马尔可夫链蒙特卡罗方法。《地球物理学》，84

(6): R1003–R1020 ，2019。

[STG21] Chris Sherlock、Alexandre H Thiery 和 Andrew Golightly。延迟接受随机游走 Metropolis 算法的效率。统计年刊，49(5)：2972–2990，2021年。[Stu10] Andrew M. Stuart。反问题：贝叶斯视角。*Acta Numerica*，19：451–559，2010年。[SWW24] Vishwak Srinivasan、Andre Wibisono 和 Ashia Wilson。使用 Metropolis 调整镜像朗之万算法从约束空间快速采样。载于第三十七届学习理论年会，第4593–4635页。PMLR，2024年。[SWW25] Vishwak Srinivasan、Andre Wibisono 和 Ashia C Wilson。使用 Metropolis 调整预条件朗之万算法从约束空间进行高精度采样。载于第三十六届国际算法学习理论会议，2025年。[Szn91] Alain-Sol Sznitman。混沌传播专题。数学讲义，1464：165–251，1991年。圣弗卢尔概率论暑期学校第十九届，1989年。[VEH14] Tim Van Erven 和 Peter Harremos。Rényi 散度与 Kullback-Leibler 散度。*IEEE* 信息论汇刊，60(7)：3797–3820，2014年。[Ver18] Roman Vershynin。高维概率：数据科学应用导论，第47卷。剑桥大学出版社，2018年。[VW23] Santosh S. Vempala 和 Andre Wibisono。未校正朗之万算法的快速收敛：等周性质已足够。载于泛函分析的几何方面，第381–438页。Springer，2023年。[Wan97] Feng-Yu Wang。非紧致黎曼流形上的对数 Sobolev 不等式。概率论及相关领域，109(3)：417–424，1997年。