机器学习研究论文集 1–29

通过景观修改改进随机优化的朗之万蒙特卡洛方法

Michael C.H. Choi
MCHCHOI@NUS.EDU.SG统计与数据科学系和耶鲁-新加坡国立大学学院，新加坡国立大学，新加坡

Youjia Wang
191840250@SMAIL.NJU.EDU.CN
中国南京大学数学系

arXiv:2302.03973v1 [math.PR] 8 Feb 2023

摘要

对于需要最小化的目标函数 H，或在低温下需要采样的目标吉布斯分布 $\pi_\beta^0 \propto e^{-\beta H}$，本文提出并分析了在由$H^f_{\beta,c,1}$ 指定的替代景观上运行、并以修正吉布斯分布$\pi^f_{\beta,c,1} \propto e^{-\beta H^f_{\beta,c,1}}$为目标的朗之万蒙特卡洛（LMC）算法。其中，$H^f_{\beta,c,1}$ 的景观是H 景观经过变换后的版本，具体取决于参数 f、β 和 c。在低温情形下，与 π_β^0 相关联的原始对数 Sobolev 常数对 β 和能垒 M 均呈指数依赖。对于 H，在适当调整这些参数并满足相关假设后，我们证明变换后景观的能垒有所降低，从而使与$\pi^f_{\beta,c,1}$相关联的修正对数索博列夫常数对 β 和 M 均呈多项式依赖。这带来了更优的全变差混合时间界，并改善了向 H全局最小值的收敛。需要强调的是，本文提出的技术并非仅适用于朗之万蒙特卡洛，也广泛适用于其他基于梯度的优化或采样算法。

关键词：朗之万蒙特卡洛；景观修改；随机优化；谱隙；对数-索博列夫常数；亚稳态；泛函不等式

致谢

第一作者感谢新加坡国立大学和耶鲁-新加坡国立大学学院提供的启动经费支持。他还感谢教育部一级资助项目“科学数据与数据科学”合作计划提供的支持。

1. 引言

本文主要关注设计改进的朗之万蒙特卡洛算法，用于低温区间的采样或随机优化。这些任务自然出现在众多与机器学习广泛相关的领域中，包括但不限于贝叶斯计算、计算物理与化学，以及理论计算机科学。具体而言，给定一个可能非凸的目标函数 $H: \mathbb{R}^d \to \mathbb{R}$ 需要对其进行最小化，我们关注从相应的吉布斯分布中进行（近似）采样 $\pi_\beta^0 \propto e^{-\beta H}$，其中 β 是逆温度参数。在温度足够低且对 H满足适当假设的情况下，由于拉普拉斯方法的作用，吉布斯分布 π_β^0 会集中于 H 的全局极小值集合（Hwang 1980）；因此，从 π_β^0 采样有助于找到 H的近似全局极小值点。

© M.C. Choi &Y. Wang.

在各种为从 π_β^0 中采样而开发的随机算法中，近年来，基于朗之万蒙特卡洛（LMC）算法的算法重新受到关注。直观地说，LMC 可以理解为一种加入高斯噪声扰动的梯度下降方法。它是过阻尼朗之万扩散的欧拉–丸山离散化，其迭代形式如下：

$$X_{k+1} = X_k - \eta \nabla H(X_k) + \sqrt{2\eta/\beta} \cdot Z_k, \tag{1.1}$$

其中 k 表示迭代步数， η 表示步长，而 $\{Z_k\}_{k\geq 0}$ 是独立同分布的标准高斯分布。人们运用多种理论工具来分析朗之万蒙特卡洛及其诸多变体的动力学，例如非渐近分析和泛函不等式（Cheng et al. 2018; Chewiet al. 2021; Raginsky etal. 2017）、命中时间分析和切格常数（Chen et al. 2020;Zhang et al. 2017; Zou et al. 2021），仅举数例。

在 H 为凸（或等价地， π_β^0 为对数凹）的情形下，朗之万蒙特卡洛及其变体的分析已相对成熟，参见例如相关论文（Brosse et al. 2018; Dalalyanand Karagulyan 2019）及其中的参考文献。另一方面，当 H 为非凸时，已知相应的过阻尼朗之万扩散收敛至平稳态的速度较慢。阻碍收敛的主要瓶颈之一，是存在所谓的能垒 M，它将局部极小值与全局最小值分隔开来。Gayrard et al.（2005）表明，低温下谱隙的量级为 $\Theta(e^{\beta M})$。从广义上说，M 可视为衡量景观 H难度的指标。值得注意的是， M 在模拟退火背景下的冷却日程设计中也会出现，参见例如（Chiang et al.1987; Holleyand Stroock 1988; Monmarch´e 2018）。这凸显了在这类具有挑战性却十分重要的情形下，设计加速或改进算法的重要性。

在文献中，加速基于朗之万动力学的随机算法的设计大致可分为两类：一是在相同的景观上提出新的动力学，二是改变底层景观或几何结构。对于第一类，研究者考察了原始 LMC 动力学的多种变体，例如镜像朗之万（Zhang 等人 2020）、模拟调温朗之万（Lee 等人 2018）或方差缩减朗之万（Kinoshita 和 Suzuki 2022）。本文采用的思路属于第二类，即提出一种新的景观修改方法。在其他相关领域中，也有人将景观修改作为加速技术，并以不同名称加以应用，例如 Catoni 提出的能量变换算法（Catoni 1998）以及量子绝热退火（Wang 等人 2016）。本文修改方法的主要动机源自（Fang 等人 1997），该文研究了具有状态依赖扩散系数的过阻尼朗之万扩散。

1.1. 我们的贡献

给定一个可能为非凸的目标函数 H，景观修改的核心思想是将 H 转换为另一个函数，该函数在由 H 指定的原始景观上具有更优的混合或优化性能。更具体地说，本文提出了一种针对 H 的新变换 $H^f_{\beta,\,c,\,1}$，其定义为

$$H^f_{\beta,c,1}(x) := H^* + \int_{H^*}^{H(x)} h^f_{\beta,c,1}(u)du, \tag{1.2}$$

其中， $H^* := \min_x H(x)$ 是 H 的全局最小值，且 $h^f_{\beta,\,c,\,1}:\ \mathbb{R} \to \mathbb{R}^+$ 是一个具有如下形式的正函数

$$h^f_{\beta,c,1}(u) := \frac{1}{\beta f(u-c)+1}. \tag{1.3}$$

注意，这种变换 $H^f_{\beta,\,c,\,1}$ 取决于三个参数，即:

- β 是动力学中的逆温度参数（1.1）。
- $f:\ \mathbb{R} \to \mathbb{R}^+$ 是一个非负非递减函数，满足 $f \in C^1$（$\mathbb{R}$） 且f（0） $= f'$（0） $= 0$。注意，这些条件意味着 f（x） $= 0$对于 $x \le 0$成立。本文重点讨论如下文（3.3）所指定的 f。

- c 称为阈值参数，并假设其选取满足 $c > H^*$。在次水平集 $\{H(x) \le c\}$中，根据（1.2）可知 $H^f_{\beta,\,c,\,1} = H$，因此变换后的景观与原始景观重合。另一方面，在超水平集 $\{H(x) > c\}$中，我们观察到，当 β 足够大时，修改后的景观 $H^f_{\beta,\,c,\,1}$ 会被压缩，其取值由 f指定。因此，参数 f 和 c 共同决定并控制阈值 c以上的变换。

当 H 可微时，由于 $h > 0$，很容易看出， $H^f_{\beta,c,1}$ 的临界点集合与 H的临界点集合一致。接下来，我们将考虑修正吉布斯分布 $\pi^f_{\beta,c,1} \propto \exp\left(-\beta H^f_{\beta,c,1}\right)$。在特殊情形 $f = 0$下，我们看到$H^0_{\beta,c,1} = H$，因此原始吉布斯分布$\pi^0_\beta = \pi^0_{\beta,c,1}$。这解释了原始吉布斯分布中上标 0 的由来。

为了说明景观修改，我们在图 $H(x) = \sin 2x + 5\cos x + x^2 - 1.38$ 中绘制了一个基准函数及其对应的 $H^f_{\beta,c,1}(x)$，见图 1，其中 δ 是选择时的一个超参数，相关说明将在第 3节给出。

基于上述景观修改思路，我们将考虑在修改后的景观 $H^f_{\beta,c,1}$上运行朗之万蒙特卡洛算法，即我们考虑

$$Y_{k+1} = Y_k - \eta \nabla H^f_{\beta,c,1}(Y_k) + \sqrt{2\eta/\beta}\cdot Z_k,$$

其中， $Y = \{Y_k\}_{k\ge 0}$ 是修改后的动力学的迭代，而修改后的目标函数的梯度可计算为

$$\nabla H^f_{\beta,c,1} = \frac{\nabla H}{\beta f(H-c)+1}.$$

请注意，每一步我们都需要梯度 ∇H 和目标函数 H 的信息，而原始朗之万蒙特卡洛算法在每次迭代中只需要梯度 ∇H。从这个角度来看， Y中的每次迭代都会产生额外的计算成本。现在我们可以陈述本文的主要贡献:

- 我们提出并分析了一种新的朗之万蒙特卡洛算法，该算法基于经过修改的景观，适用于低温下的随机优化或采样，且只需极少的额外计算成本。

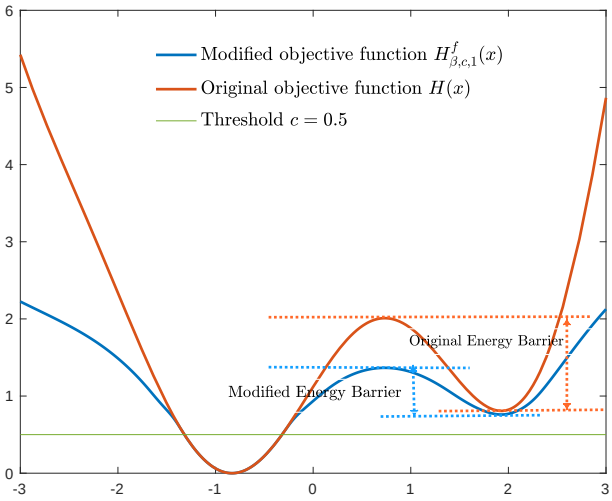

图 1：原始目标函数与修改后的目标函数的景观。此处 $H(x)$ = 正弦 $2x + \frac{5}{2}$ 余弦 $x + x^2 - 1.38$， $c = 0.5$， β =1， $\delta = 1$，以及 f 在下文的（3.3）中给出。

• 证明利用了亚稳态领域的研究成果，这些成果通过相关能垒有效地界定了谱隙和对数 Sobolev 常数等各种泛函常数，从而避免了文献中通常所需的耗散性假设（Raginsky 等人，2017）。例如，我们将证明，修正后的目标函数 $H^f_{\beta,c,1}$ 的能垒 M^f 低于原始能垒 M，由此可知，修正吉布斯分布 $\pi^f_{\beta,c,1} \propto \exp\left(-\beta H^f_{\beta,c,1}\right)$ 的对数 Sobolev 常数至多以多项式形式依赖于 β 和 M，其中 $C^f_{LSI} = \mathcal{O}\left(\beta M + {}^{\beta M^2}_{H\ m_2}(\,)_{-c}\right)$；这与原始吉布斯分布 $\pi^0_\beta \propto \exp(-\beta H)$ 的对数 Sobolev 常数形成鲜明对比，后者如 Menz 和 Schlichting 2014 所示，会随 β 和 ~~M~~ 呈指数级增长。

2. 预备知识

在介绍主要结果之前，本节将回顾一些经典结果，并确定全文将使用的一些记号。首先，我们正式定义朗之万蒙特卡洛的随机动力学。

定义 1（朗之万蒙特卡洛（LMC）， $X = \{X_k\}$）设 $\eta > 0$ 为步长，则 LMC 可写为

$$X_{k+1} = X_k - \eta \nabla H(X_k) + \sqrt{2\eta/\beta} \cdot Z_k, \tag{2.1}$$

其中 $Z_k \sim N(0, I)$ 是独立的标准高斯分布，且 $X_0 \sim \rho_0$ 其中 ρ_0可以是任意初始分布。令 ρ_k 表示 X_k 按照 LMC 演化得到的概率分布。

在满足关于 H的某些正则性条件时，可以证明，迭代（2.1）将收敛到π_η，即（2.1）的平稳分布。当 $\eta \to 0$时，朗之万蒙特卡洛在连续时间下还原为过阻尼朗之万扩散。此外，需要注意的是 $\pi_\eta \neq \pi_\beta^0$，并且$\pi_\eta$ 与 π_β^0 之间存在一个较小的步长相关偏差，其大小与 η成正比。π_η 与 π_β^0 之间的显式偏差可见（Vempala 和 Wibisono 2019）第 7 节。

接下来，我们回顾庞加莱不等式、对数索博列夫不等式和塔拉格朗德不等式的定义。

定义 2（庞加莱不等式和对数索博列夫不等式，PI 和 LSI）若博雷尔概率测度 ν在 $\mathbb{R}^d$ 上满足常数为 $C_{PI} > 0$ 的庞加莱不等式，则对所有测试函数 $\phi \in H^1(\nu)$,

$$\int (\phi - \mathbb{E}_\nu[\phi])^2 \, d\nu \leq C_{PI} \int |\nabla \phi|^2 d\nu,$$

其中，$H^1(\nu) = W^{2,\,1}(\nu)$ 是相对于测度 ν 的索博列夫空间。类似地，若对所有测试函数 ϕ，均有 $\mathbb{R}^d \to \mathbb{R}^+$ with $I(\phi\nu \mid \nu) < \infty$ 成立，则概率测度 ν 满足常数为 $C_{LSI} > 0$ 的对数索博列夫不等式

$$\int \phi \log \frac{\phi}{\mathbb{E}_\nu[\phi]} d\nu \leq \frac{C_{LSI}}{2} \int \frac{|\nabla \phi|^2}{2\phi} d\nu \, (=: I(\phi\nu|\nu))$$

假设 2.1（Talagrand不等式）定义在 $\mathbb{R}^d$ 上的博雷尔概率测度 ν 满足常数为 $T_a > 0$ 的Talagrand不等式；对于任意概率测度 ρ，有 $i.e.$

$$\frac{T_a}{2} W_2(\rho, \nu)^2 \leq D_{KL}(\rho \| \nu),$$

其中，$D_{KL}(\rho\|\nu)$ 是 ρ 相对于 ν 的 Kullback-Leiberg（KL）散度，即

$$D_{KL}(\rho\|\nu) = \mathbb{E}_\rho \left[\ln \frac{\rho}{\nu}\right] = \int_{\mathbb{R}^d} \rho(x) \ln \frac{\rho(x)}{\nu(x)} dx,$$

其中（稍微滥用一下记号），$\rho(\cdot)$ 和 $\nu(\cdot)$ 分别是 ρ 和 ν 的密度。

第三，我们引入本文将要研究的目标函数类 H。以下定义和假设取自（Menz 和 Schlichting 2014），其中给出了关于 H 的增长条件，以确保 H 的驻点集合有限。

定义 3（莫尔斯函数）光滑函数 $H\colon\ \mathbb{R}^d \to \mathbb{R}$ 是莫尔斯函数，当海森矩阵$\nabla^2 H$ 在临界点集合上非退化时，$i.e.$ 对某个 $1 \leq C < \infty$ 成立

$$\forall x \in \mathcal{S} := \left\{x \in \mathbb{R}^d : \nabla H = 0\right\} : \frac{|\xi|^2}{C} \leq \left\langle \xi, \nabla^2 H(x) \xi \right\rangle \leq C|\xi|^2.$$

假设2.2（PI 中的增长条件）$H \in C^3\left(\mathbb{R}^d, \mathbb{R}\right)$ 是一个莫尔斯函数，且存在常数 $C_H > 0$和 $D_H \geq 0$使得

$$\begin{aligned} &(\mathrm{A1_{PI}}) && \liminf_{|x|\to\infty} |\nabla H| \geq C_H, \\ &(\mathrm{A2_{PI}}) && \liminf_{|x|\to\infty} \left(|\nabla H|^2 - \Delta H\right) \geq -D_H. \end{aligned}$$

假设2.3（LSI 中的增长条件） $H \in C^3\left(\mathbb{R}^d, \mathbb{R}\right)$ 是一个莫尔斯函数，且存在常数 $C > 0$ 和 $D \geq 0$使得

$$\begin{array}{ll} (\mathrm{A1_{LSI}}) & \liminf_{|x|\to\infty} \frac{|\nabla H(x)|^2 - \Delta H(x)}{|x|^2} \geq C, \\ (\mathrm{A2_{LSI}}) & \inf_x \nabla^2 H(x) \geq -D. \end{array}$$

上述假设保证驻点集合是离散且有限的。特别地，我们定义 $\mathcal{S} := \{m_1, m_2, ..., m_S\}$ 为局部极小值的集合，其中 $S < \infty$。

第四，我们回顾鞍点高度和能垒的概念，它们广义地衡量了 H 的景观的难度。

定义 4（鞍点高度）两个局部极小值 m_i, m_j 之间的鞍点高度 $H(m_i, m_j)$定义为

$$\widehat{H}\left(m_i, m_j\right) := \inf\left\{\max_{s\in[0,1]} H(\gamma(s)) : \gamma \in C\left([0,1], \mathbb{R}^n\right), \gamma(0) = m_i, \gamma(1) = m_j\right\}.$$

假设 2.4（非退化性、唯一全局极小点和能垒）存在 $\delta > 0$使得

1. 两个局部极小值 m_i, m_j 之间的鞍点高度在唯一的指标为一的临界点 $s_{i,j} \in \mathcal{S}$处取得，$i.e.\ H(s_{i,j}) = H(m_i, m_j)$，且若 $\{\lambda_1, \ldots, \lambda_n\}$ 表示 $\nabla^2 H(s_{i,j})$的特征值，则有 $\lambda_1 < 0$ 且 $\lambda_i > 0$，其中 $i = 2, \ldots, n$。
2. 局部极小值集合 $\mathcal{S} = \{m_1, \ldots, m_S\}$ 按如下方式排序： m_1 是全局最小值，且对所有 $i \in \{3, \ldots, S\}$ 均有
$$H\left(s_{1,2}\right) - H\left(m_2\right) \geq H\left(s_{1,i}\right) - H\left(m_i\right) + \delta.$$
3. m_1 是 H的唯一全局最小值。
4. H 的能垒 M 定义为
$$M = H(s_{1,2}) - H(m_2).$$

尽管唯一全局最小值的假设 H 看起来有所限制，但这一设定确实会在实践中出现，例如研究具有唯一全局众数的高斯混合模型时，参见（Chen 和 Xi2020）。我们还将对 L 施加以下 H-光滑且海森矩阵为利普希茨连续的假设。

假设 2.5（L-光滑）H 二阶可微，且 $\|\nabla H(x) - \nabla H(y)\| \leq L\|x - y\|$。

假设2.6（利普希茨海森矩阵） $\nabla^2 H$ 是 L_1-利普希茨连续的， $i.e.\ \|\nabla^2 H(x) - \nabla^2 H(y)\| \leq L_1\|x - y\|$。

我们现在回顾亚稳态文献中的一些经典结果，本文将频繁使用这些结果。以下命题来自（Menz 和 Schlichting 2014）中的推论 2.18，该推论用能垒对各种泛函常数进行了界定。

命题 5 在假设 (2.2) 至 (2.4)下，令 $\left\{\kappa_i^2\right\}_{i=1}^S$ 由下式给出：

$$\kappa_i^2 := \det \nabla^2 H\left(m_i\right).$$

一方面，如果存在唯一全局极小点，即对于 $i \in \{2, \ldots, S\}$， $H(m_1) < H(m_i)$，则有

$$C_{PI} \lesssim \frac{1}{\kappa_2} \frac{2\pi\sqrt{\left|\det \nabla^2\left(H\left(s_{1,2}\right)\right)\right|}}{\beta\left|\lambda^-\left(s_{1,2}\right)\right|} \exp\left(\beta\left(H(s_{1,2}) - H(m_2)\right)\right),$$

$$C_{LSI} \lesssim \left(\beta\left(H(m_2) - H(m_1)\right) + \ln\left(\frac{\kappa_1}{\kappa_2}\right)\right) C_{PI}.$$

另一方面，如果对于 $i \in \{3, \ldots, M\}$， $H(m_1) = H(m_2) < H(m_i)$，则有

$$C_{PI} \lesssim \frac{1}{\kappa_1 + \kappa_2} \frac{2\pi\sqrt{\left|\det \nabla^2\left(H\left(s_{1,2}\right)\right)\right|}}{\beta\left|\lambda^-\left(s_{1,2}\right)\right|} \exp\left(\beta\left(H(s_{1,2}) - H(m_2)\right)\right),$$

$$C_{LSI} \lesssim \frac{1}{\Lambda\left(\kappa_1, \kappa_2\right)} \frac{2\pi\sqrt{\left|\det \nabla^2\left(H\left(s_{1,2}\right)\right)\right|}}{\beta\left|\lambda^-\left(s_{1,2}\right)\right|} \exp\left(\beta\left(H(s_{1,2}) - H(m_2)\right)\right),$$

其中 $\Lambda(p, q) := \frac{p-q}{\ln p - \ln q}$，且 $\lambda^-(s_{1,2})$ 是 $\nabla^2 H(s_{1,2})$ 的负特征值。符号 $\lesssim$表示 $\leq$ 在乘法误差 $1+\mathcal{O}\left(|\ln\beta|^{3/2}\beta^{-1/2}\right)$ 范围内，详见（Menz 和 Schlichting, *2014*）中的定理 2.12。

容易看出，在两种情况下， C_{PI} 和 C_{LSI} 的上界均由项 $e^{\beta M}$ 与一个有理多项式前因子相乘构成，该多项式包含与 H相关的参数。我们将重点关注主导项 $e^{\beta M}$，其中 M 是能垒，即$M = H(s_{1,2}) - H(m_2)$。因此，我们提出的景观修改旨在通过变换降低能垒，以期得到更小的 PI 和 LSI 常数上界；同时，我们还希望确保指数项前的因子不会同时对 β 和 M 呈指数依赖。

以下命题取自（Vempala 和 Wibisono 2019）中的定理 1，它给出了 k^{th} 朗之万蒙特卡洛迭代值与吉布斯分布之间 KL 散度的上界π_β^0。

命题 6 在假设 *2.5*下，对于定义 1 中的朗之万蒙特卡洛*1*，假设 $\nu = \pi_\beta^0$ 是满足 LSI 且常数为 L的 $C_{LSI} > 0$-光滑吉布斯分布。对于任意$X_0 \sim \rho_0$ 满足 $D_{KL}(\rho_0\|\nu) < \infty$，步长为 $X_k \sim \rho_k$ 的朗之万蒙特卡洛迭代值 $0 < \eta \leq \frac{\beta C_{LSI}}{4L^2}$满足

$$D_{KL}(\rho_k\|\nu) \leq e^{-\frac{\eta k}{\beta C_{LSI}}} D_{KL}(\rho_0\|\nu) + 8\eta d L^2 \cdot C_{LSI}/\beta,$$

这里回顾一下， $D_{KL}(\rho\|\nu)$ 是 ρ 关于 ν 的 KL 散度。

我们可以看到，较小的 C_{LSI} 确实可以使命题 6中的上界更小。

3. 主要结果

本节陈述本文的主要结果。景观修改的核心思想是在能垒较低的替代景观 H_β^f 上c运行1 LMC算法。我们将比较在 H_β^f、c上1 的替代景观上运行的 $Y = \{Y_k\}$ 与原始 LMC $X = \{X_k\}$，作为一种替代方案，也就是说，

$$\text{(Original LMC dynamics)} \quad X_{k+1} = X_k - \eta\nabla H(X_k) + \sqrt{2\eta/\beta}\cdot Z_k, \tag{3.1}$$

$$\text{(Modified LMC dynamics)} \quad Y_{k+1} = Y_k - \eta\nabla H^f_{\beta,c,1}(Y_k) + \sqrt{2\eta/\beta}\cdot Z_k, \tag{3.2}$$

其中 $Z_k \sim N(0, I)$，而 $X_0 \sim \rho_0$ 和 $Y_0 \sim \rho_0^f$ 分别是 X 和 Y 的初始分布。

现在，我们转而讨论景观修改中参数 f 的选择。首先，我们回顾一下，我们只考虑 $f \in C^1(\mathbb{R})$，因为这一条件确保 H_β^f,c,1 的海森矩阵连续（例如，见下文引理 13 的证明），而这对于证明主要结果是必要且至关重要的。由于 $f_1(x) := x \vee 0$ 在 $x = 0$处不可微，为了保证 $f \in C^1(\mathbb{R})$，我们将在 $x = 0$ 右侧的一个小邻域 $(0, \delta)$ 内对 $f_1(x) = x \vee 0$ 添加振荡，以定义如下 f:

$$f(x) = f(x,\delta) := \begin{cases} 0, & \text{for} \quad x \le 0, \\ \frac{20}{3\delta}x^2, & \text{for} \quad 0 < x \le \frac{\delta}{4}, \\ -\frac{20}{3\delta}(x - \frac{\delta}{2})^2 + \frac{5}{6}\delta, & \text{for} \quad \frac{\delta}{4} < x \le \frac{\delta}{2}, \\ A\exp(-\frac{1}{B(x-\frac{\delta}{2})^2}) + \frac{5}{6}\delta, & \text{for} \quad \frac{\delta}{2} < x \le \delta, \\ x, & \text{for} \quad x > \delta, \end{cases} \tag{3.3}$$

其中，δ是一个待选参数，且$A = \frac{\delta}{6}e^{3/2}$, $B = \frac{8}{3\delta^2}$，因此 $f(x) \in C^1(\mathbb{R})$。本文其余部分将固定采用这一 f 选择。在接下来的章节中，我们将说明，足够小的 δ 能够像 f_1一样降低能垒。这符合直觉，因为当 x较大时， f_1 和 f 渐近等价。为展示 f_1和 f，我们在取 $\delta = 1$时将它们绘于图2中。为说明采用这一 f时景观修改的效果，我们回顾一下：在第1.1节中，我们以一个基准目标函数为例，在图1中比较了修改后的目标函数H_β^f,c,1以及原始目标函数 H 。

从采样的角度来看，我们关注修正的 LMC Y 向原始吉布斯分布 $\pi_\beta^0 \propto e^{-\beta H}$的混合性能；而从优化的角度来看，我们将研究通过修正的 LMC Y对 H 进行随机优化。

我们将施加一个附加假设，以确保修改后的分布 π_β^f，c，1存在；等价地，在该假设下，可以保证 $\exp\left(-\beta H^f_{\beta,c,1}\right)$ 在 $\mathbb{R}^d$上可积。该结果将在第 A.1 节证明。

假设3.1 假设温度足够低，且选择 c 使得 $H^* < c < H(m_2) - \delta$，其中 $\beta > \frac{1}{\delta}$ 和 c 满足这一条件。此外，存在 $I_1 = I_1(c,\delta) < \infty$，使得

$$\int_{H>c+\delta} \frac{2\delta}{H - c}dx < I_1. \tag{3.4}$$

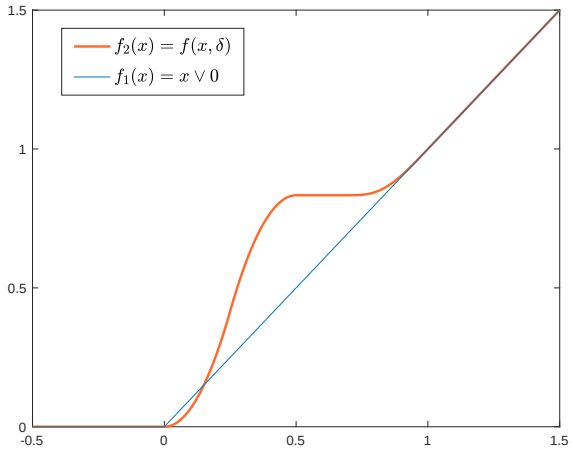

图2：在 $(0, \delta)$ 内，$f_1(x) = x \vee 0$ 振荡至 $f_2(x) = f(x, \delta)$，且 $\delta = 1$。

这一假设很容易在一大类莫尔斯函数中得到验证，例如具有二次增长的莫尔斯函数，这在模拟退火文献中十分常见（Monmarch´e 2018）。

3.1. 一些常用记号

现在，我们正式介绍本文不同部分将使用的各种记号。通常，带有上标 f 的变量与修正的 LMC Y 相关，而不带上标或带有上标 0 的对象与原始 LMC X 相关。为方便读者，我们在附录1的B表中汇总了常用记号。下面列出若干重要记号：

- 将 $G(x) = H^f_{\beta,c,1}(x)$ 记为修改后的目标函数。在假设 2.3和2.4下，我们可以发现 $x^* = m_1$ 是全局最小值，而 m_2 是第二小的局部极小值。我们还将 $H^* = H(x^*)$ 记为 H 在 x^*处的全局最小值。
- ρ_k 和 ρ^f_k 分别是原始迭代 X_k 和修正迭代 Y_k 的分布，即 $X_k \sim \rho_k, Y_k \sim \rho^f_k$。我们将 $\pi_1 = \pi^0_\beta \propto e^{-\beta H}$ 记为原始吉布斯分布，而将 $\pi_2 = \pi^f_{\beta,c,1} \propto \exp\left(-\beta H^f_{\beta,c,1}\right)$ 记为修正吉布斯分布，其中 π_1 和 π_2 是两个简写记号。我们还将 $Z_1 = \int e^{-\beta H} dx$ 和$Z_2 = \int \exp\left(-\beta H^f_{\beta,c,1}\right) dx$ 分别记为 π_1和 π_2的归一化常数。

- C^f_{LSI}是π^f_β的对数 Sobolev 常数，c、1，而 C^f_{PI}是π^f_β的 Poincaré 常数，c、1。此外，$G = H^f_\beta$、c、1 是 L^f-光滑的（将在定理 7 中证明）。π_1和 π_2的 Talagrand不等式常数分别为 T_a和 T^f_a，而 M和 M^f分别是 H和 G的能垒

。 G的局部极小值集合定义为集合 $\mathcal{S}^f = \{m_1^f, m_2^f, \ldots, m_S^f\}$，而$s_{1,2}^f$ 是 m_1^f 和 m_2^f 之间的鞍点。

• 记 $K = \nabla^2 H(m_1)$，并且 $\lambda_{min}(K) = \gamma > 0$ 表示 K 的最小特征值，因为根据假设 K，2.4， $A \in \mathbb{R}^{d\times d}$是正定矩阵。对于任意正定矩阵 $x \in \mathbb{R}^d$和 $\|x\|_A^2 = \frac{1}{2}x^T Ax$，矩阵范数 $\|B\|$ 表示矩阵B的算子范数。我们还记 ω_d 为 d维欧几里得空间中单位球的体积，并记 μ 为 $\mathbb{R}^d$上的勒贝格测度。

• 对于定义在 $\mathbb{R}^d$ 上的任意两个概率测度 ρ 和 ν，我们将 $D_{TV}(\rho,\nu) = \frac{1}{2}\int |\rho - \nu| dx$ 记为 ρ 与 ν 之间的全变差距离。将 $W_2(\rho,\nu)$ 记为 ρ 与 ν 之间的 Wasserstein距离。

3.2. 从 π_β^0 采样的全变差混合时间上界改进

在接下来的三个小节中，我们将介绍本文的主要结果。我们将看到，在我们的假设下（即通过适当调整参数），修正的 LMC 的时间复杂度关键取决于修正的 LSI C_{LSI}^f；它对 β 和 M 呈多项式依赖，而原始 LMC 对 β 和 M则呈指数依赖。

我们的第一个主要结果给出了利用修正的 LMC Y 在低温下从原始吉布斯分布 π_β^0 进行近似采样的非渐近误差界：

定理 7 在假设 2.2 至 2.6 以及 3.1成立时，则对于任意 $Y_0 \sim \rho_0^f$ 和步长 $0 < \eta \leq \frac{\beta C_{LSI}^f}{16 L^{f^2}}$，记 $\beta_0 =$ 最大值 $\left\{\frac{1}{\delta^2}, \frac{3}{20\delta}, \frac{1}{\delta}\right\}$，且当 $\beta > \beta_0$ 时，我们有

$$D_{TV}(\rho_k^f, \pi_\beta^0) \leq e^{-\frac{\eta k}{2\beta C_{LSI}^f}} \sqrt{2 D_{KL}(\rho_0^f \| \pi_{\beta,c,1}^f)} + 4\sqrt{\eta d L^{f^2} \cdot C_{LSI}^f/\beta} + \left(\frac{\beta L}{2\pi}\right)^{d/2} \left(\mu(c \leq H \leq c+\delta) e^{-\beta(c-H^*)} + e^{-\beta(c-H^*)} I_1\right),$$

其中 $C_{LSI}^f = \mathcal{O}\left(\beta M + \underline{H_{m_2}^{\beta M^2}() - c}\right)$，且 G是 L^f-光滑的，

满足

$$L^f \leq \max\left\{(2(c-H^*)+3\delta)L, \left(\frac{80}{3}(c-H^*) + \frac{92}{3}\delta\right)L, \frac{c-H^*}{2} + \frac{\delta}{8} + \frac{\delta}{2}L\right\}.$$

定理7的证明推迟至第A.2节。我们注意到，要使第一项小于或等于给定的误差容限 ϵ，所需的迭代次数 k 至多为

$$k = \mathcal{O}\left(\frac{\beta C_{LSI}^f}{\eta} \log\left(\frac{1}{\epsilon}\right)\right),$$

其对逆温度 β 和能垒 M 至多呈多项式依赖。此外，为使第二项小于 ϵ，步长 η 可选为满足 $\eta = \Theta\left(\epsilon^2\right)$，且该选择可以独立于 β。

3.3. 向全局最小值的改进收敛

在第二个主要结果中，我们通过给出达到如下形式的超水平集的概率上界，展示了迭代点 Y_k 向全局最小值 x^* 收敛率的提升：

$$\mathbb{P}\left(H(Y_k) > H^* + \epsilon\right),$$

其中， ϵ是给定的精度或误差容限。

定理8 在假设 2.2 至2.6 和 3.1下，对于任意给定的精度（满足 $0 < \epsilon < c - H^*$），步长 $0 < \eta \leq$最小值 $\left\{ \frac{\beta C_{LSI}^f}{16 L^{f^2}},\ \frac{\beta C_{LSI}}{16 L^2} \right\}$ 以及足够低的温度（满足 $\beta > \beta_0 =$最大值$\left\{ \frac{1}{\delta^2}, \frac{3}{20\delta}, \frac{1}{\delta} \right\}$），如果存在 $I_2 < \infty$ 使得

$$\int_{\mathbb{R}^d} e^{-(H - H^*)} dx < I_2,$$

则有

$$\begin{aligned}\mathbb{P}(H(X_n) > H^* + \epsilon) \leq 2e^{-\frac{\eta n}{2\beta C_{LSI}}} \sqrt{2 D_{KL}(\rho_0 \| \pi_\beta^0)} + 8\sqrt{\eta d L^2 \cdot C_{LSI}/\beta} \\ + \left(\frac{\beta L}{2\pi}\right)^{d/2} e^{-(\beta-1)\epsilon} I_2,\end{aligned}$$

当

$$\begin{aligned}\mathbb{P}(H(Y_k) > H^* + \epsilon) \leq 2e^{-\frac{\eta k}{2\beta C_{LSI}^f}} \sqrt{2 D_{KL}(\rho_0^f \| \pi_{\beta,c,1}^f)} + 8\sqrt{\eta d L^{f^2} \cdot C_{LSI}^f/\beta} \\ + \left(\frac{\beta L}{2\pi}\right)^{d/2} \left(2\mu(c \leq H \leq c + \delta) e^{-\beta(c - H^*)} + 2e^{-\beta(c-H^*)} I_1 + e^{-(\beta-1)\epsilon} I_2\right),\end{aligned}$$

其中我们回顾一下， C_{LSI}^f 和 L^f 如定理7所述。

定理8 的证明见第A.3节。由上文定理8 右侧的第三项可知，当 $\beta \to \infty$时，该项趋于零，因此在温度足够低时，这一项的作用可以忽略。若要使第一项小于或等于给定的误差容限ϵ，所需的迭代次数 k 和 n 至多为

$$k = \mathcal{O}\left(\frac{\beta C_{LSI}^f}{\eta} \log\left(\frac{1}{\epsilon}\right)\right), \quad n = \mathcal{O}\left(\frac{\beta C_{LSI}}{\eta} \log\left(\frac{1}{\epsilon}\right)\right).$$

因此，关于 n的这个上界对 β 和 M 均呈指数依赖，而关于 k 的类似上界对 β 和 M均至多呈多项式依赖。

3.4. 将超额风险降低至全局最小值

我们的最后一个主要结果涉及对差值 $\mathbb{E}_{Y_k \sim \rho_k^f}[H(Y_k)] - H^*$ 的界定，我们将其称为相对于全局最小值的超额风险。

定理 9 在假设 2.2 至 2.6、2.1 和 3.1成立的条件下，对于给定的$0 < r < \frac{3\gamma}{2L_1}$，假设 x^* 是 H 在 $B_r(x^*)$ 内的唯一局部极小值，且 $\max_{B_r(x^*)} H(x) - H^* = \delta_r \le c - H^*$。假设

$\min_{B_r'(x^*)} H(x) - H^* = \Delta_r > 0$，则对于步长 $0 < \eta \le \frac{\beta C^f_{LSI}}{16L^{f2}}$，并且当温度足够低且$\beta > \beta_0 = \max\{\frac{1}{\delta}, \frac{1}{\delta^2}, \frac{3}{20\delta}\}$时，如果存在 $0 < I_4 = I_4(c) < \infty$，使得

$$\int_{H>c} (H - H^*) e^{-(H-H^*)} dx < I_4,$$

则有

$$\begin{aligned}
\mathbb{E}_{Y_k \sim \rho_k^f}[H(Y_k)] - H^* \le\ & 4LC^f_{LSI}\left(e^{-\frac{\eta k}{\beta C^f_{LSI}}} D_{KL}(\rho_0^f \| \pi^f_{\beta,c,1}) + 8\eta d L^{f2} \cdot C^f_{LSI}/\beta\right) \\
& + 4LC^f_{LSI}\left(\frac{\beta L}{2\pi}\right)^{d/2}\left(\mu(c < H \le c+\delta)e^{-\beta(c-H^*)} + e^{-\beta(c-H^*)} I_1\right) \\
& + 3\frac{d\omega_d \Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L}{\pi}\right)^{d/2} \cdot \frac{1}{\beta} \\
& + 2\left(\frac{\beta L}{2\pi}\right)^{d/2}\left((c - H^*)\mu(H \le c)e^{-\beta\Delta_r} + e^{-(\beta-1)(c-H^*)} I_4\right),
\end{aligned}$$

其中我们回顾 C^f_{LSI} 和 L^f 如定理 7中所述。

定理 9 的证明及其直观解释见第 A.4节。接下来，我们在备注 10 中将定理9与经典结果进行比较，该备注讨论了不进行景观修改时的超额风险。我们注意到，超额风险之间的核心差异源于不同的对数 Sobolev 常数：具体而言，$C_{LSI} = \mathcal{O}\left(e^{\beta M}\right)$，而 $C^f_{LSI} = \mathcal{O}\left(\beta M + {}_H{}^{\beta M^2}{}_{m_2}() - c\right)$。尽管在计算修正的 LMC 的超额风险时，将 π^0_β 与π^f_β 之间的距离c、1 纳入其中存在一定权衡，但与对数 Sobolev 常数对时间复杂度计算的影响相比，这些因素的贡献相对较小。

注释 10 在假设 2.2 至 2.6、2.1 和 3.1下，并采用定理 9中的相同设定，对于步长 $0 < \eta \le \frac{\beta C^f_{LSI}}{16L^{f2}}$，若

$$\int_{H>c} (H - H^*) e^{-(H-H^*)} dx < I_4,$$

则由命题 6，可得

$$W_2^2\left(\rho_k, \pi^0_\beta\right) \le 2C_{LSI}\left(e^{-\frac{\eta k}{\beta C_{LSI}}} H_\nu\left(\rho_0\right) + 8\eta d L^2 \cdot C_{LSI}/\beta\right),$$

and

$$\begin{aligned}
&\mathbb{E}_{X \sim \pi_1}[H(X)] - H^* \\
&\quad \le \frac{3}{2}\frac{d\omega_d \Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L}{\pi}\right)^{d/2} \cdot \frac{1}{\beta} + \left(\frac{\beta L}{2\pi}\right)^{d/2}\left((c - H^*)\mu(H \le c)e^{-\beta\Delta_r} + e^{-(\beta-1)(c-H^*)} I_4\right),
\end{aligned}$$

这意味着

$$\begin{aligned}\mathbb{E}_{X_k\sim\rho_k}[H(X_k)]-H^* &\leq LW_2^2\left(\rho_k,\pi_\beta^0\right)+2(\mathbb{E}_{X\sim\pi_1}[H(X)]-H^*)\\ &\leq 2LC_{LSI}\left(e^{-\frac{\eta k}{\beta C_{LSI}}}H_\nu\left(\rho_0\right)+8\eta dL^2\cdot C_{LSI}/\beta\right)\\ &\quad+3\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L}{\pi}\right)^{d/2}\cdot\frac{1}{\beta}\\ &\quad+2\left(\frac{\beta L}{2\pi}\right)^{d/2}\left(c\mu(H\leq c)e^{-\beta\Delta_r}+e^{-(\beta-1)(c-H^*)}I_4\right),\end{aligned}$$

w此处 $C_{LSI}=\mathcal{O}\left(e^{\beta M}\right)$，第一条不等式的证明可见引理 *19* .

参考文献

Brosse, N., Durmus, A., & Moulines, E. (2018). 随机梯度朗之万动力学的前景与陷阱. 神经信息处理系统进展, 31. Catoni, O. (1998). Metropolis算法的能量变换方法与模拟退火的比较. 概率论及相关领域, 110, 69–89. Chen, X., Du, S. S., & Tong, X. (2020). 随机梯度朗之万动力学的驻点命中时间与遍历性. 机器学习研究杂志. Chen, Y., & Xi, X. (2020). 高斯混合模型的似然景观与局部极小值结构. arXiv预印本 arXiv:2009.13040. Cheng, X., Chatterji, N. S., Bartlett, P. L., & Jordan, M. I. (2018). 欠阻尼朗之万马尔可夫链蒙特卡洛：非渐近分析. 学习理论会议, 300–323. Chewi, S., Erdogdu, M. A., Li, M. B., Shen, R., & Zhang, M. (2021). 从 Poincar´e 到对数索博列夫不等式的朗之万蒙特卡洛分析. arXiv预印本 arXiv:2112.12662. Chiang, T.-S., Hwang, C.-R., & Sheu, S. J. (1987). R^n中的全局优化扩散. SIAM控制与优化杂志, 25(3), 737–753. Dalalyan, A. S., & Karagulyan, A. (2019). 针对梯度不准确的朗之万蒙特卡洛的易用保证. 随机过程及其应用, 129(12), 5278–5311. Fang, H., Qian, M., & Gong, G. (1997). 一种改进的退火方法及其大时间行为. 随机过程及其应用, 71(1), 55–74. Gayrard, V., Bovier, A., & Klein, M. (2005). 可逆扩散过程中的亚稳态 II：小特征值的精确渐近. 欧洲数学学会杂志, 7(1), 69–99. Holley, R., & Stroock, D. (1988). 通过索博列夫不等式实现模拟退火. 数学物理通讯, 115(4), 553–569. Hwang, C.-R. (1980). 重新审视拉普拉斯方法：概率测度的弱收敛. 概率年刊, 1177–1182. Kinoshita, Y., & S uzuki, T. (2022). 随机梯度朗之万动力学收敛率的提升（结合方差缩减）及其在优化中的应用. arXiv预印本 arXiv:2203

.16217.

Łapi´nski, T. M. (2019). 带估计误差的多元拉普拉斯近似及其在极限定理中的应用。逼近理论杂志，248，105305。

Lee, H., Risteski, A., & Ge, R. (2018). 超越对数凹性：使用模拟调温朗之万蒙特卡罗对多模态分布进行采样的可证明保证。神经信息处理系统进展，31。Menz, G., & Schlichting, A. (2014). 通过分解能量景观得到的 Poincar´e 不等式和对数索博列夫不等式。概率年刊，42(5)，1809–1884。Monmarch´e, P.(2018). 亚稳态情形下的次强制性与动力学模拟退火。概率理论及相关领域，172(3-4)，1215–1248。Otto, F., & Villani, C. (2000). Talagrand 不等式的推广及其与对数 Sobolev 不等式的联系。泛函分析杂志，173(2)，361–400。Raginsky, M., Rakhlin, A., & Telgarsky, M. (2017). 通过随机梯度朗之万动力学进行非凸学习：非渐近分析。学习理论会议，1674–1703。Vempala, S., & Wibisono, A. (2019). 未校正朗之万算法的快速收敛：等周性已足够。神经信息处理系统进展，32。Wang, Y., Wu, S., & Zou, J. (2016). 使用马尔可夫链蒙特卡罗模拟和D-Wave量子计算机进行量子退火。统计科学，362–398。Zhang, K. S., Peyr´e, Fadili, J., & Pereyra, M. (2020). 镜像朗之万蒙特卡罗的瓦瑟斯坦控制。学习理论会议，3814–3841。Zhang, Y., Liang, P., & Charikar, M. (2017). 随机梯度朗之万动力学的命中时间分析。学习理论会议，1980–2022。Zou, D., Xu, P., & Gu, Q. (2021). 非对数凹采样中随机梯度朗之万动力学的更快收敛。人工智能不确定性，1152–1162。

附录A. 主要结果的证明

A.1. 修正吉布斯分布的存在性 $\pi_{\beta,c,1}^{f}$

本小节旨在证明修正吉布斯分布 $\pi_{\beta,\,c,\,1}^{f}$ 的存在性。在下面的定理中，我们将证明，在假设 3.1 下，修正吉布斯分布存在。

定理11 在假设 *2.2* 至 *2.6* 以及 *3.1* 下，修正吉布斯分布 $\pi_{\beta,\,c,\,1}^{f}$ 存在，即

$$\int_{\mathbb{R}^d} e^{-\beta G} dx < \infty.$$

证明 首先，我们明确写出 H 和 G 之间的关系。在 $\{x:\ H(x) \le c\}$ 中，我们有 $H = G$。另一方面，在 $\{x:\ H(x) > c+\delta\}$ 中，我们有

$$\begin{aligned} G(x) &= H^* + \int_{H^*}^{H(x)} \frac{1}{\beta f(u-c)+1} du \\ &= c + \int_c^{c+\delta} \frac{1}{\beta f(u-c)+1} du + \int_{c+\delta}^{H(x)} \frac{1}{\beta(u-c)+1} du \\ &= c + \int_c^{c+\delta} \frac{1}{\beta f(u-c)+1} du + \frac{1}{\beta} \ln \frac{H-c+1/\beta}{\delta+1/\beta}. \end{aligned}$$

由于 $\beta > \frac{1}{\delta}$，因此我们有

$$c + \frac{1}{\beta}\ln\frac{H-c}{2\delta} \le G(x) \le c + \delta + \frac{1}{\beta}\ln\frac{H-c}{\delta}, \tag{A.1}$$

这意味着

$$\int_{H>c+\delta} e^{-\beta G}dx \le e^{-\beta c}\int_{H>c+\delta}\frac{2\delta}{H-c}dx < e^{-\beta c}I_1.$$

注意，$\mu(H \le c+\delta) < \infty$，所需结果即可得出。 ■

在许多一般情形下，假设3.1都可以很容易地验证。例如，对于 H的二次增长情形，若存在 $R > 0$ 和 b_1，$b_2 > 0$ 使得当 $\|x\| > R$时，$b_1\|x\|^2 < H(x) < b_2\|x\|^2$，则条件（3.4）成立。

A.2. 从 π_β^0 采样的全变差混合改进界的证明

本节将通过界定 7 与 ρ_k^f 和 π_β^0之间的全变差距离，给出定理

首先，我们根据原始能垒 M、逆温度 β 和阈值参数 c，计算修改后的目标函数 G 的能垒 M^f：

引理 12（G 的能垒降低）

在假设 *2.2* 至 *2.6* 以及 *3.1*下，修改后函数的能垒为

$$M^f = G(s_{1,2}^f) - G(m_2^f) \le \frac{1}{\beta}\cdot\ln\frac{M}{H(m_2)-c}.$$

证明 首先，回顾一下，修改后函数 G 定义为

$$G(x) = H_{\beta,c,1}^f(x) = H^* + \int_{H^*}^{H(x)}\frac{1}{\beta f(u-c)+1}du.$$

由于 $f(x) \ge 0$，可知 G 保持了 H 的驻点集合，并且对于 $\forall x$，y，$H(x) > H(y)$ 当且仅当 $G(x) > G(y)$。这意味着 $s_{1,2}^f = s_{1,2}$ 以及 $m_2^f = m_2$。此外，根据我们对 c 的选择，即 $H^* < c < H(m_2) - \delta$，有

$$\begin{aligned}
M^f &= G(s_{1,2}^f) - G(m_2^f) = G(s_{1,2}) - G(m_2)\\
&= \int_{H(m_2)}^{H(s_{1,2})}\frac{1}{\beta(u-c)+1}du\\
&= \frac{1}{\beta}\left(\ln\left(H(s_{1,2}) - c + 1/\beta\right) - \ln\left(H(m_2) - c + 1/\beta\right)\right)\\
&= \frac{1}{\beta}\cdot\ln\frac{H(s_{1,2}) - H(m_2)}{H(m_2) - c + 1/\beta}\\
&\le \frac{1}{\beta}\cdot\ln\frac{M}{H(m_2)-c}.
\end{aligned}$$

■

接下来，我们将把修正能垒 M^f代入命题 5中，以计算修正后的 PI 常数和 LSI 常数，即 C_{PI}^f和 C_{LSI}^f。

引理 13 （对数 Sobolev 常数的更小上界）在假设 2.2 至 2.6 以及3.1成立的条件下，有

$$C_{LSI}^{f}=\mathcal{O}\left(\beta M+\frac{\beta M^{2}}{H(m_{2})-c}\right).$$

证明 注意到当 $H(x)>c+\delta$ 时， G 的梯度和海森矩阵可计算为

$$\nabla G(x)=\frac{\nabla H(x)}{1+\beta(H(x)-c)},$$
$$\nabla^{2}G(x)=\frac{\nabla^{2}H(x)}{\beta(H-c)+1}-\frac{\nabla H\cdot\nabla H^{T}\cdot\beta}{(\beta(H-c)+1)^{2}},$$

由此可得

$$\kappa_{i}^{f^{2}}:=\det\nabla^{2}G(m_{i}^{f})=\frac{\kappa_{i}^{2}}{(\beta(H(m_{i})-c)+1)^{d}}.$$

此外，注意到 $G(x)\leq H(x)$，并回忆 $G(m_{1})=H^{*}$，因此有

$$G(m_{2})-G(m_{1})\leq H(m_{2})-H^{*}.$$

综合以上结果并结合命题 5 和引理13 可得

$$\begin{aligned}C_{PI}^{f}&\lesssim\frac{1}{\kappa_{2}}\frac{2\pi\sqrt{\left|\det\nabla^{2}\left(H\left(s_{1,2}\right)\right)\right|}}{\left|\lambda^{-}\left(s_{1,2}\right)\right|}\cdot\frac{\beta(H(s_{1,2})-c)+1}{\beta}\cdot\frac{M}{H(m_{2})-c}\\&=\frac{1}{\kappa_{2}}\frac{2\pi\sqrt{\left|\det\nabla^{2}\left(H\left(s_{1,2}\right)\right)\right|}}{\left|\lambda^{-}\left(s_{1,2}\right)\right|}\cdot\left(M+\frac{M^{2}}{H(m_{2})-c}+\frac{1}{\beta}\frac{M}{H(m_{2}-c)}\right)\end{aligned}$$

and

$$C_{LSI}^{f}\lesssim\left(\beta(H(m_{2})-H^{*})+\ln\frac{\kappa_{1}}{\kappa_{2}}\right)\cdot C_{PI}^{f},\tag{A.2}$$

这意味着 $C_{LSI}^{f}=\mathcal{O}\left(\beta M+\frac{\beta M^{2}}{H(m_{2})-c}\right)$。 ■

注 14 我们可以看到， C_{LSI}^{f} 相比于 C_{LSI}（原始吉布斯分布的 LSI 常数）π_{β}^{0}更小，而这一点可以直接由命题5得到。具体而言，当 β <style id='19}>较大时，我们可以写成

$$C_{LSI}=\mathcal{O}\left(e^{\beta M}\right),$$

而

$$C_{LSI}^{f}=\mathcal{O}\left(\beta M+\frac{\beta M^{2}}{H(m_{2})-c}\right).$$

换句话说， C_{LSI}同时呈指数级依赖于 β和 M，而修改后的常数C_{LSI}^{f}对这些参数至多呈多项式依赖，相比之下，这是一个显著的改进。

根据上文的命题6 和备注14 ，我们可以注意到，由于 LSI 常数降低，迭代 Y_{k} 到 $\pi_{\beta,c,1}^{f}$ 的 KL 散度更小。接下来，我们估计$\pi_{\beta,c,1}^{f}$ 在 $\beta\to\infty$时趋于 π_{β}^{0} 的渐近行为。

引理15 在假设 2.2 至 2.6 以及 3.1下，有

$$D_{TV}(\pi_\beta^0, \pi_{\beta,c,1}^f) \le \left(\frac{\beta L}{2\pi}\right)^{d/2} \left(\mu(c \le H \le c+\delta)e^{-\beta(c-H^*)} + e^{-\beta(c-H^*)} I_1\right),$$

其中我们回顾 I_1 在假设3.1中作为上界出现，而 μ 是 $\mathbb{R}^d$上的勒贝格测度。

证明 回顾 $\mu_1(x) = e^{-\beta H}, \mu_2(x) = e^{-\beta G}$ 和 $H \ge G$，我们有

$$\begin{aligned}
D_{TV}(\pi_\beta^0, \pi_{\beta,c,1}^f) &= \frac{1}{2}\int_{\mathbb{R}^d} \left|\frac{\mu_1}{Z_1} - \frac{\mu_2}{Z_2}\right| dx \\
&\le \frac{1}{2Z_1}\left(\int_{\mathbb{R}^d} |\mu_1 - \mu_2| dx + \int_{\mathbb{R}^d} \left|\frac{Z_2 - Z_1}{Z_2}\right| \mu_2 dx\right) \\
&= \frac{1}{2Z_1}\int_{\mathbb{R}^d} |\mu_1 - \mu_2| dx + \frac{1}{2Z_1}|Z_2 - Z_1| \\
&\le \frac{1}{Z_1}\int_{\mathbb{R}^d} |\mu_1 - \mu_2| dx \\
&= \frac{1}{Z_1}\int_{\mathbb{R}^d} \left|e^{-\beta H} - e^{-\beta G}\right| dx \\
&\le \frac{1}{Z_1}\int_{H \ge c} e^{-\beta G} dx \\
&= \frac{1}{Z_1}\left(\int_{c \le H \le c+\delta} e^{-\beta G} dx + \int_{H > c+\delta} e^{-\beta G} dx\right) \\
&\le \frac{1}{Z_1}\left(\mu(c \le H \le c+\delta)e^{-\beta c} + e^{-\beta c} I_1\right),
\end{aligned}$$

其中第一个不等式利用了三角不等式，最后一个不等式由以下结果得出$G \ge c$ 当 $c \le H \le c+\delta$时，以及 $G(x) \ge c + \frac{1}{\beta}\ln\frac{H-c}{2\delta}$ 根据方程（A.1）当 $H(x) > c+\delta$时。此外，注意到 $H(x) \le H^* + \frac{L}{2}\|x - x^*\|^2$，这意味着

$$Z_1 = \int_{\mathbb{R}^d} e^{-\beta H} dx \ge e^{-\beta H^*} \int_{\mathbb{R}^d} \exp\left(-\frac{\beta L}{2}\|x - x^*\|^2\right) dx = e^{-\beta H^*}\left(\frac{2\pi}{\beta L}\right)^{d/2}, \tag{A.3}$$

由此我们得到所需结果。 ■

在展示关于 ρ_k^f 与 π_β^0 之间全变差的结果之前，我们还需要一个技术引理。该引理源自（Kinoshita 和 Suzuki 2022）中的性质 C.1，并阐明了 H 与 ∇H 之间的关系。

引理16 在假设 2.5 下， H 为 L-*smooth*，我们有

$$H(x) - H^* \ge \frac{1}{2L}\|\nabla H(x)\|^2. \tag{A.4}$$

证明 记 $u(x) = H(x) - H^*$，那么 u 也是 L-*smooth*。注意，对于一个 L-*smooth* 函数H，存在

$$\forall x, y \in \mathbb{R}^d, H(y) \le H(x) + \langle \nabla H(x), y - x\rangle + \frac{L}{2}\|y - x\|^2,$$

于是我们有

$$u\left(x-\frac{1}{L}\nabla u(x)\right)\leq u(x)-\frac{1}{L}\|\nabla u(x)\|^2+\frac{1}{2L}\|\nabla u(x)\|^2=u(x)-\frac{1}{2L}\|\nabla(x)\|^2.$$

根据 $u\geq 0$，我们有

$$u(x)=H(x)-H^*\geq\frac{1}{2L}\|\nabla H(x)\|^2.$$

■

综合上述命题 6、引理12、13、15 和16 即可证明本文的一项主要结果：

证明 [定理 7] 的证明首先，利用三角不等式可得

$$D_{TV}(\rho_k^f,\pi_\beta^0)\leq D_{TV}(\rho_k^f,\pi_{\beta,c,1}^f)+D_{TV}(\pi_{\beta,c,1}^f,\pi_\beta^0). \tag{A.5}$$

我们直接应用引理15 对（A.5）右侧的第二项进行界定，从而得到所述结果。至于（A.5）中的第一项，回顾命题 6的结果，并结合皮恩斯克不等式

$$D_{TV}(\nu_1,\nu_2)\leq\sqrt{2D_{KL}(\nu_1\|\nu_2)}, \tag{A.6}$$

可得

$$\begin{aligned}D_{TV}(\rho_k^f,\pi_{\beta,c,1}^f)&\leq\sqrt{2e^{-\frac{\eta k}{\beta C_{LSI}^f}}D_{KL}(\rho_0^f\|\pi_{\beta,c,1}^f)+16\eta dL^{f^2}\cdot C_{LSI}^f/\beta}\\&\leq\sqrt{2e^{-\frac{\eta k}{\beta C_{LSI}^f}}D_{KL}(\rho_0^f\|\pi_{\beta,c,1}^f)}+4\sqrt{\eta dL^{f^2}\cdot C_{LSI}^f/\beta}.\end{aligned} \tag{A.7}$$

最后，我们继续对 $L^f=\sup\|\nabla^2G(x)\|$ 进行界定。当 $H(x)>c$ 时， G 的海森矩阵可写为

$$\nabla^2G(x)=\frac{\nabla^2H}{\beta f(H-c)+1}-\frac{\nabla H\cdot\nabla H^T\cdot\beta f'(H-c)}{(\beta f(H-c)+1)^2},$$

于是我们看到

$$\begin{aligned}\|\nabla^2G(x)\|&=\max_{\|y\|=1}\langle y,\nabla^2G(x)\cdot y\rangle\\&\leq\frac{\|\nabla^2H(x)\|}{\beta f(H-c)+1}+\frac{\beta f'(H-c)}{(\beta f(H-c)+1)^2}\cdot\max_{\|y\|=1}\langle y,\nabla H\cdot\nabla H^Ty\rangle\\&\leq\frac{L}{\beta f(H-c)+1}+\frac{\beta f'(H-c)}{(\beta f(H-c)+1)^2}\cdot\|\nabla H\|^2\\&=\frac{1}{\beta f(H-c)+1}\left(L+\frac{\beta f'(H-c)}{\beta f(H-c)+1}\|\nabla H\|^2\right).\end{aligned}$$

根据引理 16，当 $H(x) > c+\delta$，$f'(x) = 1$ 且 $\|\nabla H\|^2 \le 2L(H-H^*)$，由于 $\beta > \frac{1}{\delta^2}$，我们有

$$\begin{aligned}\|\nabla^2 G(x)\| &\le \frac{1}{\beta(H-c)}\left(L + \frac{\beta(H-H^*)}{\beta(H-c)+1}\cdot 2L\right)\\ &\le \frac{1}{\beta\delta}\left(3 + \frac{2(c-H^*)}{\delta}\right)\cdot L\\ &\le (2(c-H^*)+3\delta)\cdot L,\end{aligned}$$

当 $c+\frac{\delta}{4} \le H(x) \le c+\delta$ 时，回忆起 $f'(x) \le f'(\frac{\delta}{4}) = \frac{10}{3}$，对于 $\beta > \frac{1}{\delta^2}$，我们有 e

$$\begin{aligned}\|\nabla^2 G(x)\| &\le \frac{1}{\beta(H-c)}\left(L + \frac{10}{3}\frac{\beta(H-H^*)}{\beta(H-c)+1}\cdot 2L\right)\\ &\le \frac{4}{\beta\delta}\left(L + \frac{20}{3}L\cdot\frac{c-H^*+\delta}{\delta}\right)\\ &\le \left(\frac{80}{3}(c-H^*) + \frac{92}{3}\delta\right)\cdot L,\end{aligned}$$

并且当 $c < H(x) < c+\frac{\delta}{4}$，$f(x) = \frac{20}{3\delta}x^2$ 且 $f'(x) = \frac{40}{3\delta}x$。为简化以下计算，记 $\alpha = \frac{20}{3\delta}\beta$、$H-c=t$ 和 $c-H^*+\frac{\delta}{4} = mL$，对于 $\beta > \frac{3}{20\delta}$，我们可以写成

$$\begin{aligned}\|\nabla^2 G(x)\| &\le \frac{1}{\beta\frac{20}{3\delta}(H-c)^2+1}\left(L + \frac{\beta\frac{40}{3\delta}(H-c)}{\beta\frac{20}{3\delta}(H-c)^2+1}\left(c-H^*+\frac{\delta}{4}\right)\right)\\ &= \frac{1}{\alpha t^2+1}\left(1+\frac{2\alpha t}{\alpha t^2+1}m\right)L\\ &= \frac{\alpha t^2+2m\alpha t+1}{(\alpha t^2+1)^2}L = \frac{(\alpha t+\frac{1}{t})+2m\alpha}{(\alpha t+\frac{1}{t})^2}L \le \frac{m\alpha+\sqrt{\alpha}}{2\alpha}L\\ &= \frac{c-H^*}{2}+\frac{\delta}{8}+\frac{\delta}{2}L.\end{aligned}$$

因此，假设 H 是 L-光滑的，当 $\beta > \frac{1}{\delta^2} \vee \frac{3}{20\delta}$时，我们便有 $L^f \le \max\{(2(c-H^*)+3\delta)L, \left(\frac{80}{3}(c-H^*)+\frac{92}{3}\delta\right)L, \frac{c-H^*}{2}+\frac{\delta}{8}+\frac{\delta}{2}L\}$。所需结果得证。
■

A.3. 改进趋向全局最小值的收敛性的证明

本节重点是证明以下形式的大偏差概率上界：

$$\mathbb{P}\left(H(Y_k) > H^* + \epsilon\right),$$

如定理 8所述。首先，我们将 $\{X_k\}$和 $\{Y_k\}$的大偏差概率与 π_1 和 π_2的大偏差概率联系起来：

引理17 回顾 $\pi_1 = \pi_\beta^0 \propto e^{-\beta H}$ 和 $\pi_2 = \pi_\beta^f{}_{,\ c,\ 1} \propto e^{-\beta G}$。对于任意精度 $\epsilon > 0$，我们有

$$\left|\mathbb{P}\left(H(X_k) > H^* + \epsilon\right) - \int_{H>H^*+\epsilon} \pi_1 dx\right| \leq 2D_{TV}(\rho_k, \pi_1),$$
$$\left|\mathbb{P}\left(H(Y_k) > H^* + \epsilon\right) - \int_{H>H^*+\epsilon} \pi_2 dx\right| \leq 2D_{TV}(\rho_k^f, \pi_2).$$

其中

$$\left|\int_{H>H^*+\epsilon} \pi_1 dx - \int_{H>H^*+\epsilon} \pi_2 dx\right| \leq 2D_{TV}(\pi_1, \pi_2).$$

证明 首先，我们注意到

$$\begin{aligned}\mathbb{P}(H(X_k) > H^* + \epsilon) &= \int_{H>H^*+\epsilon} \rho_k dx \\ &= \int_{H>H^*+\epsilon} \rho_k dx - \int_{H>H^*+\epsilon} \pi_1 dx + \int_{H>H^*+\epsilon} \pi_1 dx,\end{aligned}$$

and

$$\left|\int_{H>H^*+\epsilon} \rho_k dx - \int_{H>H^*+\epsilon} \pi_1 dx\right| \leq \int_{H>H^*+\epsilon} |\rho_k - \pi_1|\, dx \leq 2D_{TV}(\rho_k, \pi_1),$$

由此可得

$$\left|\mathbb{P}\left(H(X_k) > H^* + \epsilon\right) - \int_{H>H^*+\epsilon} \pi_1 dx\right| \leq 2D_{TV}(\rho_k, \pi_1),$$

这就完成了 X_k 的情形。Y_k 的情形与 X_k 的情形类似。此外，我们看到

$$\left|\int_{H>H^*+\epsilon} \pi_1 dx - \int_{H>H^*+\epsilon} \pi_2 dx\right| \leq \int_{H>H^*+\epsilon} |\pi_1 - \pi_2|\, dx \leq 2D_{TV}(\pi_1, \pi_2),$$

这就完成了证明。 ■

引理17 给出了关于 k^{th} 迭代的分布与相应吉布斯分布之间概率差异的上界。此外，根据引理 15，我们观察到 $\int_{H>H^*+\epsilon} \pi_1 dx$ 和 $\int_{H>H^*+\epsilon} \pi_2 dx$ 渐近收敛于 0，当 $\beta \to \infty$时。为获得定量收敛率，我们给出以下引理:

引理18 在假设 2.2 至2.6 以及 3.1成立时，对于 $\beta > \frac{1}{\delta}$和 $0 < \epsilon < c$，如果存在 $I_2 < \infty$使得

$$\int_{\mathbb{R}^d} e^{-(H-H^*)} dx < I_2, \tag{A.8}$$

那么我们有

$$\int_{H>H^*+\epsilon} \pi_1 dx \leq \left(\frac{\beta L}{2\pi}\right)^{d/2} e^{-(\beta-1)\epsilon} I_2,$$
$$\int_{H>H^*+\epsilon} \pi_2 dx \leq \left(\frac{\beta L}{2\pi}\right)^{d/2} \left(2\mu(c \leq H \leq c+\delta)e^{-\beta(c-H^*)} + 2e^{-\beta(c-H^*)} I_1 + e^{-(\beta-1)\epsilon} I_2\right).$$

证明 可得

$$\begin{aligned}\int_{H>H^*+\epsilon}\pi_1 dx &= \frac{1}{Z_1}\int_{H>H^*+\epsilon} e^{-\beta H}dx\\ &= \frac{1}{Z_1 e^{H^*}}\int_{H>H^*+\epsilon} e^{-\beta(H-H^*)}dx\\ &\le \frac{1}{Z_1 e^{H^*}} e^{-(\beta-1)\epsilon}\int_{H>H^*+\epsilon} e^{-(H-H^*)}dx\\ &\le \left(\frac{\beta L}{2\pi}\right)^{d/2} e^{-(\beta-1)\epsilon}\int_{H>H^*+\epsilon} e^{-(H-H^*)}dx\\ &\le \left(\frac{\beta L}{2\pi}\right)^{d/2} e^{-(\beta-1)\epsilon} I_2,\end{aligned}$$

其中，最后一个不等式由 $Z_1 \ge e^{-\beta H^*}\left(\frac{2\pi}{\beta L}\right)^{d/2}$ 可得，如方程（A.3中先前所证明的）。此外，根据引理15 和17，我们可以直接计算得出

$$\begin{aligned}\int_{H>H^*+\epsilon}\pi_2 dx &\le \int_{H>H^*+\epsilon}\pi_1 dx + 2D_{TV}(\pi_1,\pi_2)\\ &\le \left(\frac{\beta L}{2\pi}\right)^{d/2} e^{-(\beta-1)\epsilon} I_2 + 2\left(\frac{\beta L}{2\pi}\right)^{d/2}\left(\mu(c\le H\le c+\delta)e^{-\beta(c-H^*)} + e^{-\beta(c-H^*)} I_1\right)\\ &= \left(\frac{\beta L}{2\pi}\right)^{d/2}\left(2\mu(c\le H\le c+\delta)e^{-\beta(c-H^*)} + 2e^{-\beta(c-H^*)} I_1 + e^{-(\beta-1)\epsilon} I_2\right),\end{aligned}$$

这就是所需结果。 ■

在许多一般情形下，条件（A.8）都可以验证。例如，当 H 满足二次增长时，分布 $\pi_1^0 \propto e^{-H}$ 满足次高斯尾部增长，这与（A.8）直接吻合。

利用引理17 和18，我们可以计算并界定朗之万蒙特卡洛算法在集合 $\{x : \ H(x) \ge H^* + \epsilon\}$ 时处于第 k^{th} 次迭代的概率。为定量展示景观修改带来的改进，我们将同时比较原始迭代与修正迭代的大偏差概率上界。证明 [定理 8] 作为引理17的直接推论，我们分别得到

$$\begin{aligned}\mathbb{P}(H(X_k) > H^*+\epsilon) &\le \int_{H>H^*+\epsilon}\pi_1 dx + 2D_{TV}(\rho_k,\pi_1),\\ \mathbb{P}(H(Y_k) > H^*+\epsilon) &\le \int_{H>H^*+\epsilon}\pi_2 dx + 2D_{TV}(\rho_k^f,\pi_2).\end{aligned}$$

结合引理18、命题6、方程（A.6）和（A.7），我们得到

$$\begin{aligned}\mathbb{P}(H(X_k) > H^*+\epsilon) &\le \int_{H>H^*+\epsilon}\pi_1 dx + 2D_{TV}(\rho_k,\pi_1)\\ &\le 2e^{-\frac{\eta k}{2\beta C_{LSI}}}\sqrt{2D_{KL}(\rho_0\|\pi_\beta^0)} + 8\sqrt{\eta d L^2\cdot C_{LSI}/\beta}\\ &\quad + \left(\frac{\beta L}{2\pi}\right)^{d/2} e^{-(\beta-1)\epsilon} I_2,\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(H(Y_k) > H^* + \epsilon) &\leq \int_{H>H^*+\epsilon} \pi_2 dx + 2D_{TV}(\rho_k^f, \pi_2) \\ &\leq 2e^{-\frac{\eta k}{2\beta C_{LSI}^f}} \sqrt{2D_{KL}(\rho_0^f \| \pi_{\beta,c,1}^f)} + 8\sqrt{\eta d L f^2 \cdot C_{LSI}^f/\beta} \\ &\quad + \left(\frac{\beta L}{2\pi}\right)^{d/2} \left(2\mu(c \leq H \leq c+\delta)e^{-\beta(c-H^*)} + 2e^{-\beta(c-H^*)} I_1 + e^{-(\beta-1)\epsilon} I_2\right).\end{aligned}$$

■

A.4. 约化超额风险至全局最小值的证明

本节首先研究修正迭代 $Y = \{Y_k\}$在修改后的目标函数 H_β^f、c、1 以及原始目标函数 H下的超额风险。对于修改后的目标函数下的超额风险，即 $\mathbb{E}_{Y_k\sim\rho_k^f}[H_{\beta,c,1}^f(Y_k)] - H^*$，借鉴（Raginsky 等，2017）中的思路，可将其分解为两部分：离散化误差和采样产生的近似误差，即

$$\mathbb{E}_{Y_k\sim\rho_k^f}[H_{\beta,c,1}^f(Y_k)] - H^* = \underbrace{\mathbb{E}_{Y_k\sim\rho_k^f}[H_{\beta,c,1}^f(Y_k)] - \mathbb{E}_{Y\sim\pi_2}[H_{\beta,c,1}^f(Y)]}_{\text{discretization error}} + \underbrace{\mathbb{E}_{Y\sim\pi_2}[H_{\beta,c,1}^f(Y)] - H^*}_{\text{sampling error}}. \tag{A.9}$$

在原始目标函数下修正迭代的超额风险，即 $\mathbb{E}_{Y_k\sim\rho_k^f}[H(Y_k)] - H^*$，使我们能够在相同标准下比较原始 LMC $X = \{X_k\}$ 与修正的 LMC $Y = \{Y_k\}$的超额风险。具体而言，我们将研究

$$\mathbb{E}_{Y_k\sim\rho_k^f}[H(Y_k)] - H^* = \underbrace{\mathbb{E}_{Y_k\sim\rho_k^f}[H(Y_k)] - \mathbb{E}_{X\sim\pi_1}[H(X)]}_{\text{modified discretization error}} + \underbrace{\mathbb{E}_{X\sim\pi_1}[H(X)] - H^*}_{\text{sampling error}}. \tag{A.10}$$

关于方程 (A.9) 中的离散化误差，我们可以用以下引理对其进行界定，该引理借鉴自 (Kinoshita 和 Suzuki 2022) 的定理 C.1。

引理19 假设 H 是 L-*smooth*，那么对于任意分布 ν_1 和 ν_2、随机变量$X \sim \nu_1$ 和 $Y \sim \nu_2$，有

$$\mathbb{E}_{X\sim\nu_1}[H(X)] - \mathbb{E}_{Y\sim\nu_2}[H(Y)] \leq L W_2^2(\nu_1, \nu_2) + \mathbb{E}_{Y\sim\nu_2}[H(Y)] - H^*.$$

证明 令 $X \sim \nu_1$和 $Y \sim \nu_2$为最优耦合，使得 $W_2^2(\nu_1,\nu_2) = \mathbb{E}[\|X - Y\|^2]$，且 $\lambda \in \Pi(\nu_1,\nu_2)$ 为最优分布。这里，$\Pi(\nu_1,\nu_2)$ 是满足 $\xi\left(A, \mathbb{R}^d\right) = \nu_1(A)$ 和 $\xi\left(\mathbb{R}^d, B\right) = \nu_2(B)$ 的 ξ 上的概率测度集合，其中 $\mathcal{B}\left(\mathbb{R}^d\right) \otimes \mathcal{B}\left(\mathbb{R}^d\right)$ 对于 $\forall A, B \in \mathcal{B}\left(\mathbb{R}^d\right)$,

其中我们记 $\mathcal{B}\left(\mathbb{R}^d\right)$ 为 $\mathbb{R}^d$ 上的博雷尔σ代数。对 Y 进行泰勒展开，可得 e

$$\begin{aligned} H(X)-H(Y) &\leq \|\nabla H(Y)\|\|X-Y\|+\frac{L}{2}\|X-Y\|^2 \\ &\leq \frac{L}{2}\|X-Y\|^2+\frac{1}{2L}\|\nabla H(Y)\|^2+\frac{L}{2}\|X-Y\|^2 \\ &= L\|X-Y\|^2+\frac{1}{2L}\|\nabla H(Y)\|^2 \\ &\leq L\|X-Y\|^2+H(Y)-H^*, \end{aligned}$$

其中最后一个不等式利用了引理 16。对 λ 两边取期望，可得

$$\mathbb{E}_{X\sim\nu_1}[H(X)]-\mathbb{E}_{Y\sim\nu_2}[H(Y)]\leq LW_2^2(\nu_1,\nu_2)+\mathbb{E}_{Y\sim\nu_2}[H(Y)]-H^*.$$

■

注意，当 $\beta>\beta_0$ =最大值$\left\{\frac{1}{\delta^2},\frac{3}{20\delta},\frac{1}{\delta}\right\}$时，根据定理 7，我们知道修改后的函数 $H_{\beta,c,1}^f$ 是 L^f-光滑的，因此由引理 19可直接得到推论。

引理 20 在假设 2.2 至 2.6 以及 3.1下，对于 $\beta>\beta_0$ =最大值$\left\{\frac{1}{\delta^2},\frac{3}{20\delta},\frac{1}{\delta}\right\}$，我们有推论 e

$$\mathbb{E}_{Y_k\sim\rho_k^f}[H_{\beta,c,1}^f(Y_k)]-H^*\leq L^fW_2^2\left(\rho_k^f,\pi_{\beta,c,1}^f\right)+2\left(\mathbb{E}_{Y\sim\pi_2}[H_{\beta,c,1}^f(Y)]-H^*\right),$$

其中 L^f =最大值 $\left\{(2(c-H^*)+3\delta)L,\left(\frac{80}{3}(c-H^*)+\frac{92}{3}\delta\right)L,\frac{c-H^*}{2}+\frac{\delta}{8}+\frac{\delta}{2}L\right\}$。

证明 由方程 (A.9)，可得

$$\begin{aligned} \mathbb{E}_{Y_k\sim\rho_k^f}[H_{\beta,c,1}^f(Y_k)]-H^* &= \mathbb{E}_{Y_k\sim\rho_k^f}[H_{\beta,c,1}^f(Y_k)]-\mathbb{E}_{Y\sim\pi_2}[H_{\beta,c,1}^f(Y)]+\mathbb{E}_{Y\sim\pi_2}[H_{\beta,c,1}^f(Y)]-H^* \\ &\leq L^fW_2^2\left(\rho_k^f,\pi_{\beta,c,1}^f\right)+2\left(\mathbb{E}_{Y\sim\pi_2}[H_{\beta,c,1}^f(Y)]-H^*\right), \end{aligned}$$

其中，该不等式可通过将 $\nu_1=\rho_k^f$ 和 $\nu_2=\pi_2=\pi_\beta^f$，c，1 代入引理19得到。 ■

关于 $H_{\beta,c,1}^f$ 相对于 $\pi_{\beta,c,1}^f$的期望，我们有以下引理。

引理 21 在假设 2.2 至 2.6 以及 3.1下，对于给定的 $0<r<\frac{3\gamma}{2L_1}$，且在 $B_r(x^*)$ 中除 x^* 外不存在其他局部极小值，设 $\max_{B_r(x^*)}H(x)-H^*=\delta_r\leq c-H^*$，且 $\min_{B_r'(x^*)}H(x)-H^*=\Delta_r>0$。那么对于步长 $0<\eta\leq\frac{\beta C_{LSI}^f}{16L^{f^2}}$，$\beta>\beta_0$ =最大值 $\left\{\frac{1}{\delta},\frac{1}{\delta^2},\frac{3}{20\delta}\right\}$，若存在 $I_3=I_3(c,\delta)<\infty$ 使得

$$\int_{H>c+\delta}\left(c-H^*+\delta+\delta\ln\frac{H-c}{\delta}\right)\cdot\frac{2\delta}{H-c}dx<I_3,$$

于是我们有

$$\begin{aligned} \mathbb{E}_{Y\sim\pi_2}[H_{\beta,c,1}^f(Y)]-H^* \leq{}& \frac{3}{2}\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L^f}{\pi}\right)^{d/2}\cdot\frac{1}{\beta} \\ &+\left(\frac{\beta L^f}{2\pi}\right)^{d/2}\left((c-H^*+\delta)\mu(H\leq c+\delta)e^{-\beta\Delta_r}+e^{-\beta(c-H^*)}\cdot I_3\right), \end{aligned}$$

其中 L^f =最大值 $\left\{(2(c-H^*)+3\delta)L,\left(\frac{80}{3}(c-H^*)+\frac{92}{3}\delta\right)L,\frac{c-H^*}{2}+\frac{\delta}{8}+\frac{\delta}{2}L\right\}$。

证明 利用（Łapi´nski 2019）中的思路，我们写作

$$\begin{aligned}\mathbb{E}_{Y\sim\pi_2}[H^f_{\beta,c,1}(Y)] - H^* &= \frac{1}{Z_2}\int_{\mathbb{R}^d} Ge^{-\beta G}dx - H^* \\ &= \frac{1}{Z_2}\int_{\mathbb{R}^d}(G-H^*)e^{-\beta G}dx \\ &= \frac{1}{Z_2}\left(\int_{B_r(x^*)}(G-H^*)e^{-\beta G}dx + \int_{B_r'(x^*)}(G-H^*)e^{-\beta G}dx\right),\end{aligned} \tag{A.11}$$

其中

$$Z_2 = \int_{\mathbb{R}^d} e^{-\beta G}dx \geq e^{-\beta H^*}\int_{\mathbb{R}^d}\exp\left(-\frac{\beta L^f}{2}\|x-x^*\|^2\right)dx = e^{-\beta H^*}\left(\frac{2\pi}{\beta L^f}\right)^{d/2},$$

因为 $G(x) \leq H^* + \frac{L^f}{2}\|x-x^*\|^2$。

注意，当 $\delta_r \leq c$时， $B_r(x^*) \subset \{x:\ H(x) \leq c\}$，因此，当 $x \in B_r(x^*)$ 时， H 和 G的三阶导数界保持不变，这意味着 L_1 是$B_r(x^*)$中的共享参数。因此，我们有

$$\left|G - H^* - \frac{1}{2}\|x-x^*\|_K^2\right| \leq \frac{L_1}{6}\|x-x^*\|^3, \quad \forall x \in B_r(x^*).$$

我们接下来分别处理右侧的两个项（A.11）。对于第一项，我们看到

$$\begin{aligned}&\int_{B_r(x^*)}(G-H^*)e^{-\beta G}dx \\ &\quad\leq e^{-\beta H^*}\int_{B_r(x^*)}\left(\frac{1}{2}\|x-x^*\|_K^2 + \frac{L_1}{6}\|x-x^*\|^3\right)\exp\left(-\beta\left(\frac{1}{2}\|x-x^*\|_K^2 - \frac{L_1}{6}\|x-x^*\|^3\right)\right)dx \\ &\quad\leq e^{-\beta H^*}\int_{B_r(x^*)}\left(\frac{3}{4}\|x-x^*\|_K^2\right)\exp\left(-\frac{\beta}{4}\|x-x^*\|_K^2\right)dx \\ &\quad\leq e^{-\beta H^*}\frac{3}{4}\int_{\mathbb{R}^d}\|x-x^*\|_K^2\exp\left(-\frac{\beta}{4}\|x-x^*\|_K^2\right)dx \\ &\quad= e^{-\beta H^*}\frac{3}{4}\frac{d\omega_d}{\sqrt{\det K}}\int_0^\infty r^{d+1}e^{-\frac{\beta}{4}r^2}dr = e^{-\beta H^*}\frac{3}{8}\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{4}{\beta}\right)^{d/2+1},\end{aligned}$$

其中， ω_d 是 $\mathbb{R}^d$中 d维单位球的体积。第二个不等式由$r < \frac{3\gamma}{2L_1}$得出。因此，

$$\frac{1}{Z_2}\int_{B_r(x^*)}(G-H^*)e^{-\beta G}dx \leq \frac{3}{2}\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L^f}{\pi}\right)^{d/2}\cdot\frac{1}{\beta} = \mathcal{O}(\beta^{-1}).$$

对于右侧的第二项（A.11），对 $B_r'(x^*)$ 的积分可分为以下三部分：

$$\begin{aligned}
&\int_{B_r'(x^*)}(G-H^*)e^{-\beta G}dx \\
&\quad=\int_{\{H<c\}\cap B_r'(x^*)}(G-H^*)e^{-\beta G}dx+\int_{c\le H\le c+\delta}(G-H^*)e^{-\beta G}dx+\int_{H>c+\delta}(G-H^*)e^{-\beta G}dx \\
&\quad\le(c-H^*)\mu(H<c)e^{-\beta(\Delta_r+H^*)}+(c-H^*+\delta)\mu(c\le H\le c+\delta)e^{-\beta c}+\int_{H>c+\delta}(G-H^*)e^{-\beta G}dx \\
&\quad\le(c-H^*+\delta)\mu(H\le c+\delta)e^{-\beta(\Delta_r+H^*)}+\int_{H>c+\delta}(G-H^*)e^{-\beta G}dx.
\end{aligned}$$

注意，当 $H(x)>c+\delta$时，由于 $\beta>\beta_0\ge\frac{1}{\delta}$，根据方程 (A.1) 可得

$$\begin{aligned}
c+\frac{1}{\beta}\ln\frac{H-c}{2\delta}\le G(x)&\le c+\delta+\frac{1}{\beta}\ln\frac{H-c}{\delta}\\
&\le c+\delta+\delta\ln\frac{H-c}{\delta}.
\end{aligned}$$

将这些结果代入上面的积分，得到

$$\begin{aligned}
\int_{H>c+\delta}(G-H^*)e^{-\beta G}dx&\le\int_{H>c+\delta}\left(c-H^*+\delta+\delta\ln\frac{H-c}{\delta}\right)e^{-\left(\beta c+\ln\frac{H-c}{2\delta}\right)}dx\\
&=e^{-\beta c}\int_{H>c+\delta}\left(c-H^*+\delta+\delta\ln\frac{H-c}{\delta}\right)\cdot\frac{2\delta}{H-c}dx\\
&<e^{-\beta c}\cdot I_3,
\end{aligned}$$

因此

$$\frac{1}{Z_2}\int_{B_r'(x^*)}(G-H^*)e^{-\beta G}dx\le\left(\frac{\beta L^f}{\pi}\right)^{d/2}\left((c-H^*+\delta)\mu(H\le c+\delta)e^{-\beta\Delta_r}+e^{-\beta(c-H^*)}\cdot I_3\right).$$

结合前一部分，即可得到所需结果。 ■

注意，根据假设2.1中的 Talagrand不等式，可以通过 KL 散度估计修正的 LMC 的第 k^{th} 次迭代 ρ_k^f 与其对应的修正吉布斯分布 $\pi_{\beta,c,1}^f$ 之间的 Wasserstein距离；进而利用命题6，即可得到$H_{\beta,c,1}^f$下的超额风险。

定理22 在假设2.2 至2.6、2.1 和3.1下，对于给定的$0<r<\frac{3\gamma}{2L_1}$，x^* 是 $B_r(x^*)$中 H 的唯一局部极小值，设 $\max_{B_r(x^*)}H(x)-H^*=\delta_r\le c-H^*$。假设

$\min_{B_r'(x^*)}H(x)-H^*=\Delta_r>0$，则对于步长 $0<\eta\le\frac{\beta C_{LSI}^f}{16L^{f^2}}$, $\beta>\beta_0=\max\left\{\frac{1}{\delta},\frac{1}{\delta^2},\frac{3}{20\delta}\right\}$，如果存在 $I_3=I_3(c,\delta)<\infty$ 使得

$$\int_{H>c+\delta}\left(c-H^*+\delta+\delta\ln\frac{H-c}{\delta}\right)\cdot\frac{2\delta}{H-c}dx<I_3,$$

则有

$$\mathbb{E}_{Y_k\sim\rho_k^f}[H^f_{\beta,c,1}(Y_k)] - H^* \le 2L^f \cdot C^f_{LSI}\left(e^{-\frac{\eta k}{\beta C^f_{LSI}}} D_{KL}(\rho_0^f\|\pi^f_{\beta,c,1}) + 8\eta d {L^f}^2 \cdot C^f_{LSI}/\beta\right)$$
$$+ 3\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L^f}{\pi}\right)^{d/2}\cdot\frac{1}{\beta}$$
$$+ 2\left(\frac{\beta L^f}{2\pi}\right)^{d/2}\left((c-H^*+\delta)\mu(H\le c+\delta)e^{-\beta\Delta_r} + e^{-\beta(c-H^*)}\cdot I_3\right),$$

其中 $C^f_{LSI} = \mathcal{O}\left(\beta M + \underline{H^{\frac{\beta M^2}{m_2}}()-c}\right)$，并

且

$$L^f \le \max\left\{(2(c-H^*)+3\delta)L, \left(\frac{80}{3}(c-H^*)+\frac{92}{3}\delta\right)L, \frac{c-H^*}{2}+\frac{\delta}{8}+\frac{\delta}{2}L\right\}.$$

证明 首先，根据推论 20、命题 6 和引理21，利用假设 2.1中的 Talagrand不等式，我们有

$$\mathbb{E}_{Y_k\sim\rho_k^f}[H^f_{\beta,c,1}(Y_k)] - H^* \le \frac{2L^f}{T_a^f}\left(e^{-\frac{\eta k}{\beta C^f_{LSI}}} D_{KL}(\rho_0^f\|\pi^f_{\beta,c,1}) + 8\eta d {L^f}^2 \cdot C^f_{LSI}/\beta\right) \tag{A.12}$$
$$+ 3\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L^f}{\pi}\right)^{d/2}\cdot\frac{1}{\beta}$$
$$+ 2\left(\frac{\beta L^f}{2\pi}\right)^{d/2}\left((c-H^*+\delta)\mu(H\le c+\delta)e^{-\beta\Delta_r} + e^{-\beta(c-H^*)}\cdot I_3\right),$$

其中 T_a^f满足：对于任意分布 ρ，都存在

$$\frac{T_a^f}{2}W_2^2(\rho,\pi^f_{\beta,c,1}) \le D_{KL}(\rho\|\pi^f_{\beta,c,1}).$$

此外，根据（Otto 和 Villani，2000）中的定理 1，我们注意到，若分布 ν满足常数为 C_{LSI}的对数 Sobolev 不等式，则也满足常数为 $T_a = \frac{1}{C_{LSI}}$的 Talagrand 不等式。因此，我们将 $\frac{1}{C^f_{LSI}}$代入 T_a^f方程（A.12），即可得到所需结果。■

接下来，我们按照方程（ $\mathbb{E}_{Y_k\sim\rho_k^f}[H(Y_k)] - H^*$ A.10）中的规定，推导 的上界，并将这些结果与给出 $\mathbb{E}_{X_k\sim\rho_k}[H(X_k)] - H^*$ 上界的经典结果进行比较。在此背景下，我们首先量化原始吉布斯分布与修正后的吉布斯分布之间的 KL 散度和 Wasserstein 距离：

引理 23 在假设 2.2 至 2.6、2.1 和 3.1下，对于给定的 $0 < r < \frac{3\gamma}{2L_1}$，且 x^* 是 H 在 B_r (x^*) 中的唯一局部极小值，假设 $\max_{B_r\ (x^*)} H(x) - H^* = \delta_r \le c - H^*$。设 $\min_{B_r'\ (x^*)} H(x) - H^* = \Delta_r > 0$，则有

$$D_{KL}(\pi^0_\beta\|\pi^f_{\beta,c,1}) \le \left(\frac{\beta L}{2\pi}\right)^{d/2}\left(\mu(c<H\le c+\delta)e^{-\beta(c-H^*)} + e^{-\beta(c-H^*)}I_1\right),$$

and

$$W_2^2(\pi_\beta^0, \pi_{\beta,c,1}^f) \le 2C_{LSI}^f \left(\frac{\beta L}{2\pi}\right)^{d/2} \left(\mu(c < H \le c+\delta)e^{-\beta(c-H^*)} + e^{-\beta(c-H^*)} I_1\right),$$

其中 $C_{LSI}^f = \mathcal{O}\left(\beta M + \underline{{}_{H\ m_2}^{\beta M^2}\ (\ )}_{-c}\right)$。

证明 回顾 $\pi_1 = \pi_\beta^0 \propto e^{-\beta H}$, $\pi_2 = \pi_{\beta,c,1}^f \propto e^{-\beta G}$ 和 $G \le H$，我们有

$$\begin{aligned}
D_{KL}(\pi_1 \| \pi_2) &= \int_{\mathbb{R}^d} \pi_1 \ln\left(\frac{\pi_1}{\pi_2}\right) dx \\
&= \int_{\mathbb{R}^d} \frac{e^{-\beta H}}{Z_1} \ln\left(\frac{e^{-\beta H}/Z_1}{e^{-\beta G}/Z_2}\right) dx \\
&= \ln\left(\frac{Z_2}{Z_1}\right) - \beta \int_{\mathbb{R}^d} (H-G)\frac{e^{-\beta H}}{Z_1} dx \\
&\le \ln\left(\frac{Z_2}{Z_1}\right) \le \frac{Z_2 - Z_1}{Z_1},
\end{aligned}$$

其中，在最后一个不等式中我们利用了 $Z_2/Z_1 \ge 1$。现在，我们观察到 Z_1 和 Z_2 均可分解为

$$\begin{aligned}
Z_1 &= \int_{H \le c} e^{-\beta H} dx + \int_{H > c} e^{-\beta H} dx, \\
Z_2 &= \int_{H \le c} e^{-\beta G} dx + \int_{c < H \le c+\delta} e^{-\beta G} dx + \int_{H > c+\delta} e^{-\beta G} dx.
\end{aligned}$$

对于上述方程的第一项，由于 $r < \frac{3\gamma}{2L_1}$,

$$\begin{aligned}
\int_{H \le c} e^{-\beta G} dx = \int_{H \le c} e^{-\beta H} dx &= \int_{B_r(x^*)} e^{-\beta H} dx + \int_{B_r'(x^*) \cap \{H \le c\}} e^{-\beta H} dx \\
&\le e^{-\beta H^*} \left(\int_{B_r(x^*)} \exp\left(-\frac{\beta}{4}\|x - x^*\|_K^2\right) dx + \mu(H \le c) e^{-\beta \Delta_r}\right) \\
&\le e^{-\beta H^*} \left(\int_{\mathbb{R}^d} \exp\left(-\frac{\beta}{4}\|x - x^*\|_K^2\right) dx + \mu(H \le c) e^{-\beta \Delta_r}\right) \\
&= e^{-\beta H^*} \left(\frac{d\omega_d}{\sqrt{\det K}} \int_0^\infty r^{d-1} e^{-\frac{\beta}{4} r^2} dr + \mu(H \le c) e^{-\beta \Delta_r}\right) \\
&= e^{-\beta H^*} \left(\frac{1}{2} \frac{d\omega_d \Gamma(\frac{d}{2})}{\sqrt{\det K}} \left(\frac{4}{\beta}\right)^{d/2} + \mu(H \le c) e^{-\beta \Delta_r}\right),
\end{aligned}$$

这给出了 $\int_{H \le c} e^{-\beta H}\, dx$ 的上界。对于 Z_1 的下界，根据方程（A.3），我们有

$$Z_1 = \int_{\mathbb{R}^d} e^{-\beta H} dx \ge e^{-\beta H^*} \left(\frac{2\pi}{\beta L}\right)^{d/2}.$$

此外，我们注意到

$$\begin{aligned}\int_{c<H\leq c+\delta} e^{-\beta G}dx+\int_{H>c+\delta} e^{-\beta G}dx &\leq \mu(c<H\leq c+\delta)e^{-\beta c}+e^{-\beta c}\int_{H>c+\delta}\frac{2\delta}{H-c}dx\\ &\leq \mu(c<H\leq c+\delta)e^{-\beta c}+e^{-\beta c}I_1.\end{aligned}$$

综合以上结果可得

$$\begin{aligned}D_{KL}(\pi_1\|\pi_2)\leq\frac{Z_2-Z_1}{Z_1}&\leq\frac{\int_{c<H\leq c+\delta}e^{-\beta G}dx+\int_{H>c+\delta}e^{-\beta G}dx}{Z_1}\\ &\leq\left(\frac{\beta L}{2\pi}\right)^{d/2}\left(\mu(c<H\leq c+\delta)e^{-\beta(c-H^*)}+e^{-\beta(c-H^*)}I_1\right)\\ &=\mathcal{O}\left(\beta^{d/2}e^{-\beta(c-H^*)}\right).\end{aligned}$$

再次使用Talagrand不等式，并注意到 T_a^f可以用 $\frac{1}{C_{LSI}^f}$替代，即可类似地得到Wasserstein距离的情形。 ■

由引理16 和 23，我们对 $H(Y_k)$的期望进行界定。

证明 [定理9]的证明 代入 $\nu_1=\rho$ $\overset{f}{k}$ 和 $\nu_2=\pi_\beta^0$ 代入引理 19，可得

$$\begin{aligned}\mathbb{E}_{Y_k\sim\rho_k^f}[H(Y_k)]-H^*&=\mathbb{E}_{Y_k\sim\rho_k^f}[H(Y_k)]-\mathbb{E}_{X\sim\pi_1}[H(X)]+\mathbb{E}_{X\sim\pi_1}[H(X)]-H^*\\ &\leq LW_2^2(\rho_k^f,\pi_\beta^0)+2\left(\mathbb{E}_{X\sim\pi_1}[H(X)]-H^*\right).\end{aligned}\tag{A.13}$$

根据命题 6 和引理 23，利用关于 π_{β}^{f}、c、1的 Talagrand不等式，我们得到

$$\begin{aligned}W_2^2(\rho_k^f,\pi_\beta^0)&\leq 2\left(W_2^2(\rho_k^f,\pi_{\beta,c,1}^f)^2+W_2^2(\pi_\beta^0,\pi_{\beta,c,1}^f)^2\right)\\ &\leq 4C_{LSI}^f\left(e^{-\frac{\eta k}{\beta C_{LSI}^f}}D_{KL}(\rho_0^f\|\pi_{\beta,c,1}^f)+8\eta dL^{f2}\cdot C_{LSI}^f/\beta\right)\\ &\quad+4C_{LSI}^f\left(\frac{\beta L}{2\pi}\right)^{d/2}\left(\mu(c<H\leq c+\delta)e^{-\beta(c-H^*)}+e^{-\beta(c-H^*)}I_1\right).\end{aligned}$$

至于 $\mathbb{E}_{X\sim\pi_1}[H(X)],$，计算方法与引理21的证明中所述方法类似。具体而言，可以计算得出

$$\begin{aligned}\mathbb{E}_{X\sim\pi_1}[H(X)]-H^*&=\frac{1}{Z_1}\left(\int_{H\leq c}(H-H^*)e^{-\beta H}dx+\int_{H>c}(H-H^*)e^{-\beta H}dx\right)\\ &\leq\frac{3}{2}\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L}{\pi}\right)^{d/2}\cdot\frac{1}{\beta}\\ &\quad+\left(\frac{\beta L}{2\pi}\right)^{d/2}\left((c-H^*)\mu(H\leq c)e^{-\beta\Delta_r}+\int_{H>c}(H-H^*)e^{-\beta(H-H^*)}dx\right)\\ &\leq\frac{3}{2}\frac{d\omega_d\Gamma(\frac{d}{2}+1)}{\sqrt{\det K}}\left(\frac{2L}{\pi}\right)^{d/2}\cdot\frac{1}{\beta}\\ &\quad+\left(\frac{\beta L}{2\pi}\right)^{d/2}\left((c-H^*)\mu(H\leq c)e^{-\beta\Delta_r}+e^{-(\beta-1)(c-H^*)}I_4\right).\end{aligned}$$

将上述结果代入方程（A.13），即可得到所需结果。 ■

附录 B。常用记号表

表 1总结了本文中经常使用的重要记号。

记号	描述	定义
$G(x)$	修改后的目标函数	$G(x) = H^f_{\beta,c,1}(x)$
X_k	k^{th} 原始 LMC 的迭代	
Y_k	k^{th} 修正的 LMC 的迭代	
ρ_k	X_k的分布	$X_k \sim \rho_k$
ρ_k^f	Y_k的分布	$Y_k \sim \rho_k^f$
π_β^0	原始动力学的吉布斯分布	$\pi_\beta^0 \propto e^{-\beta H}$
$\pi^f_{\beta,c,1}$	修改后的动力学的吉布斯分布	$\pi^f_{\beta,c,1} \propto e^{-\beta G}$
π_1	π_β^0的简写记号	$\pi_1 = \pi_\beta^0$
π_2	简写记号$\pi^f_{\beta,c,1}$	$\pi_2 = \pi^f_{\beta,c,1}$
Z_1	π_1的归一化常数	$Z_1 = \int e^{-\beta H} dx$
Z_2	π_2的归一化常数	$Z_2 = \int e^{-\beta G} dx$
C_{PI}	π_1的 Poincar´e 常数	
C_{LSI}	π_1的对数 Sobolev 常数	
C_{PI}^f	π的 Poincar´e 常数$_2$	
C_{LSI}^f	π_2的对数 Sobolev 常数	
L	H 是 L-光滑的	$\|\nabla H(x) - \nabla H(y)\| \le L\|x-y\|$
L^f	G 是 L^f-光滑的	$\|\nabla G(x) - \nabla G(y)\| \le L^f\|x-y\|$
L_1	$\nabla^2 H$ 是 L_1-Lipschitz连续的	$\|\nabla^2 H(x) - \nabla^2 H(y)\| \le L_1\|x-y\|$
T_a	π_1的塔拉格朗日常数	
T_a^f	π_2的塔拉格朗日常数	
$\mathcal{S} = \{m_k\}_{k=1}^S$	H的局部极小值	$H(m_1) < H(m_2) \le H(m_i)\ 3 \le i \le S$
$\mathcal{S}^f = \{m_k^f\}_{k=1}^{S^f}$	G的局部极小值	已证明，$m_k^f = m_k, \forall k$
$s_{1,2}$	m_1 与 m_2之间的鞍点	
$s_{1,2}^f$	m_1^f 与 m_2^f 之间的鞍点	已证明 $s_{1,2}^f = s_{1,2}$ 和 $S = S^f$
x^*	H （以及 G）的唯一全局极小点	$x^* = m_1$
H^*	H的全局最小值	$H^* = H(x^*)$
M	H的能垒	$M = H(s_{1,2}) - H(m_2)$
M^f	G的能垒	$M^f = G(s'_{1,2}) - G(m'_2)$
K	H 在 m_1处的海森矩阵	$K = \nabla^2 H(m_1)$
γ	$\nabla^2 H(m_1)$的最小特征值	$\gamma = \lambda_{min}(K)$
$\|x\|_A$	x 相对于矩阵 A的矩阵范数	$\|x\|_A^2 = \frac{1}{2}x^T Ax$
$\|B\|$	矩阵 B的算子范数	
ω_d	$\mathbb{R}^d$中的单位球体积	
μ	$\mathbb{R}^d$中的勒贝格测度	
$\Pi(\rho,\nu)$	ρ 与 ν之间可能的耦合集合	
$D_{KL}(\rho\|\nu)$	ρ 相对于 ν的 KL 散度	$D_{KL}(\rho\|\nu) = \int \rho \ln \frac{\rho}{\nu} dx$
$D_{TV}(\rho,\nu)$	ρ 与 ν之间的全变差距离	$D_{TV}(\rho,\nu) = \frac{1}{2}\int \|\rho - \nu\| dx$
$W_2(\rho,\nu)$	ρ与 ν之间的Wasserstein距离	$W_2^2(\rho,\nu) = \inf_{\xi\in\Pi(\rho,\nu)} \int \|x-y\|^2 \xi(dx,dy)$

表 1：常用记号表