arXiv:2301.01409v1 [stat.ME] 4 Jan 2023

黎曼流形与拉格朗日蒙特卡洛的修正变体中的几何遍历性

James A. Brofos[1*]、Vivekananda Roy[2] 和 Roy R. Lederma n[1]

[1]耶鲁大学统计与数据科学系，美国康涅狄格州纽黑文市 Hillhouse 大道 24 号，邮编 06510。[2]爱荷华州立大学统计系，美国爱荷华州艾姆斯市 Osborn 大道 2438 号，邮编 50011。

*通讯作者。电子邮箱：james.brofos@yale.edu；贡献者：vroy@iastate.edu；roy.lederman@yale.edu；

摘要

黎曼流形哈密顿蒙特卡罗（RMHMC）和拉格朗日蒙特卡罗（LMC）已成为强大的贝叶斯推断方法。与欧几里得哈密顿蒙特卡罗（EHMC）和梅特罗波利斯调整朗之万算法（MALA）不同，这些黎曼算法的几何遍历性尚未得到广泛研究。另一方面，已有研究近期证明，在某些条件下，流形梅特罗波利斯调整朗之万算法（MMALA）具有几何遍历性。本研究探讨将 LMC 和 RMHMC 的转移核与 MMALA 混合，以使所得方法具备“继承而来”的几何遍历性理论。我们基于单步 HMC 与 MALA 之间的类比，说明采用这种混合核的动机。随后，我们在若干贝叶斯推断基准任务上评估原始转移核和修改后的转移核。

关键词：马尔可夫链蒙特卡罗、黎曼流形哈密顿蒙特卡罗、统计计算、拉格朗日蒙特卡洛

1 引言

贝叶斯推断旨在将主观信息来源与观测数据结合起来。通过指定先验信念，并正确捕捉随机系统中的不确定性来源，可以采用贝叶斯方法来刻画和推理后验分布中的不确定性。形式上，后验分布是一个概率分布$\Pi : \mathfrak{B}(\mathbb{R}^m) \to [0,1]$，其密度为 $\pi : \mathbb{R}^m \to \mathbb{R}_+$。在贝叶斯推断的实际应用中，人们关注的是关于 Π 的可积函数 $h: \mathbb{R}^m \to \mathbb{R}$ 的期望计算；也就是说，我们希望计算 $\mathbb{E}_{q\sim\Pi}[h(q)]$。由于 π 通常难以处理，这些期望无法以闭式形式求得，因此常使用基于 π 样本的蒙特卡罗方法来近似这些均值。例如，可以通过选取适当的期望来计算后验分布的均值和方差。因此，贝叶斯推断中的一个核心问题是从后验分布中生成样本。

一种常用的蒙特卡洛方法是利用马尔可夫链蒙特卡罗（MCMC）技术生成后验分布的近似样本。在 MCMC 中，给定链的某个初始状态 q^0，通过归纳方式生成一系列随机变量（q^1，q^2，.....）：条件分布 $q^n|q^{n-1}$，.....，q^0 满足马尔可夫性质。在某些理想的正则性条件下，随机变量 q^n 的分布将渐近收敛到 Π。确定收敛速率，或至少给出该速率的上界，有助于量化马尔可夫链（q^1，q^2，.....）向目标分布混合的速度。当收敛速率呈几何级数般快速时，马尔可夫链具有很强的稳定性，此时估计量 $n^{-1}\sum_{i=1}^{n} h(q^i)$ of $\mathbb{E}_{q\sim\Pi}[h(q)]$ 可以配备

在额外的技术假设下，中心极限定理成立。该中心极限定理具有重要的实际意义，因为它使我们能够使用 MCMC 估计的渐近有效标准误，而这些标准误又可用于决定马尔可夫链需要运行多长时间 [Roy，2020]。

最受欢迎的 MCMC 方法之一是欧几里得哈密顿蒙特卡罗（EHMC）[Neal，2010b]。在 EHMC 中，人们使用数值积分器计算哈密顿运动方程的近似解，并确保由此得到的马尔可夫链在由相空间构成的位置变量 $q \in \mathbb{R}^m$ 和辅助动量变量 $p \in \mathbb{R}^m$组成的空间中满足细致平衡。哈密顿蒙特卡罗的收敛速率由Livingstone 等人 [2016] 在对数密度 [的条件下确立（另见 Durmus 等人，2020]）。然而，传统 EHMC 难以处理具有多个空间尺度的分布 [Pourzanjani 和 Petzold，2019年，Betancourt，2012]这一观察促成了 HMC 几何方法的发展，具体而言，就是Girolami 和 Calderhead [2011]提出的黎曼流形哈密顿蒙特卡罗（RMHMC）方法。

与欧几里得哈密顿蒙特卡洛不同，黎曼流形哈密顿蒙特卡洛会适应后验分布的二阶结构，从而使其提议能够沿着后验分布的方向对齐，

表现出最大的局部变化。然而，RMHMC 所采用的哈密顿量形式复杂，因此需要使用复杂的数值积分器，其成本显著高于 EHMC 所用的数值积分器。在 Lan 等人 [2015], 引入拉格朗日蒙特卡洛（LMC）后，这一问题得到了一定程度的缓解；该方法能够构建数值积分过程更高效的马尔可夫链蒙特卡洛方法，同时继续利用二阶几何信息，以便高效地遍历后验分布。RMHMC 和 LMC 均可视为欧几里得哈密顿蒙特卡洛算法的推广，它们将有关后验分布的二阶几何信息纳入马尔可夫链转移核。纳入二阶信息使这些马尔可夫链蒙特卡洛几何方法能够更高效地探索目标分布的典型区域。

然而，迄今为止，RMHMC 和 LMC 均尚未建立几何遍历性理论。本文致力于确立修正版本的 RMHMC 和 LMC 的几何遍历性。在 RMHMC 和 LMC 的实现中，通常会将积分步数随机化，并以正概率执行一步积分。对于欧几里得哈密顿蒙特卡洛，在 Livingstone 等人 [2016]的研究中，单步哈密顿蒙特卡洛与梅特罗波利斯调整朗之万算法（MALA）之间的等价关系，对于确立哈密顿蒙特卡洛的几何遍历性至关重要。我们的方法分为两部分。首先，我们针对 RMHMC 和 LMC 提出一种简单的修正，其动机来自单步欧几里得哈密顿蒙特卡洛与 MALA 之间的独特对应关系。具体而言，我们提出，不采用 RMHMC 或 LMC 对应的转移核执行单个积分步，而改为采用流形梅特罗波利斯调整朗之万算法（MMALA）的转移核。近期 Roy 和 Zhang [2023] 的研究给出了 MMALA 几何遍历的条件。一旦证明 RMHMC 和 LMC 的转移核（积分步数固定且非随机）在所需意义下可逆，便可通过继承性质，为这些修正后的 RMHMC 和 LMC 变体赋予几何遍历性理论。第 3 节将介绍这一构造。第 4 节将对所提出的 RMHMC 和 LMC 修正进行数值评估，并特别关注应用 MMALA 转移核的概率。不过，我们先在第 2 节概述数值积分和马尔可夫链所需的数学概念。

2 预备知识

在第 2.1 节中，我们回顾几何马尔可夫链蒙特卡洛方法中用于贝叶斯推断的广义蛙跳法和拉格朗日蛙跳法。第 2.2节讨论基于对合以及基于朗之万扩散离散化的马尔可夫链。我们还将讨论几何遍历性、混合转移核，以及欧几里得情形下单步 HMC 与 MALA 的等价性。

2.1 数值积分器

哈密顿力学和拉格朗日力学是等价的，动量和速度通过勒让德变换相关联 [Marsden and Ratiu, 2010]。本节将考察这些运动方程的数值积分器。在整个讨论过程中，我们考虑如下形式的哈密顿量函数。

定义 1 设 $\pi\colon \mathbb{R}^m \to \mathbb{R}_+$ 是一个光滑概率密度，且 $\mathbf{G}\colon \mathbb{R}^m \to \mathrm{PD}(m)$。*Riemannian Hamiltonian* 是由下式定义的函数 $H\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$

$$H(q,p) = -\log \pi(q) + \frac{1}{2}\log\det(\mathbf{G}(q)) + \frac{1}{2}p^\top \mathbf{G}^{-1}(q)p. \tag{1}$$

此外，假设 $\int_{\mathbb{R}^m}\int_{\mathbb{R}^m} \exp(-H(q,p))\,\mathrm{d}q\,\mathrm{d}p < +\infty$，则 *Riemannian density* 是概率密度 $\pi(q,p) \propto \exp(-H(q,p))$。

Remark 1 函数定义中出现的项 $\frac{1}{2}\log\det(\mathbf{G}(q)) + \frac{1}{2}p^\top\mathbf{G}^{-1}(q)p$ 是这样选取的，使得条件密度 H 在式 (1) 中满足 $\pi(q,p)$

$$\pi(p|q) = \mathrm{Normal}(0, \mathbf{G}(q)). \tag{2}$$

容易看出，将 p 积分消去后，$\pi(q,\ p)$ 的 q 边缘密度为 $\pi(q)$。

哈密顿量函数 $H\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ 生成哈密顿运动方程，按定义，这些方程是以下耦合微分方程的解

$$\dot{q}_t = \nabla_p H(q_t, p_t) \tag{3}$$
$$\dot{p}_t = -\nabla_q H(q_t, p_t). \tag{4}$$

除特殊情况外，映射 $t \mapsto (q_t,\ p_t) \in \mathbb{R}^m \times \mathbb{R}^m$ 不存在闭式解；因此，需要使用数值积分器来近似满足式 (3) 和 (4) 的运动方程。广义蛙跳法就是这样一种方法，现定义如下。

定义 2 步长为 $\epsilon \in \mathbb{R}$ 的数值积分器应用于式 (1) 中的黎曼哈密顿量，得到映射 $(q,\ p) \mapsto \hat{}\ \Phi_\epsilon(q,\ p)$，其定义如下：

$$\breve{p} = p - \frac{\epsilon}{2}\nabla_q H(q, \breve{p}) \tag{5}$$
$$\tilde{q} = q + \frac{\epsilon}{2}\left(\nabla_p H(q,\breve{p}) + \nabla_p H(\tilde{q},\breve{p})\right) \tag{6}$$
$$= q + \frac{\epsilon}{2}\left(\mathbf{G}^{-1}(q) + \mathbf{G}^{-1}(\tilde{q})\right)\breve{p} \tag{7}$$
$$\tilde{p} = \breve{p} - \frac{\epsilon}{2}\nabla_q H(\tilde{q},\breve{p}) \tag{8}$$
$$= p - \frac{\epsilon}{2}\nabla_q H(q,\breve{p}) - \frac{\epsilon}{2}\nabla_q H(\tilde{q},\breve{p}), \tag{9}$$

其中 $\hat{}\ \Phi_\epsilon(q,p) = (\tilde{q},\tilde{p})$.

Remark 2 可以证明 [Leimkuhler 和 Reich，2005，Hairer 等人，2006] 广义蛙跳积分器是一种辛变换，因此 $|\det(\nabla\Phi(q,\ p))| = 1$。

Remark 3 现在我们必须对符号作一番说明。数值积分在我们的分析中起着核心作用，因此我们特别关注多步积分。同时，我们也将讨论马尔可夫链，它同样由多个步骤组成。为了区分这两种步骤，我们用下标表示数值积分的步骤，而用上标表示马尔可夫链的步骤。给定初始数据 $(q_0, p_0) \in \mathbb{R}^m \times \mathbb{R}^m$，我们将广义蛙跳积分器的 k 步表示为 $(q_k, p_k) = \hat{\Phi}_\epsilon^k (q_0, p_0) = \hat{\Phi}_\epsilon (q_{k-1}, p_{k-1})$。类似地，我们令 $\check{p}_k = p_{k-1} - \frac{\epsilon}{2}\nabla_q H (q_{k-1}, \check{p}_k)$，并称其为 *intermediate momentum at the k-th step*。

式（1）中形式的哈密顿量可变换为如下拉格朗日量

$$L(q, \dot{q}) = -\log \pi(q) + \frac{1}{2}\log\det(\mathbf{G}(q)) - \frac{1}{2}\dot{q}^\top \mathbf{G}(q)\dot{q}. \tag{10}$$

如本小节开头所述，拉格朗日力学与哈密顿力学在形式上等价，其中动量 $p \in \mathbb{R}^m$ 与速度 $\dot{q} \in \mathbb{R}^m$ 之间通过勒让德变换 $p = \mathbf{G}(q)\dot{q}$相关联。拉格朗日量会导出等价的运动方程，称为欧拉-拉格朗日方程，其形式如下：

$$\nabla_q L(q_t, \dot{q}_t) = \frac{\mathrm{d}}{\mathrm{d}t}\nabla_{\dot{q}} L(q_t, \dot{q}_t). \tag{11}$$

与哈密顿方法相比，拉格朗日形式的一个优势在于，可以找到运动方程的显式数值积分器 [Lan 等人，2015] 例如拉格朗日蛙跳法。

定义 3 将步长为 $\epsilon \in \mathbb{R}$ 的拉格朗日蛙跳法应用于式（1）中的黎曼哈密顿量，得到映射 $(q,\ p) \mapsto \tilde{\Phi}_\epsilon (q,\ p)$，定义如下：

$$v = \mathbf{G}^{-1}(q)p \tag{12}$$

$$\check{v} = v - \frac{\epsilon}{2}\mathbf{G}^{-1}(q)\nabla_q U(q) - \frac{\epsilon}{2}\Omega(q, v)\check{v} \tag{13}$$

$$\tilde{q} = q + \epsilon\check{v} \tag{14}$$

$$\tilde{v} = \check{v} - \frac{\epsilon}{2}\mathbf{G}^{-1}(\tilde{q})\nabla_q U(\tilde{q}) - \frac{\epsilon}{2}\Omega(\tilde{q}, \check{v})\tilde{v} \tag{15}$$

$$= v - \frac{\epsilon}{2}\mathbf{G}^{-1}(q)\nabla_q U(q) - \frac{\epsilon}{2}\Omega(q, v)\check{v} - \frac{\epsilon}{2}\mathbf{G}^{-1}(\tilde{q})\nabla_q U(\tilde{q}) - \frac{\epsilon}{2}\Omega(\tilde{q}, \check{v})\tilde{v} \tag{16}$$

$$\tilde{p} = \mathbf{G}^{-1}(\tilde{q})\tilde{v} \tag{17}$$

其中 $(\tilde{q},\ \tilde{p}) = \Phi_\epsilon(q,\ p)$ 和

$$U(q) = -\log \pi(q) + \frac{1}{2}\log\det(\mathbf{G}(q)) \tag{18}$$

$$\Omega_{ij}(q,v)=\sum_{k=1}^{m}\Gamma_{kj}^{i}(q)v_k \tag{19}$$

$$\Gamma_{kj}^{i}(q)=\frac{1}{2}\sum_{l=1}^{m}\mathbf{G}_{kl}^{-1}(q)\left(\frac{\partial}{\partial q_i}\mathbf{G}_{lj}(q)+\frac{\partial}{\partial q_j}\mathbf{G}_{il}(q)-\frac{\partial}{\partial q_l}\mathbf{G}_{ij}(q)\right). \tag{20}$$

˜变换的雅可比行列式 Φ_ϵ（定义见定义3）为,

$$|\det(\nabla\tilde{\Phi}_\epsilon(q,p))|=\det(\mathbf{G}^{-1}(q))\det(\mathbf{G}(\tilde{q}))\frac{|\det(\mathrm{Id}+\epsilon\Omega(q,v)/2)\det(\mathrm{Id}-\epsilon\Omega(\tilde{q},\tilde{v})/2)|}{|\det(\mathrm{Id}+\epsilon\Omega(\tilde{q},\bar{v})/2)\det(\mathrm{Id}-\epsilon\Omega(q,\bar{v})/2)|}. \tag{21}$$

Remark 4 给定初始数据（q_0,p_0）$\in\mathbb{R}^m\times\mathbb{R}^m$，我们将广义（或拉格朗日量）蛙跳积分器的 k 步表示为（q_k,p_k）$=\tilde{\Phi}_\epsilon^k$（q_0,p_0）$=\tilde{\Phi}_\epsilon$（q_{k-1},p_{k-1}）。

当$\mathbf{G}$ 是关于 q 的常函数时，广义蛙跳积分器退化为标准蛙跳积分器，其定义如下。

定义 4 设 $\pi:\mathbb{R}^m\to\mathbb{R}_+$ 为光滑概率密度。*EuclideanHamiltonian* 是一个光滑映射 $H:\mathbb{R}^m\times\mathbb{R}^m\to\mathbb{R}$，其形式为,

$$H(q,p)=-\log\pi(q)+\frac{1}{2}p^\top\mathbf{G}^{-1}p, \tag{22}$$

其中$\mathbf{G}\in\mathrm{PD}(m)$。

定义 5 考虑定义 4中定义的欧几里得哈密顿量。*Euclidean leapfrog integrator*是映射 $(q,p)\mapsto\breve{}\Phi_\epsilon(q,p)$ 定义为

$$\breve{p}=p+\frac{\epsilon}{2}\nabla\log\pi(q) \tag{23}$$

$$\tilde{q}=q+\epsilon\mathbf{G}^{-1}\breve{p} \tag{24}$$

$$\tilde{p}=\breve{p}+\frac{\epsilon}{2}\nabla\log\pi(\tilde{q}), \tag{25}$$

其中 $\breve{}\,\Phi_\epsilon(q,p)=(\tilde{q},\tilde{p})$。

欧几里得蛙跳积分器是事实上的标准数值积分器，应用于欧几里得哈密顿蒙特卡洛。这是因为它具有较高的精度和计算效率；相比之下，定义2中的广义蛙跳积分器并非全显式。然而，将广义蛙跳法和欧几里得蛙跳法应用于定义4所示形式的哈密顿量时，两者完全等价；以下结果将对此作出精确说明。

命题 1 *When the generalized leapfrog (definition 2) or Lagrangian leapfrog (definition 3) is applied to a Euclidean Hamiltonian in the form of definition 4 (in the sense that we take* $q \equiv$ *for every* $q \in \mathbb{R}^m$*) the resulting map is equivalent* $\mathbf{G}()$ $\mathbf{G}$ *, to the Euclidean leapfrog method in definition 5.*

Proof 对于广义蛙跳积分器，由于 $\nabla_q H(q,p) = -\nabla \log \pi(q)$ 不依赖于 p，且 $\nabla_p H(q,p) = \mathbf{G}p$ 不依赖于 q，因此结论显然成立。

对于拉格朗日蛙跳法，欧几里得哈密顿量的形式意味着 $\Omega(q,v)$ 恒为零（因为度量的所有导数都为零）。拉格朗日蛙跳积分器可以改写为

$$v = \mathbf{G}^{-1}p \tag{26}$$
$$\breve{v} = v + \frac{\epsilon}{2}\mathbf{G}^{-1}\nabla \log \pi(q) \tag{27}$$
$$= \mathbf{G}^{-1}\breve{p} \tag{28}$$
$$\tilde{q} = q + \epsilon\breve{v} \tag{29}$$
$$= q + \epsilon\mathbf{G}^{-1}\breve{p} \tag{30}$$
$$\tilde{v} = \breve{v} + \frac{\epsilon}{2}\mathbf{G}^{-1}\nabla \log \pi(\tilde{q}) \tag{31}$$
$$= \mathbf{G}^{-1}\tilde{p}, \tag{32}$$

其中 $\breve{p}$ 和 $\tilde{p}$ 分别如公式（23）和（25）所示。 □

2.2 马尔可夫链

本节旨在回顾马尔可夫链的基础知识，包括其构造、收敛性质和中心极限定理。马尔可夫链是一类随机过程，可以通过其转移核归纳定义。由于转移核给出的是仅依赖于当前状态的概率测度，因此显然可见，马尔可夫链满足马尔可夫性质。

定义 6 $\mathbb{R}^m$ 上的马尔可夫链转移核是一个函数 $K : \mathbb{R}^m \times \mathfrak{B}(\mathbb{R}^m) \to [0,1]$，满足：(i) 对于 $q \in \mathbb{R}^m$，$K(q,\cdot)$ 是一个概率测度；(ii) 对于固定的 $A \in \mathfrak{B}(\mathbb{R}^m)$，函数 $q \mapsto K(q,A)$ 是可测的。

在基于模拟的推断，尤其是贝叶斯推断的背景下，人们关注的是构造一个收敛到指定目标概率分布的马尔可夫链；正如引言中所述，该目标分布通常仅通过其密度（差一个归一化常数）而知。现在我们定义两种收敛概念。

定义 7 给定一个马尔可夫链转移核 K，马尔可夫链是由 $q^{n+1}|q^n \sim K(q^n, \cdot)$ 归纳定义的随机变量序列。

定义 8 两个概率测度 Π :$\mathfrak{B}(\mathbb{R}^m) \to [0,1]$ 和 Ξ: $\mathfrak{B}(\mathbb{R}^m) \to [0,1]$ 之间的 *total variation distance* 定义为

$$\|\Pi - \Xi\|_{\mathrm{TV}} = 2 \sup_{A \in \mathfrak{B}(\mathbb{R}^m)} |\Pi(A) - \Xi(A)|. \tag{33}$$

定义 9 具有转移核 K 的马尔可夫链称为遍历的，若

$$\lim_{n \to \infty} \|K^n(q,\cdot) - \Pi(\cdot)\|_{\mathrm{TV}} = 0, \tag{34}$$

其中，对于 $A \in \mathfrak{B}(\mathbb{R}^m)$， $K^n(q,\ A) = \mathrm{Pr}(q^n \in A | q^0 = q)$ 是 n 步马尔可夫转移概率。

事实证明，若能分别证明马尔可夫链具有以下三种性质：不可约性、非周期性和平稳性，就可以证明其遍历性。现在我们来定义这些概念。

定义 10 若对于每个满足 $\Pi(A) > 0$ 的集合 $A \in \mathfrak{B}(\mathbb{R}^m)$，都存在 $n \in \mathbb{N}$，使得对所有 $q \in \mathbb{R}^m$， $K^n(q,\ A) > 0$，则称马尔可夫链转移核 K 是 Π-不可约的。

定义 11 给定一个马尔可夫链转移核 K: $\mathbb{R}^m \times \mathfrak{B}(\mathbb{R}^m) \to [0,1]$，若存在 $n \in \mathbb{N}$、$\delta > 0$ 以及一个概率测度 ν: $\mathfrak{B}(\mathbb{R}^m) \to [0,1]$，使得对于任意 $q \in C$ 和 $A \in \mathfrak{B}(\mathbb{R}^m)$，均有 $K^n(q, A) \geq \delta \cdot \nu(A)$，则称集合 $C \subseteq \mathbb{R}^m$ 为 *small* 。

因此，我们可以说集合 C 是 (n, δ, ν)-小集。

定义 12 若存在一个小集 C，使集合 : 的最大公约数为

$$\{n' \in \mathbb{N}, \text{ 以及概率测度 } \nu, \text{ 使得 } C \text{ 是 } (n', \delta, \nu)\text{-小集}, \text{ 则称马尔可夫链转移核 } \nu \text{ 非周期。} \tag{35}$$

定义 13 若对于任意博雷尔集 ( )，均有 (,) ()，则称具有转移核 K 的马尔可夫链关于概率测度 Π $A \in \mathfrak{B}\ \mathbb{R}^m$ $\mathop{\mathbb{E}}_{q \sim \Pi} K\ q\ A\ =$ 平稳。 $\Pi\ A$

定义 14 若存在 $\rho \in$ （0，1）及函数 V: $\mathbb{R}^m \to \mathbb{R}_+$，则称转移核为 K 的马尔可夫链是几何遍历的。

$$\|K^n(q,\cdot) - \Pi(\cdot)\|_{\mathrm{TV}} \leq V(q)\rho^n. \tag{36}$$

从定义可以明显看出，几何遍历性是比单纯遍历性更强的一种收敛形式。不过，对于后一种情况，可以给出一些简单条件，以判定马尔可夫链是否遍历。

定理 2 [1994] *Suppose that K is a Markov chain transition kernel (that is Π-irreducible* Tierney *), aperiodic and for which Π is the stationary probability measure. , Then K produces an ergodic Markov chain.*

然而，几何遍历性除了能为马尔可夫链向目标概率测度的混合提供上界之外，还有助于建立基于序列 $(q^1, q^2, \ldots)$ 计算的期望的中心极限定理。我们将其具体表述如下。

定理 3 [1993] *Let $h \mathbb{R}^m \to \mathbb{R}$ be a function for which* (Meyn 和 Tweedie)*: ( ) $\mathbb{E}_{q\sim\Pi}|h\ q|^{2+\gamma} < \infty$ for some $\gamma > 0$. Let K be a Markov chain transition kernel that , , is aperiodic Π-irreducible and for which Π is the unique stationary distribution. ()where $q^{i+1}|q^i \sim K(q^i, \cdot)$. Then, Assume further that K converges geometrically to Π. Define $\mathcal{S}_K[h] = n^{-1}\sum_{i=1}^n h\ q^i$ as $n \to \infty$,*

$$\sqrt{n}\left(\mathcal{S}_K[h] - \mathop{\mathbb{E}}_{q\sim\Pi}[h(q)]\right) \xrightarrow{d} \text{Normal}(0, \tau_h^2) \tag{37}$$

where τ_h^2 is a constant less than infinity that depends on h (and K). The quantity , , τ_h^2 has a closed-form given by

$$\tau_h^2 = \mathop{\text{Var}}_{q\sim\Pi}(h(q)) + 2\sum_{k=1}^{\infty}\mathop{\text{Cov}}_{q\sim\Pi}\left(h(q), h(q^k)\right). \tag{38}$$

要证明马尔可夫链具有 Π 作为平稳分布，通常只需证明该链相对于 Π满足细致平衡即可。在第3节讨论黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛的边缘转移核时，我们需要用到细致平衡，因此现在定义这一概念。

定义 15 如果马尔可夫链转移核相对于概率分布 Π满足 *detailed balance*（等价地，称该马尔可夫链相对于 Π满足 *reversible*），则对于任意集合 $Q, Q' \in \mathfrak{B}(\mathbb{R}^m)$,

$$\int_Q K(q, Q')\ \Pi(\mathrm{d}q) = \int_{Q'} K(q, Q)\ \Pi(\mathrm{d}q). \tag{39}$$

Remark 5 当马尔可夫链转移核 K满足细致平衡时，可知 Π 是转移核13（定义 K）的平稳概率测度。

2.2.1 Metropolis-Hastings 核与广义朗之万算法

本节回顾 Metropolis-Hastings 马尔可夫链。我们首先对这些对象进行形式化定义，然后介绍一种基于朗之万扩散的 Metropolis-Hastings 方法特例；该方法回顾了 Roy 和 Zhang [2023].

10 修正黎曼流形与拉格朗日蒙特卡洛

定义 16（Metropolis-Hastings算法）设 $\pi: \mathbb{R}^m \to \mathbb{R}_+$ 为概率密度，并且对每个 $q \in \mathbb{R}^m$，令 $\tilde{\pi}(\cdot|q)$: $\mathbb{R}^m \to \mathbb{R}_+$ 为依赖于 q的概率密度。提议密度为 $\tilde{\pi}(\cdot|q)$ 的Metropolis-Hastings转移核为:

$$\begin{aligned} K(q, A) = \int_A \min\left\{1, \frac{\pi(\tilde{q})\tilde{\pi}(q|\tilde{q})}{\pi(q)\tilde{\pi}(\tilde{q}|q)}\right\} \tilde{\pi}(\tilde{q}|q)\ \mathrm{d}\tilde{q} \\ + \left(1 - \int_{\mathbb{R}^m} \min\left\{1, \frac{\pi(\tilde{q})\tilde{\pi}(q|\tilde{q})}{\pi(q)\tilde{\pi}(\tilde{q}|q)}\right\} \tilde{\pi}(\tilde{q}|q)\ \mathrm{d}\tilde{q}\right) 1\{q \in A\}, \end{aligned} \tag{40}$$

对于 $q \in \mathbb{R}^m$ 和 $A \in \mathfrak{B}(\mathbb{R}^m)$。

定义 17 $\epsilon \in \mathbb{R} \setminus \{0\}$ c_ϵ $\mathbb{R}^m \to \mathbb{R}^m$ $\mathbb{R}^m \to$ m 义 Metropolis 调整朗之万算法）固定 . 令 : 且令 $\mathbf{A}$: PD( )。广义 Metropolis 调整朗之万算法是 Metropolis-Hastings算法的一种，其提议密度为

$$\tilde{\pi}(\tilde{q}|q) = \mathrm{Normal}(\tilde{q};\ c_\epsilon(q), \epsilon^2\mathbf{A}(q)). \tag{41}$$

我们将步长为 ϵ 的广义 Metropolis 调整朗之万算法的马尔可夫链转移核记为 J_ϵ : $\mathbb{R}^m \times \mathfrak{B}(\mathbb{R}^m) \to [0, 1]$。

定义 18（黎曼流形Metropolis调整朗之万算法）在以下特殊情形下

$$c_\epsilon(q) = \frac{\epsilon^2}{2}\mathbf{A}(q)\nabla \log \pi(q) + \frac{\epsilon^2}{2}\Gamma(q) \tag{42}$$

$$\Gamma_i(q) = \sum_{j=1}^{m} \frac{\partial}{\partial q_j}\mathbf{A}_{ij}(q) \tag{43}$$

我们将由此得到的方法称为黎曼流形Metropolis调整朗之万算法（MMALA）。

Remark 6 方程（42）中的提议形式基于在黎曼度量$\mathbf{G}(q) = \mathbf{A}^{-1}(q)$ Xifara等人 [2014] 的流形上朗之万扩散的Euler-Maruyama离散化；详情请参见相关文献。总而言之，方程（42）是将步长为 ϵ^2的Euler-Maruyama离散化应用于以下随机微分方程后得到的漂移分量:

$$\mathrm{d}X_t = \frac{1}{2}\mathbf{G}^{-1}(X_t)\nabla \log \pi(X_t)\ \mathrm{d}t + \Gamma(X_t)\ \mathrm{d}t + \mathbf{G}^{-1/2}(X_t)\ \mathrm{d}B_t, \tag{44}$$

其中， B_t 是在时间 $t \in \mathbb{R}$的欧几里得布朗运动。当G(q)为费希尔信息矩阵时，项G^{-1}(X_t)∇log π(X_t)可识别为函数log π 在由Fisher度量[Amari和Nagaoka， 2000] 于位置 X_t所生成几何结构下的 *naturalgradient*。另一方面，项 $\Gamma(X_t)\,\mathrm{d}t$ +G$^{-1/2}$(X_t) dB_t 对应于 *manifold Brownian motion*，因为其无穷小生成元是流形 [上的Laplace-Beltrami算子Hsu, 2002]。因此，由于方程（44）中的随机微分方程由一个包含对数密度的流形梯度的项和另一个包含流形布朗运动的项组成，所以它被称为 *Riemannian Langevin equation*，对应于欧几里得情形。

定义 19 （简化黎曼流形梅特罗波利斯调整朗之万算法）计算函数 Γ 在定义18 中可能不便于求值。忽略该函数仍可得到一个马尔可夫链转移核，从而得到以下特殊情形：

$$c_\epsilon(q) = \frac{\epsilon^2}{2}\mathbf{A}(q)\nabla \log \pi(q). \tag{45}$$

由此得到的方法称为简化黎曼流形 Metropolis 调整朗之万算法，该方法由Girolami 和 Calderhead [2011]提出。

我们用缩写 SMALA 指代简化的 Metropolis 调整朗之万算法。

Remark 7 该函数$\mathbf{A}$有三种常见的选取方式。第一种是令$\mathbf{A}$为常函数，此时我们简单地写作$\mathbf{A} \in \mathrm{PD}(m)$。第二种方式是令$\mathbf{A}$ 取 *inverse* 费希尔信息矩阵与对数先验的负海森矩阵之和；这可以捕捉似然和先验的二阶几何。将费希尔信息与对数先验的负海森矩阵之和的逆作为预条件器，是Girolami 和 Calderhead [2011]所倡导的方法。第三种方式是令$\mathbf{A}$ 为 SoftAbs 度量的逆；SoftAbs 度量是目标分布对数密度的海森矩阵的一种平滑变换。详情请参阅Betancourt [2012]。

2.2.2 贝叶斯推断的几何方法

定义20 (对合蒙特卡罗 [Neklyudov 等人， 2020]) 设 $\Phi : \mathbb{R}^m \to \mathbb{R}^m$ 是一个光滑对合，且设 $\pi : \mathbb{R}^m \to \mathbb{R}_+$ 是定义在 $\mathbb{R}^m$上的概率密度。目标密度为 π 的对合蒙特卡罗马尔可夫链转移核为

$$\begin{aligned} K(q, A) = &\min\left\{1, \frac{\pi(\Phi(q))}{\pi(q)}\left|\det(\nabla\Phi(q))\right|\right\}\mathbf{1}\left\{\Phi(q) \in A\right\} \\ &+ \left(1 - \min\left\{1, \frac{\pi(\Phi(q))}{\pi(q)}\left|\det(\nabla\Phi(q))\right|\right\}\right)\mathbf{1}\left\{q \in A\right\}, \end{aligned} \tag{46}$$

对于 $q \in \mathbb{R}^m$ 和 $A \in \mathfrak{B}\left(\mathbb{R}^m\right)$ 。

容易验证，对合蒙特卡罗的转移核相对于密度 π 满足细致平衡。对合蒙特卡罗可导出两个特殊的转移核，分别对应于黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛。

定义21 设 $\hat{} \Phi_\epsilon$ 如定义2所述，并设 H 如定义1所述。步长为 ϵ 且采用 k个积分步的黎曼流形哈密顿蒙特卡罗对合为：

$$\Phi = \mathbf{F} \circ \underbrace{\hat{\Phi}_\epsilon \circ \cdots \circ \hat{\Phi}_\epsilon}_{k \text{ times}}, \tag{47}$$

其中 $\mathbf{F}(q, p) = (q, -p)$。黎曼流形哈密顿蒙特卡罗的目标密度为 $\pi(q, p) \propto \exp(-H(q, p))$。对合蒙特的转移核

12 修正黎曼流形与拉格朗日蒙特卡洛

卡洛（定义20）与式（ Φ 中给出的对合47）称为步长为 ϵ 、具有 k 个积分步数的黎曼流形哈密顿蒙特卡罗转移核。

Remark 8 可以证明，若 Φ 是可逆函数，且$\mathbf{F}\circ\Phi$ 是对合，那么$\mathbf{F}\circ\Phi\circ\cdots\circ\Phi$ 也是对合。这解释了为什么式（47）也是对合。更多细节请参见命题15。

在特殊情况下，若$\mathbf{G}$ 是关于 q的常函数，则所得的马尔可夫链称为欧几里得哈密顿蒙特卡罗。

定义22 设 $\check{}\ \Phi_\epsilon$ 如定义5 所述，并设 H 如定义4所述。步长为 ϵ 、具有 k个积分步数的欧几里得哈密顿蒙特卡罗（EHMC）对合为:

$$\Phi = \mathbf{F}\circ\underbrace{\check{\Phi}_\epsilon\circ\cdots\circ\check{\Phi}_\epsilon}_{k\text{ times}}, \tag{48}$$

其中 $\mathbf{F}(q,p) = (q,-p)$ 。欧几里得哈密顿蒙特卡罗的目标密度为 $\pi(q,p) \propto \exp(-H(q,p))$。带有对合 Φ （见式（48））的对合蒙特卡罗（定义 20）的转移核称为步长为 ϵ 且具有 k 个积分步数的欧几里得哈密顿蒙特卡罗转移核。

定义23 设 $\tilde{\Phi}_\epsilon$ 如定义 3所述，并设 H 如定义 1所述。步长为 ϵ 且具有 k 个积分步数的拉格朗日蒙特卡罗的对合为:

$$\Phi = \mathbf{F}\circ\underbrace{\tilde{\Phi}_\epsilon\circ\cdots\circ\tilde{\Phi}_\epsilon}_{k\text{ times}}, \tag{49}$$

其中 $\mathbf{F}(q,p) = (q,-p)$。拉格朗日哈密顿蒙特卡罗的目标密度为 $\pi(q,p) \propto \exp(-H(q,p))$。我们将其称为步长为 ϵ 且具有 k 个积分步数的拉格朗日蒙特卡罗转移核。

Remark 9 与黎曼流形哈密顿蒙特卡罗所采用的转移核不同，拉格朗日蒙特卡罗中出现的对合的雅可比行列式必须通过 k 应用式（21）来计算，其中，与之前一样， k 表示积分步数。

在$\mathbf{G}$ 是 q的常函数这一特殊情况下，黎曼流形哈密顿蒙特卡罗和拉格朗日蒙特卡罗会产生一个转移核，该转移核等价于采用常质量矩阵$\mathbf{G}$的欧几里得哈密顿蒙特卡罗。

命题 4 *The transition kernels of RMHMC (or LMC) with step-size ϵ and k integration steps when applied to the Euclidean Hamiltonian in definition 4 (in the sense that we take $q \equiv$ for every $q \in \mathbb{R}^m$) are both equivalent to the EHMC* $\mathbf{G}()$ $\mathbf{G}$ *transition kernel.*

Proof 由命题 1 可知，黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛所生成的提议与欧几里得哈密顿蒙特卡洛完全相同。还需验证其接受决策也相同。注意到，式（1）中的黎曼哈密顿量与式（22）中的欧几里得哈密顿量只相差一个加法常数（因为 $\mathbf{G}$ 与位置无关），即可验证这一点。因此，欧几里得密度与黎曼密度成正比，接受概率也将完全相同。□

现在我们进一步阐述欧几里得哈密顿蒙特卡洛采用单步时 与梅特罗波利斯调整朗之万算法之间的联系：具体而言，这两种方法可以构造为在 q空间中完全等价，产生相同的提议和完全相同的接受概率。

命题 5 [2010] *Let* $\in \mathbb{R}^{m\times m}$ *be a fixed positive definite matrix. Consider EHMC with the Hamiltonian* $H\ q\ p\ =\ -\ \pi\ q\ +\ \frac{1}{2}p_{-}^{\top}\ ^{-1}p$ *and* (Neal a) $\mathbf{A}\ (,)\log()\ \mathbf{A}$*MALA with proposal distribution* $\tilde{q}|q \sim$正态 $\left(q+\frac{\epsilon^2}{2}\mathbf{A}^{-1}\nabla\log\ \pi(q)\ \epsilon^2\mathbf{A}^{-1}\right)$*. The marginal transition kernel of EHMC with step-size* ϵ *and a single integration step can be constructed to be exactly equivalent to the transition kernel of MALA.*

证明见第A.1节。混合转移核的分析将是我们分析的核心。在标准的黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛实现中，通常会对积分步数进行随机化。这样做是为了避免使用固定积分步数可能导致的任何马尔可夫链病态（例如不可约性失效）。因此，我们考虑由混合构成的马尔可夫链转移核。

定义24 令 $K_{\epsilon,k}:(\mathbb{R}^m\times\mathbb{R}^m)\times\mathfrak{B}(\mathbb{R}^m\times\mathbb{R}^m)\ \to\mathbb{R}_+$ 表示步长为 ϵ 且积分步数为 k 的黎曼流形哈密顿蒙特卡洛（或拉格朗日蒙特卡洛）的马尔可夫链转移核。令 $\tilde{K}_{\epsilon,k}$ 表示 $K_{\epsilon,k}$的边缘转移核（定义见引理6）。定义*mixture transition kernel of RMHMC (or LMC)*为

$$\tilde{K}_\epsilon=\sum_{k=1}^{\infty}\alpha_k\tilde{K}_{\epsilon,k},\tag{50}$$

其中（$\alpha_1,\alpha_2,$ .....）是一个概率向量。

EHMC、RMHMC 和 LMC 在（q,p）空间中均满足细致平衡，因此我们考察它们在对动量进行 Gibbs 重采样步骤之间，单独的 q 状态演变。这会得到如下形式的边缘马尔可夫链转移核：

引理 6 *Let* $K\ \ \mathbb{R}^m\times\mathbb{R}^m\ \times\mathfrak{B}\ \mathbb{R}^m\times\mathbb{R}^m\ \to\mathbb{R}$ *be a Markov chain tran-*: ( ) ( ) (,) ( ) ( ) *sition kernel. Suppose that* K *satisfies detailed balance with respect to the density* $\pi\ q\ p$ *with q-marginal distribution* $\pi\ q$ *and conditional density* $\pi\ p|q$ *; i.e.* $\pi\ q\ p\ =$ (,)

π q π sample $p^n \sim \pi(p|q)$ sample q^{n+1} $p^{n+1} \sim K(q^n, p^n, \cdot)$ and discard both momenta. () $(p|q)$

. *Consider the marginal chain constructed as follows. Given* $q^n = q$, ( ) (,) ((,)) *The transition kernel of the marginal chain satisfies*

$$\tilde{K}(q, Q) = \int_{\mathbb{R}^m} K((q,p),(Q,\mathbb{R}^m))\ \pi(p|q)\ \mathrm{d}p \tag{51}$$

where $Q \in \mathfrak{B}(\mathbb{R}^m)$。

证明见第 A.2

命题 7 *Let* $\tilde{K}$ *be the marginal transition kernel described in lemma 6. The marginal chain satisfies detailed balance with respect to the distribution whose density is* π (q) .

证明见第 A.3。

2.2.3 遍历性定理

MALA、MMALA 和 SMALA 的几何遍历性已在马尔可夫链文献中得到研究。在第 3 节中，我们将看到，混合马尔可夫链转移核的单个分量具有几何遍历性，如何意味着混合核本身也具有几何遍历性。带着这一目标，我们现在回顾两个相关的重要结果。

定理8 (Roberts 和 Tweedie [1996]) *Consider the transition kernel* K *of the Metropolis-adjusted Langevin algorithm with* $\mathbf{A} = \mathrm{Id}$ *and* $c_\epsilon(q) = x + \frac{\epsilon^2}{2}\nabla \log \pi(q)$。 *Then* K *is geometrically ergodic under the following conditions:*

1. 我们有

$$\liminf_{\|q\|\to\infty} \left(\|q\| - \left\|q + \frac{\epsilon^2}{2}\nabla \log \pi(q)\right\| \right) > 0. \tag{52}$$

2. 我们有

$$\lim_{\|q\|\to\infty} \int_{(A(q)\cup I(q))\cap(A(q)\cap I(q))} \tilde{\pi}(\tilde{q}|q)\ \mathrm{d}\tilde{q} = 0, \tag{53}$$

其中

$$A(q) = \{\tilde{q} \in \mathcal{X} : \pi(q)\tilde{\pi}(\tilde{q}|q) \le \pi(y)\tilde{\pi}(q|\tilde{q})\} \tag{54}$$
$$I(q) = \{\tilde{q} \in \mathcal{X} : \|\tilde{q}\| \le \|q\|\}\,. \tag{55}$$

定理 9 [2023] *Consider the transition kernel* K *of the generalized Metropolis-adjusted Langevin algorithm. Then* K *is geometrically ergodic under* (Roy 和 Zhang) *the following conditions:*

1. 存在矩阵 $\mathbf{A}_l \in \mathrm{PD}(m)$ 和 $\mathbf{A}_u \in \mathrm{PD}(m)$，使得 $\mathbf{A}_l \leq \mathbf{A}(q) \leq \mathbf{A}_u$。*2.* 当 $A \subset \mathbb{R}^m$ 有界时，函数 c_ϵ： $\mathbb{R}^m \to \mathbb{R}^m$ 在 A上有界。*3.* 存在一个量

$$C = \limsup_{\|q\| \to \infty} \int_{R(q)} \left(1 - \tilde{\pi}(\tilde{q}|q) \min\left\{1, \frac{\pi(\tilde{q})\tilde{\pi}(q|\tilde{q})}{\pi(q)\tilde{\pi}(\tilde{q}|q)}\right\}\right) \mathrm{d}y \tag{56}$$

严格小于一，其中

$$R(q) = \{\tilde{q} \in \mathbb{R}^m : \pi(q)\tilde{\pi}(\tilde{q}|x) > \pi(\tilde{q})\tilde{\pi}(q|\tilde{q})\}. \tag{57}$$

4. 存在 $s > 0$，使得，

$$\liminf_{\|q\| \to \infty} \left(\|\mathbf{A}_u^{-1/2}(q)x\| - \|\mathbf{A}_u^{-1/2}(q)c_\epsilon(q)\|\right) > \frac{\log(D(s)) - \log(1-C)}{s}, \tag{58}$$

其中，

$$D(s) = \epsilon^{-m/2} \left(\frac{\pi}{2}\right)^{m/2-1} \left(\frac{\det(\mathbf{A}_u)}{\det(\mathbf{A}_l)}\right)^{1/2} \exp(\epsilon s^2/2) \int_0^\infty \exp\left(-(r-\epsilon s)^2/2\epsilon\right) r^{m-1} \mathrm{d}r \tag{59}$$

在第 3 节中，我们将考察混合转移核及其几何遍历性。若一个转移核是几何遍历的，且所有转移核均可逆，则根据以下结果，混合转移核也是几何遍历的。

定理 10 *Let α_1 α_2 be a sequence satisfying* (Lee 和 [2014]) *$\alpha_i \geq 0$ and $\sum_{i=1}^\infty \alpha_i = 1$. Let $\tilde{K} = \sum_{i=1}^\infty \alpha_i K_i$ be a mixture of reversible transition kernels with invariant distribution* Π. *Suppose that K_1 satisfies the properties: (i)* Π *is the unique invariant distribution* (Latuszynski) (,,..) , *and (ii) K_1 is geometrically ergodic. Then $\tilde{K}$ is geometrically ergodic.*

˜我们将 K 从 K_1 推导出的几何遍历性称为“继承的”几何遍历性。

3 继承的几何遍历性

本节讨论“继承的”几何遍历性的概念：当 MMALA 具有几何遍历性时，我们对黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛进行修改，使其也具有几何遍历性。在欧几里得情形下，根据命题 5，在 Metropolis 调整的 Langevin 算法具有几何遍历性的情形下，哈密顿蒙特卡洛也具有几何遍历性

通过调用定理 10。几何情形更为复杂，因为没有类似于命题 5 的结果，可以证明单步黎曼流形哈密顿蒙特卡洛或拉格朗日蒙特卡洛的边缘转移核，与在适当选择均值函数时广义 Metropolis 调整朗之万算法的马尔可夫链转移核完全等价。因此，我们提出如下马尔可夫链转移核。

定义25 设 $K_{\epsilon,\ k}:(\mathbb{R}^m\times\mathbb{R}^m)\times\mathfrak{B}(\mathbb{R}^m\times\mathbb{R}^m)\to\mathbb{R}_+$ 为步长为 ϵ 且积分步数为 k 的黎曼流形哈密顿蒙特卡洛（或拉格朗日蒙特卡洛）马尔可夫链转移核。设 $\tilde{K}_{\epsilon,\ k}$ 为边缘转移核（定义见引理 6）。设 (>为满足 $\alpha_i\geq 0$ 且 $\sum_{i=1}^{\infty}\alpha_i=1$的序列。设 J_ϵ 为广义 Metropolis 调整朗之万算法的马尔可夫链转移核（定义见定义 17）。 *Langevin mixture transition kernel of RMHMC (or LMC)*记作 LMRMHMC（或 LMLMC），其定义如下：

$$\tilde{K}_\epsilon=\alpha_1 J_\epsilon+\sum_{k=2}^{\infty}\alpha_k\tilde{K}_{\epsilon,k}. \tag{60}$$

我们将 α_1 称为 MMALA混合权重。

修正转移核只是用广义 Metropolis 调整朗之万算法的转移核，替换 RMHMC（或 LMC）的一个单步转移。为了应用定理10，有必要验证 *RMHMC* 和 *LMC* 的边缘转移核是可逆的。幸运的是，这一点很容易证明，如下所示。

推论 11 *For fixed $\epsilon\in\mathbb{R}$ and $k\in\mathbb{N}$, the RMHMC and LMC transition kernels with step size ϵ and k integration steps satisfy detailed balance with respect to the distribution π. It follows that their marginal chains are reversibly with respect to the distribution π by proposition 7.*

命题 12 *Suppose $\alpha_1>0$. Under the conditions of theorem 9, the Markov chain transition kernel of LMRMHMC (or LMLMC) in eq. 60 () is geometrically ergodic.*

Proof 这可由定理 10 直接得出，因为根据推论 11，RMHMC（或 LMC）的边缘转移核是可逆的。

□

Remark 10 在实际应用中，我们按以下方式选择混合概率（$\alpha_1,\ \alpha_2,\ \ldots\ldots$）。对于给定的目标分布，我们会选择一个最大积分步数 $k_{\max}$，使得 $\alpha_k=0$ 对于 $k>k_{\max}$。然后，给定对 $\alpha_1\in$ 的特定选择 $(0,1],$，我们将剩余的概率质量平均分配给每个 $k\in\{2,\ldots,k_{\max}\}$；也就是说，该分数可表示为 $\alpha_k=\frac{1-\alpha_1}{k_{\max}-1}$，其中 $k=2,\ \ldots\ldots$，$k_{\max}$。在我们的实验中，我们将考虑以下可变选择：

混合参数 α_1。特殊情形 $\alpha_1 = 0$按照约定，对应于未经修正的 RMHMC 和 LMC 转移核，其单步转移使用规定的对合计算，详见定义 24。

与 MMALA 转移核进行混合，也可立即确立不可约性、非周期性以及所有紧集都是小集，正如下一个结果所示。

引理 13 *Let $\pi(q) \propto \mathcal{L}(q)$ be continuous and bounded on compact sets and denote by Π the probability measure with density π. Suppose $\alpha_1 > 0$. The marginal transition kernel of LMRMHMC (or LMLMC) is Π-irreducible aperiodic and all non-negligible compact sets are small.*

Proof 以概率 α_1，马尔可夫链根据采用正态提议分布的 Metropolis-Hastings 接受—拒绝决策进行转移。因此，对于任意集合 $A \subset \mathfrak{B}(\mathbb{R}^m)$，若满足 $\int_A \pi(q)\,\mathrm{d}q > 0$，则有，

$$\tilde{K}_\epsilon(q, A) \geq \alpha_1 J_\epsilon(q, A) \tag{61}$$

$$> 0 \tag{62}$$

因为正态提议分布在 $\mathbb{R}^m$上处处非零。事实˜ K_ϵ 是非周期的，且所有非可忽略的紧集都是小集，这直接由Mengersen 和 Tweedie [1996]的引理 1.2 得出。□

推论 14 *From lemma 13, we have that the modified Markov chain of RMHMC (or LMC) is Π-irreducible aperiodic and from proposition 7 Π is the stationary distribution. Hence it follows from theorem 2 that the modified Markov chain of RMHMC (or LMC) produces an ergodic Markov chain.*

4 实验

现在，我们开始研究所提出的 RMHMC 和 LMC 的改进变体。这些示例旨在代表范围广泛的后验分布。为了使用最大均值差异度量计算马尔可夫链的收敛情况，我们比较目标分布的 10,000 个独立同分布样本与数量相同的独立马尔可夫链样本；在此情形下，我们计算前一百个采样步内最大均值差异的无偏估计量。对于期望平方跳跃距离（ESJD）和有效样本量（ESS）这两项度量，我们对由 1,000,000 个样本组成的长马尔可夫链进行十次重复实验，但对于 FitzHugh-Nagumo 后验分布，我们只采样 100,000 次。在每个实验中，我们考察采用定义 18所述转移核的混合 LMC 和 RMHMC；但对于 Neal 漏斗分布，我们考察与定义 19相结合的混合方法。我们首先在第 4.1 节介绍用于评估相应马尔可夫链性能的度量指标

这些度量可用于评估 LMRHMC 和 LMLMC 对应的马尔可夫链的性能。

用于复现这些实验结果的代码可见于 https://tinyurl.com/29kz7krz。

4.1 度量与指标

现在，我们详细介绍用于比较马尔可夫链的评估指标。首先回顾 Gretton 等人 [2012]提出的最大均值差异方法。

定义 26 设 $k:\ \mathbb{R}^m\times\mathbb{R}^m\to\mathbb{R}$ 是关于其参数对称的正定函数。设 Π 和 Π' 是定义在 $\mathbb{R}^m$上的两个概率分布。 Π与 Π' 之间的平方最大均值差异定义为:

$$\mathrm{MMD}^2(k,\Pi,\Pi')=\mathop{\mathbb{E}}_{q,q'\sim\Pi}k(q,q')+\mathop{\mathbb{E}}_{q,q'\sim\Pi'}k(q,q')-2\mathop{\mathbb{E}}_{q\sim\Pi,q'\sim\Pi'}k(q,q'). \tag{63}$$

设 $(q_*^1,\ldots,q_*^r)\ \sim\Pi$和 $(q^1,\ldots,q^s)\ \sim\Pi'$。平方最大均值差异的一个无偏估计量为:

$$\mathrm{MMD}_{\mathrm{u}}^2(k,\left\{q_*^i\right\}_{i=1}^r,\left\{q^i\right\}_{i=1}^s)=\frac{1}{r(r-1)}\sum_{i=1}^r\sum_{j\neq i}^r k(q_*^i,q_*^j)+\frac{1}{s(s-1)}\sum_{i=1}^s\sum_{j\neq i}^s k(q^i,q^j)-\frac{2}{rs}\sum_{i=1}^r\sum_{j=1}^s k(q_*^i,q^j). \tag{64}$$

在评估中，我们采用平方指数正定核$k(q,q')\ =\exp(-\|q-q'\|^2/2h)$，其中 $h\in\mathbb{R}_+$ 是一个称为核带宽的参数。在实验中，我们将 h 设为目标分布的独立同分布样本之间距离的中位数。回顾定义 14 中给出的几何遍历性定义。收敛的度量是全变差范数，我们可以从理论上讨论它，但无法在计算环境中对其进行评估。相反，我们可以通过式（63）测量目标分布Π 的收敛情况，将其表示为步数 n的函数；如果我们有来自 $K^n(q^0,\cdot)$的样本以及来自 Π的样本，就可以利用式（64）得到无偏估计。事实上，如定义 26 所示，令$(q_*^1,\ldots,q_*^r)\ \sim\Pi$，对于固定的初始位置 q^0和步数 n，令$(q^1,\ldots,q^s)\ \sim K^n(q^0,\cdot)$，由此我们可以计算平方最大均值差异的无偏估计。在后一种情况下，$(q^1,\ldots,q^s)$可以通过运行 s 条相互独立的马尔可夫链 n 步获得；在某些基准案例中，可能可以获得来自目标分布Π 的独立样本。在我们的实验中，目标分布的独立同分布样本可以从香蕉形后验分布、Neal 漏斗分布、菲茨休-长门后验分布以及多尺度 Student-t分布中生成。

Remark 11 回顾定义 14中给出的几何遍历性定义。取对数可得

$$\log \|K^n(q,\cdot) - \Pi(\cdot)\|_{\mathrm{TV}} \leq n \log \rho + \log V(q). \tag{65}$$

因此，可以通过以下现象来判断 *evidence of geometric ergodicity*：若总变差距离随 n 变化而在对数尺度上呈线性下降。当然，我们无法直接计算总变差距离，但可以观察在对数尺度上绘制 $|$最大均值差异$_u^2|$ 时是否呈现类似的负线性趋势。

作为衡量遍历性的另一项指标，我们考虑将马尔可夫链样本与独立同分布样本投影到一维子空间上进行比较。设 $(q_*^1, \ldots\ldots, q_*^r) \sim \Pi$，并设 $q^n \sim K^n(q^0, \cdot)$，其中 $n = 1, \ldots\ldots, s$。设 u 为一个随机单位向量。我们计算投影 $(u^\top q_*^1, \ldots\ldots, u^\top q_*^r)$ 和 $(u^\top q^1, \ldots\ldots, u^\top q^s)$ 的柯尔莫哥洛夫-斯米尔诺夫统计量（KS 统计量）。对一百个随机生成的单位向量重复此过程，可得到柯尔莫哥洛夫–斯米尔诺夫统计量的分布。该分布越集中在零附近，马尔可夫链迭代值 $(q^1, \ldots\ldots, q^s)$ 的分布就越接近来自目标分布的独立同分布样本 $(q_*^1, \ldots\ldots, q_*^r)$。

我们还考虑期望平方跳跃距离（ESJD）[Gelman 和 Pasarica, 2007]，它用于衡量马尔可夫链相邻状态之间的差异。直观而言，在样本空间中移动得更有效率（ESJD 更高）的马尔可夫链，其样本自相关会更小。形式上，ESJD 定义为 $\mathbb{E}_{q\sim\Pi}\|q' - q\|^2$ 其中 $q'|q \sim K(q, \cdot)$。注意范数 $\|\cdot\|$ 的选择由实践者决定，也可以根据目标分布的几何性质进行选取。通常按如下方式近似该期望：令 q^k 表示马尔可夫链在第 k 步的状态，并在第 $k+1$ 步生成一个候选状态，记为 q_{prop}^{k+1}；以概率 α_k 接受该提议状态；随后计算以下经验均值，作为 ESJD 的近似值

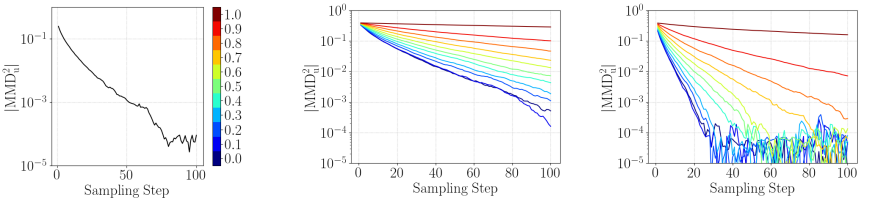
$$\mathrm{ESJD}(\{(q^k, q_{\mathrm{prop}}^{k+1}, \alpha_k)\}_{k=1}^n) = \frac{1}{n}\sum_{k=1}^n \alpha_k \|q_{\mathrm{prop}}^{k+1} - q^k\|^2. \tag{66}$$

在第4.8节的实验中，我们发现，期望平方跳跃距离作为一种度量可能会产生误导。因此，我们还引入了中位数平方跳跃距离（MSJD），其定义如下：

$$\mathrm{MSJD}(\{(q^k, q_{\mathrm{prop}}^{k+1}, \alpha_k)\}_{k=1}^n) = \mathrm{Median}\left(\{\alpha_k \|q_{\mathrm{prop}}^{k+1} - q^k\|^2\}_{k=1}^n\right). \tag{67}$$

作为中位数，我们预计 MSJD 对离群值的敏感性低于期望平方跳跃距离。

此外，我们还将有效样本量（ESS）作为衡量马尔可夫链过程的指标。我们采用 Kumar等人 [2019], 的方法计算有效样本量，他们对该方法作了如下简要描述。对于 $i = 1, \ldots, p$,



(a) 欧几里得哈密顿蒙特卡洛

(b) 拉格朗日修正黎曼流形哈密顿蒙特卡洛

(c) LMLMC

图 1：我们考察香蕉形后验分布中马尔可夫链的遍历性随采样步长的变化。我们发现，采用较小的步长会降低黎曼流形哈密顿蒙特卡洛的遍历性，尤其是与 LMC 算法相比时。在黎曼流形哈密顿蒙特卡洛和 LMC 中，我们都观察到，采用 MMALA 进行激进混合会使马尔可夫链混合得更慢；另一方面，相对较小的混合概率会使其遍历性与未经修改的黎曼流形哈密顿蒙特卡洛和 LMC 实现几乎没有差别。

令 $(q^{i,1}, q^{i,2}, \ldots, q^{i,n})$ 为一列取值于 $\mathbb{R}^m$ 的参数。整数 p 称为链数。第 j 个参数的有效样本量按以下方式计算，

$$\mathrm{ESS}\left(\left\{q_j^{i,1}, \ldots, q_j^{i,n}\right\}_{i=1}^{p}\right) = \frac{pn}{\hat{\tau}} \tag{68}$$

$$\hat{\tau} = \left(2\sum_{k=1}^{r} \hat{\rho}_{2k} + \hat{\rho}_{2k+1}\right) - 1, \tag{69}$$

其中 $\hat{\rho}_k$ 是根据所有 p 个序列，对 j个参数在滞后 k步时的自相关所作的估计（详情请参阅Vehtari 等人 [2021]），而 r 是使 $\hat{\rho}_{2(r+1)} + \hat{\rho}_{2(r+1)+1} \leq 0$成立的最小整数。在我们的实验中，我们将一条较长的马尔可夫链从中间分成两半，以此设定 $p = 2$ 。作为实际考量，我们还将报告后验分布所有参数中的最小有效样本量（每秒），以表示该方法的计算效率。

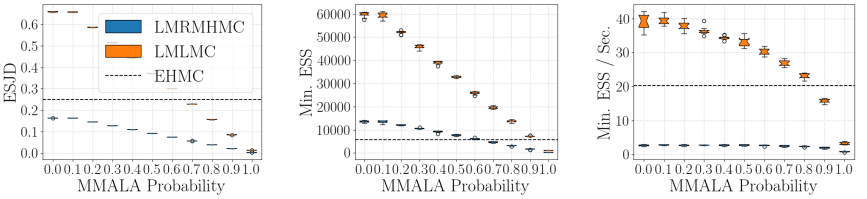
4.2 香蕉形分布

第一个示例考察以下生成模型，它展示了一个不可识别参数的例子。

$$\theta_1, \theta_2 \overset{\text{i.i.d.}}{\sim} \mathrm{Normal}(0, \sigma_\theta^2) \tag{70}$$

$$y_i | \theta_1, \theta_2 \overset{\text{i.i.d.}}{\sim} \mathrm{Normal}(\theta_1 + \theta_2^2, \sigma_y^2) \quad \text{for } i = 1, \ldots, N. \tag{71}$$

给定观测值 $\{y_i\}_{i=1}^N$，我们希望对 (θ_1, θ_2) 的后验分布进行采样，这是一个二维后验分布。在我们的实验中，我们考虑 $n =$



(a) 期望平方跳跃距离

(b) 最小有效样本量

(c) 最小有效样本量/秒

图 2：我们展示了香蕉形后验分布推断的期望平方跳跃距离、最小有效样本量，以及按时间归一化的最小有效样本量。我们观察到，对于黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛，采用 MMALA 进行激进混合会导致期望平方跳跃距离下降；相应地，最小有效样本量也会降低。对于拉格朗日蒙特卡洛，按时间归一化的最小有效样本量是混合概率的递减函数。

100。我们采用步长为 $\epsilon = 0.1$ 的欧几里得哈密顿蒙特卡洛；黎曼流形哈密顿蒙特卡洛的步长设为 0.04；在拉格朗日蒙特卡洛中，我们使用积分步长 $\epsilon = 0.1$。在每种情况下，我们都设定 $k_{\max} = 10$。对于几何方法，我们将费希尔信息与对数先验的负海森矩阵之和用作黎曼度量。在黎曼流形哈密顿蒙特卡洛中采用较小积分步长的原因是，对于较大的步长，定义对合中的式（5）将无法求得 $\breve{p}$ 的解，因此必须采用较小的积分步。

在图 1中，我们展示了统计量 $|\text{最大均值差异}_u^2|$ 在欧几里得 HMC（EHMC）、RMHMC 和 LMC 的马尔可夫链运行一百多步期间的变化；我们根据 RMHMC 和 LMC 马尔可夫链与 MMALA 的混合程度，为其着色。EHMC 和 RMHMC 均未在该目标分布中表现出明显的几何遍历性证据；另一方面，在较大范围的混合概率下，LMC 的 n上的 $|\text{最大均值差异}_u^2|$ 在对数尺度上随步数线性下降。图 2 展示了三种马尔可夫链蒙特卡洛算法的期望平方跳跃距离、最小有效样本量和每秒最小有效样本量。我们观察到，LMC 在这些指标上的表现远胜其他算法；而 RMHMC 则因步长较小，以及广义蛙跳积分器应用于该后验分布时出现的病态问题而表现不佳。

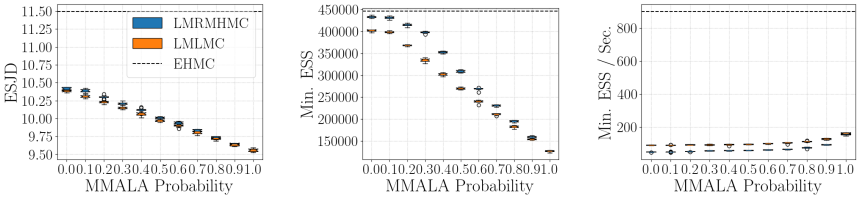
4.3 分层贝叶斯逻辑回归

We 考虑从分层贝叶斯逻辑回归模型中进行抽样 el

$$\alpha \sim \text{Gamma}(\omega, \theta) \tag{72}$$

$$\beta_i | \alpha \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, \alpha^{-1}) \quad \text{for } i = 1, \ldots, m \tag{73}$$

$$y_i | x_i, \beta \overset{\text{indep.}}{\sim} \text{Bernoulli}\left(\frac{1}{1 + \exp(-x_i^\top \beta)}\right) \quad \text{for } i = 1, \ldots, N. \tag{74}$$



(a) 期望平方跳跃距离

(b) 最小有效样本量

(c) 每秒最小有效样本量

图 3：我们展示了分层贝叶斯逻辑回归后验分布推断中的期望平方跳跃距离、最小有效样本量以及按时间归一化的最小有效样本量。我们观察到，在这项推断任务中，两种几何方法的表现都不如标准哈密顿蒙特卡洛。对于 MMALA 的每个混合权重，与 LMC 相比，黎曼流形哈密顿蒙特卡洛具有更大的期望平方跳跃距离和最小有效样本量；然而，由于 LMC 的计算效率更高，其按时间归一化的有效样本量更大。不过，在所有这些指标上，欧几里得哈密顿蒙特卡洛都优于几何推断算法。

我们考虑一个包含 $N = 270$ 个观测值和$m = 14$ 个协变量的逻辑回归数据集。在实验中，我们设定 $\omega = 10$ 和 $\theta = 2$ 。我们采用 Metropolis-within-Gibbs 抽样程序，在以下后验分布之间交替抽样： $\alpha|\beta,\ k,\ \theta$和 $\beta|\alpha,\ \{(x_i,\ y_i)\}_{i=1}^n$；前者可以通过解析方法进行抽样，而后者则采用欧几里得哈密顿蒙特卡洛、黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛进行抽样。在欧几里得哈密顿蒙特卡洛中，我们设定 $\epsilon = 0.1$；在黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛中，我们设定 $\epsilon = 0.8$。我们将 $k_{\max} = 10$ 设为每种情况下积分步数的上限。实现 LMRMHMC 和 LMLMC 时，我们使用费希尔信息与对数先验的负海森矩阵之和作为度量。

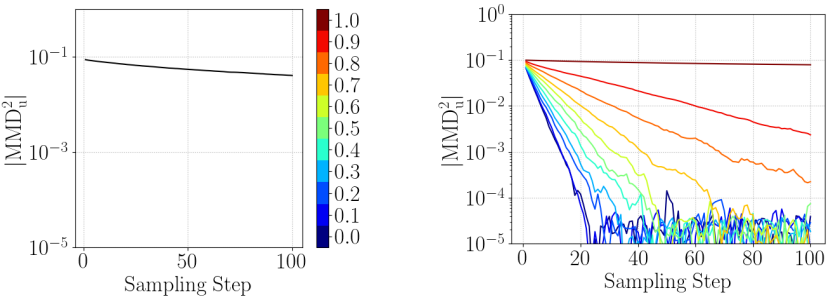
在图 3 中，我们展示了分层贝叶斯逻辑回归后验分布的 ESJD、最小有效样本量和每秒最小有效样本量。这两种几何方法在该后验分布上的表现都不佳，始终逊于欧几里得哈密顿蒙特卡洛。对 LMC 的一种批评是，它在高维情况下的性能会显著下降 [Betancourt 等人， 2014]；LMC 产生的较小有效样本量体现了这一现象。不过，由于省去了不动点迭代而节省的计算开销，LMC 的时间归一化有效样本量仍略胜 RMHMC。

4.4 Neal 漏斗分布

Neal 漏斗分布 [Neal， 2003] 是按以下层次方式定义的密度。

$$v \sim \text{Normal}(0, 9) \tag{75}$$

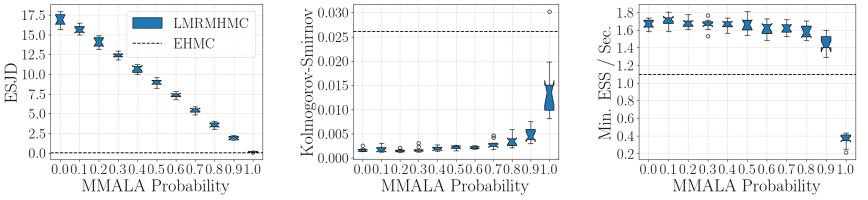
$$x_i|v \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, \exp(-v)) \quad \text{for } i = 1, \ldots, N. \tag{76}$$



(a) 欧几里得哈密顿蒙特卡洛

(b) LMRMHMC

图 4：我们考察马尔可夫链在 Neal 漏斗分布中随采样步长变化的遍历性。我们观察到，由于漏斗形分布固有的多个空间尺度，EHMC 在该分布上的表现不佳。此外，我们发现，使用 SMALA 对 RMHMC 进行激进混合，会使最大均值差异的收敛越来越慢；纯 SMALA算法的收敛速度则比 EHMC 更慢。另一方面，适度地与 SMALA 混合，所产生的遍历性与未经修改的 RMHMC 实现几乎无法区分。



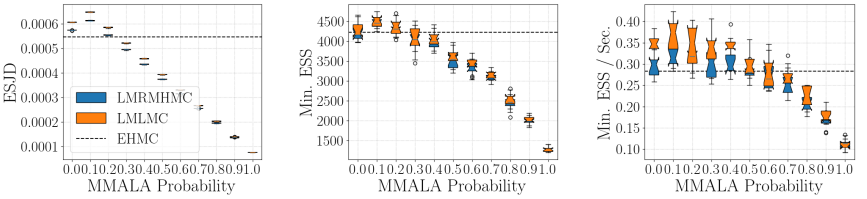
(a) 期望平方跳跃距离

(b) KS

(c) 最小有效样本量 / 秒

图 5：我们展示了 Neal 漏斗分布推断中的期望平方跳跃距离、柯尔莫哥洛夫–斯米尔诺夫统计量的分布，以及按时间归一化的最小有效样本量。尽管黎曼流形哈密顿蒙特卡洛与欧几里得哈密顿蒙特卡洛的按时间归一化有效样本量相当，但从柯尔莫哥洛夫–斯米尔诺夫统计量的分布可以明显看出，黎曼流形哈密顿蒙特卡洛生成的样本比欧几里得哈密顿蒙特卡洛更接近目标分布。

该分布形似漏斗，其中“颈部”的粗细由随机变量 v 控制。该模型反映了稀疏数据层次模型中遇到的后验分布。在此任务中，目标是对 $(v,\ x_1,\ \ldots\ldots,\ x_N)$ 进行联合采样，从而得到一个 $(N+1)$ 维目标分布。在欧几里得哈密顿蒙特卡洛中，我们采用步长为 $\epsilon = 0.1$，并且 $k_{\max} = 10$。我们的黎曼流形哈密顿蒙特卡洛实现使用 $k_{\max} = 20$ 个积分步，步长为 $\epsilon = 0.1$，并采用 SoftAbs度量。此任务中我们不考虑拉格朗日蒙特卡洛，因为我们发现，在保留.....的同时，如何将 SoftAbs 结构扩展到拉格朗日蒙特卡罗框架并不显而易见



(a) 期望平方跳跃距离

(b) 最小有效样本量

(c) 最小有效样本量 / 秒

图 6：我们展示了随机波动率模型后验分布推断中的期望平方跳跃距离、最小有效样本量以及经时间归一化的最小有效样本量。我们观察到，两种几何方法在与 MMALA 采用适度混合概率时，所产生的经时间归一化有效样本量与欧几里得哈密顿蒙特卡洛相当。对于与 MMALA 搭配的每种混合权重，我们观察到，与黎曼流形哈密顿蒙特卡洛相比，拉格朗日蒙特卡洛具有更大的期望平方跳跃距离。

每一步的计算成本为立方级。在本实验中，我们使用 SoftAbs度量。

在图 4 中，我们展示了 |最大均值差异^2_u|，比较了欧几里得哈密顿蒙特卡洛和黎曼流形哈密顿蒙特卡洛。正如预期的那样，由于多尺度现象，欧几里得哈密顿蒙特卡洛难以从 Neal 漏斗分布中采样。另一方面，黎曼流形哈密顿蒙特卡洛表现出更强的收敛性质；在与 SMALA 的多种混合概率下，其误差均呈线性下降。然而，我们发现，与 SMALA 进行过于激进的混合可能适得其反，因为单步方法对目标分布的遍历效率较低。在图5 中，我们展示了期望平方跳跃距离、柯尔莫哥洛夫–斯米尔诺夫统计量以及每秒最小有效样本量。在这些指标上，黎曼流形哈密顿蒙特卡洛都明显优于欧几里得哈密顿蒙特卡洛。

4.5 随机波动率模型

We 考虑如下生成模式的随机波动率模型 l.

$$\frac{\phi+1}{2} \sim \text{Beta}(20, 3/2) \tag{77}$$

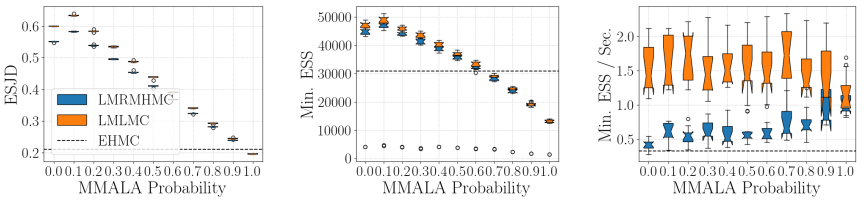
$$\sigma^2 \sim \text{InverseChiSquared}(10, 1/20) \tag{78}$$

$$x_1|\phi,\sigma^2 \sim \text{Normal}(0, \sigma^2/(1-\phi^2)) \tag{79}$$

$$x_{t+1}|x_t,\phi,\sigma^2 \sim \text{Normal}(\phi x_t, \sigma^2) \quad \text{for } t = 2,\ldots,T-1 \tag{80}$$

$$y_t|\beta,x_t \sim \text{Normal}(0, \beta^2\exp(x_t)) \quad \text{for } t = 1,\ldots,T. \tag{81}$$

此外，参数 β 配备了一个与 β^{-1}成比例的非正规先验。在此示例中，我们旨在根据观测值 $y_1,\ldots,y_T$生成联合分布 $x_1,\ldots,x_T,\beta,\phi,\sigma^2$ 的样本。我们采用类似 Metropolis-within-Gibbs 的策略，在以下两种采样之间交替进行：$(x_1,\ldots,x_T)|(y_1,\ldots,y_T)\phi,\sigma^2,\beta$ 和 $(\phi,\sigma^2,\beta)|\{(x_i,y_i)\}_{i=1}^T$；前一种情况下使用欧几里得哈密顿蒙特卡洛，后一种情况下则比较欧几里得



（a）期望平方跳跃距离　（b）最小有效样本量　（c）最小有效样本量/秒

图 7：我们展示了对数高斯 Cox–泊松模型后验分布进行推断时的期望平方跳跃距离、最小有效样本量以及时间归一化的最小有效样本量。对于任何混合概率，几何方法的时间归一化有效样本量均高于欧几里得哈密顿蒙特卡洛，其中拉格朗日蒙特卡洛在这些指标上优于黎曼流形哈密顿蒙特卡洛。对于几乎所有混合概率，我们还观察到，拉格朗日蒙特卡洛和黎曼流形哈密顿蒙特卡洛的期望平方跳跃距离均大于欧几里得哈密顿蒙特卡洛。

哈密顿蒙特卡洛、黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛。对任一分布进行采样时，黎曼度量取为 Fisher 信息与对数先验的负海森矩阵之和。在实验中，我们设定 $T = 1{,}000$。使用欧几里得哈密顿蒙特卡洛对 (ϕ, σ^2, β) 进行采样时，我们使用 $k_{\max} = 50$ 个积分步，步长为 $\epsilon = 0.01$；使用黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛时，我们使用 $k_{\max} = 6$ 个积分步，步长为 $\epsilon = 0.5$。

在图 6 中，我们展示了期望平方跳跃距离、最小有效样本量以及每秒最小有效样本量。我们观察到，MMALA 使用适度的混合概率时，产生的马尔可夫链略优于欧几里得哈密顿蒙特卡洛。

4.6 对数高斯 Cox–泊松过程

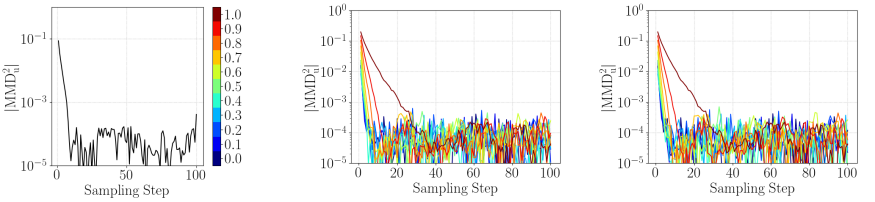
我们考虑对数高斯 Cox–泊松模型中的推断，其生成模型如下：

$$\Sigma_{(i,j),(i',j')}|\sigma^2, \beta = \sigma^2 \exp\left(-\sqrt{(i-i')^2 + (j-j')^2}/(N\beta)\right) \tag{82}$$

$$\text{vec}(\mathbf{x})|\Sigma \sim \text{MultivariateNormal}(\mu\mathbf{1}, \Sigma) \tag{83}$$

$$y_{ij}|x_{ij} \sim \text{Poisson}(\exp(x_{ij})/N^2), \tag{84}$$

先验分布为 β ∼伽马分布$(2, 1/2)$和 σ^2 ∼伽马分布$(2, 1/2)$。在这个例子中，目标是根据观测值 $\{y_{ij}\}_{i}^{N}{}_{,j=1}$ 对联合分布 $(\beta, \sigma^2, \{x_{ij}\}_{i}^{N}{}_{,j=1})$ 进行采样。与随机波动率模型的情况一样，我们交替从给定 $\{y_{ij}\}_{i}^{N}{}_{,j=1}$, σ^2 和 β 时的 $\{x_{ij}\}_{i}^{N}{}_{,j=1}$，以及给定 $\{x_{ij}\}_{i}^{N}{}_{,j=1}$ 时的 (β, σ^2) 中采样。两种情况下，度量均由费希尔信息与对数先验的负海森矩阵之和给出。在实验中，我们设定 $N = 16$。在前一种情况下，我们采用欧几里得哈密顿蒙特卡洛，而在



(a) 欧几里得哈密顿蒙特卡洛

(b) 拉格朗日修正黎曼流形哈密顿蒙特卡洛

(c) LMLMC

图 8：我们考察菲茨休-长门后验分布中马尔可夫链的遍历性随采样步长的变化。我们发现，采用适度的混合概率时，修正黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛的转移核比欧几里得哈密顿蒙特卡洛的转移核混合得更有效，且与未经修正的黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛转移核几乎无法区分。使用较大的混合概率时，我们发现混合速率会降低。

后一种情况下，我们比较欧几里得哈密顿蒙特卡洛、黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛。实现欧几里得哈密顿蒙特卡洛时，我们采用 $k_{\max} = 50$ 个积分步数，步长为 $\epsilon = 0.01$；在黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛中，我们采用 $k_{\max} = 6$ 个积分步数，步长为 $\epsilon = 0.5$。

在图 7 中，我们展示了 ESJD、最小有效样本量以及每秒最小有效样本量。我们观察到，在一定范围的混合概率下，LMRMHMC 和 LMLMC 中采用 MMALA 进行混合的表现优于 EHMC。RMHMC 和 LMC 的最小有效样本量指标相近，但由于 LMC 具有计算优势，因此其每秒最小有效样本量更高。

4.7 菲茨休-南云模型

给定取值于 $\mathbb{R}$ 的参数 a、b 和 c，菲茨休-南云微分方程定义如下：

$$\dot{v}_t = c\left(v_t - \frac{v_t^3}{3} + r_t\right) \tag{85}$$

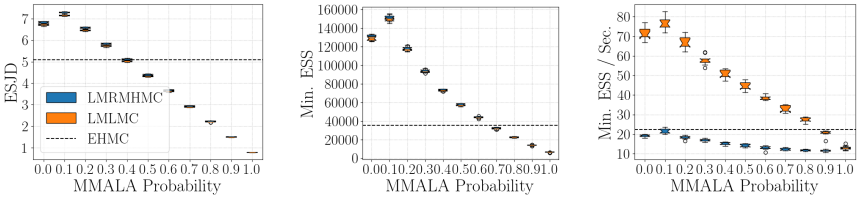
$$\dot{r}_t = -\left(\frac{v_t - a + br_t}{c}\right). \tag{86}$$

Giv在初始条件 v_0 和 r_0下，我们考虑以下生成模式 l:

$$(a, b, c) \overset{\text{i.i.d.}}{\sim} \text{Normal}(0, 1) \tag{87}$$

$$\hat{r}_{t_k} | a, b, c, r_{t_k}, \sigma^2 \overset{\text{indep.}}{\sim} \text{Normal}(r_{t_k}, \sigma^2) \quad \text{for } k = 1, \ldots, n \tag{88}$$

$$\hat{v}_{t_k} | a, b, c, v_{t_k}, \sigma^2 \overset{\text{indep.}}{\sim} \text{Normal}(v_{t_k}, \sigma^2) \quad \text{for } k = 1, \ldots, n, \tag{89}$$



(a) 期望平方跳跃距离

(b) 最小有效样本量

(c) 最小有效样本量 / 秒

图 9：我们展示了针对菲茨休-南云后验分布进行推断时的期望平方跳跃距离、最小有效样本量，以及按时间归一化的最小有效样本量。在这项推断任务中，按照时间归一化的有效样本量指标，拉格朗日蒙特卡洛凭借更高的计算效率优于黎曼流形哈密顿蒙特卡洛和欧几里得哈密顿蒙特卡洛。若不考虑时间因素，则可观察到黎曼流形哈密顿蒙特卡洛与拉格朗日蒙特卡洛的期望平方跳跃距离和最小有效样本量相近。我们还观察到，对于黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛，使用 MMALA 进行激进混合会导致期望平方跳跃距离减小；相应地，最小有效样本量也会降低。

其中 t_1、……、t_n 是 $[0, T]$之间等间距分布的点。在我们的实验中，我们设定 $v_0 = 1$、$r_0 = -1$、$\sigma^2 = 1/4$、$T = 10$和 $n = 200$。我们的度量由费希尔信息与对数先验的负海森矩阵之和给出。在实现中，我们使用步长为 $\epsilon = 0.01$ 、积分步数为$k_{\max} = 10$ 的欧几里得哈密顿蒙特卡洛。在黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛中，我们采用积分步长 $\epsilon = 0.5$ 和积分步数 $k_{\max} = 6$ 。

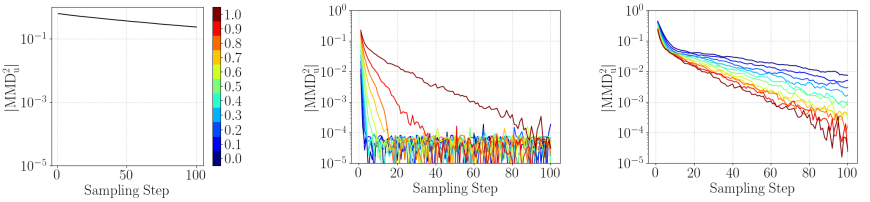
图8 展示了 EHMC、RMHMC 和 LMC 的遍历性度量。我们观察到，这些马尔可夫链蒙特卡洛算法的|最大均值差异^2_u| 均按对数尺度线性下降。在图9中，我们展示了这三种马尔可夫链蒙特卡洛算法的期望平方跳跃距离、最小有效样本量和每秒最小有效样本量。我们观察到，就每秒最小有效样本量而言，LMC 的表现最为出色；而由于复杂度较高，RMHMC 的时间归一化表现仅接近 EHMC。

4.8 多尺度斯图登特分布

固定 $m \in \mathbb{N}$ ，令 $\Sigma \in \mathrm{PD}(m)$ 和 $\nu > 2$。多元斯图登特-t 分布的密度函数为，

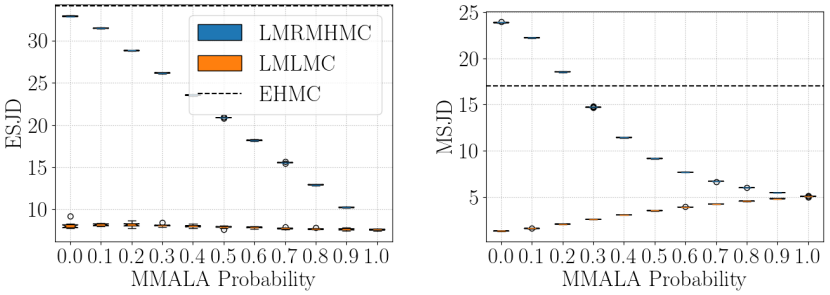
$$\pi(q) \propto \left(1 + \frac{1}{\nu} x^\top \Sigma^{-1} x\right)^{-(m+\nu)/2}. \tag{90}$$

在我们的实验中，我们考虑 $m = 20$、$\nu = 5$和 $\Sigma = \mathrm{diag}(1, \ldots, 1, 10_4) \in \mathbb{R}^{m \times m}$。多元 Student-$t$ 分布中存在差异极大的空间尺度，这会导致采用单位质量矩阵的欧几里得哈密顿蒙特卡洛表现出高度振荡的行为 [Pourzanjani 和 Petzold, 2019]，从而限制该方法在遍历性和有效样本量方面的效率。我们采用步长为 $\epsilon = 0.8$的欧几里得哈密顿蒙特卡洛，



(a) 欧几里得哈密顿蒙特卡洛　(b) 拉格朗日修正黎曼流形哈密顿蒙特卡洛　(c) LMLMC

图 10：我们考察多尺度 Student-t 分布中马尔可夫链的遍历性随采样步长的变化。可以看出，对于 LMC，从遍历性的角度而言，与 MMALA 进行大量混合是有益的；而对于 RMHMC，只有适度的混合概率就能产生一种与未修正版本相竞争的修正马尔可夫链。由于后验分布具有多个空间尺度，欧几里得哈密顿蒙特卡洛的表现不佳。

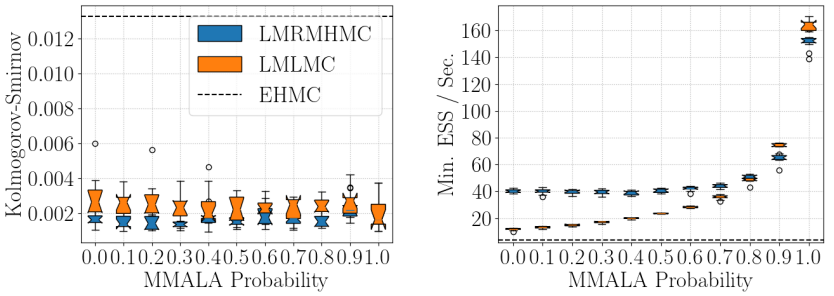


(a) 期望平方跳跃距离　(b) MSJD

图 11：我们展示了在多尺度 Student-t 分布中进行推断时的期望平方跳跃距离和 MSJD。有趣的是，欧几里得哈密顿蒙特卡洛的期望平方跳跃距离最大，但这并不意味着采样效率更高，这一点可从其较差的柯尔莫哥洛夫–斯米尔诺夫统计量分布中看出。为进一步考察这一特性，我们还测量了 MSJD，结果表明，与黎曼流形哈密顿蒙特卡洛几何方法相比，欧几里得哈密顿蒙特卡洛在遍历参数空间方面效果较差。

并采用步长为 $\epsilon = 0.7$的黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛；每种情况下均采用$k_{\max} = 20$ 个积分步数。对于黎曼流形哈密顿蒙特卡洛和拉格朗日蒙特卡洛，我们采用度量，该度量由多元 Student-t 分布的对数密度海森矩阵中的正定项给出。

在图 10 中，我们展示了 $|$最大均值差异$_u^2|$ 随马尔可夫链步数变化的情况。所有马尔可夫链蒙特卡洛算法在对数尺度下，最大均值差异估计值均呈线性下降；然而，这些方法在线性关系的斜率 方面差异显著。可以看到，黎曼流形哈密顿蒙特卡洛的收敛速度明显最快。值得注意的是，拉格朗日蒙特卡洛的表现



(a) KS

(b) 每秒最小有效样本量

图 12：我们展示了多尺度 Student-t 分布推断中柯尔莫哥洛夫–斯米尔诺夫统计量的分布，以及按时间归一化的最小有效样本量。对于几何方法，我们观察到 RMHMC 的期望平方跳跃距离远大于 LMC；然而，就按时间归一化的性能而言，最好将任一方法与 MMALA 完全混合。事实上，我们发现，单独使用 MMALA 所产生的柯尔莫哥洛夫–斯米尔诺夫统计量分布，足以与任何混合核相媲美。

在该目标分布上的收敛速度较慢 于 MMALA；这是由于后验分布的维数较高，拉格朗日蒙特卡洛难以维持接受提议所需的哈密顿能量。在图 12 中，我们展示了期望平方跳跃距离、KS 和每秒最小有效样本量。欧几里得哈密顿蒙特卡洛的期望平方跳跃距离最大，但由于其提议机制呈现振荡行为，这并未转化为较大的有效样本量；尽管计算复杂度较高，黎曼流形哈密顿蒙特卡洛仍取得了最大的每秒最小有效样本量。

5 结论

本文探讨了如何为 RMHMC 和 LMC 建立几何遍历性理论。我们采用的基本技术是，将由单个积分步构成的 RMHMC（或 LMC）转移核替换为 MMALA 转移核。这一改进的灵感来自欧几里得情形，在该情形下，单步 HMC 和 MALA 可以构造为完全等价。通过证明边缘转移核的可逆性，便可继承 MMALA 的几何遍历性。我们在一系列贝叶斯推断任务上评估了 RMHMC 和 LMC 的改进变体，分别称为 LMRMHMC 和 LMLMC。我们发现，使用 MMALA 转移核进行过于频繁的混合，会损害马尔可夫链在多种指标上的表现；但适度混合则能产生与原始 RMHMC 或 LMC 方法相当甚至更优的表现，同时为这些方法提供几何遍历性理论支持。

致谢。 我们感谢耶鲁大学研究计算中心提供研究计算基础设施。本文研究工作得到美国国家科学基金会研究生研究奖学金（1752134号资助）的支持。本文所表达的任何观点、研究发现、结论或建议均属于作者，并不一定反映美国国家科学基金会的观点。本研究还部分得到美国国立卫生研究院／美国国立普通医学科学研究所 R01GM136780 项目和美国空军科学研究办公室 FA9550-21-1-0317 项目的支持。

声明与陈述

无利益冲突需要声明。

附录A 证明

A.1 命题的证明 5

Proof MALA 的提议分布为

$$\tilde{q}|q \sim \text{Normal}\left(q + \frac{\epsilon^2}{2}\mathbf{A}^{-1}\nabla \log \pi(q), \epsilon^2\mathbf{A}^{-1}\right) \tag{A1}$$

$$\tilde{\pi}(\tilde{q}|q) \propto \exp\left(-\frac{1}{2\epsilon^2}\left(\tilde{q} - q - \frac{\epsilon^2}{2}\mathbf{A}^{-1}\nabla \log \pi(q)\right)^\top \mathbf{A}\left(\tilde{q} - q - \frac{\epsilon^2}{2}\mathbf{A}^{-1}\nabla \log \pi(q)\right)\right) \tag{A2}$$

$$= \exp\left(-\frac{1}{2\epsilon^2}\left(\mathbf{A}\tilde{q} - \mathbf{A}q - \frac{\epsilon^2}{2}\nabla \log \pi(q)\right)^\top \mathbf{A}^{-1}\left(\mathbf{A}\tilde{q} - \mathbf{A}q - \frac{\epsilon^2}{2}\nabla \log \pi(q)\right)\right). \tag{A3}$$

因此，MALA 的接受概率为，

$$\min\left\{1, \frac{\pi(\tilde{q})\tilde{\pi}(q|\tilde{q})}{\pi(q)\tilde{\pi}(\tilde{q}|q)}\right\}$$

$$= \min\left\{1, \frac{\pi(\tilde{q})}{\pi(q)} \cdot \frac{\exp\left(-\frac{1}{2\epsilon^2}\left(\mathbf{A}q - \mathbf{A}\tilde{q} - \frac{\epsilon^2}{2}\nabla \log \pi(\tilde{q})\right)^\top \mathbf{A}^{-1}\left(\mathbf{A}q - \mathbf{A}\tilde{q} - \frac{\epsilon^2}{2}\nabla \log \pi(\tilde{q})\right)\right)}{\exp\left(-\frac{1}{2\epsilon^2}\left(\mathbf{A}\tilde{q} - \mathbf{A}q - \frac{\epsilon^2}{2}\nabla \log \pi(q)\right)^\top \mathbf{A}^{-1}\left(\mathbf{A}\tilde{q} - \mathbf{A}q - \frac{\epsilon^2}{2}\nabla \log \pi(q)\right)\right)}\right\}. \tag{A4}$$

给定 q 时的提议 $\tilde{q}$ 可通过生成 $z \sim \text{Normal}(0, \text{Id})$ 并设定

$$\tilde{q} = q + \frac{\epsilon}{2}\mathbf{A}^{-1}\nabla \log \pi(q) + \epsilon\sqrt{\mathbf{A}^{-1}}z. \tag{A5}$$

在哈密顿蒙特卡洛中，蛙跳积分器应用于哈密顿量

$H\ q\ p\ =-\quad \pi\ q\ +(.)\log(\ )\mathbf{A}$。更新序列如下:

$\frac{1}{2}p^\top\ \ ^{-1}p$

$$\bar{p} = p + \frac{\epsilon}{2}\nabla \log \pi(q) \tag{A6}$$

$$\tilde{q}' = q + \epsilon\mathbf{A}^{-1}\bar{p} \tag{A7}$$

$$= q + \frac{\epsilon^2}{2}\mathbf{A}^{-1}\nabla \log \pi(q) + \epsilon\mathbf{A}^{-1}p \tag{A8}$$

$$\tilde{p}' = \bar{p} + \frac{\epsilon}{2}\nabla \log \pi(\tilde{q}') \tag{A9}$$

$$= p + \frac{\epsilon}{2}\nabla \log \pi(q) + \frac{\epsilon}{2}\nabla \log \pi(\tilde{q}'). \tag{A10}$$

从密度正比于 $\exp(-H(q,p))$ 的分布中采样时，显然可知 p 与 q 相互独立，且 $p \sim$ 服从 $\text{Normal}(0, \mathbf{A})$。因此，使用与公式 (A5) 中相同的 z，我们可以通过令 $_p$ 满足 $p = \sqrt{\mathbf{A}}z$ 来对其采样，使得 $\epsilon\mathbf{A}^{-1}p = \epsilon\sqrt{\mathbf{A}^{-1}}z$；因此可见，在这种情况下，HMC 和 MALA 产生完全相同的提议（即 $\tilde{q}' = \tilde{q}$）。回顾一下，HMC 的接受概率为：

$$\min\left\{1, \frac{\pi(\tilde{q})}{\pi(q)} \cdot \frac{\exp\left(-\frac{1}{2}(\tilde{p}')^\top \mathbf{A}^{-1}(\tilde{p}')\right)}{\exp\left(-\frac{1}{2}p^\top \mathbf{A}^{-1}p\right)}\right\}. \tag{A11}$$

重新整理公式 (A7) 可得：

$$p = \frac{1}{\epsilon}\left(\mathbf{A}\tilde{q} - \mathbf{A}q - \frac{\epsilon^2}{2}\nabla \log \pi(q)\right) \tag{A12}$$

$$\tilde{p}' = \frac{1}{\epsilon}\left(\mathbf{A}\tilde{q} - \mathbf{A}q + \frac{\epsilon^2}{2}\nabla \log \pi(\tilde{q})\right) \tag{A13}$$

$$= -\frac{1}{\epsilon}\left(\mathbf{A}q - \mathbf{A}\tilde{q} - \frac{\epsilon^2}{2}\nabla \log \pi(\tilde{q})\right). \tag{A14}$$

将这些结果代入公式 (A11)，并与公式 (A4) 比较可知，不仅 $\tilde{q}$ 和 $\tilde{q}'$ 完全相同，而且接受概率也完全相同。因此，单步 HMC 的边缘链与 MALA 链完全等价。

□

A.2 引理的证明6

Proof

$$\Pr(q^{n+1} \in Q | q^n = q) = \Pr(q^{n+1} \in Q \text{ and } p^{n+1} \in \mathbb{R}^m | q^n = q) \tag{A15}$$

$$= \int_{\mathbb{R}^m} \Pr(q^{n+1} \in Q \text{ and } p^{n+1} \in \mathbb{R}^m | q^n = q, p^n = p)\ \pi(p|q)\ \mathrm{d}p \tag{A16}$$

$$= \int_{\mathbb{R}^m} K((q,p),(Q,\mathbb{R}^m))\ \pi(p|q)\ \mathrm{d}p. \tag{A17}$$

□

A.3 命题的证明7

Proof 给定 $Q, Q' \in \mathfrak{B}(\mathbb{R}^m)$，我们有

$$\int_Q \tilde{K}(q,Q')\ \pi(q)\ \mathrm{d}q = \int_Q \int_{\mathbb{R}^m} K((q,p),(Q',\mathbb{R}^m))\ \pi(p|q)\ \mathrm{d}p\ \pi(q)\ \mathrm{d}q \tag{A18}$$

$$= \int_Q \int_{\mathbb{R}^m} K((q,p),(Q',\mathbb{R}^m))\ \pi(q,p)\ \mathrm{d}p\ \mathrm{d}q \tag{A19}$$

$$= \int_{Q'} \int_{\mathbb{R}^m} K((q,p),(Q,\mathbb{R}^m))\ \pi(q,p)\ \mathrm{d}p\ \mathrm{d}q \tag{A20}$$

$$= \int_{Q'} \int_{\mathbb{R}^m} K((q,p),(Q,\mathbb{R}^m))\ \pi(p|q)\ \mathrm{d}p\ \pi(q)\ \mathrm{d}q \tag{A21}$$

$$= \int_{Q'} \tilde{K}(q,Q)\ \pi(q)\ \mathrm{d}q, \tag{A22}$$

其中，在式（A20）中，我们利用了相空间链相对于 π（q，p）满足细致平衡这一事实。这验证了边缘链满足细致平衡。□

A.4 对合复合的证明

命题 **15** *Suppose that* $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ *is an invertible function, and let* $\mathbf{F}$ *be the momentum flip function given in definition 21. Suppose that* $\mathbf{F} \circ \Phi$ $\mathbf{F}$ *is an involution then* $\Phi^k \circ \mathbf{F}$ *is also an involution.*

Proof 这将通过归纳法证明，基础情形由假设成立。作为归纳假设，假设$\mathbf{F} \circ \Phi^k$ 是一个对合。利用$\mathbf{F} \circ \Phi$ 是一个对合这一事实，我们立即得到 $\Phi^{-1} \circ \mathbf{F} = \mathbf{F} \circ \Phi$。利用归纳假设，还可得到 $\Phi^{-k} \circ \mathbf{F} = \mathbf{F} \circ \Phi^k$。因此，我们得到：

$$\Phi^{-1} \circ \mathbf{F} = \mathbf{F} \circ \Phi \tag{A23}$$

$$\implies \Phi^{-1} \circ \mathbf{F} \circ \Phi^k = \mathbf{F} \circ \Phi^{k+1} \tag{A24}$$

$$\implies \Phi^{-1} \circ \Phi^{-k} \circ \mathbf{F} = \mathbf{F} \circ \Phi^{k+1} \tag{A25}$$

$$\implies \Phi^{-k-1} \circ \mathbf{F} = \mathbf{F} \circ \Phi^{k+1} \tag{A26}$$

$$\implies \mathbf{F} \circ \Phi^{k+1} \circ \mathbf{F} \circ \Phi^{k+1} = \mathrm{Id}. \tag{A27}$$

这验证了$\mathbf{F} \circ \Phi^{k+1}$ 也是一个对合。 □

数据可用性声明

本研究期间生成和/或分析的数据集可在 GitHub 仓库中获取，https://tinyurl.com/29kz7krz。

参考文献

S. Amari 和 H. Nagaoka。信息几何方法。数学专著译丛。美国数学会，2000年。ISBN ISBN9780821843024。网址 https://books.google.com/books?id=vc2FWSo7wLUC。

M. Betancourt。黎曼流形哈密顿蒙特卡罗的一般度量，12 2012年。

M. Betancourt、Simon Byrne、Samuel Livingstone 和 Mark Girolami。《哈密顿蒙特卡洛的几何基础》。《统计科学》，23，2014年10月。

´Alain Durmus、Eric Moulines 和 Eero Saksman。哈密顿蒙特卡洛的不可约性与几何遍历性。《统计年鉴》，48(6): 3545–3564，2020年。

Andrew Gelman 和 Cristian Pasarica。使用期望平方跳跃距离自适应调整 Metropolis 算法的尺度。《统计学报》，2007年。网址：http://dx.doi.org/10.2139/ssrn.1010403。

Mark Girolami 和 Ben Calderhead。黎曼流形朗之万与哈密顿蒙特卡洛方法。《皇家统计学会学报：*B*辑（统计方法学）》，73(2):123–214，2011年。网址：https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00765.x。

Arthur Gretton、Karsten M. Borgwardt、Malte J. Rasch、Bernhard Sch¨olkopf 和 Alexander Smola。核双样本检验。*J. Mach. Learn. Res.*，13（null）：723–773，2012年3月。ISSN 1532-4435。

Ernst Hairer、ChristianLubich 和GerhardWanner。几何数值积分，施普林格计算数学丛书第31卷。Springer-Verlag，柏林，第二版，2006年。ISBN 3-540-30663-3；978- 3-540-30663-4。常微分方程的保结构算法。

E.P. Hsu。流形上的随机分析。《当代数学》。美国数学会，2002年。ISBN 9780821808023。网址https://books.google.com/books?id=GDEPCgAAQBAJ。

Ravin Kumar、Colin Carroll、Ari Hartikainen 和 Osvaldo A. Martin。ArviZ：用于贝叶斯模型探索性分析的统一 Python 库。《开放源代码软件杂志》，2019年。网址http://joss.theoj.org/papers/10.21105/joss.01143。

Shiwei Lan、Vasileios Stathopoulos、Babak Shahbaba 和 Mark Girolami。基于拉格朗日动力学的马尔可夫链蒙特卡罗。《计算与图形统计学》，24(2):357–378，2015年4月。ISSN 1537-2715。网址：http://dx.doi.org/10.1080/10618600.2014.902764。

Anthony Lee 和 Krzysztof ´Latuszy´nski。近似贝叶斯计算中马尔可夫链蒙特卡洛核的方差界与几何遍历性。《生物统计学》，101(3):655–671，2014年。ISSN 00063444。网址：http://www.jstor.org/stable/43304674。

Benedict Leimkuhler 和 Sebastian Reich。《模拟哈密顿动力学》。剑桥应用与计算数学专著丛书。剑桥大学出版社，2005。

34 修正黎曼流形与拉格朗日蒙特卡洛Samuel Livingstone、Michael Betancourt、Simon Byrne 和 Mark Girolami。《哈密顿蒙特卡洛的几何遍历性》。 *Bernoulli*，25，2016 年 01 月。doi：10.3150/18-BEJ1083。Jerrold E. Marsden 和 Tudor S. Ratiu。 力学与对称性导论：经典力学系统基础论述。Springer Publishing Company, Incorporated，2010。ISBN 1441931430。K. L. Mengersen 和 R. L. Tweedie。Hastings 算法和 Metropolis 算法的收敛速率。 《统计年鉴》，24(1):101 – 121，1996 年。doi：10.1214/aos/1033066201。URLhttps://doi.org/10.1214/aos/1033066201。S.P. Meyn 和 R.L. Tweedie。马尔可夫链与随机稳定性。Springer-Verlag，伦敦，1993。Radford M. Neal。切片采样。 《统计年鉴》，31(3):705 – 767，2003。URLhttps://doi.org/10.1214/aos/1056562461。Radford M. Neal。使用哈密顿动力学的马尔可夫链蒙特卡洛。 马尔可夫链蒙特卡罗手册，54:113–162，2010a。Radford M. Neal。使用哈密顿动力学的马尔可夫链蒙特卡洛。 马尔可夫链蒙特卡罗手册，54:113–162，2010b。

Kirill Neklyudov、Max Welling、Evgenii Egorov 和 Dmitry P. Vetrov。《对合式马尔可夫链蒙特卡洛：一个统一框架》。载于 第*37*届国际机器学习会议论文集，*ICML 2020*，*2020*年*7*月*13*日至*18*日，线上会议，第119卷，机器学习研究论文集，第7273–7282页。PMLR，2020年。网址：http://proceedings.mlr.press/v119/neklyudov20a.html。

Arya A. Pourzanjani 和 Linda R. Petzold。用于多尺度分布采样的隐式哈密顿蒙特卡罗，2019年。Gareth O. Roberts 和 Richard L. Tweedie。朗之万分布及其离散近似的指数收敛。*Bernoulli*，2(4)：341 – 363，1996 年。网址 https://doi.org/。Vivekananda Roy。马尔可夫链蒙特卡罗的收敛诊断。统计学及其应用年度评论，7：387–412，2020年。Vivekananda Roy 和 Lijin Zhang。位置依赖型MALA的收敛及其在广义线性混合模型条件模拟中的应用。《计算与图形统计学杂志》，2023。即将出版。Luke Tierney。用于探索后验分布的马尔可夫链。《统计年鉴》，22(4)：1701 – 1728，1994。doi：10.1214/aos/1176325750。

URL https://doi.org/10.1214/aos/1176325750.

Aki Vehtari、Andrew Gelman、Daniel Simpson、Bob Carpenter 和 Paul-Christian B¨urkner。秩正态化、折叠与局部化：一种用于评估马尔可夫链蒙特卡洛收敛性的改进 r^（附讨论）。《贝叶斯分析》，16(2)，2021年6月。ISSN 1936-0975。doi：10.1214/20-ba1221。网址http://dx.doi.org/10.1214/20-BA1221。

T. Xifara、C. Sherlock、S. Livingstone、S. Byrne 和 M. Girolami。朗之万扩散与梅特罗波利斯调整朗之万算法。《统计与概率快报》，91：14 – 19，2014年。ISSN 0167-7152。网址http://www.sciencedirect.com/science/article/pii/S0167715214001333。