arXiv:2302.01378v1 [math.OC] 2 Feb 2023

有限状态上的最优里奇曲率马尔可夫链蒙特卡罗方法

Wuchen Li 和 Linyuan Lu

Abstract. 我们在有限状态空间上构造了一种新的马尔可夫链蒙特卡罗方法，并对接受-拒绝比率函数进行了最优选择。我们证明，所构造的连续时间马尔可夫跳跃过程具有全局时间收敛速率，且该速率以 L^1 距离衡量。收敛速率不低于二分之一，且与目标分布无关。例如，我们的方法在两点状态空间上可还原为 Metropolis–Hastings（MH）算法。此外，该方法还构成了一种用于对一般目标分布进行采样的新算法。我们通过数值示例展示了所提算法的有效性。

1. Introduction

马尔可夫链蒙特卡罗（MCMC）方法 [13]是科学计算、统计学和贝叶斯反问题中的重要计算算法，并应用于机器学习领域 [6, 17]。MCMC方法从目标分布中生成随机样本，无论采样空间维数很高，还是目标分布的表达式难以处理。典型的MCMC方法是Metropolis–Hastings（MH）算法。该算法构造一个接受-拒绝型马尔可夫链过程，再据此从目标分布中生成样本。一般的MCMC方法已得到广泛研究 [16, 17]。

MCMC 算法的收敛性分析是一个关键问题 [17]。近年来，人们已知可逆马尔可夫过程在最优传输型度量空间中构成梯度流；连续状态的情形见 [1, 3, 19], ，离散状态的情形见 [4, 12, 14], ，在统计物理中也称为 Onsager 梯度流[7, 18]。最优传输型度量空间中相对熵（散度函数/自由能）的 Hessian 算子，为建立 MCMC 方法的收敛速率提供了凸分析框架。例如，在有限状态空间上，马尔可夫跳跃过程的收敛速率由 Hessian 矩阵的最小特征值决定 [5, 15]。这个最小特征值称为有限状态马尔可夫过程的熵里奇曲率下界。我们也将其称为平均场信息海森矩阵的最小特征值 [11]。

从现在起，我们着重构造连续时间有限状态马尔可夫跳跃过程。由此自然产生了一个“逆问题”。我们能否将最优传输型度量空间中的收敛性分析应用于构造一种马尔可夫跳跃过程，以采样目标分布，并达到最优（最大）的收敛速率？

Key words and phrases. 传输信息几何；概率单纯形中的海森矩阵；广义 Metropolis–Hastings 算法；MSC 分类代码：46N10；05C21；60J27。W Li 获得 AFOSR MURI FA9550-18-1-0502 和 AFOSR YIP 奖（2023 年）的资助。W. Li 和 L. Lu 均获得 NSF RTG: 2038080 的资助。

具有最优里奇曲率下界的有限状态 *MCMC* 算法是什么？

本文通过求解最优里奇曲率下界问题，设计了一种有限状态马尔可夫跳跃过程。我们构造了该马尔可夫跳跃过程的一个特定 Q-矩阵（生成元）。随后，我们对所构造的连续时间 MCMC 方法在 ϕ-散度下进行了指数收敛性分析，其中库尔贝克-莱布勒散度是一个重要示例。我们还证明，对于任意目标分布，全局时间收敛速率在 L^1 距离下至少可以达到二分之一。我们还通过数值实验验证了所提出的算法比 Metropolis–Hastings 算法收敛更快。

主要结果概述如下。给定一个有限状态 $I = \{1, \cdots, n\}$，假设存在一个目标概率分布 $\pi = (\pi_i)_{i=1}^n \in \mathbb{R}_+^n$，$n \in \mathbb{N}$，且 $\sum_{i=1}^n \pi_i = 1$，我们构造一个 Q-矩阵，它是一个连续时间马尔可夫跳跃过程的生成元：

$$Q_{ij} = \begin{cases} \dfrac{\pi_j}{1 - \min_{k \in I} \pi_k}, & \text{if } j \neq i; \\ -\dfrac{1 - \pi_i}{1 - \min_{k \in I} \pi_k}, & \text{if } j = i. \end{cases} \tag{1}$$

由 Q-矩阵 (1) 诱导的马尔可夫跳跃过程满足以下收敛结果。

定理 1 (非正式)。记 $p(t) = (p_i(t))_{i=1}^n \in \mathbb{R}_+^n$, $t \geq 0$ 为由 Q-矩阵 (1) 所定义的马尔可夫跳跃过程的概率分布。换言之，$p(t)$ 满足 *Kolomogrov* 前向方程：

$$\frac{dp_i(t)}{dt} = \sum_{j=1}^n \Big[Q_{ji} p_j(t) - Q_{ij} p_i(t) \Big],$$

初始分布为 $p(0) \in \mathbb{R}_+^n$, $\sum_{i=1}^n p_i(0) = 1$。那么

$$\sum_{i=1}^n |p_i(t) - \pi_i| \leq C e^{-\kappa t},$$

其中 $C = \sqrt{2 \sum_{i=1}^n p_i(0) \log \frac{p_i(0)}{\pi_i}} > 0$。且 $\kappa > 0$ 是满足以下条件的常数

$$\kappa \geq \frac{1}{1 - \min_{k \in I} \pi_k} \cdot \min_{i,j \in I} \left(1 - \frac{1}{2} (\sqrt{\pi_i} - \sqrt{\pi_j})^2 \right) \geq \frac{1}{2}.$$

我们指出，若 $n = 2$，则两点状态空间上的 Q-矩阵 (1) 恰好是 Metropolis–Hastings 算法中的生成元。换言之，$Q_{12} = \min\left\{1, \frac{\pi_2}{\pi_1}\right\}$。当 $n \geq 3$ 时，Q-矩阵 (1) 不同于 Metropolis–Hastings 算法中的生成元；参见第 4 节中的例 5。文献中已对 MCMC 算法的收敛性分析进行了研究 [20]。与以往工作相比，我们构造了一种新的 MCMC 采样生成元 Q-矩阵 (1)，用于求解有限状态马尔可夫过程收敛速率优化中的一个逆问题。并且，在 L_1 距离下，其收敛速率不低于二分之一，且与目标分布 π 的任意选择无关。我们的工作是输运信息几何凸分析 [8, 9, 10, 11] 的自然进展，旨在设计并计算具有快速收敛速率保证的 MCMC 算法。

本文结构安排如下。第 2 节回顾可逆跳跃马尔可夫过程收敛性分析的一些事实。基于李雅普诺夫泛函海森矩阵的最小特征值，我们利用 ϕ-散度（亦称相对熵或自由能）推导收敛速率。我们还给出了里奇曲率下界的优化问题，旨在设计用于目标分布采样的可逆马尔可夫跳跃过程。利用这一最优性条件，我们构造了一个生成元为 Q-矩阵 (1) 的马尔可夫跳跃过程。第 3 节介绍主要结果。我们证明了所构造的马尔可夫跳跃过程具有全局时间指数收敛性，并给出了解析形式的“最优收敛速率”。证明见第 3.1 节。我们还在第 4 节给出了若干两点和三点状态空间上的 Q-矩阵示例。在 250、500、1000 和 2000 个状态下进行的数值实验表明，所提出的算法比 Metropolis–Hastings 算法收敛更快。

2. Optimal Ricci Curvature problems for finite state MCMC methods

本节介绍本文的研究动机。首先，我们简要回顾可逆马尔可夫跳跃过程及其在连续时间 MCMC 方法中的构造。接着，我们给出一个常数，该常数决定了马尔可夫跳跃过程的收敛速率。该常数由最优传输型度量空间中 ϕ-散度的 Hessian 矩阵的最小特征值（里奇曲率下界）导出。上述两个步骤是“正问题”：针对给定的目标分布构造 MCMC 方法，并推导其收敛速率。最后，我们介绍一个逆问题。这是一类优化问题，旨在设计具有最大收敛速率的马尔可夫跳跃过程。通过求解一个“局部”收敛速率问题，我们导出了 Q-矩阵 (1)。

2.1. 可逆 MCMC 方法与对称加权图。

本小节中，我们使用对称加权图设计一种 MCMC 方法。它是一个可逆马尔可夫跳跃过程，其平稳分布为给定的目标分布。假设存在一个满足以下条件的目标概率分布函数

$$\pi = (\pi_i)_{i=1}^n \in \mathbb{R}_+^n, \qquad \sum_{i=1}^n \pi_i = 1.$$

我们构造一个生成元为 Q-矩阵的马尔可夫跳跃过程，其平稳分布满足 π。

定义 1. 考虑一个带自环的对称加权图(I, ω, E)，其中$I = \{1, \cdots, n\}$是顶点集，$\omega = (\omega_{ij})_{1\le i,j\le n} \in \mathbb{R}_+^{n\times n}$是一个对称权重矩阵，满足

$$\omega_{ij} = \omega_{ji} \ge 0, \quad \omega_{ii} \ge 0,$$

且 $E = \{(i, j)\colon\ \omega_{ij} > 0\}$ 是一个边集。假设

$$\sum_{j=1}^n \omega_{ij} = \pi_i, \quad \textit{for any } i \in I.$$

定义一个矩阵 $Q \in \mathbb{R}^{n\times n}$，使得

$$Q_{ij} = \begin{cases} \dfrac{\omega_{ij}}{\pi_i}, & for\ j \neq i; \\ -\displaystyle\sum_{k=1,k\neq i}^{n} \dfrac{\omega_{ik}}{\pi_i}, & for\ j = i. \end{cases} \tag{2}$$

Q-矩阵 (2) 是有限状态 $\{1, 2, \cdots, n\}$上的连续时间可逆马尔可夫链的生成元。换言之， Q-矩阵满足行和为零条件:

$$Q_{ij} \geq 0, \quad \text{for } j \neq i, \quad Q_{ii} = -\sum_{j=1, j\neq i}^{n} Q_{ij}.$$

该连续时间马尔可夫链是可逆的，因为它满足细致平衡关系:

$$Q_{ij}\pi_i = Q_{ji}\pi_j = \omega_{ij} = \omega_{ji}. \tag{3}$$

马尔可夫过程的概率律 $p_i(t) \in \mathbb{R}_+, i \in I$ 满足柯尔莫哥洛夫前向方程:

$$\frac{dp_i(t)}{dt} = \sum_{j=1}^{n} \Big[Q_{ji}p_j(t) - Q_{ij}p_i(t)\Big] = \sum_{j=1}^{n} \omega_{ij}\Big[\frac{p_j(t)}{\pi_j} - \frac{p_i(t)}{\pi_i}\Big]. \tag{4}$$

注意， π 是方程 (4) 的一个驻点。换言之，对于任意 $i \in I$,

$$\frac{d\pi_i}{dt} = \sum_{j=1}^{n} \Big[Q_{ji}\pi_j - Q_{ij}\pi_i\Big] = \sum_{j=1}^{n} \omega_{ij}\Big[\frac{\pi_j}{\pi_j} - \frac{\pi_i}{\pi_i}\Big] = 0.$$

2.2. **李雅普诺夫方法与收敛速率**。接下来，我们应用李雅普诺夫方法研究方程 (4) 的收敛行为。在有限状态概率空间上定义 ϕ–散度（相对熵/自由能）:

$$\mathrm{D}_\phi(p\|\pi) := \sum_{i=1}^{n} \phi\left(\frac{p_i}{\pi_i}\right)\pi_i,$$

其中， $\phi \in C^2(\mathbb{R}_+^1; \mathbb{R})$ 是一个凸函数，且 $\phi(1) = 0$ 和 $\phi'(1) = 0$。将 ϕ-散度作为李雅普诺夫函数，方程 (4) 构成一个梯度流，称为昂萨格梯度流 [18]:

$$\begin{aligned} \frac{dp_i}{dt} &= \sum_{j=1}^{n} \omega_{ij}\Big[\frac{p_j}{\pi_j} - \frac{p_i}{\pi_i}\Big] \\ &= \sum_{j=1}^{n} \omega_{ij} \frac{\frac{p_j}{\pi_j} - \frac{p_i}{\pi_i}}{\phi'\left(\frac{p_j}{\pi_j}\right) - \phi'\left(\frac{p_i}{\pi_i}\right)}\Big[\phi'\left(\frac{p_j}{\pi_j}\right) - \phi'\left(\frac{p_i}{\pi_i}\right)\Big] \\ &= \sum_{j=1}^{n} \theta_{ij}(\omega, p)\Big[\phi'\left(\frac{p_j}{\pi_j}\right) - \phi'\left(\frac{p_i}{\pi_i}\right)\Big] \\ &= \sum_{j=1}^{n} \theta_{ij}(\omega, p)(\partial_{p_j} - \partial_{p_i})\mathrm{D}_\phi(p\|\pi), \end{aligned} \tag{5}$$

其中，我们利用了这样一个事实： $\partial_{p_i}\mathrm{D}_\phi(p\|\pi) = \phi'(\frac{p_i}{\pi_i})$，其中

$$\theta_{ij}(\omega,p) := \omega_{ij}\theta\left(\frac{p_i}{\pi_i},\frac{p_j}{\pi_j}\right) \geq 0, \quad \text{and} \quad \theta(x,y) := \frac{x-y}{\phi'(x)-\phi'(y)}.$$

沿着动力学 (4)， ϕ-散度按如下方式衰减：

$$\begin{aligned}\frac{d}{dt}\mathrm{D}_\phi(p(t)\|\pi) =& \sum_{i=1}^n \phi'\left(\frac{p_i(t)}{\pi_i}\right)\cdot\frac{dp_i(t)}{dt}\\ =& \sum_{i=1}^n \phi'\left(\frac{p_i(t)}{\pi_i}\right)\sum_{j=1}^n \theta_{ij}(\omega,p(t))\left[\phi'\left(\frac{p_j(t)}{\pi_j}\right)-\phi'\left(\frac{p_i(t)}{\pi_i}\right)\right]\\ =& -\frac{1}{2}\sum_{i,j=1}^n \theta_{ij}(\omega,p(t))\left[\phi'\left(\frac{p_j(t)}{\pi_j}\right)-\phi'\left(\frac{p_i(t)}{\pi_i}\right)\right]^2 \leq 0.\end{aligned}$$

接下来，我们给出沿着动力学 (4) 的 ϕ-散度的海森矩阵，并据此推导方程 (4) 的收敛速率。计算李雅普诺夫泛函海森矩阵的方法称为图上的伽马演算或平均场信息海森矩阵；参见 [11]。而收敛速率通常称为“广义里奇曲率下界”，或最优传输型度量空间中的测地凸性；参见 [12, 15]。

定义 2。对于任意 $f\in\mathbb{R}^n$，定义一个伽马一算子：

$$\Gamma_1(\omega,p,f,f) = \frac{1}{2}\sum_{i,j=1}^n (f_i-f_j)^2\theta_{ij}(\omega,p).$$

定义一个伽马二算子：

$$\Gamma_2(\omega,p,f,f) = \frac{1}{2}\sum_{i,j=1}^n (f_i-f_j)^2 a_{ij}(\omega,p),$$

其中

$$\begin{aligned}a_{ij}(\omega,p) :=& \frac{1}{2}\sum_{k=1}^n\Big[\frac{\partial\theta_{ij}}{\partial p_i}\eta_{ki}+\frac{\partial\eta_{ij}}{\partial p_i}\theta_{ki}+\frac{\partial\eta_{jk}}{\partial p_j}\theta_{ij}-\frac{\partial\eta_{ki}}{\partial p_k}\theta_{jk}\\ &-\frac{\partial\theta_{ij}}{\partial p_j}\eta_{jk}-\frac{\partial\eta_{ij}}{\partial p_j}\theta_{jk}-\frac{\partial\eta_{ki}}{\partial p_i}\theta_{ij}+\frac{\partial\eta_{jk}}{\partial p_k}\theta_{ki}\Big],\end{aligned}$$

with

$$\eta_{ij}(\omega,p) := \theta_{ij}(\omega,p)(\partial_{p_j}-\partial_{p_i})\mathrm{D}_\phi(p\|\pi) = \omega_{ij}\left(\frac{p_j}{\pi_j}-\frac{p_i}{\pi_i}\right).$$

定义 3 (收敛速率/里奇曲率**)**。 定义一个尽可能大的标量$\kappa(\omega,p)\in\mathbb{R}$，使得

$$\Gamma_2(\omega,p,f,f)\geq\kappa(\omega,p)\Gamma_1(\omega,p,f,f),$$

对于任意 $f\in\mathbb{R}^n$，以下估计成立：

$$\kappa(\omega,p)\geq\min_{(i,j)\in E}\frac{a_{ij}(\omega,p)}{\theta_{ij}(\omega,p)}.$$

ϕ-散度对于动力学 (4) 的以下收敛结果成立。

推论 1 (收敛分析 [11])。 假设存在一个正标量$\kappa > 0$，使得 $\kappa(\omega, p) > \kappa > 0$ 对任意 $p \in \mathbb{R}^n_+$ 且满足 $\sum_{i=1}^n p_i = 1$ 均成立，并且 $p(t)$ 满足方程 (4)，则

$$\mathrm{D}_\phi(p(t)\|\pi) \le e^{-2\kappa t}\mathrm{D}_\phi(p(0)\|\pi),$$

对于任意初始条件 $p(0) \in \mathbb{R}^n_+$，满足 $\sum_{i=1}^n p_i(0) = 1$。

证明。 由 [11], 中的引理 5 可得

$$\frac{d}{dt}\mathrm{D}_\phi(p(t)\|\pi) = -\Gamma_1(\omega, p(t), f(t), f(t))|_{f(t)=\phi'(\frac{p(t)}{\pi})},$$

and

$$\frac{d^2}{dt^2}\mathrm{D}_\phi(p(t)\|\pi) = 2\Gamma_2(\omega, p(t), f(t), f(t))|_{f(t)=\phi'(\frac{p(t)}{\pi})}.$$

根据常数 κ 的定义，可得

$$\frac{d^2}{dt^2}\mathrm{D}_\phi(p(t)\|\pi) \ge -2\kappa\frac{d}{dt}\mathrm{D}_\phi(p(t)\|\pi).$$

在时间区间 $[t, \infty)$ 上积分，其中 $t > 0$，并利用 $\phi(1) = 0$、$\phi'(1) = 0$ 这一事实，可得

$$\frac{d}{dt}\mathrm{D}_\phi(p(t)\|\pi) \le -2\kappa\mathrm{D}_\phi(p(t)\|\pi).$$

根据格朗沃尔不等式，我们证明了动力学(4)的指数收敛结果。□

2.3. **最优里奇曲率问题**。现在我们来介绍最优里奇曲率问题。我们设计一个加权矩阵函数 $\omega \in \mathbb{R}^{n\times n}_+$，使其最大化方程 (4) 的收敛速率 κ。

我们提出以下优化问题，以最大化动力学 (4) 的全局时间收敛速率。

定义 4 （最优全局收敛速率问题）。考虑以下极小极大问题:

$$\max_{\omega\in\mathbb{R}^{n\times n}_+}\min_{p\in\mathbb{R}^n_+}\ \kappa(\omega, p), \tag{6}$$

其中，最大化是在所有可能的权重矩阵 $\omega \in \mathbb{R}^{n\times n}_+$ 上进行的，最小化是在满足以下条件的所有离散概率函数 $p \in \mathbb{R}^n_+$ 上进行的:

$$\sum_{i=1}^n p_i = 1, \quad p_i \ge 0, \quad i \in I,$$
$$\sum_{j=1}^n \omega_{ij} = \pi_i, \quad \omega_{ij} = \omega_{ji} \ge 0, \quad \omega_{ii} \ge 0, \quad i, j \in I.$$

求解极小极大问题 (6)，可以得到一个“最优”权重矩阵函数，使马尔可夫链蒙特卡罗方法的概率密度以理想的最大化全局时间速率收敛到目标分布。一般而言，从极小极大问题 (6) 推导解析最优性条件并非易事。

从现在起，我们提出一种替代方法。我们得到一个简单的变分问题，它只最大化动力学（4）在平稳分布附近的局部收敛速率。换言之，我们最大化平稳分布 $p = \pi$ 处的收敛速率 $^{a_{ij}}{}_{(\omega,p)}\ \theta_{ij(\omega,p)}$。

定义 5 (局部最优收敛速率问题**)**。 考虑以下最大化问题：

$$\max_{\omega\in\mathbb{R}_+^{n\times n}} \min_{(i,j)\in E} \frac{a_{ij}(\omega,\pi)}{\theta_{ij}(\omega,\pi)}, \tag{7}$$

其中，最大化是在所有可能的权重矩阵 $\omega \in \mathbb{R}_+^{n\times n}$ 上进行的，使得

$$\sum_{j=1}^{n}\omega_{ij} = \pi_i, \quad \omega_{ij} = \omega_{ji} \geq 0, \quad \omega_{ii} \geq 0, \quad i,j \in I.$$

接下来，我们推导一类权重函数，用于求解变分问题 **(7)** 的临界点。具体如下。

定义 6. 记 $\omega^* \in \mathbb{R}_+^{n\times n}$，满足以下条件：存在一个常数 $c > 0$，使得

$$c = \frac{1}{1-\min_{k\in I}\pi_k},$$

and

$$\omega_{ij}^* = \begin{cases} c\pi_i\pi_j, & \text{for } i \neq j; \\ (1-c)\pi_i + c\pi_i^2, & \text{for } i = j. \end{cases}$$

在这种情况下，Q 矩阵 (2) 满足

$$Q_{ij} = \begin{cases} c\pi_j, & \text{for } j \neq i; \\ -c(1-\pi_i), & \text{for } j = i. \end{cases}$$

这正是 Q 矩阵 (1)。

权重矩阵的推导：我们首先证明以下命题。
命题： 以下恒等式成立：

$$a_{ij}(\omega,\pi) = \omega_{ij}\Big[\frac{\sum_{k\neq i}\omega_{ik}}{\pi_i} + \frac{\sum_{k\neq j}\omega_{jk}}{\pi_j}\Big] - \sum_{k\neq i, k\neq j}\frac{\omega_{ik}\omega_{jk}}{\pi_k}.$$

命题 *2* 的证明。我们回顾一下， $\Gamma_2(p,f,f)$ 可以写成二次型 $\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(\omega,p)(f_i-f_j)^2$。注意到 $\frac{\partial\eta_{ij}}{\partial p_i} = \frac{\omega_{ij}}{\pi_i}$ 和 $\frac{\partial\eta_{ij}}{\partial p_j} = -\frac{\omega_{ij}}{\pi_j}$。因此

$$\begin{aligned} a_{ij}(\omega,p) =& \frac{1}{2}\sum_{k=1}^{n}\Big(\frac{\partial\theta_{ij}}{\partial p_i}\eta_{ki} + \frac{\partial\eta_{ij}}{\partial p_i}\theta_{ki} + \frac{\partial\eta_{jk}}{\partial p_j}\theta_{ij} - \frac{\partial\eta_{ki}}{\partial p_k}\theta_{jk} \\ & - \frac{\partial\theta_{ij}}{\partial p_j}\eta_{jk} - \frac{\partial\eta_{ij}}{\partial p_j}\theta_{jk} - \frac{\partial\eta_{ki}}{\partial p_i}\theta_{ij} + \frac{\partial\eta_{jk}}{\partial p_k}\theta_{ki}\Big) \\ =& -\frac{1}{2}\Big(\frac{\partial\theta_{ij}}{\partial p_i} - \frac{\partial\theta_{ij}}{\partial p_j}\Big)\eta_{ij} + \Big(\frac{\partial\eta_{ij}}{\partial p_i} - \frac{\partial\eta_{ij}}{\partial p_j}\Big)\theta_{ij} \\ & + \frac{1}{2}\sum_{k\neq i,j}\Big(\frac{\partial\eta_{ij}}{\partial p_i}\theta_{ki} + \frac{\partial\eta_{jk}}{\partial p_j}\theta_{ij} - \frac{\partial\eta_{ki}}{\partial p_k}\theta_{jk} \\ & - \frac{\partial\eta_{ij}}{\partial p_j}\theta_{jk} - \frac{\partial\eta_{ki}}{\partial p_i}\theta_{ij} + \frac{\partial\eta_{jk}}{\partial p_k}\theta_{ki}\Big). \end{aligned}$$

注意，当 $p = \pi$ 时，我们有

$$\frac{\partial\theta_{ij}(\omega,p)}{\partial p_i}\eta_{ki}(\omega,p)|_{p=\pi} = 0, \quad \theta_{ij}(\omega,\pi) = \omega_{ij}, \quad \text{for any } i,j,k \in I.$$

因此我们得到下面的 $a_{ij}(\omega,\ \pi)$:

$$a_{ij}(\omega,\pi)=\omega_{ij}^2\Big(\frac{1}{\pi_i}+\frac{1}{\pi_j}\Big)+\sum_{k\neq i,j}\Big(\frac{\omega_{ij}\omega_{ik}}{\pi_i}+\frac{\omega_{ij}\omega_{jk}}{\pi_j}-\frac{\omega_{ik}\omega_{jk}}{\pi_k}\Big).$$

证明至此完成。 □

根据该命题，我们定义

$$F_{ij}(\omega):=\frac{a_{ij}(\omega,\pi)}{\theta_{ij}(\omega,\pi)}=\frac{\sum_{k\neq i}\omega_{ik}}{\pi_i}+\frac{\sum_{k\neq j}\omega_{jk}}{\pi_j}-\sum_{k\neq i,k\neq j}\frac{\omega_{ik}\omega_{jk}}{\pi_k\omega_{ij}}.$$

对于 $i,\ j,\ k\in I$，其中 $i\neq j$、 $i\neq k$，我们令

$$\frac{\partial}{\partial\omega_{ik}}F_{ij}(\omega)=\frac{1}{\pi_i}-\frac{\omega_{jk}}{\pi_k\omega_{ij}}=0.$$

这意味着

$$\frac{\pi_k}{\pi_i}=\frac{\omega_{jk}}{\omega_{ij}},\quad \text{for any } i\neq j,\ i\neq k,\ i,j,k\in I. \tag{8}$$

方程 (8) 的一个解满足

$$\omega_{ij}^*=\omega_{ij}=c\pi_i\pi_j,\quad \text{for } i\neq j,$$

其中 c 是一个常数。而常数 c 是以下优化问题的解:

$$\max_{c\in\mathbb{R}_+}\ c\quad \text{s.t.}\quad c\geq 0,\quad (1-c)\pi_i+c\pi_i^2\geq 0,\quad \text{for } i\in I.$$

我们有

$$c\leq\frac{1}{1-\pi_i},\quad \text{for any } i\in I.$$

因此，最优常数满足 $c=\min_{k\in I}\{\frac{1}{1-\pi_k}\}$，由此完成推导。□

注 1. 我们指出，两个变分问题(6) 和(7) 在二点空间上都有闭式临界点。可以证明，这相当于寻找权重矩阵的最优取值:

$$\max_{\omega_{12}}\Big\{\omega_{12}\colon\quad \omega_{11}+\omega_{12}=\pi_1,\quad \omega_{21}+\omega_{22}=\pi_2,\quad \omega_{ij}\geq 0,\quad i,j\in\{1,2\}\Big\}.$$

因此，最优权重函数满足 $\omega_{12}^*=$ 最小值$\{\pi_1,\pi_2\}$。于是， $Q_{12}=$最小值$\{1,\frac{\pi_2}{\pi_1}\}$，这就是两点状态空间上 *Metropolis-Hastings*算法的生成元。换言之，两点状态空间上的 *Metropolis-Hastings*算法使里奇曲率/收敛速率达到最大。然而，当$n\geq 3$ 时，情况并非如此。稍后在第 *4* 节中，我们将证明，对于 $n=3$， Q-矩阵 (1) 不同于 *Metropolis-Hastings*算法的生成元。

3. Main results

在本节中，我们始终考虑 Q-矩阵 (1) 及其相应的 Kolmogorov 前向方程 (4)。我们使用 ϕ-散度作为李雅普诺夫函数，并研究动力学 (4) 的指数收敛结果。

定理 2。 记 $p(t) = (p_i(t))_{i=1}^n \in \mathbb{R}_+^n$, $t \geq 0$ 满足柯尔莫哥洛夫前向方程 (4)。假设 $\phi \in C^2(\mathbb{R}_+;\mathbb{R})$、$\phi(x)$ 关于 x 是凸的，且 $\phi''(x) \geq 0$，以及 $\phi(1) = \phi'(1) = 0$。那么，ϕ-散度随时间指数快速收敛至零：

$$\mathrm{D}_\phi(p(t)\|\pi) \leq e^{-2\kappa t}\mathrm{D}_\phi(p(0)\|\pi),$$

对于任意初始条件 $p(0) \in \mathbb{R}_+^n$，其中 $\sum_{i=1}^n p_i(0) = 1$。并且，全局时间收敛速率 $\kappa > 0$ 是一个常数，满足

$$\kappa = \frac{1}{1-\min_{k\in I}\pi_k} \cdot \min_{i,j\in I}\left(1 - \frac{1}{2}(\pi_i+\pi_j) + \frac{1}{2}\xi_\phi(\pi_i,\pi_j)\right).$$

这里，$\xi_\phi(\pi_i,\pi_j)$ 是一个定义如下的正二元函数

$$\xi_\phi(\pi_i,\pi_j) = \inf_{(p_i,p_j)\in[0,1]^2} \left(\phi''\left(\frac{p_i}{\pi_i}\right)\pi_j + \phi''\left(\frac{p_j}{\pi_j}\right)\pi_i\right) \cdot \frac{\frac{p_i}{\pi_i} - \frac{p_j}{\pi_j}}{\phi'(\frac{p_i}{\pi_i}) - \phi'(\frac{p_j}{\pi_j})} \geq 0.$$

特别地，全局时间收敛速率 κ 的下界为 1 2：

$$\kappa \geq \frac{1}{2}.$$

从现在起，我们将 ϕ-散度视为 α-散度，参见 [2]。即，

$$\phi(x) = \begin{cases} \dfrac{x^\alpha - 1 - \alpha(x-1)}{\alpha(\alpha-1)}, & \alpha \neq 0,1; \\ 1 - x + x\log x, & \alpha = 0; \\ x - 1 - \log x, & \alpha = 1. \end{cases} \tag{9}$$

在上述公式中，log 表示自然对数函数。在这种情况下，我们将详细分析二元函数 ξ_ϕ 以及全局时间收敛速率 κ。

命题 1。 假设 ϕ 定义于(9)中。如果 $\alpha \in [0,2]$，则

$$\xi_\phi(\pi_i,\pi_j) \geq 2\sqrt{\pi_i\pi_j},$$

并且全局时间收敛速率 κ满足

$$\kappa \geq \frac{1}{1-\min_{k\in I}\pi_k} \cdot \min_{i,j\in I}\left(1 - \frac{1}{2}(\sqrt{\pi_i} - \sqrt{\pi_j})^2\right).$$

随后我们给出若干 ϕ-散度的例子。典型例子包括卡方散度（χ^2）、逆向 Kullback–Leibler 散度（KL）和 KL 散度。利用这些例子，我们证明了动力学（4）的全局时间收敛结果，并得到解析收敛率。

例 1。 设 $\alpha = 2$ 和 $\phi(x) = \frac{x^2-x}{2}$。则 ϕ-散度构成 χ^2-散度：

$$\mathrm{D}_\phi(p\|\pi) = \frac{1}{2}\sum_{i=1}^n \frac{p_i^2}{\pi_i}.$$

在这种情况下，我们有

$$\xi_\phi(\pi_i,\pi_j)=\pi_i+\pi_j.$$

根据定理 *2*，在动力学 (4)下， χ -散度以指数收敛速率趋于零:

$$\kappa=\frac{1}{1-\min_{k\in I}\pi_k}.$$

例2. 令 $\alpha=1$ 和 $\phi(x)=x-1-\log x$。则 ϕ-散度构成反向 *Kullback–Leibler* 散度:

$$\mathrm{D}_\phi(p\|\pi)=\sum_{i=1}^n \pi_i\log\frac{\pi_i}{p_i}.$$

在这种情况下，我们有

$$\xi_\phi(\pi_i,\pi_j)=\inf_{(p_i,p_j)\in[0,1]^2}\Big(\frac{p_j\pi_i}{p_j}+\frac{p_i\pi_j}{p_j}\Big)=2\sqrt{\pi_i\pi_j}.$$

根据定理 *2*，在动力学 (4)下，反向 *KL* 散度以指数收敛速率趋于零:

$$\kappa=\frac{1}{1-\min_{k\in I}\pi_k}\cdot\min_{i,j\in I}\left(1-\frac{1}{2}(\sqrt{\pi_i}-\sqrt{\pi_j})^2\right).$$

例3. 令 $\alpha=0$ 和 $\phi(x)=1-x+x\log x$。 ϕ-散度构成 *KL* 散度:

$$\mathrm{D}_\phi(p\|\pi)=\mathrm{D}_{\mathrm{KL}}(p\|\pi):=\sum_{i=1}^n p_i\log\frac{p_i}{\pi_i}.$$

并且，函数 $\xi_{\mathrm{KL}}(\pi_i,\pi_j):=\xi_\phi(\pi_i,\pi_j)$ 定义了如下的一个特定对称二元函数。定义函数 $\xi\colon\ \mathbb{R}_+^2\to\mathbb{R}_+$ 为

$$\xi(s,t):=\inf_{x\in(0,1)\cup(1,\infty)}\frac{(\frac{s}{x}+t)(x-1)}{\log x}.$$

引理1。 我们有以下性质。

(i) $\xi(\pi_i,\pi_j)=\xi_{\mathrm{KL}}(\pi_i,\pi_j)$;

(ii) $2\sqrt{st}\le\xi(s,t)\le\frac{2(s-t)}{\log(s)-\log(t)}$.

此外，

$$\xi_{\mathrm{KL}}(\pi_i,\pi_j)\ge 2\sqrt{\pi_i\pi_j}.$$

根据引理*1*，我们证明动力学(4)的指数收敛结果。随后，我们推导出所提出的权重函数 ω_{ij} 的解析收敛率，该收敛率不低于[1 2]。

定理3。 设 $p(t)=(p_i(t))_{i=1}^n$, $t\ge 0$ 满足柯尔莫哥洛夫前向方程(4)。那么，

(i) *KL* 散度随时间呈指数级收敛至零:

$$\mathrm{D}_{\mathrm{KL}}(p(t)\|\pi)\le e^{-2\kappa t}\mathrm{D}_{\mathrm{KL}}(p(0)\|\pi);$$

(ii) L_1 距离随时间呈指数衰减至零:

$$\sum_{i=1}^n|p_i(t)-\pi_i|\le\sqrt{2\mathrm{D}_{\mathrm{KL}}(p(0)\|\pi)}e^{-\kappa t},$$

对于任意初始条件 $p(0) \in \mathbb{R}_+^n$，且满足 $\sum_{i=1}^n p_i(0) = 1$。全局时间收敛速率 $\kappa > 0$ 是一个常数，满足

$$\kappa \geq \frac{1}{1-\min_{k\in I}\pi_k} \cdot \min_{i,j\in I}\left(1-\frac{1}{2}(\sqrt{\pi_i}-\sqrt{\pi_j})^2\right) \geq \frac{1}{2}.$$

3.1. 证明。 本小节给出所有证明。

定理 *2* 的证明。给定一个概率分布 $\pi = (\pi_1, \ldots\ldots, \pi_n)^{\mathsf{T}}$，将$\phi$-散度（能量函数）记为 $E(p) = \mathrm{D}_\phi(p\|\pi)$。由于

$$\omega_{ij} = c\pi_i\pi_j, \qquad c = \frac{1}{1-\min_{k\in I}\pi_k},$$

我们有

$$\begin{aligned}\theta_{ij} &= c\pi_i\pi_j \frac{\frac{p_i}{\pi_i} - \frac{p_j}{\pi_j}}{\frac{\partial E(p)}{\partial p_i} - \frac{\partial E(p)}{\partial p_j}} \\ &= c\frac{\pi_j p_i - \pi_i p_j}{\phi'(\frac{p_i}{\pi_i}) - \phi'(\frac{p_j}{\pi_j})}.\end{aligned}$$

因此

$$\eta_{ij} = c(\pi_j p_i - \pi_i p_j).$$

于是，$\Gamma_2(p, f, f)$ 可以写成二次型 $\frac{1}{2}\sum_{i,j=1}^n a_{ij}(f_i - f_j)^2$，，其中

$$\begin{aligned}a_{ij} =& \frac{1}{2}\sum_{k=1}^n \Big(\frac{\partial\theta_{ij}}{\partial p_i}\eta_{ki} + \frac{\partial\eta_{ij}}{\partial p_i}\theta_{ki} + \frac{\partial\eta_{jk}}{\partial p_j}\theta_{ij} - \frac{\partial\eta_{ki}}{\partial p_k}\theta_{jk} \\ &\quad - \frac{\partial\theta_{ij}}{\partial p_j}\eta_{jk} - \frac{\partial\eta_{ij}}{\partial p_j}\theta_{jk} - \frac{\partial\eta_{ki}}{\partial p_i}\theta_{ij} + \frac{\partial\eta_{jk}}{\partial p_k}\theta_{ki}\Big) \\ =& -\frac{1}{2}\Big(\frac{\partial\theta_{ij}}{\partial p_i} - \frac{\partial\theta_{ij}}{\partial p_j}\Big)\eta_{ij} + \Big(\frac{\partial\eta_{ij}}{\partial p_i} - \frac{\partial\eta_{ij}}{\partial p_j}\Big)\theta_{ij} \\ &+ \frac{1}{2}\sum_{k\neq i,j}\Big(\frac{\partial\eta_{ij}}{\partial p_i}\theta_{ki} + \frac{\partial\eta_{jk}}{\partial p_j}\theta_{ij} - \frac{\partial\eta_{ki}}{\partial p_k}\theta_{jk} \\ &\quad - \frac{\partial\eta_{ij}}{\partial p_j}\theta_{jk} - \frac{\partial\eta_{ki}}{\partial p_i}\theta_{ij} + \frac{\partial\eta_{jk}}{\partial p_k}\theta_{ki}\Big) \\ =& -\frac{1}{2}\Big(\frac{\partial\theta_{ij}}{\partial p_i} - \frac{\partial\theta_{ij}}{\partial p_j}\Big)\eta_{ij} + \Big(\frac{\partial\eta_{ij}}{\partial p_i} - \frac{\partial\eta_{ij}}{\partial p_j}\Big)\theta_{ij} \\ &+ \frac{1}{2}\sum_{k\neq i,j}\Big(c\pi_j\theta_{ki} + c\pi_k\theta_{ij} - c\pi_i\theta_{jk} \\ &\quad + c\pi_i\theta_{jk} + c\pi_k\theta_{ij} - c\pi_j\theta_{ki}\Big) \\ =& -\frac{1}{2}\Big(\frac{\partial\theta_{ij}}{\partial p_i} - \frac{\partial\theta_{ij}}{\partial p_j}\Big)\eta_{ij} + c(\pi_j+\pi_i)\theta_{ij} + c\sum_{k\neq i,j}\pi_k\theta_{ij} \\ =& -\frac{1}{2}\Big(\frac{\partial\theta_{ij}}{\partial p_i} - \frac{\partial\theta_{ij}}{\partial p_j}\Big)\eta_{ij} + c\theta_{ij}\sum_{k=1}^n\pi_k \\ =& -\frac{c}{2}\Big(\frac{\partial\theta_{ij}}{\partial p_i} - \frac{\partial\theta_{ij}}{\partial p_j}\Big)(\pi_j p_i - \pi_i p_j) + c\theta_{ij}.\end{aligned}$$

因此，我们有

$$\frac{a_{ij}}{\theta_{ij}} = c\left(1 - \frac{1}{2}\left(\frac{\partial \log \theta_{ij}}{\partial p_i} - \frac{\partial \log \theta_{ij}}{\partial p_j}\right)(\pi_j p_i - \pi_i p_j)\right).$$

记 $\frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_j}$ 为 $\partial_{\vec{ij}}$。则

$$\log \theta_{ij} = \log(c) + \log(\pi_j p_i - \pi_i p_j) - \log(\partial_{\vec{ij}} E),$$

and

$$\partial_{\vec{ij}} \log \theta_{ij} = \frac{\pi_j + \pi_i}{\pi_j p_i - \pi_i p_j} - \frac{\partial^2_{\vec{ij}} E}{\partial_{\vec{ij}} E}.$$

我们有

$$\frac{a_{ij}}{\theta_{ij}} = c\left(1 - \frac{\pi_i + \pi_j}{2}\right) + \frac{1}{2}\theta_{ij}\partial^2_{\vec{ij}} E. \tag{10}$$

若 $E(p)$ 是凸的，则 $\theta_{ij} > 0$ 和 $\partial^2_{\vec{ij}} E \geq 0$。因此

$$\kappa \geq c\left(1 - \frac{\pi_i + \pi_j}{2}\right). \tag{11}$$

由于 $c \geq 1$ 和 $\frac{\pi_i + \pi_j}{2} \leq \frac{1}{2}$，我们有 $\kappa \geq \frac{1}{2}$。根据推论 1，我们有

$$\mathrm{D}_\phi(p(t)\|\pi) \leq e^{-2\kappa t}\mathrm{D}_\phi(p(0)\|\pi).$$

□

命题 *1* 的证明。由于 ϕ 由方程 (9) 定义，因此

$$\phi'(x) = \frac{x^{\alpha-1}}{\alpha - 1}, \qquad \phi''(x) = x^{\alpha-2}.$$

因此，二元函数 $\xi_\phi(\pi_i, \pi_j)$ 满足

$$\xi_\phi(\pi_i, \pi_j) = \inf_{(p_i,p_j)\in[0,1]^2} \left(\left(\frac{p_i}{\pi_i}\right)^{\alpha-2}\pi_j + \left(\frac{p_j}{\pi_j}\right)^{\alpha-2}\pi_i\right) \cdot \frac{\frac{p_i}{\pi_i} - \frac{p_j}{\pi_j}}{\frac{1}{\alpha-1}\left(\left(\frac{p_i}{\pi_i}\right)^{\alpha-1} - \left(\frac{p_j}{\pi_j}\right)^{\alpha-1}\right)}.$$

为证明 $\xi_\phi(\pi_i, \pi_j) \geq 2\sqrt{\pi_i \pi_j}$，我们需要证明

$$u^{\alpha-2}\pi_j + v^{\alpha-2}\pi_i \geq \frac{2\sqrt{\pi_i\pi_j}}{\alpha - 1}\frac{u^{\alpha-1} - v^{\alpha-1}}{u - v}, \tag{12}$$

其中 $u = \frac{p_i}{\pi_i}$ 和 $v = \frac{p_j}{\pi_j}$。对式（12）左端应用柯西-施瓦茨不等式，我们将证明以下不等式：

$$2\sqrt{\pi_i\pi_j}(uv)^{\frac{\alpha-2}{2}} \geq \frac{2\sqrt{\pi_i\pi_j}}{\alpha - 1}\frac{u^{\alpha-1} - v^{\alpha-1}}{u - v}. \tag{13}$$

记 $w = \left(\frac{u}{v}\right)^{\frac{1}{2}}$。在不等式（13）两边除以 $2\sqrt{\pi_i\pi_j}(uv)^{\frac{\alpha-2}{2}}$，我们需要证明

$$1 \geq \frac{1}{\alpha - 1}\frac{w^{\alpha-1} - \frac{1}{w^{\alpha-1}}}{w - \frac{1}{w}}.$$

记 $w = e^z$，且 $\beta = \alpha - 1$。由于 $\alpha \in [0, 2]$，因此 $\beta \in [-1, 1]$。于是，上述不等式 (13) 满足

$$1 \geq \frac{1}{\beta}\frac{e^{\beta z} - e^{-\beta z}}{e^z - e^{-z}} = \frac{1}{|\beta|}\frac{\sinh(|\beta|z)}{\sinh(z)}, \qquad \text{for } |\beta| \in [0, 1].$$

注意，$\sinh(a)=\frac{e^a-e^{-a}}{2}$。我们将证明

$$H(z) := |\beta|\sinh(z) - \sinh(|\beta|z) \geq 0, \quad \text{for } |\beta| \in [0,1] \text{ and } z \geq 0.$$

当 $|\beta| \leq 1$ 时，我们有

$$H'(z) = |\beta|\big(\cosh(z) - \cosh(|\beta|z)\big) \geq 0.$$

因此，$H(z)$ 是一个递增函数，且 $H(0) = 0$。于是 $H(z) \geq 0$。证明完毕。□

引理*1*的证明。如果 $\phi(x) = x\log x + 1 - x$，那么 $\phi'(x) = \log x$ 且 $\phi''(x) = \frac{1}{x}$。因此

$$\pi_i\pi_j\left(\frac{1}{p_i}+\frac{1}{p_j}\right)\frac{\frac{p_i}{\pi_i}-\frac{p_j}{\pi_j}}{\log\frac{p_i}{\pi_i}-\log\frac{p_j}{\pi_j}} = \left(\pi_j\frac{\pi_i p_j}{p_i\pi_j}+\pi_i\right)\frac{\frac{p_i\pi_j}{\pi_i p_j}-1}{\log\frac{p_i\pi_j}{\pi_i p_j}}.$$

记 $x=\frac{p_i\pi_j}{p_j\pi_i}$。接下来证明 (i)。

接下来证明 (ii)。定义如下函数:

$$F(x,s,t) = \begin{cases} \dfrac{(\frac{s}{x}+t)(x-1)}{\log x}, & \text{if } x \neq 1; \\ s+t, & \text{if } x = 1. \end{cases}$$

显然，$\xi(s,\ t) = \inf_{x\in\mathbb{R}_+} F(x,\ s,\ t)$。因此

$$\xi(s,t) \leq F\left(\frac{s}{t}, s, t\right) = \frac{2(s-t)}{\log s - \log t}.$$

接下来证明 $F(x,\ s,\ t) \geq 2\sqrt{s}t$。记

$$G(x) = \left(\frac{s}{x}+t\right)(x-1) - 2\sqrt{st}\log x.$$

注意到，当 $x \in \mathbb{R}_+$ 时，$G(x)$ 是增函数。由于

$$G'(x) = \frac{s}{x^2} + t - 2\frac{\sqrt{st}}{x} = \left(\frac{\sqrt{s}}{x} - \sqrt{t}\right)^2 \geq 0.$$

由于 $G(1) = 0$ 且 G 是增函数，我们有

$$\begin{cases} G(x) \leq 0, & \text{if } x < 1; \\ G(x) > 0, & \text{if } x > 1. \end{cases}$$

这就完成了(ii)的证明。

由 (i) 和 (ii)，我们得到

$$\xi_{\mathrm{KL}}(\pi_i,\pi_j) = \xi(\pi_i,\pi_j) \geq 2\sqrt{\pi_i\pi_j}.$$

□

定理*3*的证明。当 $\phi(x) = x\log\, x$ 且 $\mathrm{D}_\phi(p\|\pi) = \mathrm{D}_{\mathrm{KL}}(p||\pi)$ 时。由引理**1**，我们得到

$$\begin{aligned}\frac{a_{ij}}{\theta_{ij}} &= c\left(1-\frac{1}{2}(\pi_i+\pi_j)+\frac{\pi_i\pi_j}{2}\left(\frac{1}{p_i}+\frac{1}{p_j}\right)\frac{\frac{p_i}{\pi_i}-\frac{p_j}{\pi_j}}{\log\frac{p_i}{\pi_i}-\log\frac{p_j}{\pi_j}}\right)\\ &\geq c\left(1-\frac{1}{2}(\pi_i+\pi_j)+\frac{1}{2}\xi_{\mathrm{KL}}(\pi_i,\pi_j)\right)\\ &\geq c\left(1-\frac{1}{2}(\pi_i+\pi_j)+\sqrt{\pi_i\pi_j}\right)\\ &= c\left(1-\frac{1}{2}(\sqrt{\pi_i}-\sqrt{\pi_j})^2\right).\end{aligned}$$

因此

$$\kappa \geq c\left(1-\frac{1}{2}(\sqrt{\pi_i}-\sqrt{\pi_j})^2\right).$$

由推论 **1**，我们有

$$\mathrm{D}_{\mathrm{KL}}(p(t)\|\pi) \leq e^{-2\kappa t}\mathrm{D}_{\mathrm{KL}}(p(0)\|\pi).$$

由Pinsker不等式，我们得到

$$\sum_{i=1}^{n}|p_i(t)-\pi_i| \leq \sqrt{2\mathrm{D}_{\mathrm{KL}}(p(t)\|\pi)} \leq e^{-\kappa t}\sqrt{2\mathrm{D}_{\mathrm{KL}}(p(0)\|\pi)}.$$

证明至此完成。 □

注 2. 我们的证明表明，全局收敛速率取决于对函数 ξ_ϕ 的估计，该函数构成一类二元函数；引理*1*中给出了一个示例。我们仅使用了一个粗略估计，即 $\xi_{\mathrm{KL}}(s,t) \geq 2\sqrt{st}$。在未来的工作中，我们将研究一般 ϕ-散度函数中 ξ_ϕ 的性质和下界。

4. Examples

本节介绍几个 Q-矩阵（**1**）的解析示例。我们还提供数值示例，以验证动力学（**4**）的指数收敛结果。

4.1. **解析示例**。

例4（两点状态）。给定 $\pi = (\pi_1, \pi_2)^{\mathsf{T}} \in \mathbb{R}^2_+$，其中 $\pi_1+\pi_2 = 1$，记$c = \min\left\{\frac{1}{1-\pi_1}, \frac{1}{1-\pi_2}\right\} = \min\left\{\frac{1}{\pi_2}, \frac{1}{\pi_1}\right\}$。则 Q-矩阵 （**1**）满足

$$\begin{aligned}Q &= \begin{pmatrix}-c(1-\pi_1) & c\pi_2\\ c\pi_1 & -c(1-\pi_2)\end{pmatrix}\\ &= \begin{pmatrix}-\min\{1,\frac{\pi_2}{\pi_1}\} & \min\{1,\frac{\pi_2}{\pi_1}\}\\ \min\{1,\frac{\pi_1}{\pi_2}\} & -\min\{1,\frac{\pi_1}{\pi_2}\}\end{pmatrix},\end{aligned}$$

这正是 *Metropolis–Hastings*算法中的 Q-矩阵。

例 5 (三点状态**).** 给定 $\pi = (\pi_1, \pi_2, \pi_3)^{\mathsf{T}} \in \mathbb{R}^3_+$ 且 $\pi_1+\pi_2+\pi_3 = 1$，记 $c =$ 最小值 $\left\{\frac{1}{1-\pi_1}, \frac{1}{1-\pi_2}, \frac{1}{1-\pi_3}\right\}$。$Q$-矩阵 (**1**) 满足

$$Q = \begin{pmatrix}-c(1-\pi_1) & c\pi_2 & c\pi_3\\ c\pi_1 & -c(1-\pi_2) & c\pi_3\\ c\pi_1 & c\pi_2 & -c(1-\pi_3)\end{pmatrix}.$$

我们还将所提出的 Q-矩阵 (1) 与 *Metropolis-Hastings* 算法中的生成元进行比较，记为 Q^{MH}。Q^{MH}-矩阵可按下式计算：

$$\frac{1}{2}\begin{pmatrix} -\min\{1,\frac{\pi_2}{\pi_1}\}-\min\{1,\frac{\pi_3}{\pi_1}\} & \min\{1,\frac{\pi_2}{\pi_1}\} & \min\{1,\frac{\pi_3}{\pi_1}\} \\ \min\{1,\frac{\pi_1}{\pi_2}\} & -\min\{1,\frac{\pi_1}{\pi_2}\}-\min\{1,\frac{\pi_3}{\pi_2}\} & \min\{1,\frac{\pi_3}{\pi_2}\} \\ \min\{1,\frac{\pi_1}{\pi_3}\} & \min\{1,\frac{\pi_2}{\pi_3}\} & -\min\{1,\frac{\pi_1}{\pi_3}\}-\min\{1,\frac{\pi_2}{\pi_3}\} \end{pmatrix}.$$

我们注意到 $Q \neq Q^{\mathrm{MH}}$。

4.2. 数值示例。在本小节中，我们还对所提出的指数收敛结果进行了数值验证。我们随机选择 $p^0(\omega_k) \neq \pi(\omega_k) \in \mathbb{R}^n_+$，其中 $\sum_{i=1}^n p_i^0(\omega_k) = \sum_{i=1}^n \pi(\omega_k) = 1$，其中 ω_k 是一个随机采样实现，且满足 $k = 1, \cdots, K = 100$。对于每个实现 ω_k，我们计算常微分方程（4）的前向欧拉时间离散化：

$$\frac{p_i^{N+1}(\omega_k) - p_i^N(\omega_k)}{\Delta t} = \sum_{j=1}^n \Big[Q_{ji}p_j^N(\omega_k) - Q_{ij}p_i^N(\omega_k)\Big], \tag{14}$$

其中 $p^0(\omega_k) \in \mathbb{R}^n_+$ 是一个随机初始点，$\Delta t = 0.01$ 是步长。我们考虑两种矩阵选择。一种是 Q-矩阵（1），另一种是$Q^{\mathrm{MH}} = (Q_{ij}^{\mathrm{MH}})_{1\le i,\ j\le n}$，其中当 $j \neq i$ 时，$Q_{ij}^{\mathrm{MH}} = \frac{1}{n-1}\min\{1, \frac{\pi_j}{\pi_i}\}$，并且 $Q_{ii}^{\mathrm{MH}} = -\sum_{k=1,\ k\neq i}^n Q_{ik}^{\mathrm{MH}}$。我们还使用矩阵 Q 和矩阵 Q^{MH} 运行前向欧拉方法（14），其中 $N = 0, 1, \cdots, \frac{T}{\Delta t}$，以及 $T = 10$。计算完所有时间步和实现后，我们计算 p^N 与 π之间的样本平均距离 L^1：

$$\frac{1}{K}\sum_{k=1}^K \sum_{i=1}^n |p^N(\omega_k) - \pi(\omega_k)|.$$

在图1中，我们绘制了 p^N 和π 之间平均 L^1 距离随时间变量变化的收敛结果。结果表明，采用所提出的Q矩阵(1)的动力学(4)，比 Metropolis-Hastings算法中的 Q^{MH}矩阵收敛更快。

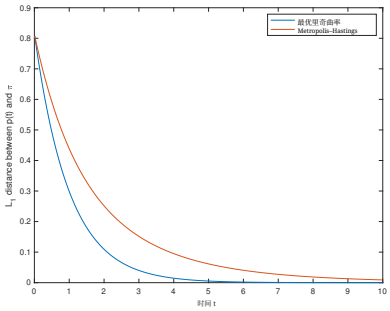
5. Discussions

我们还从 Q矩阵(1)的特征值中获得了一些见解。我们将 Q矩阵(1)改写如下：

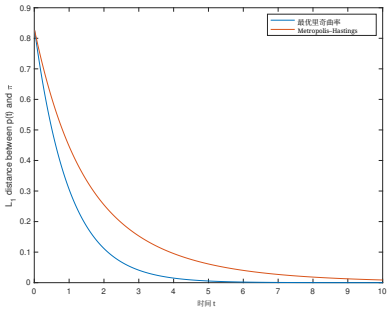
$$Q = c\Big(I - (1, \cdots, 1)(\pi_1, \cdots, \pi_n)^{\mathsf{T}}\Big),$$

其中，$I \in \mathbb{R}^{n\times n}$ 是单位矩阵，$(1, \cdots, 1)(\pi_1, \cdots, \pi_n)^{\mathsf{T}}$ 是一个秩一矩阵。因此，Q 矩阵(1)是单位矩阵的秩一修正，其大多数特征值均为 -1。此外，正如本文所示，我们证明了采用 Q矩阵(1)的主方程(4)的收敛结果，其中收敛速率保证不依赖于任何目标分布 π。

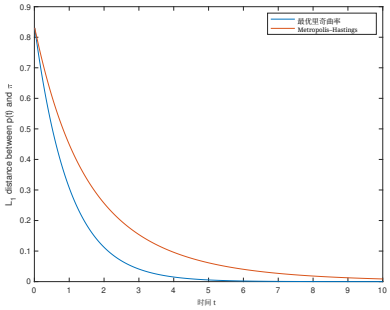
我们注意到，目前构造的权重函数是稠密的，本质上是一个完全图。这给具有极大状态空间的马尔可夫链蒙特卡罗计算带来了挑战。未来，我们将设计具有最大里奇曲率下界的稀疏图权重函数。我们也在研究离散时间马尔可夫链的收敛性分析，这同样基于平均场信息矩阵的计算 [11]。



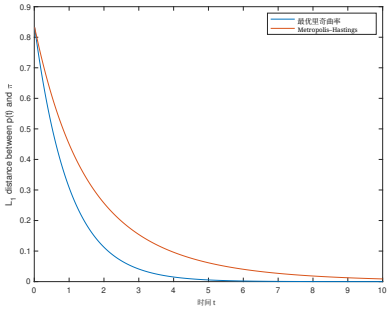
(A) $n = 250$。



(B) $n = 500$。



(C) $n = 1000$。



(D) $n = 2000$。

FIGURE 1. 上述四幅图展示了动力学 (4) 在 $n = 250$,500, 1000,2000 时的收敛结果。x 轴表示时间，y 轴表示 L_1 与 t 及 π 之间的距离。蓝色曲线表示所提出的 Q 矩阵 (1) 的收敛行为，而红色曲线展示了 Metropolis-Hastings算法的收敛行为。

REFERENCES

[1] L. Ambrosio、N. Gigli 和 G. Savar´e. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media，2006。[2] A. Cichocki 和 S.I. Amari. Alpha、Beta 和 Gamma 散度族：灵活且稳健的相似性度量。*Entropy*，12，1532-1568，2010。[3] D. Bakry 和 M. Émery. 超收缩扩散。*Seminaire de probabilites de Strasbourg* 19 177 206, 1985，：–。

[4] S.N. Chow、W. Huang、Y. Li 和 H. Zhou. 图上自由能泛函或马尔可夫过程的福克-普朗克方程。*Archive for Rational Mechanics and Analysis*，203(3):969–1008，2012。[5] M. Erbar 和 J. Maas. 通过熵的凸性研究有限马尔可夫链的里奇曲率。*Archive for Rational Mechanics and Analysis*，206(3):997–1038，2012。[6] W.K. Hastings. 使用马尔可夫链的蒙特卡洛抽样方法及其应用。*Biometrika*，第57卷，第1期，第97–109页，1970。[7] Y. Gao、W. Li 和 J.G. Liu. 具有非线性激活函数的有限状态平均场博弈主方程。*arXiv:2212.05675*，2022。[8] W. Li. 通过传输信息几何得到的海森度量。*Journal of Mathematical Physics*，62 033301 2021，。

[9] W. Li. 图上的传输信息几何：黎曼微积分。 *Information Geometry* 5 161 207, 2022 , : – .

[10] W. Li. 通过广义密度流形实现扩散超收缩性。 *arXiv:1907.12546*, 2019.[11] W. Li, and L. Lu. 图上的平均场信息海森矩阵。 *arXiv:2203.06307*, 2022.[12] J. Maas. 有限马尔可夫链的熵梯度流。 *Journal of Functional Analysis* 261 8 2250 2292, 2011 , ( ): – .[13] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. 利用高速计算机进行状态方程计算。 *Journal of Chemical Physics* 21 1087 1092, , , – . 1953

[14] A. Mielke. 反应扩散系统与能量漂移扩散的梯度结构。*Nonlinearity*, 24(4): 13-29, 2011. [15] A. Mielke. 可逆马尔可夫链中相对熵的测地凸性。 *Calculus of Variations and Partial Differential Equations* 48 1 2 1 31, 2013 , (-): – .[16] G.O. Roberts, and T.L. Richard. 朗之万分布及其离散近似的指数收敛。 *Bernoulli*, 2(4), 341-363, 1996. [17] G.O. Roberts, and A.F.M. Smith. 吉布斯采样器和 Metropolis-Hastings算法收敛的简单条件。 *Stochastic Processes and their Applications*, vol. 49, issue 2, 207-216, 1994.[18] L. Onsager. 不可逆过程中的互易关系， *I+II. Physical Review*, 37 426, 1931 405, – .[19] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften丛书第338卷。Springer，柏林，2009.[20] F. Weber, and R. Zacher. 马尔可夫链离散情形下，Bakry-´Emery思想中的曲率-维数条件下的熵方法。 *Journal of FunctionalAnalysis*, 第281卷，(5), 0022-1236，2021.*Email address*: wuchen@mailbox.sc.eduDepartment of Mathematics, University of South Carolina, Columbia, SC 29208.*Email address*: lu@math.sc.edu Department of Mathematics, University of South Carolina, Columbia, SC 29208.