J. Appl. Prob. **54,** 627–637 (2017)
*doi:*10.1017/jpr.2017.21

ON SOME MIXING TIMES FOR NONREVERSIBLE FINITE MARKOV CHAINS

LU-JING HUANG * AND
YONG-HUA MAO,* ** *Beijing Normal University*

Abstract

By adding a vorticity matrix to the reversible transition probability matrix, we show that the commute time and average hitting time are smaller than that of the original reversible one. In particular, we give an affirmative answer to a conjecture of Aldous and Fill (2002). Further quantitive properties are also studied for the nonreversible finite Markov chains.

Keywords: Markov chain; nonreversible; vorticity matrix; mixing time

2010 Mathematics Subject Classification: Primary 60J27
Secondary 60G20

1. Introduction

In many applications of Markov chains, such as the Metropolis–Hasting algorithm and Gibbs sampler, Markov chains are constructed to be reversible. But are reversible chains more efficient than nonreversible chains? Recently, it has been shown that the nonreversible chains are better in some respects. For example, the authors in [3] and [4] proved that the asymptotic variances of nonreversible chains are smaller than the corresponding reversible ones; in [6] the authors showed that a nonreversible walk converges more rapidly than the reversible walk, and in [8] and [9] it was also found that nonreversible diffusions converge more rapidly in terms of the spectral gap. As a main result of our paper, we prove that the average hitting times of nonreversible chains are also smaller than original reversible ones.

In the paper we will study the nonreversible finite Markov chains, whose probability transition matrices are obtained by adding vorticity matrices to the reversible probability transition matrices. We show that the commute times, and the average hitting times of a nonreversible chain become smaller. Our method is based on the relation of the commute time and the capacity [1], and the variational formulas of the nonreversible electrical network [7]. As an application, we give an affirmative answer to a conjecture in Aldous and Fill [1, Chapter 9].

Let us begin with some notations. Let V be a finite state space, on which $P = (P(i, j): i, j \in V)$ is the probability transition matrix of an irreducible discrete-time Markov chain $X = \{X_n : n \geq 0\}$. The chain X or P has the unique stationary distribution $\mu = \{\mu(i): i \in V\}$, i.e.

$$\sum_{i \in V} \mu(i) P(i, j) = \mu(j), \qquad j \in V.$$

Let P^* be the time-reversal chain of P, i.e.

$$P^*(i, j) = \frac{\mu(j) P(j, i)}{\mu(i)}, \qquad i, j \in V.$$

Received 17 May 2016; revision received 7 December 2016.
* Postal address: Laboratory of Mathematics and Complex Systems, Ministry of Education, School of Mathematical Sciences, Beijing Normal University, Beijing 100875, P. R. China.
** Email address: maoyh@bnu.edu.cn

Henceforth, we define A^* for a matrix A in the same way. We say that P is reversible with respect to μ if $P = P^*$. Let $\mathbb{P}_i(\cdot)$ and $\mathbb{E}_i(\cdot)$ be the probabilities and expectations for the chain X started at state i. Let

$$\tau_i = \inf\{n \geq 0\colon X_n = i\}$$

be the first hitting time to the state $i \in V$. For any pair of points i, j in V, let $T_{ij}(P) = \mathbb{E}_i\tau_j + \mathbb{E}_j\tau_i$ be the commute time. Denote by

$$T(P) = \max_{i,j} T_{ij}(P), \quad \text{and} \quad T_0(P) = \sum_i \sum_j \mu(i)\mu(j)\mathbb{E}_i\tau_j$$

the maximal commute time and the average hitting time, respectively. It is easy to see that

$$T_0(P) = \frac{1}{2}\sum_i \sum_j \mu(i)\mu(j)T_{ij}(P).$$

Now we let $K = (P+P^*)/2$, the reversible part of P, and $\Gamma = (\mathrm{diag}(\mu)P - \mathrm{diag}(\mu)P^*)/2$, where $\mathrm{diag}(\mu)$ is the diagonal matrix for vector μ. Then we can decompose P into

$$P = K + \mathrm{diag}(\mu)^{-1}\Gamma.$$

It is easy to see that the matrix Γ satisfies

$$\Gamma 1 = 0 \quad \text{and} \quad \Gamma^\top = -\Gamma, \tag{1.1}$$

where $\Gamma^\top$ is the transpose of Γ. Following [3], a matrix satisfying (1.1) is called a vorticity matrix. The above decomposition suggests adding a vorticity matrix to a reversible probability transition matrix.

Generally, assume that $K = (K(i,j)\colon i, j \in V)$ is an irreducible probability transition matrix, which is reversible with respect to a probability measure μ, i.e.

$$\mu(i)K(i,j) = \mu(j)K(j,i), \qquad i, j \in V.$$

This implies that μ is the stationary distribution of K. For every vorticity matrix Γ, we define

$$P_\Gamma = K + \mathrm{diag}(\mu)^{-1}\Gamma.$$

Then by the definition of the vorticity matrix, we have

$$P_\Gamma 1 = K1 + \mathrm{diag}(\mu)^{-1}\Gamma 1 = 1 \quad \text{and} \quad \mu P_\Gamma = \mu K + \mu\,\mathrm{diag}(\mu)^{-1}\Gamma = \mu.$$

To keep P_Γ as a probability transition matrix, we assume that Γ satisfies:

$$\Gamma(i,j) \geq -\mu(i)K(i,j), \qquad i, j \in V. \tag{1.2}$$

In our first result, we compare the mixing times of the chains K and P_Γ.

Theorem 1.1. *Let Γ be a vorticity matrix satisfying (1.2). Fix every pair of points $i \neq j$ in V, let $T_{ij}(K)$, $T_{ij}(P_\Gamma)$ respectively be the commute time between i, j of chains K and P_Γ. Then*

$$T_{ij}(P_\Gamma) \leq T_{ij}(K).$$

Consequently, the maximal commute times and the average hitting times of the chains satisfy

$$T(P_\Gamma) \leq T(K) \quad \text{and} \quad T_0(P_\Gamma) \leq T_0(K).$$

https://doi.org/10.1017/jpr.2017.21 Published online by Cambridge University Press

Remark 1.1. In general, for an irreducible finite Markov chain P with stationary distribution μ, if $T_0(P) < \infty$ then the chain has strong ergodicity: there exist $C < \infty$ and $\rho < 1$ such that

$$\sup_i \sum_j |P^n(i, j) - \mu(j)| \leq C\rho^n;$$

see, e.g. [5]. Moreover, we have $\rho < 1 - 1/T_0(P)$. Hence, by Theorem 1.1, we can see that the nonreversible Markov chains whose probability transition matrices are obtained by adding vorticity matrices to the reversible probability transition matrices have a smaller upper bound on the convergence rate than the corresponding reversible ones. Furthermore, some results on the effect of adding vorticity matrices to the reversible Markov chains in terms of asymptotic variance can be found in [3] and [4], and we will further study the effect of nonreversibility to the mixing times and asymptotic variance in a forthcoming work.

Next, we give an affirmative answer to a conjecture in Aldous and Fill [1, Chapter 9, Conjecture 22].

Let P be an irreducible probability transition matrix on V with the stationary distribution μ. For $0 \leq \lambda \leq 1$, define $P(\lambda) = (1-\lambda)P + \lambda P^*$, where P^* is the time-reversal of P. Then $P(\lambda)$ is a probability transition matrix, and $P(\frac{1}{2})$ is reversible. They all have the same stationary distribution μ. For any pair of points i, j in V, let $T_{ij}(\lambda)$ be the commute time between i, j of chain $P(\lambda)$. Similarly, write $T(\lambda)$, $T_0(\lambda)$ respectively the maximal commute time and average hitting time. Next, let

$$Z(i, j) = \sum_{n=0}^{\infty} [P^n(i, j) - \mu(j)]$$

be the fundamental matrix of P. In fact, the fundamental matrix Z can be viewed as the inverse of the operator $I - P$ on the linear space of functions $f\colon V \to \mathbb{R}$ satisfying $\mu(f) := \sum_{i \in V} \mu(i) f(i) = 0$. From [1, Sections 2 and 3], we can see that it has an intimate relation with mixing times and asymptotic variance.

In [1], Aldous and Fill conjectured that

$$\mathrm{trace}[Z^2(P^* - P)] \geq 0. \tag{1.3}$$

They also proved that (1.3) implies that

$$T_0(\lambda) \leq T_0\big(\tfrac{1}{2}\big);$$

see Corollary 24 in [1, Chapter 9]. Indeed, from their proof, it follows that when $0 \leq \lambda \leq \frac{1}{2}$,

$$\frac{\mathrm{d}T_0(\lambda)}{\mathrm{d}\lambda} = (1 - 2\lambda)^{-1} \mathrm{trace}[Z(\lambda)^2 (P^*(\lambda) - P(\lambda)], \tag{1.4}$$

where $Z(\lambda)$ is the fundamental matrix of $P(\lambda)$ and $P^*(\lambda) = (1 - \lambda)P^* + \lambda P$.

In Theorem 3.1 below, we will show that the mixing times of the nonreversible chains whose vorticity part is controlled by a parameter have monotone and symmetry properties. And as a corollary of it, we see that the left-hand side of (1.4) is indeed nonnegative for $0 \leq \lambda \leq \frac{1}{2}$.

Corollary 1.1. *Assume P is an irreducible probability transition matrix on V with the stationary distribution μ. Let P^* be the time-reversal of P and Z be the fundamental matrix of P. Then (1.3) holds.*

https://doi.org/10.1017/jpr.2017.21 Published online by Cambridge University Press

Remark 1.2. For the circulant transition matrix, the authors in [2] proved a more general result: this inequality holds for the average hitting time with any power order.

In the final part of our paper, Section 4, we introduce the decomposition of the vorticity matrix. By the decomposition, we generalize the result of Theorem 3.1 to the case of multiple parameters.

2. Proof of Theorem 1.1

Recall that P is the irreducible probability transition matrix of chain X with stationary distribution μ. Define the scalar product $\langle f, g\rangle = \sum_{k\in V}\mu(k)f(k)g(k)$ for $f, g\colon V \to \mathbb{R}$. For two different states $i, j \in V$, the capacity of chain P between i and j is defined as

$$C_{ij}(P) = \mu(i)\mathbb{P}_i(\tau_j < \tau_i^+), \tag{2.1}$$

where $\tau_i^+ = \inf\{n \geq 1\colon X_n = i\}$ is the first return time. The following lemma gives us the relation between the capacity and the commute time.

Lemma 2.1. *For $i \neq j$ in V,*

$$T_{ij}(P) = C_{ij}(P)^{-1}. \tag{2.2}$$

Proof. From [1, Chapter 2, Corollary 8], we have

$$\mathbb{P}_i(\tau_j < \tau_i^+) = \frac{1}{\mu(i)T_{ij}(P)}.$$

This yields (2.2) by definition (2.1). □

Next, for any $i \neq j \in V$, let $U_{ij}\colon V \to \mathbb{R}$ be the equilibrium potential of chain X defined by

$$U_{ij}(k) = \mathbb{P}_k(\tau_i < \tau_j), \qquad k \in V. \tag{2.3}$$

In fact, U_{ij} is the unique solution of the harmonic equation

$$[(I-P)U](k) = 0, \quad k \neq i, j, \qquad U(i) = 1, \qquad U(j) = 0.$$

The following variational formula of capacity is from [7, Lemma 3.1], we translate it into the discrete-time case.

Lemma 2.2. *Let P be an irreducible probability transition matrix. For every pair of points $i \neq j$ in V,*

$$C_{ij}(P) = \inf\{\langle f, (I-P)(I-K)^{-1}(I-P)^* f\rangle\colon f(i) = 1, f(j) = 0\},$$

where $K = (P + P^*)/2$ is the reversible part of P. Moreover, the infimum is attained by $f_{ij} = (U_{ij} + U_{ij}^*)/2$, where U_{ij} and U_{ij}^* are the harmonic functions defined in (2.3) of chains P and P^*, respectively.

Proof. Let $L = P - I$ and $S = K - I$ as in [7, Lemma 3.1]; the result follows. □

Remark 2.1. (i) When P is reversible, i.e. $P = K$, we have the classical form of the variational formula:

$$C_{ij}(P) = \inf\{\langle f, (I-P)f\rangle\colon f(i) = 1, f(j) = 0\}.$$

(ii) On the subspace $\{f : \mu(f) = 0\}$, $(I - K)^{-1}$ is well defined and understood as: for f with $\mu(f) = 0$,

$$(I - K)^{-1} f = (I - K - \mu)^{-1} f = \sum_{n=0}^{\infty} (K^n f - \mu(f)) = \sum_{n=0}^{\infty} K^n f.$$

Now we are ready to prove Theorem 1.1.

Proof of Theorem 1.1. Fix $i \neq j$. Let $C_{ij}(P_\Gamma)$ and $C_{ij}(K)$ be the capacities for P_Γ and K, respectively. By Lemma 2.1, we need to prove that $C_{ij}(P_\Gamma) \geq C_{ij}(K)$. For this, note that, by definition, $I - P_\Gamma = I - K - \mathrm{diag}(\mu)^{-1}\Gamma$ and $(\mathrm{diag}(\mu)^{-1}\Gamma)^* = -\mathrm{diag}(\mu)^{-1}\Gamma$. We have

$$\begin{aligned}
&(I - P_\Gamma)(I - K)^{-1}(I - P_\Gamma)^* \\
&\quad = I - K - \mathrm{diag}(\mu)^{-1}\Gamma - (\mathrm{diag}(\mu)^{-1}\Gamma)^* + (\mathrm{diag}(\mu)^{-1}\Gamma)(I - K)^{-1}(\mathrm{diag}(\mu)^{-1}\Gamma)^* \\
&\quad = I - K + (\mathrm{diag}(\mu)^{-1}\Gamma)(I - K)^{-1}(\mathrm{diag}(\mu)^{-1}\Gamma)^*.
\end{aligned}$$

Then, for any $f : V \to \mathbb{R}$, we have

$$\mu((\mathrm{diag}(\mu)^{-1}\Gamma)^* f) = -\sum_{k,\ell} \Gamma_{k\ell} f_\ell = \sum_{k,\ell} \Gamma_{\ell k} f_\ell = 0,$$

so that

$$\begin{aligned}
&\langle f, (I - P_\Gamma)(I - K)^{-1}(I - P_\Gamma)^* f \rangle \\
&\quad = \langle f, (I - K) f \rangle + \langle (\mathrm{diag}(\mu)^{-1}\Gamma)^* f, (I - K)^{-1}(\mathrm{diag}(\mu)^{-1}\Gamma)^* f \rangle \\
&\quad \geq \langle f, (I - K) f \rangle.
\end{aligned}$$

Thus, $C_{ij}(P_\Gamma) \geq C_{ij}(K)$ by Lemma 2.2. □

3. The monotone and symmetry properties of mixing times

In this section we introduce a parameter to control the vorticity matrix and show that the mixing times are viewed as functions of this parameter.

Let K be an irreducible reversible probability transition matrix with respect to μ. Assume that there exists a nonzero vorticity matrix Γ satisfying (1.2). Define

$$P(\lambda) = K + \lambda \mathrm{diag}(\mu)^{-1}\Gamma, \qquad -1 \leq \lambda \leq 1, \tag{3.1}$$

to be a family of probability transition matrices, all having the same stationary distribution μ.

For any pair of points i, j in V, let $T_{ij}(\lambda)$ be the commute time between i, j of chain $P(\lambda)$. Similarly, let $T(\lambda)$, $T_0(\lambda)$ be the maximal commute time and average hitting time, respectively. The following result tells us that as functions of the variable λ, the mixing times have the monotone and symmetry properties.

Theorem 3.1. *For the reversible chain K, let $P(\lambda)$ be defined by (3.1) with nonzero vorticity matrix Γ satisfying (1.2). For any pair of points $i \neq j$ in V, denote by $S(\lambda)$ any one of $T_{ij}(\lambda)$, $T(\lambda)$, and $T_0(\lambda)$. Then,*

(i) *for every $\lambda \in [-1, 1]$, $S(\lambda) = S(-\lambda)$;*

https://doi.org/10.1017/jpr.2017.21 Published online by Cambridge University Press

(ii) *$S(\lambda)$ is nondecreasing on* $[-1,0]$. *In particular,*

$$\max_{-1\le\lambda\le 1} S(\lambda)=S(0) \quad \textit{and} \quad \min_{-1\le\lambda\le 1} S(\lambda)=S(1)=S(-1).$$

If, furthermore, the matrix Γ has a row with only two nonzero elements, then $T_0(\lambda)$ is strictly increasing.

Proof. (i) Let $C(\lambda)$ be the capacity of $P(\lambda)$. For any pair of points $i \ne j$ in V, and $f\colon V\to\mathbb{R}$ with $f(i)=1$, $f(j)=0$, from the proof of Theorem 1.1, we have

$$\begin{aligned}&\langle f,(I-P(\lambda))(I-K)^{-1}(I-P(\lambda))^*f\rangle\\ &\quad=\langle f,(I-K)f\rangle+\lambda^2\langle(\operatorname{diag}(\mu)^{-1}\Gamma)^*f,(I-K)^{-1}(\operatorname{diag}(\mu)^{-1}\Gamma)^*f\rangle.\end{aligned}$$

By Lemmas 2.1 and 2.2, it is obvious that $S(\lambda)$ is symmetric in $[-1,1]$ and increasing in $[-1,0]$.

(ii) Suppose that the vorticity matrix Γ has a row that only has two nonzero elements, i.e. there exist i_0,i_1,i_2 in V such that $a:=\Gamma(i_0,i_1)>0$, $\Gamma(i_0,i_2)=-a<0$, and $\Gamma(i_0,j)=0$ for any $j\ne i_1,i_2$ in V. For every $\lambda\in[-1,0]$, let $f^\lambda_{i_1i_2}=(U^\lambda_{i_1i_2}+U^{\lambda*}_{i_1i_2})/2$, where $U^\lambda_{i_1i_2}$, $U^{\lambda*}_{i_1i_2}$ are the harmonic functions defined in (2.3) of chains $P(\lambda)$ and $P^*(\lambda)$, respectively. Then $f^\lambda_{i_1i_2}(i_1)=1$ and $f^\lambda_{i_1i_2}(i_2)=0$. By the assumption, we have

$$\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2}(i_0)=\frac{a}{\mu(i_0)}>0; \tag{3.2}$$

thus, $\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2}\ne 0$. Now we claim that $\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2}$ is not a constant vector. Indeed, if there exists a constant α such that $\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2}=\alpha\mathbf{1}$, then

$$\Gamma(\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2})=\alpha\Gamma\mathbf{1}=0.$$

Hence,

$$(f^\lambda_{i_1i_2})^\top\Gamma\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2}=0.$$

Since $\Gamma^\top=-\Gamma$, we obtain

$$(f^\lambda_{i_1i_2})^\top\Gamma^\top\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2}=(\operatorname{diag}(\mu)^{-1/2}\Gamma f^\lambda_{i_1i_2})^\top(\operatorname{diag}(\mu)^{-1/2}\Gamma f^\lambda_{i_1i_2})=0.$$

So $\operatorname{diag}(\mu)^{-1/2}\Gamma f^\lambda_{i_1i_2}=0$, i.e. $\operatorname{diag}(\mu)^{-1}\Gamma f^\lambda_{i_1i_2}=0$, which contradicts (3.2). By the above analysis, we know that

$$\langle f^\lambda_{i_1i_2},(\operatorname{diag}(\mu)^{-1}\Gamma)(I-K)^{-1}(\operatorname{diag}(\mu)^{-1}\Gamma)^*f^\lambda_{i_1i_2}\rangle>0,\qquad \lambda\in[-1,0]. \tag{3.3}$$

Now, for $-1\le\lambda_1<\lambda_2\le 0$, by Lemma 2.2 and (3.3), we have

$$\begin{aligned}C_{i_1i_2}(\lambda_1)&=\langle f^{\lambda_1}_{i_1i_2},[I-K+\lambda_1^2(\operatorname{diag}(\mu)^{-1}\Gamma)(I-K)^{-1}(\operatorname{diag}(\mu)^{-1}\Gamma)^*]f^{\lambda_1}_{i_1i_2}\rangle\\ &>\langle f^{\lambda_1}_{i_1i_2},[I-K+\lambda_2^2(\operatorname{diag}(\mu)^{-1}\Gamma)(I-K)^{-1}(\operatorname{diag}(\mu)^{-1}\Gamma)^*]f^{\lambda_1}_{i_1i_2}\rangle\\ &\ge C_{i_1i_2}(\lambda_2).\end{aligned}$$

It follows from Lemma 2.1 that $T_{i_1i_2}(\lambda_1)<T_{i_1i_2}(\lambda_2)$. Hence, $T_0(\lambda_1)<T_0(\lambda_2)$. □

https://doi.org/10.1017/jpr.2017.21 Published online by Cambridge University Press

As an application, we return to consider the case that probability transition matrices $P(\lambda) = \lambda P + (1-\lambda)P^*(0 \le \lambda \le 1)$ from Section 1. Indeed, we have the following results which yield a proof of the conjecture of Aldous and Fill.

Corollary 3.1. *For any pair of points $i \neq j$ in V and every $\lambda \in [0,1]$, let $S(\lambda)$ be any one of $T_{ij}(\lambda)$, $T(\lambda)$, and $T_0(\lambda)$ of $P(\lambda)$. Then*

(i) *for every $\lambda \in [0,1]$, $S(\lambda) = S(1-\lambda)$;*

(ii) *$S(\lambda)$ is nondecreasing on $[0, \frac{1}{2}]$. In particular,*

$$\max_{0\le\lambda\le1} S(\lambda) = S\big(\tfrac{1}{2}\big) \quad \textit{and} \quad \min_{0\le\lambda\le1} S(\lambda) = S(0) = S(1).$$

Proof. Let $K = (P+P^*)/2 = P_{1/2}$ and $\Gamma = \mathrm{diag}(\mu)P - \mathrm{diag}(\mu)P^*$. Then Γ is a vorticity matrix and $P(\lambda) = K + (\lambda - \frac{1}{2})\mathrm{diag}(\mu)^{-1}\Gamma\ (0 \le \lambda \le 1)$. From Theorem 3.1, we obtain the results easily. □

4. Decomposition of the vorticity matrix

To introduce multiple parameters to control the vorticity matrix, we consider the decomposition of the vorticity matrix in this section.

For the irreducible reversible chain K, let $G = (V, E)$ be the graph associated to K, where V is the state space and $E = \{(i,j) \in V \times V : i \neq j, K(i,j) > 0\}$ is the set of edges. Note that we distinguish edges (i,j) and (j,i). For $i_0, i_1, \ldots, i_{n-1}, i_n (n \ge 3)$ in V, if $(i_k, i_{k+1}) \in E$, $k = 0, 1, \ldots, n-1, i_0 = i_n$, then $c := (i_0, i_1, \ldots, i_{n-1}, i_n)$ is called a cycle on G, (i_k, i_{k+1}) and $(i_{k+1}, i_k)(k = 0, 1, \ldots, n-1)$ are called the edges of c, and $i_k (k = 0, 1, \ldots, n-1)$ are the vertices of c. We define the unit vorticity matrix $\Gamma^{(c)} = (\Gamma^{(c)}(i,j) : i, j \in V)$ associated with the cycle $c = (i_0, i_1, \ldots, i_{n-1}, i_n)$ as

$$\Gamma^{(c)}(i,j) = \begin{cases} 1, & i = i_k,\ j = i_{k+1},\ k = 0, 1, \ldots, n-1, \\ -1, & i = i_{k+1},\ j = i_k,\ k = 0, 1, \ldots, n-1, \\ 0 & \text{otherwise.} \end{cases}$$

We remark that the definition of the unit vorticity matrix is independent of the choice of starting point in a cycle. We have the following decomposition of vorticity matrices.

Proposition 4.1. *Assume that Γ is a vorticity matrix such that $P_\Gamma = K + \mathrm{diag}(\mu)^{-1}\Gamma$ is a transition matrix. Then there exist cycles $c_1, c_2, \ldots, c_m$ on G and positive $\lambda_1, \lambda_2, \ldots, \lambda_m$ $(m \ge 1)$ such that*

$$\Gamma = \lambda_1\Gamma^{(c_1)} + \lambda_2\Gamma^{(c_2)} + \cdots + \lambda_m\Gamma^{(c_m)}.$$

Furthermore, $\Gamma^{(c_1)}, \Gamma^{(c_2)}, \ldots, \Gamma^{(c_m)}$ can be chosen to be linearly independent in the sense that if there exists $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$ such that

$$\alpha_1\Gamma^{(c_1)} + \alpha_2\Gamma^{(c_2)} + \cdots + \alpha_m\Gamma^{(c_m)} = 0,$$

then $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$.

Proof. We need only consider the $\Gamma \neq 0$ case.

(i) Since $\Gamma(\neq 0)$ is a vorticity matrix, there exists a pair (i_0, i_1) such that $\Gamma(i_0, i_1) > 0$, so that $\Gamma(i_1, i_0) < 0$ by definition. Because P_Γ is a transition matrix, we have $K(i_1, i_0) > 0$,

https://doi.org/10.1017/jpr.2017.21 Published online by Cambridge University Press

i.e. $(i_0, i_1) \in E$. As every row of Γ sums to 0, there must exist $i_2 \neq i_0$ such that $\Gamma(i_1, i_2) > 0$ and $(i_1, i_2) \in E$. Because V is finite, we can repeat this procedure until we encounter a vertex i_k that we already obtained, i.e. there exist $i_l = i_k$ for some $0 < l < k-1$ such that $\Gamma(i_r, i_{r+1}) > 0, l \leq r \leq k-1$, and $c_1 := (i_l, i_{l+1}, \dots, i_{k-1}, i_k)$ is a cycle on G. Define

$$\lambda_1 = \min\{\Gamma(i_l, i_{l+1}), \Gamma(i_{l+1}, i_{l+2}), \dots, \Gamma(i_{k-1}, i_k)\} > 0 \quad \text{and} \quad \Gamma_1 = \Gamma - \lambda_1 \Gamma^{(c_1)},$$

where $\Gamma^{(c_1)}$ is the unit vorticity matrix associated with c_1. Then $\Gamma = \Gamma_1 + \lambda_1 \Gamma^{(c_1)}$ and Γ_1 is also a vorticity matrix. If $\Gamma_1 = 0$ then the proof of the theorem is completed. Otherwise, we can repeat the above procedure. Note that in each step, at least one positive entry in Γ will be deleted (reset to be null), so the procedure will stop after finite steps. Hence, we can obtain a decomposition of Γ, i.e. there exist cycles $c_1, c_2, \dots, c_m$ and positive numbers $\lambda_1, \lambda_2, \dots, \lambda_m$ such that

$$\Gamma = \lambda_1 \Gamma^{(c_1)} + \lambda_2 \Gamma^{(c_2)} + \cdots + \lambda_m \Gamma^{(c_m)}.$$

(ii) Now assume that there exist $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{R}$ such that

$$\alpha_1 \Gamma^{(c_1)} + \alpha_2 \Gamma^{(c_2)} + \cdots + \alpha_m \Gamma^{(c_m)} = 0.$$

By the definition of λ_1, there exists an edge (j_1, j_2) of c_1 such that

$$\Gamma(j_1, j_2) = \lambda_1 \Gamma^{(c_1)}(j_1, j_2) \neq 0, \qquad \Gamma^{(c_2)}(j_1, j_2) = \cdots = \Gamma^{(c_m)}(j_1, j_2) = 0.$$

Hence, $\alpha_1 = 0$. Similarly, we can inductively prove that $\alpha_2 = \cdots = \alpha_m = 0$. □

As a corollary of Proposition 4.1 and [3, Proposition 4.3], we obtain the relation of the vorticity matrices that satisfies (1.2) and cycles.

Corollary 4.1. *For an irreducible reversible Markov chain K with stationary distribution μ, there exists a vorticity matrix $\Gamma \neq 0$ such that $P_\Gamma = K + \operatorname{diag}(\mu)^{-1}\Gamma$ is a probability transition matrix if and only if there exists at least one cycle in its graph G.*

5. The case of multiple parameters

The decomposition of the vorticity matrix tells us that the unit vorticity matrices are fundamental. We will introduce parameters corresponding to the unit vorticity matrices in the decomposition and the mixing times are viewed as the multivariate functions.

Assume that the graph G of the chain K has exactly m cycles $c_r = (i_{r0}, i_{r1}, \dots, i_{rk_r})$ $(r = 1, 2, \dots, m)$ which do not have any edge in common with each other. For every $1 \leq r \leq m$, let $\Gamma^{(c_r)}$ be the unit vorticity matrix associated with c_r, and

$$\widehat{\lambda}_r = \min_{0 \leq s \leq k_r - 1} \{\mu(i_{rs}) K(i_{rs}, i_{r(s+1)})\}.$$

Let $\widehat{\lambda} = (\widehat{\lambda}_1, \dots, \widehat{\lambda}_m)$ and

$$P(\lambda) = K + \sum_{r=1}^{m} \lambda_r \operatorname{diag}(\mu)^{-1} \Gamma^{(c_r)}, \qquad |\lambda_r| \leq \widehat{\lambda}_r \ (1 \leq r \leq m),$$

where $\lambda = (\lambda_1, \dots, \lambda_m)$. Then, for any λ with $|\lambda_r| \leq \widehat{\lambda}_r$ $(r = 1, 2, \dots, m)$, $P(\lambda)$ is a probability transition matrix and has the same stationary distribution μ. For any i, j in V, let $T_{ij}(\lambda)$ be the commute time between i and j of $P(\lambda)$. Similarly, $T(\lambda)$ and $T_0(\lambda)$ are the maximal commute time and the average hitting time, respectively. We have the following monotone and symmetry properties.

https://doi.org/10.1017/jpr.2017.21 Published online by Cambridge University Press

Theorem 5.1. *Assume that the graph G has exactly m cycles $c_r = (i_{r0}, i_{r1}, \dots, i_{rk_r} = i_{r0})$ ($r = 1, 2, \dots, m$), which have no common edges with each other. For any pair of points $i \neq j$ in V, denote by $S(\lambda)$ any one of $T_{ij}(\lambda)$, $T(\lambda)$, and $T_0(\lambda)$. Then,*

(i) *for every λ with $|\lambda_r| \leq \widehat{\lambda_r}(r = 1, 2, \dots, m)$, we have $S(\lambda) = S(|\lambda|)$, where $|\lambda| = (|\lambda_1|, \dots, |\lambda_m|)$,*

(ii) *$T_0(\lambda)$ strictly increases in $[-\widehat{\lambda}, 0]$. In particular,*

$$\max_{|\lambda_r| \leq \widehat{\lambda}_r, \forall r} T_0(\lambda) = T_0(0), \qquad \min_{|\lambda_r| \leq \widehat{\lambda}_r, \forall r} T_0(\lambda) = T_0(\widehat{\lambda}).$$

To prove Theorem 5.1, we need a new variational formula in [7] which gives us the relation between the commute time and flows. For an irreducible Markov chain P on V with stationary distribution μ, let $K = (P + P^*)/2$ and $\Gamma = (\mathrm{diag}(\mu)P - \mathrm{diag}(\mu)P^*)/2$. Recall the graph $G = (V, E)$ of K. A flow on G is by definition an antisymmetric function $\varphi \colon E \to \mathbb{R}$, i.e. $\varphi(i, j) = -\varphi(j, i)$ for any $(i, j) \in E$. Denote by $\mathcal{F}$ the set of flows endowed with the scalar product

$$\langle\langle \varphi, \psi \rangle\rangle = \frac{1}{2} \sum_{(i,j)\in E} \frac{1}{\mu(i)K(i, j)} \varphi(i, j)\psi(i, j),$$

and let $\| \cdot \|$ be the norm associated with this scalar product.

For a function $f \colon V \to \mathbb{R}$, let $\Psi_f(i, j) = \mu(i)K(i, j)[f(i) - f(j)]$ be the gradient flow associated with f. Write $\mathcal{G} = \{\Psi_f | f \colon V \to \mathbb{R}\}$. It is easy to check that $\| \Psi_f \|^2 = \langle (I - P)f, f \rangle$. For a cycle $c = (i_0, i_1, \dots, i_{k-1}, i_k = i_0)$ in G, define the flow $\mathcal{X}_c$ associated with cycle c as

$$\mathcal{X}_c(e) = \sum_{r=0}^{k-1} \{\delta_{(i_r, i_{r+1})} - \delta_{(i_{r+1}, i_r)}\}(e), \qquad e \in E,$$

where δ is the Kronecker delta. Denote by $\mathcal{C}$ the subspace of $\mathcal{F}$ spanned by flows associated with cycles. In [7], it was shown that on space $(\mathcal{F}, \langle\langle \cdot, \cdot \rangle\rangle)$,

$$\mathcal{F} = \mathcal{G} \oplus \mathcal{C}, \qquad \mathcal{G} \perp \mathcal{C}. \tag{5.1}$$

For any $f \colon V \to \mathbb{R}$, we also denote the flow Υ_f as

$$\Upsilon_f(i, j) = \Gamma(i, j)[f(i) + f(j)], \qquad (i, j) \in E. \tag{5.2}$$

The following variational formula is [7, Lemma 4.4].

Lemma 5.1. *Assume that P is an irreducible Markov chain with stationary distribution μ. For a pair of points $i \neq j$ in V, the capacity of P between i and j satisfies*

$$C_{ij}(P) = \inf\{\langle f, (I - P)f \rangle + \inf_{\varphi \in \mathcal{C}} \| \Upsilon_f - \varphi \|^2 \colon f(i) = 1, f(j) = 0\}.$$

Remark 5.1. In fact, for any $f \colon V \to \mathbb{R}$ with $f(i) = 1$ and $f(j) = 0$, $\inf_{\varphi \in \mathcal{C}} \| \Upsilon_f - \varphi \|$ is the length of flow that is the projection of the flow Υ_f on the space of gradient flows $\mathcal{G}$.

Proof of Theorem 5.1. We prove only the case of $m = 2$, the proof of the case of $m \geq 3$ is essentially the same.

(i) Fix a pair of points $i \neq j$ in V. For any $f\colon V \to \mathbb{R}$ with $f(i) = 1$ and $f(j) = 0$, by (5.2), we have

$$\begin{aligned}\Upsilon_f(k,l) &= \lambda_1 \Gamma^{(c_1)}(k,l)[f(k)+f(l)] + \lambda_2 \Gamma^{(c_2)}(k,l)[f(k)+f(l)]\\ &=: \lambda_1 \Upsilon_{f1}(k,l) + \lambda_2 \Upsilon_{f2}(k,l), \qquad (k,l) \in E.\end{aligned}$$

Then Υ_{f1} and Υ_{f2} are flows. By (5.1), there exist functions $g_1, g_2\colon V \to \mathbb{R}$ and flows $\Delta_{f1}, \Delta_{f2} \in \mathfrak{C}$ such that

$$\Upsilon_{f1} = \Psi_{g_1} + \Delta_{f1}, \qquad \Upsilon_{f2} = \Psi_{g_2} + \Delta_{f2}.$$

According to Remark 5.1, we have

$$\begin{aligned}\inf_{\varphi\in\mathfrak{C}} \| \Upsilon_f - \varphi \|^2 &= \langle\langle \lambda_1\Psi_{g_1} + \lambda_2\Psi_{g_2}, \lambda_1\Psi_{g_1} + \lambda_2\Psi_{g_2}\rangle\rangle\\ &= \lambda_1^2\langle\langle\Psi_{g_1},\Psi_{g_1}\rangle\rangle + \lambda_2^2\langle\langle\Psi_{g_2},\Psi_{g_2}\rangle\rangle + 2\lambda_1\lambda_2\langle\langle\Psi_{g_1},\Psi_{g_2}\rangle\rangle. \qquad (5.3)\end{aligned}$$

(ii) Recall that for a flow φ, the support of φ is $\{(k,l)\colon \varphi(k,l) > 0\}$. We claim that the supports of Ψ_{g_1} and Ψ_{g_2} are separated, so that, by (5.3),

$$\inf_{\varphi\in\mathfrak{C}} \| \Upsilon_f - \varphi \|^2 = \lambda_1^2\langle\langle\Psi_{g_1},\Psi_{g_1}\rangle\rangle + \lambda_2^2\langle\langle\Psi_{g_2},\Psi_{g_2}\rangle\rangle.$$

Then by Lemma 5.1, Lemma 2.1, and the proof of Theorem 3.1, we complete the proof. □

Now we prove that the supports of Ψ_{g_1} and Ψ_{g_2} are separated. By the definition of flow Υ_{f1}, the support of Υ_{f1} is included in the edge set of cycle c_1.

We claim that the support of Ψ_{g_1} is also included in the edge set of cycle c_1. In fact, let $(k,l) \in E$ be outside of the cycle c_1. If (k,l) is also outside of the cycle c_2, by the definition of Δ_{f1}, we have $\Delta_{f1}(k,l) = 0$. Since $\Upsilon_{f1}(k,l) = 0$, we have

$$\Psi_{g_1}(k,l) = \Upsilon_{f1}(k,l) - \Delta_{f1}(k,l) = 0.$$

Otherwise, assume that (k,l) is on the cycle c_2. Since cycles c_1 and c_2 have no common edges and G have no more cycles, there exists a vertex v such that all paths connecting c_1 and c_2 must go through v. So v separates G into two subgraphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, such that c_1 is on G_1 and c_2 is on G_2. Furthermore, $V_1 \bigcap V_2 = \{v\}$ and $\Upsilon_{f1} \mid_{E_2}$, $\Psi_{g_1} \mid_{E_2}$, $\Delta_{f1} \mid_{E_2}$ are flows on G_2. In particular, $\Psi_{g_1} \mid_{E_2}$ is a gradient flow on G_2 and $\Delta_{f1} \mid_{E_2}$ is on the flow space that is spanned by cycle flows on G_2. By (5.1), we have $\Psi_{g_1} \mid_{E_2} \perp \Delta_{f1} \mid_{E_2}$. Then, by the definition of Υ_{f1},

$$\Upsilon_{f1} \mid_{E_2} = \Psi_{g_1} \mid_{E_2} + \Delta_{f1} \mid_{E_2} = 0,$$

so $\Psi_{g_1} \mid_{E_2} = \Delta_{f1} \mid_{E_2} = 0$. In particular, $\Psi_{g_1}(k,l) = 0$. We have thus proved that the support of Ψ_{g_1} is included in the edge set of cycle c_1.

By a similar argument, we can prove that the support of Ψ_{g_2} is included in the edge set of cycle c_2. Thus, the supports of Ψ_{g_1} and Ψ_{g_2} are separated.

Acknowledgements

The authors would like to thank the anonymous referees for their helpful comments and suggestions on an earlier version of the paper. Research supported in part by the NSFC (grant numbers 11131003, 11571043, and 11626245).

References

[1] Aldous, D. J. and Fill, J. A. (2002). *Reversible Markov Chains and Random Walks on Graphs*. Available at https://www.stat.berkeley.edu/~aldous/RWG/book.html.
[2] Avrachenkov, K., Cottatellucci, L., Maggi, L. and Mao, Y.-H. (2013). Maximum entropy mixing time of circulant Markov processes. *Statist. Prob. Lett.* **83,** 768–773.
[3] Bierkens, J. (2016). Non-reversible Metropolis-Hastings. *Statist. Comput.* **26,** 1213–1228.
[4] Chen, T.-L. and Hwang, C.-R. (2013). Accelerating reversible Markov chains. *Statist. Prob. Lett.* **83,** 1956–1962.
[5] Cui, H. and Mao, Y.-H. (2010). Eigentime identity for asymmetric finite Markov chains. *Front. Math. China* **5,** 623–634.
[6] Diaconis, P., Holmes, S. and Neal, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Prob.* **10,** 726–752.
[7] Gaudillière, A. and Landim, C. (2014). A Dirichlet principle for non reversible Markov chains and some recurrence theorems. *Prob. Theory Relat. Fields* **158,** 55–89.
[8] Hwang, C.-R., Hwang-Ma, S.-Y. and Sheu, S. J. (1993). Accelerating Gaussian diffusions. *Ann. Appl. Prob.* **3,** 897–913.
[9] Hwang, C.-R., Hwang-Ma, S.-Y. and Sheu, S.-J. (2005). Accelerating diffusions. *Ann. Appl. Prob.* **15,** 1433–1444.