Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 117 (2007) 961–979
stochastic processes and their applications
www.elsevier.com/locate/spa

Convergence of some time inhomogeneous Markov chains via spectral techniques

L. Saloff-Coste, J. Zúñiga*

Department of Mathematics, Cornell University, 310 Malott Hall, 14853 Ithaca, NY, United States

Received 2 September 2006; received in revised form 10 November 2006; accepted 18 November 2006
Available online 11 December 2006

Abstract

We consider the problem of giving explicit spectral bounds for time inhomogeneous Markov chains on a finite state space. We give bounds that apply when there exists a probability π such that each of the different steps corresponds to a nice ergodic Markov kernel with stationary measure π. For instance, our results provide sharp bounds for models such as semi-random transpositions and semi-random insertions (in these cases π is the uniform probability on the symmetric group).
© 2006 Elsevier B.V. All rights reserved.

Keywords: Time inhomogeneous Markov chains; Singular values; Spectral techniques

1. Introduction

Spectral theory is one of the basic quantitative techniques for studying time homogeneous ergodic finite Markov chains. See, e.g., [2,5,7,9,22]. This paper shows how spectral theory can be used to study the convergence of time inhomogeneous finite Markov chains under the strong assumption that there is a (positive) probability measure π which is invariant for each individual step.

One of the first relevant references concerning such Markov chains is a note of Emile Borel [4] where a Doeblin type criterion is derived for time inhomogeneous card shuffling models. Ergodicity for time inhomogeneous finite Markov chains in general is discussed in [17, 21,25] where further references can be found. However, there seems to be very little in the

* Corresponding author. Tel.: +1 607 2559351; fax: +1 607 2557149.
E-mail address: zuniga@math.cornell.edu (J. Zúñiga).

0304-4149/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2006.11.004

literature concerning explicit convergence bounds for specific models of time inhomogeneous chains in the spirit of the work of Aldous, Diaconis, and their collaborators (e.g., [2,7]) for time homogeneous chains. One such result is found in [19,20] where Mironov and Mossel et al. study a time inhomogeneous shuffling process defined as follows. Consider a deck of n cards. For any infinite sequence $r = (r_k)_1^\infty$ with $r_k \in \{1, \ldots, n\}$, the r-semi-random transposition chain evolves as follows: At time k, transpose the card at position r_k with the card at a uniformly chosen random position. This model was considered earlier in [1,3]. Mironov uses the cyclic-to-random shuffle to study the attacks on the RC4 stream cipher. Using a strong stationary time argument, due to Broder, he shows that the cyclic-to-random shuffle mixes the deck in order $n \log n$ shuffles. Mossel et al. generalize the results of Mironov to show that for any r, order $n \log n$ semi-random transpositions suffice to mix up the deck. Their main result is to prove that order $n \log n$ cyclic-to-random transpositions are necessary, an improvement to the lower bound of order n given by Mironov (see [20]).

The present work is motivated in part by the following question. What can be said if semi-random transpositions are replaced by other similar models, for instance semi-random insertions? More precisely, for each sequence r as above, the r-semi-random insertion chain is the time inhomogeneous Markov chain which, at time k, inserts the card at position r_k into a uniformly chosen random position (a more formal definition will be given later). We do not see how to apply the strong stationary time technique of [20] to semi-random insertions, yet the spectral technique developed in this paper applies to both semi-random transpositions and semi-random insertions. In either case, it shows that a deck of n cards is mixed up after order $n \log n$ shuffles. The same technique applies to many further examples.

After this work was completed, Yuval Peres informed us that the improved upper bound obtained here for semi-random transpositions was derived independently and by a similar argument by Murali Ganapathy in [15] where time inhomogeneous Markov chains are interpreted as models for adversarially modified Markov chains.

2. Time inhomogeneous Markov chains

2.1. Basic notation

Let V be a finite set equipped with a sequence of kernels $(K_n)_1^\infty$ such that, for each n, $K_n(x, y) \geq 0$ and $\sum_y K_n(x, y) = 1$. An associated Markov chain is a V-valued random process $X = (X_n)_0^\infty$ such that

$$P(X_n = x | X_{n-1} = y, X_{n-2} = x_{n-2}, \ldots, X_0 = x_0) = P(X_n = x | X_{n-1} = y) = K_n(x, y).$$

The distribution μ_n of X_n is determined by the initial distribution μ_0 and given by

$$\mu_n(x) = \sum_{x \in V} \mu_0(x) K_{0,n}(x, y)$$

where $K_{n,m}(x, y)$ is defined inductively for each n and each $m \geq n$ by

$$K_{n,m}(x, y) = \sum_{z \in V} K_{n,m-1}(x, z) K_m(z, y)$$

with $K_{n,n} = I$ (the identity). If we interpret the K_n's as matrices then this definition means that $K_{n,m} = K_{n+1} \cdots K_m$. This paper is mostly concerned with the behavior of the measures

$K_{0,n}(x,\cdot)$ as n tends to infinity. In the case of time homogeneous chains where all $K_i = Q$ are equal, we write $K_{0,n} = Q^n$.

Definition 2.1. We say that a measure π is invariant for the sequence $(K_n)_1^\infty$ if, for each n, we have

$$\sum_{x\in V} \pi(x)K_n(x,y) = \pi(y).$$

We say that a measure π is reversible for the sequence $(K_n)_1^\infty$ if, for each n, we have

$$\pi(x)K_n(x,y) = \pi(y)K_n(y,x).$$

Recall that a Markov kernel K on V is irreducible if for any $x, y \in V$ there exists $n = n(x,y)$ and a finite sequence $(x_i)_1^n$ with $x_0 = x, x_n = y$ and $K(x_i, x_{i+1}) > 0$, $i = 0,\dots,n-1$. An irreducible Markov kernel on a finite set V admits a unique invariant probability measure and this measure is positive (i.e., gives positive mass to every element of V).

Obviously, most sequences of Markov kernels do not admit any invariant measure and the existence of such a measure is a very special assumption on the sequence $(K_n)_1^\infty$. However, a large class of examples is provided by (time inhomogeneous) random walks on groups. Namely, let G be a finite group. Then, for any probability measure p on G, the Markov kernel $K(x,y) = p(x^{-1}y)$ admits the uniform measure $\pi_G : \pi_G(A) = |A|/|G|$ ($|A| = \#A$) as an invariant measure. Thus, any sequence $(p_i)_1^\infty$ of probability measures on G yields a sequence $(K_i)_1^\infty$ of Markov kernels having π_G as an invariant measure. The measure π_G is reversible if and only if each p_i satisfies the symmetry condition $p_i(x) = p_i(x^{-1})$, $x \in G$. The iterated kernels $K_{n,m}$ are then given by the convolution product

$$K_{n,m}(x,y) = p_{n+1} * \cdots * p_m(x^{-1}y)$$

where

$$u * v(x) = \sum_{y\in G} u(y)v(y^{-1}x).$$

The problem treated by Borel [4] as well as the examples of semi-random transpositions and semi-random insertions all fall into this category with the group G being the symmetric group. The uniform measure is a reversible measure for semi-random transpositions but not for semi-random insertions. The following example shows that a good choice of kernels can lead to a very efficient mixing process.

Example. On the symmetric group S_n, consider the kernels K_j defined by

$$K_j(x,y) = \begin{cases} 1/(n-j+1) & \text{if } x^{-1}y = (j,k) \text{ for some } k \in \{j,\dots,n\} \\ 0 & \text{otherwise.} \end{cases}$$

Thus K_j corresponds to transposing the card in position j with the card at a uniformly chosen position in $\{j,\dots,n\}$. With this notation, the sequence $K_1,\dots,K_{n-1}$ leads to a uniformly chosen permutation, that is,

$$K_{0,n-1}(x,\cdot) = \pi$$

where π denotes the uniform measure on S_n. This is a special case of the subgroup algorithm of Diaconis and Shahshahani [12]. Note that, except for K_1, the K_i's are not irreducible.

Definition 2.2. Fix a probability measure π on V. Let $\mathcal{Q} = \{Q_1, \ldots, Q_n\}$ be a finite set of Markov kernels on a finite set V, all admitting π as an invariant measure. We say that $(\mathcal{Q}, \pi)$ is ergodic if, for any sequence $(K_i)_1^\infty$ of Markov kernels with invariant measure π such that $K_i \in \mathcal{Q}$ for infinitely many i's, we have

$$\lim_{n\to\infty} K_{0,n}(x, z) - K_{0,n}(y, z) = 0 \tag{2.1}$$

for all $x, y, z \in V$.

We will give in Theorem 3.4 a necessary and sufficient condition for the ergodicity of a finite family $\mathcal{Q}$ of Markov kernels sharing a given positive invariant measure π.

Remarks. 1. Let $\mathcal{Q} = \{Q_1, \ldots, Q_n\}$ and π be as in Definition 2.2. If $(\mathcal{Q}, \pi)$ is ergodic then for any sequence $(K_i)_1^\infty$ of Markov kernels with invariant measure π such that $K_i \in \mathcal{Q}$ for infinitely many i's, we have

$$\forall x \in V, \quad \lim_{n\to\infty} K_{0,n}(x, \cdot) - \pi = 0.$$

2. Let $\mathcal{Q} = \{Q_1, \ldots, Q_n\}$ and π be as in Definition 2.2. If $(\mathcal{Q}, \pi)$ is ergodic then there exists $V_0 \subset V$ such that V_0 is the unique recurrent class for any $Q_i \in \mathcal{Q}$. Moreover, $V_0 = \{x : \pi(x) > 0\}$. In particular, if π is positive and $(\mathcal{Q}, \pi)$ is ergodic then each kernel in $\mathcal{Q}$ must be irreducible.

3. Fix an irreducible Markov kernel Q with invariant measure π. Set $\mathcal{Q} = \{Q\}$. We will see that the property that $(\mathcal{Q}, \pi)$ is ergodic in the sense of Definition 2.2 is stronger than the property that

$$\forall x, y, z \in V, \quad \lim_{n\to\infty} Q^n(x, z) - Q^n(y, z) = 0$$

which is satisfied if and only if Q is aperiodic. See Theorem 3.4.

4. Condition (2.1) is an example of what [6] calls a merging of measures. Such conditions are classical in the literature of inhomogeneous Markov chains; see [17,21]. Note that remark 1 shows that if a sequence $(K_i)_1^\infty$ has invariant measure π then the merging of measures property in (2.1) yields the stronger result of converging to a distribution.

2.2. Borel–Doeblin ergodicity theorem

In this short section we present one of the simplest quantitative convergence results that we know for time inhomogeneous finite Markov chains admitting a stationary distribution. It essentially captures (in a slightly more general form) the content of Borel's note [4] and is based on a Doeblin type hypothesis. Although the result is quantitative, it usually gives very poor estimates.

Proposition 2.3. *Let $(K_n)_1^\infty$ be a sequence of Markov kernels on a finite set V. Assume that it admits an invariant measure π. For any increasing sequence of integers $(n_j)_0^\infty$ we have*

$$\sup_{x,y\in V} \{|K_{0,n_k}(x, y) - \pi(y)|\} \le \prod_0^{k-1} (1 - c_j) \tag{2.2}$$

where, for each j, c_j is the largest real c such that

$$K_{n_j,n_{j+1}}(x, y) \ge c\pi(y). \tag{2.3}$$

Proof. Observe that $P_j(x,y) = (1-c_j)^{-1}\left(K_{n_j,n_{j+1}}(x,y) - c_j\pi(y)\right)$ is a Markov kernel with invariant measure π and that

$$K_{0,n_k}(x,\cdot) - \pi = \left(\prod_0^{k-1}(1-c_j)\right)\left(\prod_0^{k-1}(P_j(x,\cdot)-\pi)\right).$$

The result follows. □

Proposition 2.4. *Let $(K_n)_1^\infty$ be a sequence of Markov kernels on a finite set V. Assume that it admits an invariant measure π. If there exists an increasing sequence of integers $(n_j)_0^\infty$ such that $\sum_0^\infty c_j = \infty$ where c_j is defined at (2.3) then*

$$\lim_{n\to\infty} |K_{0,n}(x,y) - \pi(y)| = 0.$$

Furthermore, if there exists $c > 0$ such that for $n_j = mj$ we have $c_j > c$ then

$$\sup_{x,y\in V}\{|K_{0,n}(x,y) - \pi(y)|\} \le (1-c)^{\lfloor n/m \rfloor}.$$

Example. Let G be a finite group. Fix a sequence of generating sets $(S_j)_1^\infty$ and assume that each S_j contains the identity element of G. Let $K_j(x,y) = |S_j|^{-1}$ if $x^{-1}y \in S_j$ and $K_j(x,y) = 0$ otherwise. Thus K_j is the Markov kernel of the simple random walk on G associated with S_j. We claim that the time inhomogeneous Markov chain with kernel sequence $(K_j)_1^\infty$ converges to the uniform distribution. To see this, observe that

$$\forall x, \quad xS_j \cdots S_{j+|G|} = G$$

because the sequence of sets $xS_j \cdots S_{j+k}$, $k = 0, 2, \ldots,$ is (strictly) increasing under inclusion. It follows that, for any j, x, y, $K_{j,j+|G|}(x,y) \ge |G|^{-|G|}$ and Proposition 2.4 applies.

3. Spectral analysis

3.1. Singular values

Recall that the singular values of a given linear map A acting on a finite dimensional Euclidean space are the square roots of the eigenvalues of the self-adjoint linear map AA^* where A^* is the adjoint of A (note that A^*A and AA^* have the same eigenvalues).

Given a Markov kernel K with a positive invariant measure π, we can consider K as a linear map

$$Ku = \sum_{y\in V} K(\cdot,y)u(y)$$

defined on the Euclidean space $L^2(V,\pi)$ with scalar product

$$\langle u, v\rangle = \sum_{x\in V} u(x)v(x)\pi(x).$$

The adjoint K^* is associated with the Markov kernel

$$K^*(x,y) = \pi(y)K(y,x)/\pi(x)$$

which also has π as invariant measure. Note that in what follows we always assume that the invariant measure π is positive, that is $\pi(x) > 0$ for each $x \in V$.

Definition 3.1. Let K be a Markov kernel K with positive invariant measure π. For $i \in \{0, \dots, |V| - 1\}$, we denote by $\sigma_i(K)$ the i-th singular value of K on $L^2(V, \pi)$ arranged in non-increasing order.

Note that for any Markov kernel K, $\sigma_0(K) = 1$ and $\sigma_i(K) \in [0, 1]$, $i \in \{1, \dots, |V| - 1\}$. If K is normal (i.e., $K^*K = KK^*$) then K and K^* are diagonalizable in the same basis and the singular values of K are the moduli of the eigenvalues of K counted with multiplicity and arranged in non-increasing order.

Given two probability measures μ, π on V with π positive, set

$$d_2(\mu, \pi) = \left(\sum_y \left| \frac{\mu(y)}{\pi(y)} - 1 \right|^2 \pi(y) \right)^{1/2} \tag{3.1}$$

and

$$d_{\mathrm{TV}}(\mu, \pi) = \sup_{A \subset V} |\mu(A) - \pi(A)| . \tag{3.2}$$

By Jensen's inequality we get that $d_2(\mu, \pi)$ controls $d_{\mathrm{TV}}(\mu, \pi)$ in the following way

$$2d_{\mathrm{TV}}(\mu, \pi) \le d_2(\mu, \pi).$$

Proposition 3.2. *Let K be a Markov kernel with positive invariant measure π on a finite set V. Let $\sigma_i(K)$, $i = 0, \dots, |V| - 1$, be its singular values as introduced in Definition 3.1. Let $(\psi_i)_0^{|V|-1}$ be an orthonormal basis of $L^2(\pi)$ such that ψ_i is an eigenfunction of KK^* with eigenvalue $\sigma_i(K)^2$ (without loss of generality, we always assume that ψ_0 is the constant function 1). Then we have*

$$d_2(K(x, \cdot), \pi)^2 = \sum_{i=1}^{|V|-1} \sigma_i(K)^2 |\psi_i(x)|^2. \tag{3.3}$$

Proof. Set

$$\delta_x(y) = \begin{cases} 1/\pi(x) & \text{if } y = x \\ 0 & \text{otherwise.} \end{cases}$$

Then $\delta_x(y) = \sum_i \overline{\psi_i(x)} \psi_i(y)$ and

$$\begin{aligned} \sum_{y \in V} \left| \frac{K(x, y)}{\pi(y)} - 1 \right|^2 \pi(y) &= \langle K^* \delta_x, K^* \delta_x \rangle = \langle KK^* \delta_x, \delta_x \rangle \\ &= \sum_{i=1}^{|V|-1} \sigma_i(K)^2 |\psi_i(x)|^2. \end{aligned} \tag{3.4}$$

Remark. 1. In the context of time inhomogeneous chains, one would like to apply this result to $K = K_{0,n} = K_1 \cdots K_n$. This is generally not practically feasible because neither the eigenvalues nor the eigenfunctions of $K_1 \cdots K_n K_n^* \cdots K_1^*$ are available. In the next subsection, we use well known singular value inequalities to extract useful estimates from (3.3) when $K = K_1 \cdots K_n$.

2. In the finite Markov chains literature, the use of KK^* was introduced by [18,13] under the name of multiplicative reversibilization.

3.2. Ergodicity via singular values

The main technical result of this paper is the following theorem.

Theorem 3.3. *Let $(K_i)_1^\infty$ be a sequence of Markov kernels on V admitting the positive probability measure π as an invariant distribution. For each j, let $\sigma_i(K_j)$, $i = 0, \dots, |V| - 1$, be the singular values of K_j on $L^2(\pi)$ as in Definition 3.1. Then we have*

$$d_2(K_{0,n}(x, \cdot), \pi) \le \left(\pi(x)^{-1} - 1\right)^{1/2} \prod_1^n \sigma_1(K_j) \tag{3.5}$$

and

$$\sum_{x \in V} d_2(K_{0,n}(x, \cdot), \pi)^2 \pi(x) \le \sum_{i=1}^{|V|-1} \prod_{j=1}^n \sigma_i(K_j)^2. \tag{3.6}$$

Proof. For clarity, we break the proof into two steps. In the first step, we eliminate eigenvectors. For (3.5), apply (3.3) with $K = K_{0,n}$ and note that, by definition, $\sigma_j(K_{0,n}) \le \sigma_1(K_{0,n})$, $j = 1, \dots, |V| - 1$. Thus

$$\begin{aligned} d_2(K_{0,n}(x, \cdot), \pi) &\le \sigma_1(K_1 \cdots K_n) \left(\sum_j |\psi_j(x)|^2\right)^{1/2} \\ &\le \sigma_1(K_1 \cdots K_n) \left(\frac{1}{\pi(x)} - 1\right)^{1/2}. \end{aligned} \tag{3.7}$$

The second inequality is, in fact, an equality and follows from the identity $\delta_x = \sum_0^{|V|-1} \overline{\psi_i(x)} \psi_i$ which implies $\sum_0^{|V|-1} |\psi_i(x)|^2 = \|\delta_x\|_2^2 = \pi(x)^{-1}$.

For (3.6), write

$$\begin{aligned} \sum_{x \in V} d_2(K_{0,n}(x, \cdot), \pi)^2 \pi(x) &= \sum_{x \in V} \sum_{j=1}^{|V|-1} \sigma_i(K_1 \cdots K_n)^2 |\psi_i(x)|^2 \pi(x) \\ &= \sum_{j=1}^{|V|-1} \sigma_i(K_1 \cdots K_n)^2. \end{aligned} \tag{3.8}$$

The second step uses the following singular value inequalities with $k = 1$ for (3.5) and $k = |V| - 1$ for (3.6):

$$\forall\, k = 1, \dots, |V| - 1, \quad \sum_{j=1}^k \sigma_j(K_1 \cdots K_n)^2 \le \sum_{j=1}^k \prod_{i=1}^n \sigma_j(K_i)^2. \tag{3.9}$$

These inequalities follow from Theorem 3.3.4 and Corollary 3.3.10 of [16] after observing that in the case at hand, all the largest singular values (denoted here by $\sigma_0(\cdot)$) are equal to 1. In fact, in the present setting, [16, Theorem 3.3.4] yields the interesting inequalities

$$\forall\, k = 1, \dots, |V| - 1, \quad \prod_{j=1}^k \sigma_j(K_1 \cdots K_n) \le \prod_{j=1}^k \prod_{i=1}^n \sigma_j(K_i). \tag{3.10}$$

Theorem 3.3 now follows from (3.7)–(3.9). □

Remarks. 1. Let K be a Markov kernel with positive invariant measure π. It follows from (3.5) applied to the constant sequence $K_i = K$ that $\sigma_1(K) = 1$ whenever K is either non-irreducible or periodic. Indeed, if $\sigma_1(K) < 1$ then $K^n(x, \cdot)$ converges to π for all x, which implies that K is irreducible and aperiodic. The converse holds true if π is a reversible measure for K but is false in general. On the symmetric group S_n, consider the Markov kernel K corresponding to inserting the top card at one of the two bottom positions (picked with equal probability). This is known as the Rudvalis shuffle and is discussed in [27]. This kernel K yields an irreducible aperiodic chain but K^*K corresponds to either transposing the two bottom cards or doing nothing, each with probability 1/2. In particular K^*K is very far from being irreducible and 1 is a singular value for K with high multiplicity.

2. Let us point out that there are many examples of time inhomogeneous random walks that converge but for which the present techniques may fail to apply. For instance, let each K_i correspond to transposing the cards in positions i and $i + 1$ or doing nothing each with equal probability. Note that in this case $\mathcal{Q} = \{K_1, \ldots, K_n\}$ is not ergodic in the sense of Definition 2.2 and while $\sigma_1(K_n \cdots K_1) < 1$ each K_i has $\sigma_1(K_i) = 1$. See also the example of Section 2.1 (subgroup algorithm).

One basic application of (3.5) is to give a necessary and sufficient condition for a finite family $\mathcal{Q}$ of Markov kernels to be ergodic in the sense of Definition 2.2.

Theorem 3.4. *Let $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ be a finite family of Markov kernels on a finite set V with positive invariant measure π. The pair $(\mathcal{Q}, \pi)$ is ergodic in the sense of Definition 2.2 if and only if $\sigma_1(Q_j) < 1$ for each $j \in \{1, \ldots, k\}$.*

Moreover, if $(\mathcal{Q}, \pi)$ is ergodic then for any sequence $(K_i)_1^\infty$ with invariant measure π such that infinitely many K_i are in $\mathcal{Q}$, we have

$$\forall x \in V, \quad \lim_{n\to\infty} K_{0,n}(x, \cdot) - \pi = 0.$$

Proof. Assume that

$$\sigma = \max_{\{1,\ldots,k\}} \sigma_1(Q_j) < 1.$$

Let $(K_i)_1^\infty$ be a sequence of Markov kernels with invariant measure π such that

$$N_n = \#\{i \in \{1, \ldots, n\} : K_i \in \mathcal{Q}\}$$

tends to infinity with k. By (3.5),

$$d_2(K_{0,n}(x, \cdot), \pi) \le \pi(x)^{-1/2}\sigma^{N_n}.$$

Hence

$$\lim_{n\to\infty} K_{0,n}(x, y) - \pi(y) = 0$$

for all $x, y \in V$.

Conversely, assume that one of the Q_i, say, Q_1, satisfies $\sigma_1(Q_1) = 1$. Then consider the sequence $K_{2i+1} = Q_1$, $K_{2i} = Q_1^*$, $i = 1, 2, \ldots$. As $\sigma_1(Q_1) = 1$, the reversible chain with kernel $Q_1 Q_1^*$ is not irreducible. It follows that there exists x, y, z such that

$$\lim_{n\to\infty} K_{0,2n}(x, y) = \lim_{n\to\infty} [Q_1 Q_1^*]^n(x, y) = 0$$

and

$$\lim_{n\to\infty} K_{0,2n}(z,y) = \lim_{n\to\infty} [Q_1 Q_1^*]^n(z,y) > 0.$$

This shows that

$$\lim_{n\to\infty} K_{0,2n}(x,y) - K_{0,2n}(z,y) \neq 0$$

as desired. □

Remarks. 1. Note that the condition that $\sigma_1(Q_i) < 1$ in Theorem 3.4 cannot be replaced by the hypothesis that the Q_i's are irreducible and aperiodic. For instance, on the symmetric group, let K correspond to inserting the top card into one of the two bottom positions with equal probability. This K is irreducible and aperiodic but the pair $(\{K\}, \pi)$ (where π is the uniform measure) is not ergodic in the sense of Definition 2.2.

2. If the Q_i's are all reversible with respect to π then the condition that $\sigma_1(Q_i) < 1$ for each i is equivalent to the fact that each Q_i is irreducible aperiodic. Thus Theorem 3.4 implies that, for any finite family $\mathcal{Q} = \{Q_1, \dots Q_k\}$ of Markov kernels that are all reversible with respect to a given distribution π and irreducible aperiodic, the pair $(\mathcal{Q}, \pi)$ is ergodic in the sense of Definition 2.2.

We end this section with an observation that often yields control on

$$d_\infty(\mu,\pi) = \sup_{x,y\in V} \left|\frac{\mu(y)}{\pi(y)} - 1\right|. \tag{3.11}$$

Note that Proposition 2.3 is concerned with the quantity $d_\infty(K_{0,n},\pi)$ whereas Theorem 3.3 controls the smaller quantity $d_2(K_{0,n}(x,\cdot),\pi)$. The following simple inequality shows how to bound $d_\infty(K_{0,n},\pi)$ using Theorem 3.3 and other similar results. Namely, for any decomposition $n = (n-m) + m$ for $m < n$, we have

$$\left|\frac{K_{0,n}(x,y)}{\pi(y)} - 1\right| \le d_2(K_{0,m}(x,\cdot),\pi) d_2(K^*_{m,n}(y,\cdot),\pi) \tag{3.12}$$

where $K^*_{m,n} = [K_{m,n}]^* = K^*_n K^*_{n-1} \cdots K^*_{m+1}$. Indeed, with the notation used in the proof of Proposition 3.2, we have

$$\begin{aligned}
\left|\frac{K_{0,n}(x,y)}{\pi(y)} - 1\right| &= \langle (K_{0,n}-\pi)\delta_y, \delta_x\rangle = \langle (K_{0,m}-\pi)(K_{m,n}-\pi)\delta_y, \delta_x\rangle \\
&= \langle (K_{m,n}-\pi)\delta_y, (K^*_{0,m}-\pi)\delta_x\rangle \\
&\le \langle (K_{m,n}-\pi)\delta_y, (K_{m,n}-\pi)\delta_y\rangle^{1/2} \langle (K^*_{0,m}-\pi)\delta_x, (K^*_{0,m}-\pi)\delta_x\rangle^{1/2} \\
&= d_2(K_{0,m}(x,\cdot),\pi) d_2(K^*_{m,n}(y,\cdot),\pi).
\end{aligned}$$

Often, $d_2(K^*_{m,n}(y,\cdot),\pi)$ can be controlled just like $d_2(K_{0,m}(x,\cdot),\pi)$. For instance, using (3.5) and the fact that $\sigma_1(K_j) = \sigma_1(K^*_j)$, we obtain

$$\left|\frac{K_{0,n}(x,y)}{\pi(y)} - 1\right| \le (\pi(x)\pi(y))^{-1/2} \prod_1^n \sigma_1(K_j). \tag{3.13}$$

3.3. Application to time inhomogeneous random walks

Theorem 3.3 simplifies in the case of a random walk. This subsection spells this out and provides a few simple applications. More sophisticated examples will be discussed in Section 4 below.

Let G be a group and let $\pi(x) \equiv 1/|G|$ be the uniform measure on G. For any sequence $(p_i)_1^\infty$ of probability measures on G we let $(K_i)_1^\infty$ be the sequence of Markov kernels with

$$K_i(x, y) = p_i(x^{-1}y).$$

As noted earlier, the iterated kernels $K_{n,m}$ are then given by the convolution product

$$K_{n,m}(x, y) = p_{n+1} * \cdots * p_m(x^{-1}y)$$

where $u * v(x) = \sum_{y \in G} u(y)v(y^{-1}x)$. Set

$$p_{n,m} = p_{n+1} * \cdots * p_m$$

so that

$$K_{n,m}(x, y) = p_{n,m}(x^{-1}y).$$

If the p_i's are equal, say $p_i = p$, we write $p_{0,n} = p * \cdots * p = p^{(n)}$. Finally, set

$$\sigma_j(p_i) = \sigma_j(K_i)$$

where all singular values are with respect to the uniform measure π on G. The following result is a direct application of Theorem 3.3 in the case of random walks. Note that, for random walks as above, $d_2(K_{0,n}(x, \cdot), \pi) = d_2(p_{0,n}, \pi)$ is independent of the starting point x. Hence (3.8) becomes

$$d_2(p_{0,n}, \pi)^2 = \sum \sigma_j(p_1 * \cdots * p_n)^2. \tag{3.14}$$

To obtain (3.17) below, use (3.12).

Theorem 3.5. *Let $(p_i)_1^\infty$ be a sequence of probability measures on a finite group G as above. Then we have*

$$d_2(p_{0,n}, \pi) \le (|G| - 1)^{1/2} \prod_1^n \sigma_1(p_j), \tag{3.15}$$

$$d_2(p_{0,n}, \pi) \le \left(\sum_{i=1}^{|G|-1} \prod_{j=1}^{n} \sigma_i(p_j)^2 \right)^{1/2} \tag{3.16}$$

and, for any $m < n$,

$$d_\infty(p_{0,n}, \pi) \le \left(\sum_{i=1}^{|G|-1} \prod_{j=1}^{m} \sigma_i(p_j)^2 \right)^{1/2} \left(\sum_{i=1}^{|G|-1} \prod_{j=m+1}^{n} \sigma_i(p_j)^2 \right)^{1/2}. \tag{3.17}$$

Example 1. Let G be a finite group. Fix a sequence of generating set $(S_j)_1^\infty$ and assume that each S_j contains the identity element of G. Let $K_j(x, y) = |S_j|^{-1}$ if $x^{-1}y \in S_j$ and $K_j(x, y) = 0$ otherwise. Thus K_j is the Markov kernel of the simple random walk on G associated with S_j.

For each j, consider the Cayley graph associated with the generating set $S_j^\sharp = S_j \cup S_j^{-1}$ and let d_j be the diameter of this Cayley graph. Applying the well known eigenvalue estimate stated in [23, Theorem 6.2] to the reversible walk associated with $K_j K_j^*$, we obtain

$$\sigma_1(K_j)^2 \le 1 - \epsilon_j d_j^{-2}$$

where

$$\epsilon_j = \min_{s \in S_j^\sharp} K_j K_j^*(x, xs) \ge |S_j|^{-2}.$$

Thus the bound (3.15) yields

$$d_2(K_{0,n}(x,\cdot),\pi) \le |G|^{1/2} \prod_1^n \left(1 - \frac{1}{|S_j|^2 d_j^2}\right)^{1/2}.$$

Even if we use the trivial bounds $|S_j| \le |G|$, $d_j \le |G|$, this result is much better than the one obtained in the same situation in Section 2.2.

Example 2. This example illustrates what can be lost going from (3.14) to (3.16). Let $G = \mathbb{Z}/p\mathbb{Z}$, p prime, equipped with its uniform probability measure π. For any $n \in \{1, \dots, p-1\}$, let q_n be the probability measure such that $q_n(\pm n) = 1/2$. Because any $n \in \{1, \dots, p-1\}$ is a generator of the cyclic group G, all the q_n have precisely the same behavior (see, e.g., [23]). Namely, they are irreducible aperiodic and

$$c_1\left(1 + \sqrt{p^2/k}\right) \mathrm{e}^{-C_1 k/p^2} \le d_2(q_n^{(k)}, \pi) \le C_2\left(1 + \sqrt{p^2/k}\right) \mathrm{e}^{-c_2 k/p^2}.$$

Thus it takes order p^2 steps for the walk associated with q_n to converge. Let us now consider the time inhomogeneous chain for which the k-th step is associated with q_n if $k = a(p-1) + n$, $n \in \{1, \dots, p-1\}$, $a = 0, 1, \dots$. In particular, for $k < p$, the distribution after k steps is

$$q_{0,k} = q_1 * \cdots * q_k.$$

All the q_n's share the same orthonormal eigenvectors

$$x \mapsto \exp(-2\pi i j x/p), \quad j = 0, \dots, p-1,$$

with associated eigenvalue $\beta_j(q_n) = \cos(2\pi jn/p)$ (to get the singular values $\sigma_j(q_n)$, simply take the absolute value and enumerate in non-increasing order). Note that the parametrization by $j \in \{0, 1, \dots, p-1\}$ is not the non-increasing enumeration introduced earlier. In any case, the list of all singular values of $q_1 * \cdots * q_k$ (counted with multiplicity) is

$$\prod_{n=1}^k |\cos(2\pi jn/p)|, \quad j = 1, \dots, p-1.$$

Observe that ordering these in non-increasing order is not a simple task! In the present case, (3.14) is equivalent to

$$d_2(q_{0,k},\pi)^2 = \sum_{j=1}^{p-1} \prod_{n=1}^k |\cos(2\pi jn/p)|^2 \tag{3.18}$$

whereas a moment of thought reveals that (3.16) reads

$$d_2(q_{0,k},\pi)^2 \le \sum_{j=1}^{p-1} |\cos(2\pi j/p)|^{2k} = d_2(q_1^{(k)},\pi)^2. \tag{3.19}$$

This last estimate shows that Ap^2 steps with A large enough suffice to reach approximate stationarity. This is very far from optimal. Indeed, we now proceed to show that the correct answer is close to $p^{2/3}$. Since p is prime, for each j, multiplication by $j \pmod p$ is a bijection. Thus the values of $jn \pmod p$, $n = 1, \dots, k$, are all distinct when $k < p$. In particular

$$d_2(q_{0,p-1},\pi)^2 = (p-1)\prod_{j=1}^{p-1} |\cos(2\pi j/p)|^2 \le p\mathrm{e}^{-c_1 p}$$

is very small for large p. In fact, by the argument above, for $k < p/4$, we have

$$\begin{aligned} d_2(q_{0,k},\pi)^2 &\le (p-1)\prod_{j=1}^{k} |\cos(2\pi j/p)|^2 \le p\exp\left(-2\frac{\pi^2}{p^2}\sum_1^k j^2\right) \\ &\le p\exp\left(-\frac{2\pi^2 k^3}{3p^2}\right). \end{aligned}$$

Thus $d_2(q_{0,k},\pi)$ is small after order $p^{2/3}(\log p)^{1/3}$ steps. Now, (3.18) also yields

$$d_2(q_{0,k},\pi) \ge \prod_{j=1}^{k} |\cos(2\pi j/p)|$$

and the left-hand side is larger than e^{-ck^3/p^2} for $k < p/8$. This shows that $d_2(q_{0,k},\pi)$ is not small after order $p^{2/3}$ steps.

4. Examples

4.1. Semi-random transpositions

On the symmetric group S_n, let π be the uniform probability measure. For each $i \in \{1, \dots, n\}$, set

$$q_i(x) = \begin{cases} 1/n & \text{if } x = (i,j) \text{ for some } j,\ 1 \le j \le n, \\ 0 & \text{otherwise.} \end{cases}$$

Here we use the convention that $(i,i) = \mathrm{id}$. Using the method of [11] and results from [14], one proves the following result. See [24] for details.

Theorem 4.1. *Let q_1 be the measure corresponding to transposing top to random on the symmetric group S_n.*

1. *For any $n > 1$ and all $k \ge n(\log n + c)$, $c > 0$, we have*
$$d_2(q_1^{(k)},\pi) \le \sqrt{2}\mathrm{e}^{-c}.$$
2. *For any sequence k_n such that $(k_n - n\log n)/n$ tends to $-\infty$ as n tends to infinity, we have*
$$\lim_{n\to\infty} d_2(q_1^{(k_n)},\pi) = \infty, \qquad \lim_{n\to\infty} d_{\mathrm{TV}}(q_1^{(k_n)},\pi) = 1.$$

Note that $q_i(x) = q_i(x^{-1})$ for all $x \in G$ and all i. Thus these chains are reversible with respect to the uniform measure π. Furthermore, for any i, the measures q_1, q_i are images of each other under some inner automorphisms of the symmetric group. It follows that all q_i's have the same singular values (counted with multiplicity),

$$d_2(q_i^{(k)}, \pi)^2 = d_2(q_1^{(k)}, \pi)^2 = \sum_{m=1}^{n!-1} \sigma_m(q_1)^{2k}. \tag{4.1}$$

Theorem 4.1 applies to q_i, $i \neq 1$, as well.

Definition 4.2 (*Semi-Random Transpositions*). For any sequence $r = (r_i)_1^\infty$, $r_i \in \{1, \dots, n\}$, the r-semi-random transposition Markov chain is the chain associated with the sequence $(K_i)_1^\infty$ where $K_i(x, y) = p_i(x^{-1}y)$, $x, y \in S_n$, with $p_i = q_{r_i}$. We let $p_{0,k}^r = p_1 * \cdots * p_k$ be distribution of this chain after k steps, starting from the identity element.

Our main result about semi-random transpositions is the following.

Theorem 4.3. *For any n and any $r = (r_i)_1^\infty$, $r_i \in \{1, \dots, n\}$, we have*

$$d_2(p_{0,k}^r, \pi) \le d_2(q_1^{(k)}, \pi)$$

for all k. In particular,

$$d_2(p_{0,k}^r, \pi) \le \sqrt{2}\mathrm{e}^{-c}$$

for all $k \ge n(\log n + c)$, $c > 0$.

Proof. By Theorem 3.5, we have

$$d_2(p_{0,k}^r, \pi)^2 \le \sum_{m=1}^{n!-1} \prod_{i=1}^{k} \sigma_m(q_{r_i})^2.$$

As $\sigma_m(q_{r_i}) = \sigma_m(q_1)$ for all m, i, the last inequality together with (4.1) gives

$$d_2(p_{0,k}^r, \pi)^2 \le \sum_{m=1}^{n!-1} \sigma_m(q_1)^{2k} = d_2(q_1^{(k)}, \pi)^2.$$

This and Theorem 4.1 yield the desired results. □

Remark. Mossel et al. [20] prove that $d_{\mathrm{TV}}(p_{0,k}^r, \pi)$ tends to 0 if $k > C_1 n \log n$ and n tends to infinity. Theorem 4.3 is stronger in that it gives an L^2 bound (we always have $2d_{\mathrm{TV}}(\mu, \pi) \le d_2(\mu, \pi)$) and it gives $C_1 = 1$. By Theorem 4.1, this is optimal in the case of transposing top and random. They also prove a very interesting lower bound of order $n \log n$ for the case of cyclic semi-random transpositions, that is, for the case where $r_{i+kn} = i$ for all $1 \le i \le n$ and $k = 0, 1, \dots$.

Set

$$q(x) = \begin{cases} 1/n & \text{if } x = \mathrm{id}, \\ 2/n^2 & \text{if } x = (i, j) \text{ for some } i, j,\ 1 \le i < j \le n, \\ 0 & \text{otherwise.} \end{cases}$$

This is the measure driving the random transposition shuffle studied by Diaconis and Shahshahani in [11]. For completeness, we record the following simple result relating the average behavior of semi-random transpositions to the behavior of random transpositions.

Proposition 4.4. *Let* $\mathbf{E}$ *be the expectation over independent uniform random choices of the entries* $r_i \in \{1, \ldots, n\}$ *of* $r = (r_i)_1^\infty$. *Then, for any n,*

$$\mathbf{E}(d_2(p^r_{0,k}, \pi)) \geq d_2(q^{(k)}, \pi), \qquad \mathbf{E}(d_{\mathrm{TV}}(p^r_{0,k}, \pi)) \geq d_{\mathrm{TV}}(q^{(k)}, \pi)$$

where q denotes the random transposition measure.

Proof. Observe that $\mathbf{E}(p^r_{0,k}(x)) = q^{(k)}(x)$ and use the Minkovski inequality to move the expectation inside the norms. □

Remark. Let k_n be such that $(2k_n - n \log n)/n$ tends to $-\infty$ as n tends to infinity. By [7, p. 43–44], for any $\epsilon \in (0, 1)$, there exists $N(\epsilon)$ such that $d_{\mathrm{TV}}(q^{(k_n)}, \pi) \geq 1 - \epsilon$ for $n > N(\epsilon)$. Thus $\mathbf{E}(d_{\mathrm{TV}}(p^r_{0,k_n}, \pi)) \geq 1 - \epsilon$ for $n > N(\epsilon)$. It follows that

$$\mathbf{P}(\{r : 1 - d_{\mathrm{TV}}(p^r_{0,k_n}, \pi) \geq \eta\}) \leq \epsilon/\eta.$$

This gives $\mathbf{P}(\{r : d_{\mathrm{TV}}(p^r_{0,k_n}, \pi) > 1 - \eta\}) \geq 1 - \epsilon/\eta$. Thus, "most" semi-random transposition schemes take at least k_n steps to mix up a deck of n cards.

4.2. Semi-random insertions

Keeping the notation of the previous section, let $c_{i,j}$ be the permutation corresponding to picking the card in position i and inserting it so that its new position is j, that is,

$$c_{i,j} = \begin{cases} \mathrm{id} & \text{if } i = j \\ (j, j-1, \ldots, i+1, i) & \text{if } 1 \leq i < j \leq n \\ (j, j+1, \ldots, i-1, i) & \text{if } 1 \leq j < i \leq n. \end{cases}$$

Note that $c_{i,j}^{-1} = c_{j,i}$ and $c_{i,j} = c_{j,i}$ if and only if $|j - i| \leq 1$. The random insertion measure on S_n is defined by

$$\widetilde{q}(x) = \begin{cases} 1/n & \text{if } x = \mathrm{id}, \\ 2/n^2 & \text{if } x = c_{i,j} \text{ for some } i, j, 1 \leq i \neq j \leq n, |i - j| = 1 \\ 1/n^2 & \text{if } x = c_{i,j} \text{ for some } i, j, 1 \leq i \neq j \leq n, |i - j| > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note that another description of this measure is that $\widetilde{q}$ is the image of the uniform measure on $\{1, \ldots, n\}^2$ under the map $(i, j) \mapsto c_{i,j}$.

Theorem 4.5 (*Diaconis and Saloff-Coste [8]*). *Let* $\widetilde{q}$ *denote the random insertion measure on* S_n.

1. *For all* $n > 28$ *and all* $k \geq 2n(\log n + c)$, $c > 2$, *we have*

$$d_2(\widetilde{q}^{(k)}, \pi) \leq 2\mathrm{e}^{-(c-2)}.$$

2. *For any sequence* k_n *such that* $(2k_n - n \log n)/n$ *tends to* $-\infty$ *as n tends to infinity, we have*

$$\lim_{n\to\infty} d_2(\widetilde{q}^{(k_n)}, \pi) = \infty, \qquad \lim_{n\to\infty} d_{\mathrm{TV}}(\widetilde{q}^{(k_n)}, \pi) = 1.$$

This result is proved (in a slightly less precise form) in [8] using comparison with random transpositions. For the present version, see [24]. The lower bounds come from the work of Uyemura-Reyes in [26].

For each $i \in \{1, \ldots, n\}$, set

$$\widetilde{q}_i(x) = \begin{cases} 1/n & \text{if } x = c_{i,j} \text{ for some } j, 1 \le j \le n, \\ 0 & \text{otherwise.} \end{cases}$$

This measure is associated with the shuffling scheme "insert the card in position i at random". In particular, $\widetilde{q}_1$ is the well known "top in at random". The measures $\widetilde{q}_i$ are not symmetric and hence π is not reversible for these chains. It also must be emphasized that there is no automorphism of the symmetric group conjugating $\widetilde{q}_i$ to $\widetilde{q}_j$ for $j \notin \{i, n-i\}$. On the one hand, the probabilistic techniques used to study top in at random such as coupling and strong stationary time (see [7, Chap. 4]) do not seem to easily apply for $i \notin \{1, n\}$. On the other hand, the adjoint $\widetilde{q}_i^*$ of $\widetilde{q}_i$ is given by $\widetilde{q}_i^*(x) = \widetilde{q}_i(x^{-1})$ and can be described as "insert a uniformly chosen card in position i". From this description it is obvious that

$$\widetilde{q}_i^* * \widetilde{q}_i = \widetilde{q}, \tag{4.2}$$

that is, "random to i" followed by "i in at random" is exactly "random insertion". The next statement gathers results concerning these shuffles. The upper bound follows from the present argument. The lower bound is obtained as for top to random in [7]. See also [24].

Theorem 4.6. *On the symmetric group S_n, let $\widetilde{q}_i$ denote the i in at random measure, $i \in \{1, \ldots, n\}$, and let $\widetilde{q}$ be the random insertion measure.*

1. *For all $i \in \{1, \ldots, n\}$ and $j = 0, \ldots, n! - 1$, we have $\sigma_j(\widetilde{q}_i) = \sigma_j(\widetilde{q})^{1/2}$.*
2. *For all $i \in \{1, \ldots, n\}$ and $k = 0, 1, \ldots,$*
$$d_2(\widetilde{q}_i^{(2k)}, \pi) \le d_2(\widetilde{q}^{(k)}, \pi).$$
In particular, for all $n > 28$, $c > 2$ and $k \ge 4n(\log n + c)$, we have
$$d_2(\widetilde{q}_i^{(k)}, \pi) \le 2\mathrm{e}^{-(c-2)}.$$
3. *For any sequence k_n such that $(2k_n - n\log n)/n$ tends to $-\infty$ as n tends to infinity, we have*
$$\lim_{n\to\infty} d_2(\widetilde{q}_i^{(k_n)}, \pi) = \infty, \qquad \lim_{n\to\infty} d_{\mathrm{TV}}(\widetilde{q}_i^{(k_n)}, \pi) = 1.$$

Definition 4.7 (*Semi-Random Insertions*). For any sequence $r = (r_i)_1^\infty$, $r_i \in \{1, \ldots, n\}$, the r-semi-random insertion Markov chain is the chain associated with the sequence $(K_i)_1^\infty$ where $K_i(x, y) = \widetilde{p}_i(x^{-1}y)$, $x, y \in S_n$, with $\widetilde{p}_i = \widetilde{q}_{r_i}$. We let $\widetilde{p}_{0,k}^r = \widetilde{p}_1 * \cdots * \widetilde{p}_k$ be the distribution of this chain after k steps, starting from the identity element.

Our main result about semi-random insertions is the following.

Theorem 4.8. *For any n and any $r = (r_i)_1^\infty$, $r_i \in \{1, \ldots, n\}$, we have*

$$d_2(\widetilde{p}_{0,2k}^r, \pi) \le d_2(\widetilde{q}^{(k)}, \pi)$$

for all k. Moreover, for all $n > 28$, $c > 2$ and $k \ge 4n(\log n + c)$,

$$d_2(\widetilde{p}_{0,k}^r, \pi) \le 2\mathrm{e}^{-(c-2)}.$$

This follows directly from Theorems 4.5 and 4.6.

Remark. The techniques used here are robust and can be used to treat many other problems. To mention one example, consider a process for which each step is either transpose i to random or insert i in at random (i varying between 1 and n). This process will converge to uniform in order $n \log n$.

4.3. Variable mixing times

In the examples of the last two sections, the individual steps forming the time inhomogeneous chain of interest all had the same mixing time. In this section, we show how to deal with examples of time inhomogeneous chains for which individual steps have possibly different mixing times.

Let G be a finite group equipped with its uniform measure π. Consider a sequence of probability measures $(p_i)_1^\infty$ on G. We are interested in the convergence of the k step measure $p_{0,k} = p_1 * \cdots * p_k$.

Fix $B \in (0, \infty)$. Let $p_i^\#$ denote either $p_i * p_i^*$ or $p_i^* * p_i$ whichever is more convenient (for each i, independently). These are the so called multiplicative reversibilizations of p_i. We assume that we are given the following data concerning each individual step p_i:

(D1) For each i, we have an upper bound $\beta_i \in [0, 1]$ on $\sigma_1(p_i)$, that is,

$$\sigma_1(p_i) \le \beta_i.$$

Note that this upper bound is trivial if $\beta_i = 1$.

(D2) For each i, we have an upper bound $N_i \in [1, +\infty]$ on the number k of steps needed to insure that $d_2([p_i^\#]^{(k)}, \pi) \le B$, that is,

$$\inf\{k : d_2([p_i^\#]^{(k)}, \pi) \le B\} \le N_i.$$

Roughly speaking, N_i estimates the mixing time of the time homogeneous chain $p_i^\#$. Note that this upper bound is trivial if $N_i = +\infty$.

Theorem 4.9. *Fix $B > 0$. Referring to the notation* (D1)–(D2) *above, let*

$$N = \inf\left\{m : \sum_{i=1}^{m} 1/N_i \ge 2\right\}.$$

For $k \ge N$, let I_k be any subset of $\{1, \dots, k\}$ such that $\sum_{I_k} 1/N_i \ge 2$. Then

$$d_2(p_{0,k}, \pi) \le B \prod_{I_k^c} \beta_i$$

where $I_k^c = \{1, \dots, k\} \setminus I_k$.

Proof. By Theorem 3.5, we have

$$d_2(p_{0,k}, \pi)^2 \le \sum_{j=1}^{|G|-1} \prod_{i=1}^{k} \sigma_j(p_i)^2.$$

Because of the definition of N, for each $k \ge N$ there is a least one $I_k \subset \{1, \dots, k\}$ such that $\sum_{I_k} N_i^{-1} \ge 2$. Fix such a subset I_k and choose reals t_i such that $\sum_{I_k} t_i^{-1} = 1$ and $t_i \ge 2N_i$, $i \in I_k$. Then by the generalized Hölder inequality

$$\left\| \prod_I f_i \right\|_1 \le \prod_I \|f_i\|_{t_i}, \qquad \sum_I 1/t_i = 1,$$

we have

$$\begin{aligned} d_2(p_{0,k},\pi)^2 &\le \left(\prod_{I_k^c}\sigma_1(p_i)^2\right)\sum_{j=1}^{|G|-1}\prod_{I_k}\sigma_j(p_i)^2 \\ &\le \left(\prod_{I_k^c}\sigma_1(p_i)^2\right)\prod_{I_k}\left(\sum_{j=1}^{|G|-1}\sigma_j(p_i)^{2t_i}\right)^{1/t_i} \\ &\le \left(\prod_{I_k^c}\sigma_1(p_i)^2\right)\prod_{I_k}\left(\sum_{j=1}^{|G|-1}\sigma_j(p_i)^{4N_i}\right)^{1/t_i} \\ &\le B^2\prod_{I_k^c}\sigma_1(p_i)^2. \end{aligned}$$

To obtain the last inequality we have used (D2) which gives

$$\sum_1^{|G|-1}\sigma_j(p_i)^{4N_i} = d_2([p_i^{\#}]^{(N_i)},\pi)^2 \le B^2$$

and the fact that $\sum_{I_k} 1/t_i = 1$.

Before illustrating Theorem 4.9 with some examples, let us emphasize some of its main features. The point of Theorem 4.9 is to estimate for the mixing time of a time inhomogeneous random walk based only on information on each individual step and not on any knowledge of how the different steps interact. To be more precise, define

$$\tau(p_i^{\#}) = \inf\left\{k : d_2([p_i^{\#}]^{(k)},\pi) \le 1\right\}$$

to be the L^2 mixing time of the random walk driven by $p_i^{\#}$ and let

$$\tau = \inf\left\{k : d_2(p_{0,k},\pi) \le 1\right\}$$

be the L^2 mixing time of the time inhomogeneous random walk driven by $(p_i)_1^{\infty}$. Then Theorem 4.9 asserts that

$$\tau \le \inf\left\{m : \sum_1^m 1/\tau(p_i^{\#}) \ge 2\right\}.$$

It is useful and interesting that this result does not require knowledge of the singular values of each individual step beyond the "global" information contained in $\tau(p_i^{\#})$. Assume for instance that all individual steps p_i are drawn from a finite set $\mathcal{Q} = \{q_1,\dots,q_s\}$ and that q_i appears with frequency f_i, $\sum_1^s f_i = 1$ (by this we mean that for m large enough the number of steps driven by q_i in the first m steps is very close to mf_i). Then τ is controlled by

$$2\left(\sum_1^s f_i/\tau(q_i^{\#})\right)^{-1},$$

i.e., twice the harmonic mean (with weights f_i) of the $\tau(p_i^{\#})$. $\square$

Example 1. Let G be the symmetric group S_n. Let q be the random transposition measure and q_j be the transpose j with random measure. Let (p_i) be a sequence of probability measure on

S_n such $p_i = q$ if $i = 0 \pmod 5$, $p_i \in \{q_1, \dots q_n\}$ if $i = 2, 3 \pmod 5$, and p_i is arbitrary otherwise. In this case we have $\tau(q^\#) = \tau(q)/2 = (\frac{1}{4} n \log n)(1 + o(1))$, $\tau(q_i^\#) = \tau(q_i)/2 = (\frac{1}{2} n \log n)(1 + o(1))$. Theorem 4.9 yields

$$d_2(p_{0,k}, \pi) \le 1$$

for $k \ge (\frac{5}{4} n \log n)(1 + o(1))$ where $5/4 = 2/(4\frac{1}{5} + 2\frac{2}{5} + 0\frac{2}{5})$.

Example 2. Let G be the Heisenberg group (mod p), $p > 2$ prime, i.e.,

$$G = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{Z}/p\mathbb{Z} \right\}.$$

Let X (resp. Y) be the matrix with $x = 1, y = z = 0$ (resp. $y = 1, x = z = 0$) and, for any $n \in \{1, \dots, p - 1\}$, set

$$S_n = \{I, X^n, X^{-n}, Y^n, Y^{-n}\}$$

where I denotes the identity matrix. Then S_n is a generating set. Let $q_n(g) = |S_n|^{-1}$ if $g \in S_n$ and 0 otherwise. Let π be the uniform measure on G. Then [10, Theorem 1.1] yields constants c_1, c_2 such that, for all n, $\sigma_1(q_n) \le 1 - c_1/p^2$ and $d_2(q_n^{(k)}, \pi) \le 1$ for $k \ge c_2 p^2$. In this case, the most efficient way to see that the constants $c_1 c_2$ are independent of n is to observe that the different q_n's are images of each other under some group automorphisms. Now, Theorem 4.9 shows that, for any sequence $(p_i)_1^\infty$ with $p_i \in \{q_1 \dots, q_{p-1}\}$, we have

$$d_2(p_{0,k}, \pi) \le \mathrm{e}^{-c_1 m/p^2} \quad \text{for any } k \ge c_2 p^2 + m.$$

This is sharp without further hypotheses on the sequence (p_i). However, for the sequence that goes cyclically through $q_1, \dots q_{p-1}$, one expects a much faster convergence. In fact, in view of Example 2 in Section 3.3, we conjecture that the cyclic sequence above converges in order $p^{2/3}$, up to a logarithmic factor. To prove this seems to be an interesting and possibly quite challenging open problem.

Acknowledgement

This research was partially supported by the NSF grants DMS 0102126 and DMS 0306194.

References

[1] D. Aldous, P. Diaconis, Shuffling cards and stopping times, Amer. Math. Monthly 93 (1986) 333–348.

[2] D. Aldous, J. Fill, Reversible Markov chains and random walks on graphs. http://www.stat.berkeley.edu/users/aldous/.

[3] E. Bolker, D. Robbins, The bias of three pseudo-random shuffles, Aequationes Math. 22 (1981) 268–292.

[4] E. Borel, Sur le battage des cartes, C. R. Acad. Sc. 1 154 (1912) 23–25. Also in Oeuvres de Émile Borel. Tome II. Centre National de la Recherche Scientifique, Paris, 1972.

[5] F.R.K. Chung, Spectral Graph Theory, in: CBMS Regional Conference Series in Mathematics, vol. 92, American Mathematical Society, Providence, RI, 1997.

[6] A. D'Aristotile, P. Diaconis, D. Freedman, On merging of probabilities, Sankhya: The Indian J. Statist. 50 (1988) 363–380. Series A, Pt. 3.

[7] P. Diaconis, Group Representations in Probability and Statistics, in: Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 11, Institute of Mathematical Statistics, Hayward, CA, 1988.

[8] P. Diaconis, L. Saloff-Coste, Comparison techniques for random walks on finite groups, Ann. Probab. 21 (1993) 2131–2156.
[9] P. Diaconis, L. Saloff-Coste, Random walks on finite groups: A survey of analytic techniques, in: Probability Measures on Groups and Related Structures XI, World Scientific, Oberwolfach, 1994, pp. 44–75.
[10] P. Diaconis, L. Saloff-Coste, Moderate growth and random walk on finite groups, Geom. Funct. Anal. 4 (1994) 1–36.
[11] P. Diaconis, M. Shahshahani, Generating a random permutation with random transpositions, Z. Wahrsch. Verw. Gebiete 57 (1981) 159–179.
[12] P. Diaconis, M. Shahshahani, The subgroup algorithm for generating uniform random variables, Probab. Eng. Inf. Sci. 1 (1987) 15–32.
[13] J. Fill, Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process, Ann. Appl. Probab. 1 (1) (1991) 62–87.
[14] L. Flatto, A. Odlyzko, D. Wales, Random shuffles and group representations, Ann. Probab. 13 (1985) 152–178.
[15] M. Ganapathy, Robust mixing time, in: Proceedings of RANDOM'06.
[16] R. Horn, C. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, New York, 1991.
[17] M. Iosifescu, Finite Markov processes and their applications, in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Ltd., Chichester, 1980. Editura Tehnică, Bucharest.
[18] M. Mihail, Combinatorial aspects of expanders, Ph.D. Dissertation, Dept. Computer Science, Harvard University, 1989.
[19] I. Mironov, (Not so) random shuffles of RC4, in: Advances in Cryptology—CRYPTO, in: Lecture Notes in Comput. Sci., vol. 2442, Springer, Berlin, 2002, pp. 304–319.
[20] E. Mossel, Y. Peres, A. Sinclair, Shuffling by semi-random transpositions, in: 45th Symposium on Foundations of Comp. Sci., arXiv:math.PR/0404438, 2004.
[21] U. Păun, Ergodic theorems for finite Markov chains, Math. Rep. (Bucur.) 3 (53) (2001) 383–390. 2002.
[22] L. Saloff-Coste, Lectures on finite Markov chains, in: Lectures on Probability Theory and Statistics (Saint-Flour, 1996), in: Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413.
[23] L. Saloff-Coste, Random walks on finite groups, in: H. Kesten (Ed.), Probability on Discrete Structures, in: Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 263–346.
[24] L. Saloff-Coste, J. Zúñiga, Refined estimates for three basic shuffling schemes.
[25] E. Seneta, On strong ergodicity of inhomogeneous products of finite stochastic matrices, Studia Math. 46 (1973) 241–247.
[26] J. Uyemura-Reyes, Random walk, semi-direct products, and card shuffling, Ph.D. Thesis, Stanford University, 2002.
[27] D. Wilson, Mixing time of the Rudvalis shuffle, Electron. Comm. Probab. 8 (2003) 77–85.