The Annals of Applied Probability
2010, Vol. 20, No. 5, 1831–1853
DOI: 10.1214/09-AAP661

TIME INHOMOGENEOUS MARKOV CHAINS WITH WAVE-LIKE BEHAVIOR

BY L. SALOFF-COSTE[1] AND J. ZÚÑIGA[2]

Cornell University and Stanford University

Starting from a given Markov kernel on a finite set V and a bijection g of V, we construct and study a time inhomogeneous Markov chain whose kernel at time n is obtained from K by transport of g^{n-1}. We show that this construction leads to interesting examples, and we obtain quantitative results for some of these examples.

1. Introduction. In [15, 17, 18], we considered the problem of obtaining quantitative results describing the ergodic behavior of time inhomogeneous finite Markov chains. In general, a time inhomogeneous Markov chain, say on a finite set V, is described by a sequence of Markov kernels $(K_i)_1^\infty$. At time n, the distribution of the chain started at x is denoted by $K_{0,n}(x,\cdot)$. More generally, for $n \le m$, we define $K_{n,m}$ inductively by $K_{n,n} = I$ (the identity matrix) and

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

If each K_i is irreducible and aperiodic, one expects that, in many cases, the Markov chain driven by this sequence will have the property that

$$\forall x,y \qquad \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} \to 0 \qquad \text{as } n \to \infty.$$

We call this property *total variation merging* and say that the chain driven by the sequence $(K_i)_1^\infty$ is merging. Note that, in general, $K_{0,n}(x,\cdot)$ does not tend to a limiting distribution. However, when merging occurs, the chain does forget where it started: asymptotically, the distribution sequence evolves in time following a well-defined pattern which is independent of the starting distribution.

In this paper, we will mostly discuss a stronger notion which we call relative-sup merging. By definition, the sequence $(K_i)_1^\infty$ is merging in relative-sup if

$$\max_{x,y,z\in V}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1\right|\right\} \to 0 \qquad \text{as } n \to \infty.$$

Received January 2009; revised July 2009.
[1]Supported in part by NSF Grant DMS-06-03886.
[2]Supported in part by NSF Grants DMS-06-03886, DMS-03-06194 and DMS-08-03018.
AMS 2000 subject classifications. 60J05, 60J10.
Key words and phrases. Time inhomogeneous Markov chains, wave like behavior, singular values.

In general, the relative-sup distance between two measures μ and ν (on a finite or countable state space) is defined by (note the asymmetry)

$$\max_{x\in V}\left\{\left|\frac{\mu(x)}{\nu(x)}-1\right|\right\}.$$

In particular, for a time inhomogeneous chain driven by a sequence $(K_i)_1^\infty$ of Markov kernels, we will consider quantities such as

$$\max_{x,z\in V}\left\{\left|\frac{K_{0,n}(x,z)}{\mu_n(z)}-1\right|\right\},$$

where $\mu_n=\mu_0K_{0,n}$ for some starting measure μ_0. For $\varepsilon>0$, we also define the ε relative-sup merging time $T_\infty(\varepsilon)$ by

$$T_\infty(\varepsilon)=\min\left\{n:\max_{x,y,z\in V}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)}-1\right|\right\}<\varepsilon\right\}.$$

See [17] for more details.

Background and general results concerning time inhomogeneous Markov chains are described in [10, 14, 19] where further references can be found. It turns out that the study of merging is difficult, both at the qualitative and the quantitative level, except in the special but interesting case when all the kernels in the sequence $(K_i)_1^\infty$ share the same stationary probability measure. See, for example, [3, 8, 13, 15]. Only a small set of examples have been treated in the literature mostly because proving anything about concrete time inhomogeneous Markov chains is difficult.

This paper describes a special class of examples whose structure is, in itself, quite interesting and for which some results can be obtained. The set up is as follows. On a finite or countable set V, we are given a Markov kernel K and a bijection $g:V\to V$. We then consider the time inhomogeneous Markov chain driven by the sequence of the kernels

$$K_i(x,y)=K(g^{i-1}x,g^{i-1}y),\qquad x,y\in V, i=1,2,\dots.$$

The problem is to study this time inhomogeneous chain and its merging properties. As we shall see, this covers some interesting examples and leads to interesting results as well as difficult open problems.

The examples discussed in this paper can serve to illustrate the techniques developed in [17, 18]. In particular, we will make use of the following basic singular value technique. See [1] and Theorem 3.2 of [17].

THEOREM 1.1. *Given a sequence of Markov kernels K_i, $i=1,2,\dots,$ on a set V and a positive probability measure μ_0, set $\mu_n=\mu_0K_{0,n}$ and let $\sigma_1(i)$ be the second largest singular value of the operator $K_i:\ell^2(\mu_i)\to\ell^2(\mu_{i-1})$. Then*

$$\left|\frac{K_{0,n}(x,z)}{\mu_n(z)}-1\right|\le\left(\frac{1}{\mu_0(x)}-1\right)^{1/2}\left(\frac{1}{\mu_n(z)}-1\right)^{1/2}\prod_1^n\sigma_1(i).$$

This good-looking result is deceptive because, unless one can get some control on the sequence of measures μ_n, it is essentially useless. Note in particular that $\sigma_1(n)$ depends very much on μ_{n-1} and μ_n.

2. Stability. It is well established that the stationary distribution of an irreducible aperiodic time homogeneous Markov chain plays a crucial part in the analysis of the ergodic properties of the chain. Not much can be said unless one can get some control on the stationary distribution. Moreover, unless the chain is reversible or some algebraic miracle occurs, the computation of the stationary measure is a difficult problem.

The situation for time inhomogeneous Markov chains is much worse. In order to understand how the chain behaves when started from an arbitrary distribution, it is crucial to find (at least) *one* initial distribution μ_0 such that sequence of probability measures $\mu_n = \mu_0 K_{0,n}$ is somewhat well behaved. The ideal situation is when there is a π such $\pi K_{0,n} = \pi$. This occurs if an only if all K_i admit the same invariant measure π, a rather fortunate but rare circumstance. The next definition, taken from [17], introduces a property that is an obvious weakening of the existence of a common invariant measure.

DEFINITION 2.1. Fix $c \geq 1$. A sequence of Markov kernels $(K_n)_1^\infty$ on a finite set V is c-stable if there exists a measure μ_0 such that

$$\forall n \geq 0,\ x \in V \qquad c^{-1} \leq \frac{\mu_n(x)}{\mu_0(x)} \leq c, \tag{2.1}$$

where $\mu_n = \mu_0 K_{0,n}$. If this holds, we say that $(K_n)_1^\infty$ is c-stable with respect to the measure μ_0.

We refer the reader to [17, 18], for examples, and results involving c-stability. The idea behind this definition is that, if a sequence is c-stable with respect to a probability measure μ_0, then one can study the merging of this sequence *more or less* as one would study the ergodicity of a time homogeneous chain with invariant measure μ_0. Why this is true is not obvious and the required technical details are quite intricate. Precise results in this direction are described in [17, 18]. We think that c-stability is an interesting property in itself and that it deserves some attention. Note also that, even for a fixed sequence $(K_i)_1^\infty$ on a fixed finite state space, c-stability is a nontrivial property. The case of the two point space is treated in [17].

A special case of interest to us here is when the time inhomogeneous Markov chain is driven by a sequence $(K_i)_1^\infty$ that is periodic in the sense that there is an integer k such that

$$\forall i \qquad K_{i+k} = K_i.$$

In such case, there is an obvious candidate for a "good" starting distribution μ_0, namely, the invariant measure π of $K_1 \cdots K_k = K_{0,k}$. Indeed, if we pick $\mu_0 = \pi$ then the sequence $\mu_n = \mu_0 K_{0,n}$ is also periodic of period k. If we can compute π, this might allow us to investigate the property of the sequence μ_n including c-stability. Note however that in many examples of interest, the period k will grow with the size of the state space V so that, even in that case, investigating c-stability in a meaningful way is difficult.

An example of this type is cyclic to random transpositions. On $V = S_n$, the symmetric group, let Q_i be the Markov kernel $Q_i(x, y) = 1/n$ if $y = x$ or if $y = x(i, j)$ for some $j \neq i$ and $Q_i(x, y) = 0$ otherwise. Here (i, j) stands for the corresponding transposition. This kernel corresponds to "transpose the card in position i with the card in a uniformly chosen position." The cyclic-to-random transposition chain is driven by the sequence of kernels $(K_i)_1^\infty$ with $K_i = Q_{i \bmod n}$ (by definition, $Q_0 = Q_n$). See [8, 13, 15]. Of course, in this example, the uniform measure is invariant for all Q_i. Other examples of periodic time inhomogeneous chains are discussed in [3].

3. Periodic waves. We now describe in detail the construction outlined in the introduction. This construction is of a rather general nature and produces periodic time inhomogeneous Markov chains that reduce, in a sense, to time homogeneous chains.

Let K be a Markov kernel on a finite state space V, and let $g : V \to V$, $x \mapsto g(x) = gx$ be a bijection. The order of the map g is

$$k = \min\{n \in \mathbb{N} : \forall x \in V\, g^n x = x\}, \qquad g^n = g \circ g \circ \cdots \circ g.$$

For all $x, y \in V$, set

$$K_i(x, y) = K(g^{i-1}x, g^{i-1}y) \tag{3.1}$$

so that $K = K_1$. Consider the inhomogeneous Markov chain driven by the sequence $(K_i)_1^\infty$ defined above. It is easy to see that all K_i are irreducible aperiodic kernels if and only if K is. Moreover, if K has stationary distribution π then K_i has stationary distribution π_i where $\pi_i(x) = \pi(g^{i-1}x)$. Obviously, the sequence $(K_i)_1^\infty$ is periodic of period k. Examples are discussed below after we discuss some general properties of these chains. Given this definition, the obvious question we face is the following: How are the (quantitative) merging properties of the chain driven by $(K_i)_1^\infty$ related to the (quantitative) ergodic properties of the chain driven by K?

PROPOSITION 3.1. *Set*

$$\widetilde{K}(x, y) = K(x, g^{-1}y), \tag{3.2}$$

where $g^{-1} : V \to V$ is the inverse of the map g. Then $K_{0,n}$ is given by

$$K_{0,n}(x, y) = \widetilde{K}^n(x, g^n y).$$

PROOF. We proceed by induction. For $n = 1$ the result holds by definition. Assume that $\widetilde{K}^n(x, y) = K_{0,n}(x, g^{-n}y)$. Then we have

$$\begin{aligned}\widetilde{K}^{n+1}(x, y) &= \sum_{z\in V} \widetilde{K}^n(x, z)\widetilde{K}(z, y)\\ &= \sum_{z\in V} K_{0,n}(x, g^{-n}z)K_{n+1}(g^{-n}z, g^{-n-1}y)\\ &= K_{0,n+1}\bigl(x, g^{-(n+1)}y\bigr).\end{aligned}$$

This gives the desired result. □

COROLLARY 3.2. *The kernel $\widetilde{K}$ is irreducible aperiodic if and only if there exists an integer $n_0 > 0$ such that for all $x, y \in V$, $K_{0,n_0}(x, y) > 0$.*

The following examples illustrate some of the subtleties of this construction.

EXAMPLE 3.1. Let K be irreducible, periodic of period k, with periodicity classes $C_0, \dots, C_{k-1}$ so that $K(x, y) > 0$ if and only if $x \in C_i$ and $y \in C_{i+1 \bmod k}$. Assume that $|C_0| = \cdots = |C_{k-1}|$, that is, all the periodicity classes have the same cardinality. Let $g : V \to V$ be a bijection such that $g(C_i) = C_{i-1 \bmod k}$. Let $K_i(x, y) = K(g^{i-1}x, g^{i-1}y)$, $\widetilde{K}(x, y) = K(x, g^{-1}y)$ as above. It is clear that $\widetilde{K}(x, y) > 0$ if and only if x, y are in the same class C_i for some i. That is, $\widetilde{K}$ is not irreducible. One the other hand, for any x, y there exists $n = n(x, y)$ such that $K_{0,n}(x, y) > 0$.

EXAMPLE 3.2. On $V = \{1, 2, 3, 4\}$, consider the irreducible aperiodic reversible kernel K given by $K(1, 1) = K(1, 2) = K(2, 1) = K(2, 3) = K(3, 2) = K(3, 4) = 1/2$, $K(4, 3) = 1$ and $K(x, y) = 0$, otherwise. Let g be the map that transposes 3 and 4. Then $K_2(1, 1) = K_2(1, 2) = K(2, 1) = K_2(2, 4) = K_2(4, 2) = K_2(4, 3) = 1/2$, $K_2(3, 4) = 1$ and $K_2(x, y) = 0$, otherwise. The graph structure for kernels K and K_2 is illustrated in Figure 1. It follows that

$$K_{0,2n}(4, 4) = 1, \qquad K_{0,2n+1}(4, 3) = 1.$$

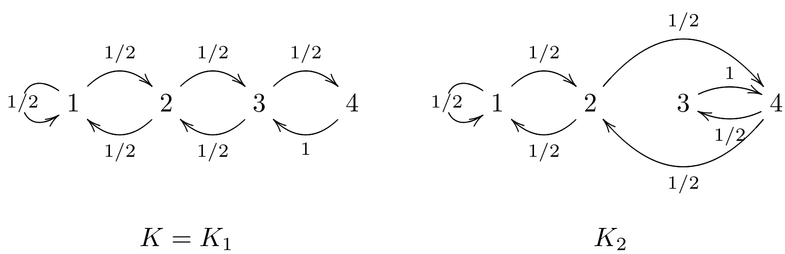

FIG. 1. *Graph structure for kernels K and K_2.*

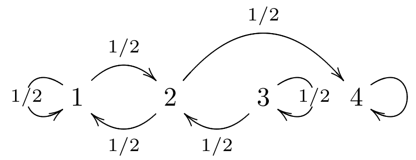

FIG. 2. *Graph structure for $\widetilde{K}$.*

This shows that the property that K is irreducible aperiodic does not imply that for each x, y there is an $n = n(x, y)$ such that $K_{0,n}(x, y) > 0$. Further, $\widetilde{K}(1,1) = \widetilde{K}(1,2) = \widetilde{K}(2,1) = \widetilde{K}(2,4) = \widetilde{K}(3,2) = \widetilde{K}(3,3) = 1/2$, $\widetilde{K}(4,4) = 1$. Hence, $\widetilde{K}$ is not irreducible and has a unique absorbing state, namely, the point 4 as illustrated by Figure 2.

This implies that the sequence $K_1, K_2, K_1, K_2, \dots$ is merging in total variation, that is, $K_{0,n}(x, z) - K_{0,n}(y, z) \to 0$ for any x, y, z. Note that for $z \neq 4$, we have $K_{0,2n}(x, z) \to 0$ for any x. However, this same sequence is not merging in relative-sup distance. Indeed,

$$T_\infty(\varepsilon) = \min\left\{n : \max_{x,y,z}\left\{\left|\frac{K_{0,n}(x, z)}{K_{0,n}(y, z)} - 1\right|\right\} < \varepsilon\right\} = \infty$$

since $K_{0,2n}(4, 1) = 0$ and $K_{0,2n}(1, 1) > 0$.

This gives an example of a pair K_1, K_2 of reversible, irreducible and aperiodic Markov kernels such the sequence $K_1, K_2, K_1, K_2, \dots$ is not merging in relative-sup distance.

EXAMPLE 3.3. On the symmetric group S_n, set σ and σ' to be the cycles $\sigma = (n, n-1, \dots, 1)$ and $\sigma' = (n-1, n-2, \dots, 1)$ and a to be the permutation defined by $a(i) = n - i + 1$. In terms of a deck of n cards, σ takes the top card to the bottom, σ' takes the top card to the second to last position whereas a reverses the order of the deck. Consider the kernel $K(x, y) = 1/2$ if $x^{-1}y \in \{\sigma, \sigma'\}$ and 0 otherwise, and the bijection $g(x) = axa^{-1}$, which is of order 2. Observe that K is irreducible and aperiodic. Note that $g(\sigma) = \sigma^{-1}$ (take the bottom card and put it on top) and $g(\sigma') = (2, 3, \dots, n)$ (take the bottom card and put it in second position). From this it follows that

$$\begin{aligned} K_{0,2}(x, y) &= \sum_z K(x, z)K(g(z), g(y)) \\ &= \begin{cases} 1/4, & \text{if } x^{-1}y \in \{e, (1,2), (1,n), (1,n,2)\}, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

This shows that, for all n, $K_{0,2n}(e, x) = 0$ unless $x \in B = \{e, (1,2), (1,n), (1,n,2)\}$, and $K_{0,2n+1}(e, x) = 0$ unless $x \in \sigma B \cup \sigma' B$. We note that describing $\widetilde{K}$ is difficult.

PROPOSITION 3.3. *Let $\widetilde{\pi}$ be an invariant measure for $\widetilde{K}$. Set*

$$\forall x \in V, i = 1, 2, \dots \qquad \mu_i(x) = \widetilde{\pi}(g^i x).$$

Then $\mu_{i-1}K_i = \mu_i$.

PROOF. Indeed, we have

$$\begin{aligned}\mu_{i-1}K_i(x) &= \sum_{z\in V} \mu_{i-1}(z)K_i(z,x) = \sum_{z\in V} \widetilde{\pi}(g^{i-1}z)K_1(g^{i-1}z, g^{i-1}x)\\ &= \sum_{z\in V} \widetilde{\pi}(g^{i-1}z)\widetilde{K}(g^{i-1}z, g^i x) = \widetilde{\pi}(g^i x) = \mu_i(x).\end{aligned}$$

□

The "wave" appearing in the title of this paper corresponds to the distribution $\widetilde{\pi}$. The time inhomogeneous chain driven by the sequence $(K_i)_1^\infty$ produces the wave $\widetilde{\pi}$, moving around in a periodic fashion under the action of the bijection g on the set V. Despite the similarity in names, we do not claim any connection of this paper with the subject of traveling waves.

COROLLARY 3.4. *Assume that $\widetilde{K}$ admits a positive invariant measure $\widetilde{\pi}$. Then the sequence $(K_n)_1^\infty$ is c-stable with respect to the measure $\mu_0 = \widetilde{\pi}$ with*

$$c = \max_{x,i}\{\widetilde{\pi}(g^i x)/\widetilde{\pi}(x)\}.$$

The next proposition discusses the singular value decompositions of various operators appearing in this construction. The proof is by inspection. We use the following notation. We assume that $\widetilde{\pi}$ is an invariant measure for $\widetilde{K}$ and that $\widetilde{\pi}(x) > 0$ for all $x \in V$. Let $\widetilde{\sigma}_j$, $j = 0, \dots, |V| - 1$, be the singular values of $\widetilde{K} : \ell^2(\widetilde{\pi}) \to \ell^2(\widetilde{\pi})$ in nonincreasing order, and let $(\widetilde{\phi}_j)_0^{|V|-1}$, $(\widetilde{\psi}_j)_0^{|V|-1}$, be orthonormal bases of $\ell^2(\widetilde{\pi})$ such that $\widetilde{K}\widetilde{\phi}_j = \widetilde{\sigma}_j\widetilde{\psi}_j$ (with $\widetilde{\sigma}_0 = 1, \widetilde{\phi}_0 = \widetilde{\psi}_0 \equiv 1$). We refer the reader to [17] for a detailed discussion. The orthonormal bases $(\widetilde{\phi}_j)_0^{|V|-1}$, $(\widetilde{\psi}_j)_0^{|V|-1}$ are, respectively, eigenbases for K^*K and KK^*.

PROPOSITION 3.5. *For any $i = 1 \in \{1, \dots\}$, $\phi_j^i(x) = \widetilde{\phi}_j(g^i x)$, $j = 0, \dots, |V| - 1$, and $\psi_j^i(x) = \widetilde{\psi}_j(g^{i-1}x)$, $j = 0, \dots, |V| - 1$, are orthonormal bases of $\ell^2(\mu_i)$ and $\ell^2(\mu_{i-1})$, respectively, which provide a singular value decomposition of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$ in the sense that $K_i\phi_j^i = \widetilde{\sigma}_j\psi_j^i$. In particular, the singular values $\sigma_j(K_i, \mu_{i-1})$ of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$ are given by $\sigma_j(K_i, \mu_{i-1}) = \widetilde{\sigma}_j$, $j = 0, \dots, |V| - 1$.*

If $\widetilde{\alpha}$ is an eigenvalue of $\widetilde{K}$ with eigenfunction $\widetilde{\omega}$ and k is the order of g then $\widetilde{\alpha}^k$ is an eigenvalue of $K_1 \cdots K_k$ with the same eigenfunction.

This proposition illustrates clearly the difficulties that appear in relating the ergodic properties of the kernel K (that serves as the basic ingredient of this construction) to the merging properties of the sequence $(K_i)_1^\infty$. Indeed, it is rather unclear how the ergodic properties of K and the properties of its stationary measure π relate to $(\widetilde{K}, \widetilde{\pi})$.

In the following two examples, $\pi = \widetilde{\pi}$ is the uniform measure on V. Even in these cases, the above construction is quite interesting and nontrivial. Examples with $\pi \neq \widetilde{\pi}$ will be discussed in the next two sections.

EXAMPLE 3.4 (Cycling for binary vectors). In this example, the kernel K is not irreducible. Take $V = \{0, 1\}^N$ with π being the uniform distribution on V. Let e_i be the binary vector with a unique 1 in position i. Let $K(x, y) = 0$ except if $y = x$ or $y = x + e_1$ in which case $K(x, y) = 1/2$ (K randomizes the first binary entry of x). Let $gx = (x_2, \dots, x_N, x_1)$ if $x = (x_1, \dots, x_N)$ (shift to the left). Using the definition, one checks that K_i is the Markov kernel that randomizes the ith coordinate. Hence, $K_1 \cdots K_N = \pi$ (after N steps, we have a binary vector picked uniformly at random).

The kernel $\widetilde{K}$ corresponds to randomizing the first entry and shifting left. Its invariant measure $\widetilde{\pi}$ is uniform. One recovers immediately the fact that the uniform distribution is reached after exactly N steps. The singular values (=eigenvalues) of K (which is reversible) are 1 (multiplicity $2^N - 1$) and 0 (multiplicity 1). The kernel $\widetilde{K}$ has the property that $\widetilde{K}^*\widetilde{K} = K = K^2$ so that it has the same singular values. The operator $\widetilde{K}$ has two eigenvalues, 0 and 1, and is not diagonalizable, but $\widetilde{K} - \pi$ is nilpotent since $(\widetilde{K} - \pi)^N = 0$.

EXAMPLE 3.5 (Cyclic-to-random transposition). See, for example, [13, 15]. On the symmetric group S_n, let $K(x, y) = 1/n$ if $y = x(1, j)$, $j = 1, 2, \dots, n$, and $K(x, y) = 0$ otherwise (this is called "transpose top with random"). Let σ be the cycle $(1, 2, \dots, n)$ and $g : S_n \to S_n$, $x \mapsto g(x) = \sigma x \sigma^{-1}$. Observe that $g^i((1, j)) = (i, j + i \bmod n)$ so that K_i is "transpose i with random." Hence, we recover the cyclic-to-random transposition chain.

Because $\widetilde{\pi} = \pi$ in this case, it follows that the singular values of $\widetilde{K}$ are equal to the singular values of K which can be computed by using the representation theory of S_n. Note that, as K is reversible, the singular values of K are the square roots of the square of its eigenvalues, that is, the absolute value of the eigenvalues. In particular, $\widetilde{\sigma}_1 = 1 - 1/n$ and thus $\sigma_1(K_i, \pi) = \widetilde{\sigma}_1 = 1 - 1/n$ for all i (see [2, 7, 15, 16]). The eigenvalues of $\widetilde{K}$ are rather mysterious, and it is not clear that $\widetilde{K}$ is diagonalizable. See [13] where the eigenvalues of $K_1 \cdots K_n$ (hence, indirectly, the eigenvalues of $\widetilde{K}$) are investigated and used to obtain a very interesting lower bound on the mixing time of cyclic to random transposition.

Propositions 3.1 and 3.5 reduce the study of the merging of the sequence $(K_i)_1^\infty$ to the study of the ergodicity of the time homogeneous Markov chain driven by $\widetilde{K}$. More precisely, we have the following result.

THEOREM 3.6. *Fix V, K, g, $\widetilde{K}$ and $(K_i)_1^\infty$ as above.*

(1) *The sequence $(K_i)_1^\infty$ is merging in relative-sup if and only if the kernel $\widetilde{K}$ is irreducible and aperiodic.*

(2) *If $\widetilde{K}$ is irreducible and aperiodic, let $\widetilde{\pi}$ be its unique invariant probability measure and set $\mu_i(x) = \widetilde{\pi}(g^i x)$, $x \in V$. Then*

$$\left|\frac{K_{0,n}(x,z)}{\mu_n(z)} - 1\right| \le \left(\frac{1}{\widetilde{\pi}(x)} - 1\right)^{1/2}\left(\frac{1}{\widetilde{\pi}(g^n z)} - 1\right)^{1/2}\widetilde{\sigma}_1^n,$$

where $\widetilde{\sigma}_1$ is the second largest singular value of $\widetilde{K}$ acting on $\ell^2(\widetilde{\pi})$.

PROOF. Use Propositions 3.1 and 3.5. To obtain the last inequality, use Theorem 1.1. Theorem 3.2 of [17] also yields additional inequality for the chi-square distance between $K_{0,n}(x,\cdot)$ and μ_n. □

REMARK 3.7. Example 3.2 gives an example where total variation merging occurs, but $\widetilde{K}$ is not irreducible.

PROPOSITION 3.8. *Assume that K is irreducible and*

$$\min_{x\in V}\{K(x,x)\} > 0.$$

Then, for any bijection g of V, $\widetilde{K}$ is irreducible and aperiodic, and $(K_i)_1^\infty$ is merging in relative-sup.

PROOF. By Example 3.6 of [17] we have $K_{0,|V|}(x,y) > 0$ for all $x, y \in V$. By Corollary 3.2, this implies that $\widetilde{K}$ is irreducible aperiodic. By Theorem 3.6(1), we conclude that $(K_i)_1^\infty$ is merging. □

The proof of the proposition above illustrates the surprising fact that it is not always advantageous to study $\widetilde{K}$ instead of the sequence $(K_i)_1^\infty$. In Proposition 3.8, we use the sequence (K_i) to study $\widetilde{K}$! Indeed, the chain $\widetilde{K}$ seems often difficult to study. For one thing, $\widetilde{K}$ is not necessarily reversible even if K is. In general, this means that computing $\widetilde{\pi}$ may be difficult. Even when we can compute $\widetilde{\pi}$, it might be difficult to study the ergodicity of $\widetilde{K}$ from its definition. Consider, for instance, the case of cyclic-to-random transposition. In this case, $\widetilde{\pi}$ is the uniform distribution, but $\widetilde{K}$ is not invariant under the action of S_n. In other words, the chain driven by $\widetilde{K}$ is not a random walk on S_n. This makes studying $\widetilde{K}$ and its powers directly rather difficult (and, indeed, mysterious). The results obtained in [8, 13, 15] concerning the cyclic-to-random transposition chain are essentially obtained by considering the sequence $(K_i)_1^\infty$, not $\widetilde{K}$ (which, for one thing, does not appear in those papers).

4. Perturbations of symmetric kernels. Let Q be a symmetric Markov kernel on a finite set V, that is, $Q(x,y)=Q(y,x)$ for all $x,y\in V$. This kernel has the uniform distribution $u\equiv 1/|V|$ as its reversible measure. Fix an $\varepsilon\in(0,1)$ and a set $A\subset V$, and consider the kernel

$$K=Q+\Delta_A, \tag{4.1}$$

where Δ_A is some perturbation kernel such that for all $x,y\in V$:

(a) $\sum_z \Delta_A(x,z)=0$,
(b) $\Delta_A(x,y)\geq -\varepsilon Q(x,y)$ and
(c) $x\notin A\Longrightarrow \Delta_A(x,y)=0$.

Let g be a permutation of the vertex set V and consider the sequence $(K_i)_1^\infty$ defined by $K_i(x,y)=K(g^{i-1}x,g^{i-1}y)$. Set $\widetilde{K}(x,y)=K(x,g^{-1}y)$, as before. Let $\widetilde{\pi}$ be an invariant probability measure for $\widetilde{K}$ and set

$$\mu_i(x)=\widetilde{\pi}(g^i x),\qquad x\in V, i=0,1,2,\dots.$$

Define also the symmetric kernel

$$Q_g(x,y)=Q(g^{-1}x,g^{-1}y).$$

Consider the following two assumptions on the kernel $\widetilde{K}$:

(A1) (Irreducibility of $\widetilde{K}$) For all $x,y\in V$ there exists an $n=n(x,y)$ such that $\widetilde{K}^n(x,y)>0$.

(A2) (Aperiodicity of $\widetilde{K}$) There exists a number N such that, for all $m\geq N$ and all $x\in V$, $\widetilde{K}^m(x,x)>0$.

Recall (see Theorem 3.6) that these properties are necessary for the relative-sup merging of the sequence $(K_i)_1^\infty$. In general, it is not obvious at all how they can be checked. However, if the permutation g is an automorphism of the graph structure on V with edge set $E=\{(x,y):K(x,y)>0\}$, then these properties reduce to the similar properties for K (see Proposition 3.1).

The most useful technical result concerning such time inhomogeneous perturbations of Q is the following comparison lemma. For more on comparison techniques see [4].

LEMMA 4.1. *Referring to the above setting, assume that*

$$\exists c>0\qquad \max_{x\in V}\{\widetilde{\pi}(x)\}\leq c\min_{x\in V}\{\widetilde{\pi}(x)\}. \tag{4.2}$$

Consider the operators $Q_g,\widetilde{K}$ acting respectively on $\ell^2(u),\ell^2(\widetilde{\pi})$. Then the Dirichlet forms $\mathcal{E}_{Q_g^*Q_g,u}$ of $Q_g^*Q_g$ on $\ell^2(u)$ and $\mathcal{E}_{\widetilde{K}^*\widetilde{K},\widetilde{\pi}}$ of $\widetilde{K}^*\widetilde{K}$ on $\ell^2(\widetilde{\pi})$ satisfy

$$\mathcal{E}_{Q_g^*Q_g,u}(f,f)\leq \frac{c}{(1-\varepsilon)^2}\mathcal{E}_{\widetilde{K}^*\widetilde{K},\widetilde{\pi}}(f,f) \tag{4.3}$$

for any function f defined on V.

PROOF. Working on $\ell^2(\widetilde{\pi})$ and $\ell^2(u)$, respectively, we compare the kernel $\widetilde{K}^*\widetilde{K}$ to the kernel $Q_g^*Q_g$, that is, Q^*Q moved by g^{-1}. Write

$$\begin{aligned}\widetilde{\pi}(x)\widetilde{K}^*\widetilde{K}(x,y) &\geq \frac{1}{c}\sum_z u(z)K(z,g^{-1}x)K(z,g^{-1}y)\\ &\geq \frac{(1-\varepsilon)^2}{c}\sum_z u(z)Q(z,g^{-1}x)Q(z,g^{-1}y)\\ &= \frac{(1-\varepsilon)^2}{c}u(x)Q_g^*Q_g(x,y).\end{aligned}$$

The third line uses the fact that for any z, $u(g^{-1}z)=u(z)=1/|V|$. $\square$

The importance of this lemma comes from the fact that Q_g is simply Q transported by g^{-1} and thus has the same properties as Q. For instance, Q_g has the same eigenvalues and singular values as Q (the eigenvectors of Q_g are the eigenvectors of Q transported by g^{-1}, etc.). Similarly, Q_g satisfies the same Nash and logarithmic Sobolev inequalities on $\ell^2(u)$ as Q itself. By Lemma 4.1, these properties will be transferred to $(\widetilde{K},\widetilde{\pi})$. The following two propositions and assorted remarks are based on this observation.

PROPOSITION 4.2. *Referring to the above setting, assume that* (4.2) *holds, that is,*

$$\max_{x\in V}\{\widetilde{\pi}(x)\}\leq c\min_{x\in V}\{\widetilde{\pi}(x)\}.$$

Let σ_1 be the second largest singular value of Q on $\ell^2(u)$. Then the second largest singular value $\widetilde{\sigma}_1$ of $\widetilde{K}$ on $\ell^2(\widetilde{\pi})$ is bounded by

$$\widetilde{\sigma}_1\leq 1-\frac{(1-\varepsilon)^2}{c^2}(1-\sigma_1).$$

Furthermore by Theorem 3.6 we obtain

$$\max_{x,z\in V}\left\{\left|\frac{K_{0,n}(x,z)}{\mu_n(z)}-1\right|\right\}\leq c|V|\left(1-\frac{(1-\varepsilon)^2}{c^2}(1-\sigma_1)\right)^n.$$

REMARK 4.3. If instead of using σ_1 we use the logarithmic Sobolev constant $l(Q^*Q)$ of Q^*Q (see [6, 18] for the definition; we follow the notation of [18]); then we get

$$l(\widetilde{K}^*\widetilde{K})\geq\frac{(1-\varepsilon)^2}{c^2}l(Q^*Q).$$

In cases where a good estimate on $l(Q^*Q)$ is known, this can, potentially, improved upon the merging bound stated in the corollary above. See [6, 18].

In the next corollary, we make use of one of the main results of [5, 18] which concerns the use of the Nash inequalities. In applications, the constants c, c_1, C_1, D appearing in the statement below are indeed taking fixed values whereas the parameter T grows with the size of the underlying state space. It is, in general, equal to the square of the diameter of the state space V equipped with the graph structure induced by the symmetric kernel Q. For an introduction to the use of Nash inequality in the study of ergodic Markov chains, see [5].

PROPOSITION 4.4. *Referring to the above setting, assume that there are constants $c, c_1, C_1, D \in (0, \infty)$ and a parameter $T > 1$ such that*:

- *Condition* (4.2) *holds, that is,*

$$\max_{x\in V}\{\widetilde{\pi}(x)\} \le c \min_{x\in V}\{\widetilde{\pi}(x)\}.$$

- *The second largest singular value $\sigma_1(Q)$ of Q on $\ell^2(u)$ satisfies*

$$\sigma_1(Q) \le 1 - \frac{c_1}{T}.$$

- *The kernel Q satisfies the Nash inequality* (*all norms are w.r.t. u*)

$$\forall f : V \to V \qquad \|f\|_2^{2+1/D} \le C_1 T\left(\mathcal{E}_{Q^*Q}(f, f) + \frac{1}{T}\|f\|_2^2\right)\|f\|_1^{1/D}.$$

Then, for any $n > 2T$ and $x, z \in V$, we have

$$\left|\frac{K_{0,n}(x, z)}{\mu_n(z)} - 1\right| \le \left(\frac{16(1 + 4D)C_1 c^{2+3/(2D)}}{(1 - \varepsilon)^2}\right)^{2D} e^{-2c_1(1-\varepsilon)^2(n-2T)/c^2T}.$$

PROOF. Let $u \equiv 1/|V|$. For any function $f : V \to V$ we have $\mathcal{E}_{Q^*Q,u}(f, f) = \mathcal{E}_{Q_g^*Q_g,u}(f \circ g^{-1}, f \circ g^{-1})$ and $\|f\|_p = \|f \circ g^{-1}\|_p$ for $p = 1, 2$. Thus $(\mathcal{E}_{Q_g^*Q_g}, u)$ satisfies the same Nash inequality as $(\mathcal{E}_{Q^*Q}, u)$. By Lemma 4.1 and (4.2), this yields the Nash inequality,

$$\|f\|_{\ell^2(\widetilde{\pi})}^{2+1/D} \le \frac{C_1 T c^{2+3/(2D)}}{(1 - \varepsilon)^2}\left(\mathcal{E}_{\widetilde{K}^*\widetilde{K},\widetilde{\pi}}(f, f) + \frac{1}{T}\|f\|_{\ell^2(\widetilde{\pi})}^2\right)\|f\|_{\ell^1(\widetilde{\pi})}^{1/D}$$

for $(\mathcal{E}_{\widetilde{K}^*\widetilde{K}}, \widetilde{\pi})$. The desired result now follows by applying Propositions 3.1, 4.2 and the results of [5]. (See also Theorem 2.5 of [18].) $\square$

Observe that the conclusion can be rephrased by saying that, under the hypotheses made, the time inhomogeneous chain driven by $(K_i)_1^\infty$ has a relative-sup merging time at most of order T. This will be illustrated below in concrete examples.

Assuming (as is natural) that we understand well the finite Markov chain driven by the symmetric kernel Q, the main difficulty that remains in studying the time inhomogeneous chain $(K_i)_1^\infty$ considered in this section is to verify the condition (4.2) for some (explicit) constant c. The following lemma is useful in this regard.

LEMMA 4.5. *Assume that $\widetilde{\pi} \neq u$ and that $\widetilde{K}$ satisfies the irreducibility condition* (A1) *above. Let $M = \max_x\{\widetilde{\pi}(x)\}$ and $m = \min_x\{\widetilde{\pi}(x)\}$. Let*

$$A_+^* = \left\{x \in V : \sum_y \widetilde{K}(y, x) > 1\right\}, \qquad A_-^* = \left\{x \in V : \sum_y \widetilde{K}(y, x) < 1\right\}.$$

Then there are points $x_+ \in A_+^*, x_- \in A_-^*$ such that $\widetilde{\pi}(x_+) = M, \widetilde{\pi}(x_-) = m$.

PROOF. Let $B = \{z : \sum_y \widetilde{K}(y, z) = 1\}$. Let $x \in V$ be a point such that $\widetilde{\pi}(x) = M$. Then we must have $\sum_y \widetilde{K}(y, x) \geq 1$. If $\sum_y \widetilde{K}(y, x) > 1$, we are done. Otherwise, $x \in B$ and we must have $\widetilde{\pi}(y) = M$ for all y such that $\widetilde{K}(y, x) > 0$. Either one of these points y satisfies $\sum_z \widetilde{K}(z, y) > 1$ and we are done, or we repeat the argument. Since $\widetilde{K}$ satisfies (A1) and $\widetilde{\pi} \neq u$, this process necessarily yields a point x_+ such that $\widetilde{\pi}(x) = M$ and $x_+ \notin B$. Of course, we must then have $x_+ \in A_+^*$. The same line of reasoning proves the existence of the desired point $x_- \in A_-^*$. □

REMARK 4.6. Note that A_+^*, A_-^* are contained in the "$\widetilde{K}$-boundary" of A, that is in the set $A^* = \{z : \exists y \in A, \widetilde{K}(y, z) > 0\}$. Indeed, if $x \notin A^*$ then

$$\sum_y \widetilde{K}(y, x) = \sum_y Q(y, g^{-1}x) = \sum_y Q(g^{-1}x, y) = 1.$$

(a) If we can find n_0 such that $\inf\{\widetilde{K}^{n_0}(x, y) : x, y \in A^*\} > \delta > 0$, then since $\widetilde{\pi} = \widetilde{\pi}\widetilde{K}^{n_0}$, one obtains $\widetilde{\pi}(x_+) = \max\{\widetilde{\pi}\} \leq \delta^{-1}\min\{\widetilde{\pi}\} = \delta^{-1}\widetilde{\pi}(x_-)$. Unfortunately, the nature of the kernel $\widetilde{K}$ makes it difficult to find a suitable n_0.

(b) A variation on this idea is as follows. Assume that, for any $(x, y) \in A_+^* \times A_-^*$, we can find an element $b = b(x, y)$ such that

$$\frac{\widetilde{K}(b, x)}{1 - \sum_{z \neq b} \widetilde{K}(z, x)} \in (0, \infty) \quad \text{and} \quad \frac{1 - \sum_{z \neq b} \widetilde{K}(z, y)}{\widetilde{K}(b, y)} \in (0, \infty).$$

Then for $x, y \in A_+^* \times A_-^*$ such that $\widetilde{\pi}(x) = M$ and $\widetilde{\pi}(y) = m$ as defined in Lemma 4.5 we have

$$\widetilde{\pi}(x) \leq \left(\frac{\widetilde{K}(b, x)(1 - \sum_{z \neq b} \widetilde{K}(z, y))}{\widetilde{K}(b, y)(1 - \sum_{z \neq b} \widetilde{K}(z, x))}\right)\widetilde{\pi}(y).$$

This gives $\max\{\widetilde{\pi}\} \leq C \min\{\widetilde{\pi}\}$ with

$$C = \max_{(x,y) \in A_+^* \times A_-^*} \left\{\frac{\widetilde{K}(b, x)(1 - \sum_{z \neq b} \widetilde{K}(z, y))}{\widetilde{K}(b, y)(1 - \sum_{z \neq b} \widetilde{K}(z, x))}\right\}.$$

Note that C depends on the choice of the $b(x, y)$ for each $(x, y) \in A_+^* \times A_-^*$. Different choices of allowed bs may yield a different constant C. If the location of $\max \widetilde{\pi}$ and $\min \widetilde{\pi}$ can be determined, then there is no need to calculate C over all $A_+^* \times A_-^*$. Examples using this remark are in the next two sections.

5. Cyclic edge perturbation on the circle. This section examines some examples of a moving wave on the circle graph. On the circle graph on $N = 2l + 1$ vertices and for $\varepsilon > 0$ fixed, let K be the reversible Markov kernel corresponding to putting weight 1 on all edges except the (0, 1) edge which has weight $1 + \varepsilon$. Hence

$$K(x, y) = \begin{cases} 0, & \text{if } |x - y| \neq 1, \\ 1/2, & \text{if } |x - y| = 1 \text{ and } x \notin \{0, 1\}, \\ (1+\varepsilon)/(2+\varepsilon), & \text{if } (x, y) \in \{(0, 1), (1, 0)\}, \\ 1/(2+\varepsilon), & \text{if } (x, y) \in \{(0, -1), (1, 2)\}. \end{cases} \tag{5.1}$$

This has reversible measure

$$\pi(x) = \begin{cases} 1/(N+\varepsilon), & \text{if } x \neq 0, 1, \\ (1+\varepsilon/2)/(N+\varepsilon), & \text{if } x = 0, 1. \end{cases}$$

Note that this can be written as a perturbation (see Section 4) of the symmetric kernel Q of simple random walk, $Q(x, y) = 1/2$ if $|x - y| = 1$ and $Q(x, y) = 0$ otherwise. The perturbation set A is $A = \{0, 1\}$ and $\Delta_A = 0$ except for the following values:

$$\Delta_A(0, 1) = \Delta_A(1, 0) = \varepsilon/(4 + 2\varepsilon), \qquad \Delta_A(0, -1) = \Delta_A(1, 2) = -\varepsilon/(4 + 2\varepsilon).$$

Because $N = 2l + 1$ is odd, the chain driven by Q is ergodic with relative-sup mixing time of order N^2. Its singular values (i.e., eigenvalues) on $\ell^2(u)$ are

$$\cos\left(\frac{2\pi j}{N}\right), \qquad j = 0, 1, \ldots, N - 1.$$

In particular, the second largest is attained at $j = (N - 1)/2$ and equals

$$\beta_1 = \cos\frac{\pi}{N}. \tag{5.2}$$

Moreover, Q satisfies the Nash inequality

$$\forall f : V \to V \qquad \|f\|_2^6 \leq 2^7 N^2 \left(\mathcal{E}_{Q^*Q}(f, f) + \frac{1}{4(N+1)^2}\|f\|_2^2\right)\|f\|_1^4. \tag{5.3}$$

See, for example, Theorem 5.2 and Lemma 5.3 in [5].

We will investigate the general construction described earlier based on the kernel K above and various bijections including $x \mapsto x - 1$ and $x \mapsto x + 2$. In these two cases, we prove a merging time estimate of the type

$$T_\infty(\eta) \leq C(\varepsilon) N^2 (1 + \log_+ 1/\eta) \qquad \forall \eta > 0$$

for the associated periodic time inhomogeneous chain, but there are interesting differences in the analysis of the two chains.

First, consider $g(x) = x - 1$. Then K_i is the reversible kernel corresponding to putting weight $1 + \varepsilon$ on the edge $(i - 1, i) \bmod N$. The graphs for Q and K_2 are

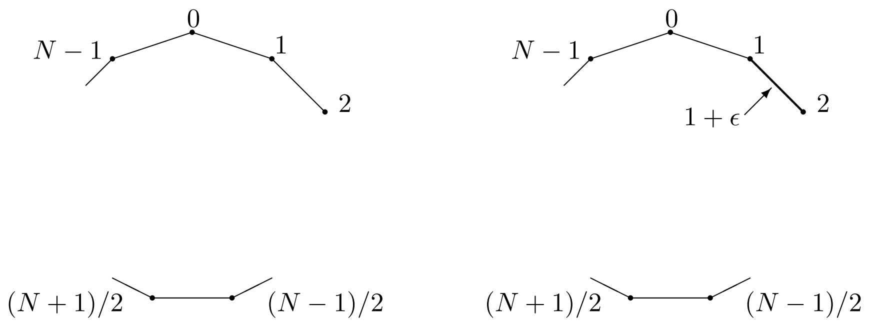

FIG. 3. *The cycling edge perturbation of Q.*

given in Figure 3. The kernel $\widetilde{K}(x,y)=K(x,g^{-1}y)$ is given by

$$\widetilde{K}(x,y)=\begin{cases}0, & \text{if } y\notin\{x,x-2\},\\ 1/2, & \text{if } y\in\{x,x-2\} \text{ and } x\notin\{0,1\},\\ (1+\varepsilon)/(2+\varepsilon), & \text{if } (x,y)\in\{(0,0),(1,-1)\},\\ 1/(2+\varepsilon), & \text{if } (x,y)\in\{(0,-2),(1,1)\}.\end{cases}$$

A simple calculation shows that $\widetilde{\pi}$ is constant away from 0, 1 and that

$$\widetilde{\pi}(x)=\begin{cases}2(1+\varepsilon)/(\varepsilon^2+2N\varepsilon+2N), & \text{if } x\neq 0,1,\\ (\varepsilon+1)(\varepsilon+2)/(\varepsilon^2+2N\varepsilon+2N), & \text{if } x=0,\\ (\varepsilon+2)/(\varepsilon^2+2N\varepsilon+2N), & \text{if } x=1.\end{cases}$$

This proves c-stability of the sequence $(K_i)_1^\infty$ with respect to $\mu_0=\widetilde{\pi}$ with $c=1+\varepsilon$. This distribution yields the wave $\mu_i(x)=\widetilde{\pi}(g^i x)$ created by the time inhomogeneous Markov chain driven by $(K_i)_1^\infty$.

Using Proposition 4.2 and (5.2), this proves that the relative-sup merging time for the sequence $(K_i)_1^\infty$ is bounded by $T_\infty(\eta)\le C(\varepsilon)N^2(\log N+\log_+ 1/\eta)$. An improved result showing relative-sup merging in time of order N^2 is obtained using Proposition 4.4 and the Nash inequality (5.3) of the circle graph.

Let us now consider what happens if we choose $g(x)=x+2$. In terms of the sequence K_i, this means that K_i now has the same perturbation as K but at the edge $(-2i,-2i+1) \bmod N$. The kernel $\widetilde{K}$ is given by

$$\widetilde{K}(x,y)=\begin{cases}0, & \text{if } y-x\notin\{1,3\},\\ 1/2, & \text{if } y-x\in\{1,3\} \text{ and } x\notin\{0,1\},\\ (1+\varepsilon)/(2+\varepsilon), & \text{if } (x,y)\in\{(0,3),(1,2)\},\\ 1/(2+\varepsilon), & \text{if } (x,y)\in\{(0,1),(1,4)\}.\end{cases}$$

Contrary to what happens with $g: x\mapsto x-1$, in the present case, there is no simple formula for $\widetilde{\pi}$ (in particular, $\widetilde{\pi}$ is not constant away from the perturbation). Figure 4 presents a simulation of the stationary measure $\widetilde{\pi}$ for $N=41$ and $\varepsilon=1$.

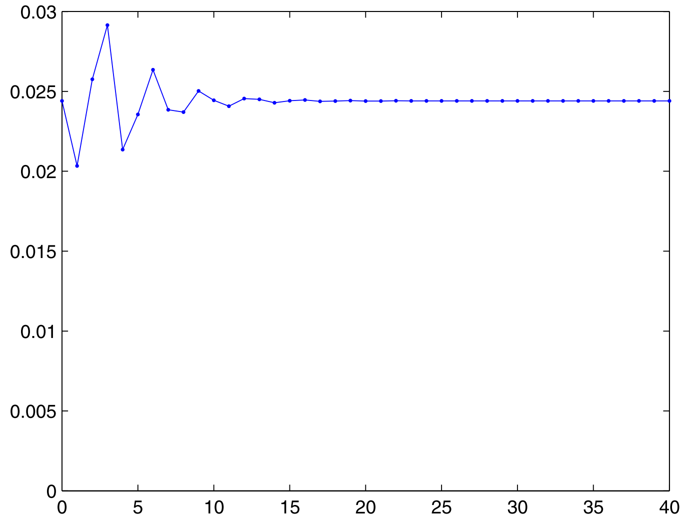

FIG. 4. $\tilde{\pi}$ *for* $N = 41$ *and* $\varepsilon = 1$.

However, it is easy to see from the linear equations defining $\widetilde{\pi}$ (i.e., from Lemma 4.5) that $\max\{\widetilde{\pi}\}$ must be attained at either 2 or 3, and $\min\{\widetilde{\pi}\}$ must be attained at either 1 or 4. Suppose they are attained at 2 and 1. As

$$\widetilde{\pi}(2) = \left(\frac{1+\varepsilon}{2+\varepsilon}\right)\widetilde{\pi}(1) + \frac{1}{2}\widetilde{\pi}(4)$$

we must have

$$\widetilde{\pi}(2) \le \left(\frac{1+\varepsilon}{1+\varepsilon/2}\right)\widetilde{\pi}(1).$$

Suppose instead the max and min are attained at 2 and 4. Then, the same equation gives

$$\left(1 - \frac{1+\varepsilon}{2+\varepsilon}\right)\widetilde{\pi}(2) \le \frac{1}{2}\widetilde{\pi}(4),$$

that is,

$$\widetilde{\pi}(4) \ge \left(\frac{1}{1+\varepsilon/2}\right)\widetilde{\pi}(2).$$

The case where the max and min are attained at 3 and 2 is treated similarly. The remaining case where the max and min are attained at 3 and 1 is slightly different because there is no direct relation between $\widetilde{\pi}(3)$ and $\widetilde{\pi}(1)$. However, the same line of reasoning yields

$$\widetilde{\pi}(3) \le \left(\frac{1+\varepsilon}{1+\varepsilon/2}\right)\widetilde{\pi}(0) \quad \text{and} \quad \widetilde{\pi}(0) \le (1+\varepsilon/2)\widetilde{\pi}(1).$$

This shows that

$$\max\{\widetilde{\pi}\} \le (1+\varepsilon)\min\{\widetilde{\pi}\}. \tag{5.4}$$

Because of this and Corollary 3.4, the sequence $(K_i)_1^\infty$ is $(1+\varepsilon)$-stable with respect to $\widetilde{\pi}$. Applying Proposition 4.4 and (5.3) yield again a relative merging time of order N^2 for the sequence $(K_i)_1^\infty$. The following theorem records this result in more general form.

THEOREM 5.1. *Let $V_N = \{0, \dots, N\}$. Fix $\varepsilon > 0$ and let K be as in (5.1). Fix a permutation $g = g_N$ of V_N and let K_i, $\widetilde{K}, \widetilde{\pi}, \mu_i$ be associated to K, g as in Section 3. Assume that there exists $c \ge 1$ such that*

$$\max_{x\in V_N}\{\widetilde{\pi}(x)\} \le c \min_{x\in V_N}\{\widetilde{\pi}(x)\}. \tag{5.5}$$

Then there is a constant $C(\varepsilon, c)$ such that the relative-sup merging time for $(K_i)_1^\infty$ is bounded by

$$T_\infty(\eta) \le C(\varepsilon, c)N^2(1+\log_+ 1/\eta).$$

REMARK 5.2. For which permutations g of the set $V_N = \{0, \dots, N\}$ does the conclusion of the theorem above hold? According to the theorem, it suffices to check that condition (5.5) is satisfied. For instance, (5.5) is satisfied if $g(x) = x-1$ or $g(x) = x+2$ [in fact, by symmetry, for $g(x) = x \pm 1$, $g(x) = x \pm 2$]. It is very plausible that (5.5) is always satisfied, whatever the permutation g is. However, this does not follow directly from an argument similar to the one used for $g(x) = x-1$ and $g(x) = x+2$. In fact, the argument already fails miserably for $g(x) = x+3$. The reader may want to convince herself of that. In general, we want to compare the min and max of $\widetilde{\pi}$. It is easy to see that the max is attained at either $g(0)$ or $g(1)$ and the min at either $g(-1)$ or $g(2)$. The case where the max and min are attained at either $(g(0), g(2))$ or $(g(1), g(-1))$ can be treated as above because the values of $\widetilde{\pi}$ at $g(0), g(2)$ [resp., at $g(-1), g(1)$] are both related to the value at 1 (resp., 0). But, in the other cases, it becomes much more tricky to compare the max and min without further hypotheses.

Let P be the lazy version of the kernel defined in (5.1) with

$$P(x,y) = \begin{cases} 1/2, & \text{if } x = y, \\ 1/4, & \text{if } |x-y| = 1 \text{ and } x \ne \{0,1\}, \\ (1+\varepsilon)/2(2+\varepsilon), & \text{if } (x,y) \in \{(0,1),(1,0)\}, \\ 1/2(2+\varepsilon), & \text{if } (x,y) \in \{(0,-1),(1,2)\}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

Let g be any permutation of the set $V_N = \{0, \dots, N\}$, and define $P_i(x,y) = P(g^{i-1}x, g^{i-1}y)$ for all $i = 1, 2, \dots$ and $\widetilde{P}(x,y) = P(x, g^{-1}y)$. In this case, we can show that condition (4.2) holds which implies a relative-sup merging time of order N^2 for any permutation g.

THEOREM 5.3. *Let $V_N = \{0, \dots, N\}$. Fix $\varepsilon > 0$ and let P be as in* (5.6). *Fix a permutation $g = g_N$ of V_N and let P_i, $\widetilde{P}, \widetilde{\pi}, \mu_i$ be associated to P, g as in Section* 3 (*replacing K by P*). *Then*

$$\max_{x \in V_N} \{\widetilde{\pi}(x)\} \le (1+\varepsilon) \min_{x \in V_N} \{\widetilde{\pi}(x)\}. \tag{5.7}$$

Furthermore, there is a constant $C(\varepsilon)$ such that the relative-sup merging time for $(P_i)_1^\infty$ is bounded by

$$T_\infty(\eta) \le C(\varepsilon) N^2 (1 + \log_+ 1/\eta).$$

PROOF. By Proposition 4.4 and (5.2)–(5.3), it suffices to prove (5.7). Fix a permutation $g = g_N$ of $V_N = \{0, \dots, N\}$. The kernel $\widetilde{P}$ is given by

$$\widetilde{P}(x,y) = \begin{cases} 1/2, & \text{if } x = g^{-1}y, \\ 1/4, & \text{if } |x - g^{-1}y| = 1 \text{ and } x \neq \{0,1\}, \\ (1+\varepsilon)/2(2+\varepsilon), & \text{if } (x, g^{-1}y) \in \{(0,1),(1,0)\}, \\ 1/2(2+\varepsilon), & \text{if } (x, g^{-1}y) \in \{(0,-1),(1,2)\}, \\ 0, & \text{otherwise.} \end{cases}$$

By Lemma 4.5, the maximum value of $\widetilde{\pi}$ is attained at either $g(0)$ or $g(1)$ and the minimum at $g(-1)$ or $g(2)$. Moreover,

$$\begin{aligned} \widetilde{\pi}(g(-1)) &= \frac{\widetilde{\pi}(-1)}{2} + \frac{\widetilde{\pi}(-2)}{4} + \frac{\widetilde{\pi}(0)}{2(2+\varepsilon)}, \\ \widetilde{\pi}(g(2)) &= \frac{\widetilde{\pi}(2)}{2} + \frac{\widetilde{\pi}(3)}{4} + \frac{\widetilde{\pi}(1)}{2(2+\varepsilon)}, \\ \widetilde{\pi}(g(0)) &= \frac{\widetilde{\pi}(0)}{2} + \frac{\widetilde{\pi}(-1)}{4} + \frac{(1+\varepsilon)\widetilde{\pi}(1)}{2(2+\varepsilon)}, \\ \widetilde{\pi}(g(1)) &= \frac{\widetilde{\pi}(1)}{2} + \frac{\widetilde{\pi}(2)}{4} + \frac{(1+\varepsilon)\widetilde{\pi}(0)}{2(2+\varepsilon)}. \end{aligned}$$

Note that for any of the four possible max/min pairs, the max and min values can be both compared via the equations above to either $\widetilde{\pi}(0)$ or $\widetilde{\pi}(1)$. See Remark 4.6(b). For instance, suppose the max/min pair is $(g(0), g(-1))$. Then

$$\widetilde{\pi}(g(0)) \le \frac{4+2\varepsilon}{4+\varepsilon} \widetilde{\pi}(0) \quad \text{and} \quad \widetilde{\pi}(0) \le \frac{2+\varepsilon}{2} \widetilde{\pi}(g(-1)).$$

Hence,

$$\widetilde{\pi}(g(0)) \le \frac{(2+\varepsilon)^2}{4+\varepsilon} \widetilde{\pi}(g(-1)).$$

The other cases are similar, and it follows that $\max\{\widetilde{\pi}\} \le (1+\varepsilon)\min\{\widetilde{\pi}\}$. $\square$

6. Further examples: Single point perturbations. In the next two examples, we consider perturbations of a symmetric kernel as described in Section 4 but with $A = \{o\}$ for some $o \in V$, that is, the perturbation occurs at a single point. In the second example, we make an additional assumption on the structure of the perturbation. In these cases, we are able to obtain easily applicable bounds.

EXAMPLE 6.1. Let Q be be a symmetric kernel as in Section 4. Fix $\varepsilon \in (0, 1)$, and let $K = Q + \Delta_o$ where $\Delta_o = \Delta_{\{o\}}$ satisfies

$$-\varepsilon Q(o, y) \le \Delta_o(o, y), \qquad \sum_y \Delta_o(o, y) = 0 \quad \text{and} \quad \Delta_o(x, y) = 0 \qquad \text{if } x \ne o.$$

Note that $K(x, y) \ge (1 - \varepsilon)Q(x, y)$, and K satisfies the properties (a)–(c) listed at the beginning of Section 4. Fix a permutation g of V and assume that $\widetilde{K}$ is irreducible. Then Lemma 4.5 says that the min and max of $\widetilde{\pi}$ are attained respectively on A_+^*, A_-^* and Remark 4.6(b) gives

$$\max_{x\in V}\{\widetilde{\pi}\} \le C \min_{x\in V}\{\widetilde{\pi}\}, \tag{6.1}$$

where

$$\begin{aligned} C &= \max_{(x,y)\in A_+^*\times A_-^*}\left\{\frac{\widetilde{K}(o, x)(1 - \sum_{z\ne o}\widetilde{K}(z, y))}{\widetilde{K}(o, y)(1 - \sum_{z\ne o}\widetilde{K}(z, x))}\right\} \\ &\le \max_{x\in A_+^*}\left\{\frac{K(o, g^{-1}x)}{(1 - \varepsilon)Q(o, g^{-1}x)}\right\} \\ &= \frac{1}{(1 - \varepsilon)\theta}, \qquad \theta = \max_{x\in A_+^*}\left\{\frac{K(o, g^{-1}x)}{Q(o, g^{-1}x)}\right\}. \end{aligned}$$

Equation (6.1) and Proposition 4.2 now imply that the relative-sup η merging time of the sequence $(K_i)_1^\infty$ is at most

$$\frac{D}{1 - \sigma_1}(\log|V| + \log_+ 1/\eta), \tag{6.2}$$

where σ_1 is the second largest singular value of the kernel Q on $\ell^2(u)$, and $D = D(\varepsilon, \theta)$ is a constant that depends only on $\varepsilon \in (0, 1)$ and θ (the constant D can easily be made explicit).

EXAMPLE 6.2 (Perturbation of expander graphs). Fix an integer r and consider a sequence $\mathcal{G}_N = (V_N, E_N)$ of regular graphs with vertex set V_N of size $|V_N|$ tending to infinity and symmetric edge set $E_N \subset V_N \times V_N$ with $(x, x) \in E_N$ for all $x \in V_N$. On each graph, consider the symmetric Markov kernel $Q = Q_N$ corresponding to the simple random walk on $\mathcal{G}_N$. Hence, $Q_N(x, y) = 1/r$ if

$(x, y) \in E_N$ and $Q_N(x, y) = 0$ otherwise. Let $\sigma_1(N)$ be the second largest singular value of Q_N on $\ell^2(u_N)$ where u_N is the uniform probability measure on V_N. Assume that there is a constant $a \in (0, 1)$ such that

$$\forall N \qquad 1 - \sigma_1(N) \geq a. \tag{6.3}$$

This property is a strong form of the property that defines the so-called expander graphs (see, e.g., [11, 12] and the references therein).

Fix an origin $o = o_N$ in V_N and consider a perturbation K_N of Q_N as in Example 6.1. Fix also a bijection $g_N : V_N \to V_N$. For each N, consider the time inhomogeneous chain on V_N driven by $(K_{N,i})_1^\infty$ where $K_{N,i}(x, y) = K_N(g_N^{i-1}x, g_N^{i-1}y)$. In this situation, (6.2) yields merging for the sequence $(K_{N,i})_1^\infty$ in order $\log|V_N|$ steps, uniformly in N. Note that this result requires the degree r of the graph to be fixed (or, at least, bounded from above, uniformly in N).

EXAMPLE 6.3. Here we strengthened the hypotheses and the conclusion in the previous example. Namely, we assume that there exists $\delta \in (0, 1 - Q(o, o))$ such that

$$0 < \Delta_o(o, o) \leq \delta, \qquad -\delta\left(\frac{Q(o, y)}{1 - Q(o, o)}\right) \leq \Delta_o(o, y) < 0 \qquad \text{if } y \neq o, \tag{6.4}$$

and

$$\Delta(x, y) = 0 \qquad \text{if } x \neq o.$$

Set

$$\varepsilon = \frac{\delta}{1 - Q(o, o)}. \tag{6.5}$$

A careful analysis of this example yields a much improved estimate for c-stability and the relative sup merging time when compared to the previous example. The difference lies in the fact that the perturbation is positive only at o.

LEMMA 6.1. *Assume that $\widetilde{K}$ is irreducible. Let $m = \min_x\{\widetilde{\pi}(x)\}$ and $M = \max_x\{\widetilde{\pi}(x)\}$. We have that $\widetilde{\pi}(o) = M$ and for ε as in (6.5)*

$$m \geq (1 - \varepsilon)\widetilde{\pi}(o).$$

PROOF. Lemma 4.5 tells us that $M = \widetilde{\pi}(o)$ and that there exists $m = \widetilde{\pi}(x_-)$ for some x_- with $\widetilde{K}(o, x_-) > 0$. Further,

$$\begin{aligned}
\widetilde{\pi}(x_-) &= \sum_x \widetilde{\pi}(x)\widetilde{K}(x, x_-) \\
&\geq \widetilde{\pi}(o)\widetilde{K}(o, x_-) + \widetilde{\pi}(x_-)\sum_{x \neq o} Q(x, g^{-1}x_-) \\
&\geq (1 - \varepsilon)\widetilde{\pi}(o)Q(o, g^{-1}x_-) + \widetilde{\pi}(x_-)\big(1 - Q(o, g^{-1}x_-)\big).
\end{aligned}$$

So we get $\widetilde{\pi}(x_-) \geq (1 - \varepsilon)\widetilde{\pi}(o)$ as desired. □

EXAMPLE 6.4. Let $\mathcal{G}_N = (V_N, E_N)$ be a sequence of regular expander graphs as in Example 6.2 but with degree $r_N \geq 3$ that might depend on N. Fix $\delta \in (0, 2/3)$ and bijections $g_N : V_N \to V_N$. Consider a perturbation K_N of the simple random walk Q_N on $\mathcal{G}_N$ as in Example 6.3. The constant ε at (6.5) is $\varepsilon_N = \delta(r_N/(r_N - 1)) < 3\delta/2$ and the measure $\widetilde{\pi}_N$ satisfies

$$\max_{V_N}\{\widetilde{\pi}_N\} \leq (1 - 3\delta/2)^{-1} \min_{V_N}\{\widetilde{\pi}_N\}.$$

It follows from this and Proposition 4.2 that the associated sequence of perturbed kernels $(K_{N,i})_1^\infty$ merges in order $\log|V_N|$ steps.

EXAMPLE 6.5 (Sticky permutation). The following is a particular case of Example 6.3. It is treated in more detail in [18]. On $V = S_n$, the symmetric group, let

$$Q(x, y) = \begin{cases} 1/2n, & \text{if } y = x(1, j),\ j \in \{2, \dots, n\}, \\ (n+1)/(2n), & \text{if } x = y. \\ 0, & \text{otherwise.} \end{cases}$$

This is the kernel of the lazy version of the random walk called "transpose top and random." Fix a permutation $\rho_n \in S_n$, $\delta \in (0, (n-1)/(2n))$ and let

$$K(x, y) = \begin{cases} Q(x, y), & \text{if } x \neq \rho_n, \\ Q(x, y) + \delta, & \text{if } x = y = \rho_n, \\ Q(x, y) - \delta/(n-1), & \text{if } x = \rho_n \text{ and } y = x(1, j) \\ & \quad \text{for } j \in \{2, \dots, n\}. \end{cases}$$

In words, K is obtained from Q by adding extra holding probability at ρ_n, making ρ_n "sticky." Next, if σ is the cycle $(1, \dots, n)$, let

$$K_i(x, y) = K(\sigma^{i-1} x \sigma^{-i+1}, \sigma^{i-1} y \sigma^{-i+1}).$$

Hence K_i is Q_i with some added holding at $\rho_i = \sigma^{-i+1}\rho\sigma^{i-1}$. This is obviously a special case of Example 6.3, and we thus have

$$\max\{\widetilde{\pi}\} \leq c \min\{\widetilde{\pi}\}, \qquad c = \bigl(1 - 2n\delta/(n-1)\bigr)^{-1}. \tag{6.6}$$

Hence Proposition 4.2 applies. The second largest singular value of Q is known to be $\sigma_1 = 1 - 1/(2n)$ (see, e.g., [2, 7, 16]). This yields an upper bound of order $n(n \log n + \log_+ 1/\eta)$ for the relative-sup merging time $T_\infty(\eta)$ of the sequence $(K_i)_1^\infty$. This result can be improved by using the logarithmic Sobolev inequality technique of [18], (6.6) and Lemma 4.3. The logarithmic Sobolev constant $l(Q^2)$ of Q^2 is of order $1/n \log n$ (see [6]). This yields a relative-sup merging time upper bound of order $n((\log n)^2 + \log_+ 1/\eta)$. This result holds also if we replace the lazy random walk Q above by its nonlazy version, the usual "transpose top with random."

A total variation merging time estimate of order $n(\log n + \log_+ 1/\eta)$ is obtained in [18] by using Lemma 6.1 together with the modified logarithmic Sobolev inequality technique. The crucial point is that the modified logarithmic Sobolev constant $l'(Q^2)$ of Q^2 is of order $1/n$ (see [9, 18]). We do not know how to prove this improved estimate for the nonlazy version of this example.

REFERENCES

[1] DEL MORAL, P., LEDOUX, M. and MICLO, L. (2003). On contraction properties of Markov kernels. *Probab. Theory Related Fields* **126** 395–420. MR1992499

[2] DIACONIS, P. (1991). Finite Fourier methods: Access to tools. In *Probabilistic Combinatorics and Its Applications* (*San Francisco*, *CA*, *1991*). *Proc. Sympos. Appl. Math.* **44** 171–194. Amer. Math. Soc., Providence, RI. MR1141927

[3] DIACONIS, P. and RAM, A. (2000). Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. *Michigan Math. J.* **48** 157–190. MR1786485

[4] DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730. MR1233621

[5] DIACONIS, P. and SALOFF-COSTE, L. (1996). Nash inequalities for finite Markov chains. *J. Theoret. Probab.* **9** 459–510. MR1385408

[6] DIACONIS, P. and SALOFF-COSTE, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112

[7] FLATTO, L., ODLYZKO, A. M. and WALES, D. B. (1985). Random shuffles and group representations. *Ann. Probab.* **13** 154–178. MR770635

[8] GANAPATHY, M. (2007). Robust mixing time. *Electron. J. Probab.* **12** 262–299.

[9] GOEL, S. (2004). Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.* **114** 51–79. MR2094147

[10] IOSIFESCU, M. (1980). *Finite Markov Processes and Their Applications*. Wiley, Chichester. MR587116

[11] LUBOTZKY, A. (1994). *Discrete Groups*, *Expanding Graphs and Invariant Measures*. *Progress in Mathematics* **125**. Birkhäuser, Basel. MR1308046

[12] LUBOTZKY, A. (1995). Cayley graphs: Eigenvalues, expanders and random walks. In *Surveys in Combinatorics*. *London Mathematical Society Lecture Note Series* **218** 155–189. Cambridge Univ. Press, Cambridge. MR1358635

[13] MOSSEL, E., PERES, Y. and SINCLAIR, A. (2004). Shuffling by semi-random transpositions. In *45th Symposium on Foundations of Comp. Sci.* Available at arXiv:math.PR/0404438.

[14] PĂUN, U. (2001). Ergodic theorems for finite Markov chains. *Math. Rep.* (*Bucur.*) **3** 383–390. MR1990903

[15] SALOFF-COSTE, L. and ZÚÑIGA, J. (2007). Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stochastic Process. Appl.* **117** 961–979. MR2340874

[16] SALOFF-COSTE, L. and ZÚÑIGA, J. (2008). Refined estimates for some basic random walks on the symmetric and alternating groups. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 359–392.

[17] SALOFF-COSTE, L. and ZÚÑIGA, J. (2009). Merging of time inhomogeneous Markov chains, part I: Singular values and stability. *Electron. J. Probab.* **14** 1456–1494.

[18] SALOFF-COSTE, L. and ZÚÑIGA, J. (2010). Merging of time inhomogeneous Markov chains, part II: Nash and log-Sobolev inequalities. To appear.

[19] SENETA, E. (1973). On strong ergodicity of inhomogeneous products of finite stochastic matrices. *Studia Math*. **46** 241–247. MR0332843

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
MALOTT HALL
ITHACA, NEW YORK 14853
USA
E-MAIL: lsc@math.cornell.edu

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
BUILDING 380
STANFORD, CALIFORNIA 94305
USA
E-MAIL: jzuniga@math.stanford.edu