Merging and stability for time inhomogeneous finite Markov chains

Laurent Saloff-Coste* and Jessica Zúñiga**

1 Introduction

As is apparent from most text books, the definition of a Markov process includes, in the most natural way, processes that are time inhomogeneous. Nevertheless, most modern references quickly restrict themselves to the time homogeneous case by assuming the existence of a time homogeneous transition function, a case for which there is a vast literature.

The goal of this paper is to point out some interesting problems concerning the quantitative study of time inhomogeneous Markov processes and, in particular, time inhomogeneous Markov chains on finite state spaces. Indeed, almost nothing is known about the quantitative behavior of time inhomogeneous chains. Even the simplest examples resist analysis. We describe some precise questions and examples, and a few results. They indicate the extent of our lack of understanding, illustrate the difficulties and, perhaps, point to some hope for progress.

We think the problems discussed below have an intrinsic mathematical interest (indeed, some of them appear quite hard to solve) and are very natural. Nevertheless, it is reasonable to ask whether or not time inhomogeneous chains are relevant in some applications. Most of the recent interest in Markov chains is related to Monte Carlo Markov Chain algorithms. In this context, one seeks a Markov chain with a given stationary distribution. Hence, time homogeneity is rather natural. See, e.g., [26]. Still, one of the popular algorithms of this sort, the Gibbs sampler, can be viewed as a time inhomogeneous chain (one that, despite huge amount of attention, is still resisting analysis). Time inhomogeneity also appears in the so-called simulated annealing algorithms. See [12] for a discussion that is close in spirit to the present work and for older references. However, certain special features of each of these two algorithms distinguish them from the more basic time inhomogeneous problems we want to discuss here. Namely, in the Gibbs sampler, each individual step is not ergodic (it involves only one coordinate) whereas, in the simulated annealing context, the time inhomogeneity vanishes asymptotically. Other interesting stochastic algorithms that present time inhomogeneity are discussed in [10].

In many applications of finite Markov chains, the kernel describes transitions between different classes in a population of interest. Assuming that these transition probabilities can be observed empirically, one application is to compute the stationary

*Research partially supported by NSF grant DMS 0603886.
**Research partially supported by NSF grants DMS 0603886, DMS 0306194 and DMS 0803018.

measure which describes the steady state of the system. Examples of this type include models for population migrations between countries, models for credit scores used to study the default risk of certain loan portfolios, etc. In such examples, it is natural to consider cases when the Markov kernel describing the evolution of the system depends on time in either a deterministic or a random manner. The reason for the time inhomogeneity may come, for example, from seasonal factors. Or it may model various external events that are independent of the state of the system. Even if one decides that time homogeneity is warranted, one may wish to study the possible effects of small but non-vanishing time dependent perturbations of the model. It seems rather important to understand whether or not such perturbations can drastically alter the behavior of the underlying model. This type of practical questions fit nicely with the theoretical problems discussed below.

A large class of natural examples of time inhomogeneous chains comes from time inhomogeneous random walks on groups. These are discussed in [2], [28]. A special case is the semi-random transpositions model discussed in [14], [21], [22], [28].

2 Merging and stability

This section introduces the two main properties we want to focus on: merging (in total variation or relative-sup) and stability. Given two Markov kernels K_1, K_2, we set

$$K_1K_2(x,y) = \sum_z K_1(x,z)K_2(z,y).$$

Given a sequence $(K_i)_1^\infty$ and $0 \le m \le n$, we set

$$K_{m,n} = K_{m+1} \dots K_n, \quad K_{m,m} = I.$$

2.1 Merging. Recall that an aperiodic irreducible Markov kernel K on a finite state space admits a unique invariant probability measure π. Further, for any starting measure μ_0 and any large time n, the distribution $\mu_n = \mu_0 K^n$ at time n is both essentially independent from the starting distribution μ_0 and well approximated by π.

Consider now the evolution of a system started according to an initial distribution μ_0 and driven by a sequence $(K_i)_1^\infty$ of Markov kernels so that, at time n, the distribution is $\mu_n = \mu_0 K_1 K_2 \dots K_n$. In [1], [4] such a sequence $(\mu_n)_1^\infty$ of probability measures is called a "set of absolute probabilities" but we will not use this terminology here. In many cases, for very large n, the distribution μ_n will be essentially independent of the initial distribution μ_0. Namely, if μ_0, μ_0' are two initial distributions and $\mu_n = \mu_0 K_1 \dots K_n$, $\mu_n' = \mu_0' K_1 \dots K_n$, then it will often be the case that

$$\lim_{n\to\infty} \|\mu_n - \mu_n'\|_{\mathrm{TV}} = 0.$$

We call this loss of memory property *merging* (total variation merging, to be more precise).

One may also want to know whether or not

$$\lim_{n\to\infty} \sup_x \left\{ \left| \frac{\mu'_n(x)}{\mu_n(x)} - 1 \right| \right\} = 0.$$

We call this later property relative-sup merging. Total variation merging is often discussed under the name of "weak ergodicity". See, e.g., [1], [4], [6], [15], [16], [18], [24]. We think "merging" is more appropriate.

If there is merging, then one may want to ask quantitative questions about the merging time. For any $\epsilon \in (0, 1)$, we set

$$T_1(\epsilon) = \inf \{n : \forall \mu_0, \mu'_0,\ \|\mu_n - \mu'_n\|_{\mathrm{TV}} \le \epsilon\} \tag{2.1}$$

and

$$T_\infty(\epsilon) = \inf \left\{ n : \forall \mu_0, \mu'_0,\ \left\| \frac{\mu'_n}{\mu_n} - 1 \right\|_\infty \le \epsilon \right\}. \tag{2.2}$$

The next definition introduces the collective notions of merging and merging time for a given set $\mathcal{Q}$ of Markov kernels.

Definition 2.1. Let $\mathcal{Q}$ be a set of Markov kernels on a finite state space. We say that $\mathcal{Q}$ is merging in total variation (resp. relative-sup) if any sequence $(K_i)_1^\infty$ of kernels in $\mathcal{Q}$ is merging in total variation (resp. relative-sup). We say that $\mathcal{Q}$ has total-variation (resp. relative-sup) ϵ-merging time at most $T(\epsilon)$ if the total variation (resp. relative-sup) ϵ-merging time (2.1) (resp. (2.2)) is bounded above by $T(\epsilon)$, for any sequence $(K_i)_1^\infty$ of kernels in $\mathcal{Q}$.

Let us emphasize that, from the view point of the present work, it is more natural to think in terms of properties shared by all sequences drawn from a set of kernels than in terms of properties of some particular sequence.

2.2 Stability. In the previous section, the notion of merging was introduced as a natural generalization of the loss of memory property in the time inhomogeneous context. The notion of *stability* introduced below is a generalization of the existence of a positive invariant distribution.

Definition 2.2. Fix $c \ge 1$. Given a Markov chain driven by a sequence of Markov kernels $(K_i)_1^\infty$, we say that a probability measure π is c-stable (for $(K_i)_1^\infty$) if there exists a positive measure μ_0 such that the sequence $\mu^n = \mu_0 K_{0,n}$ satisfies

$$c^{-1}\pi \le \mu_n \le c\pi.$$

When such a measure π exists, we say that $(K_i)_1^\infty$ is c-stable.

Example 2.3. Let K be an irreducible aperiodic kernel. Then the chain driven by K is 1-stable. Indeed, it admits a positive invariant measure π and $\pi K^n = \pi$. Further, for any probability measure μ_0 with $\|(\mu_0/\pi) - 1\|_\infty \le \epsilon$, the sequence $\mu_n = \mu_0 K^n$, $n = 1, 2, \ldots$, satisfies $(1-\epsilon)\pi \le \mu_n \le (1+\epsilon)\pi$. Indeed, in the space of signed measures, the linear map $\mu \mapsto \mu K$ is a contraction for the distance $d(\mu, \nu) = \|(\mu/\pi) - (\nu/\pi)\|_\infty$.

In the next definition, we consider the notion of c-stability for a family $\mathcal{Q}$ of Markov kernels on a fixed state space. This definition is of interest even in the case when $\mathcal{Q} = \{Q_1, Q_2\}$ is a pair.

Definition 2.4. Fix $c \geq 1$. Given a set $\mathcal{Q}$ of Markov kernels on a fixed state space, we say that a probability measure π is a c-stable measure for $\mathcal{Q}$ if there exists a positive measure μ_0 such that for any choice of sequence $(K_i)_1^\infty$ in $\mathcal{Q}$, the sequence $\mu_n = \mu_0 K_{0,n}$ satisfies

$$c^{-1}\pi \leq \mu_n \leq c\pi.$$

When such a measure π exists, we say that $\mathcal{Q}$ is c-stable.

Example 2.5. Assume the state space is a group G and let $\mathcal{Q}$ be the set of all Markov kernels Q such that $Q(zx, zy) = Q(x, y)$ for all $x, y, z \in G$. This set is 1-stable with 1-stable measure u, the uniform measure on G.

Example 2.6. On the two-point space, a finite set $\mathcal{Q}$ of Markov kernels is c-stable if and only if it contains no pairs $\{Q_1, Q_2\}$ with $Q_i = \begin{pmatrix} a_i & 1-a_i \\ 1-b_i & b_i \end{pmatrix}$ such that $Q_1 \neq Q_2$, $a_1 = 0$, $b_2 = 0$. This condition is clearly necessary. It is not immediately obvious that it is sufficient. See [29].

Remark 2.7. Consider the problem of deciding whether or not a pair $\mathcal{Q} = \{Q_1, Q_2\}$ of two irreducible ergodic Markov kernels with invariant measure π_1, π_2, respectively, is c-stable. This can be pictured by considering a rooted infinite binary tree with edges labeled Q_1 (= left) and Q_2 (= right) as in Figure 1. Obviously, any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}$ corresponds uniquely to an end $\omega \in \Omega$ where Ω denotes the set of the ends the tree. Given an initial measure μ_0 (placed at the root), the measure $\mu_n^\omega = \mu_0 K_{0,n}$ is obtained by following ω from the root down to level n. Thus, for each choice of μ_0, we obtain a tree with vertices labeled with measures.

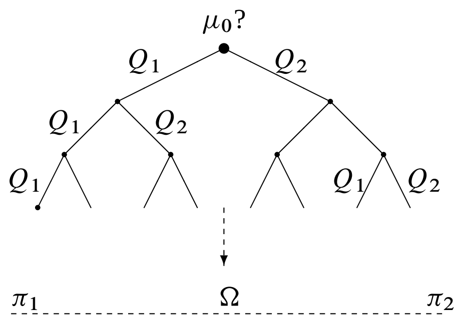

Figure 1. The Q_1, Q_2 tree.

The question of c-stability is the problem of finding an initial measure μ_0 which, in some sense, minimizes the variations among the μ_n^ω's. At the left-most and right-most ends ω_1, ω_2, we get $\mu_n^{\omega_i} \to \pi_i$. Note that, if Q_1, Q_2 share the same invariant measure

$\pi_1 = \pi_2 = \pi$, then the choice $\mu_0 = \pi$ yields a tree all of whose vertices are labeled by π. The existence of a c-stable measure μ_0 can be viewed as a weakening of this. The difficulty is that the existence of an invariant measure and thus the equality between π_1 and π_2 can be viewed as an algebraic property whereas there seems to be no algebraic tools to study c-stability.

2.3 Simple results and examples. We are interested in finding conditions on the individual kernels K_i of a sequence $(K_n)_1^\infty$ that imply merging. This is not obvious even if we consider the very special case when all the K_i's are drawn from a finite set of kernels $\mathcal{Q} = \{Q_0, \dots, Q_m\}$ or even from a pair $\mathcal{Q} = \{Q_0, Q_1\}$.

- Suppose that Q_0, Q_1 are irreducible and aperiodic. Does it imply any sequence $(K_i)_1^\infty$ drawn from $\mathcal{Q} = \{Q_0, Q_1\}$ is merging?

The answer is no. Let π_0 be the invariant measure of Q_0 and let $Q_1 = Q_0^*$ be the adjoint of Q_0 on $\ell^2(\pi_0)$. If (Q_0, π_0) is not reversible (i.e., Q_0 is not self-adjoint on $\ell^2(\pi_0)$) then it is possible that $Q_0 Q_0^*$ is not irreducible. When $Q_0 Q_0^*$ is not irreducible, the sequence $K_i = Q_{i \bmod 2}$ is not merging.

- Suppose that Q_0, Q_1 are reversible, irreducible and aperiodic. Does it imply any sequence $(K_i)_1^\infty$ drawn from $\mathcal{Q} = \{Q_0, Q_1\}$ is merging in relative-sup?

The answer is no, even on the two point space! On the two point space, $\mathcal{Q} = \{Q_0, Q_1\}$ is merging in total variation as long as Q_0, Q_1 are irreducible aperiodic but relative sup merging fails for the irreducible aperiodic pairs of the type

$$Q_0 = \begin{pmatrix} 0 & 1 \\ 1-a & a \end{pmatrix}, \quad Q_1 = \begin{pmatrix} b & 1-b \\ 1 & 0 \end{pmatrix},$$

with $0 < a, b < 1$. See [29].

The following examples are instructive.

Example 2.8. On $S = \{1, \dots, 5\}$ consider the reversible kernels Q_0, Q_1 corresponding to the graphs in Figure 2 (all edges have weight 1). Consider the sequence $K_i = Q_{i \bmod 2}$ so that $K_1 = Q_1, K_2 = Q_0, K_3 = Q_1, \dots$. If, at an even time $n = 2\ell$, the chain is at states 2 or 5 then from that time on, the chain will be in $\{2, 5\}$ at even times and in $\{3, 4\}$ at odd times. In this example, the chain driven by $(K_i)_1^\infty$ is merging in total variation but is not merging in relative-sup.

Example 2.9. The kernels depicted in Figure 3 yield an example where total variation (hence, a fortiori, relative-sup) merging fails. In this example, the sequence $(K_i)_1^\infty$ with $K_i = Q_{i \bmod 2}$ fails to be merging in total variation because the chain will eventually end up oscillating either between 2 and 1, or between $\{4, 7\}$ and $\{5, 6\}$, with a preference for one or the other depending on the starting distribution μ_0.

Let us give two simple results concerning merging.

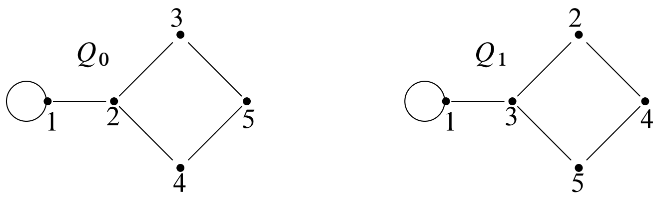

Figure 2. A five-point example.

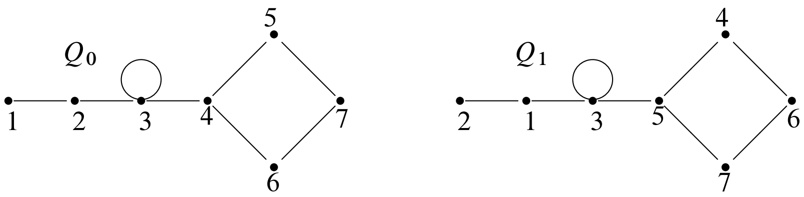

Figure 3. A seven-point example.

Proposition 2.10. *Assume that, for each i, there exists a state y_i and a real $\epsilon_i \in (0,1)$ such that*

$$\forall\, x, \quad K_i(x, y_i) \geq \epsilon_i .$$

If $\sum_i \epsilon_i = \infty$ then the sequence $(K_i)_1^\infty$ is merging in total variation. If, in addition, each K_i is irreducible then the sequence $(K_i)_1^\infty$ is also merging in relative-sup.

Proof. For total variation, this can be proved by a well-known Doeblin's coupling argument (see, e.g., [13], [29]) and irreducibility of the kernels is not needed. Of course, the mass might ultimately concentrate on a fraction of the state space.

Merging in relative-sup is a bit more subtle and irreducibility is needed for that conclusion to hold (even in the time homogeneous case). A proof using singular values can be found in [29]. □

Remark 2.11. Under the much stronger hypothesis $\forall\, x, y, \quad K_i(x,y) \geq \epsilon_i > 0$, one gets an immediate control of any sequence $\mu_n = \mu_0 K_{0,n}$, $n = 1, 2, \dots,$ in the form

$$\forall\, z, \quad \epsilon_n \leq \min_{x,y}\{K_n(x,y)\} \leq \mu_n(z) \leq \sup_{x,y}\{K_n(x,y)\} \leq 1 - (N-1)\epsilon_n$$

where N is the size of the state space.

Remark 2.12. The hypothesis $\exists\, y_i, \forall\, x, \; K_i(x, y_i) \geq \epsilon_i > 0$, is obviously too strong in many cases but it can often be applied to study a time inhomogeneous chain $(K_i)_1^\infty$ by grouping terms and considering the sequence $Q_i = K_{n_i, n_{i+1}}$ for an appropriately chosen increasing sequence n_i. In the simplest case, for a given sequence $(K_i)_1^\infty$, one seeks $\epsilon \in (0,1)$ and an integer m such that $K_{\ell m, \ell m + m}(x,y) \geq \epsilon$ for all x, y, ℓ. When such a lower bound holds, one concludes that (1) the chain is merging in total variation and relative-sup and (2) there exists $c \in (0,1)$ such that for any starting measure μ_0 and n large enough, the measures $\mu_n = \mu_0 K_{0,n}$ satisfy $c \le \mu_n(z) \le 1-c$. However, this type of argument is bound to yield very poor quantitative results in most cases.

For the next result, recall that an adjacency matrix A is a matrix whose entries are either 0 or 1.

Proposition 2.13. *On a finite state space let $(K_i)_1^\infty$ be a sequence of Markov kernels. Assume that:*

(1) (Uniform irreducibility) *There exist ℓ, $\epsilon \in (0,1)$ and adjacency matrices $(A_i)_1^\infty$, such that, $\forall\, i, x, y,\ A_i^\ell(x,y) > 0$ and $K_i(x,y) \ge \epsilon A_i(x,y)$.*

(2) (Uniform laziness) *There exists $\eta \in (0,1)$ such that, $\forall\, i, x,\ K_i(x,x) \ge \eta$.*

Then the chain driven by $(K_i)_1^\infty$ is merging in total variation and relative-sup norm. Moreover, there exist n_0 and $c \in (0,1)$ such that for any starting distribution μ_0, all $n \ge n_0$ and all z, $\mu_n = \mu_0 K_{0,n}$ satisfies $\mu_n(z) \in (c, 1-c)$.

Proof. Let N be the size of the state space. Using (1)–(2), one can show (see [29]) that $K_{n,n+N}(x,y) \ge (\min\{\epsilon,\eta\})^{N-1}$. The desired result follows from Proposition 2.10 and Remark 2.12. □

Note that this argument can only give very poor quantitative results!

3 A short review of the literature

The largest body of literature concerning time inhomogeneous Markov processes come, perhaps, from the analysis of Partial Differential Equations where time dependent coefficients are allowed. The book [36] can serve as a basic reference. Unfortunately, it seems that the results developed in that context are local in nature and are not very relevant to the quantitative problems we are interested in. The literature on (finite) time inhomogeneous Markov chains can be organized under three basic headings: Weak ergodicity, asymptotic structure, and products of stochastic matrices. We now briefly review each of these directions.

3.1 Weak ergodicity. One of the earliest references concerning the asymptotic behavior of time inhomogeneous chains is a note of Emile Borel [2] where he discusses time inhomogeneous card shufflings. In the context of general time inhomogeneous chains on finite state spaces, *weak ergodicity*, which we call *total variation merging*, i.e., the tendency to forget the distant past, was introduced in [19] and is the main subject of [16]. See also [5] and the reference to the work of Doeblin given there. A sample of additional old and not so old references in this direction is [15], [18], [23], [24], [25], [32]. An historical review is given in [33]. The main tools developed in these references to prove weak ergodicity are the use of ergodic coefficients and couplings. A modern perspective, close in spirit to our interests, is in [10], [11], [13]. It may be worth pointing out that, by design, ergodic coefficients mostly capture some asymptotic properties and are not well suited for quantitative results, even in the time homogeneous case.

3.2 Asymptotic structure. One of the basic results in the theory of time homogeneous finite Markov chains describes the decomposition of the state space into non-essential (or transient) states, essential classes and periodic subclasses. It turns out that, perhaps surprisingly, there exists a completely general version of this result for time inhomogeneous chains. This result is rather more subtle than its time homogeneous counterpart. Sonin [34], Theorem 1, calls it the Decomposition-Separation Theorem and reviews its history which starts with a paper of Kolmogorov [19], with further important contributions by Blackwell [1], Cohn [4] and Sonin [34].

Fix a sequence $(K_n)_1^\infty$ of Markov kernels on a finite state space Ω. The Decomposition-Separation Theorem yields a sequence $(\{S_n^k, k = 0, \dots c\})_{n=1}^\infty$ of partitions of Ω so that:

(a) With probability one, the trajectories of any Markov chain (X_n) driven by $(K_n)_1^\infty$ will, after a finite number of steps, enter one of the sequence $S^k = (S_n^k)_{n=1}^\infty$, $k = 1, \dots, c$, and stay there forever. Further, for each k,

$$\sum_{n=1}^{\infty} \mathbf{P}(X_n \in S_n^k; X_{n+1} \notin S_{n+1}^k) + \mathbf{P}(X_n \notin S_n^k; X_{n+1} \in S_{n+1}^k) < \infty.$$

(b) For each $k = 1, \dots, c$, and for any two Markov chains $(X_n^1)_1^\infty, (X_n^2)_1^\infty$ driven by $(K_n)_1^\infty$ such that $\lim_{n\to\infty} \mathbf{P}(X_n^i \in S_n^k) > 0$, and any sequence of states $x_n \in S_n^k$,

$$\lim_{n\to\infty} \frac{\mathbf{P}(X_n^1 = x_n | X_n^1 \in S_n^k)}{\mathbf{P}(X_n^2 = x_n | X_n^2 \in S_n^k)} = 1.$$

The sequence $(S_n^0)_1^\infty$ describes "non-essential states" and a chain is weakly ergodic (i.e., merging in total variation) if and only if $c = 1$, i.e., there is only one essential class. We refer the reader to [34] for a detailed discussion and connections with other problems.

The Decomposition-Separation Theorem can be illustrated (albeit, in a rather trivial way) using Example 2.9 of Figure 3 above. In this case, $\Omega = \{1, \dots, 7\}$. We consider the sequence of partitions (S_n^k), $k \in \{0, 1, 2\}$, where $S_{2n}^0 = \{1, 3, 5, 6\}$, $S_{2n+1}^0 = \{2, 3, 4, 7\}$, $S_{2n}^1 = \{2\}$, $S_{2n+1}^1 = \{1\}$ and $S_{2n}^2 = \{4, 7\}$, $S_{2n+1}^2 = \{5, 6\}$. Any chain driven by $Q_1, Q_0, Q_1, \dots$ will eventually end up staying either in S_n^1 or in S_n^2 forever.

The Decomposition-Separation Theorem is a very general result which holds without any hypothesis on the kernels K_n. We are instead interested in finding hypotheses, perhaps very restrictive ones, on the individual kernels K_n that translate into strong quantitative results concerning the merging property of the chain.

3.3 Products of stochastic matrices. There is a rather rich literature on the study of products of stochastic matrices. Recall that stochastic matrices are matrices with non-negative entries and row sums equal to 1. This last assumption, which breaks the row/column symmetry, implies that there is significant differences between forward and backward products of stochastic matrices. Given a sequence K_i of stochastic matrices The forward products form the sequence

$$K^f_{0,n} = K_1 K_2 \dots K_n, \quad n = 1, \dots,$$

whereas the backward products form the sequence

$$K^b_{0,n} = K_n \dots K_2 K_1, \quad n = 1, \dots.$$

There is a crucial difference between these two sequences: The entries $K^f_{0,n}(x, y)$ do not have any general monotonicity properties but, for any y,

$$n \mapsto M(n, y) = \max_x \{K^b_{0,n}(x, y)\}$$

is monotone non-increasing and

$$n \mapsto m(n, y) = \min_x \{K^b_{0,n}(x, y)\}$$

is monotone non-decreasing. These properties are obvious consequences of the fact that the matrices K_i are stochastic matrices. Of course, both $\lim_{n\to\infty} M(n, y)$ and $\lim_{n\to\infty} m(n, y)$ exist for all y.

If, for some reason, we know that

$$\forall\, x, x', \quad \lim_{n\to\infty} \sum_y |K^b_{0,n}(x, y) - K^b_{0,n}(x', y)| = 0$$

then it follows that the backward products converge to a row-constant matrix Π, i.e.,

$$\forall\, x, x', y, \quad \Pi(x, y) = \lim_{n\to\infty} K^b_{0,n}(x, y), \quad \Pi(x, y) = \Pi(x', y).$$

The references [16], [19], [23], [25], [35], [38] form a sample of old and recent works dealing with this observation.

Changing viewpoint and notation somewhat, consider all finite products of matrices drawn from a set $\mathcal{Q}$ of $N \times N$ stochastic matrices. For $\omega = (\dots, K_{i-1}, K_i, K_{i+1}, \dots) \in \mathcal{Q}^{\mathbb{Z}}$ a doubly infinite sequence of matrices and $m \le n \in \mathbb{Z}$, set

$$K^\omega_{m,n} = K_{m+1} \dots K_n \quad (K_{m,m} = I).$$

A stochastic matrix is called (SIA) if its products converge to a constant row matrix. Here, (SIA) stands for stochastic, irreducible and aperiodic although "irreducible" really means that the matrix has a unique recurrent class (transient states are allowed so that the constant row limit matrix may have some 0 columns). A central result in this area (e.g., [35], [38]) is that, if $\mathcal{Q}$ is finite and all finite products of matrices in $\mathcal{Q}$ are (SIA) then, for any doubly infinite sequence $\omega \in \mathcal{Q}^{\mathbb{Z}}$,

$$\lim_{n-m\to\infty} \sum_y |K^\omega_{m,n}(x, y) - K^\omega_{m,n}(x', y)| = 0 \tag{3.1}$$

and

$$\lim_{m\to-\infty} K^{\omega}_{m,n} = \Pi^{\omega}_n \tag{3.2}$$

where Π^{ω}_n is a row-constant matrix. Let π^{ω}_n be the probability measure corresponding to the rows of row-constant matrix Π^{ω}_n. Observe that (3.1) and (3.2) imply

$$\lim_{n\to\infty} \sum_y |K^{\omega}_{0,n}(x,y) - \pi^{\omega}_n(y)| = 0.$$

The following proposition establishes some relations between these considerations, total variation merging and stability.

Proposition 3.1. *Let $\mathcal{Q}$ be a set of $N \times N$ stochastic matrices. Assume that $\mathcal{Q}$ is merging (in total variation) and c-stable w.r.t. a positive measure π. Then*

(1) *Any finite product P of matrices in $\mathcal{Q}$ is irreducible aperiodic and its unique positive invariant measure π_P satisfies $c^{-1}\pi \le \pi_P \le c\pi$.*

(2) *For any $\omega \in \mathcal{Q}^{\mathbb{Z}}$ and any $n \in \mathbb{Z}$, π^{ω}_n satisfies $c^{-1}\pi \le \pi^{\omega}_n \le c\pi$, i.e., any limit row π' of backward products of matrices in $\mathcal{Q}$ satisfies $c^{-1}\pi \le \pi' \le c\pi$.*

Proof. (1) As $\mathcal{Q}$ is c-stable w.r.t. π, there exists a positive measure μ_0 such that for any finite product P of matrices in $\mathcal{Q}$ and any n, $c^{-1}\pi \le \mu_0 P^n \le c\pi$. Since $\mathcal{Q}$ is merging, we must have $\lim_{n\to\infty} P^n = \Pi_P$ with Π_P having constant rows, call them π_P. This implies $c^{-1}\pi \le \pi_P \le c\pi$. Since π is positive, π_P must be positive and $\lim_{n\to\infty} P^n = \Pi_P$ implies that P is irreducible aperiodic. We note that (1) is, in fact, a sufficient condition for stability. See [29], Proposition 4.9. Under the hypothesis that $\mathcal{Q}$ is merging, (1) is thus a necessary and sufficient condition for c-stability.

(2) Fix $\omega \in \mathcal{Q}^{\mathbb{Z}}$. By hypothesis, on the one hand, there exists a positive probability measure μ_0 such that $c^{-1}\pi \le \mu_0 K^{\omega}_{m,n} \le c\pi$. On the other hand, merging imply that $\lim_{m\to-\infty} K^{\omega}_{m,n} = \Pi^{\omega}_n$ and thus, $\lim_{m\to-\infty} \mu_0 K^{\omega}_{m,n} = \pi^{\omega}_n$. The desired result follows. □

3.4 Product of random stochastic matrices. For pointers to the literature on products of random stochastic matrices and Markov chains in a random environment, see, e.g., [3], [6], [27], [37] and the references therein. We end this section with short comments regarding the simplest case of products of random stochastic matrices, i.e., the case where the matrices K_i form an i.i.d sequence of stochastic matrices. The backward and forward products $K^b_{0,n} = K_n \dots K_1$, $K^f_{0,n} = K_1 \dots K_n$ become random variables taking values in the set of all $N \times N$ stochastic matrices. Although these two sequences of random variables have very different behavior as n varies, $K^b_{0,n}$ and $K^f_{0,n}$ have the same law. Takahashi [37] proves that if

$$\forall\, x, x', \quad \lim_{n\to\infty} \sum_y |K^f_{0,n}(x,y) - K^f_{0,n}(x',y)| = 0 \quad \text{almost surely}$$

then $K_{0,n}^f$ converges in law and the limit law is that of the limit random variable $\lim_{n\to\infty} K_{0,n}^b$. Rosenblatt [27] applies the theory of random walks on semigroups to show that the Cesaro sums $n^{-1}\sum_1^n K_{0,j}^f(x,y)$ always converge to a constant almost surely. The articles [3], [6] discuss similar results under more general hypotheses on the nature of the random sequence $(K_i)_1^\infty$. Unfortunately, these interesting results concerning random environments do not shed much light on the quantitative questions emphasized here.

4 Quantitative results and examples

Informally, the question we want to focus on is the following. Let (K,π) be an irreducible aperiodic Markov kernel and its stationary probability measure. Let $(K_i)_1^\infty$ be a sequence of Markov kernels so that, for each i, K_i is a perturbation of K with invariant measure π_i that is a perturbation of π (what "perturbation" means here is left open on purpose). For an initial distribution μ_0, consider the associated sequence of measures defined by $\mu_n = \mu_0 K_1 \dots K_n$, $n = 1, 2, \dots$.

Problem 4.1. (1) Does total variation merging hold?

(2) Does relative-sup merging hold?

(3) Does there exists $c \geq 1$ such that, for n large enough,

$$\forall\, x, \quad c^{-1} \leq \frac{\mu_n(x)}{\pi(x)} \leq c?$$

Obviously, these questions call for quantitative results describing the merging times, the constant c and the "large" time n in terms of bounds on the allowed perturbations.

To understand what is meant by quantitative results, it is easier to consider a family of problems depending on a parameter representing the size and complexity of the problem. So, one starts with a family (Ω_N, K_N, π_N) of ergodic Markov kernels depending on the parameter N whose mixing time sequence $(T_1(N,\epsilon))_1^\infty$ (say, in total variation) is understood. Then, for each N, we consider perturbations $(K_{N,i})_{i=1}^\infty$ of K_N with stationary measure $\pi_{N,i}$ close to π_N and ask if the merging time of $(K_{N,i})_{i=1}^\infty$ can be controlled in terms of $T_1(N,\epsilon)$.

Problem 4.2. Let $\Omega_N = \{0,\dots,N\}$. Let $\mathcal{Q}_N$ be the set of all birth and death chains Q on V_N with $Q(x, x+\epsilon) \in [1/4, 3/4]$ for all $x, x+\epsilon \in V_N$, $\epsilon \in \{-1,0,1\}$ and with reversible measure π satisfying $1/4 \leq (N+1)\pi(x) \leq 4$, $x \in V_N$.

(1) Prove or disprove that there exists a constant A independent of N such that $\mathcal{Q}_N$ has total variation ϵ-merging time at most $AN^2(1+\log_+ 1/\epsilon)$.

(2) Prove or disprove that there exists a constant A independent of N such that $\mathcal{Q}_N$ has relative-sup ϵ-merging time at most $AN^2(1+\log_+ 1/\epsilon)$.

(3) Prove or disprove that there exist constants $A, C \geq 1$, such that, for any N and any sequence $(K_i)_1^\infty \in \mathcal{Q}_N$, we have

$$\forall\, x, y \in \Omega_N,\ \forall\, n \geq AN^2,\ \ \frac{1}{C(N+1)} \leq K_{0,n}(x,y) \leq \frac{C}{N+1}.$$

Here the time homogeneous model is the birth and death chain K_N with constant rates $p = q = r = 1/3$ and $\pi_N = 1/(N+1)$, so that $K_N(x,y) = 0$ unless $|x-y| \leq 1$, $K(0,0) = K(N,N) = 2/3$ and $K(x,x) = K(x, x\pm 1) = 1/3$ otherwise. Of course, it is well known that $T_1(K_N,\epsilon) \simeq T_\infty(K_N,\epsilon) \simeq N^2(1+\log_+(1/\epsilon))$ for small $\epsilon > 0$. Problem 1.2 asks whether or not these mixing/merging times are stable under suitable time inhomogeneous perturbations of K_N and whether or not the limiting behavior stays comparable to that of the model chain. To the best of our knowledge the answer is not known and this innocent looking problem should be taken seriously.

There appears to be only a small number of papers that attempt to prove quantitative results for time inhomogeneous chains. These include [11], [13], [14], [21], [22] and the authors' works [28], [29], [30], [31]. The works [14], [21], [22], [28] treat only examples of time inhomogeneous chains that admit an invariant measure. Technically, this is a very specific hypothesis and, indeed, these works show that many of the well developed techniques that have been used to study time homogeneous chains can be successfully applied under this hypothesis.

4.1 Singular values. A typical qualitative result about finite Markov chains is that an irreducible aperiodic chain is ergodic. We do not know of any quantitative versions of this statement. Let K be an irreducible aperiodic Markov kernel with stationary measure π so that $\mu_n = \mu_0 K^n \to \pi$ as n tends to infinity, for any starting distribution μ_0.

If (K,π) is reversible (i.e., $\pi(x)K(x,y) = \pi(y)K(y,x)$) and if β denotes the second largest absolute value of the eigenvalues of K acting on $\ell^2(\pi)$ then $\beta < 1$ and

$$2\|\mu_n - \pi\|_{\mathrm{TV}} \leq \|\mu_0/\pi\|_2 \beta^n \tag{4.1}$$

where $\|\mu_0/\pi\|_2$ is the norm of $f_0 = \mu_0/\pi$ in $\ell^2(\pi)$. This can be considered as a quantitative result although it involves the perhaps unknown reversible measure π.

If (K,π) is not reversible, the inequality still holds with β being the second largest singular value of K on $\ell^2(\pi)$ (i.e., the square root of the second largest eigenvalue of KK^* where K^* is the adjoint of K on $\ell^2(\pi)$). However, it is then possible that $\beta = 1$, in which case the inequality fails to capture the qualitative ergodicity of the chain.

Inequality (4.1) has an elegant generalization to the time inhomogeneous setting. Let $(K_i)_1^\infty$ be a sequence of irreducible Markov kernels (on a finite state space). Fix a positive probability measure μ_0 (by positive we mean here that $\mu_0(x) > 0$ for all x) and set

$$\mu_n = \mu_0 K_{0,n}.$$

In the time inhomogeneous setting, we want to compare this sequence of measures $(\mu_n)_1^\infty$ to the sequence of measures $(K_{0,n}(x,\cdot))_1^\infty$ describing the distribution at time n of the chain started at an arbitrary point x.

To state the result, for each i, consider K_i as a linear operator acting from $\ell^2(\mu_i)$ to $\ell^2(\mu_{i-1})$. One easily checks that this operator is a contraction. Its singular values are the square roots of the eigenvalues of the operator $P_i = K_i^* K_i : \ell^2(\mu_i) \to \ell^2(\mu_i)$ where $K_i^* : \ell^2(\mu_{i-1}) \to \ell^2(\mu_i)$ is the adjoint operator which is a Markov operator with kernel

$$K_i^*(x, y) = \frac{K_i(y, x)\mu_{i-1}(y)}{\mu_i(x)}.$$

We let

$$\sigma_i = \sigma(K_i, \mu_i, \mu_{i-1})$$

be the second largest singular value of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$. It is the square root of the second largest eigenvalue of the Markov kernel

$$P_i(x, y) = \frac{1}{\mu_i(x)} \sum_z K_i(z, x) K_i(z, y) \mu_{i-1}(z). \tag{4.2}$$

Theorem 4.3. *With the notation introduced above, we have*

$$\|K_{0,n}(x, \cdot) - \mu_n\|_{\mathrm{TV}} \leq \mu_0(x)^{-1/2} \prod_1^n \sigma_i$$

and

$$\left| \frac{K_{0,n}(x, y)}{\mu_n(y)} - 1 \right| \leq [\mu_0(x)\mu_n(y)]^{-1/2} \prod_1^n \sigma_i$$

For the proof, see [11], [29]. The proofs given in [11] and [29] are rather different in spirit, with [11] avoiding the explicit use of singular values. Introducing singular values allows for further refinements and is useful for practical estimates. See [28], [29]. When coupled with the hypothesis of c-stability, the above result becomes a powerful and very applicable tool. See, e.g., [29], Theorem 4.11, and the examples treated in [29], [30]. Unfortunately, proving c-stability is not an easy task.

A good example of application of Theorem 4.3 is the following result taken from [29]. We refer the reader to [29] for the proof.

Theorem 4.4. *Fix $1 < a < A < \infty$. Let $\mathcal{Q}_N(a, A)$ be the set of all constant rate birth an death chains on $\{0, \dots, N\}$ with parameters p, q, r satisfying $p/q \in [a, A]$. The set $\mathcal{Q}_N(a, A)$ is merging in relative-sup with relative-sup ϵ-merging time bounded above by*

$$T_\infty(\epsilon) \leq C(a, A)(N + \log_+ 1/\epsilon).$$

In contrast, note that the set $\mathcal{Q} = \{Q_1, Q_2\}$ where Q_i is the p_i, q_i constant rate birth and death chain on $\{0, \dots, N\}$ and $p_1 = q_2, q_1 = p_2$ cannot be merging faster than N^2 because the product $K = Q_1 Q_2$ is, essentially, a simple random walk on a circle with almost uniform invariant measure. See [29], Example 2.17.

It may be illuminating to point out that Theorem 4.3 is of some interest even in the time homogeneous case. Suppose K is irreducible aperiodic kernel with stationary measure π and second largest singular value σ on $\ell^2(\pi)$. Then we have

$$\left|\frac{K^n(x,y)}{\pi(y)} - 1\right| \le [\pi(x)\pi(y)]^{-1/2}\sigma^n. \tag{4.3}$$

One difficulty attached to this estimate is that both $[\pi(x)\pi(y)]^{-1/2}$ and σ depends on the perhaps unknown stationary measure π.

Consider instead an initial measure $\mu_0 > 0$ and set $\mu_n = \mu_0 K^n$. Then we also have

$$\left|\frac{K^n(x,y)}{\mu_n(y)} - 1\right| \le [\mu_0(x)\mu_n(y)]^{-1/2}\prod_1^n \sigma_i \tag{4.4}$$

where σ_i is the second largest singular value of $K\colon \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$. In particular, setting $\mu_0^* = \min_x\{\mu_0(x)\}$,

$$\left|\frac{\pi(y)}{\mu_n(y)} - 1\right| \le [\mu_0^*\mu_n(y)]^{-1/2}\prod_1^n \sigma_i. \tag{4.5}$$

The estimates (4.4)–(4.5) have the disadvantage that each σ_i depends on μ_0 through μ_{i-1} and μ_i. They have the advantage that they do not depend in any direct way of π. From a computational viewpoint, they offer a dynamical estimate of the error in the approximation of π by μ_n.

4.2 An example where stability fails. In this section, we present a simple example that indicates why stability is a difficult property to study from a quantitative viewpoint. Let $\Omega_N = \{0, 1, \ldots, N\}$, $N = 2n + 1$. Fix $p, q, r \ge 0$ with $p + q + r = 1$, $p \ne q$, and $\eta_1 \in [0, 1)$. Consider the Markov kernels Q_1 given by

$$\begin{aligned} Q_1(2x, 2x+1) &= p, \quad x = 0, \ldots, n,\\ Q_1(2x, 2x-1) &= q, \quad x = 1, \ldots, n,\\ Q_1(2x-1, 2x) &= q, \quad x = 1, \ldots, n,\\ Q_1(2x+1, 2x) &= p, \quad x = 0, \ldots, n-1,\\ Q_1(x, x) &= r, \quad x = 1, \ldots, 2n, \end{aligned}$$

and

$$Q_1(0,0) = q + r, \quad Q_1(N,N) = \eta_1, \quad Q_1(N, N-1) = 1 - \eta_1.$$

This chain has reversible measure π_1 given by

$$\pi_1(0) = \cdots = \pi_1(N-1) = (1-\eta_1)p^{-1}\pi_1(N) = \frac{(1-\eta_1)p^{-1}}{N(1-\eta_1)p^{-1} + 1}.$$

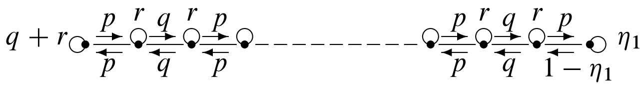

Figure 4. The chain with kernel Q_1.

Next, we let Q_2 be the kernel obtained by exchanging the roles of p and q and replacing η_1 by $\eta_2 \in [0, 1)$. Obviously, this kernel has reversible measure π_2 given by

$$\pi_2(0) = \cdots = \pi_2(N-1) = (1-\eta_2)q^{-1}\pi_2(N) = \frac{(1-\eta_2)q^{-1}}{N(1-\eta_2)q^{-1}+1}.$$

As long as p, q are bounded away from 0 and 1 and η_1, η_2 are bounded away from 1 these kernels Q_1, Q_2 can be viewed as perturbations of the simple random walk on a stick (with loops at the ends). Their respective invariant measures are close to uniform. In fact, they are uniform if $\eta_1 = q + r$, $\eta_2 = p + r$.

It is clear that, even if $r\eta_1\eta_2 = 0$, for any sequence $(K_i)_1^\infty$ with $K_i \in \{Q_1, Q_2\}$ we have

$$\min_{x,y\in\Omega_N} \{K_{m,m+2N+1}(x, y)\} \geq (\min\{p, q\})^{2N+1} > 0.$$

Hence, if we let $\mu_0 = u$ be the uniform measure and set $\mu_n = \mu_0 K_{0,n}$ then there exists a constant $c = c(p, q, N) \in (1, \infty)$ such that

$$\forall\, n, \quad c^{-1} \leq \mu_n(x) \leq c.$$

Further, it follows that any such sequence $(K_i)_1^\infty$ is merging in total variation and in relative-sup.

Nevertheless, we are going to show that the stability property fails at the quantitative level as N tends to infinity. For this purpose, we compute the kernel of $K = Q_1 Q_2$. To understand K, it is useful to imagine that the elements of $\{0, \ldots, N\}$ arranged on a circle with the even points in the upper half of the circle and the odd points on the lower half of the circle. The only points on the horizontal diameter of the circle are 0 and N.

The kernel K is given by the formulae:

$$\begin{aligned}
K(2x, 2x+2) &= p^2,\ K(2x+2, 2x) = q^2, && x = 0, \ldots n-2,\\
K(2x+1, 2x+3) &= q^2,\ K(2x+3, 2x+1) = p^2, && x = 0, \ldots, n-2,\\
K(0,0) &= 2pq + r,\quad K(x,x) = 2pq + r^2, && x = 1, \ldots, N-2,\\
K(x, x+1) &= K(x+1, x) = r(p+q) && x = 1, \ldots, N-2,\\
K(0,1) &= q^2 + r(1-r),\ K(1,0) = p^2 + r(1-r),\\
K(N-1, N) &= p\eta_2 + rq,\\
K(N, N-1) &= (1-\eta_2)\eta_1 + (1-\eta_1)r,\\
K(N-2, N) &= q^2,\ K(N, N-2) = (1-\eta_1)p,
\end{aligned}$$

$$K(N-1,N-1) = p(q+1-\eta_2)+r^2,$$
$$K(N,N) = \eta_1\eta_2+(1-\eta_1)q.$$

The following special cases are of interest.

(i) $r=0$, $\eta_1 = q, \eta_2 = p$. In this case $\pi_1 = \pi_2$ is uniform and K is the kernel of a nearest-neighbors random walk on the circle with transition probabilities p^2, q^2 and holding $2pq$. Of course, this chain admits the uniform measure as invariant measure.

(ii) $r=0$, $\eta_1=\eta_2=0$. In this case, K is essentially the kernel of a $p'=p^2, q'=q^2, r'=2pq$ birth and death chain. More precisely, after writing $x_0=N, x_1=N-2,\dots,x_{n-1}=1, x_n=0, x_{n+1}=2,\dots,x_{N-1}=N-3, x_N=N-1$, we have

$$K(x_i,x_{i+1}) = p^2,\quad K(x_i,x_{i-1}) = q^2,\quad K(x_i,x_i)=2pq$$

except for $K(x_0,x_1)=p$, $K(x_0,x_0)=q$, $K(x_N,x_N)=p+pq$. This chain has invariant measure

$$\pi(x_i) = \pi(x_0)p^{-1}(p/q)^{2i},\quad i=1,\dots,N.$$

Using the same notation as in (ii) above, we can compute the invariant measure π of K when $r=0$ for arbitrary values of η_1, η_2. Indeed, π must satisfy the following equations:

$$\begin{aligned}
\pi(x_i) &= 2pq\pi(x_i)+p^2\pi(x_{i-1})+q^2\pi_i(x_{i+1}),\quad i=2,\dots,N-1,\\
\pi(x_1) &= 2pq\pi(x_1)+(1-\eta_1)p\pi(x_0)+q^2\pi(x_2),\\
\pi(x_0) &= (\eta_1\eta_2+(1-\eta_1)q)\pi(x_0)+q^2\pi(x_1)+p\eta_2\pi(x_N),\\
\pi(x_N) &= p(q+1-\eta_2)\pi(x_N)+(1-\eta_2)\eta_1\pi(x_0)+p^2\pi(x_{N-1}).
\end{aligned}$$

Because of the first equation, we set $\pi(x_i)=\alpha+\beta(p/q)^{2i}$ for $i=1,\dots,N$. This gives

$$\begin{aligned}
(1-\eta_1)p\pi(x_0) &= (\beta+\alpha)p^2,\\
(p-\eta_1(\eta_2-q))\pi(x_0) &= q^2(\alpha+\beta(p/q)^2)+p\eta_2(\alpha+\beta(p/q)^{2N}),\\
(1-\eta_2)\eta_1\pi(x_0) &= \alpha(q^2+p(\eta_2-p))+p\eta_2\beta(p/q)^{2N}.
\end{aligned}$$

Since the equations of the system $\pi=\pi K$ are not independent, the three equations above are not either. Indeed, subtracting the last equation from the second yields the first. So the previous system is equivalent to

$$\begin{aligned}
(1-\eta_1)p^{-1}\pi(x_0) &= \beta+\alpha,\\
(1-\eta_2)\eta_1\pi(x_0) &= \alpha(q^2+p(\eta_2-p))+p\eta_2\beta(p/q)^{2N}.
\end{aligned}$$

Hence, recalling that $q^2 - p^2 = q - p$ since $p + q = 1$,

$$\beta = \frac{(1-\eta_1)(q/p) - (1-\eta_2)}{q - p + p\eta_2(1-(p/q)^{2N})}\pi(x_0)$$

and

$$\alpha = \frac{(1-\eta_2)\eta_1 - (1-\eta_1)\eta_2(p/q)^{2N}}{q - p + p\eta_2(1-(p/q)^{2N})}\pi(x_0).$$

When $\eta_1 = \eta_2 = 0$ (resp. $\eta_1 = q, \eta_2 = p$), we recover $\alpha = 0$, $\beta = p^{-1}\pi(x_0)$ (resp. $\alpha = \pi(x_0)$, $\beta = 0$).

The denominator $q - p + p\eta_2(1-(p/q)^{2N})$ is positive or negative depending on whether $q > p$ or $q < p$. By inspection of these formulae, one easily proves the following facts (the notation x_i refers to the relabelling of the state space introduced in (ii) above).

- Assume that $q > p$, $r = 0$. For any fixed $\eta_1 > 0$, there is a constant $c = c(p, q, \eta_1, \eta_2) \in (1, \infty)$ such that, for all large enough N, we have

  $$\forall\, x,\ \ c^{-1} \le (N+1)\pi(x) \le c.$$

  If $\eta_1 = 0$ then there is a constant $c = c(p, q, \eta_2) \in (1, \infty)$ such that, for all large enough N, we have

  $$\forall\, x_i,\ \ c^{-1} \le (q/p)^{2i}\pi(x_i) \le c.$$

- Assume that $q < p$, $r = 0$. For any fixed $\eta_2 > 0$, there is a constant $c = c(p, q, \eta_1, \eta_2) \in (1, \infty)$ such that, for all large enough N, we have

  $$\forall\, x,\ \ c^{-1} \le (N+1)\pi(x) \le c.$$

  If $\eta_2 = 0$ then there is a constant $c = c(p, q, \eta_1) \in (1, \infty)$ such that, for all large enough N, we have

  $$\forall\, x_i,\ \ c^{-1} \le (q/p)^{2(i-N)}\pi(x_i) \le c.$$

On the one hand, when $r = \eta_1 = \eta_2 = 0$ and $0 < p \neq q < 1$ are fixed, there are no constants c independent of N for which the set $\mathcal{Q} = \{Q_1, Q_2\}$ is c-stable. One can even take p_N, q_N so that $p_N/q_N = 1 + aN^{-\alpha} + o(N^{-1})$ as N tends to infinity with $a > 0$ and $0 < \alpha < 1$. Then Q_1 and Q_2 are asymptotically equal but there are no constants c independent of N for which $\mathcal{Q} = \{Q_1, Q_2\}$ is c-stable.

On the other hand, when $0 < p, q, r < 1$, $\eta_1 = q + r$ and $\eta_2 = p + r$, the uniform measure is invariant for both kernels and $\mathcal{Q}$ is 1-stable.

It seems likely that for fixed η_1, η_2, r, p, q with $0 < p, q < 1$ and either $r > 0$ or $\eta_1\eta_2 > 0$ the set $\mathcal{Q}$ is c-stable but we do not know how to prove that.

5 Time dependent edge weights

In this section, we consider a family of graphs $\mathcal{G}_N = (\Omega_N, E_N)$. These graphs are non-oriented with no multiple edges (edges are pairs of vertices $e = \{x, y\}$ or singletons $e = \{x\}$). We assume connectedness. We let $d(x)$ be the degree of x, i.e., $d(x) = \#\{e \in E : e \ni x\}$ and set

$$\delta(x) = \frac{d(x)}{\sum_x d(x)}.$$

For simplicity, we assume that these graphs have bounded degree, i.e.,

$$\forall\, N, \quad \forall\, x \in \Omega_N, \quad d(x) \le D,$$

uniformly in N. A simple example is the lazy stick of length $(N+1)$ as in Problem 4.2 and Figure 5.

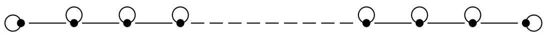

Figure 5. The lazy stick.

5.1 Adapted kernels. For any choice of positive weights $\mathbf{w} = (w_e)_{e \in E}$ on $\mathcal{G}_n$, we obtain a reversible Markov kernel $K(\mathbf{w})$ with support on pairs (x, y) such that $\{x, y\} \in E$, in which case

$$K(\mathbf{w})(x, y) = \frac{w_{\{x,y\}}}{\sum_{e \ni x} w_e}.$$

The reversible measure is

$$\pi(\mathbf{w})(x) = c(\mathbf{w})^{-1} \sum_{e \ni x} w_e, \quad c(\mathbf{w}) = \sum_x \sum_{e \ni x} w_e.$$

For instance, picking $\mathbf{w} = \mathbf{1}$, i.e., $w_e = 1$ for all $e \in E$, we obtain the kernel $K_{\text{sr}}(x, y) = K(\mathbf{1})(x, y) = 1_E(\{x, y\})/d(x)$ of the simple random walk on the given graph. The reversible measure for K_{sr} is $\pi(\mathbf{1}) = \delta$.

Set

$$R(\mathbf{w}) = \max\left\{w_e/w_{e'} : e, e' \in E\right\}.$$

Observe that $R(\mathbf{w}) \le b$ implies

$$\forall\, x, \quad b^{-1}\delta(x) \le \pi(\mathbf{w})(x) \le b\delta(x). \tag{5.1}$$

For instance, to prove the upper bound, let $w_0 = \min\{w_e\}$ and write

$$\pi(\mathbf{w})(x) = c(\mathbf{w})^{-1} \sum_{e \ni x} w_e \le \frac{1}{\sum_x d(x)} \sum_{e \ni x} \frac{w_e}{w_0} \le b\delta(x).$$

The proof of the lower bound is similar. Further, we also have

$$\forall\, x, y, \quad (Db)^{-1}\pi(\mathbf{w})(y) \le \pi(\mathbf{w})(x) \le Db\pi(\mathbf{w})(y). \tag{5.2}$$

Indeed, $\sum_{e\ni x} w_e \le Dbw_0 \le Db\sum_{e\ni y} w_e$.

For any N and $b > 1$, set

$$\mathcal{Q}(\mathcal{G}_N, b) = \{K(\mathbf{w}) : R(\mathbf{w}) \le b\}.$$

For any $N, b > 1$ and fixed probability measure π on Ω_N, set

$$\mathcal{Q}(\mathcal{G}_N, b, \pi) = \{K(\mathbf{w}) : R(\mathbf{w}) \le b,\ \pi(\mathbf{w}) = \pi\}.$$

The set of weight $\mathcal{Q}(\mathcal{G}_N, b, \pi)$ may well be empty. However, we can use the Metropolis algorithm construction to prove the following lemma.

Lemma 5.1. *Assume that $\{x\} \in E$ for all x (i.e, the graphs $\mathcal{G}_N$ have a loop at each vertex) and that $a^{-1} \le \pi(x)/\delta(x) \le a$. Then the set $\mathcal{Q}(\mathcal{G}_N, a^2(b^3 + bD), \pi)$ is non-empty for any $b \ge 1$. It contains a continuum of kernels $K(\mathbf{w})$ for any $b > 1$.*

Proof. Starting from any weight $\mathbf{v}$ with $R(\mathbf{v}) \le b$, we define a new weight $\mathbf{w}$ by setting

$$\forall\, \{x,y\} \in E, x \ne y,\quad w_{\{x,y\}} = v_{\{x,y\}} \min\left\{\frac{\pi(x)}{\pi(\mathbf{v})(x)}, \frac{\pi(y)}{\pi(\mathbf{v})(y)}\right\}$$

and

$$w_{\{x\}} = c(\mathbf{v})\pi(x) - \sum_{y\ne x} v_{\{x,y\}} \min\left\{\frac{\pi(x)}{\pi(\mathbf{v})(x)}, \frac{\pi(y)}{\pi(\mathbf{v})(y)}\right\}.$$

It is clear that $\pi(\mathbf{w}) = \pi$ (Indeed, $K(\mathbf{w})$ is the kernel of the Metropolis algorithm chain for π with proposal based on $K(\mathbf{v})$). Further, since

$$\sum_{y\ne x} v_{\{x,y\}} \min\left\{\frac{\pi(x)}{\pi(\mathbf{v})(x)}, \frac{\pi(y)}{\pi(\mathbf{v})(y)}\right\} \le \pi(x)\left(c(\mathbf{v}) - \frac{v_{\{x\}}}{\pi(\mathbf{v})(x)}\right),$$

we have

$$\frac{\pi(x)v_{\{x\}}}{\pi(\mathbf{v})(x)} \le w_{\{x\}} \le c(\mathbf{v})\pi(x).$$

Now, since $a^{-1}\delta(x) \le \pi(x) \le a\delta(x)$ and $\mathbf{v} \in \mathcal{Q}(\mathcal{G}_N, b)$, we obtain

$$\forall\, x \ne y, x' \ne y',\quad \frac{w_{\{x,y\}}}{w_{\{x',y'\}}} \le b^3a^2.$$

and

$$\forall\, \{x,y\} \in E, x',\quad \max\left\{\frac{w_{\{x,y\}}}{w_{\{x'\}}}, \frac{w_{\{x'\}}}{w_{\{x,y\}}}\right\} \le a^2bD.$$

Hence $R(\mathbf{w}) \le a^2(b^3 + bD)$ and $K(\mathbf{w}) \in \mathcal{Q}(\mathcal{G}_N, a^2(b^3 + bD), \pi)$ as desired. □

5.2 Time homogeneous results. For each N, let σ_N be the second singular value of (K_{sr}, δ), i.e., the second largest eigenvalue in absolute value of the simple random walk on $\mathcal{G}_N$. For instance, for the "lazy stick" of Figure 5, $1-\sigma_N$ is of order $1/N^2$. For any $\mathbf{w}$, let $\sigma(\mathbf{w})$ be the second largest singular value of $(K(\mathbf{w}), \pi(\mathbf{w}))$. The following lemma concerns the time homogeneous chains associated with kernels in $\mathcal{Q}(\mathcal{G}_N, b)$.

Proposition 5.2. *For any $b \geq 1$ and any $K(\mathbf{w}) \in \mathcal{Q}(\mathcal{G}_N, b)$,*

$$b^{-2}(1-\sigma_N) \leq 1-\sigma(\mathbf{w}).$$

In particular, uniformly over $w \in \mathcal{Q}(\mathcal{G}_N, b)$,

$$\left|\frac{K(\mathbf{w})^n(x,y)}{\pi(\mathbf{w})(y)} - 1\right| \leq b d_*^{-1} \Delta_N (1-b^{-2}(1-\sigma_N))^n, \tag{5.3}$$

with $\Delta_N = \sum_x d(x)$, $d_* = \min_x\{d(x)\}$.

Proof. This is based on the basic comparison techniques of [7]. In the present case, it is best to compare the lowest and second largest eigenvalues of K_{sr}, call them β_- and β_1, respectively, with the same quantities $\beta_-(\mathbf{w})$ and $\beta_1(\mathbf{w})$ relative to $K(\mathbf{w})$. The relation with the singular value $\sigma(\mathbf{w})$ is given by $\sigma(\mathbf{w}) = \max\{-\beta_-(\mathbf{w}), \beta_1(\mathbf{w})\}$. For comparison purpose, one uses the Dirichlet forms (recall that edges here are (non-oriented) pairs $\{x,y\}$)

$$\mathcal{E}_{\mathbf{w}}(f,f) = \frac{1}{c(\mathbf{w})} \sum_{e=\{x,y\}} |f(x)-f(y)|^2 w_e$$

and

$$\mathcal{E}_{\text{sr}}(f,f) = \mathcal{E}_{\mathbf{1}}(f,f) = \frac{1}{\Delta_N} \sum_{e=\{x,y\}} |f(x)-f(y)|^2.$$

Clearly, for any f,

$$\mathcal{E}_{\text{sr}}(f,f) \leq \frac{c(\mathbf{w})b}{\Delta_N}\mathcal{E}_{\mathbf{w}}(f,f), \quad \operatorname{Var}_{\pi(\mathbf{w})}(f) \leq \frac{\Delta_N b}{c(\mathbf{w})}\operatorname{Var}_{\delta}(f). \tag{5.4}$$

This yields $1-\beta_1 \leq b^2(1-\beta_1(\mathbf{w}))$. A similar argument using (the sum here is over all x, y with $\{x,y\} \in E$, which explains the $\frac{1}{2}$ factor)

$$\mathcal{F}_{\mathbf{w}}(f,f) = \frac{1}{2c(\mathbf{w})} \sum_{x,y:\{x,y\}\in E} |f(x)+f(y)|^2 w_{\{x,y\}}$$

yields $1+\beta_- \leq b^2(1+\beta_-(\mathbf{w}))$. This gives the desired result. □

Example 5.3. For our present purpose, call "(d,ϵ)-expander family" any infinite family of regular graphs $\mathcal{G}_N$ of fixed degree d, with $|\Omega_N| = \#\Omega_N$ tending to infinity with N and satisfying $\sigma_N \leq 1-\epsilon$. See [17], [20] for various related definitions and discussions of particular examples. Proposition 5.2 shows that for any $K(\mathbf{w}) \in \mathcal{Q}(\mathcal{G}_N, b)$, we have

$$\left|\frac{K(\mathbf{w})^n(x,y)}{\pi(\mathbf{w})(y)} - 1\right| \leq b|\Omega_N|(1-\epsilon/b^2)^n,$$

Let us point out that, beside singular values , there are further related techniques that yield complementary results. They include the use of Nash and logarithmic Sobolev inequalities (modified or not). See [8], [9], [28], [30]. For instance, to show that on the "lazy stick" $\mathcal{G}_N$ of Figure 5, any chains with kernel in $\mathcal{Q}(\mathcal{G}_N, b)$ converges to stationarity in order N^2, one uses the Nash inequality technique of [8].

5.3 Time inhomogeneous chains. A fundamental question about time inhomogeneous Markov chains is whether or not a result similar to (5.3) holds true for time inhomogeneous chains with kernels in $\mathcal{Q}_N(\mathcal{G}_N, b)$. Little is known about this.

Fix $b > 1$. Let $(K_i)_1^\infty$ be a sequence of Markov kernels in $\mathcal{Q}(\mathcal{G}_N, b)$ and $K_{m,n}$ be the associated iterated kernel. Recall that the property "$\sigma_N < 1$" is equivalent to the irreducibility and aperiodicity of K_{sr}. Because all the kernels in $\mathcal{Q}(\mathcal{G}_N, b)$ are (uniformly) adapted to the graph structure $\mathcal{G}_N$, there exists $\ell = \ell(N, b)$ and $\epsilon = \epsilon(N, b) > 0$ such that, for all n, $K_{n,n+\ell}(x, y) \geq \epsilon$. As explained in Section 2.3, this implies relative-sup merging for any such time inhomogeneous chain. However, this result is purely qualitative. No acceptable quantitative result can be obtain by such an argument.

Problem 5.4. Fix reals $D, b > 1$. Prove or disprove that there exists a constant A such that for any family $\mathcal{G}_N$ with maximal degree at most D, any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}(\mathcal{G}_N, b)$, any initial distributions μ_0, μ_0' and any $\epsilon > 0$, if

$$n \geq A(1 - \sigma_N)^{-1}(\log |\Omega_N| + \log_+(1/\epsilon))$$

then $\mu_n = \mu_0 K_{0,n}$ and $\mu_n' = \mu_0' K_{0,n}$ satisfy

$$\max_{x \in \Omega_N} \left\{ \left| \frac{\mu_n'(x)}{\mu_n(x)} - 1 \right| \right\} \leq \epsilon.$$

This is an open problem, even for the "lazy stick" of Figure 5. It seems rather unclear whether one should except a positive answer or not.

Next, we consider another question, quite interesting but, a priori, of a different nature. Recall that, given $\mathcal{G}_N$, δ denotes the normalized reversible measure of K_{sr}.

Problem 5.5. Fix reals $D, b > 1$. Prove or disprove that there exists a constant $A \geq 1$ such that for any family $\mathcal{G}_N$ with maximal degree at most D, any sequence $(K_i)_1^\infty$with $K_i \in \mathcal{Q}(\mathcal{G}_N, b)$ and any initial distributions μ_0, if

$$n \geq A(1 - \sigma_N)^{-1}(\log |\Omega_N|)$$

then $\mu_n = \mu_0 K_{0,n}$ satisfies

$$\forall\, x \in \Omega_N, \quad A^{-1} \leq \frac{\mu_n(x)}{\delta(x)} \leq A.$$

In words, a positive solution to Problem 5.4 yields the relative-sup merging in time of order at most $A(1-\sigma_N)^{-1}\log|\Omega_N|$, uniformly for any time inhomogeneous chain with kernels in $\mathcal{Q}(\mathcal{G}_N, b)$ whereas a positive solution to Problem 5.5 would indicate that, after a time of order at most $A(1-\sigma_N)^{-1}\log|\Omega_N|$, uniformly for any time inhomogeneous chain with kernels in $\mathcal{Q}(\mathcal{G}_N, b)$ and for any initial distribution μ_0, the measure $\mu_n = \mu_0 K_{0,n}$ is comparable to δ. In fact, because of the uniform way in which Problem 5.5 is formulated, a positive answer implies that the measure δ is A-stable for $\mathcal{Q}(\mathcal{G}_N, b)$.

At this writing, the best evidence for a positive answer to these problems is contained in the following two partial results. The first result concerns sequences whose kernels share the same invariant distribution. For the proof, see [28].

Theorem 5.6. *Fix reals $D, b > 1$ and measures π_N on Ω_N. Assume that $\mathcal{G}_N$ has maximal degree at most D and that $\mathcal{Q}(\mathcal{G}_N, b, \pi_N)$ is non-empty. Under these circumstances, there is a constant $A = A(D, b)$ such that for any $\epsilon > 0$, any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}(\mathcal{G}_N, b, \pi_N)$ and any pair μ_0, μ_0' of initial distributions, if*

$$n \geq A(1-\sigma_N)^{-1}(\log|\Omega_n| + \log_+(1/\epsilon))$$

then $\mu_n = \mu_0 K_{0,n}$ and $\mu_n' = \mu_0' K_{0,n}$ satisfy

$$\max_{x\in\Omega_N}\left\{\left|\frac{\mu_n'(x)}{\mu_n(x)} - 1\right|\right\} \leq \epsilon.$$

Note the hypothesis that $\mathcal{Q}(\mathcal{G}_N, b, \pi_N)$ is non-empty implies that $b^{-1} \leq \pi_N/\delta \leq b$. The second result assumes c-stability. For the proof, see [30].

Theorem 5.7. *Fix reals $D, b, c > 1$. Assume that $\mathcal{G}_N$ has maximal degree at most D. Let $(K_i)_1^\infty$ be a sequence of kernels on Ω_N with $K_i \in \mathcal{Q}(\mathcal{G}_N, b)$. Assume that the distribution δ on Ω_N is c-stable for $(K_i)_1^\infty$. Then there exists a constant $A = A(D, b, c)$ such that for any $\epsilon > 0$ and pair μ_0, μ_0' of initial distributions, if*

$$n \geq A(1-\sigma_N)^{-1}(\log|\Omega_n| + \log_+(1/\epsilon))$$

then $\mu_n = \mu_0 K_{0,n}$ and $\mu_n' = \mu_0' K_{0,n}$ satisfy

$$\max_{x\in\Omega_N}\left\{\left|\frac{\mu_n'(x)}{\mu_n(x)} - 1\right|\right\} \leq \epsilon.$$

Theorem 5.6 can be viewed as a special case of Theorem 5.7. Indeed, if $\mathcal{Q}(\mathcal{G}_N, b, \pi_N)$ is not empty then we must have $b^{-1}\delta \leq \pi_N \leq b^1\delta$ so that δ is a b-stable measure for any sequence of kernels in $\mathcal{Q}(\mathcal{G}_N, b, \pi_N)$. By Lemma 5.1, it is not difficult to produce examples where Theorem 5.6 applies. Finding examples of application of Theorem 5.7 (where the K_i's do not all share the same invariant distribution) is a difficult problem.

Under the stability hypothesis of Theorem 5.7, methods such as Nash inequalities and logarithmic Sobolev inequality can also be applied. See [30].

Remark 5.8. Consider the kernels Q_1, Q_2 of Section 4.2, with fixed p, q, r, η_1, η_2 with $r = \eta_1 = \eta_2 = 0$ and $0 < p \neq q < 1$. The kernels Q_1, Q_2 are adapted to the graph structure of Figure 6. We proved in Section 4.2 that stability fails for

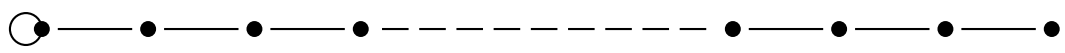

Figure 6. The underlying graph for the kernels Q_1, Q_2 of Section 4.2.

$\mathcal{Q} = \{Q_1, Q_2\}$. Even on the "lazy stick" of Figure 5, we do not understand whether stability holds or not. An interesting example of stability on the lazy stick is proved in [29]. This example involves perturbations that are localized at the ends of the stick. Further examples are discussed in [31].

References

[1] D. Blackwell, Finite non-homogeneous chains. *Ann. of Math.* **46** (1945), 594–599. 128, 129, 134

[2] E. Borel, Sur le battage des cartes. *Oeuvres de Émile Borel*, Tome II, Centre National de la Recherche Scientifique, Paris, 1972. 128, 133

[3] R. Cogburn, The ergodic theory of Markov chains in random environments. *Z. Wahrsch. Verw. Gebiete* **22** (1984), 109–128. 136, 137

[4] H. Cohn, Finite non-homogeneous Markov chains: asymptotic behaviour. *Adv. in Appl. Probab.* **8** (1976), 502–516. 128, 129, 134

[5] H. Cohn, On a paper by Doeblin on non-homogeneous Markov chains. *Adv. in Appl. Probab.* **13** (1981), 388–401. 133

[6] J. Cohen, Ergodic theorems in demography. *Bull. Amer. Math. Soc. (N.S.)* **1** (1979), 275–295. 129, 136, 137

[7] P. Diaconis and L. Saloff-Coste, Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** (1993), 696–730. 146

[8] P. Diaconis and L. Saloff-Coste, Nash inequalities for Finite Markov chains. *J. Theoret. Probab.* **9** (1996), 459–510. 147

[9] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** (1996), 695–750. 147

[10] P. Del Moral and A. Guionnet, On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.* **37** (2001), 155–194. 127, 133

[11] P. Del Moral, M. Ledoux, and L. Miclo, On contraction properties of Markov kernels. *Probab. Theory Related Fields* **126** (2003), 395–420. 133, 138, 139

[12] P. Del Moral and L. Miclo, Dynamiques recuites de type Feynman-Kac: résultats précis et conjectures. *ESAIM Probab. Stat.* **10** (2006), 76–140. 127

[13] R. Douc, E. Moulines, and J. S. Rosenthal, Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.* **14** (2004), 1643–1665. 132, 133, 138

[14] M. Ganapathy, Robust mixing. In *Approximation, randomization and combinatorial optimization*, Proceedings of RANDOM'06, Lecture Notes in Comput. Sci. 4110, Springer-Verlag, Berlin 2006, 351–362. 128, 138

[15] D. Griffeath, Uniform coupling of non-homogeneous Markov chains. *J. Appl. Probab.* **12** (1975), 753–762. 129, 133

[16] J. Hajnal, Weak ergodicity in non-homogeneous Markov chains. *Proc. Cambridge Philos. Soc.* **54** (1958), 233–246. 129, 133, 135

[17] S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their applications. *Bull. Amer. Math. Soc.* **43** (2006), 439–561. 146

[18] M. Iosifescu, *Finite Markov processes and their applications*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Ltd., Chichester; Editura Tehnică, Bucharest 1980. 129, 133

[19] A. N. Kolmogoroff, Zur Theorie der Markoffchen Ketten. *Math. Ann.* **112** (1936), 155–160. 133, 134, 135

[20] A. Lubotzky, Cayley graphs: eigenvalues, expanders and random walks. In *Surveys in combinatorics, 1995*, London Math. Soc. Lecture Note Ser. 218, Cambridge University Press, Cambridge 1995, 155–189. 146

[21] I. Mironov, (Not so) random shuffles of RC4. In *Advances in cryptology – WCRYPTO 2002*, Lecture Notes in Comput. Sci. 2442, Springer-Verlag, Berlin 2002, 304–319. 128, 138

[22] E. Mossel, Y. Peres, and A. Sinclair, Shuffling by semi-random transpositions. *Proceedings of the 45th annual IEEE symposium on foundations of computer science* (FOCS'04), IEEE Computer Society Washington, DC, 2004, 572–581. arXiv:math.PR/0404438 128, 138

[23] M. Neumann and H. Schneider, The convergence of general products of matrices and the weak ergodicity of Markov chains. *Linear Algebra Appl.* **287** (1999), 307–314. 133, 135

[24] U. Păun, Ergodic theorems for finite Markov chains. *Math. Rep.* (*Bucur.*) **3** (53) (2001), 383–390. 129, 133

[25] A. Rhodius, On the maximum of ergodicity coefficients, the Dobrushin ergodicity coefficient, and products of stochastic matrices. *Linear Algebra Appl.* **253** (1997), 141–154. 133, 135

[26] G. Roberts and J. Rosenthal, General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** (2004), 20–71. 127

[27] M. Rosenblatt, Products of independent identically distributed stochastic matrices. *J. Math. Anal. Appl.* **11** (1965), 1–10. 136, 137

[28] L. Saloff-Coste and J. Zúñiga, Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stochastic Process. Appl.* **117** (2007), 961–979. 128, 138, 139, 147, 148

[29] L. Saloff-Coste and J. Zúñiga, Merging for time inhomogeneous Markov chains, part I: Singular values and stability. *Electron. J. Probab.* **14** (2009), 1456–1494. 130, 131, 132, 133, 136, 138, 139, 149

[30] L. Saloff-Coste and J. Zúñiga, Merging for time inhomogeneous Markov chains, part II: Nash and log-Sobolev inequalities. *Ann. Probab.* **39** (2011), 1161–1203. 138, 139, 147, 148

[31] L. Saloff-Coste and J. Zúñiga, Time inhomogeneous Markov chains with wave like behavior. *Ann. Appl. Probab.* **20** (2010), 1831–1853. 138, 149

[32] E. Seneta, On strong ergodicity of inhomogeneous products of finite stochastic matrices. *Studia Math.* **46** (1973), 241–247. 133

[33] E. Seneta, On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Cambridge Philos. Soc.* **74** (1973), 507–513. 133

[34] I. Sonin, The asymptotic behaviour of a general finite nonhomogeneous Markov chain (the decomposition-separation theorem). In *Statistics, probability and game theory*, IMS Lecture Notes Monogr. Ser. 30, Inst. Math. Statist., Hayward, CA, 1996, 337–346. 134

[35] Ö. Stenflo, Perfect sampling from the limit of deterministic products of stochastic matrices. *Electron. Commun. Probab.* 13 (2008) 474–481. 135

[36] D. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*. Grundlehren Math. Wiss. 233, Springer-Verlag, Berlin 1997. 133

[37] Y. Takahashi, Markov chains with random transition matrices. *Kōdai Math. Sem. Rep.* **21** (1969) 426–447. 136

[38] J. Wolfowitz, Products of indecomposable aperiodic stochastic matrices. *Proc. Amer. Math. Soc.* **14** (1963), 733–737. 135