Electronic Journal of Probability

Vol. 14 (2009), Paper no. 49, pages 1456–1494.

Journal URL
http://www.math.washington.edu/~ejpecp/

Merging for time inhomogeneous finite Markov chains, Part I: singular values and stability

L. Saloff-Coste*
Department of mathematics
Cornell University
310 Malott Hall
Ithaca, NY 14853
lsc@math.cornell.edu

J. Zúñiga†
Department of mathematics
Stanford University
building 380
Stanford, CA 94305
jzuniga@math.stanford.edu

Abstract

We develop singular value techniques in the context of time inhomogeneous finite Markov chains with the goal of obtaining quantitative results concerning the asymptotic behavior of such chains. We introduce the notion of c-stability which can be viewed as a generalization of the case when a time inhomogeneous chain admits an invariant measure. We describe a number of examples where these techniques yield quantitative results concerning the merging of the distributions of the time inhomogeneous chain started at two arbitrary points.

Key words: Time inhomogeneous Markov chains, merging, singular value inequalities.

AMS 2000 Subject Classification: Primary 60J10.

Submitted to EJP on January 28, 2009, final version accepted May 24, 2009.

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

1 Introduction

The quantitative study of time inhomogeneous Markov chains is a very broad and challenging task. Time inhomogeneity introduces so much flexibility that a great variety of complex behaviors may occur. For instance, in terms of ergodic properties, time inhomogeneity allows for the construction of Markov chains that very efficiently and exactly attain a target distribution in finite time. An example is the classical algorithm for picking a permutation at random. Thinking of a deck of n cards, one way to describe this algorithm is as follows. At step $i \mod n$, pick a card uniformly at random among the bottom $n-i+1$ cards and insert it in position i. After $n-1$ steps, the deck is distributed according to the uniform distribution. However, it is not possible to recognize this fact by inspecting the properties of the individual steps. Indeed, changing the order of the steps destroys the neat convergence result mentioned above.

In this article, we are interested in studying the the ergodic properties of a time inhomogeneous chain through the individual ergodic properties of the one step Markov kernels. The works [4; 23; 25] consider similar problems. To illustrate what we have in mind, consider the following. Given a sequence of irreducible Markov kernels $(K_i)_1^\infty$ on a finite set V, let K_i^n be the usual iterated kernel of the chain driven by K_i alone, and let $K_{0,n}(x,\cdot)$ be the distribution of the chain $(X_t)_1^\infty$ driven by the sequence $(K_i)_1^\infty$ with $X_0 = x$. Let π_i be the invariant probability measure of the kernel K_i. Suppose we understand well the convergence $K_i^n(x,\cdot) \to \pi_i(\cdot) \;\; \forall x$ and that this convergence is, in some sense, uniform over i. For instance, assume that there exists $\beta \in (0,1)$ and $T > 0$ such that, for all i and $n \geq T + m$, $m > 0$

$$\max_{x,y}\left\{\left|\frac{K_i^n(x,y)}{\pi_i(y)} - 1\right|\right\} \leq \beta^m. \tag{1}$$

We would like to apply (1) to deduce results concerning the proximity of the measures

$$K_{0,n}(x,\cdot), K_{0,n}(y,\cdot),\; x,y \in V.$$

These are the distributions at time n for the chain started at two distinct points x,y. To give a precise version of the types of questions we would like to consider, we present the following open problem.

Problem 1.1. Let $(K_i,\pi_i)_1^\infty$ be a sequence of irreducible reversible Markov kernels on a finite set V satisfying (1). Assume further that there exists a probability measure π and a constant $c \geq 1$ such that

$$\forall i = 1,2,\dots,\;\; c^{-1}\pi \leq \pi_i \leq c\pi \;\text{ and }\; \min_{x\in V}\{K_i(x,x)\} \geq c^{-1}. \tag{2}$$

1. Prove (or disprove) that there exist $B(T,V,\pi,\beta,c)$ and $\alpha = \alpha(T,V,\pi,\beta)$ such that, for all $n \geq B(T,V,\pi,\beta) + m$,

$$\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\text{TV}} \leq \alpha^m \quad \text{or} \quad \max_{x,y,z}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1\right|\right\} \leq \alpha^m.$$

2. Prove (or disprove) that there exists a constant $A = A(c)$ such that, for all $n \geq A(T+m)$,

$$\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\text{TV}} \leq \beta^m \quad \text{or} \quad \max_{x,y,z}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1\right|\right\} \leq \beta^m.$$

3. Consider problems (1)-(2) above when $K_i \in \mathcal{Q} = \{Q_1, Q_2\}$ where Q_1, Q_2 are two irreducible Markov kernels with reversible measures π_1, π_2 satisfying $c^{-1}\pi_1 \leq \pi_2 \leq c\pi_1$, for all $x \in V$.

Concerning the formulation of these problems, let us point out that without the two hypotheses in (2), there are examples showing that (1) is not sufficient to imply any of the desired conclusions in item 1, even under the restrictive condition of item 3.

The questions presented in Problem 1.1 are actually quite challenging and, at the present time, we are far from a general solution, even in the simplest context of item 3. In fact, one of our aims is to highlight that these types of questions are not understood at all! We will only give partial results for Problem 1.1 under very specific additional hypotheses. In this respect, item 3 offers a very reasonable open problem for which evidence for a positive answer is still scarce but a counter example would be quite surprising.

In Section 6 (Remark 6.17), we show that the strongest conclusion in (2) holds true on the two-point space. The proof is quite subtle even in this simple setting. The best evidence supporting a positive answer to Problem 1.1(1) is the fact that the conclusion

$$\|K_{0,n}(x,) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} \leq (\min_z\{\pi(z)\})^{-1/2}\beta^n$$

holds if we assume that all $\pi_i = \pi$ are equal (a similar relative-sup result also holds true). This means we can take $B(T,V,\pi,\beta) = (\log \pi_*^{-1/2})/(\log \beta^{-1})$, $\pi_* = \min_z\{\pi(z)\}$ and $\alpha = \beta$. This result follows, for instance, from [23] which provides further quantitative estimates but falls short of the general statement

$$\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} \leq \beta^m, \;\; n \geq A(T+m).$$

The very strong conclusion above is known only in the case when V is a group, the kernels K_i are invariant, and $T \geq \alpha^{-1}\log V$. See [23, Theorem 4.9].

In view of the difficulties mentioned above, the purpose of this paper (and of the companion papers [25; 26]) is to develop techniques that apply to some instances of Problem 1.1 and some of its variants. Namely, we show how to adapt tools that have been successfully applied to time homogeneous chains to the study of time inhomogeneous chains and provide a variety of examples where these tools apply. The most successful techniques in the quantitative study of (time homogeneous) finite Markov chains include: coupling, strong stationary time, spectral methods, and functional inequalities such as Nash or log-Sobolev inequalities. This article focuses on spectral methods, more precisely, singular values methods. The companion paper [25] develops Nash and log-Sobolev inequalities techniques. Two papers that are close in spirit to the present work are [4; 11]. In particular, the techniques developed in [4] are closely related to those we develop here and in [25]. We point out that the singular values and functional inequalities techniques discussed here and in [4; 25] have the advantage of leading to results in distances such as ℓ^2-distance (i.e., chi-square) and relative-sup norm which are stronger than total variation.

The material in this paper is organized as follows. Section 2 introduces our basic notation and the concept of merging (in total variation and relative-sup distances). See Definitions 2.1, 2.8, 2.11. Section 3 shows how singular value decompositions can be used, theoretically, to obtain merging bounds. The main result is Theorem 3.2. An application to time inhomogeneous constant rate birth and death chains is presented. Section 4 introduces the fundamental concept of stability (Definition 4.1), a relaxation of the very restrictive hypothesis used in [23] that the kernels driving the time inhomogeneous chain under investigation all share the same invariant distribution. If the stability hypothesis is satisfied then the singular value analysis becomes much easier to apply in practice. See Theorems 4.10 and 4.11. Section 4.2 offers our first example of stability concerning end-point perturbations of simple random walk on a stick. A general class of birth and death examples where stability holds is studied in Section 5. Further examples of stability are described in [25; 26]. The final section, Section 6, gives a complete analysis of time inhomogeneous chains on the two-point space. We characterize total variation merging and study stability and relative-sup merging in this simple but fundamental case.

We end this introduction with some brief comments regarding the coupling and strong stationary time techniques. Since, typically, time inhomogeneous Markov chains do not converge to a fixed distribution, adapting the technique of strong stationary time poses immediate difficulties. This comment seems to apply also to the recent technique of evolving sets [17], which is somewhat related to strong stationary times. In addition, effective constructions of strong stationary times are usually not very robust and this is likely to pose further difficulties. An example of a strong stationary time argument for a time inhomogeneous chain that admits a stationary measure can be found in [18].

Concerning coupling, as far as theory goes, there is absolutely no difficulties in adapting the coupling technique to time inhomogeneous Markov chains. Indeed, the usual coupling inequality

$$\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\text{TV}} \leq \mathbf{P}(T > n)$$

holds true (with the exact same proof) if T is the coupling time of a coupling (X_n, Y_n) adapted to the sequence $(K_i)_1^\infty$ with starting points $X_0 = x$ and $Y_0 = y$. See [11] for practical results in this direction and related techniques. Coupling is certainly useful in the context of time inhomogeneous chains but we would like to point out that time inhomogeneity introduces very serious difficulties in the construction and analysis of couplings for specific examples. This seems related to the lack of robustness of the coupling technique. For instance, in many coupling constructions, it is important that past progress toward coupling is not destroyed at a later stage, yet, the necessary adaptation to the changing steps of a time inhomogeneous chain makes this difficult to achieve.

2 Merging

2.1 Different notions of merging

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

$$\begin{aligned} P(X_n = y | X_{n-1} = x, \ldots, X_0 = x_0) &= P(X_n = y | X_{n-1} = x) \\ &= K_n(x,y). \end{aligned}$$

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

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

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

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

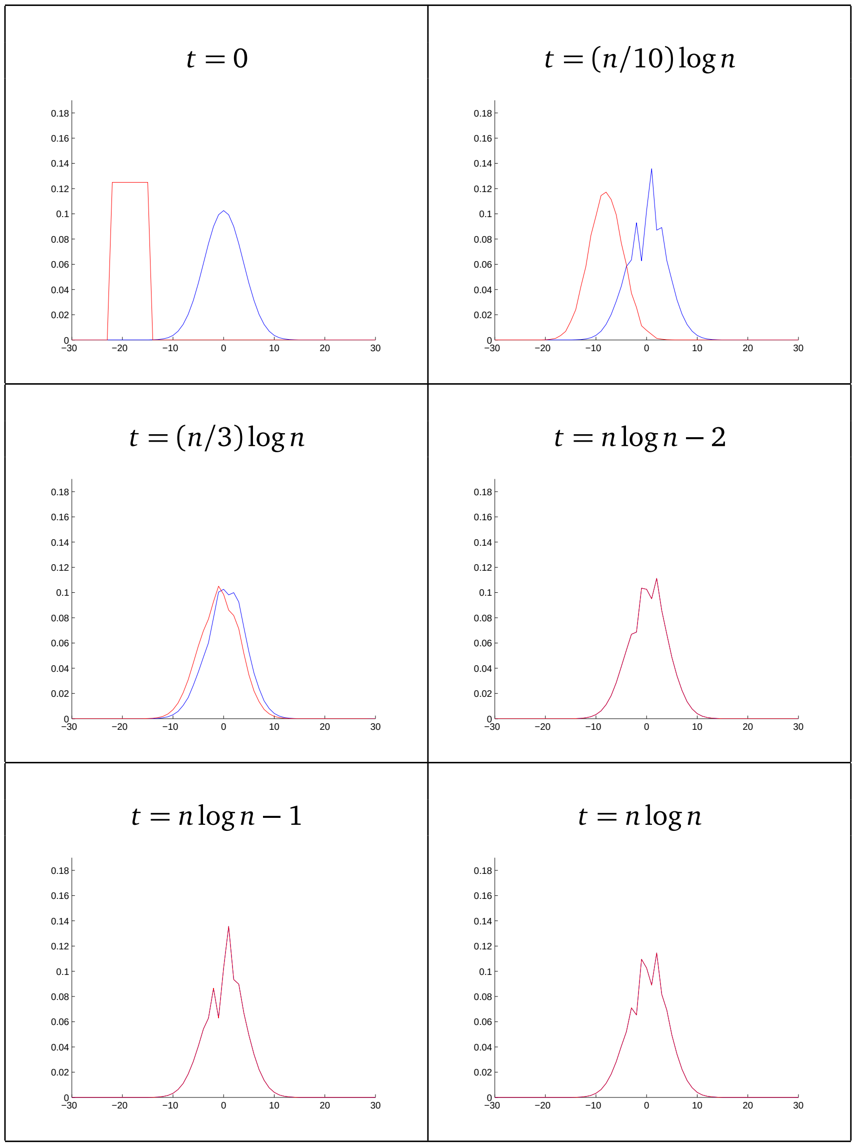


Figure 1: Illustration of merging (both in total variation and relative-sup) based on the binomial example studied in Section 5.2. The first frame shows two particular initial distributions, one of which is the binomial. The other frames show the evolution under a time inhomogeneous chain driven by a deterministic sequence involving two kernels from the set $\mathcal{Q}_N(Q, \epsilon)$ of Section 5.2, a set consisting of perturbations of the Ehrenfest chain kernel. In the fourth frame, the distributions have merged. The last two frames illustrate the evolution after merging and the absence of a limiting distribution. Here $N = 30$ and the total number of points is $n = 61$.

with $K_{n,n} = I$ (the identity). If we view the K_n's as matrices then this definition means that $K_{n,m} = K_{n+1} \cdots K_m$. In the case of time homogeneous chains where all $K_i = Q$ are equal, we write $K_{0,n} = Q^n$.

Our main interest is understanding mixing type properties of time inhomogeneous Markov chains. However, in general, $\mu_n = \mu_0 K_{0,n}$ does not converge toward a limiting distribution. Instead, the natural notion to consider is that of merging defined below. For a discussion of this property and its variants, see, e.g., [3; 16; 19; 27].

Definition 2.1 (Total variation merging)**.** Fix a sequence $(K_i)_1^\infty$ of Markov kernels on a finite set V. We say the sequence $(K_i)_1^\infty$ is merging in total variation if for any $x, y \in V$,

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

A rather trivial example that illustrates merging versus mixing is as follows.

Example 2.2. Fix two probability distributions π_i, $i = 1, 2$. Let $\mathcal{Q} = \{Q_1, Q_2\}$ with $Q_i(x, y) = \pi_i(y)$. Then for any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}$, $K_{0,n} = \pi_{i(n)}$ where $i(n) = i$ if $K_n = Q_i$, $i = 1, 2$.

Remark 2.3. If the sequence $(K_i)_1^\infty$ is merging then, for any two starting distributions μ_0, ν_0, the measures $\mu_n = \mu_0 K_{0,n}$ and $\nu_n = \nu_0 K_{0,n}$ are merging, that is, $\forall A \subset V$, $\mu_n(A) - \nu_n(A) \to 0$.

Our goal is to develop quantitative results for time inhomogeneous chains in the spirit of the work concerning homogeneous chains of Aldous, Diaconis and others who obtain precise estimates on the mixing time of ergodic chains that depend on size of the state space in an explicit way. In these works, convergence to stationary is measured in terms of various distances between measures μ, ν such as the total variation distance

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A\subset V}\{\mu(A) - \nu(A)\},$$

the chi-square distance w.r.t. ν and the relative-sup distance $\|\frac{\mu}{\nu} - 1\|_\infty$. See, e.g., [1; 5; 6; 12; 21]. Given an irreducible aperiodic chain with kernel K on a finite set V, there exists a unique probability measure $\pi > 0$ such that $K^n(x,\cdot) \to \pi(\cdot)$ as $n \to \infty$, for all x. This qualitative property can be stated equivalently using total variation, the chi-square distance or relative-sup distance. However, if we do not assume irreducibility, it is possible that there exists a unique probability measure π (with perhaps $\pi(y) = 0$ for some y) such that, for all x, $K^n(x,\cdot) \to \pi(\cdot)$ as n tends to infinity (this happens when there is a unique absorbing class with no periodicity). In such a case, $K^n(x,\cdot)$ does converge to π in total variation but the chi-square and relative-sup distances are not well defined (or are equal to $+\infty$). This observation has consequences in the study of time inhomogeneous Markov chains. Since there seems to be no simple natural property that would replace irreducibility in the time inhomogeneous context, one must regard total variation merging and other notions of merging as truly different properties.

Definition 2.4 (Relative-sup merging)**.** Fix a sequence of $(K_i)_1^\infty$ of Markov kernels. We say the sequence is merging in relative-sup distance if

$$\lim_{n\to\infty} \max_{x,y,z} \left\{ \left| \frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1 \right| \right\} = 0.$$

The techniques discussed in this paper are mostly related to the notion of merging in relative-sup distance. A graphic illustration of merging is given in Figure 1.

Remark 2.5. On the two-point space, consider the reversible irreducible aperiodic kernels

$$K_1 = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}, \; K_2 = \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}.$$

Then the sequence $K_1, K_2, K_1, K_2, \dots$ is merging in total variation but is not merging in relative-sup distance. See Section 6 for details.

When focusing on the relation between ergodic properties of individual kernels K_i and the behavior of an associated time inhomogeneous chain, it is intuitive to look at the K_i as a set instead of a sequence. The following definition introduces a notion of merging for sets of kernels.

Definition 2.6. Let $\mathcal{Q}$ be a set of Markov kernels on a finite state space V. We say that $\mathcal{Q}$ is merging in total variation if, for any sequence $(K_i)_1^\infty$, $K_i \in \mathcal{Q}$, we have

$$\forall\, x \in V, \quad \lim_{n\to\infty} \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} = 0.$$

We say that $\mathcal{Q}$ is merging in relative-sup if, for any sequence $(K_i)_1^\infty$, $K_i \in \mathcal{Q}$, we have

$$\lim_{n\to\infty} \max_{x,y,z} \left\{ \left| \frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1 \right| \right\} = 0.$$

One of the goals of this work is to describe some non-trivial examples of merging families $\mathcal{Q} = \{Q_1, Q_2\}$ where Q_1 and Q_2 have distinct invariant measures.

Example 2.7. Many examples (with all $Q_i \in \mathcal{Q}$ sharing the same invariant distribution) are given in [23], with quantitative bounds. For instance, let $V = G$ be a finite group and S_1, S_2 be two symmetric generating sets. Assume that the identity element e belongs to $S_1 \cap S_2$. Assume further that $\max\{\#S_1, \#S_2\} = N$ and that any element of G is the product of at most D elements of S_i, $i = 1, 2$. In other words, the Cayley graphs of G associated with S_1 and S_2 both have diameter at most D. Let $Q_i(x,y) = (\#S_i)^{-1}\mathbf{1}_{S_i}(x^{-1}y)$, $i = 1, 2$. Then $\mathcal{Q} = \{Q_1, Q_2\}$ is merging. Moreover, for any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}$, we have

$$\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} \le |G|^{1/2} \left(1 - \frac{1}{(ND)^2}\right)^{n/2}$$

where $|G| = \#G$. In fact, $K_{0,n}(x,\cdot) \to \pi$ where π is the uniform distribution on G and [23] gives $2\|K_{0,n}(x,\cdot) - \pi\|_{\mathrm{TV}} \le |G|^{1/2}\left(1 - \frac{1}{(ND)^2}\right)^{n/2}$.

2.2 Merging time

In the quantitative theory of ergodic time homogeneous Markov chains, the notion of mixing time plays a crucial role. For time inhomogeneous chains, we propose to consider the following corresponding definitions.

Definition 2.8. Fix $\epsilon \in (0,1)$. Given a sequence $(K_i)_1^\infty$ of Markov kernels on a finite set V, we call max total variation ϵ-merging time the quantity

$$T_{\mathrm{TV}}(\epsilon) = \inf\left\{ n : \max_{x,y\in V} \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} < \epsilon \right\}$$

Definition 2.9. Fix $\epsilon \in (0,1)$. We say that a set $\mathscr{Q}$ of Markov kernels on V has max total variation ϵ-merging time at most T if for any sequence $(K_i)_1^\infty$ with $K_i \in \mathscr{Q}$ for all i, we have $T_{\mathrm{TV}}(\epsilon) \leq T$, that is,

$$\forall t > T, \max_{x,y \in V} \left\{ \|K_{0,t}(x,\cdot) - K_{0,t}(y,\cdot)\|_{\mathrm{TV}} \right\} \leq \epsilon.$$

Example 2.10. If $\mathscr{Q} = \{Q_1, Q_2\}$ is as in Example 2.7 the total variation ϵ-merging time for $\mathscr{Q}$ is at most $(ND)^2(\log|G| + 2\log 1/\epsilon)$.

As noted earlier, merging can be defined and measured in ways other than total variation. One very natural and much stronger notion than total variation distance is relative-sup distance used in Definitions 2.4-2.6 and in the definitions below.

Definition 2.11. Fix $\epsilon \in (0,1)$. Given a sequence $(K_i)_1^\infty$ of Markov kernels on a finite set V, we call relative-sup ϵ-merging time the quantity

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

Definition 2.12. Fix $\epsilon \in (0,1)$. We say that a set $\mathscr{Q}$ of Markov kernels on V has relative-sup ϵ-merging time at most T if for any sequence $(K_i)_1^\infty$ with $K_i \in \mathscr{Q}$ for all i, we have $T_\infty(\epsilon) \leq T$, that is,

$$\forall t > T, \max_{x,y,z \in V} \left\{ \left| \frac{K_{0,t}(x,\cdot)}{K_{0,t}(y,\cdot)} - 1 \right| \right\} < \epsilon.$$

Remark 2.13. If the sequence $(K_i)_1^\infty$ is merging in total variation or relative-sup then, for any initial distribution μ_0 the sequence $\mu_n = \mu_0 K_{0,n}$ must merge with the sequence ν_n where ν_n is the invariant measure for $K_{0,n}$. In total variation, we have

$$\|\mu_n - \nu_n\|_{\mathrm{TV}} \leq \max_{x,y} \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}}.$$

In relative-sup, for $\epsilon \in (0,1/2)$, inequality (4) below yields

$$\max_{x,y,z} \left| \frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1 \right| \leq \epsilon \implies \max_x \left| \frac{\mu_n(x)}{\nu_n(x)} - 1 \right| \leq 4\epsilon.$$

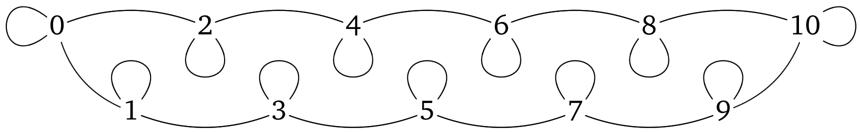
Example 2.14. If $\mathscr{Q} = \{Q_1, Q_2\}$ is as in Example 2.7 the relative-sup ϵ-merging time for $\mathscr{Q}$ is at most $2(ND)^2(\log|G| + \log 1/\epsilon)$. This follows from [23].

The following simple example illustrates how the merging time of a family of kernels $\mathscr{Q}$ may differ significantly form the merging time of a particular sequence $(K_i)_1^\infty$ with $K_i \in \mathscr{Q}$ for all $i \geq 1$.

Example 2.15. Let Q_1 be the birth and death chain kernel on $V_N = \{0, \dots, N\}$ with constant rates p, q, $p+q = 1$, $p > q$. This means here that $Q_1(x, x+1) = p$, $Q_1(x, x-1) = q$ when these are well defined and $Q_1(0,0) = q$, $Q_1(N,N) = p$. The reversible measure for Q_1 is $\pi_1(x) = c(p,q,N)(q/p)^{N-x}$ with $c(p,q,N) = (1-q/p)/(1-(q/p)^{N+1})$. The chain driven by this kernel is well understood. In particular, the mixing time is of order N starting from the end where π_1 attains its minimum.

Let Q_2 be the birth and death chain with constant rates q, p. Hence, $Q_2(x, x+1) = q$, $Q_2(x, x-1) = p$ when these are well defined and $Q_2(0,0) = p$, $Q_2(N,N) = q$. The reversible measure for K_2 is

$\pi_2(x) = c(p,q,N)(q/p)^x$. It is an interesting problem to study the merging property of the set $\mathcal{Q} = \{Q_1, Q_2\}$. Here, we only make a simple observation concerning the behavior of the sequence $K_i = Q_{i \bmod 2}$. Let $Q = K_{0,2} = Q_1Q_2$. The graph structure of this kernel is a circle. As an example, below we give the graph structure for $N = 10$.



Edges are drawn between points x and y if $Q(x,y) > 0$. Note that $Q(x,y) > 0$ if and only if $Q(y,x) > 0$, so that all edges can be traversed in both directions (possibly with different probabilities).

For the Markov chain driven by Q, there is equal probability of going from a point x to any of its neighbors as long as $x \neq 0, N$. Using this fact, one can compute the invariant measure π of Q and conclude that

$$\max_{V_N}\{\pi\} \leq (p/q)^2 \min_{V_N}\{\pi\}.$$

It follows that $(q/p)^2 \leq (N+1)\pi(x) \leq (p/q)^2$. This and the comparison techniques of [8] show that the sequence $Q_1, Q_2, Q_1, Q_2, \ldots,$ is merging in relative sup in time of order N^2. Compare with the fact that each kernel K_i in the sequence has a mixing time of order N.

3 Singular value analysis

3.1 Preliminaries

We say that a measure μ is positive if $\forall x,\ \mu(x) > 0$. Given a positive probability measure μ on V and a Markov kernel K, set $\mu' = \mu K$. If K satisfies

$$\forall\, y \in V,\ \sum_{x \in V} K(x,y) > 0 \tag{1}$$

then μ' is also positive. Obviously, any irreducible kernel K satisfies (1).

Fix $p \in [1,\infty]$ and consider K as a linear operator

$$K : \ell^p(\mu') \to \ell^p(\mu),\ \ Kf(x) = \sum_y K(x,y)f(y). \tag{2}$$

It is important to note, and easy to check, that for any measure μ, the operator $K : \ell^p(\mu') \to \ell^p(\mu)$ is a contraction.

Consider a sequence $(K_i)_1^\infty$ of Markov kernels satisfying (1). Fix a positive probability measure μ_0 and set $\mu_n = \mu_0 K_{0,n}$. Observe that $\mu_n > 0$ and set

$$d_p(K_{0,n}(x,\cdot),\mu_n) = \left(\sum_y \left|\frac{K_{0,n}(x,y)}{\mu_n(y)} - 1\right|^p \mu_n(y)\right)^{1/p}.$$

Note that $2\|K_{0,n}(x,\cdot)-\mu_n\|_{\text{TV}} = d_1(K_{0,n}(x,\cdot),\mu_n)$ and, if $1 \le p \le r \le \infty$, $d_p(K_{0,n}(x,\cdot),\mu_n) \le d_r(K_{0,n}(x,\cdot),\mu_n)$. Further, one easily checks the important fact that

$$n \mapsto d_p(K_{0,n}(x,\cdot),\mu_n)$$

is non-increasing.

It follows that we may control the total variation merging of a sequence $(K_i)_0^\infty$ with K_i satisfying (1) by

$$\|K_{0,n}(x,\cdot)-K_{0,n}(y,\cdot)\|_{\text{TV}} \le \max_{x\in V}\{d_2(K_{0,n}(x,\cdot),\mu_n)\}. \tag{3}$$

To control relative-sup merging we note that if

$$\max_{x,y,z}\left\{\left|\frac{K_{0,n}(x,z)}{\mu_n(z)}-1\right|\right\} \le \epsilon \le 1/2 \quad \text{then} \quad \max_{x,y,z}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)}-1\right|\right\} \le 4\epsilon.$$

The last inequality follows from the fact that if $1-\epsilon \le a/b, c/b \le 1+\epsilon$ with $\epsilon \in (0,1/2)$ then

$$1-2\epsilon \le \frac{1-\epsilon}{1+\epsilon} \le \frac{a}{c} \le \frac{1+\epsilon}{1-\epsilon} \le 1+4\epsilon. \tag{4}$$

3.2 Singular value decomposition

The following material can be developed over the real or complex numbers with little change. Since our operators are Markov operators, we work over the reals. Let $\mathscr{H}$ and $\mathscr{G}$ be (real) Hilbert spaces equipped with inner products $\langle\cdot,\cdot\rangle_{\mathscr{H}}$ and $\langle\cdot,\cdot\rangle_{\mathscr{G}}$ respectively. If $u:\mathscr{H}\times\mathscr{G}\to\mathbb{R}$ is a bounded bilinear form, by the Riesz representation theorem, there are unique operators $A:\mathscr{H}\to\mathscr{G}$ and $B:\mathscr{G}\to\mathscr{H}$ such that

$$u(h,k) = \langle Ah,k\rangle_{\mathscr{G}} = \langle h,Bk\rangle_{\mathscr{H}}. \tag{5}$$

If $A:\mathscr{H}\to\mathscr{G}$ is given and we set $u(h,k)=\langle Ah,k\rangle_{\mathscr{G}}$ then the unique operator $B:\mathscr{G}\to\mathscr{H}$ satisfying (5) is called the *adjoint* of A and is denoted as $B=A^*$. The following classical result can be derived from [20, Theorem 1.9.3].

Theorem 3.1 (Singular value decomposition). *Let $\mathscr{H}$ and $\mathscr{G}$ be two Hilbert spaces of the same dimension, finite or countable. Let $A:\mathscr{H}\to\mathscr{G}$ be a compact operator. There exist orthonormal bases (ϕ_i) of $\mathscr{H}$ and (ψ_i) of $\mathscr{G}$ and non-negative reals $\sigma_i=\sigma_i(\mathscr{H},\mathscr{G},A)$ such that $A\phi_i=\sigma_i\psi_i$ and $A^*\psi_i=\sigma_i\phi_i$. The non-negative numbers σ_i are called the singular values of A and are equal to the square root of the eigenvalues of the self-adjoint compact operator $A^*A:\mathscr{H}\to\mathscr{H}$ and also of $AA^*:\mathscr{G}\to\mathscr{G}$.*

One important difference between eigenvalues and singular values is that the singular values depend very much on the Hilbert structures carried by $\mathscr{H},\mathscr{G}$. For instance, a Markov operator K on a finite set V may have singular values larger than 1 when viewed as an operator from $\ell^2(\nu)$ to $\ell^2(\mu)$ for arbitrary positive probability measure ν,μ (even with $\nu=\mu$).

We now apply the singular value decomposition above to obtain an expression of the ℓ^2 distance between $\mu'=\mu K$ and $K(x,\cdot)$ when K is a Markov kernel satisfying (1) and μ a positive probability measure on V. Consider the operator $K = K_\mu : \ell^2(\mu') \to \ell^2(\mu)$ defined by (2). Then the adjoint $K^*_\mu : \ell^2(\mu) \to \ell^2(\mu')$ has kernel $K^*_\mu(x,y)$ given by

$$K^*_\mu(y,x) = \frac{K(x,y)\mu(x)}{\mu'(y)}.$$

By Theorem 3.1, there are eigenbases $(\varphi_i)_0^{|V|-1}$ and $(\psi_i)_0^{|V|-1}$ of $\ell^2(\mu')$ and $\ell^2(\mu)$ respectively such that

$$K_\mu \varphi_i = \sigma_i \psi_i \text{ and } K^*_\mu \psi_i = \sigma_i \varphi_i$$

where $\sigma_i = \sigma_i(K,\mu)$, $i = 0, \dots |V|-1$ are the singular values of K, i.e., the square roots of the eigenvalues of $K^*_\mu K_\mu$ (and $K_\mu K^*_\mu$) given in non-increasing order, i.e.

$$1 = \sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{|V|-1}$$

and $\psi_0 = \varphi_0 \equiv 1$. From this it follows that, for any $x \in V$,

$$d_2(K(x,\cdot),\mu')^2 = \sum_{i=1}^{|V|-1} |\psi_i(x)|^2 \sigma_i^2. \tag{6}$$

To see this, write

$$\begin{aligned} d_2(K(x,\cdot),\mu')^2 &= \left\langle \frac{K(x,\cdot)}{\mu'} - 1, \frac{K(x,\cdot)}{\mu'} - 1 \right\rangle_{\mu'} \\ &= \left\langle \frac{K(x,\cdot)}{\mu'}, \frac{K(x,\cdot)}{\mu'} \right\rangle_{\mu'} - 1. \end{aligned}$$

With $\tilde{\delta_y} = \delta_y/\mu'(y)$, we have $K(x,y)/\mu'(y) = K\tilde{\delta_y}(x)$. Write

$$\tilde{\delta_y} = \sum_0^{|V|-1} a_i \varphi_i \text{ where } a_i = \langle \tilde{\delta_y}, \varphi_i \rangle_{\mu'} = \varphi_i(y)$$

so we get that

$$\frac{K(x,y)}{\mu'(y)} = \sum_{i=0}^{|V|-1} \sigma_i \psi_i(x) \varphi_i(y).$$

Using this equality yields the desired result. This leads to the main result of this section. In what follows we often write K for K_μ when the context makes it clear that we are considering K as an operator from $\ell^2(\mu')$ to $\ell^2(\mu)$ for some fixed μ.

Theorem 3.2. *Let $(K_i)_1^\infty$ be a sequence of Markov kernels on a finite set V, all satisfying* (1). *Fix a positive starting measure μ_0 and set $\mu_i = \mu_0 K_{0,i}$. For each $i = 0, 1, \dots$, let*

$$\sigma_j(K_i, \mu_{i-1}), \;\; j = 0, 1, \dots, |V|-1,$$

be the singular values of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$ in non-increasing order. Then

$$\sum_{x \in V} d_2(K_{0,n}(x,\cdot),\mu_n)^2 \mu_0(x) \leq \sum_{j=1}^{|V|-1} \prod_{i=1}^{n} \sigma_j(K_i, \mu_{i-1})^2.$$

and, for all $x \in V$,

$$d_2(K_{0,n}(x,\cdot),\mu_n)^2 \le \left(\frac{1}{\mu_0(x)}-1\right)\prod_{i=1}^{n}\sigma_1(K_i,\mu_{i-1})^2.$$

Moreover, for all $x, y \in V$,

$$\left|\frac{K_{0,n}(x,y)}{\mu_n(y)}-1\right| \le \left(\frac{1}{\mu_0(x)}-1\right)^{1/2}\left(\frac{1}{\mu_n(y)}-1\right)^{1/2}\prod_{i=1}^{n}\sigma_1(K_i,\mu_{i-1}).$$

Proof. Apply the discussion prior to Theorem 6 with $\mu = \mu_0$, $K = K_{0,n}$ and $\mu' = \mu_n$. Let $(\psi_i)_0^{|V|-1}$ be the orthonormal basis of $\ell^2(\mu_0)$ given by Theorem 3.1 and $\tilde{\delta_x} = \delta_x/\mu_0(x)$. Then $\tilde{\delta_x} = \sum_0^{|V|-1}\psi_i(x)\psi_i$. This yields

$$\sum_{i=0}^{|V|-1}|\psi_i(x)|^2 = \|\tilde{\delta_x}\|^2_{\ell^2(\mu_0)} = \mu_0(x)^{-1}.$$

Furthermore, Theorem 3.3.4 and Corollary 3.3.10 in [15] give the inequality

$$\forall\, k = 1,\dots,|V|-1,\quad \sum_{j=1}^{k}\sigma_j(K_{0,n},\mu_0)^2 \le \sum_{j=1}^{k}\prod_{i=1}^{n}\sigma_j(K_i,\mu_{i-1})^2.$$

Using this with $k = |V|-1$ in (6) yields the first claimed inequality. The second inequality then follows from the fact that $\sigma_1(K_{0,n},\mu_0) \ge \sigma_j(K_{0,n},\mu_0)$ for all $j = 1\dots|V|-1$. The last inequality follows from writing

$$\left|\frac{K_{0,n}(x,y)}{\mu_n(y)}-1\right| \le \sigma(K_{0,n},\mu_0)\sum_{1}^{|V|-1}|\psi_i(x)\phi_i(y)|$$

and bounding $\sum_1^{|V|-1}|\psi_i(x)\phi_i(y)|$ by $(\mu_0(x)^{-1}-1)^{1/2}(\mu_n(y)^{-1}-1)^{1/2}$. □

Remark 3.3. The singular value $\sigma_1(K_i,\mu_{i-1}) = \sqrt{\beta_1(i)}$ is the square root of the second largest eigenvalue $\beta_1(i)$ of $K_i^*K_i : \ell^2(\mu_i) \to \ell^2(\mu_i)$. The operator $P_i = K_i^*K_i$ has Markov kernel

$$P_i(x,y) = \frac{1}{\mu_i(x)}\sum_{z\in V}\mu_{i-1}(z)K_i(z,x)K_i(z,y) \tag{7}$$

with reversible measure μ_i. Hence

$$1-\beta_1(i) = \min_{f\not\equiv\mu_i(f)}\left\{\frac{\mathcal{E}_{P_i,\mu_i}(f,f)}{\mathrm{Var}_{\mu_i}(f)}\right\}$$

with

$$\mathcal{E}_{P_i,\mu_i}(f,f) = \frac{1}{2}\sum_{x,y\in V}|f(x)-f(y)|^2P_i(x,y)\mu_i(x).$$

The difficulty in applying Theorem 3.2 is that it usually requires some control on the sequence of measures μ_i. Indeed, assume that each K_i is aperiodic irreducible with invariant probability measure

π_i. One natural way to put quantitative hypotheses on the ergodic behavior of the individual steps (K_i, π_i) is to consider the Markov kernel

$$\widetilde{P}_i(x,y) = \frac{1}{\pi_i(x)} \sum_{z\in V} \pi_i(z)K_i(z,x)K_i(z,y)$$

which is the kernel of the operator $K_i^* K_i$ when K_i is understood as an operator acting on $\ell^2(\pi_i)$ (note the difficulty of notation coming from the fact that we are using the same notation K_i to denote two operators acting on different Hilbert spaces). For instance, let $\widetilde{\beta}_i$ be the second largest eigenvalue of $(\widetilde{P}_i, \pi_i)$. Given the extreme similarity between the definitions of P_i and $\widetilde{P}_i$, one may hope to bound β_i using $\widetilde{\beta}_i$. This however requires some control of

$$M_i = \max_z \left\{ \frac{\pi_i(z)}{\mu_{i-1}(z)}, \frac{\mu_i(z)}{\pi_i(z)} \right\}.$$

Indeed, by a simple comparison argument (see, e.g., [7; 9; 21]), we have

$$\beta_i \le 1 - M_i^{-2}(1-\widetilde{\beta}_i).$$

One concludes that

$$d_2(K_{0,n}(x,\cdot),\mu_n)^2 \le \left(\frac{1}{\mu_0(x)} - 1\right)\prod_{i=1}^n (1 - M_i^{-2}(1-\widetilde{\beta}_i)).$$

and

$$\left|\frac{K_{0,n}(x,y)}{\mu_n(y)} - 1\right| \le \left(\frac{1}{\mu_0(x)} - 1\right)^{1/2}\left(\frac{1}{\mu_n(y)} - 1\right)^{1/2}\prod_{i=1}^n (1 - M_i^{-2}(1-\widetilde{\beta}_i))^{1/2}.$$

Remark 3.4. The paper [4] studies certain contraction properties of Markov operators. It contains, in a more general context, the observation made above that a Markov operator is always a contraction from $\ell^p(\mu K)$ to $\ell^p(\mu)$ and that, in the case of ℓ^2 spaces, the operator norm $\|K - \mu'\|_{\ell^2(\mu K)\to\ell^2(\mu)}$ is given by the second largest singular value of $K_\mu : \ell^2(\mu K) \to \ell^2(\mu)$ which is also the square root of the second eigenvalue of the Markov operator P acting on $\ell^2(\mu K)$ where $P = K_\mu^* K_\mu$, $K_\mu^* : \ell^2(\mu) \to \ell^2(\mu K)$. This yields a slightly less precise version of the last inequality in Theorem 3.2. Namely, writing

$$(K_{0,n} - \mu_n) = (K_1 - \mu_1)(K_2 - \mu_2)\cdots(K_n - \mu_n)$$

and using the contraction property above one gets

$$\|K_{0,n} - \mu_n\|_{\ell^2(\mu_n)\to\ell^2(\mu_0)} \le \prod_1^n \sigma_1(K_i, \mu_{i-1}).$$

As $\|I - \mu_0\|_{\ell^2(\mu_0)\to\ell^\infty(\mu_0)} = \max_x(\mu_0(x)^{-1} - 1)^{1/2}$, it follows that

$$\max_{x\in V} d_2(K_{0,n}(x,\cdot),\mu_n) \le \left(\frac{1}{\min_x\{\mu_0(x)\}} - 1\right)^{1/2}\prod_1^n \sigma_1(K_i, \mu_{i-1}).$$

Example 3.5 (Doeblin's condition)**.** Assume that, for each i, there exists $\alpha_i \in (0,1)$, and a probability measure π_i (which does not have to have full support) such that

$$\forall\, i,\ x,y \in V,\ \ K_i(x,y) \geq \alpha_i \pi_i(y).$$

This is known as a Doeblin type condition. For any positive probability measure μ_0, the kernel P_i defined at (7) is then bounded below by

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

This implies that $\beta_1(i)$, the second largest eigenvalue of P_i, is bounded by $\beta_1(i) \leq 1 - \alpha_i/2$. Theorem 3.2 then yields

$$d_2(K_{0,n}(x,\cdot),\mu_n) \leq \mu_0(x)^{-1/2} \prod_{i=1}^{n} (1-\alpha_i/2)^{1/2}.$$

Let us observe that the very classical coupling argument usually employed in relation to Doeblin's condition applies without change in the present context and yields

$$\max_{x,y}\{\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}}\} \leq \prod_{1}^{n} (1-\alpha_i).$$

See [11] for interesting developments in this direction.

Example 3.6. On a finite state space V, consider a sequence of edge sets $E_i \subset V \times V$. For each i, assume that

1. For all $x \in V$, $(x,x) \in E_i$.
2. For all $x,y \in V$, there exist $k = k(i,x,y)$ and a sequence $(x_j)_0^k$ of elements of V such that $x_0 = x$, $x_k = y$ and $(x_j, x_{j+1}) \in E_i$, $j \in \{0,\dots,k-1\}$.

Consider a sequence $(K_i)_1^\infty$ of Markov kernels on V such that

$$\forall\, i,\ \forall\, x,y \in V,\ \ K_i(x,y) \geq \epsilon \mathbf{1}_{E_i}(x,y) \tag{8}$$

for some $\epsilon > 0$. We claim that the sequence $(K_i)_1^\infty$ is merging, that is,

$$K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot) \to 0 \quad \text{as} \quad n \to \infty.$$

This easily follows from Example 3.5 after one remarks that the hypotheses imply

$$\forall\, m,\ \ \forall\, x,y \in V,\ \ \ K_{m,m+|V|}(x,y) \geq \epsilon^{|V|-1}.$$

To prove this, for any fixed m, let $\Omega_{m,i}(x) = \{y \in V : K_{m,m+i}(x,y) > 0\}$. Note that $\{x\} \subset \Omega_{m,i}(x) \subset \Omega_{m,i+1}(x)$ because of condition 1 above. Further, because of condition 2, no nonempty subset $\Omega \subset V$, $\Omega \neq V$, can have the property that $\forall\, x \in \Omega, y \in V \setminus \Omega,\ K_i(x,y) = 0$. Hence $\Omega_{m,i}(x)$, $i = 1,2,\dots$ is a strictly increasing sequence of sets until, for some k, it reaches $\Omega_{m,k}(x) = V$. Of course, the integer k is at most $|V|-1$. Now, hypothesis (8) implies that $K_{m,m+|V|}(x,y) \geq \epsilon^{|V|-1}$ as desired. Of course, this line of reasoning can only yield poor quantitative bounds in general.

3.3 Application to constant rates birth and death chains

A constant rate birth and death chain Q on $V = V_N = \{0, 1 \dots, N\}$ is determined by parameters $p, q, r \in [0, 1]$, $p + q + r = 1$, and given by $Q(0, 0) = 1 - p$, $Q(N, N) = 1 - q$, $Q(x, x + 1) = p$, $x \in \{0, N - 1\}$, $Q(x, x - 1) = q$, $x \in \{1, N\}$, $Q(x, x) = r$, $x \in \{1, \dots, N - 1\}$. It has reversible measure (assuming $q \neq p$)

$$\pi(x) = c(p/q)^{x-N}, \;\; c = c(N, p, q) = (1 - (q/p))/(1 - (q/p)^{N+1}).$$

For any $A \geq a \geq 1$, let $\mathscr{Q}_N^{\uparrow}(a, A)$ be the collection of all constant rate birth and death chains on V_N with $p/q \in [a, A]$. Let $\mathscr{M}_N^{\uparrow}(a, A)$ be the set of all probability measures on V_N such that

$$a\mu(x) \leq \mu(x + 1) \leq A\mu(x), \;\; x \in \{0, \dots, N - 1\}.$$

For such a probability measure, we have

$$\mu(0) \geq \frac{A - 1}{A^{N+1} - 1}, \;\; \mu(N) \geq \frac{1 - 1/a}{1 - 1/a^{N+1}}.$$

Lemma 3.7. *Fix $A \geq a \geq 1$. If $\mu \in \mathscr{M}_N^{\uparrow}(a, A)$ and $Q \in \mathscr{Q}_N^{\uparrow}(a, A)$ then*

$$\mu' = \mu Q \in \mathscr{M}_N^{\uparrow}(a, A).$$

Proof. This follows by inspection. The end-points are the more interesting case. Let us check for instance that $a\mu'(0) \leq \mu'(1) \leq A\mu'(0)$. We have

$$\mu'(0) = (r + q)\mu(0) + q\mu(1), \;\; \mu'(1) = p\mu(0) + r\mu(1) + q\mu(2).$$

Hence $a\mu'(0) \leq \mu'(1)$ because $a\mu(1) \leq \mu(2)$ and $aq \leq p$. We also have $\mu'(1) \leq A\mu'(0)$ because

$$\begin{aligned} \mu'(1) &= (p/q)q\mu(0) + r\mu(1) + q\mu(2) \\ &\leq \max\{p/q, A\}((r + q)\mu(0) + q\mu(1)) \leq A\mu'(0). \end{aligned}$$

□

For $\eta \in (0, 1/2)$, let $\mathscr{H}_N(\eta)$ be the set of all Markov kernels K with $K(x, x) \in (\eta, 1 - \eta)$, $x \in V_N$.

Theorem 3.8. *Fix $A \geq a > 1$ and $\eta \in (0, 1/2)$. Set*

$$\alpha = \alpha(\eta, a, A) = \frac{\eta^2(1 - a^{-1/2})^2}{A + A^2}.$$

Then, for any initial distribution $\mu_0 \in \mathscr{M}_N^{\uparrow}(a, A)$ and any sequence $(K_i)_1^\infty$ with $K_i \in \mathscr{Q}_M^{\uparrow}(a, A) \cap \mathscr{H}_N(\eta)$, we have

$$d_2(K_{0,n}(x, \cdot), \mu_n)^2 \leq \frac{A^{N+1} - A}{A - 1}(1 - \alpha)^n,$$

and

$$\max_{x,z}\left\{\left|\frac{K_{n,0}(x, z)}{\mu_n(z)} - 1\right|\right\} \leq \frac{A^{N+1} - 1}{A - 1}(1 - \alpha)^{n/2}.$$

In particular, the relative-sup ϵ-merging time of the family $\mathscr{Q}_N^{\uparrow}(a, A) \cap \mathscr{H}_N(\eta)$ is at most

$$\frac{2\log(4[(A - 1)\epsilon]^{-1}) + 2(N + 1)\log A}{-\log(1 - \alpha)}.$$

Remark 3.9. This is an example where one expects the starting point to have a huge influence on the merging time between $K_{0,n}(x,\cdot)$ and $\mu_n(\cdot)$. And indeed, the proof given below based on the last two inequalities in Theorem 3.2 shows that the uniform upper bound given above can be drastically improved if one starts from N (or close to N). This is because, if starting from N, the factor $\mu_0(N)^{-1}-1$ is bounded above by $1/(a-1)$. Using this, the proof below shows approximate merging after a constant number of steps if starting from N. To obtain the uniform upper bound of Theorem 3.8, we will use the complementary fact that $\mu_0(0)^{-1}-1 \le (A^{N+1}-1)/(A-1)$.

Proof. To apply Theorem 3.2, we use Remark 3.3 and compute the kernel of $P_i = K_i^* K_i$ given by

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

We will use that (P_i, μ_i) is reversible with

$$\mu_i(x) P_i(x,x+1) \ge \frac{\eta^2}{1+A} \mu_i(x), x \in \{0,\dots,N-1\}. \tag{9}$$

To obtain this, observe first that

$$\mu_i(x) = \sum_z \mu_{i-1}(z) K_i(z,x) \le (1+A)\mu_{i-1}(x).$$

Then write

$$\begin{aligned}
\mu_i(x) P_i(x,x+1) &\ge \mu_{i-1}(x) K_i(x,x) K_i(x,x+1) \\
&\quad + \mu_{i-1}(x+1) K_i(x+1,x) K_i(x+1,x+1) \\
&\ge \frac{\eta}{1+A}(p+q)\mu_i(x) \ge \frac{\eta^2}{1+A}\mu_i(x).
\end{aligned}$$

The fact that $\mu_i(x) \ge a\mu_i(x-1)$ is equivalent to saying that $\mu_i(x) = z_i a^{h_i(x)}$ with $h_i(x+1) - h_i(x) \ge 1$, $x \in \{0, N-1\}$. In [10, Proposition 6.1], the Metropolis chain $M = M_i$ for any such measure $\mu = \mu_i$, with base simple random walk, is studied. There, it is proved that the second largest eigenvalue of the Metropolis chain is bounded by

$$\beta_1(M_i) \le 1 - (1-a^{1/2})^2/2.$$

The Metropolis chain has

$$\mu_i(x) M_i(x,x+1) = \frac{1}{2}\mu_i(x+1).$$

Hence, (9) and $\mu_i \in \mathscr{M}^{\uparrow}(a,A)$ give

$$\mu_i(x) P_i(x,x+1) \ge \frac{2\eta^2}{(1+A)A} \mu_i(x) M_i(x,x+1).$$

Now, a simple comparison argument yields

$$\beta_1(i) \le 1-\alpha, \quad \alpha = \frac{\eta^2(1-a^{-1/2})^2}{A+A^2}.$$

□

Remark 3.10. The total variation merging of the chains studied above can be obtained by a coupling argument. Indeed, for any staring points $x < y$, construct the obvious coupling that have the chains move in parallel, except when one is at an end point. The order is preserved and the two copies couple when the lowest chain hits N. A simple argument bounds the upper tail of this hitting time and shows that order N suffices for total variation merging from any two starting states. For this coupling argument (and thus for the total variation merging result) the upper bound $p/q \leq A$ is irrelevant.

4 Stability

This section introduces a concept, c-stability, that plays a crucial role in some applications of the singular values techniques used in this paper and in the functional inequalities techniques discussed in [25].

In a sense, this property is a straightforward generalization of the property that a family of kernels share the same invariant measure. We believe that understanding this property is also of independent interest.

Definition 4.1. 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 $\mu_0 > 0$ such that

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

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.

Definition 4.2. A set $\mathcal{Q}$ of Markov kernels is c-stable with respect to a probability measure $\mu_0 > 0$ if any sequence $(K_i)_1^\infty$ such that $K_i \in \mathcal{Q}$ is c-stable with respect to $\mu_0 > 0$.

Remark 4.3. If all K_i share the same invariant distribution π then $(K_i)_1^\infty$ is 1-stable with respect to π.

Remark 4.4. Suppose a set $\mathcal{Q}$ of Markov kernels is c-stable with respect to a measure μ_0. Let π be the invariant measure for some irreducible aperiodic $Q \in \mathcal{Q}$. Then we must have (consider the sequence $(K_i)_1^\infty$ with $K_i = Q$ for all i)

$$x \in V,\;\; \frac{1}{c} \leq \frac{\pi(x)}{\mu_0(x)} \leq c.$$

Hence, $\mathcal{Q}$ is also c^2-stable with respect to π and the invariant measures π, π' for any two aperiodic irreducible kernels $Q, Q' \in \mathcal{Q}$ must satisfy

$$x \in V,\;\; \frac{1}{c^2} \leq \frac{\pi(x)}{\pi'(x)} \leq c^2. \tag{1}$$

Remark 4.5. It is not difficult to find two Markov kernels K_1, K_2 on a finite state space V that are reversible, irreducible and aperiodic with reversible measures π_1, π_2 satisfying (1) so that $\{K_1, K_2\}$ is not c-stable. See, e.g., Remark 2.5. This shows that the necessary condition (1) for a set $\mathcal{Q}$ of Markov kernels to be c-stable is not a sufficient condition.

Example 4.6. On $V_N = \{0,\ldots,N\}$, consider two birth and death chains $Q_{N,1}, Q_{N,2}$ with $Q_{N,i}(x,x+1) = p_i$, $x \in \{0,\ldots,N-1\}$, $Q_{N,i}(x,x-1) = q_i$, $x \in \{1,\ldots,N\}$, $Q_{N,i}(x,x) = r_i$, $x \in \{1,\ldots,N-1\}$, $Q_{N,i}(0,0) = 1-p_i$, $Q_{N,i}(N,N) = 1-q_i$, $p_i+q_i+r_i = 1$, $i = 1,2$. Assume that $r_i \in [1/4,3/4]$. Recall that the invariant (reversible) measure for $Q_{N,i}$ is

$$\pi_{N,i}(x) = \left(\sum_{y=0}^{N}(p_i/q_i)^y\right)^{-1} (p_i/q_i)^x.$$

For any choice of the parameters p_i, q_i with $1 < p_1/q_1 < p_2/q_2$ there are no constants c such that the family $\mathcal{Q}_N = \{Q_{N,1}, Q_{N,2}\}$ is c-stable with respect to some measure $\mu_{N,0}$, uniformly over N because

$$\lim_{N\to\infty} \frac{\pi_{N,1}(0)}{\pi_{N,2}(0)} = \infty.$$

However, the sequence $(K_i)_1^\infty$ defined by $K_{2i+1} = Q_1$, $K_{2i} = Q_2$, is c-stable. Indeed, consider the chain with kernel $Q = Q_1Q_2$. This chain is irreducible and aperiodic and thus has an invariant measure μ_0. Set $\mu_n = \mu_0 K_{0,n}$. Then $\mu_{2n} = \mu_0$ and $\mu_{2n+1} = \mu_0 Q_1 = \mu_1$. It is easy to check that

$$Q(x,x\pm 1) \geq m = \min\{(p_1+p_2)/4,(q_1+q_2)/4\} > 0.$$

Since $r_i > m$, $\mu_i = \mu_0 Q_1$ and $\mu_0 = \mu_1 Q_2$ it follows that

$$m\mu_0(x) \leq \mu_1(x) \leq m^{-1}\mu_0(x).$$

Hence the sequence $(K_i)_1^\infty$ is $(1/m)$-stable.

Theorem 4.7. *Let $V_N = \{0,\ldots,N\}$. Let $(K_i)_1^\infty$ be a sequence of birth and death Markov kernels on V_N. Assume that $K_i(x,x), K_i(x,x\pm 1) \in [1/4,3/4]$, $x \in V_n$, $i = 1,\ldots$, and that $(K_i)_1^\infty$ is c-stable with respect to the uniform measure on V_N, for some constant $c \geq 1$ independent of N. Then there exists a constant $A = A(c)$ (in particular, independent of N) such that the relative-sup ϵ-merging time for $(K_i)_1^\infty$ on V_N is bounded by*

$$T_\infty(\epsilon) \leq AN^2(\log N + \log_+ 1/\epsilon).$$

This will be proved later as a consequence of a more general theorem. The estimate can be improved to $T_\infty(\epsilon) \leq AN^2(1+\log_+ 1/\epsilon)$ with the help of the Nash inequality technique of [25].

We close this section by stating an open question that seems worth studying.

Problem 4.8. Let $\mathcal{Q}$ be a set of irreducible aperiodic Markov kernels on a finite set V. Assume that there is a constant $a \geq 1$ such that $\min\{K(x,x) : x \in V, K \in \mathcal{Q}\} \geq a^{-1}$, and that for any two kernels K, K' in $\mathcal{Q}$ the associate invariant measures π, π' satisfy $a^{-1}\pi \leq \pi' \leq a\pi$. Prove or disprove that this implies that $\mathcal{Q}$ is c-stable (ideally, with a constant c depending only on a).

Getting positive results in this direction under strong additional restrictions on the kernels in $\mathcal{Q}$ is of interest. For instance, assume further that the kernels in $\mathcal{Q}$ are all birth and death kernels and that for any two kernels $K, K' \in \mathcal{Q}$, $a^{-1}K(x,y) \leq K'(x,y) \leq aK(x,y)$. Prove (or disprove) that c-stability holds in this case.

In this general direction, we only have the following (not very practical) simple result.

Proposition 4.9. *Assume that there is a constant $a \geq 1$ such that for any two finite sequences $Q_1, \dots, Q_i$ and $Q'_1, \dots, Q'_j$ of kernels from $\mathcal{Q}$ the stationary measures π, π' of the products $Q_1 \cdots Q_i$, $Q'_1 \cdots Q'_j$ satisfy $a^{-1}\pi \leq \pi' \leq a\pi$. Then $\mathcal{Q}$ is a^2-stable with respect to the invariant measure π_K of any kernel $K \in \mathcal{Q}$.*

Proof. Let $(K_i)_1^\infty$ be a sequence of kernels from $\mathcal{Q}$. By hypothesis, for any fixed n, if π'_n denotes the invariant measure of $K_{0,n} = K_1 \cdots K_n$, we have $a^{-1}\pi_K \leq \pi'_n \leq a\pi_K$. Hence, $a^{-1}\pi'_n \leq \pi_K K_{0,n} \leq a\pi'_n$. This implies $a^{-2}\pi_K \leq \pi_K K_{0,n} \leq a^2\pi_K$ as desired. □

4.1 Singular values and c-stability

Suppose Q has invariant measure π and second largest singular value σ_1. Then $d_2(Q_n(x,\cdot),\pi) \leq \pi(x)^{-1/2}\sigma_1^n$. See, e.g., [12] or [23, Theorem 3.3]. The following two statements can be viewed as a generalization of this inequality and illustrates the use of c-stability. The first one uses c-stability of the sequence $(K_i)_1^\infty$ whereas the second assumes the c-stability of a set of kernels. In both results, the crucial point is that the unknown singular values $\sigma(K_i, \mu_{i-1})$ are replaced by expressions that depend on singular values that can, in many cases, be estimated. Theorem 4.7 is a simple corollary of the following theorem.

Theorem 4.10. *Fix $c \in (1,\infty)$. Let $(K_i)_1^\infty$ be a sequence of irreducible Markov kernels on a finite set V. Assume that $(K_i)_1^\infty$ is c-stable with respect to a positive probability measure μ_0. For each i, set $\mu_0^i = \mu_0 K_i$ and let $\sigma_1(K_i,\mu_0)$ be the second largest singular value of K_i as an operator from $\ell^2(\mu_0^i)$ to $\ell^2(\mu_0)$. Then*

$$d_2(K_{0,n}(x,\cdot),\mu_n) \leq \mu_0(x)^{-1/2}\prod_1^n (1 - c^{-2}(1-\sigma_1(K_i,\mu_0)^2))^{1/2}.$$

In addition,

$$\left|\frac{K_{0,n}(x,y)}{\mu_n(y)} - 1\right| \leq c^{1/2}(\mu_0(x)\mu_0(y))^{-1/2}\prod_1^n (1 - c^{-2}(1-\sigma_1(K_i,\mu_0)^2))^{1/2}.$$

Proof. First note that since $\mu_{i-1}/\mu_0 \in [1/c, c]$, we have $\mu_0^i/\mu_i \in [1/c, c]$. Consider the operator P_i of Remark 3.3 and its kernel

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

By assumption

$$\mu_i(x)P_i(x,y) \geq c^{-1}\mu_0^i(x)\left[\frac{1}{\mu_0^i(x)}\sum_z \mu_0(z)K_i(z,x)K_i(z,y)\right]$$

where the term in brackets on the right-hand side is the kernel of $K_i^* K_i$ where $K_i : \ell^2(\mu_0^i) \to \ell^2(\mu_0)$. This kernel has second largest eigenvalue $\sigma(K_i,\mu_0)^2$. A simple eigenvalue comparison argument yields

$$1 - \sigma_1(K_i,\mu_{i-1})^2 \geq \frac{1}{c^2}(1-\sigma_1(K_i,\mu_0)^2).$$

Together with Theorem 3.2, this gives the stated result. The last inequality in the theorem is simply obtained by replacing μ_n by μ_0 using c-stability. □

Theorem 4.11. *Fix $c \in (1,\infty)$. Let $\mathcal{Q}$ be a family of irreducible aperiodic Markov kernels on a finite set V. Assume that $\mathcal{Q}$ is c-stable with respect to some positive probability measure μ_0. Let $(K_i)_1^\infty$ be a sequence of Markov kernels with $K_i \in \mathcal{Q}$ for all i. Let π_i be the invariant measure of K_i. Let $\sigma_1(K_i)$ be the second largest singular value of K_i as an operator on $\ell^2(\pi_i)$. Then, we have*

$$d_2(K_{0,n}(x,\cdot),\mu_n) \le \mu_0(x)^{-1/2}\prod_1^n(1-c^{-4}(1-\sigma_1(K_i)^2))^{1/2}.$$

In addition,

$$\left|\frac{K_{0,n}(x,y)}{\mu_n(y)}-1\right| \le c^{1/2}(\mu_0(x)\mu_0(y))^{-1/2}\prod_1^n(1-c^{-4}(1-\sigma_1(K_i)^2))^{1/2}.$$

Proof. Recall that the hypothesis that $\mathcal{Q}$ is c-stable implies $\pi_i/\mu_i \in [1/c^2, c^2]$. Consider again the operator P_i of Remark 3.3 and its kernel

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

By assumption

$$\mu_i(x)P_i(x,y) \ge c^{-2}\pi_i(x)\left[\frac{1}{\pi_i(x)}\sum_z \pi_i(z)K_i(z,x)K_i(z,y)\right]$$

where the term in brackets on the right-hand side is the kernel of $K_i^*K_i$ where $K_i : \ell^2(\pi_i) \to \ell^2(\pi_i)$. As $\mu_i(x) \le c^2\pi_i(x)$, a simple eigenvalue comparison argument yields

$$1-\sigma_1(K_i,\mu_{i-1})^2 \ge \frac{1}{c^4}(1-\sigma_1(K_i)^2).$$

Together with Theorem 3.2, this gives the desired result. □

The following result is an immediate corollary of Theorem 4.11. It gives a partial answer to Problem 1.1 stated in the introduction based on the notion of c-stability.

Corollary 4.12. *Fix $c \in (1,\infty)$ and $\lambda \in (0,1)$. Let $\mathcal{Q}$ be a family of irreducible aperiodic Markov kernels on a finite set V. Assume that $\mathcal{Q}$ is c-stable with respect to some positive probability measure μ_0 and set $\mu_0^* = \min_x\{\mu_0(x)\}$. For any K in $\mathcal{Q}$, let π_K be its invariant measure and $\sigma_1(K)$ be the singular value of K on $\ell^2(\pi_K)$. Assume that, $\forall K \in \mathcal{Q}$, $\sigma_1(K) \le 1-\lambda$. Then, for any $\epsilon > 0$, the relative-sup ϵ-merging time of $\mathcal{Q}$ is bounded above by*

$$\frac{2c^4}{\lambda(2-\lambda)}\log\left(\frac{c^{1/2}}{\epsilon\mu_0^*}\right).$$

Remark 4.13. The kernels in Remark 2.5 are two reversible irreducible aperiodic kernels K_1, K_2 on the 2-point space so that the sequence obtained by alternating K_1, K_2 is not merging in relative-sup distance. While these two kernels satisfy $\max\{\sigma_1(K_1),\sigma_2(K_2)\} < 1$, $\mathcal{Q} = \{K_1, K_2\}$ fails to be c-stable for any $c > 0$. The family of kernels

$$\mathcal{Q} = \left\{M_p = \begin{pmatrix} p & 1-p \\ p & 1-p \end{pmatrix}, p \in (0,1)\right\}$$

has relative-sup merging time bounded by 1 but is not c-stable for any $c \geq 1$. To see that c-stability fails for $\mathcal{Q}$, note that we may choose a sequence $(K_n)_1^\infty$ such that $K_n = M_{p_n}$ where $p_n \to 0$. This shows that the c-stability hypothesis is not a necessary hypothesis for the conclusion to hold for certain $\mathcal{Q}$.

The following proposition describes a relation of merging in total variation to merging in the relative-max distance under c-stability. Note that we already noticed that without the hypothesis of c-stability the properties listed below are not equivalent.

Proposition 4.14. *Let V be a state space equipped with a finite family $\mathcal{Q}$ of irreducible Markov kernels. Assume that $\mathcal{Q}$ is c-stable and that either of the following conditions hold for each given $K \in \mathcal{Q}$:*

(i) K is reversible with respect to some positive probability measure $\pi > 0$.

(ii) There exists $\epsilon > 0$ such that K satisfies $K(x,x) > \epsilon$ for all x.

Then the following properties are equivalent:

1. *Each $K \in \mathcal{Q}$ is irreducible aperiodic (this is automatic under (ii)).*
2. *$\mathcal{Q}$ is merging in total variation.*
3. *$\mathcal{Q}$ is merging in relative-sup.*

Proof. Clearly the third listed property implies the second which implies the first. We simply need to show that the first property implies the third. Let $(K_n)_1^\infty$ be a sequence with $K_n \in \mathcal{Q}$ for all $n \geq 1$. Let π_n be the invariant measure for K_n and $\sigma_1(K_n)$ be its second largest singular value as an operator on $\ell^2(\pi_n)$.

Either of the conditions (i)-(ii) above implies that $\sigma_1(K_n) < 1$. Indeed, if (i) holds, $K_n = K_n^*$ and $\sigma_1(K_n) = \max\{\beta_1(K_n), -\beta_{|V|-1}(K_n)\}$ where $\beta_1(K_n)$ and $\beta_{|V|-1}(K_n)$ are the second largest and smallest eigenvalues of K_n respectively. It is well-known, for a reversible kernel K_n, the fact that K_n is irreducible aperiodic is equivalent to

$$\max\{\beta_1(K_n), -\beta_{|V|-1}(K_n)\} < 1.$$

If (ii) holds, Lemma 2.5 of [8] tells us that

$$\sigma_1(K_n) \leq 1 - \epsilon(1 - \beta_1(\tilde{K}_n))$$

where $\tilde{K}_n = 1/2(K_n + K_n^*)$. If K_n is irreducible aperiodic, so is $\tilde{K}_n$ and it follows that $\beta_1(\tilde{K}_n) < 1$. Hence $\sigma_1(K_n) < 1$. Since $\mathcal{Q}$ is a finite family, the $\sigma_1(K_n)$ can be bounded uniformly away from 1. Theorem 4.11 now yields that $\mathcal{Q}$ is merging in relative-sup. □

4.2 End-point perturbations for random walk on a stick

This section provides the first non-trivial example of c-stability.

Let $p,q,r \geq 0$ be fixed with $p+q+r=1$, $p \geq q$ and $\min\{q,r\} > 0$. On $V_N = \{0,\dots,N\}$, set

$$Q(x,y) = \begin{cases} p & \text{if } y = x+1 \text{ and } x \neq N \\ 1-p & \text{if } y = x \text{ and } x = 0 \\ q & \text{if } y = x-1 \text{ and } x \neq 0 \\ 1-q & \text{if } y = x \text{ and } x = N \\ r & \text{if } y = x \text{ and } x \neq 0, N, \end{cases} \tag{2}$$

and let ν be the associated reversible probability measure given by

$$\nu(x) = c_N (p/q)^{x-N}, \ \ c_N = \frac{1-(q/p)}{1-(q/p)^{N+1}} \ \text{ if } p \neq q,$$

and $\nu \equiv 1/(N+1)$ if $p = q$.

Fix $\epsilon \in (0,1)$ and let $\mathcal{Q}(p,q,r,\epsilon)$ be the set of all birth and death kernels $Q_{a,b}$ on $V_N = \{0,\dots,N\}$ of the form

$$Q_{a,b}(x,y) = \begin{cases} Q(x,y) & \text{if } x \neq 0, N \text{ or } |x-y| \geq 2 \\ a & \text{if } x = 0 \text{ and } y = 1 \\ 1-a & \text{if } x = y = 0 \\ b & \text{if } x = N \text{ and } y = N-1 \\ 1-b & \text{if } x = y = N, \end{cases} \tag{3}$$

with

$$p \leq a \leq 1-\epsilon, \ \ q \leq b \leq 1-\epsilon.$$

Proposition 4.15. *For any $\epsilon \in (0,1/2)$ and $\epsilon \leq p,q,r \leq 1-2\epsilon$, the set $\mathcal{Q}(p,q,r,\epsilon)$ is c-stable with respect to the measure ν with $c = 2 + 1/\epsilon$.*

This example is interesting in that it seems quite difficult to handle by inspection and algebra. Proving c-stability involves keeping track of huge amounts of cancellations, which appears to be rather difficult. We will use an extremely strong coupling property to obtain the result.

Proof. Consider a sequence $(Q_k)_1^\infty$ such that $Q_k = Q_{a_k,b_k} \in \mathcal{Q}(p,q,r,\epsilon)$. We construct a coupling $(W_t^1, W_t^2)_0^\infty$ such that marginally W_t^1 is driven by Q and W_t^2 is driven by $(Q_k)_1^\infty$.

For all $t \geq 0$, let W_t^1 be driven by Q. Set $W_0^1 = W_0^2$ and construct W_t^2 with the property that

$$\forall\, t \geq 0, \ \ |W_t^2 - W_t^1| \leq 1,$$

inductively as follows.

Case 1 $W_t^1, W_t^2 \in [1, N-1]$. Then

$$W_{t+1}^2 = \begin{cases} W_t^2 + 1 & \text{if } W_{t+1}^1 = W_t^1 + 1 \\ W_t^2 & \text{if } W_{t+1}^1 = W_t^1 \\ W_t^2 - 1 & \text{if } W_{t+1}^1 = W_t^1 - 1. \end{cases}$$

Note that in this case $|W_{t+1}^1 - W_{t+1}^2| = |W_t^1 - W_t^2|$.

Case 2 $W_t^1 = W_t^2 \in \{0, N\}$. Pick W_{t+1}^2 according to $Q_t(W_t^2, \cdot)$. Since W_t^1 and W_t^2 are at an end-point then it follows that $|W_{t+1}^1 - W_{t+1}^2| \le 1$.

Case 3(a) $W_t^1 = 1$ and $W_t^2 = 0$. Then

$$W_{t+1}^2 = \begin{cases} 1 & \text{if } W_{t+1}^1 = 2 \\ W_t^2 + Z_t & \text{otherwise} \end{cases} \tag{4}$$

where Z_t is the random variable with distribution (recall that $a_t > p$)

$$\mathbb{P}(Z_t = 1) = \frac{a_{t+1} - p}{1 - p} \text{ and } \mathbb{P}(Z_t = 0) = \frac{1 - a_{t+1}}{1 - p}.$$

Case 3(b) $W_t^1 = 0$ and $W_t^2 = 1$. Then

$$W_{t+1}^2 = \begin{cases} 2 & \text{if } W_{t+1}^1 = 1 \\ W_t^2 + Y_t & \text{otherwise} \end{cases}$$

where Y_t is a random variable with distribution

$$\mathbb{P}(Y_t = -1) = \frac{q}{1 - p} \text{ and } \mathbb{P}(Y_t = 0) = \frac{r}{1 - p}.$$

Case 4(a) $W_t^1 = N - 1$ and $W_t^2 = N$. Then

$$W_{t+1}^2 = \begin{cases} N - 1 & \text{if } W_{t+1}^1 = N - 2 \\ W_t^2 + \widetilde{Z}_t & \text{otherwise} \end{cases}$$

where $\widetilde{Z}_t$ is a random variable with distribution

$$\mathbb{P}(\widetilde{Z}_t = -1) = \frac{b_{t+1} - q}{1 - q} \text{ and } \mathbb{P}(\widetilde{Z}_t = 0) = \frac{1 - b_{t+1}}{1 - q}.$$

Case 4(b) $W_t^1 = N$ and $W_t^2 = N - 1$. Then

$$W_{t+1}^2 = \begin{cases} N - 2 & \text{if } W_{t+1}^1 = N - 1 \\ W_t^2 + \widetilde{Y}_t & \text{otherwise} \end{cases}$$

where $\widetilde{Y}_t$ has distribution

$$\mathbb{P}(\widetilde{Y}_t = 1) = \frac{p}{1 - q} \text{ and } \mathbb{P}(Y_t = 0) = \frac{r}{1 - q}.$$

One easily checks that, for all $t > 0$, given $W_0^2, \ldots, W_{t-1}^2, W_t^2 = x$, W_{t+1}^2 is distributed according to $Q_{t+1}(x, \cdot)$ and that $|W_t^2 - W_t^1| \le 1$.

Armed with this coupling, we now prove the stated stability property. For $x \in V_N$, set $A_x = \{z \in V_N : |z - x| \le 1\}$. We have

$$W_t^2 \in A_{W_t^1}, W_t^1 \in A_{W_t^2}.$$

If $W_0^1 = W_0^2$ have initial distribution ν then W_t^1 has distribution $\nu Q^t = \nu$ and W_t^2 has distribution $\nu_t = \nu K_{0,t}$. Moreover

$$\begin{aligned} \nu(x) &= \mathbb{P}_\nu(W_t^1 = x) \le \mathbb{P}(W_t^2 \in A_x) = \nu_t(A_x), \\ \nu_t(x) &= \mathbb{P}_\nu(W_t^2 = x) \le \mathbb{P}(W_t^1 \in A_x) = \nu(A_x). \end{aligned}$$

The second inequality and the explicit formula for ν gives

$$\nu_t(x) \le ((q/p) + 1 + (p/q))\nu(x) \le (2 + \epsilon^{-1})\nu(x).$$

Since $K_t(y,z) = Q_{a_t,b_t}(y,z) \ge \epsilon$ for all z and $y \in A_z$, we also have

$$\nu_t(x) = \nu_{t-1}K_t(x) \ge \epsilon \nu_{t-1}(A_x) \ge \epsilon \nu(x).$$

Hence

$$\forall\, t \ge 0,\ \ \epsilon \le \frac{\nu_t(x)}{\nu(x)} \le (2 + \epsilon^{-1}).$$

□

Having proved Proposition 4.15, Theorem 4.11, together with well-known results concerning the p, r, q simple random walk on the stick, yields the following result. The details are left to the reader.

Theorem 4.16. *Fix $\epsilon \in (0, 1/2)$ and $\epsilon \le p, r, q \le 1 - 2\epsilon$.*

1. *Assume that $p/q \ge 1 + \epsilon$. Then there is a constant $D(\epsilon)$ such that the family $\mathcal{Q}(p, q, r, \epsilon)$ has a total variation η-merging time bounded above by*

$$D(\epsilon)(N + \log_+ 1/\eta).$$

   Furthermore, for any sequence $K_i \in \mathcal{Q}(p, r, q, \epsilon)$, we have

$$n \ge D(\epsilon)(N + \log_+ 1/\eta) \Longrightarrow \max_{x,y,z} \left\{ \left| \frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1 \right| \right\} \le \eta.$$

2. *Assume that $p = q$. Then there is a constant $D(\epsilon)$ such that the family $\mathcal{Q}(p, q, r, \epsilon)$ has a total variation η-merging time bounded above by*

$$D(\epsilon)N^2(\log N + \log_+ 1/\eta).$$

   Furthermore, for any sequence $K_i \in \mathcal{Q}(p, r, q, \epsilon)$, we have

$$n \ge D(\epsilon)N^2(\log N + \log_+ 1/\eta) \Longrightarrow \max_{x,y,z} \left\{ \left| \frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1 \right| \right\} \le \eta.$$

Remark 4.17. In case 2, i.e., when $p = q$, the result stated above can be improved by using the Nash inequality techniques developed in [25]. This leads to bounds on the merging times of order N^2 instead $N^2 \log N$. Using singular values, we can get a hint that this is the case by considering the average ℓ^2 distance and the first inequality in Theorem 3.2. Indeed, in the case $p = q$, comparison with the simple random walk on the stick yields control not only of the top singular value, but of most singular values. Namely, if $(K_i)_1^\infty$ is a sequence in $\mathcal{Q}(p, p, r, \epsilon)$, $p, r \in [\epsilon, 1 - 2\epsilon]$ and $\mu_0 = \nu \equiv 1/(N+1)$, $\mu_n = \nu K_{0,n}$, then the j-th singular value $\sigma_j(K_i, \mu_{i-1})$ is bounded by

$$\sigma_j(K_i, \mu_{i-1}) \le 1 - C(\epsilon)^{-1} \left(\frac{j}{N}\right)^2.$$

Using this in the first inequality stated in Theorem 3.2 (together with stability) yields

$$\frac{1}{(N+1)^2} \sum_{x,y} \left| \frac{K_{0,n}(x,y)}{\mu_n(y)} - 1 \right|^2 \le \eta \quad \text{if} \quad n \ge D(\epsilon) N^2 (1 + \log_+(1/\eta)).$$

5 Stability of some inhomogeneous birth and death chains

Recall that a necessary condition for a family $\mathcal{Q}$ of irreducible aperiodic Markov kernels to be c-stable is that, for any pair of kernels in $\mathcal{Q}$, the associated stationary measures π, π' satisfy $c^{-2} \le \pi/\pi' \le c^2$. In less precise terms, all the stationary measures must have a similar behavior which we refer to as the stationary measure behavior of the family.

The goal of this section is to provide examples of c-stable families of Markov chains that allow for a great variety of stationary measure behaviors. Because we lack techniques to study c-stability, providing such examples is both important and not immediate. The examples presented in this section are sets of "perturbations" of birth and death chains having a center of symmetry. Except for this symmetry, the stationary measure of the birth and death chain that serves as the basis for the family is arbitrary. Hence, this produces examples with a wide variety of behaviors.

5.1 Middle edge perturbation for birth and death chains with symmetry

For $N \ge 1$, a general birth a death chain on $[-N, N]$ is described by

$$Q(x,y) = \begin{cases} p_x & \text{if } y = x+1 \\ q_x & \text{if } y = x-1 \\ r_x & \text{if } y = x \text{ and } y \ne -N, N \\ 0 & \text{otherwise} \end{cases}$$

with $p_N = q_{-N} = 0$ and $1 = p_x + q_x + r_x$ for all $x \in [-N, N]$. We consider the case when for all $x \in [-N, N]$

$$p_x = q_{-x} \text{ and } r_x = r_{-x}. \tag{1}$$

This immediately implies that $q_0 = p_0$. The kernel Q has reversible stationary distribution,

$$\nu(x) = c \prod_{y=-N+1}^{x} \frac{p_{y-1}}{q_y}$$

where c is the appropriate normalizing constant. Moreover, one checks that

$$\nu(-x) = \nu(x), \ \ x \in [-N, N].$$

For $\epsilon \in \mathbb{R}$, let Δ_ϵ be the (non-Markovian) kernel defined by

$$\Delta_\epsilon(x, y) = \begin{cases} \epsilon & \text{if } (x, y) = (0, 1) \\ -\epsilon & \text{if } (x, y) = (0, -1) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

For $\epsilon \in (-q_0, q_0)$, the perturbation $Q_\epsilon = Q + \Delta_\epsilon$ of Q is a Markov kernel and has stationary distribution

$$\nu_\epsilon(x) = \begin{cases} c_\epsilon \prod_{y=-N+1}^{z} \frac{p_{y-1}}{q_y} \frac{q_0-\epsilon}{q_0-\epsilon} & \text{if } z < 0 \\ c_\epsilon \prod_{y=-N+1}^{0} \frac{p_{y-1}}{q_y} \frac{q_0}{q_0-\epsilon} & \text{if } z = 0 \\ c_\epsilon \prod_{y=-N+1}^{z} \frac{p_{y-1}}{q_y} \frac{q_0+\epsilon}{q_0-\epsilon} & \text{if } z > 0 \end{cases} \tag{3}$$

where c_ϵ is a normalizing constant. Using the facts the $\sum \nu_\epsilon(x) = 1$ and (1) it follows that

$$\frac{q_0}{q_0 - \epsilon} c_\epsilon = c.$$

This implies that

$$\nu_\epsilon(x) = \begin{cases} \nu(z) - \frac{\epsilon \nu(z)}{q_0} & \text{if } z < 0 \\ \nu(0) & \text{if } z = 0 \\ \nu(z) + \frac{\epsilon \nu(z)}{q_0} & \text{if } z > 0. \end{cases}$$

Definition 5.1. Fix $\epsilon \in [0, q_0)$ and set $\mathscr{Q}_N(Q, \epsilon) = \{Q_\delta : \delta \in [-\epsilon, \epsilon]\}$.

Definition 5.2. Fix $\epsilon \in [0, q_0)$. Let $\mathscr{S}_N(\nu, \epsilon)$ be the set of probability measures μ on $[-N, N]$ satisfying the following two properties:

(1) for $x \in [-N, N]$, there exist constants $a_{\mu,x}$ such that $a_{\mu,x} = a_{\mu,-x}$ and

$$\mu(x) = \begin{cases} \nu(x) + a_{\mu,x} & \textit{if } x > 0 \\ \nu(x) - a_{\mu,x} & \textit{if } x < 0 \\ \nu(0) & \textit{if } x = 0; \end{cases}$$

(2) for all $x \in [0, N]$,

$$\nu_\epsilon(x) \geq \mu(x) \geq \nu_{-\epsilon}(x) \ \text{ and } \ \nu_\epsilon(-x) \leq \mu(-x) \leq \nu_{-\epsilon}(-x).$$

where ν_ϵ and $\nu_{-\epsilon}$ are defined in (3).

Theorem 5.3. *Fix $\epsilon \in [0, q_0)$. Let $\mu \in \mathscr{S}_N(\nu, \epsilon)$. Then for any $\delta \in [-\epsilon, \epsilon]$ we have $\mu Q_\delta \subseteq \mathscr{S}_N(\nu, \epsilon)$.*

Proof. Let $\mu \in \mathscr{S}_N(\nu, \epsilon)$.

(1) Set $a_0 = 0$, for any $x \in [1, N-1]$ we have that

$$\begin{aligned} \mu Q(x) &= p_{x-1}\mu(x-1) + r_x\mu(x) + q_{x+1}\mu(x+1) \\ &= \nu(x) + p_{x-1}a_{x-1} + r_x a_x + q_{x+1}a_{x+1} \end{aligned}$$

and

$$\begin{aligned} \mu Q(-x) &= p_{-x-1}\mu(-x-1) + r_{-x}\mu(-x) + q_{-x+1}\mu(-x+1) \\ &= \nu(-x) - p_{-x-1}a_{-x-1} - r_{-x}a_{-x} - q_{-x+1}a_{-x+1} \\ &= \nu(-x) - \left(p_{x-1}a_{x-1} + r_x a_x + q_{x+1}a_{x+1}\right). \end{aligned}$$

At the end-points $-N$ and N we get that

$$\begin{aligned} \mu Q(N) &= \mu(N)(1-q_N) + \mu(N-1)p_{N-1} \\ &= \nu(N) + a_N(1-q_N) + a_{N-1}p_{N-1} \end{aligned}$$

and

$$\begin{aligned} \mu Q(-N) &= \mu(-N)(1-p_{-N}) + \mu(-N+1)q_{-N+1} \\ &= \nu(-N) - a_{-N}(1-p_{-N}) - a_{-N+1}q_{-N+1} \\ &= \nu(N) - \left(a_N(1-q_N) + a_{N-1}p_{N-1}\right). \end{aligned}$$

Using similar arguments, it is easy to verify that $\mu Q(0) = \nu(0)$.
To check that μQ_δ satisfies (1) of Definition 5.2 we note that

$$\mu\Delta_\delta(x) = \begin{cases} \delta\mu(0) & \text{if } x = 1 \\ -\delta\mu(0) & \text{if } x = -1 \\ 0 & \text{otherwise.} \end{cases}$$

The desired result now follows from the fact that for any $x \in [-N, N]$

$$\mu Q_\delta(x) = \mu Q(x) + \mu\Delta_\delta(x).$$

(2) For any $\delta \in [-\epsilon, \epsilon]$ we have that

$$Q_\delta = Q_\epsilon - \Delta_\epsilon + \Delta_\delta = Q_\epsilon - \Delta_{\epsilon-\delta}.$$

It follows that $\nu_\epsilon Q_\delta = \nu_\epsilon - \nu_\epsilon \Delta_{\epsilon-\delta}$. We get that

$$\nu_\epsilon Q_\delta(x) = \begin{cases} \nu_\epsilon(1) - (\epsilon-\delta)\nu_\epsilon(0) & \text{if } x = 1 \\ \nu_\epsilon(-1) + (\epsilon-\delta)\nu_\epsilon(0) & \text{if } x = -1 \\ \nu_\epsilon(x) & \text{otherwise.} \end{cases}$$

and

$$\nu_{-\epsilon}Q_\delta(x) = \begin{cases} \nu_{-\epsilon}(1) + (\epsilon+\delta)\nu_{-\epsilon}(0) & \text{if } x = 1 \\ \nu_{-\epsilon}(-1) - (\epsilon+\delta)\nu_{-\epsilon}(0) & \text{if } x = -1 \\ \nu_{-\epsilon}(x) & \text{otherwise.} \end{cases}$$

Since $\delta \in [-\epsilon, \epsilon]$, we have that for $x \in [0, N]$

$$\nu_\epsilon Q_\delta(x) \le \nu_\epsilon(x) \text{ and } \nu_{-\epsilon}Q_\delta(x) \ge \nu_{-\epsilon}(x) \tag{4}$$

and also that,

$$\nu_\epsilon Q_\delta(-x) \ge \nu_\epsilon(-x) \text{ and } \nu_{-\epsilon}Q_\delta(-x) \le \nu_{-\epsilon}(-x). \tag{5}$$

By property (2) of μ being in $\mathscr{S}_N(\nu, \epsilon)$ we get that for any $x \in [1, N]$

$$\nu_\epsilon Q_\delta(x) \ge \mu Q_\delta(x) \ge \nu_{-\epsilon}Q_\delta(x) \text{ and } \nu_\epsilon Q_\delta(-x) \le \mu Q_\delta(-x) \le \nu_{-\epsilon}Q_\delta(-x).$$

Equations (5) and (4) imply that for $x \in [1, N]$

$$\nu_\epsilon(x) \ge \mu Q_\delta(x) \ge \nu_{-\epsilon}(x) \text{ and } \nu_\epsilon(-x) \le \mu Q_\delta(-x) \le \nu_{-\epsilon}(-x).$$

The inequalities above, along with the first part of the proof give the desired result. $\square$

Theorem 5.4. *Fix $\epsilon \in [0, q_0)$. The set $\mathscr{Q}_N(Q, \epsilon)$ is $\frac{q_0+\epsilon}{q_0-\epsilon}$-stable with respect to any measure $\mu_0 \in \mathscr{S}_N(\epsilon, \nu)$.*

Proof. Let $\mu_0 \in \mathscr{S}_N(\nu, \epsilon)$, and set $\mu_n = \mu_0 K_{0,n}$. By Theorem 5.3 it follows that for any $n \ge 0$ and any $x \in [0, N]$

$$\nu_\epsilon(x) \ge \mu_n(x) \ge \nu_{-\epsilon}(x) \text{ and } \nu_\epsilon(-x) \le \mu_n(-x) \le \nu_{-\epsilon}(-x).$$

So we have that for any $x \in [-N, N]$

$$\min_z \frac{\nu_\epsilon(z)}{\nu(z)} \le \frac{\mu_n(x)}{\nu(x)} \le \max_z \frac{\nu_\epsilon(z)}{\nu(z)}.$$

Recall that

$$c_\epsilon \left(\frac{q_0}{q_0 - \epsilon}\right) = c.$$

It follows that

$$\max_z \frac{\nu_\epsilon(z)}{\nu(z)} = \frac{c_\epsilon}{c}\left(\frac{q_0+\epsilon}{q_0-\epsilon}\right) = \frac{q_0+\epsilon}{q_0}, \quad \min_z \frac{\nu_\epsilon(z)}{\nu(z)} = \frac{c_\epsilon}{c} = \frac{q_0-\epsilon}{q_0}.$$

We get

$$\frac{q_0-\epsilon}{q_0} \le \frac{\mu_n(x)}{\nu(x)} \le \frac{q_0+\epsilon}{q_0}$$

which implies,

$$\frac{q_0-\epsilon}{q_0+\epsilon} \le \frac{\mu_n(x)}{\mu_0(x)} \le \frac{q_0+\epsilon}{q_0-\epsilon}.$$

$\square$

5.2 Example: the binomial chain

We now illustrate the above construction on a classical example, the (centered) binomial distribution $\pi(x) = 2^{-2N}\binom{2N}{N+x}$ on $V_N = \{-N,\dots,N\}$. The birth and death chain Q given by

$$Q(x,x+1) = \frac{N-x}{2N+1},\ Q(x,x-1) = \frac{N+x}{2N+1},\ Q(x,x) = \frac{1}{2N+1}$$

admits the binomial distribution π as reversible measure. It is obviously symmetric with respect to 0. Its second largest singular value(=eigenvalue) is $1 - 2/(2N+1)$. By Theorem 5.4, the set $\mathcal{Q}_N(\epsilon) = \mathcal{Q}_N(Q,\epsilon)$ with $\epsilon \in [0, q_0)$ is $(q_0+\epsilon)/(q_0-\epsilon)$-stable. Here, $q_0 = N/(2N+1)$. Hence we can apply Theorem 4.11 which yields a constant A (independent of N) such that, for any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}_N(q_0/2)$,

$$\max_{x,y,z}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1\right|\right\} \le \eta \text{ if } n \ge AN(N + \log 1/\eta). \tag{6}$$

This is a good example to point out that the present singular value technique is most precise when applied to bound the average ℓ^2-distance

$$\left(\sum_{x,y}\left|\frac{K_{0,n}(x,y)}{\mu_n(y)} - 1\right|^2 \mu_n(y)\mu_0(x)\right)^{1/2}$$

(here $\mu_0 = \pi$, $\mu_n = \pi K_{0,n}$ which, by c-stability, is comparable to π). Indeed, to bound this quantity, Theorem 3.2 allows us to use information on all singular values, not just the second largest. The singular values of Q not equal to 1 are the numbers $|2N+1-2i|/(2N+1)$ with multiplicity $\binom{2N}{i} + \binom{2N}{2N+1-i}$, $i = 1,\dots,N$ (this follows from the classical fact that the eigenvalues are $(2N+1-2i)/(2N+1)$ with multiplicity $\binom{2N}{i}$, $i = 0,\dots,2N$). The technique behind the proof of Theorem 4.11 yields

$$\sum_{x,y}\left|\frac{K_{0,n}(x,y)}{\mu_n(y)} - 1\right|^2 \mu_n(y)\mu_0(x) \le \eta^2$$

for $n \ge AN(\log N + \log 1/\eta)$. This indicates merging after order $N\log N$ steps whereas the bound (6) requires N^2 steps. Singular values alone do not yield a bound of order $N\log N$ for the ℓ^2-distance from a fixed starting point or for the relative-sup merging time. In [25], such an improved upper bound is obtained by using a logarithmic Sobolev inequality. Figure 1 in the introduction illustrates the merging in time of order $N\log N$ of a time inhomogeneous chain of this type driven by a sequence $(K_i)_1^\infty$ with $K_i \in \{Q_{\delta_1}, Q_{\delta_2}\} \subset \mathcal{Q}_N(Q,\delta_1)$ with

$$\delta_1 = \frac{N}{2(2N+1)} \text{ and } \delta_2 = -\frac{N}{4(2N+1)}.$$

In Figure 1, $N = 30$.

6 Two-point inhomogeneous Markov chains

This final section examines time inhomogeneous Markov chains on the two-point space. We characterize total variation merging and discuss stability and relative-sup merging.

6.1 Parametrization

Let $V = \{0,1\}$ be the two-point space. Let $0 \leq p \leq 1$, and set

$$M_p(x,y) = \begin{cases} p & \text{if } y = 0 \\ 1-p & \text{if } y = 1. \end{cases} \tag{1}$$

Obviously, this chain has stationary measure

$$\nu_p(z) = \begin{cases} p & \text{if } x = 0 \\ 1-p & \text{if } x = 1. \end{cases}$$

In fact, for any probability measure μ, $\mu M_p = \nu_p$. Let I denote the 2×2 identity matrix and consider the family of kernels

$$\mathcal{M}_p = \left\{ K[\alpha,p] : K[\alpha,p] = \left(\frac{1}{1+\alpha}\right)\left(\alpha I + M_p\right) : \alpha \in \left[-\min\{p,1-p\},\infty\right] \right\}.$$

By convention we have that $I = K[\infty,p]$ for all $0 \leq p \leq 1$. Any kernel in $\mathcal{M}_p$ has invariant measure ν_p and $\mathcal{M}_p$ is exactly the set of all Markov kernels on V with invariant measure ν_p. Indeed, if $q_1, q_2 \in [0,1]$ and

$$K = \begin{pmatrix} q_1 & 1-q_1 \\ q_2 & 1-q_2 \end{pmatrix}$$

then solving the equation $(1+\alpha)K = \alpha I + M_p$ yields

$$\alpha = \frac{q_1 - q_2}{1 + q_2 - q_1} \text{ and } p = \frac{q_2}{1+q_2-q_1}.$$

Note that $K[\alpha,p]$ has no holding in at least one point in V when $\alpha = -\min\{p,1-p\}$. Also, the only non irreducible kernels are I and those with $\min\{p,1-p\} = 0$ whereas the only irreducible periodic kernel is $K[-1/2,1/2]$. For later purpose, we note that the nontrivial eigenvalue $\beta_1(K[\alpha,p])$ of $K[\alpha,p]$ is given by

$$\beta_1(K[\alpha,p]) = \frac{\alpha}{1+\alpha}. \tag{2}$$

6.2 Total variation merging

Let $(K_i)_1^\infty$ with $K_i = K[\alpha_i, p_i]$, $p_i \in [0,1]$, $\alpha_i \in [-\min\{p_i,1-p_i\},\infty) \subset [-1/2,\infty)$. The following statement identifies $K_{0,n}$ using the (α,p)-parametrization.

Lemma 6.1.

$$K_{0,n} = \left(\frac{1}{1+\alpha_{0,n}}\right)\left(\alpha_{0,n} I + M_{p_{0,n}}\right) = K[\alpha_{0,n}, p_{0,n}]$$

where

$$\alpha_{0,n} = \frac{\prod_{i=1}^n \alpha_i}{\prod_{i=1}^n (1+\alpha_i) - \prod_{i=1}^n \alpha_i} \tag{3}$$

and

$$p_{0,n} = \frac{\sum_{i=1}^n \left(\prod_{j=1}^{i-1}(1+\alpha_j)\right)\left(\prod_{j=i+1}^n \alpha_j\right) p_i}{\prod_{i=1}^n (1+\alpha_i) - \prod_{i=1}^n \alpha_i}. \tag{4}$$

Proof. This can be shown by induction. Note that $K_{0,2} = K_1K_2$ is equal to

$$\frac{1}{1+\alpha_{0,2}}(\alpha_{0,2}I + M_{p_{0,2}})$$

with

$$\alpha_{0,2} = \frac{\alpha_1\alpha_2}{1+\alpha_1+\alpha_2}, \;\; p_{0,2} = \frac{\alpha_2 p_1 + (1+\alpha_1)p_2}{1+\alpha_1+\alpha_2}.$$

Using the two step formula above for $K_{0,n+1} = K_{0,n}K_{n+1}$, we have

$$\alpha_{0,n+1} = \frac{\alpha_{0,n}\alpha_{n+1}}{1+\alpha_{0,n}+\alpha_{n+1}} \;\; \text{and} \;\; p_{0,n+1} = \frac{\alpha_{n+1}p_{0,n} + (1+\alpha_{0,n})p_{n+1}}{1+\alpha_{0,n}+\alpha_{n+1}}. \tag{5}$$

To see that $\alpha_{0,n+1}$ and $p_{0,n+1}$ can be written in the forms of (3) and (4), we use the induction hypothesis along with the equality

$$1+\alpha_{0,n}+\alpha_{n+1} = \frac{\prod_{i=1}^{n+1}(1+\alpha_i) - \prod_{i=1}^{n+1}\alpha_i}{\prod_{i=1}^{n}(1+\alpha_i) - \prod_{i=1}^{n}\alpha_i}.$$

□

To study the merging properties of $(K_i)_1^\infty$, observe that, for any two initial probability measures μ_0 and ν_0 on V, we have

$$\begin{aligned} \mu_n - \nu_n &= \mu_0 K_{0,n} - \nu_0 K_{0,n} \\ &= \left(\frac{1}{1+\alpha_{0,n}}\right)\left(\alpha_{0,n}(\mu_0-\nu_0)I + (\mu_0-\nu_0)M_{p_{0,n}}\right) \\ &= \left(\frac{\alpha_{0,n}}{1+\alpha_{0,n}}\right)(\mu_0-\nu_0) = \left(\prod_{i=1}^{n}\frac{\alpha_i}{1+\alpha_i}\right)(\mu_0-\nu_0). \end{aligned} \tag{6}$$

This, together with elementary considerations, yields the following proposition.

Proposition 6.2. *The sequence $(K_i)_1^\infty$ is not merging in total variation if and only if the following three conditions hold:*

1. *For all i, $\alpha_i \neq 0$;*
2. *$\sum_{i:\alpha_i<0}(1+2\alpha_i) < \infty$;*
3. *$\sum_{i:\alpha_i>0}(1+\alpha_i)^{-1} < \infty$.*

Remark 6.3. The meaning of this proposition is that, in order to avoid merging we must prevent $\alpha_i = 0$ (i.e., $K_i = M_{p_i}$) for some i and have K_i approaching either I (no moves) or $K[-1/2, 1/2]$ (periodic chain) at fast enough rates. For instance, for $i = 1, 2, \ldots$, take

$$K_{2i+1} = \begin{pmatrix} 1-i^{-2} & i^{-2} \\ (2i)^{-2} & 1-(2i)^{-2} \end{pmatrix}, \;\; K_{2i} = \begin{pmatrix} i^{-2} & 1-i^{-2} \\ 1-(3i)^{-2} & (3i)^{-2} \end{pmatrix}.$$

Then $\alpha_{2i+1} = (4/5)i^2 - 1 \to \infty$ and $\alpha_{2i} = -(1-10i^{-2}/9)/(2-10i^{-2}/9) \to -1/2$ and there is no merging.

Proposition 6.4. *Let $\mathcal{Q} = \{K[\alpha_1, p_1], \dots, K[\alpha_m, p_m]\}$ be a finite family of irreducible Markov kernels on $V = \{0,1\}$. This family is merging in total variation if and only if $K[-1/2, 1/2] \notin \mathcal{Q}$, that is $(\alpha_i, p_i) \neq (-1/2, 1/2)$, $i = 1, \dots, m$.*

Proof. Recall that, by definition, $\mathcal{Q}$ is merging if any sequence made of kernels from $\mathcal{Q}$ is merging. If $K[-1/2, 1/2] \notin \mathcal{Q}$ then for any infinite sequence from $\mathcal{Q}$ either condition 2 or condition 3 of Proposition 6.2 is violated and we have merging. □

Remark 6.5. The family $\mathcal{Q}$ considered in Proposition 6.4 may not be merging in relative-sup distance even when $K[-1/2, 1/2] \notin \mathcal{Q}$. Indeed, Remark 6.9 gives an example of kernels $K[\alpha_i, p_i], K[\alpha_j, p_j] \neq K[-1/2, 1/2]$ such that the product $K[\alpha_i, p_i]K[\alpha_j, p_j]$ has an absorbing state at 0 and 1 is not absorbing. For any measure $\mu_0 > 0$, the sequence $(K_n)_1^\infty$, with $K_n = K[\alpha_i, p_i]$ if n is odd and $K_n = K[\alpha_j, p_j]$ otherwise, satisfies

$$\lim_{n\to\infty} \mu_{2n}(1) = 0.$$

Remark 6.6. Consider a sequence $(K_i)_1^\infty$ of Markov kernels on the two-point space such that (1) is satisfied. Write $K_i = K[\alpha_i, p_i]$. Because the kernels are reversible, (1) and formula (2) yield

$$|\alpha_i|/(1+\alpha_i) \leq \beta < 1.$$

This implies that the α_i's stay uniformly away from $-1/2$ and $+\infty$. By Proposition 6.2 and (6) we get

$$\|K_{0,n}(0,\cdot) - K_{0,n}(1,\cdot)\|_{\text{TV}} \leq \beta^n.$$

6.3 Stability

The study of stability on the 2-point space turns out to be quite interesting. We prove the following result which readily follows from the more precise statement in Proposition 6.10 below.

Theorem 6.7. *Fix $0 < \epsilon \leq \eta \leq 1/2$. Let $\mathcal{Q}(\epsilon, \eta)$ be the set of all Markov kernels $K[\alpha, p]$ on $\{0,1\}$ with*

$$p \in [\eta, 1-\eta] \text{ and } \alpha \in [-\min(p, 1-p) + \epsilon, \infty).$$

Then $\mathcal{Q}(\epsilon, \eta)$ is $(\epsilon^2\eta)^{-1}$-stable with respect to any measure μ_0 with $\mu_0(0) \in [\eta, 1-\eta]$.

The special case in the following proposition is easy but important for the treatment of the general case in Theorem 6.7.

Proposition 6.8. *Fix $\eta \in (0, 1/2)$ and let*

$$\mathcal{Q}(+, \eta) = \{K[\alpha, p] : \alpha \geq 0,\ p \in [\eta, 1-\eta]\}.$$

Then $\mathcal{Q}(+, \eta)$ is η^{-1}-stable with respect to any measure μ_0 with $\mu_0(0) \in [\eta, 1-\eta]$.

Proof. The crucial point is that when all α_i are non-negative then $p_{0,n}$ is a convex combination of the p_i, $1 \leq i \leq n$. Hence $p_{0,n} \in [\eta, 1-\eta]$ and $\mu_n(0) \in [\eta, 1-\eta]$. This gives the stated result. □

When the α_i are not all positive it is still possible to show c-stability but the proof is bit subtle. This illustrates in this simple case the intrinsic difficulties related to the notion of stability.

Remark 6.9. Consider the product $K[\alpha_1, p_1]K[\alpha_2, p_2] = K[\alpha_{0,2}, p_{0,2}]$ when $p_1 \neq p_2$, $\min(p_1, 1 - p_1) = 1 - p_1$, $\min(p_2, 1 - p_2) = p_2$, $\alpha_1 = -(1 - p_1)$ and $\alpha_2 = -p_2$. Then

$$p_{0,2} = \frac{\alpha_2 p_1 + (1+\alpha_1)p_2}{1+\alpha_1+\alpha_2} = \frac{-p_2 p_1 + p_1 p_2}{p_2 - p_1} = 0.$$

Proposition 6.10. *Let $K_i = K[\alpha_i, p_i]$, $i = 1, 2, \ldots$, be a sequence of Markov kernels on $\{0, 1\}$ with $K_i \in \mathcal{Q}(\epsilon, \eta)$. Then $p_{0,n} \in [\epsilon^2\eta, 1 - \epsilon^2\eta]$ for all $n \geq 1$.*

In order to prove this proposition, we need the following technical lemma.

Lemma 6.11. *Fix $0 \leq \epsilon \leq \eta \leq 1/2$. Let $K_1 = K[\alpha_1, p_1]$ and $K_2 = K[\alpha_2, p_2]$ be two Markov kernels on $\{0, 1\}$. Assume that $p_i \in [\eta, 1 - \eta]$, $i = 1, 2$.*

(1) If $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ then $p_{0,2} \in [\eta, 1 - \eta]$.

(2) If $\alpha_1 \in [-\min(p_1, 1 - p_1) + \epsilon, 0]$ and $\alpha_2 \geq 0$ then

$$\alpha_{0,2} \in [-\min(p_{0,2}, 1 - p_{0,2}) + \epsilon, 0] \text{ and } p_{0,2} \in [\eta, 1 - \eta].$$

(3) If $\alpha_1 \in [-\min(p_1, 1 - p_1) + \epsilon, \infty)$ and $\alpha_2 \in [-\min(p_2, 1 - p_2) + \epsilon, 0]$ then $p_{0,2} \in [\epsilon\eta, 1 - \epsilon\eta]$.

Proof. **(1)** The fact that $p_{0,2} \in [\eta, 1 - \eta]$ follows since $p_{0,2}$ is a convex combination of p_1 and p_2 and $\alpha_{0,2} \geq 0$ follows by Lemma 6.1.

(2) Since $1 + \alpha_1 \geq 0$ then $p_{0,2}$ is a convex combination of p_1 and p_2 so we get $p_{0,2} \in [\eta, 1 - \eta]$. To see that $\alpha_{0,2} \in [0, -\min(p_{0,2}, 1 - p_{0,2}) + \epsilon]$, Lemma 6.1 implies that we just need to check the following two inequalities:

(a) $|\alpha_1\alpha_2| \leq p_1\alpha_2 + (1 + \alpha_1)p_2 - \epsilon(1 + \alpha_1 + \alpha_2)$. This inequality follows from

$$\begin{aligned} p_1\alpha_2 + (1+\alpha_1)p_2 - \epsilon(1+\alpha_1+\alpha_2) + \alpha_1\alpha_2 \\ &= \alpha_2(p_1+\alpha_1) + (1+\alpha_1)p_2 - \epsilon(1+\alpha_1+\alpha_2) \\ &\geq \epsilon\alpha_2 + (1+\alpha_1)p_2 - \epsilon(1+\alpha_1+\alpha_2) \\ &= (1+\alpha_1)(p_2-\epsilon) \geq (1+\alpha_1)(\eta-\epsilon) \geq 0. \end{aligned}$$

(b) $|\alpha_1\alpha_2| \leq q_1\alpha_2 + (1 + \alpha_1)q_2 - \epsilon(1 + \alpha_1 + \alpha_2)$ where $q_i = 1 - p_i$. This inequality follows from the same calculations as in part (a) and the facts that $q_i \in [\eta, 1 - \eta]$ and $\alpha_i \geq -q_i + \epsilon$.

(3) Write

$$p_{0,2} = \frac{\alpha_2 p_1 + (1+\alpha_1)p_2}{1+\alpha_1+\alpha_2} = p_2 - |\alpha_2| \frac{(p_1 - p_2)}{1+\alpha_1+\alpha_2} \tag{7}$$

and observe that

$$\left| \frac{p_1 - p_2}{1+\alpha_1+\alpha_2} \right| = \frac{|p_1 - p_2|}{1+\alpha_1+\alpha_2} \leq \frac{|p_1 - p_2|}{|p_1 - p_2| + 2\epsilon}$$

The last inequality follows from $1+\alpha_1+\alpha_2 \geq p_1-p_2+2\epsilon$ and $1+\alpha_1+\alpha_2 \geq p_2-p_1+2\epsilon$. Note that $x/(x+2\epsilon)$ is an increasing function for $x \geq 0$, since $|p_1-p_2| \leq 1$ we get that

$$\left|\frac{p_1-p_2}{1+\alpha_1+\alpha_2}\right| \leq \frac{1}{1+2\epsilon} \leq 1-\epsilon.$$

If $p_1 \leq p_2$ then (7) implies $p_{0,2} \geq p_2 \geq \eta \geq \epsilon\eta$. With $q_2 = 1-p_2$, we get

$$p_{0,2} \leq 1-q_2+q_2\,\frac{p_2-p_1}{1+\alpha_1+\alpha_2} \leq 1-q_2+q_2(1-\epsilon) \leq 1-\epsilon q_2 \leq 1-\epsilon\eta.$$

If $p_2 \leq p_1$ then (7) implies $p_{0,2} \leq p_2 \leq 1-\eta \leq 1-\epsilon\eta$. For the lower bound we note that

$$p_{0,2} \geq p_2 - p_2\,\frac{p_1-p_2}{1+\alpha_1+\alpha_2} \geq p_2-p_2(1-\epsilon) \geq \epsilon p_2 \geq \epsilon\eta.$$

□

Proof of Proposition 6.10. To introduce convenient notation, if $K[\alpha,p] \in \mathscr{M}_p$ we say that $\alpha = \alpha(K)$ and $p = p(K)$. Let $(K_i)_1^\infty$ be the sequence of kernels considered in Proposition 6.10. Fix $n \geq 1$, and let $\{i_j\}_{j=1}^m$ be the set of numbers $1 \leq i_j \leq n$ such that $\alpha(K_{i_j}) < 0$. Set $i_0 = 0$ and consider the kernels

$$Q_j = K_{i_j, i_{j+1}-1} \quad \text{for } 0 \leq j \leq m$$

Note that for any $j \in [1,m]$ and $l \in [i_j, i_{j+1}-1]$ we have that $\alpha(K_l) \geq 0$ and $p(K_l) \in [\eta, 1-\eta]$. Note that either $Q_j = I$ or by Lemma 6.11 we have that $\alpha(Q_j) \geq 0$ and $p(Q_j) \in [\eta, 1-\eta]$. Now we write,

$$K_{0,n} = Q_0 K_{i_1} Q_1 \dots K_{i_j} Q_j \dots K_{i_m} Q_m.$$

For any $j \in [1,m]$, consider the kernel $M_j = K_{i_j} Q_j$. Since

$$\alpha(K_j) \in [-\min(p_i, 1-p_i)+\epsilon, 0] \quad \text{and} \quad \alpha(Q_j) \geq 0 \text{ or } Q_j = I,$$

it follows by Lemma 6.11 that

$$\alpha(M_j) \in [-\min(p(M_j), 1-p(M_j))+\epsilon, 0] \quad \text{and} \quad p(M_j) \in [\eta, 1-\eta]. \tag{8}$$

So now we write

$$K_{0,n} = Q_0 M_1 M_2 \dots M_m.$$

Consider the kernels $\widetilde{M}_i = M_{2i-1}M_{2i}$. It follows by Lemma 6.11 and (8) that $\alpha(\widetilde{M}_i) \geq 0$ and $p(\widetilde{M}_j) \in [\epsilon\eta, 1-\epsilon\eta]$. Let $B = Q_0\widetilde{M_1}$. Since $Q_0 = I$ or $\alpha(Q_0) \geq 0$ and $p(Q_0) \in [\eta, 1-\eta]$ by Lemma 6.11 we have that $\alpha(B) \geq 0$ and $p(B) \in [\epsilon\eta, 1-\epsilon\eta]$. If $m = 2k$ then we write

$$K_{0,n} = B\widetilde{M}_2 \dots \widetilde{M}_k.$$

For any $K \in \{B, \widetilde{M}_2, \dots, \widetilde{M}_k\}$ we have that $\alpha(K) \geq 0$ and $p(K) \in [\epsilon\eta, 1-\epsilon\eta]$, so by Lemma 6.11 we get $\alpha(K_{0,n}) \geq 0$ and $p(K_{0,n}) \in [\epsilon\eta, 1-\epsilon\eta]$.

If $m = 2k+1$ the we write

$$K_{0,n} = B\widetilde{M}_2 \dots \widetilde{M}_k M_m = M' M_m$$

where $M' = B\widetilde{M}_2 \dots \widetilde{M}_k$. By the same arguments as above we have $\alpha(M') \geq 0$ and $p(M') \in [\epsilon\eta, 1-\epsilon\eta]$. Lemma 6.11 and (8) we get that

$$p(K_{0,n}) \in [\epsilon^2\eta, 1-\epsilon^2\eta]$$

as desired. □

6.4 Stability and relative-sup merging

Given a sequence $K_i = K[\alpha_i, p_i]$ of Markov kernel on the two-point space, Lemma 6.1 indicates that relative-sup merging is equivalent to

$$\lim_{n\to\infty}\left|\sum_{i=1}^{n}\prod_{j=1}^{i-1}\left(1+\frac{1}{\alpha_j}\right)\frac{p_i}{\alpha_i}\right| = \infty \text{ and } \lim_{n\to\infty}\left|\sum_{i=1}^{n}\prod_{j=1}^{i-1}\left(1+\frac{1}{\alpha_j}\right)\frac{1-p_i}{\alpha_i}\right| = \infty. \tag{9}$$

Indeed, with the convention that these expressions are ∞ if $\alpha_i = 0$ for some $i \le n$, this is the same as saying that

$$\max\left\{\frac{|\alpha_{0,n}|}{p_{0,n}}, \frac{|\alpha_{0,n}|}{1-p_{0,n}}\right\} \to 0.$$

It does not appear easy to decide when (9) holds. Hence, even on the two-point space, c-stability is useful when studying relative-sup merging. This is illustrated by the results obtained below. The reader should note that this section falls short of providing statement analogous to the clean definitive result – Proposition 6.2 – obtained for merging in total variation.

Theorem 6.12. *Fix $0 < \epsilon < \eta \le 1/2$. The set $\widetilde{\mathcal{Q}}(\epsilon,\eta)$ of all Markov kernels $K[\alpha, p]$ on $\{0,1\}$ with*

$$p \in [\eta, 1-\eta] \text{ and } \alpha \in [-\min(p, 1-p) + \epsilon, \epsilon^{-1}).$$

is merging in relative-sup distance.

Proof. Fix $0 < \epsilon < \eta \le 1/2$. The assumptions above imply that the α_1's are bounded uniformly away from both $-1/2$ and $+\infty$. Proposition 6.2 implies that $\widetilde{\mathcal{Q}}(\epsilon,\eta)$ is merging in total variation. Proposition 4.14 and Theorem 6.7 now yield the desired result. □

Next, we address the case of sequences drawn from a finite family of kernels. We will need the following two simple results.

Lemma 6.13. *Let $\mathcal{Q} = \{K[\alpha_1, p_1], \ldots, K[\alpha_m, p_m]\}$ be a finite set of irreducible Markov kernels on $V = \{0,1\}$. Then $\mathcal{Q}$ is c-stable with respect to a positive probability measure μ_0 for some $c \in [1,\infty)$ if and only if*

$$\widetilde{\mathcal{Q}} = \{\widetilde{K} = KK' : K, K' \in \mathcal{Q}\}$$

is $\widetilde{c}$-stable with respect to μ_0 for some $\widetilde{c} \in [1,\infty)$.

Proof. Obviously, c-stability of $\mathcal{Q}$ implies c-stability of $\widetilde{\mathcal{Q}}$. In the other direction, fix a sequence $(K_i)_1^\infty$ from kernels in $\mathcal{Q}$ and consider the sequence $(\widetilde{K}_i)_1^\infty$ with $\widetilde{K}_i = K_{2i-1}K_{2i} \in \widetilde{\mathcal{Q}}$. By hypothesis, this sequence is $\widetilde{c}$-stable with respect μ_0. Because the $K_i's$ are irreducible and drawn from a finite set of kernels, this implies that the sequence $(K_i)_1^\infty$ is c-stable with respect to μ_0 for some $c \in [1,\infty)$. □

Lemma 6.14. *Let $K = K[\alpha, p], K' = K[\alpha', p']$ be two irreducible Markov kernels on $V = \{0,1\}$ and set $KK' = \widetilde{K}$. Then one of the following three alternatives is satisfied:*

1. *$\widetilde{K} = I$, which happens if and only if $K = K' = K[-1/2, 1/2]$.*

2. $\widetilde{K}$ *has a unique absorbing state, which happens if and only if* $K \neq K'$ *and either* $\alpha = -p, \alpha' = -(1-p')$ *or* $\alpha = -(1-p), \alpha' = -p'$.

3. $\widetilde{K} = K[\widetilde{\alpha}, \widetilde{p}]$ *is irreducible with* $\widetilde{p} \in (0,1)$ *and* $\widetilde{\alpha} > -\min\{\widetilde{p}, 1-\widetilde{p}\}$.

Proof. The condition $\widetilde{p} \in (0,1)$ and $\widetilde{\alpha} > -\min\{\widetilde{p}, 1-\widetilde{p}\}$ is equivalent to saying that $\widetilde{K}$ has only strictly positive entries. Since $K, K' \neq I$, $\widetilde{K} = I$ can only occur when $K = K' = K[-1/2, 1/2]$. Further, $\widetilde{K}(0,1) > 0$ unless $K(0,1) = K'(1,0) = 1$. Similarly, $\widetilde{K}(1,0) > 0$ unless $K(1,0) = K'(0,1) = 1$. This implies condition 2. If $\widetilde{K}$ does not have absorbing states then condition 3 must hold. □

Theorem 6.15. *Let* $\mathscr{Q} = \{K[\alpha_1, p_1], \dots, K[\alpha_m, p_m]\}$ *be a finite set of irreducible Markov kernels on* $V = \{0,1\}$. *This finite set is* c*-stable if and only if there are no pairs of distinct indices* $i, j \in \{1, \dots, m\}$ *such that*

$$\alpha_i = -p_i \text{ and } \alpha_j = -(1-p_j). \tag{10}$$

Note that the set $\mathscr{Q}$ in Theorem 6.15 might very well contained the kernel $K[-1/2, 1/2]$. Note also that (10) implies that $\max\{p_i, 1-p_j\} \leq 1/2$.

Proof of Theorem 6.15. The fact that there are no pairs of distinct $i, j \in \{1, \dots, m\}$ such that for $K[\alpha_i, p_i], K[\alpha_j, p_j] \in \mathscr{Q}$ condition (10) is satisfied is necessary for c-stability. Indeed, if there is such a pair, consider the sequence $(K_n)_1^\infty$ with $K_n = K[\alpha_i, p_i]$ if n is odd and $K_n = K[\alpha_j, p_j]$ otherwise. Note that $K_{0,2n} = Q^n$ where $Q = K[\alpha_i, p_i]K[\alpha_j, p_j]$. Equation (5) yields $Q = K[\tilde{\alpha}, \tilde{p}]$ with

$$\tilde{\alpha} = \frac{p_i(1-p_j)}{p_j - p_i} \text{ and } \tilde{p} = 1.$$

Since Q has a unique absorbing state at 0, for any $\mu_0 > 0$ we have that

$$\lim_{n\to\infty} \mu_{2n}(1) = \lim_{n\to\infty} \mu_0 K_{0,2n}(1) = \lim_{n\to\infty} \mu_0 Q^n(1) = 0.$$

Hence, the family $\mathscr{Q}$ cannot by c-stable according to Definition 4.2.

Assume now that there are no pairs of distinct $i, j \in \{1, \dots, m\}$ such that (10) is satisfied. By Lemma 6.13, to prove stability with respect to the uniform measure (or any positive probability), it suffices to prove the stability of $\widetilde{\mathscr{Q}} = \{K_i K_j : K_i, K_j \in \mathscr{Q}\}$. By Lemma 6.14 the kernels in $\widetilde{\mathscr{Q}}$ are either equal to I or are in the set $Q(\epsilon, \eta)$ of Theorem 6.7. Hence, Theorem 6.7 yields the c-stability of $\tilde{\mathscr{Q}}$. □

Corollary 6.16. *Let* $\mathscr{Q} = \{K[\alpha_1, p_1], \dots, K[\alpha_m, p_m]\}$ *be a finite set of irreducible Markov kernels on* $V = \{0,1\}$. $\mathscr{Q}$ *is merging in the relative-sup distance if and only if the following two conditions are satisfied.*

(1) $K[-1/2, 1/2] \notin \mathscr{Q}$.

(2) There are no pairs of distinct indicies $i, j, \in \{1, \dots, m\}$ *satisfying* (10).

Proof. Assume that $K[-1/2, 1/2] \in \mathscr{Q}$ and consider the sequence $(K_n)_1^\infty$ with $K_n = K[-1/2, 1/2]$ for all $n \geq 1$. For any $n \geq 1$, $K[-1/2, 1/2]^n$ is the either the identity or $K[-1/2, 1/2]$ depending on the parity of n. This chain clearly does not merge in the relative-sup distance according to Definition 2.4.

Assume that condition 2 above is not satisfied and there exists a pair of distinct indicies i, j such that (10) holds. By the proof of Theorem 6.15 we get that the kernel $Q = K[\alpha_i, p_i]K[\alpha_j, p_j]$ has an absorbing state at 0. Consider the sequence $(K_n)_1^\infty$ where $K_n = K[\alpha_i, p_i]$ if n is odd and $K_n = K[\alpha_j, p_j]$ otherwise. It follows that the family $\mathcal{Q}$ is not merging in relative-sup distance since, for any $n \geq 1$, $Q^n(1,1) > 0$ and

$$\max_{x,y,z}\left\{\left|\frac{K_{0,2n}(x,z)}{K_{0,2n}(y,z)} - 1\right|\right\} = \max_{x,y,z}\left\{\left|\frac{Q^n(x,z)}{Q^n(y,z)} - 1\right|\right\} = \left|\frac{Q^n(1,1)}{Q^n(0,1)} - 1\right| = \infty.$$

If a family $\mathcal{Q}$ satisfies both conditions 1 and 2 above, Propositions 4.14 and 6.4 along with Theorem 6.15 yield the desired relative-sup merging. □

Remark 6.17 (Solution of Problem 1.1(2) on the 2-point space). In Problem 1.1, we make three main hypotheses:

- (H1) All reversible measure π_i are comparable (with comparison constant $c \geq 1$). In the case of the 2-point space, this means there exists $\eta(c) \in (0, 1/2)$ such that $p \in [\eta(c), 1 - \eta(c)]$.
- (H2) Inequality (1) holds. Because all the kernels involved are reversible, this implies that the second largest eigenvalue in modulus, $\sigma_1(K_i, \pi_i) = \sigma_1(K_i)$, satisfies $\sigma_1(K_i) \leq \beta$. Formula (2) shows that $\beta_1(K[\alpha, p]) = |\alpha|/(1+\alpha)$. So, on the 2-point space, (1) yields $\alpha \in [-\frac{1}{2} + \epsilon, \epsilon^{-1}]$ for some $\epsilon \in (0, 1)$.
- (H3) Uniform holding, i.e., $\min_x\{K(x,x)\} \geq c^{-1}$. One the 2-point space,

  $$\min_x\{K[\alpha, p](x,x)\} = \frac{\alpha + \min(p, 1-p)}{1+\alpha}.$$

  Hence, we get that $\alpha \geq -\min(p, 1-p) + (2c)^{-1}$.

The discussion above shows that the hypotheses (H1)-(H2)-(H3) imply that each K_i in $(K_i)_1^\infty$ belongs to the set $\mathcal{Q}(\eta(c), (2c)^{-1})$ of Theorem 6.7 which provides the stability of the family in question. Thus Theorem 6.7 and 4.12 yield the desired relative-sup merging property.

References

[1] Aldous, D. *Random walks on finite groups and rapidly mixing Markov chains.* Seminar on probability, XVII, 243–297, Lecture Notes in Math., 986, Springer, 1983. MR0770418

[2] Cohen, J. *Ergodicity of age structure in populations with Markovian vital rates. III. Finite-state moments and growth rate. An illustration.* Advances in Appl. Probability 9 (1977), no. 3, 462–475. MR0465278

[3] D'Aristotile, A., Diaconis, P. and Freedman, D. *On merging of probabilities.* Sankhya: The Indian Journal of Statistics, 1988, Volume 50, Series A, Pt. 3, pp. 363-380. MR1065549

[4] Del Moral, P., Ledoux, M. and Miclo, L. *On contraction properties of Markov kernels* Probab. Theory Related Fields 126 (2003), 395–420. MR1992499

[5] Diaconis, P. *Group representations in probability and statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988. MR0964069

[6] Diaconis, P. and Shahshahani, M. *Generating a random permutation with random transpositions.* Z. Wahrsch. Verw. Gebiete 57 (1981), 159–179. MR0626813

[7] Diaconis, P. and Saloff-Coste, L. *Comparison theorems for reversible Markov chains.* Ann. Appl. Probab. 3 (1993), 696–730. MR1233621

[8] Diaconis, P. and Saloff-Coste, L. *Nash inequalities for Finite Markov chains.* J. Theoret. Probab. 9 (1996), 459–510. MR1385408

[9] Diaconis, P. and Saloff-Coste, L. *Logarithmic Sobolev inequalities for finite Markov chains*. Ann. Appl. Probab. 6 (1996), no. 3, 695–750. MR1410112

[10] Diaconis, P. and Saloff-Coste, L. *What do we know about the Metroplis algorithm?* J. Comp. and System Sci. 57 (1998) 20–36. MR1649805

[11] Douc, R.; Moulines, E. and Rosenthal, Jeffrey S. *Quantitative bounds on convergence of time-inhomogeneous Markov chains.* Ann. Appl. Probab. 14 (2004), 1643–1665. MR2099647

[12] Fill, J. *Eigenvalue bounds on convergence to stationarity for non-reversible Markov chains, with an application to the exclusion process.* Annals Appl. Probab. 1 (1991) 62–87. MR1097464

[13] Ganapathy, M. *Robust mixing time.* Proceedings of RANDOM'06.

[14] Hajnal, J. *The ergodic properties of non-homogeneous finite Markov chains.* Proc. Cambridge Philos. Soc. 52 (1956), 67–77. MR0073874

[15] Horn, R and Johnson, C. *Topics in matrix analysis.* Cambridge; New York: Cambridge University Press, 1991. MR1091716

[16] Iosifescu, M. *Finite Markov processes and their applications.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester; Editura Tehnică, Bucharest, 1980. MR0587116

[17] Morris, B. and Peres, Y. *Evolving sets, mixing and heat kernel bounds.* Probab. Theory Related Fields 133 (2005) 245–266. MR2198701

[18] Mossel, E, Peres, Y. and Sinclair A. *Shuffling by semi-random transpositions*. 45th Symposium on Foundations of Comp. Sci. (2004) arXiv:math.PR/0404438

[19] Păun, U. *Ergodic theorems for finite Markov chains.* Math. Rep. (Bucur.) 3(53) (2001), 383–390 (2002). MR1990903

[20] J. R. Ringrose. *Compact non-self-adjoint operators.* London: Van Nostrand Reinhold Math. Studies 1971

[21] Saloff-Coste, L. *Lectures on finite Markov chains.* Lectures on probability theory and statistics (Saint-Flour, 1996), 301–413, Lecture Notes in Math., 1665, Springer, Berlin, 1997. MR1490046

[22] Saloff-Coste, L. *Random walks on finite groups.* In Probability on discrete structures (H. Kesten, Ed.), 261-346, Encyclopaedia Math. Sci., 110, Springer, Berlin, 2004. MR2023654

[23] Saloff-Coste, L. and Zúñiga, J. *Convergence of some time inhomogeneous Markov chains via spectral techniques.* Stocahstic Processes and their Appl. 117 (2007) 961–979. MR2340874

[24] Saloff-Coste, L. and Zúñiga, J. *Refined estimates for some basic random walks on the symmetric and alternating groups.* Latin American Journal of Probability and Mathematical Statistics (ALEA) 4, 359-392 (2008). MR2461789

[25] Saloff-Coste, L. and Zúñiga, J. *Merging of time inhomogeneous Markov chains, part II: Nash and log-Sobolev inequalities* (Submitted.)

[26] Saloff-Coste, L. and Zúñiga, J. *Time inhomogeneous Markov chains with wave like behavior.* (Submitted.)

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