HAL
open science

On the spectral analysis of second-order Markov chains

Persi Diaconis, Laurent Miclo

To cite this version:

Persi Diaconis, Laurent Miclo. On the spectral analysis of second-order Markov chains. Annales de la Faculté des Sciences de Toulouse. Mathématiques., 2013, 22 (3), pp.573-621. hal-00719047

HAL Id: hal-00719047

https://hal.science/hal-00719047v1

Submitted on 18 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the spectral analysis of second-order Markov chains

Persi Diaconis[†] and Laurent Miclo[‡]

[†]Department of Mathematics
Stanford University, California, USA

[‡]Institut de Mathématiques de Toulouse, UMR 5219
Université de Toulouse and CNRS, France

Abstract

Second order Markov chains which are trajectorially reversible are considered. Contrary to the reversibility notion for usual Markov chains, no symmetry property can be deduced for the corresponding transition operators. Nevertheless and even if they are not diagonalizable in general, we study some features of their spectral decompositions and in particular the behavior of the spectral gap under appropriate perturbations is investigated. Our quantitative and qualitative results confirm that the resort to second order Markov chains is an interesting option to improve the speed of convergence to equilibrium.

Keywords: second order Markov chains, trajectorial reversibility, spectrum of non-reversible Markov kernels, spectral gap, improvement of speed of convergence, discrete position-speed Markov chains, homogeneous and Abelian second order Markov chains.

MSC2010: first: 60J10, secondary: 37A30, 47A55, 15A18, 60B15, 65C40.

1 Introduction

Resorting to nonreversible Markov chains or to second order Markov chains may present some important advantages from a practical simulation point of view. This feature was emphatized by Neal [10] who showed how the asymptotical variance can be diminished for empirical estimators of expectations, when one uses stochastic chains which have a tendency of not returning back where they were just coming from. Our main results here go in the same direction, but we will rather be interested in the asymptotic speed of convergence to equilibrium, measured through the spectral gap. This will lead us to study the whole spectrum of some second order Markov chains and its behavior under perturbations designed to improve them.

A stochastic chain $(X_n)_{n\in\mathbb{N}}$ is said to be a second order Markov chain if at any time $n \in \mathbb{N}$, its future only depends on its past through the present position and the previous one, namely if we have the following almost sure equality for the conditional laws

$$\mathcal{L}((X_m)_{m\geq n+1}|(X_m)_{m\in[\![0,n]\!]}) \quad = \quad \mathcal{L}((X_m)_{m\geq n+1}|X_n, X_{n-1})$$

(by convention $X_{-1} = X_0$). Equivalently, this means that the stochastic chain $(X_n, X_{n+1})_{n\in\mathbb{N}}$ is Markovian.
We will only consider time-homogeneous second order finite Markov chains, i.e. the X_n, for $n \in \mathbb{N}$, take value in a finite state space V and for any $n \in \mathbb{N} \setminus \{0\}$ and $x, x', x'' \in V$, the quantity

$$M(x, x'; x'') \quad := \quad \mathbb{P}[X_{n+1} = x''|X_n = x', X_{n-1} = x] \tag{1}$$

does not depend on n (when it is well-defined, i.e. when the event $\{X_n = x', X_{n-1} = x\}$ is not negligeable with respect to the underlying probability $\mathbb{P}$). Conversely, if we are given a Markov kernel $M := (M(x, x'; x''))_{x,x',x''\in V}$ from V^2 to V and a probability distribution $m_0 = (m_0(x, x'))_{x,x'\in V}$ on V^2, there exists a unique (in law) stochastic chain $(X_n)_{n\in\mathbb{N}}$ on V such that the equality (1) holds and such that the initial law of (X_0, X_1) is m_0. The probability distribution m on V^2 is said to be trajectorially reversible with respect to the kernel M, if the corresponding second order Markov chain starting with the initial law m satisfies for any $n \in \mathbb{N}$, the identity in law

$$\mathcal{L}(X_0, X_1, \cdots, X_n) \quad = \quad \mathcal{L}(X_n, X_{n-1}, \cdots, X_0) \tag{2}$$

This notion of reversibility is different from the usual reversibility for the Markov chain $(X_n, X_{n+1})_{n\in\mathbb{N}}$, which is not satisfied in general. As it was shown by Bacallado and Pande [2], it is sufficient to check the equality (2) for $n = 1$ and $n = 2$. In particular the distribution m is symmetrical in its first and second variables. Let us write

$$\begin{aligned} \forall\, x, x' \in V, \qquad m(x, x') &= \nu(x)L(x, x') \\ &=: (\nu \times L)(x, x') \end{aligned} \tag{3}$$

where ν is the law of X_0 and L is a Markov kernel corresponding to the conditional law under m of X_1 when X_0 is given (it is unique if ν is positive). By symmetry, ν is reversible (in the usual sense) with respect to L. Let $(\widetilde{X}_n)_{n\in\mathbb{N}}$ be a Markov chain on V, starting with ν as initial distribution and whose transition kernel is given by L. It is not difficult to see that the distribution m is invariant for the Markov chain $(X_n, X_{n+1})_{n\in\mathbb{N}}$ and by consequence we are insured of the identities in law

$$\forall\, n \in \mathbb{N}, \qquad \mathcal{L}(X_n, X_{n+1}) \quad = \quad \mathcal{L}(\widetilde{X}_n, \widetilde{X}_{n+1})$$

In particular, appropriate assumptions of aperiodicity and irreducibility on L and M will insure that whatever the initial distributions,

$$\begin{aligned} \lim_{n\to\infty} \mathcal{L}(X_n, X_{n+1}) &= m \\ &= \lim_{n\to\infty} \mathcal{L}(\widetilde{X}_n, \widetilde{X}_{n+1}) \end{aligned}$$

and our main goal in this paper is to compare the corresponding speeds of convergence. They will be measured via spectral gaps, which give the asymptotical convergence rates to equilibrium. Let M be extended as a Markov kernel on V^2 through the formula

$$\forall\ (x,x'),(x'',x''')\in V^2,\qquad M((x,x'),(x'',x''')) \ \coloneqq\ \delta_{x'}(x'')M(x,x';x''') \tag{4}$$

(where $\delta_{x'}$ is the Dirac function at x'), i.e. we are just reinterpreting M as the transition kernel of the Markov chain $(X_n,X_{n+1})_{n\in\mathbb{N}}$. This enables us to see M as a right operator on the space $\mathcal{F}(V^2)$ of real-valued functions defined on V^2 (or on the space $\mathcal{F}(V^2,\mathbb{C})$ of complex plane-valued functions defined on V^2) and as a left operator on the space $\mathcal{M}(V^2)$ of signed measures on V^2. We have for instance

$$\forall\ F\in\mathcal{F}(V^2),\ \forall\ (x,x')\in V^2,\qquad M[F](x,x') \ \coloneqq\ \sum_{x''\in V} M(x,x';x'')F(x',x'')$$

By invariance of m, the subspace $\mathcal{F}_*(m,\mathbb{C})\coloneqq\{F\in\mathcal{F}(V^2,\mathbb{C})\ :\ m[F]=0\}$ is stable by M, so let $\Theta_*(M)\subset\mathbb{C}$ be the spectrum of M_*, the restriction of M to $\mathcal{F}_*(m,\mathbb{C})$. The spectral gap of M is

$$\lambda(M) \ \coloneqq\ 1-\max\{|\theta|\ :\ \theta\in\Theta_*(M)\}\ \in\ [0,1]$$

Then, if $\|\cdot\|$ is a norm on $\mathcal{M}(V^2)$, we have for any initial distribution m_0 on V^2,

$$\lim_{n\to\infty}\frac{1}{n}\ln\left(\|m_0M^n-m\|\right) \ \leq\ \ln(1-\lambda(M))$$

and there exists an initial point $(x,x')\in V^2$ such that

$$\lim_{n\to\infty}\frac{1}{n}\ln\left(\left\|\delta_{(x,x')}M^n-m\right\|\right) \ =\ \ln(1-\lambda(M))$$

Similar definitions can be given for the spectral gap of L, by replacing the state space V^2 by V. One can alternatively consider the second order Markov kernel $M^{(0)}$ naturally associated to L:

$$\forall\ x,x',x''\in V,\qquad M^{(0)}(x,x';x'') \ \coloneqq\ L(x',x'') \tag{5}$$

(it corresponds to the transition kernel of the Markov chain $(\widetilde{X}_n,\widetilde{X}_{n+1})_{n\in\mathbb{N}}$). We will check that the spectrum of $M^{(0)}$ is just the spectrum of L (with the same multiplicities, except for the eigenvalue 0) plus the eigenvalue 0, so that $\lambda(L)=\lambda(M^{(0)})$. Note that since ν is reversible with respect to L, this operator is diagonalizable in $\mathbb{R}$ and we have $\Theta_*(M^{(0)})\subset\mathbb{R}$, but in general $M^{(0)}$ won't be diagonalizable: we will see that it contains 2×2-Jordan blocks if 0 is an eigenvalue of L.

Our first result states that except in the cases where $L=\nu$ (where the probability measure ν is seen as a Markov kernel via the formulas: $\forall\ x,x'\in V,\ \nu(x,x')=\nu(x')$, this convention will be assumed enforced in the whole paper) or where L is aperiodic (which by reversibility means that the graph induced by L is bipartite), it is always possible to find a second order Markov chain M which admits $\nu\times L$ as trajectorial reversible measure and which has a larger spectral gap than L. Note that the largest possible spectral gap is 1 and that the equality $L=\nu$ is equivalent to the fact that $\lambda(L)=1$, so in this situation the spectral gap cannot be improved by any means. In the aperiodic case, the spectral gap cannot be improved either: let V' and V'' be proper subsets of V such that $V=V'\sqcup V''$ and such that the $V'\times V'$ and $V''\times V''$ entries of L are null. Then if M admits $\nu\times L$ as trajectorial reversible measure, necessary M is also periodic, since the Markov chain $(X_n,X_{n+1})_{n\in\mathbb{N}}$ goes from the set $V'\times V''$ to the set $V''\times V'$ and conversely.

Indeed, let $M^{(1)}$ be a second order Markov kernel admitting $\nu\times L$ as trajectorial reversible measure and satisfying

$$\forall\ x,x',x''\in V,\qquad M^{(1)}(x,x';x'') \ >\ L(x',x'')\qquad \text{, if } x''\neq x,\ L(x,x')>0 \text{ and } L(x',x'')>0$$

(in particular $M^{(1)}$ has less tendency to come back where it just came from than $M^{(0)}$). It is not difficult to see that such kernels exist. For any $a \in [0,1]$, define

$$M^{(a)} \quad := \quad (1-a)M^{(0)} + aM^{(1)}$$

which is a second order Markov kernel still admitting $\nu \times L$ as trajectorial reversible measure. Then we have

Theorem 1 *Assume that L is irreducible, that $L \neq \nu$ and that L is aperiodic. The mapping $[0,1] \ni a \mapsto \lambda(M^{(a)})$ is differentiable at 0 and we have*

$$\frac{d\lambda(M^{(a)})}{da}(0) \quad > \quad 0$$

In fact we will show a better result, saying that under the above assumptions the whole spectrum of L outside 0 has an initial tendency to improve: let $[0,a_0] \ni a \mapsto \theta(a)$ be a differentiable function such that $\theta(a) \in \Theta_*(M^{(a)})$ for $a \in [0,a_0]$ (where $a_0 \in (0,1)$ will be small enough in practice) and such that $\theta(0) \neq 0$. Since $\theta(0) \in \Theta_*(L)$, it is real. Then we will prove that $d\theta(a)/da(0)$ is real and that its sign is the opposite sign of $\theta(0)$, namely initially, when a starts to increase from 0, $\theta(a)$ is going in the direction of 0.
But if instead of improving the whole spectrum outside 0, one only wants to improve the spectral gap, there is a more efficient way to do it, which uses the ordering of the space created by an eigenfunction associated to the eigenvalue of L corresponding to the spectral gap (under the assumption that such an eigenfunction is unique up to a factor) as it will be explained in detail in Section 3. Nevertheless, it should be kept in mind that to get quantitative bounds on the distance to equilibrium (for instance in the chi-square sense), not just an asymptotical result, the knowledge of the whole spectrum is preferable to the spectral gap.

In order to get a global picture of how the spectral gap (and indeed the whole spectrum) depends on the parameter $a \in [0,1]$, we consider a family of homogeneous-space simple examples. Assume that V is a group G. Let $(X_n)_{n\in\mathbb{N}}$ be a second order Markov chain with kernel M and define the corresponding speed process $(Y_n)_{n\in\mathbb{N}}$ by

$$\forall\ n \in \mathbb{N}, \qquad Y_n \quad := \quad X_n^{-1}X_{n+1}$$

The position-speed process $(X_n, Y_n)_{n\in\mathbb{N}}$ is Markovian and its transition kernel can be written as

$$\forall\ (x,y),(x',y') \in G^2, \qquad P((x,y),(x',y')) \quad = \quad \delta_{xy}(x')K_{x'}(y,y') \tag{6}$$

where the Markov kernel of transition from speed y to speed y' above the point x' is given by

$$K_{x'}(y,y') \quad := \quad M(x'y^{-1}, x'; x'y') \tag{7}$$

In the examples we will consider, the group G is Abelian (in which case we adopt an additive notation) and the speed variable only takes values in a symmetrical subset $S \subset G$ of generators of G, which will be small in comparison to G. Furthermore the situation will be assumed to be homogeneous: the kernels K_x do not depend on the support point $x \in G$. Then under additional hypotheses (of trajectorial reversibility and a kind of irreducibility, see next section), the probability measure ν is uniform and the transition kernel L corresponds to the Markov kernel of a reversible random walk: there exists a symmetric probability measure μ whose support is S such that

$$\forall\ x,x' \in G, \qquad L(x,x') \quad = \quad \mu(x'-x) \tag{8}$$

In this context, the second order Markov kernel $M^{(0)}$ corresponds to the Markov kernel $K^{(0)}$ for the speed variables given by

$$\forall\ y,y' \in S, \qquad K^{(0)}(y,y') \quad := \quad \mu(y')$$

To define a natural perturbation, assume furthermore that μ is the uniform distribution over S (so our framework is parametrized by the data (G,S) consisting of a finite Abelian group and a symmetric subset of generators). Consider the Markov kernel $K^{(1)}$ which forbids to reverse speed and transfers the corresponding weight to all the other directions:

$$\forall\, y,y' \in S, \qquad K^{(1)}(y,y') \;:=\; \begin{cases} \frac{1}{|S|-1} & \text{, if } y' \neq -y \\ 0 & \text{, if } y' = -y \end{cases}$$

Of course, this definition makes sense only if $|S| \geq 2$ (due to the symmetry of S, $|S| = 1$ corresponds to two trivial situations: $G = S = \mathbb{Z}_1$ or $G = \mathbb{Z}_2$ and $S = \{1\}$). Define as well the interpolating kernels:

$$\forall\, a \in [0,1], \qquad K^{(a)} \;:=\; (1-a)K^{(0)} + aK^{(1)} \tag{9}$$

We associate to them the second order Markov kernels $M^{(a)}$ for $a \in [0,1]$, which are all trajectorially reversible with respect to the probability measure

$$\forall\, x,x' \in G, \qquad m(x,x') \;:=\; \begin{cases} \frac{1}{|G||S|} & \text{, if } x'-x \in S \\ 0 & \text{, otherwise} \end{cases} \tag{10}$$

Then we have the following behavior for the spectral gap of $M^{(a)}$ as a function of $a \in [0,1]$:

Theorem 2 *Assume that G is a finite Abelian group and that S is a generating set of G with $|S| \geq 2$. Let μ be the uniform distribution on S and define*

$$a_0 \;:=\; (|S|-1)\frac{1-\sqrt{\lambda(2-\lambda)}}{1+\sqrt{\lambda(2-\lambda)}} \;\geq\; 0 \tag{11}$$

where $\lambda := \lambda(L)$ is the spectral gap of the random walk transition kernel defined in (8). Then under the framework described above, $\lambda(M^{(a)})$ is nondecreasing for $a \in [0, a_0 \wedge 1]$ and nonincreasing on $[a_0 \wedge 1, 1]$. So the best choice of $a \in [0,1]$ to get the largest possible spectral gap $\Lambda := \max_{a\in[0,1]} \lambda(M^{(a)})$ is $a = a_0 \wedge 1$ and we have, if $a_0 \leq 1$,

$$\Lambda \;=\; \sqrt{\frac{1-\sqrt{\lambda(2-\lambda)}}{1+\sqrt{\lambda(2-\lambda)}}}$$

while if $a_0 > 1$,

$$\Lambda \;=\; 1 - \frac{|S|}{2(|S|-1)}\left(1-\lambda+\sqrt{\left(\frac{|S|-2}{|S|}\right)^2 - \lambda(2-\lambda)}\right) \tag{12}$$

The example treated in Section 4 will show that when $|S| = 2$, the improvement of the spectral gap can be important and erases the diffusive effect which usually slows down the convergence to equilibrium of Monte-Carlo methods. Unfortunately, as soon as $|S| \geq 3$, the improvement of the spectral gap is quite weak, as we will see in Section 5. In fact our analysis will also furnish all the eigenvalues of the operators $M^{(a)}$, for $a \in [0,1]$, see Section 5 for details.

When $a_0 \geq 1$, Formula (12) is a particular case of a result due to Alon, Benjamini, Lubetzky and Sodin [1]: they consider a random walk on a connected, non bipartite and d-regular graph V, with $d \geq 3$, whose spectral gap is still denoted by λ (be careful of the changes of notations with the article [1], what they call λ, $\widetilde{\rho}$ and d correspond here respectively to $|S|(1-\lambda)$, $1-\Lambda$ and $|S|$), and they want to compare it with the second order random walk $(X_n)_{n\in\mathbb{N}}$ which does not backtrack: at any time $n \in \mathbb{N}\setminus\{0\}$ and for any positions $y, x_n, x_{n-1}, ..., x_0$,

$$\mathbb{P}[X_{n+1} = y | X_n = x_n, ..., X_0 = x_0] \;=\; \frac{1}{d-1}\mathbb{1}_{y\in N(x_n)\setminus\{x_{n-1}\}}$$

where $N(x_n)$ is the set of neighbors of x_n (for $n=0$, their convention is that $\mathbb{P}[X_1=y|X_0=x_0]=\frac{1}{d}\mathbb{1}_{y\in N(x_0)}$, but this is not very important). Then they show that whatever the initial condition, if V is a Ramanujan graph (which means that $\lambda\leq 1-2\sqrt{d-1}/d$, namely $a_0\geq 1$),

$$\limsup_{n\to\infty}\max_{u\in V}\left|\mathbb{P}[X_n=u]-\frac{1}{n}\right|^{1/n} \;=\; 1-\Lambda \tag{13}$$

where Λ is defined as in (12) (the value given by this formula is attained for $a=1$, so it does amount in our restricted framework to forbid backtracking, note also that our statement is slightly more precise, since we get the asymptotic rate of convergence for the couple (X_n,X_{n+1})). Their approach is completely different from ours and it is based on the recurrence relations satisfied by the numbers of nonbacktracking paths of a given length and going from a particular point to another particular point. These recurrence relations enable the authors to write explicitely the matrix of these numbers in terms of Chebyshev polynomials (applied to the adjacency matrix) and an asymptotic analysis of some linear combinations of these polynomials leads to (13). This result of Alon, Benjamini, Lubetzky and Sodin [1] is valid in a more general setting and its proof is more straightforward (admitting only a limited knowledge of Chebyshev polynomials) than ours, but they do not consider the case of weak forbidden backtracking (corresponding to $0<a<1$), nor are they interested in the behavior of the whole spectrum. In particular, they are not always exhibiting a faster chain. But more important for us is the following conjecture, which cannot be stated for d-regular graph, since the sentence "keep going in the same direction" has then no (at least immediate) meaning.

Indeed, we believe that the qualitative part of Theorem 2 concerning the variation of the spectral gap could be extended to more general situations. More importantly, we think that it is always possible to strongly improve the spectral gap, under the assumptions of Theorem 2, by considering a more appropriate perturbation.

Conjecture 3 Let G be a finite Abelian group and S be a generating set of G with $|S|\geq 2$. Let $K^{(1)}$ be the identity $S\times S$ matrix (seen as a speed transfer kernel, it forces to keep going always in the same direction) and consider again the interpolating kernels given in (9), as well as the associated second order Markov kernels $M^{(a)}$ for $a\in[0,1]$, which are also trajectorially reversible with respect to the probability measure m defined in (10). There exists $a_0\in[0,1]$ such that $\lambda(M^{(a)})$ is nondecreasing for $a\in[0,a_0]$ and nonincreasing on $[a_0,1]$. Furthermore there exists two values $0<c_1<c_2$ (maybe depending on $|S|$), such that

$$c_1\sqrt{\lambda(L)} \;\leq\; \lambda(M^{(a_0)}) \;\leq\; c_2\sqrt{\lambda(L)}$$

□

The main interest of second order Markov chains is maybe that they enable to introduce a notion of speed in a discrete setting. In addition to the fact that a global reversibility property can be imposed to them, this feature could be the reason why they provide so good models for molecular dynamics as it was put forward by the results of Bacallado and Pande [2]. In this context, a general motivation for us was to investigate the possible spectral implications of the trajectorial reversibility assumption. The results presented here should be seen as first attempts in this direction.

The plan of the paper is as follows: in next section we will study the structure of second order Markov kernels and the trajectorial reversibility property, in the general case and in the position-speed point of view for groups. In Section 3, we prove Theorem 1 and we present some links with the work of Neal [10]. In Section 4, we investigate the particular homogeneous example of the discrete circle $\mathbb{Z}_N$, with $N\in\mathbb{N}\setminus\{0,1\}$, when the set of allowed speeds is $S=\{-1,+1\}$. This will

lead us to revisit the analysis of a similar model done by Diaconis, Holmes and Neal [3], see also Gade and Overton [5]. In Section 5, we extend these results to prove Theorem 2 and we give some examples.

2 Trajectorial reversibility

We investigate here the trajectorially reversible second order Markov kernels, in order to be able to easily perturb them in the sequel. Their structure is not without recalling the framework of differentiable manifolds, feature which should not be so surprising, since second order Markov chains should be seen as discrete position-speed Markov processes (at any time, the knowledge of the position and of the speed of the chain is sufficient to determine the law of the future of the trajectory). In the homogeneous Abelian group case, we also exhibit a skew Kronecker product property of the second order Markov kernels which simplifies their spectral decomposition.

Let M be a second order Markov kernel on a finite state space V as in the introduction. We want to characterize the fact that M is trajectory reversible with respect to some probability measure m on V^2. As it was mentioned in the introduction, necessarily m is of the form (3), where L is reversible with respect to ν. Let $V' := \{x \in V \,:\, \nu(x) > 0\}$ be the support of ν. By reversibility, if $x \in V'$ and $x' \in V$ satisfies $L(x, x') > 0$, then $x' \in V'$. For $x' \in V'$, consider $V'_{x'} := \{x \in V' \,:\, L(x', x) > 0\}$, and $M_{x'}$ the Markov kernel defined by

$$\forall\, x, x'' \in V, \qquad M_{x'}(x, x'') \;\; := \;\; M(x, x'; x'')$$

Let us check that the restriction of $M_{x'}$ to $V'_{x'} \times V'_{x'}$ is still a Markov kernel (which will also be denoted by $M_{x'}$). Indeed, the trajectorial reversibility of M with m can be written has

$$\forall\, x, x' \in V', \forall\, x'' \in V, \qquad \nu(x')L(x', x)M_{x'}(x, x'') \;\; = \;\; \nu(x')L(x', x'')M_{x'}(x'', x) \tag{14}$$

so it appears that if $x \in V'_{x'}$ and $x'' \notin V'_{x'}$, then we have $M_{x'}(x, x'') = 0$. Thus only what happens on V' is relevant: in the sequel we assume that $V' = V$ and we remove the primes from the notations. By considering (14) for $x' \in V$ and $x, x'' \in V_{x'}$ and by dividing this identity by $\nu(x')$, it follows that the kernel $M_{x'}$ must be reversible with respect to the probability measure $L(x', \cdot)$ on $V_{x'}$. Let denote by $\mathcal{R}(L(x', \cdot))$ the convex set consisting of the kernels on $V_{x'}$ which are reversible with respect to $L(x', \cdot)$ and let $\mathcal{R}(L)$ be the "kernel-bundle" $\prod_{x \in V} \mathcal{R}(L(x', \cdot))$ over V. The above arguments show the direct implication in the following characterization and the reverse implication is proven in the same way through (14).

Lemma 4 *The second order Markovian kernel M is trajectorially reversible with respect to $m = \nu \times L$ if and only if $(M_x)_{x \in V} \in \mathcal{R}(L)$, namely if M is a section of the bundle $\mathcal{R}(L)$ and from now on this will be indicated by the notation $M \in \mathcal{R}(L)$.*

This result makes it clear that it is easy to perturbe a second order Markov kernel while keeping it trajectorially reversible with respect to a fixed measure $m = \nu \times L$, since this can be done independently in the different fibers of $\mathcal{R}(L)$. Note also that the particular second order Markov kernel $M^{(0)}$ considered in the introduction corresponds to the section $(L(x, \cdot))_{x \in V} \in \mathcal{R}(L)$.

Contrary to reversibility, the notions of invariance, irreducibility and aperiodicity relative to second order Markovian kernel M will be the usual ones, when M is interpreted as a transition kernel from V^2 to V^2, see (4). Nevertheless usually V^2 is not an appropriate state space and we restrict M to a subset of V^2, for instance in the previous situation the natural state space is $\{(x, x') \in V^2 \,:\, m(x, x') > 0\}$ (or a recurrent component of this subset). We now check an assertion made in the introduction:

Lemma 5 *If the second order Markovian kernel M is trajectorially reversible with respect to m, then m is invariant for M.*

Proof
By definition a probability measure m on V^2 is invariant by M if

$$\forall\ (x',x'')\in V^2,\qquad \sum_{x\in V} m(x,x')M(x,x';x'') \;=\; m(x',x'')$$

Equivalently, it means that for any $F\in\mathcal{F}(V^2)$,

$$\mathbb{E}[F(X_1,X_2)] \;=\; \mathbb{E}[F(X_0,X_1)] \tag{15}$$

where $(X_n)_{n\in\mathbb{N}}$ is a second order Markov chain with M as kernel and starting from the initial distribution m. But if m is reversible, we have

$$\begin{aligned}\mathbb{E}[F(X_1,X_2)] &= \mathbb{E}[F(X_1,X_0)]\\ &= \mathbb{E}[F(X_0,X_1)]\end{aligned}$$

where the second equality comes from the symmetry of m.

■

Remark 6 One can give a characterization similar to Lemma 4 of second order Markov kernels which leave a measure m invariant. First note that if m is left invariant by a second order Markov kernel M, then its marginals on the first and second factor of V^2 are equal. Indeed, just apply (15) to a function F only depending on the first variable. Let us call ν this common marginal distribution on V and as before assume that its support is V. Let L_1 and L_2 the Markov kernels defined by

$$\forall\ x,x'\in V,\qquad m(x,x') \;=\; \nu(x)L_2(x,x') \;=\; \nu(x')L_1(x',x)$$

(ν is invariant by L_1 and L_2).
As above, represent M as a family of Markov kernels $(M_x)_{x\in V}$. It follows from the invariance of m by M, that for any $x\in V$, M_x can be seen as a Markov kernel from V_x^- to V_x^+, where

$$\begin{aligned}V_x^- &:= \{x'\in V\,:\,m(x',x)>0\}\\ V_x^+ &:= \{x'\in V\,:\,m(x,x')>0\}\end{aligned}$$

Then it is not difficult to see that m is invariant for M if and only if for any $x\in V$, M_x transports the probability measure $L_1(x,\cdot)$ into the probability measure $L_2(x,\cdot)$. In particular Lemma 4 enables to see that a probability measure m is trajectorially reversible with respect to M if and only it is invariant and symmetric.

□

Assume again that the second order Markov kernel M is trajectorially reversible with respect to m and consider M^* the dual operator of M in $\mathbb{L}^2(m)$. Since m is invariant for M, it is well-known that M^* is Markovian in the abstract sense: it is nonnegativity preserving and the function $\mathbb{1}$ taking everywhere the value 1 is preserved. But as we compute it in next result, in general M^* has not the form of a second order Markov kernel:

Lemma 7 *Define the tilde operation on $\mathcal{F}(V^2)$ as the exchange of coordinates:*

$$\forall\ F\in\mathcal{F}(V^2),\ \forall\ x,x'\in V,\qquad \widetilde{F}(x,x') \;:=\; F(x',x)$$

Then we have

$$\forall\ F\in\mathcal{F}(V^2),\qquad M^*[F] \;=\; \widetilde{M[\widetilde{F}]}$$

or more explicitely,

$$\forall\ x,x'\in V,\qquad M^*[F](x,x') \;=\; \sum_{x''\in V} M(x',x;x'')F(x'',x)$$

Proof

Consider again $(X_n)_{n\in\mathbb{N}}$ a second order Markov chain with M as kernel and starting from the initial distribution m. For any $F, G \in \mathcal{F}(V^2)$, we have

$$\begin{aligned}
m[FM[G]] &= \mathbb{E}[F(X_0,X_1)G(X_1,X_2)] \\
&= \mathbb{E}[F(X_2,X_1)G(X_1,X_0)] \\
&= \mathbb{E}[\widetilde{G}(X_0,X_1)\widetilde{F}(X_1,X_2)] \\
&= \mathbb{E}[\widetilde{G}(X_0,X_1)M[\widetilde{F}](X_0,X_1)] \\
&= m[\widetilde{G}M[\widetilde{F}]] \\
&= m[\widetilde{\widetilde{G}M[\widetilde{F}]}] \\
&= m[\widetilde{GM[\widetilde{F}]}]
\end{aligned}$$

where the last but one equality comes from the symmetry of m.

■

We now come to the situation where the state space is a group G and we consider the family of Markov kernels $(K_x)_{x\in G}$ associated to M as in (7). In addition define the Markov kernels

$$\forall\ x, y, y' \in G, \qquad \widehat{K}_x(y,y') \ := \ K_x(y^{-1},y')$$

which will be called the modified speed kernels associated to M. If a symmetric probability measure $m = \nu \times L$ is given on V^2 with ν giving a positive mass to all the elements of G, we consider the family of probability distributions $(\widehat{\mu}_x)_{x\in G}$ given by

$$\forall\ x, y \in G, \qquad \widehat{\mu}_x(y) \ := \ L(x,xy)$$

Then Lemma 4 leads to the following result:

Lemma 8 *The second order Markovian kernel M is trajectorially reversible with respect to the symmetric distribution m if and only if for any $x \in G$, $\widehat{K}_x$ is reversible with respect to $\widehat{\mu}_x$. If furthermore the kernels $\widehat{K}_x$ do not depend on $x \in G$ and admit a unique recurrence class S which generates G, then ν is the uniform distribution and L is the Markov kernel of the random walk on G whose increments follow the symmetric law $\widehat{\mu}$ on S which is reversible with respect to $\widehat{K}$.*

Proof

The first part follows immediately from the fact that

$$\forall\ x, y, y' \in G, \qquad \widehat{K}_x(y,y') \ = \ M_x(xy,xy')$$

For the second part, assume that for all $x \in G$, $\widehat{K}_x = \widehat{K}$, where $\widehat{K}$ admits a unique recurrence class S which generates G and that $\widehat{K}$ is reversible with respect to a probability measure $\widehat{\mu}$. Then the support of $\widehat{\mu}$ is S and by uniqueness of the invariant measure on a recurrence class, we have

$$\forall\ x, x' \in G, \qquad L(x,x') \ = \ \widehat{\mu}_x(x^{-1}x') \ = \ \widehat{\mu}(x^{-1}x')$$

So it appears that L is the Markov kernel of the random walk on G whose increments follow the law $\widehat{\mu}$. Since S generates G, L is irreducible and admits a unique invariant probability measure which is necessarily the uniform distribution ν, because, as L, it must be invariant by translations. By reversibility of ν with respect to L, we get that $\widehat{\mu}$ is symmetric:

$$\forall\ y \in S, \qquad \widehat{\mu}(y^{-1}) \ = \ \widehat{\mu}(y)$$

and in particular S must be symmetric.

■

Note that in practice, it is convenient to rather start with a Markov kernel $\widehat{K}$ reversible with respect to a symmetric probability measure $\widehat{\mu}$. It is then immediate to check that the associated second order Markov kernel M defined by

$$\forall\ x, x', x'' \in G, \qquad M(x, x'; x'') \ := \ \widehat{K}((x')^{-1}x, (x')^{-1}x'')$$

is trajectorially reversible with respect to the measure m defined by

$$\forall\ x, x' \in G, \qquad m(x, x') \ = \ \frac{1}{|G|}\widehat{\mu}(x^{-1}x')$$

The second order Markov kernel M will then be said to be (spatially) homogeneous (this situation can be extended to state spaces which are not groups, under appropriate homogeneity assumptions, but we won't enter in the description of such a setting here).

Remark 9 In general $\widehat{K}$ may have several recurrent classes and the link with the recurrent classes of M is not so clear. Consider for instance the case where $G = \mathbb{Z}_N$, with $N \in \mathbb{N} \setminus \{0, 1\}$ and where the modified speed kernel $\widehat{K}$ is given by the matrix $\frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ on the set $S := \{-1, 1\} \subset \mathbb{Z}_N$. The speed kernel K is then the identity kernel, which means that once a speed has been chosen in $\{-1, 1\}$, it will never change. So M has two recurrent classes: $\{(x, x-1) \ : \ x \in \mathbb{Z}_N\}$ and $\{(x, x+1) \ : \ x \in \mathbb{Z}_N\}$. Conversely, if we take for $\widehat{K}$ the identity matrix, then all the sets of the form $\{(x, x+1), (x+1, x)\}$, where $x \in \mathbb{Z}_N$, are recurrent classes for M.
Indeed, even the link between the recurrent classes of K and those of $\widehat{K}$ can be quite involved. Nevertheless, if a set I is symmetric and invariant for $\widehat{K}$ (namely $x \in I$ and $y \notin I$ imply that $\widehat{K}(x, y) = 0$), then it is also invariant for K.

□

We now concentrate on the homogeneous case. We will always denote by S the support of the measure $\widehat{\mu}$ entering in the definition of this situation. The transition kernels $\widehat{K}$ and K will be understood as $S \times S$ matrices. Without loss of generality, we assume that G is generated by S (replace otherwise G by the subgroup generated by S). By considering the change of variables

$$G^2 \ni (x, x') \quad \mapsto \quad (x, y) := (x, x^{-1}x') \in G^2$$

we transform the transition kernel M of the Markov chain $(X_n, X_{n+1})_{n\in\mathbb{N}}$ into the position-speed variables transition kernel P given in (6). Note that the natural state space for this kernel is rather $G \times S$ instead of G^2. In the next result we exhibit the spectral decomposition of kernels whose structure generalizes that of P.

Proposition 10 *Let S and V be finite sets with respective cardinals s and v. Let K be a $S \times S$ matrix and for any $y \in S$, let Q_y be a $V \times V$ matrix. Assume there is a basis $(\varphi_l)_{l\in[\![v]\!]}$ of $\mathcal{F}(V, \mathbb{C})$ consisting of eigenfunctions for Q_y, independent of $y \in S$ (where $[\![v]\!] := \{1, 2, ..., v\}$). The corresponding eigenvalues are allowed to depend on $y \in S$ and are denoted by $(\sigma_l(y))_{l\in[\![v]\!]}$. For fixed $l \in [\![v]\!]$, consider K_l the $S \times S$ matrix defined by*

$$\forall\ y, y' \in S, \qquad K_l(y, y') \ := \ \sigma_l(y)K(y, y')$$

Assume that it is diagonalizable (in $\mathbb{C}$) and denote by $(\theta_{l,k})_{k\in[\![s]\!]}$ and $(\psi_{l,k})_{k\in[\![s]\!]}$ its eigenvalues and corresponding eigenvectors.
Next consider P the $(V \times S) \times (V \times S)$ matrix defined by

$$\forall\ (x, y), (x', y') \in V \times S, \qquad P((x, y), (x', y')) \ = \ K(y, y')Q_y(x, x')$$

(this is the skew Kronecker product of Q by K). Then P is diagonalizable, its eigenvalues are the $\theta_{l,k}$ for $(l, k) \in [\![v]\!] \times [\![s]\!]$ and a corresponding family of eigenvectors is $(\varphi_l \otimes \psi_{l,k})_{(l,k)\in[\![v]\!]\times[\![s]\!]}$.

Proof

This result is in fact more difficult to state than to prove. For $(l,k) \in [\![v]\!] \times [\![s]\!]$, consider $F_{l,k} := \varphi_l \otimes \psi_{l,k}$. Since the family $(\varphi_l)_{l\in[\![v]\!]}$ is linearly independent and the same is true for the family $(\psi_{l,k})_{k\in[\![s]\!]}$ for any fixed $l \in [\![v]\!]$, the family $(F_{l,k})_{(l,k)\in[\![v]\!]\times[\![s]\!]}$ is also linearly independent and is thus a basis of $\mathcal{F}(V \times S, \mathbb{C})$. Next for given $(l,k) \in [\![v]\!] \times [\![s]\!]$ and $(x,y) \in V \times S$, we compute that

$$\begin{aligned}
P[F_{l,k}](x,y) &= \sum_{x'\in V, y'\in S} K(y,y')Q_y(x,x')F_{l,k}(x',y') \\
&= \sum_{y'\in S} K(y,y')\psi_{l,k}(y') \sum_{x'\in V} Q_y(x,x')\varphi_l(x') \\
&= \sum_{y'\in S} K(y,y')\psi_{l,k}(y')\sigma_l(y)\varphi_l(x) \\
&= \varphi_l(x) \sum_{y'\in S} K_l(y,y')\psi_{l,k}(y') \\
&= \varphi_l(x)\theta_{l,k}\psi_{l,k}(y) \\
&= \theta_{l,k}F_{l,k}(x,y)
\end{aligned}$$

so $F_{l,k}$ is an eigenvector for the eigenvalue $\theta_{l,k}$ relatively to P.

■

Coming back to our homogeneous setting, we would like to apply the above proposition with $V = G$ and the translation kernels $(Q_y)_{y\in S}$ defined by

$$\forall\ x, x' \in G, \qquad Q_y(x,x') \ := \ \delta_{xy}(x')$$

If they were jointly diagonalizable, they would commute and since G is generated by S, we would end up with

$$\forall\ y, y' \in G, \qquad Q_yQ_{y'} \ = \ Q_{y'}Q_y$$

and these identities imply that G is Abelian. Conversely, if G is Abelian, then we can find $r \in \mathbb{N}\backslash\{0\}$ and $N_1, ..., N_r \in \mathbb{N} \setminus \{0,1\}$ such that G is isomorphe to the group product $\prod_{l\in[\![1,r]\!]} \mathbb{Z}_{N_l}$, so a generical element $x \in G$ will be written under the form $(x_l)_{l\in[\![1,r]\!]}$, with $x_l \in \mathbb{Z}_{N_l}$ for $l \in [\![r]\!]$. For $k = (k_l)_{l\in[\![1,r]\!]} \in \prod_{l\in[\![1,r]\!]}[\![0, N_l - 1]\!]$, consider the mapping $\rho_k \in \mathcal{F}(G,\mathbb{C})$ defined by

$$\forall\ x = (x_l)_{l\in[\![1,r]\!]} \in G, \qquad \rho_k(x) \ = \ \exp\left(2\pi i \sum_{l\in[\![1,r]\!]} k_l x_l/N_l\right)$$

Since the mappings ρ_k, for $k \in \prod_{l\in[\![1,r]\!]}[\![0, N_l - 1]\!]$, correspond to the irreducible representations of G, we get that for any $y \in S$,

$$Q_y[\rho_k] \ = \ \rho_k(y)\rho_k$$

So it appears that $(\rho_k)_{k\in\prod_{l\in[\![1,r]\!]}[\![0,N_l-1]\!]}$ is a basis of $\mathcal{F}(G,\mathbb{C})$ which enables to simultaneously diagonalize the Q_y, for $y \in S$, and we are in position to exploit Proposition 10. To do so conveniently, we introduce some notations:

$$G^* \ := \ \left\{\rho_k \ : \ k \in \prod_{l\in[\![1,r]\!]} [\![0, N_l - 1]\!]\right\} \tag{16}$$

and for $\rho \in G^*$, $\triangle(\rho)$ is the $S \times S$ diagonal matrix whose diagonal entries are given by $(\rho(y))_{y\in S}$. We also consider the $S \times S$ matrix $A_\rho := \triangle(\rho)K$. Then Proposition 10 implies the following result:

Proposition 11 *Let G be a finite Abelian group and M a homogeneous second order Markov kernel on G with speed transfer matrix K. If all the matrices A_ρ, with $\rho \in G^*$, are diagonalizable, then M is diagonalizable and its spectrum is the union of the spectra of the A_ρ, with $\rho \in G^*$ (with multiplicities).*

One can go further: if one of the matrices A_ρ is not diagonalizable, one can consider its decomposition into Jordan blocks and an immediate extension of Proposition 10 enables to transfer these Jordan blocks into Jordan blocks for M. In particular the spectrum of M is always the union of the spectra of the A_ρ, with $\rho \in G^*$ (with obvious relations between the dimensions of the corresponding Jordan blocks). These observations do not require M to be trajectorially reversible, but we believe that the latter property could be helpful to understand the spectral structure of the matrices A_ρ, with $\rho \in G^*$. Nevertheless this reversibility condition won't be sufficient to insure that they are diagonalizable, even under the stronger assumption made in Lemma 12 below (see Section 4).

We end this section with a discussion of the structural properties which can be deduced from a natural reinforcement of the trajectorial reversibility. Indeed, let M be a trajectorially reversible homogeneous second order Markov kernel, with modified speed kernel $\widehat{K}$ reversible with respect to $\widehat{\mu}$. By definition of K, we have for any $y, y' \in S$, $\widehat{\mu}(y)K(-y, y') = \widehat{\mu}(y')K(-y', y)$, so taking into account the symmetry of $\widehat{\mu}$, we get $\widehat{\mu}(-y)K(-y, y') = \widehat{\mu}(-y')K(-y', y)$. Summing over $y' \in S$, it follows that $\widehat{\mu}(-y) = \widehat{\mu}K(-y)$, namely $\widehat{\mu}$ is invariant for K. So the assumption of the next result is not so farfetched, note that it is in particular satisfied by the families of examples considered in Theorem 2 and Conjecture 3.

Lemma 12 *Assume that $\widehat{\mu}$ is also reversible with respect to K, then the matrices A_ρ, with $\rho \in G^*$, are conjugate to symmetric and centro-Hermitian matrices.*

Recall that a $S \times S$ matrix B (where S is a symmetrical set) is said to be centro-Hermitian if $\widetilde{B} = B$, where

$$\forall\ y, y' \in S, \qquad \widetilde{B}(y, y') \ \coloneqq\ \overline{B(-y, -y')}$$

see for instance Lee [8] or Pressman [12]. Weaver [14] gave a nice review of the corresponding notion of a (real) centro-symmetric matrix, property which is satisfied by the speed transition kernel under the above assumption.

Proof

Under the hypothesis of this lemma, we simplify the notations and write μ instead of $\widehat{\mu}$. For $\rho \in G^*$, let $\rho^{1/2} \coloneqq (\rho^{1/2}(y))_{y \in S}$ be a vector of square roots of $(\rho(y))_{y \in S}$, chosen in such a way that $\rho^{1/2}(-y) = \overline{\rho^{1/2}(y)}$ (recall that we have $\rho(-y) = \overline{\rho(y)}$ and that this complex number has modulus 1). Define next

$$\begin{aligned} B_\rho \ &\coloneqq\ \triangle(\sqrt{\mu}\rho^{-1/2})A_\rho\triangle(\rho^{1/2}/\sqrt{\mu}) \\ &=\ \triangle(\sqrt{\mu}\rho^{1/2})K\triangle(\rho^{1/2}/\sqrt{\mu}) \end{aligned} \tag{17}$$

(where for any vector v, $\triangle(v)$ is the diagonal matrix with corresponding entries). Since A_ρ is conjugate to B_ρ, both matrix have the same spectral structure, but B_ρ has two practical advantages: is is symmetric and centro-Hermitian. Indeed, we have

$$\forall\ y, y' \in S, \qquad B_\rho(y, y') \ =\ \frac{\sqrt{\mu(y)}}{\sqrt{\mu(y')}}K(y, y')\rho^{1/2}(y)\rho^{1/2}(y')$$

so the symmetry of B_ρ is a consequence of the reversibility of K with respect to μ. The centro-Hermitian property comes from the fact that

$$\forall\ y \in S, \qquad \mu(-y) = \mu(y) \quad \text{and } \rho^{1/2}(-y) = \overline{\rho^{1/2}(y)}$$

■

A consequence of the above result is that the spectrum of A_ρ, for $\rho \in G^*$, is stable by conjugacy, since this is true for any centro-Hermitian matrix B (see for instance Lee [8]): to any $f = (f(y))_{y\in S} \in \mathcal{F}(S,\mathbb{C})$ (seen as a column vector), associate the vector $\widetilde{f} := (\overline{f(-y)})_{y\in S} \in \mathcal{F}(S,\mathbb{C})$. An immediate computation shows that

$$\forall\ f \in \mathcal{F}(S,\mathbb{C}), \qquad \widetilde{Bf} \ = \ \widetilde{B}\widetilde{f}$$

Thus if B is centro-Hermitian and if f is an eigenvector of B associated to eigenvalue $\theta \in \mathbb{C}$, then $\widetilde{f}$ is an eigenvector of B associated to eigenvalue $\bar{\theta}$.

Note that we knew a priori that the spectrum of M is stable by conjugacy, because its entries are real. But this property corresponds to the fact that the spectrum of A_ρ is conjugate to the spectrum of $A_{\bar{\rho}} = \overline{A_\rho}$, for any $\rho \in G^*$. This leads to the following observation: if θ is a non-real eigenvalue of A_ρ, with $\rho \neq \bar{\rho} \in G^*$, then its multiplicity as an eigenvalue of M is at least the double of its multiplicity as an eigenvalue of A_ρ.

Remark 13 In view of the above considerations, it is not clear how to deduce spectral informations from the trajectorial reversibility assumption. Nevertheless in the homogeneous case, one can use it to deduce information on the convergence to equilibrium in the following way. For simplicity consider the Abelian case. Let $(X_n, Y_n)_{n\in\mathbb{N}}$ be the position-speed Markov chain associated to the second order Markov kernel M, whose modified speed kernel is denoted $\widehat{K}$ as above. Consider the Markov chain $(\widehat{Y}_n)_{n\in\mathbb{N}}$ defined by

$$\forall\ n \in \mathbb{N}, \qquad \widehat{Y}_n \ := \ (-1)^n Y_n$$

This chain has $\widehat{K}$ as transition kernel and thus is reversible with respect to $\widehat{\mu}$, so to understand its convergence to equibrium, we can resort to the usual spectral techniques. The chain $(X_n, Y_n)_{n\in\mathbb{N}}$ can be deduced from $(\widehat{Y}_n)_{n\in\mathbb{N}}$ via a deterministic procedure: for any $n \in \mathbb{N}$, we have

$$\begin{aligned} X_n &= \widehat{Y}_0 - \widehat{Y}_1 + \widehat{Y}_2 - \cdots + (-1)^{n-1}\widehat{Y}_{n-1} \\ Y_n &= (-1)^n \widehat{Y}_n \end{aligned}$$

So $(X_n)_{n\in\mathbb{N}}$ appears as a time-inhomogeneous, but periodic, additive functional of $(\widehat{Y}_n)_{n\in\mathbb{N}}$. To come back to the situation of a time-homogeneous functional, it is better to consider the Markov chain $(\widehat{Y}_{2n}, \widehat{Y}_{2n+1})_{n\in\mathbb{N}}$, whose convergence to equilibrium can easily be deduced from that of $(\widehat{Y}_n)_{n\in\mathbb{N}}$. Then one can resort to the well-known large time behavior of additive functionals (see for instance Maxwell and Woodroofe [9] and the references given therein), at least on the covering lattice of G, based on the central limit theorem for martingales. Next it remains to "project" this result on G: once the variance enables to cover a fundamental domain for G, the stochastic chain $(X_n)_{n\in\mathbb{N}}$ will be close to equilibrium, namely uniformly distributed.
But we won't developp this approach here, because it is not in the spirit of the paper.

□

3 Infinitesimal improvement of spectrum

Our purpose here is to show that the spectral gap (and even the whole spectrum outside 0) of a reversible Markov chain can always been improved (except in the situation of an i.i.d. sequence) by resorting to related second order Markov chains, which "have a little less tendency to hesitate in their way to equilibrium" than the original chain. The rigorous formulation of this heuristic is enclosed in Theorem 1, and this section is mainly devoted to his proof. But we will also investigate

the best way to improve only the spectral gap, when the underlying first order kernel L is the transition matrix of a random walk on a regular graph and under an additional assumption of multiplicity one for the spectral gap.

Our starting point is a Markov kernel L which admits a reversible probability measure ν, positive on the finite set V. Since we will only be interested in second order Markov chains which are trajectorially reversible with respect to the probability measure $\nu \times L$ (defined in (3)), up to a restriction of the state space V, there is no loss of generality in assuming that L is furthermore irreducible. We associate to L the second order Markov kernel $M^{(0)}$ defined in (5). We begin by verifying an assertion made in the introduction:

Proposition 14 *The spectrum of the transition operator $M^{(0)}$ is equal to the spectrum of L with the same multiplicities, except for the eigenvalue 0, whose multiplicity is $N^2 - N$, where $N = |V|$. In particular $M^{(0)}$ is diagonalizable if and only if 0 is not an eigenvalue of L. When 0 is an eigenvalue of L, let d be the dimension of the corresponding eigenspace. Then the canonical form of $M^{(0)}$ contains d 2×2-Jordan blocks $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.*

It follows that $(M^{(0)})^2$ is always diagonalizable. This is not true for $M^{(0)}$, as we will see on an example in next section.

Proof

Since ν has positive weight everywhere and is reversible with respect to L, the latter operator is diagonalizable in $\mathbb{R}$. Let $(\theta_i)_{i\in[\![N]\!]}$ be its spectrum and $(f_i)_{i\in[\![N]\!]}$ be a corresponding family of eigenvectors. For $i \in [\![N]\!]$, define $F_i \in \mathcal{F}(V^2)$ by

$$\forall\ (x,x') \in V^2, \qquad F_i(x,x') \ \ := \ \ f_i(x')$$

We compute that for any $(x,x') \in V^2$,

$$\begin{aligned} M^{(0)}[F_i](x,x') &= \sum_{x''\in V} M(x,x';x'')F_i(x',x'') \\ &= \sum_{x''\in V} L(x',x'')f_i(x'') \\ &= L[f_i](x') \\ &= \theta_i f_i(x') \\ &= \theta_i F_i(x,x') \end{aligned}$$

so that F_i is an eigenfunction associated to the eigenvalue θ_i with respect to $M^{(0)}$. We denote by $\mathcal{F}$ the vector space generated by the functions F_i, with $i \in [\![N]\!]$. It is in fact the subspace of $\mathcal{F}(V^2)$ consisting of functions only depending on the second variable. Let us also denote by $\mathcal{F}'$ the subspace of $\mathcal{F}$ generated by the functions F_i, with $i \in [\![N]\!]$ satisfying $\theta_i = 0$.
Next for $x_0 \in V$, consider $\mathcal{F}_{x_0}$ the space of functions F from $\mathcal{F}(V^2)$ which can be written under the following form

$$\forall\ (x,x') \in V^2, \qquad F(x,x') \ \ = \ \ \delta_{x_0}(x)f(x')$$

where $f \in \mathcal{F}(V)$ satisfies $L[f](x_0) = 0$. For such a function $F \in \mathcal{F}_{x_0}$, we compute that for any $(x,x') \in V^2$,

$$\begin{aligned} M^{(0)}[F](x,x') &= \sum_{x''\in V} L(x',x'')F(x',x'') \\ &= \delta_{x_0}(x') \sum_{x''\in V} L(x',x'')f(x'') \\ &= \delta_{x_0}(x')L[f](x_0) \\ &= 0 \end{aligned}$$

so that F is an eigenfunction associated to the eigenvalue 0 with respect to $M^{(0)}$. Next consider $\mathcal{F}''$ the subspace of $\mathcal{F}(V^2)$ generated by all the $\mathcal{F}_{x_0}$, with $x_0 \in V$. By the above computation, $\mathcal{F}''$ is included in the eigenspace associated to the eigenvalue 0 of $M^{(0)}$. Clearly the subspaces $\mathcal{F}_{x_0}$, for $x_0 \in V$, are mutually linearly independent, since the corresponding functions vanish on disjoint sets. Furthermore the mapping $\mathcal{F}(V) \ni f \mapsto L[f](x_0)$ is a non degenerate linear form, so the dimension of $\mathcal{F}_{x_0}$ is $N-1$. Thus we get that the dimension of $\mathcal{F}''$ is $N^2 - N$. Note that the dimension of $\mathcal{F}$ is N, so $M^{(0)}$ would be diagonalizable if we had $\mathcal{F} \cap \mathcal{F}'' = \{0\}$. But in general, we only have $\mathcal{F} \cap \mathcal{F}'' = \mathcal{F}'$. Indeed, let $F \in \mathcal{F} \cap \mathcal{F}''$, we can find some coefficients $(a_{x_0})_{x_0 \in V}$ and some functions $f_{x_0} \in \mathcal{F}_{x_0}$, for $x_0 \in V$, such that

$$F = \sum_{x_0 \in V} a_{x_0} \delta_{x_0} \otimes f_{x_0}$$

Further there exists $f \in \mathcal{F}(V)$ such that

$$\forall\ (x, x') \in V^2, \qquad \begin{aligned} F(x, x') &= f(x') \\ &= a_x f_x(x') \end{aligned}$$

It follows that for any $x \in V$,

$$L[f](x) = a_x L[f_x](x) = 0$$

so that $F \in \mathcal{F}'$. Conversely, if $F \in \mathcal{F}'$, just choose $a_{x_0} = 1$ and $f_{x_0} = f$ for all $x_0 \in V$ in the above description to see that $F \in \mathcal{F}''$ and by consequence $F \in \mathcal{F} \cap \mathcal{F}''$.
Thus in general, we have only found $N^2 - d$ independent eigenvectors for $M^{(0)}$. For $i \in [\![N]\!]$, with $\theta_i = 0$, consider the function $G_i \in \mathcal{F}(V^2)$ defined by

$$\forall\ (x, x') \in V^2, \qquad G_i(x, x') := f_i(x)$$

We compute that for any $(x, x') \in V^2$,

$$\begin{aligned} M^{(0)}[G_i](x, x') &= \sum_{x'' \in V} M(x, x'; x'') G_i(x', x'') \\ &= \sum_{x'' \in V} L(x', x'') f_i(x') \\ &= f_i(x') \\ &= F_i(x, x') \end{aligned}$$

Thus the 2-dimensional vector space generated by F_i and G_i is stable by $M^{(0)}$ and the matrix associated to the operator $M^{(0)}$ in the basis (G_i, F_i) is the 2×2-Jordan block $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. The announced results follow at once.

■

In the above proof the reversibility of L was not really necessary, it was sufficient to assume that it is diagonalizable. One can even go further by revisiting the above proof:

Remark 15 Let L be any Markov kernel on S and construct the kernel $M^{(0)}$ as in (5). Then as above the spectrum of $M^{(0)}$ is equal to the spectrum of L plus the value 0. More precisely, the Jordan blocks associated to an eigenvalue $\theta \neq 0$ of L are in exact correspondance with those of $M^{(0)}$. Concerning the eigenvalue 0, let d be the number of 1×1-Jordan blocks for L. All of them are transformed into 2×2 blocks for $M^{(0)}$. But for $r \geq 2$, all the $r \times r$-Jordan blocks of L are transformed into $r \times r$-Jordan blocks for $M^{(0)}$. And there is furthermore the creation of $N^2 - N - d$

1×1-Jordan blocks for $M^{(0)}$. This comes from the fact that the $(r+1) \times (r+1)$ matrix A defined by

$$\forall\ i,j \in [\![r+1]\!], \qquad A(i,j) \ := \ \begin{cases} 1 & \text{, if } j = i+1 \text{ and } i \in [\![r-1]\!] \\ 1 & \text{, if } i = 0 \text{ and } j = r+1 \\ 0 & \text{, otherwise} \end{cases}$$

admits a decomposition into a 1×1 block and a $r \times r$ block (both associated to the eigenvalue 0). $\square$

In Proposition 14 we have considered the spectrum of $M^{(0)}$ as an operator on $\mathcal{F}(V^2)$, but in general we are only interested on its action on a subspace of $\mathcal{F}(V^2)$: since the probability measure $\nu \times L$ is invariant for $M^{(0)}$, we prefer to see $M^{(0)}$ as an operator on $\mathbb{L}^2(\nu \times L)$, whose dimension is the cardinal of the set $\{(x,y) \in V^2 \ : \ L(x,y) > 0\}$. By our assumption $\nu > 0$, we have $\mathcal{F}(V) = \mathbb{L}^2(\nu)$ and there is no ambiguity for the spectrum of L. The proof of Proposition 14 can then be extended to this situation: the spectrum of $M^{(0)}$ is the spectrum of L plus maybe the eigenvalue 0. Indeed the eigenvalue 0 is necessary in the spectrum of $M^{(0)}$ if $\dim(\mathbb{L}^2(\nu \times L)) > N$. Note that under our reversibility and irreducibility assumptions on L, the only situations when this inegality is not satisfied is when $|V| = 1$ or $|V| = 2$ and L is given by the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Similar observations also hold in the more general setting of the above remark (but assume furthermore that L is irreducible, so that it admits a unique positive invariant measure).

In order to go in the direction of Theorem 1, consider a second order Markov kernel $M^{(1)}$ admitting $\nu \times L$ as trajectorial reversible measure and satisfying for any $x, x', x'' \in V$,

$$M^{(1)}(x,x';x'') \ > \ L(x',x'') \qquad \text{, if } x'' \neq x,\ L(x,x') > 0 \text{ and } L(x',x'') > 0 \tag{18}$$

Such transition kernels exists, since according to Lemma 4, the only requirement is that for any fixed $x' \in V$, the kernel $(M^{(1)}_{x'}(x,x''))_{x,x'' \in V}$ is reversible with respect to the probability $L(x',\cdot)$. Next, as in the introduction, for any $a \in [0,1]$, we consider the interpolation kernel between $M^{(0)}$ and $M^{(1)}$:

$$M^{(a)} \ := \ (1-a)M^{(0)} + aM^{(1)}$$

which is a second order Markov kernel still admitting $\nu \times L$ as trajectorial reversible measure. We need some information about the behavior of the spectral decomposition of $M^{(a)}$ as a function of a, at least for small a. From Chapter 2 §1.8 of Kato [7], we easily deduce the following result:

Lemma 16 *There exists $a_0 \in (0,1)$ such that the number s of distinct eigenvalues of $M^{(a)}$ does not depend on $a \in (0,a_0)$. It is possible to parameterize these eigenvalues by $(\theta_l(a))_{l \in [\![s]\!]}$ in such a way that for any $l \in [\![s]\!]$, the mapping $(0,a_0) \ni a \mapsto \theta_l(a)$ is analytical. Furthermore these mappings admit continuous extension to $[0,a_0)$ and the set $\{\theta_l(0) \ : \ l \in [\![s]\!]\}$ coincides with the spectrum of $M^{(0)}$ (note nevertheless that the cardinal of this spectrum can be stricly less than s, when occurs a splitting phenomenon of some eigenvalues at 0). Moreover for any $l \in [\![s]\!]$, we can find a continuous mapping $[0,a_0) \ni a \mapsto F_l(a) \in \mathcal{F}(V^2,\mathbb{C}) \setminus \{0\}$ such that for any $a \in [0,a_0)$, F_l is an eigenfunction associated to the eigenvalue $\theta_l(a)$ of $M^{(a)}$, and such that the mapping $(0,a_0) \ni a \mapsto F_l(a) \in \mathcal{F}(V^2,\mathbb{C}) \setminus \{0\}$ is analytical.*

In fact the above lemma holds for general second order Markov kernel $M^{(1)}$. But the assumption (18) is need for the next result.

Proposition 17 *With the notations introduced in the previous lemma, assume that $l \in [\![s]\!]$ is such that $\theta_l(0) \neq 0$. Then the limit $\lim_{a \to 0_+} \theta_l'(a)$ exists and belongs to $\mathbb{R}$. By consequence $\theta_l'(0)$ exists and coincides with the previous limit. Furthermore if $|\theta_l(0)| \neq 1$, then the sign of $\theta_l'(0)$ is the opposite sign of $\theta_l(0)$ (recall that $\theta_l(0) \in \mathbb{R} \setminus \{0\}$). If $|\theta_l(0)| = 1$, we have $\theta_l'(0) = 0$.*

Proof

By definition, we have for any $a \in [0, a_0)$,

$$M^{(a)} F_l(a) \quad = \quad \theta_l(a) F_l(a)$$

and we can differentiate this equality for $a \in (0, a_0)$ to get

$$(M^{(1)} - M^{(0)}) F_l(a) + M^{(a)} F_l'(a) \quad = \quad \theta_l'(a) F_l(a) + \theta_l(a) F_l'(a) \tag{19}$$

where the primes represent differentiation with respect to $a \in (0, a_0)$.

Our goal is to get rid of the term $F_l'(a)$ that we don't control as a is going to 0_+. Lemma 7 implies that the fonction $\widetilde{F}_l(a)$ defined over there is an eigenvector of $(M^{(a)})^*$ associated to the eigenvalue $\theta(a)$. Indeed, we have

$$\begin{aligned} (M^{(a)})^*[\widetilde{F}_l(a)] &= \widetilde{M^{(a)}[F_l}(a)] \\ &= \widetilde{\theta(a) F_l}(a) \\ &= \theta(a) \widetilde{F}_l(a) \end{aligned}$$

Multiply (19) by $\widetilde{F}_l(a)$ and integrate the obtained equation with respect to m, to get

$$m[\widetilde{F}_l(a)(M^{(1)} - M^{(0)}) F_l(a)] + m[\widetilde{F}_l(a) M^{(a)} F_l'(a)] \quad = \quad \theta_l'(a) m[\widetilde{F}_l(a) F_l(a)] + \theta_l(a) m[\widetilde{F}_l(a) F_l'(a)]$$

By definition of the dual operator $(M^{(a)})^*$ we have

$$\begin{aligned} m[\widetilde{F}_l(a) M^{(a)} F_l'(a)] &= m[(M^{(a)})^*[\widetilde{F}_l(a)] F_l'(a)] \\ &= \theta(a) m[\widetilde{F}_l(a) F_l'(a)] \end{aligned}$$

Thus we deduce the following equation where the term $F_l'(a)$ has disappeared:

$$m[\widetilde{F}_l(a)(M^{(1)} - M^{(0)}) F_l(a)] \quad = \quad \theta_l'(a) m[\widetilde{F}_l(a) F_l(a)]$$

We can now use the continuity properties put forward in Lemma 16 to see that

$$\begin{aligned} \lim_{a \to 0_+} m[\widetilde{F}_l(a)(M^{(1)} - M^{(0)}) F_l(a)] &= m[\widetilde{F}_l(0)(M^{(1)} - M^{(0)}) F_l(0)] \\ \lim_{a \to 0_+} m[\widetilde{F}_l(a) F_l(a)] &= m[\widetilde{F}_l(0) F_l(0)] \end{aligned}$$

so the limit $\lim_{a \to 0_+} \theta_l'(a)$ exists if we have $m[\widetilde{F}_l(0) F_l(0)] \neq 0$. To compute this term, we note that since $F_l(0)$ is an eigenvector of $M^{(0)}$ associated to the eigenvalue $\theta_l(0)$, according to the proof of Proposition 14, there exists an eigenfunction $f \in \mathcal{F}(V)$ of $M^{(0)}$ associated to the eigenvalue $\theta_l(0)$ such that

$$\forall\ x, x' \in V, \qquad F_l(0)(x, x') \quad = \quad f(x')$$

It follows that

$$\begin{aligned} m[\widetilde{F}_l(0) F_l(0)] &= \sum_{x, x' \in V} \nu(x) L(x, x') f(x) f(x') \\ &= \nu[f L[f]] \\ &= \theta_l(0) \nu[f^2] \end{aligned}$$

which is different from 0. Thus $\lim_{a \to 0_+} \theta_l'(a)$ exists and by consequence the mapping $[0, a_0) \ni a \mapsto \theta_l(a)$ is differentiable at 0 and $\theta_l'(0) = \lim_{a \to 0_+} \theta_l'(a)$. To evaluate this derivative, it remains to compute

$$m[\widetilde{F}_l(0)(M^{(1)} - M^{(0)}) F_l(0)] \quad = \quad \sum_{x, x' \in V} \nu(x) L(x, x') f(x) \sum_{x'' \in V} D_{x'}(x, x'') f(x'') \tag{20}$$

where for any $x, x', x'' \in V$, we have defined

$$\begin{aligned} D_{x'}(x,x'') &:= M^{(1)}_{x'}(x,x'') - M^{(0)}_{x'}(x,x'') \\ &= M^{(1)}(x,x';x'') - L(x',x'') \end{aligned}$$

These quantities are real, so we already get that $\theta_l'(0) \in \mathbb{R}$. To go further, note that for any $x' \in V$, $D_{x'}$ is a generator matrix: outside the diagonal its entries are nonnegative (even positive in the present case) and the sum of the entries of any row is zero. Furthermore this matrix is reversible (in the usual sense) with respect to the (row) vector $(L(x',x))_{x\in V}$. It is then easy to see that (cf. for instance the lecture notes of Saloff-Coste [13])

$$\begin{aligned} \sum_{x,x''\in V} L(x',x)D_{x'}(x,x'')f(x)f(x'') &= -\frac{1}{2}\sum_{x\neq x''\in V} L(x',x)D_{x'}(x,x'')(f(x'')-f(x))^2 \\ &\leq 0 \end{aligned}$$

Since by reversibility of L the r.h.s. of (20) can also be written as

$$\sum_{x'\in V}\nu(x')\sum_{x,x''\in V} L(x',x)D_{x'}(x,x'')f(x)f(x'')$$

it appears that $m[\widetilde{F}_l(0)(M^{(1)}-M^{(0)})F_l(0)] \leq 0$ and by consequence $\theta_l(0)\theta_l'(0) \leq 0$.
Let us now consider the situation where $\theta_l(0)\theta_l'(0) = 0$. According to the previous computations, this is equivalent to the fact that

$$\forall\ x,x',x'' \in V, \qquad D_{x'}(x,x'')(f(x'')-f(x)) = 0$$

namely, for any $x, x', x'' \in V$ such that $D_{x'}(x,x'') \neq 0$, we must have $f(x'') = f(x)$. Due to our assumption (18), this amounts to the condition that for any $x \neq x'' \in V$ such that there exists $x' \in V$ with $L(x,x') > 0$ and $L(x',x'') > 0$ (inequalities which are equivalent to $M^{(1)}(x,x';x'') > 0$), we must have $f(x'') = f(x)$. This is equivalent to the fact that f must be constant on the recurrence classes of L^2, namely $L^2[f] = f$. Thus, since $L[f] = \theta_l(0)f$, this means that $\theta_l^2(0) = 1$, hence the announced results.

■

We can now come to the

Proof of Theorem 1

Since L is assumed to be irreducible, the eigenspace associated to the eigenvalue 1 is just $\mathrm{Vect}(\mathbb{1})$, so that $1 \notin \Theta_*(L)$. Furthermore $-1 \notin \Theta_*(L)$, otherwise L would be periodic, thus we have $\lambda(L) > 0$. On the other hand, as it was mentioned in the introduction, the condition that $L \neq \nu$ insures that $\lambda(L) \neq 1$, so in fact $\lambda(M^{(0)}) \in (0,1)$. With the notations of Lemma 16, define

$$I := \{l \in [\![s]\!] : |\theta_l(0)| = 1 - \lambda(L)\}$$

By continuity of the eigenvalues of $M^{(a)}$ as functions of a, we can find $a_0' \in (0, a_0)$ such that

$$\forall\ a \in [0,a_0'], \qquad \lambda(M^{(a)}) = \max\{1 - |\theta_l(a)| : l \in I\}$$

By considering a first order expansion at 0_+ of the functions θ_l, for $l \in I$, Proposition 17 easily implies that $[0,1] \ni a \mapsto \lambda(M^{(a)})$ is differentiable at 0 and that

$$\begin{aligned} \frac{d\lambda(M^{(a)})}{da}(0) &= \min\{\left|\theta_l'(a)\right| : l \in I\} \\ &> 0 \end{aligned}$$

■

Next consider the situation where instead of infinitesimally improving the nonnull part of the spectrum of $M^{(0)}$, we just want to improve its spectral gap. What is the best way to do it, at least infinitesimally? More precisely, let $\mathcal{D}$ be the set of signed kernels $D = (D(x, x'; x'')_{x,x',x''\in V}$ from V^2 to V, such that for $\epsilon > 0$ small enough, $M^{(0)} + \epsilon D$ is a second order Markov kernel which is trajectorially reversible with respect to $\nu \times M$ and such that the following normalization condition is satisfied:

$$\forall\ x, x' \in V, \qquad \sum_{x''\in V} \left|D(x, x'; x'')\right| \quad \leq \quad 2$$

As above, we can show that $\lambda(M^{(0)} + \epsilon D)$ is differentiable at the right of 0 and we are wondering what are the perturbations $D^\dagger \in \mathcal{D}$ which maximize the improvement of the spectral gap, namely such that

$$\frac{d}{d\epsilon}\lambda(M^{(0)} + \epsilon D^\dagger)\Big|_{\epsilon=0} \quad = \quad \sup_{D\in\mathcal{D}} \frac{d}{d\epsilon}\lambda(M^{(0)} + \epsilon D)\Big|_{\epsilon=0} \tag{21}$$

To simplify the analysis of this problem, assume that there is only one eigenvalue θ of L such that $\lambda(L) = 1 - |\theta|$ (θ can be positive or negative) and that the multiplicity of θ is 1. By a slight misuse of language, in this situation we will say that the spectral gap is of multiplicity 1. Let $f \in \mathcal{F}(V)$ be an associated eigenfunction (different from 0).
To any signed kernel D from V^2 to V, associate the family $(D_{x'})_{x'\in V}$ of signed kernels from V to V defined by

$$\forall\ x', x, x'' \in V, \qquad D_{x'}(x, x'') \quad \coloneqq \quad D(x, x'; x'')$$

For fixed $x' \in V$, let $\mathcal{D}_{x'}$ be the set of signed kernels $(D(x, x''))_{x,x''\in V}$ from V to V which are reversible with respect to $L(x', \cdot)$ (i.e. $\forall\ x, x'' \in V$, $L(x', x)D(x, x'') = L(x', x'')D(x'', x)$) and which satisfy $\sum_{x''\in V} |D(x, x'; x'')| \leq 2$ and $\sum_{x''\in V} D(x, x'') = 0$ for all $x \in V$. It is easy to check (see the proof of Lemma 4) that $D \in \mathcal{D}$ if and only if $D_{x'} \in \mathcal{D}_{x'}$ for all $x' \in V$. Define for any $x' \in V$, the mapping $H_{x'} : \mathcal{D}_{x'} \to \mathbb{R}$ via the formula

$$\forall\ D = (D(x, x''))_{x,x''\in V} \in \mathcal{D}_{x'}, \qquad H_{x'}(D) \quad = \quad \sum_{x,x''\in V} L(x', x)D(x, x'')f(x)f(x'')$$

Then the arguments given in the proofs of Proposition 17 and Theorem 1 also show the

Lemma 18 *A signed kernel $D^\dagger \in \mathcal{D}$ satisfies (21) if and only if for any $x' \in V$, $D^\dagger_{x'}$ minimizes the mapping $H_{x'}$ on $\mathcal{D}_{x'}$. As a consequence, by compactness of $\mathcal{D}_{x'}$ and continuity of $H_{x'}$, such an optimal $D^\dagger \in \mathcal{D}$ exists.*

In order to get an idea of an optimal $D^\dagger \in \mathcal{D}$, let us make the following additional assumption: for any $x' \in V$, the probability measure $L(x', \cdot)$ is uniform over $V_{x'}$ (recall that this set was defined as the support of $L(x', \cdot)$ at the beginning of section 2) and that $n \coloneqq |V_{x'}|$ does not depend on $x' \in V$: L is the Markov transition kernel of the simple random walk on the connected and regular graph on V whose edge set is $\{(x, x') \in V^2 : L(x, x') > 0\}$. The reversibility condition on the elements of $\mathcal{D}_{x'}$ amounts to asking them to be symmetric. Denote by $x_1, x_2, ..., x_n$ the elements of $V_{x'}$ (as we work with fixed $x' \in V$, we won't indicate that everything is depending on this point), ordered such that $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_n)$. To simplify notations, any element $D \in \mathcal{D}_{x'}$ is identified with a $n \times n$ matrix $(D(i, j))_{i,j\in[\![n]\!]}$ where $D(i, j)$ stands for $D(x_i, x_j)$. Then under the above assumption we have:

Proposition 19 *Let $D^\dagger_{x'} := J_n - I_n$, where I_n is the $n \times n$ identity matrix and J_n is the $n \times n$ reverse diagonal matrix ($\forall\ i,j \in [\![n]\!]$, $J_n(i,j) := \delta_{n-j+1}(i)$). The matrix $D^\dagger_{x'}$ is a minimizer of the functional $H_{x'}$ on $\mathcal{D}_{x'}$. Thus, when L is the transition kernel of a random walk on a regular graph whose spectral gap is of multiplicity 1, $D^\dagger = (D^\dagger_{x'})_{x' \in V}$ corresponds to the best way to perturb $M^{(0)}$ in order to improve infinitesimally its spectral gap.*

The proof of this proposition will be based on the following technical result:

Lemma 20 *Consider $\mathcal{SP}_n$ the set of $n \times n$ doubly substochastic matrices and the functional G defined on $\mathcal{SP}_n$ by*

$$\forall\ P = (P(i,j))_{i,j \in [\![n]\!]} \in \mathcal{SP}_n, \qquad G(P) \ = \ \sum_{i,j \in [\![n]\!]} P(i,j) g(i) h(j) \tag{22}$$

where $g, h : [\![n]\!] \to \mathbb{R}_+$ are two nondecreasing functions. Then I_n is a maximizer of G on $\mathcal{SP}_n$.

Proof

Note that for any $P \in \mathcal{SP}_n$ we can find a $Q \in \mathcal{P}_n$, the set of $n \times n$ doubly stochastic matrices, such that all the entries of Q are larger or equal to those of P. Since g and h are nonnegative functions, it follows that $G(Q) \geq G(P)$, and it is sufficient to prove the above lemma with $\mathcal{SP}_n$ replaced by $\mathcal{P}_n$. To do so, we consider an induction on $n \in \mathbb{N} \setminus \{0\}$. For $n = 1$ the result is obvious, since $\mathcal{P}_1 = \{I_1\}$. Next assume that the above lemma is true for some $n \in \mathbb{N} \setminus \{0\}$ (and for all g, h as in the lemma) and consider P a maximizer on $\mathcal{P}_{n+1}$ of a functional G which has the form described in (22) with n replaced by $n+1$. If there exists $k \in [\![n]\!]$ such that $P(k, n+1) > 0$, then $P(n+1, n+1) < 1$ and we can find $l \in [\![n]\!]$ such that $P(n+1, l) > 0$. Assume for instance that $P(k, n+1) \leq P(n+1, l)$ (otherwise work with the transposed matrix) and introduce the matrix $P' \in \mathcal{P}_{n+1}$ defined by

$$\forall\ i,j \in [\![n+1]\!],\ P'(i,j) \ = \ \begin{cases} 0 & \text{, if } (i,j) = (k, n+1) \\ P(n+1,l) - P(k,n+1) & \text{, if } (i,j) = (n+1, l) \\ P(i,j) + P(k,n+1) & \text{, if } (i,j) = (n+1, n+1) \text{ or } (i,j) = (k,l) \\ P(i,j) & \text{, otherwise} \end{cases}$$

We compute that

$$\begin{aligned} G(P) - G(P') \ &= \ (g(n+1)h(n+1) + g(k)h(l) - g(n+1)h(l) - g(k)h(n+1))P(n+1,k) \\ &= \ (g(n+1) - g(k))(h(n+1) - h(l))P(n+1,k) \end{aligned}$$

quantity which is nonnegative due to the assumptions $g(k) \leq g(n+1)$ and $h(l) \leq h(n+1)$. Thus P' is also a maximizer of G on $\mathcal{P}_{n+1}$. Iterating this procedure, we end up with a maximizer $Q \in \mathcal{P}_{n+1}$ such that $Q(n+1, i) = Q(i, n+1) = 0$ for all $i \in [\![n]\!]$ and $G(Q) = G'(Q') + g(n+1)h(n+1)$, where $Q' := (Q(i,j))_{i,j \in [\![n]\!]} \in \mathcal{P}_n$ and G' is the mapping on $\mathcal{P}_n$ defined by the rhs of (22). So it remains to apply the iterative assumption to get the result.

■

Let us now proceed to the

Proof of Proposition 19

We first assume that n is even, say $n = 2m$. Let $l \in \mathbb{R}$ be given, for any $D = (D(i,j))_{i,j \in [\![n]\!]} \in \mathcal{D}_{x'}$, we have by definition of $\mathcal{D}_{x'}$ and due to the fact that $L(x', \cdot) \equiv 1/n$ on $V_{x'}$ (set which was defined

at the beginning of Section 2), we have

$$\begin{aligned}
nH_{x'}(D) &= \sum_{x,x''\in V} D(x,x'')(f(x)-l+l)(f(x'')-l+l) \\
&= \sum_{x,x''\in V} D(x,x'')(f(x)-l)(f(x'')-l) \\
&\quad +l\sum_{x,x''\in V} D(x,x'')(f(x)+f(x'')) - l^2\sum_{x,x''\in V} D(x,x'') \\
&= \sum_{x,x''\in V} D(x,x'')(f(x)-l)(f(x'')-l) \\
&\quad +2l\sum_{x,x''\in V} D(x,x'')f(x) \\
&= \sum_{x,x''\in V} D(x,x'')(f(x)-l)(f(x'')-l) \qquad (23)
\end{aligned}$$

Choose

$$l = \frac{f(m)+f(m+1)}{2} \qquad (24)$$

so that $(f(x)-l)(f(x'')-l)$ is nonpositive for $(x,x'')\in [\![m]\!]\times[\![m+1,n]\!]\sqcup[\![m+1,n]\!]\times[\![m]\!]$ and nonnegative for $(x,x'')\in[\![m]\!]\times[\![m]\!]\sqcup[\![m+1,n]\!]\times[\![m+1,n]\!]$. It follows that to minimizes $H_{x'}$ on $\mathcal{D}_{x'}$, it is enough to consider $D\in\mathcal{D}_{x'}$ such that $D_{-,-}:=(D(i,j))_{i,j\in[\![m]\!]}$ and $D_{+,+}:=(D(i,j))_{i,j\in[\![m+1,n]\!]}$ have nonpositive entries and such that $D_{-,+}:=(D(i,j))_{(i,j)\in[\![m]\!]\times[\![m+1,n]\!]}$ and $D_{+,-}:=(D(i,j))_{(i,j)\in[\![m+1,n]\!]\times[\![m]\!]}$ have nonnegative entries. Indeed, for instance if $D(i,j)>0$ with $i,j\in[\![m]\!]$, consider the matrix $D'\in\mathcal{D}_{x'}$ defined by

$$\forall\ i',j'\in[\![n]\!],\quad D'(i',j') = \begin{cases} 0 & \text{, if } (i',j')=(i,j) \text{ or } (i',j')=(j,i) \\ D(i',j')+D(i,j) & \text{, if } (i',j')=(m+1,i) \text{ or } (i',j')=(m+1,j) \\ D(i',j')+D(i,j) & \text{, if } (i',j')=(i,m+1) \text{ or } (i',j')=(j,m+1) \\ D(i',j')-2D(i,j) & \text{, if } (i',j')=(m+1,m+1) \\ D(i',j') & \text{, otherwise} \end{cases}$$

We have

$$\begin{aligned}
H_{x'}(D') &= H_{x'}(D) - 2(\widetilde{f}(i)\widetilde{f}(j) - (\widetilde{f}(i)+\widetilde{f}(j))\widetilde{f}(m+1) + \widetilde{f}^2(m+1))D(i,j) \\
&\leq H_{x'}(D)
\end{aligned}$$

where we have taken $\widetilde{f}:=f-l$.
The case where there exists $(i,j)\in[\![m]\!]\times[\![m+1,n]\!]$ with $D(i,j)<0$ is even easier to treat: just transfer the negative weights at the entries (i,j) and (j,i) to the entries (i,i) and (j,j).
Thus to find a minimizer $D\in\mathcal{D}_{x'}$ for $H_{x'}$, we can assume that the entries of $D_{-,-}$ and $D_{+,+}$ are nonpositive while those of $D_{-,+}$ and $D_{+,-}$ are nonnegative. Define $P_{-,-}$ as the matrix obtained from $-D_{-,-}$ by reversing the order of the rows and of the columns, $P_{+,+}:=-D_{+,+}$, $P_{-,+}$ as the matrix obtained from $D_{-,+}$ by inverting the order of the rows, and $P_{+,-}$ as the matrix obtained from $D_{+,-}$ by inverting the order of the columns, so that these four matrices are doubly substochastic. We can write naturally

$$nH_{x'}(D) = -(G_{-,-}(P_{-,-}) + G_{+,+}(P_{-,-}) + G_{-,+}(P_{-,+}) + G_{+,-}(P_{+,-}))$$

where for instance $G_{-,-}$ is the functional on $\mathcal{SP}_m$ defined by

$$\forall\ P=(P(i,j))_{i,j\in[\![m]\!]}\in\mathcal{SP}_m,\qquad G_{-,-}(P) = \sum_{i,j\in[\![m]\!]} P(i,j)g(i)h(j)$$

where $g = h$ is the nonnegative and nondecreasing function given by

$$\forall\, i \in [\![m]\!], \qquad g(i) \;:=\; -f(m-i+1)$$

In fact the four functionals $G_{-,-}$, $G_{+,+}$, $G_{-,+}$ and $G_{+,-}$ are of the same type as the map G considered in Lemma 20. It follows that I_m is a maximizer for all these functional. Coming back to the description of D in terms of $P_{-,-}$, $P_{+,+}$, $P_{-,+}$ and $P_{+,-}$, we get that $D = J_n - I_n$ is in fact a minimizer for $H_{x'}$.

It remains to treat the case where n is odd, say $n = 2m+1$. In this situation, we rather define $l = f(m+1)$ instead of (24), so (23) enables to erase the $(m+1)^{\text{th}}$ row and the $(m+1)^{\text{th}}$ column. Then the above arguments can be applied to the remaining matrix.

■

The proof of Lemma 20 also shows that the maximizer is unique if and only if $g(1) < g(2) < \cdots < g(n)$ and $h(1) < h(2) < \cdots < h(n)$. From the above proof we deduce that the minimizer of $H_{x'}$ is unique if and only if f is one-to-one on $V_{x'}$. If this is true for all $x' \in V$, we get from Lemma 18 that there exists a unique $D^\dagger \in \mathcal{D}$ such that (21) is satisfied. In particular this condition is fulfilled if the eigenvector f is globally one-to-one.

Remark 21 The random walk assumption is quite restrictive, but we believe that the qualitative fact that the best infinitesimal perturbation corresponds to jumping from smallest values to largest values of an eigenvector associated to the spectral gap (on the neighborhood V_x, for $x \in V$) can be extended, except that in the general case, one would have to use the repartition function for the values of the restriction of this eigenfunction to the neighborhood V_x endowed with the probability measure $L(x,\cdot)$. We won't go further in this direction, since our purpose was just to indicate that the local landscape induced by such an eigenvector can be used to optimally improve the spectral gap, at least infinitesimally.

□

Under some circumstances, one can get nice informations about the speed of convergence to equilibrium by studying singular values, see for instance Fill [4]. This is not the case in our situation: the singular values associated to a trajectorially reversible second order Markov kernel are relatively easy to compute but they don't provide pertinent quantitative informations:

Proposition 22 *Let M be a second order Markov kernel trajectorially reversible with respect to $\nu \times L$ (with $\nu > 0$) and consider its dual operator M^* in $\mathbb{L}^2(\nu \times L)$. The spectrum of the Markovian operator M^*M in $\mathbb{L}^2(\nu \times L)$ is the union, with multiplicities, of the spectra of the M_x^2, for $x \in V$, where M_x is the transition kernel on V_x defined at the beginning of Section 2.*

We knew a priori that the kernels M_x, for $x \in V$, were diagonalizable in $\mathbb{R}$, since they are reversible with respect to the probability measures $L(x,\cdot)$. The diagonalizability in $\mathbb{R}$ of M^*M (in $\mathbb{L}^2(\nu \times L)$) comes from its self-adjointness in $\mathbb{L}^2(\nu \times L)$. The examples of next section show that usually the information contained in the M_x, for $x \in V$, is of a local nature (it just describes how the speed evolves above the point $x \in V$) and is not sufficient to deduce the global rate of convergence to equilibrium. That is why in our context the singular value approach is not very promising.

Proof

Taking into account Lemme 7, we compute that for any $F \in \mathcal{F}(V^2)$ and any $(x,x') \in V^2$,

$$\begin{aligned} M^*M[F](x,x') &= \sum_{x'',x''' \in V} M(x',x;x'')M(x'',x;x''')F(x,x''') \\ &= \sum_{x''' \in V} \left(\sum_{x'' \in V} M_x(x',x'')M_x(x'',x''') \right) F(x,x''') \\ &= \sum_{x''' \in V} M_x^2(x',x''')F(x,x''') \end{aligned}$$

It follows that for any fixed $x \in V$, the set $\{(x,x') \,:\, x' \in V_x\}$ is stable by the Markov kernel M^*M. Furthermore the restriction of M^*M to $\{(x,x') \,:\, x' \in V_x\}$ is given by the matrix M_x^2. The announced results follow, once one has identified $\mathbb{L}^2(\nu \times L)$ with the set of functions defined on the support of $\nu \times L$, which is $\sqcup_{x \in V}\{(x,x') \,:\, x' \in V_x\}$.

■

To end up this section, we make some connections with the work of Neal [10]. We begin by recalling the notion of asymptotic variance. Let P an irreducible Markov kernel on V admitting μ as invariant probability. Consider $(X_n)_{n\in\mathbb{N}}$ a Markov chain whose transition kernel is P and let $f \in \mathcal{F}(V)$ a function satisfying $\mu[f] = 0$. Then for $n \in \mathbb{N}$ large, the variance of the random variable $(f(X_1) + \cdots + f(X_n))/\sqrt{n}$ admits a limit $\sigma(P,f)$ called the asymptotic variance of f (it does not depend on the initial distribution of X_0). This result can be extended to second order Markov chains: let $(X_n)_{n\in\mathbb{N}}$ be such a chain with second order transition kernel M. Assume that M is irreducible on a subset of V^2 and let m be the associated invariant measure. Let $f \in \mathcal{F}(V)$ be a function whose integral with respect to the second marginal of m (or with respect to the first one, see Remark 6) is null. Then the above limit also exists and again it is denoted by $\sigma(M,f)$.

Next consider a second order Markov kernel $M^{(1)}$, trajectorially reversible with respect to $\nu \times L$, with $\nu > 0$ and L irreducible. Assume that the relaxed version of (18) is fulfilled:

$$\forall\, x'' \neq x \in V, \qquad M^{(1)}(x,x';x'') \;\geq\; L(x',x'')$$

and that $M^{(1)}$ is irreducible on the support of $\nu \times L$ (in his Theorem 2, Neal [10] stated this irreducibility condition is a consequence of the previous assumptions, but this is wrong in general, consider the example of next section with $a = 1$). Then Neal [10] has proven that

$$\sigma(M^{(1)},f) \;\leq\; \sigma(M^{(0)},f) \tag{25}$$

for any function $f \in \mathcal{F}(V)$. The link of this bound with relations between the spectrum of $M^{(1)}$ and that of $M^{(0)}$ is not obvious. To try to clarify the situation, let us recall Peskun's theorem [11]. Indeed, Neal [10] also gave a new and very elegant proof of Peskun's theorem, based as that of (25), on a decomposition of the trajectories. So consider two irreducible Markov kernels $P^{(0)}$ and $P^{(1)}$, both reversible with respect to the same probability measure μ on V. Assume that $P^{(1)}$ has less tendency to stay at the same place than $P^{(0)}$:

$$\forall\, x \neq x' \in V, \qquad P^{(1)}(x,x') \;\geq\; P^{(0)}(x,x')$$

Then Peskun's theorem asserts that

$$\forall\, f \in \mathcal{F}(V), \qquad \sigma(P^{(1)},f) \;\leq\; \sigma(P^{(0)},f)$$

In this situation, this bound leads to relations between the spectra of $P^{(1)}$ and $P^{(0)}$. Indeed, one have, for $i = 1,2$, that (see for instance Peskun [11])

$$\sigma(P^{(i)},f) \;=\; 2\mu[(f-\mu[f])(\mathrm{Id}-P^{(i)})^{-1}[f-\mu[f]]] - \mu[(f-\mu[f])^2]$$

where $(\mathrm{Id}-P^{(i)})^{-1}$ is the inverse of the operator $\mathrm{Id}-P^{(i)} \,:\, \mathcal{F}_0(\mu) \to \mathcal{F}_0(\mu)$, with $\mathcal{F}_0(\mu)$ the subspace of $\mathcal{F}(V)$ of functions f such that $\mu[f] = 0$. Using the variational formulation of eigenvalues (see for instance the book of Horn and Johnson [6]), it follows that

$$\forall\, j \in [\![|V|]\!], \qquad \theta_j^{(1)} \;\leq\; \theta_j^{(0)}$$

where $1 = \theta_1^{(i)} \geq \theta_2^{(i)} \geq \cdots \geq \theta_{|V|}^{(i)} \geq -1$ is the spectrum of $P^{(i)}$. This observation enables to see that if for $a \in [0,1]$ and $j \in [\![|V|]\!]$, $\theta_j^{(a)}$ is the jth eigenvalue (in nonincreasing order) of $P^{(a)} := (1-a)P^{(0)} + aP^{(1)}$, then the mapping $[0,1] \ni a \mapsto \theta_j^{(a)} \in [-1,1]$ is nonincreasing. Next,

with the notations of the introduction, consider $1-\lambda(P^{(a)}) = \max\{|\theta_2^{(a)}|, |\theta_{|V|}^{(a)}|\}$. Let $a_0 = \inf\{a \in [0,1] \,:\, 1-\lambda(a) = -\theta_{|V|}^{(a)}\}$, with the convention that $a_0 = 1$ if the previous set is empty. It appears that the mapping $[0,1] \ni a \mapsto 1-\lambda(a)$ is nonincreasing on $[0,a_0]$ and nondecreasing on $[a_0,1]$. Note that this corresponds to the qualitative behavior we are hoping for in Conjecture 3. Nevertheless, there is a strong difference between the behavior of the spectrum in the situation of Peskun's theorem and in the setting of Theorem 1: initially at $a=0$, the nonnull eigenvalues begin by going toward 0 in Theorem 1, while they are decreasing in Peskun's theorem (so for the eigenvalues starting from a negative value, they are going outward 0).

This discussion cannot be extended to (25), first because this equation is only valid for functions depending on one variable and second because we don't know if the variational formulation of eigenvalues is true for second order kernels, which usually are not reversible in the usual sense. So despite the assumptions of Theorem 1 are similar to those considered by Neal [10] and the fact that we are working in a similar spirit, it seems there is no obvious relation between (25) and our results.

4 A case study

We are interested here in the simplest family of examples in the homogeneous setting: the group G is the cyclic group $\mathbb{Z}_N$, with $N \in \mathbb{N} \setminus \{0,1,2\}$ and the speed only takes values in the symmetric set $S = \{-1,+1\} \subset \mathbb{Z}_N$. In particular we will check that Theorem 2 is true in this case.

The natural state space of the associated second order Markov chain is the set of nearest neighbors $G^{(2)} := \{(x,x') \in \mathbb{Z}_N^2 \,:\, |x'-x| = 1\}$, and for any $a \in [0,1]$, consider

$$\forall\ (x,x') \in G^{(2)},\ \forall\ x'' \in G, \qquad M^{(a)}(x,x';x'') \ := \ \begin{cases} (1+a)/2 & \text{, if } x' = x+1 \text{ and } x'' = x'+1 \\ (1+a)/2 & \text{, if } x' = x-1 \text{ and } x'' = x'-1 \\ (1-a)/2 & \text{, if } x' = x+1 \text{ and } x'' = x'-1 \\ (1-a)/2 & \text{, if } x' = x-1 \text{ and } x'' = x'+1 \\ 0 & \text{, otherwise} \end{cases}$$

This is a homogeneous situation since the speed transition kernel does not depend on the underlying position and is given by

$$K^{(a)} \ = \ \begin{pmatrix} \frac{1+a}{2} & \frac{1-a}{2} \\ \frac{1-a}{2} & \frac{1+a}{2} \end{pmatrix}$$

The set G^* defined in (16) is the collection of the mappings $\rho \,:\, G \ni x \mapsto \exp(2\pi i k x/N) \in \mathbb{C}$, where $k \in [\![0, N-1]\!]$. Thus according to the comments made after Proposition 11 and before (17), the spectral decomposition of $M^{(a)}$ into Jordan blocks corresponds to the union of the spectral decompositions of the matrices

$$B_\rho^{(a)} \ := \ \begin{pmatrix} \frac{1+a}{2}\rho(1) & \frac{1-a}{2} \\ \frac{1-a}{2} & \frac{1+a}{2}\rho(-1) \end{pmatrix}$$

where $\rho \in G^*$. For the rest of this section, let us identify ρ with $\rho(1)$ and G^* with $\{\exp(2\pi i k/N) \,:\, k \in [\![0, N-1]\!]\} \subset \mathbb{C}$. Consider more generally, for $\rho \in \mathbb{T}$, the unit circle in $\mathbb{C}$, the matrix defined by

$$B_\rho^{(a)} \ := \ \begin{pmatrix} \frac{1+a}{2}\rho & \frac{1-a}{2} \\ \frac{1-a}{2} & \frac{1+a}{2}\bar{\rho} \end{pmatrix} \tag{26}$$

These matrices were already used in Diaconis, Holmes and Neal [3] to deduce the spectral decomposition of a related model, where the discrete circle $\mathbb{Z}_N$ is replaced by the path $[\![N]\!]$ (but

the authors did not interpret the chain they studied as a second order Markov chain, nor as a position-speed Markov chain, furthermore their reduction to the above matrices is not deduced in the same manner as here).

A direct computation shows the following result:

Lemma 23 *Let write $\rho =: C + iS$, with $C = \Re(\rho), S = \Im(\rho) \in [-1,1]$ the respective real and imaginary parts of ρ, and let $a_0(\rho) := (1-|S|)/(1+|S|) \in [0,1]$. The matrix $B_\rho^{(a_0(\rho))}$ is not diagonalizable and is in the conjugacy class of a 2×2-Jordan block associated to the eigenvalue $(1+a)C/2$ (which will also conveniently denoted $\theta_{\rho,\pm}^{(a_0(\rho))}$). If $a \in [0, a_0(\rho))$, the matrix $B_\rho^{(a)}$ is diagonalizable in $\mathbb{R}$ and its eigenvalues are*

$$\theta_{\rho,\pm}^{(a)} = \frac{1+a}{2}C \pm \frac{\sqrt{(1-a)^2 - S^2(1+a)^2}}{2}$$

If $a \in (a_0(\rho), 1]$, the matrix $B_\rho^{(a)}$ is diagonalizable in $\mathbb{C}$ and its eigenvalues are

$$\theta_{\rho,\pm}^{(a)} = \frac{1+a}{2}C \pm i\frac{\sqrt{S^2(1+a)^2 - (1-a)^2}}{2}$$

Thus it appears that for $a \in [0,1]$, $M^{(a)}$ is diagonalizable if and only if $a \notin \{a_0(\rho) : \rho \in G^*\}$. In any case, the spectrum of $M^{(a)}$ coincides with the set $\{\theta_{\rho,\epsilon}^{(a)} : \rho \in G^*, \epsilon \in \{-,+\}\}$. Note that the eigenvalue 1 of $M^{(a)}$ associated to the constant eigenvectors correspond to the eigenvalue 1 of the matrix $B_1^{(a)}$. The other eigenvalue of $B_1^{(a)}$ is a, so with the notations of the introduction we have

$$\Theta_*(M^{(a)}) = \{\theta_{\rho,\epsilon}^{(a)} : \rho \in G^* \setminus \{1\}, \epsilon \in \{-,+\}\} \cup \{a\}$$

In particular the spectral gap of $M^{(a)}$ is given by

$$\lambda(M^{(a)}) = 1 - \max(\{\theta_{\rho,*}^{(a)} : \rho \in G^* \setminus \{1\}\} \cup \{a\}) \tag{27}$$

where for any $\rho \in \mathbb{T}$,

$$\theta_{\rho,*}^{(a)} := \max\left\{\left|\theta_{\rho,-}^{(a)}\right|, \left|\theta_{\rho,+}^{(a)}\right|\right\} \tag{28}$$

Note that the quantities $\theta_{\rho,\pm}^{(a)}$ only depend on the real part of $\rho \in \mathbb{T}$, i.e. $\theta_{\rho,*}^{(a)} = \theta_{\bar{\rho},*}^{(a)}$, and that $\{\theta_{\rho,-}^{(a)}, \theta_{\rho,+}^{(a)}\} = \{-\theta_{-\rho,-}^{(a)}, -\theta_{-\rho,+}^{(a)}\}$. It follows that $\theta_{\rho,*}^{(a)}$ is a function of the absolute value of the real part of $\rho \in \mathbb{T}$. A first step toward Theorem 2 is the following study:

Lemma 24 *Let $\rho \in \mathbb{T} \setminus \mathbb{R}$ be given. On $[0, a_0(\rho)]$ the mapping $a \mapsto \theta_{\rho,*}^{(a)}$ is strictly concave and decreasing. For $a \in [a_0(\rho), 1]$, we have $\theta_{\rho,*}^{(a)} = \sqrt{a}$.*

Proof

According to the above observations, we can assume that the argument of ρ belongs to $(0, \pi/2]$. In this situation, for $a \in [0, a_0(\rho)]$, we have

$$\begin{aligned}\theta_{\rho,*}^{(a)} &= \theta_{\rho,+}^{(a)} \\ &= \frac{1+a}{2}C + \frac{\sqrt{(1-a)^2 - S^2(1+a)^2}}{2}\end{aligned}$$

where $C = \Re(\rho)$, $S = \Im(\rho)$. It is clear that $\theta_{\rho,*}^{(a)}$ is a smooth function of $a \in [0, a_0(\rho))$ and its second derivative is given by

$$\frac{d^2}{d^2a}\theta_{\rho,*}^{(a)} = -\frac{2S^2}{((1-a)^2 - S^2(1+a)^2)^{\frac{3}{2}}}$$

Thus it appears that $a \mapsto \theta_{\rho,*}^{(a)}$ is strictly concave. Either by direct computation or by using Theorem 1, we get that $\frac{d}{da}\theta_{\rho,*}^{(0)} < 0$, so $\theta_{\rho,*}^{(a)}$ is a decreasing function of $a \in [0, a_0(\rho)]$.
For $a \in [a_0(\rho), 1]$, we have $\left|\theta_{\rho,+}^{(a)}\right| = \left|\theta_{\rho,-}^{(a)}\right|$ and

$$\begin{aligned} \theta_{\rho,*}^{(a)} &= \left|\theta_{\rho,+}^{(a)}\right| \\ &= \frac{1}{2}\sqrt{(1+a)^2C^2 + S^2(1+a)^2 - (1-a)^2} \\ &= \sqrt{a} \end{aligned}$$

■

So the look of the mapping $[0,1] \ni a \mapsto \theta_{\rho,*}^{(a)}$ is: it decreases until it reaches the curve $[0,1] \ni a \mapsto \sqrt{a}$ (at $a_0(\rho)$) that it follows afterward (and thus is increasing). Note that just before $a_0(\rho)$, $\theta_{\rho,*}^{(a)}$ is strongly decreasing, since $\lim_{a \to a_0(\rho)-} \frac{d}{da}\theta_{\rho,*}^{(a)} = -\infty$, as it can be easily be seen. In particular we have

$$\forall\ a \in [0,1], \qquad \theta_{\rho,*}^{(a)} \ \geq\ \sqrt{a} \ \geq\ a \tag{29}$$

so (27) can be transformed into

$$\lambda(M^{(a)}) \ = \ 1 - \max\{\theta_{\rho,*}^{(a)} \ : \ \rho \in G^* \setminus \{1\}\}$$

Furthermore, since the maximum of curves which are first nonincreasing and next nondecreasing has the same appearance, we already deduce that the mapping $[0,1] \ni a \mapsto \lambda(a)$ first nondecreases and next nonincreases. To get more information in this direction, we need the following simple result.

Lemma 25 *Let $\rho, \rho' \in \mathbb{T}$ be such that their respective real parts satisfy $C > C' \geq 0$. Then for any $a \in [0,1]$ we have,*

$$\theta_{\rho,*}^{(a)} \ \geq\ \theta_{\rho',*}^{(a)}$$

Proof
The assumption $0 \leq C' < C$ clearly implies $S^2 < (S')^2$ for the corresponding imaginary parts and $a_0(\rho) > a_0(\rho')$, so in view of the above behaviors, it is sufficient to prove the wanted inequality for $a \in [0, a_0(\rho')]$. But in this interval it can be written

$$\frac{1+a}{2}C + \frac{\sqrt{(1+a)^2C^2 - 4a}}{2} \ \geq\ \frac{1+a}{2}C' + \frac{\sqrt{(1+a)^2(C')^2 - 4a}}{2}$$

and so it is obvious (the strict inequality even holds there).

■

This result leads us to consider $\rho_0 \in G^*$ whose real part is the largest possible in absolute value, since we get

$$\forall\ a \in [0,1], \qquad \lambda(M^{(a)}) \ = \ 1 - \theta_{\rho_0,*}^{(a)}$$

A picture enables to convince oneself that

$$\rho_0 \ = \ \exp(2\pi i \lfloor N/2 \rfloor / N)$$

where $\lfloor \cdot \rfloor$ stands for the integer part. Note that ρ_0 is also the unique element of G^* such that $\lambda(M^{(0)}) = 1 - \theta_{\rho_0,*}^{(0)}$ and recall that $\lambda(M^{(0)}) = \lambda(L)$, where in the present example L is the

transition kernel of the nearest neighbor random walk on $\mathbb{Z}_N$. According to the parity of N we have different results:

• When N is even, we have $\rho_0 = -1$ and $\lambda(L) = 0$, this reflects the periodicity of the corresponding random walk. We obtain that $\lambda(M^{(a)}) = 0$, for any $a \in [0,1]$, fact we could also have deduced a priori from periodicity (see the introduction).

• When N is odd, let $C_N := \cos(\pi/N)$ be the absolute value of the real part of ρ_0. We have $\lambda(L) = 1 - C_N$ and define $a_N := a_0(\rho_0) = (1+\sin(\pi/N))^{-1}(1-\sin(\pi/N))$. Note that for N large (and still odd) this number is close to 1 and that the beginning of its expansion in terms of $1/N$ is $a_N \simeq 1 - 2\pi/N$. We can also express it in terms of the spectral gap of L:

$$\begin{aligned} a_N &= \frac{1-\sqrt{1-C_N^2}}{1+\sqrt{1-C_N^2}} \\ &= \frac{1-\sqrt{\lambda(L)(2-\lambda(L))}}{1+\sqrt{\lambda(L)(2-\lambda(L))}} \end{aligned}$$

The spectral gap of $\lambda(M^{(a)})$ for $a \in [0,1]$ is then given by

$$\lambda(M^{(a)}) = \begin{cases} 1 - \frac{1+a}{2}C_N - \frac{\sqrt{(1+a)^2C_N^2-4a}}{2} & \text{, if } a \in [0, a_N] \\ 1-\sqrt{a} & \text{, if } a \in [a_N, 1] \end{cases} \tag{30}$$

Similarly to what we remarked for $\theta^{(a)}_{\rho,*}$ after the proof of Lemma 24, $\lambda(M^{(a)})$ is improving a lot just before a_N. The largest spectral gap of $M^{(a)}$ correspond to the choice $a = a_N$ and we get

$$\begin{aligned} \lambda(M^{(a_N)}) &= 1-\sqrt{a_N} \\ &= 1-\sqrt{\frac{1-\sqrt{\lambda(L)(2-\lambda(L))}}{1+\sqrt{\lambda(L)(2-\lambda(L))}}} \end{aligned}$$

Thus for N large and odd, we have

$$\lambda(M^{(a_N)}) \sim \frac{\pi}{N}$$

which is an important improvement with respect to the spectral gap of L satisfying $\lambda(L) \sim \frac{\pi^2}{2N^2}$. The diffusive effect has been erased.

Theorem 2 amounts to the above statements in the case of our example (since $S = \{-1,1\}$, replace $|S|$ in the formulas by 2 and note that the alternative $a_0 > 1$ does not occur in the present case).

More generally, the previous considerations (see also Gade and Overton [5], when the discrete circle is replaced by a path) enable to get the following picture for the evolution of the spectrum of $M^{(a)}$ (seen as an operator on $\mathcal{F}(G^{(2)})$) as a goes from 0 to 1, for N odd (a similar behavior holds for N even, except for that the eigenvalue -1 does not move). First note that for any $a \in [0,1]$, the spectrum of $M^{(a)}$ is included into $[-1,1] \cup \mathcal{C}(\sqrt{a})$ where $\mathcal{C}(\sqrt{a}) \subset \mathbb{C}$ is the circle centred at 0 and of radius $\sqrt{a}$. Next the eigenvalues go by pairs. At $a = 0$, one starts from zero and the other one from an eigenvalue of L and they go in the direction of each other (there is one particular case which is the pair containing the eigenvalue 1: this value does not move and just looks its partner, the eigenvalue a, coming in its direction). When they meet, they begin to leave the real line and go on the circle $\mathcal{C}(\sqrt{a}) \subset \mathbb{C}$. They will keep on moving on the circle (which itself is increasing), in a conjugate way. Putting aside the exceptional pair containing 1, the last pair to leave the real line is the one containing the eigenvalue corresponding to the spectral gap. Once this pair has left the interval $[-1,1]$, all the eigenvalues (except the two exceptional ones) have the same modulus and

so all of them correspond to the spectral gap. At $a = 1$, all eigenvalues are regularly distributed on the unit circle $\mathcal{C}(1)$ and are all of multipicity 2. Indeed $M^{(1)}$ has two irreducible classes, which correspond respectively to the determinist walks which always turn around $\mathbb{Z}_N$ in the clockwise or in the anti-clockwise direction and both of them have G^* as spectrum.

Remark 26 Up to a real factor, the matrices defined in (26) with $a \in [-1, 1]$ and $\rho \in \mathbb{T}$ correspond to the general 2×2 matrices which are symmetrical and centro-Hermitian. These properties enable to get some of the above results without entering into the computations of the eigenvalues. This approach, as well as a better exploitation of Lemma 12, would have to be developed further in order to go in the direction of Conjecture 3. But here let us just deduce the first part of Lemma 24. Recall the tilde operation on $\mathcal{F}(\{-1, +1\}, \mathbb{C})$ which was defined after the proof of Lemma 12: if $f = (f(-1), f(1)) \in \mathcal{F}(\{-1, +1\}, \mathbb{C})$, we have $\widetilde{f} := (\overline{f(1)}, \overline{f(-1)})$. When the matrix $B_\rho^{(a)}$ defined in (26) has two distinct real eigenvalues, the corresponding eigenvectors, say $\varphi_{\rho,-}^{(a)}$ and $\varphi_{\rho,+}^{(a)}$, can be assumed to be real. According to the remark after the proof of Lemma 12, for $\epsilon \in \{-, +\}$, we have furthermore that $\widetilde{\varphi}_{\rho,\epsilon}^{(a)}$ is proportional to $\varphi_{\rho,\epsilon}^{(a)}$. Consider the usual bilinear (not sesquilinear) form $\langle \cdot, \cdot \rangle$ defined by

$$\forall\ f = (f(-1), f(1)),\ g = (g(-1), g(1)) \in \mathcal{F}(\{-1, +1\}, \mathbb{C}), \qquad \langle f, g \rangle \ := \ f(-1)g(-1) + f(1)g(1)$$

The above properties of the eigenvectors imply that for $\epsilon \in \{-, +\}$,

$$\begin{aligned} \left\langle \widetilde{\varphi}_{\rho,\epsilon}^{(a)}, \varphi_{\rho,\epsilon}^{(a)} \right\rangle &= 2\varphi_{\rho,\epsilon}^{(a)}(-1)\varphi_{\rho,\epsilon}^{(a)}(1) \\ &\neq 0 \end{aligned}$$

otherwise, due to the proportionality of $\widetilde{\varphi}_{\rho,\epsilon}^{(a)}$ with $\varphi_{\rho,\epsilon}^{(a)}$, we would end up with $\varphi_{\rho,\epsilon}^{(a)} = (0, 0)$. Furthermore, since $B_\rho^{(a)}$ is symmetric with respect to $\langle \cdot, \cdot \rangle$, we get that

$$\theta_{\rho,+}^{(a)} \left\langle \widetilde{\varphi}_{\rho,-}^{(a)}, \varphi_{\rho,+}^{(a)} \right\rangle = \left\langle \widetilde{\varphi}_{\rho,-}^{(a)}, B_\rho^{(a)}[\varphi_{\rho,+}^{(a)}] \right\rangle = \left\langle B_\rho^{(a)}[\widetilde{\varphi}_{\rho,-}^{(a)}], \varphi_{\rho,+}^{(a)} \right\rangle = \overline{\theta_{\rho,-}^{(a)}} \left\langle \widetilde{\varphi}_{\rho,-}^{(a)}, \varphi_{\rho,+}^{(a)} \right\rangle = \theta_{\rho,-}^{(a)} \left\langle \widetilde{\varphi}_{\rho,-}^{(a)}, \varphi_{\rho,+}^{(a)} \right\rangle$$

and thus $\left\langle \widetilde{\varphi}_{\rho,-}^{(a)}, \varphi_{\rho,+}^{(a)} \right\rangle = 0$. Note that since we can find $t \in \mathbb{R}$ such that $(\varphi_{\rho,-}^{(a)} + t\varphi_{\rho,+}^{(a)})(-1) = 0$ and by consequence $\left\langle \widetilde{\varphi}_{\rho,-}^{(a)} + t\widetilde{\varphi}_{\rho,+}^{(a)}, \varphi_{\rho,-}^{(a)} + t\varphi_{\rho,+}^{(a)} \right\rangle = 0$, we have that $\left\langle \widetilde{\varphi}_{\rho,-}^{(a)}, \varphi_{\rho,-}^{(a)}) \right\rangle$ and $\left\langle \widetilde{\varphi}_{\rho,+}^{(a)}, \varphi_{\rho,+}^{(a)}) \right\rangle$ are of opposite signs. We can renormalize the eigenvectors so that these quantities are 1 and -1 or vice-versa. Taking into account the elementary fact that for any $c_1, c_2 \in \mathbb{R}$, we have

$$\inf_{s,t \geq 0\,:\, t-s \neq 0} \frac{c_1 t - c_2 s}{t - s} \ = \ \max(c_1, c_2)$$

we deduce that

$$\begin{aligned} \theta_{\rho,+}^{(a)} &= \inf_{f \in \mathcal{F}(\{-1,+1\},\mathbb{C})\,:\, \langle \widetilde{f}, f \rangle \neq 0} \frac{\left\langle \widetilde{f}, B_\rho^{(a)}[f] \right\rangle}{\left\langle \widetilde{f}, f \right\rangle} \\ \theta_{\rho,-}^{(a)} &= \sup_{f \in \mathcal{F}(\{-1,+1\},\mathbb{C})\,:\, \langle \widetilde{f}, f \rangle \neq 0} \frac{\left\langle \widetilde{f}, B_\rho^{(a)}[f] \right\rangle}{\left\langle \widetilde{f}, f \right\rangle} \end{aligned}$$

Due to the fact that for $a \in [0, a_0(\rho)]$, these eigenvalues have the same sign (because $\det(B_\rho^{(a)}) = 4a$), the first part of Lemma 24 is then a consequence of the affine nature of the map $[0, 1] \ni a \mapsto B_\rho^{(a)}$.

When the matrix $B_\rho^{(a)}$ has conjuguate nonreal eigenvalues, note that the tilde operation enables to pass from one eigenvector to the other eigenvector. Contrary to the above situation, we then have $\left\langle \widetilde{\varphi}_{\rho,\epsilon}^{(a)}, \varphi_{\rho,\epsilon}^{(a)} \right\rangle = 0$, for $\epsilon \in \{+,-\}$, because

$$\theta_{\rho,\epsilon}^{(a)} \left\langle \widetilde{\varphi}_{\rho,\epsilon}^{(a)}, \varphi_{\rho,\epsilon}^{(a)} \right\rangle = \left\langle \widetilde{\varphi}_{\rho,\epsilon}^{(a)}, B_\rho^{(a)}[\varphi_{\rho,\epsilon}^{(a)}] \right\rangle = \left\langle B_\rho^{(a)}[\widetilde{\varphi}_{\rho,\epsilon}^{(a)}], \varphi_{\rho,\epsilon}^{(a)} \right\rangle = \overline{\theta_{\rho,\epsilon}^{(a)}} \left\langle \widetilde{\varphi}_{\rho,\epsilon}^{(a)}, \varphi_{\rho,\epsilon}^{(a)} \right\rangle$$

□

5 Global improvement of the spectral gap

Here we reduce the proof of Theorem 2 to the computations made in the previous section. But we will also see that when the generating set S contains strictly more than two elements, then the resort to the second order Markov chains considered here only provides weak improvements of the spectral gap. Three examples will illustrate this unfortunate fact.

Let us recall the framework of Theorem 2: G is a finite Abelian group and S is a symmetric subset of generators of G, with $s := |S| \geq 2$. For any $a \in [0,1]$, we consider the homogeneous speed transition kernel given by

$$\forall\ y, y' \in S, \qquad K^{(a)}(y,y') \ := \ \begin{cases} \left(1 + \frac{a}{s-1}\right)\mu(y') & \text{, if } y' \neq -y \\ (1-a)\mu(y') & \text{, if } y' = -y \end{cases}$$

where μ is the uniform distribution on S. Of course, if G is a cyclic group and $S = \{-1, 1\}$, we recover the example treated in the preceding section.
Consider G^* the set defined in (16), where G has been identified with a product group $\prod_{l \in [\![1,r]\!]} \mathbb{Z}_{N_l}$ (with $r \in \mathbb{N} \setminus \{0\}$ and $N_1, ..., N_r \in \mathbb{N} \setminus \{0,1\}$). For $a \in [0,1]$, let $M^{(a)}$ be the second order Markov kernel associated to the homogeneous speed transition kernel $K^{(a)}$. According to the comments following Proposition 11, the spectral decomposition of $M^{(a)}$ into Jordan blocks corresponds to the union of the spectral decompositions of the matrices $A_\rho^{(a)} := \triangle(\rho) K^{(a)}$ (recall that $\triangle(\rho)$ is the diagonal matrix whose diagonal entries are given by ρ), for $\rho \in G^*$. Since S is generating G, any $\rho \in G^*$ is uniquely determined by $(\rho(y))_{y \in S}$, we can identify ρ with $(\rho(y))_{y \in S}$. More generally, let $\mathcal{R}$ be the set of functions $\rho : S \to \mathbb{T}$ which satisfies $\rho(-y) = \overline{\rho(y)}$ for any $y \in S$ and $\rho(0) = 1$ if $0 \in S$. For any $\rho \in \mathcal{R}$, we define $A_\rho^{(a)} := \triangle(\rho) K^{(a)}$ in the same way as above. The main step to understand their spectral structure is contained in the next result:

Proposition 27 *Let $a \in [0,1]$ and $\rho \in \mathcal{R}$ be fixed and assume that ρ is not proportional to $\mathbb{1}$, the element of $\mathcal{R}$ always taking the value 1 on S. Denote $b^{(a)} := a/(s-1)$, $c_\rho^{(a)} := (1+b^{(a)})s^{-1} \sum_{y \in S} \rho(y)$ and consider the matrix*

$$C_\rho^{(a)} \ := \ \begin{pmatrix} 0 & -b^{(a)} \\ 1 & c_\rho^{(a)} \end{pmatrix}$$

The matrix $A_\rho^{(a)}$ is in the same conjugacy class as a block diagonal matrix whose diagonal blocks are the 2×2 block $C_\rho^{(a)}$, with multiplicity one, complemented with some 1×1 block(s) $(b^{(a)})$ and/or $(-b^{(a)})$.

Proof
Let J be the reverse diagonal $S \times S$ matrix: $\forall\ y, y' \in S$, $J(y,y') := \delta_{-y}(y')$ (it was already encountered in Lemma 20). It appears that for any $a \in [0,1]$, we have

$$K^{(a)} \ = \ (1 + b^{(a)})K^{(0)} - b^{(a)} J \tag{31}$$

Using this formula, we compute that

$$\begin{aligned} A_\rho^{(a)}[\mathbb{1}] &= (1+b^{(a)})\triangle(\rho)K^{(0)}[\mathbb{1}] - b^{(a)}\triangle(\rho)J[\mathbb{1}] \\ &= (1+b^{(a)})\triangle(\rho)\mathbb{1} - b^{(a)}\triangle(\rho)\mathbb{1} \\ &= \rho \end{aligned}$$

On the other hand we have

$$\begin{aligned} A_\rho^{(a)}[\rho] &= (1+b^{(a)})\triangle(\rho)K^{(0)}[\rho] - b^{(a)}\triangle(\rho)J[\rho] \\ &= c_\rho^{(a)}\rho - b^{(a)}\mathbb{1} \end{aligned}$$

Thus the vector space W generated by $\mathbb{1}$ and ρ is left stable by the operator $A_\rho^{(a)}$. By assumption, W is of dimension 2 and in the basis $(\mathbb{1},\rho)$, the matrix associated to the restriction of $A_\rho^{(a)}$ to W is just $C_\rho^{(a)}$.

Next consider

$$W' \;:=\; \{f \in \mathcal{F}(S,\mathbb{C}) \,:\, \mu[f]=0 \text{ and } \mu[\bar\rho f]=0\}$$

and let us check that W' is also left stable by $A_\rho^{(a)}$. Indeed, using (31), we get that for any $f \in W'$,

$$A_\rho^{(a)}[f] \;=\; -b^{(a)}\triangle(\rho)J[f] \tag{32}$$

So we compute first that

$$\begin{aligned} \mu[A_\rho^{(a)}[f]] &= -b^{(a)}\mu[\rho J[f]] \\ &= -b^{(a)}s^{-1}\sum_{y\in S}\rho(y)f(-y) \\ &= -b^{(a)}s^{-1}\sum_{y\in S}\rho(-y)f(y) \\ &= -b^{(a)}\mu[\bar\rho f] \\ &= 0 \end{aligned}$$

and next that

$$\begin{aligned} \mu[\bar\rho A_\rho^{(a)}[f]] &= -b^{(a)}\mu[\bar\rho\rho J[f]] \\ &= -b^{(a)}\mu[J[f]] \\ &= -b^{(a)}\mu[f] \\ &= 0 \end{aligned}$$

The stability of W' by $A_\rho^{(a)}$ follows from these two equalities.

In order to go further in the spectral decomposition of $A_\rho^{(a)}$, we must verify that $\mathcal{F}(S,\mathbb{C}) = W \oplus W'$, which amounts here to check that $W \cap W' = \{0\}$ (because the liberty of $\mathbb{1}$ and $\bar\rho$ in the dual space of $\mathbb{L}^2[\mu,\mathbb{C}]$ implies that the dimension of W' is $|S|-2$). So let $f \in W \cap W'$ and write it under the form $v\mathbb{1} + w\rho$, with $v, w \in \mathbb{C}$. Let $r := \mu[\rho]$, which is also equal to $\mu[\bar\rho]$ by symmetry, thus r is a real number. We compute that

$$\begin{aligned} \mu[f] &= v + rw \\ \mu[\bar\rho f] &= rv + w \end{aligned}$$

Thus the system of equations $\mu[f]=0$, $\mu[\bar\rho f]=0$ admits a non trivial solution in v, w, if and only if $r^2=1$. But if this is true, then we have $|\mu[\rho]| = \mu[|\rho|] = 1$ and necessarily ρ has to be proportional to $\mathbb{1}$. This is excluded by our assumption and it follows that $v=w=0$, i.e. $f=0$.

To conclude the proof of the above proposition, it remains to show that the restriction $A_\rho^{(a)}|_{W'}$ of $A_\rho^{(a)}$ to W' is diagonalizable and that its only possible eigenvalues are $b^{(a)}$ and $-b^{(a)}$. Coming back to (32), we compute that for any $f \in W'$ and $y \in S$,

$$\begin{aligned}(A_\rho^{(a)})^2[f](y) &= -b^{(a)}\rho(y)A_\rho^{(a)}[f](-y) \\ &= -b^{(a)}\rho(y)(-b^{(a)})\rho(-y)f(y) \\ &= (b^{(a)})^2 f(y)\end{aligned}$$

The announced results are then a consequence of the identity $(A_\rho^{(a)}|_{W'})^2 = (b^{(a)})^2\mathrm{Id}$.

■

To finish the spectral analysis of $A_\rho^{(a)}$, say for fixed $a \in [0,1]$ and $\rho \in \mathcal{R}$ such that $\mathbb{1}$ are ρ are not proportional, we must study more precisely $C_\rho^{(a)}$. To come back to the situation of the previous section, we just apply the arguments of the above proof to the case where $s = 2$:

Lemma 28 *Let $a' = b^{(a)} \in [0,1]$ and $\rho' \in \mathbb{T}$ be such that its real part is given by $s^{-1}\sum_{y\in S}\rho(y)$. Then $C_\rho^{(a)}$ is in the same conjugacy class as*

$$\widehat{B}_{\rho'}^{(a')} := \begin{pmatrix} \frac{1+a'}{2}\rho' & \frac{1-a'}{2} \\ \frac{1-a'}{2} & \frac{1+a'}{2}\overline{\rho'} \end{pmatrix}$$

Proof

Let us starts from the matrix $\widehat{B}_{\rho'}^{(a')}$, where $a' \in [0,1]$ and $\rho' \in \mathbb{T}$. According to formula (17), this matrix is in the same conjugacy class as

$$\widehat{A}_{\rho'}^{(a')} := \begin{pmatrix} \frac{1+a'}{2}\rho' & \frac{1-a'}{2}\rho' \\ \frac{1-a'}{2}\overline{\rho'} & \frac{1+a'}{2}\overline{\rho'} \end{pmatrix}$$

The latter matrix has the same structure as the matrices $A_\rho^{(a)}$ considered above, but with a set S of cardinal 2. So by the computations of the proof of Proposition 27, we have

$$\begin{pmatrix} 0 & -a' \\ 1 & (1+a')\frac{\rho+\overline{\rho'}}{2} \end{pmatrix} = U_{\rho'}^{(a')}\widehat{A}_{\rho'}^{(a')}(U_{\rho'}^{(a')})^{-1}$$

where

$$U_{\rho'}^{(a')} := \begin{pmatrix} 1 & \rho' \\ 1 & \overline{\rho'} \end{pmatrix}$$

Thus if we take $a' = b^{(a)} \in [0,1]$ and $\rho' \in \mathbb{T}$ such that its real part is given by $s^{-1}\sum_{y\in S}\rho(y)$, we recover the matrix $C_\rho^{(a)}$ and the wanted result follows.

■

The matrix $\widehat{B}_{\rho'}^{(a')}$ coincides with the matrix defined in (26), where a and ρ have to be replaced by a' and ρ'. As a consequence the spectral decomposition of $C_\rho^{(a)}$ is given by Lemma 23, where a and ρ have to be replaced by a' and ρ'. In particular, the bound (29) implies that the largest modulus of the eigenvalues of $C_\rho^{(a)}$ is larger than $a' = b^{(a)}$. Then Proposition 27 leads to the following result:

Proposition 29 *Let be given $a \in [0,1]$ and $\rho \in \mathcal{R}$ such that $\mathbb{1}$ are ρ are not proportional and define $a' \in [0,1]$ and $\rho' \in \mathbb{T}$ as in Lemma 28. Consider $\Theta_{\rho,*}^{(a)}$ the largest modulus of the eigenvalues of $A_\rho^{(a)}$. Then we have*

$$\Theta_{\rho,*}^{(a)} = \theta_{\rho',*}^{(a')}$$

where the rhs was defined in (28) (where a and ρ have to be replaced by a' and ρ').

We can now come to the

Proof of Theorem 2

To complete the previous investigation of the spectrum of the operator $M^{(a)}$, for $a \in [0,1]$, we must study the matrix $A_\rho^{(a)}$ when $\rho \in G^*$ is proportional to $\mathbb{1}$. So let $\rho \in G^*$ be such that there exists $z \in \mathbb{C}$ with

$$\forall\, y \in S, \qquad \rho(y) \;=\; z$$

By symmetry of S and due to the fact that $\rho(-y) = \overline{\rho(y)}$ for any $y \in S$, we get that $z \in \mathbb{R}$. Thus there is only two possibilities: either $z = 1$ or $z = -1$. In the latter case, namely if there exists $\rho \in G^*$ such that $\rho(y) = -1$ for all $y \in S$, we get that the underlying random walk kernel L is periodic: it only allows transitions from G_+ to G_- and from G_- to G_+, where for $\epsilon \in \{-,+\}$, $G_\epsilon \coloneqq \{x \in G \,:\, \rho(x) = \epsilon 1\}$. Then -1 is an eigenvalue of L (an associated eigenvector is $\mathbb{1}_{G_+} - \mathbb{1}_{G_-}$) and by consequence the spectral gap λ of L is zero. By periodicity, the spectral gap $\lambda(M^{(a)})$ of $M^{(a)}$, for $a \in [0,1]$, also vanishes and we get that $\Lambda \coloneqq \max_{a\in[0,1]} \lambda(M^{(a)}) = 0$. In this situation, Theorem 2 asserts that $a_0 = s - 1 \geq 1$ and formula (12) is true, because its rhs also vanishes.

Let us now consider the case $z = 1$, namely $\rho = \mathbb{1}$ (identifying elements of G^* with their representative in $\mathcal{R}$). The corresponding matrix $A_\rho^{(a)}$ is just $K^{(a)}$. The arguments used in the proof of Proposition 27 become simpler to deduce the spectral decomposition of $K^{(a)}$: the function $\mathbb{1}_S$ is an eigenvector of $K^{(a)}$ associated to the eigenvalue 1, the restriction of $K^{(a)}$ to $W' \coloneqq \{f \in \mathcal{F}(S,\mathbb{C}) \,:\, \mu[f] = 0\}$ is diagonalizable and its spectrum is included into $\{-b^{(a)}, b^{(a)}\}$.

It follows from these considerations, that the largest modulus of the eigenvalues of $M^{(a)}$, outside the eigenvalue 1 associated to the eigenvector $\mathbb{1}_G$ (which according to Proposition 10, corresponds to the eigenvector $\mathbb{1}_S$ of $A_{\mathbb{1}}^{(a)}$), is the largest of the quantities $\Theta_{\rho,*}^{(a)}$, where $\rho \in G^* \backslash \{\mathbb{1}\}$. Furthermore, when L is aperiodic, which is the situation we will consider for the rest of this proof, all $\rho \in G^* \backslash \{\mathbb{1}\}$ are not proportional to $\mathbb{1}$. Due to Proposition 29 and Lemma 25, the $\rho \in G^* \setminus \{\mathbb{1}\}$ maximizing $\Theta_{\rho,*}^{(a)}$ do not depend on the value of $a \in [0,1]$. In particular they are the $\rho_0 \in G^* \setminus \{\mathbb{1}\}$ such that the spectral gap λ of L satisfies $1 - \lambda = \Theta_{\rho_0,*}^{(0)}$. Let us define

$$C_0 \;\coloneqq\; \left| s^{-1} \sum_{y\in S} \rho_0(y) \right|$$

According to Proposition 29 and to the computations of the previous section, this quantity does not depend on the particular choice of the $\rho_0 \in G^* \setminus \{\mathbb{1}\}$ such that $1 - \lambda = \Theta_{\rho_0,*}^{(0)}$ and indeed we have

$$1 - \lambda \;=\; C_0 \tag{33}$$

Furthermore, Proposition 29 leads to a result similar to (30): the spectral gap of $\lambda(M^{(a)})$ for $a \in [0,1]$ is then given by

$$\lambda(M^{(a)}) \;=\; \begin{cases} 1 - \frac{1+b^{(a)}}{2} C_0 - \frac{\sqrt{(1+b^{(a)})^2 C_0^2 - 4b^{(a)}}}{2} & \text{, if } b^{(a)} \in [0, b_0] \\ 1 - \sqrt{b^{(a)}} & \text{, if } b^{(a)} \in [b_0, 1] \end{cases} \tag{34}$$

where $b_0 \in [0,1]$ is defined by the equation

$$\frac{1+b_0}{2} C_0 - \frac{\sqrt{(1+b_0)^2 C_0^2 - 4b_0}}{2} \;=\; \sqrt{b_0}$$

if such a value exists in $[0,1]$ and $b_0 = 1$ otherwise. Recalling that $b^{(a)} = a/(s-1)$ and taking into account the relation (33), as well as the results of the previous section, the announced results follow at once.

■

Due to the fact that in Proposition 29 the quantity $a' = a/(s-1)$ is independent of $\rho \in G^*$, the behavior of the spectrum of $M^{(a)}$ described at the end of Section 4 is also valid here with three differences:
- the increasing circle containing the nonreal eigenvalues has radius $\sqrt{a/(s-1)}$,
- if $a_0 > 1$, not the all eigenvalues issued from the spectrum of L will encounter an eigenvalue issued from 0 before a reaches 1,
- there are some "unpaired" eigenvalues, but their value is either $-a/(s-1)$ or $a/(s-1)$, note in particular that no eigenvalue stays in 0 (for $a \neq 0$).

Unfortunately, in general Theorem 2 does not provide great improvements of the spectral gap, except if $|S| = 2$. This phenomenon was already observed by Neal [10], relatively to the improvement of the asymptotic variance (recall the discussion at the end of Section 3). To get an idea of the improvement brought by the resort to the second order Markov chains introduced in Theorem 2, let us consider an abstract family of examples parametrized by N belonging to a set $\mathcal{N}$, endowed with a filter so that the notion $N \to \infty$ has a meaning. Thus for any $N \in \mathcal{N}$, we are given a finite Abelian group G_N and a symmetric set of generators S_N, with $|S_N| \geq 2$. Denote by λ_N the spectral gap of the underlying random walk and more generally, put the index N to all the notations introduced in Theorem 2 to indicate their dependence on $N \in \mathcal{N}$.

Proposition 30 *Assume that there exists $s \in \mathbb{N} \setminus \{0,1,2\}$ such that for any $N \in \mathcal{N}$, $|S_N| = s$. Under the assumption that $\lim_{N\to\infty} \lambda_N = 0$, we have as $N \to \infty$,*

$$\begin{aligned} \Lambda_N &= \lambda(M^{(1)}) \\ &\sim \frac{s}{2(s-1)}\left(1+\left(\frac{s}{s-2}\right)^2\right)\lambda_N \end{aligned}$$

Proof

Clearly we have for the quantity a_N defined in (11), with λ replaced by λ_N,

$$\lim_{N\to\infty} a_N = s-1 > 1$$

So for N "large" enough, we get

$$\begin{aligned} \Lambda_N &= 1 - \frac{s}{2(s-1)}\left(1-\lambda_N+\sqrt{\left(\frac{s-2}{s}\right)^2 - \lambda_N(2-\lambda_N)}\right) \\ &= 1 - \frac{s}{2(s-1)}\left(1-\lambda_N+\frac{s-2}{s}\sqrt{1-\left(\frac{s}{s-2}\right)^2\lambda_N(2-\lambda_N)}\right) \\ &= 1 - \frac{s}{2(s-1)}\left(1-\lambda_N+\frac{s-2}{s}\left(1-\left(\frac{s}{s-2}\right)^2\lambda_N\right)\right) + o(\lambda_N) \\ &= \frac{s}{2(s-1)}\left(1+\left(\frac{s}{s-2}\right)^2\right)\lambda_N + o(\lambda_N) \end{aligned}$$

■

Let $s \in \mathbb{N} \setminus \{0,1,2\}$ be fixed. One can take for $\mathcal{N}$ the set of all couples (G,S) where G is a finite Abelian group and S a symmetric set of generators of S, with $|S| = s$. The assumptions of the previous result are then all satisfied if we endow $\mathcal{N}$ with the filter base corresponding to the subsets $\{(G,S) \in \mathcal{N} : \lambda_{(G,S)} \leq \epsilon\}$, for $\epsilon \in (0,1)$. Nevertheless, let us end this section with three more

concrete families of examples. The first family is similar to the one considered in the previous section, except that we add 0 to the generating set S.

Example 31 So let $\mathcal{N} = \mathbb{N} \setminus \{0,1,2\}$ and for $N \in \mathcal{N}$, we take $G_N = \mathbb{Z}_N$ and $S_N = \{-1,0,1\}$, so that $s := |S_N| = 3$. According to equation (33), we have

$$\begin{aligned} 1-\lambda_N &= \max\left\{\frac{1}{3}\left|1+\exp(2\pi ik/N)+\exp(-2\pi ik/N)\right| \,:\, k \in [\![1,N-1]\!]\right\} \\ &= \max\left\{\frac{1}{3}\left|1+2\cos(2\pi k/N)\right| \,:\, k \in [\![1,N-1]\!]\right\} \\ &= \frac{1}{3}(1+2\cos(2\pi/N)) \end{aligned}$$

because we easily check that here the maximum is attained for $k=1$ (this is an important difference with the previous section, in particular no periodicity phenomenon occurs now). Note that for $N=3$, $\lambda_N = 1$: this is a situation where the spectral gap cannot be improved and Theorem 2 gives us $a_3 = 0$ and $\Lambda_3 = \lambda_3$. Indeed for $N=3$, the spectral gap is worsening with $a \in [0,1]$, since (34) is always valid (in this particular case with $C_0 = b_0 = 0$) and we get $\lambda(M^{(a)}) = 1-\sqrt{a/2}$. But for $N \in \mathcal{N} \setminus \{3\}$, we get a strict improvement of the spectral gap. Nevertheless, it remains quite modest, because we compute that for $N \in \mathcal{N}$ large,

$$\lambda_N \sim \frac{4\pi^2}{3N^2}$$

and Proposition 30 gives us

$$\begin{aligned} \Lambda_N &\sim \frac{15}{2}\lambda_N \\ &\sim \frac{10\pi^2}{N} \end{aligned}$$

Note that we can also consider the case $N=2$, then $G_2 = S_2 = \mathbb{Z}_2$. As for $N=3$, this is a situation where the spectral gap cannot be improved: $\lambda_2 = 1$. Proposition 27 enables to see that for any $a \in [0,1]$, the spectrum of $M^{(a)}$ (acting here on $\mathbb{Z}_2^2$) is equal to $\{1, a, -\sqrt{a}, \sqrt{a}\}$ so that the spectral gap is worsening with a: $\lambda(M^{(a)}) = 1-\sqrt{a}$.

□

The second family of examples corresponds to the nearest neighbor multidimensional torus models.

Example 32 Let $r \in \mathbb{N}\setminus\{0,1\}$ be fixed. We denote $\mathcal{N} = (\mathbb{N}\setminus\{0,1,2\})^r$ and for $N = (N_1,...,N_r) \in \mathcal{N}$, consider the group $G_N = \prod_{l\in[\![1,r]\!]} \mathbb{Z}_{N_l}$ and the generating set

$$S_N := \{\pm e_l \,:\, l \in [\![1,r]\!]\}$$

where $e_l = (\delta_l(k))_{k\in[\![1,r]\!]}$. Note that $|S_N| = 2r$. Let us say that $N \to \infty$ if all its coordinates N_1, ..., N_r go to infinity.
Taking into account the parametrization of G^* given in (16), equation (33) gives

$$\begin{aligned} 1-\lambda_N &= \max\left\{\frac{1}{r}\left|\sum_{l\in[\![1,r]\!]}\cos(2\pi k_l/N_l)\right| \,:\, (k_1,...,k_r) \in \prod_{l\in[\![1,r]\!]}[\![0,N_l-1]\!]\setminus\{(0,...,0)\}\right\} \\ &= \frac{1}{r}\left|\sum_{l\in[\![1,r]\!]}\cos(2\pi\lfloor N_l/2\rfloor/N_l)\right| \end{aligned}$$

Then as $N \to \infty$, we have

$$\lambda_N \sim \frac{\pi^2}{2r}\sum_{l\in l_N}\frac{1}{N_l^2}$$

where $l_N := \{l \in [\![1,r]\!] \,:\, N_l \text{ is odd}\}$. In particular $\lim_{N\to\infty} \lambda_N = 0$ and Proposition 30 implies that as $N \to \infty$,

$$\Lambda_N \quad \sim \quad \frac{r}{2r-1}\left(1+\left(\frac{r}{r-1}\right)^2\right)\lambda_N$$

The case where some of the N_i, $i \in [\![1,r]\!]$, are equal to 2 can be treated in a similar way except that now

$$|S_N| \quad = \quad \sum_{i=1}^{r} \mathbb{1}_{\{N_i=2\}} + 2\mathbb{1}_{\{N_i>2\}}$$

In particular we can deal with the situation $G = \mathbb{Z}_2^2$, $S = \{(0,1),(1,0)\}$, which, up to an isomorphism of G, is the only case with $|S| = 2$ which was not treated in Section 4.

□

The third family of examples comes back to the cyclic group, but with a different set of generators, with cardinal 4.

Example 33 We take

$$\mathcal{N} \quad := \quad \{n^2 \,:\, n \in \mathbb{N}\setminus\{0,1\} \text{ and } n \text{ odd}\}$$

so that $\sqrt{N}$ is an odd integer for $N \in \mathbb{N}$. The notion $N \to \infty$ is the usual one.
For $N \in \mathbb{N}$, we consider $G_N := \mathbb{Z}_N$ and $S_N := \{\pm 1, \pm\sqrt{N}\}$. If one had allowed squares of even numbers in the definition of $\mathcal{N}$, the corresponding underlying random walks would be periodic, so that the study of the spectral gap is trivial. Indeed, for $\sqrt{N} \in \mathbb{N}\setminus\{0,1\}$, we always have, according to equation (33),

$$1-\lambda_N \quad = \quad \max\left\{\frac{1}{2}\left|\cos(2\pi k/N)+\cos(2\pi k/\sqrt{N})\right| \,:\, k \in [\![1,N-1]\!]\right\}$$

and if $\sqrt{N}$ is even, we get $\lambda_N = 0$.
For $N \in \mathcal{N}$ large enough, the maximum is attained for $k = (N-1)/2$ (and for $k = (N+1)/2$) and it appears that

$$\begin{aligned}\lambda_N \quad &= \quad 1-\frac{1}{2}\left(\cos(\pi/N)+\cos(\pi/\sqrt{N})\right)\\ &\sim \quad \frac{\pi^2}{4N}\end{aligned}$$

for N large. Again Proposition 30 shows only a modest improvement for the spectral gap of $M^{(1)}$: for N large,

$$\begin{aligned}\Lambda_N \quad &\sim \quad \frac{10}{3}\lambda_N\\ &\sim \quad \frac{5\pi^2}{6N}\end{aligned}$$

□

References

[1] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Commun. Contemp. Math.*, 9(4):585–603, 2007.

[2] S. Bacallado and V. Pande. Bayesian analysis of higher-order, reversible, Markov chains. Preprint, Chemistry Dept., Stanford University, 2009.

[3] Persi Diaconis, Susan Holmes, and Radford M. Neal. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3):726–752, 2000.

[4] James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1(1):62–87, 1991.

[5] Kranthi K. Gade and Michael L. Overton. Optimizing the asymptotic convergence rate of the Diaconis-Holmes-Neal sampler. *Adv. in Appl. Math.*, 38(3):382–403, 2007.

[6] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[7] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[8] Anna Lee. Centro-Hermitian and skew-centro-Hermitian matrices. *Linear Algebra Appl.*, 29:205–210, 1980.

[9] Michael Maxwell and Michael Woodroofe. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.*, 28(2):713–724, 2000.

[10] R. M. Neal. Improving asymptotic variance of MCMC estimators: Non-reversible chains are better. Technical Report No. 0406, Dept. of Statistics, University of Toronto, 2004.

[11] P. H. Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60:607–612, 1973.

[12] Irwin S. Pressman. Matrices with multiple symmetry properties: applications of centro-Hermitian and per-Hermitian matrices. *Linear Algebra Appl.*, 284(1-3):239–258, 1998. ILAS Symposium on Fast Algorithms for Control, Signals and Image Processing (Winnipeg, MB, 1997).

[13] Laurent Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin, 1997.

[14] James R. Weaver. Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors. *Amer. Math. Monthly*, 92(10):711–717, 1985.

Stanford University
Department of Mathematics
Building 380, Sloan Hall
Stanford, California 94305, USA

[‡]miclo@math.univ-toulouse.fr

Institut de Mathématiques de Toulouse
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse Cedex 9, France