On the Markovian similarity

Laurent Miclo

Institut de Mathématiques de Toulouse, UMR 5219
Université de Toulouse and CNRS, France

Abstract

Two finite Markov generators L and $\widetilde{L}$ are said to be intertwined if there exists a Markov kernel Λ such that $L\Lambda = \Lambda\widetilde{L}$. The goal of this paper is to investigate the equivalence relation between finite Markov generators obtained by imposing mutual intertwinings through invertible Markov kernels, in particular its links with the traditional similarity relation. Some consequences on the comparison of speeds of convergence to equilibrium for finite irreducible Markov processes are deduced. The situation of infinite state spaces is also quickly mentioned, by showing that the Laplacians of isospectral compact Riemannian manifolds are weakly Markov-similar.

1 Introduction

Intertwining of Markov processes is an old subject, coming back to Rogers and Pitman [17] (or even to the book of Dynkin [8] for the deterministic version), which lately has attracted a renewed interest, see for instance the paper of Pal and Shkolnikov [15] and the references therein. Very recently, Patie and Savov [16] have used intertwinings between reversible Laguerre diffusions and certain non-local and non-reversible Markov processes to get information on the spectral decompositions of the latters. This arises a natural question: when are two Markov processes intertwined? To avoid a trivial answer, we will introduce a notion of Markov-similarity, where a Markovian requirement is imposed on the relation of similitude. Indeed, we will begin by investigating its links with the usual similarity in the framework of general finite Markov processes. Next we will discuss the consequences for the comparison of mixing speeds of Markov-similar finite ergodic Markov processes. Then we will only scratch the surface of the corresponding question in the non-finite setting, in particular by checking that a weak Markov-similarity holds for isospectral Riemannian manifolds.

As announced, we first study the finite state space situation. Let V be a finite set, endowed with a **Markov generator** L: it is a $V \times V$ matrix $(L(x,y))_{x,y \in V}$ whose off-diagonal entries are non-negative and whose row sums vanish:

$$\forall\ x \in V, \qquad \sum_{y \in V} L(x,y) \ = \ 0$$

Consider $\widetilde{L}$ another Markov generator on a finite set $\widetilde{V}$ (more generally, all objects associated to $\widetilde{L}$ will receive a tilde). A **Markov kernel** Λ from V to $\widetilde{V}$ is a $V \times \widetilde{V}$ matrix $(\Lambda(x,\widetilde{x}))_{(x,\widetilde{x}) \in V \times \widetilde{V}}$ whose entries are non-negative and whose row sums are all equal to 1. We say that L is **intertwined** with $\widetilde{L}$, if there exists a Markov kernel Λ from V to $\widetilde{V}$ such that $L\Lambda = \Lambda\widetilde{L}$. If furthermore there exists a Markov kernel $\widetilde{\Lambda}$ from $\widetilde{V}$ to V such that $\widetilde{L}\widetilde{\Lambda} = \widetilde{\Lambda}L$, then L and $\widetilde{L}$ are said to be **mutually intertwined**. This notion is not very interesting, because any finite Markov generators L and $\widetilde{L}$ are always mutually intertwined. Indeed, any finite Markov generator L admits an **invariant probability measure** μ, namely satisfying $\mu[L[f]] = 0$ for all functions f defined on V (where we used the traditional matrix notations: any measure is seen as a row vector and any function as a column vector). Let $\widetilde{\mu}$ be an invariant measure for $\widetilde{L}$ and define two Markov kernels Λ and $\widetilde{\Lambda}$ by

$$\forall\ (x,\widetilde{x}) \in V \times \widetilde{V}, \qquad \Lambda(x,\widetilde{x}) \ :=\ \widetilde{\mu}(\widetilde{x})$$
$$\forall\ (\widetilde{x},x) \in \widetilde{V} \times V, \qquad \widetilde{\Lambda}(\widetilde{x},x) \ :=\ \mu(x)$$

By using these Markov kernels, it is immediate to check that L and $\widetilde{L}$ are mutually intertwined.

So let us add a more stringent requirement. A Markov kernel Λ from V to $\widetilde{V}$ is said to be a **link**, if it is invertible (as a matrix). Here we depart from the terminology introduced by Diaconis and Fill [5], since for them a link is just a Markov kernel. In particular, V and $\widetilde{V}$ have the same cardinality, which will be denoted $|V|$. The Markov generators L and $\widetilde{L}$ are said to be **Markov-similar** if there exist two links Λ and $\widetilde{\Lambda}$, respectively from V to $\widetilde{V}$ and from $\widetilde{V}$ to V such that

$$L\Lambda \ =\ \Lambda\widetilde{L} \quad \text{and} \quad \widetilde{L}\widetilde{\Lambda} \ =\ \widetilde{\Lambda}L \tag{1}$$

The first motivation of this paper stems from the natural question: when are two finite Markov generators Markov-similar?

Of course, two finite Markov-similar Markov generators are linked by a similitude relation, so they are similar in the usual sense, namely they have the same eigenvalues (in $\mathbb{C}$) and the corresponding Jordan blocks have the same dimensions. But despite the results presented in this introduction, the reverse implication is not always true, as we will see in Section 3.

Recall the usual notion of transience for the points of V relatively to L. Let $x, y \in V$, we say that x **leads to** y, if there exists a finite sequence $x = x_0, x_1, x_2, ..., x_l = y$, with $l \in \mathbb{Z}_+$, such that

$L(x_{k-1}, x_k) > 0$ for all $k \in [\![l]\!] := \{1, 2, ..., n\}$. A point $x \in V$ is said to be **transient**, if there exists $y \in S$ such that x leads to y but y does not lead to x. The finite Markov generator L is said to be **non-transient**, if there is no transient point. In particular, if L is **irreducible** (namely, any point $x \in V$ leads to any point $y \in S$), then L is non-transient.

Theorem 1 *Two non-transient Markov generators L and $\widetilde{L}$ are Markov-similar if and only if they are similar.*

It is well-known that the number of irreducible classes (whose definition will be recalled in the beginning of Section 3) of a non-transient Markov generator is an information included into the spectrum of L, since it is the multiplicity of the eigenvalue 0. So according to the above result, two finite Markov-similar non-transient Markov generators have the same number of irreducible classes. Nevertheless the cardinalities of these classes can be different. This may first sound strange (but this is the deep reason behind the aggregation (8) considered in the transient setting, see Section 3) and is illustrated by the example below.

Example 2 Assume that the finite set V is partitioned into $V = \sqcup_{n \in [\![\mathfrak{n}]\!]} C_n$, with $\mathfrak{n} \in \mathbb{N}$. For $n \in [\![\mathfrak{n}]\!]$, let be given μ_n a probability measure whose support is C_n. On each C_n, consider the generator $L_n := \mu_n - I_{C_n}$ where I_{C_n} is the $C_n \times C_n$ identity matrix and where μ_n stands for the matrix whose rows are all equal to μ_n. The spectrum of $-L_n$ consists of the simple eigenvalue 0 and of the eigenvalue 1 with (geometric) multiplicity $|C_n| - 1$. Next, define the generator L on V which acts as L_n on C_n for all $n \in [\![\mathfrak{n}]\!]$, namely $L := \oplus_{n \in [\![\mathfrak{n}]\!]} L_n$. Then the spectrum of $-L$ has the eigenvalue 0 with multiplicity $\mathfrak{n}$ and the eigenvalue 1 with multiplicity $|V| - \mathfrak{n}$. Thus L is diagonalizable and its similarity class is the set of diagonalizable matrices which are isospectral to L. In particular a generator $\widetilde{L}$ defined in a similar fashion will be Markov-similar to L if and only if $|\widetilde{V}| = |V|$ and $\widetilde{\mathfrak{n}} = \mathfrak{n}$. It follows that $\{|C_n| : n \in [\![\mathfrak{n}]\!]\}$ can be different from $\{|\widetilde{C}_n| : n \in [\![\mathfrak{n}]\!]\}$ (as multisets), for instance we can have $\mathfrak{n} = 2$, $|C_1| = 1$, $|C_2| = 3$, $|\widetilde{C}_1| = 2$ and $|\widetilde{C}_2| = 2$.

□

Proposition 18 of Section 3 gives an extension of Theorem 1 to subMarkov generators, which corresponds to Markov processes which can be absorbed.

Remark 3 (a) A more stringent requirement in (1) would impose that the links Λ and $\widetilde{\Lambda}$ are inverse of each other: $\Lambda\widetilde{\Lambda} = I$, the identity kernel, as in the usual similarity relation. But this implies that Λ is a **deterministic kernel**, in the sense there exists a bijection $\sigma : V \to \widetilde{V}$ such that

$$\forall \ (x, \widetilde{x}) \in V \times \widetilde{V}, \qquad \Lambda(x, \widetilde{x}) \ = \ \delta_{\sigma(x)}(\widetilde{x})$$

(see for instance [14]). The link $\widetilde{\Lambda}$ is then the deterministic kernel associated to σ^{-1}. It follows that

$$\forall \ (x, y) \in V^2, \qquad L(x, y) \ = \ \widetilde{L}(\sigma(x), \sigma(y))$$

namely, L can be identified with $\widetilde{L}$, up to a permutation of the state space.

Under this form, it would correspond to a discrete and non symmetric version of the question "can one hear the shape of a drum?" popularized by Kac in [12], where Laplace operators on two-dimensional compact domains with Dirichlet condition on the boundary (assumed to be smooth or polygonal), should be replaced by finite Markovian generators.

(b) Consider links $\Lambda, \widetilde{\Lambda}$ such that (1) is satisfied with respect to some Markov generators $L, \widetilde{L}$. Then $\Lambda\widetilde{\Lambda}$ is an invertible Markov kernel from V to V which commutes with L:

$$L\Lambda\widetilde{\Lambda} \ = \ \Lambda\widetilde{L}\widetilde{\Lambda} \ = \ \Lambda\widetilde{\Lambda}L$$

and symmetrically, $\widetilde{\Lambda}\Lambda$ is an invertible Markov kernel from $\widetilde{V}$ to $\widetilde{V}$ commuting with $\widetilde{L}$. In [14], the convex set of Markov kernels commuting with a given Markov generator was studied, in particular in correlation with the notion of weak hypergroup, on which we will come back at the end of Section 4.

$\square$

One of the interest of Markov-similarity of two generators is that it should enable the comparison between their speeds of convergence to equilibrium or to absorption. Assume that L is a finite irreducible Markov generator and let μ be its unique invariant probability. If m_0 is a given initial probability on V, define for any $t > 0$, $m_t := m_0 \exp(tL)$, the distribution at time t of the Markov process starting from m_0 and whose evolution is dictated by L. As t goes to infinity, m_t converges toward μ and there are several ways to measure the discrepancy between m_t and μ. Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a convex function such that $\psi(1) = 0$. The set of such functions will be denoted Ψ. For $\psi \in \Psi$, the ψ-entropy of a probability measure m with respect to μ is given by

$$E_\psi[m|\mu] \ := \ \sum_{x \in V} \psi\left(\frac{m(x)}{\mu(x)}\right) \mu(x)$$

Consider the worst cases over the initial conditions, namely

$$\forall \ \psi \in \Psi, \ \forall \ t \geqslant 0, \qquad E(\psi, t) \ := \ \sup\{E_\psi[m_0 \exp(tL)|\mu] \ : \ m_0 \in \mathcal{P}(V)\}$$

where $\mathcal{P}(V)$ stands for the set of all probability measures on V. Then we have:

Proposition 4 *Let L and $\widetilde{L}$ be two Markov-similar generators. Then there exists a constant $T \geqslant 0$ such that*

$$\forall \ \psi \in \Psi, \ \forall \ t \geqslant 0, \qquad E(\psi, T + t) \ \leqslant \ \widetilde{E}(\psi, t) \quad and \quad \widetilde{E}(\psi, T + t) \ \leqslant \ E(\psi, t)$$

where $\widetilde{E}(\psi, t)$ is defined as $E(\psi, t)$, but with L replaced by $\widetilde{L}$ and μ by $\widetilde{\mu}$.

So in some sense, after the warming up time T, the convergences to equilibrium are similar for the Markov processes generated by L and $\widetilde{L}$. More precise results in this direction will be given in Section 4, in particular for some initial distributions no warming up period is necessary, but the crucial quantitative estimation of T will remain to be investigated.

To extend the previous considerations to infinite state spaces, one must begin by choosing an appropriate notion of "non degeneracy" of the links. Recall that in general, a **Markov kernel** Λ from a measurable space $(V, \mathcal{V})$ to a measurable space $(\widetilde{V}, \widetilde{\mathcal{V}})$, is a mapping from $V \times \widetilde{\mathcal{V}}$ such that:
• for any $x \in V$, $\Lambda(x, \cdot)$ is a probability measure on $(\widetilde{V}, \widetilde{\mathcal{V}})$,
• for any $A \in \widetilde{\mathcal{V}}$, $\Lambda(\cdot, A)$ is a $(V, \mathcal{V})$-measurable mapping.
When V and $\widetilde{V}$ are finite (it is then understood that they are endowed with their full σ-algebras), one recovers the above definition, namely Λ can be identified with a $V \times \widetilde{V}$ matrix whose entries are non-negative and whose row sums are all equal to 1.

Let $\mathcal{B}$ (respectively $\widetilde{\mathcal{B}}$) the vector space of bounded measurable functions on $(V, \mathcal{V})$ (resp. $(\widetilde{V}, \widetilde{\mathcal{V}})$). A Markov kernel Λ from $(V, \mathcal{V})$, to $(\widetilde{V}, \widetilde{\mathcal{V}})$ induces an operator from $\widetilde{\mathcal{B}}$ from $\mathcal{B}$ via

$$\forall \ x \in V, \ \forall \ \widetilde{f} \in \widetilde{\mathcal{B}}, \qquad \Lambda[\widetilde{f}](x) \ := \ \int \widetilde{f}(\widetilde{x}) \, \Lambda(x, d\widetilde{x}) \tag{2}$$

The Markov kernel is then said to be a **weak link** if it is a one-to-one operator.

Let L be a **Markov generator** on $(V, \mathcal{V})$, in the sense that it is defined on a subspace $\mathcal{D}(L)$ of $\mathcal{B}$ such that the corresponding martingale problems are well-posed for any initial condition (for a thorough exposition of these concepts, see e.g. the book of Ethier and Kurtz [9]). If V is finite, it corresponds to the definition given in the beginning of this introduction. The Markovian generators

L and $\widetilde{L}$ are said to be **weakly Markov-similar**, if there exist two links, Λ from $(V, \mathcal{V})$ to $(\widetilde{V}, \widetilde{\mathcal{V}})$ and $\widetilde{\Lambda}$ from $(\widetilde{V}, \widetilde{\mathcal{V}})$ to $(V, \mathcal{V})$, such that (1) is satisfied, in particular, these relations require that $\Lambda(\mathcal{D}(\widetilde{L})) \subset \mathcal{D}(L)$ and $\widetilde{\Lambda}(\mathcal{D}(L)) \subset \mathcal{D}(\widetilde{L})$.

In these general definitions, we did not mention invariant probabilities, since when V is infinite, they may not exist. Nevertheless, if we are given a probability μ invariant for L, in the sense that

$$\forall\ f \in \mathcal{D}(L), \qquad \mu[L[f]]\ =\ 0$$

then the above notions can be slightly modified to be given a $\mathbb{L}^2$ flavor: the Markov operator defined in (2) can be extended into an operator from $\mathbb{L}^2(\mu)$ to $\mathbb{L}^2(\widetilde{\mu})$, with $\widetilde{\mu} := \mu\Lambda$ and $(\mathcal{D}(L), L)$ can be replaced by its $\mathbb{L}^2(\mu)$-closure. The operator Λ is then Markovian, in the abstract sense that it preserves non-negativity and the functions taking only the value 1 (respectively μ- and $\widetilde{\mu}$-a.s.). Conversely, if the measurable spaces $(V, \mathcal{V})$ and $(\widetilde{V}, \widetilde{\mathcal{V}})$ are the Borelian spaces associated to Polish topological spaces, then any abstract Markovian operator from $\mathbb{L}^2(\widetilde{\mu})$ to $\mathbb{L}^2(\mu)$ corresponds to a Markov kernel. We are thus led naturally to the notions of weak (abstract) $\mathbb{L}^2(\mu)$-link and of weak (abstract) $\mathbb{L}^2$-Markov-similarity between L and $\widetilde{L}$, when $\widetilde{\mu}$ is left invariant by $\widetilde{L}$. Despite the fact that this subject would deserve a general investigation, here we restrict our attention to a very particular situation. We say that the Markov generator L with invariant probability μ is **nice**, if:

- The measurable space $(V, \mathcal{V})$ is the Borelian space associated to a Polish topological space.
- The operator L admits a unique invariant probability μ, which is in fact reversible, in the sense that

$$\forall\ f, g \in \mathcal{D}(L), \qquad \mu[fL[g]]\ =\ \mu[gL[f]] \tag{3}$$

This assumption enables to consider the (Friedrich) minimal extension of L as a self-adjoint operator on $\mathbb{L}^2(\mu)$, with $\mathcal{D}(L)$ as new domain.

- The spectral decomposition of $-L$ only consists of eigenvalues, say $(\lambda_l)_{l \in \mathbb{Z}_+}$, with multiplicities.
- It is possible to choose a family $(\varphi_l)_{l \in \mathbb{Z}_+}$ of eigenvectors associated to the eigenvalues $(\lambda_l)_{l \in \mathbb{Z}_+}$, such that for any $l \in \mathbb{Z}_+$, the function φ_l is bounded (this is always true for the eigenvalue 0, since by the preceding point, its eigenspace is the set of the constant functions).

The interest of this notion is:

Proposition 5 *Two nice Markov generators L and $\widetilde{L}$ are weakly $\mathbb{L}^2$-Markov-similar if and only if they are isospectral.*

A typical example of a nice Markov generator is that of a reversible elliptic diffusion with regular coefficients on a compact manifold V. In this situation, one can endow V with a Riemannian structure and find a smooth function U, such that the underlying Markov generator L has the following form (known as Witten Laplacian)

$$L\cdot\ =\ \triangle\cdot - \langle \nabla U, \nabla \cdot \rangle$$

where $\triangle$ is the Laplace-Beltrami operator, $\langle \cdot, \cdot \rangle$ is the scalar product and ∇ is the gradient operator (see e.g. the book of Ikeda and Watanabe [11]). The corresponding reversible probability μ admits as density with respect to the Riemannian measure the one proportional to $\exp(-U)$. The compactness of V implies that the spectrum of $-L$ consists only of non-negative eigenvalues with finite multiplicities and without accumulation point. Denote them by

$$0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \ldots$$

As solutions to elliptic equations, the corresponding eigenvectors are smooth and thus bounded.

Let $\widetilde{L}$ be another diffusion generator of the same kind (i.e. associated to a compact Riemannian manifold $\widetilde{V}$ and to a potential $\widetilde{U}$). Let $0 = \widetilde{\lambda}_0 < \widetilde{\lambda}_1 \leqslant \widetilde{\lambda}_2 \leqslant \ldots$ be its eigenvalues. As a consequence of Proposition 5, L and $\widetilde{L}$ are $\mathbb{L}^2$-Markov-similar if and only if we have $\lambda_n = \widetilde{\lambda}_n$ for all $n \in \mathbb{Z}_+$.

In particular, the Laplace-Beltrami operators corresponding to isospectral compact Riemannian manifolds are $\mathbb{L}^2$-Markov-similar. This result enables the coupling by intertwining (generalizing the coupling constructed by Diaconis and Fill [5] for discrete time finite Markov chains) of the Brownian motions on such manifolds, suggesting that the question of isospectrality for compact Riemannian manifolds (see e.g. the review of Bérard [1] and references therein) could be revisited from a probabilistic point of view. The study of the links between the mixing speed of such Brownian motions, as in Proposition 5, is out of the scope of this paper.

The paper is organized as follows: the next section contains the proof of Theorem 1. Section 3 investigates the transient situation, where the characterization of Markov-similarity is not complete. The subMarkovian case will also be dealt with there. Section 4 collects the considerations about mixing speeds. The proof of Proposition 5 is given in the final section.

2 The finite non-transient setting

This section is devoted to the proof of Theorem 1. It will be shown gradually, starting with the case of irreducible and reversible generators and ending with the general non-transient case.

Recall that a finite Markov generator L is said to be **reversible** with respect to a probability measure μ on V, if

$$\forall\ x, y \in V, \qquad \mu(x)L(x,y) \quad = \quad \mu(y)L(y,x)$$

This property is equivalent to the symmetry of L in $\mathbb{L}^2(\mu)$ mentioned in (3) for the general case. We begin by assuming that the generator L is irreducible and reversible. By irreducibility, the invariant measure μ is unique and positive on V. The reversibility of L with respect to μ implies that L is diagonalizable. Denote the eigenvalues (with multiplicities) of $-L$ by

$$0 = \lambda_1 < \lambda_2 \leqslant \lambda_3 \leqslant \cdots \leqslant \lambda_{|V|} \tag{4}$$

(the strict inequality comes from irreducibility).

Consider another irreducible and reversible Markov generator $\widetilde{L}$. In this case, the similarity of L and $\widetilde{L}$ reduces the fact that L and $\widetilde{L}$ are isospectral, i.e. V and $\widetilde{V}$ have the same cardinality and

$$\forall\ k \in [\![|V|]\!], \qquad \widetilde{\lambda}_k \quad = \quad \lambda_k$$

Here is the first step in the direction of Theorem 1, it corresponds to Proposition 5 in the finite case.

Lemma 6 *Two finite, irreducible and reversible Markov generators are Markov-similar if and only if they are similar.*

Proof

Let L and $\widetilde{L}$ be finite, irreducible and reversible Markov generators. If they are Markov-similar, there is a similarity relation between them, for instance $\widetilde{L} = \Lambda^{-1} L \Lambda$, so they are similar.

Conversely, assume that L and $\widetilde{L}$ are similar. Denote by (4) the common spectrum of $-L$ and $-\widetilde{L}$. Let $(\varphi_k)_{k \in [\![|V|]\!]}$ and $(\widetilde{\varphi}_k)_{k \in [\![|V|]\!]}$ be orthonormal bases of $\mathbb{L}^2(\mu)$ and $\mathbb{L}^2(\widetilde{\mu})$ consisting of corresponding eigenvectors. Without loss of generality, we can assume that $\widetilde{V} = V$ and that $\varphi_1 = \widetilde{\varphi}_1 = \mathbb{1}$ (the function always taking the value 1). To construct an invertible Markov kernel Λ from V to V such that $L\Lambda = \Lambda \widetilde{L}$, consider the operator A defined by

$$\forall\ k \in [\![|V|]\!], \qquad A[\widetilde{\varphi}_k] \ := \ \begin{cases} \varphi_k & ,\text{ if } k \geqslant 2 \\ 0 & ,\text{ if } k = 1 \end{cases}$$

For $\epsilon \in \mathbb{R}$, we are interested in the operator

$$\Lambda \ := \ \widetilde{\mu} + \epsilon A \tag{5}$$

where $\widetilde{\mu}$ is again interpreted as the matrix whose rows are all equal to the probability $\widetilde{\mu}$. It is immediately checked that

$$\forall \ k \in [\![|V|]\!], \qquad \Lambda[\widetilde{\varphi}_k] \ := \ \begin{cases} \epsilon \varphi_k & \text{, if } k \geqslant 2 \\ \varphi_1 & \text{, if } k = 1 \end{cases}$$

since by orthogonality, $\widetilde{\mu}[\widetilde{\varphi}_k] = \widetilde{\mu}[\widetilde{\varphi}_1 \widetilde{\varphi}_k] = 0$. It implies the relation $L\Lambda = \Lambda \widetilde{L}$ and that Λ is invertible as soon as $\epsilon \neq 0$.

From the relation $\Lambda[\mathbb{1}] = \mathbb{1}$, it appears that the row sums of Λ are all equal to 1. Thus it remains to find $\epsilon \neq 0$ such that all the entries of Λ are non-negative. It is sufficient to take

$$0 \ < \ |\epsilon| \ \leqslant \ \min_{x,y \in V} \frac{|A(x,y)|}{\widetilde{\mu}(y)} \tag{6}$$

By exchanging the roles of L and $\widetilde{L}$, one constructs an invertible Markov kernel $\widetilde{\Lambda}$ such that $\widetilde{L}\widetilde{\Lambda} = \widetilde{\Lambda}L$ and this ends the proof of the lemma.

∎

To extend the above lemma to all finite irreducible Markov generators, we need to recall more precisely the characteristic invariants for the similarity relation and to introduce the corresponding notation. Let R be a $N \times N$ real finite matrix. Seen as a complex matrix, it is similar to a block matrix, whose blocks are of Jordan type $(\lambda_1, n_1), (\lambda_2, n_2), ..., (\lambda_r, n_r)$, where $\lambda_1, \lambda_2, ..., \lambda_r \in \mathbb{C}$ are the eigenvalues of R (with geometric multiplicities) and $r \in \mathbb{N}$, $n_1, n_2, ..., n_r \in \mathbb{N}$ with $n_1 + n_2 + \cdots + n_r = N$. Recall that a Jordan block of type (λ, n) is a $n \times n$ matrix whose diagonal entries are equal to λ, whose first above diagonal entries are equal to 1 and whose other entries vanish. The set $\{(\lambda_k, n_k) \ : \ k \in [\![r]\!]\}$ is a characteristic invariant for the similarity class of R and will be called the **characteristic set** of R. Note that this characteristic set of R is equal to $\{(\lambda_k, n_k) \ : \ k \in [\![r]\!]\}$, if and only if one can find a (complex) basis $(\varphi_{(k,l)})_{(k,l) \in S}$ of $\mathbb{C}^N$, where $S := \{(k,l) \ : \ k \in [\![r]\!]$ and $l \in [\![n_k]\!]\}$, such that

$$\forall \ (k,l) \in S, \qquad R[\varphi_{(k,l)}] \ = \ \lambda_k \varphi_{(k,l)} + \varphi_{(k,l-1)}$$

where by convention, $\varphi_{(k,0)} = 0$ for all $k \in [\![r]\!]$. Such a basis will be said to be **adapted** to R.

Lemma 6 extends into:

Lemma 7 *Two finite and irreducible Markov generators are Markov-similar if and only if they are similar.*

Proof

It is sufficient to adapt the arguments given in the reversible situation. Again we just need to show the direct implication. Let L and $\widetilde{L}$ be two finite and irreducible Markov generators which are similar. Up to a permutation, we identify the index set $\widetilde{S}$ with S in the above notation (with $R = -\widetilde{L}$ or $R = L$). Let $(\varphi_{(k,l)})_{(k,l) \in S}$ and $(\widetilde{\varphi}_{(k,l)})_{(k,l) \in S}$ be adapted bases associated to $-L$ and $-\widetilde{L}$. By irreducibility, 0 is an eigenvalue of multiplicity 1, so we can assume that $(\lambda_1, n_1) = (0,1)$ and $\varphi_{(1,1)} = \mathbb{1}$. We begin by proving that

$$\forall \ (k,l) \in S \backslash \{(0,1)\}, \qquad \mu[\varphi_{(k,l)}] = 0$$

Indeed, for any $k \in [\![r]\!]$, we have $L[\varphi_{(k,1)}] = -\lambda_k \varphi_{(k,1)}$ with $\lambda_k \neq 0$. Integrating the previous relation with respect to μ, we obtain

$$\lambda_k \mu[\varphi_{(k,1)}] \ = \ 0$$

so that $\mu[\varphi_{(k,1)}] = 0$. Next we show that

$$\mu[\varphi_{(k,l)}] \quad = \quad 0 \tag{7}$$

by iteration on l, with $k \in [\![r]\!]$ fixed. If (7) is true for some $l \in [\![n_k - 1]\!]$, then integrating with respect to μ the relation

$$L[\varphi_{(k,l+1)}] = -\lambda_k \varphi_{(k,l+1)} + \varphi_{(k,l)}$$

we get $\lambda_k \mu[\varphi_{k,l+1}] = 0$, namely (7) with l replaced by $l + 1$.

Let $\mathcal{F}$ be the vector space generated by the family $(\varphi_{(k,l)})_{(k,l) \in S \setminus \{(0,1)\}}$, i.e. the vector space of functions f defined on V such that $\mu[f] = 0$. Define similarly $\widetilde{\mathcal{F}}$ and an operator B from $\widetilde{\mathcal{F}}$ to $\mathcal{F}$ by

$$\forall \ (k,l) \in S \setminus \{(0,1)\}, \qquad B[\widetilde{\varphi}_{(k,l)}] \quad := \quad \varphi_{(k,l)}$$

Consider two bases of $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ made up of real functions, a priori the entries of the matrix (still denoted B) associated to B in these bases are complex numbers. But we have that B is invertible and that on $\widetilde{\mathcal{F}}$, $LB = B\widetilde{L}$. Since the entries of L and $\widetilde{L}$ are real numbers, it follows that $L\Re(B) = \Re(B)\widetilde{L}$ and $L\Im(B) = \Im(B)\widetilde{L}$, where $\Re$ and $\Im$ stands for the real and imaginary parts. Furthermore, there exists a real number s such that the rank of $A := \Re(B) + s\Im(B)$ is $|V| - 1$ (use e.g. the polynomial mapping $\mathbb{C} \ni z \mapsto \det(\Re(B) + z\Im(B))$). Extend A into an operator from $\mathbb{L}^2(\widetilde{\mu})$ to $\mathbb{L}^2(\mu)$ by imposing $A[\mathbb{1}] = 0$ and note that $LA = A\widetilde{L}$ and that in the the usual basis $(\mathbb{1}_x)_{x \in V}$ formed of the indicator functions of the points, the entries of A are real numbers. For $\epsilon \neq 0$, we consider again the operator Λ given by (5). The proof goes on as before, Λ being an invertible Markov kernel if (6) is satisfied.

■

It remains to relax the irreducibility assumption to prove Theorem 1. Recall that a finite Markov generator is non-transient, if and only if it admits a invariant probability measure which gives positive weights to all the points of V. This state space can then be partitioned into parts which ignore themselves two by two and are of the type considered in Lemma 7. Nevertheless, Example 2 suggests that Theorem 1 cannot be a direct consequence of Lemma 7.

Proof of Theorem 1

The difference with the proof of Lemma 7 is that the eigenvalue 0 of a finite and non-transient Markov generator is no longer necessarily simple. Its multiplicity is the number $\mathfrak{n} \in \mathbb{N}$ of irreducible classes and the dimension of the Jordan blocks associated to each of the eigenvalue(s) 0 is 1. The arguments of the proof of Lemma 7 can be adapted by doing the following. If L and $\widetilde{L}$ are two finite and non-transient Markov generators, begin by choosing corresponding positive invariant probabilities measures μ and $\widetilde{\mu}$. Next choose an orthonormal (in $\mathbb{L}^2(\mu)$) basis $(\varphi_1, \varphi_2, ..., \varphi_{\mathfrak{n}})$ of the kernel of L, with $\varphi_1 = \mathbb{1}$ and similarly an orthonormal (in $\mathbb{L}^2(\widetilde{\mu})$) basis $(\widetilde{\varphi}_1, \widetilde{\varphi}_2, ..., \widetilde{\varphi}_{\mathfrak{n}})$ of the kernel of $\widetilde{L}$ with $\widetilde{\varphi}_1 = \mathbb{1}$. Complete these families of vectors into adapted bases for L and $\widetilde{L}$, with the convention that the index associated to the eigenvectors φ_1 and $\widetilde{\varphi}_1$ is $(1, 1)$. The argument goes on as before, since both φ_1 and $\widetilde{\varphi}_1$ are orthogonal to all the other eigenvectors in their respective bases.

■

3   On the finite transient setting

As alluded to in the introduction, in general similarity does not imply Markov-similarity. Remaining in the finite state space framework, we investigate here in more detail the transient situation

where this phenomenon appears, by obtaining a necessary condition, of spatial-spectral nature, for Markov-similarity. On a example, we will check that this condition is not sufficient. Thus the problem of finding characterizing invariants for Markov-similarity is still open and seems quite challenging.

We begin by recalling a traditional classification of the points of V according to L, refining the notion of transience defined in the introduction. Consider $x, y \in V$, if x leads to y and y to x, we say that x and y **communicate**. This defines an equivalence (**irreducibility**) relation.

Let $C_1, C_2, ..., C_{\mathfrak{n}}$, with $\mathfrak{n} \in \mathbb{N}$, be the associated equivalence classes. For $k, l \in [\![\mathfrak{n}]\!]$, we write $C_k \preceq_0 C_l$ if there exist $x \in C_l$ and $y \in C_k$ such that x leads to y. The relation $\preceq_0$ defines a partial order on $\mathcal{C}_0 := \{C_1, ..., C_{\mathfrak{n}}\}$. Consider

$$
\begin{aligned}
A_1 &:= \{a \in [\![\mathfrak{n}]\!] : C_a \text{ is minimal for } \preceq_0\} \\
B_1 &:= [\![\mathfrak{n}]\!] \backslash A_1 \\
V_1 &:= \bigsqcup_{a \in A_1} C_a
\end{aligned}
$$

From a probabilistic point of view, V_1 is the maximal subset of V supporting an invariant probability for L and $V \backslash V_1$ is the set of transient points. If $V_1 = V$, the construction stops here and L is non-transient. Otherwise, consider $\mathcal{C}_1 := \{C_a : a \in B_1\}$ and denote by $\preceq_1$ the restriction of the partial order $\preceq_0$ to $\mathcal{C}_1$. Define

$$
\begin{aligned}
A_2 &:= \{a \in B_1 : C_a \text{ is minimal for } \preceq_1\} \\
B_2 &:= B_1 \backslash A_2
\end{aligned}
$$

For each $a \in A_2$, consider L_{C_a}, the $C_a \times C_a$ matrix extracted from L (also named the Dirichlet restriction of L to C_a), it is a **subMarkovian generator**, in the sense that the off-diagonal entries are non-negative and the row sums are non-positive. The Perron-Frobenius' theorem can be applied to show that $-L_{C_a}$ admits a smallest eigenvalue (in modulus), $\lambda_1(C_a) \geqslant 0$, called the **first Dirichlet eigenvalue** of L_{C_a}. Order the elements of the set $\{\lambda_1(C_a) : a \in A_2\}$ into $\lambda_{2,1} < \lambda_{2,2} < \cdots < \lambda_{2,\kappa_2}$, with $\kappa_2 \in [\![|A_2|]\!]$. For $l \in [\![2, 1 + \kappa_2]\!] := \{2, 3, ..., 1 + \kappa_2\}$, we denote

$$
V_l := \bigcup_{a \in A_2 \,:\, \lambda_1(C_a) = \lambda_{2,l-1}} C_a \tag{8}
$$

so that $V_2 \sqcup V_3 \sqcup \cdots \sqcup V_{\kappa_2}$ forms a partition of $\bigsqcup_{a \in A_2} C_a$. It was not necessary to explicit such a partition in our first step (defining V_1), because for any $a \in A_1$, one has that L_{C_a} is still a Markov generator, so that $\lambda_1(C_a) = 0$ and we would have end up with the unique set V_1 (i.e. $\kappa_1 = 1$). The procedure goes on by iteration, in the second step (if B_2 is not empty), we construct some disjoint non-empty subsets $V_{\kappa_2+1}, V_{\kappa_2+2}, ..., V_{\kappa_2+\kappa_3}$, for some $\kappa_3 \in [\![|A_3|]\!]$, where A_3 is the set of minimal elements of B_2 for the restriction of $\preceq_1$, and so on. At the end of the construction, we have disjoint non-empty subsets $V_1, V_2, ..., V_{\mathfrak{m}}$, with $\mathfrak{m} \in [\![\mathfrak{n}]\!]$, such that $V = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_{\mathfrak{m}}$, as well as a finite sequence of positive integers $\kappa := (\kappa_1, \kappa_2, ..., \kappa_\eta)$, where η is the number of iterations of the previous procedure, in particular $\kappa_1 + \kappa_2 + \cdots + \kappa_\eta = \mathfrak{m}$. For $m \in [\![\mathfrak{m}]\!]$, let L_m be the $V_m \times V_m$ submatrix of L, it is a subMarkovian generator.

Consider another generator $\widetilde{L}$ and construct as above the subMarkovian generators $\widetilde{L}_1, \widetilde{L}_2, ..., \widetilde{L}_{\widetilde{\mathfrak{m}}}$, as well as the finite sequence $\widetilde{\kappa} := (\widetilde{\kappa}_k)_{k \in [\![\widetilde{\eta}]\!]}$. We say that L and $\widetilde{L}$ satisfy **Condition (C)** if $\widetilde{\kappa} = \kappa$ (in particular $\widetilde{\eta} = \eta$ and $\widetilde{\mathfrak{m}} = \mathfrak{m}$) and if for any $l \in [\![\mathfrak{m}]\!]$, L_l and $\widetilde{L}_l$ are similar. Example 19 at the end of this section shows that Condition (C) does not imply similarity in general. So let us call **Hypothesis (H)** the conjunction of (C) with similarity.

Proposition 8 *If two Markov generators L and $\widetilde{L}$ are Markov-similar then they satisfy (H).*

The following very simple example (which nevertheless played an important role in the study of certain Markov intertwinings in [6] and [13]) illustrates the above construction and the difference between Hypothesis (H) and the similarity relation.

Example 9 Consider on $V := [\![|V|]\!]$ a finite Markov generator $L := (L(x,y))_{x,y \in V}$ which is lower diagonal, whose diagonal entries are all different and such that the first lower diagonal is positive (namely, for all $k \in [\![2, |V|]\!]$, $L(k, k-1) > 0$). Then in the above decomposition, we have $\mathfrak{n} = \mathfrak{m} = |V|$ and for all $k \in [\![|V|]\!]$, $V_k = \{k\}$ and L_k is reduced to the real number $L(k,k)$, which is also its unique eigenvalue (note furthermore that necessarily, $L_1 = 0$). We have $\eta = |V|$ and $\kappa_k = 1$ for all $k \in [\![|V|]\!]$. Consider another Markov generator $\widetilde{L}$. The generators L and $\widetilde{L}$ satisfy Condition (C) if and only if $\widetilde{L}$ is of the same type (up to a permutation of the state space) and if for any $k \in [\![|V|]\!]$, $\widetilde{L}(k,k) = L(k,k)$. As a consequence L and $\widetilde{L}$ are similar, since they are diagonalizable, the $L(k,k)$, for $k \in [\![|V|]\!]$, being their distinct eigenvalues. Nevertheless, the mere similarity of L and $\widetilde{L}$ is a much weaker requirement, it does not imply that $\widetilde{L}$ is of the same type, and even if it is, it only asks for the equality of the spectra, i.e. of the sets $\{L(k,k) : k \in [\![|V|]\!]\} = \{\widetilde{L}(k,k) : k \in [\![|V|]\!]\}$.

Another example in the same spirit is obtained by considering a finite Markov generator L satisfying

$$\forall\, x \in [\![|V|]\!], \qquad L(x, x-1) \;\; = \;\; -L(x,x) \;\; > \;\; 0$$

(with $L(1,0) = 0 = L(1,1)$), so that all the entries outside the main and first lower diagonals vanish. Any eigenvalue λ of $-L$ is geometrically simple, because a corresponding eigenvector φ is completely determined by λ and by the value $\varphi(|V|)$ (by iteration on $k \in [\![|V|]\!]$, one computes $\varphi(|V| - k + 1)$ via the relation $L(|V| - k + 1, |V| - k)(\varphi(|V| - k) - \varphi(|V| - k + 1)) = -\lambda\varphi(|V| - k + 1))$. The dimension of the corresponding Jordan block is the cardinal of the set $\{x \in [\![|V|]\!] : L(x,x) = \lambda\}$. As above, another Markov generator $\widetilde{L}$ and L satisfy Condition (C) if and only if $\widetilde{L}$ is of the same type (up to a permutation of the state space) and if for any $k \in [\![|V|]\!]$, $\widetilde{L}(k,k) = L(k,k)$. Again (C) implies (H), due to the previous observation on the dimension of the Jordan blocks. The same remark about the mere similarity is equally valid, except that the last equality $\{L(k,k) : k \in [\![|V|]\!]\} = \{\widetilde{L}(k,k) : k \in [\![|V|]\!]\}$ must be understood in the sense of multi-sets.

$\square$

In the definition of Condition (C), it is important not to forget the equality of the finite sequences $\kappa = \widetilde{\kappa}$, as shown by

Example 10 On $V := \{1, 2, 3\}$, consider the two generators

$$L \;:=\; \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 2 & -2 \end{pmatrix} \quad \text{and} \quad \widetilde{L} \;:=\; \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & 0 & -2 \end{pmatrix}$$

With the notation of the above decomposition, we have, for all $k \in [\![3]\!]$, $V_k = \{k\} = \widetilde{V}_k$ and $L_k = (k-1) = \widetilde{L}_k$. So the fact that L and $\widetilde{L}$ are not satisfying Condition (C) comes from $\kappa = (1, 1) \neq (2) = \widetilde{\kappa}$. This also provides a very simple example of Markov generators which are similar but not Markov-similar.

$\square$

Here is a simple consequence of Proposition 8:

Corollary 11 *Let the two Markov generators L and $\widetilde{L}$ be Markov-similar. If L is non-transient, then the same is true for $\widetilde{L}$.*

Proof

10

Indeed, the non transience of a Markov generator L is equivalent to the fact that $\kappa = (\kappa_1) = (1)$.
∎

As an extension of the observation made after the statement of Theorem 1, note that two finite Markov-similar Markov generators have the same number of irreducible classes. Indeed, for a general Markov generator L, this number is the sum of the multiplicities of the first Dirichlet eigenvalues of the subMarkovian generators L_1, L_2, ..., $L_{\mathfrak{m}}$, with the notation of the above decomposition (which will be enforced for the remaining part of this section).

The proof of Proposition 8 asks for several steps. We start with

Lemma 12 *Let two Markov generators L, $\widetilde{L}$ and a link Λ be such that $L\Lambda = \Lambda\widetilde{L}$. Then we have $|V_1| \geqslant |\widetilde{V}_1|$ and $\Lambda_{V_1 \times \widetilde{V}_1}$ (the submatrix of Λ indexed by $V_1 \times \widetilde{V}_1$) is a Markov kernel.*

Proof

Consider μ an invariant probability for L whose support is V_1 (constructed as a mixture with positive weights of the invariant probabilities associated to the irreducibility classes forming V_1). The intertwining relation implies that $\mu\Lambda\widetilde{L} = 0$, namely $\widetilde{\mu} := \mu\Lambda$ is an invariant probability for $\widetilde{L}$. The support of $\widetilde{\mu}$ is included into $\widetilde{V}_1$, since it is the largest subset of $\widetilde{V}$ supporting an invariant probability for $\widetilde{L}$. The equality

$$\forall\, \widetilde{x} \notin \widetilde{V}_1, \qquad \sum_{x\in V} \mu(x)\Lambda(x,\widetilde{x}) \;=\; \widetilde{\mu}(\widetilde{x}) \;=\; 0$$

implies that $\Lambda_{V_1 \times (\widetilde{V}\setminus\widetilde{V}_1)} = 0$, namely $\Lambda_{V_1 \times \widetilde{V}_1}$ is a Markov kernel. Another consequence of the fact that $\Lambda_{V_1 \times (\widetilde{V}\setminus\widetilde{V}_1)}$ vanishes is that $|V_1| \geqslant |\widetilde{V}_1|$, otherwise Λ could not be invertible.
∎

In particular, we get

Corollary 13 *Assume that the two Markov generators L and $\widetilde{L}$ are Markov-similar. Then the two Markov generators L_1 and $\widetilde{L}_1$ are Markov-similar.*

Proof

Applying the previous lemma to the two intertwining relations (1), we get that $|V_1| = |\widetilde{V}_1|$, namely $\Lambda_{V_1,\widetilde{V}_1}$ can be seen as a square matrix. Denote $W_1 := V\setminus V_1$ and $\widetilde{W}_1 := V\setminus\widetilde{V}_1$. Since $L_{V_1,W_1} = 0$, $\widetilde{L}_{\widetilde{V}_1,\widetilde{W}_1} = 0$ and $\Lambda_{V_1,\widetilde{W}_1} = 0$, we deduce from the intertwining $L\Lambda = \Lambda\widetilde{L}$ that $L_1\Lambda_{V_1,\widetilde{V}_1} = L_{V_1,V_1}\Lambda_{V_1,\widetilde{V}_1} = \Lambda_{V_1,\widetilde{V}_1}\widetilde{L}_{\widetilde{V}_1,\widetilde{V}_1} = \Lambda_{V_1,\widetilde{V}_1}\widetilde{L}_1$. Furthermore $\Lambda_{V_1,\widetilde{V}_1}$ must be invertible, if we want Λ to be invertible. Applying the same considerations to the intertwining $\widetilde{L}\widetilde{\Lambda} = \widetilde{\Lambda}L$, it follows that L_1 and $\widetilde{L}_1$ are Markov-similar.
∎

To extend by iteration the above result to all the subMarkov generators L_l and $\widetilde{L}_l$, for $l \in [\![\mathfrak{m}]\!]$, we must adapt the arguments to the subMarkovian setting. First note that the decomposition of the state space into the partition $V = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_{\mathfrak{m}}$ can be applied verbatim to a subMarkovian generator L (with the difference that the first step can already produce several subset V_1, V_2, ..., V_{κ_1}, with $\kappa_1 \in \mathbb{N}$). The probabilist interpretation of V_1 has to be slightly modified, with respect to the strict Markovian case:

Lemma 14 *Consider L a subMarkovian generator and let $\mathcal{L}$ be the set of real numbers λ such that there exists a probability measure μ with $\mu L = -\lambda\mu$ (then λ is necessarily non-negative). We have $\mathcal{L} = \{\lambda_1(C_a) : a \in A_1\}$. Denote $\lambda_1 < \lambda_2 < \cdots < \lambda_{\kappa_1}$ the elements of $\mathcal{L}$. For any $k \in [\![\kappa_1]\!]$, V_k is the largest subset of V supporting a probability measure satisfying $\mu L = -\lambda_k\mu$.*

11

Proof

Since the classes $C_1, C_2, ..., C_{\mathfrak{n}}$ are irreducible, we can apply to each of them the Perron-Frobenius' theorem, to get for $a \in [\![m]\!]$, a probability measure μ_a (called the quasi-stationary measure associated to L_{C_a}) whose support is C_a and which is such that $\mu_a L_{C_a} = -\lambda_1(C_a)\mu_a$, where $\lambda_1(C_a) \geqslant 0$ is the first Dirichlet eigenvalue of L_{C_a}. The particularity of the set of indices A_1 is that for each $a \in A_1$ and for any probability measure ν whose support is included in C_a, we have $\nu L = \nu L_{C_a}$ (with the identification of a measure whose support is included into C_a with a measure on C_a). It follows that for $a \in A_1$, we have $\mu_a L = -\lambda_1(C_a)\mu_a$, so that

$$\{\lambda_1(C_a) \,:\, a \in A_1\} \;\subset\; \mathcal{L}$$

Conversely, consider $\lambda \in \mathcal{L}$ and a probability measure μ satisfying $\mu L = -\lambda\mu$. Let us first check that $\mathrm{supp}(\mu) \subset \sqcup_{a \in A_1} C_a$. We begin by remarking that if $x, y \in V$ are such that $\mu(x) > 0$ and $L(x,y) > 0$, then $\mu(y) > 0$. Indeed, otherwise in the equality

$$\sum_{z \in V \setminus \{y\}} \mu(z)L(z,y) \;=\; -\mu(y)L(y,y) - \lambda\mu(y)$$

the l.h.s. would be positive and the r.h.s. would vanish. It follows by iteration that if $\mu(x) > 0$ and if x leads to y, then $\mu(y) > 0$. In particular, the support of μ is an union of irreducibility classes and at least one of them is included into $\sqcup_{a \in A_1} C_a$. If all the irreducibility classes forming $\mathrm{supp}(\mu)$ are included into $\sqcup_{a \in A_1} C_a$, we get that $\mathrm{supp}(\mu) \subset \sqcup_{a \in A_1} C_a$. Otherwise, we can find $a \in A_1$ and $b \notin A_1$, with $C_a \sqcup C_b \subset \mathrm{supp}(\mu)$ and there exist $x_0 \in C_b$ and $y_0 \in C_a$ with $L(x_0, y_0) > 0$. The restriction to C_a of $\mu L = -\lambda\mu$ writes down $\mu_{C_a} L_{C_a} = -\lambda\mu_{C_a}$, where μ_{C_a} is the restriction of μ to C_a. Since μ_{C_a} is positive, it follows from the uniqueness statement in Perron-Frobenius' theorem, that μ_{C_a} is proportional to the quasi-stationary measure μ_a associated to L_{C_a} and $\lambda = \lambda_1(C_a)$. Due to the property satisfied by x_0, y_0, we have that $\mu(x_0) > 0$ and $((\mu - \mu_{C_a})L)(y_0) > 0$. We deduce that

$$
\begin{aligned}
-\lambda\mu(y_0) \;&=\; (\mu L)(y_0) \\
&=\; ((\mu - \mu_{C_a} + \mu_{C_a})L)(y_0) \\
&=\; ((\mu - \mu_{C_a})L)(y_0) + (\mu_{C_a}L)(y_0) \\
&>\; (\mu_{C_a}L)(y_0) \\
&=\; -\lambda\mu(y_0)
\end{aligned}
$$

which is a contradiction.

The above arguments also show that μ is a mixture of the quasi-stationary measures associated to the irreducible classes included into $\sqcup_{a \in A_1} C_a$. Furthermore, the classes C_a, with $a \in A_1$, which are such that $\mu(C_a) > 0$ must satisfy $\lambda_1(C_a) = \lambda$. It follows that if $\mu L = -\lambda_1\mu$, then the support of μ is included into V_k, where $k \in [\![\kappa_1]\!]$ is such that $\lambda = \lambda_k$, and is equal to V_k if μ is chosen to be a non-degenerate convex combination of the quasi-stationary measures associated to the C_a included into V_k.

∎

This result allows us to adapt the proof of Lemma 12 and Corollary 13 to get the following generalization, where a sublink stands for an invertible subMarkov kernel (i.e. a matrix with non-negative entries whose row sums are bounded above by 1). We also say that two subMarkovian generators L and $\widetilde{L}$ are subMarkov-similar if there exist two sublinks Λ and $\widetilde{\Lambda}$ such that (1) is valid.

Lemma 15 *Let L, $\widetilde{L}$ be two subMarkov generators and Λ a sublink such that $L\Lambda = \Lambda\widetilde{L}$. Then we have $\mathcal{L} \subset \widetilde{\mathcal{L}}$. Assume furthermore that L and $\widetilde{L}$ are subMarkov-similar. Then $\mathcal{L} = \widetilde{\mathcal{L}}$ and the*

subMarkov generators L_m and $\widetilde{L}_m$ are subMarkov-similar, for $m \in [\![\kappa_1]\!]$, as well as the subMarkov generators $L_W := L_{W,W}$ and $\widetilde{L}_{\widetilde{W}} := \widetilde{L}_{\widetilde{W},\widetilde{W}}$, where $W := V \backslash (V_1 \sqcup \cdots \sqcup V_{\kappa_1})$ and $\widetilde{W} := \widetilde{V} \backslash (\widetilde{V}_1 \sqcup \cdots \sqcup \widetilde{V}_{\kappa_1})$.

Proof

With the notation of Lemma 14, consider $\lambda \in \mathcal{L}$ and a probability measure μ on V such that $\mu L = -\lambda\mu$. The measure $\mu\Lambda$ is non-negative and cannot be 0, because Λ is invertible. We can thus define the probability measure $\widetilde{\mu} := \mu\Lambda / \mu\Lambda(\widetilde{V})$. By multiplying on the left the relation $L\Lambda = \Lambda\widetilde{L}$ by μ, we get that $\widetilde{\mu}\widetilde{L} = -\lambda\widetilde{\mu}$, so that $\lambda \in \widetilde{\mathcal{L}}$.

So if L and $\widetilde{L}$ are subMarkov-similar, we get $\mathcal{L} = \widetilde{\mathcal{L}}$. The arguments of the proofs of Corollary 13 and Lemma 12 can now be repeated, with the notion of invariant measure replaced by that of eigen-probability measure associated to $\lambda \in \mathcal{L}$ (with respect to $-L$ and $-\widetilde{L}$). Indeed, the subMarkov-similarity of the subMarkov generators L_W and $\widetilde{L}_{\widetilde{W}}$ is also valid in Corollary 13, using the sublinks $\Lambda_{W,\widetilde{W}}$ and $\widetilde{\Lambda}_{\widetilde{W},W}$. It was not asserted there, just because the subMarkov-similarity between subMarkov generators had not yet been defined.

∎

Remark 16 From the above proof, it also follows that for all $m \in [\![\mathfrak{m}]\!]$, we have $\Lambda_{V_m, \bar{V}_m} = 0$, where

$$
\begin{aligned}
\bar{V}_m &:= \bigsqcup_{m \in J_m} V_m \\
J_m &:= [\![\kappa_1 + \cdots + \kappa_{j_m} + 1, \mathfrak{m}]\!] \backslash \{m\}
\end{aligned}
$$

where $j_m \in [\![0, \eta - 1]\!]$ is such that $m \in [\![\kappa_1 + \cdots + \kappa_{j_m} + 1, \kappa_1 + \cdots + \kappa_{j_m+1}]\!]$.

□

Proposition 8 is now a simple consequence of the previous lemma. Indeed, extending naturally Conditions (C) and (H) to subMarkovian generators, we get:

Proposition 17 *Consider two subMarkov generators L and $\widetilde{L}$. If they are subMarkov-similar, then they satisfy Hypothesis (H).*

Proof

Applying iteratively Lemma 15, we end up with the conclusion that for $l \in [\![\mathfrak{m}]\!]$, L_l is subMarkov-similar to $\widetilde{L}_l$ and $\mathcal{L}_l = \widetilde{\mathcal{L}}_l$. SubMarkov-similarity implying similarity, we conclude to the announced validity of Hypotheses (C) and (H).

∎

Let us now mention an extension of Theorem 1 to the present subMarkov framework. In some sense, the following result is the "Dirichlet condition" analogue of Theorem 1 (whose "Neumann condition" corresponds to the fact that Markov processes are conservative). We say that a subMarkov generator L is isotransient, if $L = L_1$ (this appellation amounts to non-transcience for Markov generators). Note in particular that for any subMarkov generator L, L_m is isotransient for all $m \in [\![\mathfrak{m}]\!]$.

Proposition 18 *Two isotransient subMarkov generators L and $\widetilde{L}$ are subMarkov-similar if and only if they are similar.*

Proof

As usual, the direct implication is obvious. We begin by showing the subMarkovian extension of Lemma 7, namely that two finite and irreducible subMarkov generators are similar if and only if they are subMarkov-similar. Consider two similar and irreducible subMarkov generators L and $\widetilde{L}$.

By Perron-Frobenius' theorem, there a exists positive eigenvector ψ associated to the first Dirichlet eigenvalue λ_1 of L. The operator $L^\dagger \cdot := \psi^{-1}(L - \lambda_1 I)[\psi \cdot]$ is a Markov generator (sometimes called the Doob transform of L, see e.g. [7]). Its spectrum is the spectrum of L shifted by λ_1. If $\widetilde{L}^\dagger$ is constructed similarly for $\widetilde{L}$, it appears that $L^\dagger$ and $\widetilde{L}^\dagger$ are similar irreducible Markov generators, so from Lemma 7, there exist two links $\Lambda^\dagger$ and $\widetilde{\Lambda}^\dagger$ such that

$$L^\dagger \Lambda^\dagger \;=\; \Lambda^\dagger \widetilde{L}^\dagger \quad \text{and} \quad \widetilde{L}^\dagger \widetilde{\Lambda}^\dagger \;=\; \widetilde{\Lambda}^\dagger L^\dagger$$

It remains to define the non-negative kernels

$$\Lambda[\cdot] \;:=\; \psi \Lambda^\dagger \left[\frac{1}{\widetilde{\psi}} \cdot \right]$$

$$\widetilde{\Lambda}[\cdot] \;:=\; \widetilde{\psi} \widetilde{\Lambda}^\dagger \left[\frac{1}{\psi} \cdot \right]$$

to ensure that (1) is satisfied. To get sublinks, divide Λ and $\widetilde{\Lambda}$ by a sufficiently large constant.

When L and $\widetilde{L}$ are isotransient subMarkov generators, let $C_1, C_2, ..., C_{\mathfrak{n}}$, with $m \in \mathbb{N}$, be the irreducibility class(es) of L and let λ_1 be the common first Dirichlet eigenvalue of the corresponding restrictions L_{C_l}. For $l \in [\![\mathfrak{n}]\!]$, let ψ_l be a positive eigenvector on C_l associated to λ_1. Let ψ the function on V coinciding with these eigenvectors on each of the C_l, for $l \in [\![\mathfrak{n}]\!]$. Do the same with $\widetilde{L}$, remarking that $\widetilde{\lambda}_1 = \lambda_1$ by similarity. The previous arguments are then still valid.
∎

We end this section with an example on four points illustrating that Hypothesis (H) is not sufficient for Markov-similarity.

Example 19 On $V := [\![4]\!]$, consider for any $p \in [0, 1)$, the Markov generator

$$L^{(p)} \;:=\; \begin{pmatrix} 0 & 0 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 2p & 2(1-p) & -2 \end{pmatrix}$$

and we denote $L := L^{(0)}$.

For all $p \in [0, 1)$, $-L^{(p)}$ has three eigenvalues: $\lambda_1 := 0$, $\lambda_2 := 1$ and $\lambda_3 := 2$. The similarity class of $L^{(p)}$ depends on the geometric multiplicity of λ_3, which is either 2 or 1 (with then a Jordan block of dimension 2 associated to λ_3). Computing the eigenspace of λ_3, it appears that there is a Jordan block of dimension 2 associated to λ_3 if and only if $p \neq 1/2$.

Moreover the spatial decomposition of $L^{(p)}$ is immediate to obtain for all $p \in [0, 1)$: for all $k \in [\![4]\!]$, we have $V_k = \{k\}$, $L_k^{(p)} = L(k, k)$ and $\kappa^{(p)} = (1, 1, 1, 1)$. It follows that if $\widetilde{L} := L^{(p)}$, with $p \in [0, 1)$, then L and $\widetilde{L}$ satisfy Condition (C). In particular (C) does not imply similarity for $p = 1/2$ and Hypothesis (H) is true if and only if $p \neq 1/2$. From now on, we assume that $\widetilde{L} := L^{(p)}$, with a fixed $p \in [0, 1) \backslash \{1/2\}$ and we are wondering if L and $\widetilde{L}$ are Markov-similar. We show below that this is the case if and only if $p \in [0, 1/2)$.

Denote $A := (0, 0, 2)$, $\widetilde{A} := (0, 2p, 2(1-p))$ and

$$K \;:=\; \begin{pmatrix} 0 & 0 & 0 \\ 2 & -2 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

so we can write

$$L \;=\; \begin{pmatrix} K & 0 \\ A & -2 \end{pmatrix} \quad \text{and} \quad \widetilde{L} \;=\; \begin{pmatrix} K & 0 \\ \widetilde{A} & -2 \end{pmatrix}$$

14

Let Λ a link such that $L\Lambda = \Lambda\widetilde{L}$, then necessarily it can be written under the form

$$\Lambda = \begin{pmatrix} Q & 0 \\ B & d \end{pmatrix}$$

where Q is a link, $B = (a, b, c)$ and $d = 1 - a - b - c$, with $a, b, c, d \in [0, 1]$. Indeed, $\widetilde{L}\mathbb{1}_{\{4\}} = -2\mathbb{1}_{\{4\}}$, so that $\Lambda\mathbb{1}_{\{4\}}$ is an eigenfunction associated to the eigenvalue -2 of L. Since $p \neq 1/2$, such an eigenfunction is proportional to $\mathbb{1}_{\{4\}}$, which amounts to the above form of Λ. This form can also be deduced from Remark 16, which enables to see a priori that Q must be lower diagonal.

The intertwining relation $L\Lambda = \Lambda\widetilde{L}$ is then equivalent to

$$KQ = QK \quad \text{and} \quad AQ - 2B = BK + d\widetilde{A} \tag{9}$$

Define

$$\varphi_1 := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \varphi_2 := \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \qquad \varphi_3 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

which are eigenvectors of K associated respectively to the eigenvalues $0, -2, -1$. The intertwining relation is equivalent to the existence of $x, y, z \in \mathbb{R}$ such that

$$Q\varphi_1 = x\varphi_1, \qquad Q\varphi_2 = y\varphi_2, \qquad Q\varphi_3 = z\varphi_3$$

and this means that

$$Q = \begin{pmatrix} x & 0 & 0 \\ x - y & y & 0 \\ x + y - 2z & z - y & z \end{pmatrix}$$

The fact that Λ is required to be an invertible Markov kernel is then equivalent to the constraints

$$x = 1, \qquad 0 < y \leqslant 1, \qquad y \leqslant z \leqslant (1 + y)/2$$

It follows that Condition (9) is equivalent to

$$\begin{cases} 1 + y - 2z &= 2a \\ 2(z - y) &= c + 2dp \\ 2z - c &= 2(1 - p)d \end{cases}$$

itself equivalent to

$$\begin{cases} a &= (1 + y - 2z)/2 \\ b &= (1 + y - 2z)/2 \\ c &= \frac{2}{2p-1}((1 - p)y + (2p - 1)z) \end{cases}$$

Summing these equations, we get

$$a + b + c = 1 + \frac{1}{2p - 1}y$$

so the requirement $a + b + c \leqslant 1$ is equivalent to $p < 1/2$ (recall that $y > 0$). Conversely, if $p \in [0, 1/2)$, taking e.g. $z = 1/4$ and $y > 0$ small enough leads to a solution for the link Λ.

Similar considerations show that there exists a link $\widetilde{\Lambda}$ such that $\widetilde{\Lambda}L = \widetilde{L}\widetilde{\Lambda}$ if and only if $p \in [0, 1/2)$.

□

Remark 20 In addition to Remark 16, in general the link Λ is not such that Λ_{V_m, V_m} is itself a link, for $m \in [\![\mathfrak{m}]\!]$. Indeed, in the above example under this restriction, we would have end up with $\Lambda = I$, the identity matrix, which does not enable to intertwine L and $L^{(p)}$ for $p \in (0, 1/2)$.

□

4 Comparisons of mixing speeds

The goal of this section is to discuss the consequences of Markov-similarity on speeds of convergence to equilibrium, and especially to prove Proposition 4. We introduce some sets of and Markov kernels and probability measures associated to finite Markov-similar generators which play a crucial role.

More precisely, if L and $\widetilde{L}$ are finite Markov-similar generators as in the introduction, denote $\mathcal{K}(L, \widetilde{L})$ the set of Markov kernels Λ from V to $\widetilde{V}$ such that $L\Lambda = \Lambda\widetilde{L}$ and $\mathcal{P}(L, \widetilde{L}) \coloneqq \mathcal{P}(V)\mathcal{K}(L, \widetilde{L})$, namely the set of probability measures $\widetilde{m}$ on $\widetilde{V}$ such that there exists $m \in \mathcal{P}(V)$ (recall that $\mathcal{P}(V)$ stands for the set of all probability measures on V) and $\Lambda \in \mathcal{K}(L, \widetilde{L})$ such that $\widetilde{m} = m\Lambda$. The sets $\mathcal{K}(\widetilde{L}, L)$ and $\mathcal{P}(\widetilde{L}, L)$ are defined symmetrically, by inverting the roles of L and $\widetilde{L}$. Here is the advantage of Markov-similarity:

Lemma 21 *Assume furthermore that L is irreducible. Then the set $\mathcal{P}(L, \widetilde{L})$ is a neighborhood of $\widetilde{\mu}$ the invariant measure of $\widetilde{L}$, which is necessarily also irreducible.*

We don't know if $\mathcal{P}(L, \widetilde{L})$ is always convex.

Proof

Since L and $\widetilde{L}$ are Markov-similar, we deduce from Corollary 11 that the non-transience of L implies that of $\widetilde{L}$. Furthermore the number of irreducible class(es) of $\widetilde{L}$ is that of L, so $\widetilde{L}$ is irreducible. It follows that $\widetilde{L}$ has a unique invariant measure and it is given by $\widetilde{\mu} = \mu\Lambda$, for all $\Lambda \in \mathcal{K}(L, \widetilde{L})$. Note that μ belongs to the interior of $\mathcal{P}(V)$ (as a subset of the space of all signed measures on V), because it gives a positive weight to all the points. By Markov-similarity, we can find a link Λ in $\mathcal{K}(L, \widetilde{L})$. Its invertibility implies that it transforms a neighborhood of μ into a neighborhood of $\widetilde{\mu}$. So $\mathcal{P}(L, \widetilde{L})$ is a neighborhood of $\widetilde{\mu}$.

∎

Remark that $\mathcal{P}(L, \widetilde{L})$ is left invariant by the Markov semi-group $(\exp(tL))_{t \geqslant 0}$ generated by L. Indeed, for any initial conditions $m_0 \in \mathcal{P}(V)$, $\widetilde{m}_0 \in \mathcal{P}(\widetilde{V})$ and any time $t \geqslant 0$, denote

$$m_t \coloneqq m_0\exp(tL) \quad \text{and} \quad \widetilde{m}_t \coloneqq \widetilde{m}_0\exp(t\widetilde{L})$$

the marginal distribution at time t obtained respectively through the evolutions generated by L and $\widetilde{L}$. Assume $\widetilde{m}_0 \in \mathcal{P}(L, \widetilde{L})$, so there exist $m_0 \in \mathcal{P}(V)$ and $\Lambda \in \mathcal{K}(L, \widetilde{L})$ such that $\widetilde{m}_0 = m_0\Lambda$. For any $t \geqslant 0$, we get

$$
\begin{aligned}
\widetilde{m}_t &= \widetilde{m}_0\exp(t\widetilde{L}) \\
&= m_0\Lambda\exp(t\widetilde{L}) \\
&= m_0\exp(tL)\Lambda \\
&= m_t\Lambda
\end{aligned}
$$

This implies that $\widetilde{m}_t \in \mathcal{P}(L, \widetilde{L})$, as announced. Since we also have $\widetilde{\mu} = \mu\Lambda$, it follows that

$$
\begin{aligned}
E_\psi(\widetilde{m}_t|\widetilde{\mu}) &= E_\psi(m_t\Lambda|\mu\Lambda) \\
&\leqslant E_\psi(m_t|\mu)
\end{aligned}
\tag{10}
$$

because ψ-entropies decrease under the action of Markov kernels, for any $\psi \in \Psi$. This well-known property, which holds on general measurable spaces (see e.g. Proposition 1.1 of [4]), is an important reason behind our interest in considering intertwining kernels which are Markovian. Thus, seen through the ψ-entropy, the convergence of $\widetilde{m}_t$ toward $\widetilde{\mu}$ for large $t \geqslant 0$ is dominated by that of m_t toward μ. In particular we deduce that

$$\widetilde{m}_0 \in \mathcal{P}(L, \widetilde{L}) \quad \Rightarrow \quad \forall\, t \geqslant 0,\, \forall\, \psi \in \Psi, \quad E_\psi(\widetilde{m}_t|\mu) \leqslant E(\psi, t)$$

Proposition 4 is now a simple consequence of

Lemma 22 *Under the assumption of Lemma 21, there exists $T \geqslant 0$, such that for any $\widetilde{m}_0 \in \mathcal{P}(\widetilde{V})$, $\widetilde{m}_T \in \mathcal{P}(L, \widetilde{L})$.*

Of course in practice, the problem will be to describe $\mathcal{P}(L, \widetilde{L})$ and to estimate T.

Proof

As in the proof of Lemma 21, consider a link $\Lambda \in \mathcal{K}(L, \widetilde{L})$. The set $\mathcal{P}(V)\Lambda \subset \mathcal{P}(L, \widetilde{L})$ is convex and left invariant by the semigroup $(\exp(t\widetilde{L}))_{t \geqslant 0}$. Thus it is sufficient to see that for any $\widetilde{x} \in \widetilde{V}$, there exists $T_{\widetilde{x}} \geqslant 0$ such that $\delta_{\widetilde{x}} \exp(T_{\widetilde{x}}\widetilde{L}) \in \mathcal{P}(V)\Lambda$. Indeed, by stability of $\mathcal{P}(L, \widetilde{L})$ by the semi-group generated by $\widetilde{L}$, we get

$$\forall\ \widetilde{x} \in \widetilde{V}, \qquad \delta_{\widetilde{x}} \exp(T\widetilde{L}) \ \in\ \mathcal{P}(L, \widetilde{L})$$

with $T := \max\{T_{\widetilde{x}} : \widetilde{x} \in \widetilde{V}\}$. By convexity of the mapping $\mathcal{P}(\widetilde{V}) \ni \widetilde{m}_0 \mapsto \widetilde{m}_T$ and of the set $\mathcal{P}(V)\Lambda$, it appears then that

$$\forall\ \widetilde{m}_0 \in \mathcal{P}(\widetilde{V}), \qquad \widetilde{m}_0 \exp(T\widetilde{L}) \ \in\ \mathcal{P}(V)\Lambda \ \subset\ \mathcal{P}(L, \widetilde{L})$$

But for any fixed $\widetilde{x} \in \widetilde{V}$, we have that $\delta_{\widetilde{x}} \exp(t\widetilde{L})$ converges toward $\widetilde{\mu}$ for large t, so for large enough $T_{\widetilde{x}} \geqslant 0$, $\delta_{\widetilde{x}} \exp(T_{\widetilde{x}}\widetilde{L})$ belongs to the neighborhood $\mathcal{P}(L, \widetilde{L})$ of $\widetilde{\mu}$. $\blacksquare$

In this paper, we adopted an equivalence relation point of view on Markov intertwinings, through the Markov-similarity. But the order relation aspect of the Markov intertwinings is also very interesting and maybe more relevant for applications. Such considerations can be found in [14], but let us slightly modify the definitions given there by saying that the Markov generator $\widetilde{L}$ on the finite set $\widetilde{V}$ is dominated by the Markov generator L on the finite set V (written $\widetilde{L} \prec L$), if there exists a injective Markov kernel Λ from V to $\widetilde{V}$ such that $L\Lambda = \Lambda\widetilde{L}$. The requirement that Λ is one-to-one (with respect to the functional interpretation (2)) means that $|V| \geqslant |\widetilde{V}|$ and that Λ has maximal rank as a matrix. Note that two Markov generators L and $\widetilde{L}$ are Markov-similar if and only if $\widetilde{L} \prec L$ and $L \prec \widetilde{L}$. Most of the results presented up to now have variants for the domination relation $\prec$. In this spirit, Lemma 21 can be strengthened into

Lemma 23 *Assume that the two Markov generators L and $\widetilde{L}$ are such that $\widetilde{L} \prec L$ and L is irreducible. Then $\widetilde{L}$ is irreducible and if $\widetilde{\mu}$ is its invariant probability, $\mathcal{P}(L, \widetilde{L})$ is a neighborhhood of $\widetilde{\mu}$.*

Proof

We begin by proving that $\widetilde{L}$ is irreducible. Let $\Lambda \in \mathcal{K}(L, \widetilde{L})$ be injective. Let $\widetilde{f}$ be a function on $\widetilde{V}$ such that $\widetilde{L}[\widetilde{f}] = 0$. By the intertwining relation, we get that $L[\Lambda[\widetilde{f}]] = 0$, so that by irreducibility of L, $\Lambda[\widetilde{f}]$ is constant and by injectivity of Λ, $\widetilde{f}$ is constant (since $\Lambda[\mathbb{1}] = \mathbb{1}$). This property implies that if $\widetilde{V}$ is decomposed into irreducible classes with respect to $\widetilde{L}$, then there is only one terminal class (namely $\widetilde{A}_1$ is a singleton, with the notation introduced at the beginning of Section 3). So to prove that $\widetilde{L}$ is irreducible, it is sufficient that to show that $\widetilde{L}$ admits an invariant probability whose support is $\widetilde{V}$. By the intertwining relation, we get that $\widetilde{\mu} := \mu\Lambda$ is an invariant probability of $\widetilde{L}$, if μ is the invariant probability of L. It remains to see that $\widetilde{\mu}$ gives a positive weight to all the elements of $\widetilde{V}$. Let $\mathcal{M}(V)$ be the set of signed measures on V. The Markov kernel Λ can be seen as an operator from $\mathcal{M}(V)$ to $\mathcal{M}(\widetilde{V})$ via:

$$\forall\ m \in \mathcal{M}(V), \forall\ \widetilde{x} \in \widetilde{V}, \qquad m\Lambda(\widetilde{x}) \ := \ \sum_{x \in V} m(x)\Lambda(x, \widetilde{x})$$

It corresponds to the dual operator of Λ seen as an operator on functions, through the natural duality between functions and signed measures on V. In particular, seen as an operator on signed

measures, Λ is onto. As a consequence, for any $\widetilde{x} \in \widetilde{V}$, we can find $m_{\widetilde{x}} \in \mathcal{M}(V)$ such that $m_{\widetilde{x}}\Lambda = \delta_{\widetilde{x}}$. Since μ gives a positive weight to all elements of V, we can find $\epsilon > 0$ small enough so that for all real numbers $(a_x)_{x \in V}$ with $|a_x| \leqslant \epsilon$ for all $x \in V$, $\mu + \sum_{x \in V} a_x m_x$ is a non negative measure. It follows that

$$\widetilde{\mu} + \sum_{x \in V} a_x \delta_x \quad = \quad (\mu + \sum_{x \in V} a_x m_x)\Lambda$$

is a non-negative measure. This is only possible, for all $(a_x)_{x \in V}$ as above, if and only if $\widetilde{\mu}$ gives a positive weight to all the elements of $\widetilde{V}$.

The same argument shows that Λ transforms neighborhoods of μ into neighborhoods of $\widetilde{\mu}$, so $\mathcal{P}(L, \widetilde{L})$ is a neighborhood of $\widetilde{\mu}$.

■

The proof of Lemma 22 can now be adapted (replacing the link Λ by an injective Markov kernel) to show:

Proposition 24 *Under the assumption of Lemma 22, there exists $T \geqslant 0$ such that $\widetilde{m}_T \in \mathcal{P}(L, \widetilde{L})$ for all $\widetilde{m}_0 \in \mathcal{P}(\widetilde{V})$. It follows that*

$$\forall \ \psi \in \Psi, \ \forall \ t \geqslant 0, \qquad \widetilde{E}(\psi, T + t) \quad \leqslant \quad E(\psi, t)$$

with the notation introduced in Proposition 4, because for any $\psi \in \Psi$, $\widetilde{m}_0 \in \mathcal{P}(L, \widetilde{L})$ and $t \geqslant 0$, we have

$$E_\psi(\widetilde{m}_t | \widetilde{\mu}) \quad \leqslant \quad E(\psi, t)$$

according to (10).

Let us illustrate the previous considerations on the simplest example.

Example 25 Consider the two points set $V := \{0, 1\}$. Any generator L on V can be written under the form $L = l(\mu - \text{Id})$, where Id is the 2×2-identity matrix, $l \geqslant 0$ and μ is "the" invariant measure of L (note that except if $L = 0$, which corresponds to $l = 0$, L has a unique invariant measure μ). It appears that $-L$ is diagonalizable and its eigenvalues are 0 and l. The generator L is non-transient if and only if $a > 0$ and $\mu > 0$ (in the sense that $\mu(0) > 0$ and $\mu(1) > 0$). The generator $L \neq 0$ is transient if and only if $l > 0$ and μ is a Dirac mass. The left case is $L = 0$. Consider another generator $\widetilde{L} = \widetilde{l}(\widetilde{\mu} - \text{Id})$ on $\{0, 1\}$. According to Corollary 11 and Theorem 1, it is Markov-similar to the non-transient L if and only if $l = \widetilde{l}$ and $\widetilde{\mu} > 0$. From Corollary 11, we also deduce that the generator $\widetilde{L} \neq 0$ is Markov-similar to the transient $L \neq 0$ if and only if $\widetilde{l} = l$ and $\widetilde{\mu}$ is a Dirac mass. Finally the unique generator Markov similar to $L = 0$ is 0 itself.

From now on, we assume that L and $\widetilde{L}$, as above, are non-transient and Markov-similar. Let Λ be a Markov kernel on $\{0, 1\}$ such that $L\Lambda = \Lambda\widetilde{L}$. This amounts to $\mu\Lambda = \widetilde{\mu}$, namely $\mathcal{K}(L, \widetilde{L})$ is the set of Markov kernels transporting μ on $\widetilde{\mu}$ (in general, it is only a subset of those Markov kernels). Since $\Lambda - \text{Id}$ is a Markov generator, we can find $a \geqslant 0$ and a probability measure ν on $\{0, 1\}$ such that

$$\Lambda \quad = \quad (1 - a)\text{Id} + a\nu$$

This is not sufficient to insure that Λ is a Markov kernel: to get that the entries are non-negative, we need furthermore that $a \in [0, 1/(1 - \min(\nu))]$, but it will not be convenient to work directly with this condition. The relation $\mu\Lambda = \widetilde{\mu}$ is equivalent to

$$a\nu \quad = \quad \widetilde{\mu} - (1 - a)\mu$$

For the l.h.s. to be non-negative, we must have $a \geqslant 1 - \min(\widetilde{\mu}/\mu)$. The kernel Λ can be written under the form

$$\Lambda_a \ :=\ (1-a)(\mathrm{Id}-\mu) + \widetilde{\mu}$$

For the entries of this matrix to be non-negative, we must have:
- For $a > 1$: for all $x \in \{0,1\}$,

$$(1-a)(1-\mu(x)) + \widetilde{\mu}(x) \ \geqslant\ 0$$

i.e.

$$a \ \leqslant\ 1 + \frac{\widetilde{\mu}(x)}{1-\mu(x)}$$

Let $a_+ := 1 + \min(\widetilde{\mu}/(1-\mu))$, this condition is $a \leqslant a_+$.
- For $a = 1$, $\Lambda_1 = \widetilde{\mu}$ has non-negative entries.
- For $a < 1$: for all $x \in \{0,1\}$,

$$(a-1)\mu(x) + \widetilde{\mu}(x) \ \geqslant\ 0$$

and we recover the condition $a \geqslant a_- := 1 - \min(\widetilde{\mu}/\mu)$.

Thus we get

$$\mathcal{K}(L,\widetilde{L}) \ =\ \{\Lambda_a \ :\ a \in [a_-, a_+]\}$$

Since the mapping $a \mapsto \Lambda_a$ is affine and that the set of probability measures on $\{0,1\}$ is of dimension 1, it appears that $\mathcal{P}(L,\widetilde{L})$ is the segment generated by the four probabilities $\eta_{\star,y} := \delta_y \Lambda_{a_\star}$, with $y \in \{0,1\}$ and $\star \in \{-,+\}$. Let $x_0, x_1 \in \{0,1\}$ be respectively such that

$$\frac{\widetilde{\mu}}{\mu}(x_0) \ =\ \min\left(\frac{\widetilde{\mu}}{\mu}\right)$$

$$\frac{\widetilde{\mu}}{1-\mu}(x_1) \ =\ \min\left(\frac{\widetilde{\mu}}{1-\mu}\right)$$

We have, for any $y \in \{0,1\}$,

$$\eta_{-,y} \ =\ \frac{\widetilde{\mu}(x_0)}{\mu(x_0)}(\delta_y - \mu) + \widetilde{\mu}$$

$$\eta_{+,y} \ =\ \frac{\widetilde{\mu}(x_1)}{1-\mu(x_1)}(\mu - \delta_y) + \widetilde{\mu}$$

So, denoting $\bar{x} := 1 - x$ for all $x \in \{0,1\}$, we compute that $\eta_{-,\bar{x}_0} = \delta_{\bar{x}_0}$ and $\eta_{+,x_1} = \delta_{\bar{x}_1}$. We also get

$$\eta_{-,x_0}(x_0) \ =\ \frac{\widetilde{\mu}(x_0)}{\mu(x_0)} \quad\text{and}\quad \eta_{+,\bar{x}_1}(x_1) \ =\ \frac{\widetilde{\mu}(x_1)}{\mu(\bar{x}_1)} \tag{11}$$

So η_{-,x_0} is a Dirac mass if and only if $\widetilde{\mu} = \mu$ and $\eta_{+,\bar{x}_1}$ is a Dirac mass if and only if $\widetilde{\mu}$ is the image of μ by the involution of $\{0,1\}$, $x \mapsto \bar{x}$. Without loss of generality, assume that $\mu(1) \geqslant \mu(0)$ and $\widetilde{\mu}(1) \geqslant \widetilde{\mu}(0)$. In particular $\widetilde{\mu}$ is the image of μ by the involution of $\{0,1\}$, $x \mapsto \bar{x}$ if and only if μ and $\widetilde{\mu}$ are the uniform measure. Next let us dismiss the cases where $\mu = \widetilde{\mu}$, i.e. $L = \widetilde{L}$, because it is clear then that $\mathcal{P}(L,L) = \mathcal{P}(V)$. From the above considerations, it follows that for $L \neq \widetilde{L}$, we will have $\mathcal{P}(L,\widetilde{L}) = \mathcal{P}(V)$ if and only if the convex hull of $\{\delta_{\bar{x}_0}, \delta_{\bar{x}_1}\}$ is $\mathcal{P}(V)$, i.e. if $x_0 \neq x_1$. But we just assumed that μ and $\widetilde{\mu}$ are non-decreasing, we have $x_1 = 0$. Thus $\mathcal{P}(L,\widetilde{L}) = \mathcal{P}(V)$ if and

only if $\frac{\widetilde{\mu}(1)}{\mu(1)} \geqslant \frac{\widetilde{\mu}(0)}{\mu(0)}$, i.e. $\widetilde{\mu}(1) \leqslant \mu(1)$. Note that $\mathcal{P}(L,\widetilde{L}) = \mathcal{P}(V) = \mathcal{P}(\widetilde{L},L)$ is in fact equivalent to $L = \widetilde{L}$. A first conclusion is that if $\widetilde{\mu}(1) \leqslant \mu(1)$, then

$$\forall \ \psi \in \Psi, \ \forall \ t \geqslant 0, \qquad \widetilde{E}(\psi,t) \ \leqslant \ E(\psi,t)$$

Assume next that $\widetilde{\mu}(1) > \mu(1)$. Then we have $x_0 = x_1 = 0$ and from (11) (taking into account that $\widetilde{\mu}(0)/\mu(0) \geqslant \widetilde{\mu}(0)/\mu(1)$) we deduce that

$$\mathcal{P}(L,\widetilde{L}) \ = \ \left[\frac{\widetilde{\mu}(0)}{\mu(0)}\delta_0 + \left(1 - \frac{\widetilde{\mu}(0)}{\mu(0)}\right)\delta_1, \delta_1 \right]$$

It follows that we can take in Lemma 22,

$$
\begin{aligned}
T \ &= \ \min\left\{t \geqslant 0 \ : \ \forall \ \widetilde{m}_0 \in \mathcal{P}(\{0,1\}), \widetilde{m}_t(1) \geqslant 1 - \frac{\widetilde{\mu}(0)}{\mu(0)}\right\} \\
&= \ \min\left\{t \geqslant 0 \ : \ \delta_0 \exp(t\widetilde{L})(1) \geqslant 1 - \frac{\widetilde{\mu}(0)}{\mu(0)}\right\} \\
&= \ \min\left\{t \geqslant 0 \ : \ (1 - \exp(-lt))\widetilde{\mu}(1) \geqslant 1 - \frac{\widetilde{\mu}(0)}{\mu(0)}\right\} \\
&= \ -\frac{1}{l}\ln\left(1 - \frac{\widetilde{\mu}(1) - \mu(1)}{\widetilde{\mu}(1)(1 - \mu(1))}\right)
\end{aligned}
$$

$\square$

We end this section by pointing out the links between the objects introduced above with the notion of weak hypergroup. Let be given a Markov generator L on the finite set V. The set $\mathcal{K}(L,L)$ was called the Markov commutator of L in [14], since it consists of the Markov kernels commuting with L (as already mentioned in Remark 3). Following this previous paper, the generator L is said to be a **weak hypergroup** with respect to $x_0 \in V$ if for any $m \in \mathcal{P}(V)$, there exists $K \in \mathcal{K}(L,L)$ such that $K(x_0,\cdot) = m(\cdot)$. Taking advantage of the fact that for any Markov generators $L, \widetilde{L}$ and $\widehat{L}$, we have the inclusion $\mathcal{K}(L,\widetilde{L})\mathcal{K}(\widetilde{L},\widehat{L}) \subset \mathcal{K}(L,\widehat{L})$, we deduce the following criterion:

Proposition 26 *Assume that L and $\widetilde{L}$ are two Markov generators on V and $\widetilde{V}$ respectively, such that $\widetilde{L}$ is a weak hypergroup with respect to $\widetilde{x}_0$ and there exists $x_0 \in V$ and $\Lambda \in \mathcal{K}(L,\widetilde{L})$ with $\Lambda(x_0,\cdot) = \delta_{\widetilde{x}_0}$. Then we have $\mathcal{P}(L,\widetilde{L}) = \mathcal{P}(\widetilde{V})$ and by consequence,*

$$\forall \ \psi \in \Psi, \ \forall \ t \geqslant 0, \qquad \widetilde{E}(\psi,t) \ \leqslant \ E(\psi,t)$$

This condition generalizes the deduction of $\mathcal{P}(L,\widetilde{L}) = \mathcal{P}(\widetilde{V})$ given in Example 25, which is continued below:

Example 27 We come back to the two point case, with the notation introduced in Example 25. Consider $L = l(\mu - \mathrm{Id})$ a non-transient generator, where $\mu(0) \leqslant \mu(1)$. Let us check that L is a weak hypergroup with respect to 0. We begin by computing the commutator $\mathcal{K}(L,L)$. We have seen that any Markov kernel K on $\{0,1\}$ can be written under the form $(1 - k)\mathrm{Id} + k\nu$, where $\nu \in \mathcal{P}(\{0,1\})$ and $k \in [0, 1/(1 - \min(\nu))]$. It appears that K commutes with L if and only if $k\nu$ commutes with μ, namely if $k = 0$ or $\nu = \mu$. So we get

$$\mathcal{K}(L,L) \ = \ \{K_k := (1 - k)\mathrm{Id} + k\mu \ : \ k \in [0, 1/(1 - \min(\mu))]\}$$

Since $K_0(0,\cdot) = \delta_0$ and $K_{1/(1-\mu(0))}(0,\cdot) = \delta_1$, we get that L is a weak hypergroup.

Consider another non-transient Markov generator $\widetilde{L}$, to fulfill the assumptions of Proposition 26, we are wondering if we can find $x_0 \in \{0,1\}$ and $\Lambda \in \mathcal{K}(L,\widetilde{L})$ such that $\Lambda(x_0,\cdot) = \delta_0$. As we have already deduced it from (11), this is equivalent to $\widetilde{\mu}(1) \leqslant \mu(1)$.

$\square$

5 On infinite state spaces

The goal of this short section is to prove Proposition 5 and to suggest that the infinite state space situation would deserve to be investigated further.

Let L and $\widetilde{L}$ be two Markov generators, respectively on the measurable spaces $(V, \mathcal{V})$ and $(\widetilde{V}, \widetilde{\mathcal{V}})$. The simple implication in Proposition 5 holds under weaker assumptions than L and $\widetilde{L}$ being nice:

Lemma 28 *Assume that L and $\widetilde{L}$ admit unique invariant probabilities μ and $\widetilde{\mu}$ which are reversible, and that their spectra consist of eigenvalues, respectively in $\mathbb{L}^2(\mu)$ and $\mathbb{L}^2(\widetilde{\mu})$. If L and $\widetilde{L}$ are weakly Markov-similar in the abstract sense, then they are isospectral.*

Proof

Let Λ be an abstract weak $\mathbb{L}^2$-link such that $L\Lambda = \Lambda\widetilde{L}$. As the operator Λ is Markovian, it has norm 1. Denote by $(\widetilde{\lambda}_l)_{l\in\mathbb{Z}_+}$ and $(\widetilde{\varphi}_l)_{l\in\mathbb{Z}_+}$ the eigenvalues and respective orthonormal eigenvectors of $\widetilde{L}$. For any $l \in \mathbb{Z}_+$, $\Lambda[\widetilde{\varphi}_l]$ belongs to $\mathbb{L}^2(\mu)$ (its norm is less or equal to 1) and from the intertwining relation we deduce that it is an eigenfunction of L associated to the eigenvalue $\widetilde{\lambda}_l$. Taking into account that Λ is one-to-one, we deduce that the spectrum of $\widetilde{L}$ is included into that of L. Conversely, considering $\widetilde{\Lambda}$ an abstract weak $\mathbb{L}^2$-link such that $\widetilde{L}\widetilde{\Lambda} = \widetilde{\Lambda}L$, we get the reverse inclusion.

∎

The proof of the reciprocal implication is an extension of that of Lemma 6.

Lemma 29 *If the two nice generators L and $\widetilde{L}$ are isospectral, then they are weakly Markov-similar.*

Proof

Let $(\varphi_l)_{l\in\mathbb{Z}_+}$ and $(\widetilde{\varphi}_l)_{l\in\mathbb{Z}_+}$ be bounded orthonormal eigenvectors of L and $\widetilde{L}$, respectively, associated to the same family of eigenvalues $(\lambda_l)_{l\in\mathbb{Z}_+}$. We can and will assume that $\varphi_0 = \mathbb{1}$ and $\widetilde{\varphi}_0 = \mathbb{1}$ are the constant eigenvectors associated to the eigenvalue 0. We will construct an operator Λ such that $L\Lambda = \Lambda\widetilde{L}$ by requiring that

$$\forall\, l \in \mathbb{Z}_+, \qquad \Lambda[\widetilde{\varphi}_l] \;=\; a_l \varphi_l$$

for a conveniently chosen sequence $(a_l)_{l\in\mathbb{Z}_+}$. First we impose that $a_0 = 1$, so that $\Lambda[\mathbb{1}] = \mathbb{1}$. Next we choose the remaining coefficients positive and satisfying

$$\sum_{l\in\mathbb{N}} a_l \, \|\varphi_l\|_\infty \, \|\widetilde{\varphi}_l\|_\infty \;\leqslant\; 1 \tag{12}$$

This is possible, since the eigenvectors are bounded. Let us check that such an operator Λ preserves non-negativity. It is sufficient to show that if $\widetilde{f}$ is a measurable fonction defined on $\widetilde{V}$ and taking values in $[0,1]$, then $\Lambda[\widetilde{f}] \geqslant 0$ μ-a.s. Since $\widetilde{f} \in \mathbb{L}^2(\widetilde{\mu})$, we can decompose it on the orthonormal basis $(\widetilde{\varphi}_l)_{l\in\mathbb{Z}_+}$:

$$\widetilde{f} \;=\; \sum_{l\in\mathbb{Z}_+} b_l \widetilde{\varphi}_l$$

where a priori the coefficients $(b_l)_{l\in\mathbb{Z}_+}$ belong to $\mathbb{R}$. We have $b_0 = \widetilde{\mu}[\mathbb{1}\,\widetilde{f}] \in [0,1]$ and

$$\begin{aligned}
\forall\, l \in \mathbb{N}, \qquad |b_l| &= |\widetilde{\mu}[\widetilde{\varphi}_l \widetilde{f}]| \\
&\leqslant \|\widetilde{\varphi}_l\|_\infty \, \widetilde{\mu}[\widetilde{f}] \\
&= \|\widetilde{\varphi}_l\|_\infty \, b_0
\end{aligned}$$

Thus we get

$$
\begin{aligned}
\Lambda[\widetilde{f}] \;&=\; \Lambda\left[\sum_{l\in\mathbb{Z}_+} b_l\widetilde{\varphi}_l\right]\\
&=\; \sum_{l\in\mathbb{Z}_+} a_l b_l \varphi_l\\
&=\; b_0\mathbb{1} + \sum_{l\in\mathbb{N}} a_l b_l \varphi_l\\
&\geqslant\; b_0 - \sum_{l\in\mathbb{N}} a_l |b_l|\,\|\varphi_l\|_\infty\\
&\geqslant\; \left(1 - \sum_{l\in\mathbb{N}} a_l\,\|\varphi_l\|_\infty\,\|\widetilde{\varphi}_l\|_\infty\right) b_0\\
&\geqslant\; 0
\end{aligned}
$$

according to (12). It follows that Λ is a Markov operator in the abstract sense. It comes from a Markov kernel, due to the assumption on the state spaces.

It remains to show that $\mu\Lambda$ is equal to the invariant probability $\widetilde{\mu}$ of $\widetilde{L}$. By the intertwining relation, $\mu\Lambda$ is an invariant probability of $\widetilde{L}$, thus by uniqueness of the latter, we have $\mu\Lambda = \widetilde{\Lambda}$.
∎

It is natural to imagine a strong version of Proposition 5. The Markov operator $\Lambda : \mathbb{L}^2(\widetilde{\mu}) \to \mathbb{L}^2(\mu)$ is said to be a **strong link** if it is invertible and its inverse is bounded. This notion leads to the definition: two Markov generators L and $\widetilde{L}$ are **strongly Markov-similar** if they can be mutually intertwined through strong links. We are wondering if two nice isospectral Markov generators would not be strongly Markov-similar if their eigenvectors are uniformly bounded, namely with the above notation, if

$$
\sup_{l\in\mathbb{Z}_+} \|\varphi_l\|_\infty \;<\; +\infty \quad\text{and}\quad \sup_{l\in\mathbb{Z}_+} \|\widetilde{\varphi}_l\|_\infty \;<\; +\infty
$$

Note that the examples of isospectral flat manifolds presented in the review of Gordon [10] can be shown to be strongly Markov-similar, by transforming the transplantation maps (i.e. unitary instead of Markovian intertwining maps, see the papers of Bérard [2, 3]) into strong links. More precisely, it is sufficient to take $a, b > 0$ such that $4a + 3b = 1$ in the matrix T displayed page 763 of Gordon [10].

Acknowledgments:

References

[1] Pierre Bérard. Variétés riemanniennes isospectrales non isométriques. *Astérisque*, (177-178):Exp. No. 705, 127–154, 1989. Séminaire Bourbaki, Vol. 1988/89.

[2] Pierre Bérard. Transplantation et isospectralité. I. *Math. Ann.*, 292(3):547–559, 1992.

[3] Pierre Bérard. Transplantation et isospectralité. II. *J. London Math. Soc. (2)*, 48(3):565–576, 1993.

[4] P. Del Moral, M. Ledoux, and L. Miclo. On contraction properties of Markov kernels. *Probab. Theory Related Fields*, 126(3):395–420, 2003.

[5] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.

[6] Persi Diaconis and Laurent Miclo. On times to quasi-stationarity for birth and death processes. *J. Theoret. Probab.*, 22(3):558–586, 2009.

[7] Persi Diaconis and Laurent Miclo. On quantitative convergence to quasi-stationarity. Available at `http://hal.archives-ouvertes.fr/hal-01002622`, June 2014.

[8] E. B. Dynkin. *Markov processes. Vols. I, II*, volume 122 of *Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen Wissenschaften, Bände 121*. Academic Press Inc., Publishers, New York, 1965.

[9] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[10] Carolyn S. Gordon. Survey of isospectral manifolds. In *Handbook of differential geometry, Vol. I*, pages 747–778. North-Holland, Amsterdam, 2000.

[11] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1989.

[12] Mark Kac. Can one hear the shape of a drum? *Amer. Math. Monthly*, 73(4, part II):1–23, 1966.

[13] Laurent Miclo. On absorption times and Dirichlet eigenvalues. *ESAIM Probab. Stat.*, 14:117–150, 2010.

[14] Laurent Miclo. On the Markov commutator. Preprint available at `https://hal.archives-ouvertes.fr/hal-01143511`, April 2015.

[15] S. Pal and M. Shkolnikov. Intertwining diffusions and wave equations. *ArXiv e-prints*, June 2013.

[16] P. Patie and M. Savov. Spectral expansions of non-self-adjoint generalized Laguerre semigroups. *ArXiv e-prints*, June 2015.

[17] L. C. G. Rogers and J. W. Pitman. Markov functions. *Ann. Probab.*, 9(4):573–582, 1981.

miclo@math.univ-toulouse.fr

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