ANNALES
DE LA FACULTÉ
DES SCIENCES
DE
TOULOUSE
Mathématiques

Persi Diaconis and Laurent Miclo

On quantitative convergence to quasi-stationarity

Tome XXIV, nº 4 (2015), p. 973-1016.

<http://afst.cedram.org/item?id=AFST_2015_6_24_4_973_0>

Annales de la Faculté des Sciences de Toulouse Vol. XXIV, n° 4, 2015
pp. 973-1016

On quantitative convergence to quasi-stationarity

Persi Diaconis[1], Laurent Miclo[2]

Résumé. — On cherche à quantifier le comportement en temps long des processus de Markov finis, absorbés et supposés irréductibles en dehors du point absorbant. Par le biais de transformations de Doob, on montre qu'il est juste besoin du ratio maximal des valeurs prises par le premier vecteur propre de Dirichlet associé pour se ramener à la situation bien plus étudiée de la convergence à l'équilibre des processus de Markov finis. On obtient ainsi des estimées explicites de convergence à la quasi-stationnarité, en particulier via l'utilisation d'inégalités fonctionnelles. Quand le processus est de plus réversible, on retrouve le taux optimal de convergence exponentielle donné par le trou spectral entre les deux premières valeurs propres de Dirichlet. Plusieurs exemples simples illustrent les bornes obtenues.

Abstract. — The quantitative long time behavior of absorbing, finite, irreducible Markov processes is considered. Via Doob transforms, it is shown that only the knowledge of the ratio of the values of the underlying first Dirichlet eigenvector is necessary to come back to the well-investigated situation of the convergence to equilibrium of ergodic finite Markov processes. This leads to explicit estimates on the convergence to quasi-stationarity, in particular via functional inequalities. When the process is reversible, the optimal exponential rate consisting of the spectral gap between the two first Dirichlet eigenvalues is recovered. Several simple examples are provided to illustrate the bounds obtained.

[1] Stanford University, Department of Mathematic, Building 380, Sloan Hall, Stanford, California 94305, USA
[2] Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex 9, France
miclo@math.univ-toulouse.fr

1. Introduction

This paper begins to develop a quantitative theory of rates of convergence to quasi-stationarity, as in the following example. Consider the simple symmetric random walk on $\{0, ..., N\}$ with holding $1/2$ at N and absorbing at 0. Let $\bar{X}_t$ be the position of the walk at time $t \in \mathbb{Z}_+$ and T be the absorption time at 0. Let $\mu_t(x) := \mathbb{P}[\bar{X}_t = x | T > t]$, for $x \in \{1, ..., N\}$. Classical theory, reviewed below, shows that

$$\lim_{t\to+\infty} \mu_t(x) \quad = \quad \nu(x) \ := \ Z^{-1} \cos\left(\frac{(2N+1-2x)\pi}{2(2N+1)}\right)$$

with $Z^{-1} := 2\tan\left(\frac{\pi}{2(2N+1)}\right)$, the normalizing constant. The measure ν is called a quasi-stationary distribution. How large does t have to be so that these asymptotics are useful? In Section 3, which is devoted to explicit computations, we prove for the continuous time counterpart of the above process that for any starting distribution on $\{1, ..., N\}$ and for all $s \geqslant 0$,

$$\|\mu_t - \nu\|_{\mathrm{tv}} \quad \leqslant \quad \frac{2\sqrt{2}}{\pi^2}(1 + \mathcal{O}(N^{-1}))\exp(-s) \tag{1.1}$$

for

$$t \quad = \quad \frac{5}{4\pi^2} N^2 \ln(N) + \frac{s}{2\pi^2} N^2$$

Thus the quasi-stationary asymptotics takes hold for t larger than $N^2\ln(N)$. In (1.1), μ_t and ν depend on N but the bounds are uniform in N.

We will work mainly in the continuous time setting, which is more convenient to deal with. We will come back to the discrete time framework in Section 4. Generally, a quasi-stationary distribution of an absorbing Markov process $\bar{X} := (\bar{X}_t)_{t\geqslant 0}$ is a probability measure ν on the state space S (where the absorbing points have been removed) such that starting from this distribution, the time marginal laws $\mathcal{L}(\bar{X}_t)$ remain proportional to ν on S, for all $t \geqslant 0$. For nice processes $\bar{X}$, the quasi-stationary distribution is unique and starting from any distribution on S, the conditional (to non-absorption) law $\mu_t := \mathcal{L}(\bar{X}_t | \bar{X}_t \in S)$ converges toward ν for large times $t \geqslant 0$. The purpose of this article is to investigate this convergence quantitatively when S is finite.

More precisely, the framework is as follows. The whole finite state space is $\bar{S} := S \sqcup \{\infty\}$, where ∞ is the absorbing point. There is no loss of generality in assuming there is only one such point, up to lumping together all the absorbing points. Let $\bar{L}$ be the generator of the process $\bar{X}$ on $\bar{S}$, seen as a matrix $(\bar{L}(x,y))_{x,y\in\bar{S}}$. To any given probability measure m_0 on $\bar{S}$, there is a unique (in law) Markov process $\bar{X}$ whose generator is $\bar{L}$ and whose initial distribution $\mathcal{L}(\bar{X}_0)$ is m_0. For any $t \geqslant 0$, let $m_t = \mathcal{L}(\bar{X}_t)$. Using matrix notation, where measures are seen as row vectors (and functions as column vectors), we have

$$\forall\, t \geqslant 0, \qquad m_t \quad = \quad m_0\bar{P}_t$$

where $(\bar{P}_t)_{t\geqslant 0}$ is the semi-group $(\exp(t\bar{L}))_{t\geqslant 0}$ associated to $\bar{L}$. Except if m_0 is the Dirac mass on ∞, for any $t \geqslant 0$, $m_t(S) > 0$ and we can define the probability measures μ_t as the restriction to S of $m_t/m_t(S)$. They will be our main objects of interest here. By definition, we have

$$\forall\, t \geqslant 0, \forall\, f \in \mathcal{F}, \qquad \mu_t[f] \quad = \quad \frac{\mu_0[\bar{P}_t[f]]}{\mu_0[\bar{P}_t[\mathbb{1}_S]]} \tag{1.2}$$

where $\mathcal{F}$ is the space of real functions defined on S, also seen as functions defined on $\bar{S}$ which vanish at ∞ (Dirichlet condition at ∞). A probability measure ν on S is said to be a quasi-stationary measure for $\bar{L}$ if $\mu_0 = \nu$ implies that $\mu_t = \nu$ for all $t \geqslant 0$. We will recall below a convenient assumption ensuring there is a unique quasi-invariant measure ν associated to $\bar{L}$. The objective of this paper is to quantify the convergence of μ_t toward ν for large times $t \geqslant 0$, whatever the initial distribution μ_0.

For any $x \in S$, denote $V(x) = \bar{L}(x,\infty) \geqslant 0$, the killing rate at x and recall that by assumption, V does not vanish identically. The symbol V will designate the function $S \ni x \mapsto V(x)$ as well as the $S \times S$ diagonal matrix whose values on the diagonal are given by V, namely the multiplication operator by V on $\mathcal{F}$. Let L be the Markov generator on S which is such that the $S\times S$ minor of $\bar{L}$ can be written $L-V$. Our main assumption is that L is irreducible. At some point, this hypothesis will be strengthened by a reversibility assumption, in order to get more explicit results. A traditional application of the Perron-Frobenius theorem (see for instance the book [7] of Collet, Martínez and San Martín) to $L-V$ or to the associated semi-group, seen as operators on measures on S, ensures that there exists a unique quasi-invariant measure ν associated to $\bar{L}$. The probability measure ν gives a positive weight to any point of S. Furthermore there exists $\lambda_1 > 0$ such that $\nu(L-V) = -\lambda_1\nu$, λ_1 is the eigenvalue of $V-L$ which is strictly less than the real parts of the remaining eigenvalues (in $\mathbb{C}$). In the same manner, there exists a unique invariant measure η for L, charging all points of S. To see the relation between ν and η, consider the operator L^* which is adjoint to L in $\mathbb{L}^2(\eta)$. As a matrix, it is given by

$$\forall\, x,y \in S, \qquad L^*(x,y) \quad = \quad \frac{\eta(y)}{\eta(x)}L(y,x)$$

The fact that η is invariant is equivalent to the fact that L^* is a Markovian generator. We can thus apply the Perron-Frobenius theorem to $L^* - V$, seen as an operator on $\mathcal{F}$ to find a positive function φ^* on S such that $(L^* - V)[\varphi^*] = -\lambda_1\varphi^*$. Let us renormalize φ^* so that $\eta[\varphi^*] = 1$. Then $\nu = \varphi^* \cdot \eta$, the probability measure admitting the density φ^* with respect to η. Indeed, for any test function $f \in \mathcal{F}$, we have

$$\begin{aligned}(\varphi^* \cdot \eta)[(L-V)[f]] &= \eta[\varphi^*(L-V)[f]]\\ &= \eta[(L^*-V)[\varphi^*]f]\\ &= -\lambda_1\eta[\varphi^* f]\\ &= -\lambda_1(\varphi^*\cdot\eta)[f]\end{aligned}$$

so that $(\varphi^* \cdot \eta)(L - V) = -\lambda_1(\varphi^* \cdot \eta)$ and by consequence $(\varphi^* \cdot \eta)\bar{P}_t = \exp(-\lambda_1 t)(\varphi^* \cdot \eta) + (1 - \exp(-\lambda_1 t))\delta_\infty$.

This relation implies that if the process $\bar{X}$ is started from the quasi-distribution ν, then the absorption time $\tau := \inf\{t \geqslant 0 \,:\, \bar{X}_t = \infty\}$ is distributed as an exponential distribution of parameter λ_1. Indeed, we have for any $t \geqslant 0$,

$$\begin{aligned}\mathbb{P}_\nu[\tau > t] &= \nu\bar{P}_t[S]\\ &= \exp(-\lambda_1 t)\end{aligned}$$

where $\mathbb{P}_\nu$ is the underlying probability measure, when $\bar{X}_0$ is distributed according to ν. More generally, from this identity, it is not difficult to deduce that for any initial distribution m_0 not equal to δ_∞, we have

$$\lim_{t\to+\infty}\frac{\ln(\mathbb{P}_{m_0}[\tau > t])}{t} = -\lambda_1$$

showing that λ_1 is the exponential rate of absorption.

Furthermore, we can find a positive function $\varphi \in \mathcal{F}$ such that $(L-V)\varphi = -\lambda_1\varphi$, but we rather normalize it through the relation $\eta[\varphi^2] = 1$. For any positive function $f \in \mathcal{F}$, we note $f_\wedge := \min_{x\in S} f(x)$ and $f_\vee := \max_{x\in S} f(x)$.

Finally, consider the Markovian operator $\widetilde{L}$ on S which is defined by its off-diagonal entries via

$$\forall\, x \neq y \in S, \qquad \widetilde{L}(x,y) := L(x,y)\frac{\varphi(y)}{\varphi(x)} \tag{1.3}$$

(the diagonal entries are such the row sums vanish).

Let $(\widetilde{P}_t)_{t\geqslant 0}$ be the associated Markovian semi-group. Since $\widetilde{L}$ is irreducible, it admits an invariant probability $\widetilde{\eta}$. In the next section we will check that it is given by

$$\forall\, x \in S, \qquad \widetilde{\eta}(x) \;=\; \frac{\varphi(x)\varphi^*(x)\eta(x)}{\sum_{y\in S}\varphi(y)\varphi^*(y)\eta(y)} \tag{1.4}$$

To give a first estimate on the convergence of μ_t toward ν, let us recall that the total variation of a signed measure m on S satisfying $m(S) = 0$ is given equivalently by

$$\begin{aligned} \|m\|_{\mathrm{tv}} &:= 2\sup_{A\subset S} |m(A)| \\ &= \sup_{f\in\mathcal{F},\, \|f\|_\infty\leqslant 1} m(f) \\ &= \sum_{x\in S} |m(x)| \end{aligned}$$

(where as usual, $\|f\|_\infty$ designates the supremum norm of f). Note this definition differs by a factor of 2 from the probabilist version.

Theorem 1.1. — *For any probability measure μ_0 on S and for any $t \geqslant 0$, we have*

$$\frac{\varphi_\wedge}{2\varphi_\vee}\left\|\widetilde{\mu}_0\widetilde{P}_t - \widetilde{\eta}\right\|_{\mathrm{tv}} \;\leqslant\; \|\mu_t - \nu\|_{\mathrm{tv}} \;\leqslant\; 2\frac{\varphi_\vee}{\varphi_\wedge}\left\|\widetilde{\mu}_0\widetilde{P}_t - \widetilde{\eta}\right\|_{\mathrm{tv}}$$

where $\widetilde{\mu}_0$ is the probability on S whose density with respect to μ_0 is proportional to φ. In particular the asymptotic exponential rate of convergence of $\|\mu_t - \nu\|_{\mathrm{tv}}$ and $\left\|\widetilde{\mu}_0\widetilde{P}_t - \widetilde{\eta}\right\|_{\mathrm{tv}}$ are the same.

Note that in the trivial case where there is no absorption, namely $V \equiv 0$, we have $\varphi \equiv 1 \equiv \varphi^*$, $(\widetilde{P}_t)_{t\geqslant 0} = (P_t)_{t\geqslant 0}$, the Markovian semi-group generated by L, $\nu = \widetilde{\eta}$ and $\mu_t = \mu_0\widetilde{P}_t$ for all $t \geqslant 0$, so that the above bounds are optimal, up to the factor 2.

The ratio $\varphi_\vee/\varphi_\wedge$ has already appeared in the literature about absorbing Markov processes, see for instance Lemma 2.3 of the paper of Jacka and Roberts [21], where they studied the process conditioned to have never been absorbed. In a forthcoming paper, we will investigate this quantity $\varphi_\vee/\varphi_\wedge$, providing different upper bounds via path and spectral considerations. This is a first step toward the extension of the results presented here to certain denumerable chains.

Theorem 1.1 reduces the study of convergence to quasi-stationarity to the much more well-studied situation of the convergence to equilibrium. One can for instance resort to functional inequality techniques (see for instance the lecture notes of Saloff-Coste [28]), the simplest of them being the $\mathbb{L}^2$ approach. Let $\widetilde{L}^\diamond$ be the additive symmetrization of $\widetilde{L}$ in $\mathbb{L}^2(\widetilde{\eta})$: it is equal to $(\widetilde{L}+\widetilde{L}^*)/2$, where $\widetilde{L}^*$ is the adjoint operator of $\widetilde{L}$ in $\mathbb{L}^2(\widetilde{\eta})$. This self-adjointness implies that $\widetilde{L}^\diamond$ is diagonalizable in $\mathbb{R}$. Let $\widetilde{\lambda}>0$ stand for the smallest non-zero eigenvalue of $-\widetilde{L}^\diamond$. Since $\widetilde{L}^\diamond$ is irreducible, the eigenvalue 0 has multiplicity 1 (with eigenspace consisting of the constant functions) and $\widetilde{\lambda}$ is the spectral gap of $\widetilde{L}^\diamond$. Then we get:

Theorem 1.2. — *For any $t\geqslant 0$, we have*

$$\sup_{\mu_0\in\mathcal{P}}\|\mu_t-\nu\|_{\mathrm{tv}} \quad\leqslant\quad \sqrt{\frac{\eta[\varphi\varphi^*]}{(\varphi\varphi^*\eta)_\wedge}\frac{\varphi_\vee}{\varphi_\wedge}}\exp(-\widetilde{\lambda}t)$$

where $\mathcal{P}$ stands for the set of probability measures on S.

The spectral gap $\widetilde{\lambda}$ is the highest constant such that $\widetilde{\lambda}\,\widetilde{\eta}[(f-\widetilde{\eta}[f])^2]\leqslant -\widetilde{\eta}[f\widetilde{L}^\diamond[f]]$ for any $f\in\mathcal{F}$. Equivalently, $\widetilde{\lambda}=A^{-1}$, where A is the smallest positive constant such that the following Poincaré inequality is satisfied for all $f\in\mathcal{F}$,

$$\sum_{x\in S}(f(x)-\widetilde{\eta}[f])^2\,\varphi^*(x)\varphi(x)\eta(x)\leqslant$$

$$\frac{A}{2}\sum_{x,y\in S}(f(y)-f(x))^2\,\varphi^*(x)\varphi(y)\eta(x)L(x,y) \tag{1.5}$$

This variational formulation enables comparison of $\widetilde{\lambda}$ with λ (see for instance Diaconis and Saloff-Coste [12] and Fill [17]), the spectral gap of the additive symmetrization of L in $\mathbb{L}^2(\eta)$:

$$\widetilde{\lambda} \quad\geqslant\quad \frac{\varphi_\wedge\varphi^*_\wedge}{\varphi_\vee\varphi^*_\vee}\lambda \tag{1.6}$$

We will put these considerations into practice in Example 3.4.

Let us now assume that η is reversible for L. Then $-(L-V)$ is self-adjoint in $\mathbb{L}^2(\eta)$ and so is diagonalizable in $\mathbb{R}$. As was already mentioned for the general case, its smallest eigenvalue is $\lambda_1>0$. Consider its next eigenvalue $\lambda_2>\lambda_1$ (the strict inequality is a consequence of the irreducibility of L in the Perron-Frobenius theorem). The next result shows that to get a useful understanding of the convergence of μ_t toward ν for large $t\geqslant 0$, only the knowledge of η, of the ratio of the extrema of φ and of $\lambda_2-\lambda_1$ is required.

Theorem 1.3. — *Under the reversibility assumption, for any $t \geqslant 0$, we have*

$$\begin{aligned}\sup_{\mu_0\in\mathcal{P}} \|\mu_t - \nu\|_{\mathrm{tv}} &\leqslant \sqrt{\frac{1}{(\varphi^2\eta)_\wedge}\frac{\varphi_\vee}{\varphi_\wedge}}\exp(-(\lambda_2-\lambda_1)t)\\ &\leqslant \sqrt{\frac{1}{\eta_\wedge}\left(\frac{\varphi_\vee}{\varphi_\wedge}\right)^2}\exp(-(\lambda_2-\lambda_1)t)\end{aligned}$$

Note that (1.2) can be written in terms of Feynman-Kac integrals. Let $(X_t)_{t\geqslant 0}$ be a Markov process starting from the initial law μ_0 and admitting L as generator. We have

$$\forall\, t \geqslant 0, \forall\, f \in \bar{\mathcal{F}}, \qquad \mu_t[f] \;=\; \frac{\mathbb{E}_{\mu_0}\left[f(X_t)\exp\left(-\int_0^t V(X_s)\,ds\right)\right]}{\mathbb{E}_{\mu_0}\left[\exp\left(-\int_0^t V(X_s)\,ds\right)\right]}$$

The stability for large times of such expressions have been extensively studied by Del Moral and his coauthors (see for instance his recent book [9] and the references given there). They also use estimates on the convergence to equilibrium of Markov processes. Since their assumptions are based on Dobrushin type conditions on the underlying Markov process (or on some of its modifications, see e.g. Del Moral and Miclo [10]), the deduced bounds are often quite coarse. While we work here in the same spirit, we will rather resort to spectral techniques, which lead to relatively sharp estimates, as will be illustrated by several examples. In particular, we obtain in the reversible case the optimal asymptotical rate $\lambda_2-\lambda_1$ (see e.g. the review paper of Méléard and Villemonais [24], with a non-quantified pre-exponential factor). Under appropriate conditions, this rate was deduced asymptotically for birth and death processes by van Doorn [29] (see also van Doorn and Zeifman [31] for another example), which are outside the scope of the present note, because the state space is not finite. We hope that in a future work, we will be able to extend the above quantitative bounds to more general situations of appropriate denumerable Markov processes or diffusions, requiring at least the condition there is a unique quasi-invariant measure (usually this requires that the process comes in from infinity fast enough, see for instance Collet, Martínez and San Martín [7]). For Brownian motion absorbed on the boundary of a compact domain in Euclidean spaces, one may see Gyrya and Saloff-Coste [18] and Lierl and Saloff-Coste [23]

The literature on quasi-stationarity is substantial and we are able to call on several comprehensive surveys. One short readable survey, close in spirit to our paper, is by Van Doorn and Pollett [30] (discrete state space,

continuous time). More general state spaces and applications in biology are emphasized by Méléard and Villemonais [24]. A recent book length treatment by Collet, Martínez and San Martín [7] treats all aspects. All of these review the history (Yaglom, Bartlett, Darroch-Seneta, ...). A most useful adjunct to these surveys is the annotated online bibliography kept up to date by Phil Pollett, see
`//http://www.maths.uq.edu.au/~pkp/papers/qsds/qsds.html`.

We have not found very much literature on the kind of quantitative questions treated here. A useful review of previous quantitative efforts is in Section 4 of Van Doorn and Pollett [30]. This is along the lines of spectral gap estimates without consideration of the size of the state space or the starting distribution. Some quantitative bounds are also deduced in the recent papers of Barbour and Pollett [2, 3], of Cloez and Thai [6] and of Champagnat and Villemonais [5].

The plan of the paper is very simple: the next section presents the proof of the above theorems, as well as an alternative bound based on logarithmic Sobolev inequalities, the next section contains some illustrative examples. The final section gives further examples in discrete time.

2. Proofs

The following arguments are based on a simple use of Doob's transforms, which by a conjugation by φ, replace V by a constant killing rate.

2.1. Proof of Theorem 1.1

Let Φ be the diagonal matrix corresponding to the multiplication by φ operating on $\mathcal{F}$. Thus Φ^{-1} is just the diagonal matrix corresponding to the multiplication by $1/\varphi$. We begin by checking that the generator matrix $\widetilde{L}$ defined in (1.3) satisfies

$$\widetilde{L} \quad = \quad \Phi^{-1}(L - V + \lambda_1 I)\Phi \tag{2.1}$$

where I is the identity matrix. Indeed, the off-diagonal entries of the r.h.s. coincide with those of $\Phi^{-1}L\Phi$ which are those of $\widetilde{L}$ by (1.3). Thus it is sufficient to check that the sums of the rows of $\Phi^{-1}(L - V + \lambda_1 I)\Phi$ vanish. The sum corresponding to the row indexed by $x \in S$ is

$$\frac{1}{\varphi(x)}(L[\varphi](x) - V(x)\varphi(x)) + \lambda_1 \quad = \quad 0$$

since by definition, φ is an eigenfunction of $L-V$ associated to the eigenvalue $-\lambda_1$.

It is now easy to check (1.4): it must be seen that

$$\forall\, f \in \mathcal{F}, \qquad \widetilde{\eta}[\widetilde{L}[f]] \;=\; 0$$

From (2.1), the l.h.s. is equal to

$$\begin{aligned}
\widetilde{\eta}[\varphi^{-1}(L-V+\lambda_1)[\varphi f]] &= \eta[\varphi^*(L-V+\lambda_1)[\varphi f]]/\eta[\varphi\varphi^*] \\
&= \eta[\varphi f(L^*-V+\lambda_1)[\varphi^*]]/\eta[\varphi\varphi^*] \\
&= 0
\end{aligned}$$

because φ^* is an eigenfunction of L^*-V associated to the eigenvalue $-\lambda_1$.

Next rewrite (2.1) in the form

$$\Phi(\widetilde{L}-\lambda_1 I)\Phi^{-1} \;=\; L-V \tag{2.2}$$

and exponentiate this identity to find

$$\forall\, t \geqslant 0, \qquad \exp(-\lambda_1 t)\Phi\widetilde{P}_t\Phi^{-1} \;=\; \bar{P}_t$$

(the r.h.s. is to be understood as the restriction of $\bar{P}_t$ to $\mathcal{F}$, as explained after (1.2)). Thus for any $\mu_0 \in \mathcal{P}$ and $f \in \mathcal{F}$, we have

$$\forall\, t \geqslant 0, \qquad \exp(-\lambda_1 t)\mu_0[\varphi]\widetilde{\mu}_0[\widetilde{P}_t[f/\varphi]] \;=\; \mu_0[\bar{P}_t[f]]$$

(recall that $\widetilde{\mu}_0$ is the probability on S whose density with respect to μ_0 is proportional to φ). We deduce from (1.2) that

$$\forall\, t \geqslant 0, \qquad \mu_t[f] \;=\; \frac{\widetilde{\mu}_0[\widetilde{P}_t[f/\varphi]]}{\widetilde{\mu}_0[\widetilde{P}_t[1/\varphi]]} \tag{2.3}$$

Since $\widetilde{P}_t$ converges to $\widetilde{\eta}$ as t goes to infinity, we get that

$$\begin{aligned}
\lim_{t\to+\infty} \mu_t[f] &= \frac{\widetilde{\eta}[f/\varphi]}{\widetilde{\eta}[1/\varphi]} \\
&= \nu[f]
\end{aligned}$$

due to the proportionality between the measures ν, $\varphi^*\cdot\eta$ and $\varphi^{-1}\cdot\widetilde{\eta}$. Thus the convergence toward quasi-stationarity has been recovered.

To get an estimate on the speed of convergence, we need the two following basic lemmas.

On a general measurable space, consider two probability measures $\widetilde{\mu} \ll \widetilde{\nu}$, as well as a measurable function $\psi > 0$. Define

$$\begin{aligned}
\mu &:= \frac{\psi}{Z_{\widetilde{\mu}}}\cdot\widetilde{\mu} \quad \text{with} \quad Z_{\widetilde{\mu}} := \widetilde{\mu}[\psi] \\
\nu &:= \frac{\psi}{Z_{\widetilde{\nu}}}\cdot\widetilde{\nu} \quad \text{with} \quad Z_{\widetilde{\nu}} := \widetilde{\nu}[\psi]
\end{aligned}$$

Let $\widetilde{f}$ and f stand for the Radon-Nikodym densities of $\widetilde{\mu}$ with respect to $\widetilde{\nu}$ and of μ with respect to ν. Obviously, we have

$$f = \frac{Z_{\widetilde{\nu}}}{Z_{\widetilde{\mu}}}\widetilde{f}$$

Finally, choose $\widetilde{m}$ and m to be medians of $\widetilde{f}$ and f with respect to $\widetilde{\nu}$ and ν. The following result is well-known.

Lemma 2.1. — *We have*

$$\int |f-m|\,d\nu \leqslant \|\mu-\nu\|_{\mathrm{tv}} \leqslant 2\int |f-m|\,d\nu$$
$$\int \left|\widetilde{f}-\widetilde{m}\right|\,d\widetilde{\nu} \leqslant \|\widetilde{\mu}-\widetilde{\nu}\|_{\mathrm{tv}} \leqslant 2\int \left|\widetilde{f}-\widetilde{m}\right|\,d\widetilde{\nu}$$

Proof. — Of course it is sufficient to show the bounds for $\|\mu-\nu\|_{\mathrm{tv}}$. They are a consequence of

$$\|\mu-\nu\|_{\mathrm{tv}} = \int |f-1|\,d\nu$$

and of the following characterization of a median:

$$\int |f-m|\,d\nu = \inf\left\{\int |f-r|\,d\nu \,:\, r\in\mathbb{R}\right\} \tag{2.4}$$

So the lower bound is immediate and for the upper bound, just note that

$$\begin{aligned}|1-m| &= \left|\int (f-m)\,d\nu\right| \\ &\leqslant \int |f-m|\,d\nu\end{aligned}$$

□

The interest of the introduction of the medians comes from:

Lemma 2.2. — *We have*

$$\int |f-m|\,d\nu \leqslant \frac{\psi_\wedge}{\psi_\vee}\int \left|\widetilde{f}-\widetilde{m}\right|\,d\widetilde{\nu}$$

and it follows from the previous lemma that

$$\frac{\psi_\vee}{2\psi_\wedge}\|\widetilde{\mu}-\widetilde{\nu}\|_{\mathrm{tv}} \leqslant \|\mu-\nu\|_{\mathrm{tv}} \leqslant 2\frac{\psi_\wedge}{\psi_\vee}\|\widetilde{\mu}-\widetilde{\nu}\|_{\mathrm{tv}}$$

Proof. — From (2.4), we have

$$\begin{aligned}
\int \left|\widetilde{f}-\widetilde{m}\right| d\widetilde{\nu} &= \inf\left\{\int \left|\widetilde{f}-r\right| d\widetilde{\nu} \,:\, r\in\mathbb{R}\right\}\\
&= \inf\left\{\int \left|\frac{Z_{\widetilde{\mu}}}{Z_{\widetilde{\nu}}}f-r\right| d\widetilde{\nu} \,:\, r\in\mathbb{R}\right\}\\
&= \frac{Z_{\widetilde{\mu}}}{Z_{\widetilde{\nu}}}\inf\left\{\int |f-r|\,\frac{Z_{\widetilde{\nu}}}{\psi}\,d\nu \,:\, r\in\mathbb{R}\right\}\\
&\geqslant \frac{Z_{\widetilde{\mu}}}{Z_{\widetilde{\nu}}}\frac{Z_{\widetilde{\nu}}}{\mathrm{esssup}_{\widetilde{\nu}}\psi}\inf\left\{\int |f-r|\,d\nu \,:\, r\in\mathbb{R}\right\}\\
&= \frac{Z_{\widetilde{\nu}}}{\mathrm{esssup}_{\widetilde{\nu}}\psi}\int |f-m|\,d\nu\\
&\geqslant \frac{\mathrm{essinf}_{\widetilde{\mu}}\psi}{\mathrm{esssup}_{\widetilde{\nu}}\psi}\int |f-m|\,d\nu\\
&\geqslant \frac{\psi_\wedge}{\psi_\vee}\int |f-m|\,d\nu
\end{aligned}$$

□

For any fixed $t\geqslant 0$, it remains to apply these general bounds with ν the quasi-stationary probability measure, $\widetilde{\nu}:=\widetilde{\eta}$

$$\begin{aligned}
\mu &:= \mu_t\\
\widetilde{\mu} &:= \widetilde{\mu}_0\widetilde{P}_t\\
\psi &:= 1/\varphi
\end{aligned}$$

Since $\psi_\wedge/\psi_\vee=\varphi_\vee/\varphi_\wedge$, the conclusion of Lemma 2.2 implies the wanted bound.

Remark 2.3. — From (2.3), we could have been tempted to write that for any $f\in\mathcal{F}$,

$$\begin{aligned}
\mu_t[f]-\nu[f] &= \frac{\widetilde{\mu}_0[\widetilde{P}_t[f/\varphi]]}{\widetilde{\mu}_0[\widetilde{P}_t[1/\varphi]]}-\frac{\widetilde{\eta}[f/\varphi]}{\widetilde{\eta}[1/\varphi]}\\
&= \frac{1}{\widetilde{\mu}_0[\widetilde{P}_t[1/\varphi]]}(\widetilde{\mu}_0[\widetilde{P}_t[f/\varphi]]-\widetilde{\eta}[f/\varphi])\\
&\qquad\qquad +\frac{\widetilde{\eta}[f/\varphi]}{\widetilde{\eta}[1/\varphi]\widetilde{\mu}_0[\widetilde{P}_t[1/\varphi]]}(\widetilde{\eta}[1/\varphi]-\widetilde{\mu}_0[\widetilde{P}_t[1/\varphi]])\\
&\leqslant \varphi_\vee\left|\widetilde{\mu}_0[\widetilde{P}_t[f/\varphi]]-\widetilde{\eta}[f/\varphi]\right|+\frac{\varphi_\vee^2}{\varphi_\wedge}\,\|f\|_\infty\left|\widetilde{\eta}[1/\varphi]-\widetilde{\mu}_0[\widetilde{P}_t[1/\varphi]])\right|
\end{aligned}$$

Taking the supremum of $f \in \mathcal{F}$ satisfying $\|f\|_\infty \leqslant 1$, it appears that

$$\begin{aligned}
\|\mu_t - \nu\|_{\mathrm{tv}} &\leqslant \left(\left(\frac{\varphi_\vee}{\varphi_\wedge}\right) + \left(\frac{\varphi_\vee}{\varphi_\wedge}\right)^2\right) \left\|\widetilde{\mu}_0 \widetilde{P}_t - \widetilde{\eta}\right\|_{\mathrm{tv}} \\
&\leqslant 2\left(\frac{\varphi_\vee}{\varphi_\wedge}\right)^2 \left\|\widetilde{\mu}_0 \widetilde{P}_t - \widetilde{\eta}\right\|_{\mathrm{tv}}
\end{aligned}$$

which is worse than the bound of Theorem 1.1 by a factor $\varphi_\vee/\varphi_\wedge$.

2.2. Proof of Theorem 1.2

Since Theorem 1.1 brings us back to the situation of convergence to equilibrium of Markov processes, it is sufficient to use the argument of Fill [17]) for non-reversible processes. We recall them below for the sake of completeness.

To gain a factor 2, it is in fact better not to use Theorem 1.1, but to directly make a comparison between $\mathbb{L}^2$ quantities. More precisely, for given $\mu_0 \in \mathcal{P}$ and $t \geqslant 0$, denote by f_t (respectively $\widetilde{f}_t$) the density of the probability μ_t with respect to ν (resp. $\widetilde{\mu}_t := \widetilde{\mu}_0 \widetilde{P}_t$ with respect to $\widetilde{\eta}$). We have by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\|\mu_t - \nu\|_{\mathrm{tv}} &= \sum_{x \in S} |f_t(x) - 1|\, \nu(x) \\
&\leqslant \sqrt{I_t}
\end{aligned}$$

where

$$I_t := \sum_{x \in S} (f_t(x) - 1)^2\, \nu(x)$$

Let us also define

$$\widetilde{I}_t := \sum_{x \in S} (\widetilde{f}_t(x) - 1)^2\, \widetilde{\eta}(x)$$

It is easy to compare these quantities:

Lemma 2.4. — *For any $t \geqslant 0$, we have*

$$I_t \leqslant \left(\frac{\varphi_\vee}{\varphi_\wedge}\right)^2 \widetilde{I}_t$$

Proof. — One recognizes in I_t the variance of f_t with respect to ν, so that

$$I_t \quad = \quad \inf\left\{\sum_{x\in S}(f_t(x)-r)^2\,\nu(x)\ :\ r\in\mathbb{R}\right\}$$

Similarly we have

$$\widetilde{I}_t \quad = \quad \inf\left\{\sum_{x\in S}(\widetilde{f}_t(x)-r)^2\,\widetilde{\eta}(x)\ :\ r\in\mathbb{R}\right\}$$

The same arguments as those used in Lemma 2.2 give the conclusion without difficulty. □

Putting together these estimates, we end up with

$$\|\mu_t-\nu\|_{\mathrm{tv}} \quad \leqslant \quad \frac{\varphi_\vee}{\varphi_\wedge}\sqrt{\widetilde{I}_t} \tag{2.5}$$

To study the evolution of $\widetilde{I}_t$ with respect to the time $t\geqslant 0$, recall that

$$\forall\ x\in S,\ \forall\ t\geqslant 0,\qquad \partial_t\widetilde{f}_t(x) \quad = \quad \widetilde{L}^*[\widetilde{f}_t](x)$$

This comes from the relation $\widetilde{f}_t=\widetilde{P}_t^*[\widetilde{f}_0]$, where $\widetilde{P}_t^*$ the adjoint operator of $\widetilde{P}_t$ in $\mathbb{L}^2(\widetilde{\eta})$. Thus we get that for all $t\geqslant 0$,

$$\begin{aligned}
\partial_t\widetilde{I}_t &= 2\widetilde{\eta}[(\widetilde{f}_t-1)\partial_t\widetilde{f}_t]\\
&= 2\widetilde{\eta}[(\widetilde{f}_t-1)\widetilde{L}^*[\widetilde{f}_t]]\\
&= 2\widetilde{\eta}[(\widetilde{f}_t-1)\widetilde{L}^*[\widetilde{f}_t-1]]\\
&= 2\widetilde{\eta}[\widetilde{L}[\widetilde{f}_t-1](\widetilde{f}_t-1)]\\
&= 2\widetilde{\eta}[\widetilde{L}^\diamond[\widetilde{f}_t-1](\widetilde{f}_t-1)]
\end{aligned} \tag{2.6}$$

By definition of $\widetilde{\lambda}$, the r.h.s. is bounded above by $-2\widetilde{\lambda}\widetilde{I}_t$, which leads to the ordinary differential inequality

$$\forall\ t\geqslant 0,\qquad \partial_t\widetilde{I}_t \quad \leqslant \quad -2\widetilde{\lambda}\widetilde{I}_t$$

Gronwall's lemma implies that

$$\forall\ t\geqslant 0,\qquad \widetilde{I}_t \quad \leqslant \quad \exp(-2\widetilde{\lambda}t)\widetilde{I}_0 \tag{2.7}$$

so it remains to bound $\widetilde{I}_0$ above. But note that

$$\begin{aligned}
\widetilde{I}_0 &= \widetilde{\eta}[\widetilde{f}_0^2]-1\\
&\leqslant \widetilde{\eta}[\widetilde{f}_0^2]\\
&= \widetilde{\mu}_0[\widetilde{f}_0]\\
&\leqslant \left\|\widetilde{f}_0\right\|_\infty\\
&\leqslant \frac{1}{\widetilde{\eta}_\wedge}\\
&= \frac{\eta[\varphi\varphi^*]}{(\varphi\varphi^*\eta)_\wedge}
\end{aligned}$$

which, in conjunction with (2.5) and (2.7), lead to the bound of Theorem 1.2.

From

$$\begin{aligned}
\forall\, x\neq y,\qquad \widetilde{L}^*(x,y) &= \frac{\widetilde{\eta}(y)}{\widetilde{\eta}(x)}\widetilde{L}(y,x)\\
&= \frac{\varphi^*(y)\eta(y)}{\varphi^*(x)\eta(x)}L(y,x) \qquad (2.8)
\end{aligned}$$

we deduce that the matrix of the Markov generator $\widetilde{L}^\diamond$ is described by its off-diagonal entries:

$$\forall\, x\neq y\in S,\qquad \widetilde{L}^\diamond(x,y) = \frac{1}{2}\left(L(y,x)\frac{\varphi^*(y)\eta(y)}{\varphi^*(x)\eta(x)}+L(x,y)\frac{\varphi(y)}{\varphi(x)}\right)$$

The Poincaré formulation (1.5) then comes from the variational characterization of the eigenvalues and from the equality

$$\forall\, g\in\mathcal{F},\qquad \widetilde{\eta}[g\widetilde{L}^\diamond[g]] = \widetilde{\eta}[g\widetilde{L}[g]]$$

already used in (2.6).

Remark 2.5. — Similarly to the lower bound in Theorem 1.1, we have also in Lemma 2.4

$$\forall\, t\geqslant 0,\qquad I_t \geqslant \left(\frac{\varphi_\wedge}{\varphi_\vee}\right)^2\widetilde{I}_t$$

In particular $\sqrt{I_t}$ and $\sqrt{\widetilde{I}_t}$ have the same asymptotic exponential rate of convergence. This common rate is the smallest real part of the non-zero eigenvalues of $-\widetilde{L}$, but since this operator is not assumed to be reversible, this rate may be larger than $\widetilde{\lambda}$.

Remark 2.6. — It is possible to improve the pre-exponential factor $\sqrt{\frac{\eta[\varphi\varphi^*]}{(\varphi\varphi^*\eta)_\wedge}\frac{\varphi_\vee}{\varphi_\wedge}}$ in Theorem 1.2, but at the expense of the rate $\widetilde{\lambda}$, via the logarithmic Sobolev inequalities associated to the symmetrization $\widetilde{L}^\diamond$ of $\widetilde{L}$.

Let $\widetilde{\alpha} > 0$ be the largest constant such that for all $g \in \mathcal{F}$,

$$\widetilde{\alpha}\sum_{x\in S} g^2(x)\ln\left(\frac{g^2(x)}{\widetilde{\eta}[g^2]}\right)\varphi^*(x)\varphi(x)\eta(x) \leqslant$$
$$\sum_{x,y\in S}(g(y)-g(x))^2\,\varphi^*(x)\varphi(y)\eta(x)L(x,y) \qquad (2.9)$$

Then we have

$$\sup_{\mu_0\in\mathcal{P}}\|\mu_t-\nu\|_{\mathrm{tv}} \ \leqslant\ \sqrt{2\ln\left(\frac{\eta[\varphi\varphi^*]}{(\varphi\varphi^*\eta)_\wedge}\right)\frac{\varphi_\vee}{\varphi_\wedge}}\exp(-(\widetilde{\alpha}/2)t) \qquad (2.10)$$

The proof of this bound has the same structure as the one of Theorem 1.2, with the quantities I_t and $\widetilde{I}_t$ replaced by the relative entropies

$$\begin{aligned} J_t &:= \sum_{x\in S} f_t(x)\ln(f_t(x))\,\nu(x)\\ \widetilde{J}_t &:= \sum_{x\in S}\widetilde{f}_t(x)\ln(\widetilde{f}_t(x))\,\widetilde{\eta}(x)\end{aligned}$$

Indeed, Pinsker's inequality gives the bound

$$\|\mu_t-\nu\|_{\mathrm{tv}} \ \leqslant\ \sqrt{2}\sqrt{J_t}$$

Next, taking into account the relations (see Holley and Stroock [19])

$$J_t = \inf\left\{\sum_{x\in S}(f_t(x)\ln(f_t(x)) - f_t(x)\ln(r) - f_t(x)+r)\,\nu(x)\ :\ r\in\mathbb{R}_+\right\}$$
$$\widetilde{J}_t = \inf\left\{\sum_{x\in S}(\widetilde{f}_t(x)\ln(\widetilde{f}_t(x)) - \widetilde{f}_t(x)\ln(r) - \widetilde{f}_t(x)+r)\,\widetilde{\eta}(x)\ :\ r\in\mathbb{R}_+\right\}$$

we deduce as in Lemma 2.2 that

$$\forall\ t\geqslant 0,\qquad \frac{\varphi_\wedge}{\varphi_\vee}\widetilde{J}_t\ \leqslant\ J_t\ \leqslant\ \frac{\varphi_\vee}{\varphi_\wedge}\widetilde{J}_t$$

As a consequence, we get

$$\forall\ t \geqslant 0, \qquad \|\mu_t - \nu\|_{\mathrm{tv}} \ \leqslant\ \sqrt{2\frac{\varphi_\vee}{\varphi_\wedge}}\sqrt{\widetilde{J}_t}$$

which reduces our task to the investigation of the time evolution of $\widetilde{J}_t$.

By differentiation, it appears that

$$\begin{aligned}
\partial_t \widetilde{J}_t &= \sum_{x\in S}(1+\ln(\widetilde{f}_t(x)))\partial_t \widetilde{f}_t(x)\,\widetilde{\eta}(x) \\
&= \sum_{x\in S}(1+\ln(\widetilde{f}_t(x)))\widetilde{L}^*[\widetilde{f}_t(x)]\,\widetilde{\eta}(x) \\
&= \sum_{x,y\in S}(1+\ln(\widetilde{f}_t(x)))(\widetilde{f}_t(y)-\widetilde{f}_t(x))\,\widetilde{\eta}(x)\widetilde{L}^*(x,y)
\end{aligned}$$

To proceed, note (cf. for instance Miclo [25]) that for all $x, y \in S$,

$$(1+\ln(\widetilde{f}_t(x)))(\widetilde{f}_t(y)-\widetilde{f}_t(x)) \leqslant \widetilde{f}_t(y)\ln(\widetilde{f}_t(y)) - \widetilde{f}_t(x)\ln(\widetilde{f}_t(x)) - \left(\sqrt{\widetilde{f}_t(y)}-\sqrt{\widetilde{f}_t(x)}\right)^2$$

and that by invariance of $\widetilde{\eta}$ with respect to $\widetilde{L}^*$,

$$\sum_{x,y\in S}\left(\widetilde{f}_t(y)\ln(\widetilde{f}_t(y)) - \widetilde{f}_t(x)\ln(\widetilde{f}_t(x))\right)\widetilde{\eta}(x)\widetilde{L}^*(x,y) \ =\ 0$$

Thus we end up with the refined Jensen type bound:

$$\begin{aligned}
\partial_t \widetilde{J}_t &\leqslant -\sum_{x,y\in S}\left(\sqrt{\widetilde{f}_t(y)}-\sqrt{\widetilde{f}_t(x)}\right)^2\widetilde{\eta}(x)\widetilde{L}^*(x,y) \\
&= -\sum_{x,y\in S}\left(\sqrt{\widetilde{f}_t(y)}-\sqrt{\widetilde{f}_t(x)}\right)^2\widetilde{\eta}(y)\widetilde{L}^*(y,x) \\
&= -\sum_{x,y\in S}\left(\sqrt{\widetilde{f}_t(y)}-\sqrt{\widetilde{f}_t(x)}\right)^2\frac{\varphi^*(x)\varphi(y)\eta(x)}{\eta[\varphi^*\varphi]}L(x,y)
\end{aligned}$$

where we used (2.8). The logarithmic Sobolev inequality (2.9), with $g := f_t$, allows comparison of the r.h.s. with $\widetilde{J}_t$ to give the differential inequality

$$\forall\ t \geqslant 0, \qquad \partial_t \widetilde{J}_t \ \leqslant\ -\widetilde{\alpha}\widetilde{J}_t$$

Gronwall's lemma implies again that

$$\begin{array}{rcl}\forall\, t\geqslant 0, \qquad \widetilde{J}_t & \leqslant & \exp(-\widetilde{\alpha}t)\widetilde{J}_0\\ & \leqslant & \exp(-\widetilde{\alpha}t)\ln((\widetilde{f}_0)_\vee)\\ & \leqslant & \exp(-\widetilde{\alpha}t)\ln(1/\widetilde{\eta}_\wedge)\end{array}$$

The announced bound (2.10) follows.

Despite the deterioration of exponential rate in (2.10), this bound can be interesting for not too large times $t\geqslant 0$, especially when one looks for "quasi-mixing times". Diaconis and Saloff-Coste [13] have shown the following general bound between the logarithmic Sobolev constant $\widetilde{\alpha}$ and the spectral gap $\widetilde{\lambda}$:

$$\widetilde{\alpha} \quad\geqslant\quad \frac{1-2\widetilde{\eta}_\wedge}{\ln(1/\widetilde{\eta}_\wedge-1)}\widetilde{\lambda} \tag{2.11}$$

(where the factor on the r.h.s. is taken to be $1/2$ in the particular case where $\widetilde{\eta}_\wedge=1/2$). But this relation is not very pertinent for quasi-mixing times estimates: if $\tau_{\widetilde{\lambda}}\geqslant 0$ and $\tau_{\widetilde{\alpha}}\geqslant 0$ are the times $t\geqslant 0$ in Theorem 1.2 and (2.10) such that the corresponding upper bounds are equal to 1, we get

$$\begin{array}{rcl}\tau_{\widetilde{\lambda}} & = & \frac{1}{\widetilde{\lambda}}(\ln(\varphi_\vee/\varphi_\wedge)+\ln(1/\widetilde{\eta}_\wedge)/2)\\ \tau_{\widetilde{\alpha}} & = & \frac{1}{\widetilde{\alpha}}(\ln(\varphi_\vee/\varphi_\wedge)+\ln(\ln(1/\widetilde{\eta}_\wedge))+\ln(2))\end{array}$$

and the injection of (2.11) leads to the disappointing $\tau_{\widetilde{\lambda}}\ll\tau_{\widetilde{\alpha}}$ for small $\widetilde{\eta}_\wedge>0$. Indeed, the interest of (2.10) appears when one has good estimates on $\widetilde{\alpha}$ (by tensorization for instance) and $\widetilde{\eta}_\wedge$ is very small. Simple examples on product spaces are provided in Subsection 3.5. Nevertheless, we believe that modified logarithmic Sobolev inequalities (see e.g. the article of Bobkov and Tetali [4]), namely the consideration of the best constant $\widehat{\alpha}>0$ such that for all $g\in\mathcal{F}$,

$$\begin{array}{l}\widehat{\alpha}\displaystyle\sum_{x\in S}g^2(x)\ln\left(\frac{g^2(x)}{\widetilde{\eta}[g^2]}\right)\varphi^*(x)\varphi(x)\eta(x)\\ \qquad\leqslant\displaystyle\sum_{x,y\in S}(|g(y)|-|g(x)|)(\ln(|g(y)|)-\ln(|g(x)|))\,\varphi^*(x)\varphi(y)\eta(x)L(x,y)\end{array}$$

is better suited to the above entropic approach.

2.3. Proof of Theorem 1.3

Under the assumption that ν is reversible for L, we have that $L^*=L$. The equations for φ and φ^* are thus the same and only the corresponding renormalizations are different. If follows that φ and φ^* are proportional and since only ratios enter the pre-exponential factor of Theorem 1.2, it can be replaced by the pre-exponential factor of Theorem 1.3 (recall the normalization $\eta[\varphi^2] = 1$).

But the main advantage of Theorem 1.3 is the explicit rate $\lambda_2 - \lambda_1$. It is a consequence of the conjugacy relation (2.2). It shows first that $\widetilde{L}$ must be reversible with respect to $\widetilde{\eta}$ (but this can also be checked directly from the expressions (1.3) and (1.4)) and second that the spectrum of $\widetilde{L}$ is obtained from the spectrum of $L - V$ by subtracting the value λ_1. In particular the spectral gap $\widetilde{\lambda}$ of $\widetilde{L}^\diamond = \widetilde{L}$ is equal to $\lambda_2 - \lambda_1$.

Remarks 2.7. — (a) The fact that the spectrum of $\widetilde{L}$ is obtained from the spectrum of $L-V$ by subtracting the value λ_1 is always true, but in the non-reversible case it is not clear how to use this possibly complex valued spectrum to deduce a bound on $\widetilde{\lambda}$. In the reversible situation Remark 2.5 can be made more precise: the common asymptotic exponential rate of $\sqrt{I_t}$ and $\sqrt{\widetilde{I}_t}$ is $\lambda_2 - \lambda_1$.

(b) The logarithmic Sobolev inequality approach is equally valid in the reversible case, we get

$$\sup_{\mu_0 \in \mathcal{P}} \|\mu_t - \nu\|_{\mathrm{tv}} \quad \leqslant \quad \sqrt{2 \ln \left(\frac{1}{(\varphi^2 \eta)_\wedge} \right) \frac{\varphi_\vee}{\varphi_\wedge}} \exp(-(\widetilde{\alpha}/2)t)$$

where $\widetilde{\alpha}$ is the logarithmic Sobolev constant associated to the symmetric operator $\widetilde{L}$ in $\widetilde{\eta}$ (in particular (2.11) is satisfied with $\widetilde{\lambda}$ replaced by $\lambda_2 - \lambda_1$).

3. Examples

Several basic examples are provided here, which in particular serve to illustrate some assertions made in the previous theoretical developments.

3.1. A finite birth and death example with $\lambda_1 \approx \lambda_2 - \lambda_1$

This example and the next two are birth and death processes on $\bar{S} := [\![0, N]\!]$, with $N \in \mathbb{N}$, absorbed in 0. So $S = [\![1, N]\!]$, $\infty = 0$ and L gives positive rates only to the oriented edges $(x, x+1)$ and $(x+1, x)$ where $x \in [\![1, N-1]\!]$. In this one-dimensional setting, L admits a unique reversible probability η. Let us assume that the killing rate at 1 is 1, namely $V(1) = \bar{L}(1, 0) = 1$. The other values of V are taken to be zero.

Specifically for this example, we choose

$$\forall\, x \in [\![1, N-2]\!], \qquad L(x, x+1) \;:=\; L(x+1, x) \;:=\; 1 \tag{3.1}$$

$$L(N-1, N) \;=\; 1 \text{ and } L(N, N-1) \;=\; 2 \tag{3.2}$$

(the value 2 simplifies the analysis of the reflection at N by replacing the forbidden jump to $N+1$ by a supplementary jump at $N-1$). The reversible probability η is then given by

$$\forall\, x \in S, \qquad \eta(x) \;=\; \begin{cases} \frac{2}{2N-1} & \text{, if } x \in [\![N-1]\!] \\ \frac{1}{2N-1} & \text{, if } x = N \end{cases}$$

Let φ be the function defined by

$$\forall\, x \in S, \qquad \varphi(x) \;:=\; \frac{1}{Z}\sin(\pi x/(2N)) \tag{3.3}$$

where Z is the renormalization constant such that $\eta[\varphi^2] = 1$. Due to the value 2 in (3.2), it is easy to check that $(L-V)[\varphi] = 2(\cos(\pi/(2N)) - 1)\varphi$. The positivity of φ and Perron-Frobenius theorem imply that φ is indeed the function considered in the introduction and that

$$\lambda_1 \;=\; 2(1 - \cos(\pi/(2N)))$$

The density of the quasi-invariant probability measure ν with respect to η is proportional to φ.

More generally, define for $k \in [\![1, N-1]\!]$, the function φ_k by

$$\forall\, x \in S, \qquad \varphi_k(x) \;:=\; \sin((2k+1)\pi x/(2N))$$

By straightforward calculation, $(L-V)[\varphi_k] = 2(\cos((2k+1)\pi/(2N)) - 1)\varphi_k$. Thus the spectrum of $L-V$ is $\{2(\cos((2k+1)\pi/(2N)) - 1) \,:\, k \in [\![0, N-1]\!]\}$. In particular

$$\begin{aligned} \lambda_2 - \lambda_1 &= 2(\cos(\pi/(2N)) - \cos(3\pi/(2N)) \\ &= 4\sin(\pi/N)\sin(\pi/(2N)) \\ &= 2\frac{\pi^2}{N^2}(1 + \mathcal{O}(N^{-2})) \end{aligned}$$

as N goes to infinity. Since

$$\lambda_1 \;=\; \frac{\pi^2}{4N^2}(1 + \mathcal{O}(N^{-2}))$$

in this situation λ_1 and $\lambda_2 - \lambda_1$ are of the same order, meaning that absorption and convergence to quasi-stationarity happen at similar rates.

From (3.3), we deduce that

$$\begin{aligned}\frac{\varphi_\vee}{\varphi_\wedge} &= \frac{1}{\sin(\pi/(2N))}\\ &= \frac{2N}{\pi}(1+\mathcal{O}(N^{-2}))\end{aligned}$$

Taking into account the classical Riemann sum approximation, we furthermore get

$$\begin{aligned}Z^2 &= \frac{2}{2N-1}\sum_{x\in[\![1,N-1]\!]}\sin^2(\pi x/(2N))+\frac{1}{2N-1}\\ &= (1+\mathcal{O}(N^{-1}))\int_0^1 \sin^2(\pi u/2)\,du\\ &= \frac{1}{2}(1+\mathcal{O}(N^{-1}))\end{aligned}$$

The first bound of Theorem 1.3 asserts that

$$\sup_{\mu_0\in\mathcal{P}}\|\mu_t-\nu\|_{\mathrm{tv}} \leqslant \frac{4}{\pi^2}N^{5/2}\exp\left(-\frac{2\pi^2}{N^2}t(1+\mathcal{O}(N^{-2}))\right)(1+\mathcal{O}(N^{-1}))$$

(the second bound of Theorem 1.3, which doesn't need the estimate on Z, leads to a similar bound with 4 replaced by $4\sqrt{2}$). It follows that for any given $s>0$, if

$$t = \frac{5}{4\pi^2}N^2\ln(N)+\frac{s}{2\pi^2}N^2$$

then

$$\sup_{\mu_0\in\mathcal{P}}\|\mu_t-\nu\|_{\mathrm{tv}} \leqslant \frac{4}{\pi^2}(1+\mathcal{O}(N^{-1}))\exp(-s)$$

3.2. A finite birth and death example with $\lambda_1 \ll \lambda_2-\lambda_1$

The setting is as in the previous example, except that for some $r>1$, we replace (3.1) and (3.2) by

$$\forall\, x\in[\![1,N-2]\!],\qquad \begin{cases} L(x,x+1) := r\\ L(x+1,x) := 1\end{cases} \tag{3.4}$$

$$L(N-1,N) = r \text{ and } L(N,N-1) = 1+r \tag{3.5}$$

The reversible probability η is then given by

$$\forall\, x\in S,\qquad \eta(x) = \begin{cases} \frac{r^2-1}{2r^N-r-1}r^{x-1} & \text{, if } x\in[\![N-1]\!]\\ \frac{r-1}{2r^N-r-1}r^{N-1} & \text{, if } x=N\end{cases} \tag{3.6}$$

Contrary to the previous example, it seems more difficult to derive explicit formulas for the eigenvalues and eigenfunctions associated to $L - V$. To describe them, consider the rational fraction in X,

$$P_N(X) \quad := \quad \frac{X^{2(N+1)} - X^{2N} + r^{1-N}X^2 - r^{-N-1}}{X^2 - r^{-1}}$$

Lemma 3.1. — *For $N \geqslant 1$, P_N is a polynomial which admits $2N$ distinct zeros. Denote by R the set of zeros.*

Let Λ be the image of R by the mapping

$$\Psi \, : \, \rho \mapsto \frac{(1+r)\rho - 1 - r\rho^2}{\rho}$$

For $N > (1+r)/(r-1)$, the spectrum of $V - L$ is Λ and for any $\lambda \in \Lambda$, an associated eigenfunction φ_λ is defined by

$$\forall \ x \in S, \qquad \varphi_\lambda(x) \quad := \quad \rho_+^x - \rho_-^x$$

where

$$\rho_\pm \quad := \quad \frac{1}{2r}(r + 1 - \lambda \pm \sqrt{(\lambda - 1 - r)^2 - 4r}) \tag{3.7}$$

(with $\sqrt{\cdot}$ standing for the principal value of the complex square root) are the reciprocal images of λ by Ψ.

Proof. — Let λ be an eigenvalue of $V - L$ and φ be an associated eigenfunction on S. With the convention that $\varphi(0) = 0$, the values of φ satisfy the recursive formula

$$\forall \ x \in [\![1, N-1]\!], \qquad \varphi(x+1) \quad = \quad \frac{(1 + r - \lambda)\varphi(x) - \varphi(x-1)}{r} \tag{3.8}$$

It follows that on $[\![1, N]\!]$, φ is necessarily proportional to the functions φ_λ defined above, where $\rho_\pm$ are the solutions of the quadratic equation in X,

$$rX^2 + (\lambda - 1 - r)X + 1 \quad = \quad 0 \tag{3.9}$$

except if this equation admits a double solution ρ_*, in which case φ must be proportional to the function φ_* defined by

$$\forall \ x \in S, \qquad \varphi_*(x) \quad := \quad x\rho_*^x$$

Whatever the case, we have that

$$\forall \ x \in [\![1, N-1]\!], \qquad (L - V)[\varphi](x) \quad = \quad -\lambda\varphi(x)$$

This relation is also satisfied at $x=N$ if and only if $\varphi(N+1)=\varphi(N-1)$ (as in the previous example, this justifies the simplifying choice of $L(N,N-1)=1+r$).

• Let us first consider the situation where (3.9) admits a double solution. One computes immediately that this corresponds to $\lambda=(1\pm\sqrt{r})^2$ and $\rho_*=\mp 1/\sqrt{r}$. The condition $\varphi_*(N+1)=\varphi_*(N-1)$ is equivalent to $\rho_*=\pm\sqrt{(N-1)/(N+1)}$. The assumption $N>(1+r)/(r-1)$ forbids that $\sqrt{(N+1)/(N-1)}=\sqrt{r}$, so that we are led to a contradiction. Only the next case is possible.

• Assume that (3.9) has two distinct solutions ρ_+ and ρ_-, they are given in the statement of the above lemma. The condition $\varphi(N+1)=\varphi(N-1)$ amounts to

$$\rho_+^{N+1}-\rho_+^{N-1}-\rho_-^{N+1}+\rho_-^{N-1} \quad = \quad 0$$

But from (3.9) we see that $\rho_-=1/(r\rho_+)$, so ρ_+ is a solution of

$$X^{2(N+1)}-X^{2N}+r^{1-N}X^2-r^{-N-1} \quad = \quad 0$$

This equation admits two obvious solutions, $X=1/\sqrt{r}$ and $X=-1/\sqrt{r}$, so that P_N is indeed a polynomial. But these values are not allowable for ρ_+, because we would have $\rho_+=\rho_-$. It follows that ρ_+ is a root of P_N. Note that if $\rho\in\mathbb{C}$ is a root of P_N, the same is true for $1/(r\rho)$ and that $1/\sqrt{r}$ and $-1/\sqrt{r}$ are the only fixed points of the involutive mapping $\xi\,:\,\mathbb{C}\setminus\{0\}\ni\rho\mapsto 1/(r\rho)$. As a consequence, we can group the roots of P_N by pairs stable by ξ, say $\{\rho_1,\xi(\rho_1)\}$, $\{\rho_2,\xi(\rho_2)\}$, ..., $\{\rho_N,\xi(\rho_N)\}$. Moreover, notice that the mapping Ψ defined in the above lemma is constant on each of these pairs, so that the cardinality of $\Lambda:=\Psi(R)$ is at most N. But it appears from (3.9) and from the previous discussion that all the eigenvalues of $V-L$ are elements of Λ. From (3.8) we deduce that all the eigenvalues of $V-L$ are simple and since by reversibility $V-L$ is known to be diagonalizable, it follows $V-L$ admits N distinct eigenvalues. Thus Λ must be of cardinality N and exactly consists of the eigenvalues of $V-L$. It is interesting to remark that it is relatively difficult to check directly that all the ρ_1, ρ_2, ..., ρ_N are different, or equivalently that all the roots of P_N are distinct (try to compute its discriminant). □

By the Perron-Frobenius theorem, the smallest eigenvalue λ_1 of $V-L$ is characterized by the fact that the associated eigenfunction has a fixed sign. This observation in conjunction with Lemma 3.1 lead to

PROPOSITION 3.2. — *For large N we have*

$$\lambda_1 \quad \sim \quad \frac{1}{2}(r+1)(r-1)^2\frac{1}{r^{N+1}}$$

moreover, if φ is an associated eigenvector,

$$\frac{\varphi_\vee}{\varphi_\wedge} = \frac{r}{r-1}(1+\mathcal{O}(r^{-N}))$$

Proof. — With the notation of Lemma 3.1, we have $\rho_+\rho_- = 1/r$ (recall (3.9)). So if $\rho_+ > 0$, then we get $0 < \rho_- < \rho_+$. It follows that the mapping $\mathbb{R}_+^* \ni u \mapsto \rho_+^u - \rho_-^u$ does not vanish and in particular φ_λ only takes positive values. By consequence, the corresponding $\lambda \in \Lambda$ is the first Dirichlet eigenvalue λ_1. To work out this program, we begin by showing that for N large enough, there exists $\rho_1 \in R \cap \mathbb{R}_+$ satisfying

$$\frac{1}{r^2} \leqslant \rho_1^2 \leqslant \frac{1}{r^2} + \frac{1}{r^{N+1}} \tag{3.10}$$

It is enough to show that there exists $\rho_1 \in [1/r, \sqrt{1/r^2 + r^{-1-N}}]$ such that

$$Q(\rho_1^2) := \rho_1^{2(N+1)} - \rho_1^{2N} + r^{1-N}\rho_1^2 - r^{-N-1} = 0 \tag{3.11}$$

Write $h_1 = r^{N+1}(\rho_1^2 - 1/r^2)$ and for all $h \geqslant 0$,

$$\begin{aligned} f(h) &:= r^{2N} Q\left(\frac{1}{r^2} + \frac{h}{r^{N+1}}\right) \\ &= \left(1 + \frac{h}{r^{N-1}}\right)^N \left(\frac{1}{r^2} - 1 + \frac{h}{r^{N+1}}\right) + h \end{aligned}$$

we just need to check that $f(0) \leqslant 0$ and $f(1) \geqslant 0$. The former inequality is immediate and the latter one is satisfied for N large enough, since $\lim_{N\to\infty} f(1) = 1/r^2$.
Next, injecting the a priori bound (3.10) in (3.11), it follows that

$$\lim_{N\to\infty} h_1 = 1 - \frac{1}{r^2}$$

Replacing $\rho_1 = \sqrt{r^{-2} + (1-r^{-2})r^{-(N+1)}(1+\circ(1))} = r^{-1} + (1-r^{-2})r^{-N}(1/2 + \circ(1))$ in

$$\lambda_1 = \frac{(1+r)\rho_1 - 1 - r\rho_1^2}{\rho_1}$$

we deduce the first announced behavior. Let ρ_{1-} and ρ_{1+} be the corresponding values of ρ_- and ρ_+, from $\rho_{1-}\rho_{1+} = 1/r$, we obtain

$$\rho_{1-} = \rho_1 = \frac{1}{r} + \mathcal{O}(r^{-N}) \quad \text{and} \quad \rho_{1+} = 1 + \mathcal{O}(r^{-N}) \tag{3.12}$$

Taking into account the expression of $\varphi := \varphi_{\lambda_1}$ given in Lemma 3.1, we get

$$\begin{aligned} \frac{\varphi_\vee}{\varphi_\wedge} &= \frac{\varphi(N)}{\varphi(1)} \\ &= \frac{\rho_{1+}^N - \rho_{1-}^N}{\rho_{1+} - \rho_{1-}} \\ &= \frac{r}{r-1}(1+\mathcal{O}(r^{-N})) \end{aligned}$$

□

To be in position to use Theorem 1.3, it remains to evaluate $\lambda_2 - \lambda_1$.

From the previous proof, it appears there is only one eigenvalue $\lambda \in \Lambda$ such that $\rho_- > 0$. Moreover there is at most one eigenvalue $\lambda \in \Lambda$ such that $\rho_- < 0$. Indeed, in this case we have $\rho_- < \rho_+ < 0$ and it follows from Lemma 3.1 that $\varphi_\lambda(x) > 0$ for $x \in S$ odd and $\varphi_\lambda(x) < 0$ for $x \in S$ even, in particular φ_λ has the maximal number of sign changes. The discrete version of Sturm's theorem (see for instance Miclo [26]) then implies that λ must be λ_N, the largest eigenvalue of $V - L$. Since R is symmetrical with respect to zero, $-\rho_{1-}$ and $-\rho_{1+}$ (with the notation of (3.12)) also belong to R and this leads to the estimate

$$\lambda_N = 2(1+r) + \mathcal{O}(r^{-N})$$

The previous arguments show that except for ρ_{1-}, ρ_{1+}, $-\rho_{1-}$ and $-\rho_{1+}$, all the other elements of R are complex numbers which are not real. It follows from Lemma 3.1 that for $\lambda \in \Lambda \setminus \{\lambda_1, \lambda_N\}$,

$$(\lambda - 1 - r)^2 - 4r < 0$$

so that

$$\lambda > 1 + r - 2\sqrt{r} = (1-\sqrt{r})^2$$

In particular, for $N > 2$, we get

$$\lambda_2 > (1-\sqrt{r})^2$$

and as announced, for N large,

$$\lambda_2 - \lambda_1 \sim \lambda_2 \gg \lambda_1$$

meaning that convergence to quasi-stationarity happens at a much faster rate than absorption.

Theorem 1.3 shows that for any $t \geqslant 0$,

$$\begin{aligned}
&\sup_{\mu_0\in\mathcal{P}} \|\mu_t-\nu\|_{\mathrm{tv}} \leqslant \\
&\leqslant \sqrt{\frac{2r^N-r-1}{r^2-1}}\left(\frac{r}{r-1}\right)^2 (1+\mathcal{O}(r^{-N}))\exp\left(-(1-\sqrt{r}+\mathcal{O}(r^{-N}))^2 t\right)\\
&\leqslant \sqrt{\frac{r^N}{r-1}}\left(\frac{r}{r-1}\right)^2 (1+\mathcal{O}(r^{-N}))\exp\left(-(1-\sqrt{r}+\mathcal{O}(r^{-N}))^2 t\right)
\end{aligned}$$

(where $\mathcal{O}(r^{-N})$ is with respect to N, uniformly in $t \geqslant 0$). It follows that for any fixed $s \geqslant 0$, if for N large enough we consider the time

$$t \;:=\; \frac{1}{2(1-\sqrt{r})^2}\left(\ln(r)N+2s\right)$$

then

$$\sup_{\mu_0\in\mathcal{P}} \|\mu_t-\nu\|_{\mathrm{tv}} \;\leqslant\; \frac{r^2}{(r-1)^{5/2}}\,(1+\circ(1))\exp(-s)$$

Notice that the relaxation time to quasi-stationarity needs to be at least of order N, since it is already the order of time required by the semi-group associated to $\widetilde{L}$ to get from 1 to N, which supports a non-negligible part of $\widetilde{\eta}$ (but starting from N, it can be shown that the relaxation time to quasi-stationarity is bounded independently from N, using Theorem 1.1 and Proposition 3.2).

3.3. A finite birth and death example with $\lambda_1 \gg \lambda_2 - \lambda_1$

The setting is as in the previous example, except that $r < 1$ in (3.4) and (3.5).

The beginning of Subsection 3.2 is still valid: the reversible probability η is given by (3.6) and Lemma 3.1 is true, without the condition $N > (1+r)/(r-1)$, which is now void. The difference starts with Proposition 3.2, which must be replaced by

Proposition 3.3. — *For $N \geqslant 4$, we have*

$$(1-\sqrt{r})^2+4\sqrt{r}\sin^2((1-r)/(4N+4)) \leqslant \lambda_1 \leqslant (1-\sqrt{r})^2+4\sqrt{r}\sin^2(\pi/(2N))$$
$$(1-\sqrt{r})^2+4\sqrt{r}\sin^2(\pi/(2N)) \leqslant \lambda_2 \leqslant (1-\sqrt{r})^2+4\sqrt{r}\sin^2(\pi/N)$$

Furthermore, if φ is an eigenvector associated to λ_1, we have

$$\frac{\varphi_\vee}{\varphi_\wedge} \;\leqslant\; r^{-(N-1)/2}\frac{1}{\sin((1-r)/(2N+2))}$$

Proof. — First we show that none of the roots of P_N is a real number. Consider the function

$$f\,:\,\mathbb{R}_+\ni x\mapsto x^{N+1}-x^N+r^{1-N}x-r^{-N-1}$$

According to the arguments of Lemma 3.1, it is sufficient to show that f only vanishes at $1/r$. Its second derivative is given by $f''(x)=(N+1)Nx^{N-1}-N(N-1)x^{N-2}$, for $x\geqslant 0$. Thus f'' is negative on $(0,(N-1)/(N+1))$ and positive on $((N-1)/(N+1),+\infty)$. Furthermore, we compute that

$$f'\left(\frac{N-1}{N+1}\right) \quad=\quad -\left(1-\frac{2}{N+1}\right)^{N-1}+r^{1-N}$$

This quantity is positive for $N\geqslant 2$. Thus f is increasing on $(0,+\infty)$ and can only vanish at $1/r$.

Since we know that the roots of P_N are given by (3.7) for $\lambda\in\Lambda\subset\mathbb{R}$, we deduce that

$$\forall\,\lambda\in\Lambda,\qquad (\lambda-1-r)^2\quad<\quad 4r$$

It follows that the modulus of $\rho_\pm$ in (3.7) is given by $1/\sqrt{r}$, independently of $\lambda\in\Lambda$. More precisely, there exists a set $\Theta\subset(0,\pi)$, such that the roots of P_N are given by

$$\left\{\frac{1}{\sqrt{r}}\exp(\pm i\theta)\,:\,\theta\in\Theta\right\}$$

By using the mapping Ψ of Lemma 3.1, we get that the spectrum of L is

$$\Lambda\quad=\quad\{l(\theta):=1+r-2\sqrt{r}\cos(\theta)\,:\,\theta\in\Theta\}$$

and that corresponding eigenvectors are given by

$$\forall\,x\in[\![1,N]\!],\qquad \varphi_\theta(x)\quad=\quad r^{-x/2}\sin(\theta x) \tag{3.13}$$

for $\theta\in\Theta$ (note the slight modification of notation with respect to Lemma 3.1, indexing by elements of Θ instead of Λ). Ordering Θ into $0<\theta_1<\theta_2<\cdots<\theta_N<\pi$, it appears that

$$\lambda_1=l(\theta_1)\quad\text{and}\quad\lambda_2=l(\theta_2) \tag{3.14}$$

From [26], we deduce that φ_{θ_1} is non-decreasing and that φ_{θ_2} changes sign once (more generally φ_{θ_k} changes sign $k-1$ times, for $k\in[\![1,N]\!]$). This remark and (3.13) lead to the bounds

$$\theta_1\;\leqslant\;\pi/N\quad\text{and}\quad\pi/N\;\leqslant\;\theta_2\;\leqslant\;2\pi/N \tag{3.15}$$

Taking into account that

$$\begin{aligned}\forall\ \theta\in\Theta,\qquad l(\theta) &= (1-\sqrt{r})^2+2\sqrt{r}(1-\cos(\theta))\\ &= (1-\sqrt{r})^2+4\sqrt{r}\sin^2(\theta/2)\end{aligned}$$

and that sinus is positive and increasing on $(0,\pi/2)$, we get that for $N\geqslant 4$,

$$\begin{aligned}\lambda_1 &\leqslant (1-\sqrt{r})^2+4\sqrt{r}\sin^2(\pi/(2N))\\ (1-\sqrt{r})^2+4\sqrt{r}\sin^2(\pi/(2N)) \leqslant \lambda_2 &\leqslant (1-\sqrt{r})^2+4\sqrt{r}\sin^2(\pi/N)\end{aligned}$$

To obtain a lower bound of the same kind for λ_1, recall that the elements $\theta\in\Theta$ satisfy the equation

$$\frac{1}{r^{N+1}}\exp(i2(N+1)\theta)-\frac{1}{r^N}\exp(i2N\theta)+\frac{1}{r^N}\exp(i2\theta)-\frac{1}{r^{N+1}}=0 \quad (3.16)$$

and in particular $g(\theta)=0$, where the mapping g is defined by

$$\forall\ \theta\in\mathbb{R},\qquad g(\theta) := \sin(2(N+1)\theta)-r\sin(2N\theta)+r\sin(2\theta) \quad (3.17)$$

One computes that $g'(0)=2N(1-r)+2+2r$ and that

$$\forall\ \theta\in\mathbb{R},\qquad |g''(\theta)| \leqslant 8(N+1)^2 \quad (3.18)$$

By consequence, the first zero of g after 0 is larger than $(2N(1-r)+2+2r)/(4(N+1)^2)$ and in particular

$$\begin{aligned}\theta_1 &\geqslant \frac{2(1-r)(N+1)}{4(N+1)^2}\\ &= \frac{1-r}{2(N+1)}\end{aligned} \quad (3.19)$$

leading to the announced lower bound on λ_1.

Furthermore, if $\varphi:=\varphi_{\theta_1}$, we have

$$\begin{aligned}\frac{\varphi_\vee}{\varphi_\wedge} &= \frac{\varphi_{\theta_1}(N)}{\varphi_{\theta_1}(1)}\\ &\leqslant r^{-(N-1)/2}\frac{1}{\sin(\theta_1)}\\ &\leqslant r^{-(N-1)/2}\frac{1}{\sin((1-r)/(2N+2))}\end{aligned}$$

□

Working for fixed $r \in (0,1)$ in the asymptotic $N \to \infty$, we deduce that

$$\begin{aligned}(1-\sqrt{r})^2 + \frac{\sqrt{r}(1-r)^2}{4}\frac{1}{N^2}(1+\circ(1)) \;\leqslant\; \lambda_1 \;\leqslant\; (1-\sqrt{r})^2 + \frac{\pi^2\sqrt{r}}{N^2}(1+\circ(1))\\ (1-\sqrt{r})^2 + \frac{\pi^2\sqrt{r}}{N^2}(1+\circ(1)) \;\leqslant\; \lambda_2 \;\leqslant\; (1-\sqrt{r})^2 + \frac{4\pi^2\sqrt{r}}{N^2}(1+\circ(1))\end{aligned}$$

and

$$\frac{\varphi_\vee}{\varphi_\wedge} \quad\leqslant\quad \frac{2N}{(1-r)r^{(N-1)/2}}(1+\circ(1))$$

In particular, we get

$$\begin{aligned}\lambda_2 - \lambda_1 &\quad\leqslant\quad \frac{16\pi^2-(1-r)^2}{4N^2}\sqrt{r}(1+\circ(1))\\ \lambda_1 &\quad\sim\quad (1-\sqrt{r})^2\end{aligned}$$

and, as announced, for N large,

$$\lambda_2 - \lambda_1 \quad\ll\quad \lambda_1$$

meaning that absorption happens at a much faster rate than convergence to quasi-stationarity.

To exhibit a quantitative estimate for the latter convergence, we need a lower bound on $\lambda_2 - \lambda_1$.

Lemma 3.4. — *For N large enough, we have*

$$\lambda_2 - \lambda_1 \quad\geqslant\quad \frac{(1-r)^2\sqrt{r}}{2N^2}(1+\circ(1))$$

Proof. — With the notation of the proof of Proposition 3.3, let us begin by obtaining a lower bound on $\theta_2 - \theta_1$. Considering the function g defined in (3.17), θ_2 is larger than the zero of g following θ_1. We have

$$g'(\theta_1) \quad:=\quad 2(N+1)\cos(2(N+1)\theta_1) - r2N\cos(2N\theta_1) + 2r\cos(2\theta_1)$$

From (3.16), we also obtain

$$\cos(2(N+1)\theta_1) - r\cos(2N\theta_1) + r\cos(2\theta_1) - 1 \quad=\quad 0$$

so that

$$\begin{aligned}g'(\theta_1) &\quad=\quad 2N(1-r\cos(2\theta_1)) + 2\cos(2(N+1)\theta_1) + 2r\cos(2\theta_1)\\ &\quad\geqslant\quad 2N(1-r) - 2(1+r)\end{aligned}$$

Taking into account (3.18), we deduce that

$$\begin{aligned}\theta_2-\theta_1 &\geqslant \frac{N(1-r)-(1+r)}{2(N+1)^2}\\ &= \frac{1-r}{2N}(1+\circ(1))\end{aligned}$$

Next, we have (recall (3.15)),

$$\begin{aligned}-\cos(\theta_2) &\geqslant -\cos(\theta_1)+\min_{\theta\in[\theta_1,\theta_2]}\sin(\theta)(\theta_2-\theta_1)\\ &= -\cos(\theta_1)+(1+\circ(1))\theta_1(\theta_2-\theta_1)\\ &\geqslant -\cos(\theta_1)+\frac{1-r}{2(N+1)}\frac{1-r}{2N}(1+\circ(1))\end{aligned}$$

where we used (3.19). The announced bound is now a consequence of (3.14). □

Putting together the previous estimates, with $\eta_\wedge\sim(1-r)r^{N-1}/(r+1)$, we get

$$\sup_{\mu_0\in\mathcal{P}}\|\mu_t-\nu\|_{\mathrm{tv}}\leqslant\frac{4\sqrt{r+1}N^2}{(1-r)^{5/2}r^{3(N-1)/2}}(1+\circ(1))\exp\left(-\frac{(1-r)^2\sqrt{r}}{2N^2}(1+\circ(1))t\right)$$

In particular, for any given $\epsilon>0$, if we consider

$$t_N \quad := \quad 4(1+\epsilon)\frac{N^2\ln(N)}{(1-r)^2\sqrt{r}}$$

then

$$\lim_{N\to\infty}\sup_{\mu_0\in\mathcal{P}}\|\mu_{t_N}-\nu\|_{\mathrm{tv}} \quad = \quad 0$$

3.4. A non-reversible example

Let $N\in\mathbb{N}$ be fixed. We consider $\bar{S}=S\sqcup\{\infty\}$, with $S=\mathbb{Z}_N$. The generator $\bar{L}$ allows with rate 1 jumps adding 1 in $\mathbb{Z}_N$ and a jump at rate 1 from $0\in\mathbb{Z}_N$ to ∞, the absorbing point. Namely, the generator L is given by

$$\forall\, x,y\in\mathbb{Z}_N,\qquad L(x,y) \quad := \quad \begin{cases}1 & \text{, if } y=x+1\\ -1 & \text{, if } y=x\\ 0 & \text{, otherwise}\end{cases}$$

whose invariant probability measure η is the uniform distribution. The potential V takes the value 1 at 0 and 0 otherwise. The spectral decomposition of the highly non-reversible operator $L-V$ is given by:

Lemma 3.5. — *Let $\mathcal{C}$ be the set of (complex) solutions of the equation $X^N + X^{N-1} - 1 = 0$. Its cardinality is N (i.e. all the solutions of the equation are distinct), the set of eigenvalues of $L - V$ is $\{c-1 : c \in \mathcal{C}\}$ and corresponding eigenvectors are given by the functions φ_c, for $c \in \mathcal{C}$, defined by*

$$\forall\, x \in [\![0, N-1]\!], \qquad \varphi_c(x) \quad := \quad \begin{cases} 1 & , \textit{if } x = 0 \\ c^{x-N} & , \textit{otherwise} \end{cases}$$

(where $\mathbb{Z}_N$ is naturally identified with $[\![0, N-1]\!]$).

Proof. — We begin by checking that all the roots of the polynomial $X^N + X^{N-1} - 1$ are simple. Indeed, if $c \in \mathcal{C}$ had multiplicity at least two, it would also satisfy $Nc^{N-1} + (N-1)c^{N-2} = 0$, namely $c = (1-N)/N$ (because 0 does not belong to $\mathcal{C}$). The equation $c^N + c^{N-1} = 1$ could then be rewritten

$$\frac{1}{N}\left(\frac{1-N}{N}\right)^{N-1} \quad = \quad 1$$

but this is impossible, because the absolute value of the l.h.s. is strictly less than 1.

Next we compute that for $c \in \mathcal{C}$,

$$\forall\, x \in \mathbb{Z}_N, \qquad L[\varphi_c](x) \quad = \quad \begin{cases} c^{1-N} - 1 & , \text{if } x = 0 \\ (c-1)\varphi_c(x) & , \text{otherwise} \end{cases}$$

Note that

$$\begin{aligned} c^{1-N} - 1 \quad &= \quad c^{1-N} - c + (c-1)\varphi_c(0) \\ &= \quad 1 + (c-1)\varphi_c(0) \\ &= \quad V(0)\varphi_c(0) + (c-1)\varphi_c(0) \end{aligned}$$

Thus it appears that on $\mathbb{Z}_N$,

$$(L - V)[\varphi_c] \quad = \quad (c-1)\varphi_c$$

which is the wanted result, since we have exhibited exactly N eigenvalues. $\square$

Necessarily $\mathcal{C}$ contains some real numbers, due to the Perron-Frobenius theorem which asserts that the smallest eigenvalue λ_1 of $V - L$ satisfies

$$\lambda_1 \quad = \quad 1 - \max\{c : c \in \mathcal{C} \cap \mathbb{R}\}$$

By the strong irreversibility of L, the set $\mathcal{C} \cap \mathbb{R}$ is in fact very restricted, an observation which enables easy deduction of the asymptotic behavior of λ_1 for N large:

Lemma 3.6. — *If N is odd, $\mathcal{C}\cap\mathbb{R}=\{1-\lambda_1\}$ and if N is even, $\mathcal{C}\cap\mathbb{R}$ consists of two points. In both cases, $\mathcal{C}\cap\mathbb{R}_+=\{1-\lambda_1\}$ and we have for N large*

$$\lambda_1 \quad\sim\quad \frac{\ln(2)}{N}$$

Proof. — Consider the function

$$g\,:\,\mathbb{R}\ni x \quad\mapsto\quad x^N+x^{N-1}-1$$

The study of its variations leads to the two first announced results by differentiating it twice. Indeed, if N is odd, g is increasing on $(-\infty,(1-N)/N)$, decreasing on $((1-N)/N,0)$ and increasing on $(0,+\infty)$. As was already seen in the proof of the previous lemma, $g((1-N)/N)<0$, so that g admits a unique real root contained in $(0,+\infty)$. For N even, g is decreasing on $(-\infty,(1-N)/N)$ and increasing on $((1-N)/N,+\infty)$. Since $g((1-N)/N)<0$ and $\lim_{\pm\infty} g=+\infty$, g admits two real roots, the largest one being the unique one belonging to $(0,+\infty)$, since $g(0)=-1$.

Let $y>0$ be given and for $N>y$ consider $x_N=1-y/N$. It appears that

$$\lim_{N\to\infty} g(x_N) \quad=\quad 2\exp(-y)-1$$

It follows that the unique root c_N of g in $(0,+\infty)$ satisfies for N large

$$c_N-1 \quad\sim\quad -\frac{\ln(2)}{N}$$

which amounts to the last announced result. □

Let $\varphi=\varphi_{1-\lambda_1}$, with the notation of Lemma 3.5, be an eigenvector associated to λ_1. We have, for N large

$$\begin{aligned}\frac{\varphi_\vee}{\varphi_\wedge} &= \frac{\varphi(1)}{\varphi(0)}\\ &= (1-\lambda_1)^{1-N}\\ &\sim \exp(\ln(2)) \;=\; 2\end{aligned}$$

In addition, note that L^*, the dual operator of L in $\mathbb{L}^2(\eta)$, is given by

$$\forall\, x,y\in\mathbb{Z}_N,\qquad L^*(x,y) \quad:=\quad \begin{cases} 1 & \text{, if } y=x-1\\ -1 & \text{, if } y=x\\ 0 & \text{, otherwise}\end{cases}$$

It corresponds to the conjugation of L with the involutive transformation of $\mathbb{Z}_N$ given by $\iota : \mathbb{Z}_N \ni x \mapsto -x$ (or $[\![1, N-1]\!] \ni x \mapsto N-x$ and $\iota(0) = 0$). It follows that the function φ^* considered in the introduction is proportional to $\varphi \circ \iota$, so that the mapping $\varphi\varphi^*$ is constant. In particular the probability $\widetilde{\eta}$ defined in (1.4) is equal to η, the uniform distribution on $\mathbb{Z}_N$. Furthermore, we compute that the generator $\widetilde{L}$ defined in (1.3) is given by

$$\forall\, x, y \in \mathbb{Z}_N, \qquad \widetilde{L}(x,y) \quad := \quad \begin{cases} (1-\lambda_1) & \text{, if } x \neq 0 \text{ and } y = x+1 \\ -(1-\lambda_1) & \text{, if } x \neq 0 \text{ and } y = x \\ (1-\lambda_1)^{1-N} & \text{, if } x = 0 \text{ and } y = 1 \\ -(1-\lambda_1)^{1-N} & \text{, if } x = 0 \text{ and } y = 0 \\ 0 & \text{, otherwise} \end{cases}$$

Its additive symmetrization $\widetilde{L}^\diamond$ in $\mathbb{L}^2(\eta)$ gives the rate $(1-\lambda_1)/2$ to any oriented edge $(x, x+1)$ or $(x+1, x)$ of $\mathbb{Z}_N$, except to the edges $(0,1)$ and $(1,0)$, which have the rate $(1-\lambda_1)^{1-N}/2$. By comparison with the usual continuous-time random walk on $\mathbb{Z}_N$, we deduce that the spectral gap $\widetilde{\lambda}$ of $\widetilde{L}^\diamond$ satisfies

$$(1-\cos(2\pi/N))(1-\lambda_1) \;\leqslant\; \widetilde{\lambda} \;\leqslant\; (1-\cos(2\pi/N))(1-\lambda_1)^{1-N}$$

namely, asymptotically for N large,

$$\frac{2\pi^2}{N^2}(1+\circ(1)) \;\leqslant\; \widetilde{\lambda} \;\leqslant\; \frac{4\pi^2}{N^2}(1+\circ(1))$$

Relying on (1.6), we would have obtained

$$\widetilde{\lambda} \quad \geqslant \quad \frac{1+\circ(1)}{4}\lambda$$

where λ is the spectral gap of the additive symmetrization of L in $\mathbb{L}^2(\eta)$, which is the usual continuous-time random walk on $\mathbb{Z}_N$, so that $\lambda \sim 2\pi^2/N^2$. Thus it only leads to a slight deterioration on the estimate of $\widetilde{\lambda}$ obtained by working directly with (1.5).

For large N, Theorem 1.2 leads to

$$\sup_{\mu_0 \in \mathcal{P}} \|\mu_t - \nu\|_{\mathrm{tv}} \quad \leqslant \quad 2\sqrt{N}(1+\circ(1)) \exp\left(\frac{2\pi^2}{N^2}(1+\circ(1))t\right)$$

In particular, for any given $\epsilon > 0$, if we consider

$$t_N \quad := \quad (1+\epsilon)\frac{N^2 \ln(N)}{4\pi^2}$$

then

$$\lim_{N\to\infty} \sup_{\mu_0 \in \mathcal{P}} \|\mu_{t_N} - \nu\|_{\mathrm{tv}} \quad = \quad 0$$

3.5. A product example

Let us first come back to the general setting of the introduction (which is then tensorized). Let $d \in \mathbb{N}$, be given. On S^d, consider the Markovian generator

$$L^{(d)} \quad := \quad \frac{1}{d} \sum_{k \in [\![1,d]\!]} L_k$$

where L_k acts like L on the k-th coordinate of S^d. Define furthermore the potential $V^{(d)}$ by

$$\forall\, x := (x_1, ..., x_d) \in S^d, \qquad V^{(d)} \quad := \quad \frac{1}{d} \sum_{k \in [\![1,d]\!]} V(x_k)$$

Note that the associated $\bar{L}^{(d)}$ is not of the form $(1/d) \sum_{k \in [\![1,d]\!]} \bar{L}_k$, because the underlying state space would be $(\bar{S})^d$ and not $S^d \sqcup \{\infty\}$ as it should be. One recovers the subMarkovian generator $L^{(d)} - V^{(d)}$ by modifying $(1/d) \sum_{k \in [\![1,d]\!]} \bar{L}_k$ so that all the points of $\{x := (x_1, ..., x_d) \in (\bar{S})^d \,:\, \exists\, k \in [\![1, d]\!]$ with $x_k = \infty\}$ become absorbing.

The invariant measure $\eta^{(d)}$ associated to $L^{(d)}$ is $\eta^{\otimes d}$ and we have

$$L^{(d)} - V^{(d)} \quad = \quad \frac{1}{d} \sum_{k \in [\![1,d]\!]} (L - V)_k$$

It appears in particular that the first eigenvalue of $V^{(d)} - L^{(d)}$ is λ_1, the same as that of $V - L$ and the associated quasi-stationary distribution (respectively first eigenfunction) is $\nu^{\otimes d}$ (resp. $\varphi^{\otimes d}$). It follows that $\widetilde{L}^{(d)}$, the Doob transform of $L^{(d)} - V^{(d)}$ by $\varphi^{\otimes d}$, satisfies

$$\widetilde{L}^{(d)} \quad = \quad \frac{1}{d} \sum_{k \in [\![1,d]\!]} \widetilde{L}_k$$

and that its invariant probability $\widetilde{\eta}^{(d)}$ is $\widetilde{\eta}^{\otimes d}$. In a similar way, we have that

$$L^{*(d)} \quad = \quad \frac{1}{d} \sum_{k \in [\![1,d]\!]} L_k^*$$

and the first eigenvector of $-L^{*(d)}$ is $(\varphi^*)^{\otimes d}$. Finally $\widetilde{L}^{\diamond(d)}$, the additive symmetrization of $\widetilde{L}^{(d)}$ in $\mathbb{L}^2(\widetilde{\eta}^{\otimes d})$, is equal to $(1/d) \sum_{k \in [\![1,d]\!]} \widetilde{L}_k^{\diamond}$, so that its spectral gap $\widetilde{\lambda}$ (respectively its logarithmic Sobolev constant $\widetilde{\alpha}$) is equal to that of $\widetilde{L}^\diamond$ (for such tensorization properties, see for instance the book [1] of Ané et al.).

With obvious notation, Theorem 1.2 then leads to the fact that for any $t \geqslant 0$, we have

$$\sup_{\mu_0^{(d)} \in \mathcal{P}^{(d)}} \left\| \mu_t^{(d)} - \nu^{\otimes d} \right\|_{\mathrm{tv}} \leqslant \left(\sqrt{\frac{\eta[\varphi\varphi^*]}{(\varphi\varphi^*\eta)_\wedge}} \frac{\varphi_\vee}{\varphi_\wedge} \right)^d \exp(-\widetilde{\lambda} t)$$

Under the reversibility condition of Theorem 1.3, we get that for any $t \geqslant 0$,

$$\sup_{\mu_0^{(d)} \in \mathcal{P}^{(d)}} \left\| \mu_t^{(d)} - \nu^{\otimes d} \right\|_{\mathrm{tv}} \leqslant \left(\sqrt{\frac{1}{\eta_\wedge}} \left(\frac{\varphi_\vee}{\varphi_\wedge} \right)^2 \right)^d \exp(-(\lambda_2 - \lambda_1)t) \tag{3.20}$$

The bound (2.10) can be rewritten in the form

$$\sup_{\mu_0^{(d)} \in \mathcal{P}^{(d)}} \left\| \mu_t^{(d)} - \nu^{\otimes d} \right\|_{\mathrm{tv}} \leqslant \sqrt{2d \ln \left(\frac{\eta[\varphi\varphi^*]}{(\varphi\varphi^*\eta)_\wedge} \right) \left(\frac{\varphi_\vee}{\varphi_\wedge} \right)^d} \exp(-(\widetilde{\alpha}/2)t)$$

or under the reversibility condition,

$$\sup_{\mu_0^{(d)} \in \mathcal{P}^{(d)}} \left\| \mu_t^{(d)} - \nu^{\otimes d} \right\|_{\mathrm{tv}} \leqslant \sqrt{2d \ln \left(\frac{1}{\eta_\wedge \varphi_\wedge^2} \right) \left(\frac{\varphi_\vee}{\varphi_\wedge} \right)^d} \exp(-(\widetilde{\alpha}/2)t) \tag{3.21}$$

It is easy to construct an example showing that (3.21) can lead to a better estimate than (3.20). Take

$$S := \{1, 2\}, \quad L := \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad V := \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

for which $\varphi \equiv 1$, $\eta = (1/2, 1/2)$, $\lambda_2 - \lambda_1 = 2$ and $\widetilde{\alpha} = 1$ (recall the convention after (2.11), which is an equality in this two-points case, see Diaconis and Saloff-Coste [13]). The r.h.s. of (3.20) and (3.21) are respectively $2^{d/2} \exp(-2t)$ and $\sqrt{2d \ln(2)} \exp(-t/2)$. The first bound leads to a mixing time (the first time $t > 0$ the quantity $\sup_{\mu_0^{(d)} \in \mathcal{P}^{(d)}} \left\| \mu_t^{(d)} - \nu^{\otimes d} \right\|_{\mathrm{tv}}$ goes below a fixed level such as 1) of order d, while the second bound rather gives order $\ln(d)$.

For a little less artificial example, one can come back to Subsection 3.3, with $N = 2$ and $r > 0$ very small. Indeed, one computes that

$$\begin{aligned}
\lambda_2 - \lambda_1 &= 2\sqrt{r(1+r)} \\
\frac{\varphi_\vee}{\varphi_\wedge} &= \sqrt{\frac{1-r}{r}} \\
\eta_\wedge &= \frac{r}{1+r} \\
\widetilde{\eta}_\wedge &= \frac{1-r}{2}
\end{aligned}$$

It follows from (2.11) that for $0 < r \ll 1$,

$$\begin{aligned}
\widetilde{\alpha} &\geqslant \frac{r}{\ln((1+r)/(1-r))}(\lambda_2 - \lambda_1) \\
&\sim \frac{\lambda_2 - \lambda_1}{2}
\end{aligned}$$

For $r > 0$ small, we get from (3.20) and (3.21) that the leading term in $d \in \mathbb{N}$ in the deduced upper bounds on the mixing time are respectively $3d/(4\sqrt{r})\ln(1/r)$ and $d/(2\sqrt{r})\ln(1/r)$, showing thus a little advantage for the estimate coming from (3.21).

4. Some discrete time models

Of course the theory can be developed in discrete time as well. We briefly carry this out here and treat some higher dimensional examples where all the spectral information is available. Let $\bar{S} := S \sqcup \{\infty\}$ be the extended state space with ∞ the absorbing state. Denote by N the cardinality of S. The transition matrix $\bar{Q}$ can be written

$$\bar{Q} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_1 & & & \\ \vdots & & Q & \\ a_N & & & \end{pmatrix}$$

with Q an $N \times N$ matrix, here assumed to be irreducible. Let ψ and φ be positive left and right eigenvectors of Q with eigenvalue $\beta > 0$ of largest size. Set

$$\forall\, x, y \in S, \qquad K(x,y) := Q(x,y)\frac{\varphi(y)}{\beta\varphi(x)}$$

This is a Markov transition matrix on S with stationary distribution π given by

$$\forall\, x \in S, \qquad \pi(x) \;\; := \;\; \frac{\varphi(x)\psi(x)}{\sum_{y\in S}\varphi(y)\psi(y)}$$

It has the probabilistic interpretation of the transition probabilities for the original chain conditioned on non-absorption (for all time). The quasi-stationary distribution is given by

$$\forall\, x \in S, \qquad \nu(x) \;\; := \;\; \frac{\psi(x)}{\sum_{y\in S}\psi(y)}$$

observe that the ratio $r := \varphi_\vee/\varphi_\wedge$ allows the bounds

$$\forall\, x \in S, \qquad r^{-1}\nu(x) \;\leqslant\; \pi(x) \;\leqslant\; r\nu(x)$$

If Q above is diagonalizable, with right eigenfunctions $(f_i)_{i\in[\![N]\!]}$ and left eigenfunctions $(g_i)_{i\in[\![N]\!]}$ for eigenvalues $(\beta_i)_{i\in[\![N]\!]}$, normalized so that $\sum_{x\in S} g_i(x)f_j(x) = \delta_{i,j}$ for any $i, j \in [\![N]\!]$, then

$$\forall\, l \in \mathbb{Z}_+, \forall\, x, y \in S, \qquad Q^l(x,y) \;\; = \;\; \sum_{k=1}^{N} \beta_k^l f_k(x) g_k(y)$$

thus

$$P[X_l = y|X_0 = x, T > l] \;\; = \;\; \frac{Q^l(x,y)}{Q^l(x)}$$

with $Q^l(x) = \sum_{y\in S} Q^l(x,y)$ and where $(X_n)_{n\in\mathbb{Z}_+}$ is the underlying absorbing Markov chain and T is its absorbing time.

The argument for Theorem 1.1 works in the discrete case just as well. With notation as above, it shows for any starting state x,

$$\forall\, l \in \mathbb{Z}_+, \quad \frac{\varphi_\wedge}{2\varphi_\vee}\left\|K^l(x,.) - \pi\right\|_{\mathrm{tv}} \leqslant \left\|Q^l(x,.) - \nu\right\|_{\mathrm{tv}} \leqslant \frac{2\varphi_\vee}{\varphi_\wedge}\left\|K^l(x,.) - \pi\right\|_{\mathrm{tv}}$$

Explicit diagonalizations are available surprisingly often. For example, for a birth and death chain on $\{0, 1, ..., 2N\}$, symmetric with respect to N, take the starting point to be zero and the absorbing point to be N. If $(\varphi_i)_{i\in[\![0,2N]\!]}$ are the right eigenvectors of the original chain, often available as orthogonal polynomials, φ_1, φ_3, ..., φ_{2N-1} all vanish at N and so restrict to the needed $(f_i)_{i\in[\![N]\!]}$. Because birth and death chains are reversible, these determine the family $(g_i)_{i\in[\![N]\!]}$ and the ingredients for analysis are available. The Ehrenfest urn and the example at the end of this section are two cases where we have carried this approach out to get sharp answers (matching upper and lower bounds for convergence to quasi-stationarity). It is only fair to report that the analysis involved can require substantial effort.

4.1. Example of rock breaking

In this example the matrix Q is not irreducible, nevertheless the above results can be applied, because the function φ is (strictly) positive. To justify this observation, for $\epsilon \in (0,1)$, replace Q by $(1-\epsilon)Q+\epsilon J$, where J has all its entries equal to $1/N$, apply the previous results and let ϵ go to zero.

Let $n \in \mathbb{N}$ be given and $\bar{S} := \mathcal{P}(n)$, the set of all partitions of n. Thus if $n=4$, $\bar{S}=\{4,31,22,211,1111\}$. An absorbing Markov chain on $\bar{S}$, modeled on a rock breaking Markov chain studied by Kolmogorov, is developed in Diaconis, Pang and Ram [11]. Briefly, if $\lambda=(\lambda_1,\lambda_2,\dots,\lambda_l)$, with $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_l > 0$, $\lambda_1+\cdots+\lambda_l = n$, the chain proceeds from λ by independently choosing, for $i \in [\![1,l]\!]$, binomial variables $\lambda_i^{(1)}$ of parameters $(\lambda_i, 1/2)$, so that we can write $\lambda_i \quad \lambda_i^{(1)}+\lambda_i^{(2)}$. Next, after discarding any zeros and reordering the $\lambda_i^{(1)}, \lambda_i^{(2)}$, for $i\in[\![1,l]\!]$, we get the new position of the chain. It is absorbing at (1^n). The natural starting place is (n).

In [11], the eigenvalues are shown to be $1, 1/2, 1/4, \dots, 1/2^n$, with $1/2^{n-l}$ having multiplicity $p(n,l)$, the number of partitions of n into l parts. In particular the second eigenvalue is $1/2$, with multiplicity 1. The eigenvectors are given explicitly and these restrict to give explicit left and right eigenbases of Q. With notation as above, for $\beta=1/2$, for all $\lambda=(\lambda_1,\lambda_2,\dots,\lambda_l)\in\mathcal{P}(n)$,

$$
\begin{aligned}
\varphi(\lambda) &= \sum_{i\in[\![1,l]\!]} \binom{\lambda_i}{2} \\
\psi(\lambda) &= \begin{cases} 1 & \text{, if } \lambda=(1^{n-2},2) \\ 0 & \text{, otherwise} \end{cases}
\end{aligned}
$$

Thus $\varphi_\vee/\varphi_\wedge = \binom{n}{2}/1 = \binom{n}{2}$. When $n=4$, the original transition matrix is

$$
\begin{array}{c c}
 & \begin{array}{ccccc} 1^4 & 1^2 2 & 2^2 & 13 & 4 \end{array} \\
\begin{array}{c} 1^4 \\ 1^2 2 \\ 2^2 \\ 13 \\ 4 \end{array} &
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
1/2 & 1/2 & 0 & 0 & 0 \\
1/4 & 1/2 & 1/4 & 0 & 0 \\
0 & 3/4 & 0 & 1/4 & 0 \\
0 & 0 & 3/8 & 1/2 & 1/8
\end{pmatrix}
\end{array}
$$

The left (right) eigenvectors are given as the rows (columns) of the two arrays

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 & 0 \\ 1 & 6 & 3 & 4 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 2 & -3 & 0 & 1 & 0 \\ -6 & 12 & -3 & -4 & 1 \end{pmatrix}$$

So $\psi = (1,0,0,0)$, $\varphi = (1,2,3,6)^{\mathrm{t}}$. The adjusted transition matrix K is given by

$$\begin{array}{cc} & \begin{array}{cccc} 1^2 2 & 2^2 & 13 & 4 \end{array} \\ \begin{array}{c} 1^2 2 \\ 2^2 \\ 13 \\ 4 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \end{pmatrix} \end{array}$$

The reader may check that discarding the top row and first column of the eigenvector arrays gives the eigenvectors of K.

In this example the quasi-stationary distribution ν is the stationary distribution π of K, both are the Dirac mass at $1^{n-2}2$. The chain K is itself absorbing. This rock breaking chain is a special case of a host of explicitly diagonalizable Markov chains derived from Hopf algebras [11]. Some other algebraic constructions leading to explicit quasi-stationary calculations may be found in Defosseux [8] (fusion coefficients and random walks in alcoves of affine Lie algebras). Symmetric function theory, in various deformations (Sekiguchi-Debiard operators) leads to further explicit diagonalizations in the work of Jiang [22]. Turning either of these last sets of examples into sharp bounds seems like a fascinating research project.

4.2. Geometric theory

The basic path arguments of Holley and Stroock [19], Jerrum and Sinclair [20] and Diaconis and Stroock [16] can be applied to absorbing chains. This was done in a sophisticated context in Miclo [27]. The following paragraph develops a simple version in the discrete context. Let S be a finite set, ∞ an absorbing point and $\bar{Q}$ a Markov chain on $\bar{S} := S \sqcup \{\infty\}$. We suppose as above that the chain is absorbing with probability one and that the chain restricted to S is connected. Suppose that q is a probability on S and consider $\mathbb{L}^2(q)$ endowed with its usual inner product $\langle f, g\rangle_q := \sum_{x\in S} f(x)g(x)q(x)$, for $f, g \in \mathbb{L}^2(q)$. Suppose too that $q(x)\bar{Q}(x,y) = q(y)\bar{Q}(y,x)$ for $x, y \in S$. When needed, define $q(\infty) = 0$ and the functions from $\mathbb{L}^2(q)$ are extended on $\bar{S}$ by making them vanish at ∞. Let β_1 be the largest eigenvalue of Q the restriction of $\bar{Q}$ to S. The minimax characterization gives

Lemma 4.1. — *If the Poincaré inequality $\|f\|_q^2 \leqslant A \left\langle (I-Q)f, f\right\rangle_q$ holds for all $f \in \mathbb{L}^2(q)$, then $\beta_1 \leqslant 1 - 1/A$.*

Remark 4.2. Of course, the analogue λ_1 of Section 1 satisfies $\lambda_1 = 1-\beta_1$.

Define a Dirichlet form $\mathcal{E}$ on $\mathbb{L}^2(q)$, by

$$\forall\, f \in \mathbb{L}^2(q), \qquad \mathcal{E}(f,f) \quad := \quad \frac{1}{2}\sum_{x,y\in\bar{S}} (f(y)-f(x))^2\, q(x)\bar{Q}(x,y)$$

Lemma 4.3. *For $f \in \mathbb{L}^2(q)$, we have*

$$\begin{aligned}\mathcal{E}(f,f) \quad &= \quad \left\langle (I-Q)f,f\right\rangle_q - \frac{1}{2}\sum_{x\in S} f^2(x)q(x)\bar{Q}(x,\infty)\\ &\leqslant \quad \left\langle (I-Q)f,f\right\rangle_q\end{aligned}$$

Proof. — This is simple by directly computing both sides of the equality, separating the cases where $x, y \in S$ from the cases where $x \in S$ and $y = \infty$. □

To bring in geometry, for $x \in S$, let γ_x be a path starting at x and ending at ∞ with steps possible with respect to $\bar{Q}$. If there are many absorbing points, γ_x may connect x to any of them. Thus $\gamma_x = (x_0 = x, x_1, ..., x_l = \infty)$ with $\bar{Q}(x_i, x_{i+1}) > 0$ for $0 \leqslant i \leqslant l-1$. Let the length l of the path be denoted $|\gamma_x|$.

Proposition 4.4. — *With the notation as above, A in Lemma 4.1 may be taken as*

$$A \quad = \quad \max_{x\in S,\, y\in\bar{S}\,:\,K(x,y)>0} \frac{2}{q(x)\bar{Q}(x,y)} \sum_{z\in S\,:\,(x,y)\in\gamma_z} |\gamma_z|\, q(z)$$

Proof. — Let $x \in S$ be given and write $\gamma_x = (x_0, x_1, ..., x_l)$. The idea is to expand

$$\begin{aligned}f^2(x) \quad &= \quad ((f(x_0)-f(x_1)) + (f(x_1)-f(x_2)) + \cdots + (f(x_{l-1})-f(x_l)))^2\\ &\leqslant \quad |\gamma_x| \sum_{i=0}^{l-1} (f(x_i)-f(x_{i+1}))^2\end{aligned}$$

Thus for $f \in \mathbb{L}^2(q)$,

$$\begin{aligned}
\sum_{x\in S} f^2(x)q(x) &= \sum_{x\in S}\left(\sum_{e\in\gamma_x} f(e^-)-f(e^+)\right)^2 q(x)\\
&\leqslant \sum_{x\in S}|\gamma_x|\; q(x)\sum_{e\in\gamma_x}(f(e^-)-f(e^+))^2\\
&= \sum_{x\in S,\, y\in\bar{S}\,:\,K(x,y)>0}(f(x)-f(y))^2\frac{q(x)\bar{Q}(x,y)}{q(x)\bar{Q}(x,y)}\sum_{z\in S\,:\,(x,y)\in\gamma_z}|\gamma_z|\,q(z)\\
&\leqslant A\mathcal{E}(f,f)
\end{aligned}$$

□

Remark 4.5. *Path technology has evolved: with many choices of paths, one may choose randomly, see Diaconis and Saloff-Coste [14], weights may be used in the Cauchy-Schwarz bound, as in Diaconis and Saloff-Coste [15]. This can be important when the stationary distributions varies a lot. Paths may be used locally, see Diaconis and Saloff-Coste [14]. Any such variation is easy to adapt in the above proposition.*

4.3. Other examples in discrete time

The following calculations are classical. The neat form presented here is borrowed from the thesis work of Zhou [32] and provides an alternative approach to Examples 3.1, 3.2 and 3.3. Consider a birth and death chain on $[\![0,N]\!]$ with transition matrix

$$\begin{pmatrix}
r & 1-r & & & \\
p & 0 & q & & \\
 & & \ddots & & \\
 & & p & 0 & q\\
 & & & 1-s & s
\end{pmatrix}$$

with $p\in(0,1)$, $q=1-p$, $r,s\in[0,1]$.

Proposition 4.6. — *The eigenvalues and right eigenfunctions are of form*

$$\beta := 2\sqrt{pq}\cos(\theta), \qquad \forall\; x\in[\![0,N]\!],\quad \varphi(x) := \left(\frac{p}{q}\right)^{x/2}\cos(\theta x+c)$$

where θ and c are determined by the boundary values:

$$\begin{aligned}
r\varphi(0)+(1-r)\varphi(1) &= \beta\varphi(0)\\
(1-s)\varphi(N-1)+s\varphi(N) &= \beta\varphi(N)
\end{aligned}$$

Proof. — This follows from the trigonometric identity

$$\forall\ \alpha,\beta\in\mathbb{R},\qquad \cos(\alpha)+\cos(\beta)\ =\ 2\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)$$

Since for any $\theta, c\in\mathbb{R}$,

$$q\left(\frac{p}{q}\right)^{(x+1)/2}\cos(\theta(x+1)+c)+p\left(\frac{p}{q}\right)^{(x-1)/2}\cos(\theta(x-1)+c)$$
$$=2\sqrt{pq}\cos(\theta)\left(\frac{p}{q}\right)^{x/2}\cos(\theta x+c)$$

Zhou [32] has shown that the above boundary conditions lead indeed to $N+1$ eigenvalues. □

As an example, take $p=q=1/2=r$, $s=1$. This gives the simple random walk on $[\![0,N]\!]$ absorbing at N with holding at 0. The above proposition gives the equations

$$\cos(c)+\cos(\theta+c)\ =\ 2\cos(\theta)\cos(c),\quad \cos(N\theta+c)\ =\ 0\ \text{ or }\ \theta\ =\ 0$$

These have solutions $c=\theta/2$, $\theta=j\pi/(2N+1)$, for $j=0,1,3,...,2N-1$. It follows that the chain has eigenvalues $\beta_j:=\cos(j\pi/(2N+1))$, for $j=0,1,3,...,2N-1$ with right eigenfunctions φ_j given by

$$\forall\ x\in[\![0,N]\!],\qquad \varphi_j(x)\ :=\ \cos\left(\frac{(2x+1)j\pi}{2(2N+1)}\right)$$

The left eigenfunctions are $\psi_0(x)=\delta_N(x)$ and for $j=1,3,...,2N-1$,

$$\forall\ x\in[\![0,N]\!],\quad \psi_j(x)\ :=\ \begin{cases}\varphi_j(x) & \text{, if } x\in[\![0,N-1]\!]\\ \frac{(-1)^{(j+1)/2}}{2}\cot\left(\frac{j\pi}{2(2N+1)}\right) & \text{, if } x=N\end{cases}$$

In particular, the quasi-stationary distribution has probability density given by

$$\forall\ x\in[\![0,N-1]\!],\qquad \nu(x)\ :=\ \frac{\psi_1(x)}{\sum_{y\in[\![0,N-1]\!]}\psi_1(y)}$$
$$=\ 2\tan\left(\frac{\pi}{2(2N+1)}\right)\cos\left(\frac{(2x+1)\pi}{2(2N+1)}\right)$$

Consider the geometric bound from Proposition 4.4. From the discussion above,

$$\beta_1\ =\ \cos\left(\frac{\pi}{2N+1}\right)\ =\ 1-\frac{\pi^2}{2(2N+1)^2}+\mathcal{O}\left(\frac{1}{N^4}\right)$$

The reversing probability q on $[\![0, N-1]\!]$ is the uniform distribution. There is a unique choice of (not self-intersecting) paths from x to N. The quantity A is obviously maximized at the edge $(N-1, N)$. Then, it is

$$A \;=\; \frac{4N}{N}\sum_{x\in[\![0,N-1]\!]} N-x \;=\; 2N(N+1)$$

This gives $\beta_1 \leqslant 1 - \frac{1}{4N(N+1)}$ which compares reasonably with the correct answer.

In this problem, $\varphi_\vee/\varphi_\wedge$ is of order N and our bounds show that order $N^2\ln(N)$ steps suffice for convergence to quasi-stationarity. Using all of the spectrum, classical analysis shows that order N^2 steps are necessary and sufficient. Zhou [32] gives similar exact formulae for reflecting and absorbing boundaries at zero. He also derives the exact spectral data for some absorbing birth and death chains from biology (Morans model with various types of mutation).

Acknowledgments. — We are grateful to Bertrand Cloez for discussions on the subject, as well as for pointing out some related references. Thanks also to Hua Zhou for help with Section 4 and to the ANR STAB and to the LABEX CIMI for their support.

Bibliography

[1] Ané (C.), Blachère (S.), Chafaï (D.), Fougères (P.), Gentil (I.), Malrieu (F.), Roberto (C.), and Scheffer (G.). — Sur les inégalités de Sobolev logarithmiques, volume 10 of Panoramas et Synthèses [Panoramas and Syntheses]. Société Mathématique de France, Paris (2000). With a preface by Dominique Bakry and Michel Ledoux.

[2] Barbour (A. D.) and Pollett (P. K.). — Total variation approximation for quasi-stationary distributions. J. Appl. Probab., 47(4), p. 934-946 (2010).

[3] Barbour (A. D.) and Pollett (P. K.). — Total variation approximation for quasi-equilibrium distributions, II. Stochastic Process. Appl., 122(11), p. 3740-3756 (2012).

[4] Bobkov (S. G.) and Tetali (P.). — Modified logarithmic Sobolev inequalities in discrete settings. J. Theoret. Probab., 19(2), p. 289-336 (2006).

[5] Champagnat (N.) and Villemonais (D.). — Exponential convergence to quasi-stationary distribution and Q-process. ArXiv e-prints, April 2014.

[6] Cloez (B.) and Thai (M. N.). — Quantitative results for the Fleming-Viot particle system in discrete space. ArXiv e-prints, December 2013.

[7] Collet (P.), Martínez (S.), and San Martín (J.). Quasi-stationary distributions. Markov chains, diffusions and dynamical systems. Probability and its Applications (New York). Springer, Heidelberg (2013).

[8] Defosseux (M.). — Fusion coefficients and random walks in alcoves. ArXiv e-prints, July 2013.

[9] Del Moral (P.). — Mean field simulation for Monte Carlo integration, volume 126 of Monographs on Statistics and Applied Probability. CRC Press, Boca Raton, FL (2013).

[10] Del Moral (P.) and Miclo (L.). — Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman-Kac semigroups. ESAIM Probab. Stat., 7, p. 171-208 (2003).

[11] Diaconis (P.), Amy Pang (C. Y.), and Ram (A.). — Hopf algebras and Markov chains: two examples and a theory. J. Algebraic Combin., 39(3), p. 527-585 (2014).

[12] Diaconis (P.) and Saloff-Coste (L.). — Comparison theorems for reversible Markov chains. Ann. Appl. Probab., 3(3), p. 696-730 (1993).

[13] Diaconis (P.) and Saloff-Coste (L.). — Logarithmic Sobolev inequalities for finite Markov chains. Ann. Appl. Probab., 6(3), p. 695-750 (1996).

[14] Diaconis (P.) and Saloff-Coste (L.). — Nash inequalities for finite Markov chains. J. Theoret. Probab., 9(2), p. 459-510 (1996).

[15] Diaconis (P.) and Saloff-Coste (L.). — What do we know about the Metropolis algorithm? J. Comput. System Sci., 57(1), p. 20-36 (1998). 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV).

[16] Diaconis (P.) and Stroock (D.). — Geometric bounds for eigenvalues of Markov chains. Ann. Appl. Probab., 1(1), p. 36-61 (1991).

[17] Fill (J. A.). — Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. Ann. Appl. Probab., 1(1), p. 62-87 (1991).

[18] Gyrya (P.) and Saloff-Coste (L.). — Neumann and Dirichlet heat kernels in inner uniform domains. Astérisque, (336), p. viii+144 (2011).

[19] Holley (R.) and Stroock (D.). — Simulated annealing via Sobolev inequalities. Comm. Math. Phys., 115(4), p. 553-569 (1988).

[20] Jerrum (M.) and Sinclair (A.). — Approximating the permanent. SIAM J. Comput., 18(6), p. 1149-1178 (1989).

[21] Jacka (S. D.) and Roberts (G. O.). — Weak convergence of conditioned processes on a countable state space. J. Appl. Probab., 32(4), p. 902-916 (1995.)

[22] Jiang (Y.). — Mixing Time of Metropolis Chain Based on Random Transposition Walk Converging to Multivariate Ewens Distribution. ArXiv e-prints, April 2012.

[23] Lierl (J.) and Saloff-Coste (L.). — The Dirichlet heat kernel in inner uniform domains: local results, compact domains and non-symmetric forms. ArXiv e-prints, October 2012.

[24] Méléard (S.) and Villemonais (D.). — Quasi-stationary distributions and population processes. Probab. Surv., 9, p. 340-410 (2012).

[25] Miclo (L.). — Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In Séminaire de Probabilités, XXXI, volume 1655 of Lecture Notes in Math., pages 136-167. Springer, Berlin (1997).

[26] Miclo (L.). — On eigenfunctions of Markov processes on trees. Probab. Theory Related Fields, 142(3-4), p. 561-594 (2008).

[27] Miclo (L.). — On hyperboundedness and spectrum of Markov operators. Available at `http://hal.archives-ouvertes.fr/hal-00777146`, January 2013.

[28] Saloff-Coste (L.). — Lectures on finite Markov chains. In Lectures on probability theory and statistics (Saint-Flour, 1996), volume 1665 of Lecture Notes in Math., p. 301-413. Springer, Berlin (1997).

[29] van Doorn (E. A.). — Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. Adv. in Appl. Probab., 23(4), p. 683-700 (1991).

[30] van Doorn (E. A.) and Pollett (P. K.). — Quasi-stationary distributions for discrete-state models. European J. Oper. Res., 230(1), p. 1-14 (2013).

[31] van Doorn (E. A.) and Zeifman (A. I.). — On the speed of convergence to stationarity of the Erlang loss system. Queueing Syst., 63(1-4), p. 241-252 (2009).

[32] Zhou (H.). — Examples of multivariate Markov chains with orthogonal polynomial eigenfunctions. Ph.D. thesis, dept. of statistics, Stanford University (2008).