A probabilistic approach to Carne's bound

Rémi Peyre

October 15, 2007

Abstract

Carne's bound is a sharp inequality controlling the transition probabilities for a discrete reversible Markov chain (§ 1). Its ordinary proof (§ 2) uses spectral techniques which look as efficient as miraculous. Here we present a new proof, comparing a "drift" for ways "out" and "back", to get the gaussian part of the bound (§ 3), and using a conditioning technique to get the flight factor (§ 5). Moreover we show how our proof is more "supple" than Carne's one and may generalize (§ 4.2).

1 Introduction

1.1 The Markov chain

Let V be a finite or countable set of points. Let us consider an irreducible Markov chain $(X_t)_{t\in\mathbb{N}}$ on V, with transition kernel $(p(x,y))_{x,y\in V}$, and whose law is denoted by $\mathbb{P}_x$ when starting at x. That chain is supposed to be *reversible*, i.e. we suppose that there exists a measure μ on V such that, for all $x \in V$ $0 < \mu(x) < \infty$, and

$$\forall x, y \in V \quad \mu(x)p(x,y) = \mu(y)p(y,x). \tag{1.1}$$

By irreducibility, μ is then uniquely determined up to a multiplicative factor; in the sequel, we shall suppose it fixed. Note that we do not demand μ to be finite.

Then one may associate to the kernel a (non-oriented) graph (V,E) with vertices set V by defining the set of edges through

$$\{x,y\} \in E \Leftrightarrow p(x,y) \neq 0. \tag{1.2}$$

(*A priori* that definition should determine an *oriented* graph, but actually $p(x,y) \neq 0 \Leftrightarrow p(y,x) \neq 0$ by reversibility). As usual, we shall write $z \sim z'$ to mean that $\{z,z'\} \in V^{(1)}$. The graph distance, denoted by d, will stand for the length of the shortest path(s) in E joining two points. Speaking in terms of probability, one has:

$$d(x,y) = \inf\big\{t \in \mathbb{N}\,;\, p^t(x,y) \neq 0\big\}, \tag{1.3}$$

where p^t denotes the t-th convolution power of the kernel p.

This paper aims at explaining by probabilistic arguments an inequality due to Carne to sharply bound $p^t(x,y)$ above when $d(x,y) \gtrsim \sqrt{t}$. Indeed to the best of our knowledge, all the methods developed so far to get that kind of bounds used spectral analysis techniques [1, 2]. We shall also show how our probabilistic approach allows us to generalize Carne–Varopoulos type bounds for more "flexible" distances than the graph distance.

(1) Look out for the fact that $\sim$ is *not* an equivalence relation.

1.2 Carne's bound and its history

In 1985, N. Th. Varopoulos [1] was the first to give a concentration result bounding $p^t(x,y)$ above for a reversible Markov chain, whose leading term was $\exp\left(-\frac{d(x,y)^2}{Ct}\right)$, $C>0$ being an explicit constant depending on the transition kernel p. His method introduced a time-continuous Markov process on the cabled graph associated to (V,E), and studied the spectral properties of that process in an L^2 space. Moreover that proof required extra assumptions about the transition kernel.

The same year, T. K. Carne [2], by a simpler spectral method, got a finer result under the general assumptions stated in § 1.1:

Theoreme 1.1 (Carne 1985). *Suppose the hypotheses of § 1.1 are satisfied. Denote by P the $L^2(\mu)$-operator associated to the transition kernel p and let $|P|$ stand for its norm, which is always $\leqslant 1$ (see a more precise definition in § 2.1). Then:*

$$p^t(x,y) \leqslant 2\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P|^t \exp\left(\frac{-d(x,y)^2}{2t}\right). \tag{1.4}$$

My work was motivated by two goals: first, find a proof of theorem 1.1 which would be more natural than the original proof of Carne, then, adapt Carne–Varopoulos type bounds to distances which depend continuously on the transition kernel (see § 4.2).

2 Carne's proof

We give here the proof of [2] as it was exposed in [3].

2.1 Norm of the transition kernel

Let us first give a precise definition of P:

Definition 2.1. *P is the operator induced by $\mathbb{P}$ on $L^2(\mu)$ through:*

$$Pf(x) = \mathbb{E}_x[f(X_1)] = \sum_{y\sim x} p(x,y)f(y), \tag{2.1}$$

Then we define $|P|$ as the operator norm of P in $L^2(\mu)$, i.e. $|P| = \sup_{\|f\|_{L^2(\mu)}=1} \|Pf\|_{L^2(\mu)}$. Note that P is self-adjoint by reversibility of μ, and $|P| \leqslant 1$ by Jensen's inequality.

A more intrinsic defintion of $|P|$ is given by the following classical

Lemma 2.2 ([3, chap. 5-2]). *For any $x \in V$,*

$$|P| = \limsup_{t\longrightarrow\infty} \left(p^t(x,x)\right)^{1/t} = \sup_{t\geqslant 1} \left(p^t(x,x)\right)^{1/t}. \tag{2.2}$$

2.2 Chebychev's polynomials

Since $P^n f(x) = \mathbb{E}_x[f(X_n)]$, one can write

$$p^t(x,y) = \left\langle \frac{\boldsymbol{\delta}_x}{\mu(x)}, P^t\boldsymbol{\delta}_y \right\rangle_{L^2(\mu)} = \frac{|P|^t}{\mu(x)} \left\langle \boldsymbol{\delta}_x, \left(\frac{P}{|P|}\right)^t \boldsymbol{\delta}_y \right\rangle_{L^2(\mu)}. \tag{2.3}$$

The trick then consists in decomposing the polynomial Z^t in the basis of Chebychev's polynomials. The following results are classical:

Lemma 2.3. *For any $k \in \mathbb{Z}$, there exists a unique polynomial $Q_k(Z)$ satisfying*

$$\forall \theta \in \mathbb{C} \quad Q_k(\cos\theta) = \cos(k\theta), \tag{2.4}$$

called the k-th (first type) Chebychev polynomial. *It satisfies:*

1. $\deg Q_k = |k|$*;*
2. $|x| \leqslant 1 \quad \Rightarrow \quad |Q(x)| \leqslant 1$*;*
3. $\forall t \in \mathbb{N} \quad Z^t = \frac{1}{2^t}\sum_{k\in\mathbb{Z}} \binom{t}{(t+k)/2} Q_k(Z)$*, where by convention* $\binom{t}{p} = 0$ *whenever* $p \notin \{0, 1, \ldots, t\}$.

By property 3 in Lemma 2.3, formula (2.3) gives

$$p^t(x,y) = \frac{|P|^t}{2^t \mu(x)} \sum_{k\in\mathbb{Z}} \binom{t}{(t+k)/2} \left\langle \boldsymbol{\delta}_x, Q_k\left(\frac{P}{|P|}\right)\boldsymbol{\delta}_y \right\rangle_{L^2(\mu)}. \tag{2.5}$$

The linear operator $\frac{P}{|P|}$ on $L^2(\mu)$ is self-adjoint and its norm is 1 by construction; so it decomposes onto a countable orthonormal basis of eigenvectors as

$$\frac{P}{|P|}\left(\sum_{\lambda\in\mathrm{Spec}(P/|P|)} a_\lambda \boldsymbol{v}_\lambda\right) = \sum_{\lambda\in\mathrm{Spec}(P/|P|)} \lambda a_\lambda \boldsymbol{v}_\lambda, \tag{2.6}$$

where $\boldsymbol{v}_\lambda$ is the eigenvector associated to the eigenvalue λ, the eigenvalues being counted with multiplicity. By definition of $|P|$ we have Spec $\left(\frac{P}{|P|}\right) \subset [-1,1]$. So

$$Q_k\left(\frac{P}{|P|}\right)\left(\sum_{\lambda\in\mathrm{Spec}(P/|P|)} a_\lambda \boldsymbol{v}_\lambda\right) = \sum_{\lambda\in\mathrm{Spec}(P/|P|)} Q_k(\lambda) a_\lambda \boldsymbol{v}_\lambda, \tag{2.7}$$

where the $Q_k(\lambda)$ are all of absolute value less than or equal to 1 by lemma 2.3, property 2. Then, the operator norm of $Q_k\left(\frac{P}{|P|}\right)$ on $L^2(\mu)$ is at most 1, so to write:

$$\left\langle \boldsymbol{\delta}_x, Q_k\left(\frac{P}{|P|}\right)\boldsymbol{\delta}_y \right\rangle_{L^2(\mu)} \leqslant \|\boldsymbol{\delta}_x\|_{L^2(\mu)} \cdot \|\boldsymbol{\delta}_y\|_{L^2(\mu)} = \sqrt{\mu(x)\mu(y)}. \tag{2.8}$$

Now, we notice that for $|k| < d(x,y)$, $Q_k\left(\frac{P}{|P|}\right)\boldsymbol{\delta}_y$ is a linear combination of the $P^u\boldsymbol{\delta}_y, 0 \leqslant u < d(x,y)$, by property 1 in Lemma 2.3; then $Q_k\left(\frac{P}{|P|}\right)\boldsymbol{\delta}_y$ is a function supported by the $z \in V$ satisfying $d(z,y) < d(x,y)$, in particular $\left\langle \boldsymbol{\delta}_x, Q_k\left(\frac{P}{|P|}\right)\boldsymbol{\delta}_y \right\rangle_{L^2(\mu)} = 0$. In the end,

$$\begin{aligned} p^t(x,y) &= \frac{|P|^t}{2^t\mu(x)} \sum_{|k|\geqslant d(x,y)} \binom{t}{(t+k)/2} \left\langle \boldsymbol{\delta}_x, Q_k\left(\frac{P}{|P|}\right)\boldsymbol{\delta}_y \right\rangle_{L^2(\mu)} \\ &\overset{(2.8)}{\leqslant} \frac{|P|^t}{2^t\mu(x)} \sqrt{\mu(x)\mu(y)} \sum_{|k|\geqslant d(x,y)} \binom{t}{(t+k)/2} \\ &\leqslant 2|P|^t \left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} \frac{1}{2^t} \sum_{k\geqslant d(x,y)} \binom{t}{(t+k)/2} \end{aligned} \tag{2.9}$$

(where the last inequality is an equality as soon as $d(x,y) > 0$).

To conclude, it only remains to prove the relation

$$\frac{1}{2^t}\sum_{k\geqslant d(x,y)} \binom{t}{(t+k)/2} \leqslant \exp(-d(x,y)^2/2t). \tag{2.10}$$

To do that, we notice that, if X is a random variable equidistributed on $\{-1,1\}$, then, the law of X^{*t} (which denotes the t-th convolution power of X) is $\frac{1}{2^t}\sum_{k\in\mathbb{Z}}\binom{t}{(t+k)/2}\delta_k$, so

$$\frac{1}{2^t}\sum_{k\geqslant d(x,y)}\binom{t}{(t+k)/2}=\mathbb{P}\left(X^{*t}\geqslant d(x,y)\right). \tag{2.11}$$

Now, we check by direct computation that for all $\lambda>0$, $\mathbb{E}\left[e^{\lambda X}\right]\leqslant e^{\lambda^2/2}$, hence $\mathbb{E}\left[e^{\lambda X^{*t}}\right]\leqslant e^{t\lambda^2/2}$, and then by Chebychev's inequality:

$$\mathbb{P}\left(X^{*t}\geqslant d(x,y)\right)=\mathbb{P}\left(e^{\lambda X^{*t}}\geqslant e^{\lambda d(x,y)}\right)\leqslant\frac{e^{t\lambda^2/2}}{e^{\lambda d(x,y)}}, \tag{2.12}$$

hence we get (2.10) by taking $\lambda=d(x,y)/t$, which ends the proof.

3 The Gaussian factor

As told before, this article presents a new, probabilistic proof of Carne's bound. In this section, only the Gaussian part of the bound will be considered. The fundamental estimate is the

Theoreme 3.1. *Let $\mathbb{P}$ be a Markov chain as described in § 1.1; let $t\geqslant 2$; let $x\neq y\in V$; then*

$$p^t(x,y)\leqslant\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2}\exp\left(-\frac{(d(x,y)-1)^2}{2(t-1)}\right). \tag{3.1}$$

The following immediate corollary yields a more pleasant formula:

Corollary 3.2. *For $t\geqslant 1$ and $x,y\in V$,*

$$p^t(x,y)\leqslant\sqrt{e}\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2}\exp\left(-\frac{d(x,y)^2}{2t}\right). \tag{3.2}$$

Remark 3.3. The first factor in the bound (3.2) is slightly better than that of (1.4), but actually one could replace the 2 by a $\sqrt{e}$ in the proof of § 2 by refining the bound (2.10).

Proof. Denote $d=d(x,y)$. First, note that by reversibility of the chain, one has

$$p^t(y,x)=\frac{\mu(x)}{\mu(y)}p^t(x,y); \tag{3.3}$$

so to prove (3.1) it suffices to show:

$$p^t(x,y)\,p^t(y,x)\leqslant\exp\left(-\frac{(d-1)^2}{t-1}\right). \tag{3.4}$$

Now, rather than reasoning on the graph, which is a "complicated" object, we shall introduce a function $\xi:V\longrightarrow\mathbb{R}$ which measures how much the random walk X is closer to x or to y. ξ must satisfy:

Assumption 3.4.

- *$\xi(x)=0$; $\xi(y)=d$;*
- *ξ is 1-Lipschitz, i.e. for $z\sim v$ we have $|\xi(v)-\xi(z)|\leqslant 1$.*

Such a map ξ always exists since, for instance, the map $d(x, \cdot)$ always satisfies assumption 3.4. Each point of V tends to make ξ increase or decrease, depending on the values of the transition kernel. Let us denote by $m(z)$ the expected value for the variation of ξ after the particle having visited z, i.e.

$$m(z) = \mathbb{E}_z\left[\xi(X_1)\right] - \xi(z). \tag{3.5}$$

Now, let $(M_u)_{u \geqslant 1}$ be the process defined by:

$$M_u = \xi(X_u) - \xi(X_1) - \sum_{s=1}^{u-1} m(X_s) = \sum_{s=1}^{u-1} \left(\xi(X_{s+1}) - \xi(X_s) - m(X_s)\right) ; \tag{3.6}$$

obviously M is a martingale starting at 0. Let us look at the chain starting at x. On the event $\{X_t = y\}$, one trivially has $\xi(X_t) - \xi(X_1) \geqslant d - 1$, hence

$$\mathbb{E}_x\left[M_t \middle| X_t = y\right] \geqslant d - 1 - \mathbb{E}_x\left[\sum_{u=1}^{t-1} m(X_u) \middle| X_t = y\right]. \tag{3.7}$$

One may carry out the same reasoning starting at y, which gives:

$$\mathbb{E}_y\left[M_t | X_t = x\right] \leqslant -(d-1) - \mathbb{E}_y\left[\sum_{u=1}^{t-1} m(X_u) \middle| X_t = x\right]. \tag{3.8}$$

What can we see? If the terms $\mathbb{E}_x\left[\sum_{u=1}^{t-1} m(X_u) \middle| X_t = y\right]$, resp. $\mathbb{E}_y\left[\sum_{u=1}^{t-1} m(X_u) \middle| X_t = x\right]$ were not present in (3.7) and (3.8), these formulae would reduce to $\mathbb{E}_x\left[M_t | X_t = y\right] \geqslant d-1$, resp. $\mathbb{E}_y\left[M_t | X_t = x\right] \leqslant -(d-1)$, so that we would observe a large deviation phenomenon on martingales, which would yield a control respectively on $p^t(x, y)$ and $p^t(y, x)$. Unfortunately, that phenomenon seems to be wiped out because of the terms $m(X_s)$. The key idea then consists in noticing that, by reversibility, these $m(X_s)$ are the same for the "way out" as for the "way back"; subsequently, if the $m(X_s)$ tend to make the right hand side of (3.7) diminish (which would *damp* the large deviation phenomenon), then they tend to make the right hand side of (3.8) increase, which this time translates into a *strengthening* of the large deviation phenomenon. So, $p^t(x, y)$ and $p^t(y, x)$ cannot be large simultaneously, which will lead us to (3.4).

So, we consider X^x, X^y two independent chains with respective laws $\mathbb{P}_x$ and $\mathbb{P}_y$; let $\mathbb{P}_{x \otimes y}$ be their joint law. The respective realizations of $(M_u)_{u \geqslant 1}$ for the paths starting at x and at y are denoted by $(M_u^x)_{u \geqslant 1}$ and $(M_u^y)_{u \geqslant 1}$. By reversibility,

$$\forall u \in \{1, \dots, t-1\} \quad \mathbb{E}_x\left[m(X_u) \middle| X_t = y\right] = \mathbb{E}_y\left[m(X_{t-u}) \middle| X_t = x\right]. \tag{3.9}$$

Hence by combining (3.7) and (3.8),

$$\mathbb{E}_{x \otimes y}\left[M_t^x - M_t^y \middle| X_t^x = y \text{ et } X_t^y = x\right] \geqslant 2(d-1). \tag{3.10}$$

It remains to control the deviations of $M_t^x - M_t^y$. We remark that this random variable may be interpreted as the final value of a $2(t-1)$ steps martingale, whose steps satisfy the assumptions of the following lemma:

Lemma 3.5. *Let $(\mathcal{F}_t)_{t \in \mathbb{N}}$ be a filtration; let $(X_t)_{t \geqslant 1}$ be an adapted real-valued process with $\mathbb{E}[X_{t+1} | \mathcal{F}_t] = 0$[(2)]. We suppose that, for all $t \in \mathbb{N}$, $\mathcal{L}aw(X_{t+1} | \mathcal{F}_t)$ is supported by an interval of length 2 almost surely. Then, letting $u \geqslant 0$ be a fixed time, we have for all $\lambda \in \mathbb{R}$:*

$$\mathbb{E}\left[\exp\left(\lambda \sum_{t=1}^{u} X_t\right)\right] \leqslant \exp\left(u \frac{\lambda^2}{2}\right). \tag{3.11}$$

[(2)]In other words, the X_t's are the increments of a martingale.

Proof. The proof relies on Hoeffding's inequality, whose statement is recalled below:

Lemma 3.6 (Hoeffding). *Let X be a centered real-valued random variable, supported by an interval of length 2, then*

$$\forall\lambda\in\mathbb{R}\quad \mathbb{E}\left[e^{\lambda X}\right]\leqslant e^{\lambda^2/2}. \tag{3.12}$$

That point being taken for granted, we prove lemma 3.5 by induction on u:

- For $u=0$ the result is trivial.
- Let $u\geqslant 1$; suppose the result to be true for $u-1$. Let $\lambda\in\mathbb{R}$; we write

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{t=1}^{u}X_t\right)\right] &= \mathbb{E}\left[\exp\left(\lambda\sum_{t=1}^{u-1}X_t\right)\mathbb{E}\left[e^{\lambda X_u}|\mathcal{F}_{u-1}\right]\right]\\ &\leqslant \underbrace{\mathbb{E}\left[\exp\left(\lambda\sum_{t=1}^{u-1}X_t\right)\right]}_{\leqslant e^{(u-1)\lambda^2/2}\text{ by induction}}\cdot\underbrace{\left\|\mathbb{E}\left[e^{\lambda X_u}|\mathcal{F}_{u-1}\right]\right\|_\infty}_{\leqslant e^{\lambda^2/2}\text{ by (3.12)}}\leqslant e^{u\lambda^2/2}.\end{aligned} \tag{3.13}$$

□

To conclude, it only remains to us to prove the following measure concentration lemma:

Lemma 3.7. *Let X be a centered real-valued random variable satisfying for some $k>0$:*

$$\forall\lambda\in\mathbb{R}\quad \mathbb{E}[e^{\lambda X}]\leqslant e^{k\lambda^2/2}. \tag{3.14}$$

If $\mathcal{A}$ is an event such that

$$\mathbb{E}[X|\mathcal{A}]\geqslant C \tag{3.15}$$

for some $C\geqslant 0$, then

$$\mathbb{P}(\mathcal{A})\leqslant\exp\left(-\frac{C^2}{2k}\right). \tag{3.16}$$

Proof. To lighten notations, let us denote $p=\mathbb{P}(\mathcal{A})$. Let us fix $\lambda>0$, then we have

$$\mathbb{E}\left[e^{\lambda X}|\mathcal{A}\right]=\frac{\mathbb{E}\left[\mathbb{1}_{\mathcal{A}}e^{\lambda X}\right]}{\mathbb{P}(\mathcal{A})}\overset{(3.14)}{\leqslant}\frac{1}{p}e^{k\lambda^2/2}. \tag{3.17}$$

It follows, by Jensen's inequality, that

$$\mathbb{E}[X|\mathcal{A}]\leqslant\frac{1}{\lambda}\ln\left(\frac{1}{p}e^{k\lambda^2/2}\right), \tag{3.18}$$

hence by assumption (3.15):

$$\frac{1}{\lambda}\ln\left(\frac{1}{p}e^{k\lambda^2/2}\right)\geqslant C, \tag{3.19}$$

whence finally

$$p\leqslant e^{-C\lambda+k\lambda^2/2}. \tag{3.20}$$

Then it suffices to take $\lambda=C/k$ to get the announced result. □

Now we conclude the proof of Theorem 3.1. Lemma 3.5 permits us to control the Laplace transform of $M_t^x-M_t^y$ under $\mathbb{P}_{x\otimes y}$:

$$\forall\lambda\in\mathbb{R}\quad \mathbb{E}_{x\otimes y}\left[e^{\lambda(M_t^x-M_t^y)}\right]\leqslant e^{(t-1)\lambda^2}, \tag{3.21}$$

and Formula (3.10) then gives, *via* Lemma 3.7:

$$\mathbb{P}_{x\otimes y}(X_t^x=y\text{ et }X_t^y=x)\leqslant\exp\left(-\frac{(d-1)^2}{t-1}\right), \tag{3.22}$$

i.e. (3.4). □

4 Generalization to a larger class of distances

4.1 Statement of the generalized theorem

Now we will show that the reasoning made above can in fact adapt to a whole class of distances. So let us consider a new distance on V, which we will also call d —to avoid confusions, the graph distance that we had defined by (1.3) will be denoted d^{G} from now on. We have:

Theoreme 4.1. *Suppose that d is built so that, if $\xi : V \longrightarrow \mathbb{R}$ is any 1-Lipschitz function with respect to d, one has, for all $x \in V$:*

- $$\big|\mathbb{E}_x\left[\xi(X_1)\right] - \xi(x)\big| \leqslant B\,; \tag{4.1}$$

- $$\forall \lambda \geqslant 0 \quad \mathbb{E}_x\left[e^{\lambda\{\xi(X_1)-\mathbb{E}_x[\xi(X_1)]\}}\right] \leqslant e^{A\lambda^2}, \tag{4.2}$$

for some constants A and B independent of ξ. Then, for all $x, y \in V$, one has

$$p^t(x,y) \leqslant \left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} \exp\left(-\frac{(d(x,y)-B)_+{}^2}{4At}\right) \leqslant e^{B/2A}\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} \exp\left(-\frac{d(x,y)^2}{4At}\right). \tag{4.3}$$

Remark 4.2. We can already point out that, in the case when the distance is d^{G}, it is impossible to get anything better than $A = 1/2$ and $B = 1$. Subsequently, the result will be worsened by a $\sqrt{e}$ factor compared to (3.2) —which remains negligible compared to the exponential part of the bound—: as we will see later, it is due to the difference in treating the first steps, since the specific argument that we have used for d^{G} in the proof of Theorem 3.1 may not generalize.

Proof. We follow again the proof of theorem 3.1: denote $d = d(x,y)$, let ξ satisfy assumption 3.4, define m by (3.5) and let $\mathbb{P}_{x\otimes y}$ be the joint law of two independent chains of respective laws $\mathbb{P}_x$ and $\mathbb{P}_y$; we want to bound $\mathbb{P}_{x\otimes y}[X_t^x = y \text{ et } X_t^y = x]$ above to conclude by formula (3.3).

The first difference lies in the defintion of M: now, the martingale starts at time 0. So it is defined by:

$$M_u = \xi(X_u) - \xi(X_0) - \sum_{s=0}^{u-1} m(X_s) = \sum_{s=0}^{u-1} \big(\xi(X_{s+1}) - \xi(X_s) - m(X_s)\big). \tag{4.4}$$

Then we get:

$$\begin{cases} \mathbb{E}_x\left[M_t \middle| X_t = y\right] = d - \mathbb{E}_x\left[\sum_{u=0}^{t-1} m(X_u)\middle| X_t = y\right], \\ \mathbb{E}_y\left[M_t \middle| X_t = x\right] = -d - \mathbb{E}_y\left[\sum_{u=0}^{t-1} m(X_u)\middle| X_t = x\right] \end{cases} \tag{4.5}$$

When we want to combine these two formulae as we did in (3.10), we observe that all the terms $\mathbb{E}[m(X_u)]$ will cancel pairwise, *except* the terms corresponding to the first steps, i.e. to $u = 0$ in the two respective formulae. But we know exactly what these terms are, since under $\mathbb{P}_x$, we have $X_0 = x$ a.s. (by definition!), resp. $X_0 = y$ a.s. under $\mathbb{P}_y$. Thus

$$\mathbb{E}_{x\otimes y}\left[M_t^x - M_t^y \middle| X_t^x = y \text{ et } X_t^y = x\right] = 2d - m(x) + m(y) \geqslant 2d - 2B. \tag{4.6}$$

Taking into account assumption (4.2) —which plays here the role played before by Hoeffding's inequality—, we copy off the proof of Lemma 3.5 to get:

$$\forall \lambda \geqslant 0 \quad \mathbb{E}_{x\otimes y}\left[e^{\lambda(M_t^x+M_t^y)}\right] \leqslant e^{2At\lambda^2}, \tag{4.7}$$

and it only remains to conclude by Lemma 3.7[(3)]. □

4.2 More flexible distances

Now we will show how one may build distances statisfying Theorem 4.1, such that the metric structure of V continuously depends on the transition kernel. The method developed below is certainly neither the best nor the most elegant, but it has the advantage to be of relative pedagogical simplicity.

We keep on the principle of putting a length to each edge, but that time all the edges will not have the same size: indeed we will put a larger length to the edges that are more difficult to visit, in order to ensure that the metric structure of the graph will not be too much perturbed when we add a very "unlikely" edge.

Let $\alpha > 0$ be an arbitrary parameter. To each couple $(x, y) \in V \times V$, we associate a length $\ell(x, y)$ such that:

$$\forall a \geqslant 0 \quad \mathbb{P}_x\left(\ell(x, X_1) \geqslant 1 + a\right) \leqslant e^{-a^2/\alpha}, \tag{4.8}$$

and we define the length of the edge $[xy]$ by $|[xy]| = \min\{\ell(x, y), \ell(y, x)\}$. Then, for any 1-Lipschitz function ξ on V, we have:

Claim 4.3.

$$\forall x \in V \quad \forall a \geqslant 0 \quad \mathbb{P}_x\left(|\xi(X_1) - \xi(x)| \geqslant 1 + a\right) \leqslant e^{-a^2/\alpha}. \tag{4.9}$$

Now we give a formula for $\ell(x, y)$ satisfying (4.8). First we define what we will call the β-entropy of a probability law:

Definition 4.4. *Let $\beta \in]0, 1]$; let p be a probability measure on a discrete space $\mathcal{X}$. We call β-entropy of p the (possibly infinite) number:*

$$H_\beta(p) = \sum_{x\in\mathcal{X}} p(x)^{1-\beta\text{(4)}}. \tag{4.10}$$

A transition kernel p on V being given, we will also denote, for $x \in V$, $H_\beta(x) = H_\beta\left(p(x, \cdot)\right)$.

The β-entropy permits us to control the probability that the observed event is rare:

Claim 4.5. *Let $\beta \in]0, 1]$; let $\mathbb{P}$ be a probability law on a discrete state space $\mathcal{X}$. We suppose $H_\beta(p) < \infty$. Then, for all $\varpi \in]0, 1]$, one has:*

$$\mathbb{P}\left(p(x) \leqslant \varpi\right) \leqslant H_\beta(p)\varpi^\beta. \tag{4.11}$$

Proof. Use the identity $H_\beta(p) = \mathbb{E}[p(x)^{-\beta}]$ and the fact that the map $\varpi \mapsto \varpi^{-\beta}$ is decreasing, then apply Markov's inequality. □

So, one may choose the following expression for $\ell(x, y)$ to satisfy (4.8), where we set that, for $a < 0$, $a^{1/2} = 0$:

$$\ell(x, y) = 1 + \sqrt{\alpha}\Big(\beta \ln(p(x, y)^{-1}) - \ln H_\beta(x)\Big)^{1/2}. \tag{4.12}$$

Now, we want to show that (4.9) permits us to get (4.1) and (4.2) indeed. Let us begin with an easy observation:

(3) If it occurs that $B \geqslant d$, Lemma 3.7 cannot be applied and we may only bound $p^t(x, y)p^t(y, x)$ above by 1; this explains the positive part appearing in (4.3).

(4) For $\beta = 1$ we set by continuity $H_\beta(p) = \#\{x \in \mathcal{X} \,;\, p(x) > 0\}$.

Claim 4.6. *Let ξ be a 1-Lipchitz function on V for a distance d built as above. Then there exists a random variable Y whose repartition map satisfies $\forall a \geqslant 0$ $\mathbb{P}(Y \geqslant 1+a) = e^{-a^2/\alpha}$, i.e. Y has a law with density*

$$\mathrm{d}\mathbb{P}(Y=y) = \mathbb{1}_{y\geqslant 1} \frac{2(y-1)}{\alpha} e^{-(y-1)^2/\alpha} \mathrm{d}y, \tag{4.13}$$

such that one has:

$$\mathbb{P}_x\text{-p.s.} \quad |\xi(X_1) - \xi(x)| \leqslant Y. \tag{4.14}$$

So, we easily find that we can take $B = \mathbb{E}[Y] = 1 + \frac{\sqrt{\pi}}{2}\alpha$ into (4.1).

To get a formula for A in (4.2), things are a bit more complicated. The tool which we will use is the

Lemma 4.7. *Let Y be a positive random variable whose Laplace tranform $\widehat{Y}(\lambda) = \mathbb{E}\left[e^{\lambda Y}\right]$ is supposed to be finite for all $\lambda \geqslant 0$, and let us denote $\overline{Y} = \mathbb{E}[Y]$. Let X be a real-valued random variable satisfying $|X| \leqslant Y$ a.s.; let us denote $\overline{X} = \mathbb{E}[X]$ and $\widetilde{X} = X - \mathbb{E}[X]$. Then the Laplace transform of $\widetilde{X}$ satisfies:*

$$\forall \lambda \in \mathbb{R} \quad \widehat{\widetilde{X}}(\lambda) \leqslant e^{|\lambda|\overline{Y}} \widehat{Y}(|\lambda|) - 2|\lambda|\overline{Y}. \tag{4.15}$$

Remark 4.8. The bound (4.15) is of quite poor quality close to 0, as it may be particularily striking in the case $Y \equiv 1$, where we get the bound $e^{2|\lambda|} - 2|\lambda|$, while we know (Hoeffding's lemma 3.6) that $e^{\lambda^2/2}$ would work. In fact, we can compute that in a neighborhood of 0, the right hand side of (4.15) takes the form:

$$1 + \left(\mathbb{E}[Y^2] + 3\mathbb{E}[Y]^2\right)\frac{\lambda^2}{2} + o(\lambda^2), \tag{4.16}$$

while a variance calculation proves that in fact,

$$\widehat{\widetilde{X}}(\lambda) \leqslant 1 + \mathbb{E}[Y^2]\frac{\lambda^2}{2} + o(\lambda^2). \tag{4.17}$$

Proof. We may restrict ourselves to the case $\lambda \geqslant 0$, the case $\lambda \leqslant 0$ being then treated by using the result for $-X$.
We write

$$\widehat{\widetilde{X}}(\lambda) = \mathbb{E}\left[e^{\lambda\widetilde{X}} - \lambda\widetilde{X}\right]. \tag{4.18}$$

But, since $|X| \leqslant Y$, we have $|\overline{X}| \leqslant \overline{Y}$, hence $\widetilde{X} \in [-Y-\overline{Y}, Y+\overline{Y}]$. And since, on an interval of the form $[-a,a]$, the map $x \mapsto e^{\lambda x} - \lambda x$ takes its maximum at a, it follows that

$$\widehat{\widetilde{X}}(\lambda) \leqslant \mathbb{E}\left[e^{\lambda(Y+\overline{Y})} - \lambda(Y+\overline{Y})\right] = e^{\lambda\overline{Y}}\widehat{Y}(\lambda) - 2\lambda\overline{Y}. \tag{4.19}$$

□

Now we have the following control on the Laplace transform of the random variable Y defined by (4.13):

Proposition 4.9. *If Y is a random variable whose law is given by (4.13), then for $\lambda \geqslant 0$ one has:*

$$\mathbb{E}\left[e^{\lambda Y}\right] \leqslant \left(1 + \frac{\sqrt{\pi}\alpha}{2}\lambda + \frac{\alpha^2}{4}\lambda^2\right) e^{\lambda + \alpha^2\lambda^2/4}. \tag{4.20}$$

Moreover, that bound is sharp close to 0, by which we mean that, if we rewrite (4.20) under the form $\mathbb{E}[e^{\lambda Y}] \leqslant f(\lambda)$, we have $\mathbb{E}[Y] = \frac{\mathrm{d}}{\mathrm{d}\lambda}\big|_{\lambda=0}\mathbb{E}[e^{\lambda Y}] = f'(0) = 1 + \frac{\sqrt{\pi}}{2}\alpha$.

Proof. We begin with noticing that we can write $Y = 1 + \alpha Z$, where Z is a random variable with law

$$\mathrm{d}\mathbb{P}(Z = z) = \mathbb{1}_{z \geqslant 0} 2ze^{-z^2}\mathrm{d}z. \tag{4.21}$$

Then it suffices to prove that

$$\forall \lambda \geqslant 0 \quad \mathbb{E}\left[e^{\lambda Z}\right] \leqslant \left(1 + \frac{\sqrt{\pi}}{2}\lambda + \frac{\lambda^2}{4}\right) e^{\lambda^2/4}. \tag{4.22}$$

To do that, we write:

$$\mathbb{E}\left[e^{\lambda Z}\right] = e^{\lambda^2/4} \int_0^\infty 2ze^{-(z-\lambda/2)^2}\mathrm{d}z \underset{t=z-\lambda/2}{=} e^{\lambda^2/4} \underbrace{\int_{-\lambda/2}^\infty (2t+\lambda)e^{-t^2}\mathrm{d}t}_{I(\lambda)}, \tag{4.23}$$

where $I(0) = 0$ and, by the theorem of differenciation under the integral,

$$I'(\lambda) = \int_{-\lambda/2}^\infty e^{-t^2}\mathrm{d}t \leqslant \frac{\sqrt{\pi}}{2} + \frac{\lambda}{2}, \tag{4.24}$$

whence $I(\lambda) \leqslant 1 + \frac{\sqrt{\pi}}{2}\lambda + \frac{\lambda^2}{4}$, and (4.20). □

From that we deduce the existence of a suitable value for A:

Proposition 4.10. *For all $\alpha > 0$, there exists a constant $A(\alpha) < \infty$ such that, if d satisfies condition (4.9) for the value α, then (4.2) is satisfied for the value $A(\alpha)$.*

Proof. For Y with law (4.13), denoting $\overline{Y} = \mathbb{E}[Y] = 1 + \frac{\sqrt{\pi}}{2}\alpha$, Proposition 4.9 shows that $\ln\left(e^{\lambda\overline{Y}}\mathbb{E}[e^{\lambda Y}] - 2\lambda\overline{Y}\right)/\lambda^2$ is bounded for $\lambda \xrightarrow{\geqslant} 0$ and $\lambda \longrightarrow \infty$. By continuity, this function is thus bounded on the whole half-line $[0, +\infty)$. Lemma 4.7 then gives the existence of A. □

Remark 4.11. We have not found any simple bound for $A(\alpha)$, but, for a given value of α, it is easy to compute numerically the maximum of the map $\lambda \mapsto \ln(e^{\lambda\overline{Y}}\mathbb{E}[e^{\lambda Y}] - 2\lambda\overline{Y})/\lambda^2$, which gives a suitable value for A.

4.3 A concrete example

We shall illustrate the preceding subsection by showing how our generalization gives some results in cases when the usual Carne bound is irrelevant.

Here we consider V to be the set of vertices of an inifinite 3-tree. The tree distance is denoted by d^{A}, *which will not be the same as the graph distance* d^{G}. Let L be an integer devised to become large and $\varepsilon > 0$ a real number devised to become small. We consider as the Markov chain on V the process which, from point x, jumps on each neighbor of x with probability $(1-\varepsilon)/3$, and, with probability ε, chooses uniformly an arrival point in the (closed) ball centered on x with radius L. As one may check immediately, that chain is reversible and its invariant measure is the counting measure. Let us sum up: our process looks much like the simple random walk on a 3-tree, but sometimes the mobile may jump by roughly L units. We would like to say that, even if L is large, it suffices for ε to be small enough to get an exponential bound where L does not appear.

If we naively apply formula (1.4) to this transition kernel, we will not get anything interesting: indeed, small as ε might be, the graph distance is the same:

$$\forall x, y \in G \quad d^{\mathrm{G}}(x, y) = \left\lceil \frac{d^{\mathrm{A}}(x, y)}{L} \right\rceil. \tag{4.25}$$

Then, bounding below d^{G} by d^{A}/L, and merely bounding $|P|$ by $1^{(5)}$, Carne's bound yields

$$p^t(x,y) \leqslant 2\exp\left(-\frac{d^{\mathrm{A}}(x,y)}{2tL^2}\right). \tag{4.26}$$

Concretely, if $L = 17$ and $\varepsilon = 1/2^{2^{30}}$, it will give a bound with a $d^{\mathrm{A}}(x,y)/578t$ in the exponential, which is strongly worse than the $d^{\mathrm{A}}(x,y)/2t$ of the case $\varepsilon = 0$. Yet it is obvious that the influence of large jumps should be nearly zero: the bound (4.26) thus must be improvable!

So we will apply the techniques exposed in § 4.2. Here we have chosen arbitrarily $\alpha = 1$ and $\beta = 1/2$. We suppose that L is large enough; in fact our computations will be valid as soon as $L \geqslant 2$. Let us denote by $N = 3 \cdot 2^L - 2$ the cardinality of a ball of radius L. We have:

$$H_{1/2}(x) = 3\left(\frac{1-\varepsilon}{3} + \frac{\varepsilon}{N}\right)^{1/2} + (N-3)\left(\frac{\varepsilon}{N}\right)^{1/2} \leqslant \sqrt{3} + \sqrt{N}\varepsilon^{1/2}, \tag{4.27}$$

hence $H_{1/2}(x) \leqslant 2$ for $\varepsilon \leqslant 1/(42 \cdot 2^L)$. So, if we bound over $H_{1/2}(x)$ by 2, we get that for $2 \leqslant d^{\mathrm{A}}(x,y) \leqslant L$, one has:

$$\ell(x,y) \geqslant 1 + \left(\ln(N/\varepsilon)/2 - \ln 2\right)^{1/2}, \tag{4.28}$$

in particular $\ell(x,y) \geqslant 1 + \sqrt{\ln(\varepsilon^{-1})/2}$.

Our observation is that, if ε is small enough, d coincides with d^{A}: indeed one has $\ell(x,y) = 1$ for $x \sim y$, $\ell(x,y) = \infty$ if $d^{\mathrm{A}}(x,y) > L$, and $\ell(x,y) \geqslant L$ for $2 \leqslant d^{\mathrm{A}}(x,y) \leqslant L$, as soon as:

$$\varepsilon \leqslant e^{-2L^2} \tag{4.29}$$

—which is however a quite strong condition. (Note that for L large enough, condition (4.29) implies that $\varepsilon \leqslant 1/(42 \cdot 2^L)$ indeed.)

Numerical computations for $\alpha = 1$ give $A = 8.09\ldots$, resp. $B = 1.88\ldots$; so we have for $\varepsilon \leqslant e^{-2L^2}$

$$p^t(x,y) \leqslant \frac{9}{8}\exp\left(-\frac{d^{\mathrm{A}}(x,y)^2}{33t}\right). \tag{4.30}$$

Although that bound undoubtedly improves Carne's bound (4.26) in "extreme" cases like that mentioned above, and even if it is certainly possible to get some better results by a more subtle choice of α and β, I find that bound rather disappointing in the sense that we still remain far from Carne's bound for the smallest values of ε. Anyway, Theorem 4.1 is theoretically interesting and may have better applications; in particular Lemma 4.7 can certainlty be improved.

5 The flight factor

5.1 Frame of the proof

The Gaussian bound (3.2) has got a disadvantage with respect to Carne's bound (1.4): in the case when $|P| < 1$, it does not show the exponential decreasing of $p^t(x,y)$ in the variable t. In fact, Lemma 2.2 implies $p^t(x,y)p^t(y,x) \leqslant p^{2t}(x,x) \leqslant |P|^{2t}$, whence

$$p^t(x,y) \leqslant \left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P|^t, \tag{5.1}$$

but that is not enough to get again (1.4). The present section precisely aims at doing this. Here we will exclusively focus on the case when $d = d^{\mathrm{G}}$, cf. Remark 5.8 below.

(5) Actually, here we could do better, but in this subsection we are only interested in the Gaussian part of the bound.

Let $x, y \in V$. For u a time devised to go to infinity, denote by

$$\mathcal{R}_u = \{\exists s \geqslant u\,;\; X_s = x\} \tag{5.2}$$

the event which tells that the particle comes back at x at least once after time u. The strategy of our proof then consists in looking at our Markov chain *conditioned to the event $\mathcal{R}_u$*. Why this? Well, the fact that $|P| < 1$ expresses a possibility for the particle to "flee to infinity". That flight is responsible for the exponential decay with respect to t of the quantity $p^t(x,y)p^t(y,x)$ introduced in (3.4), which measures the probability that the particle, starting at x, goes to y at time t and then comes back to x at time $2t$. Conditioning with respect to $\mathcal{R}_u$ then aims, in a way, at preventing the particle from going to infinity, which will give us a Markov chain for which $|P| = 1$, where the bound (3.2) will be relevant. Then it will remain to show that this conditioning selects sufficiently well the cases when the particle makes a return trip to get back a factor $|P|^t$ in (3.2).

Our proof will use a kind of density argument: in a first step we will add some more assumptions on our Markov chain to carry out the reasoning, then in a second step we will prove that we can get rid of these extra assumptions by slightly perturbing the original Markov chain.

5.2 Proof under extra assumptions

We will use the following notation:

Definition 5.1. *We denote by τ_x the hitting time of x by a walk on V, i.e. $\tau_x = \inf\{t \geqslant 0\,;\; X_t = x\}$. For all $z \in V$, we denote:*

$$R(z) = \mathbb{E}_z\left[\mathbb{1}_{\tau_x < \infty} |P|^{-\tau_x}\right]. \tag{5.3}$$

In this first part of the proof, we add to the assumptions of § 1.1 the following conditions:

Assumption 5.2.

1. *V is finite;*
2. *There exists a cemetery point $\partial \in V$ such that $p(\partial, \cdot) = \delta_\partial$. We will denote $\widetilde{V}$ for $V \setminus \{\partial\}$;*
3. *The chain $\mathbb{P}$ is aperiodic on $\widetilde{V}$.*

Remark 5.3. Under the assumption 5.2, the chain will just be required to be irreducible and reversible on $\widetilde{V}$; moreover the definition of $|P|$ will be that given by the formula (2.2) of lemma 2.2, for arbitrary $x \in \widetilde{V}$.

Assumption 5.2 permits to obtain sharp results about the recurrence behaviour of the chain:

Lemma 5.4. *Suppose we have an irreducible and reversible Markov chain satisfying Assumption 5.2. Then:*

1. *There exist two constants $0 < c_1 \leqslant c_2 < \infty$ such that*

$$\forall t \geqslant 0 \quad c_1 |P|^t \leqslant \mathbb{P}_x(\mathcal{R}_T) \leqslant c_2 |P|^t. \tag{5.4}$$

2. $\mathbb{P}_x(\mathcal{R}_{t+1}) / \mathbb{P}_x(\mathcal{R}_t) \underset{t \longrightarrow \infty}{\longrightarrow} |P|$.
3. *For all $z \in \widetilde{V}$, $R(z) < \infty$.*

The proof of this lemma, which is rather technical, is postponed to the Appendix.
Now we are armed to prove Carne's bound for a Markov chain satisfying Assumption 5.2. Let us fix $t \geqslant 0$; we have the key proposition:

Proposition 5.5. *The law of* $(X_s)_{0\leqslant s\leqslant 2t}$ *under* $\mathbb{P}_x(\cdot|\mathcal{R}_u)$ *converges when* $u \longrightarrow \infty$ *(for the total variation norm on* $V^{\{0,\dots,2t\}}$*) to the law* $\mathbb{P}'_x$ *of the Markov chain on* $\widetilde{V}$ *starting at* x*, with transition probabilities:*

$$\forall z, v \in \widetilde{V} \quad p'(z,v) = \frac{R(v)p(z,v)}{\sum_{w\sim z} R(w)p(z,w)}^{(6)}. \tag{5.5}$$

Proof. It is true in a general framework that $\mathbb{P}(\cdot|\mathcal{R}_u)$ is a time-inhomogeneous Markov chain with

$$\mathbb{P}\left(X_{s+1} = v \middle| X_s = z \text{ and } \mathcal{R}_u\right) = \frac{S_{s+1,u}(v)p(z,v)}{\sum_{w\sim z} S_{s+1,u}(w)p(z,w)}, \tag{5.6}$$

where we let $S_{s+1,u}(z) = \mathbb{P}(\mathcal{R}_u|X_{s+1} = z)$. Our attack will consist in proving that, for all $s \in \{0,\dots,2t-1\}$, for all $z \in V$, we have $S_{s+1,u}(z)/\mathbb{P}_x(\mathcal{R}_{u-(s+1)}) \underset{u\longrightarrow\infty}{\longrightarrow} R(z)$.

To begin with, let us notice that $S_{s+1,u}(z) = \mathbb{P}_z(\mathcal{R}_{u-(s+1)})$, which we shall denote by $S_{u-(s+1)}(z)$; then what we want to prove can be written:

$$\forall z \in \widetilde{V} \quad \frac{S_u(z)}{\mathbb{P}_x(\mathcal{R}_u)} \underset{u\longrightarrow\infty}{\longrightarrow} R(z). \tag{5.7}$$

The idea consists in splitting the probability space according to the value of τ_x, thanks to the strong Markov property:

$$S_u(z) = \sum_{s=0}^{u} \mathbb{P}_z(\tau_x = s)\mathbb{P}_x(\mathcal{R}_{u-s}) + \sum_{s\geqslant u+1} \mathbb{P}_z(\tau_x = s), \tag{5.8}$$

whence

$$\frac{S_u(z)}{\mathbb{P}_x(\mathcal{R}_u)} = \sum_{s=0}^{u} \mathbb{P}_z(\tau_x = s)\frac{\mathbb{P}_x(\mathcal{R}_{u-s})}{\mathbb{P}_x(\mathcal{R}_u)} + \sum_{s\geqslant u+1} \frac{\mathbb{P}_z(\tau_x = s)}{\mathbb{P}_x(\mathcal{R}_u)}. \tag{5.9}$$

Let us fix an arbitrarily small $\varepsilon > 0$. Since $\sum_s \mathbb{P}_z(\tau_x = s)|P|^{-s}$ converges (cf. Lemma 5.4-3), we may introduce a time u_0 for which one has $\sum_{s>u_0} \mathbb{P}_z(\tau_x = s) \leqslant \varepsilon$. Then, for $u \geqslant u_0$ one has on the one hand,

$$\sum_{s=u_0+1}^{u} \mathbb{P}_z(\tau_x = s)\frac{\mathbb{P}_x(\mathcal{R}_{u-s})}{\mathbb{P}_x(\mathcal{R}_u)} + \sum_{s>u} \frac{\mathbb{P}_z(\tau_x = s)}{\mathbb{P}_x(\mathcal{R}_u)} \overset{\text{lemma 5.4-1}}{\leqslant} \frac{c_2 \vee 1}{c_1} \sum_{s>u_0} \mathbb{P}_z(\tau_x = s)|P|^{-s} = \frac{c_2 \vee 1}{c_1}\varepsilon, \tag{5.10}$$

on the other hand,

$$\sum_{s=0}^{u_0} \mathbb{P}_z(\tau_x = s)\frac{\mathbb{P}_x(\mathcal{R}_{u-s})}{\mathbb{P}_x(\mathcal{R}_u)} \underset{u\longrightarrow\infty}{\overset{\text{lemma 5.4-2}}{\longrightarrow}} \sum_{s=0}^{u_0} \mathbb{P}_z(\tau_x = s)|P|^{-s}. \tag{5.11}$$

It follows that

$$\limsup_{u\longrightarrow\infty} \left| R(z) - \frac{S_u(v)}{\mathbb{P}_x(\mathcal{R}_u)} \right| \leqslant \left(1 + \frac{c_2 \vee 1}{c_1}\right)\varepsilon, \tag{5.12}$$

hence (5.7) by letting $\varepsilon \longrightarrow 0$. □

(6) That defines a Markov chain on $\widetilde{V}$ indeed because $R(\partial) = 0$.

Now we want to look at the chain $\mathbb{P}'$. First, $\mathbb{P}'$ is clearly irreducible. Then, one has:

Proposition 5.6. *The chain $\mathbb{P}'$ is reversible, and its invariant measure is:*

$$\forall z \in \widetilde{V} \quad \mu'(z) = \begin{cases} R(z)^2\mu(z) & \text{if } z \neq x; \text{ (7)} \\ R(x)R^+(x)\mu(x) & \text{if } z = x, \end{cases} \tag{5.13}$$

where $R^+(x)$ is defined by:

Definition 5.7. *We denote by τ_x^+ the* return time *to x, i.e.:*

$$\tau_x^+ = \inf\{s \geqslant 1\,;\, X_s = x\}. \tag{5.14}$$

Then $R^+(x)$ is defined by:

$$R^+(x) = \mathbb{E}_x\left[\mathbb{1}_{\tau_x^+<\infty}|P|^{-\tau_x^+}\right]. \tag{5.15}$$

Proof. Let $z, v \in \widetilde{V}$ with $z \neq x$. We can lighten the expression of $p'(z,v)$, since by Markov's property,

$$R(z) = \sum_{w\sim z} p(z,w)\mathbb{E}_w\left[\mathbb{1}_{\tau_x<\infty}|P|^{-(\tau_x+1)}\right] = |P|^{-1}\sum_{w\sim z} p(z,w)R(w), \tag{5.16}$$

thus (5.5) can be rewritten as

$$p'(z,v) = \frac{p(z,v)R(v)}{|P|R(z)}. \tag{5.17}$$

In the case when $z = x$, the same argument leads to

$$p^*(x,v) = \frac{p(x,v)R(v)}{|P|R^+(x)}. \tag{5.18}$$

So, it only remains to use (5.17), (5.18) and the reversibility of μ under $\mathbb{P}$ to get the reversibility of μ' under $\mathbb{P}'$. □

Now we are ready to end the proof. We observe that, by Markov's property, $p^t(x,y)p^t(y,x) = \mathbb{P}_x(X_t = y \text{ and } X_{2t} = x)$, which we will denote by $\mathbb{P}_x(\mathcal{A})$. For $u \geqslant 2t$, Bayes' formula gives:

$$\frac{\mathbb{P}_x(\mathcal{A})}{\mathbb{P}_x(\mathcal{A}|\mathcal{R}_u)} = \frac{\mathbb{P}_x(\mathcal{R}_u)}{\mathbb{P}_x(\mathcal{R}_u|\mathcal{A})} \overset{\text{Markov}}{=} \frac{\mathbb{P}_x(\mathcal{R}_u)}{\mathbb{P}_x(\mathcal{R}_{u-2t})}, \tag{5.19}$$

hence, letting u go to infinity:

$$\frac{\mathbb{P}_x(\mathcal{A})}{\mathbb{P}'_x(\mathcal{A})} = |P|^{2t}. \tag{5.20}$$

Now, the Markov chain $\mathbb{P}'$ satisfies the assumptions of theorem 3.1, hence $\mathbb{P}'_x(\mathcal{A}) \leqslant e\cdot\exp\left(-d(x,y)^2/t\right)$, so $\mathbb{P}_x(\mathcal{A}) \leqslant e|P|^{2t}\exp\left(-d(x,y)^2/t\right)$, and finally we get the desired formula:

$$p^t(x,y) \leqslant \sqrt{e}\left(\frac{\mu(x)}{\mu(y)}\right)^{1/2}|P|^t\exp\left(-\frac{d(x,y)^2}{2t}\right). \tag{5.21}$$

Remark 5.8. The reasoning carried out above cannot apply to other distances than d^{G}: indeed, the distance which appears in (5.21) in fact comes as the distance *associated to the process* $\mathbb{P}'$. When one works with the graph distance, that distance is the same for $\mathbb{P}$ and for $\mathbb{P}'$, but this is no more true if d depends in a more subtle way on the transition kernel.

(7) The careful reader may have noticed that $R(x) = 1$; we let that factor appear for ease of understanding.

5.3 The density argument

Now, we want to get rid of Assumption 5.2. We will proceed in two steps: first we will just relax Assumption 5.2-3, then we will deal with the general case.

Relaxing the aperiodicity condition

We consider a finite set V with a transition law $(p(x,y))_{x,y\in V}$, such that there exists a cemetery point $\partial \in V$ satisfying Assumption 5.2-2. We suppose that the Markov chain defined by p is irreducible and reversible on $\widetilde{V} = V \smallsetminus \{\partial\}$, with a reversible measure μ. We denote by n the cardinality of $\widetilde{V}$, and by M the matrix $\big((p(y,x))\big)_{x,y\in\widetilde{V}}$. The following lemma gives an algebraic characterization of the value $|P|$ defined in (2.2):

Lemma 5.9 ([3, chap. 5-2]). *$|P|$ is the spectral radius of M.*

For $\varepsilon \in [0,1[$, let p_ε be the transition kernel defined by:

$$\forall x,y \in V \quad p_\varepsilon(x,y) = \begin{cases} p(x,x) + \varepsilon(1-p(x,x)) & \text{if } y = x; \\ (1-\varepsilon)p(x,y) & \text{if } y \neq x. \end{cases} \tag{5.22}$$

The Markov chain $\mathbb{P}^\varepsilon$ generated by p_ε is an irreducible reversible chain whose graph and reversible measure are the same as for $p = p_0$, and which satisfies thye whole of Assumption 5.2 as soon as $\varepsilon > 0$. Thus, for $x,y \in \widetilde{V}$, $t > 0$ and $\varepsilon > 0$, we have

$$p_\varepsilon^t(x,y) \leqslant \sqrt{e}\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P_\varepsilon|^t \exp\left(-\frac{d(x,y)^2}{2t}\right). \tag{5.23}$$

To conclude, we just have to notice that $p_\varepsilon^t(x,y)$, resp. P_ε, are functions of ε continuous at 0. Indeed, the finite-sized matrix M_ε varies continuously with ε, thus its spectral radius $|P_\varepsilon|$ also varies continuously, as well as $p_\varepsilon^t(x,y)$ which is the coefficient number (y,x) of M_ε^t.

Infinite graphs

Now we turn to the general case, i.e. we consider a chain that merely satisfies the assumptions of § 1.1. Let us give a mark $\nu(z) > 0$ to each vertice z of V, in such a way that for all $\varepsilon > 0$, $\#\{z \in V\,;\, \nu(z) > 0\}$ is finite.

Let us fix $x,y \in V$, and let us take $\varepsilon > 0$ arbitrarily small (we shall always suppose $\varepsilon < \nu(x), \nu(y)$ to avoid certain problems). We define a finite set V_ε equipped with a transition kernel $(p_\varepsilon(z,v))_{z,v\in V_\varepsilon}$ through the following way:

Definition 5.10. *V_ε is obtained by identifying all the points with ν-mass less than ε to a cemetery point ∂:*

$$V_\varepsilon = \{z \in V\,;\, \nu(z) \geqslant \varepsilon\} \cup \{\partial\}. \tag{5.24}$$

From now on we will denote the points of V in the same way as their images on V_ε. Then p_ε is the kernel p projected on V_ε, with the requirement that ∂ is a cemetery point:

$$\forall z,v \in V_\varepsilon \quad p_\varepsilon(z,v) = \begin{cases} p(z,v) & \text{if } z,v \in \widetilde{V_\varepsilon}; \\ 0 & \text{if } z = \partial \text{ and } v \in \widetilde{V_\varepsilon}; \\ 1 & \text{if } z = v = \partial; \\ \sum_{\nu(w)<\varepsilon} p(z,w) & \text{if } z \in \widetilde{V_\varepsilon} \text{ and } w = \partial. \end{cases} \tag{5.25}$$

Then the chain $\mathbb{P}^\varepsilon$ satisfies points 1 and 2 of Assumption 5.2, and it is reversible with measure $\mu_{|\widetilde{V_\varepsilon}}$. This chain may not be irreducible, but we can suppose that such is the case by keeping only the irreducible component of $\widetilde{V_\varepsilon}$ containing x.

So the relation (5.21) is satisfied for V_ε equipped with $\mathbb{P}^\varepsilon$; it only remains to prove that $p_\varepsilon^t(x,y) \underset{\varepsilon \longrightarrow 0}{\longrightarrow} p^t(x,y)$, resp. $|P_\varepsilon| \underset{\varepsilon \longrightarrow 0}{\longrightarrow} |P|$.

Let us deal immediately with the operator norm. The very construction of p_ε ensures that for all $z, v \in \widetilde{V_\varepsilon}$, we have $p_\varepsilon^t(z,v) \leqslant p^t(z,v)$. Taking $z = v = x$, the characterization (2.2) of $|P|$ immediately gives that $|P_\varepsilon| \leqslant |P|$, which is enough for us (but convergence when $\varepsilon \longrightarrow 0$ is also true).

Now, we observe that the law of the t first steps of the chain generated by p_ε converges to the law of the initial chain in the sense of total variation. Indeed, given the way how V_ε and p_ε are constructed respectively from V and p, we have a canonical map which associates a walk on $V_{(\varepsilon)}$ to a walk on V, so that the law $\mathbb{P}_x$ maps into the law $\mathbb{P}_x^\varepsilon$. That map is defined as follows: the points of the walk on V are sent onto their projections on V_ε until the image walk hits ∂, and from that time on the walk stays at ∂. In particular, if a realization of the original chain stays in $\widetilde{V_\varepsilon}$ up to time t, its image by our map is kept safe on $\{0, \ldots, t\}$, and so

$$\begin{aligned} \left\| \mathbb{P}_{x\,|V_\varepsilon^{\{0,\ldots,t\}}}^\varepsilon - \mathbb{P}_{x\,|V^{\{0,\ldots,t\}}} \right\|_{\mathrm{TV}} &\leqslant \mathbb{P}_x \left(\exists u \in \{0, \ldots, t\}\ X_u \notin \widetilde{V_\varepsilon} \right) \\ &\leqslant \sum_{u=0}^{t} \mathbb{P}_x \left(\nu(X_u) < \varepsilon \right) \underset{\varepsilon \longrightarrow 0}{\overset{\mathrm{DCV}}{\longrightarrow}} 0. \end{aligned} \tag{5.26}$$

In particular, $p_\varepsilon^t(x,y) \underset{\varepsilon \longrightarrow 0}{\longrightarrow} p^t(x,y)$, and so (5.21) is satisfied for V equipped with p, QED.

A Appendix: Finite sub-Markov chains

This appendix aims at proving Lemma 5.4. Let us recall that in that lemma, we consider a sub-Markov chain on a finite graph $\widetilde{V}$, given by a kernel p, which is irreducible and aperiodic (the fact that the chain is reversible is not used in the proof of lemma 5.4). Denote by n the cardinality of $\widetilde{V}$.

The study of the chain may be expressed into matricial terms: we introduce the matrix

$$M = \big(\big(p(v,z)\big)\big)_{z,v \in \widetilde{V}}. \tag{A.1}$$

Then the aperiodicity condition translates into the existence of a time t_0 such that, for all $t \geqslant t_0$, M^{t_0} has strictly positive coefficients (actually $t_0 = n^2$ would always do). On the other hand, Lemma 5.9 above permits us to consider $|P|$ as the spectral radius of M.

So we have in hand all the assumptions to apply the strongest form of the Perron–Frobenius theorem, whose general statement and proof the reader can find in [4, chap. 5]:

Proposition A.1 (Perron–Frobenius). *$|P|$ is a simple eigenvalue of M, and all the other eigenvalues of M have an absolute value strictly less than $|P|$. Moreover, the eigenvector v associated to the eigenvalue $|P|$ has all its entries strictly positive.*

Now, let us begin with proving point 1 of Lemma 5.4, and in a first step let us prove the second inequality. Markov's strong property gives us the over-muliplicativity relation:

$$\forall t, u \geqslant 0 \quad \mathbb{P}_x(\mathcal{R}_{t+u}) \geqslant \mathbb{P}_x(\mathcal{R}_t)\mathbb{P}_x(\mathcal{R}_u). \tag{A.2}$$

We deduce that, for all $t \geqslant 1$, one has $\mathbb{P}_x(\mathcal{R}_t)^{1/t} \leqslant \limsup_{u \longrightarrow \infty} \mathbb{P}_x(\mathcal{R}_u)^{1/u}$. Moreover, if $|P| < 1$, we have, for all $t \geqslant 1$,

$$\mathbb{P}(\mathcal{R}_t) \leqslant \sum_{u \geqslant t} p^u(x,x) \overset{(2.2)}{\leqslant} \sum_{u \geqslant t} |P|^u = \frac{|P|^t}{1-|P|}, \tag{A.3}$$

hence $\limsup_{u \longrightarrow \infty} \mathbb{P}_x(\mathcal{R}_u)^{1/u} \leqslant |P|$, that last relation being also trivially true in the case when $|P| = 1$. Finally, the second inequality of (5.4) is satisfied for $t \geqslant 1$ with $c_2 = 1$, the case $t = 0$ being trivial.

For the lower bound, we will only show that there exists a constant $c_3 > 0$ such that one has, for t large enough:

$$p^t(x,x) \geqslant c_3 |P|^t, \tag{A.4}$$

the first inequality of (5.4) then will follow for t large enough, and the case when t is small then will be dealt with by finiteness, thanks to noticing that irreducibility ensures that $\mathbb{P}_x(\mathcal{R}_t) > 0$ for all $t \in \mathbb{N}$. To prove (A.4), let us consider the eigenvector $\boldsymbol{v} = (v_i)_{i \in \widetilde{V}}$ associated to the eigenvalue $|P|$. We shall keep in mind that, by Proposition A.1, $v_x > 0$. Denote $\overline{v} = \max_i v_i$; the relation $M^t v = |P|^t v$ then gives for all $t \geqslant 0$:

$$|P|^t v_x = \sum_{z \in \widetilde{V}} p^t(z,x) v_z, \tag{A.5}$$

hence

$$\max_{z \in \widetilde{V}} p^t(z,x) \geqslant \frac{v_x}{n\overline{v}} |P|^t. \tag{A.6}$$

Let now t_0 be such that one has $\forall z,v \in \widetilde{V}\ p^{t_0}(z,v) > 0$ (such a t_0 exists by aperiodicity, as we noticed it *supra*), and let us denote $\eta = \min_{z,v \in \widetilde{V}} p^{t_0}(z,v) > 0$. For $t \geqslant t_0$, by (A.6) we can fix z_1 such that $p^{t-t_0}(z_1,x) \geqslant v_x |P|^{t-t_0}/n\overline{v}$. It follows that

$$p^t(x,x) \geqslant \frac{\eta v_x}{n\overline{v}} |P|^{t-t_0}, \tag{A.7}$$

hence (A.4) with $c_3 = \eta v_x / n\overline{v}|P|^{t_0}$.

Now, let us look at the fine behaviour of the sequence $\mathbb{P}_x(\mathcal{R}_t)$ when $t \longrightarrow \infty$. By Markov's property,

$$\mathbb{P}_x(\mathcal{R}_t) = \sum_{z \in \widetilde{V}} p^t(x,z) \mathbb{P}_z(\mathcal{R}_0), \tag{A.8}$$

subsequently to prove property 2 in Lemma 5.4, we just have to show that we have $p^{t+1}(x,z)/p^t(x,z) \underset{t \longrightarrow \infty}{\longrightarrow} |P|$ for all $z \in \widetilde{V}$. More precisely, we will show that there exists a constant $c_4(z) > 0$ such that $p^t(x,z)/|P|^t \underset{t \longrightarrow \infty}{\longrightarrow} c_4(z)$.

$p^t(x,z)$ can be rewritten in matricial terms as ${}^T\boldsymbol{\delta}_z M^t \boldsymbol{\delta}_x$. Now, by A.1, if $\dot{M}$ stands for the matrix of the projection on $\mathbb{R}\boldsymbol{v}$ relatively to the sum of the characteristic spaces for the eigenvalues of M other than $|P|$:

$$\frac{1}{|P|^t} M^t \underset{t \longrightarrow \infty}{\longrightarrow} \dot{M}. \tag{A.9}$$

Subsequently, $p^t(x,z)/|P|^t$ tends to the value $c_4(z) = {}^T\boldsymbol{\delta}_y \dot{M} \boldsymbol{\delta}_x$ when t tends to infinity. The non-nullity of $c_4(z)$ then is a consequence of point 1 which we have proved a few lines above: indeed, taking again the notations t_0 and η used above, we have by Markov's property:

$$\forall t \geqslant t_0 \quad p^t(x,z) \geqslant p^{t-t_0}(x,x) p^{t_0}(x,z) \geqslant c_2 |P|^{t-t_0} \eta, \tag{A.10}$$

hence $c_4 \geqslant c_2 > 0$.

Point 3 may be the most subtle. In fact we will prove that for all z in $\widetilde{V}$, $\mathbb{P}_z(\tau_x = t)$ decreases exponentially *with an exponential decay factor strictly less than* $|P|$. More precisely, we will estimate the decay factor of $\mathbb{P}_z(\tau_x \geqslant t \text{ and } \tau_\partial \geqslant t) = \mathbb{P}_z\big((\forall u \in \{0,\dots,t-1\})(X_u \neq x, \partial)\big)$. In other words, we have to look at the decay speed of the sub-Markov chain associated to the transition kernel p, *but restricted to* $\widetilde{V} \smallsetminus \{x\}$. In matricial words, it is the spectral radius of the matrix M^*, which is the matrix M where the x-th line has been replaced by zeroes. Let us denote by $|P^*|$ its spectral radius. The weak form of Perron–Frobenius theorem (cf. [4]) claims that there exists a $|P^*|$-eigenvector $\boldsymbol{v}^*$ with positive or zero entries for M^*. Each entry of M^* is less than or equal to the corresponding entry of M, and moreover $M^* \neq M$; since M is the matrix of an irreducible aperiodic chain, it follows that, for t sufficiently large, each entry of $(M^*)^t$ is *strictly* less than the correponding entry of M^t. So, for t sufficiently large:

$$|P^*|^t \boldsymbol{v}^* = (M^*)^t\, \boldsymbol{v}^* < M^t \boldsymbol{v}^*, \tag{A.11}$$

which means that each entry of $|P^*|^t\boldsymbol{v}^*$ is strictly less than the corresponding entry of $M^t\boldsymbol{v}^*$. Now let us reason by contradiction by supposing that $|P^*| \geqslant |P|$, then (A.11) shows that we can find $t_1 > 0$ and $\rho_1 > |P|$ such that $M^{t_1}\boldsymbol{v}^* \geqslant \rho_1^{t_1}\boldsymbol{v}^*$, hence by iterating:

$$\limsup_{t \longrightarrow \infty} \left|M^t \boldsymbol{v}^*\right|^{1/t} \geqslant \rho_1 > |P|. \tag{A.12}$$

But that is absurd, since the spectral radius of M is actually $|P|$. This implies that $|P^*| < |P|$, as we wanted.

Further readings and acknowledgements

The use of the martingales introduced in § 3 can actually be seen as a discrete variant of the forward/backward martingale decomposition technique of Lyons and Weian [5], as was pointed out to me by Laurent Saloff-Coste. In fact, paper [6] gives a bound for continuous diffusions whose spirit is quite close to that of (3.1).

The present work was launched by informal discussions with my colleagues Yann Ollivier and Vincent Beffara, who usefully encouraged and helped me when necessary.

References

[1] N. Th. Varopoulos. Long range estimates for Markov chains. *Bul. Sci. Math. (2)*, 109:225–252, 1985.

[2] T. K. Carne. A transmutation formula for Markov chains. *Bul. Sci. Math. (2)*, 109:399–405, 1985.

[3] R. Lyons. Probabilities on trees and networks. `http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html`.

[4] D. Serre. *Matrices: Theory and Applications*. Springer, 2002.

[5] T. J. Lyons and Z. Weian. A crossing estimate for the canonical process on a Dirichlet space and a tightness result. In *Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987)*, pages 249–271, 1988. *Astérisque* # 157–158.

[6] J. Lunt, T. J. Lyons, and T. S. Zhang. Integrability of functionals of Dirichlet processes, probabilistic representations of semigroups, and estimates of heat kernels. *J. Funct. Anal.*, 153(2):320–342, 1998.