HAL
open science

A concentration theorem for the equilibrium measure of Markov chains with nonnegative coarse Ricci curvature

Laurent Veysseire

To cite this version:

Laurent Veysseire. A concentration theorem for the equilibrium measure of Markov chains with nonnegative coarse Ricci curvature. 2012. hal-00678728

HAL Id: hal-00678728

https://hal.science/hal-00678728

Preprint submitted on 13 Mar 2012

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

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

A concentration theorem for the equilibrium measure of Markov chains with nonnegative coarse Ricci curvature

Laurent Veysseire

Abstract

In this article, we prove a concentration inequality of the order of the exponential of a double integral of the coarse Ricci curvature for the equilibrium measure of a Markov chain, in the case when this curvature is nonnegative. This is, to the author's knowledge, the first concentration result in a discrete setting using a non-constant curvature instead of its infimum.

Introduction

For a Markov chain on a Polish space, a nonnegative coarse Ricci curvature means that the distributions after one step of the chain are closer (in the sense of the W_1 distance) than their starting points are [4]. Remind that the W_1 (Wasserstein) metric between two probability measures is the infimum over the set of couplings between this two probability measures of the expectation of the distance between the two points.

In the case when the space is ε-geodesic (see Definition 2), a nonnegative coarse Ricci curvature allows to extend the local attractiveness of a point x_0 to a global one (see [4] or Lemma 7). The attractiveness of a point implies exponential concentration of the equilibrium probability measure around this point, if the Markov chain does not spread out too quickly.

One of the simplest example is the random walk on $\mathbb{N}$ where we jump from n to $n+1$ with probability p and to $(n-1)_+$ with probability $1-p$. In this case, the coarse Ricci curvature is 0. If $p < \frac{1}{2}$, then 0 is attractive and we have exponential concentration. If $p \geq \frac{1}{2}$, then we don't have any attractive point, neither do we have any invariant probability measure.

Here we prove that the concentration of the equilibrium measure around an attractive point behaves at least like the exponential of a double integral of the coarse Ricci curvature.

We may remark that this is the right behaviour of the invariant distribution for diffusion processes on the real line, as we see in the example below.

Example 1 *Let us consider a diffusion process on the real line whose generator takes the form:*

$$Lf = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2} - \frac{\mathrm{d}V}{\mathrm{d}x}\frac{\mathrm{d}f}{\mathrm{d}x}$$

where the energy $V(x)$ is smooth. Then the coarse Ricci curvature is $\frac{\mathrm{d}^2 V}{\mathrm{d}x^2}$, and the measure $\mathrm{e}^{-V(x)}\mathrm{d}x$ is reversible. We see that the density of the invariant measure is exacly a double integral of the coarse Ricci curvature.

1 The concentration Theorems

We define ε-geodesic spaces as in [4].

Definition 2 *Let $\varepsilon > 0$. A metric space (X, d) is said to be ε-geodesic if for each $(x, y) \in X^2$, there exists $n \in \mathbb{N}$ and a sequence $x = x_0, x_1, \ldots, x_n = y \in X$ such that $d(x_i, x_{i+1}) \leq \varepsilon$ for each $0 \leq i \leq n-1$ and $d(x, y) = \sum_{i=0}^{n} d(x_i, x_{i+1})$.*

For a Markov chain with transition kernel P on a ε-geodesic space, we will denote by $K_\varepsilon(x)$ the local coarse Ricci curvature at x:

$$K_\varepsilon(x) := \inf_{y \in X | 0 < d(x,y) \leq \varepsilon} \kappa(x, y)$$

Where $\kappa(x, y) := 1 - \frac{W_1(P_x, P_y)}{d(x,y)}$ is the coarse Ricci curvature between x and y as defined in [4].

Here we will prove the following concentration result for the equilibrium measure of Markov Chains:

Theorem 3 *Let X be an ε-geodesic metric space and P be the transition kernel of a Markov chain on X. Assume that:*

- *there exists $\rho > 0$ and a point x_0 such that x_0 is ρ-attractive for the Markov chain in the sense that*

$$\forall x | \varepsilon < d(x, x_0) \leq 2\varepsilon, W_1(\delta_{x_0}, P_x) \leq d(x_0, x) - \rho,$$

- *there exists a non-increasing function* $K : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *satisfying:*

$$K_\varepsilon(x) \geq K(d(x, x_0)),$$

- *there exists* $s > 0$ *such that for any* $x \in X$*, any* 1*-lipschitz function* $f : X \mapsto \mathbb{R}$ *and any* $\lambda \in \mathbb{R}$*, we have:*

$$\mathbb{E}_{P_x}\left[\mathrm{e}^{\lambda f}\right] \leq \mathrm{e}^{\lambda \mathbb{E}_{P_x}[f] + \frac{\lambda^2 s^2}{2}}.$$

Then we have, for every $l > 2\varepsilon + \frac{\ln(2)s^2}{\rho}$ *and any equilibrium measure* π*:*

$$\mathbb{P}_{x\sim\pi}(d(x, x_0) \geq l) \leq C_0 \mathrm{e}^{-\frac{1}{2s^2}\Phi(l)}$$

with

$$\Phi(l) := \rho l + \int_{2\varepsilon}^{l} \left(\int_{2\varepsilon}^{u} K(v)\mathrm{d}v \right) \mathrm{d}u$$

and

$$C_0 = \frac{\mathrm{e}^{\frac{3\varepsilon}{2s^2}\max(3\varepsilon, \rho + \frac{\ln(2)s^2}{\rho}) - \frac{\rho^2}{4s^2} + \frac{1}{2s^2}\left(\rho(2\varepsilon + \frac{\ln(2)s^2}{\rho}) + \int_{2\varepsilon}^{2\varepsilon + \frac{\ln(2)s^2}{\rho}} \int_{2\varepsilon}^{u} K(v)\mathrm{d}v\mathrm{d}u\right)}}{1 - \mathrm{e}^{-\frac{\rho^2}{4s^2}}}$$

Remark 4 *If* $K = 0$*, we obtain exponential concentration, as proved in [4].*

Proposition 5 *If closed balls are compact, then under the hypotheses of Theorem 3, there exists an equilibrium measure.*

Remark 6 *In the case when* $\int_0^\infty K(r)\mathrm{d}r = \infty$*, and for some (hence any)* $x_0 \in X$*,* $W_1(\delta_{x_0}, P_{x_0}) < \infty$*, then for any* $\rho > 0$*, there exists a* $\varepsilon > 0$ *large enough such that* x_0 *is* ρ*-attractive. This is a trivial consequence of the Lemma below.*

Lemma 7 *Let* X *be an* ε*-geodesic metric space and* P *be the transition kernel of a Markov chain such that there exists a non-increasing function* $K : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *and a point* $x_0 \in X$ *satisfying:*

$$K_\varepsilon(x) \geq K(d(x, x_0)).$$

Then we have

$$\mathbb{E}_{y\sim P(x)}[d(x_0, y)] \leq d(x_0, x) - F(d(x_0, x)),$$

where

$$F(l) := \begin{cases} \rho + \int_{2\varepsilon}^{l} K(u)\mathrm{d}u & \text{if } 2\varepsilon \leq l \\ \rho & \text{if } \varepsilon < l \leq 2\varepsilon \\ -J(0) & \text{if } l \leq \varepsilon \end{cases}$$

with $\rho := \inf_{x|\varepsilon < d(x,x_0) \leq 2\varepsilon} d(x, x_0) - W_1(P(x), \delta_{x_0})$ *and* $J(x_0) = W_1(P(x_0), \delta_{x_0})$*.*

Proof :

If $\varepsilon < d(x,x_0) \leq 2\varepsilon$, this is just the definition of ρ. If $d(x,x_0) \leq \varepsilon$, we have $K_\varepsilon(x_0) \geq 0$, so $W_1(P(x),\delta_{x_0}) \leq W_1(P(x),P(x_0)) + W_1(P(x_0),\delta_{x_0}) \leq d(x,x_0) + J(x_0)$. If $d(x,x_0) \geq 2\varepsilon$, there exists $x_1,\ldots,x_n = x$ such that $d(x_i,x_{i+1}) \leq \varepsilon$, $\varepsilon < d(x_1,x_0) \leq 2\varepsilon$ and $d(x,x_0) = d(x_1,x_0) + \sum_{i=1}^{n-1} d(x_i,x_{i+1})$. We have then

$$\begin{aligned}
W_1(P(x),\delta_{x_0}) &\leq W_1(P(x_1),\delta_{x_0}) + \sum_{i=1}^{n-1} W_1(P(x_i),P(x_{i+1})) \\
&\leq d(x_1,x_0) - \rho + \sum_{i=1}^{n-1}(1 - K(d(x_i,x_0)))d(x_i,x_{i+1}) \\
&\leq d(x,x_0) - \rho - \sum_{i=1}^{n-1} \int_{d(x_i,x_0)}^{d(x_{i+1},x_0)} K(l)\mathrm{d}l \\
&\leq d(x,x_0) - \rho - \int_{d(x_1,x_0)}^{d(x,x_0)} K(l)\mathrm{d}l \\
&\leq d(x,x_0) - F(d(x,x_0)).\square
\end{aligned}$$

Lemma 8 *Let μ be a probability measure on X and $s > 0$ be such that for any 1-lipschitz function f, we have the following inequality:*

$$\mathbb{E}_\mu\left[\mathrm{e}^{\lambda f}\right] \leq \mathrm{e}^{\lambda \mathbb{E}_\mu[f] + \frac{\lambda^2 s^2}{2}}.$$

Then, for each $\mathcal{C}^1$ function $g : \mathbb{R} \mapsto \mathbb{R}$ such that g' is Lipschitz and $\|g'\|_{lip} < \frac{1}{s^2}$ and for each 1-lipschitz function f, we have:

$$\mathbb{E}_\mu\left[\mathrm{e}^{g\circ f}\right] \leq \frac{\mathrm{e}^{g(\mathbb{E}_\mu[f]) + \frac{s^2 g'^2(\mathbb{E}_\mu[f])}{2\left(1 - s^2\|g'\|_{lip}\right)}}}{\sqrt{1 - s^2\|g'\|_{lip}}}.$$

Proof :

For each $x \in X$, we have

$$\mathrm{e}^{g\circ f(x)} \leq \mathrm{e}^{g(\mathbb{E}_\mu[f]) + (f(x) - \mathbb{E}_\mu[f])g'(\mathbb{E}_\mu[f]) + \frac{(f(x)-\mathbb{E}_\mu[f])^2}{2}\|g'\|_{lip}}.$$

Now we use the fact that the Laplace transform of a Gaussian measure $\mathcal{N}(M,\sigma^2)$ is:

$$\int_{-\infty}^{\infty} \mathrm{e}^{\lambda u}\mathrm{e}^{-\frac{(u-M)^2}{2\sigma^2}} \frac{\mathrm{d}u}{\sqrt{2\pi\sigma^2}} = \mathrm{e}^{\lambda M + \frac{\lambda^2\sigma^2}{2}}.$$

So, taking $\lambda = f(x) - \mathbb{E}_\mu[f]$, $M = g'(\mathbb{E}_\mu[f])$ and $\sigma^2 = \|g'\|_{lip}$, we get:

$$\mathrm{e}^{g(f(x))} \leq \mathrm{e}^{g(\mathbb{E}_\mu[f])} \int_{-\infty}^{\infty} \mathrm{e}^{u(f(x)-\mathbb{E}_\mu[f])}\mathrm{e}^{-\frac{(u-g'(\mathbb{E}_\mu[f]))^2}{2\|g'\|_{lip}}} \frac{\mathrm{d}u}{\sqrt{2\pi\|g'\|_{lip}}}.$$

Integrating this inequality with respect to μ and using our assumption yields:

$$\mathbb{E}_\mu\left[\mathrm{e}^{g\circ f}\right] \leq \mathrm{e}^{g(\mathbb{E}_\mu[f])}\int_{-\infty}^{\infty}\mathrm{e}^{\frac{u^2s^2}{2}}\mathrm{e}^{-\frac{(u-g'(\mathbb{E}_\mu[f]))^2}{2\|g'\|_{lip}}}\frac{\mathrm{d}u}{\sqrt{2\pi\|g'\|_{lip}}} = \mathrm{e}^{g(\mathbb{E}_\mu[f])}\frac{\mathrm{e}^{\frac{s^2g'^2(\mathbb{E}_\mu[f])}{2(1-s^2\|g'\|_{lip}}}}{\sqrt{1-s^2\|g'\|_{lip}}}$$

as needed.□

Theorem 9 *Let X be a ε-geodesic metric space and P be the transition kernel of a Markov chain. Assume that there exists a non-increasing function $K:\mathbb{R}_+\mapsto\mathbb{R}_+$ and a point $x_0\in X$ satisfying:*

$$K_\varepsilon(x)\geq K(d(x,x_0))$$

and that there exists $s>0$ such that for any $x\in X$, any 1-lipschitz function $f:X\mapsto\mathbb{R}$ and any $\lambda\in\mathbb{R}$, we have:

$$\mathbb{E}_{P_x}\left[\mathrm{e}^{\lambda f}\right]\leq \mathrm{e}^{\lambda\mathbb{E}_{P_x}[f]+\frac{\lambda^2s^2}{2}}.$$

Let F be defined as in Lemma 7. Then, for every pair $(\alpha,d_0)\in\mathbb{R}_+^2$ satisfying:

- $d_0\geq 2\varepsilon$
- $F(d_0)>\frac{s^2K(d_0)}{2}$
- $\alpha<\frac{1}{s^2K(d_0)}$
- $C_{\alpha,d_0}:=\dfrac{\mathrm{e}^{-\alpha F(d_0)^2\left(1-\frac{\alpha s^2}{2(1-\alpha s^2K(d_0))}\right)}}{\sqrt{1-\alpha s^2K(d_0)}}<1$

we have the following concentration inequality for any equilibrium measure π of the Markov chain and any $l\geq d_0$:

$$\mathbb{P}_{x\sim\pi}(d(x,x_0)\geq l)\leq C'_{\alpha,d_0}\frac{C_{\alpha,d_0}}{1-C_{\alpha,d_0}}\mathrm{e}^{-\alpha(\varphi(l)-\varphi(d_0))}$$

where $\varphi(l)=\int_0^l F(u)\mathrm{d}u$, and $C'_{\alpha,d_0}:=\mathrm{e}^{\alpha\mathbf{1}_{J(x_0)+\varepsilon>d_0-F(d_0)}\int_{d_0-F(d_0)}^{J(x_0)+\varepsilon}\sup(F(d_0),F(u))\mathrm{d}u}$.

Proof : we set $\psi(x)=\alpha\varphi(x)$ if $x\geq d_0$ and $\psi(x)=\alpha(\psi(d_0)-(d_0-x)F(d_0))$ if $x<d_0$. Under our assumptions, ψ is convex and increasing, and we have $\|\psi'\|_{lip}=\alpha K(d_0)<\frac{1}{s^2}$. Our goal is to bound the quantity $\mathbb{E}_{x\sim\pi}\left[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,x_0)\geq d_0}\right]$. We have:

$$\begin{aligned}\mathbb{E}_{x\sim\pi}\left[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,x_0)\geq d_0}\right] &= \mathbb{E}_{x\sim\pi}\left[\mathbb{E}_{y\sim P_x}\left[\mathrm{e}^{\psi(d(y,x_0))}\mathbf{1}_{d(y,x_0)\geq d_0}\right]\right]\\ &\leq \mathbb{E}_{x\sim\pi}\left[\mathbb{E}_{y\sim P_x}\left[\mathrm{e}^{\psi(d(y,x_0))}\right]\right].\end{aligned}$$

Using Lemma 8 with $\mu = P_x$ and $g = \psi$, and Lemma 7, we get:

$$\mathbb{E}_{x\sim\pi}\left[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,x_0)\geq d_0}\right] \leq$$
$$\mathbb{E}_{x\sim\pi}\left[\frac{\mathrm{e}^{\psi(d(x,x_0)-F(d(x,x_0)))+\frac{\alpha^2 s^2 F(d_0)^2}{2(1-\alpha s^2 K(d_0))}}}{\sqrt{1-\alpha s^2 K(d_0)}}\mathbf{1}_{d(x,x_0)<d_0}\right.$$
$$\left.+\frac{\mathrm{e}^{\psi(d(x,x_0)-F(d(x,x_0)))+\frac{\alpha^2 s^2 F(d(x,x_0))^2}{2(1-\alpha s^2 K(d_0))}}}{\sqrt{1-\alpha s^2 K(d_0)}}\mathbf{1}_{d(x,x_0)\geq d_0}\right].$$

The function $l \mapsto l - F(l)$ is nondecreasing on $[0,\varepsilon]$ and on (ε, d_0), and ψ is an increasing function. Then, for $d(x,x_0) < d_0$, we have $\psi(d(x,x_0) - F(d_0)) \leq \psi(\max(J(x_0)+\varepsilon, d_0 - F(d_0)) = \ln(C'_{\alpha,d_0}) + \alpha(\varphi(d_0) - F^2(d_0))$.

For $d(x,x_0) \geq d_0$, we have $\psi(d(x,x_0) - F(d(x,x_0))) \leq \psi(d(x,x_0)) - \alpha F^2(d_0)$. So we get:

$$\mathbb{E}_{x\sim\pi}\left[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,x_0)\geq d_0}\right] \leq C'_{\alpha,d_0}C_{\alpha,d_0}\mathrm{e}^{\alpha\varphi(d_0)} + C_{\alpha,d_0}\mathbb{E}_{x\sim\pi}\left[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,x_0)\geq d_0}\right].$$

And then, since $C_{\alpha,d_0} < 1$, we finally obtain:

$$\mathbb{E}_{x\sim\pi}\left[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,x_0)\geq d_0}\right] \leq \frac{C'_{\alpha,d_0}C_{\alpha,d_0}}{1-C_{\alpha,d_0}}\mathrm{e}^{\alpha\varphi(d_0)}.$$

Now we just have to use the Markov inequality to derive the desired inequality.□

Remark 10 *In the previous proof, we didn't fully use the hypothesis $F(d_0) > \frac{s^2K(d_0)}{2}$. In fact, for a fixed d_0, $\ln(C_{\alpha,d_0})$ is a convex function of α on the interval $[0, \frac{1}{s^2K(d_0)})$. We have $C_{0,d_0} = 1$ and $\frac{\partial}{\partial\alpha}\ln(C_{\alpha,d_0})|_{\alpha=0} < 0$ if and only if $F(d_0) > \frac{s^2K(d_0)}{2}$. So if $0 < F(d_0) \leq \frac{s^2K(d_0)}{2}$, there doesn't exist any α such that $C_{\alpha,d_0} < 1$ and so the theorem wouldn't tell us anything at all.*

Remark 11 *If $K(d_0) \leq \frac{1}{2}$ have $C_{\frac{2}{s^2},d_0} \geq 1$, so we must have $\alpha < \frac{2}{s^2}$. Under the hypothesis that $\kappa(x) \xrightarrow[x\to\infty]{} 0$ and $F(x) \xrightarrow[x\to\infty]{} +\infty$, we can find for any $0 < \alpha < \frac{2}{s^2}$ a d_0 such that $C_{\alpha,d_0} < 1$. Of course we need a greater d_0 when α gets closer to $\frac{2}{s^2}$.*

One way to choose α and d_0 is given by the following proof of Theorem 3:

Proof of Theorem 3: we use Theorem 9 with $\alpha = \frac{1}{2s^2}$ and $d_0 = 2\varepsilon + \frac{\ln(2)s^2}{\rho}$. We only have to check that in this case, $C_{\alpha,d_0} \leq \mathrm{e}^{-\frac{\rho^2}{4s^2}}$ and $C'_{\alpha,d_0} \leq \mathrm{e}^{\frac{3\varepsilon}{2s^2}\max(3\varepsilon,\rho+\frac{\ln(2)s^2}{\rho})}$.

We have

$$\ln(C_{\alpha,d_0}) = \left(-\alpha + \frac{\alpha^2 s^2}{2(1-\alpha s^2 K(d_0))}\right) F^2(d_0) - \frac{1}{2}\ln(1-\alpha s^2 K(d_0)).$$

Since $K(d_0) \leq 1$, we have $-\alpha + \frac{\alpha^2 s^2}{2(1-\alpha s^2 K(d_0))} \leq -\frac{1}{4s^2}$. We have $F(d_0) \geq \rho + \frac{\ln(2)s^2}{\rho}K(d_0)$, and then $F(d_0)^2 \geq \rho^2 + 2\ln(2)s^2 K(d_0)$. Using the concavity of ln on $[\frac{1}{2}, 1]$, we get $\ln(1-\alpha s^2 K(d_0)) \geq -\ln(2)K(d_0)$. Thus we get:

$$\ln(C_{\alpha,d_0}) \leq -\frac{1}{4s^2}(\rho^2 + 2\ln(2)s^2 K(d_0)) + \frac{\ln(2)}{2}K(d_0) = -\frac{\rho^2}{4s^2}.$$

For C'_{α,d_0}, we have

$$\begin{aligned}\ln(C'_{\alpha,d_0}) &= \frac{1}{2s^2}\mathbf{1}_{J(x_0)+\varepsilon > d_0 - F(d_0)} \int_{d_0-F(d_0)}^{J(x_0)+\varepsilon} \max(F(d_0), F(u))\mathrm{d}u \\ &\leq \frac{1}{2s^2}((J(x_0)+\varepsilon) - (d_0 - F(d_0)))_+ \max(F(d_0), F(J(x_0)+\varepsilon)).\end{aligned}$$

By the triangular inequality for W_1, we have $J(x_0) \leq W_1(\delta_{x_0}, P(x)) + W_1(P(x), P(x_0)) \leq W_1(\delta_{x_0}, P(x)) + d(x_0, x)$, for any x because the coarse Ricci curvature is nonnegative. If we take x such that $\varepsilon < d(x_0, x) \leq 2\varepsilon$, we have $J(x_0) \leq 2d(x_0, x) - \rho \leq 4\varepsilon - \rho$. We have $F(d_0) \leq \rho + \frac{\ln(2)s^2}{\rho}$, so $d_0 - F(d_0) \geq 2\varepsilon - \rho$ and then $((J(x_0)+\varepsilon) - (d_0 - F(d_0)))_+ \leq 3\varepsilon$. And finally, $F(J(x_0)+\varepsilon) \leq F(5\varepsilon - \rho) \leq 3\varepsilon$. Putting that together give us the desired bound for C'_{α,d_0}.□

Proof of Proposition 5: We take α and d_0 as in the proof of Theorem 3. We consider the sequence of probability measures $P^n_{x_0}$. Then, doing as in the proof of Theorem 9, we have :

$$\mathbb{E}_{x\sim P^{n+1}_{x_0}}[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,0)\geq d_0}] \leq C'_{\alpha,d_0}C_{\alpha,d_0} + C_{\alpha,d_0}\mathbb{E}_{x\sim P^n_{x_0}}[\mathrm{e}^{\psi(d(x,x_0))}\mathbf{1}_{d(x,0)\geq d_0}].$$

From that, we can conclude that there exists $C < +\infty$ such that for all n, we have $\mathbb{E}_{P^n_{x_0}}[\mathrm{e}^{\psi(d(x,x_0))}] < C$. So the sequence $P^n_{x_0}$ is tight, and then, so is the sequence $\pi_n = \frac{1}{n+1}\sum_{i=0}^n P^i_{x_0}$. Because closed balls are compact, we can extract a weakly convergent subsequence $\pi_{\theta(n)}$, and we denote by π its limit. The W_1 distance metrizes the weak convergence on the set of probability measures on X satisfying $\mathbb{E}[\mathrm{e}^{\psi(d(x,x_0))}] < C$ (see [6]). Thus the subsequence $\pi_{\theta(n)}$ converges to π

for the W_1 distance. Furthermore, we have $W_1(\pi_n, P\pi_n) \leq \frac{C'}{n+1}$ with $C' < \infty$ a constant. We have then

$$W_1(\pi, P_\pi) \leq W_1(\pi, \pi_{\theta(n)}) + W_1(\pi_{\theta(n)}, P\pi_{\theta(n)}) + W_1(P\pi_{\theta(n)}, P\pi).$$

The nonnegative coarse Ricci curvature implies that P contracts the W_1 distance ([4]), so the third term of the right hand side is at most the first one. We have already seen that the first two terms tend to 0 when n tends to $+\infty$. So the right hand side tends to 0 when n tends to $+\infty$. Thus $W_1(\pi, P\pi) = 0$, and then π is an invariant measure.□

2 Some examples

Let us see which concentration we can get with Theorem 3 and Theorem 9 in some examples below.

Example 12 (Discrete time M/M/k queue (see, for example [2])) *Let $0 < n_0 < k$ be two integers. We consider here the Markov chain on integers with transition kernel:*

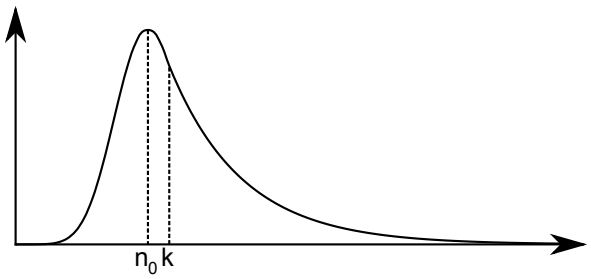
$$\begin{aligned} p(n, n+1) &= \frac{n_0}{n_0 + k} \\ p(n, n) &= \frac{(k-n)_+}{n_0 + k} \\ p(n, n-1) &= \frac{\min(n, k)}{n_0 + k} \\ p(n, m) &= 0 \quad \textit{if } |n - m| > 1. \end{aligned}$$

The origin x_0 we will consider to apply Theorem 3 is n_0, the only point at which the probability to jump at left equals the probability to jump at right (that is why we chose n_0 integer). Hoeffding's Lemma (see [3]) states that for a random variable X such that $a \leq X \leq b$ almost surely, we have $\mathbb{E}\left[e^{\lambda}(X - \mathbb{E}[X])\right] \leq e^{\frac{\lambda^2(b-a)^2}{8}}$. So we can take $s = 1$ in theorem 3. To compute the coarse Ricci curvature, we remark that if $x < y$, the measure P_y dominates stochastically the measure P_x, and thus the W_1 distance between them is the difference of their expectations. For $x < y$, the coarse Ricci curvature $K(x, y)$ is then $\frac{1}{n_0+k}$ if $y \leq k$, $\frac{k-x}{y-x}\frac{1}{n_0+k}$ if $x < k < y$ and 0 if $x \geq k$. If we take $\varepsilon = 1$, we have $\rho = \frac{1}{n_0+k}$, and $K(r) = \frac{\mathbf{1}_{r<k-n_0}}{n_0+k}$.

Applying Theorem 3 should give a Gaussian then exponential concentration, but, as ρ is very small, d_0 is large $(2 + (n_0 + k)\ln(2))$. If $k - n_0 \leq \frac{2\ln(2)n_0+2}{1-\ln(2)}$, we get only the exponential part. If k is too large,

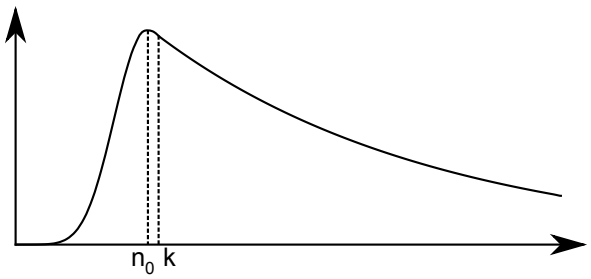
d_0 is large too, and the gaussian-then-exponential bounds starts far away from n_0. We can try to take a larger ε to get a better ρ. Indeed, we get $\rho = \frac{\min(\varepsilon, k-n_0)}{n_0+k}$, but we pay that by a worse curvature $K(r) = \frac{1}{n_0+k}\min(1, \max(0, \frac{k-n_0-r}{\varepsilon}))$. We distinguish 3 cases depending on how $k - n_0$ is tall with respect to n_0.

When $k - n_0$ is between $\sqrt{n_0}$ and n_0, the equilibrium measure is well approximated by a Gaussian between 0 and k.



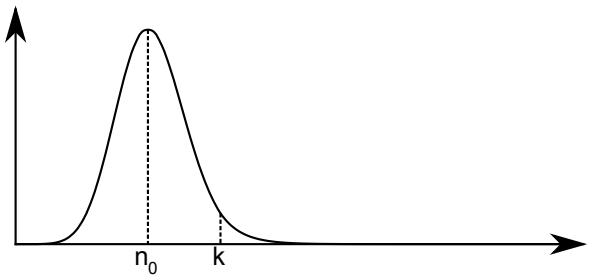
The optimal ε is $O(\sqrt{n_0})$, the coefficient of the Gaussian part of the concentration inequality is $O(\frac{1}{n_0})$, which is good, and the coefficient of the exponential part is $O(\frac{k-n_0}{n_0})$, like the right one.

When $k - n_0$ is $o(\sqrt{n_0})$, the mass of $[0, k]$ under the equilibrium measure is negligible with respect to the one of $[k, \infty)$.



The optimal ε and d_0 are $O(k - n_0)$, this time, we have no Gaussian part because d_0 is too large (and indeed, there is no Gaussian part in the equilibrium measure), and the coefficient of the exponential part is about one half of the right one.

When $k - n_0$ is greater than n_0, the equilibrium measure is almost the Poissonian one with parameter n_0, the density of the equilibrium measure is illustrated below:



The optimal ε and d_0 are $O(\sqrt{k})$, the coefficient appearing in the Gaussian part is $O(\frac{1}{k})$, instead of an expected $\frac{1}{n_0}$, and the coefficient of the exponential part is $O(1)$, which is clearly not optimal, so Theorem 3 gives a rather bad concentration inequality.

Example 13 (Discrete time Ornstein Uhlenbeck) *Let $0 < \alpha \leq 1$ be a real parameter. Here we consider the Markov Chain on $\mathbb{R}$ given by the transition kernel:*

$$P_x = \mathcal{N}((1-\alpha)x, 1).$$

It is shown in [1] that in the Gaussian case, we can take the variance of the distribution for s^2. So we take $s^2 = 1$, and for every $\varepsilon > 0$, the curvature is constant $K = \alpha$. We have $\rho = \varepsilon - \sqrt{\frac{2}{\pi}}\left(\mathrm{e}^{-\frac{(1-\alpha)^2\varepsilon^2}{2}} + \int_0^{(1-\alpha)\varepsilon} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\right) \geq -\sqrt{2}\pi + \alpha\varepsilon$. Theorem 3 applied with $\varepsilon = \frac{\sqrt{2\ln(2)\alpha}+\sqrt{\frac{8}{\pi}}}{\alpha}$ gives us Gaussian concentration with coefficient $\frac{\alpha}{4}$ instead of $\frac{\alpha(2-\alpha)}{2}$ (so we have a loss of a factor between 2 and 4), and $d_0 = O(\sqrt{\frac{1}{\alpha}})$

The bad behaviour of s^2 prevents to easily generalize Theorem 3 or Theorem 9 to continuous time. The following example of a continuous time processes, whose generator merges a diffusive part and a jump part, shows that a generalization of those theorems does not hold, even if the jump rate is uniformly bounded.

Example 14 *Consider a continuous time process on $\mathbb{R}_+$ with a linear drift towards 0 and a random jump to the right of size 1 and rate 1. The generator of this process is given by $Lf(x) = -\alpha x f'(x) + f(x+1) - f(x)$, with $\alpha > 0$ a constant which quantifies the drift.*

In this example, the coarse Ricci curvature is α. Indeed, using the coupling of the processes X_t and Y_t starting at x and y such that X_t and Y_t jump at the same times shows that the law of Y_t is the translation of the law of X_t by $(y-x)\mathrm{e}^{-\alpha t}$. If something like Theorem 3 or Theorem 9 did hold, we would have Gaussian concentration. But

actually there is only Poissonian concentration. Let us prove there is Poissonian concentration and no better. We denote by X_t the value of the process at the time T. Let $T_1, T_2, \dots$ be the successive times of the jumps. For all $T > 0$, let $N(T)$ be the number of jumps between 0 and T. We have

$$X_T = \mathrm{e}^{-\alpha T} X_0 + \sum_{i=1}^{N(T)} \mathrm{e}^{-\alpha(T-T_i)}.$$

If we take $X_0 = 0$ then $\mathbb{E}[X_T] \leq T$, and since the coarse Ricci curvature is greater than $\alpha > 0$, there exists an unique invariant probability measure π (see [5]).

Now take X_0 with the law π. Then X_1 has the law π, and is greater than $\mathrm{e}^{-\alpha} N(1)$, which has a Poissonian concentration since $N(1)$ follows precisely a Poisson law of parameter 1. So we cannot have a better concentration than a Poissonian one.

It remains to prove that π has Poissonian concentration. We take $X_0 = 0$. Let us consider the Laplace/Fourier transform of X_T, that is $G_T(\lambda) := \mathbb{E}[\mathrm{e}^{\lambda X_T}]$ for $\lambda \in \mathbb{C}$. $N(T)$ has the law $\mathcal{P}(T)$, and the repartition of the T_i's knowing $N(T)$ is the one of $N(T)$ independent random variables uniformly distributed in $[0, T]$. So we have:

$$\begin{aligned}
G_T(\lambda) &= \sum_{k=0}^{\infty} \frac{T^k \mathrm{e}^{-T}}{k!} \left(\frac{1}{T} \int_0^T \mathrm{e}^{\lambda \mathrm{e}^{-\alpha t}} \mathrm{d}t \right)^k \\
&= \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-T}}{k!} \left(\int_0^T \left(\sum_{n=0}^{\infty} \frac{\lambda^n \mathrm{e}^{-n\alpha t}}{n!} \right) \mathrm{d}t \right)^k \\
&= \mathrm{e}^{\sum_{n=1}^{\infty} \left[\frac{-\lambda^n \mathrm{e}^{-n\alpha t}}{n\alpha n!} \right]_{t=0}^{T}} \\
&= \mathrm{e}^{\frac{I(\lambda) - I(\lambda \mathrm{e}^{-\alpha T})}{\alpha}}
\end{aligned}$$

with $I(\lambda) = \sum_{n=1}^{\infty} \frac{\lambda^n}{nn!} = \int_0^{\lambda} \frac{(\mathrm{e}^z - 1)\mathrm{d}z}{z}$. We see that $G_T(\lambda)$ tends to a limit $G(\lambda)$, which is the Laplace/Fourier transform of π, when T tends to $+\infty$.

We have $G(\lambda) = \mathrm{e}^{\frac{I(\lambda)}{\alpha}}$. An integration by parts gives us $I(\lambda) = \frac{\mathrm{e}^{\lambda} - \lambda - 1}{\lambda} + \int_0^{\lambda} \frac{\mathrm{e}^z - z - 1}{z^2} \mathrm{d}z$, so $I(\lambda) \sim \frac{\mathrm{e}^{\lambda}}{\lambda}$. For $l > 1$, we use the Markov inequality on $\mathrm{e}^{\ln(l) X}$ and get:

$$\mathbb{P}_{\pi}[X \geq l] \leq \mathrm{e}^{\frac{I(\ln(l))}{\alpha} - l \ln(l)},$$

and we have $\frac{I(\ln(l))}{\alpha} \sim \frac{l}{\alpha \ln(l)} = o(l \ln(l))$. So we have Poissonian concentration.

References

[1] S. G. Bobkov and F. Götze. Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities. *Journal of Functional Analysis*, 163(1):1–28, 1999.

[2] G. Grimmett and D. Stirzaker. *Probability and random processes.* Oxford University Press, third edition, 2003.

[3] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[4] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256:810–864, February 2009.

[5] L. Veysseire. Coarse Ricci curvature for continuous-time Markov processes. Available on ArXiv, 2012.

[6] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften.* Springer, 2009.