HAL open science

Coarse Ricci curvature for continuous-time Markov processes

Laurent Veysseire

▶ To cite this version:

Laurent Veysseire. Coarse Ricci curvature for continuous-time Markov processes. 2012. hal-00665610

HAL Id: hal-00665610
https://hal.science/hal-00665610v1

Preprint submitted on 2 Feb 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.

Coarse Ricci curvature for continuous-time Markov processes

Laurent Veysseire

Abstract

In this paper, we generalize Ollivier's notion of coarse Ricci curvature for Markov chains to continuous time Markov processes. We prove Wasserstein contraction and a Lichnerowicz-like spectral gap bound for reversible Markov processes with positive coarse Ricci curvature.

Introduction

In [7], Ollivier defines the coarse Ricci curvature for Markov chains on metric spaces, in a discrete time framework. Here we extend this notion to continuous time Markov processes. We define the curvatures κ and $\bar{\kappa}$ (see Definition 6), and prove (Theorem 9) that a control of $\bar{\kappa}$ (or κ) implies that the Markov process contracts the W_1 Wasserstein distance between measures exponentially fast (or that the W_1 distance des not explode faster than exponentially, if we have negative curvature). Note that the definition of the coarse Ricci curvature is local. It is natural to think that positive curvature gives global contraction, but it is not a trivial consequence.

We also show that the coarse Ricci curvature allows to generalize the Lichnerowicz Theorem (see [6]) that we recall below.

Theorem 1 (Lichnerowicz) *Let $(\mathcal{M}, g)$ be a n-dimensional Riemannian manifold. If there exists $K > 0$ such that for each $x \in \mathcal{M}$, for each $u \in \mathrm{T}_x\mathcal{M}$, we have $\mathrm{Ric}_x(u,u) \geq K g_x(u,u)$, then the spectral gap λ_1 of the Laplace operator Δ acting on L^2 satisfies*

$$\lambda_1 \geq \frac{n}{n-1} K.$$

Here we denote by Ric *the Ricci curvature tensor of $\mathcal{M}$.*

Using the contraction of the Wasserstein distance given by Theorem 9, we can prove we have a spectral gap. What we can get using coarse Ricci curvature is the following:

Theorem 2 *Let (P^t) be the semi-group of a reversible and ergodic Markov process on a Polish space (E,d), admitting a left continuous modification. Assume that for all (x,y) in E^2 with $x \neq y$, the coarse Ricci curvature $\kappa(x,y)$ between x and y (see Definition 6) is bounded below by a constant $K > 0$. Also assume that for some (then any) x_0, $\int d^2(x,x_0)\mathrm{d}\pi(x) < +\infty$, where π is the reversible measure.*

Then the operator norm of P^t acting on the space $L^2_0(E,\pi)$ of 0-mean $L^2(\pi)$ functions is at most e^{-Kt}. In the case when the Markov process admits a generator L, this means L has a spectral gap $\lambda_1(L) \geq K$.

This Theorem looks like Theorem 1.9 of [3]. The difference is that the contraction hypothesis was global, and some assumption about the first eigenfunction was required. Here in Theorem 9, the coarse Ricci curvature is local.

In the special case of diffusion processes on Riemannian manifolds, we can get better lower bounds for the spectral gap, depending on the harmonic mean of the Ricci curvature instead of its infimum, as shown in [9] and [10].

1 Coarse Ricci curvature: definition and examples

The coarse Ricci curvature of a Markov process on a Polish (metric, complete, separable) space (E,d) is defined thanks to the Wasserstein metric, which is based on optimal coupling:

Definition 3 *The Wasserstein distance between two probability measures is the (possibly infinite) quantity defined by:*

$$W_1(\mu,\nu) = \inf_{\xi \in \Phi(\mu,\nu)} \int d(x,y)\mathrm{d}\xi(x,y).$$

Here $\Phi(\mu,\nu)$ is the set of all couplings between μ and ν, that is, the set of probability measures on E^2 whose marginal laws are μ and ν.

The duality theorem of Kantorovitch (see [11]) gives another interpretation of this distance and allows to extend it to finite measures provided they have the same total mass, and makes the triangular inequality for W_1 easier to check.

Theorem 4 (Kantorovitch–Rubinstein) *We have the following equality:*

$$W_1(\mu,\nu) = \sup_{f \text{ bounded}, 1-\text{Lipschitz}} \int f \mathrm{d}(\mu - \nu).$$

In [7], Ollivier defines the coarse Ricci curvature between two different points for discrete time Markov chains in the following way:

Definition 5 (Ollivier) *If P is the transition kernel of a Markov chain on a metric space (E, d), the coarse Ricci curvature between x and y is defined by*

$$\kappa(x, y) = 1 - \frac{W_1(\delta_x.P, \delta_y.P)}{d(x, y)}.$$

A natural generalization of this quantity for continuous-time Markov processes is the following:

Definition 6 *The coarse Ricci curvature between x and y is defined by:*

$$\kappa(x, y) = \varliminf_{t \to 0} \frac{1}{t} \left(1 - \frac{W_1(P_x^t, P_y^t)}{d(x, y)} \right)$$

where $P_x^t = \delta_x.P^t$. We also denote

$$\bar{\kappa}(x, y) = \varlimsup_{t \to 0} \frac{1}{t} \left(1 - \frac{W_1(P_x^t, P_y^t)}{d(x, y)} \right).$$

Remark 7 *For every $(x, y, z) \in E^3$, we have the inequality*

$$\kappa(x, z) \geq \frac{d(x, y)\kappa(x, y) + d(y, z)\kappa(y, z)}{d(x, z)}$$

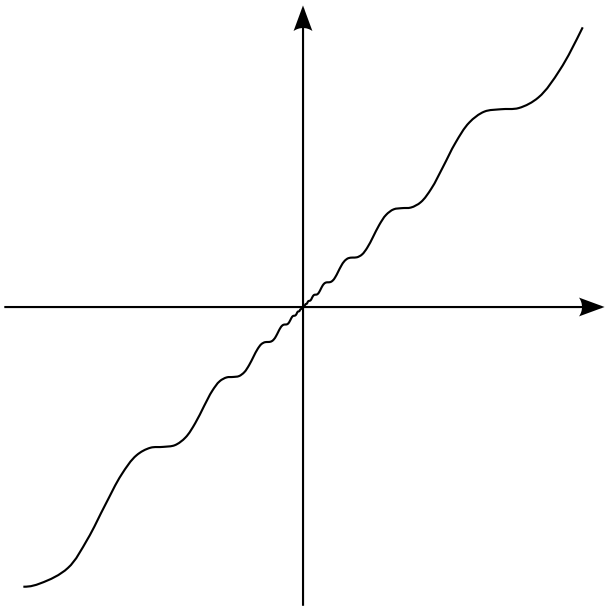
(this trivially comes from the triangular inequality for W_1). This property is not always satisfied by $\bar{\kappa}$ as we will see in the example below. This is the reason why we choose the liminf in the definition of Ricci curvature.

This inequality is particularly interesting when $d(x, z) = d(x, y) + d(y, z)$ because in this case, the right-hand term of the inequality is a convex combination of $\kappa(x, y)$ and $\kappa(y, z)$, so we have $\kappa(x, z) \geq \min(\kappa(x, y), \kappa(y, z))$.

So in the case of ε-geodesic spaces (see Proposition 19 in [7]), the infimum of $\kappa(x, y)$ on E^2 equals the infimum of $\kappa(x, y)$ for the couples (x, y) such that $d(x, y) \leq \varepsilon$. So we only have to pay attention to "local" curvature.

Example 8 *Let us illustrate the difference between κ and $\bar{\kappa}$. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be an increasing continuous function. Then the deterministic kernel defined on the space $f(\mathbb{R})$ by $P_x^t = \delta_{f(f^{-1}(x)+t)}$ is Markovian. We choose f such that there exist $t_1 < t_2 < t_3$ such that*

$$f(t) = \begin{cases} f(t_1) + (t - t_1)(1 + \frac{1}{\sqrt{2}} \sin(\ln(|t - t_1|))) & \textit{in a neighborhood of } t_1 \\ f(t_2) + t - t_2 & \textit{in a neighborhood of } t_2 \\ f(t_3) + (t - t_3)(1 + \frac{1}{\sqrt{2}} \sin(\ln(|t - t_3|))) & \textit{in a neighborhood of } t_3 \end{cases}$$



The graph above is the one of the function $x \mapsto x(1 + \frac{\sin(\frac{2\pi \ln(|x|)}{\ln(2)})}{\sqrt{1+\frac{4\pi^2}{\ln^2(2)}}})$*, it illustrates the behaviour of* f *on the neighborhoods of* t_1 *and* t_3*.*

Then, if we note $x = f(t_1)$*,* $y = f(t_2)$*,* $z = f(t_3)$*, the curvatures are* $\kappa(x,y) = -\frac{1}{\sqrt{2}(y-x)}$*,* $\kappa(y,z) = -\frac{1}{\sqrt{2}(z-y)}$ *and* $\kappa(x,z) = 0$*, whereas* $\bar{\kappa}(x,y) = \frac{1}{\sqrt{2}(y-x)}$*,* $\bar{\kappa}(y,z) = \frac{1}{\sqrt{2}(z-y)}$ *and* $\bar{\kappa}(x,z) = 0$*. We have* $\kappa(x,z) \geq \inf(\kappa(x,y), \kappa(y,z))$*, as stated in Remark 7, and the same is not true for* $\bar{\kappa}$*.*

2 W_1 contraction in positive coarse Ricci curvature

It is known ([1, 3, 4, 5, 7]) that for Markov Chains, a positive coarse Ricci curvature implies that the Markov operator acting on measures is contractive for the W^1 distance. The following Theorem is a generalization of this result to continuous-time Markov processes:

Theorem 9 *Let* P^t *be the semigroup of a left-continuous Markov process satisfying* $\bar{\kappa}(x,y) \geq K > -\infty$ *for all* (x,y) *in* E^2 *with* $x \neq y$*. Then we have:*

$$\forall (x,y) \in E^2, W_1(P_x^t, P_y^t) \leq d(x,y)\mathrm{e}^{-Kt}.$$

The hypothesis of this Theorem states that for every $\varepsilon > 0$ and $(x,y) \in E^2$, there exist $t < \varepsilon$ such that $W_1(P_x^t, P_y^t) \leq d(x,y)\mathrm{e}^{-(K-\varepsilon)t}$,

but we do not control how this t depends on x and y. The infimum of this t on every neighborhood of every pair of points could be 0, so we have to refine Ollivier's proof for the discrete time case.

The hypothesis of this Theorem may seem difficult to check on concrete examples, but we can hope to compute κ, or $\bar{\kappa}$ thanks to the generator of the Markov process in classical cases, under some assumption about the growth in t of the first momentum of P_x^t, as we did in [10] for diffusion processes on manifolds.

Remark 10 *In particular, the same inequality* $W_1(P_x^t, P_y^t) \leq d(x,y)\mathrm{e}^{-Kt}$ *holds if* $\kappa(x,y) \geq K$.

Corollary 11 *We have* $\inf_{x,y} \bar{\kappa}(x,y) = \inf_{x,y} \kappa(x,y)$ *for left-continuous Markov processes.*

Proof of the Corollary : Since $\bar{\kappa} \geq \kappa$, we trivially have $\inf_{x,y} \bar{\kappa}(x,y) \geq \inf_{x,y} \kappa(x,y)$. Now set $K = \inf_{x,y} \bar{\kappa}(x,y)$. Theorem 9 tells us that for any (x,y), $W_1(P_x^t, P_y^t) \leq d(x,y)\mathrm{e}^{-Kt}$, so the definition of $\kappa(x,y)$ implies

$$\kappa(x,y) \geq \varliminf_{t\to 0} \frac{1}{t}\left(1 - \frac{d(x,y)\mathrm{e}^{-Kt}}{d(x,y)}\right) = \varliminf_{t\to 0} \frac{1-\mathrm{e}^{-Kt}}{t} = K.$$

Thus $\inf_{x,y} \bar{\kappa}(x,y) \leq \inf_{x,y} \kappa(x,y)$.$\square$

Remark 12 *Usually in the literature, the processes are chosen right-continuous, but Theorem 9 also works when the process admits a left-continuous modification. Indeed, the conclusion of the theorem only depends on the law of the process.*

So Theorem 9 does apply to diffusion processes and to minimal jump processes as defined in [2], when they do not explode in a finite time. Indeed, such processes admit left-continuous modifications. In the case of minimal jump processes, we just have to replace the value of the process at the time of the jump with the value of the process just before the jump to make it left-continuous, and this is a modification, because for every t, the probability that the process jumps at time t is 0.

Theorem 9 also applies to some jump processes with an infinite number of jumps in a finite time, provided the locations of the jumps tend to one state in E, from which the jump process restarts, as in the following example.

Example 13 *Take the process on* $E = \{0\} \cup \{2^{-n}, n \in \mathbb{N}\}$ *defined in the following way: jump from state* 2^{-n} *to state* $2^{-(n+1)}$ *after a time*

of exponential law $\mathcal{E}(2^n+1)$. As $\sum_{n=0}^{\infty}\frac{1}{2^n+1}<\infty$, the sum of the times of the jumps converges almost surely. After this infinite number of jumps, the process restarts at 0 and then jumps to 1 after a time of law $\mathcal{E}(\frac{1}{2})$.

In this example, we have $\kappa(0,2^{-n})=\frac{3}{2}$ and $\kappa(2^{-n},2^{-(n+1)})=\frac{1}{2}$, so thanks to Remark 7, we have $\inf(\kappa(x,y))=\frac{1}{2}$, so we can use Theorem 9.

Corollary 14 *Let (P^t) be the semigroup of a Markov process on a Polish space admitting a left-continuous modification. Assume that $\kappa(x,y)>K>0$ for all $(x,y)\in E^2$ with $x\neq y$, and that there exist some $x_0\in E$, $t_0>0$ and $M>0$ such that $W_1(\delta_{x_0},P^t_{x_0})<M$ for every $0\leq t\leq t_0$.*

Then the Markov process admits an unique equilibrium probability measure. This equilibrium measure has a finite first moment.

Proof of the Corollary: We consider the process starting at x_0, restricted to times which are integer multiples of t_0. Using the W_1 contraction implied by Theorem 9 ,we can easily prove by induction that $W_1\left(P^{nt_0}_{x_0},P^{(n+1)t_0}_{x_0}\right)\leq M\mathrm{e}^{-Knt_0}$. So the sequence $(P^{nt_0}_{x_0})$ is a Cauchy sequence for W_1, and then it converges to a limit π in the W_1 distance, and π admits a finite first moment. Now if t is not an integer multiple of t_0, we have $W_1\left(P^{\left\lfloor\frac{t}{t_0}\right\rfloor t_0}_{x_0},P^t_{x_0}\right)\leq M\mathrm{e}^{-K\left\lfloor\frac{t}{t_0}\right\rfloor t_0}$, and the right hand term tends to 0 when t tends to infinity. Thus the family $(P^t_{x_0})$ also tends to π.

Now we have for every $T>0$,

$$W_1(\pi.P^t,\pi)\leq W_1(\pi.P^t,P^{T+t}_{x_0})+W_1(P^{T+t}_{x_0},\pi)\leq \mathrm{e}^{-Kt}W_1(\pi,P^T_{x_0})+W_1(P^{T+t}_{x_0},\pi)$$

and the right hand term tends to 0 when T tends to the infinity, so $W_1(\pi.P^t,\pi)=0$ and thus π is invariant.

Since $W_1(P^t_x,P^t_{x_0})\leq \mathrm{e}^{-Kt}d(x,x_0)$, (P^t_x) converges to π in W_1 distance and thus in weak convergence topology for every $x\in E$. Then $\mu.P^t$ converges weakly to π for every probability measure μ, including any invariant one. Thus π is the unique equilibrium probability measure.□

Theorem 9 implies Theorem 2 as follows.

Proof of Theorem 2: Let π be the unique reversible probability measure. Theorem 9 implies that the operator P^t acting on the space $\mathrm{Lipsch}_0(\pi)$ of Lipschitz functions with mean 0 (with respect to π) has a norm smaller than e^{-Kt}. Under the hypothesis of Theorem 2, the L^2 norm is controlled by the Lipschitz norm because

$\mathrm{Var}_\pi(f) \leq \mathbb{E}_\pi[(f(x)-f(x_0))^2] \leq \|f\|^2_{\mathrm{Lipsch}}\mathbb{E}_\pi[d(x,x_0)^2]$ (keep in mind that $\mathbb{E}_\pi[d(x,x_0)^2]$ is assumed to be finite). Now, for any self-adjoint operator S on a Hilbert space H, for any $x \in H, x \neq 0$, we have $\frac{\|Sx\|}{\|x\|} \leq \sqrt{\frac{\|S^2x\|}{\|x\|}}$, because $\frac{\|S^2x\|^2}{\|x\|^2} - \frac{\|Sx\|^4}{\|x\|^4} = \frac{\|S^2x - \frac{\|Sx\|^2}{\|x\|^2}x\|^2}{\|x\|^2} \geq 0$. So by induction, we get $\frac{\|Sx\|}{\|x\|} \leq \left(\frac{\|S^{2^n}x\|}{\|x\|}\right)^{\frac{1}{2^n}}$. As π is reversible, P^t is self-adjoint on $L^2_0(\pi)$, so we use this result with $S = P^t$ and $x = f \in \mathrm{Lipsch}_0(\pi) \subset L^2_0(\pi)$: we get

$$\frac{\|P^t f\|_{L^2}}{\|f\|_{L^2}} \leq \left(\frac{\sqrt{\mathbb{E}_\pi[d(x,x_0)^2]}\|f\|_{\mathrm{Lipsch}}\mathrm{e}^{-2^n Kt}}{\|f\|_{L^2}}\right)^{\frac{1}{2^n}}.$$

The right hand term tends to e^{-Kt} when n tends to infinity. So we have shown that $\|P^t f\|_{L^2} \leq \mathrm{e}^{-Kt}\|f\|_{L^2}$ for any $f \in \mathrm{Lipsch}_0(\pi)$. The probability measure π is regular (see, for example [8]), so indicator functions can be approximated in $L^2(\pi)$ norm by Lipschitz functions, so Lipschitz functions are dense in $L^2(\pi)$. Thus, $\mathrm{Lipsch}_0(\pi)$ is dense in $L^2_0(\pi)$. The operator P^t is 1-Lipschitz on $L^2(\pi)$, so it is continuous on $L^2_0(\pi)$, and then $\|P^t f\|_{L^2} \leq \mathrm{e}^{-Kt}\|f\|_{L^2}$ for any $f \in L^2_0(\pi)$.□

Example 15 *Consider the Brownian motion on the circle $\mathbb{R}/2\pi\mathbb{Z}$, equipped with the "Euclidean" distance:*

$$d(\theta_1,\theta_2) = 2\left|\sin\left(\frac{\theta_2-\theta_1}{2}\right)\right|.$$

This distance is not geodesic, so we have to compute $\kappa(\theta_1,\theta_2)$ for each (θ_1,θ_2) such that $\theta_2 \neq \theta_1 + 2k\pi$, not only for those such that $|\theta_1-\theta_2| < \varepsilon$. The distance is smooth and bounded, so we can use the formula in [10] to compute the coarse Ricci curvature. The Taylor expansion of the distance is:

$$d(\theta_1+\varepsilon v, \theta_2+\varepsilon w) = d(\theta_1,\theta_2)\left(1 + \frac{w-v}{2\tan\left(\frac{\theta_2-\theta_1}{2}\right)} - \frac{(w-v)^2}{8} + O(\varepsilon^3)\right).$$

So from [10], the coarse Ricci curvature is $\kappa(\theta_1,\theta_2) = 0 - \frac{-\frac{1}{4}-\frac{1}{4}}{2} + \sqrt{\frac{1}{4}\times\frac{1}{4}} = \frac{1}{2}$.

The process has a positive curvature with this non-geodesic distance, and Theorem 2 gives the right spectral gap for the generator $\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}$. If we use the geodesic distance, we get nothing because curvature is 0.

To prove Theorem 9, we need a generalization of stopping times, which we call weak stopping times:

Definition 16 *Let X_t be a random process on a probability space Ω, and $\mathcal{F}_t$ be its natural filtration. Let $\mathcal{F}_\infty = \mathfrak{S}((\mathcal{F}_t)_{t\in\mathbb{R}_+})$ be the σ-algebra generated by all the $\mathcal{F}_t$. A random variable T is a* weak stopping time *for X_t if there exists a σ-algebra $\mathcal{G}$ independent of $\mathcal{F}_\infty$ such that T is a stopping time for the filtration $\mathcal{G}_t = \mathfrak{S}(\mathcal{G}, \mathcal{F}_t)$, ie a positive real-valued random variable T such that $\forall t \geq 0, \{\omega \in \Omega | T(\omega) \leq t\} \in \mathcal{G}_t$.*

Lemma 17 *Let T be a random variable having the form $T = \varphi(\omega, v)$, where ω and v are independent and ω is a left-continuous process, and $\forall v, \forall t, \mathbf{1}_{\varphi(\omega,v)\leq t}$ only depends on $\omega|_{[0,t]}$. Then T is a weak stopping time for the process ω, with $\mathcal{G}$ the σ-algebra generated by v.*

Conversely, let X_t be a left-continuous random process, and T be any weak stopping time for X_t. Let ω and u be two independent random variables on another probability space, having the law of X and the uniform law on $[0, 1]$ (and assume that ω is left-continuous). Then there exists a measurable function $\varphi(\omega, u)$ such that $(\omega_t, \varphi(\omega, u))$ has the law of (X_t, T).

Proof: The first part of the Lemma is trivial once we note that the measurable sets which depend on $\omega|_{[0,t]}$ and v are in $\mathfrak{S}(\mathcal{G}, \mathcal{F}_t)$, with $\mathcal{G}$ the σ-algebra of events only depending on v and $\mathcal{F}_t$ the natural filtration of ω.

So let us prove the other part of the Lemma. Take $Y_t = \mathbb{E}[\mathbf{1}_{T<t}|\mathcal{F}_\infty]$. It is $\mathcal{F}_t$-measurable because T is a weak stopping time, so there exists a measurable function f_t such that $Y_t = f_t(X|_{[0,t]})$. If $t_1 < t_2$, we have $\mathbf{1}_{T<t_1} \leq \mathbf{1}_{T<t_2}$, then $Y_{t_1} \leq Y_{t_2}$ almost surely by taking the conditional expectation with respect to $\mathcal{F}_\infty$. So for all ω outside an exceptional set N of mesure 0, the function $t \mapsto f_t(\omega|_{[0,t]})$ is non-decreasing on the subset of rational times, and bounded by 1. We define the events $A_t = \{\omega | \exists \omega' \notin N, \omega'|_{[0,t]} = \omega|_{[0,t]}\}$. We take

$$f'_t(\omega|_{[0,t]}) = \sup_{t'<t, t\in\mathbb{Q}} (f_{t'}(\omega|_{[0,t']})\mathbf{1}_{A_t} + \mathbf{1}_{A_t^c}).$$

Then for all ω, $t \mapsto f'_t(\omega|_{[0,t]})$ is non-decreasing and left continuous. Furthermore, we can write $f'_t(X|_{[0,t]}) = \mathbb{E}[\mathbf{1}_{T<t}|\mathcal{F}_\infty]$ by using the fact that $\mathbf{1}_{T<t}$ is the limit of the increasing sequence $\mathbf{1}_{T<t'_i}$ for any increasing rational sequence t'_i converging to t, and the monotone convergence theorem. Here, $t \mapsto f_t(\omega|_{[0,t]})$ is a kind of repartition function of the conditional law of T knowing ω. We just have to take $\varphi(\omega, u) = \sup\{t \in \bar{\mathbb{R}}_+ | f'_t(\omega|_{[0,t]}) < u\}$. $\square$

Definition 18 *If $(S, \mathcal{A})$ is a measurable space, a* kernel *on S will be a measurable application from S to the set of probability measures on S.*

If k is a kernel on S and μ is a finite measure on S, $\mu.k$ is the finite measure defined by $\mu.k(A) = \int k(x)(A)\mathrm{d}\mu(x)$. If k_1 and k_2 are two kernels on S, $k_1 * k_2$ is the kernel defined by $k_1 * k_2(x)(A) = \int k_2(y)(A)\mathrm{d}k_1(x)(y)$.

Proof of Theorem 9: Let $t > 0$ and $\varepsilon > 0$. We will show that for every x and y,

$$W_1(P_x^t, P_y^t) \leq d(x, y)\mathrm{e}^{-(K-\varepsilon)t}.$$

We denote by $M^{(x)}$ the Markov process starting at point $x \in E$, and $M_t^{(x)}$ its value at time t. We consider the set $\mathcal{K}$ of kernels k on $E^2 \times [0, t]$ satisfying:

- $\forall (x, y, s), \int d(X, Y)\mathrm{e}^{(K-\varepsilon)S}\mathrm{d}k((x, y, s))(X, Y, S) \leq d(x, y)\mathrm{e}^{(K-\varepsilon)s}$.
- $\forall (x, y, s), k((x, y, s))(E^2 \times [0, s[) = 0$ (i.e. k is a time increasing kernel)
- there exist weak stopping times T and T' for the Markov process starting at x and y, depending measurably on (x, y, s), such that for any random variable (X, Y, S) having the law $k(x, y, s)$, we have $(X, S) \sim (M_T^{(x)}, T + s)$ and $(Y, S) \sim (M_{T'}^{(y)}, T' + s)$.

Let $(x_0, y_0) \in E^2$, and $\mathcal{I}$ be the set $\{k((x_0, y_0, 0)), k \in \mathcal{K}\}$. Our goal is to prove that there exists an element (X, Y, S) of $\mathcal{I}$ satisfying $S = t$ almost-surely, because this would provide us a coupling between $M_t^{(x_0)}$ and $M_t^{(y_0)}$ satisfying

$$\mathbb{E}[d(X, Y)] \leq d(x_0, y_0)\mathrm{e}^{-(K-\varepsilon)t}.$$

We will prove that $\mathcal{I}$ is an inductive set for a well-chosen order relation, and that any maximal element of $\mathcal{I}$ (whose existence is guaranteed by Zorn's Lemma) satisfies $S = t$ almost surely.

To do this, we will prove some nice properties of $\mathcal{K}$:

Lemma 19 *The set $\mathcal{K}$ is stable under $*$, and any sequence $(k_i)_{i\in\mathbb{N}^*}$ of elements of $\mathcal{K}$ satisfies that the sequence of the products $(k_1 * k_2 * \cdots * k_n)_{n\in\mathbb{N}^*}$ has a limit k_∞ in $\mathcal{K}$, in the sense that for all $(x, y, s) \in E^2 \times [0, t]$, the sequence $(k_1 * k_2 * \cdots * k_n)((x, y, s))$ weakly converges to $k_\infty((x, y, s))$.*

Proof of the Lemma:

Let $(k_i)_{i\in\mathbb{N}}$ be a sequence of elements of $\mathcal{K}$. Let $(x,y,s)\in E^2\times[0,t]$, and $(X_i,Y_i,S_i)_{i\in\mathbb{N}^*}$ be a Markov chain with non-stationary kernel k_i. Then (X_n,Y_n,S_n) has law $(k_1*k_2*\cdots*k_n)(x,y,s)$. The quantity $\mathbb{E}[d(X_i,Y_i)\mathrm{e}^{(K-\varepsilon)S_i}]$ is non-increasing in i. Indeed, we have

$$\begin{aligned}\mathbb{E}[d(X_{i+1},Y_{i+1})\mathrm{e}^{(K-\varepsilon)S_{i+1}}] &= \mathbb{E}[\mathbb{E}[d(X_{i+1},Y_{i+1})\mathrm{e}^{(K-\varepsilon)S_{i+1}}|(X_i,Y_i,S_i)]]\\ &\leq \mathbb{E}[d(X_i,Y_i)\mathrm{e}^{(K-\varepsilon)S_i}].\end{aligned}$$

Thus $\mathbb{E}[d(X_i,Y_i)\mathrm{e}^{(K-\varepsilon)S_i}]\leq d(x,y)\mathrm{e}^{(\kappa-\varepsilon)s}$.

Taking $i=2$ in the previous expression just says that k_1*k_2 satisfies the first condition in the definition of $\mathcal{K}$. We will prove below that the sequence (X_i,Y_i,S_i) converges almost surely, and we note $(X_\infty,Y_\infty,S_\infty)$ the limit of this sequence. Then, thanks to the monotone convergence theorem, we will get $\mathbb{E}[d(X_\infty,Y_\infty)\mathrm{e}^{(K-\varepsilon)S_\infty}]\leq d(x,y)\mathrm{e}^{(K-\varepsilon)s}$, which is the first condition to check for proving that k_∞ belongs to $\mathcal{K}$.

Now we construct variables (X_i',Y_i',S_i') having the same law as (X_i,Y_i,S_i) over the appropriate probability spaces to prove we have the weak stopping times required by the definition of $\mathcal{K}$.

Let Ω^0 be the space of left-continuous functions from $\mathbb{R}_+$ to E. Let us apply Lemma 17 to the weak stopping times T coming from the third condition of the definition of $\mathcal{K}$ applied to k_i: there exist measurable functions φ_i from $E^2\times[0,t]\times\Omega^0\times[0,1]$ to $\mathbb{R}_+$ such that for every $(x',y',s')\in E^2\times[0,t]$, $\mathbf{1}_{\varphi_i(x',y',s',\omega,u)\leq s''}$ does not depend on the values of ω for times greater than s'', and if we choose ω and u two independent random variables with laws $\mathbb{P}_{x'}$ (the law of the Markov process starting at x') and the uniform law on $[0,1]$, then $(\omega(\varphi_i(x',y',s',\omega,u)),\varphi_i(x',y',s',\omega,u))$ has the law of $(X',S'-s')$ where (X',Y',S') has the law $k_i(x',y',s')$.

Using disintegration of measure on $k_i(x',y',s')$ gives us the existence of a conditional law of Y' knowing $X'=x''$ and $S'=s''$ when (X',Y',S') has the law $k_i(x',y',s')$, and this conditional law depends measurably on (x',y',s',x'',s''). Furthermore, as E is a Polish space, any probability measure on E is the law of $f(u)$ with $f:[0,1]\mapsto E$ a measurable function, depending measurably on the probability measure on E and u is a random variable with the uniform law on $[0,1]$. Then there exist measurable functions $\psi_i:E^2\times[0,t]\times E\times[0,t]\times[0,1]\mapsto E$ such that the law of $\psi_i(x',y',s',x'',s'',u)$ with u a uniform random variable on $[0,1]$ is the conditional law of Y' knowing $X'=x'',S'=s''$, where X',Y',S' has the law $k_i(x',y',s')$.

We take $\Omega=\Omega_0\times[0,1]^{\mathbb{N}}\times[0,1]^{\mathbb{N}}$, and we will denote by $(\omega,(u_i)_{i\in\mathbb{N}},(v_j)_{j\in\mathbb{N}})$ the typical element of this set. We put on the space Ω the probability measure $\mathbb{P}_x\otimes\mathcal{U}([0,1])^{\otimes\mathbb{N}}\otimes\mathcal{U}([0,1])^{\otimes\mathbb{N}}$, which depends on (x,y,s) in a

measurable way. Now we define the following random variables over Ω:

$$\begin{aligned}
&X'_0 = x\\
&Y'_0 = y\\
&S'_0 = s\\
&\omega_0 = \omega\\
&X'_{i+1} = \omega(\varphi_{i+1}(X'_i, Y'_i, S'_i, \omega_i, u_i))\\
&S'_{i+1} = S_i + \varphi_{i+1}(X'_i, Y'_i, S'_i, \omega_i, u_i)\\
&Y'_{i+1} = \psi_{i+1}(X'_i, Y'_i, S'_i, X'_{i+1}, S'_{i+1}, v_i)\\
&\forall s', \omega_{i+1}(s') = \omega(s' + \varphi_{i+1}(X'_i, Y'_i, S'_i, \omega_i, u_i))
\end{aligned}$$

(in other words, ω_{i+1} is ω_i "shifted" by $S'_{i+1} - S'_i$).

We prove by induction that for all n, $S'_n - s$ is a weak stopping time with $\mathcal{G} = \mathfrak{S}((u_i)_{i\in\mathbb{N}}, (v_j)_{j\in\mathbb{N}})$ and the conditional law of $(\omega_n, (u_{n+i}), (v_{n+j}))$ knowing $(X'_0, Y'_0, S'_0, \ldots, X'_n, Y'_n, S'_n)$ is $\mathbb{P}_{X'_n} \otimes \mathcal{U}([0,1])^{\otimes\mathbb{N}} \otimes \mathcal{U}([0,1])^{\otimes\mathbb{N}}$.

The case $n = 0$ is trivial. If we fix $u_0, \ldots, u_n, v_0, \ldots, v_n$ and take $s \leq s' \leq t$, we have to show that the function $\mathbf{1}_{S'_{n+1} \leq s'}$ does not depend on the values of ω for times greater than $s' - s$. Because of the property of φ_{n+1}, we know that if $S'_n < s'$, the function $\mathbf{1}_{S'_{n+1} \leq s'}$ does not depend on the values of ω_n for times greater than $s' - S'_n$, that is, on the values of ω for times greater than $s' - s$. The induction hypothesis tells that $S'_n - s$ is a weak stopping time with $\mathcal{G} = \mathfrak{S}((u_i)_{i\in\mathbb{N}}, (v_j)_{j\in\mathbb{N}})$, and then the event $S'_n > s'$ does not depend on values of ω for times greater than s'. So S'_{n+1} is a weak stopping time with $\mathcal{G} = \mathfrak{S}((u_i)_{i\in\mathbb{N}}, (v_j)_{j\in\mathbb{N}})$. We can use the Markov property, so the conditional law of ω_{n+1} knowing $(X'_0, Y'_0, S'_0, \ldots, X'_{n+1}, Y'_{n+1}, S'_{n+1})$ is $\mathbb{P}_{X'_{n+1}}$. As (X'_k, Y'_k, S'_k) only depends on ω and the u_i's and v_j's with i and j smaller than $k-1$, so the subsequence (u_{n+1+i}, v_{n+1+j}) is independent of $(X'_0, Y'_0, S'_0, \ldots, X'_{n+1}, Y'_{n+1}, S'_{n+1})$.

So the conditional law of $(X'_{i+1}, Y'_{i+1}, S'_{i+1})$ knowing $(X'_0, Y'_0, S'_0, \ldots, X'_i, Y'_i, S'_i)$ is $k_{i+1}((X'_i, Y'_i, S'_i))$. Thus $((X'_i, Y'_i, S'_i))_{i\in\mathbb{N}}$ and $((X_i, Y_i, S_i))_{i\in\mathbb{N}}$ have the same law.

The sequence (S'_i) is non-decreasing and bounded by t, so it converges almost surely to a limit S'_∞, which is also a weak stopping time, because the supremum of a family of stopping times for the filtration $\mathcal{G}_t$ is a stopping time for the filtration $\mathcal{G}_t$. Because of the left continuity of ω, the sequence (X'_i) converges to $X'_\infty = \omega(S'_\infty - s)$.

Of course, we can do the same thing by swapping the roles of x and y to define (X''_i, Y''_i, S''_i), and then we have the convergence of (S''_i) to a weak stopping time S''_∞ and the convergence of (Y''_i) to $Y''_\infty = \omega'(S''_\infty - s)$. So we have proved that $((X_i, Y_i, S_i))$ converges almost surely, thus we have the existence of a limit k_∞ of $k_1 * k_2 * \cdots * k_n$.

The fact that S_2', S_2'', S_∞' and S_∞'' are weak stopping times show us that $k_1 * k_2$ and k_∞ satisfy the last two points of the definition of $\mathcal{K}$, so they belong to $\mathcal{K}$.$\square$

End of the proof of Theorem 9:

Let us put the following partial order relation on $\mathcal{I}$: $\mu_1 \preceq \mu_2$ if and only if there exists $k \in \mathcal{K}$ so that $\mu_2 = \mu_1.k$. First we check that $\preceq$ is an order relation. Transitivity of $\preceq$ is due to the fact that $\mathcal{K}$ is stable under $*$. Reflexivity is a consequence that $1_* \in \mathcal{K}$, with $1_* : (x, y, s) \to \delta_{(x,y,s)}$ the trivial kernel. Antisymmetry is a bit harder to check. Suppose $\mu \preceq \nu \preceq \mu$. We have $\nu = k_1.\mu$ and $\mu = k_2.\nu$ with $(k_1, k_2) \in \mathcal{K}^2$. We construct (X_0, Y_0, S_0) of law μ, (X_1, Y_1, S_1) of conditional law $k_1(X_0, Y_0, S_0)$ knowing (X_0, Y_0, S_0), and (X_2, Y_2, S_2) of conditional law $k_2(X_1, Y_1, S_1)$ knowing $(X_0, Y_0, S_0, X_1, Y_1, S_1)$. Then we have $S_0 \leq S_1 \leq S_2$ almost surely, so $\mathbb{E}[S_0] \leq \mathbb{E}[S_1] \leq \mathbb{E}[S_2] = \mathbb{E}[S_0]$ (S_0 and S_2 have the same law). As $S_1 - S_0 \geq 0$ almost surely and $\mathbb{E}[S_1 - S_0] = 0$, we have $S_0 = S_1$ almost surely. Since $k_1 \in \mathcal{K}$, we then have $k_1(x, y, s) = \delta_{(x,y,s)}$, μ-almost surely, and so $\nu = \mu$.

Now we will prove that $\mathcal{I}$ is an inductive set. Let $A \subset \mathcal{K}$ be a totally ordered subset. If A is empty, then $\delta_{(x_0,y_0,0)} \in \mathcal{I}$ is an upper bound of A. Otherwise, we consider $M = \sup_{\mu \in A} \mathbb{E}_\mu[S] \in [0, t]$. If there exists $\mu \in A$ such that $\mathbb{E}_\mu[S] = M$, then μ is the maximum of A. In the remaining case, there exists an increasing sequence $(\mu_i)_{i\in\mathbb{N}} \in A^{\mathbb{N}}$ such that $\mathbb{E}_{\mu_i}[S] \nearrow M$, and so for every $\mu \in A$, there exists $i \in \mathbb{N}$ so that $\mu \preceq \mu_i$, because A is totally ordered and $\mu \preceq \mu' \Rightarrow \mathbb{E}_\mu[S] \leq \mathbb{E}_{\mu'}[S]$. Any upper bound of all the μ_i will be an upper bound of A. For each i, there exists $k_i \in \mathcal{K}$ so that $\mu_{i+1} = \mu_i.k_i$. Then by lemma 19 $\mu_\infty = \lim_{i\to\infty} \mu_i$ exists, belongs to $\mathcal{I}$ and we have for each i, $\mu_\infty = \mu_i.(k_i * k_{i+1} * \dots)$. So μ_∞ is an upper bound of A.

We can apply Zorn's lemma to $\mathcal{I}$ to get a maximal element $\mu_{\max}$. Then we have $\mu_{\max}.k = \mu_{\max}$ for every $k \in \mathcal{K}$. Let us prove that under $\mu_{\max}$, $s = t$ almost surely. To do so, we will construct a particular $k \in \mathcal{K}$ such that for all (x, y, s), we have $s = t$ or $\mathbb{P}_{(x',y',s')\sim k(x,y,s)}(s' > s) = 1$, and then the fact that $\mu_{\max}.k = \mu_{\max}$ implies that $s = t$ almost surely under $\mu_{\max}$.

By definition of $\bar{\kappa}$, for each (x, y, s) with $s < t$, there exists $0 < \eta(x, y, s) \leq t-s$ such that $W_1(\mathbb{P}_x^{\eta(x,y,s)}, \mathbb{P}_y^{\eta(x,y,s)}) \leq d(x, y)\mathrm{e}^{-(K-\varepsilon)\eta(x,y,s)}$ (because $\bar{\kappa}(x, y) > K - \varepsilon$). So we have a coupling $\xi(x, y, s)$ between $\mathbb{P}_x^{\eta(x,y,s)}$ and $\mathbb{P}_y^{\eta(x,y,s)}$ such that $\mathbb{E}_{\xi(x,y,s)}[d(X, Y)] \leq d(x, y)\mathrm{e}^{-(K-\varepsilon)\eta(x,y,s)}$.

It remains to prove that we can choose $\eta(x, y, s)$ and $\xi(x, y, s)$ in a measurable way to get our k. A simple choice for $\eta(x, y, s)$ is the maximal one

$$\sup(\{\eta \in]0, t - s] | W_1(\mathbb{P}_x^\eta, \mathbb{P}_y^\eta) \leq d(x, y)\mathrm{e}^{-(K-\varepsilon)\eta}\}),$$

which is measurable. The fact that this supremum is actually a maximum is due to the existence of a left continuous modification. Indeed, let $(\eta_i)_{i\in\mathbb{N}}$ be a maximizing sequence for the expression above, and η be the supremum. Let f be any bounded 1-lipschitz function from E to $\mathbb{R}$. Because of the left continuous modification, $M_{\eta_i}^{(x)}$ converges to $M_\eta^{(x)}$ and $M_{\eta_i}^{(y)}$ converges to $M_\eta^{(y)}$. So by the dominated convergence theorem, $\mathbb{E}[f(M_{\eta_i}^{(x)})]$ converges to $\mathbb{E}[f(M_\eta^{(x)})]$ and $\mathbb{E}[f(M_{\eta_i}^{(y)})]$ converges to $\mathbb{E}[f(M_\eta^{(y)})]$. Thus $\int f \mathrm{d}(\mathbb{P}_x^{\eta_i} - \mathbb{P}_y^{\eta_i})$ converges to $\int f \mathrm{d}(\mathbb{P}_x^{\eta} - \mathbb{P}_y^{\eta})$ and this latter is smaller than $d(x,y)\mathrm{e}^{-(K-\varepsilon)\eta}$. Then, there exists a measurable way to choose an optimal coupling between two probability measures (Corollary 5.22 in [11]), thus we can get a measurable $\xi(x,y,s)$.

We can then set $k(x,y,s) = \xi(x,y,s) \otimes \delta_{s+\eta(x,y,s)}$ for $s < t$, and $k(x,y,t) = \delta_{(x,y,t)}$ (because $\eta(x,y,s) \geq 0$ is trivially a weak stopping time). Since $\eta(x,y,s) > 0$, and $\mu_{\max}.k = \mu_{\max}$, we have $s = t$ $\mu_{\max}$-almost surely, so $\mu_{\max}$ provides a coupling between $\mathbb{P}_{x_0}^t$ and $\mathbb{P}_{y_0}^t$ which satisfies $\mathbb{E}[d(X,Y)] \leq d(x_0,y_0)\mathrm{e}^{-(\kappa-\varepsilon)t}$, so $W_1(\mathbb{P}_{x_0}^t, \mathbb{P}_{y_0}^t) \leq d(x_0,y_0)\mathrm{e}^{-(\kappa-\varepsilon)t}$ as needed. Letting ε decrease to 0 gives the conclusion of Theorem 9.□

References

[1] R. Bubley, M. E. Dyer, *Path coupling: a technique for proving rapid mixing in Markov chains*, FOCS 1997, 223-231.

[2] M.F.Chen, *From Markov chains to non-equilibrium particle systems*, Singapore: Word Scientific, 2004

[3] M.F.Chen, F.Y.Wang, *Application of coupling method to the first eigenvalue on manifold*, Sci.Sin. Ser A 37 1–14, 1994

[4] R.L.Dobrushin*Prescribing a system of random variables by conditional expectations*, Theory of Probability and its Applications **15** (1970) n° 3, 458-486.

[5] R.Dobrushin, *Perturbation methods of the theory of Gibbsian fields*, in R. Dobrushin, P. Groeneboom, M. Ledoux, *Lectures on probability theory and statistics*, Lectures from the 24th Saint-Flour Summer School held July 7-23, 1994, edited by P. Bernard, Lecture Notes in Mathematics **1648**, Springer, Berlin (1996), 1-66.

[6] A.Lichnerowicz, *Géométrie des groupes de transformations*, Dunod, 1958

[7] Y.Ollivier, *Ricci curvature of Markov chains on metric spaces*, J.Funct.Anal.256, n°3,p 810–864, 2009

[8] K.R.Parthasarathy, *Probability measures on metric spaces*, Academic press, New York, 1967

[9] L.Veysseire, *A harmonic mean bound for the spcetral gap of the Laplacian on Riemannian manifolds*, Comptes rendus mathematique, 2010 vol 438, p 1319–1322

[10] L.Veysseire, *Improved spectral gap bounds on positively curved Manifolds*

[11] C.Villani, *Optimal transport, old and new*, Springer-Verlag, 2009