HAL open science

Courbure de Ricci grossière de processus markoviens

Laurent Veysseire

To cite this version:

Laurent Veysseire. Courbure de Ricci grossière de processus markoviens. Mathématiques générales [math.GM]. Ecole normale supérieure de lyon - ENS LYON, 2012. Français. NNT : 2012ENSL0732 . tel-00737099

HAL Id: tel-00737099

https://theses.hal.science/tel-00737099v1

Submitted on 1 Oct 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.

UNIVERSITÉ DE LYON – ÉCOLE NORMALE SUPÉRIEURE DE LYON
N° d'ordre : 2012ENSL0732

THÈSE

en vue d'obtenir le grade de

Docteur de l'École Normale Supérieure de Lyon – Université de Lyon

Spécialité : **Mathématiques**

préparée au laboratoire **Unité de Mathématiques Pures et Appliquées**

dans le cadre de l'École doctorale **INFOMATH**

présentée et soutenue publiquement par

Laurent VEYSSEIRE

le 16 juillet 2012

Courbure de Ricci grossière de processus markoviens

Thèse dirigée par **Yann OLLIVIER**

Après avis de :

Franck BARTHE	Rapporteur
Anton THALMAIER	Rapporteur

Devant le jury composé de

Franck BARTHE	Rapporteur
Christophe GARBAN	Membre
Christian LEONARD	Membre
Yann OLLIVIER	Directeur de thèse
Anton THALMAIER	Rapporteur

Résumé

La courbure de Ricci grossière d'un processus markovien sur un espace polonais est définie comme un taux de contraction local de la distance de Wasserstein W_1 entre les lois du processus partant de deux points distincts.

La première partie de cette thèse traite de résultats valables dans le cas d'espaces polonais quelconques. Le plus simple d'entre eux est que l'infimum de la courbure de Ricci grossière est un taux de contraction global du semi-groupe du processus pour la distance W_1 entre mesures de probabilité. Quoiqu'intuitif, ce résultat est difficile à démontrer en temps continu. La preuve de ce résultat, et les conséquences qui en découlent pour le trou spectral du générateur font l'objet du chapitre 1.

Un autre résultat intéressant, faisant intervenir les valeurs de la courbure de Ricci grossière en différents points, et pas seulement son infimum, est un résultat de concentration des mesures d'équilibre, valable uniquement en temps discret. Il sera traité dans le chapitre 2

La seconde partie de cette thèse traite du cas particulier des diffusions sur les variétés riemanniennes. Des calculs explicites peuvent alors être effectués. Une formule est donnée permettant d'obtenir la courbure de Ricci grossière à partir du générateur de la diffusion. Dans le cas où la métrique est adaptée à la diffusion, nous montrons l'existence d'un couplage entre les trajectoires issues de deux points distincts tel que la courbure de Ricci grossière est exactement le taux de décroissance de la distance entre ces trajectoires. On arrive alors à prouver que le trou spectral du générateur de la diffusion est plus grand que la moyenne harmonique de la courbure de Ricci. Ce résultat peut être généralisé lorsque la métrique n'est pas celle induite par le générateur, mais il nécessite une hypothèse très contraignante, et la courbure que l'on doit choisir n'est plus la courbure de Ricci grossière, mais une autre qui est toujours plus petite.

Remerciements

En tout premier lieu, je souhaite adresser mes remerciements à mon directeur de thèse, Yann Ollivier, qui a investi beaucoup de son temps pour encadrer cette thèse, et dont l'aide m'a été précieuse à de nombreuses reprises.

Je tiens également à remercier Vincent Beffara, Olivier Druet et Cédric Villani, qui m'ont apporté leur aide lorsque je les ai sollicités.

Enfin, je remercie Franck Barthe et Anton Thalmaier d'avoir accepté d'être les rapporteurs de cette thèse, ainsi que Christophe Garban et Christian Leonard d'en être membres du jury.

Table des matières

Résumé 1
Remerciements 1

Introduction **4**
0.1 Un peu de géométrie riemannienne 4
0.1.1 Variétés riemanniennes et géodésiques 4
0.1.2 Dérivation des champs de tenseurs et transport parallèle 5
0.1.3 Tenseur de Riemann, courbure sectionelle et courbure de Ricci 6
0.2 Inégalités de Poincaré 8
0.2.1 Générateur, mesure d'équilibre, mesure réversible . . . 9
0.2.2 L'opérateur carré du champ 10
0.2.3 Inégalité de Poincaré et convergence vers l'équilibre . . 10
0.3 La courbure de Bakry–Émery 11
0.3.1 Processus de diffusion 11
0.3.2 Opérateur carré du champ itéré et inégalité de courbure-dimension 12
0.4 Distances de Wasserstein 14
0.5 La courbure de Lott–Sturm–Villani 15
0.5.1 Définition et lien avec la courbure de Bakry–Émery . . 15
0.5.2 Conséquences d'une courbure positive au sens de Lott–Sturm–Villani 16
0.6 La courbure de Ricci grossière 16
0.6.1 Définition et premières propriétés 17
0.6.2 La courbure de Ricci grossière sur quelques exemples . 17
0.6.3 Les conséquences d'une courbure de Ricci grossière positive 18
0.7 Les nouveaux résultats de cette thèse 19
0.7.1 Courbure de Ricci grossière en temps continu et contraction de la distance W_1 19
0.7.2 Courbure de Ricci grossière et trou spectral 19
0.7.3 Courbure de Ricci grossière et inégalités de concentration 21
0.8 Problèmes ouverts 22

I Courbure de Ricci grossière dans le cas d'un espace quelconque **23**

1 Coarse Ricci curvature for continuous-time Markov processes **24**

1.1 Coarse Ricci curvature: definition and examples 25
1.2 W_1 contraction in positive coarse Ricci curvature 27

2 A concentration theorem for the equilibrium measure of Markov chains with nonnegative coarse Ricci curvature 35
2.1 The concentration Theorems 36
2.2 Some examples 41

II Le cas particulier des diffusions sur les variétés riemanniennes 45

3 Improved spectral gap bounds on positively curved manifolds 46
3.1 The harmonic mean in a nutshell 49
3.2 Coarse Ricci curvature for diffusions on Riemannian manifolds 50
3.2.1 Coarse Ricci curvature: definition and calculation . . . 51
3.2.2 The limit of $\kappa(x, y)$ when x and y are close 58
3.2.3 Construction of the coupling 61
3.2.4 The (H) condition and the curvature $\tilde{\kappa}$ 63
3.3 New bounds for the spectral gap 66

Introduction

La courbure de Ricci grossière que nous allons étudier ici sera définie dans le cadre de processus markoviens sur des espaces métriques, cadre quelque peu éloigné de celui de la géométrie riemannienne dont la courbure de Ricci est issue. Nous commencerons donc dans cette introduction par rappeler quelques notions de géométrie riemannienne et par introduire la courbure de Ricci.

Nous introduirons ensuite différentes extensions de la courbure de Ricci : la courbure de Bakry–Émery, la courbure de Sturm–Lott–Villani et la courbure de Ricci grossière, ainsi que quelques résultats obtenus grâce à ces extensions, tels que inégalités de Sobolev logarithmiques, inégalités de Poincaré ou encore inégalités de concentration.

Nous présenterons ensuite les résultats nouveaux obtenus dans cette thèse : des inégalités de Poincaré pour les diffusions dans les variétés riemanniennes avec une moyenne harmonique de la courbure de Ricci, une généralisation en temps continu de la courbure de Ricci grossière, et une inégalité de concentration obtenue grâce à la courbure de Ricci grossière en temps discret, impliquant une intégrale double de la courbure plutôt que l'infimum de cette courbure.

0.1 Un peu de géométrie riemannienne

La géométrie riemannienne traite de variétés riemanniennes, qui sont des variétés différentiables munies d'une métrique qui ressemble localement à celle d'un espace euclidien. Cette métrique permet de définir une façon de dériver des champs de tenseurs sur la variété. Le défaut de commutativité de deux dérivations successives permet de définir le tenseur de Riemann. Le tenseur de Ricci est alors défini à partir du tenseur de Riemann.

0.1.1 Variétés riemanniennes et géodésiques

Rappelons tout d'abord la définition d'une variété riemannienne.

Définition 0.1 *Une variété riemannienne est un couple $(\mathcal{M}, g)$ où $\mathcal{M}$ est une variété différentiable munie d'une métrique g, qui est un champ de formes bilinéaires symétriques définies positives sur l'espace tangent. C'est à dire pour tout x dans $\mathcal{M}$, $g_x(.,.)$ est une forme bilinéaire symétrique définie positive sur $\mathrm{T}_x\mathcal{M}$, l'espace tangent à $\mathcal{M}$ au point x.*

Pour simplifier, on se placera dans le cadre où $\mathcal{M}$ est une variété C^∞ et g est C^∞.

Grâce à la métrique g, on peut définir la longueur des courbes de la façon suivante : si $\gamma : [0,T] \mapsto \mathcal{M}$ est une courbe C^1, sa longueur est

$$l(\gamma) = \int_0^T \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))}\mathrm{d}t.$$

Étant donnés deux points x et y sur $\mathcal{M}$, on définit la distance $d(x,y)$ comme l'infimum des longueurs des courbes C^1 joignant x et y. Une courbe réalisant cet infimum sera appelée *géodésique*, elles existent et sont uniques pour x et y assez proches. La définition usuelle des géodésiques en géométrie riemannienne est plus générale, nous la verrons un peu plus loin. On peut remarquer que les géodésiques parcourues à vitesse constante (c'est à dire telles que $\sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))} = cste = \frac{l(\gamma)}{T}$) sont exactement les courbes qui minimisent la fonctionelle

$$L(\gamma) = \int_0^T g_{\gamma(t)}(\gamma'(t), \gamma'(t))\mathrm{d}t,$$

à extrémités fixées $\gamma(0) = x$ et $\gamma(T) = y$, et à T fixé. En effet, pour une courbe γ quelconque, si on note $\tilde{\gamma}$ la courbe γ reparamétrée pour que sa vitesse soit constante (donc $\gamma(t) = \tilde{\gamma}\left(\int_0^t \sqrt{g_\gamma(u)(\gamma'(u), \gamma'(u))}\mathrm{d}u\right)$), on obtient

$$L(\gamma) = \int_0^T g_{\gamma(t)}(\gamma'(t), \gamma'(t))\mathrm{d}t \geq \frac{1}{T}\left(\int_0^T \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))}\mathrm{d}t\right)^2 = L(\tilde{\gamma})$$

avec égalité si et seulement si γ est à vitesse constante, auquel cas $L(\gamma) = \frac{l(\gamma)^2}{T}$.

Les équations d'Euler–Lagrange associées à la minimisation de L peuvent s'écrire dans une carte :

$$\ddot{x}^k(t) = -\Gamma_{ij}{}^k(x(t))\dot{x}^i(t)\dot{x}^j(t) \tag{EG}$$

où les coefficients $\Gamma_{ij}^k = \frac{g^{kl}}{2}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})$ sont les coefficients de Christoffel, qui dépendent de la carte choisie et ne représentent pas un objet canonique sur $\mathcal{M}$.

Les courbes qui satisfont ces équations seront également appelées géodésiques, et elles minimisent localement la longueur dans le sens où la restriction de γ aux sous-intervalles $[t_1, t_2]$ minimisent la longueur des courbes joignant $\gamma(t_1)$ et $\gamma(t_2)$ pour $t_2 - t_1$ assez petit.

0.1.2 Dérivation des champs de tenseurs et transport parallèle

Grâce aux géodésiques, on peut définir des cartes canoniques : les cartes exponentielles. Ainsi, on pourra dériver des champs de tenseur, ou même un tenseur défini seulement le long d'une courbe, et définir ainsi le transport parallèle le long d'une courbe.

Définition 0.2 *Soit $x \in \mathcal{M}$, l'application exponentielle au point x est celle qui associe à tout $p \in \mathrm{T}_x\mathcal{M}$ dans un voisinage de 0 la valeur au temps 1, si elle existe, de l'unique solution de l'équation des géodésiques (EG) avec les conditions initiales $x(0) = x$ et $\dot{x}(0) = p$.*

La variété riemannienne est dite *complète* si l'application exponentielle est définie partout (voir le théorème de Hopf–Rinow et son corollaire dans [15] pour d'autres propositions équivalentes).

Cette application exponentielle est un difféomorphisme entre un voisinage de 0 dans $\mathrm{T}_x\mathcal{M}$ et un voisinage de x dans $\mathcal{M}$. Elle définit donc une carte, que l'on appelle la carte exponentielle. Soit $T^{j_1\dots j_l}_{i_1\dots i_k}$ un champ de tenseurs C^∞. Le champ de tenseurs $\nabla_i T^{j_1\dots j_l}_{i_1\dots i_k}$ est défini comme celui qui a pour coordonnées

$$\nabla_i T^{j_1\dots j_l}_{i_1\dots i_k}(x) = \partial_i T^{j_1\dots j_l}_{i_1\dots i_k}(0)$$

dans la carte exponentielle au point x. Ses coordonnées dans une carte quelconque sont alors :

$$\nabla_i T^{j_1\dots j_l}_{i_1\dots i_k} = \partial_i T^{j_1\dots j_l}_{i_1\dots i_k} - \sum_{\alpha=1}^{k} \Gamma_{ii_\alpha}{}^{m} T^{j_1\dots j_l}_{i_1\dots i_{\alpha-1} m i_{\alpha+1}\dots i_k} + \sum_{\alpha=1}^{l} \Gamma_{im}{}^{j_\alpha} T^{j_1\dots j_{\alpha-1} m j_{\alpha+1}\dots j_l}_{i_1\dots i_k}.$$

avec $\Gamma_{ij}{}^{k}$ les coefficients de Christoffel de la carte. L'opérateur ∇ est appelé connexion de Levi–Civita, il définit le gradient des champs de tenseurs de façon canonique.

Supposons que l'on ait un tenseur défini sur une courbe $\gamma(t)$: $T(t) \in \mathrm{T}^*_{\gamma(t)}\mathcal{M}^{\otimes k} \otimes \mathrm{T}_{\gamma(t)}\mathcal{M}^{\otimes l}$. On peut alors dériver T le long de γ :

$$\begin{aligned}(\frac{\mathrm{d}}{\mathrm{d}t}T)^{j_1\dots j_l}_{i_1\dots i_k} = \frac{\mathrm{d}}{\mathrm{d}t}(T^{j_1\dots j_l}_{i_1\dots i_k}) + \Bigg(&- \sum_{\alpha=1}^{k} \Gamma_{ii_\alpha}{}^{m} T^{j_1\dots j_l}_{i_1\dots i_{\alpha-1} m i_{\alpha+1}\dots i_k} \\ &+ \sum_{\alpha=1}^{l} \Gamma_{im}{}^{j_\alpha} T^{j_1\dots j_{\alpha-1} m j_{\alpha+1}\dots j_l}_{i_1\dots i_k} \Bigg) \frac{\mathrm{d}\gamma^i(t)}{\mathrm{d}t}.\end{aligned}$$

Cette façon de dériver T ne dépend pas de la carte choisie car on obtient $\frac{\mathrm{d}}{\mathrm{d}t}T = \nabla_i T \dot{\gamma}^i$ en prolongeant T à $\mathcal{M}$.

Définition 0.3 *Soit $\gamma : [0,T] \mapsto \mathcal{M}$ une courbe C^1. Le transport parallèle de $p \in \mathrm{T}_{\gamma(0)}\mathcal{M}$ le long de γ est l'unique solution $p(t) \in \mathrm{T}_{\gamma(t)}\mathcal{M}$ de $\frac{\mathrm{d}}{\mathrm{d}t}p(t) = 0$ avec la condition initiale $p(0) = p$.*

Le transport parallèle définit des isométries entre les espaces tangents $\mathrm{T}_{\gamma(t)}\mathcal{M}$.

0.1.3 Tenseur de Riemann, courbure sectionelle et courbure de Ricci

La non-commutation des dérivations successives permet de définir le tenseur de Riemann, à partir duquel sont définies courbure sectionelle et courbure de Ricci.

Contrairement au cas de l'espace euclidien, on n'obtient généralement pas un tenseur symétrique en appliquant plusieurs fois l'opérateur ∇. La non-commutation des dérivées successives est caractérisée par le tenseur de Riemann.

Définition 0.4 *Le tenseur de Riemann est l'unique champ de tenseurs* ${R_{ij}}^k{}_l$ *satisfaisant*

$$(\nabla_i \nabla_j - \nabla_j \nabla_i) V^k = {R_{ij}}^k{}_l V^l$$

pour tout champ de vecteurs V *de classe* C^2.

Le tenseur de Riemann s'annule sur toute la variété si et seulement si elle est localement isométrique à l'espace euclidien, c'est pourquoi on l'appelle aussi courbure de Riemann.

Son expression dans une carte quelconque est la suivante :

$${R_{ij}}^k{}_l = \partial_i {\Gamma_{jl}}^k - \partial_j {\Gamma_{il}}^k + {\Gamma_{im}}^k {\Gamma_{jl}}^m - {\Gamma_{jm}}^k {\Gamma_{il}}^m.$$

Dans la carte exponentielle, la métrique est constante à l'ordre 1 et on obtient une expression simplifiée du tenseur de Riemann :

$$R_{ijkl}(x) := g_{kk'}(x) {R_{ij}}^{k'}{}_l(x) = \frac{1}{2}\left(\partial^2_{il} g_{jk} - \partial^2_{ik} g_{jl} - \partial^2_{jl} g_{jk} + \partial^2_{jk} g_{il}\right)(0).$$

On peut voir grâce à cette expression les symétries du tenseur de Riemann : R_{ijkl} est antisymétrique en (i,j) et (k,l), est invariant si on permute simultanément i avec k et j avec l, et satisfait l'identité de Bianchi :

$$R_{ijkl} + R_{jkil} + R_{kijl} = 0.$$

Le tenseur de Riemann caractérise la variation de la métrique à l'ordre 2 dans la carte exponentielle. En utilisant les symétries de $\partial^2_{ij} g_{kl}(0)$ dans la carte exponentielle, on montre que :

$$\partial^2_{ij} g_{kl}(0) = \frac{1}{3}\left(R_{iklj} + R_{ilkj}\right)(x).$$

D'où le développement limité :

$$g_{ij}(\varepsilon u) = g_{ij}(x) - \frac{\varepsilon^2}{3} R_{ikjl}(x) u^k u^l + O(\varepsilon^3). \tag{1}$$

On définit à partir du tenseur de Riemann la courbure sectionnelle de la façon suivante.

Définition 0.5 *Soit* $x \in \mathcal{M}$, *et* $P \subset \mathrm{T}_x\mathcal{M}$ *un sous-espace vectoriel de dimension* 2. *La courbure sectionnelle de* P *est la quantité définie par :*

$$K(P) = R(u,v,u,v) := R_{ijkl}(x) u^i v^j u^k v^l,$$

où (u,v) *est une base orthonormée quelconque de* P.

Vérifions que cette quantité ne dépend pas de la base orthonormée choisie. Soient (u,v) et (u',v') deux bases orthonormées de P. Il existe alors une matrice

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in O_2(\mathbb{R})$$

telle que $u' = a.u + b.v$ et $v' = c.u + d.v$. On a alors

$$\begin{aligned} R(u', v', u', v') =& R(a.u + b.v, c.u + d.v, a.u + b.v, c.u + d.v) \\ =& a^2 d^2 R(u, v, u, v) + abcdR(u, v, v, u) \\ &+ abcdR(v, u, u, v) + b^2 c^2 R(v, u, v, u) \end{aligned}$$

les 12 autres termes sont nuls compte tenu de l'antisymétrie de R en les 2 premières et les 2 dernières variables. On obtient alors $R(u', v', u', v') = (ad - bc)^2 R(u, v, u, v) = R(u, v, u, v)$ car le déterminant d'une matrice de $O_2(\mathbb{R})$ est ± 1.

La connaissance de la courbure sectionnelle de tous les 2-plans de $\mathrm{T}_x\mathcal{M}$ équivaut à celle du tenseur de Riemann au point x.

On a le développement limité suivant pour u et v dans $\mathrm{T}_x\mathcal{M}$:

$$\begin{aligned} d^2(\exp_x(\varepsilon u), \exp_x(\varepsilon v)) &= \varepsilon^2 \|v - u\|^2 - \frac{\varepsilon^4}{3} R(u, v, u, v) \\ &= \varepsilon^2 \|v - u\|^2 - \frac{\varepsilon^4}{3} (\|u\|^2 \|v\|^2 - \langle u, v\rangle^2) K(P) \end{aligned}$$

avec P le 2-plan engendré par u et v (la quantité dans les parenthèses s'annulle si u et v ont l'idée saugrenue d'être colinéaires).

On définit également la courbure de Ricci à partir du tenseur de Riemann.

Définition 0.6 *La courbure de Ricci de $\mathcal{M}$ au point x est le tenseur d'ordre 2 donné par :*

$$\mathrm{Ric}_{ij}(x) := R_{ki}{}^{k}{}_{j}(x) = g^{kl}(x) R_{kilj}(x)$$

Si u est un vecteur unitaire de $\mathrm{T}_x\mathcal{M}$, alors $\mathrm{Ric}(u) := \mathrm{Ric}_{ij}(x) u^i u^j$ est $n - 1$ fois la moyenne de la courbure sectionnelle sur l'ensemble des 2-plans de $\mathrm{T}_x\mathcal{M}$ qui contiennent u, avec n la dimension de $\mathcal{M}$.

Pour ε assez petit, $\exp_x$ est un difféomorphisme entre la boule de rayon ε et de centre 0 de $\mathrm{T}_x\mathcal{M}$ et la boule de rayon ε et de centre x dans $\mathcal{M}$. En utilisant (1), la mesure image de la mesure de volume sur la boule de centre x et de rayon ε par la réciproque de ce difféomorphisme a une densité $(1 - \frac{\varepsilon^2}{6}\mathrm{Ric}(u, u) + O(\varepsilon^3))\mathbf{1}_{\|u\|<1}$ au point u par rapport à la mesure de Lebesgue. On en déduit que le volume riemannien de la boule de centre x et de rayon ε dans $\mathcal{M}$ diffère du cas euclidien d'un facteur $1 - \frac{\varepsilon^2}{6(n+2)} R(x)$, où $R(x)$ est la courbure scalaire au point x, soit la trace de Ricci, ou n fois la moyenne de $\mathrm{Ric}(u, u)$ sur tous les vecteurs unitaires de $\mathrm{T}_x\mathcal{M}$

0.2 Inégalités de Poincaré

Une inégalité de Poincaré est une inégalité fonctionnelle. En probabilités, elle caractérise la vitesse de convergence exponentielle vers la mesure d'équilibre d'un processus markovien.

En analyse, une inégalité de Poincaré est une inégalité de la forme :

$$\int_\Omega f^2 \mathrm{d\,vol} \leq C \int_\Omega \|\nabla f\|^2 \mathrm{d\,vol},$$

pour toute fonction f suffisament régulière telle que $\int_\Omega f \mathrm{d}\,\mathrm{vol} = 0$, avec Ω un ouvert de $\mathbb{R}^n$, ou d'une variété riemannienne.

Dans le cadre des processus markoviens en temps continu, ces inégalités se généralisent en remplaçant $\|\nabla f\|^2$ par $\Gamma(f, f)$, avec Γ l'opérateur carré du champ associé au processus, et en remplaçant $\mathrm{d}\,\mathrm{vol}$ par la mesure d'équilibre du processus.

0.2.1 Générateur, mesure d'équilibre, mesure réversible

Un processus markovien en temps continu est caractérisé par un semi-groupe P^t d'opérateurs agissant sur les fonctions mesurables bornées de l'espace des états E. $P^t f(x)$ est l'espérance de $f(X_t)$, avec X_t la valeur à l'instant t du processus issu de x. Dans la plupart des cas classiques, le processus admet un générateur L, qui est un opérateur agissant sur un domaine dense d'un espace fonctionnel qui contient les fonctions mesurables bornées, et qui satisfait $Lf = \lim_{t\to 0} \frac{P^t f - f}{t}$. Souvent, le générateur suffira à définir le semi-groupe markovien. Il est toutefois possible que des processus markoviens différents aient le même générateur.

Définition 0.7 *On appelle mesure invariante du processus une mesure μ qui reste stable sous l'action de P^t, c'est à dire qui satisfait :*

$$\int P^t f \mathrm{d}\mu = \int f \mathrm{d}\mu$$

pour toute fonction f μ-intégrable et pour tout $t > 0$. En particulier, si f est dans le domaine de L, on a

$$\int Lf \mathrm{d}\mu = 0.$$

Définition 0.8 *On appelle mesure réversible du processus une mesure μ telle que P^t est autoadjoint dans $L^2(\mu)$, c'est à dire qui satisfait :*

$$\int f P^t g \mathrm{d}\mu = \int g P^t f \mathrm{d}\mu$$

pour toutes les fonctions f, g dans $L^2(\mu)$ et tout $t > 0$. En particulier, si f et g sont dans le domaine de L, on a

$$\int f L g \mathrm{d}\mu = \int g L f \mathrm{d}\mu.$$

Si le processus est défini de façon unique par le générateur, on peut se contenter des formules ne faisnt intervenir que L et non P^t pour définir les mesures invariantes et réversibles.

Le cas qui nous intéressera sera celui où μ est une mesure de probabilité. On remarque que dans ce cas, une mesure réversible est une mesure d'équilibre, en prenant $g = 1$ dans la définition 0.8. Intuitivement, la réversibilité signifie que la loi du processus X_t dont le point de départ X_0 est choisi selon la mesure réversible est la même que la loi du processus en temps inversé X_{T-t} entre 0 et T, et ce pour tout $T > 0$.

0.2.2 L'opérateur carré du champ

Supposons que L stabilise une algèbre $\mathcal{A}$ de fonctions sur l'espace des états E, dense dans le domaine de L. On définit ainsi l'opérateur carré du champ.

Définition 0.9 *L'opérateur carré du champ est l'application bilinéaire définie sur $\mathcal{A}$ par :*

$$\Gamma(f,g) = \frac{1}{2}(L(fg) - fL(g) - gL(f)).$$

Un exemple simple est le mouvement brownien sur une variété riemannienne, dont le générateur est :

$$L(f) = \frac{1}{2}\Delta(f) := \frac{1}{2}g^{ij}\nabla_i\nabla_j f.$$

Dans ce cas, l'opérateur carré du champ est $\Gamma(f,g) = \frac{1}{2}g^{ij}\nabla_i f\nabla_j g$, et donc $\Gamma(f,f) = \frac{1}{2}\|\nabla f\|^2$, ce qui justifie l'appellation « carré du champ ».

Un autre exemple classique est la marche aléatoire sur un graphe localement fini, dont le générateur est :

$$Lf(x) = \sum_{y\sim x}(f(y) - f(x)).$$

On a alors $\Gamma(f,g)(x) = \frac{1}{2}\sum_{y\sim x}(f(y)-f(x))(g(y)-g(x))$, et donc $\Gamma(f,f)(x) = \frac{1}{2}\sum_{y\sim x}(f(y) - f(x))^2$.

En fait, $\Gamma(f,f)(x)$ est toujours positif ou nul, car on a

$$\Gamma(f,f)(x) = \lim_{t\to 0}\frac{\mathrm{Var}_{P_x^t}[f]}{2t} \geq 0.$$

0.2.3 Inégalité de Poincaré et convergence vers l'équilibre

Une inégalité de Poincaré est définie, dans le cadre des processus markoviens en temps continu, de la façon suivante :

Définition 0.10 *On dit qu'un processus markovien en temps continu, admettant une mesure de probabilité μ pour mesure d'équilibre, satisfait une inégalité de Poincaré si on a pour toute fonction f telle que $\Gamma(f,f)$ est μ-intégrable :*

$$\mathrm{Var}_\mu[f] \leq C\mathbb{E}_\mu[\Gamma(f,f)]$$

avec C une constante positive.

Si on a une inégalité de Poincaré, alors on a une convergence exponentielle de la mesure à l'instant t vers la mesure d'équilibre [1].

Théorème 0.11 *Soit P^t le semi-groupe d'un processus markovien en temps continu, μ une mesure de probabilité invariante, et ν une mesure de probabilité absolument continue par rapport à μ, et telle que $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^2(\mu)$. Supposons qu'on ait une inégalité de Poincaré avec constante C. Alors on a :*

$$\left\|\frac{\mathrm{d}(\nu P^t)}{\mathrm{d}\mu} - 1\right\|_{L^2(\mu)} \leq \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu} - 1\right\|_{L^2(\mu)} \mathrm{e}^{-\frac{t}{C}}.$$

En effet, on a pour toute fonction $f \in L^2(\mu)$ dans le domaine de L :

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Var}_\mu[P^t f] &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\int (P^t f)^2 \mathrm{d}\mu - \left(\int f \mathrm{d}\mu\right)^2\right) \\ &= \int 2P^t f L(P^t f)\mathrm{d}\mu \\ &= -2\int \Gamma(P^t f, P^t f)\mathrm{d}\mu \\ &\leq -\frac{2}{C}\mathrm{Var}_\mu[P^t f]\end{aligned}$$

On en déduit que pour toute fonction $f \in L^2(\mu)$

$$\mathrm{Var}_\mu[P^t f] \leq \mathrm{Var}_\mu[f]\mathrm{e}^{-\frac{2t}{C}}$$

On a alors

$$\begin{aligned}\int f \mathrm{d}(\nu P^t - \mu) &= \int P^t(f - \int f \mathrm{d}\mu)\mathrm{d}(\nu - \mu) \\ &\leq \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu} - 1\right\|_{L^2(\mu)} \|P^t(f - \int f \mathrm{d}\mu)\|_{L^2(\mu)} \\ &\leq \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu} - 1\right\|_{L^2(\mu)} \mathrm{e}^{-\frac{t}{C}} \|f\|_{L^2(\mu)}.\end{aligned}$$

Le théorème de représentation de Riesz permet d'en déduire que $\frac{\mathrm{d}(\nu P^t)}{\mathrm{d}\mu}$ est dans $L^2(\mu)$ et que sa norme est plus petite que $\left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu} - 1\right\|_{L^2(\mu)} \mathrm{e}^{-\frac{t}{C}}$.

Dans le cas où μ est réversible, le trou spectral de L est $\frac{1}{C}$, avec C la constante optimale dans l'inégalité de Poincaré, et $\frac{1}{C}$ est alors le coefficient optimal de la convergence exponentielle de νP^t vers μ en norme $L^2(\mu)$.

0.3 La courbure de Bakry–Émery

Dans cette section, nous verrons une généralisation de la courbure de Ricci introduite par Bakry et Émery, dans le cadre de processus markoviens de diffusion [4, 5]. Dans le cas des diffusions sur les variétés, on a une expression simple de cette courbure, et on retrouve la courbure de Ricci (à un facteur $\frac{1}{2}$ près) dans le cas du mouvement brownien. L'inégalité de courbure-dimension de Bakry–Émery est une généralisation d'une condition de courbure de Ricci minorée uniformément par une constante positive. Cette inégalité implique une inégalité de Sobolev logarithmique, qui implique en particulier une inégalité de Poincaré (voir [1]), et on obtient donc une borne inférieure pour le trou spectral du générateur de la diffusion.

0.3.1 Processus de diffusion

On dit que le processus markovien est une *diffusion* si l'opérateur carré du champ (voir définition 0.9) est une dérivation dans le sens où il suit une règle de Leibniz :

$$\Gamma(fg, h) = f\Gamma(g, h) + g\Gamma(f, h).$$

Nous considèrerons par la suite des processus de diffusion dans les variétés, dont le générateur prend la forme suivante :

$$Lf = \frac{1}{2}A^{ij}\nabla_i\nabla_j f + B^i\nabla_i f.$$

Où $A^{ij}(x)$ est un tenseur symétrique et positif de $\mathrm{T}_x\mathcal{M}^{\otimes 2}$ et $B^i(x) \in \mathrm{T}_x\mathcal{M}$. Si A^{ij} et B^i sont suffisament réguliers, il existe un unique processus continu maximal (au sens où il est défini jusqu'au temps où la trajectoire est sortie de tout compact) dont le générateur est L.

Dans ce cas, on obtient $\Gamma(f,g) = \frac{1}{2}A^{ij}\nabla_i f\nabla_j g$, et donc c'est bien une dérivation, et le processus est une diffusion dans le sens qu'on a défini plus haut.

0.3.2 Opérateur carré du champ itéré et inégalité de courbure-dimension

L'inégalité de courbure-dimension de Bakry–Émery fait intervenir l'opérateur carré du champ itéré, dont la définition est la suivante.

Définition 0.12 *L'opérateur carré du champ itéré est l'application bilinéaire sur $\mathcal{A}$ définie par :*

$$\Gamma_2(f,g) := \frac{1}{2}(L(\Gamma(f,g)) - \Gamma(f,L(g)) - \Gamma(L(f),g)).$$

L'opérateur Γ_2 est donc à Γ ce que Γ est au produit de 2 fonctions, d'où son appellation.

Dans le cas d'une diffusion sur une variété dont le générateur est de la forme

$$Lf = g^{ij}\nabla_i\nabla_j f + B^i\nabla_i f,$$

l'expression de Γ_2 est la suivante :

$$\Gamma_2(f,h) = \langle \mathrm{Hess}(f)|\mathrm{Hess}(h)\rangle_{HS} + \left(\mathrm{Ric}^{ij} - \frac{1}{2}\left(g^{jk}\nabla_k B^i + g^{ik}\nabla_k B^j\right)\right)\nabla_i f\nabla_j h.$$

Pour une diffusion plus générale sur une variété, on peut se ramener au cas $A^{ij} = 2g^{ij}$ si L est elliptique, en choisissant convenablement la métrique de la variété.

L'inégalité de courbure-dimension de Bakry–Émery est définie de la façon suivante :

Définition 0.13 *On dit que L satisfait une inégalité de courbure-dimension $CD(\rho,n')$ au sens de Bakry–Émery si*

$$\Gamma_2(f,f) \geq \rho\Gamma(f,f) + \frac{1}{n'}(Lf)^2$$

pour toute fonction $f \in \mathcal{A}$.

Proposition 0.14 *Dans le cas du mouvement brownien (accéléré deux fois) sur une variété ($L = \Delta$), on a une inégalité de courbure-dimension de Bakry–Émery $CD(K, n')$, avec K et $n' > 0$ constants, si et seulement si* $\mathrm{Ric}(x)(u,u) \geq K\|u\|^2$ *pour tout $x \in \mathcal{M}$ et $u \in \mathrm{T}_x\mathcal{M}$, et $n' > n$, avec n la dimension de $\mathcal{M}$.*

Cette proposition justifie l'appellation d'inégalité de courbure-dimension.
Preuve : On a ici $\Gamma_2(f,f) = \left(\|\mathrm{Hess}(f)\|_{HS}^2 + \mathrm{Ric}^{ij}\nabla_i f \nabla_j f\right)$ et $\Gamma(f,f) = \|\nabla f\|^2$. En outre, pour les matrices $n \times n$, on a $\|M\|_{HS}^2 = \mathrm{tr}(M^T M) \geq \frac{1}{n}(\mathrm{tr}(M))^2$, avec égalité pour les matrices scalaires. Puisque $\Delta(f) = \mathrm{tr}(\mathrm{Hess}(f))$, on a bien une inégalité $CD(\rho, n)$ avec n la dimension de $\mathcal{M}$ et ρ l'infimum de la courbure de Ricci. Et donc également une inégalité $CD(K, n')$ avec $K \leq \rho$ et $n' \geq n$.

Réciproquement, supposons qu'on ait une inégalité $CD(K, n')$.

Alors si on choisit $x \in \mathcal{M}$ et f telle que $\nabla f(x) = 0$ et $\mathrm{Hess}(f)(x) = g(x)$, l'inégalité $CD(K, n')$ au point x nous donne $\frac{1}{n} \geq \frac{1}{n'}$, et donc $n' \geq n$ si $n' > 0$.

Pour la courbure, pour tout $\varepsilon > 0$, il existe $x \in \mathcal{M}$ et $u \in \mathrm{T}_x\mathcal{M}$ tels que $\|u\| = 1$ et $\mathrm{Ric}(x)(u,u) \leq \rho + \varepsilon$, avec ρ l'infimum de la courbure de Ricci, comme précédemment. On choisit alors f telle que $\nabla f(x) = g(x)^{-1}u$ et $\mathrm{Hess}(f)(x) = 0$, et l'inégalité $CD(K, n')$ au point x nous donne $\rho + \varepsilon \geq K$. Ceci étant vrai pour tout $\varepsilon > 0$, on en déduit $K \leq \rho$. $\square$

Reprenons notre exemple $Lf = g^{ij}\nabla_i\nabla_j f + B^i \nabla_i f$. Pour un point $x \in \mathcal{M}$, on cherche à minimiser la quantité $(\Gamma_2(f,f) - \frac{1}{n'}(Lf)^2)(x)$ pour trouver le $\rho(x)$ optimal (le plus grand possible). Cette quantité dépend de $\nabla f(x)$ et $\nabla^2 f(x)$. Pour f telle que $\nabla f(x) = 0$ et $\nabla_i\nabla_j f(x) = g_{ij}(x)$, l'inégalité $CD(\rho, n')$ impose $\frac{1}{n'} \leq \frac{1}{n}$, avec n la dimension de $\mathcal{M}$. À $\nabla f(x)$ fixé, le minimum est atteint pour $\nabla_i\nabla_j f(x) = \frac{g_{ij}}{n'-n} B^k \nabla_k f(x)$, et vaut

$$\left(\mathrm{Ric}^{ij} - \frac{1}{2}\left(g^{jk}\nabla_k B^i + g^{ik}\nabla_k B^j\right) - \frac{B^i B^j}{n'-n}\right)(x)\nabla_i f(x)\nabla_j f(x).$$

Cette quantité est bilinéaire en ∇f, on en déduit que le $\rho(x)$ optimal cherché est la plus petite valeur propre de

$$\left(\mathrm{Ric}^{ij} - \frac{1}{2}\left(g^{jk}\nabla_k B^i + g^{ik}\nabla_k B^j\right) - \frac{B^i B^j}{n'-n}\right)(x) g_{jk}(x).$$

Sous une hypothèse de réversibilité, une inégalité de courbure-dimension implique une inégalité de Sobolev logarithmique [5] :

Théorème 0.15 *Soit π une mesure de probabilité réversible d'un processus de diffusion satisfaisant une inégalité courbure-dimension au sens de Bakry–Émery $CD(\rho, n)$, avec $\rho > 0$ et $n \geq 1$ constants. Alors on a l'inégalité de Sobolev logarithmique :*

$$\mathrm{Ent}_\pi(f^2) \leq \frac{2(n-1)}{n\rho}\mathbb{E}_\pi[\Gamma(f,f)],$$

avec $\mathrm{Ent}_\pi(f^2) := \mathbb{E}_\pi[f^2 \ln(f^2)] - \mathbb{E}_\pi[f^2]\ln(\mathbb{E}_\pi[f^2])$ *l'entropie de f^2 par rapport à π.*

Sur notre exemple $Lf = g^{ij}\nabla_i\nabla_j f + B^i\nabla_i f$, l'hypothèse de réversibilité est équivalente à $B^i = -g^{ij}\nabla_j\varphi$, avec $\pi = \frac{\mathrm{e}^{-\varphi}}{\int_{\mathcal{M}} \mathrm{e}^{-\varphi}\mathrm{d}\,\mathrm{vol}}$.

Puisqu'une inégalité de Sobolev logarithmique implique une inégalité de Poincaré, sous les hypothèses du théorème 0.15, L admet un trou spectral $\lambda_1 \geq \frac{n\rho}{n-1}$.

0.4 Distances de Wasserstein

Pour introduire la courbure de Lott–Sturm–Villani et la courbure de Ricci grossière, on a besoin des distances de Wasserstein W_1 et W_2. Ces distances entre mesures de probabilité ont pour origine le problème de Monge–Kantorovitch.

On se donne deux mesures μ_1 et μ_2 de même masse totale sur un même espace E, et on cherche à minimiser le coût d'un transport de μ_1 vers μ_2. Un transport de μ_1 vers μ_2 est donné par une mesure ξ sur $E \times E$ dont les marginales sont μ_1 et μ_2, où $\mathrm{d}\xi(x,y)$ représente la quantité transportée de x vers y. Le coût du transport à minimiser est de la forme $C(\xi) = \int_{E\times E} c(x,y)\mathrm{d}\xi(x,y)$, où la fonction c est le coût unitaire de transport de x vers y.

Les distances de Wasserstein sont définies grâce à la solution de ce problème de minimisation pour $c(x,y) = d(x,y)^p$.

Définition 0.16 *Soit (E,d) un espace polonais, et μ_1 et μ_2 deux mesures de probabilité sur E. Soit $1 \leq p < +\infty$. La distance de Wasserstein d'ordre p entre μ_1 et μ_2 est définie par :*

$$W_p(\mu_1,\mu_2) := \left(\inf_{\xi\in\Phi(\mu_1,\mu_2)} \int_{E\times E} d(x,y)^p \mathrm{d}\xi(x,y)\right)^{\frac{1}{p}},$$

avec $\Phi(\mu_1,\mu_2)$ l'ensemble des couplages de μ_1 et μ_2, c'est-à-dire des mesures de probabilité sur $E \times E$ dont les marginales sont μ_1 et μ_2.

Cette distance peut éventuellement valoir $+\infty$, par exemple si μ_1 a un moment d'ordre p fini et pas μ_2. Vérifions que W_p satisfait l'inégalité triangulaire. Soient μ_0, μ_1 et μ_2 trois mesures de probabilité sur E. Soit $\varepsilon > 0$, il existe alors des couplages $\xi_1 \in \Phi(\mu_0,\mu_1)$ et $\xi_2 \in \Phi(\mu_1,\mu_2)$ tels que $\int d(x,y)^p\mathrm{d}\xi_1(x,y) \leq (W_p(\mu_0,\mu_1)+\varepsilon)^p$ et $\int d(x,y)^p\mathrm{d}\xi_2(x,y) \leq (W_p(\mu_1,\mu_2)+\varepsilon)^p$. On peut alors construire 3 variables aléatoires X_0, X_1 et X_2 sur un même espace de probabilités Ω, telles que la loi de (X_0,X_1) soit ξ_1 et la loi de (X_1,X_2) soit ξ_2 (voir « Gluing Lemma » dans [31]). La loi de (X_0,X_2) fournit un couplage entre μ_0 et μ_2. On a :

$$\begin{aligned}
\mathbb{E}[d(X_0,X_2)^p]^{\frac{1}{p}} &= \|d(X_0,X_2)\|_{L^p(\Omega)} \leq \|d(X_0,X_1)+d(X_1,X_2)\|_{L^p(\Omega)} \\
&\leq \|d(X_0,X_1)\|_{L^p(\Omega)} + \|d(X_1,X_2)\|_{L^p(\Omega)} \\
&\leq W_p(\mu_0,\mu_1) + W_p(\mu_1,\mu_2) + 2\varepsilon.
\end{aligned}$$

On a donc $W_p(\mu_0,\mu_2) \leq W_p(\mu_0,\mu_1) + W_p(\mu_1,\mu_2)$.

Dans le cas particulier $p = 1$, le théorème de dualité de Kantorovitch–Rubinstein donne une autre interprétation de la distance W_1.

Théorème 0.17 *Soient μ_1 et μ_2 deux mesures de probabilité sur un espace polonais (E, d). Alors, on a :*

$$W_1(\mu_1, \mu_2) = \sup_{f \text{ 1-Lipschitz}} \int_E f \mathrm{d}\mu_1 - \int_E f \mathrm{d}\mu_2.$$

0.5 La courbure de Lott–Sturm–Villani

La courbure de Lott–Sturm–Villani est définie dans le cadre des espaces métriques mesurés. Elle correspond à la convexité de l'entropie relative dans l'espace des mesures de probabilités. Elle est une extension de la courbure de Bakry–Émery.

0.5.1 Définition et lien avec la courbure de Bakry–Émery

La courbure de Lott–Sturm–Villani d'un espace métrique mesuré admet la définition suivante (voir [26]) :

Définition 0.18 *Soit (M, d, μ) un espace métrique mesuré, où μ est une mesure borélienne. On dit que (M, d, μ) a une courbure de Lott–Sturm–Villani supérieure à $K \in \mathbb{R}$ si pour tout couple de mesures de probabilité ν_0, ν_1 sur M, ayant un moment d'ordre 2 et une entropie relative par rapport à μ finis, il existe une géodésique $\gamma(t)$ pour la distance W_2 joignant ν_0 à ν_1 telle que*

$$\mathrm{Ent}(\gamma(t)|\mu) \leq (1-t)\mathrm{Ent}(\nu_0|\mu) + t\mathrm{Ent}(\nu_1|\mu) - \frac{K}{2}t(1-t)W_2^2(\nu_0, \nu_1)$$

avec

$$\mathrm{Ent}(\nu|\mu) := \int \ln\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\nu$$

l'entropie relative de ν par rapport à μ.

La courbure de Lott–Sturm–Villani de (M, d, μ) est alors le plus grand K possible dans la définition ci-dessus.

Dans le cas où (M, d) est une variété riemannienne, et où $\mathrm{d}\mu = \mathrm{e}^{-V}\mathrm{d}\,\mathrm{vol}$ avec $V : M \mapsto \mathbb{R}$ une fonction assez régulière, il existe une diffusion « canonique » dont μ est une mesure réversible. Le générateur de cette diffusion est le suivant :

$$Lf = \Delta f - \langle \nabla V, \nabla f \rangle.$$

On a alors le résultat suivant (Théorème 4.9 de [26]) :

Théorème 0.19 *Soit (M, d) une variété riemannienne, et $V : M \mapsto \mathbb{R}$ une fonction de classe $\mathcal{C}^2$. L'espace métrique mesuré $(M, d, \mathrm{e}^{-V}.\mathrm{vol})$ a une courbure de Lott–Sturm–Villani supérieure à K si et seulement si la diffusion de générateur*

$$L = \delta - \langle \nabla V, \nabla \rangle$$

vérifie une inégalité de courbure-dimension $CD(K, \infty)$ au sens de Bakry–Émery.

0.5.2 Conséquences d'une courbure positive au sens de Lott–Sturm–Villani

Dans leur article [22], Lott et Villani utilisent une définition plus forte de cette courbure. En effet, pour que la courbure de (M, d, μ) soit plus grande que K, il faut qu'une famille de fonctionnelles U_μ soient convexes par déplacement, dans le sens où pour tout couple (μ_0, μ_1) de mesures de probabilité sur M, il existe une géodésique $(\mu_t)_{t\in[0,1]}$ pour W_2 telle que toutes les fonctionnelles U_μ satifont :

$$U_\mu(\mu_t) \leq (1-t)U_\mu(\mu_0) + tU_\mu(\mu_1) - \lambda(U)t(1-t)W_2^2(\mu_0, \mu_1).$$

Les fonctionnelles U_μ sont définies par

$$U_\mu(\nu) = \int_M U(\rho(x))\mathrm{d}\mu(x) + U'(\infty)\nu_s(M)$$

pour toute fonction U sur $\mathbb{R}^+$ convexe (d'où l'existence de $U'(\infty)$) s'annulant en 0 et telles que $\lambda \to \mathrm{e}^{-\lambda}U(\mathrm{e}^\lambda)$ soit convexe, avec ρ et ν_s telles que

$$\nu = \rho\mu + \nu_s$$

soit la décomposition de Lebesgue de ν par rapport à μ. La constante $\lambda(U)$ vaut $\frac{1}{2}\inf_{r>0} K(U'_+(r)) - \frac{U(r)}{r})$.

C'est une condition plus forte que celle de Sturm, qui correspond au seul cas $U(x) = x\ln(x)$. Toutefois, elle y est équivalente dans le cas des variétés riemanniennes.

On a alors

Théorème 0.20 *Si un espace métrique mesuré (M, d, μ) a une courbure supérieure à $K > 0$ au sens de Lott et Villani, alors il vérifie une inégalité de Sobolev logarithmique*

$$\mathrm{Ent}_\mu[f^2] \leq \frac{2}{K}\int \|\nabla f\|^2 \mathrm{d}\mu$$

et donc en particulier une inégalité de Poincaré

$$\int f^2 \mathrm{d}\mu - \frac{(\int f \mathrm{d}\mu)^2}{\mu(M)} \leq \frac{1}{K}\int \|\nabla f\|^2$$

avec $\|\nabla f(x)\| = \overline{\lim}_{y\to x} \frac{|f(y)-f(x)|}{d(x,y)}$ le gradient local de f au point x.

0.6 La courbure de Ricci grossière

La courbure de Ricci grossière est une notion locale, définie pour des chaînes de Markov sur un espace métrique. Un contrôle de cette courbure permet d'obtenir une inégalité de Poincaré, ou encore des inégalités de concentration pour la mesure d'équilibre.

0.6.1 Définition et premières propriétés

Dans le cadre des chaînes de Markov (donc en temps discret), la courbure de Ricci grossière est définie ([24], Définition 3) entre deux points distincts, x et y, par la formule :

$$\kappa(x,y) = 1 - \frac{W_1(P_x, P_y)}{d(x,y)}.$$

Une interprétation intuitive de cette courbure est qu'elle est positive si et seulement si les distributions P_x et P_y sont plus proches, pour la métrique W_1, que les diracs δ_x et δ_y.

En fait, l'infimum de la courbure de Ricci grossière est donné par le taux de contraction pour la distance W_1 de l'opérateur markovien P agissant sur les mesures. Si $K = \inf_{x \neq y \in X} \kappa(x,y)$, alors pour tout couple de mesures de probabilité (μ, ν), on a :

$$W_1(\mu P, \nu P) \leq (1-K) W_1(\mu, \nu).$$

Cette condition de contraction de la métrique de Wasserstein estaussi connue sous le nom de *critère de Dobrushin–Shlosman* [13, 14, 10], ou encore *couplage de trajectoires* [8].

La courbure de Ricci grossière étant définie pour un couple de points, elle n'est pas aussi locale que la courbure de Ricci sur une variété. Toutefois, pour un espace ε-géodésique, un contrôle local de cette courbure (seulement pour les couples de points distants de moins de ε) est suffisant. Un espace est dit *ε-géodésique* si pour tout couple de points x et y, il existe une suite de points $x = x_0, x_1, \ldots, x_n = y$ telle que la distance entre deux points successifs de cette suite soit toujours onférieure à ε, et satisfaisant

$$d(x,y) = d(x_0, x_1) + d(x_1, x_2) + \ldots + d(x_{n-1}, x_n).$$

Si (X, d) est ε-géodésique, alors on a :

$$\inf_{x \neq y} \kappa(x,y) = \inf_{0 < d(x,y) \leq \varepsilon} \kappa(x,y).$$

Les espaces métriques géodésiques sont ε-géodésiques pour tout $\varepsilon > 0$. Un autre exemple d'espace ε-géodésique est l'ensemble des sommets d'un graphe non-orienté, muni de la distance associée au graphe, qui est ε-géodésique pour tout $\varepsilon \geq 1$.

0.6.2 La courbure de Ricci grossière sur quelques exemples

Un des avantages de la courbure de Ricci grossière par rapport à la courbure de Lott–Sturm–Villani est qu'elle est plus facile à calculer. Nous donnons ici quelques exemples sur lesquels on peut calculer ou estimer la courbure de Ricci grossière.

Si une chaîne de Markov sur $\mathbb{R}^n$ muni d'une norme quelconque satisfait une invariance par translation (dans le sens où pour tout x et h dans $\mathbb{R}^n$, la mesure P_{x+h} est la mesure image de P_x par la translation de vecteur h), alors

la courbure de Ricci grossière $\kappa(x,y)$ est nulle pour tout $x \neq y$. En effet, on a $W_1(P_x, P_y) \leq d(x,y)$ puisque la translation de vecteur $y - x$ définit un couplage entre P_x et P_y tel que la distance soit $d(x,y)$ presque sûrement. Donc $\kappa(x,y) \geq 0$. Pour l'inégalité inverse, on utilise le théorème de dualité de Kantorovitch–Rubinstein. On pose $f(z) = \zeta(z)$, avec ζ une forme linéaire de norme 1 telle que $\zeta(y-x) = \|y-x\|$. La fonction f est 1-lipschitzienne, et on a $\int f \mathrm{d}(P_y - P_x) = d(x,y)$, donc $W_1(P_x, P_y) \geq d(x,y)$. Donc $\kappa(x,y) \leq 0$.

L'exemple précédent inclut également les marches aléatoires sur $\mathbb{Z}^n$ invariantes par translation.

La deuxième partie de la preuve précédente s'applique également dans le cas de chaînes de Markov sur $\mathbb{R}^n$ qui sont des martingales. Ces martingales ont donc une courbure de Ricci négative ou nulle.

Un exemple classique est la marche aléatoire paresseuse sur $\{0,1\}^n$ muni de la distance de Hamming, dont le noyau de transition peut être décrit comme suit : partant de l'état $(e_i)_{i=1}^n$, on choisit un indice i uniformément entre 1 et n, puis on change la valeur de e_i avec probabilité $\frac{1}{2}$. La courbure de Ricci grossière de cette chaîne de Markov entre deux points distants de 1 est $\frac{1}{2n}$.

Soit (X,d) une variété riemannienne, on considère la chaîne de Markov dont les transitions sont des sauts uniformes dans une boule de petit rayon ε centrée autour du point de départ. La courbure de Ricci grossière de cette chaîne de Markov entre x et $\exp_x(\varepsilon u)$ se comporte comme $\frac{\varepsilon^2}{2(n+2)}\mathrm{Ric}_x(u,u)$ lorsque ε tend vers 0, ce qui justifie l'appellation « courbure de Ricci grossière ».

0.6.3 Les conséquences d'une courbure de Ricci grossière positive

Si la courbure de Ricci grossière d'une chaîne de Markov est minorée par une constante $K > 0$, alors l'opérateur P est $(1-K)$-contractant pour W_1. La suite des mesures $\delta_x P^n$ est une suite de Cauchy pour W_1, et il existe alors une unique mesure de probabilité invariante.

En outre, on obtient une inégalité de Poincaré, avec constante K, si la chaîne de Markov est réversible ou si l'espace des états X est fini.

Sous des hypothèses raisonnables sur le noyau de transition P, une courbure de Ricci grossière minorée par $K > 0$ implique que, sous la mesure d'équilibre de la chaîne de Markov, les fonctions 1-Lipschitziennes sont concentrées autour de leur moyenne. La borne sur $\mathbb{P}[f - \mathbb{E}[f] > t]$ est gaussienne jusqu'à un certain $t_{\max}$, puis exponentielle ensuite ([24], Théorème 33).

Le même type de concentration peut être obtenu avec la mesure empirique sur un bout de trajectoire ([19], Théorèmes 4 et 5).

Si on sait seulement que la courbure est positive ou nulle, l'existence d'un point attractif implique une concentration exponentielle de toute mesure d'équilibre autour de ce point ([24], Théorème 49).

0.7 Les nouveaux résultats de cette thèse

Les chapitres 1, 2 et 3 de ce manuscrit reprennent des article éponymes [28], [29] et [30] dans lesquels sont exposés les nouveaux résultats de cette thèse.

On trouve dans le chapitre 2 une preuve de la contraction pour la distance W_1 pour les processus markoviens en temps continu dont la courbure de Ricci grossière est minorée par une constante strictement positive.

Des bornes inférieures du trou spectral sont démontrées au chapitre 1 dans un cadre général, et au chapitre 3 dans le cas de diffusions sur les variétés riemanniennes.

Des inégalités de concentration sont obtenues au chapitre 2, en utilisant la courbure de Ricci grossière locale et pas seulement son infimum.

0.7.1 Courbure de Ricci grossière en temps continu et contraction de la distance W_1

On définit la courbure de Ricci grossière en temps continu comme suit. Pour un processus markovien en temps continu sur un espace polonais (X, d), de semi-groupe P^t, la courbure de Ricci grossière entre deux points x et y de X est la quantité

$$\kappa(x,y) = \underline{\lim}_{t\to 0} \frac{1}{t}\left(1 - \frac{W_1(\delta_x P^t, \delta_y P^t)}{d(x,y)}\right).$$

On choisit $\underline{\lim}$ dans la définition car cela permet d'avoir une notion de courbure locale en espace dans le cas où X est géodésique, ou ε-géodésique.

Théorème 0.21 *Soit P^t le semi-groupe d'un processus markovien continu à gauche. Supposons qu'il existe $K > -\infty$ tel que pour tout $x \neq y$ on ait $\kappa(x,y) \geq K$. On a alors l'inégalité suivante :*

$$W_1(\delta_x P^t, \delta_y P^t) \leq d(x,y)\mathrm{e}^{-Kt},$$

pour tout x et y dans X, et pour tout $t \geq 0$.

En fait, ce résultat reste vrai si on avait pris $\overline{\lim}$ au lieu de $\underline{\lim}$ dans la définition de la courbure de Ricci grossière.

Si $K > 0$, alors le semi-groupe P^t contracte la distance W_1.

Ce théorème semble très simple, mais il est bien plus difficile à prouver que son homologue en temps discret. La difficulté de ce résultat réside dans le fait qu'à priori, on n'a pas de $t_0 > 0$ uniforme tel que le processus soit contractant pout tout temps $t \leq t_0$ et tout couple x, y de points de départ. Les hypothèses donnent seulement un t_0 local, et l'existence d'un t_0 global est la conclusion du théorème.

0.7.2 Courbure de Ricci grossière et trou spectral

Dans un cadre général, on montre au chapitre 1 que le trou spectral du générateur d'un processus markovien dont la courbure de Ricci grossière est minorée par une constante $K > 0$ est supérieur à cette constante.

Théorème 0.22 *Soit L le générateur d'un processus markovien continu à gauche admettant une mesure de probabilité réversible π. Supposons que la courbure de Ricci grossière de ce processus est uniformément minorée par $K > 0$, et que la mesure π admet un moment d'ordre 2 fini.*

Alors L admet un trou spectral dans $L^2(\pi)$, et il est supérieur à K.

Ce théorème est une généralisation d'un théorème de Lichnerowicz [21] (avec une perte d'un facteur $\frac{n}{n-1}$), valable pour le laplacien sur une variété riemannienne à courbure de Ricci minorée par une constante positive. Il ressemble au théorème 1.9 de [10], mais diffère de celui-ci du fait que l'hypothèse de contraction est locale.

Au chapitre 3, on traite le cas particulier des diffusions sur les variétés riemanniennes.

Le calcul de la courbure de Ricci grossière à partir du générateur montre que la courbure de Ricci grossière coïncide avec la courbure de Bakry–Émery (avec dimension infinie) dans le cas où la métrique de la variété est adaptée à la diffusion.

On montre alors que dans ce cas, si la diffusion est réversible et que la courbure de Ricci grossière est positive, le trou spectral du générateur est minoré par la moyenne harmonique de la courbure de Ricci grossière par rapport à la mesure réversible.

Théorème 0.23 *Soit $L = \Delta - \nabla\varphi.\nabla$ le générateur d'une diffusion réversible sur une variété riemannienne compacte X. On pose, pour tout $x \in X$,*

$$\kappa(x) := \inf_{u \in \mathrm{T}_x X \,|\, \|u\|=1} \mathrm{Ric}_x(u,u) + \mathrm{Hess}(\varphi)(u,u).$$

On note π la mesure de probabilité réversible ($\mathrm{d}\pi = \frac{\mathrm{e}^{-\varphi}\mathrm{d\,vol}}{\int_X \mathrm{e}^{-\varphi}\mathrm{d\,vol}}$). Si la courbure $\kappa(x)$ est positive, alors le trou spectral λ_1 de L dans $L^2(\pi)$ satisfait :

$$\lambda_1 \geq \frac{1}{\int_X \frac{\mathrm{d}\pi(x)}{\kappa(x)}} \geq \inf_{x \in X} \kappa(x).$$

Ce résultat n'est pas le premier à utiliser une courbure non-constante au lieu de son infimum pour obtenir une minoration du trou spectral. La proposition 1.5 de [3] donne une estimation du trou spectral dans le cas où la courbure est proche d'une constante positive, au sens de la norme $L^p(\pi)$, avec p dépendant de la dimension de la variété ($p > \frac{n}{2}$), ce qui autorise un peu de courbure négative.

En adaptant une deuxième preuve du théorème 0.23, on obtient un raffinement de celui-ci pour la courbure de Bakry–Émery associée à une dimension finie.

Théorème 0.24 *Soit $L = \Delta - \nabla\varphi.\nabla$ le générateur d'un processus de diffusion réversible sur une variété riemannienne compacte. Supposons que ce processus de diffusion satisfait une inégalité $CD(\rho(x), n)$ au sens de Bakry–Émery, avec $\rho(x) \geq 0$ pour tout $x \in X$. Alors le trou spectral λ_1 de L sur $L^2(\pi)$ satisfait l'inégalité :*

$$\lambda_1 \geq \frac{nc}{n-1} + \frac{1}{\int_X \frac{\mathrm{d}\pi(x)}{\rho(x)-c}},$$

pour tout $0 \leq c \leq \inf_X \rho$.

On retrouve le théorème 0.23 en prenant $c = 0$. En prenant $c = \inf_X \rho$, on retrouve la borne de Bakry–Émery, avec un terme supplémentaire toujours positif.

Une généralisation du théorème 0.23 a été obtenue dans le cas où la métrique de la variété n'est plus adaptée à la diffusion. Elle nécessite une hypothèse assez restrictive, qui permet de définir une autre courbure, notée $\tilde{\kappa}$, en utilisant un infimum sur un sous-ensemble de couplages au lieu de tous les couplages possibles (dans la définition de la distance W_1). La courbure $\tilde{\kappa}$ est alors toujours inférieure à la courbure de Ricci grossière. On montre alors que trou spectral du générateur est supérieur à la moyenne harmonique de la courbure $\tilde{\kappa}$, dans le cas d'une diffusion réversible et si $\tilde{\kappa} \geq 0$.

0.7.3 Courbure de Ricci grossière et inégalités de concentration

L'utilisation de la courbure de Ricci grossière permet d'obtenir des inégalités de concentration pour la mesure d'équilibre de chaînes de Markov. Ces résultats, développés au chapitre 2, ne sont valables qu'en temps discret, mais ils ne requièrent plus l'hypothèse de réversibilité, qui était nécessaire pour les résultats sur le trou spectral. L'espace (X, d) doit être ε-géodésique (ce qui est le cas des graphes, par exemple).

La concentration obtenue pour une mesure d'équilibre est l'exponentielle d'une intégrale seconde de la courbure, ce qui est comparable à la densité de la mesure réversible d'une diffusion sur $\mathbb{R}$. En effet, si une diffusion sur $\mathbb{R}$ a pour générateur $L(f) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2} - b\frac{\mathrm{d}f}{\mathrm{d}x}$, la courbure est $\frac{\mathrm{d}b}{\mathrm{d}x}$, tandis que la densité de la mesure réversible est proportionnelle à $\mathrm{e}^{-\int b}$.

On a le théorème suivant :

Théorème 0.25 *Soit P le noyau de transition d'une chaîne de Markov à courbure de Ricci grossière positive sur un espace des états ε-géodésique.*

Supposons qu'il existe un point attractif x_0, dans le sens où

$$\mathbb{E}_{P_x} d(x_0, y) \leq d(x_0, x) - \rho,$$

avec $\rho > 0$ et pour tout x tel que $\varepsilon < d(x, x_0) \leq 2\varepsilon$.

Supposons également que les sauts de la chaîne de Markov ne sont pas trop étalés, dans le sens qu'il existe $s > 0$ tel que pour tout $x \in X$ et toute fonction 1-lipschitzienne sur X, on ait

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

On note K la fonction sur $\mathbb{R}_+$ définie grâce à la courbure de Ricci grossière :

$$K(r) := \inf_{d(x,x_0) \leq r, 0 < d(x,y) \leq \varepsilon} \kappa(x, y).$$

On a alors une inégalité de concentration de la forme :

$$\mathbb{P}_\pi[d(x, x_0) \geq l] \leq C\mathrm{e}^{\alpha\left(-\rho l - \int_{u=2\varepsilon}^{l} \int_{r=2\varepsilon}^{l} K(r)\mathrm{d}r\mathrm{d}u\right)},$$

pour toute mesure de probabilité invariante π, et tout $l \geq d_0$ avec C, α et d_0 des constantes explicites dépendant de s, ρ et ε.

Il semble que ce soit le premier résultat de concentration pour des processus discrets qui utilise une courbure non-constante. L'hypothèse d'existence d'un point attractif est indispensable, sinon il se peut qu'il n'y ait pas de mesure de probabilité invariante, comme dans le cas de la marche aléatoire sur $\mathbb{Z}$, qui a une courbure nulle.

0.8 Problèmes ouverts

Les nouveaux résultats de cette thèse semblent pouvoir être améliorés.

Le théorème de concentration ne fonctionne qu'en temps discret. Il est toutefois possible en bricolant un peu de montrer la concentration poissonnienne de la mesure d'équilibre du processus $M/M/\infty$. Cela nécessite de ralentir le processus loin du point attractif, et l'accélérer plus près pour s'assurer que le taux de saut soit égal à 1. On utilise ensuite une chaîne de Markov avec la même mesure d'équilibre que le processus ralenti, à laquelle on applique une version plus raffinée du théorème de concentration, où on autorise s^2 à dépendre de la distance au point attractif. Enfin, il faut corriger l'estimation obtenue pour le processus avec vitesse modifiée pour obtenir la concentration du processus $M/M/\infty$ de départ. Ce genre de méthode ne peut fonctionner qu'avec des processus de saut, et on n'est pas assuré que la courbure du processus avec vitesse modifiée ait le comportement souhaité, quand bien même ce serait le cas pour le processus original. Une version en temps continu du théorème de concentration est donc souhaitable.

Les estimations du trou spectral par la moyenne harmonique ne fonctionnent que pour des diffusions sur les variétés. En effet, pour les démontrer, on a besoin d'un couplage tel que la distance décroisse comme $\mathrm{e}^{-\int_0^t \kappa(X_t)}$ presque sûrement. On peut espérer obtenir une borne un peu moins bonne en contrôlant la variation de $d(X_{\delta t}, Y_{\delta t})$ autour de sa moyenne $\sim \mathrm{e}^{-\delta t K(X_0,Y_0)}$ lorsque δt est petit. On pourrait ainsi se passer de l'hypothèse $\nabla_u A(u,u)$ dans le cas des diffusions sur les variétés. Dans le cas d'espaces plus généraux, il devient nécessaire d'estimer $K(x,y)$ avec x et y éloignés, de façon plus subtile qu'en prenant l'infimum de la courbure, ce qui peut se faire en contrôlant la norme Lipschitz (ou Hölder) de la courbure.

Les résultats obtenus nécessitent presque tous l'hypothèse que la courbure est partout positive, alors qu'une faible quantité de courbure négative ne devrait pas beaucoup changer le comportement du processus markovien.

D'autres questions qui n'ont pas été abordées dans cette thèse peuvent aussi se poser concernant la courbure de Ricci grossière. Par exemple l'existence d'un analogue du théorème de Bishop–Gromov, qui donne une borne supérieure du volume des boules ou de celui des sphères dans les espaces dont la courbure est minorée par K.

On peut aussi se demander s'il y a une implication , ou une équivalence, entre avoir une courbure de Ricci grossière minorée par K et avoir une courbure supérieure à K au sens de Lott–Sturm–Villani.

Première partie

Courbure de Ricci grossière dans le cas d'un espace quelconque

Chapter 1

Coarse Ricci curvature for continuous-time Markov processes

In this chapter, 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.

Similar results are stated in an independant work [12].

Introduction

In [24], 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 1.6), and prove (Theorem 1.9) 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 [21]) that we recall below.

Theorem 1.1 (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 1.9, we can prove we have a spectral gap. What we can get using coarse Ricci curvature is the following:

Theorem 1.2 *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 1.6) 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 [10]. The difference is that the contraction hypothesis was global, and some assumption about the first eigenfunction was required. Here in Theorem 1.9, 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 [27] and in the chapter 3 of this thesis.

1.1 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 1.3 *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 [31]) 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 1.4 (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 [24], Ollivier defines the coarse Ricci curvature between two different points for discrete time Markov chains in the following way:

Definition 1.5 (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 1.6 *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 1.7 *For every $(x, y, z) \in E^3$, we have the inequality*

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

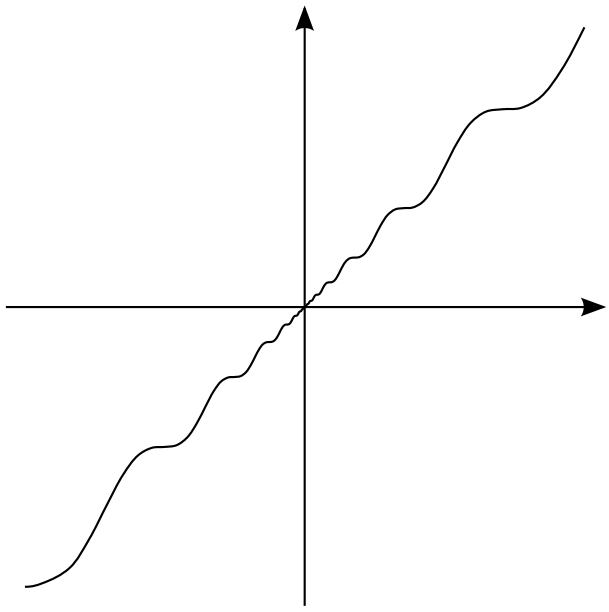
(this trivially comes from the triangular inequality for W_1). This property is not always satisfied by $\bar{\kappa}$ as we will see in the example below. This is the reason why we choose the liminf in the definition of Ricci curvature.

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 [24]), 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 1.8 *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 1.7, and the same is not true for* $\bar{\kappa}$*.*

1.2 W_1 contraction in positive coarse Ricci curvature

It is known ([8, 10, 13, 14, 24]) 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 1.9 *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 do in chapter 3 for diffusion processes on manifolds.

Remark 1.10 *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 1.11 *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 1.9 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 $\inf_{x,y} \bar{\kappa}(x,y) \leq \inf_{x,y} \kappa(x,y)$.□

Remark 1.12 *Usually in the literature, the processes are chosen right-continuous, but Theorem 1.9 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 1.9 does apply to diffusion processes and to minimal jump processes as defined in [9], 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 1.9 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 1.13 *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 1.7, we have $\inf(\kappa(x,y)) = \frac{1}{2}$, so we can use Theorem 1.9.

Corollary 1.14 *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* $W_1(\delta_{x_0}, P_{x_0}^t) < M$ *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 1.9 ,we can easily prove by induction that

$W_1\left(P_{x_0}^{nt_0}, P_{x_0}^{(n+1)t_0}\right) \leq M\mathrm{e}^{-Knt_0}$. So the sequence $(P_{x_0}^{nt_0})$ 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 $W_1\left(P_{x_0}^{\left\lfloor \frac{t}{t_0}\right\rfloor t_0}, P_{x_0}^t\right) \leq M\mathrm{e}^{-K\left\lfloor \frac{t}{t_0}\right\rfloor t_0}$, and the right hand term tends to 0 when t tends to infinity. Thus the family $(P_{x_0}^t)$ also tends to π.

Now we have for every $T > 0$,

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

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 $W_1(P_x^t, P_{x_0}^t) \leq \mathrm{e}^{-Kt} d(x, x_0)$, (P_x^t) 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 1.9 implies Theorem 1.2 as follows.

Proof of Theorem 1.2: Let π be the unique reversible probability measure. Theorem 1.9 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 1.2, the L^2 norm is controlled by the Lipschitz norm because $\mathrm{Var}_\pi(f) \leq \mathbb{E}_\pi[(f(x) - f(x_0))^2] \leq \|f\|_{\mathrm{Lipsch}}^2 \mathbb{E}_\pi[d(x, x_0)^2]$ (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 $\|Sx\|^2 = \langle Sx, Sx\rangle = \langle x, S^2x\rangle \leq \|x\|\|S^2x\|$ by Cauchy-Schwarz. 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 $L_0^2(\pi)$, so we use this result with $S = P^t$ and $x = f \in \mathrm{Lipsch}_0(\pi) \subset L_0^2(\pi)$: 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 [25]), 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 $L_0^2(\pi)$. The operator P^t is 1-Lipschitz on $L^2(\pi)$, so it is continuous on $L_0^2(\pi)$, and then $\|P^t f\|_{L^2} \leq \mathrm{e}^{-Kt}\|f\|_{L^2}$ for any $f \in L_0^2(\pi)$.□

Example 1.15 *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 of Theorem 3.13 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 Theorem 3.13, 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 1.2 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 1.9, we need a generalization of stopping times, which we call weak stopping times:

Definition 1.16 *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 1.17 *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 1.18 *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 1.9: Let $t > 0$ and $\varepsilon > 0$. We will show that for every x and y,

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

We denote by $M^{(x)}$ the Markov process starting at point $x \in E$, and $M^{(x)}_t$ 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 $(X, S) \sim (M^{(x)}_T, T + s)$ and $(Y, S) \sim (M^{(y)}_{T'}, T' + s)$.

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 $M^{(x_0)}_t$ and $M^{(y_0)}_t$ 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 1.19 *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 1.17 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 $(X'_0, Y'_0, S'_0, \dots, X'_n, Y'_n, S'_n)$ 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 $u_0, \dots, u_n, v_0, \dots, v_n$ 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 $(X'_0, Y'_0, S'_0, \dots, X'_{n+1}, Y'_{n+1}, S'_{n+1})$ 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 $(X'_0, Y'_0, S'_0, \dots, X'_{n+1}, Y'_{n+1}, S'_{n+1})$.

So the conditional law of $(X'_{i+1}, Y'_{i+1}, S'_{i+1})$ knowing $(X'_0, Y'_0, S'_0, \dots, X'_i, Y'_i, S'_i)$ 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 1.9:

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 1.19 $\mu_\infty = \lim_{i\to\infty} \mu_i$ 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 [31]), 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 1.9.$\square$

Chapter 2

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

In this chapter, 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 [24]. 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 2.2), a nonnegative coarse Ricci curvature allows to extend the local attractiveness of a point x_0 to a global one (see [24] or Lemma 2.7). 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 2.1 *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.

2.1 The concentration Theorems

We define ε-geodesic spaces as in [24].

Definition 2.2 *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 [24].

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

Theorem 2.3 *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 2.4 *If $K = 0$, we obtain exponential concentration, as proved in [24].*

Proposition 2.5 *If closed balls are compact, then under the hypotheses of Theorem 2.3, there exists an equilibrium measure.*

Remark 2.6 *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 2.7 *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 2.8 *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

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

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 $\lambda = f(x) - \mathbb{E}_\mu[f]$, $M = g'\left(\mathbb{E}_\mu[f]\right)$ 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:

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

as needed.□

Theorem 2.9 *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:

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

Let F be defined as in Lemma 2.7. Then, for every pair $(\alpha, d_0) \in \mathbb{R}_+^2$ satisfying:

- $d_0 \geq 2\varepsilon$
- $F(d_0) > \frac{s^2 K(d_0)}{2}$

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

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 $\psi(x) = \alpha(\varphi(d_0) - (d_0 - x)F(d_0))$ 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 2.8 with $\mu = P_x$ and $g = \psi$, and Lemma 2.7, we get:

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

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 $d(x,x_0) \geq d_0$, let us denote $\delta = d(x,x_0) - F(d(x,x_0))$. Then we have

$$\psi(\delta) \leq \psi(d(x,x_0)) - F(d)\psi'(\delta) \leq \psi(d(x,x_0)) - \frac{1}{\alpha}\psi'(\delta).$$

The hypothesis $C_{\alpha,d_0} < 1$ implies $\frac{1}{\alpha} - \frac{s^2}{2(1-\alpha s^2 K(d_0))} \geq 0$. The minimum of ψ'^2 is $\alpha^2 F(d_0)^2$, 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 2.10 *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 2.11 *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 2.3:

Proof of Theorem 2.3: we use Theorem 2.9 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

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

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

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

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 2.5: We take α and d_0 as in the proof of Theorem 2.3. We consider the sequence of probability measures $P^n_{x_0}$. Then, doing as in the proof of Theorem 2.9, 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 $\pi_n = \frac{1}{n+1}\sum_{i=0}^{n} P_{x_0}^i$. 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 [31]). 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 ([24]), 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.□

2.2 Some examples

Let us see which concentration we can get with Theorem 2.3 and Theorem 2.9 in some examples below.

Example 2.12 (Discrete time M/M/k queue (see, for example [16])) *Let $0 < n_0 < k$ be two integers. We consider here the Markov chain on integers with transition kernel:*

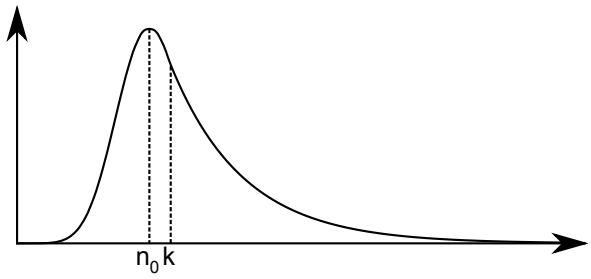
$$\begin{aligned} p(n, n+1) &= \frac{n_0}{n_0 + k} \\ p(n, n) &= \frac{(k-n)_+}{n_0 + k} \\ p(n, n-1) &= \frac{\min(n, k)}{n_0 + k} \\ p(n, m) &= 0 \quad \textit{if } |n - m| > 1. \end{aligned}$$

The origin x_0 we will consider to apply Theorem 2.3 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 [18]) states that for a random variable X such that $a \leq X \leq b$ almost surely, we have $\mathbb{E}\left[\mathrm{e}^{\lambda}(X - \mathbb{E}[X])\right] \leq \mathrm{e}^{\frac{\lambda^2(b-a)^2}{8}}$. So we can take $s = 1$ in theorem 2.3. 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 2.3 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 $\rho = \frac{\min(\varepsilon, k-n_0)}{n_0+k}$,

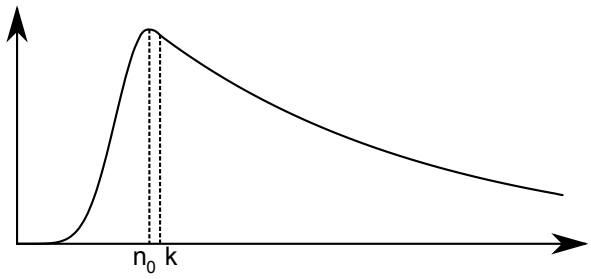
but we pay that by a worse curvature $K(r) = \frac{1}{n_0+k}\min(1, \max(0, \frac{k-n_0-r}{\varepsilon}))$. We distinguish 3 cases depending on how $k - n_0$ is tall with respect to n_0.

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.



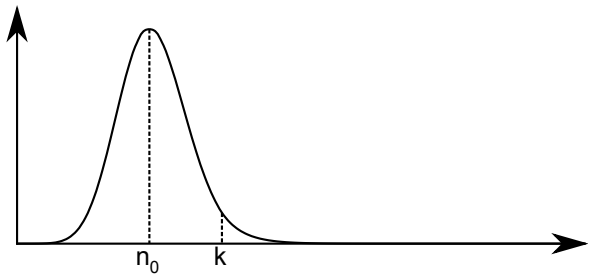
The optimal ε is $O(\sqrt{n_0})$, the coefficient of the Gaussian part of the concentration inequality is $O(\frac{1}{n_0})$, which is good, and the coefficient of the exponential part is $O(\frac{k-n_0}{n_0})$, like the right one.

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 2.3 gives a rather bad concentration inequality.

Example 2.13 (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 [7] 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 2.3 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 2.3 or Theorem 2.9 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 2.14 *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 2.3 or Theorem 2.9 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 $T_1, T_2, \ldots$ 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 $\alpha > 0$, according to Corollary 1.14, there exists an unique invariant probability measure π.

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.

Part II

Le cas particulier des diffusions sur les variétés riemanniennes

Chapter 3

Improved spectral gap bounds on positively curved manifolds

A coupling method and an analytic one allow us to prove new lower bounds for the spectral gap of reversible diffusions on compact manifolds. Those bounds are based on the a notion of curvature of the diffusion, like the coarse Ricci curvature or the Bakry–Émery curvature-dimension inequalities. We show that when this curvature is nonnegative, its harmonic mean is a lower bound for the spectral gap.

Introduction

The study of the spectrum of the Laplace Operator on Riemannian manifolds has many applications in various domains of mathematics. A whole chapter of [6] is devoted to this issue. In this article, we take the convention

$$\Delta = g^{ij}\nabla_i\nabla_j$$

for the Laplace operator. The spectral gap of Δ is the opposite of the greatest non-zero eigenvalue of Δ (the spectrum of Δ is discrete and non-positive).

One way to estimate this spectral gap is to use the Ricci curvature, as we see it in the Lichnerowicz theorem (see [21]).

Theorem 3.1 (Lichnerowicz) *Let $\mathcal{M}$ 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 Δ satisfies*

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

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

Chen and Wang improved this result in [11], using the diameter of the manifold in their estimates:

Theorem 3.2 *Let $\mathcal{M}$ be a compact connected n-dimensional Riemannian manifold, K be the infimum of the Ricci curvature on $\mathcal{M}$ and D be the diameter of $\mathcal{M}$. Then if $K \geq 0$, we have the following bounds:*

$$\lambda_1 \geq \frac{\pi^2}{D^2} + \max\left(\frac{\pi}{4n}, 1 - \frac{2}{\pi}\right)$$

and if $n > 1$,

$$\lambda_1 \geq \frac{nK}{(n-1)\left(1-\cos^n\left(\frac{D\sqrt{K(n-1)}}{2}\right)\right)}.$$

And if $K \leq 0$, *we have the following bounds:*

$$\lambda_1 \geq \frac{\pi^2}{D^2} + \left(\frac{\pi}{2} - 1\right) K$$

and if $n > 1$,

$$\lambda_1 \geq \frac{\pi^2\sqrt{1-\frac{2D^2K}{\pi^4}}}{D^2 \operatorname{ch}\left(\frac{D\sqrt{-K(n-1)}}{2}\right)}$$

In [3], E.Aubry gives a lower bound for λ_1 when the curvature

$$\underline{\mathrm{Ric}}(x) := \inf_{u \in \mathrm{T}_x\mathcal{M}} \frac{\mathrm{Ric}_x(u,u)}{\|u\|^2}$$

is close to a positive constant in the sense of L^p norm with p large enough:

Theorem 3.3 *Let* $\mathcal{M}$ *be a complete* n*-dimensional Riemannian Manifold,* $p > \frac{n}{2}$ *and* $K > 0$, *such that*

$$\int_{\mathcal{M}} (\underline{\mathrm{Ric}} - K)_-^p < +\infty.$$

Then $\mathcal{M}$ *has a finite volume and the spectral gap of* Δ *on* $\mathcal{M}$ *satisfies:*

$$\lambda_1 \geq \frac{n}{n-1} K \left(1 - \frac{C(p,n)}{K} \|(\underline{\mathrm{Ric}} - K)_-\|_p\right)$$

where $C(p,n)$ *is a constant only depending on* p *and* n, *and* $\|f\|_p = \left(\frac{\int_{\mathcal{M}} |f|^p}{\mathrm{vol}(\mathcal{M})}\right)^{\frac{1}{p}}$.

This allows a little negative curvature, which is not the case of our results.

This article recapitulates and extends the results already stated in [27] and presents a coupling method, more adapted to discrete spaces than the analytic one.

We show by a coupling method that another bound for λ_1 is the harmonic mean of the Ricci curvature.

Theorem 3.4 *Let* $\mathcal{M}$ *be a compact Riemannian manifold with positive Ricci curvature. Then we have*

$$\frac{1}{\lambda_1} \leq \int_{\mathcal{M}} \frac{\mathrm{d}\mu(x)}{\underline{\mathrm{Ric}}(x)},$$

with $\mathrm{d}\mu = \frac{\mathrm{d}\,\mathrm{vol}}{\mathrm{vol}(\mathcal{M})}$, *where* vol *is the Riemannian volume measure on* $\mathcal{M}$.

This bound is often better than the Lichnerowicz one because the harmonic mean is better (and can be much better) than the infimum. But unfortunately we lose the $\frac{n}{n-1}$ factor.

Merging the proof of Theorem 3.1 and an analytic proof of Theorem 3.4 gives us the following improvement:

Theorem 3.5 *Let $\mathcal{M}$ be a Riemannian manifold with positive Ricci curvature and $K = \inf_{x\in\mathcal{M}} \underline{\mathrm{Ric}}(x)$. Then for every $0 \le c \le K$, we have:*

$$\lambda_1 \ge \frac{n}{n-1}c + \frac{1}{\int_{\mathcal{M}} \frac{\mathrm{d\,vol}}{\underline{\mathrm{Ric}}(x)-c}}.$$

Taking $c = K$ gives us the Lichnerowicz bound or even better, while $c = 0$ gives us Theorem 3.4.

Our coupling approach is based on a notion of coarse Ricci curvature, introduced by Yann Ollivier in [24], which uses the Wasserstein distance W_1. A major step in our proof is the use of the coupling given by the following theorem:

Theorem 3.6 *Let $\mathcal{M}$ be a smooth Riemannian manifold, and F^i be a smooth vector field on $\mathcal{M}$. Assume that there exists a diffusion process associated with the generator $Lf = \Delta f + F^i\nabla_i f$. Let $\kappa(x, y)$ be the coarse Ricci curvature of the diffusion between x and y (see Definition 3.10). Then for any two distinct points x and y of $\mathcal{M}$, there exists a coupling $(x(t), y(t))$ between the paths of the diffusion process starting at x and y which satisfies:*

$$d(x(t), y(t)) = d(x, y)\mathrm{e}^{-\int_0^t \kappa(x(s),y(s))\mathrm{d}s}$$

on the event that for any $s \in [0, t]$, $(x(s), y(s))$ does not belong to the cut-locus of $\mathcal{M}$.

The contraction rate $\kappa(x, y)$ of this coupling behaves like the one of the coupling derived from the diffusion in $\mathcal{C}^1$ path space defined by M.Arnaudon, K.A.Coulibaly and A.Thalmaier in [2] when x and y are close. We have a cut-locus problem that we will avoid by making a compactness assumption, which was anyway necessary to replace $\kappa(x, y)$ by its limit when x and y are infinitely close.

The coupling method and the analytic one keep working when we add a drift to the Brownian motion, provided the diffusion is reversible. In this case, the generator takes the following form:

$$L = \frac{1}{2}g^{ij}(\nabla_i\nabla_j - (\nabla_j\varphi)\nabla_i)$$

with φ a smooth function on $\mathcal{M}$, and $\mathrm{e}^{-\varphi}\mathrm{d\,vol}$ is a reversible measure. We have then the following generalization of Theorem 3.5:

Theorem 3.7 *Let $\mathcal{M}$ be a compact Riemannian manifold and $L = \frac{1}{2}g^{ij}(\nabla_i\nabla_j - (\nabla_j\varphi)\nabla_i)$ be the operator associated with a reversible diffusion process on $\mathcal{M}$.*

Suppose that we have a curvature-dimension inequality in the sense of Bakry-Émery (see [4] or [5]) with a positive curvature ρ and a constant and positive dimension n',which is

$$\Gamma_2(f)(x) \geq \rho(x)\Gamma(f)(x) + \frac{1}{n'}L(f)(x)^2.$$

Let R be the infimum of ρ. Then for every $0 \leq c < R$, we have

$$\lambda_1(L) \geq \frac{n'}{n'-1}c + \frac{1}{\int_{\mathcal{M}} \frac{\mathrm{d}\pi(x)}{\rho(x)-c}}$$

with $\mathrm{d}\pi = \frac{\mathrm{e}^{-\varphi}\mathrm{d\,vol}}{\int_{\mathcal{M}} \mathrm{e}^{-\varphi}\mathrm{d\,vol}}$ the reversible probability measure.

We try to generalize our coupling method to diffusions which are not adaptated to the metric g, that is, whose generator takes the more general form:

$$L = \frac{1}{2}A^{ij}\nabla_i\nabla_j + F^i\nabla_i$$

without having necessarily $A^{ij} = g^{ij}$ anymore. We have a generalization of Theorem 3.6 only on the very restrictive condition:

$$(H) \Leftrightarrow \forall u \in \mathrm{T}\mathcal{M}, u^i g_{jk} u^j g_{lm} u^l \nabla_i A^{km} = 0 \Leftrightarrow g^{il}\nabla_l A^{jk} + g^{jl}\nabla_l A^{ki} + g^{kl}\nabla_l A^{ij} = 0$$

and with a lower $\tilde{\kappa}(x,y)$ instead of $\kappa(x,y)$. Note that (H) is true for $A^{ij} = g^{ij}$, in which case we have $\tilde{\kappa} = \kappa$.

We have the following generalization of Theorem 3.4:

Theorem 3.8 *Consider a diffusion process on a compact Riemannian manifold $\mathcal{M}$ which is reversible and satisfies (H). For every x in $\mathcal{M}$, we set $\tilde{\kappa}(x) = \inf_{u\in \mathrm{T}_x\mathcal{M}} \tilde{\kappa}(x,u)$. If we have $\tilde{\kappa}(x) \geq \varepsilon > 0$, then the spectral gap of L is at least the harmonic mean of $\tilde{\kappa}$ (with respect to the reversible probability measure π):*

$$\frac{1}{\lambda_1(L)} \leq \int_{\mathcal{M}} \frac{\mathrm{d}\pi(x)}{\tilde{\kappa}(x)}.$$

In section 3.1, we present a short argument which shows how we can derive the harmonic mean from Theorem 3.6. In section 3.2, we define the coarse Ricci curvature for diffusions and construct our couplings, so it's where Theorem 3.6 is proved. In section 3.3, we present the proofs using the couplings and purely analytical ones for the harmonic mean bounds for the spectral gap.

3.1 The harmonic mean in a nutshell

The result and its proof presented in this section are a shortcut found by Y.Ollivier to obtain a harmonic mean from Theorem 3.6.

Using a classical method, we will prove thanks to Theorem 3.6 the following result, which is a weaker version of Theorem 3.4:

Theorem 3.9 *Let $\mathcal{M}$ be a compact Riemannian manifold with positive Ricci curvature, and f be any 1-Lipschitz function on $\mathcal{M}$. Then the variance of f is at most the average of $\frac{1}{\underline{\mathrm{Ric}}}$.*

Indeed, the Poincaré inequality states that $\mathrm{Var}_\mu(f) \leq \frac{1}{\lambda_1}\int \|\nabla f\|^2 \mathrm{d}\mu$, and the integral on the right hand side is at most 1 for 1-Lipschitz functions. In [23], E.Milman shows that the converse is true i.e a control on the variance of Lipschitz functions (and even on the L^1 norm of 0-mean Lipschitz functions) implies a Poincaré inequality, with a universal loss in the constants, under the hypothesis of a Bakry–Émery $CD(0,\infty)$ curvature-dimension inequality.
Proof of theorem 3.9: We only have to prove the result for f regular enough, and use a density argument to get the result for non-regular f. We consider the semi-group P^t generated by the Laplacian operator. Then the limit of P^t when t tends to infinity is the operator which associates to f the constant function equals to the mean of f (respect to the normalized Riemannian volume measure). So the variance of f is the limit of the mean of $P^t(f^2) - (P^t(f))^2$ when t tends to infinity. We have

$$P^t(f^2)-(P^t(f))^2 = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\left(P^s((P^{t-s}(f))^2)\right)\mathrm{d}s = \int_0^t P^s(2\|\nabla(P^{t-s}(f))\|^2)\,\mathrm{d}s$$

Integrating over $\mathcal{M}$ yields

$$\int_{\mathcal{M}} (P^t(f^2)(x)-(P^t(f)(x)))^2\,\mathrm{d}\,\mathrm{vol}(x) = \int_{\mathcal{M}}\int_0^t 2\|\nabla(P^{t-s}(f))(x)\|^2\,\mathrm{d}s\,\mathrm{d}\,\mathrm{vol}(x).$$

Thanks to Theorem 3.6, by taking y very close to x, we have $\|\nabla(P^{t-s}(f))(x)\| \leq \mathbb{E}_{\mathbb{P}_x}[\mathrm{e}^{-\int_0^{t-s}\underline{\mathrm{Ric}}(X_u)\mathrm{d}u}\|\nabla f(X_{t-s})\|]$, where the right hand side is the expectation of the term inside the brackets when X has the law $\mathbb{P}_x$ of the twice accelerated Brownian motion on $\mathcal{M}$ starting at x. Using the convexity of the exponential function, and the fact that f is 1-Lipschitz, we get then

$$\begin{aligned}
\int_{\mathcal{M}}(P^t(f^2)(x) - (P^t(f)(x)))\mathrm{d}\,\mathrm{vol}(x) &\leq 2\int_{\mathcal{M}}\int_0^t \left(\mathbb{E}_{\mathbb{P}_x}\left[\mathrm{e}^{-\int_0^{t-s}\underline{\mathrm{Ric}}(X_u)\mathrm{d}u}\right]\right)^2 \mathrm{d}s\,\mathrm{d}\,\mathrm{vol}(x)\\
&\leq 2\int_{\mathcal{M}}\int_0^t \mathbb{E}_{\mathbb{P}_x}\left[\mathrm{e}^{-2\int_0^{t-s}\underline{\mathrm{Ric}}(X_u)\mathrm{d}u}\right]\mathrm{d}s\,\mathrm{d}\,\mathrm{vol}(x)\\
&\leq 2\int_0^t\int_{\mathcal{M}} \mathbb{E}_{\mathbb{P}_x}\left[\int_0^1 \mathrm{e}^{-2(t-s)\underline{\mathrm{Ric}}(X_{(t-s)u})}\mathrm{d}u\right]\mathrm{d}\,\mathrm{vol}(x)\,\mathrm{d}s\\
&= 2\int_0^t\int_{\mathcal{M}} \mathrm{e}^{-2(t-s)\underline{\mathrm{Ric}}(x)}\mathrm{d}\,\mathrm{vol}(x)\,\mathrm{d}s\\
&= \int_{\mathcal{M}} \frac{1-\mathrm{e}^{-2t\underline{\mathrm{Ric}}(x)}}{\underline{\mathrm{Ric}}(x)}\mathrm{d}\,\mathrm{vol}(x).
\end{aligned}$$

We just have to take the limit when t tends to infinity and to divide by $\int_{\mathcal{M}}\mathrm{d}\,\mathrm{vol}(x)$ to get the theorem.□

3.2 Coarse Ricci curvature for diffusions on Riemannian manifolds

In this section, we introduce the Coarse Ricci curvature κ for general diffusions and give an explicit formula. Then we construct the coupling of Theorem 3.6, we show why the (H) condition is needed and we define $\tilde{\kappa}$ when it is satisfied.

3.2.1 Coarse Ricci curvature: definition and calculation

Following what is done in [24] for Markov chains, we define the coarse Ricci curvature of diffusions as the rate of decay of the Wasserstein distance W_1 between the measures associated with the diffusion and starting at two different points:

Definition 3.10 *Let $\mathcal{M}$ be a Riemannian manifold and P^t be the semi-group of a diffusion on $\mathcal{M}$. The coarse Ricci curvature between two different points x and y is the following quantity:*

$$\kappa(x,y) = \varliminf_{t\to 0} \frac{d(x,y) - W_1(\delta_x.P^t, \delta_y.P^t)}{td(x,y)}.$$

The Wasserstein distance W_1 between two measures is the infimum on all the couplings of the expectation of the distance. Our coupling will be consrtucted thanks to optimal ones.

To get an expression of this curvature only depending on the coefficients of the generator of the diffusion, we need to make sure that the diffusion does not move far away too fast.

Definition 3.11 *A diffusion on $\mathcal{M}$ is said to be locally uniformly L^1-bounded at x if $\exists M > 0, \exists \eta > 0, \forall y \in \mathcal{M} | d(x,y) < \eta, \forall 0 < t < \eta, \int d(x,z) \mathrm{d}(\delta_y.P^t)(z) < M$.*

Remark 3.12 *If $\mathcal{M}$ is compact, any diffusion is locally uniformly L^1-bounded at each point (it suffices to take M equals to the diameter of $\mathcal{M}$ in the previous definition).*

The following theorem gives an expression of $\kappa(x,y)$. Recall that the generator of the diffusion is

$$L(f) = \frac{1}{2} A^{ij} \nabla_i \nabla_j f + F^i \nabla_i f$$

with A symmetric and non-negative.

Theorem 3.13 *Take two distinct points x and y in $\mathcal{M}$, such that A and F are continuous at x and y, and that the diffusion is locally uniformly L^1-bounded at x and y. Assume that the distance between two points in the neighborhoods of x and y admits the following second-order Taylor expansion:*

$$\begin{aligned} d(\exp_x(\varepsilon v), \exp_y(\varepsilon w)) = d(x,y) \Bigg[1 + \varepsilon \left(l_i^{(1)} v^i + l_j^{(2)} w^j \right) \\ + \frac{\varepsilon^2}{2} \left(q_{i_1 i_2}^{(1)} v^{i_1} v^{i_2} + q_{j_1 j_2}^{(2)} w^{j_1} w^{j_2} + 2 q_{ij}^{(12)} v^i w^j \right) + o\left(\frac{\varepsilon^2}{|\ln(\varepsilon)|} \right) \Bigg]. \end{aligned}$$

Then the coarse Ricci curvature between x and y is:

$$\begin{aligned} \kappa(x,y) = -l_i^{(1)} F^i(x) - l_j^{(2)} F^j(y) - \frac{q_{i_1 i_2}^{(1)} A^{i_1 i_2}(x) + q_{j_1 j_2}^{(2)} A^{j_1 j_2}(y)}{2} \\ + \operatorname{tr}\left(\sqrt{A^{i_1 i_2}(x) q_{i_2 j_1}^{(12)} A^{j_1 j_2}(y) q_{i_3 j_2}^{(12)}} \right). \end{aligned}$$

Here the matrix $S^{i_1}{}_{i_3} = A^{i_1 i_2}(x) q^{(12)}_{i_2 j_1} A^{j_1 j_2}(y) q^{(12)}_{i_3 j_2}$ is diagonalizable with non-negative eigenvalues, since it is the product of two symmetric non-negative matrices, so $S^{i_1}{}_{i_2}$ admits an unique diagonalizable square root $R^{i_1}{}_{i_2}$ with non-negative eigenvalues, and the last term of the formula is simply $R^i{}_i$.

Remark 3.14 *We don't assume here that d is the usual geodesic distance on the manifold $\mathcal{M}$, but only that it admits a nice second order Taylor expansion. For example, we can take d the Euclidean distance on the sphere S^n embedded in $\mathbb{R}^{n+1}$.*

Proof: The idea is to approximate the distributions P^t_x and P^t_y for small t by Gaussian distributions in the tangent spaces $\mathrm{T}_x\mathcal{M}$ and $\mathrm{T}_y\mathcal{M}$, and to approximate the distance by its second order Taylor expansion. We can describe the process $x(t)$ starting at x in the exponential map by the equation:

$$\mathrm{d}X^i(t) = B^i{}_\alpha(X(t))\mathrm{d}W^\alpha(t) + F'^i(X(t))\mathrm{d}t$$

where $W(t)$ is a Brownian motion in $\mathbb{R}^n$, $B^{i_1}{}_{\alpha_1}(0)\delta^{\alpha_1\alpha_2}B^{i_2}{}_{\alpha_2}(0) = A^{i_1 i_2}(x)$ and $F'^i(0) = F^i(x)$, and B and F' are continuous (because of the continuity of A and F) and defined in a neighborhood of 0. Keep in mind that $X(t)$ may not be defined for every $t > 0$, but we have $x(t) = \exp_x(X(t))$ when it is. We will approximate $X^i(t)$ by

$$X^{(0)i}(t) = B^i{}_\alpha(0)W^\alpha(t) + tF^i(x),$$

which has the Gaussian law $\mathcal{N}(tF(x), tA(x))$. For small t, the ball K_t of radius $\sqrt{(2A^{i_1 i_2}(x)g_{i_1 i_2}(x) + 2)t|\ln(t)|}$ of $\mathrm{T}_x\mathcal{M}$ is included in the definition domain of B and F'. We will show that $X(s)$ remains in K_t for $0 \le s \le t$ with probability $1 - o(t)$. Let T_t be the exit time of $X(s)$ from K_t, and

$$X_t(s) = \begin{cases} X(s) & \text{if } s \le T_t \\ X^{(0)}(s) - X^{(0)}(T_t) + X(T_t) & \text{if } s > T_t. \end{cases}$$

We want to prove that $\|X_t|_{[0,t]}\|_\infty = \sup_{s\in[0,t]} \|X_t(s)\| \le \sqrt{(2A^{i_1 i_2}(x)g_{i_1 i_2}(x) + 2)t|\ln(t)|}$ with probability $1 - o(t)$.

We first prove that $\|X_t - X^{(0)}|_{[0,t]}\|_\infty = o(\sqrt{t|\ln(t)|})$ with probability $1 - o(t)$. We have $\mathrm{d}(X_t - X^{(0)})(s) = \mathbf{1}_{s<T_t}[(B^i{}_\alpha(X(s)) - B^i{}_\alpha(0))\mathrm{d}W^\alpha(s) + (F'^i(X(s)) - F^i(x))\mathrm{d}s]$. Because of the continuity of B and F', we have $\mathrm{d}(X_t - X(0))(s) = o(1)\mathrm{d}W(s) + o(1)\mathrm{d}s$. We have $\|\int_0^s o(1)\mathrm{d}u|_{0,t]}\|_\infty = o(t) = o(\sqrt{t|\ln(t)|})$. For each coordinate of the martingale $U^i(s) = \int_0^s \mathbf{1}_{s<T_t}(B^i{}_\alpha(X_t(u)) - B^i{}_\alpha(0))\mathrm{d}W^\alpha(u)$, we will apply the Doob inequality to the sub-martingale $\mathrm{e}^{\lambda U^i}$. Indeed, we have $\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[\mathrm{e}^{\lambda U^i(s)}] \le \frac{c_1(t)}{2}\lambda^2\mathbb{E}[\mathrm{e}^{\lambda U^i(s)}]$, with $c_1(t) = o(1)$ depending on the infinite norm on K_t of $(B^{i_1}{}_{\alpha_1}(X) - B^{i_1}{}_{\alpha_1}(0))\delta^{\alpha_1\alpha_2}(B^{i_2}{}_{\alpha_2}(X) - B^{i_2}{}_{\alpha_2}(0))$ so $\mathbb{E}[\mathrm{e}^{\lambda U^i(t)}] \le \mathrm{e}^{\frac{\lambda^2 c_1(t)t}{2}}$. So the Doob inequality implies, by taking $c_2(t) = \sqrt{(3c_1(t))t|\ln(t)|} = o(\sqrt{t|\ln(t)|})$, and taking $\lambda = \pm\frac{c_2(t)}{c_1(t)}$, that $\mathbb{P}(\sup_{[0,t]}|U^i(s)| \ge c_2(t)) \le 2\mathrm{e}^{-\frac{c_2(t)^2}{2tc_1(t)}} = 2\mathrm{e}^{-\frac{3|\ln(t)|}{2}} = 2t^{\frac{3}{2}} = o(t)$. We deduce $\|U|_{[0,t]}\|_\infty \le o(\sqrt{t\ln(t)})$ with probability $1 - o(t)$, so we have the same conclusion for $\|X_t - X^{(0)}|_{[0,t]}\|_\infty$.

We have $\|sF^i(0)\|_\infty = O(t) = o(\sqrt{t|\ln(t)|})$, so it remains to prove that $\|B^i{}_\alpha(0)W^\alpha(s)|_{[0,t]}\|_\infty \leq \sqrt{(2A^{i_1i_2}(x)g_{i_1i_2}(x)+1)t|\ln(t)|}$ with probability $1-o(t)$. We can suppose we are working with an orthonormal basis of eigenvectors of $A^{i_1i_2}(x)g_{i_2i_3}(x)$. In this case, using the same method as above, we prove $\mathbb{P}(\sup_{[0,t]} B^i{}_\alpha(0)W^\alpha(s) \leq \sqrt{(2\lambda_i+\frac{1}{n})t|\ln(t)|})$ with probability $1-o(t)$, and then if the inequality is true for all i, we get the announced result by summing the squares.

Now we set

$$\bar{X}_t(s) = \begin{cases} X(s) & \text{if } s \leq T_t \\ 0 & \text{otherwise.} \end{cases}$$

We have $\mathbb{E}[d(x(t), \exp_x(\bar{X}_t(t)))] = o(t)$. Indeed, if $X(t)$ does not exit from K_t ($T_t \geq t$), then the distance is 0. If $T_t < t$ (what we have shown to occur with probability $o(t)$), we apply the Markov property, and using the local uniform L^1-boundedness assumption, the conditional expectation of $d(x(t), x)$ knowing $(X(T_t), T_t)$ is smaller than M for t small enough. So $\mathbb{E}[d(x(t), \exp_x(\bar{X}(t)))] \leq Mo(t) = o(t)$. So the Wasserstein distance between the distributions of $x(t)$ and $\exp_x(\bar{X}_t(t))$ is $o(t)$.

Of course, we can do the same for the process starting at y, and define $Y^{(0)}$, K'_t, Y_t and $\bar{Y}_t$.

We denote $\tilde{d}$ the second order Taylor expansion of the distance:

$$\tilde{d}(X,Y) = d(x,y)[1+l_i^{(1)}X^i+l_j^{(2)}Y^j+\frac{1}{2}(q_{i_1i_2}^{(1)}X^{i_1}X^{i_2}+q_{j_1j_2}^{(2)}Y^{j_1}Y^{j_2}+2q_{ij}^{(12)}X^iY^j)]$$

The supremum of $\tilde{d}(X,Y)-d(\exp_x(X), \exp_y(Y))$ over $K_t \times K'_t$ is $o(\frac{\sqrt{t|\ln(t)|}^2}{|\ln(\sqrt{t|\ln(t)|})|}) = o(t)$. So the Wasserstein distance between $\exp_x(\bar{X}_t(t))$ and $\exp_y(\bar{Y}_t(t))$ differs of $o(t)$ from the minimum over all couplings of $\mathbb{E}[\tilde{d}(\bar{X}_t(t), \bar{Y}_t(t))]$.

It remains to prove that we have a difference of $o(t)$ between the solutions of the minimization problems of $\mathbb{E}[\tilde{d}(\bar{X}_t(t), \bar{Y}_t(t))]$ and $\mathbb{E}[\tilde{d}(X^{(0)}(t), Y^{(0)}(t))]$. The expectation and the covariance of $\bar{X}_t(t) - X^{(0)}(t)$ and $\bar{Y}_t(t) - Y^{(0)}(t)$ are $o(t)$. Indeed, we have $\bar{X}_t(t) - X^{(0)}(t) = (\bar{X}_t(t) - X_t(t)) + (X_t(t) - X^{(0)}(t))$. The expectation and the covariance of $X_t(t) - X^{(0)}(t)$ are $o(t)$ because this quantity is $\int_0^t o(1)\mathrm{d}W(s) + o(1)\mathrm{d}s$ (the stochastic integral is a martingale, so its expectation at time t is 0). We have $\bar{X}_t(t) - X_t(t) = 0$ when $T_t \geq t$, which occurs with probability $1-o(t)$, and the conditional expectation and covariance of $\bar{X}_t(t) - X_t(t)$ knowing $T_t < t$ are $O(\sqrt{t|\ln(t)|})$ and $O(t|\ln(t)|)$, so the expectation and covariance of $\bar{X}_t(t) - X_t(t)$ are $o(t\sqrt{t|\ln(t)|})$ and $o(t^2|\ln(t)|)$ (so $o(t)$ anyway). So the expectation and the covariance of $\bar{X}_t(t) - X^{(0)}(t)$ are $o(t)$.

The expectation and the covariance of $X^{(0)}(t)$ and $Y^{(0)}(t)$ are $O(t)$, so for any coupling of $(\bar{X}_t(t), X^{(0)}(t))$ with $(\bar{Y}(T), Y^{(0)}(t))$, we have

$$\begin{aligned}
&\mathbb{E}[\tilde{d}(\bar{X}_t(t), \bar{Y}_t(t)) - \tilde{d}(X^{(0)}(t), Y^{(0)}(t))] = \\
&\quad d(x,y)\mathbb{E}\Big[l_i^{(1)}(\bar{X}_t^i(t) - X^{(0)i}(t)) + l_j^{(2)}(\bar{Y}_t^j(t) - Y^{(0)j}(t)) \\
&\quad + \frac{1}{2}q_{i_1i_2}^{(1)}(\bar{X}_t^{i_1}(t) - X^{(0)i_1}(t))(\bar{X}_t^{i_2}(t) + X^{(0)i_2}(t)) \\
&\quad + \frac{1}{2}q_{j_1j_2}^{(2)}(\bar{Y}_t^{j_1}(t) - Y^{(0)j_1}(t))(\bar{Y}_t^{j_2}(t) + Y^{(0)j_2}(t)) \\
&\quad + q_{ij}^{(12)}(\bar{X}_t^i(t) - X^{(0)i}(t))(\bar{Y}_t^j(t) + Y^{(0)j}(t)) \\
&\quad + q_{ij}^{(12)}(\bar{X}_t^i(t) + X^{(0)i}(t))(\bar{Y}_t^j(t) - Y^{(0)j}(t))\Big] = o(t).
\end{aligned}$$

The last four terms above are $o(t)$ because of the Cauchy Schwarz inequality, which implies that for every family of random vectors V_t and W_t whose covariance matrices satisfy $\mathrm{Cov}(V_t, V_t) = o(t)$ and $\mathrm{Cov}(W_t, W_t) = O(t)$, we have $\mathrm{Cov}(V_t, W_t) = o(t)$. So we have proved that $W_1(P_x^t, P_y^t) = \inf \mathbb{E}[\tilde{d}(X^{(0)}(t), Y^{(0)}(t))] + o(t)$. The laws of $X^{(0)}(t)$ and $Y^{(0)}(t)$ are $\mathcal{N}(tF(x), tA(x))$ and $\mathcal{N}(tF(y), tA(y))$, so we have

$$\begin{aligned}
\mathbb{E}[\tilde{d}(X^{(0)}(t), Y^{(0)}(t))] &= d(x,y)\Big[1 + t(l_i^{(1)}F^i(x) + l_j^{(2)}F^j(y)) \\
&\quad + \frac{t}{2}(q_{i_1i_2}^{(1)}A^{i_1i_2}(x) + q_{j_1j_2}^{(2)}A^{j_1j_2}(y)) \\
&\quad + \frac{t^2}{2}\left(q_{i_1i_2}^{(1)}F^{i_1}(x)F^{i_2}(x) + q_{j_1j_2}^{(2)}F^{j_1}(y)F^{j_2}(y) + 2q_{ij}^{(12)}F^i(x)F^j(y)\right) \\
&\quad + \mathbb{E}[q_{ij}^{(12)}(X^{(0)i}(t) - tF^i(x))(Y^{(0)j}(t) - tF^j(y))]\Big].
\end{aligned}$$

We only have to minimize the last term, and the minimum is

$$-t\,\mathrm{tr}\left(\sqrt{A^{i_1i_2}(x)q_{i_2j_1}^{(12)}A^{j_1j_2}(y)q_{i_3j_2}^{(12)}}\right)$$

according to Lemma 3.15 below.

So we have proved that

$$\begin{aligned}
W_1(P_x^t, P_y^t) = d(x,y)\Big[1 - t\Big(&-l_i^{(1)}F^i(x) - l_j^{(2)}F^j(y) - \frac{1}{2}(q_{i_1i_2}^{(1)}A^{i_1i_2}(x) + q_{j_1j_2}^{(2)}A^{j_1j_2}(y)) \\
&+ \mathrm{tr}\left(\sqrt{A^{i_1i_2}(x)q_{i_2j_1}^{(12)}A^{j_1j_2}(y)q_{i_3j_2}^{(12)}}\right)\Big) + o(t)\Big]
\end{aligned}$$

which precisely means that

$$\begin{aligned}
\kappa(x,y) = -l_i^{(1)}F^i(x) - l_j^{(2)}F^j(y) - \frac{1}{2}(q_{i_1i_2}^{(1)}A^{i_1i_2}(x) + q_{j_1j_2}^{(2)}A^{j_1j_2}(y)) \\
+ \mathrm{tr}\left(\sqrt{A^{i_1i_2}(x)q_{i_2j_1}^{(12)}A^{j_1j_2}(y)q_{i_3j_2}^{(12)}}\right). \square
\end{aligned}$$

Lemma 3.15 *Let $A^{i_1 i_2}$ and $B^{j_1 j_2}$ be two symmetric non-negative tensors belonging to $E_1 \otimes E_1$ and $E_2 \otimes E_2$, with E_1 and E_2 two finite dimensional $\mathbb{R}$-vector spaces, not necessarily of the same dimension. Let D_{ij} be a tensor belonging to $E_1^* \otimes E_2^*$. Then the minimum of $\mathbb{E}[D_{ij} X^i Y^j]$ over all couplings between X of law $\mathcal{N}(0, A)$ and Y of law $\mathcal{N}(0, B)$ is*

$$-\operatorname{tr}\left(\sqrt{A^{i_1 i_2} D_{i_2 j_1} B^{j_1 j_2} D_{i_3 j_2}}\right).$$

Proof: The quantity to be minimized only depends on the covariance $C^{ij} = \mathbb{E}[X^i Y^j]$ between X and Y (this quantity is $C^{ij} D_{ij}$). So our problem is equivalent to minimizing $C^{ij} D_{ij}$ over the set of all possible C such that there exists a coupling between X and Y such that the covariance between X and Y is C. Since X and Y are Gaussian, C is the covariance of a coupling between X and Y if and only if

$$\begin{pmatrix} A & C \\ C^T & B \end{pmatrix}$$

is a symmetric non-negative matrix (because there exists a Gaussian coupling having this covariance). This condition is equivalent to $\forall (X_i^*, Y_j^*) \in E_1^* \times E_2^*, X_{i_1}^* A^{i_1 i_2} X_{i_2}^* + Y_{j_1}^* B^{j_1 j_2} Y_{j_2}^* + 2 X_i^* C^{ij} Y_j^* \geq 0$, which is equivalent to $\forall (X_i^*, Y_j^*), |X_i^* C^{ij} Y_j^*| \leq \sqrt{X_{i_1}^* A^{i_1 i_2} X_{i_2}^* Y_{j_1}^* B^{j_1 j_2} Y^* j_2}$. In particular, this implies $C \in \operatorname{Im}(A) \otimes \operatorname{Im}(B)$ (just take $X^* \in \operatorname{Ker}(A)$ or $Y^* \in \operatorname{Ker}(B)$ and remember $\operatorname{Im}(A^T) = (\operatorname{Ker}(A))^\perp$).

Let $n_1 = \operatorname{rk}(A)$ and $n_2 = \operatorname{rk}(B)$ be the ranks of A and B, using suitable bases of $\operatorname{Im}(A)$ and $\operatorname{Im}(B)$, we find "square roots" $A'^i{}_\alpha$ and $B'^j{}_\beta$ of A and B, in the sense that $A' I^{(n_1)} A'^T = A$ and $B' I^{(n_2)} B'^T = B$ ($I_{(n)}$ the scalar product on a canonical n-dimensional Euclidean space and $I^{(n)}$ the associated scalar product in the dual of this space). Then A' and B' admit left inverses A'^{-1} and B'^{-1} (here we don't necessarily have unicity, we just choose two left inverses). We set $C' = A'^{-1} C B'^{-1T}$. We have

$$\begin{pmatrix} A & C \\ C^T & B \end{pmatrix} \geq 0 \Leftrightarrow \begin{pmatrix} A'^{-1} & 0 \\ 0 & B'^{-1} \end{pmatrix} \cdot \begin{pmatrix} A & C \\ C^T & B \end{pmatrix} \cdot \begin{pmatrix} A'^{-1T} & 0 \\ 0 & B'^{-1T} \end{pmatrix} \geq 0$$
$$\Leftrightarrow \begin{pmatrix} I^{(n_1)} & C' \\ C'^T & I^{(n_2)} \end{pmatrix} \geq 0$$

(because $A' A'^{-1}$ restricted to $\operatorname{Im}(A)$ is the identity, and likewise for $B' B'^{-1}$). So we have reduced the problem to the case where E_1' and E_2' are Euclidean spaces of dimensions n_1 and n_2, with $A = I^{(n_1)}$, $B = I^{(n_2)}$ and $D' = A'^T D B'$ instead of D. There exist two nice orthogonal bases so that the matrix of D' in the associated dual bases has the following form:

$$D' = \begin{pmatrix} \operatorname{diag}(\lambda_1, \ldots, \lambda_r) & 0 \\ 0 & 0 \end{pmatrix}$$

with $\operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ the diagonal matrix with coefficients $\lambda_1, \ldots, \lambda_r$, $\lambda_k > 0$ and $r = \operatorname{rk}(D') = \operatorname{rk}(ADB)$, and furthermore, we have unicity of the

coefficients λ_k. This result can be proved thanks to the polar decomposition. We have

$$\begin{pmatrix} I^{(n_1)} & C' \\ C'^T & I^{(n_2)} \end{pmatrix} > 0 \Leftrightarrow \|C'\|_{op} \leq 1$$

with $\|C'\|_{op}$ the operator norm of C' associated with the Euclidean norms, hence the coefficients of C' are greater than or equals to -1. The minimum of $C'^{\alpha\beta}D'_{\alpha\beta}$ is then $-\sum_{k=1}^r \lambda_k$, and is attained when the matrix of C' in the nice bases is

$$C' = \begin{pmatrix} -I_r & 0 \\ 0 & C'' \end{pmatrix}$$

with $\|C''\|_{op} \leq 1$, and only for those ones C'.

The endomorphism $I^{(n_1)}D'I^{(n_2)}D'^T$ has the eigenvalues λ_k^2 and 0 with multiplicity $n_1 - r$ (its matrix in the nice basis is $\mathrm{diag}(\lambda_1^2, \ldots, \lambda_r^2, 0, \ldots, 0)$). We have

$$I^{(n_1)}D'I^{(n_2)}D'^T = I^{(n_1)}A'^T DB'I^{(n_2)}B'^T D^T A' = I^{(n_1)}A'^T DBD^T A'.$$

For any two matrices M, N of size $p \times q$ and $q \times p$, we have for every $n \in \mathbb{N}$, $\mathrm{tr}((MN)^n) = \mathrm{tr}((NM)^n)$, so MN and NM have the same eigenvalues with the same multiplicity, except for the eigenvalue 0, where the difference of the multiplicities is $|p - q|$. So the matrix

$$A'I^{(n_1)}A'DBD^T = ADBD^T$$

also has the eigenvalues λ_k^2 and 0 with some multiplicity. The λ_k are then the non-zero eigenvalues of $\sqrt{ADBD^T}$. So the minimum we were looking for is $-\sum_{k=1}^r \lambda_k = -\mathrm{tr}(\sqrt{ADBD^T})$ $(= -\mathrm{tr}(\sqrt{BD^TAD})$ so the symmetry between A and B is respected, which was not straightforward by looking at the formula). $\square$

The two following remarks provide a good understanding of what the set of the solutions of our minimization problem look like.

Remark 3.16 *The set of all possible covariances is convex and compact, and the quantity to minimize is linear, so the minimum is attained at an extremal point of this convex set. Suppose $n_1 \geq n_2$, then in the case of an extremal covariance, the coupling between X and Y has the form $Y^j = M^j{}_i X^i$. Indeed, $C \mapsto A'^{-1}CB'^{-1T}$ restricted to $\mathrm{Im}(A) \otimes \mathrm{Im}(B)$ is linear and bijective, so C is an extremal covariance if and only if C' is an extremal tensor of operator norm smaller than or equals to 1. we know that for any tensor C' there exists two orthogonal bases in which the matrix of C' can be written:*

$$C' = \begin{pmatrix} \mathrm{diag}(\mu_1, \ldots, \mu_{n_2}) \\ 0 \end{pmatrix}$$

with $\mu_k \geq 0$. The operator norm of C' is then $\max_{1 \leq k \leq n_2} |\mu_k|$.

So C' is an extremal tensor of norm at most 1 if and only if $\mu_k = 1$ for every k.

Indeed, if at least one μ_k is strictly smaller than 1, C' is a non-trivial convex combination of the tensors whose matrices in the same basis are

$$\begin{pmatrix} \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_{n_2}) \\ 0 \end{pmatrix}$$

with $\varepsilon_i = \pm 1$, and each of these tensors has operator norm 1. And conversely, if $\mu_k = 1$, then C' is an extremal tensor of norm at most 1. Assume that $C' = tC^{(1)} + (1-t)C^{(2)}$ with $t \in]0,1[$, and $C^{(1)}$ and $C^{(2)}$ have an operator norm smaller than or eqals to 1. Then the matrices of $C^{(1)}$ and $C^{(2)}$ have coefficients smaller than or equals to 1, so their coefficients on the "diagonal" must be 1. The coefficients outside the "diagonal" are 0 because the sum of the squared coefficients on each row and each column is less than 1. So $C^{(1)} = C^{(2)} = C'$.

If C is an extremal covariance, we have $C'^T I_{(n_1)} C' = I^{(n_2)}$ (just do the product of the matrices in the nice bases). We set then $M = C^T A'^{-1T} I_{(n_1)} A'^{-1} = B'C'^T I_{(n_1)} A'^{-1}$. The covariance of $Y - MX$ is

$$\begin{aligned} B - MC - C^T M^T + MAM^T =& B - [B'C'^T I_{(n_1)} A'^{-1}][A'C'B'^T] \\ & -[B'C'^T A'^T][A'^{-1T} I_{(n_1)} C'B'^T] \\ & +[B'C'^T I_{(n_1)} A'^{-1}][A' I^{(n_1)} A'^T][A'^{-1T} I_{(n_1)} C'B'^T] \\ =& B - B'C'^T I_{(n_1)} C'B'^T = 0. \end{aligned}$$

So $Y = MX$ as previously said.

Remark 3.17 *For any solution C of our minimization problem, we have*

$$\begin{aligned} & CD^T C = A'C'B'^T D^T A'C'B'^T = A'C'D'^T C'B'^T = A' I^{(n_1)} D' I^{(n_2)} B'^T \\ & = A' I^{(n_1)} A'^T D B' I^{(n_2)} B'^T = ADB. \end{aligned}$$

In particular, we have $(CD^T)^2 = ADBD^T$ and $(D^T C)^2 = D^T ADB$. If we take C_0 the solution which corresponds to $C'' = 0$, C_0 is the unique solution with minimal rank (hence the optimal coupling with "the least correlation" between X and Y). We have $\mathrm{rk}(C_0) = \mathrm{rk}(ADB) = \mathrm{rk}(ADBD^T)$, so $\mathrm{rk}(C_0 D^T) \leq \mathrm{rk}(ADBD^T)$, and furthermore $\mathrm{tr}(C_0 D^T) = -\mathrm{tr}(\sqrt{ADBD^T})$, hence we have $C_0 D^T = -\sqrt{ADBD^T}$, and in a similar way $D^T C_0 = -\sqrt{D^T ADB}$. Since $C_0 D^T C_0 = ADB$, we have $\mathrm{Im}(C_0) \supset \mathrm{Im}(ADB)$ and $\mathrm{Im}(C_0^T) \supset \mathrm{Im}(BD^T A)$, and we have in fact equalities because these matrices have the same rank. As ADB, $ADBD^T$ and $D^T ADB$ have the same rank, there exist E and F (which will play the role of D^{-1T}) such that $ED^T ADB = ADB = ADBD^T F$, and then C_0 is given by the formula

$$C_0 = -\sqrt{ADBD^T} F = -E\sqrt{D^T ADB}.$$

For the other solutions, we have

$$C - C_0 = A' \begin{pmatrix} 0 & 0 \\ 0 & C'' \end{pmatrix} B'^T$$

The condition $\|C''\|_{op} \leq 1$ is equivalent to the positivity of

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & I_{n_1 - r} & 0 & C'' \\ 0 & 0 & 0 & 0 \\ 0 & C''^T & 0 & I_{n_2 - r} \end{pmatrix}$$

which is equivalent to the positivity of

$$\begin{pmatrix} A - C_0 B^{-1} C_0^T & C - C_0 \\ (C - C_0)^T & B - C_0^T A^{-1} C_0 \end{pmatrix}$$

where $A^{-1} = A'^{-1T} I_{(n_1)} A'^{-1}$ *and* $B^{-1} = B'^{-1T} I_{(n_2)} B'^{-1}$. *But we would find the same results for the products* $C_0 B^{-1} C_0^T$ *and* $C_0^T A^{-1} C_0$ *by taking* A^{-1} *and* B^{-1} *such that* $AA^{-1}A = A$ *and* $BB^{-1}B = B$, *so this does not depend on the choice of* A', A'^{-1}, B' *or* B'^{-1}.

We can split $\mathrm{Im}(A)$ *as the direct sum of* $\mathrm{Im}(ADB)$ *and the orthogonal (for the quadratic form induced by* A *on* $\mathrm{Im}(A)$*) of this space (which can be written as* $\mathrm{Im}(A) \cap \mathrm{Ker}(BD^T)$*). The two matrices* $C_0 B^{-1} C_0^T$ *and* $A - C_0 B^{-1} C_0^T$ *correspond to the decomposition of* A *on this two subspaces. The similar remark is valid for the matrices* $C_0^T A^{-1} C_0$ *and* $B - C_0^T A^{-1} C_0$ *with respect to the decomposition of* $\mathrm{Im}(B)$ *as the sum of* $\mathrm{Im}(BD^T A)$ *and its orthogonal. An optimal coupling is then any coupling of* X *and* Y *satisfying that the covariance between the orthogonal projections (with respect to* A *and* B*) of* X *and* Y *on* $\mathrm{Im}(ADB)$ *and* $\mathrm{Im}(BD^T A)$ *is* C_0.

3.2.2 The limit of $\kappa(x, y)$ when x and y are close

Let us look at what the formula given by Theorem 3.13 for $\kappa(x, y)$ becomes when we take $y = \exp_x(\delta u)$, d the usual geodesic distance on Riemannian manifolds and when δ tends to 0. We have the following result that gives the second order Taylor expansion of the geodesic distance on Riemannian manifolds.

Lemma 3.18 *Let* $x \in \mathcal{M}$, $(u, v, w) \in (\mathrm{T}_x\mathcal{M})^3$ *such that* $g_{ij} u^i u^j = 1$, $y = \exp_x(\delta u)$, $w' \in \mathrm{T}_y\mathcal{M}$ *obtained from* w *by parallel transport along the geodesic* $t \mapsto \exp_x(\delta t u)$. *Then we have for fixed small enough* δ, *the following Taylor expansion in* ε:

$$\begin{aligned} d(\exp_x(\varepsilon v), \exp_y(\varepsilon w')) = \delta \Bigg(1 &+ \frac{\varepsilon}{\delta} u^i g_{ij} (w^j - v^j) \\ &+ \frac{\varepsilon^2}{2\delta^2} \left(r_{ij}^{(1)} v_i v_j + r_{ij}^{(2)} w^i w^j + 2 r_{ij}^{(12)} v^i w^j + O(\varepsilon^3) \right) \Bigg) \end{aligned}$$

with

$$\begin{aligned} r_{ij}^{(1)} &= g_{ij} - g_{ik} u^k u^l g_{lj} - \tfrac{\delta^2}{3} R_{kilj} u^k u^l + o(\delta^2) \\ r_{ij}^{(2)} &= g_{ij} - g_{ik} u^k u^l g_{lj} - \tfrac{\delta^2}{3} R_{kilj} u^k u^l + o(\delta^2) \\ r_{ij}^{(12)} &= -g_{ij} + g_{ik} u^k u^l g_{lj} - \tfrac{\delta^2}{6} R_{kilj} u^k u^l + o(\delta^2), \end{aligned}$$

where R_{kilj} *is the Riemann tensor of the manifold, and* $r_{ij}^{(1)} u^i v^j = r_{ij}^{(2)} u^i v^j = r_{ij}^{(12)} u^i v^j = r_{ij}^{(12)} u^j v^i = 0$ *(and not only* $o(\delta^2)$*).*

Proof : We will take δ small enough such that (x, y) does not belong to the cut-locus. Then the Riemannian distance is smooth on a neighborhood of (x, y).

For the term in ε, the well known fact that the sphere of center x and radius δ is orthogonal at y to the geodesic joining x to y gives us that the part

of this term depending on w is proportional to $g_{ij}u^iw^j$. A similar argument holds for the term in ε depending on v. Taking v and w proportional to u give the two constants, so we have the term in ε.

For the term in ε^2, we only show that it does only depend on the orthogonal projections of v and w on the orthogonal of u, the proof of the behaviour in δ being based on tedious calculations. We define Σ_x as the image by the exponential map at x of a small ball of the orthogonal of u, and Σ_y as the image by the exponential map at y of a small ball of the orthogonal of u'. For ε small enough, the geodesic between $x_1 = \exp_x(\varepsilon v)$ and $y_1 = \exp_y(\varepsilon w')$ intersects Σ_x and Σ_y at $x_2 = \exp_x(\varepsilon v_1)$ and $y_2 = \exp_y(\varepsilon w_1')$ (we may have to extend the geodesic of $O(\varepsilon)$ beyond x_1 and y_1). We have : $d(x_1, y_1) = \overline{d}(x_1, x_2) + d(x_2, y_2) + \overline{d}(y_2, y_1)$ with $\overline{d}(x_1, x_2) = -d(x_1, x_2)$ if we needed to extend the geodesic beyond x_1 and $d(x_1, x_2)$ otherwise, and the same for $\overline{d}(y_2, y_1)$. We also have $v_1 = v_2 + O(\varepsilon)$ and $w_1 = w_2 + O(\varepsilon)$, where $v_2 = v - \langle u, v\rangle u$, $w_2 = w - \langle u, w\rangle u$, and the $O(\varepsilon)$ are orthogonal to u. Since in the exponential map, the variation of the metric is of order 2, we have $d(x_1, x_2) = \varepsilon\|v - v_1\|(1 + O(\varepsilon^2))$, and $d(y_2, y_1) = \varepsilon\|w - w_1\|(1 + O(\varepsilon^2))$. So we get $\overline{d}(x_1, x_2) = -\varepsilon\langle u, v\rangle + O(\varepsilon^3)$ and $\overline{d}(y_2, y_1) = \varepsilon\langle u, w\rangle + O(\varepsilon^3)$, so we find the terms in ε we expected, and no terms in ε^2. As v_1 and w_1 are orthogonal to u, we get $d(x_2, y_2) = d(\exp_x(v_2), \exp_y(w_2')) + O(\varepsilon^3)$. So the ε^2 term does only depend on v_2 and w_2 as wanted. $\square$

From Theorem 3.13 and Lemma 3.18, we get:

Theorem 3.19 *Suppose we have a diffusion process on a manifold $\mathcal{M}$ such that A and F are $\mathcal{C}^1$, $\mathrm{rk}(A) = n$ everywhere, locally uniformly L^1-bounded. Then $\kappa(x, \exp_x(\delta u))$ converges to*

$$\kappa(x, u) := -u^i g_{ij} u^k \nabla_k F^j + \frac{1}{2} R_{kilj} A^{ij} u^k u^l - \frac{1}{4} \overline{u^i \nabla_i A}^{\alpha\beta} \left(\overline{g^{-1}} \otimes \overline{A} + \overline{A} \otimes \overline{g^{-1}}\right)^{-1}_{\alpha\gamma\delta\beta} \overline{u^j \nabla_j A}^{\gamma\delta}$$

when δ tends to 0.

Here, for any $M \in \mathrm{T}_x\mathcal{M} \otimes \mathrm{T}_x\mathcal{M}$, we denote by $\overline{M}$ the canonical projection of M to $(\mathrm{T}_x\mathcal{M}/\mathrm{Vect}(u)) \otimes (\mathrm{T}_x\mathcal{M}/\mathrm{Vect}(u))$, and the tensor

$$T_{ijkl} = \left(\overline{g^{-1}} \otimes \overline{A} + \overline{A} \otimes \overline{g^{-1}}\right)^{-1}_{ijkl}$$

is uniquely defined by the relationship:

$$T_{ijkl} \left(\overline{g^{-1}}^{jm} \overline{A}^{kn} + \overline{A}^{jm} \overline{g^{-1}}^{kn}\right) = \delta_i^m \delta_l^n.$$

The contraction $T_{ijkl} M^{jk}$ is the unique matrix N_{il} such that $\overline{A} N \overline{g^{-1}} + \overline{g^{-1}} N \overline{A} = M$.

Remark 3.20 *In the special case $A^{ij} = g^{ij}$, we find the usual curvature of the Bakry-Emery theory:*

$$\kappa(x, u) = -\langle u, \nabla_u F\rangle + \frac{1}{2}\mathrm{Ric}(u, u).$$

Proof: The hypothesis that A and F are $\mathcal{C}^1$ gives us that the parallel transport of $A(y)$ and $F(y)$ along the geodesic are $A^{ij}(x)+\delta u^k\nabla_k A^{ij}(x)+o(\delta)=A^{ij}(x)+\delta E^{ij}(\delta)$ where $E^{ij}(\delta)$ tends to $u^k\nabla_k A^{ij}(x)$ when δ tends to 0, and $F^i(x)+\delta u^k\nabla_k F^i(x)+o(\delta)$. The application of Theorem 3.13 and Lemma 3.18 leads to

$$\begin{aligned}\kappa(x,\exp_x(\delta u)) &= \frac{u^i g_{ij}}{\delta}(F^j-(F^j+\delta u^k\nabla_k F^j+o(\delta)))\\ &+\frac{1}{\delta^2}\Big[-(A^{ij}+\frac{\delta}{2}E^{ij}(\delta))(g_{ij}-g_{ik}u^k u^l g_{lj}-\frac{\delta^2}{3}R_{kilj}u^k u^l)\\ &\qquad+\operatorname{tr}(\sqrt{Ar^{(12)}(\delta)(A+\delta E(\delta))r^{(12)T}(\delta)})\Big]+o(1).\end{aligned}$$

The difficult point is to understand the behaviour of the square root when δ tends to 0. The quantity under the square root tends to $(A^{ik}(g_{kj}-g_{kl}u^k u^l g_{lj}))^2$, which is of rank $n-1$ (its kernel is $\mathrm{Vect}(u)$). The square root of matrices is an analytic function in a neigborhood of matrices with positive eigenvalues. This is why we quotient the space $\mathrm{T}_x\mathcal{M}$ by $\mathrm{Vect}(u)$ (thanks to Lemma 3.18, we know that $r^{(12)}\in u^\perp\otimes u^\perp$).

We need the second-order Taylor expansion of $\operatorname{tr}(\sqrt{M^2+\varepsilon N})$ with M a diagonalizable matrix with positive eigenvalues. We have $\sqrt{M^2+\varepsilon N}=M+\varepsilon H+\varepsilon^2K+O(\varepsilon^3)$, so we have $HM+MH=N$ and $H^2+KM+MK=0$. If we work in a diagonalization basis of M (with $\lambda_{(i)}$ the eigenvalues of M), we get:

$$H^i{}_j=\frac{N^i{}_j}{\lambda_{(i)}+\lambda_{(j)}}$$

and

$$K^i{}_j=-\frac{1}{\lambda_{(i)}+\lambda_{(j)}}\sum_k\frac{N^i{}_k}{\lambda_{(i)}+\lambda_{(k)}}\frac{N^k{}_j}{\lambda_{(i)}+\lambda_{(k)}}.$$

So we have:

$$\operatorname{tr}(H)=\sum_i\frac{N^i{}_i}{2\lambda_{(i)}}=\frac{1}{2}\operatorname{tr}(M^{-1}N)$$

and

$$\begin{aligned}\operatorname{tr}(K)&=-\textstyle\sum_{i,j}\frac{N^i{}_jN^j{}_i}{2\lambda_{(i)}(\lambda_{(i)}+\lambda_{(j)})^2}\\ &=-\textstyle\sum_{i,j}\frac{N^i{}_jN^j{}_i}{4(\lambda_{(i)}+\lambda_{(j)})^2}(\frac{1}{\lambda_{(i)}}+\frac{1}{\lambda_{(j)}})\\ &=-\textstyle\sum_{i,j}\frac{N^i{}_jN^j{}_i}{4\lambda_{(i)}\lambda_j(\lambda_{(i)}+\lambda_{(j)})}\\ &=-\tfrac{1}{4}\operatorname{tr}(M^{-1}N((I\otimes M+M\otimes I)^{-1}(M^{-1}N))).\end{aligned}$$

We only have to apply these results with $M=\overline{A}(g-guu^Tg)$, $\varepsilon=\delta$ and $N=\overline{A}(g-guu^Tg)\overline{E(\delta)}(g-guu^Tg)+\frac{\delta}{6}(\overline{A}R(u)\overline{A}(g-guu^Tg)+\overline{A}(g-guu^Tg)\overline{A}R(u))+o(\delta)$, where $R_{ij}(u)=R_{kijl}u^ku^l\in u^\perp\otimes u^\perp$. We obtain:

$$\begin{aligned}\operatorname{tr}(\sqrt{Ar^{(12)}(\delta)(A+\delta E(\delta))r^{(12)T}(\delta)})&=\operatorname{tr}(\overline{A}(g-guu^Tg))\\ &+\frac{\delta}{2}\operatorname{tr}((\overline{E(\delta)}+\frac{\delta}{3}R(u))(g-guu^Tg))\\ -\frac{\delta^2}{4}\operatorname{tr}\big[\overline{\nabla_uA}(g-guu^Tg)((I\otimes M+M\otimes I)^{-1}(\overline{\nabla_uA}(g-guu^Tg)))\big]&+o(\delta^2).\end{aligned}$$

We have $\mathrm{tr}(\overline{A}R(u)) = \mathrm{tr}(AR(u))$, and the last term can be written $-\frac{\delta^2}{4}\,\mathrm{tr}(\overline{\nabla_u A}((\overline{A}\otimes\overline{g^{-1}}+\overline{g^{-1}}\otimes\overline{A})^{-1}\overline{\nabla_u A}))$ because the inverse of $g-guu^Tg$ (acting on $\mathrm{T}_x\mathcal{M}/\mathrm{Vect}(u)$) is $\overline{g^{-1}}$. Replacing this expression of the trace of the square root in the expression of $\kappa(x,\exp_x(\delta u))$ cancels the terms of order $\frac{1}{\delta^2}$ and $\frac{1}{\delta}$, and we get the announced result. $\square$

Remark 3.21 *The dependency on u of the last term of the formula for the curvature is generally not quadratic (because of the complicated dependency on u of the tensor $(\overline{A}\otimes\overline{g^{-1}}+\overline{g^{-1}}\otimes\overline{A})^{-1}$), but is always non-positive and greater than or equals to the same expression without the bars (which we would have obtained by using the W_2 distance instead of the W_1 in the definition on κ, and this expression without the bars depends on u in a quadratic way).*

3.2.3 Construction of the coupling

Now we will construct a coupling between the paths of the diffusion process thanks to the optimal coupling in the tangent spaces. In the case when A is invertible everywhere on $\mathcal{M}$, we have $\mathrm{rk}(A(x)q^{(12)}(x,y)A(y)) = \mathrm{rk}(q^{(12)})(x,y)$. The following lemma shows us that this rank is $n-1$, provided (x,y) is not on the cut-locus of $\mathcal{M}$.

Lemma 3.22 *For the usual distance on $\mathcal{M}$, if $(x,y)\in\mathcal{M}^2$ is not on the cut-locus of $\mathcal{M}$, then we have*

$$q^{(12)}(x,y)(v,w) = -\frac{1}{\delta}\left[g(x)(v,(\mathrm{d}(\exp_x)(\delta u))^{-1}(w)) - g(x)(v,u)g(y)(w,u')\right],$$

with $\delta = d(x,y)$, u the unit norm vector of $\mathrm{T}_x\mathcal{M}$ such that $y=\exp_x(\delta u)$ and u' is the image of u by the parallel transport along the geodesic joining x to y.

We can remark that by a symmetry argument between x and y, this lemma implies that $g(y)\mathrm{d}(\exp_x)(\delta u) = (g(x)\mathrm{d}(\exp_y)(-\delta u'))^T$.

Proof : We alredy know from Lemma 3.18 that $q^{(12)}(x,y)(v,w)=0$ if $v=u$ or $w=u'$. So we can check that it is also true for the formula given in Lemma 3.22, and thereafter we only have to prove the formula for v orthogonal to u and w orthogonal to u'.

Assume $w=u'$, we have to check that $g(x)(v,(\mathrm{d}(\exp_x)(\delta u))^{-1}(u')) - g(x)(v,u)=0$. We trivially have $\mathrm{d}(\exp_x)(\delta u)(u)=u'$, so $\mathrm{d}(\exp_x)(\delta u))^{-1}(u') = u$, then we are done.

Assume $v=u$, we have to check that $g(x)(u,(\mathrm{d}(\exp_x)(\delta u))^{-1}(w)) - g(y)(w,u')=0$. Set $\tilde{w}=(\mathrm{d}(\exp_x)(\delta u))^{-1}(w)$. Then we have $\exp_x(\delta u+\varepsilon\tilde{w}) = \exp_y(\varepsilon w + O(\varepsilon^2))$. Thus we get

$$\|\delta u+\varepsilon\tilde{w}\| = d(x,\exp_x(\delta u+\varepsilon\tilde{w})) = d(x,\exp_y(\varepsilon w+O(\varepsilon^2))) = \delta+\varepsilon g(y)(u',w)+O(\varepsilon^2)$$

where we used Lemma 3.18 for the last equality. Identifying the first order term in ε, we get $g(x)(u,\tilde{w}) = g(y)(u',w)$, which is precisely what we had to check.

Now we can assume that v is orthogonal to u and w is orthogonal to u'. We have to prove that $q^{(12)}(x,y)(v,w) = -\frac{1}{\delta}g(x)(v,(\mathrm{d}(\exp_x)(\delta u))^{-1}(w))$ We set $\tilde{y} = \exp_y(\varepsilon w)$, $\tilde{\delta} = d(x,\tilde{y})$ and $\tilde{u}$ the unit vector of $\mathrm{T}_x\mathcal{M}$ such that $\tilde{y} = \exp_x(\tilde{\delta}\tilde{u})$. Using Lemma 3.18 between x and $\tilde{y}$, we get

$$d(\exp_x(\varepsilon v), \exp_y(\varepsilon w)) = \tilde{\delta}\left(1 - \frac{\varepsilon}{\tilde{\delta}}g(x)(\tilde{u},v) + \frac{\varepsilon^2}{2}q^{(1)}(x,\tilde{y})(v,v) + O(\varepsilon^3)\right).$$

But we have $q^{(1)}(x,\tilde{y}) = q^{(1)}(x,y) + O(\varepsilon)$, and $\tilde{u} = u + \varepsilon(\mathrm{d}(\exp_x)(\delta u))^{-1}(w) + O(\varepsilon^2)$ because $(\mathrm{d}(\exp_x)(\delta u))^{-1}(w)$ is orthogonal to u. Using Lemma 3.18 between x and y, we get

$$\tilde{\delta} = d(x, \exp_y(\varepsilon w)) = \delta\left(1 + \frac{\varepsilon^2}{2}q^{(2)}(x,y)(w,w) + O(\varepsilon^3)\right).$$

Thus, we finally get

$$d(\exp_x(\varepsilon v), \exp_y(\varepsilon w)) = \delta\left(1 + \frac{\varepsilon^2}{2}\left(q^{(1)}(x,y)(v,v) + q^{(2)}(x,y)(w,w) - \tfrac{2}{\delta}g(x)(v,(\mathrm{d}(\exp_x)(\delta u))^{-1}(w))\right) + O(\varepsilon^3)\right)$$

and then by definition of $q^{(12)}(x,y)$, we have

$$q^{(12)}(x,y)(v,w) = \frac{1}{\delta}g(x)(v,(\mathrm{d}(\exp_x)(\delta u))^{-1}(w)).\square$$

Since $\mathrm{d}(\exp_x)(\delta u(x,y))$ is invertible, the rank of $q^{(12)}(x,y)$ is $n-1$ (its kernel is u'). Thus the rank of $A(x)q^{(12)}A(y)$ is also $n-1$.

According to Remarks 3.16 and 3.17, we have two extremal covariances $C^+(x,y)$ and $C^-(x,y)$ in the set of the covariances of optimal couplings, given by the formulas:

$$C^+(x,y) = -\sqrt{A(x)q^{(12)}(x,y)A(y)q^{(12)T}(x,y)}p(x,y) + \frac{1}{\sqrt{u^TA(x)^{-1}uu'^TA(y)^{-1}u'}}uu'^T$$

$$C^-(x,y) = -\sqrt{A(x)q^{(12)}(x,y)A(y)q^{(12)T}(x,y)}p(x,y) - \frac{1}{\sqrt{u^TA(x)^{-1}uu'^TA(y)^{-1}u'}}uu'^T$$

with $y = \exp_x(\delta u)$ with δ small enough, u' the parallel transport of u and $p(x,y)$ any matrix such that $A(x)q^{(12)}(x,y)A(y)q^{(12)T}(x,y)p(x,y) = A(x)q^{(12)}(x,y)A(y)$. The extremal covariance $C^+(x,y)$ tends to $A(x)$ when y tends to x, whereas $C^-(x,y)$ tends to $A(x) - 2\frac{uu^T}{u^TA(x)^{-1}u}$ when u stays fixed and δ tends to 0, so the coupling with $C^+(x,y)$ generalizes the coupling by parallel transport, whereas the one with $C^-(x,y)$ generalizes the coupling by reflection introduced by Kendall in [20]. Here we will use C^+ to construct our coupling for Theorem 3.6, because the behaviour of C^- when δ tends to 0 is irregular.

So we can construct a coupling between the paths as a diffusion process on $\mathcal{M}\times\mathcal{M}$ (at least in the neighborhood of the diagonal), whose generator is defined by:

$$\begin{aligned} L^+(f)(x,y) = \frac{1}{2}[&A(x)^{ij}\nabla^2_{(11)ij}f(x,y) + A(y)^{ij}\nabla^2_{(22)ij}f(x,y) \\ &+ 2C^{+ij}(x,y)\nabla^2_{(12)ij}f(x,y)] + F^i(x)\nabla_{(1)i}f(x,y) + F^i(y)\nabla_{(2)i}f(x,y). \end{aligned}$$

The coupling above in the case $A = g^{-1}$ is the one of Theorem 3.6.

Proof of Theorem 3.6: Let us consider the diffusion process of infinitesimal generator L^+, which is well defined outside the cut-locus of $\mathcal{M}$. In the special case of compact Riemannian manifolds, this is true when $d(x,y)$ is strictly smaller than the injectivity radius. To get the infinitesimal variation of $d(x(t),y(t))$, we have to compute $L^+(f)$ where f has the special form $f(x,y) = \varphi(d(x,y))$ with φ regular enough ($\mathcal{C}^2$). We have:

$$\begin{aligned}
&\nabla_{(1)i} f(x,y) = \varphi'(d(x,y))\nabla_{(1)i} d(x,y)\\
&\nabla_{(2)i} f(x,y) = \varphi'(d(x,y))\nabla_{(2)i} d(x,y)\\
&\nabla^2_{(11)ij} f(x,y) = \varphi'(d(x,y))\nabla^2_{(11)ij} d(x,y) + \varphi''(d(x,y))\nabla_{(1)i} d(x,y)\nabla_{(1)j} d(x,y)\\
&\nabla^2_{(12)ij} f(x,y) = \varphi'(d(x,y))\nabla^2_{(12)ij} d(x,y) + \varphi''(d(x,y))\nabla_{(1)i} d(x,y)\nabla_{(2)j} d(x,y)\\
&\nabla^2_{(22)ij} f(x,y) = \varphi'(d(x,y))\nabla^2_{(22)ij} d(x,y) + \varphi''(d(x,y))\nabla_{(2)i} d(x,y)\nabla_{(2)j} d(x,y)
\end{aligned}$$

with, according to Lemma 3.18:

$$\begin{aligned}
&\nabla_{(1)i} d(x,y) = -g_{ij}(x)u^j(x,y)\\
&\nabla_{(2)i} d(x,y) = -g_{ij}(y)u^j(y,x)\\
&\nabla^2_{(11)ij} d(x,y) = d(x,y)q_i^{(1)} j(x,y)\\
&\nabla^2_{(12)ij} d(x,y) = d(x,y)q_i^{(12)} j(x,y)\\
&\nabla^2_{(22)ij} d(x,y) = d(x,y)q_i^{(2)} j(x,y),
\end{aligned}$$

where $u(x,y) \in \mathrm{T}_x\mathcal{M}$ is an unit norm vector such that $y = \exp_x(d(x,y)u(x,y))$, and the $q^{(\cdot)}(x,y)$ are defined thanks to the Taylor expansion of the distance, like in Theorem 3.13. Thus we get:

$$\begin{aligned}
L^+ f(x,y) = \frac{1}{2}\varphi''(d(x,y))[&A^{ij}(x)g_{ik}(x)u^k(x,y)g_{jl}(x)u^l(x,y)\\
+ A^{ij}(y)g_{ik}(k)u^k(y,x)g_{jl}(y)u^l(y,x) &+ 2C^{+ij}(x,y)g_{ik}(x)u^k(x,y)g_{jl}(y)u^l(y,x)]\\
&- d(x,y)\varphi'(x,y)\kappa(x,y).
\end{aligned}$$

Since we have $A = g^{-1}$, we get $C^+(x,y) = C_0(x,y) - u(x,y)u^T(y,x)$, with $C_0 \in g^{-1}(x)u(x,y)^\perp \otimes g^{-1}(y)u(y,x)^\perp$. So the term containing $\varphi''(d(x,y))$ is 0, which means that the variance of $d(x(t),y(t))$ is $o(t)$ when t tends to 0. So $\mathrm{d}d(x(t),y(t)) = -d(x(t),y(t))\kappa(x(t),y(t))\mathrm{d}t$. Then by integration of this equality, we get:

$$d(x(t),y(t)) = d(x(0),y(0))\mathrm{e}^{-\int_0^t \kappa(x(s),y(s))\mathrm{d}s}.\square$$

3.2.4 The (H) condition and the curvature $\tilde{\kappa}$

The variance term of this optimal coupling is generally not 0 in the case when $A \neq g^{-1}$ (nor a multiple of g^{-1}). So we can try to use another coupling, by replacing $C^+(x,y)$ with $\tilde{C}(x,y)$, which is the optimal covariance (for the distance) under the set of covariances which cancel the variance term of d (if this set is non-empty).

We will prove this set is non-empty if and only if the condition

$$(H) \Leftrightarrow \forall u \in \mathrm{T}\mathcal{M}, u^i g_{jk} u^j g_{lm} u^l \nabla_i A^{km} = 0$$

is satisfied.

Indeed, the variance term is always nonnegative, so it may vanish if and only if its minimum is 0. This is equivalent, according to Lemma 3.15, to

$$\begin{aligned}&2\operatorname{tr}(\sqrt{A(x)g(x)u(x,y)u^T(y,x)g(y)A(y)g(y)u(y,x)u^T(x,y)g(x)})\\&= u^T(x,y)g(x)A(x)g(x)u(x,y) + u^T(y,x)g(y)A(y)g(y)u(y,x),\end{aligned}$$

which is equivalent to

$$u^T(x,y)g(x)A(x)g(x)u(x,y) = u^T(y,x)g(y)A(y)g(y)u(y,x)$$

(this is the equality case in the inequality between arithmetic and geometric mean). Differentiating this condition with respect to y along the geodesic starting at x in the direction u gives the condition (H), and of course the converse implication is obtained by integration.

The hypothesis (H) is a very strong hypothesis: for a given metric, the set of the possible A which are nonnegative and satisfy (H) is a convex cone of finite dimension. Indeed, H is equivalent to: for every geodesic $\gamma(t)$, $A(\gamma(t))(g^{-1}\dot\gamma(t))^{\otimes 2}$ is constant. We choose $x \in \mathcal{M}$, and we take a family of vectors $u_{(k)}, k = 1, \dots, \frac{n(n+1)}{2}$ such that $\{u_{(k)}^{\otimes 2}\}$ is a basis of the symmetric tensors of $\mathrm{T}_x^2\mathcal{M}$. Then we take $x_(k) = \exp_x(\varepsilon u_{(k)})$, with ε small enough to have $\|\varepsilon u_{(k)}\| < r$, with r the injectivity radius of $\mathcal{M}$. Then there exists a ball B centered at x such that for every $y \in B$ and every k, there exists an unique minimal geodesic joining y and $x_{(k)}$, with velocity $v_{((k)}$ at y, and $\{v_{(k)}^{\otimes 2}\}$ is a basis of the symmetric tensors of $\mathrm{T}_y\mathcal{M}$. The knowledge of A at the points $x_{(k)}$ is sufficient to uniquely determine A on the ball B. For any $z \in \mathcal{M}$, we have $x = \exp_z(v)$ for some $v \in \mathrm{T}_z\mathcal{M}$. We can find a family of vectors $v_{(k)} \in \mathrm{T}_z\mathcal{M}$ in a neighborhood of v such that $\{v_{(k)}^{\otimes 2}\}$ is a basis of the symmetric tensors of $\mathrm{T}_z^2\mathcal{M}$, and $\exp_z(v_{(k)}) \in B$. Then the knowledge of A on the points $x_{(k)}$ uniquely determines A on $\mathcal{M}$.

This argument also shows that A is smooth, and the second order Taylor expansion of A in the neighborhood of a single point is sufficient to determine A one the whole manifold. The condition (H), and the equations obtained by differentiating it twice show that this Taylor expansion must belong to a subspace of dimension $\frac{n(n+1)^2(n+2)}{12}$.

The following examples give the set of the possible A in the cases when $\mathcal{M}$ is an Euclidean space of dimension n, the sphere of dimension n or the hyperbolic space of dimension n, providing examples where (H) is satisfied without having $A = g^{ij}$.

Example 3.23 *In all three cases mentionned below, $\mathcal{M}$ can be considered as a submanifold of $E = \mathbb{R}^{n+1}$ such that the geodesics are the intersection of $\mathcal{M}$ and a two dimensional vector subspace of E. Let $(e_1, \dots, e_{n+1})$ be the canonical basis of E and $(e_1^*, \dots, e_{n+1}^*)$ be the corresponding dual basis*

- *We take $\mathcal{M}$ equal to the affine hyperplane of equation $e_{n+1}^*(x) = 1$, equipped with the Euclidean metric $\sum_{i=1}^n e_i^{*2}$ in the first case.*
- *We put the scalar product $s = \sum_{i=1}^{n+1} e_i^{*2}$ on E, and we take $\mathcal{M}$ equal to the sphere $s(x,x) = 1$, equipped with the metric induced by s in the second case*

- *We put the quadratic form* $q = \sum_{i=1}^{n} e_i^{*2} - {e_{n+1}^*}^2$ *on* E*, and we take* $\mathcal{M} = \{x | q(x,x) = -1 \text{ and } e_{n+1}^*(x) > 0\}$*, equipped with the metric induced by* q *in the third case.*

Then we take $T \in E^{*\otimes 4}$ *any tensor with the same symmetry as a Riemann tensor, that is,* T *must satisfy* $T_{ijkl} = -T_{jikl} = -T_{ijlk} = T_{klij}$ *and the Bianchi identity* $T_{ijkl} + T_{jkil} + T_{kijl} = 0$*. We construct the tensor field* A *on* $\mathcal{M}$ *in the following way: let* $(x,v) \in \mathrm{T}\mathcal{M}$*, we want to have*

$$A(x)(g^{-1}v)^{\otimes 2} = T_{ijkl} x^i v^j x^k v^l$$

where the sense of the right hand side is given by considering x *and* v *as elements of* E*. The quadratic dependency in* v *is trivial, so* A *is well defined by the previous equation. Let us consider a unit speed geodesic on* $\mathcal{M}$*, joining two distinct points* x *and* y*, and* v *and* w *be the speed vectors of the geodesic at points* x *and* y*. As said above, the geodesic is included in a two dimensional subspace of* E*, so* (x,v) *and* (y,w) *are two bases of this subspace. Thus there exists a matrix*

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

such that $y = ax + bv$ *and* $w = cx + dv$*. Then we have* $T(y,w,y,w) = \det(M)^2 T(x,v,x,v)$ *(that is a classical property of the Riemann tensor). If* l *is the length of the geodesic, we have:*

$$M = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \qquad \textit{in the case of the Euclidean space,}$$
$$M = \begin{pmatrix} \cos(l) & \sin(l) \\ -\sin(l) & \cos(l) \end{pmatrix} \qquad \textit{in the case of the sphere,}$$
$$M = \begin{pmatrix} \mathrm{ch}(l) & \mathrm{sh}(l) \\ \mathrm{sh}(l) & \mathrm{ch}(l) \end{pmatrix} \qquad \textit{in the case of the hyperbolic space.}$$

In each of the three cases, we have $\det(M) = 1$*. Thus we have* $A(x)(g^{-1}v)^{\otimes 2} = A(y)(g^{-1}w)^{\otimes 2}$ *as wanted.*

The linear application $T \mapsto A$ *is injective, so the dimension of its image is* $\frac{n(n+1)^2(n+2)}{12}$*, which is the maximal dimension of the vector space of symmetric tensor fields on* $\mathcal{M}$ *satisfying the hypothesis* (H)*. Thus this image is exactly this vector space. But the tensor fields* A *which interest us are nonnegative on* $\mathcal{M}$*, and this implies some restrictions on* T*. In the cases of the Euclidean space and the sphere, it is true if and only if the "sectional curvature" associated to* T *is nonnegative, whereas in the case of the hyperbolic space, it is true if and only if this "sectional curvature" is nonnegative on the planes whose intersection with the cone* $q(x,x) = 0$ *is not* $\{0\}$*.*

In the case when (H) is satisfied, the covariances which cancel the variance term of d take the form: $C(x,y) = \tilde{C}_0(x,y) + C'(x,y)$, with

$$\tilde{C}_0(x,y) = -\frac{A(x)g(x)u(x,y)u^T(y,x)g(y)A(y)}{u^T(x,y)g(x)A(x)g(x)u(x,y)} = -\frac{A(x)g(x)u(x,y)u^T(y,x)g(y)A(y)}{u^T(y,x)g(y)A(y)g(y)u(y,x)}$$

and $C'(x,y)$ is such that the big matrix:

$$\begin{pmatrix} A'(x,y) & C'(x,y) \\ C'^T(x,y) & A'(y,x) \end{pmatrix}$$

is nonnegative, with $A'(x,y) = A(x) - \frac{A(x)g(x)u(x,y)u^T(x,y)g(x)A(x)}{u^T(x,y)g(x)A(x)g(x)u(x,y)}$.

Using Lemma 3.15 again gives us the following expression of $\tilde{\kappa}(x,y)$:

$$\begin{aligned}\tilde{\kappa}(x,y) &= \frac{1}{\delta}(F(y)g(y)u(y,x) + F(x)g(x)u(x,y)) \\ &- \frac{1}{2}(\mathrm{tr}(A(x)q^{(1)}(x,y)) + \mathrm{tr}(A(y)q^{(2)}(x,y))) - \mathrm{tr}(\tilde{C}_0(x,y)q^{(12)T}(x,y)) \\ &\qquad + \mathrm{tr}(\sqrt{A'(x,y)q^{(12)}(x,y)A'(y,x)q^{(12)T}(x,y)}).\end{aligned}$$

We can define $\tilde{\kappa}(x,u)$ as the limit when δ tends to 0 of $\tilde{\kappa}(x, \exp_x(\delta u))$. Then we have:

$$\begin{aligned}\tilde{\kappa}(x,u) = &-u^i g_{ij} u^k \nabla_k F^j + \tfrac{1}{2} A^{ij} R_{ikjl} u^k u^l - \tfrac{u^i g_{ij} u^k \nabla_k A^{jl} g_{lm} u^n \nabla_n A^{mo} g_{op} u^p}{2u^i g_{ij} A^{jk} g_{kl} u^l} \\ &-\tfrac{1}{4} B^{ij} (A' \otimes (g^{-1} - uu^T) + (g^{-1} - uu^T) \otimes A')^{-1}_{kijl} B^{kl}\end{aligned}$$

with

$$\begin{aligned}A' &= A - \tfrac{Aguu^T gA}{u^T gAgu} \\ B^{ij} &= \nabla_u A - \tfrac{\nabla_u Aguu^T gA + Aguu^T g\nabla_u A}{u^T gAgu},\end{aligned}$$

and as A', B and $g^{-1} - uu^T$ belong to $g^{-1}u^\perp \otimes g^{-1}u^\perp$, we take $(A' \otimes (g^{-1} - uu^T) + (g^{-1} - uu^T) \otimes A')^{-1}$ the unique inverse of $A' \otimes (g^{-1} - uu^T) + (g^{-1} - uu^T) \otimes A'$ in $(\mathrm{T}^*_x\mathcal{M}/\mathrm{Vect}(gu))^{\otimes 4}$.

And we have the equivalent of Theorem 3.6:

Lemma 3.24 *If the hypothesis (H) is satisfied, then there exists a coupling between paths such that*

$$d(X(t), Y(t)) = d(X(0), Y(0))\mathrm{e}^{-\int_0^t \tilde{\kappa}(X(s),Y(s))\mathrm{d}s}$$

almost surely on the event that for every $0 \le s \le t$, $d(x',y')^2$ is smooth in a neighborhood of $(X(s), Y(s))$.

3.3 New bounds for the spectral gap

The idea to prove Theorems 3.4 and 3.8 is to look at the exponential decay of the Lipschitz norm of $P^t f$ when f is Lipschitz with mean 0 (respect to the reversible probability measure). Then we use the reversibility assumption to conclude that the variance of $P^t f$ also decreases exponentially fast with the same rate, which is hence a lower bound for the spectral gap.

Proof of Theorem 3.8: Let x and y be two points of $\mathcal{M}$ such that $d(x,y) < r_i$ where r_i is the injectivity radius of $\mathcal{M}$. We have $P^t f(y) - P^t f(x) = \mathbb{E}[f(Y(t)) - f(X(t))]$ for any coupling between paths. If f is 1-Lipschitz, then $|f(Y(t)) - f(X(t))| \le d(Y(t), X(t))$, so Lemma 3.24 tells us that $|P^t f(y) - P^t f(x)| \le d(x,y)\mathbb{E}[\mathrm{e}^{-\int_0^t \tilde{\kappa}(X(s),Y(s))\mathrm{d}s}]$. For any $\varepsilon > \eta > 0$, there exists $\delta > 0$ such that for all (x',y') such that $d(x',y') \le \delta$, we have $\tilde{\kappa}(x',y') \ge \tilde{\kappa}(x') - \eta$, where $\tilde{\kappa}(x') = \inf_{u \in \mathrm{T}_{x'}\mathcal{M}} \tilde{\kappa}(x',u)$. Taking $T = \frac{\ln(\frac{r_i}{\delta})}{\varepsilon}$, we have $d(X(T), Y(T)) \le \delta$, and then for $t \ge T$ we have

$\mathbb{E}[\mathrm{e}^{-\int_0^t \tilde{\kappa}(X(s),Y(s))\mathrm{d}s}] \leq \frac{\delta}{r_i}\mathbb{E}[\mathrm{e}^{-\int_T^{t-T}(\tilde{\kappa}(X(s))-\eta)\mathrm{d}s}]$. Following what was done in [17], we use the Feynman-Kac semigroup F^t generated by K, with

$$Kf(x) = \frac{1}{2}A^{ij}(x)\nabla^2_{ij}f(x) + F^i(x)\nabla_i f(x) - (\tilde{\kappa}(x) - \eta)f(x).$$

Indeed we have $\mathbb{E}[\mathrm{e}^{-\int_0^t(\tilde{\kappa}(X(s))-\eta)\mathrm{d}s}] = F^t 1(x)$. The Lipschitz norm of $P^t f$ is at most $\sup_{x\in\mathcal{M}} \frac{\delta}{r_i}\mathbb{E}[\mathrm{e}^{-\int_T^{t-T}\tilde{\kappa}(X(s),Y(s))\mathrm{d}s}]$, for every $t \geq T$. This quantity is $\sup_{x\in\mathcal{M}} \frac{\delta}{r_i}\mathbb{E}_{\delta_x.P^T}F^{t-T}1(y)$, so it is smaller than or equals to

$$\frac{\delta}{r_i}\sup_{x\in\mathcal{M}} \|\frac{\mathrm{d}(\delta_x.P^T)}{\mathrm{d}\pi}\|_{L^2(\pi)}\|F^{t-T}1\|_{L^2(\pi)}.$$

Then the Lipschitz norm of $P^t f$ decreases exponentially fast with a better rate than the one of $F^t 1$.

The L^2-norm of $F^t 1$ decreases exponentially with rate $\inf_{h|\int h^2\mathrm{d}\pi=1}\int -hKh\mathrm{d}\pi \geq \inf_{h|\int h^2\mathrm{d}\pi=1}\lambda_1\mathrm{Var}_\pi(h) + \int(\tilde{\kappa}(x)-\eta)h(x)^2\mathrm{d}\pi(x) = \lambda_1 + \inf_{h|\int h^2\mathrm{d}\pi=1}\int(\tilde{\kappa}(x)-\eta)h(x)^2\mathrm{d}\pi(x) - \lambda_1(\int h\mathrm{d}\pi)^2$. The method of Lagrange multiplicators suggests to take

$$h(x) = \frac{c}{\tilde{\kappa}(x) - \eta + \alpha},$$

with α such that

$$\frac{1}{\lambda_1} = \int \frac{\mathrm{d}\pi(x)}{\tilde{\kappa}(x) - \eta + \alpha}$$

and $c = \frac{1}{\sqrt{\int \frac{\mathrm{d}\pi(x)}{(\tilde{\kappa}(x)-\eta+\alpha)^2}}}$. With this h, we have

$$\int(\tilde{\kappa}(x)-\eta)h(x)^2\mathrm{d}\pi(x) - \lambda_1\left(\int h\mathrm{d}\pi\right)^2 = -\alpha.$$

This is indeed the minimal h when λ_1 is at least the harmonic mean λ of $\tilde{\kappa} - \inf(\tilde{\kappa})$. We can see it by using Cauchy-Schwarz:

$$\int(\tilde{\kappa}(x)-\eta)h(x)^2\mathrm{d}\pi(x) - \lambda_1\left(\int h\mathrm{d}\pi\right)^2 \geq \int(\tilde{\kappa}(x)-\eta)h(x)^2\mathrm{d}\pi(x)$$
$$- \lambda_1\left(\int(\tilde{\kappa}(x)-\eta+\alpha)h^2\mathrm{d}\pi(x)\right)\left(\int\frac{\mathrm{d}\pi(x)}{\tilde{\kappa}(x)-\eta+\alpha}\right) = -\alpha.$$

In the case where $\lambda_1 < \lambda$, we take $\alpha = \eta - \inf(\tilde{\kappa})$. This time we get

$$\int(\tilde{\kappa}(x)-\eta)h(x)^2\mathrm{d}\pi(x) - \lambda_1\left(\int h\mathrm{d}\pi\right)^2 \geq \int(\tilde{\kappa}(x)-\eta)h(x)^2\mathrm{d}\pi(x)$$
$$-\lambda\left(\int(\tilde{\kappa}(x)-\eta+\alpha)h^2\mathrm{d}\pi(x)\right)\left(\int\frac{\mathrm{d}\pi(x)}{\tilde{\kappa}(x)-\eta+\alpha}\right)+(\lambda-\lambda_1)\int(h\mathrm{d}\pi)^2 \geq -\alpha.$$

A minimizing sequence $h_i(x)$ can be, for example, a sequence such that h_i^2 tends to a Dirac at a point where the minimum of $\tilde{\kappa}$ is reached.

In both cases, the exponential decay rate for zero-mean Lipschitz functions is at least $\lambda_1 - \alpha$, then by density of the Lipschitz functions on $L^2(\pi)$

and by the reversibility assumption, the exponential decay rate for zero-mean $L^2(\pi)$ functions (which is equal to λ_1) is also at least $\lambda_1 - \alpha$. Thus α is non-negative, which means that λ_1 is at least the harmonic mean of $\tilde{\kappa} - \eta$, so letting η tend to 0 yields the result. □

Purely analytical methods can also be used to prove this result in the case $A^{ij} = g^{ij}$, and they also work when $\inf_{x\in\mathcal{M}} \kappa(x) = 0$.

Lemma 3.25 *Let f be a regular enough ($\mathcal{C}^3$) function from $\mathcal{M}$ to $\mathbb{R}$. Then we have*

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \|\nabla P^t f\|^2 = h(2L(h) + u_k g^{kl}\nabla_l A^{ij}\nabla_i h u_j)$$
$$+h^2\left(u_k g^{kl}\nabla_l A^{ij}\nabla_i u_j - A^{ij}g^{kl}\nabla_i u_k \nabla_j u_l + A^{ij}R_{lij\alpha}g^{\alpha\beta}u_\beta g^{kl}u_k + 2u_k g^{kl}\nabla_l F^i u_i\right)$$

where $h = \|\nabla f\|$ and $\nabla_k f = h u_k$.

Proof: We have $\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \|\nabla P^t f\|^2 = 2\nabla_k f g^{kl}\nabla_l(Lf)$, and

$$\begin{aligned}\nabla_l(Lf) &= \tfrac{1}{2}(\nabla_l A^{ij}\nabla^2_{ij}f + A^{ij}\nabla^3_{lij}f) + \nabla_l F^i \nabla_i f + F^i \nabla^2_{li} f\\ &= \tfrac{1}{2}(\nabla_l A^{ij}\nabla^2_{ij}f + A^{ij}(\nabla^3_{ijl}f + R_{lij\alpha}g^{\alpha\beta}\nabla_\beta f)) + \nabla_l F^i\nabla_i f + F^i\nabla^2_{il}f.\end{aligned}$$

Differentiating $\nabla_i f = h u_i$, we get $\nabla^2_{ij}f = \nabla_i h u_j + h\nabla_i u_j$, and $\nabla^3_{ijl}f = \nabla^2_{ij}h u_l + \nabla_j h\nabla_i u_l + \nabla_i h\nabla_j u_l + h\nabla^2_{ij}u_l$. So we get:

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \|\nabla P^t f\|^2 = h g^{kl}u_k\big[&\nabla_l A^{ij}(\nabla_i h u_j + h\nabla_i u_j)\\ &+ A^{ij}\left(\nabla^2_{ij}h u_l + \nabla_j h\nabla_i u_l + \nabla_i h\nabla_j u_l + h\nabla^2_{ij}u_l + hR_{lij\alpha}g^{\alpha\beta}u_\beta\right)\\ &+2h\nabla_l F^i u_i + 2F^i(\nabla_i h u_l + h\nabla_i u_l)\big].\end{aligned}$$

Differentiating $g^{kl}u_k u_l = 1$ gives $g^{kl}u_k\nabla_j u_l = 0$ and $g^{kl}(\nabla_i u_k\nabla_j u_l + u_k\nabla^2_{ij}u_l) = 0$, so using these relationships, the above expression can be simplified to get the formula given in Lemma 3.25. □

Proof of Theorem 3.7: We first prove the Theorem in the case $n' = \infty$ and $c = 0$, in which case we get the result of Theorem 3.8. Indeed, in this case, the optimal $\rho(x)$ is nothing but $\kappa(x) = \inf_{u\in \mathrm{T}_x\mathcal{M}} \kappa(x,u)$.

Let f be an eigenfunction of L for the eigenvalue $-\lambda_1$. With the previous notation for h and u, we have:

$$\begin{aligned}-2\lambda_1\|h\|^2_{L^2(\pi)} = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\|\nabla P^t f\|^2_{L^2(\pi)} &= \textstyle\int 2h(x)L(h)(x) - 2h(x)^2\kappa(x, g^{-1}u(x))\\ &\quad -h(x)^2 A^{ij}(x)g^{kl}(x)\nabla_i u_k(x)\nabla_j u_l(x)\mathrm{d}\pi(x)\\ &\le -2\lambda_1(\textstyle\int h(x)^2\mathrm{d}\pi(x) - (\int h(x)\mathrm{d}\pi(x))^2)\\ &\quad -2\textstyle\int \kappa(x)h(x)^2\mathrm{d}\pi(x) + 0\end{aligned}$$

where the inequality $\int hL(h) \le -\lambda_1\mathrm{Var}(h)$ is due to the reversibility assumption. By Cauchy-Schwarz, we have

$$(\int h(x)\mathrm{d}\pi(x))^2 \le \int \frac{\mathrm{d}\pi(x)}{\kappa(x)}\int \kappa(x)h(x)^2\mathrm{d}\pi(x).$$

Finally we get:

$$\int \kappa(x)h^2(x)\mathrm{d}\pi(x)(\lambda_1\int\frac{\mathrm{d}\pi(x)}{\kappa(x)} - 1) \ge 0,$$

and if $\int \frac{\mathrm{d}\pi(x)}{\kappa(x)} < +\infty$, then $\int \kappa(x)h^2(x) > 0$, because f is nonconstant, so h can't be 0 almost everywhere. So we have

$$\lambda_1 \geq \frac{1}{\int_{\mathcal{M}} \frac{\mathrm{d}\pi(x)}{\kappa(x)}}.$$

In the general case, we have $n' \geq n$ and the optimal ρ is given by:

$$\rho(x) = \frac{1}{2} \inf_{u \in \mathrm{T}_x\mathcal{M}, \|u\|=1} \mathrm{Ric}(u,u) + \nabla^2_{u,u}\varphi - \frac{(\nabla_u \varphi)^2}{n'-n}.$$

So we have $\rho \leq \kappa$. Then with the previous notation, we have:

$$\lambda_1 (\int_{\mathcal{M}} h(x)\mathrm{d}\pi(x))^2 - \int_{\mathcal{M}} \rho(x)h^2(x)\mathrm{d}\pi(x) \geq 0$$

because we have just shown the same with κ instead of ρ.

We also have

$$\begin{aligned}\int_{\mathcal{M}} \Gamma_2(f)(x)\mathrm{d}\pi(x) &= \int_{\mathcal{M}} \tfrac{1}{2}[L(\tfrac{h^2}{2} - \langle \nabla f, \nabla(Lf)\rangle]\mathrm{d}\pi = 0 + \tfrac{\lambda_1}{2}\int_{\mathcal{M}} h^2\mathrm{d}\pi \\ &\geq \int_{\mathcal{M}} \rho\Gamma(f) + \tfrac{1}{n'}L(f)^2\mathrm{d}\pi = \tfrac{1}{2}\int_{\mathcal{M}} \rho h^2\mathrm{d}\pi + \tfrac{\lambda_1}{2n'}\int_{\mathcal{M}} h^2\mathrm{d}\pi.\end{aligned}$$

Thus for any $\theta \in [0,1]$ we have:

$$(1-\theta)\lambda_1 (\int_{\mathcal{M}} h\mathrm{d}\pi)^2 - \int_{\mathcal{M}} \left(\rho - \theta\lambda_1 \frac{n'-1}{n'}\right) h^2\mathrm{d}\pi \geq 0.$$

For $\theta = 1$, we have $0 \leq \int_{\mathcal{M}} (\lambda_1 \frac{n'-1}{n'} - \rho)h^2\mathrm{d}\pi \leq \lambda_1 \frac{n'-1}{n'} - \inf(\rho)\int_{\mathcal{M}} h^2\mathrm{d}\pi$, this proves the Bakry–Émery bound:

$$\lambda_1 \geq \frac{n'}{n'-1}\inf(\rho).$$

So for any $c \in [0, \inf(\rho)]$, we take $\theta = \frac{n'c}{(n'-1)\lambda_1} \in [0,1]$. By Cauchy–Schwarz, we get

$$(1-\theta)\lambda_1 \int (\rho - c)h^2\mathrm{d}\pi \int \frac{\mathrm{d}\pi}{\rho - c} - \int (\rho - c)h^2\mathrm{d}\pi \geq 0$$

Thus we get $(\lambda_1 - c\frac{n'}{n'-1})\int \frac{\mathrm{d}\pi}{\rho - c} - 1 \geq 0$, which leads to the desired result.□

Bibliography

[1] C. Ané, S. Blachère, D. Chafai, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*. Number 10 in Panoramas et Synthèses. Société Mathématique de France, 2000.

[2] M. Arnaudon, K. A. Coulibaly, and A. Thalmaier. Horizontal Diffusion in C^1 Path Space. In *Séminaire de Probabilités XLIII*, volume 2006 of *Lecture Notes in Mathematics*, pages 73–94. Springer, 2011.

[3] E. Aubry. Finiteness of π_1 and geometric inequalities in almost positive Ricci curvature. *Annales Scientifiques de l'École Normale Supérieure*, 40, July–August 2007.

[4] D. Bakry and M. Émery. Hypercontractivité de semi-groupes de diffusion. *Comptes Rendus des séances de l'Acedémie des Sciences. Série 1, Mathématique*, 299(15):775–778, 1984.

[5] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX, 1983/84*, volume 1123 of *Lecture notes in Mathematics*, pages 177–206. Springer, 1985.

[6] M. Berger. *A panoramic view of Riemannian geometry*. Springer, 2003.

[7] 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.

[8] R. Bubley and M. Dyer. Path Coupling: a Technique for Proving Rapid Mixing in Markov Chains. *Foundations of Computer Science*, pages 223–231, 1997.

[9] M.-F. Chen. *From Markov Chains to Non-Equilibrium Particle Systems*. World Scientific, 2004.

[10] M.-F. Chen and F.-Y. Wang. Application of coupling method to the first eigenvalue on manifold. *Science in China, Series A: Mathematics*, 37(1):1–14, 1994.

[11] M.-F. Chen and F.-Y. Wang. General formula for lower bound of the first eigenvalue on Riemannian manifolds. *Science in China, Series A: Mathematics*, 40(4):384–394, April 1997.

[12] B. Cloez. Wasserstein decay of one dimensional jump-diffusions. available on ArXiv, 2012.

[13] R. Dobrushin. Prescribing a system of random variables by conditional distribution. *Theory of Probability and its Applications*, 15(3):458–486, 1970.

[14] R. Dobrushin. Perturbation methods of the theory of Gibbsian fields. In *Lectures on Probability Theory and Statistics*, volume 1648 of *Lecture notes in Mathematics*, pages 1–66. Springer, 1996.

[15] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer, third edition, 2004.

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

[17] A. Guillin, C. Léonard, L. Wu, and N. Yao. Transportation-information inequalities for Markov processes. *Probability Theory and Related Fields*, 144:669–695, 2009.

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

[19] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *The Annals of Probability*, 38(6):2418–2442, 2010.

[20] W. S. Kendall. Nonnegative Ricci curvature and the Brownian coupling property. *Stochastics*, 19:111–129, 1986.

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

[22] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. Available on ArXiv, 2006.

[23] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Inventiones Mathematicae*, 177(1):1–43, 2009.

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

[25] K. Parthasarathy. *Probability measures on metric spaces*. Academic Press, New York, 1967.

[26] K.-T. Sturm. On the geometry of metric measure spaces. I. *Acta Mathematica*, 196(1):65–131, 2006.

[27] L. Veysseire. A harmonic mean bound for the spectral gap of the Laplacian on Riemannian manifolds. *Comptes Rendus Mathématiques*, 348:1319–1322, December 2010.

[28] L. Veysseire. Improved spectral gap bounds on positively curved manifolds. Available on ArXiv, 2011.

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

[30] L. Veysseire. A concentration theorem for the equilibrium measure of markov chains with nonnegative coarse ricci curvature. Available on ArXiv, 2012.

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