Available online at www.sciencedirect.com

ScienceDirect

Journal of Functional Analysis 256 (2009) 810–864

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

Ricci curvature of Markov chains on metric spaces

Yann Ollivier

ENS Lyon, UMPA, 46 allée d'Italie, Lyon, France

Received 20 December 2007; accepted 3 November 2008

Available online 28 November 2008

Communicated by Cédric Villani

Abstract

We define the coarse Ricci curvature of metric spaces in terms of how much small balls are closer (in Wasserstein transportation distance) than their centers are. This definition naturally extends to any Markov chain on a metric space. For a Riemannian manifold this gives back, after scaling, the value of Ricci curvature of a tangent vector. Examples of positively curved spaces for this definition include the discrete cube and discrete versions of the Ornstein–Uhlenbeck process. Moreover this generalization is consistent with the Bakry–Émery Ricci curvature for Brownian motion with a drift on a Riemannian manifold.

Positive Ricci curvature is shown to imply a spectral gap, a Lévy–Gromov–like Gaussian concentration theorem and a kind of modified logarithmic Sobolev inequality. The bounds obtained are sharp in a variety of examples.

Keywords: Ricci curvature; Markov chains; Metric geometry; Concentration of measure

0. Introduction

In Riemannian geometry, positively curved spaces in the sense of Ricci curvature enjoy numerous properties, some of them with very natural probabilistic interpretations. A basic result involving positive Ricci curvature is the Bonnet–Myers theorem bounding the diameter of the space via curvature; let us also mention Lichnerowicz's theorem for the spectral gap of the Laplacian (Theorem 181 in [7]), hence a control on mixing properties of Brownian motion; and the Lévy–Gromov theorem for isoperimetric inequalities and concentration of measure [27]. The scope of these theorems has been noticeably extended by Bakry–Émery theory [5,6], which

E-mail address: yann.ollivier@umpa.ens-lyon.fr.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2008.11.001

highlights the analytic and probabilistic significance of Ricci curvature; in particular, they show that in positive Ricci curvature, a logarithmic Sobolev inequality holds. We refer to the nice survey [30] and the references therein for a discussion of the geometric interest of lower bounds on Ricci curvature and the need for a generalized notion of positive Ricci curvature for metric measure spaces.

Here we define a notion of Ricci curvature which makes sense for a metric space equipped with a Markov chain (or with a measure), and allows to extend the results above. Namely, we compare the transportation distance between the measures issuing from two given points to the distance between these points (Definition 3), so that Ricci curvature is positive if and only if the random walk operator is contracting on the space of probability measures equipped with this transportation distance (Proposition 20). Thus, the techniques presented here are a metric version of the usual coupling method; namely, Ricci curvature appears as a refined version of Dobrushin's classical ergodic coefficient ([16,17], or e.g. Section 6.7.1 in [9]) using the metric structure of the underlying space.

Our definition is very easy to implement on concrete examples. Especially, in ε-geodesic spaces, positive curvature is a local property (Proposition 19), as can be expected of a notion of curvature. As a result, we can test our notion in discrete spaces such as graphs. An example is the discrete cube $\{0,1\}^N$, which from the point of view of concentration of measure or convex geometry [29,35] behaves very much like the sphere S^N, and is thus expected to somehow have positive curvature.

Our notion enjoys the following properties: when applied to a Riemannian manifold equipped with (a discrete-time approximation of) Brownian motion, it gives back the usual value of the Ricci curvature of a tangent vector. It is consistent with the Bakry–Émery extension, and provides a visual explanation for the curvature contribution $-\nabla^{\mathrm{sym}} b$ of the drift term b in this theory. We are able to prove generalizations of the Bonnet–Myers theorem, of the Lichnerowicz spectral gap theorem and of the Lévy–Gromov isoperimetry theorem, as well as a kind of modified logarithmic Sobolev inequality. As a by-product, we get a new proof for Gaussian concentration and the logarithmic Sobolev inequality in the Lévy–Gromov or Bakry–Émery context (although with some loss in the numerical constants). We refer to Section 1.3 for an overview of the results.

Some of the results of this text have been announced in a short note [40].

Historical remarks and related work. In the respective context of Riemannian manifolds or of discrete Markov chains, our techniques reduce, respectively, to Bakry–Émery theory or to a metric version of the coupling method. As far as I know, it had not been observed that these can actually be viewed as the same phenomenon.

From the discrete Markov chain point of view, the techniques presented here are just a version of the usual coupling method using the metric structure of the underlying space. Usually the coupling method involves total variation distance (see e.g. Section 6.7.1 in [9]), which can be seen as a transportation distance with respect to the trivial metric. The coupling method is especially powerful in product or product-like spaces, such as spin systems. The work of Marton [32,33] emphasized the relationship between couplings and concentration of measure in product-like situations, so it is not surprising that we are able to get the same kind of results. The relationship between couplings and spectral gap is thoroughly explored in the works of Chen (e.g. [11,13,14]).

The contraction property of Markov chains in transportation distance seems to make its appearance in Dobrushin's paper [18] (in which the current wide interest in transportation distances originates), and is implicit in the widely used "Dobrushin criterion" for spin systems [18,20].

It later appears sporadically in the literature, as in Chen and Wang [14] (Theorem 1.9, as a tool for spectral gap estimates, using the coupling by reflection); at the very end of Dobrushin's notes [19] (Dobrushin's study of the topic was stopped by his death); in Bubley and Dyer [4] for the particular case of product spaces, after Dobrushin; in the second edition of [12, Section 5.3]; in Djellout, Guillin and Wu [15] in the context of dependent sequences of random variables to get Gaussian concentration results; in lecture notes by Peres [42] and in [44, p. 94]. See also the related work mentioned below. However, the theorems exposed in our work are new.

From the Riemannian point of view, our approach boils down to contraction of the Lipschitz norm by the heat equation, which is one of the results of Bakry and Émery ([5,6], see also [1] and [43]). This latter property was suggested in [43] as a possible definition of a lower bound on Ricci curvature for diffusion operators in general spaces, though it does not provide an explicit value for Ricci curvature at a given point.

Another notion of lower bound on Ricci curvature, valid for length spaces equipped with a measure, has been simultaneously introduced by Sturm [46], Lott and Villani [31], and Ohta [37] (see also [43] and [41]). It relies on ideas from optimal transportation theory and analysis of paths in the space of probability measures. Their definition keeps a lot of the properties traditionally associated with positive Ricci curvature, and is compatible with the Bakry–Émery extension. However, it has two main drawbacks. First, the definition is rather involved and difficult to check on concrete examples. Second, it is infinitesimal, and difficult to adapt to discrete settings [10].

Related work. After having written a first version of this text, we learned that related ideas appear in some recent papers. Joulin [28] uses contraction of the Lipschitz constant (under the name "Wasserstein curvature") to get a Poisson-type concentration result for continuous-time Markov chains on a countable space, at least in the bounded, one-dimensional case. Oliveira [38] considers Kac's random walk on $\mathrm{SO}(n)$; in our language, his result is that this random walk has positive coarse Ricci curvature, which allows him to improve mixing time estimates significantly.

Notation. We use the symbol $\approx$ to denote equality up to a multiplicative universal constant (typically 2 or 4); the symbol $\sim$ denotes usual asymptotic equivalence. The word "distribution" is used as a synonym for "probability measure".

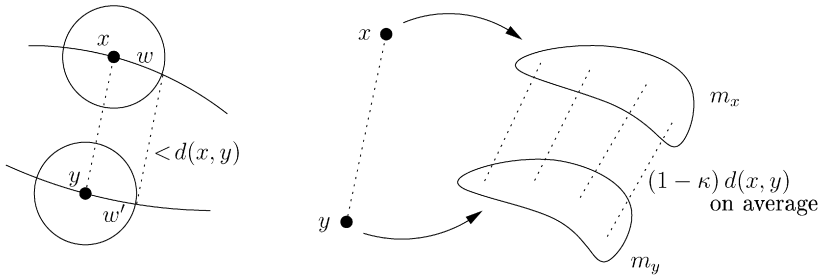
Here for simplicity we will mainly consider discrete-time processes. Similar definitions and results can be given for continuous time (see e.g. Section 3.3.4).

1. Definitions and statements

1.1. Coarse Ricci curvature

In Riemannian geometry, positive Ricci curvature is characterized [43] by the fact that "small spheres are closer (in transportation distance) than their centers are." More precisely, consider two very close points x, y in a Riemannian manifold, defining a tangent vector (xy). Let w be another tangent vector at x; let w' be the tangent vector at y obtained by parallel transport of w from x to y. Now if we follow the two geodesics issuing from x, w and y, w', in positive curvature the geodesics will get closer, and will part away in negative curvature. Ricci curvature along (xy) is this phenomenon, averaged on all directions w at x. If we think of a direction w at x as a point on a small sphere S_x centered at x, this shows that, on average, Ricci curvature controls whether the distance between a point of S_x and the corresponding point of S_y is smaller or larger than the distance $d(x, y)$.

In a more general context, we will use a probability measure m_x depending on x as an analogue for the sphere (or ball) S_x centered at x.



Definition 1. Let (X, d) be a Polish metric space, equipped with its Borel σ-algebra.

A *random walk m on X* is a family of probability measures $m_x(\cdot)$ on X for each $x \in X$, satisfying the following two technical assumptions: (i) the measure m_x depends measurably on the point $x \in X$; (ii) each measure m_x has finite first moment, i.e. for some (hence any) $o \in X$, for any $x \in X$ one has $\int d(o, y)\, \mathrm{d}m_x(y) < \infty$.

Instead of "corresponding points" between two close spheres S_x and S_y, we will use transportation distances between measures. We refer to [47] for an introduction to the topic. This distance is usually associated with the names of Kantorovich, Rubinstein, Wasserstein, Ornstein, Monge, and others (see [36] for a historical account); we stick to the simpler and more descriptive "transportation distance."

Definition 2. Let (X, d) be a metric space and let ν_1, ν_2 be two probability measures on X. The L^1 *transportation distance* between ν_1 and ν_2 is

$$W_1(\nu_1, \nu_2) := \inf_{\xi \in \Pi(\nu_1, \nu_2)} \int_{(x,y) \in X \times X} d(x, y)\, \mathrm{d}\xi(x, y)$$

where $\Pi(\nu_1, \nu_2)$ is the set of measures on $X \times X$ projecting to ν_1 and ν_2.

Intuitively, $\mathrm{d}\xi(x, y)$ represents the mass that travels from x to y, hence the constraint on the projections of ξ, ensuring that the initial measure is ν_1 and the final measure is ν_2. The infimum is actually attained (Theorem 1.3 in [47]), but the optimal coupling is generally not unique. In what follows, it is enough to choose one such coupling.

The data $(m_x)_{x \in X}$ allow to define a notion of curvature as follows: as in the Riemannian case, we will ask whether the measures m_x and m_y are closer or further apart than the points x and y are, in which case Ricci curvature will be, respectively, positive or negative.

Definition 3 *(Coarse Ricci curvature).* Let (X, d) be a metric space with a random walk m. Let $x, y \in X$ be two distinct points. The *coarse Ricci curvature of* (X, d, m) *along* (xy) is

$$\kappa(x, y) := 1 - \frac{W_1(m_x, m_y)}{d(x, y)}.$$

We will see below (Proposition 19) that in geodesic spaces, it is enough to know $\kappa(x, y)$ for close points x, y.

Geometers will think of m_x as a replacement for the notion of ball around x. Probabilists will rather think of this data as defining a Markov chain whose transition probability from x to y in n steps is

$$\mathrm{d}m_x^{*n}(y) := \int_{z\in X} \mathrm{d}m_x^{*(n-1)}(z)\,\mathrm{d}m_z(y)$$

where of course $m_x^{*1} := m_x$. Recall that a measure ν on X is *invariant* for this random walk if $\mathrm{d}\nu(x) = \int_y \mathrm{d}\nu(y)\,\mathrm{d}m_y(x)$. It is *reversible* if moreover, the detailed balance condition $\mathrm{d}\nu(x)\,\mathrm{d}m_x(y) = \mathrm{d}\nu(y)\,\mathrm{d}m_y(x)$ holds.

Other generalizations of Ricci curvature start with a metric measure space [31,46]. Here, as in Bakry–Émery theory, the measure appears as the invariant distribution of some process on the space (e.g. Brownian motion on a Riemannian manifold), which can be chosen in any convenient way. The following remark produces a random walk from a metric measure space, and allows to define the "Ricci curvature at scale ε" for any metric space.

Example 4 *(ε-step random walk).* Let (X, d, μ) be a metric measure space, and assume that balls in X have finite measure and that $\operatorname{Supp}\mu = X$. Choose some $\varepsilon > 0$. The *ε-step random walk* on X, starting at a point x, consists in randomly jumping in the ball of radius ε around x, with probability proportional to μ; namely, $m_x = \mu_{|B(x,\varepsilon)}/\mu(B(x, \varepsilon))$. (One can also use other functions of the distance, such as Gaussian kernels.)

As explained above, when (X, d) is a Riemannian manifold and m_x is the ε-step random walk with small ε, for close enough x, y this definition captures the Ricci curvature in the direction xy (up to some scaling factor depending on ε, see Example 7). In general there is no need for ε to be small: for example if X is a graph, $\varepsilon = 1$ is a natural choice.

If a continuous-time Markov kernel is given, one can also define a continuous-time version of coarse Ricci curvature by setting

$$\kappa(x, y) := -\frac{\mathrm{d}}{\mathrm{d}t}\frac{W_1(m_x^t, m_y^t)}{d(x, y)}$$

when this derivative exists (or take a lim inf), but for simplicity we will mainly work with the discrete-time version here. Indeed, for continuous-time Markov chains, existence of the process is already a non-trivial issue, even in the case of jump processes [12]. We will sometimes use our results on concrete continuous-time examples (e.g. $M/M/\infty$ queues in Section 3.3.4), but only when they appear as an obvious limit of a discrete-time approximation.

One could use the L^p transportation distance instead of the L^1 one in the definition; however, this will make positive curvature a stronger assumption, and is never needed in our theorems.

Notation. In analogy with the Riemannian case, when computing the transportation distance between measures m_x and m_y, we will think of $X \times X$ equipped with the coupling measure as a

tangent space, and for $z \in X \times X$ we will write $x + z$ and $y + z$ for the two projections to X. So in this notation we have

$$\kappa(x, y) = \frac{1}{d(x, y)} \int \big(d(x, y) - d(x + z, y + z)\big)\, \mathrm{d}z$$

where implicitly $\mathrm{d}z$ is the optimal coupling between m_x and m_y.

1.2. Examples

Example 5 *($\mathbb{Z}^N$ and $\mathbb{R}^N$).* Let m be the simple random walk on the graph of the grid $\mathbb{Z}^N$ equipped with its graph metric. Then for any two points $x, y \in \mathbb{Z}^d$, the coarse Ricci curvature along (xy) is 0.

Indeed, we can transport the measure m_x around x to the measure m_y by a translation of vector $y - x$ (and this is optimal), so that the distance between m_x and m_y is exactly that between x and y.

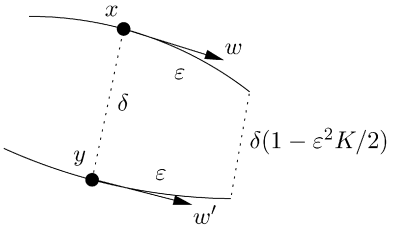
This example generalizes to the case of $\mathbb{Z}^N$ or $\mathbb{R}^N$ equipped with any distance and random walk which are translation-invariant (consistently with [31]). For example, the triangular tiling of the plane has 0 curvature.

We now justify the terminology by showing that, in the case of the ε-step random walk on a Riemannian manifold, we get back the usual Ricci curvature (up to some scaling factor).

Proposition 6. *Let (X, d) be a smooth complete Riemannian manifold. Let v, w be unit tangent vectors at $x \in X$. Let $\varepsilon, \delta > 0$. Let $y = \exp_x \delta v$ and let w' be the tangent vector at y obtained by parallel transport of w along the geodesic* $\exp_x tv$. *Then*

$$d(\exp_x \varepsilon w, \exp_y \varepsilon w') = \delta\left(1 - \frac{\varepsilon^2}{2} K(v, w) + O\big(\varepsilon^3 + \varepsilon^2 \delta\big)\right)$$

as $(\varepsilon, \delta) \to 0$. Here $K(v, w)$ is the sectional curvature in the tangent plane (v, w).



Example 7 *(Riemannian manifold).* Let (X, d) be a smooth complete N-dimensional Riemannian manifold. For some $\varepsilon > 0$, let the Markov chain m^ε be defined by

$$\mathrm{d}m_x^\varepsilon(y) := \frac{1}{\mathrm{vol}(B(x, \varepsilon))}\, \mathrm{d}\,\mathrm{vol}(y)$$

if $y \in B(x, \varepsilon)$, and 0 otherwise.

Let $x \in X$ and let v be a unit tangent vector at x. Let y be a point on the geodesic issuing from v, with $d(x, y)$ small enough. Then

$$\kappa(x, y) = \frac{\varepsilon^2 \operatorname{Ric}(v, v)}{2(N+2)} + O\big(\varepsilon^3 + \varepsilon^2 d(x, y)\big).$$

The proof is postponed to Section 8; it is a refinement of Theorem 1.5 (condition (xii)) in [43], except that therein, the infimum of Ricci curvature is used instead of its value along a tangent vector. Basically, the value of $\kappa(x, y)$ is obtained by averaging Proposition 6 for w in the unit ball of the tangent space at x, which provides an upper bound for κ. The lower bound requires use of the dual characterization of transportation distance (Theorem 1.14 in [47]).

Example 8 *(Discrete cube).* Let $X = \{0, 1\}^N$ be the discrete cube equipped with the Hamming metric (each edge is of length 1). Let m be the lazy random walk on the graph X, i.e. $m_x(x) = 1/2$ and $m_x(y) = 1/2N$ if y is a neighbor of x.

Let $x, y \in X$ be neighbors. Then $\kappa(x, y) = 1/N$.

This examples generalizes to arbitrary binomial distributions (see Section 3.3.3).

Here laziness is necessary to avoid parity problems: if no laziness is introduced, points at odd distance never meet under the random walk; in this case one would have to consider coarse Ricci curvature for points at even distance only.

Actually, since the discrete cube is a 1-geodesic space, one has $\kappa(x, y) \geqslant 1/N$ for any pair $x, y \in X$, not only neighbors (see Proposition 19).

Proof. We can suppose that $x = 00\ldots0$ and $y = 10\ldots0$. For $z \in X$ and $1 \leqslant i \leqslant N$, let us denote by z^i the neighbor of z in which the ith bit is switched. An optimal coupling between m_x and m_y is as follows: for $i \geqslant 2$, move x^i to y^i (both have mass $1/2N$ under m_x and m_y respectively). Now $m_x(x) = 1/2$ and $m_y(x) = 1/2N$, and likewise for y. So it is enough to move a mass $1/2 - 1/2N$ from x to y. All points are moved over a distance 1 by this coupling, except for a mass $1/2N$ which remains at x and a mass $1/2N$ which remains at y, and so the coarse Ricci curvature is at least $1/N$.

Optimality of this coupling is obtained as follows: consider the function $f : X \to \{0, 1\}$ which sends a point of X to its first bit. This is a 1-Lipschitz function, with $f(x) = 0$ and $f(y) = 1$. The expectations of f under m_x and m_y are $1/2N$ and $1 - 1/2N$ respectively, so that $1 - 1/N$ is a lower bound on $W_1(m_x, m_y)$.

A very short but less visual proof can be obtained with the L^1 tensorization property (Proposition 27). □

Example 9 *(Ornstein–Uhlenbeck process).* Let $s \geqslant 0$, $\alpha > 0$ and consider the Ornstein–Uhlenbeck process in $\mathbb{R}^N$ given by the stochastic differential equation

$$\mathrm{d}X_t = -\alpha X_t \,\mathrm{d}t + s \,\mathrm{d}B_t$$

where B_t is a standard N-dimensional Brownian motion. The invariant distribution is Gaussian, of variance $s^2/2\alpha$.

Let $\delta t > 0$ and let the random walk m be the flow at time δt of the process. Explicitly, m_x is a Gaussian probability measure centered at $e^{-\alpha\delta t}x$, of variance $s^2(1 - e^{-2\alpha\delta t})/2\alpha \sim s^2\delta t$ for small δt.

Then the coarse Ricci curvature $\kappa(x, y)$ of this random walk is $1 - e^{-\alpha\delta t}$, for any two $x, y \in \mathbb{R}^N$.

Proof. The transportation distance between two Gaussian distributions with the same variance is the distance between their centers, so that $\kappa(x, y) = 1 - \frac{|e^{-\alpha\delta t}x - e^{-\alpha\delta t}y|}{|x-y|}$. □

Example 10 *(Discrete Ornstein–Uhlenbeck).* Let $X = \{-N, -N+1, \dots, N-1, N\}$ and let m be the random walk on X given by

$$m_k(k) = 1/2, \qquad m_k(k+1) = 1/4 - k/4N, \qquad m_k(k-1) = 1/4 + k/4N$$

which is a lazy random walk with linear drift towards 0. The binomial distribution $\frac{1}{2^{2N}}\binom{2N}{N+k}$ is reversible for this random walk.

Then, for any two neighbors x, y in X, one has $\kappa(x, y) = 1/2N$.

Proof. Exercise. □

Example 11 *(Bakry–Émery).* Let X be an N-dimensional Riemannian manifold and F be a tangent vector field. Consider the differential operator

$$L := \frac{1}{2}\Delta + F.\nabla$$

associated with the stochastic differential equation

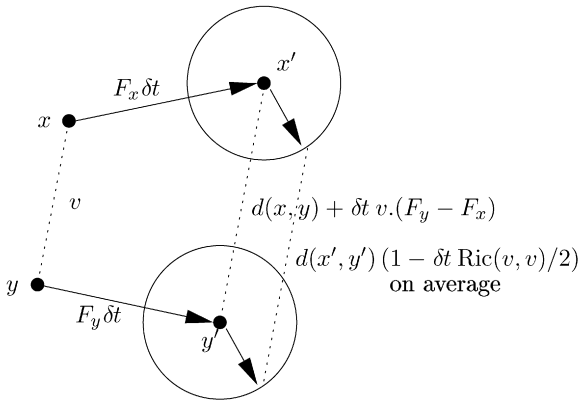
$$\mathrm{d}X_t = F\,\mathrm{d}t + \mathrm{d}B_t$$

where B_t is the Brownian motion in X. The Ricci curvature (in the Bakry–Émery sense) of this operator is $\frac{1}{2}\mathrm{Ric} - \nabla^{\mathrm{sym}}F$ where $\nabla^{\mathrm{sym}}F^{ij} := \frac{1}{2}(\nabla^i F^j + \nabla^j F^i)$ is the symmetrized of ∇F.

Consider the Euler approximation scheme at time δt for this stochastic equation, which consists in following the flow of F for a time δt and then randomly jumping in a ball of radius $\sqrt{(N+2)\delta t}$.

Let $x \in X$ and let v be a unit tangent vector at x. Let y be a point on the geodesic issuing from v, with $d(x, y)$ small enough. Then

$$\kappa(x, y) = \delta t\left(\frac{1}{2}\mathrm{Ric}(v, v) - \nabla^{\mathrm{sym}}F(v, v) + O\big(d(x, y)\big) + O(\sqrt{\delta t})\right).$$



Proof. First let us explain the normalization: jumping in a ball of radius ε generates a variance $\varepsilon^2 \frac{1}{N+2}$ in a given direction. On the other hand, the N-dimensional Brownian motion has, by definition, a variance $\mathrm{d}t$ per unit of time $\mathrm{d}t$ in a given direction, so a proper discretization of Brownian motion at time δt requires jumping in a ball of radius $\varepsilon = \sqrt{(N+2)\delta t}$. Also, as noted in [6], the generator of Brownian motion is $\frac{1}{2}\Delta$ instead of Δ, hence the $\frac{1}{2}$ factor for the Ricci part.

Now the discrete-time process begins by following the flow F for some time δt. Starting at points x and y, using elementary Euclidean geometry, it is easy to see that after this, the distance between the endpoints behaves like $d(x,y)(1+\delta t\; v.\nabla_v F + O(\delta t^2))$. Note that $v.\nabla_v F = \nabla^{\mathrm{sym}} F(v,v)$.

Now, just as in Example 7, randomly jumping in a ball of radius ε results in a gain of $d(x,y)\frac{\varepsilon^2}{2(N+2)}\,\mathrm{Ric}(v,v)$ on transportation distances. Here $\varepsilon^2 = (N+2)\delta t$. So after the two steps of the process, the distance between the endpoints is

$$d(x,y)\left(1 - \frac{\delta t}{2}\,\mathrm{Ric}(v,v) + \delta t\;\nabla^{\mathrm{sym}} F(v,v)\right)$$

as needed, up to higher-order terms. □

Maybe the reason for the additional $-\nabla^{\mathrm{sym}} F$ in Ricci curvature à la Bakry–Émery is made clearer in this context: it is simply the quantity by which the flow of X modifies distances between two starting points.

It is clear on this example why reversibility is not fundamental in this theory: the antisymmetric part of the force F generates an infinitesimal isometric displacement. With our definition, combining the Markov chain with an isometry of the space has no effect whatsoever on curvature.

Example 12 *(Multinomial distribution).* Consider the set $X = \{(x_0, x_1, \dots, x_d),\ x_i \in \mathbb{N}, \sum x_i = N\}$ viewed as the configuration set of N balls in $d+1$ boxes. Consider the process which consists in taking a ball at random among the N balls, removing it from its box, and putting it back at random in one of the $d+1$ boxes. More precisely, the transition probability from $(x_0, \dots, x_d)$ to $(x_0, \dots, x_i - 1, \dots, x_j + 1, \dots, x_d)$ (with maybe $i = j$) is $x_i/N(d+1)$. The multinomial distribution $\frac{N!}{(d+1)^N \prod x_i!}$ is reversible for this Markov chain.

Equip this configuration space with the metric $d((x_i),(x'_i)) := \frac{1}{2}\sum |x_i - x'_i|$ which is the graph distance w.r.t. the moves above. The coarse Ricci curvature of this Markov chain is $1/N$.

Proof. Exercise (see also the discussion after Proposition 27). □

Example 13 *(Geometric distribution).* Let the random walk on $\mathbb{N}$ be defined by the transition probabilities $p_{n,n+1} = 1/3$, $p_{n+1,n} = 2/3$ and $p_{0,0} = 2/3$. This random walk is reversible with respect to the geometric distribution $2^{-(n+1)}$. Then for $n \geqslant 1$ one has $\kappa(n, n+1) = 0$.

Proof. The transition kernel is translation-invariant except at 0. □

Section 5 contains more material about this latter example and how non-negative coarse Ricci curvature sometimes implies exponential concentration.

Example 14 *(Geometric distribution, 2).* Let the random walk on $\mathbb{N}$ be defined by the transition probabilities $p_{n,0} = \alpha$ and $p_{n,n+1} = 1 - \alpha$ for some $0 < \alpha < 1$. The geometric distribution $\alpha(1-\alpha)^n$ is invariant (but not reversible) for this random walk. The coarse Ricci curvature of this random walk is α.

Proof. Exercise. □

Example 15 *(δ-hyperbolic groups).* Let X be the Cayley graph of a non-elementary δ-hyperbolic group with respect to some finite generating set. Let k be a large enough integer (depending on the group) and consider the random walk on X which consists in performing k steps of the simple random walk. Let $x, y \in X$. Then $\kappa(x, y) = -2k/d(x, y)\,(1 + o(1))$ when $d(x, y)$ and k tend to infinity.

Note that $-2k/d(x, y)$ is the smallest possible value for $\kappa(x, y)$, knowing that the steps of the random walk are bounded by k.

Proof. For z in the ball of radius k around x, and z' in the ball of radius k around y, elementary δ-hyperbolic geometry yields $d(z, z') = d(x, y) + d(x, z) + d(y, z') - (y, z)_x - (x, z')_y$ up to some multiple of δ, where $(\cdot,\cdot)$ denotes the Gromov product with respect to some basepoint [24]. Since this decomposes as the sum of a term depending on z only and a term depending on z' only, to compute the transportation distance it is enough to know the expectation of $(y, z)_x$ for z in the ball around x, and likewise for $(x, z')_y$. Using that balls have exponential growth, it is not difficult (see Proposition 21 in [39]) to see that the expectation of $(y, z)_x$ is bounded by a constant, whatever k, hence the conclusion.

The same argument applies to trees or discrete δ-hyperbolic spaces with a uniform lower bound on the exponential growth rate of balls. □

Example 16 *(Kac's random walk on orthogonal matrices, after [38]).* Consider the following random walk on the set of $N \times N$ orthogonal matrices: at each step, a pair of indices $1 \leqslant i < j \leqslant N$ is selected at random, an angle $\theta \in [0; 2\pi)$ is picked at random, and a rotation of angle θ is performed in the coordinate plane i, j. Equip $\mathrm{SO}(N)$ with the Riemannian metric induced by the Hilbert–Schmidt inner product $\mathrm{Tr}(a^* b)$ on its tangent space. It is proven in a preprint by Oliveira [38] that this random walk has coarse Ricci curvature $1 - \sqrt{1 - 2/N(N-1)} \sim 1/N^2$.

This is consistent with the fact that $\mathrm{SO}(N)$ has, as a Riemannian manifold, a positive Ricci curvature in the usual sense. However, from the computational point of view, Kac's random walk above is much nicer than either the Brownian motion or the ε-step random walk of Example 7. Oliveira uses his result to prove a new estimate $O(N^2 \ln N)$ for the mixing time of this random walk, neatly improving on previous estimates $O(N^4 \ln N)$ by Diaconis–Saloff-Coste and $O(N^{2.5} \ln N)$ by Pak–Sidenko; $\Omega(N^2)$ is an easy lower bound, see [38].

Example 17 *(Glauber dynamics for the Ising model).* Let G be a finite graph. Consider the configuration space $X := \{-1, 1\}^G$ together with the energy function

$$U(S) := -\sum_{x \sim y \in G} S(x)S(y) - h\sum_x S(x) \quad \text{for } S \in X,$$

where $h \in \mathbb{R}$ is the external magnetic field. For some $\beta \geqslant 0$, equip X with the Gibbs distribution $\mu := \mathrm{e}^{-\beta U}/Z$ where as usual $Z := \sum_S \mathrm{e}^{-\beta U(S)}$. The distance between two states is defined as the number of vertices of G at which their values differ.

For $S \in X$ and $x \in G$, denote by S_{x+} and S_{x-} the states obtained from S by setting $S_{x+}(x) = +1$ and $S_{x-}(x) = -1$, respectively. Consider the following random walk on X (known as the *Glauber dynamics*): at each step, a vertex $x \in G$ is chosen at random, and a new value for $S(x)$ is picked according to local equilibrium, i.e. $S(x)$ is set to 1 or -1 with probabilities proportional to $\mathrm{e}^{-\beta U(S_{x+})}$ and $\mathrm{e}^{-\beta U(S_{x-})}$ respectively (note that only the neighbors of x influence the ratio of these probabilities). The Gibbs distribution is reversible for this Markov chain.

Then the coarse Ricci curvature of this Markov chain is at least

$$\frac{1}{|G|}\left(1 - v_{\max}\frac{\mathrm{e}^{\beta} - \mathrm{e}^{-\beta}}{\mathrm{e}^{\beta} + \mathrm{e}^{-\beta}}\right)$$

where $v_{\max}$ is the maximal valency of a vertex of G. In particular, if

$$\beta < \frac{1}{2}\ln\left(\frac{v_{\max} + 1}{v_{\max} - 1}\right)$$

then curvature is positive. Consequently, the critical β is at least this quantity.

This estimate for the critical temperature coincides with the one derived in [26]. Actually, our argument generalizes to different settings (such as non-constant/negative values of the coupling J_{xy} between spins, or continuous spin spaces), and the positive curvature condition for the Glauber dynamics exactly amounts to the well-known one-site Dobrushin criterion [18] (or to $G(\beta) < 1$ in the notation of [26, Eq. (19)]). By comparison, the exact value of the critical β for the Ising model on the regular infinite tree of valency v is $\frac{1}{2}\ln(\frac{v}{v-2})$, which shows asymptotic optimality of this criterion. When block dynamics (see [34]) are used instead of single-site updates, positive coarse Ricci curvature of the block dynamics Markov chain is equivalent to the Dobrushin–Shlosman criterion [20].

As shown in the rest of this paper, positive curvature implies several properties, especially, exponential convergence to equilibrium, concentration inequalities and a modified logarithmic Sobolev inequality. For the Glauber dynamics, the constants we get in these inequalities are essentially the same as in the infinite-temperature (independent) case, up to some factor depending

on temperature which diverges when positive curvature ceases to hold. This is more or less equivalent to the main results of the literature under the Dobrushin–Shlosman criterion (see e.g. the review [34]). Note however that in our setting we do not need the underlying graph to be $\mathbb{Z}^N$.

Proof. Using Proposition 19, it is enough to bound coarse Ricci curvature for pairs of states at distance 1. Let S, S' be two states differing only at $x \in G$. We can suppose that $S(x) = -1$ and $S'(x) = 1$. Let m_S and $m_{S'}$ be the law of the step of the random walk issuing from S and S' respectively. We have to prove that the transportation distance between m_S and $m_{S'}$ is at most $1 - \frac{1}{|G|}(1 - v_{\max}\frac{e^{\beta}-e^{-\beta}}{e^{\beta}+e^{-\beta}})$.

The measure m_S decomposes as $m_S = \frac{1}{|G|}\sum_{y\in G} m_S^y$, according to the vertex $y \in G$ which is modified by the random walk, and likewise for $m_{S'}$. To evaluate the transportation distance, we will compare m_S^y to $m_{S'}^y$.

If the step of the random walk consists in modifying the value of S at x (which occurs with probability $\frac{1}{|G|}$), then the resulting state has the same law for S and S', i.e. $m_S^x = m_{S'}^x$. Thus in this case the transportation distance is 0 and the contribution to coarse Ricci curvature is $1 \times \frac{1}{|G|}$.

If the step consists in modifying the value of S at some point y in G not adjacent to x, then the value at x does not influence local equilibrium at y, and so m_S^y and $m_{S'}^y$ are identical except at x. So in this case the distance is 1 and the contribution to coarse Ricci curvature is 0.

Now if the step consists in modifying the value of S at some point $y \in G$ adjacent to x (which occurs with probability $v_x/|G|$ where v_x is the valency of x), then the value at x does influence the law of the new value at y, by some amount which we now evaluate. The final distance between the two laws will be this amount plus 1 (1 accounts for the difference at x), and the contribution to coarse Ricci curvature will be negative.

Let us now evaluate this amount more precisely. Let $y \in G$ be adjacent to x. Set $a = e^{-\beta U(S_{y+})}/e^{-\beta U(S_{y-})}$. The step of the random walk consists in setting $S(y)$ to 1 with probability $\frac{a}{a+1}$, and to -1 with probability $\frac{1}{a+1}$. Setting likewise $a' = e^{-\beta U(S'_{y+})}/e^{-\beta U(S'_{y-})}$ for S', we are left to evaluate the distance between the distributions on $\{-1, 1\}$ given by $(\frac{a}{a+1}; \frac{1}{a+1})$ and $(\frac{a'}{a'+1}; \frac{1}{a'+1})$. It is immediate to check, using the definition of the energy U, that $a' = e^{4\beta}a$. Then, a simple computation shows that the distance between these two distributions is at most $\frac{e^{\beta}-e^{-\beta}}{e^{\beta}+e^{-\beta}}$. This value is actually achieved when y has odd valency, $h = 0$ and switching the value at x changes the majority of spin signs around y. (Our argument is suboptimal here when valency is even—a more precise estimation yields the absence of a phase transition on $\mathbb{Z}$.)

Combining these different cases yields the desired curvature evaluation. To convert this into an evaluation of the critical β, reason as follows: magnetization, defined as $\frac{1}{|G|}\sum_{x\in G} S(x)$, is a $\frac{1}{|G|}$-Lipschitz function of the state. Now let μ_0 be the Gibbs measure without magnetic field, and μ_h the Gibbs measure with external magnetic field h. Use the Glauber dynamics with magnetic field h, but starting with an initial state picked under μ_0; Corollary 22 yields that the magnetization under μ_h is controlled by $\frac{1}{|G|}W_1(\mu_0, \mu_0 * m)/\kappa$ where κ is the coarse Ricci curvature, and $W_1(\mu_0, \mu_0 * m)$ is the transportation distance between the Gibbs measure μ_0 and the measure obtained from it after one step of the Glauber dynamics with magnetic field h; reasoning as above this transportation distance is easily bounded by $\frac{1}{|G|}\frac{e^{\beta h}-e^{-\beta h}}{e^{\beta h}+e^{-\beta h}}$, so that the derivative of magnetization w.r.t. h stays bounded when $|G| \to \infty$, which is the classical criterion used to define critical temperature. (Compare Eq. (22) in [26].) □

Further examples. More examples can be found in Sections 3.3.3 (binomial and Poisson distributions), 3.3.4 ($M/M/\infty$ queues and generalizations), 3.3.5 (exponential tails), 3.3.6 (heavy tails) and 5 (geometric distributions on $\mathbb{N}$, exponential distributions on $\mathbb{R}^N$).

1.3. Overview of the results

Notation for random walks. Before we present the main results, we need to define some quantities related to the local behavior of the random walk: the *jump*, which will help control the diameter of the space, and the *coarse diffusion constant*, which controls concentration properties. Moreover, we define a notion of local dimension. The larger the dimension, the better for concentration of measure.

Definition 18 *(Jump, diffusion constant, dimension).* Let the *jump* of the random walk at x be

$$J(x) := \mathbb{E}_{m_x} d(x, \cdot) = W_1(\delta_x, m_x).$$

Let the *(coarse) diffusion constant* of the random walk at x be

$$\sigma(x) := \left(\frac{1}{2} \iint d(y,z)^2 \, \mathrm{d}m_x(y) \, \mathrm{d}m_x(z) \right)^{1/2}$$

and, if ν is an invariant distribution, let

$$\sigma := \left\| \sigma(x) \right\|_{L^2(X,\nu)}$$

be the average diffusion constant.

Let also $\sigma_\infty(x) := \frac{1}{2} \operatorname{diam} \operatorname{Supp} m_x$ and $\sigma_\infty := \sup \sigma_\infty(x)$.

Let the *local dimension* at x be

$$n_x := \frac{\sigma(x)^2}{\sup\{\operatorname{Var}_{m_x} f, f : \operatorname{Supp} m_x \to \mathbb{R} \text{ 1-Lipschitz}\}}$$

and finally $n := \inf_x n_x$.

About this definition of dimension. Obviously $n_x \geqslant 1$. For the discrete-time Brownian motion on a N-dimensional Riemannian manifold, one has $n_x \approx N$ (see the end of Section 8). For the simple random walk on a graph, $n_x \approx 1$. This definition of dimension amounts to saying that in a space of dimension n, the typical variations of a (1-dimensional) Lipschitz function are $1/\sqrt{n}$ times the typical distance between two points. This is the case in the sphere S^n, in the Gaussian measure on $\mathbb{R}^n$, and in the discrete cube $\{0,1\}^n$. So generally one could define the "statistical dimension" of a metric measure space (X, d, μ) by this formula i.e.

$$\operatorname{StatDim}(X, d, \mu) := \frac{\frac{1}{2} \iint d(x,y)^2 \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)}{\sup\{\operatorname{Var}_\mu f, f \text{ 1-Lipschitz}\}}$$

so that for each $x \in X$ the local dimension of X at x is $n_x = \operatorname{StatDim}(X, d, m_x)$. With this definition, $\mathbb{R}^N$ equipped with a Gaussian measure has statistical dimension N and local dimension $\approx N$, whereas the discrete cube $\{0,1\}^N$ has statistical dimension $\approx N$ and local dimension ≈ 1.

We now turn to the description of the main results of the paper.

Elementary properties. In Section 2 are gathered some straightforward results.

First, we prove (Proposition 19) that in an ε-geodesic space, a lower bound on $\kappa(x, y)$ for points x, y with $d(x, y) \leqslant \varepsilon$ implies the same lower bound for all pairs of points. This is simple yet very useful: indeed in the various graphs given above as examples, it was enough to compute the coarse Ricci curvature for neighbors.

Second, we prove equivalent characterizations of having coarse Ricci curvature uniformly bounded below: a space satisfies $\kappa(x, y) \geqslant \kappa$ if and only if the random walk operator is $(1 - \kappa)$-contracting on the space of probability measures equipped with the transportation distance (Proposition 20), and if and only if the random walk operator acting on Lipschitz functions contracts the Lipschitz norm by $(1-\kappa)$ (Proposition 29). An immediate corollary is the existence of a unique invariant distribution when $\kappa > 0$.

The property of contraction of the Lipschitz norm easily implies, in the reversible case, that the spectral gap of the Laplacian operator associated with the random walk is at least κ (Proposition 30); this can be seen as a generalization of Lichnerowicz's theorem, and provides sharp estimates of the spectral gap in several examples. (A similar result appears in [14].)

In analogy with the Bonnet–Myers theorem, we prove that if coarse Ricci curvature is bounded below by $\kappa > 0$, then the diameter of the space is at most $2 \sup_x J(x)/\kappa$ (Proposition 23). In case J is unbounded, we can evaluate instead the average distance to a given point x_0 under the invariant distribution ν (Proposition 24); namely, $\int d(x_0, y)\, \mathrm{d}\nu(y) \leqslant J(x_0)/\kappa$. In particular we have $\int d(x, y)\, \mathrm{d}\nu(x)\, \mathrm{d}\nu(y) \leqslant 2 \inf J/\kappa$. These are L^1 versions of the Bonnet–Myers theorem rather than generalizations: from the case of manifolds one would expect $1/\sqrt{\kappa}$ instead of $1/\kappa$. Actually this L^1 version is sharp in all our examples except Riemannian manifolds; in Section 6 we investigate additional conditions for an L^2 version of the Bonnet–Myers theorem to hold.

Let us also mention some elementary operations preserving positive curvature: composition, superposition and L^1 tensorization (Propositions 25–27).

Concentration results. Basically, if coarse Ricci curvature is bounded below by $\kappa > 0$, then the invariant distribution satisfies concentration results with variance $\sigma^2/n\kappa$ (up to some constant factor). This estimate is often sharp, as discussed in Section 3.3 where we revisit some of the examples.

However, the type of concentration (Gaussian, exponential, or $1/t^2$) depends on further local assumptions: indeed, the tail behavior of the invariant measure cannot be better than that of the local measures m_x. Without further assumptions, one only gets that the variance of a 1-Lipschitz function is at most $\sigma^2/n\kappa$, hence concentration like $\sigma^2/n\kappa t^2$ (Proposition 32). If we make the further assumption that the support of the measures m_x is uniformly bounded (i.e. $\sigma_\infty < \infty$), then we get mixed Gaussian-then-exponential concentration, with variance $\sigma^2/n\kappa$ (Theorem 33). The width of the Gaussian window depends on σ_∞, and on the rate of variation of the diffusion constant $\sigma(x)^2$.

For the case of Riemannian manifolds, simply considering smaller and smaller steps for the random walks makes the width of the Gaussian window tend to infinity, so that we recover full Gaussian concentration as in the Lévy–Gromov or Bakry–Émery context. However, for lots of discrete examples, the Gaussian-then-exponential behavior is genuine. Examples where tails are Poisson-like (binomial distribution, $M/M/\infty$ queues) or exponential are given in Sections 3.3.3 to 3.3.5. Examples of heavy tails (when $\sigma_\infty = \infty$) are given in 3.3.6.

We also get concentration results for the finite-time distributions m_x^{*k} (Remark 35).

Log-Sobolev inequality. Using a suitable non-local notion of norm of the gradient, we are able to adapt the proof by Bakry and Émery of a logarithmic Sobolev inequality for the invariant distribution. The gradient we use (Definition 41) is $(\mathrm{D}f)(x) := \sup_{y,z} \frac{|f(y)-f(z)|}{d(y,z)} \exp(-\lambda d(x,y) - \lambda d(y,z))$. This is a kind of "semi-local" Lipschitz constant for f. Typically the value of λ can be taken large at the "macroscopic" level; for Riemannian manifolds, taking smaller and smaller steps for the random walk allows to take $\lambda \to \infty$ so that we recover the usual gradient for smooth functions.

The inequality takes the form $\mathrm{Ent}\, f \leqslant C \int (\mathrm{D}f)^2/f \,\mathrm{d}\nu$ (Theorem 45). The main tool of the proof is the gradient contraction relation $\mathrm{D}(\mathrm{M}f) \leqslant (1-\kappa/2)\mathrm{M}(\mathrm{D}f)$ where M is the random walk operator (Theorem 44).

That the gradient is non-local, with a maximal possible value of λ, is consistent with the possible occurrence of non-Gaussian tails.

Exponential concentration and non-negative curvature. The simplest example of a Markov chain with zero coarse Ricci curvature is the simple random walk on $\mathbb{N}$ or $\mathbb{Z}$, for which there is no invariant distribution. However, we show that if furthermore there is a "locally attracting" point, then non-negative coarse Ricci curvature implies exponential concentration. Examples are the geometric distribution on $\mathbb{N}$ or the exponential distribution $\mathrm{e}^{-|x|}$ on $\mathbb{R}^N$ associated with the stochastic differential equation $\mathrm{d}X_t = \mathrm{d}B_t - \frac{X_t}{|X_t|}\mathrm{d}t$. In both cases we recover correct orders of magnitude.

Gromov–Hausdorff topology. One advantage of our definition is that it involves only combinations of the distance function, and no derivatives, so that it is more or less impervious to deformations of the space. In Section 7 we show that coarse Ricci curvature is continuous for Gromov–Hausdorff convergence of metric spaces (suitably reinforced, of course, so that the random walk converges as well), so that having non-negative curvature is a closed property. We also suggest a loosened definition of coarse Ricci curvature, requiring that $W_1(m_x, m_y) \leqslant (1-\kappa)d(x,y) + \delta$ instead of $W_1(m_x, m_y) \leqslant (1-\kappa)d(x,y)$. With this definition, positive curvature becomes an *open* property, so that a space close to one with positive curvature has positive curvature.

2. Elementary properties

2.1. Geodesic spaces

The idea behind curvature is to use local properties to derive global ones. We give here a simple proposition expressing that in near-geodesic spaces such as graphs (with $\varepsilon = 1$) or manifolds (for any ε), it is enough to check positivity of coarse Ricci curvature for nearby points.

Proposition 19 *(Geodesic spaces). Suppose that (X,d) is ε-geodesic in the sense that for any two points $x, y \in X$, there exists an integer n and a sequence $x_0 = x, x_1, \dots, x_n = y$ such that $d(x,y) = \sum d(x_i, x_{i+1})$ and $d(x_i, x_{i+1}) \leqslant \varepsilon$.*

Then, if $\kappa(x,y) \geqslant \kappa$ for any pair of points with $d(x,y) \leqslant \varepsilon$, then $\kappa(x,y) \geqslant \kappa$ for any pair of points $x, y \in X$.

Proof. Let (x_i) be as above. Using the triangle inequality for W_1, one has $W_1(m_x, m_y) \leqslant \sum W_1(m_{x_i}, m_{x_{i+1}}) \leqslant (1-\kappa)\sum d(x_i, x_{i+1}) = (1-\kappa)d(x,y)$. □

2.2. Contraction on the space of probability measures

Let $\mathcal{P}(X)$ by the space of all probability measures μ on X with finite first moment, i.e. for some (hence any) $o \in X$, $\int d(o,x)\,\mathrm{d}\mu(x) < \infty$. On $\mathcal{P}(X)$, the transportation distance W_1 is finite, so that it is actually a distance.

Let μ be a probability measure on X and define the measure

$$\mu * m := \int_{x\in X} \mathrm{d}\mu(x)\, m_x$$

which is the image of μ by the random walk. A priori, it may or may not belong to $\mathcal{P}(X)$.

The following proposition and its corollary can be seen as a particular case of Theorem 3 in [18] (viewing a Markov chain as a Markov field on $\mathbb{N}$). Equivalent statements also appear in [19, Proposition 14.3], in the second edition of [12, Theorem 5.22], in [15] (in the proof of Proposition 2.10), in [42] and in [38].

Proposition 20 *(W_1 contraction). Let (X,d,m) be a metric space with a random walk. Let $\kappa \in \mathbb{R}$. Then we have $\kappa(x,y) \geqslant \kappa$ for all $x,y \in X$, if and only if for any two probability distributions $\mu,\mu' \in \mathcal{P}(X)$ one has*

$$W_1(\mu * m, \mu' * m) \leqslant (1-\kappa) W_1(\mu,\mu').$$

Moreover in this case, if $\mu \in \mathcal{P}(X)$ then $\mu * m \in \mathcal{P}(X)$.

Proof. First, suppose that convolution with m is contracting in W_1 distance. For some $x,y \in X$, let $\mu = \delta_x$ and $\mu' = \delta_y$ be the Dirac measures at x and y. Then by definition $\delta_x * m = m_x$ and likewise for y, so that $W_1(m_x,m_y) \leqslant (1-\kappa) W_1(\delta_x,\delta_y) = (1-\kappa) d(x,y)$ as required.

The converse is more difficult to write than to understand. For each pair (x,y) let ξ_{xy} be a coupling (i.e. a measure on $X\times X$) between m_x and m_y witnessing for $\kappa(x,y) \geqslant \kappa$. According to Corollary 5.22 in [48], we can choose ξ_{xy} to depend measurably on the pair (x,y).

Let Ξ be a coupling between μ and μ' witnessing for $W_1(\mu,\mu')$. Then $\int_{X\times X} \mathrm{d}\Xi(x,y)\,\xi_{xy}$ is a coupling between $\mu * m$ and $\mu' * m$ and so

$$\begin{aligned}
W_1(\mu * m, \mu' * m) &\leqslant \int_{x,y} d(x,y)\,\mathrm{d}\left\{ \int_{x',y'} \mathrm{d}\Xi(x',y')\,\xi_{x'y'} \right\}(x,y) \\
&= \int_{x,y,x',y'} \mathrm{d}\Xi(x',y')\,\mathrm{d}\xi_{x'y'}(x,y)\, d(x,y) \\
&\leqslant \int_{x',y'} \mathrm{d}\Xi(x',y')\, d(x',y') \big(1-\kappa(x',y')\big) \\
&\leqslant (1-\kappa) W_1(\mu,\mu')
\end{aligned}$$

by the Fubini theorem applied to $d(x,y)\,\mathrm{d}\Xi(x',y')\,\mathrm{d}\xi_{x',y'}(x,y)$.

To see that in this situation $\mathcal{P}(X)$ is preserved by the random walk, fix some origin $o \in X$ and note that for any $\mu \in \mathcal{P}(X)$, the first moment of $\mu * m$ is $W_1(\delta_o, \mu * m) \leqslant W_1(\delta_o, m_o) + W_1(m_o, \mu * m) \leqslant W_1(\delta_o, m_o) + (1-\kappa) W_1(\delta_o, \mu)$. Now $W_1(\delta_o, \mu) < \infty$ by assumption, and $W_1(\delta_o, m_o) < \infty$ by Definition 1. □

As an immediate consequence of this contracting property we get:

Corollary 21 *(W_1 convergence). Suppose that $\kappa(x, y) \geqslant \kappa > 0$ for any two distinct $x, y \in X$. Then the random walk has a unique invariant distribution $\nu \in \mathcal{P}(X)$.*

Moreover, for any probability measure $\mu \in \mathcal{P}(X)$, the sequence $\mu * m^{*n}$ tends exponentially fast to ν in W_1 distance. Namely

$$W_1\big(\mu * m^{*n}, \nu\big) \leqslant (1-\kappa)^n W_1(\mu, \nu)$$

and in particular

$$W_1\big(m_x^{*n}, \nu\big) \leqslant (1-\kappa)^n J(x)/\kappa.$$

The last assertion follows by taking $\mu = \delta_x$ and noting that $J(x) = W_1(\delta_x, m_x)$ so that $W_1(\delta_x, \nu) \leqslant W_1(\delta_x, m_x) + W_1(m_x, \nu) \leqslant J(x) + (1-\kappa) W_1(\delta_x, \nu)$, hence $W_1(\delta_x, \nu) \leqslant J(x)/\kappa$.

This is useful to provide bounds on mixing time. For example, suppose that X is a graph; since the total variation distance between two measures μ, μ' is the transportation distance with respect to the trivial metric instead of the graph metric, we obviously have $|\mu - \mu'|_{\mathrm{TV}} \leqslant W_1(\mu, \mu')$, hence the corollary above yields the estimate $|m_x^{*t} - \nu|_{\mathrm{TV}} \leqslant (\operatorname{diam} X)(1-\kappa)^t$ for any $x \in X$. Applied for example to the discrete cube $\{0, 1\}^N$, with $\kappa = 1/N$ and diameter N, this gives the correct estimate $O(N \ln N)$ for mixing time in total variation distance, whereas the traditional estimate based on spectral gap and passage from L^2 to L^1 norm gives $O(N^2)$. Also note that the pointwise bound $|m_x^{*t} - \nu|_{\mathrm{TV}} \leqslant (1-\kappa)^t J(x)/\kappa$ depends on local data only and requires no knowledge of the invariant measure (compare [21]) or diameter; in particular it applies to infinite graphs.

Another immediate interesting corollary is the following, which allows to estimate the average of a Lipschitz function under the invariant measure, knowing some of its values. This is useful in concentration theorems, to get bounds not only on the deviations from the average, but on what the average actually is.

Corollary 22. *Suppose that $\kappa(x, y) \geqslant \kappa > 0$ for any two distinct $x, y \in X$. Let ν be the invariant distribution.*

Let f be a 1-Lipschitz function. Then, for any distribution μ, one has $|\mathbb{E}_\nu f - \mathbb{E}_\mu f| \leqslant W_1(\mu, \mu * m)/\kappa$.

In particular, for any $x \in X$ one has $|f(x) - \mathbb{E}_\nu f| \leqslant J(x)/\kappa$.

Proof. One has $W_1(\mu * m, \nu) \leqslant (1-\kappa) W_1(\mu, \nu)$. Since by the triangle inequality, $W_1(\mu * m, \nu) \geqslant W_1(\mu, \nu) - W_1(\mu, \mu * m)$, one gets $W_1(\mu, \nu) \leqslant W_1(\mu, \mu * m)/\kappa$. Now if f is a 1-Lipschitz function, for any two distributions μ, μ' one has $|\mathbb{E}_\mu f - \mathbb{E}_{\mu'} f| \leqslant W_1(\mu, \mu')$ hence the result.

The last assertion is simply the case when μ is the Dirac measure at x. □

2.3. L^1 Bonnet–Myers theorems

We now give a weak analogue of the Bonnet–Myers theorem. This result shows in particular that positivity of coarse Ricci curvature is a much stronger property than a spectral gap bound: there is no coarse Ricci curvature analogue of a family of expanders.

Proposition 23 *(L^1 Bonnet–Myers). Suppose that $\kappa(x,y) \geqslant \kappa > 0$ for all $x, y \in X$. Then for any $x, y \in X$ one has*

$$d(x,y) \leqslant \frac{J(x)+J(y)}{\kappa(x,y)}$$

and in particular

$$\operatorname{diam} X \leqslant \frac{2\sup_x J(x)}{\kappa}.$$

Proof. We have $d(x,y) = W_1(\delta_x,\delta_y) \leqslant W_1(\delta_x,m_x) + W_1(m_x,m_y) + W_1(m_y,\delta_y) \leqslant J(x) + (1-\kappa)d(x,y) + J(y)$ hence the result. □

This estimate is not sharp at all for Brownian motion in Riemannian manifolds (since $J \approx \varepsilon$ and $\kappa \approx \varepsilon^2 \operatorname{Ric}/N$, it fails by a factor $1/\varepsilon$ compared to the Bonnet–Myers theorem!), but is sharp in many other examples.

For the discrete cube $X = \{0,1\}^N$ (Example 8 above), one has $J = 1/2$ and $\kappa = 1/N$, so we get $\operatorname{diam} X \leqslant N$ which is the exact value.

For the discrete Ornstein–Uhlenbeck process (Example 10 above) one has $J = 1/2$ and $\kappa = 1/2N$, so we get $\operatorname{diam} X \leqslant 2N$ which once more is the exact value.

For the continuous Ornstein–Uhlenbeck process on $\mathbb{R}$ (Example 9 with $N = 1$), the diameter is infinite, consistently with the fact that J is unbounded. If we consider points x, y lying in some large interval $[-R; R]$ with $R \gg s/\sqrt{\alpha}$, then $\sup J \sim \alpha R \delta t$ on this interval, and $\kappa = (1 - \mathrm{e}^{\alpha\delta t}) \sim \alpha\delta t$ so that the diameter bound is $2R$, which is correct.

These examples show that one cannot replace J/κ with $J/\sqrt{\kappa}$ in this result (as could be expected from the example of Riemannian manifolds). In fact, Riemannian manifolds seem to be the only simple example where there is a diameter bound behaving like $1/\sqrt{\kappa}$. In Section 6 we investigate conditions under which an L^2 version of the Bonnet–Myers theorem holds.

In case J is not bounded, we can estimate instead the "average" diameter $\int d(x,y)\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y)$ under the invariant distribution ν. This estimate will prove very useful in several examples, to get bounds on the average of $\sigma(x)$ in cases where $\sigma(x)$ is unbounded but controlled by the distance to some "origin" (see e.g. Sections 3.3.4 and 3.3.5).

Proposition 24 *(Average L^1 Bonnet–Myers). Suppose that $\kappa(x,y) \geqslant \kappa > 0$ for any two distinct $x, y \in X$. Then for any $x \in X$,*

$$\int_X d(x,y)\,\mathrm{d}\nu(y) \leqslant \frac{J(x)}{\kappa}$$

and so

$$\int_{X\times X} d(x,y)\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y) \leqslant \frac{2\inf_x J(x)}{\kappa}.$$

Proof. The first assertion follows from Corollary 22 with $f = d(x,\cdot)$.

For the second assertion, choose an x_0 with $J(x_0)$ arbitrarily close to $\inf J$, and write

$$\begin{aligned}\int_{X\times X} d(y,z)\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(z) &\leqslant \int_{X\times X} \big(d(y,x_0)+d(x_0,z)\big)\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(z)\\ &= 2W_1(\delta_{x_0},\nu) \leqslant 2J(x_0)/\kappa\end{aligned}$$

which ends the proof. □

2.4. Three constructions

Here we describe three very simple operations which trivially preserve positive curvature, namely, composition, superposition and L^1 tensorization.

Proposition 25 *(Composition). Let X be a metric space equipped with two random walks $m = (m_x)_{x\in X}$, $m' = (m'_x)_{x\in X}$. Suppose that the coarse Ricci curvature of m (resp. m') is at least κ (resp. κ'). Let m'' be the composition of m and m', i.e. the random walk which sends a probability measure μ to $\mu * m * m'$. Then the coarse Ricci curvature of m'' is at least $\kappa + \kappa' - \kappa\kappa'$.*

Proof. Trivial when $(1-\kappa)$ is seen as a contraction coefficient. □

Superposition states that if we are given two random walks on the same space and construct a new one by, at each step, tossing a coin and deciding to follow either one random walk or the other, then the coarse Ricci curvatures mix nicely.

Proposition 26 *(Superposition). Let X be a metric space equipped with a family $(m^{(i)})$ of random walks. Suppose that for each i, the coarse Ricci curvature of $m^{(i)}$ is at least κ_i. Let (α_i) be a family of non-negative real numbers with $\sum \alpha_i = 1$. Define a random walk m on X by $m_x := \sum \alpha_i m_x^{(i)}$. Then the coarse Ricci curvature of m is at least $\sum \alpha_i\kappa_i$.*

Proof. Let $x, y \in X$ and for each i let ξ_i be a coupling between $m_x^{(i)}$ and $m_y^{(i)}$. Then $\sum \alpha_i\xi_i$ is a coupling between $\sum \alpha_i m_x^{(i)}$ and $\sum \alpha_i m_y^{(i)}$, so that

$$\begin{aligned}W_1(m_x,m_y) &\leqslant \sum \alpha_i W_1\big(m_x^{(i)}, m_y^{(i)}\big)\\ &\leqslant \sum \alpha_i(1-\kappa_i)d(x,y)\\ &= \Big(1-\sum \alpha_i\kappa_i\Big)d(x,y).\end{aligned}$$

Note that the coupling above, which consists in sending each $m_x^{(i)}$ to $m_y^{(i)}$, has no reason to be optimal, so that in general equality does not hold. □

Tensorization states that if we perform a random walk in a product space by deciding at random, at each step, to move in one or the other component, then positive curvature is preserved.

Proposition 27 *(L^1 tensorization). Let $((X_i, d_i))_{i\in I}$ be a finite family of metric spaces and suppose that X_i is equipped with a random walk $m^{(i)}$. Let X be the product of the spaces X_i, equipped with the distance $d := \sum d_i$. Let (α_i) be a family of non-negative real numbers with $\sum \alpha_i = 1$. Consider the random walk m on X defined by*

$$m_{(x_1,\dots,x_k)} := \sum \alpha_i \delta_{x_1} \otimes \cdots \otimes m_{x_i} \otimes \cdots \otimes \delta_{x_k}.$$

Suppose that for each i, the coarse Ricci curvature of $m^{(i)}$ is at least κ_i. Then the coarse Ricci curvature of m is at least $\inf \alpha_i \kappa_i$.

For example, this allows for a very short proof that the curvature of the lazy random walk on the discrete cube $\{0,1\}^N$ is $1/N$ (Example 8). Indeed, it is the N-fold product of the random walk on $\{0,1\}$ which sends each point to the equilibrium distribution $(1/2, 1/2)$, hence is of curvature 1.

Likewise, we can recover the coarse Ricci curvature for multinomial distributions (Example 12) as follows: consider a finite set S of cardinal $d+1$, representing the boxes of Example 12, endowed with an arbitrary probability distribution ν. Equip it with the trivial distance and the Markov chain sending each point of S to ν, so that coarse Ricci curvature is 1. Now consider the N-fold product of this random walk on S^N. Each component represents a ball of Example 12, and the product random walk consists in selecting a ball and putting it in a random box according to ν, as in the example. By the proposition above, the coarse Ricci curvature of this N-fold product is (at least) $1/N$. This evaluation of curvature carries down to the "quotient" Markov chain of Example 12, in which only the number of balls in each box is considered instead of the full configuration space.

The case when some α_i is equal to 0 shows why coarse Ricci curvature is given by an infimum: indeed, if $\alpha_i = 0$ then the corresponding component never gets mixed, hence curvature cannot be positive (unless this component is reduced to a single point). This is similar to what happens for the spectral gap.

The statement above is restricted to a finite product for the following technical reasons: first, to define the L^1 product of an infinite family, a basepoint has to be chosen. Second, in order for the formula above to define a random walk with finite first moment (see Definition 1), some uniform assumption on the first moments of the $m^{(i)}$ is needed.

Proof. For $x \in X$ let $\tilde{m}_x^{(i)}$ stand for $\delta_{x_1} \otimes \cdots \otimes m_{x_i} \otimes \cdots \otimes \delta_{x_k}$.

Let $x = (x_i)$ and $y = (y_i)$ be two points in X. Then

$$\begin{aligned} W_1(m_x, m_y) &\leqslant \sum \alpha_i W_1\big(\tilde{m}_x^{(i)}, \tilde{m}_y^{(i)}\big) \\ &\leqslant \sum \alpha_i \left(W_1\big(m_x^{(i)}, m_y^{(i)}\big) + \sum_{j\neq i} d_j(x_j, y_j) \right) \end{aligned}$$

$$\begin{aligned}
&\leqslant \sum \alpha_i \left((1-\kappa_i) d_i(x_i, y_i) + \sum_{j \neq i} d_j(x_j, y_j) \right) \\
&= \sum \alpha_i \left(-\kappa_i d_i(x_i, y_i) + \sum d_j(x_j, y_j) \right) \\
&= \sum d_i(x_i, y_i) - \sum \alpha_i \kappa_i d_i(x_i, y_i) \\
&\leqslant (1 - \inf \alpha_i \kappa_i) \sum d_i(x_i, y_i) \\
&= (1 - \inf \alpha_i \kappa_i) d(x, y). \qquad \square
\end{aligned}$$

2.5. Lipschitz functions and spectral gap

Definition 28 *(Averaging operator, Laplacian).* For $f \in L^2(X, \nu)$ let the averaging operator M be

$$\mathrm{M} f(x) := \int_y f(y) \, \mathrm{d}m_x(y)$$

and let $\Delta := \mathrm{M} - \mathrm{Id}$.

(This is the layman's convention for the sign of the Laplacian, i.e. $\Delta = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ on $\mathbb{R}$, so that on a Riemannian manifold Δ is a negative operator.)

The following proposition also appears in [15] (in the proof of Proposition 2.10). For the classical case of Riemannian manifolds, contraction of the norm of the gradient is one of the main results of Bakry–Émery theory.

Proposition 29 *(Lipschitz contraction). Let (X, d, m) be a random walk on a metric space. Let $\kappa \in \mathbb{R}$.*

Then the coarse Ricci curvature of X is at least κ, if and only if, for every k-Lipschitz function $f : X \to \mathbb{R}$, the function $\mathrm{M} f$ is $k(1-\kappa)$-Lipschitz.

Proof. First, suppose that the coarse Ricci curvature of X is at least κ. Then, using the notation presented at the end of Section 1.1, we have

$$\begin{aligned}
\mathrm{M} f(y) - \mathrm{M} f(x) &= \int_z f(y+z) - f(x+z) \\
&\leqslant k \int_z d(x+z, y+z) \\
&= k d(x, y) \big(1 - \kappa(x, y)\big).
\end{aligned}$$

Conversely, suppose that whenever f is 1-Lipschitz, $\mathrm{M} f$ is $(1-\kappa)$-Lipschitz. The duality theorem for transportation distance (Theorem 1.14 in [47]) states that

$$\begin{aligned} W_1(m_x, m_y) &= \sup_{f\ 1\text{-Lipschitz}} \int f \,\mathrm{d}(m_x - m_y) \\ &= \sup_{f\ 1\text{-Lipschitz}} \mathrm{M}f(x) - \mathrm{M}f(y) \\ &\leqslant (1-\kappa) d(x,y). \qquad \square \end{aligned}$$

Let ν be an invariant distribution of the random walk. Consider the space $L^2(X,\nu)/\{\text{const}\}$ equipped with the norm

$$\|f\|^2_{L^2(X,\nu)/\{\text{const}\}} := \|f - \mathbb{E}_\nu f\|^2_{L^2(X,\nu)} = \mathrm{Var}_\nu f = \frac{1}{2} \int\limits_{X\times X} \big(f(x) - f(y)\big)^2 \,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y).$$

The operators M and Δ are self-adjoint in $L^2(X,\nu)$ if and only if ν is reversible for the random walk.

It is easy to check, using associativity of variances, that

$$\mathrm{Var}_\nu f = \int \mathrm{Var}_{m_x} f \,\mathrm{d}\nu(x) + \mathrm{Var}_\nu \mathrm{M}f$$

so that $\|\mathrm{M}f\|_2 \leqslant \|f\|_2$. It is also clear that $\|\mathrm{M}f\|_\infty \leqslant \|f\|_\infty$.

Usually, spectral gap properties for Δ are expressed in the space L^2. The proposition above only implies that the spectral radius of the operator M acting on $\mathrm{Lip}(X)/\{\text{const}\}$ is at most $(1-\kappa)$. In general it is not true that a bound for the spectral radius of an operator on a dense subspace of a Hilbert space implies a bound for the spectral radius on the whole space. This holds, however, when the operator is self-adjoint or when the Hilbert space is finite-dimensional.

Proposition 30 *(Spectral gap). Let (X,d,m) be a metric space with random walk, with invariant distribution ν. Suppose that the coarse Ricci curvature of X is at least $\kappa > 0$ and that $\sigma < \infty$. Suppose that ν is reversible, or that X is finite.*

Then the spectral radius of the averaging operator acting on $L^2(X,\nu)/\{\text{const}\}$ is at most $1-\kappa$.

Compare Theorem 1.9 in [14] (Theorem 9.18 in [12]).

Proof. First, if X is finite then Lipschitz functions coincide with L^2 functions, and the norms are equivalent, so that there is nothing to prove. So we suppose that ν is reversible, i.e. M is self-adjoint.

Let f be a k-Lipschitz function. Proposition 32 below implies that Lipschitz functions belong to $L^2(X,\nu)/\{\text{const}\}$ and that the Lipschitz norm controls the L^2 norm (this is where we use that $\sigma < \infty$). Since $\mathrm{M}^t f$ is $k(1-\kappa)^t$-Lipschitz one gets $\mathrm{Var}\,\mathrm{M}^t f \leqslant Ck^2(1-\kappa)^{2t}$ for some constant C so that $\lim_{t\to\infty}(\sqrt{\mathrm{Var}\,\mathrm{M}^t f})^{1/t} \leqslant (1-\kappa)$. So the spectral radius of M is at most $1-\kappa$ on the subspace of Lipschitz functions.

Now Lipschitz functions are dense in $L^2(X,\nu)$ (indeed, a probability measure on a metric space is regular, so that indicator functions of measurable sets can be approximated by Lipschitz functions). Since M is bounded and self-adjoint, its spectral radius is controlled by its value on a dense subspace using the spectral decomposition. $\square$

Corollary 31 *(Poincaré inequality). Let (X, d, m) be an ergodic random walk on a metric space, with invariant distribution ν. Suppose that the coarse Ricci curvature of X is at least $\kappa > 0$ and that $\sigma < \infty$. Suppose that ν is reversible.*

Then the spectrum of $-\Delta$ acting on $L^2(X, \nu)/\{\text{const}\}$ is contained in $[\kappa; \infty)$. Moreover the following discrete Poincaré inequalities are satisfied for $f \in L^2(X, \nu)$:

$$\operatorname{Var}_\nu f \leqslant \frac{1}{\kappa(2-\kappa)} \int \operatorname{Var}_{m_x} f \, \mathrm{d}\nu(x)$$

and

$$\operatorname{Var}_\nu f \leqslant \frac{1}{2\kappa} \iint \big(f(y) - f(x)\big)^2 \, \mathrm{d}\nu(x) \, \mathrm{d}m_x(y).$$

Proof. These are rewritings of the inequalities $\operatorname{Var}_\nu \mathrm{M} f \leqslant (1-\kappa)^2 \operatorname{Var}_\nu f$ and $\langle f, \mathrm{M} f \rangle_{L^2(X,\nu)/\{\text{const}\}} \leqslant (1-\kappa) \operatorname{Var}_\nu f$, respectively. □

The quantities $\operatorname{Var}_{m_x} f$ and $\frac{1}{2}\int (f(y) - f(x))^2 \, \mathrm{d}m_x(y)$ are two possible analogues of $\|\nabla f(x)\|^2$ in a discrete setting. Though the latter is more common, the former is preferable when the support of m_x can be far away from x because it cancels out the "drift." Moreover one always has $\operatorname{Var}_{m_x} f \leqslant \int (f(y) - f(x))^2 \, \mathrm{d}m_x(y)$, so that the first form is generally sharper.

Reversibility is really needed here to turn an estimate of the spectral radius of M into an inequality between the norms of $\mathrm{M} f$ and f, using that M is self-adjoint. When the random walk is not reversible, applying the above to $\mathrm{M}\mathrm{M}^*$ does not work since the coarse Ricci curvature of the latter is unknown. However, a version of the Poincaré inequality with a non-local gradient still holds (Theorem 45).

As proven by Gromov and Milman ([25], or Corollary 3.1 and Theorem 3.3 in [29]), in quite a general setting a Poincaré inequality implies exponential concentration. Their argument adapts well here, and provides a concentration bound of roughly $\exp(-t\sqrt{\kappa}\,\sigma_\infty)$. We do not include the details, however, since Theorem 33 below is always more precise and covers the non-reversible case as well.

Let us compare this result to Lichnerowicz's theorem in the case of the ε-step random walk on an N-dimensional Riemannian manifold with positive Ricci curvature. This theorem states that the smallest eigenvalue of the usual Laplacian is $\frac{N}{N-1} \inf \mathrm{Ric}$, where $\inf \mathrm{Ric}$ is the largest K such that $\mathrm{Ric}(v, v) \geqslant K$ for all unit tangent vectors v. On the other hand, the operator Δ associated with the random walk is the difference between the mean value of a function on a ball of radius ε, and its value at the center of the ball: when $\varepsilon \to 0$ this behaves like $\frac{\varepsilon^2}{2(N+2)}$ times the usual Laplacian (take the average on the ball of the Taylor expansion of f). We saw (Example 7) that in this case $\kappa \sim \frac{\varepsilon^2}{2(N+2)} \inf \mathrm{Ric}$. Note that both scaling factors are the same. So we miss the $\frac{N}{N-1}$ factor, but otherwise get the correct order of magnitude.

Second, let us test this corollary for the discrete cube of Example 8. In this case the eigenbase of the discrete Laplacian is well-known (characters, or Fourier/Walsh transform), and the spectral gap of the discrete Laplacian associated with the lazy random walk is $1/N$. Since the coarse Ricci curvature κ is $1/N$ too, the value given in the proposition is sharp.

Third, consider the Ornstein–Uhlenbeck process on $\mathbb{R}$, as in Example 9. Its infinitesimal generator is $L = \frac{s^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} - \alpha x \frac{\mathrm{d}}{\mathrm{d}x}$, and the eigenfunctions are known to be $H_k(x\sqrt{\alpha/s^2})$ where H_k is

the Hermite polynomial $H_k(x) := (-1)^k e^{x^2} \frac{d^k}{dx^k} e^{-x^2}$. The associated eigenvalue of L is $-k\alpha$, so that the spectral gap of L is α. Now the random walk we consider is the flow $e^{\delta t L}$ at time δt of the process (with small δt), whose eigenvalues are $e^{-k\alpha\delta t}$. So the spectral gap of the discrete Laplacian $e^{\delta t L} - \mathrm{Id}$ is $1 - e^{-\alpha\delta t}$. Since coarse Ricci curvature is $1 - e^{-\alpha\delta t}$ too, the corollary is sharp again.

3. Concentration results

3.1. Variance of Lipschitz functions

We begin with the simplest kind of concentration, namely, an estimation of the variance of Lipschitz functions. Contrary to Gaussian or exponential concentration, the only assumption needed here is that the average diffusion constant σ is finite.

Since our Gaussian concentration result will yield basically the same variance $\sigma^2/n\kappa$, we discuss sharpness of this estimate in various examples in Section 3.3.

Proposition 32. *Let (X, d, m) be a random walk on a metric space, with coarse Ricci curvature at least $\kappa > 0$. Let ν be the unique invariant distribution. Suppose that $\sigma < \infty$.*

Then the variance of a 1-Lipschitz function is at most $\frac{\sigma^2}{n\kappa(2-\kappa)}$.

Note that since $\kappa \leqslant 1$ one has $\frac{\sigma^2}{n\kappa(2-\kappa)} \leqslant \frac{\sigma^2}{n\kappa}$.

In particular, this implies that all Lipschitz functions are in $L^2/\{\text{const}\}$; especially, $\int d(x, y)^2 \, d\nu(x) \, d\nu(y)$ is finite. The fact that the Lipschitz norm controls the L^2 norm was used above in the discussion of spectral properties of the random walk operator.

The assumption $\sigma < \infty$ is necessary. As a counterexample, consider a random walk on $\mathbb{N}$ that sends every $x \in \mathbb{N}$ to some fixed distribution ν on $\mathbb{N}$ with infinite second moment: coarse Ricci curvature is 1, yet the identity function is not in L^2.

Proof. Suppose for now that $|f|$ is bounded by $A \in \mathbb{R}$, so that $\operatorname{Var} f < \infty$. We first prove that $\operatorname{Var} \mathrm{M}^t f$ tends to 0. Let B_r be the ball of radius r in X centered at some basepoint. Using that $\mathrm{M}^t f$ is $(1-\kappa)^t$-Lipschitz on B_r and bounded by A on $X \setminus B_r$, we get $\operatorname{Var} \mathrm{M}^t f = \frac{1}{2} \iint (\mathrm{M}^t f(x) - \mathrm{M}^t f(y))^2 \, d\nu(x) \, d\nu(y) \leqslant 2(1-\kappa)^{2t} r^2 + 2A^2 \nu(X \setminus B_r)$. Taking for example $r = 1/(1-\kappa)^{t/2}$ shows that $\operatorname{Var} \mathrm{M}^t f \to 0$.

As already mentioned, one has $\operatorname{Var} f = \operatorname{Var} \mathrm{M} f + \int \operatorname{Var}_{m_x} f \, d\nu(x)$. Since $\operatorname{Var} \mathrm{M}^t f \to 0$, by induction we get

$$\operatorname{Var} f = \sum_{t=0}^{\infty} \int \operatorname{Var}_{m_x} \mathrm{M}^t f \, d\nu(x).$$

Now since f is 1-Lipschitz, by definition $\operatorname{Var}_{m_x} f \leqslant \sigma(x)^2/n_x$. Since $\mathrm{M}^t f$ is $(1-\kappa)^t$-Lipschitz, we have $\operatorname{Var}_{m_x} \mathrm{M}^t f \leqslant (1-\kappa)^{2t} \sigma(x)^2/n_x$ so that the sum above is at most $\frac{\sigma^2}{n\kappa(2-\kappa)}$. The case of unbounded f is treated by a simple limiting argument. □

3.2. Gaussian concentration

As mentioned above, positive coarse Ricci curvature implies a Gaussian-then-exponential concentration theorem. The estimated variance is $\sigma^2/n\kappa$ as above, so that this is essentially a more precise version of Proposition 32, with some loss in the constants. We will see in the discussion below (Section 3.3) that in the main examples, the order of magnitude is correct.

The fact that concentration is not always Gaussian far away from the mean is genuine, as exemplified by binomial distributions on the cube (Section 3.3.3) or $M/M/\infty$ queues (Section 3.3.4). The width of the Gaussian window is controlled by two factors. First, variations of the diffusion constant $\sigma(x)^2$ can result in purely exponential behavior (Section 3.3.5); this leads to the assumption that $\sigma(x)^2$ is bounded by a Lipschitz function. Second, as Gaussian phenomena only emerge as the result of a large number of small events, the "granularity" of the process must be bounded, which leads to the (comfortable) assumption that $\sigma_\infty < \infty$. Otherwise, a Markov chain which sends every point $x \in X$ to some fixed measure ν has coarse Ricci curvature 1 and can have arbitrary bad concentration properties depending on ν.

In the case of Riemannian manifolds, simply letting the step of the random walk tend to 0 makes the width of the Gaussian window tend to infinity, so that we recover Gaussian concentration as in the Lévy–Gromov or Bakry–Émery theorems. For the uniform measure on the discrete cube, the Gaussian width is equal to the diameter of the cube, so that we get full Gaussian concentration as well. In a series of other examples (such as Poisson measures), the transition from Gaussian to non-Gaussian regime occurs roughly as predicted by the theorem.

Theorem 33 *(Gaussian concentration). Let (X, d, m) be a random walk on a metric space, with coarse Ricci curvature at least $\kappa > 0$. Let ν be the unique invariant distribution.*

Let

$$D_x^2 := \frac{\sigma(x)^2}{n_x \kappa}$$

and

$$D^2 := \mathbb{E}_\nu D_x^2.$$

Suppose that the function $x \mapsto D_x^2$ is C-Lipschitz. Set

$$t_{\max} := \frac{D^2}{\max(\sigma_\infty, 2C/3)}.$$

Then for any 1-Lipschitz function f, for any $t \leqslant t_{\max}$ we have

$$\nu\big(\{x,\, f(x) \geqslant t + \mathbb{E}_\nu f\}\big) \leqslant \exp - \frac{t^2}{6D^2}$$

and for $t \geqslant t_{\max}$

$$\nu\big(\{x,\, f(x) \geqslant t + \mathbb{E}_\nu f\}\big) \leqslant \exp\left(- \frac{t_{\max}^2}{6D^2} - \frac{t - t_{\max}}{\max(3\sigma_\infty, 2C)} \right).$$

Remark 34. Proposition 24 or Corollary 22 often provide very sharp a priori bounds for $\mathbb{E}_\nu D_x^2$ even when no information on ν is available, as we shall see in the examples.

Remark 35. It is clear from the proof below that $\sigma(x)^2/n_x\kappa$ itself need not be Lipschitz, only bounded by some Lipschitz function. In particular, if $\sigma(x)^2$ is bounded one can always set $D^2 = \sup_x \frac{\sigma(x)^2}{n_x\kappa}$ and $C = 0$.

Remark 36 *(Continuous-time situations).* If we replace the random walk $m = (m_x)_{x\in X}$ with the lazy random walk m' whose transition probabilities are $m'_x := (1-\alpha)\delta_x + \alpha m_x$, when α tends to 0 this approximates the law at time α of the continuous-time random walk with transition rates m_x, so that the continuous-time random walk is obtained by taking the lazy random walk m' and speeding up time by $1/\alpha$ when $\alpha \to 0$. Of course this does not change the invariant distribution. The point is that when $\alpha \to 0$, both σ_x^2 and κ scale like α (and n_x tends to 1), so that D^2 has a finite limit. This means that we can apply Theorem 33 to continuous-time examples that naturally appear as limits of a discrete-time, finite-space Markov chain, as illustrated in Sections 3.3.4 to 3.3.6.

Remark 37. The condition that σ_∞ is uniformly bounded can be replaced with a Gaussian-type assumption, namely that for each measure m_x there exists a number s_x such that $\mathbb{E}_{m_x} e^{\lambda f} \leqslant e^{\lambda^2 s_x^2/2} e^{\lambda \mathbb{E}_{m_x} f}$ for any 1-Lipschitz function f. Then a similar theorem holds, with $\sigma(x)^2/n_x$ replaced with s_x^2. (When s_x^2 is constant this is Proposition 2.10 in [15].) However, this is generally not well-suited to discrete settings, because when transition probabilities are small, the best s_x^2 for which such an inequality is satisfied is usually much larger than the actual variance $\sigma(x)^2$: for example, if two points x and y are at distance 1 and $m_x(y) = \varepsilon$, s_x must satisfy $s_x^2 \geqslant 1/2\ln(1/\varepsilon) \gg \varepsilon$. Thus making this assumption will provide extremely poor estimates of the variance D^2 when some transition probabilities are small (e.g. for binomial distributions on the discrete cube), and in particular, this cannot extend to the continuous-time limit.

In Section 3.3.5, we give a simple example where the Lipschitz constant of $\sigma(x)^2$ is large, resulting in exponential rather than Gaussian behavior. In Section 3.3.6 we give two examples of positively curved process with heavy tails: one in which $\sigma_\infty = 1$ but with non-Lipschitz growth of $\sigma(x)^2$, and one with $\sigma(x)^2 \leqslant 1$ but with unbounded $\sigma_\infty(x)$. These show that the assumptions cannot be relaxed.

Proof. This proof is a variation on standard martingale methods for concentration (see e.g. Lemma 4.1 in [29], or [45]).

Lemma 38. *Let $\varphi : X \to \mathbb{R}$ be an α-Lipschitz function with $\alpha \leqslant 1$. Assume $\lambda \leqslant 1/3\sigma_\infty$. Then for $x \in X$ we have*

$$\big(\mathrm{M}e^{\lambda\varphi}\big)(x) \leqslant e^{\lambda \mathrm{M}\varphi(x) + \lambda^2\alpha^2 \frac{\sigma(x)^2}{n_x}}.$$

Note that the classical Proposition 1.16 in [29] would yield $(\mathrm{M}e^{\lambda\varphi})(x) \leqslant e^{\lambda \mathrm{M}\varphi(x) + 2\lambda^2\alpha^2\sigma_\infty^2}$, which is too weak to provide reasonable variance estimates.

Proof of Lemma 38. For any smooth function g and any real-valued random variable Y, a Taylor expansion with Lagrange remainder gives $\mathbb{E}g(Y) \leqslant g(\mathbb{E}Y) + \frac{1}{2}(\sup g'')\operatorname{Var} Y$. Applying this with $g(Y) = \mathrm{e}^{\lambda Y}$ we get

$$\big(\mathrm{M}\mathrm{e}^{\lambda\varphi}\big)(x) = \mathbb{E}_{m_x}\mathrm{e}^{\lambda\varphi} \leqslant \mathrm{e}^{\lambda\mathrm{M}\varphi(x)} + \frac{\lambda^2}{2}\Big(\sup_{\operatorname{Supp} m_x} \mathrm{e}^{\lambda\varphi}\Big)\operatorname{Var}_{m_x}\varphi$$

and note that since $\operatorname{diam}\operatorname{Supp} m_x \leqslant 2\sigma_\infty$ and φ is 1-Lipschitz we have $\sup_{\operatorname{Supp} m_x}\varphi \leqslant \mathbb{E}_{m_x}\varphi + 2\sigma_\infty$, so that

$$\big(\mathrm{M}\mathrm{e}^{\lambda\varphi}\big)(x) \leqslant \mathrm{e}^{\lambda\mathrm{M}\varphi(x)} + \frac{\lambda^2}{2}\,\mathrm{e}^{\lambda\mathrm{M}\varphi(x)+2\lambda\sigma_\infty}\,\operatorname{Var}_{m_x}\varphi.$$

Now, by definition we have $\operatorname{Var}_{m_x}\varphi \leqslant \alpha^2\,\frac{\sigma(x)^2}{n_x}$. Moreover for $\lambda \leqslant 1/3\sigma_\infty$ we have $\mathrm{e}^{2\lambda\sigma_\infty} \leqslant 2$, hence the result. □

Back to the proof of the theorem, let f be a 1-Lipschitz function and $\lambda \geqslant 0$. Define by induction $f_0 := f$ and $f_{k+1}(x) := \mathrm{M}f_k(x) + \lambda\frac{\sigma(x)^2}{n_x}(1-\kappa/2)^{2k}$.

Suppose that $\lambda \leqslant 1/2C$. Then $\lambda\frac{\sigma(x)^2}{n_x}$ is $\kappa/2$-Lipschitz. Using Proposition 29, we can show by induction that f_k is $(1-\kappa/2)^k$-Lipschitz.

Consequently, the lemma yields

$$\big(\mathrm{M}\mathrm{e}^{\lambda f_k}\big)(x) \leqslant \mathrm{e}^{\lambda\mathrm{M}f_k(x)+\lambda^2\frac{\sigma(x)^2}{n_x}(1-\kappa/2)^{2k}} = \mathrm{e}^{\lambda f_{k+1}(x)}$$

so that by induction

$$\big(\mathrm{M}^k\mathrm{e}^{\lambda f}\big)(x) \leqslant \mathrm{e}^{\lambda f_k(x)}.$$

Now setting $g(x) := \frac{\sigma(x)^2}{n_x}$ and unwinding the definition of f_k yields

$$f_k(x) = \big(\mathrm{M}^k f\big)(x) + \lambda\sum_{i=1}^{k}\big(\mathrm{M}^{k-i}g\big)(x)\,(1-\kappa/2)^{2(i-1)}$$

so that

$$\lim_{k\to\infty} f_k(x) = \mathbb{E}_\nu f + \lambda\sum_{i=1}^{\infty}\mathbb{E}_\nu g\,(1-\kappa/2)^{2(i-1)} \leqslant \mathbb{E}_\nu f + \lambda\mathbb{E}_\nu g\,\frac{4}{3\kappa}.$$

Meanwhile, $(\mathrm{M}^k\mathrm{e}^{\lambda f})(x)$ tends to $\mathbb{E}_\nu\mathrm{e}^{\lambda f}$, so that

$$\mathbb{E}_\nu\mathrm{e}^{\lambda f} \leqslant \lim_{k\to\infty}\mathrm{e}^{\lambda f_k} \leqslant \mathrm{e}^{\lambda\mathbb{E}_\nu f + \frac{4\lambda^2}{3\kappa}\mathbb{E}_\nu\frac{\sigma(x)^2}{n_x}}.$$

We can conclude by a standard Chebyshev inequality argument. The restrictions $\lambda \leqslant 1/2C$ and $\lambda \leqslant 1/3\sigma_\infty$ give the value of $t_{\max}$. □

Remark 39 *(Finite-time concentration).* The proof provides a similar concentration result for the finite-time measures m_x^{*k} as well, with variance

$$D_{x,k}^2 = \sum_{i=1}^{k} (1-\kappa/2)^{2(i-1)} \left(\mathrm{M}^{k-i} \frac{\sigma(y)^2}{n_y} \right)(x)$$

and $D_{x,k}^2$ instead of D^2 in the expression for $t_{\max}$.

3.3. Examples revisited

Let us test the sharpness of these estimates in some examples, beginning with the simplest ones. In each case, we gather the relevant quantities in a table. Recall that $\approx$ denotes an equality up to a multiplicative universal constant (typically $\leqslant 4$), while symbol $\sim$ denotes usual asymptotic equivalence (with sharp constant).

3.3.1. Riemannian manifolds

First, let X be a compact N-dimensional Riemannian manifold with positive Ricci curvature. Equip this manifold with the ε-step random walk as in Example 7. The measure $\frac{\operatorname{vol} B(x,\varepsilon)}{\operatorname{vol} B_{\mathrm{Eucl}}(\varepsilon)} \operatorname{d}\operatorname{vol}(x)$ is reversible for this random walk. In particular, when $\varepsilon \to 0$, the density of this measure with respect to the Riemannian volume is $1 + O(\varepsilon^2)$.

Let $\inf \mathrm{Ric}$ denote the largest $K > 0$ such that $\mathrm{Ric}(v,v) \geqslant K$ for any unit tangent vector v. The relevant quantities for this random walk are as follows (see Section 8 for the proofs).

Coarse Ricci curvature	$\kappa \sim \frac{\varepsilon^2}{2(N+2)} \inf \mathrm{Ric}$
Coarse diffusion constant	$\sigma(x)^2 \sim \varepsilon^2 \frac{N}{N+2} \quad \forall x$
Dimension	$n \approx N$
Variance (Lévy–Gromov thm.)	$\approx 1/\inf \mathrm{Ric}$
Gaussian variance (Theorem 33)	$D^2 \approx 1/\inf \mathrm{Ric}$
Gaussian range	$t_{\max} \approx 1/(\varepsilon \inf \mathrm{Ric}) \to \infty$

So, up to some (small) numerical constants, we recover Gaussian concentration as in the Lévy–Gromov theorem.

The same applies to diffusions with a drift on a Riemannian manifold, as considered by Bakry and Émery. To be consistent with the notation of Example 11, in the table above ε has to be replaced with $\sqrt{(N+2)\delta t}$, and $\inf \mathrm{Ric}$ with $\inf(\mathrm{Ric}(v,v) - 2\nabla^{\mathrm{sym}} F(v,v))$ for v a unit tangent vector. (In the non-compact case, care has to be taken since the solution of the stochastic differential equation of Example 11 on the manifold may not exist, and even if it does its Euler scheme approximation at time δt may not converge uniformly on the manifold. In explicit examples such as the Ornstein–Uhlenbeck process, however, this is not a problem.)

3.3.2. Discrete cube

Consider now the discrete cube $\{0,1\}^N$ equipped with its graph distance (Hamming metric) and lazy random walk (Example 8).

For a random walk on a graph one always has $\sigma \approx 1$, and $n \geqslant 1$ in full generality. The following remark allows for more precise constants.

Remark 40. Let m be a random walk on a graph. Then, for any vertex x one has $\sigma(x)^2/n_x \leqslant 1 - m_x(\{x\})$.

Proof. By definition $\sigma(x)^2/n_x$ is the maximal variance, under m_x, of a 1-Lipschitz function. So let f be a 1-Lipschitz function on the graph. Since variance is invariant by adding a constant, we can assume that $f(x) = 0$. Then $|f(y)| \leqslant 1$ for any neighbor y of x. The mass, under m_x, of all neighbors of x is $1 - m_x(\{x\})$. Hence $\mathrm{Var}_{m_x} f = \mathbb{E}_{m_x} f^2 - (\mathbb{E}_{m_x} f)^2 \leqslant \mathbb{E}_{m_x} f^2 \leqslant 1 - m_x(\{x\})$.

This value is achieved, for example, with a lazy simple random walk when x has an even number of neighbors and when no two distinct neighbors of x are mutual neighbors; in this case one can take $f(x) = 0$, $f = 1$ on half the neighbors of x and $f = -1$ on the remaining neighbors of x. □

Applying this to the lazy random walk on the discrete cube, one gets:

Coarse Ricci curvature	$\kappa = 1/N$
Coarse diffusion constant & dimension	$\sigma(x)^2/n_x \leqslant 1/2$
Estimated variance (Proposition 32)	$\sigma^2/n\kappa(2-\kappa) \sim N/4$
Actual variance	$N/4$
Gaussian variance (Theorem 33)	$D^2 \leqslant N/2$
Gaussian range	$t_{\max} = N/2$

In particular, since $N/2$ is the maximal possible value for the deviation from average of a 1-Lipschitz function on the cube, we see that $t_{\max}$ has the largest possible value.

3.3.3. Binomial distributions

The occurrence of a finite range $t_{\max}$ for the Gaussian behavior of tails is genuine, as the following example shows.

Let again $X = \{0, 1\}^N$ equipped with its Hamming metric (each edge is of length 1). Consider the following Markov chain on X: for some $0 < p < 1$, at each step, choose a bit at random among the N bits; if it is equal to 0, flip it to 1 with probability p; if it is equal to 1, flip it to 0 with probability $1 - p$. The binomial distribution $\nu((x_1, \dots, x_N)) = \prod p^{x_i}(1-p)^{1-x_i}$ is reversible for this Markov chain.

The coarse Ricci curvature of this Markov chain is $1/N$, as can easily be seen directly or using the tensorization property (Proposition 27).

Let k be the number of bits of $x \in X$ which are equal to 1. Then k follows a Markov chain on $\{0, 1, \dots, N\}$, whose transition probabilities are:

$$\begin{aligned} p_{k,k+1} &= p(1 - k/N), \\ p_{k,k-1} &= (1-p)k/N, \\ p_{k,k} &= pk/N + (1-p)(1-k/N). \end{aligned}$$

The binomial distribution with parameters N and p, namely $\binom{N}{k} p^k (1-p)^{N-k}$, is reversible for this Markov chain. Moreover, the coarse Ricci curvature of this "quotient" Markov chain is still $1/N$.

Now, fix some $\lambda > 0$ and consider the case $p = \lambda/N$. Let $N \to \infty$. It is well known that the invariant distribution tends to the Poisson distribution $e^{-\lambda}\lambda^k/k!$ on $\mathbb{N}$.

Let us see how Theorem 33 performs on this example. The table below applies either to the full space $\{0, 1\}^N$, with k the function "number of 1's," or to its projection on $\{0, 1, \dots, N\}$. Note the use of Proposition 24 to estimate σ^2 a priori, without having to resort to explicit knowledge of the invariant distribution. All constants implied in the $O(1/N)$ notation are small and completely explicit.

Coarse Ricci curvature	$\kappa = 1/N$
Coarse diffusion constant	$\sigma(k)^2 = (\lambda + k)/N + O(1/N^2)$
Estimated $\mathbb{E}k$ (Proposition 24)	$\mathbb{E}k \leqslant J(0)/\kappa = \lambda$
Actual $\mathbb{E}k$	$\mathbb{E}k = \lambda$
Average diffusion constant	$\sigma^2 = \mathbb{E}\sigma(k)^2 = 2\lambda/N + O(1/N^2)$
Dimension	$n \geqslant 1$
Estimated variance (Proposition 32)	$\sigma^2/n\kappa(2-\kappa) \leqslant \lambda + O(1/N)$
Actual variance	λ
Gaussian variance (Theorem 33)	$D^2 \leqslant 2\lambda + O(1/N)$
Lipschitz constant of D_x^2	$C = 1 + O(1/N)$
Gaussian range	$t_{\max} = 4\lambda/3$

The Poisson distribution has a roughly Gaussian behavior (with variance λ) in a range of size approximately λ around the mean; further away, it decreases like $e^{-k\ln k}$ which is not Gaussian. This is in good accordance with $t_{\max}$ the table above, and shows that the Gaussian range cannot be extended.

3.3.4. A continuous-time example: $M/M/\infty$ queues

Here we show how to apply Theorem 33 to a continuous-time example, the $M/M/\infty$ queue. These queues were brought to my attention by D. Chafaï.

The $M/M/\infty$ queue consists of an infinite number of "servers." Each server can be free (0) or busy (1). The state space consists of all sequences in $\{0, 1\}^{\mathbb{N}}$ with a finite number of 1's. The dynamics is at follows: fix two numbers $\lambda > 0$ and $\mu > 0$. At a rate λ per unit of time, a client arrives and the first free server becomes busy. At a rate μ per unit of time, each busy server finishes its job (independently of the others) and becomes free. The number $k \in \mathbb{N}$ of busy servers is a continuous-time Markov chain, whose transition probabilities at small times t are given by

$$\begin{aligned} p_{k,k+1}^t &= \lambda t + O\bigl(t^2\bigr), \\ p_{k,k-1}^t &= k\mu t + O\bigl(t^2\bigr), \\ p_{k,k}^t &= 1 - (\lambda + k\mu)t + O\bigl(t^2\bigr). \end{aligned}$$

This system is often presented as a discrete analogue of an Ornstein–Uhlenbeck process, since asymptotically the drift is linear towards the origin. However, it is not symmetric around the mean, and moreover the invariant (actually reversible) distribution ν is a Poisson distribution with parameter λ/μ, rather than a Gaussian.

This continuous-time Markov chain can be seen as a limit of the binomial Markov chain above as follows: first, replace the binomial Markov chain with its continuous-time equivalent (Remark 36); Then, set $p = \lambda/N$ and let $N \to \infty$, while speeding up time by a factor N. The analogy is especially clear in the table below: if we replace λ with λ/N and μ with $1/N$, we get essentially the same table as for the binomial distribution.

It is easy to check that Proposition 32 (with $\sigma^2/2n\kappa$ instead of $\sigma^2/n\kappa(2-\kappa)$) and Theorem 33 pass to the limit. In this continuous-time setting, the definitions become the following: $\kappa(x,y) := -\frac{\mathrm{d}}{\mathrm{d}t} W_1(m_x^t, m_y^t)/d(x,y)$ (as mentioned in the introduction) and $\sigma(x)^2 := \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\iint d(y,z)\,\mathrm{d}m_x^t(y)\,\mathrm{d}m_x^t(z)$, where m_x^t is the law at time t of the process starting at x.

Then the relevant quantities are as follows.

Coarse Ricci curvature	$\kappa = \mu$
Coarse diffusion constant	$\sigma(k)^2 = k\mu + \lambda$
Estimated $\mathbb{E}k$ (Proposition 24)	$\mathbb{E}k \leqslant J(0)/\kappa = \lambda/\mu$
Actual $\mathbb{E}k$	$\mathbb{E}k = \lambda/\mu$
Average diffusion constant	$\sigma^2 = \mathbb{E}\sigma(k)^2 = 2\lambda$
Dimension	$n \geqslant 1$
Estimated variance (Proposition 32)	$\sigma^2/2n\kappa = \lambda/\mu$
Actual variance	λ/μ
Gaussian variance (Theorem 33)	$D^2 \leqslant 2\lambda/\mu$
Lipschitz constant of D_x^2	$C = 1$
Gaussian range	$t_{\max} = 4\lambda/3\mu$

So once more Theorem 33 is in good accordance with the behavior of the random walk, whose invariant distribution is Poisson with mean and variance λ/μ, thus Gaussian-like only in some interval around this value.

An advantage of this approach is that is can be generalized to situations where the rates of the servers are not constant, but, say, bounded between $\mu_0/10$ and $10\mu_0$, and clients go to the first free server according to some predetermined scheme, e.g. the fastest free server. Indeed, the $M/M/\infty$ queue above can be seen as a Markov chain in the full configuration space of the servers, namely the space of all sequences over the alphabet {free, busy} containing a finite number of "busy." It is easy to check that the coarse Ricci curvature is still equal to μ in this configuration space. Now in the case of variable rates, the number of busy servers is generally not Markovian, so one has to work in the configuration space. If the rate of the ith server is μ_i, the coarse Ricci curvature is $\inf \mu_i$ in the configuration space, whereas the diffusion constant is controlled by $\sup \mu_i$. So if the rates vary in a bounded range, coarse Ricci curvature still provides a Gaussian-type control, though an explicit description of the invariant distribution is not available.

Let us consider more realistic queue models, such as the $M/M/k$ queue, i.e. the number of servers is equal to k (with constant or variable rates). Then, on the part of the space where some servers are free, coarse Ricci curvature is at least equal to the rate of the slowest server; whereas on the part of the space where all servers are busy, coarse Ricci curvature is 0. If, as often, an abandon rate for waiting clients is added to the model, then coarse Ricci curvature is equal to this abandon rate on the part of the space where all servers are busy (and in particular, coarse Ricci curvature is positive on the whole space).

3.3.5. An example of exponential concentration

We give here a very simple example of a Markov chain which has positive curvature but for which concentration is not Gaussian but exponential, due to large variations of the diffusion constant, resulting in a large value of C. Compare Example 14 above where exponential concentration was due to unbounded σ_∞.

This is a continuous-time random walk on $\mathbb{N}$ defined as follows. Take $0 < \alpha < \beta$. For $k \in \mathbb{N}$, the transition rate from k to $k+1$ is $(k+1)\alpha$, whereas the transition rate from $k+1$ to k is $(k+1)\beta$. It is immediate to check that the geometric distribution with decay α/β is reversible for this Markov chain.

The coarse Ricci curvature of this Markov chain is easily seen to be $\beta - \alpha$. We have $\sigma(k)^2 = (k+1)\alpha + k\beta$, so that $\sigma(k)^2$ is $(\alpha+\beta)$-Lipschitz and $C = (\alpha+\beta)/(\beta-\alpha)$.

The expectation of k under the invariant distribution can be bounded by $J(0)/\kappa = \alpha/(\beta-\alpha)$ by Proposition 24, which is actually the exact value. So the expression above for $\sigma(k)^2$ yields $\sigma^2 = 2\alpha\beta/(\beta-\alpha)$. Consequently, the estimated variance $\sigma^2/2n\kappa$ (obtained by the continuous-time version of Proposition 32) is at most $\alpha\beta/(\beta-\alpha)^2$, which is the actual value.

Now consider the case when $\beta - \alpha$ is small. If the C factor in Theorem 33 is not taken into account, we get blatantly false results since the invariant distribution is not Gaussian at all. Indeed, in the regime where $\beta - \alpha \to 0$, the width of the Gaussian window in Theorem 33 is $D^2/C \approx \alpha/(\beta-\alpha)$. This is fine, as this is the decay distance of the invariant distribution, and in this interval both the Gaussian and geometric estimates are close to 1 anyway. But without the C factor, we would get $D^2/\sigma_\infty = \alpha\beta/(\beta-\alpha)^2$, which is much larger; the invariant distribution is clearly not Gaussian on this interval.

Moreover, Theorem 33 predicts, in the exponential regime, a $\exp(-t/2C)$ behavior for concentration. Here the asymptotic behavior of the invariant distribution is $(\alpha/\beta)^t \sim (1-2/C)^t \sim \mathrm{e}^{-2t/C}$ when $\beta - \alpha$ is small. So we see that (up to a constant 4) the exponential decay rate predicted by Theorem 33 is genuine.

3.3.6. Heavy tails

It is clear that a variance control alone does not imply any concentration bound beyond the Bienaymé–Chebyshev inequality. We now show that this is still the case even under a positive curvature assumption. Namely, in Theorem 33, neither the assumption that $\sigma(x)^2$ is Lipschitz, nor the assumption that σ_∞ is bounded, can be removed (but see Remark 37).

Heavy tails with non-Lipschitz $\sigma(x)^2$. Our next example shows that if the diffusion constant $\sigma(x)^2$ is not Lipschitz, then non-exponential tails may occur in spite of positive curvature.

Consider the continuous-time random walk on $\mathbb{N}$ defined as follows: the transition rate from k to $k+1$ is $a(k+1)^2$, whereas the transition rate from k to $k-1$ is $a(k+1)^2 + bk$ for $k \geqslant 1$. Here $a, b > 0$ are fixed.

We have $\kappa = b$ and $\sigma(k)^2 = 2a(k+1)^2 + bk$, which is obviously not Lipschitz.

This Markov chain has a reversible measure ν, which satisfies $\nu(k)/\nu(k-1) = ak^2/(a(k+1)^2 + bk) = 1 - \frac{1}{k}(2 + \frac{b}{a}) + O(1/k^2)$. Consequently, asymptotically $\nu(k)$ behaves like

$$\prod_{i=1}^{k}\left(1 - \frac{1}{i}\left(2 + \frac{b}{a}\right)\right) \approx \mathrm{e}^{-(2+b/a)\sum_{i=1}^{k}\frac{1}{i}} \approx k^{-(2+b/a)}$$

thus exhibiting heavy, non-exponential tails.

This shows that the Lipschitz assumption for $\sigma(x)^2$ cannot be removed, even though $\sigma_\infty = 1$. It would seem reasonable to expect a systematic correspondance between the asymptotic behavior of $\sigma(x)^2$ and the behavior of tails.

Heavy tails with unbounded σ_∞. Consider the following random walk on $\mathbb{N}^*$: a number k goes to 1 with probability $1 - 1/4k^2$ and to $2k$ with probability $1/4k^2$. One can check that $\kappa \geqslant 1/2$. These probabilities are chosen so that $\sigma(k)^2 = (2k-1)^2 \times 1/4k^2 \times (1 - 1/4k^2) \leqslant 1$, so that the variance of the invariant distribution is small. However, let us evaluate the probability that, starting at 1, the first i steps all consist in doing a multiplication by 2, so that we end at 2^i; this probability is $\prod_{j=0}^{i-1} \frac{1}{4\cdot(2^j)^2} = 4^{-1-i(i-1)/2}$. Setting $i = \log_2 k$, we see that the invariant distribution ν satisfies

$$\nu(k) \geqslant \frac{\nu(1)}{4}\, 2^{-\log_2 k(\log_2 k - 1)}$$

for k a power of 2. This is clearly not Gaussian or exponential, though $\sigma(k)^2$ is bounded.

4. Local control and logarithmic Sobolev inequality

We now turn to control of the gradient of $\mathrm{M}f$ at some point, in terms of the gradient of f at neighboring points. This is closer to classical Bakry–Émery theory, and allows to get a kind of logarithmic Sobolev inequality.

Definition 41 *(Norm of the gradient).* Choose $\lambda > 0$ and, for any function $f : X \to \mathbb{R}$, define the λ-*range gradient* of f by

$$(\mathrm{D}f)(x) := \sup_{y,y' \in X} \frac{|f(y) - f(y')|}{d(y,y')}\, \mathrm{e}^{-\lambda d(x,y) - \lambda d(y,y')}.$$

This is a kind of "mesoscopic" Lipschitz constant of f around x, since pairs of points y, y' far away from x will not contribute much to $\mathrm{D}f(x)$. If f is a smooth function on a compact Riemannian manifold, when $\lambda \to \infty$ this quantity tends to $|\nabla f(x)|$.

It is important to note that $\log \mathrm{D}f$ is λ-Lipschitz.

We will also need a control on negative curvature: in a Riemannian manifold, Ricci curvature might be $\geqslant 1$ because there is a direction of curvature 1000 and a direction of curvature -999. The next definition captures these variations.

Definition 42 *(Unstability).* Let

$$\kappa_+(x,y) := \frac{1}{d(x,y)} \int_z \big(d(x,y) - d(x+z, y+z)\big)_+$$

and

$$\kappa_-(x,y) := \frac{1}{d(x,y)} \int_z \big(d(x,y) - d(x+z, y+z)\big)_-$$

where a_+ and a_- are the positive and negative part of $a \in \mathbb{R}$, so that $\kappa(x,y) = \kappa_+(x,y) - \kappa_-(x,y)$. (The integration over z is under a coupling realizing the value of $\kappa(x,y)$.)

The unstability $U(x,y)$ is defined as

$$U(x,y) := \frac{\kappa_-(x,y)}{\kappa(x,y)} \quad \text{and} \quad U := \sup_{x,y\in X, x\neq y} U(x,y).$$

Remark 43. If X is ε-geodesic, then an upper bound for $U(x,y)$ with $d(x,y) \leqslant \varepsilon$ implies the same upper bound for U.

In most discrete examples given in the introduction (Examples 8, 10, 12, 13, 14), unstability is actually 0, meaning that the coupling between m_x and m_y never increases distances. (This could be a possible definition of non-negative *sectional* curvature for Markov chains.) In Riemannian manifolds, unstability is controlled by the largest negative sectional curvature. Interestingly, in Example 17 (Glauber dynamics), unstability depends on temperature.

Due to the use of the gradient D, the theorems below are interesting only if a reasonable estimate for $\mathrm{D}f$ can be obtained depending on "local" data. This is not the case when f is not λ-log-Lipschitz (compare the similar phenomenon in [8]). This is consistent with the fact mentioned above, that Gaussian concentration of measure only occurs in a finite range, with exponential concentration afterwards, which implies that no true logarithmic Sobolev inequality can hold in general.

Theorem 44 *(Gradient contraction). Suppose that coarse Ricci curvature is at least $\kappa > 0$. Let $\lambda \leqslant \frac{1}{20\sigma_\infty(1+U)}$ and consider the λ-range gradient* D. *Then for any function $f : X \to \mathbb{R}$ with $\mathrm{D}f < \infty$ we have*

$$\mathrm{D}(\mathrm{M}f)(x) \leqslant (1-\kappa/2)\mathrm{M}(\mathrm{D}f)(x)$$

for all $x \in X$.

Theorem 45 *(Log-Sobolev inequality). Suppose that coarse Ricci curvature is at least $\kappa > 0$. Let $\lambda \leqslant \frac{1}{20\sigma_\infty(1+U)}$ and consider the λ-range gradient* D. *Then for any function $f : x \to \mathbb{R}$ with $\mathrm{D}f < \infty$, we have*

$$\mathrm{Var}_\nu f \leqslant \left(\sup_x \frac{4\sigma(x)^2}{\kappa n_x}\right) \int (\mathrm{D}f)^2 \,\mathrm{d}\nu$$

and for positive f,

$$\mathrm{Ent}_\nu f \leqslant \left(\sup_x \frac{4\sigma(x)^2}{\kappa n_x}\right) \int \frac{(\mathrm{D}f)^2}{f} \,\mathrm{d}\nu$$

where ν is the invariant distribution.

If moreover the random walk is reversible with respect to ν, then

$$\mathrm{Var}_\nu f \leqslant \int V(x)\,\mathrm{D}f(x)^2 \,\mathrm{d}\nu(x)$$

and

$$\operatorname{Ent}_\nu f \leqslant \int V(x) \frac{\mathrm{D}f(x)^2}{f(x)} \,\mathrm{d}\nu(x)$$

where

$$V(x) = 2\sum_{t=0}^{\infty} (1-\kappa/2)^{2t} \mathrm{M}^{t+1}\left(\frac{\sigma(x)^2}{n_x}\right).$$

The form involving $V(x)$ is motivated by the fact that, for reversible diffusions in $\mathbb{R}^N$ with non-constant diffusion coefficients, the coefficients naturally appear in the formulation of functional inequalities (see e.g. [2]). The quantity $V(x)\,\mathrm{D}f(x)^2$ is to be thought of as a crude version of the Dirichlet form associated with the random walk. It would be more satisfying to obtain inequalities involving the latter (compare Corollary 31), but I could not get a version of the commutation property $\mathrm{DM} \leqslant (1-\kappa/2)\mathrm{MD}$ involving the Dirichlet form.

Remark 46. If $\frac{\sigma(x)^2}{n_x \kappa}$ is C-Lipschitz (as in Theorem 33), then $V(x) \leqslant \frac{4\sigma^2}{\kappa n} + 2C\,\frac{J(x)}{\kappa}$.

Examples. Let us compare this theorem to classical results.

In the case of a Riemannian manifold, for any smooth function f we can choose a random walk with small enough steps, so that λ can be arbitrarily large and $\mathrm{D}f$ arbitrarily close to $|\nabla f|$. Since moreover $\sigma(x)^2$ does not depend on x for the Brownian motion, this theorem allows to recover the logarithmic Sobolev inequality in the Bakry–Émery framework, with the correct constant up to a factor 4.

Next, consider the two-point space $\{0,1\}$, equipped with the measure $\nu(0) = 1-p$ and $\nu(1) = p$. This is the space on which modified logarithmic Sobolev inequalities were introduced [8]. We endow this space with the Markov chain sending each point to the invariant distribution. Here we have $\sigma(x)^2 = p(1-p)$, $n_x = 1$ and $\kappa = 1$, so that we get the inequality $\operatorname{Ent}_\nu f \leqslant 4p(1-p)\int \frac{(\mathrm{D}f)^2}{f}\,\mathrm{d}\nu$, comparable to the known inequality [8] except for the factor 4.

The modified logarithmic Sobolev inequality for Bernoulli and Poisson measures is traditionally obtained by tensorizing this result [8]. If, instead, we directly apply the theorem above to the Bernoulli measure on $\{0,1\}^N$ or the Poisson measure on $\mathbb{N}$ (see Sections 3.3.3 and 3.3.4), we get slightly worse results. Indeed, consider the $M/M/\infty$ queue on $\mathbb{N}$, which is the limit when $N \to \infty$ of the projection on $\mathbb{N}$ of the Markov chains on $\{0,1\}^N$ associated with Bernoulli measures. Keeping the notation of Section 3.3.4, we get, in the continuous-time version, $\sigma(x)^2 = x\mu + \lambda$, which is not bounded. So we have to use the version with $V(x)$; Remark 46 and the formulas in Section 3.3.4 yield $V(x) \leqslant 8\lambda/\mu + 2(\lambda + x\mu)/\mu$ so that we get the inequality

$$\begin{aligned}\operatorname{Ent}_\nu f &\leqslant \frac{\lambda}{\mu}\int \frac{\mathrm{D}f(x)^2}{f(x)}(10 + 2x\mu/\lambda)\,\mathrm{d}\nu(x)\\ &= \frac{\lambda}{\mu}\int \frac{\mathrm{D}f(x)^2}{f(x)}\big(2\,\mathrm{d}\nu(x-1) + 10\,\mathrm{d}\nu(x)\big)\end{aligned}$$

which is to be compared to the inequality

$$\operatorname{Ent}_\nu f \leqslant \frac{\lambda}{\mu} \int \frac{\mathrm{D}_+ f(x)^2}{f(x)} \,\mathrm{d}\nu(x)$$

obtained in [8], with $\mathrm{D}_+ f(x) = f(x+1) - f(x)$. So asymptotically our version is worse by a factor $\mathrm{d}\nu(x-1)/\mathrm{d}\nu(x) \approx x$. One could say that our general, non-local notion of gradient fails to distinguish between a point and an immediate neighbor, and does not take advantage of the particular directional structure of a random walk on $\mathbb{N}$ as the use of D_+ does.

Yet being able to handle the configuration space directly rather than as a product of the two-point space allows us to deal with more general, non-product situations. Consider for example the queuing process with heterogeneous server rates mentioned at the end of Section 3.3.4, where newly arrived clients go to the fastest free server (in which case the number of busy servers is not Markovian). Then coarse Ricci curvature is equal to the infimum of the server rates, and Theorem 45 still holds, though the constants are probably not optimal when the rates are very different. I do not know if this result is new.

We now turn to the proof of Theorems 44 and 45. The proof of the former is specific to our setting, but the passage from the former to the latter is essentially a copy of the Bakry–Émery argument.

Lemma 47. *Let $x, y \in X$ with $\kappa(x,y) > 0$. Let (Z, μ) be a probability space equipped with a map $\pi : Z \to \operatorname{Supp} m_x \times \operatorname{Supp} m_y$ such that π sends μ to an optimal coupling between m_x and m_y. Let A be a positive function on Z such that $\sup A / \inf A \leqslant \mathrm{e}^\rho$ with $\rho \leqslant \frac{1}{2(1+U)}$. Then*

$$\int\limits_{z\in Z} A(z) \frac{d(x+z, y+z)}{d(x,y)} \leqslant \big(1 - \kappa(x,y)/2\big) \int\limits_z A(z)$$

and in particular

$$\int\limits_{z\in Z} A(z)\big(d(x+z, y+z) - d(x,y)\big) \leqslant 0$$

where, as usual, $x+z$ and $y+z$ denote the two projections from Z to $\operatorname{Supp} m_x$ and $\operatorname{Supp} m_y$ respectively.

Proof. The idea is the following: when A is constant, the result obviously holds since by definition $\int d(x+z, y+z)/d(x,y) = 1 - \kappa(x,y)$. Now when A is close enough to a constant, the same holds with some numerical loss.

Set $F = \sup_z A(z)$. Then

$$\int\limits_z A(z) \frac{d(x+z, y+z)}{d(x,y)} = \int\limits_z A(z) + F \int\limits_z \frac{A(z)}{F} \left(\frac{d(x+z, y+z)}{d(x,y)} - 1 \right).$$

Let $Z_- = \{z \in Z,\ d(x,y) < d(x+z, y+z)\}$ and $Z_+ = Z \setminus Z_-$. Recall that by definition, $\kappa_-(x,y) = \int_{Z_-} (d(x+z,y+z)/d(x,y) - 1)$ and $\kappa_+(x,y) = \int_{Z_+} (1 - d(x+z,y+z)/d(x,y))$, so that $\kappa = \kappa_+ - \kappa_-$. Using that $A(z) \leqslant F$ on Z_- and $A(z) \geqslant \mathrm{e}^{-\rho} F$ on Z_+, we get

$$\int_z A(z) \frac{d(x+z,y+z)}{d(x,y)} \leqslant \int_z A(z) + F\big(\kappa_-(x,y) - \mathrm{e}^{-\rho}\kappa_+(x,y)\big).$$

Now by definition of U we have $\kappa_-(x,y) \leqslant U\kappa(x,y)$. It is not difficult to check that $\rho \leqslant \frac{1}{2(1+U)}$ is enough to ensure that $\mathrm{e}^{-\rho}\kappa_+(x,y) - \kappa_-(x,y) \geqslant \kappa(x,y)/2$, hence

$$\begin{aligned}\int_z A(z) \frac{d(x+z,y+z)}{d(x,y)} &\leqslant \int_z A(z) - F\kappa(x,y)/2 \\ &\leqslant \big(1 - \kappa(x,y)/2\big) \int_z A(z)\end{aligned}$$

as needed. □

Proof of Theorem 44. Let $y, y' \in X$. Let ξ_{xy} and $\xi_{yy'}$ be optimal couplings between m_x and m_y, m_y and $m_{y'}$ respectively. Apply the gluing lemma for couplings (Lemma 7.6 in [47]) to obtain a measure μ on $Z = \operatorname{Supp} m_x \times \operatorname{Supp} m_y \times \operatorname{Supp} m_{y'}$ whose projections on $\operatorname{Supp} m_x \times \operatorname{Supp} m_y$ and $\operatorname{Supp} m_y \times \operatorname{Supp} m_{y'}$ are ξ_{xy} and $\xi_{yy'}$ respectively.

We have

$$\begin{aligned}&\frac{|\mathrm{M}f(y) - \mathrm{M}f(y')|}{d(y,y')} \mathrm{e}^{-\lambda(d(x,y)+d(y,y'))} \\ &= \left| \int_{z\in Z} f(y+z) - f(y'+z) \right| \frac{\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))}}{d(y,y')} \\ &\leqslant \int_{z\in Z} \big|f(y+z) - f(y'+z)\big| \frac{\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))}}{d(y,y')} \\ &\leqslant \int_{z\in Z} \mathrm{D}f(x+z) \frac{d(y+z,y'+z)}{\mathrm{e}^{-\lambda(d(x+z,y+z)+d(y+z,y'+z))}} \frac{\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))}}{d(y,y')} \\ &= \int_{z\in Z} A(z)B(z) \frac{d(y+z,y'+z)}{d(y,y')}\end{aligned}$$

where $A(z) = \mathrm{D}f(x+z)$ and $B(z) = \mathrm{e}^{\lambda(d(x+z,y+z) - d(x,y) + d(y+z,y'+z) - d(y,y'))}$.

Since $\operatorname{diam} \operatorname{Supp} m_x \leqslant 2\sigma_\infty$ and likewise for y, for any z, z' we have

$$\begin{aligned}\big|d(x+z,y+z) - d(x+z',y+z')\big| &\leqslant 4\sigma_\infty, \\ \big|d(y+z,y'+z) - d(y+z',y'+z')\big| &\leqslant 4\sigma_\infty\end{aligned}$$

so that B varies by a factor at most $\mathrm{e}^{8\lambda\sigma_\infty}$ on Z. Likewise, since $\mathrm{D}f$ is λ-log-Lipschitz, A varies by a factor at most $\mathrm{e}^{2\lambda\sigma_\infty}$. So the quantity $A(z)B(z)$ varies by at most $\mathrm{e}^{10\lambda\sigma_\infty}$.

So if $\lambda \leqslant \frac{1}{20\sigma_\infty(1+U)}$, we can apply Lemma 47 and get

$$\int_{z\in Z} A(z)B(z)\,\frac{d(y+z,y'+z)}{d(y,y')} \leqslant (1-\kappa/2)\int_{z\in Z} A(z)B(z).$$

Now we have $\int_z A(z)B(z) = \int_z A(z) + \int_z A(z)(B(z)-1)$. Unwinding $B(z)$ and using that $\mathrm{e}^a - 1 \leqslant a\mathrm{e}^a$ for any $a \in \mathbb{R}$, we get

$$\int_z A(z)(B(z)-1) \leqslant \lambda \int_z A(z)B(z)\big(d(x+z,y+z) - d(x,y) + d(y+z,y'+z) - d(y,y')\big)$$

which decomposes as a sum of two terms $\lambda\int_z A(z)B(z)(d(x+z,y+z)-d(x,y))$ and $\lambda\int_z A(z)B(z)(d(y+z,y'+z)-d(y,y'))$, each of which is non-positive by Lemma 47. Hence $\int_z A(z)(B(z)-1) \leqslant 0$ and $\int_z A(z)B(z) \leqslant \int_z A(z) = \int_z \mathrm{D}f(x+z) = \mathrm{M}(\mathrm{D}f)(x)$. So we have shown that for any y, y' in X we have

$$\frac{|\mathrm{M}f(y) - \mathrm{M}f(y')|}{d(y,y')}\,\mathrm{e}^{-\lambda(d(x,y)+d(y,y'))} \leqslant (1-\kappa/2)\mathrm{M}(\mathrm{D}f)(x)$$

as needed. □

Lemma 48. *Let f be a function with $\mathrm{D}f < \infty$. Let $x \in X$. Then f is $\mathrm{e}^{4\lambda\sigma_\infty}\mathrm{M}(\mathrm{D}f)(x)$-Lipschitz on $\operatorname{Supp} m_x$.*

Proof. For any $y, z \in \operatorname{Supp} m_x$, by definition of D we have $|f(y) - f(z)| \leqslant \mathrm{D}f(y)d(y,z)\mathrm{e}^{\lambda d(y,z)} \leqslant \mathrm{D}f(y)d(y,z)\mathrm{e}^{2\lambda\sigma_\infty}$. Since moreover $\mathrm{D}f$ is λ-log-Lipschitz, we have $\mathrm{D}f(y) \leqslant \mathrm{e}^{2\lambda\sigma_\infty}\inf_{\operatorname{Supp} m_x}\mathrm{D}f \leqslant \mathrm{e}^{2\lambda\sigma_\infty}\mathrm{M}(\mathrm{D}f)(x)$, so that finally

$$\big|f(y) - f(z)\big| \leqslant d(y,z)\,\mathrm{M}(\mathrm{D}f)(x)\,\mathrm{e}^{4\lambda\sigma_\infty}$$

as announced. □

Proof of Theorem 45. Let ν be the invariant distribution. Let f be a positive measurable function. Associativity of entropy (e.g. Theorem D.13 in [22] applied to the measure $f(y)\,\mathrm{d}\nu(x)\mathrm{d}m_x(y)$ on $X \times X$) states that

$$\begin{aligned}\operatorname{Ent} f &= \int_x \operatorname{Ent}_{m_x} f\,\mathrm{d}\nu(x) + \operatorname{Ent}\mathrm{M}f \\ &= \sum_{t\geqslant 0}\int_x \operatorname{Ent}_{m_x}\mathrm{M}^t f\,\mathrm{d}\nu(x)\end{aligned}$$

by induction, and similarly

$$\operatorname{Var} f = \sum_{t\geqslant 0} \int_x \operatorname{Var}_{m_x} \mathrm{M}^t f \,\mathrm{d}\nu(x).$$

Since by the lemma above, f is $\mathrm{M}(\mathrm{D}f)(x)\,\mathrm{e}^{4\lambda\sigma_\infty}$-Lipschitz on $\operatorname{Supp} m_x$ and moreover $\mathrm{e}^{8\lambda\sigma_\infty} < 2$, we have

$$\operatorname{Var}_{m_x} f \leqslant \frac{2(\mathrm{M}(\mathrm{D}f)(x))^2\,\sigma(x)^2}{n_x}$$

and, using that $a\log a \leqslant a^2 - a$, we get that $\operatorname{Ent}_{m_x} f \leqslant \frac{1}{\mathrm{M}f(x)} \operatorname{Var}_{m_x} f$ so

$$\operatorname{Ent}_{m_x} f \leqslant \frac{2(\mathrm{M}(\mathrm{D}f)(x))^2\,\sigma(x)^2}{n_x\,\mathrm{M}f(x)}.$$

Thus

$$\operatorname{Var} f \leqslant 2\sum_{t\geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \big(\mathrm{M}\big(\mathrm{D}\mathrm{M}^t f\big)(x)\big)^2 \,\mathrm{d}\nu(x)$$

and

$$\operatorname{Ent} f \leqslant 2\sum_{t\geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \frac{(\mathrm{M}(\mathrm{D}\mathrm{M}^t f)(x))^2}{\mathrm{M}^{t+1} f(x)} \,\mathrm{d}\nu(x).$$

By Theorem 44, we have $(\mathrm{D}\mathrm{M}^t f)(y) \leqslant (1-\kappa/2)^t \mathrm{M}^t(\mathrm{D}f)(y)$, so that

$$\operatorname{Var} f \leqslant 2\sum_{t\geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \big(\mathrm{M}^{t+1}\mathrm{D}f(x)\big)^2 (1-\kappa/2)^{2t} \,\mathrm{d}\nu(x)$$

and

$$\operatorname{Ent} f \leqslant 2\sum_{t\geqslant 0} \int_x \frac{\sigma(x)^2}{n_x} \frac{(\mathrm{M}^{t+1}\mathrm{D}f(x))^2}{\mathrm{M}^{t+1} f(x)} (1-\kappa/2)^{2t} \,\mathrm{d}\nu(x).$$

Now, for variance, convexity of $a \mapsto a^2$ yields

$$\big(\mathrm{M}^{t+1}\mathrm{D}f\big)^2 \leqslant \mathrm{M}^{t+1}\big((\mathrm{D}f)^2\big)$$

and for entropy, convexity of $(a,b) \mapsto a^2/b$ for $a, b > 0$ yields

$$\frac{(\mathrm{M}^{t+1}\mathrm{D}f(x))^2}{\mathrm{M}^{t+1} f(x)} \leqslant \mathrm{M}^{t+1}\left(\frac{(\mathrm{D}f)^2}{f}\right)(x).$$

Finally we get

$$\operatorname{Var} f \leqslant 2\sum_{t\geqslant 0}(1-\kappa/2)^{2t}\int_x \frac{\sigma(x)^2}{n_x}\,\mathrm{M}^{t+1}\big((\mathrm{D}f)^2\big)(x)\,\mathrm{d}\nu(x)$$

and

$$\operatorname{Ent} f \leqslant 2\sum_{t\geqslant 0}(1-\kappa/2)^{2t}\int_x \frac{\sigma(x)^2}{n_x}\,\mathrm{M}^{t+1}\left(\frac{(\mathrm{D}f)^2}{f}\right)(x)\,\mathrm{d}\nu(x).$$

Now, in the non-reversible case, simply apply the identity

$$\int g(x)\,\mathrm{M}^{t+1}h(x)\,\mathrm{d}\nu(x) \leqslant (\sup g)\int \mathrm{M}^{t+1}h(x)\,\mathrm{d}\nu(x) = (\sup g)\int h\,\mathrm{d}\nu$$

to the functions $g(x)=\frac{\sigma(x)^2}{n_x}$ and $h(x)=(\mathrm{D}f)(x)^2$ (for variance) or $h(x)=(\mathrm{D}f)(x)^2/f(x)$ (for entropy). For the reversible case, use the identity

$$\int g(x)\,\mathrm{M}^{t+1}h(x)\,\mathrm{d}\nu(x) = \int h(x)\,\mathrm{M}^{t+1}g(x)\,\mathrm{d}\nu(x)$$

instead. □

5. Exponential concentration in non-negative curvature

We have seen that positive coarse Ricci curvature implies a kind of Gaussian concentration. We now show that non-negative coarse Ricci curvature and the existence of an "attracting point" imply exponential concentration.

The basic example to keep in mind is the following. Let $\mathbb{N}$ be the set of non-negative integers equipped with its standard distance. Let $0<p<1$ and consider the nearest-neighbor random walk on $\mathbb{N}$ that goes to the left with probability p and to the right with probability $1-p$; explicitly $m_k = p\delta_{k-1}+(1-p)\delta_{k+1}$ for $k\geqslant 1$, and $m_0 = p\delta_0+(1-p)\delta_1$.

Since for $k\geqslant 1$ the transition kernel is translation-invariant, it is immediate to check that $\kappa(k,k+1)=0$; besides, $\kappa(0,1)=p$. There exists an invariant distribution if and only if $p>1/2$, and it satisfies exponential concentration with decay distance $1/\log(p/(1-p))$. For $p=1/2+\varepsilon$ with small ε this behaves like $1/4\varepsilon$. Of course, when $p\leqslant 1/2$, there is no invariant distribution so that non-negative curvature alone does not imply concentration of measure.

Geometrically, what entails exponential concentration in this example is the fact that, for $p>1/2$, the point 0 "pulls" its neighbor, and the pulling is transmitted by non-negative curvature. We now formalize this situation in the following theorem.

Theorem 49. *Let $(X,d,(m_x))$ be a metric space with random walk. Suppose that for some $o\in X$ and $r>0$ one has*:

- *$\kappa(x,y)\geqslant 0$ for all $x,y\in X$,*
- *for all $x\in X$ with $r\leqslant d(o,x)<2r$, one has $W_1(m_x,\delta_o)<d(x,o)$,*
- *X is r-geodesic,*

- *there exists $s > 0$ such that each measure m_x satisfies the Gaussian-type Laplace transform inequality*

$$\mathbb{E}_{m_x} e^{\lambda f} \leqslant e^{\lambda^2 s^2/2} e^{\lambda \mathbb{E}_{m_x} f}$$

for any $\lambda > 0$ and any 1-Lipschitz function $f : \operatorname{Supp} m_x \to \mathbb{R}$.

Set $\rho = \inf\{d(x,o) - W_1(m_x, \delta_o),\ r \leqslant d(o,x) < 2r\}$ and assume $\rho > 0$.

Then there exists an invariant distribution for the random walk. Moreover, setting $D = s^2/\rho$ and $m = r + 2s^2/\rho + \rho(1 + J(o)^2/4s^2)$, for any invariant distribution ν we have

$$\int e^{d(x,o)/D} \, d\nu(x) \leqslant \left(4 + J(o)^2/s^2\right) e^{m/D}$$

and so for any 1-Lipschitz function $f : X \to \mathbb{R}$ and $t \geqslant 0$ we have

$$\Pr\left(\left|f - f(o)\right| \geqslant t + m\right) \leqslant \left(8 + 2J(o)^2/s^2\right) e^{-t/D}.$$

So we get exponential concentration with characteristic decay distance s^2/ρ.

The last assumption is always satisfied with $s = 2\sigma_\infty$ (Proposition 1.16 in [29]).

Examples. Before proceeding to the proof, let us show how this applies to the geometric distribution above on $\mathbb{N}$. We take of course $o = 0$ and $r = 1$. We can take $s = 2\sigma_\infty = 2$. Now there is only one point x with $r \leqslant d(o,x) < 2r$, which is $x = 1$. It satisfies $m_1 = p\delta_0 + (1-p)\delta_2$, so that $W_1(m_1, \delta_0) = 2(1-p)$, which is smaller than $d(0,1) = 1$ if and only if $p > 1/2$ as was to be expected. So we can take $\rho = 1 - 2(1-p) = 2p - 1$. Then we get exponential concentration with characteristic distance $4/(2p-1)$. When p is very close to 1 this is not so good (because the discretization is too coarse), but when p is close to $1/2$ this is within a factor 2 of the optimal value.

Another example is the stochastic differential equation $dX_t = S\,dB_t - \alpha \frac{X_t}{|X_t|}\,dt$ on $\mathbb{R}^n$, for which $\exp(-2|x|\alpha/S^2)$ is a reversible measure. Take as a Markov chain the Euler approximation scheme at time δt for this stochastic differential equation, as in Example 11. Taking $r = nS^2/\alpha$ yields that $\rho \geqslant \alpha\,\delta t/2$ after some simple computation. Since we have $s^2 = S^2\delta t$ for Gaussian measures at time δt, we get exponential concentration with decay distance $2S^2/\alpha$, which is correct up to a factor 4. The additive constant in the deviation inequality is $m = r + \rho(1 + J(o)^2/4s^2) + 2s^2/\rho$ which is equal to $(n+4)S^2/\alpha + O(\delta t)$ (note that $J(o)^2 \approx ns^2$). For comparison, the actual value for the average distance to the origin under the exponential distribution $e^{-2|x|\alpha/S^2}$ is $nS^2/2\alpha$, so that up to a constant the dependency on dimension is recovered.

In general, the invariant distribution is not unique under the assumptions of the theorem. For example, start with the random walk on $\mathbb{N}$ above with geometric invariant distribution; now consider the disjoint union $\mathbb{N} \cup (\mathbb{N} + \frac{1}{2})$ where on $\mathbb{N} + \frac{1}{2}$ we use the same random walk translated by $\frac{1}{2}$: the assumptions are satisfied with $r = 1$ and $o = 0$, but clearly there are two disjoint invariant distributions. However, if $\kappa > 0$ in some large enough ball around o, then the invariant distribution will be unique.

Proof of Theorem 49. Let us first prove a lemma which shows how non-negative curvature transmits the "pulling."

Lemma 50. *Let $x \in X$ with $d(x,o) \geqslant r$. Then $W_1(m_x, o) \leqslant d(x,o) - \rho$.*

Proof. If $d(o,x) < 2r$ then this is one of the assumptions. So we suppose that $d(o,x) \geqslant 2r$.

Since X is r-geodesic, let $o = y_0, y_1, y_2, \dots, y_n = x$ be a sequence of points with $d(y_i, y_{i+1}) \leqslant r$ and $\sum d(y_i, y_{i+1}) = d(o,x)$. We can assume that $d(o,y_2) > r$ (otherwise, remove y_1). Set $z = y_1$ if $d(o,y_1) = r$ and $z = y_2$ if $d(o,y_1) < r$, so that $r \leqslant d(o,z) < 2r$. Now

$$\begin{aligned} W_1(\delta_o, m_x) &\leqslant W_1(\delta_o, m_z) + W_1(m_z, m_x) \\ &\leqslant d(o,z) - \rho + d(z,x) \end{aligned}$$

since $\kappa(z,x) \geqslant 0$. But $d(o,z) + d(z,x) = d(o,x)$ by construction, hence the conclusion. □

We are now ready to prove the theorem. The idea is to consider the function $\mathrm{e}^{\lambda d(x,o)}$. For points far away from the origin, since under the random walk the average distance to the origin decreases by ρ by the previous lemma, we expect the function to be multiplied by $\mathrm{e}^{-\lambda\rho}$ under the random walk operator. Close to the origin, the evolution of the function is controlled by the variance s^2 and the jump $J(o)$ of the origin. Since the integral of the function is preserved by the random walk operator, and it is multiplied by a quantity < 1 far away, this shows that the weight of faraway points cannot be too large.

More precisely, we need to tamper a little bit with what happens around the origin. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be defined by $\varphi(x) = 0$ if $x < r$; $\varphi(x) = (x-r)^2/kr$ if $r \leqslant x < r(\frac{k}{2} + 1)$ and $\varphi(x) = x - r - kr/4$ if $x \geqslant r(\frac{k}{2} + 1)$, for some $k > 0$ to be chosen later. Note that φ is a 1-Lipschitz function and that $\varphi'' \leqslant 2/kr$.

If Y is any random variable with values in $\mathbb{R}_+$, we have

$$\mathbb{E}\varphi(Y) \leqslant \varphi(\mathbb{E}Y) + \frac{1}{2}\operatorname{Var} Y \sup \varphi'' \leqslant \varphi(\mathbb{E}Y) + \frac{1}{kr}\operatorname{Var} Y.$$

Now choose some $\lambda > 0$ and consider the function $f : X \to \mathbb{R}$ defined by $f(x) = \mathrm{e}^{\lambda\varphi(d(o,x))}$. Note that $\varphi(d(o,x))$ is 1-Lipschitz, so that by the Laplace transform assumption we have

$$\mathrm{M} f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2} \mathrm{e}^{\lambda \mathrm{M}\varphi(d(o,x))}.$$

The Laplace transform assumption implies that the variance under m_x of any 1-Lipschitz function is at most s^2. So by the remark above, we have

$$\mathrm{M}\varphi\big(d(o,x)\big) \leqslant \varphi\big(\mathrm{M}d(o,x)\big) + \frac{s^2}{kr} = \varphi\big(W_1(m_x, \delta_o)\big) + \frac{s^2}{kr}$$

so that finally

$$\mathrm{M} f(x) \leqslant \mathrm{e}^{\lambda^2 s^2/2 + \lambda s^2/kr} \mathrm{e}^{\lambda\varphi(W_1(m_x,\delta_o))}$$

for any $x \in X$.

We will use different bounds on $\varphi(W_1(m_x,\delta_o))$ according to $d(o,x)$. First, if $d(x,o)<r$, then use non-negative curvature to write $W_1(m_x,\delta_o)\leqslant W_1(m_x,m_o)+J(o)\leqslant d(x,o)+J(o)$ so that $\varphi(W_1(m_x,\delta_o))\leqslant\varphi(r+J(o))\leqslant J(o)^2/kr$ so that

$$\mathrm{M}f(x)\leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr}=\mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr}f(x)$$

since $f(x)=1$.

Second, for any x with $d(x,o)\geqslant r$, Lemma 50 yields

$$\mathrm{M}f(x)\leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr}\mathrm{e}^{\lambda\varphi(d(x,o)-\rho)}.$$

If $d(x,o)\geqslant r(\frac{k}{2}+1)+\rho$ then $\varphi(d(x,o)-\rho)=\varphi(d(x,o))-\rho$ so that

$$\mathrm{M}f(x)\leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr-\lambda\rho}f(x).$$

If $r\leqslant d(x,o)<r(\frac{k}{2}+1)+\rho$, then $\varphi(d(x,o)-\rho)\leqslant\varphi(d(x,o))$ so that

$$\mathrm{M}f(x)\leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr}f(x).$$

Let ν be any probability measure such that $\int f\,\mathrm{d}\nu<\infty$. Let $X'=\{x\in X,\ d(x,o)<r(\frac{k}{2}+1)\}$ and $X''=X\setminus X'$. Set $A(\nu)=\int_{X'}f\,\mathrm{d}\nu$ and $B(\nu)=\int_{X''}f\,\mathrm{d}\nu$. Combining the cases above, we have shown that

$$\begin{aligned}A(\nu*m)+B(\nu*m)&=\int f\,\mathrm{d}(\nu*m)=\int \mathrm{M}f\,\mathrm{d}\nu\\&=\int_{X'}\mathrm{M}f\,\mathrm{d}\nu+\int_{X''}\mathrm{M}f\,\mathrm{d}\nu\\&\leqslant \mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr}\int_{X'}f\,\mathrm{d}\nu+\mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr-\lambda\rho}\int_{X''}f\,\mathrm{d}\nu\\&=\alpha A(\nu)+\beta B(\nu)\end{aligned}$$

with $\alpha=\mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr+\lambda J(o)^2/kr}$ and $\beta=\mathrm{e}^{\lambda^2 s^2/2+\lambda s^2/kr-\lambda\rho}$.

Choose λ small enough and k large enough (see below) so that $\beta<1$. Using that $A(\nu)\leqslant \mathrm{e}^{\lambda kr/4}$ for any probability measure ν, we get $\alpha A(\nu)+\beta B(\nu)\leqslant(\alpha-\beta)\mathrm{e}^{\lambda kr/4}+\beta(A(\nu)+B(\nu))$. In particular, if $A(\nu)+B(\nu)\leqslant\frac{(\alpha-\beta)\mathrm{e}^{\lambda kr/4}}{1-\beta}$, we get $\alpha A(\nu)+\beta B(\nu)\leqslant\frac{(\alpha-\beta)\mathrm{e}^{\lambda kr/4}}{1-\beta}$ again. So setting $R=\frac{(\alpha-\beta)\mathrm{e}^{\lambda kr/4}}{1-\beta}$, we have just shown that the set C of probability measures ν such that $\int f\,\mathrm{d}\nu\leqslant R$ is invariant under the random walk.

Moreover, if $A(\nu)+B(\nu)>R$ then $\alpha A(\nu)+\beta B(\nu)<A(\nu)+B(\nu)$. Hence, if ν is an invariant distribution, necessarily $\nu\in C$. This, together with an evaluation of R given below, will provide the bound for $\int f\,\mathrm{d}\nu$ stated in the theorem.

We now turn to existence of an invariant distribution. First, C is obviously closed and convex. Moreover, C is tight: indeed if K is a compact, say included in a ball of radius a around o, then for any $\nu\in C$ we have $\nu(X\setminus K)\leqslant R\mathrm{e}^{-\lambda a}$. So by Prokhorov's theorem, C is compact in the weak convergence topology. So C is compact convex in the topological vector space of all

(signed) Borel measures on X, and is invariant by the random walk operator, which is an affine map. By the Markov–Kakutani theorem (Theorem I.3.3.1 in [23]), it has a fixed point.

Let us finally evaluate R. We have

$$R = \frac{\alpha/\beta - 1}{1/\beta - 1}\,\mathrm{e}^{\lambda k r/4} = \frac{\mathrm{e}^{\lambda J(o)^2/kr + \lambda\rho} - 1}{\mathrm{e}^{\lambda\rho - \lambda s^2/kr - \lambda^2 s^2/2} - 1}\,\mathrm{e}^{\lambda k r/4}$$
$$\leqslant \frac{\rho + J(o)^2/kr}{\rho - s^2/kr - \lambda s^2/2}\,\mathrm{e}^{\lambda J(o)^2/kr + \lambda\rho + \lambda k r/4}$$

using $\mathrm{e}^a - 1 \leqslant a\mathrm{e}^a$ and $\mathrm{e}^a - 1 \geqslant a$.

Now take $\lambda = \rho/s^2$ and $k = 4s^2/r\rho$. This yields

$$R \leqslant \left(4 + J(o)^2/s^2\right)\mathrm{e}^{\lambda(s^2/\rho + \rho(1 + J(o)^2/4s^2))}.$$

Let ν be some invariant distribution: it satisfies $\int f\,\mathrm{d}\nu \leqslant R$. Since $d(x,o) \leqslant \varphi(d(x,o)) + r(1+k/4)$ we have $\int \mathrm{e}^{\lambda d(x,o)}\,\mathrm{d}\nu \leqslant \mathrm{e}^{\lambda r(1+k/4)} \int f\,\mathrm{d}\nu \leqslant R\mathrm{e}^{\lambda r(1+k/4)}$, hence the result in the theorem. □

6. L^2 Bonnet–Myers theorems

As seen in Section 2.3, it is generally not possible to give a bound for the diameter of a positively curved space similar to the usual Bonnet–Myers theorem involving the square root of curvature, the simplest counterexample being the discrete cube. Here we describe additional conditions which provide such a bound in two different kinds of situation.

We first give a bound on the *average* distance between two points rather than the diameter; it holds when there is an "attractive point" and is relevant for examples such as the Ornstein–Uhlenbeck process (Example 9) or its discrete analogue (Example 10).

Next, we give a direct generalization of the genuine Bonnet–Myers theorem for Riemannian manifolds. Despite lack of further examples, we found it interesting to provide an axiomatization of the Bonnet–Myers theorem in our language. This is done by reinforcing the positive curvature assumption, which compares the transportation distance between the measures issuing from two points x and y at a given time, by requiring a transportation distance inequality between the measures issuing from two given points at *different* times.

6.1. Average L^2 Bonnet–Myers

We now describe a Bonnet–Myers–like estimate on the average distance between two points, provided there is some "attractive point." The proof is somewhat similar to that of Theorem 49.

Proposition 51 *(Average L^2 Bonnet–Myers). Let $(X, d, (m_x))$ be a metric space with random walk, with coarse Ricci curvature at least $\kappa > 0$. Suppose that for some $o \in X$ and $r \geqslant 0$, one has*

$$W_1(\delta_o, m_x) \leqslant d(o,x)$$

for any $x \in X$ with $r \leqslant d(o,x) < 2r$, and that moreover X is r-geodesic.

Then

$$\int d(o,x)\,\mathrm{d}\nu(x) \leqslant \sqrt{\frac{1}{\kappa}\int \frac{\sigma(x)^2}{n_x}\,\mathrm{d}\nu(x)} + 5r$$

where as usual ν is the invariant distribution.

Note that the assumption $\int d(o,y)\,\mathrm{d}m_x(y) \leqslant d(o,x)$ cannot hold for x in some ball around o unless o is a fixed point. This is why the assumption is restricted to an annulus.

As in the Gaussian concentration theorem (Theorem 33), in case $\sigma(x)^2$ is Lipschitz, Corollary 22 may provide a useful bound on $\int \frac{\sigma(x)^2}{n_x}\,\mathrm{d}\nu(x)$ in terms of its value at some point.

As a first example, consider the discrete Ornstein–Uhlenbeck process of Example 10, which is the Markov chain on $\{-N,\dots,N\}$ given by the transition probabilities $p_{k,k} = 1/2$, $p_{k,k+1} = 1/4 - k/4N$ and $p_{k,k-1} = 1/4 + k/4N$; the coarse Ricci curvature is $\kappa = 1/2N$, and the invariant distribution is the binomial $\frac{1}{2^{2N}}\binom{2N}{N+k}$. This example is interesting because the diameter is $2N$ (which is the bound provided by Proposition 23), whereas the average distance between two points is $\approx\sqrt{N}$. It is immediate to check that 0 is attractive, namely that $o = 0$ and $r = 1$ fulfill the assumptions. Since $\sigma(x)^2 \approx 1$ and $\kappa \approx 1/N$, the proposition recovers the correct order of magnitude for distance to the origin.

Our next example is the Ornstein–Uhlenbeck process $\mathrm{d}X_t = -\alpha X_t\,\mathrm{d}t + s\,\mathrm{d}B_t$ on $\mathbb{R}^N$ (Example 9). Here it is clear that 0 is attractive in some sense, so $o = 0$ is a natural choice. The invariant distribution is Gaussian of variance s^2/α; under this distribution the average distance to 0 is $\approx\sqrt{Ns^2/\alpha}$.

At small time τ, a point $x \in \mathbb{R}^N$ is sent to a Gaussian centered at $(1-\alpha\tau)x$, of variance τs^2. The average quadratic distance to the origin under this Gaussian is $(1-\alpha\tau)^2 d(0,x)^2 + Ns^2\tau + o(\tau)$ by a simple computation. If $d(0,x)^2 > Ns^2/2\alpha$ this is less than $d(0,x)^2$, so that we can take $r = \sqrt{Ns^2/2\alpha}$. Considering the random walk discretized at time τ we have we have $\kappa \sim \alpha\tau$, $\sigma(x)^2 \sim Ns^2\tau$ and $n_x \approx N$. So in the proposition above, the first term is $\approx\sqrt{s^2/\alpha}$, whereas the second term is $5r \approx \sqrt{Ns^2/\alpha}$, which is thus dominant. So the proposition gives the correct order of magnitude; in this precise case, the first term in the proposition reflects concentration of measure (which is dimension-independent for Gaussians), whereas it is the second term $5r$ which carries the correct dependency on dimension for the average distance to the origin.

Proof. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be the function defined by $\varphi(x) = 0$ if $x \leqslant 2r$, and $\varphi(x) = (x-2r)^2$ otherwise. For any real-valued random variable Y, we have

$$\mathbb{E}\varphi(Y) \leqslant \varphi(\mathbb{E}Y) + \frac{1}{2}\operatorname{Var} Y \sup\varphi'' = \varphi(\mathbb{E}Y) + \operatorname{Var} Y.$$

Now let $f : X \to \mathbb{R}$ be defined by $f(x) = \varphi(d(o,x))$. We are going to show that

$$\mathrm{M}f(x) \leqslant (1-\kappa)^2 f(x) + \frac{\sigma(x)^2}{n_x} + 9r^2$$

for all $x \in X$. Since $\int f\,\mathrm{d}\nu = \int \mathrm{M}f\,\mathrm{d}\nu$, we will get $\int f\,\mathrm{d}\nu \leqslant (1-\kappa)^2\int f\,\mathrm{d}\nu + \int \frac{\sigma(x)^2}{n_x}\,\mathrm{d}\nu + 9r^2$ which easily implies the result.

First, suppose that $r \leqslant d(o,x) < 2r$. We have $f(x) = 0$. Now $\int d(o,y)\,\mathrm{d}m_x(y)$ is at most $d(o,y)$ by assumption. Using the bound above for φ, together with the definition of $\sigma(x)^2$ and n_x, we get

$$\mathrm{M}f(x) = \int \varphi\big(d(o,y)\big)\,\mathrm{d}m_x(y) \leqslant \varphi\bigg(\int d(o,y)\,\mathrm{d}m_x(y)\bigg) + \frac{\sigma(x)^2}{n_x} = \frac{\sigma(x)^2}{n_x}$$

since $\int d(o,y)\,\mathrm{d}m_x(y) \leqslant 2r$ by assumption.

Second, suppose that $d(x,o) \geqslant 2r$. Using that X is r-geodesic, we can find a point x' such that $d(o,x) = d(o,x') + d(x',x)$ and $r \leqslant d(o,x') < 2r$ (take the second point in a sequence joining o to x). Now we have

$$\begin{aligned}\int d(o,y)\,\mathrm{d}m_x(y) &= W_1(\delta_o, m_x)\\ &\leqslant W_1(\delta_o, m_{x'}) + W_1(m_{x'}, m_x)\\ &\leqslant W_1(\delta_o, m_{x'}) + (1-\kappa)d(x',x)\\ &\leqslant d(o,x') + (1-\kappa)d(x',x) \leqslant (1-\kappa)d(o,x) + 2\kappa r\end{aligned}$$

and as above, this implies

$$\begin{aligned}\mathrm{M}f(x) &\leqslant \varphi\bigg(\int d(o,y)\,\mathrm{d}m_x(y)\bigg) + \frac{\sigma(x)^2}{n_x}\\ &\leqslant \big((1-\kappa)d(o,x) + 2\kappa r - 2r\big)^2 + \frac{\sigma(x)^2}{n_x}\\ &= (1-\kappa)^2\varphi\big(d(o,x)\big) + \frac{\sigma(x)^2}{n_x}\end{aligned}$$

as needed.

The last case to consider is $d(o,x) < r$. In this case we have

$$\begin{aligned}\int d(o,y)\,\mathrm{d}m_x(y) &= W_1(\delta_o, m_x)\\ &\leqslant W_1(\delta_o, m_o) + W_1(m_o, m_x) = J(o) + W_1(m_o, m_x)\\ &\leqslant J(o) + (1-\kappa)d(o,x) \leqslant J(o) + r.\end{aligned}$$

So we need to bound $J(o)$. If X is included in the ball of radius r around o, the result trivially holds, so that we can assume that there exists a point x with $d(o,x) \geqslant r$. Since X is r-geodesic we can assume that $d(o,x) < 2r$ as well. Now $J(o) = W_1(m_o, \delta_o) \leqslant W_1(m_o, m_x) + W_1(m_x, \delta_o) \leqslant (1-\kappa)d(o,x) + W_1(m_x, \delta_o) \leqslant (1-\kappa)d(o,x) + d(o,x)$ by assumption, so that $J(o) \leqslant 4r$.

Plugging this into the above, for $d(o,x) < r$ we get $\int d(o,y)\,\mathrm{d}m_x(y) \leqslant 5r$ so that $\varphi(\int d(o,y)\,\mathrm{d}m_x(y)) \leqslant 9r^2$ hence $\mathrm{M}f(x) \leqslant 9r^2 + \frac{\sigma(x)^2}{n_x}$.

Combining the results, we get that whatever $x \in X$

$$\mathrm{M}f(x) \leqslant (1-\kappa)^2 f(x) + \frac{\sigma(x)^2}{n_x} + 9r^2$$

as needed. □

6.2. Strong L^2 Bonnet–Myers

As mentioned above, positive coarse Ricci curvature alone does not imply a $1/\sqrt{\kappa}$-like diameter control, because of such simple counterexamples as the discrete cube or the Ornstein–Uhlenbeck process. We now extract a property satisfied by the ordinary Brownian motion on Riemannian manifolds (without drift), which guarantees a genuine Bonnet–Myers theorem. Of course, this is of limited interest since the only available example is Riemannian manifolds, but nevertheless we found it interesting to find a sufficient condition expressed in our present language.

Our definition of coarse Ricci curvature controls the transportation distance between the measures issuing from two points x and x' at a given time t. The condition we will now use controls the transportation distance between the measures issuing from two points at two *different* times. It is based on what holds for Gaussian measures in $\mathbb{R}^N$. For any $x, x' \in \mathbb{R}^N$ and $t, t' > 0$, let m_x^{*t} and $m_{x'}^{*t'}$ be the laws of the standard Brownian motion issuing from x at time t and from x' at time t', respectively. It is easy to check that the L^2 transportation distance between these two measures is

$$W_2\big(m_x^{*t}, m_{x'}^{*t'}\big)^2 = d(x,x')^2 + N(\sqrt{t} - \sqrt{t'})^2$$

hence

$$W_1\big(m_x^{*t}, m_{x'}^{*t'}\big) \leqslant d(x,x') + \frac{N(\sqrt{t} - \sqrt{t'})^2}{2d(x,x')}.$$

The important feature here is that, when t' tends to t, the second term is of second order in $t' - t$. This is no more the case if we add a drift term to the diffusion.

We now take this inequality as an assumption and use it to copy the traditional proof of the Bonnet–Myers theorem. Here, for simplicity of notation we suppose that we are given a continuous-time Markov chain; however, the proof uses only a finite number of different values of t, so that discretization is possible (this is important in Riemannian manifolds, because the heat kernel is positive on the whole manifold at any positive time, and there is no simple control on it far away from the initial point; taking a discrete approximation with bounded steps solves this problem).

Proposition 52 *(Strong L^2 Bonnet–Myers). Let X be a metric space equipped with a continuous-time random walk m^{*t}. Assume that X is ε-geodesic, and that there exist constants $\kappa > 0, C \geqslant 0$ such that for any two small enough t, t', for any $x, x' \in X$ with $\varepsilon \leqslant d(x, x') \leqslant 2\varepsilon$ one has*

$$W_1\big(m_x^{*t}, m_{x'}^{*t'}\big) \leqslant \mathrm{e}^{-\kappa \inf(t,t')} d(x, x') + \frac{C(\sqrt{t} - \sqrt{t'})^2}{2d(x, x')}$$

with $\kappa > 0$. Assume moreover that $\varepsilon \leqslant \frac{1}{2}\sqrt{C/2\kappa}$.

Then

$$\operatorname{diam} X \leqslant \pi \sqrt{\frac{C}{2\kappa}} \left(1 + \frac{4\varepsilon}{\sqrt{C/2\kappa}}\right).$$

When $t = t'$, the assumption reduces to $W_1(m_x^{*t}, m_{x'}^{*t}) \leqslant \mathrm{e}^{-\kappa t} d(x, x')$, which is just the continuous-time version of the positive curvature assumption. The constant C plays the role of a diffusion constant, and is equal to N for (a discrete approximation of) Brownian motion on a Riemannian manifold. We restrict the assumption to $d(x, x') \geqslant \varepsilon$ to avoid divergence problems for $\frac{C(\sqrt{t}-\sqrt{t'})^2}{2d(x,x')}$ when $x' \to x$.

For Brownian motion on an N-dimensional Riemannian manifold, we can take $\kappa = \frac{1}{2} \inf \mathrm{Ric}$ by Bakry–Émery theory (the $\frac{1}{2}$ is due to the fact that the infinitesimal generator of Brownian motion is $\frac{1}{2}\Delta$), and $C = N$ as in $\mathbb{R}^N$. So we get the usual Bonnet–Myers theorem, up to a factor $\sqrt{N}$ instead of $\sqrt{N-1}$ (similarly to our spectral gap estimate in comparison with the Lichnerowicz theorem), but with the correct constant π.

Proof. Let $x, x' \in X$. Since X is ε-geodesic, we can find a sequence $x = x_0, x_1, \ldots, x_{k-1}, x_k = x'$ of points in X with $d(x_i, x_{i+1}) \leqslant \varepsilon$ and $\sum d(x_i, x_{i+1}) = d(x_0, x_k)$. By taking a subsequence (denoted x_i again), we can assume that $\varepsilon \leqslant d(x_i, x_{i+1}) \leqslant 2\varepsilon$ instead.

Set $t_i = \eta \sin(\frac{\pi d(x,x_i)}{d(x,x')})^2$ for some (small) value of η to be chosen later. Now, since $t_0 = t_k = 0$ we have

$$\begin{aligned} d(x, x') = W_1(\delta_x, \delta_{x'}) &\leqslant \sum W_1\big(m_{x_i}^{*t_i}, m_{x_{i+1}}^{*t_{i+1}}\big) \\ &\leqslant \sum \mathrm{e}^{-\kappa \inf(t_i, t_{i+1})} d(x_i, x_{i+1}) + \frac{C(\sqrt{t_{i+1}} - \sqrt{t_i})^2}{2d(x_i, x_{i+1})} \end{aligned}$$

by assumption. Now we have $|\sin b - \sin a| = |2 \sin \frac{b-a}{2} \cos \frac{a+b}{2}| \leqslant |b - a||\cos \frac{a+b}{2}|$ so that

$$\frac{C(\sqrt{t_{i+1}} - \sqrt{t_i})^2}{2d(x_i, x_{i+1})} \leqslant \frac{C\eta\pi^2 d(x_i, x_{i+1})}{2d(x, x')^2} \cos^2\left(\pi \frac{d(x, x_i) + d(x, x_{i+1})}{2d(x, x')}\right).$$

Besides, if η is small enough, one has $\mathrm{e}^{-\kappa \inf(t_i, t_{i+1})} = 1 - \kappa \inf(t_i, t_{i+1}) + O(\eta^2)$. So we get

$$\begin{aligned} d(x, x') \leqslant \sum d(x_i, x_{i+1}) &- \kappa \inf(t_i, t_{i+1}) d(x_i, x_{i+1}) \\ &+ \frac{C\eta\pi^2 d(x_i, x_{i+1})}{2d(x, x')^2} \cos^2\left(\pi \frac{d(x, x_i) + d(x, x_{i+1})}{2d(x, x')}\right) + O\big(\eta^2\big). \end{aligned}$$

Now the terms $\sum d(x_i, x_{i+1}) \cos^2(\pi \frac{d(x,x_i)+d(x,x_{i+1})}{2d(x,x')})$ and $\sum \inf(t_i, t_{i+1}) d(x_i, x_{i+1})$ are close to the integrals $d(x,x') \int_0^1 \cos^2(\pi u)\,\mathrm{d}u$ and $d(x,x')\eta \int_0^1 \sin^2(\pi u)\,\mathrm{d}u$ respectively; the relative error in the Riemann sum is easily bounded by $\pi\varepsilon/d(x,x')$ so that

$$d(x,x') \leqslant d(x,x') - \kappa\eta d(x,x')\left(\frac{1}{2} - \frac{\pi\varepsilon}{d(x,x')}\right) + \frac{C\eta\pi^2}{2d(x,x')^2} d(x,x')\left(\frac{1}{2} + \frac{\pi\varepsilon}{d(x,x')}\right) + O\big(\eta^2\big)$$

hence, taking η small enough,

$$d(x,x')^2 \leqslant \frac{C\pi^2}{2\kappa} \frac{1 + 2\pi\varepsilon/d(x,x')}{1 - 2\pi\varepsilon/d(x,x')}$$

so that either $d(x,x') \leqslant \pi\sqrt{C/2\kappa}$, or $2\pi\varepsilon/d(x,x') \leqslant 2\pi\varepsilon/\pi\sqrt{C/2\kappa} \leqslant 1/2$ by the assumption that ε is small, in which case we use $(1+a)/(1-a) \leqslant 1+4a$ for $a \leqslant 1/2$, hence the conclusion. □

7. Coarse Ricci curvature and Gromov–Hausdorff topology

One of our goals was to define a *robust* notion of curvature, not relying on differential calculus or the small-scale structure of a space. Here we first give two remarks about how changes to the metric and the random walk affect curvature. Next, in order to be able to change the underlying space as well, we introduce a Gromov–Hausdorff–like topology for metric spaces equipped with a random walk.

First, since coarse Ricci curvature is defined as a ratio between a transportation distance and a distance, we get the following remark.

Remark 53 *(Change of metric).* Let $(X, d, m = (m_x))$ be a metric space with random walk, and let d' be a metric on X which is bi-Lipschitz equivalent to d, with constant $C \geqslant 1$. Suppose that the coarse Ricci curvature of m on (X,d) is at least κ. Then the coarse Ricci curvature of m on (X,d') is at least κ' where $1-\kappa' = C^2(1-\kappa)$.

As an example, consider the ε-step random walk on a Riemannian manifold with positive Ricci curvature; κ behaves like ε^2 times the usual Ricci curvature, so that small bi-Lipschitz deformations of the metric, smaller than $O(\varepsilon^2)$, will preserve positivity of curvature of the ε-step random walk.

The next remark states that we can deform the random walk $m = (m_x)$ if the deformation depends on x in a Lipschitz way. Given a metric space (X,d), consider the space of 0-mass signed measures $\mathcal{P}_0(X) = \{\mu_+ - \mu_-\}$ where μ_+, μ_- are measures on X with finite first moment and with the same total mass. Equip this space with the norm (it is one) $\|\mu_+ - \mu_-\| := \sup_{f\ 1\text{-Lipschitz}} \int f\,\mathrm{d}(\mu_+ - \mu_-) = W_1(\mu_+, \mu_-)$. Then the following trivially holds.

Remark 54 *(Change of random walk).* Let (X,d) be a metric space and let $m = (m_x)_{x\in X}$, $m' = (m'_x)_{x\in X}$ be two random walks on X. Suppose that the coarse Ricci curvature of m is at least κ, and that the map $x \mapsto m_x - m'_x \in \mathcal{P}_0(X)$ is C-Lipschitz. Then the coarse Ricci curvature of m' is at least $\kappa - 2C$.

We now turn to changes in the space itself, for which we need to give a generalization of Gromov–Hausdorff topology taking the random walk data into account. Two spaces are close in this topology if they are close in the Gromov–Hausdorff topology and if moreover, the measures issuing from each point x are (uniformly) close in the L^1 transportation distance.

Recall [3] that two metric spaces (X, d_X) and (Y, d_Y) are at Gromov–Hausdorff distance at most $e \in [0; \infty]$ if there exists a semi-metric space (Z, d_Z) and isometries $f_X : X \hookrightarrow Z$, $f_Y : Y \hookrightarrow Z$, such that for any $x \in X$, there exists $y \in Y$ with $d_Z(f_X(x), f_Y(y)) \leqslant e$, and likewise for any $y \in Y$ (i.e. the Hausdorff distance between $f_X(X)$ and $f_Y(Y)$ is at most e). We extend this definition as follows to incorporate the random walk.

Definition 55. Let $(X, (m_x)_{x\in X})$ and $(Y, (m_y)_{y\in Y})$ be two metric spaces equipped with a random walk. For $e \in [0; \infty]$, we say that these spaces are e-close if there exists a metric space Z and two isometric embeddings $f_X : X \hookrightarrow Z$, $f_Y : Y \hookrightarrow Z$ such that for any $x \in X$, there exists $y \in Y$ such that $d_Z(f_X(x), f_Y(y)) \leqslant e$ and the L^1 transportation distance between the pushforward measures $f_X(m_x)$ and $f_Y(m_y)$ is at most $2e$, and likewise for any $y \in Y$.

It is easy to see that this is defines a semi-metric on the class of metric spaces equipped with a random walk. We say that a sequence of spaces with random walks $(X^N, (m_x^N)_{x\in X^N})$ *converges* to $(X, (m_x))$ if the semi-distance between $(X^N, (m_x^N))$ and (X, m_x) tends to 0. We say, moreover, that a sequence of points $x^N \in X^N$ tends to $x \in X$ if we can take x^N and x to be corresponding points in the definition above. We give a similar definition for convergence of tuples of points in X^N.

Coarse Ricci curvature is a continuous function in this topology. Namely, a limit of spaces with coarse Ricci curvature at least κ has coarse Ricci curvature at least κ, as expressed in the following proposition.

Proposition 56 *(Gromov–Hausdorff continuity). Let $(X^N, (m_x^N)_{x\in X^N})$ be a sequence of metric spaces with random walk, converging to a metric space with random walk $(X, (m_x)_{x\in X})$. Let x, y be two distinct points in X and let $(x^N, y^N) \in X^N \times X^N$ be a sequence of pairs of points converging to (x, y). Then $\kappa(x^N, y^N) \to \kappa(x, y)$.*

In particular, if all spaces X^N have coarse Ricci curvature at least κ, then so does X. Thus, having coarse Ricci curvature at least κ is a closed property.

Proof. We have $\kappa(x, y) = 1 - \frac{W_1(m_x, m_y)}{d(x,y)}$ and likewise for $\kappa(x^N, y^N)$. The definition ensures that $d(x^N, y^N)$ and $W_1(m_x^N, m_y^N)$ tend to $d(x, y)$ and $W_1(m_x, m_y)$ respectively, hence the result. □

Note however, that the coarse Ricci curvature of $(X, (m_x))$ may be larger than the limsup of the coarse Ricci curvatures of $(X^N, (m_x^N))$, because pairs of points in X^N, contributing to the curvature of X^N, may tend to the same point in X; for example, X may consist of a single point.

This collapsing phenomenon prevents positive curvature from being an *open* property. Yet it is possible to relax the definition of coarse Ricci curvature so as to allow any variation at small scales; with this perturbed definition, having coarse Ricci curvature greater than κ will become an open property. This is achieved as follows (compare the passage from trees to δ-hyperbolic spaces).

Definition 57. Let (X, d) be a metric space equipped with a random walk m. Let $\delta \geqslant 0$. The *coarse Ricci curvature up to δ along $x, y \in X$* is the largest $\kappa \leqslant 1$ for which

$$W_1(m_x, m_y) \leqslant (1-\kappa) d(x, y) + \delta.$$

With this definition, the following is easy.

Proposition 58. *Let $(X, (m_x))$ be a metric space with random walk with coarse Ricci curvature at least κ up to $\delta \geqslant 0$. Let $\delta' > 0$. Then there exists a neighborhood $\mathcal{V}_X$ of X such that any space $Y \in \mathcal{V}_X$ has coarse Ricci curvature at least κ up to $\delta + \delta'$.*

Consequently, the property "having curvature at least κ for some $\delta \in [0; \delta_0)$" is open.

It would be interesting to study which properties of positive coarse Ricci curvature carry to this more general setting.

8. Transportation distance in Riemannian manifolds

Here we give the proofs of Proposition 6 and of the statements of Example 7 and Section 3.3.1.

We begin with Proposition 6 and evaluation of the coarse Ricci curvature of the ε-step random walk. The argument is close to the one in [43, Theorem 1.5(xii)], except that we use the value of Ricci curvature at a given point instead of its infimum on the manifold.

Let X be a smooth N-dimensional Riemannian manifold and let $x \in X$. Let v, w be unit tangent vectors at x. Let $\delta, \varepsilon > 0$ small enough. Let $y = \exp_x(\delta v)$. Let $x' = \exp_x(\varepsilon w)$ and $y' = \exp_y(\varepsilon w')$ where w' is the tangent vector at y obtained by parallel transport of w along the geodesic $t \mapsto \exp_x(tv)$. The first claim is that $d(x', y') = \delta(1 - \frac{\varepsilon^2}{2} K(v, w) + O(\delta\varepsilon^2 + \varepsilon^3))$.

We suppose for simplicity that w and w' are orthogonal to v.

We will work in cylindrical coordinates along the geodesic $t \mapsto \exp_x(tv)$. Let $v_t = \frac{\mathrm{d}}{\mathrm{d}t} \exp_x(tv)$ be the speed of this geodesic. Let E_t be the orthogonal of v_t in the tangent space at $\exp_x(tv)$. Each point z in some neighborhood of x can be uniquely written as $\exp_{\exp_x(\tau(z)v)}(\varepsilon\zeta(z))$ for some $\tau(z) \in \mathbb{R}$ and $\zeta(z) \in E_{\tau(z)}$.

Consider the set $\exp_x(E_0)$ (restricted to some neighborhood of x to avoid topological problems), which contains x'. Let γ be a geodesic starting at some point of $\exp_x(E_0)$ and ending at y', which realizes the distance from $\exp_x(E_0)$ to y'. The distance from x' to y' is at least the length of γ. If δ and ε are small enough, the geodesic γ is arbitrarily close to the Euclidean one so that the coordinate τ is strictly increasing along γ. Let us parametrize γ using the coordinate τ, so that $\tau(\gamma(t)) = t$. Let also $w_t = \zeta(\gamma(t)) \in E_t$. In particular $w_\delta = w'$.

Consider, for each t, the geodesic $c_t : s \mapsto \exp_{\exp_x(tv)}(sw_t)$. We have $\gamma(t) = c_t(\varepsilon)$. For each given t, the vector field $\frac{D}{dt} c_t(s)$ is a Jacobi field along the geodesic $s \mapsto c_t(s)$. The initial conditions of this Jacobi field for $s = 0$ are given by $\frac{D}{\mathrm{d}t} c_t(s)_{|s=0} = v_t$ and $\frac{D}{\mathrm{d}t} \frac{\mathrm{d}}{\mathrm{d}s} c_t(s)_{|s=0} = \frac{D}{\mathrm{d}t} w_t$. Applying the Jacobi equation yields

$$\left|\frac{\mathrm{d}\gamma(t)}{\mathrm{d}t}\right|^2 = \left|\frac{\mathrm{d}c_t(\varepsilon)}{\mathrm{d}t}\right|^2 = |v_t|^2 + 2\varepsilon\langle v_t, \dot{w}_t\rangle + \varepsilon^2|\dot{w}_t|^2 - \varepsilon^2\left\langle R(w_t, v_t)w_t, v_t\right\rangle + O\left(\varepsilon^3\right)$$

where $\dot{w}_t = \frac{D}{\mathrm{d}t} w_t$. But since by definition $w_t \in E_t$, we have $\langle v_t, \dot{w}_t \rangle = 0$. Since moreover $|v_t| = 1$ we get

$$\left| \frac{\mathrm{d}\gamma(t)}{\mathrm{d}t} \right| = 1 + \frac{\varepsilon^2}{2} |\dot{w}_t|^2 - \frac{\varepsilon^2}{2} \big\langle R(w_t, v_t) w_t, v_t \big\rangle + O\big(\varepsilon^3\big).$$

Integrating from $t = 0$ to $t = \delta$ and using that $\langle R(w_t, v_t) w_t, v_t \rangle = K(w, v) + O(\delta)$ yields that the length of the geodesic γ is

$$\delta \left(1 - \frac{\varepsilon^2}{2} K(v, w) + O\big(\varepsilon^3\big) + O\big(\varepsilon^2 \delta\big) \right) + \frac{\varepsilon^2}{2} \int_{t=0}^{\delta} |\dot{w}_t|^2$$

so that the minimal value is achieved for $\dot{w}_t = 0$. But by definition $\dot{w}_t = 0$ means that the geodesic γ starts at x'. So first, we have estimated $d(x', y')$, which proves Proposition 6, and second, we have proven that the distance from y' to $\exp_x(E_0)$ is realized by x' up to the higher-order terms, which we will use below.

Let us now prove the statement of Example 7. Let μ_0, μ_1 be the uniform probability measures on the balls of radius ε centered at x and y respectively. We have to prove that

$$W_1(\mu_0, \mu_1) = d(x, y) \left(1 - \frac{\varepsilon^2}{2(N+2)} \operatorname{Ric}(v, v) \right)$$

up to higher-order terms.

Let μ_0', μ_1' be the images under the exponential map, of the uniform probability measures on the balls of radius ε in the tangent spaces at x and y respectively. So μ_0' is a measure having density $1 + O(\varepsilon^2)$ w.r.t. μ_0, and likewise for μ_1'.

If we average Proposition 6 over w in the ball of radius ε in the tangent space at x, we get that

$$W_1\big(\mu_0', \mu_1'\big) \leqslant d(x, y) \left(1 - \frac{\varepsilon^2}{2(N+2)} \operatorname{Ric}(v, v) \right)$$

up to higher-order terms, since the coupling by parallel transport realizes this value. Indeed, the average of $K(v, w)$ on the unit sphere of the tangent plane at x is $\frac{1}{N} \operatorname{Ric}(v, v)$. Averaging on the ball instead of the sphere yields an $\frac{1}{N+2}$ factor instead.

Now the density of μ_0', μ_1' with respect to μ_0, μ_1 is $1 + O(\varepsilon^2)$. More precisely write $\frac{\mathrm{d}\mu_0'}{\mathrm{d}\mu_0} = 1 + \varepsilon^2 f_0$ and $\frac{\mathrm{d}\mu_1'}{\mathrm{d}\mu_1} = 1 + \varepsilon^2 f_1$ (where f_0 and f_1 can be written very explicitly in terms of the metric and its derivatives). Note that $f_1 = f_0 + O(d(x, y))$, and that moreover f_0 integrates to 0 since both μ_0 and μ_0' are probability measures. Plugging all this in the estimate above, we get the inequality for $W_1(\mu_0, \mu_1)$ up to the desired higher-order terms.

The converse inequality is proven as follows: if f is any 1-Lipschitz function, the L^1 transportation distance between measures μ_0 and μ_1 is at least the difference of the integrals of f under μ_0 and μ_1. Consider the function f equal to the distance of a point to $\exp_x(E_0)$ (taken in some small enough neighborhood of x), equipped with a $-$ sign if the point is not on the same side of E_0 as y. Clearly f is 1-Lipschitz. We computed above a lower bound for f in cylindrical

coordinates, which after integrating yields a lower bound for $W_1(\mu_0', \mu_1')$. Arguments similar to the above turns this into the desired lower bound for $W_1(\mu_0, \mu_1)$.

Finally, let us briefly sketch the proofs of the other statements of Section 3.3.1, namely, evaluation of the diffusion constant and local dimension (Definition 18). Up to a multiplicative factor $1 + O(\varepsilon)$, these can be computed in the Euclidean space.

A simple computation shows that the expectation of the square distance of two points taken at random in a ball of radius ε in $\mathbb{R}^N$ is $\varepsilon^2 \frac{2N}{N+2}$, hence the value $\varepsilon^2 \frac{N}{N+2}$ for the diffusion constant $\sigma(x)^2$.

To evaluate the local dimension $n_x = \frac{\sigma(x)^2}{\sup \mathrm{Var}_{m_x} f, f \text{ 1-Lipschitz}}$ (Definition 18), we have to bound the maximal variance of a 1-Lipschitz function on a ball of radius ε in $\mathbb{R}^N$. We will prove that the local dimension n_x is comprised between $N-1$ and N. A projection to a coordinate axis provides a function with variance $\frac{\varepsilon^2}{N+2}$, so that local dimension is at most N. For the other bound, let f be a 1-Lipschitz function on the ball and let us compute an upper bound for its variance. Take $\varepsilon = 1$ for simplicity. Write the ball of radius 1 as the union of the spheres S_r of radii $r \leqslant 1$. Let $v(r)$ be the variance of f restricted to the sphere S_r, and let $a(r)$ be the average of f on S_r. Then associativity of variances gives

$$\mathrm{Var}\, f = \int_{r=0}^{1} v(r)\, \mathrm{d}\mu(r) + \mathrm{Var}_\mu\, a(r)$$

where μ is the measure on the interval $[0; 1]$ given by $\frac{r^{N-1}}{Z}\, \mathrm{d}r$ with $Z = \int_{r=0}^{1} r^{N-1}\, \mathrm{d}r = \frac{1}{N}$.

Since the variance of a 1-Lipschitz function on the $(N-1)$-dimensional unit sphere is at most $\frac{1}{N}$, we have $v(r) \leqslant \frac{r^2}{N}$ so that $\int_{r=0}^{1} v(r)\, \mathrm{d}\mu(r) \leqslant \frac{1}{N+2}$. To evaluate the second term, note that $a(r)$ is again 1-Lipschitz as a function of r, so that $\mathrm{Var}_\mu\, a(r) = \frac{1}{2} \iint (a(r) - a(r'))^2\, \mathrm{d}\mu(r) \mathrm{d}\mu(r')$ is at most $\frac{1}{2} \iint (r - r')^2\, \mathrm{d}\mu(r) \mathrm{d}\mu(r') = \frac{N}{(N+1)^2(N+2)}$. So finally

$$\mathrm{Var}\, f \leqslant \frac{1}{N+2} + \frac{N}{(N+1)^2(N+2)}$$

so that the local dimension n_x is bounded below by $\frac{N(N+1)^2}{N^2+3N+1} \geqslant N-1$.

Acknowledgments

I would like to thank Vincent Beffara, Fabrice Debbasch, Alessio Figalli, Pierre Pansu, Bruno Sévennec, Romain Tessera and Cédric Villani for numerous inspiring conversations about coarse geometry and Ricci curvature, as well as Thierry Bodineau, Djalil Chafaï, Aldéric Joulin, Shinichi Ohta and Roberto Imbuzeiro Oliveira for useful remarks on the manuscript and bibliographical references. Special thanks to Pierre Py for the two points x and y.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, Sur les Inégalités de Sobolev Logarithmiques, Panor. Synthèses, vol. 10, Société Mathématique de France, 2000.

[2] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations, Comm. Partial Differential Equations 26 (1–2) (2001) 43–100.

[3] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, Grad. Stud. Math., vol. 33, Amer. Math. Soc., 2001.
[4] R. Bubley, M.E. Dyer, Path coupling: A technique for proving rapid mixing in Markov chains, Found. Comput. Sci. (1997) 223–231.
[5] D. Bakry, M. Émery, Hypercontractivité de semi-groupes de diffusion, C. R. Math. Acad. Sci. Paris Sér. I 299 (15) (1984) 775–778.
[6] D. Bakry, M. Émery, Diffusions hypercontractives, in: Séminaire de Probabilités, XIX, 1983/1984, in: Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, pp. 177–206.
[7] M. Berger, A Panoramic View of Riemannian Geometry, Springer, Berlin, 2003.
[8] S. Bobkov, M. Ledoux, On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures, J. Funct. Anal. 156 (2) (1998) 347–365.
[9] P. Brémaud, Markov Chains, Texts Appl. Math., vol. 31, Springer, New York, 1999.
[10] A.-I. Bonciocat, K.-T. Sturm, Mass transportation and rough curvature bounds for discrete spaces, preprint.
[11] M.-F. Chen, Trilogy of couplings and general formulas for lower bound of spectral gap, in: Probability Towards 2000, New York, 1995, in: Lecture Notes in Statist., vol. 128, Springer, New York, 1998, pp. 123–136.
[12] M.-F. Chen, From Markov Chains to Non-Equilibrium Particle Systems, second ed., World Scientific, 2004.
[13] M.-F. Chen, S.-F. Li, Coupling methods for multidimensional diffusion processes, Ann. Probab. 17 (1) (1989) 151–177.
[14] M.-F. Chen, F.-Y. Wang, Application of coupling method to the first eigenvalue on manifold, Sci. China Ser. A 37 (1) (1994) 1–14.
[15] H. Djellout, A. Guillin, L. Wu, Transportation cost-information inequalities and applications to random dynamical systems and diffusions, Ann. Probab. 32 (3B) (2004) 2702–2732.
[16] R.L. Dobrušin, On the condition of the central limit theorem for inhomogeneous Markov chains, Dokl. Akad. Nauk SSSR (N.S.) 108 (1956) 1004–1006 (in Russian).
[17] R. Dobrushin, Central limit theorem for non-stationary Markov chains I, Teor. Veroyatnost. i Primenen. 1 (1956) 72–89 (in Russian).
[18] R.L. Dobrušin, Definition of a system of random variables by means of conditional distributions, Teor. Verojatnost. i Primenen. 15 (1970) 469–497 (in Russian).
[19] R. Dobrushin, Perturbation methods of the theory of Gibbsian fields, in: R. Dobrushin, P. Groeneboom, M. Ledoux (Eds.), Lectures on Probability Theory and Statistics, Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994, in: P. Bernard (Ed.), Lecture Notes in Math., vol. 1648, Springer, Berlin, 1996, pp. 1–66.
[20] R.L. Dobrushin, S.B. Shlosman, Constructive criterion for the uniqueness of Gibbs field, in: J. Fritz, A. Jaffe, D. Szász (Eds.), Statistical Physics and Dynamical Systems, papers from the Second Colloquium and Workshop on Random Fields: Rigorous Results in Statistical Mechanics, held in Köszeg, August 26–September 1, 1984, in: Progress in Phys., vol. 10, Birkhäuser, Boston, 1985, pp. 347–370.
[21] P. Diaconis, L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, Ann. Appl. Probab. 6 (3) (1996) 695–750.
[22] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second ed., Appl. Math., vol. 38, Springer, New York, 1998.
[23] A. Granas, J. Dugundji, Fixed Point Theory, Springer Monogr. Math., Springer, New York, 2003.
[24] É. Ghys, P. de la Harpe, Sur les Groupes Hyperboliques d'Après Mikhael Gromov, Progr. Math., vol. 83, Birkhäuser, 1990.
[25] M. Gromov, V. Milman, A topological application of the isoperimetric inequality, Amer. J. Math. 105 (1983) 843–854.
[26] R.B. Griffiths, Correlations in Ising ferromagnets III, Commun. Math. Phys. 6 (1967) 121–127.
[27] M. Gromov, Isoperimetric inequalities in Riemannian manifolds, in: V. Milman, G. Schechtman (Eds.), Asymptotic Theory of Finite Dimensional Normed Spaces, in: Lecture Notes in Math., vol. 1200, Springer, Berlin, 1986, pp. 114–129.
[28] A. Joulin, Poisson-type deviation inequalities for curved continuous time Markov chains, Bernoulli 13 (3) (2007) 782–798.
[29] M. Ledoux, The Concentration of Measure Phenomenon, Math. Surveys Monogr., vol. 89, Amer. Math. Soc., 2001.
[30] J. Lott, Optimal transport and Ricci curvature for metric-measure spaces, expository manuscript.
[31] J. Lott, C. Villani, Ricci curvature for metric-measure spaces via optimal transport, preprint.
[32] K. Marton, Bounding $\overline{d}$-distance by informational divergence: A method to prove measure concentration, Ann. Probab. 24 (2) (1996) 857–866.
[33] K. Marton, A measure concentration inequality for contracting Markov chains, Geom. Funct. Anal. 6 (3) (1996) 556–571, Geom. Funct. Anal. 7 (3) (1997) 609–613, Erratum.

[34] F. Martinelli, Relaxation times of Markov chains in statistical mechanics and combinatorial structures, in: H. Kesten (Ed.), Probability on Discrete Structures, in: Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 175–262.
[35] V. Milman, G. Schechtman, Asymptotic Theory of Finite Dimensional Normed Spaces, Lecture Notes in Math., vol. 1200, Springer, Berlin, 1986.
[36] M. Hazewinkel, The Online Encyclopaedia of Mathematics, article Wasserstein metric, http://eom.springer.de/W/w120020.htm.
[37] S.-i. Ohta, On the measure contraction property of metric measure spaces, Comment. Math. Helv. 82 (4) (2007) 805–828.
[38] R. Imbuzeiro Oliveira, On the convergence to equilibrium of Kac's random walk on matrices, preprint, arXiv: 0705.2253.
[39] Y. Ollivier, Sharp phase transition theorems for hyperbolicity of random groups, Geom. Funct. Anal. 14 (3) (2004) 595–679.
[40] Y. Ollivier, Ricci curvature of metric spaces, C. R. Math. Acad. Sci. Paris 345 (11) (2007) 643–646.
[41] F. Otto, C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. 173 (2000) 361–400.
[42] Y. Peres, Mixing for Markov chains and spin systems, Lecture Notes, 2005.
[43] M.-K. von Renesse, K.-T. Sturm, Transport inequalities, gradient estimates, and Ricci curvature, Comm. Pure Appl. Math. 68 (2005) 923–940.
[44] M.D. Sammer, Aspects of mass transportation in discrete concentration inequalities, PhD thesis, Georgia institute of technology, 2005, etd.gatech.edu/theses/available/etd-04112005-163457/unrestricted/sammer_marcus_d_200505_phd.pdf.
[45] G. Schechtman, Concentration, results and applications, in: Handbook of the Geometry of Banach Spaces, vol. 2, North-Holland, Amsterdam, 2003, pp. 1603–1634.
[46] K.-T. Sturm, On the geometry of metric measure spaces, Acta Math. 196 (1) (2006) 65–177.
[47] C. Villani, Topics in Optimal Transportation, Grad. Stud. Math., vol. 58, Amer. Math. Soc., Providence, RI, 2003.
[48] C. Villani, Optimal Transport, Old and New, Grundlehren Math. Wiss., vol. 338, Springer, 2008.