Advanced Studies in Pure Mathematics 57, 2010
Probabilistic Approach to Geometry
pp. 343–381

A survey of Ricci curvature for metric spaces and Markov chains

Yann Ollivier

Abstract.

This text is a presentation of the general context and results of [Oll07] and [Oll09], with comments on related work. The goal is to present a notion of Ricci curvature valid on arbitrary metric spaces, such as graphs, and to generalize a series of classical theorems in positive Ricci curvature, such as spectral gap estimates, concentration of measure or log-Sobolev inequalities.

The necessary background (concentration of measure, curvature in Riemannian geometry, convergence of Markov chains) is covered in the first section. Special emphasis is put on open questions of varying difficulty.

Our starting point is the following: Is there a common geometric feature between the N-dimensional sphere S^N, the discrete cube $\{0,1\}^N$, and the space $\mathbb{R}^N$ equipped with a Gaussian measure? For a start, all three spaces exhibit the *concentration of measure* phenomenon; moreover, it is known (Dvoretzky theorem) that random small-dimensional sections of the cube are close to a sphere, and small-dimensional projections of either the sphere or the cube give rise to nearly-Gaussian measures.

So one can wonder whether there exists a common underlying geometric property. A hint is given by the Gromov–Lévy theorem [Gro86], which states that Gaussian concentration occurs not only for the N-dimensional sphere, but for all Riemannian manifolds of *positive curvature* in the sense that their *Ricci curvature* is at least that of the sphere. In Riemannian geometry, Ricci curvature is the relevant notion in a series of positive-curvature theorems (see section 1.2).

One is left with the problem of finding a definition of Ricci curvature valid for spaces more general than Riemannian manifolds. Moreover,

Received December 14, 2008.
2000 *Mathematics Subject Classification.* 51F99, 53B21, 60B99.
Key words and phrases. Ricci curvature, concentration of measure, Markov chains, Wasserstein distances, metric measure spaces.

the definition should be local and not global, since the idea of curvature is to seek local properties entailing global constraints. A first step in this direction is provided by Bakry–Émery theory [BE85], which allows to define the Ricci curvature of a diffusion process on a Riemannian manifold (or equivalently, of a second-order differential operator); when the diffusion is the ordinary Brownian motion, this gives back usual Ricci curvature. When applied to the natural process on $\mathbb{R}^N$ associated with the Gaussian measure, this yields a positive curvature for the Gaussian space.

Because the Bakry–Émery definition involves differential calculus, it is not readily adaptable to discrete spaces. To deal with the next basic example, the discrete cube, one has to drop the continuity aspect and deal with more "robust" or "coarse" notions that forget the small-scale properties of the underlying space. This is similar in spirit to what has been done for a long time in the (very different) world of *negative curvature*, for which coarse notions such as δ-hyperbolicity and CAT(0) spaces have been developed.

Such a notion can be summarized as follows [Oll07, Oll09]: a metric space has positive curvature if *small balls are closer than their centers are* (Definition 18). Here one uses *transportation distances* to measure the distance between balls.

It is possible to put emphasis on a random process (consistently with Bakry–Émery theory) and replace the ball centered at a point with the transition probability of a random walk. Doing so, one finds that the property above is equivalent to a property first introduced by Dobrushin [Dob70, DS85] for Markov fields, and still known in the Ising community as the "Dobrushin criterion" (several variants of which are in use). The 1970 Dobrushin paper was actually the one to make transportation distances known to a wider audience.

Dobrushin's contraction property in transportation distance for Markov chains can be seen as a metric version of the more well-known *Dobrushin ergodic coefficient* (see e.g. Section 6.7.1 in the textbook [Bré99]). It is, by Kantorovich duality, equivalent to a Lipschitz contraction property for a semi-group, a fundamental feature of Bakry–Émery theory. Under one form or the other, this property pops out sporadically in the Markov chain literature [CW94, Dob96, BD97, Che04, Oli], generally to get rates of convergence. Note its use in [DGW04] to propagate a strong functional inequality from local to global level (thus getting concentration if this inequality holds locally). More recently, in an approach somewhat similar to ours, Joulin uses it under the name "Wasserstein

curvature" to get concentration results for the time-t and invariant measure [Jou07] as well as for the empirical measure [Jou].

Part 1. Basics: concentration, curvature, Markov chains, transportation distances

We now turn to the background material needed in this course: concentration of measure, curvature of Riemannian manifolds, convergence of discrete Markov chains, transportation distances. We will try to keep the exposition simple and informal. Good beginner's guides are as follows: [Sch01, Oll, Led01, Mas07] for concentration; [DC92] or [Pet06] for Riemannian geometry; [Mar04, ABCFGMRS00, DS96] for convergence of Markov chains; [Vil03] for optimal transport.

§1.1. Concentration of measure

The first occurrence of the concentration of measure phenomenon is generally attributed to Lévy [Lév22], who noted that, in the Euclidean unit sphere S^N of large dimension N, a neighborhood of the equator of size roughly $1/\sqrt{N}$ contains most of the mass of the sphere (for the natural volume measure).

This means that if we take a function $f : S^N \to \mathbb{R}$ which is the orthogonal projection on a coordinate axis, then for most points of S^N the value of f is close to 0 (roughly up to $1/\sqrt{N}$). But concentration of measure is much more general: indeed the above applies to any Lipschitz function f, not only the projection to a coordinate axis. The precise quantitative meaning of "most" is a Gaussian control as follows.

Theorem 1 (Concentration on the sphere, [Lév22]). *Let $S^N \subset \mathbb{R}^{N+1}$ be the Euclidean unit sphere. Let $f : S^N \to \mathbb{R}$ be a 1-Lipschitz function. Then there exists a $m \in \mathbb{R}$ such that, for any $t \geqslant 0$*

$$\nu\left(\{x \in S^N,\ |f(x) - m| \geqslant t\}\right) \leqslant 2\exp -\frac{t^2}{2D^2}$$

where $D = 1/\sqrt{N-1}$ and ν is the natural measure on S^N, normalized so that $\nu(S^N) = 1$.

Exercise 2.
Prove the theorem. (*Hint:* Use the fact that, of all parts of S^N with measure $1/2$, half-spheres are those that minimize the boundary length and the measure of their ε-neighborhood. Take for m a median of f. Then use an estimate of $\int_u^{\pi/2} \cos^{N-1}(s)\mathrm{d}s$.)

This means that something in the geometry of the sphere forces Lipschitz functions to be constant. Our goal is to convince the reader that it is positive Ricci curvature, as suggested by the Gromov–Lévy theorem.

Another space on which concentration of measure occurs is the discrete cube $X = \{0,1\}^N$. Equip this space with the uniform probability measure, which means that we pick at random a sequence of 0's and 1's with probability $1/2$. Let $f : X \to \mathbb{R}$ be the function which maps each sequence to the proportion of 1's it contains. It is well-known that f is "most of the time" equal to $1/2$, and that the deviations behave like $1/\sqrt{N}$ for large N and take a Gaussian shape. But in fact, concentration of measure states that this happens for a much wider class of functions, not only the "linear" f above.

Theorem 3 (Concentration on the cube). *Let $X = \{0,1\}^N$ be the discrete cube equipped with the uniform probability measure ν. Let $f : X \to \mathbb{R}$ be a function such that, whenever one digit of the sequence is changed, then the value of f changes by at most $1/N$. Then there exists a $m \in \mathbb{R}$ such that, for any $t \geqslant 0$*

$$\nu\left(\{x \in X,\ |f(x) - m| \geqslant t\}\right) \leqslant 2\exp -\frac{t^2}{2D^2}$$

where $D = 1/2\sqrt{N}$.

The proof uses an important Laplace transform (i.e. exponential moments) technique which is expressed in the following lemma.

Lemma 4 (Laplace transform). *Let X be a space equipped with a probability measure ν. Let $f : X \to \mathbb{R}$. Assume that there exists some $D > 0$ such that, for any $\lambda \in \mathbb{R}$ one has*

$$\mathbb{E}\mathrm{e}^{\lambda(f-\mathbb{E}f)} \leqslant \mathrm{e}^{D^2\lambda^2/2}$$

where $\mathbb{E}$ denotes integration w.r.t. ν. Then for any $t \in \mathbb{R}$ we have

$$\nu(\{x \in X,\ f(x) - \mathbb{E}f \geqslant t\}) \leqslant \mathrm{e}^{-t^2/2D^2}.$$

Exercise 5.
Prove the lemma. (*Hint:* Markov inequality applied to $\mathrm{e}^{\lambda f}$ for some λ.)

Laplace transforms may appear mysterious at first glance. Observe that for small λ we have $\mathbb{E}\mathrm{e}^{\lambda(f-\mathbb{E}f)} = 1 + \lambda\mathbb{E}(f - \mathbb{E}f) + \frac{\lambda^2}{2}\mathbb{E}(f - \mathbb{E}f)^2 +$

$O(\lambda^3) = 1 + \frac{\lambda^2}{2} \operatorname{Var} f + O(\lambda^3) = \mathrm{e}^{(\operatorname{Var} f)\,\lambda^2/2 + O(\lambda^3)}$. So the D^2 appearing in the assumption is a kind of "exponential variance" for f.

Exercise 6.
Prove the theorem. (*Hint:* Work on the Laplace transform estimate. For $N = 1$ this results from a Taylor expansion, and then use induction on N by performing the integration w.r.t. the last coordinate only.)

This tensorization property of Laplace transforms make them a very convenient tool.

The relationship between the concentration theorems on the sphere and cube may be formalized by turning $\{0,1\}^N$ into a metric space. Let us say that the distance between two length-N sequences of 0's and 1's is the number of digits to change to go from one to the other. (This is the graph metric on the edges of the hypercube, or the ℓ^1 metric on $\{0,1\}^N$, also called Hamming metric.) Then the constraint on f in the theorem above simply states that f is $1/N$-Lipschitz. If one rescales the cube metric by $1/N$ so that the diameter is 1 (for better comparison with the unit sphere), the constraint on f is simply to be 1-Lipschitz, so that the theorems on the sphere and cube parallel each other very well.

This shows why concentration is often described as a "metric measure" phenomenon, expressed in terms of Lipschitz functions. The quantity D is, in the terminology of Gromov [Gro99], the "observable diameter" of the space.

There are numerous generalizations to the theorems above. We refer to [Led01, Mas07].

Let us just mention our third basic example: $\mathbb{R}^N$ equipped with the Gaussian probability measure $\nu(\mathrm{d}x) = \mathrm{e}^{-N|x|^2/2}/Z$ where Z is the normalization constant. We have chosen the parameters such that $\mathbb{E}\,|x|^2 = 1$, for better comparison with the unit sphere. Then the same theorem holds: for any 1-Lipschitz function $f : \mathbb{R}^N \to \mathbb{R}$, deviations from the average are controlled by $\mathrm{e}^{-Nt^2/2}$.

The traditional proof uses an isoperimetric inequality stating that in Gaussian space, among all subsets of measure $1/2$, those with the smallest neighborhoods are half-spaces. The passage from isoperimetry to concentration is then similar to that for the sphere above. Another proof is possible with Bakry–Émery theory, by using the Ornstein–Uhlenbeck process, which is the natural process on $\mathbb{R}^N$ having the Gaussian measure as its invariant distribution; actually this process has positive Ricci curvature in the Bakry–Émery sense.

§1.2. Ricci curvature of Riemannian manifolds

Riemannian manifolds. Manifolds are the natural higher-dimensional generalization of curves and surfaces. Any (smooth) manifold can be seen as a subset of $M \subset \mathbb{R}^p$ such that, at each point $x \in M$, there is a N-dimensional subspace of $\mathbb{R}^p$, the *tangent space* T_xM, such that T_xM coincides with M around x up to a second order error. The number N is the *dimension* of the manifold.

Note that if $c(t)$ is a smooth curve in M, then the derivative $\mathrm{d}c(t)/\mathrm{d}t$ is an element of the tangent space at $c(t)$.

It is possible to give an abstract version of this definition without resorting to subspaces of $\mathbb{R}^p$; according to a theorem of Whitney this amounts to the same.

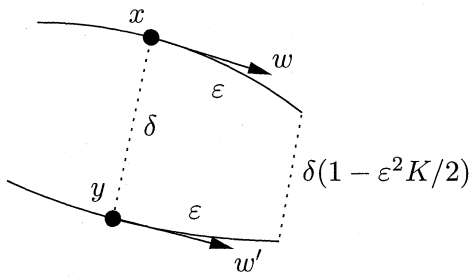
A *Riemannian manifold* is a manifold equipped with a way to measure the length of tangent vectors. Namely, suppose that for each $x \in M$ we are given (in a smooth way) a positive definite quadratic form on the tangent space T_xM; this consitutes a *metric.* (For example, if $M \subset \mathbb{R}^p$, then one can take the restriction to T_xM of the Euclidean structure on $\mathbb{R}^p$.) Then, if $c(t)$ is a curve in M, we can use the quadratic form on $T_{c(t)}M$ to define the length $\|\mathrm{d}c(t)/\mathrm{d}t\|$ of a vector tangent to c. Hence, by integration, we can define the length of a curve. One can then turn the manifold M into a metric space by defining the distance between two points to be the infimum of the lengths of the curves in M between the two points. (If $M \subset \mathbb{R}^p$ and if we use the Euclidean structure of $\mathbb{R}^p$ as the metric, then we get the usual length for curves included in M.)

We will always assume that our Riemannian manifolds are connected and complete.

A *geodesic* is a curve in M such that, for any two close enough points on the curve, the distance between these two points is realized by the curve. For example, the equator and meridians are geodesics of the sphere, but the parallels are not. Locally, geodesics between two points always exist. Moreover, given a starting point x and a tangent vector v, there is always a geodesic starting at x with tangent vector v; it is the curve which "goes straight" in M starting with direction v. We will call *endpoint of* v the point obtained after following this geodesic for a unit time and we will denote it by $\exp_x v$.

Intuition for Ricci curvature. A central notion of Riemannian geometry is *curvature.* Let $x \in M$ and let v be a tangent vector at x, with very small norm $\delta = \|v\|$. Let y be the endpoint of v, in particular the distance $d(x, y)$ is δ. Let w be another tangent vector at x, with very small norm ε; for simplicity we assume that w is orthogonal to v. Let w' be "the same" tangent vector as w, but with basepoint y; this can

be defined as the tangent vector at y whose endpoint is closest to the endpoint of w, with the constraint that w' be orthogonal to the geodesic from x to y, as w is. (Exercise: relate this to the usual definition of parallel transport and Levi–Civita connection.)



Now we are interested in the distance between the endpoint of w and that of w'. If we were in a Euclidean space, we would simply get a rectangle and so the length of the fourth side would be exactly δ. Now, consider the case of the sphere: if, say, x and y are close points on the equator, the geodesics issuing from x and y are meridians that meet at the poles. So, intuitively, in positive curvature geodesics get closer, and farther away in negative curvature. Using the distance between the endpoints of w and w' as a way to quantify this phenomenon we get:

Proposition-Definition 7 (Sectional curvature). Let (X, d) be a smooth complete Riemannian manifold. Let v, w be unit tangent vector at $x \in X$. Let $\varepsilon, \delta > 0$. Let y be the endpoint of δv and let w' be the tangent vector at y obtained by parallel transport of w along the geodesic from x to y. Then

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

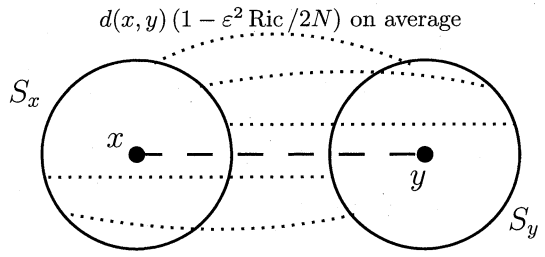
as $(\varepsilon, \delta) \to 0$. Here $K(v, w)$ is the *sectional curvature* in the directions (v, w).

Exercise 8.
Prove the proposition using the classical definition of sectional curvature.

We are now ready to define Ricci curvature.

Definition 9 (Ricci curvature). Let x be a point in a smooth N-dimensional Riemannian manifold, and let v be a tangent vector at

x. We define $\mathrm{Ric}(v)$, the *Ricci curvature along* v, as N times the average of $K(v,w)$ when w runs over the unit sphere in the tangent plane at x.



The N factor comes from the fact that Ricci curvature is traditionally defined as a trace, hence a sum on a basis instead of an average on the sphere. Moreover, generally the Ricci curvature arises as a quadratic form so that $\mathrm{Ric}(v)$ is usually denoted $\mathrm{Ric}(v,v)$.

As a consequence we get (see also condition (*xii*) in [RS05], Theorem 1.5, which uses an infimum of Ricci curvature instead):

Corollary 10 (Transport of Riemannian balls). *Let (X,d) be a smooth complete Riemannian manifold. Let v be a unit tangent vector at $x \in X$. Let $\varepsilon, \delta > 0$. Let y be the endpoint of δv.*

Let S_x be the set of endpoints of the sphere of radius ε in the tangent plane at x, and similarly for y. Then if we map S_x to S_y using parallel transport, on average points travel over a distance

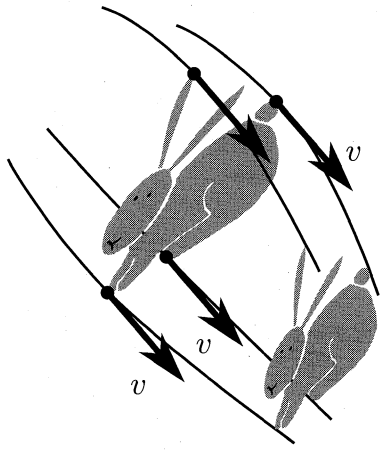
$$\delta\left(1-\frac{\varepsilon^2}{2N}\,\mathrm{Ric}(v)+O(\varepsilon^3+\varepsilon^2\delta)\right)$$

as $(\varepsilon,\delta)\to 0$. If we use balls instead of spheres, the $\frac{\varepsilon^2}{2N}$ factor becomes $\frac{\varepsilon^2}{2(N+2)}$.

This is the characterization of Ricci curvature we will use in more general spaces.

Finally, let us mention, for the record, another visual characterization of Ricci curvature. Consider once again a point $x \in M$ and a unit tangent vector v at x. Consider a very small neighborhood C of any shape around x. For each point z of C, throw a geodesic z_t starting at z with initial direction v (where v has been moved from x to z by parallel transport). As we have seen, on average these geodesics tend to get closer or farther away from the geodesic starting at x, according to the sign of curvature. Let C_t be the set obtained by "gliding" C along

these geodesics for a time t, i.e. the union $\{z_t,\ z \in C\}$. In particular $C = C_0$.



Then we have (Exercise)

$$\operatorname{vol} C_t = \operatorname{vol} C\,(1 - \tfrac{t^2}{2}\operatorname{Ric}(v) + \text{smaller terms})$$

so that Ricci curvature controls the evolution of volumes under the geodesic flow. (Note that the derivative of $\operatorname{vol} C_t$ is 0 for $t = 0$ because we choose geodesics with parallel initial speeds.)

Theorems in positive Ricci curvature. A lot of positive curvature theorems in Riemannian geometry take the form of a condition on Ricci curvature, compared to the reference positively curved space S^N. Let us mention a few of them, which will serve as benchmarks for general notions of Ricci curvature. The sentence "the Ricci curvature is at least that of S^N" means that for each unit tangent vector v, the value $\operatorname{Ric}(v)$ is at least that obtained on a unit sphere (namely $N-1$). By comparing to a sphere of different radius and rescaling, the same bounds apply whenever Ricci curvature is bounded below by any positive number.

Theorem 11 (Bonnet–Myers). *Let M be an N-dimensional Riemannian manifold. Suppose that the Ricci curvature of M is at least that of S^N. Then the diameter of M is at most that of S^N. In particular M is compact.*

Theorem 12 (Lichnerowicz). *Let M be an N-dimensional Riemannian manifold. Suppose that the Ricci curvature of M is at least*

that of S^N. Then the first non-zero eigenvalue of the Laplace–Beltrami operator on M is at least that of S^N.

Theorem 13 (Gromov–Lévy). *Let M be an N-dimensional Riemannian manifold. Suppose that the Ricci curvature of M is at least that of S^N.*

Let $A \subset M$, and let $A' \subset S^N$ be a ball of radius r around some point of S^N, where r is such that $\operatorname{vol} A / \operatorname{vol} M = \operatorname{vol} A' / \operatorname{vol} S^N$.

Then for any $\varepsilon \geqslant 0$ one has

$$\frac{\operatorname{vol} A_\varepsilon}{\operatorname{vol} M} \geqslant \frac{\operatorname{vol} A'_\varepsilon}{\operatorname{vol} S^N}$$

where A_ε is the set of points at distance at most ε from A.

In particular, if A has smooth boundary then

$$\frac{\operatorname{vol}_{N-1} \partial A}{\operatorname{vol} M} \geqslant \frac{\operatorname{vol}_{N-1} \partial A'}{\operatorname{vol} S^N}.$$

This means that relative volume growth in M is faster than in S^N. (Note that absolute volume growth is slower.)

Remember how we proved concentration of measure on S^N: a $1/\sqrt{N}$-neighborhood of a hemisphere contains almost all the mass. The Gromov–Lévy theorem implies that the same happens in any manifold with positive Ricci curvature. So we get

Corollary 14 (Concentration in positive Ricci curvature). *Let M be an N-dimensional Riemannian manifold. Suppose that the Ricci curvature of M is at least that of S^N. Let $f : M \to \mathbb{R}$ be a 1-Lipschitz function. Then there exists a $m \in \mathbb{R}$ such that, for any $t \geqslant 0$*

$$\nu(\{x \in S^n,\ |f(x) - m| \geqslant t\}) \leqslant 2 \exp -\frac{t^2}{2D^2}$$

where $D = 1/\sqrt{N-1}$ and ν is the natural measure on M, normalized so that $\nu(M) = 1$.

Let us now mention some aspects of the tools developed by Bakry and Émery. Remember that the heat equation $\frac{\partial f}{\partial t} = \Delta f$, where Δ is the Laplace–Beltrami operator generalizing the usual Laplacian $\sum \frac{\partial^2}{\partial x_i^2}$, defines a semi-group of operators $(P_t)_{t \geqslant 0}$ acting on, say, smooth functions on a Riemannian manifold.

Theorem 15 (Bakry–Émery). *Let M be an N-dimensional Riemannian manifold. Suppose that the Ricci curvature of M is at least $K > 0$. Let $(P_t)_{t \geqslant 0}$ be the heat equation semigroup on M.*

Then for any $t \geqslant 0$ and any smooth function $f : M \to \mathbb{R}$:

(i) $\sup_M \|\nabla P_t f\| \leqslant \mathrm{e}^{-Kt} \sup_M \|\nabla f\|$,
(ii) $\|\nabla P_t f\|(x) \leqslant \mathrm{e}^{-Kt}(P_t \|\nabla f\|)(x)$,
(iii) $\mathrm{Ent}\, f := \int f \log \frac{f}{\int f} \,\mathrm{d}\nu \leqslant \frac{1}{2K} \int \frac{\|\nabla f\|^2}{f} \,\mathrm{d}\nu, \quad f > 0,$

where as usual ν is the normalized Riemannian volume on M.

The first inequality states that the Lipschitz norm is exponentially decreasing under the heat equation. The second, more precise inequality states that the norm of the gradient at point x at time t is controlled by the average around x of the initial norm of the gradient. The third inequality is a so-called *logarithmic Sobolev inequality*; in this survey we will not say much about them, but there are deep links between concentration of measure, convergence of Markov chains or heat kernels, and these log-Sobolev inequalities [ABCFGMRS00]. (Compare the Poincaré inequalities mentioned below.)

The Bonnet–Myers theorem can be found in any textbook on Riemannian geometry. A probabilistic proof of the Lichnerowicz theorem using couplings of Brownian motions can be found, for example, in [Hsu02] (Theorem 6.7.3). The Gromov–Lévy theorem is proven in [Gro86] (see also e.g. [Pet06]) and Bakry–Émery theory can be found in [BE85]. Actually the concentration of measure part of the Gromov–Lévy theorem is a consequence of the logarithmic Sobolec inequality, though this was not clear at the time.

§1.3. Markov chains and their convergence

We present here basic results on Markov chains and their convergence. We refer the reader to [Mar04], [ABCFGMRS00], [DS96] for approaches focusing on convergence rates, especially in the discrete case. The presentation here is partially inspired by Laurent Veysseire's master's dissertation.

Random walks. Let X be a, say, Polish space (i.e. metrizable, separable, complete; this ensures a good behavior of measure theory). A *Markov chain kernel*, or *random walk*, on X is the following data: for each $x \in X$, let m_x be a probability measure on X (and we assume that the measure m_x depends on x in a measurable way). The Markov chain jumps from a point x to a random point picked according to m_x. The

n-step transition probability is given by

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

where of course $m_x^{*1} := m_x$.

The Markov chain defines an operator on the set of (non-negative) measures on X: if μ is a measure then we define the measure $\mu * m$ by

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

which describes, given an initial mass distribution, what is the new mass distribution after a jump. Note that mass is preserved: $(\mu * m)(X) = \mu(X)$.

In a dual way, the Markov chain defines an operator M on bounded functions on X by

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

and we check the duality $\int \mathrm{M}f\,\mathrm{d}\mu = \int f\,\mathrm{d}(\mu * m)$, at least formally. Note that $\mathrm{M}^n f(x)$ describes the expected value of f at the endpoint of n steps of the random walk. In particular, if f is constant then $\mathrm{M}f = f$. Also note that $\sup \mathrm{M}f \leqslant \sup f$.

A measure μ is said to be *invariant* if $\mu * m = \mu$. A measure μ is said to be *reversible* if $\mathrm{d}\mu(x) \otimes \mathrm{d}m_x(y) = \mathrm{d}\mu(y) \otimes \mathrm{d}m_y(x)$ as a measure on $X \times X$. One checks that a reversible measure is invariant, but reversibility describes a "local equilibrium" property much stronger than invariance, namely that, under the initial mass distribution μ, the "flow" from x to y is equal to that from y to x.

If ν is an invariant measure, then conservation of mass reads $\int \mathrm{M}f\,\mathrm{d}\nu = \int f\,\mathrm{d}\nu$.

Convergence to equilibrium, spectral gap and Poincaré inequalities. From now on we suppose that ν is an invariant *probability* measure i.e. such that $\nu(X) = 1$. We denote integration under ν by $\mathbb{E}$. Two natural questions arising in the theory and practice of Markov chains are: Starting at a given point $x \in X$, how many steps are necessary so that the law of the endpoint is close to the invariant distribution? Given a function $f : X \to \mathbb{R}$, how close to $\mathbb{E}f$ is the empirical average $\frac{1}{T}\sum_{t=1}^{T} f(x_t)$ where the x_t's are the steps of the random walk?

We will explain here the spectral answer to the first question. We will come back to empirical measures in section 2.3.4.

In practice, the most widely used criterion to compare probability measures is the *total variation distance*

$$\|\mu - \mu'\|_{\mathrm{TV}} := \sup_{A \subset X} |\mu(A) - \mu'(A)| = \frac{1}{2} \int_X \mathrm{d}\,|\mu - \mu'| = \frac{1}{2} \left\| \frac{\mathrm{d}\mu}{\mathrm{d}\mu'} - 1 \right\|_{L^1(\mu')}$$

(whenever the latter makes sense). The first expression given justifies the interest, since it controls the worst error on all possible bets on the result.

What we want to control is $\|m_x^{*t} - \nu\|_{\mathrm{TV}}$. It is often more convenient to work in L^2 than in L^1. Let us work for a moment with functions instead of measures. Let $L_0^2(\nu)$ be the quotient of $L^2(\nu)$ by the constant functions; the norm on L_0^2 is variance under ν:

$$\|f\|_{L_0^2}^2 = \|f - \mathbb{E}f\|_{L^2} = \mathrm{Var}_\nu\, f = \frac{1}{2} \iint (f(x) - f(y))^2 \,\mathrm{d}\nu(x)\mathrm{d}\nu(y).$$

One then checks (Exercise) the associativity of variances property:

$$\|f\|^2 = \|\mathrm{M}f\|^2 + \int (\mathrm{Var}_{m_x}\, f)\,\mathrm{d}\nu(x)$$

and similarly

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

where Var_{m_x} denotes variance under the measure m_x. This formula describes the "smoothing" effect of the averaging operator; it means in particular that M is a non-expanding operator on $L^2(\nu)$.

The consequence for convergence of measures is the following. Let $L_0^{2*}(\nu)$ be the set of 0-mass signed measures on X having an L^2 density w.r.t. ν (i.e. the set $\{f\nu, f \in L_0^2(\nu)\}$). Then the duality formula $\int \mathrm{M}f\,\mathrm{d}\mu = \int f\,\mathrm{d}(\mu * m)$ states that convolution by m on $L_0^{2*}(\nu)$ is the dual operator to M. In particular, the operator norm of M^* on $L_0^{2*}(\nu)$ is equal to that of M on $L_0^2(\nu)$, since both are equal to $\sup\{\int \mathrm{M}f\,\mathrm{d}\mu,\ f \in L_0^2, \|f\| \leqslant 1,\ \mu \in L_0^{2*}, \|\mu\| \leqslant 1\}$.

Assume for a moment that X is discrete, so that the Dirac measure δ_x has a density w.r.t. ν (or that, instead of starting exactly at x, we

start with an initial distribution having a density w.r.t. ν). We can write

$$\begin{aligned}\left\|\frac{\mathrm{d}m_x^{*t}}{\mathrm{d}\nu}-1\right\|_{L^1(\nu)} &\leqslant \left\|\frac{\mathrm{d}m_x^{*t}}{\mathrm{d}\nu}-1\right\|_{L^2(\nu)} = \left\|\frac{\mathrm{d}m_x^{*t}}{\mathrm{d}\nu}-1\right\|_{L^2_0(\nu)}\\ &= \left\|m_x^{*t}-\nu\right\|_{L_0^{2*}(\nu)}\\ &\leqslant \|\delta_x-\nu\|_{L_0^{2*}(\nu)}\,\|\mathrm{M}^*\|^t_{L_0^{2*}(\nu)\to L_0^{2*}(\nu)}\\ &= \sqrt{\tfrac{1-\nu(x)}{\nu(x)}}\,\|\mathrm{M}\|^t_{L_0^2(\nu)\to L_0^2(\nu)}\,.\end{aligned}$$

Remember the formula above $\mathrm{Var}_\nu f = \mathrm{Var}_\nu \mathrm{M}f + \int(\mathrm{Var}_{m_x} f)\,\mathrm{d}\nu(x)$. If we knew that, for any function f, the quantity $\int(\mathrm{Var}_{m_x} f)\,\mathrm{d}\nu(x)$ represents at least some proportion α of $\mathrm{Var}_\nu f$, then we would know that $\mathrm{Var}_\nu \mathrm{M}f \leqslant (1-\alpha)\,\mathrm{Var}_\nu f$ and so $\|\mathrm{M}\|_{L_0^2(\nu)\to L_0^2(\nu)} \leqslant \sqrt{1-\alpha}$. Setting $1-\lambda = \sqrt{1-\alpha}$, what we have shown is

Proposition 16 (Poincaré inequality 1). *The norm of the operator M on $L_0^2(\nu)$ is at most $1-\lambda$ if and only if the inequality*

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

holds for any function $f \in L_0^2(\nu)$.

In particular, in this case for any $x \in X$ with $\nu(x) > 0$ we have

$$\left\|m_x^{*t}-\nu\right\|_{\mathrm{TV}} \leqslant \tfrac{1}{2}\sqrt{\tfrac{1-\nu(x)}{\nu(x)}}\,(1-\lambda)^t.$$

Such an inequality is (a variant of) a *Poincaré inequality.* It expresses the fact that the global variance of f is controlled by the local variations $\mathrm{Var}_{m_x} f$.

Note that the $\sqrt{1/\nu(x)}$ factor is sometimes detrimental in applications, even in very simple examples. For example, using this method on the (lazy) simple random walk on the discrete cube $\{0,1\}^N$ yields mixing time estimates of $O(N^2)$ because here $1/\nu(x) = 2^N$, instead of the correct $O(N \ln N)$ (Exercise). We will see that the coarse Ricci curvature method allows to recover $O(N \ln N)$ very easily.

Continuous time. One can define a continuous-time analogue of the situation above, by deciding that in each time interval $\mathrm{d}t$, the random walk has a probability $\mathrm{d}t$ to jump from the point x to a new point picked according to m_x. This amounts to taking an "infinitely lazy" random walk and then speeding up time. Namely, we replace M with

$$\mathrm{M}_t := \lim\left((1-\varepsilon)\,\mathrm{Id} + \varepsilon \mathrm{M}\right)^{t/\varepsilon} = \mathrm{e}^{t(\mathrm{M}-\mathrm{Id})}$$

which converges as an operator on $L^2(\nu)$ since $\|\mathrm{M}\| \leqslant 1$.

The new transition probabilities at time t are given by

$$\mathrm{d}p_x^t(y) = \sum_{k\in\mathbb{N}} \mathrm{e}^{-t}\frac{t^k}{k!}\,\mathrm{d}m_x^{*t}(y)$$

i.e. the number of jumps in time t is Poissonian with parameter t.

In this situation it is nice to work with the Laplace operator

$$\Delta := \mathrm{M} - \mathrm{Id}$$

which is the discrete analogue of the Laplacian $\sum \frac{\partial^2}{\partial x_i^2}$.

In analogy with the above, we want to study the operator norm of $\mathrm{e}^{t\Delta}$ on $L_0^2(\nu)$. One can check (Exercise) that

$$\frac{\mathrm{d}}{\mathrm{d}t}_{|t=0} \mathrm{Var}_\nu(\mathrm{e}^{t\Delta}f) = 2\langle f, \Delta f\rangle_{L_0^2(\nu)} = -\iint (f(y)-f(x))^2\,\mathrm{d}\nu(x)\mathrm{d}m_x(y)$$

where the last term is called the *Dirichlet form* associated with the random walk.

Once more, if we knew that the right-hand-side represents at least some fraction α of $\mathrm{Var}_\nu f$, then we would know that $\mathrm{Var}_\nu(\mathrm{e}^{t\Delta}f)$ decreases at least exponentially with rate α. What we have proven is

Proposition 17 (Poincaré inequality 2). *The inequality* $\mathrm{Var}_\nu(\mathrm{e}^{t\Delta}f) \leqslant \mathrm{e}^{-2\lambda t}\,\mathrm{Var}_\nu f$ *holds for any* $f \in L^2(\nu)$ *if and only if*

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

holds for any function $f \in L^2(\nu)$.

In particular, in this case for any $x \in X$ *with* $\nu(x) > 0$ *we have*

$$\|p_x^t - \nu\|_{\mathrm{TV}} \leqslant \tfrac{1}{2}\sqrt{\tfrac{1-\nu(x)}{\nu(x)}}\,\mathrm{e}^{-\lambda t}.$$

This is the standard form of the Poincaré inequality. The best value of λ is called the *spectral gap*. The spirit is the same as the one above: both $\mathrm{Var}_{m_x} f$ and $\int (f(y)-f(x))^2\,\mathrm{d}m_x(y)$ quantify the variations of f around x.

The reversible case. One checks that the invariant measure ν is reversible if and only if the operator M is self-adjoint in $L^2(\nu)$. In this case, we have the duality formula

$$(f\nu) * m = (\mathrm{M}f)\nu$$

meaning that the evolution of measures is the same as the evolution of their density functions w.r.t. ν. In that case, the operator M has a real spectrum, included in $[-1;1]$, and Δ is a negative operator. The Poincaré inequality then states that the spectrum of Δ on $L^2_0(\nu)$ is contained in $[-2,-\lambda]$, hence the name *spectral gap*. Estimation of the spectral gap is easier in the reversible case, since the operator norm of M is equal to its spectral radius, which may be easier to compute.

The case of diffusions. If X is a Riemannian manifold, one can take for m_x the volume measure restricted to the ball of radius ε around X. Then, the discrete Laplace operator of the random walk is an approximation of $\frac{\varepsilon^2}{2(N+2)}$ times the Laplace–Beltrami operator of the manifold, and so the random walk will approximate Brownian motion and the heat equation. In this case for a smooth function f both $\mathrm{Var}_{m_x} f$ and $\int (f(y)-f(x))^2 \,\mathrm{d}m_x(y)$ approximate $\|\nabla f\|^2$ (up to scaling), and the Poincaré inequality reads $\mathrm{Var}\, f \leqslant \frac{1}{\lambda}\int \|\nabla f\|^2$.

§1.4. Transportation distances

Transportation distances answer the following question: One wants to move a heap of sand from an initial position to a prescribed new position, in a cost-effective way, meaning that we want grains of sand to travel over the smallest possible distance. We refer to the excellent [Vil03] or [Vil08].

This situation is formalized as follows: Let μ_1, μ_2 be two measures on a metric space X with the same total mass. A *transference plan* from μ_1 to μ_2 is a measure ξ on $X \times X$ such that $\int_y \mathrm{d}\xi(x,y) = \mathrm{d}\mu_1(x)$ and $\int_x \mathrm{d}\xi(x,y) = \mathrm{d}\mu_2(y)$. Here $\mathrm{d}\xi(x,y)$ represents the quantity of sand travelling from x to y, and the two constraints ensure that we indeed start with measure μ_1 and end up with measure μ_2. Let $\Pi(\mu_1,\mu_2)$ be the set of transference plans from μ_1 to μ_2 (also called *couplings* between μ_1 and μ_2).

The L^1 *transportation distance* (or Kantorovich–Rubinstein distance) between μ_1 and μ_2 is the best average travel distance that can be achieved. It is defined as

$$W_1(\mu_1,\mu_2) := \inf_{\xi\in\Pi(\mu_1,\mu_2)} \iint d(x,y)\,\mathrm{d}\xi(x,y).$$

In general this is only a semi-distance, as it may be infinite. The triangle inequality requires the so-called "gluing lemma" (composition of couplings), which technically imposes that X be Polish. This quantity defines a genuine distance when restricted to the set of those probability

measures μ with *finite first moment* i.e. such that $\int d(o,x)\,\mathrm{d}\mu < \infty$ for some (hence any) $o \in X$.

Wasserstein distances are a generalization obtained by optimizing $d(x,y)^p$ instead of $d(x,y)$, for some $p \geqslant 1$.

The only property of transportation distances we shall use is *Kantorovich duality*. It states that

$$W_1(\mu_1,\mu_2) = \sup_{\substack{f:X\to\mathbb{R} \\ f\,1-\text{Lipschitz}}} \int f\,\mathrm{d}\mu_1 - \int f\,\mathrm{d}\mu_2$$

and the supremum can also be restricted to bounded 1-Lipschitz functions.

The fact that the variation of the integral of a 1-Lipschitz function is bounded by the transportation distance is easy. The converse direction is not at all trivial. A non-constructive proof using minimax principles for convex functions is given in [Vil03], whereas a somewhat more constructive approach is taken in [Vil08].

Part 2. Discrete Ricci curvature

§2.1. Definition and examples

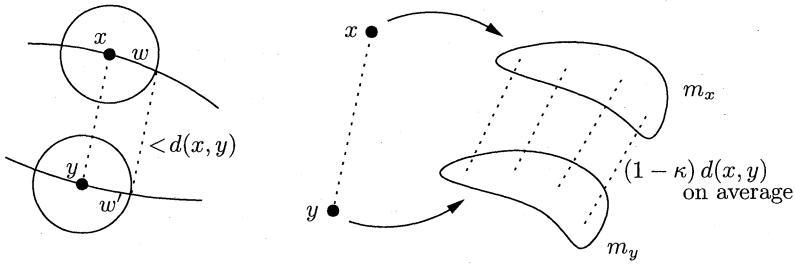
We now have all the necessary ingredients to define a coarse version of Ricci curvature, as presented in [Oll07, Oll09]. Remember that, in Riemannian manifolds, the value of Ricci curvature can be recovered by comparing the average distance between two small balls to the distance between their centers (Corollary 10). We use this characterization in more general spaces: we will say that coarse Ricci curvature is positive if *balls are closer than their centers are* in transportation distance. The difference gives the value of coarse Ricci curvature.

The notion of "small ball" depends on what is relevant in the situation considered. For example, in a graph it might be natural to take balls of radius 1, whereas in a manifold arbitrarily small balls are the natural choice. This allows for a definition of Ricci curvature "at a given scale". For example, the Earth is a reasonable approximation of an ellipsoid up to a scale of a dozen kilometers; the definition below allows to compute the Ricci curvature of the Earth at this scale and compare it to the ellipsoid.

So, to allow for a more general treatment, we will assume that for each point x in our space X, a probability measure m_x on X is given, which represents a "ball" of our own choosing around x. Of course, this is exactly the same data as a Markov chain.

Moreover, in Riemannian manifolds Ricci curvature is defined for a tangent vector. In more general spaces, the best we can do for a tangent

vector is a pair of close points. (The meaning of "close" is made precise by Exercise 22 below.)



Definition 18 (Coarse Ricci curvature). Let (X, d) be a metric space, endowed with a family $(m_x)_{x \in X}$ of probability measures on X. Let x, y be two points in X. The *coarse Ricci curvature* $\kappa(x, y)$ of X along xy is defined by the relation

$$W_1(m_x, m_y) = (1 - \kappa(x, y))\, d(x, y)$$

where W_1 is the L^1 transportation distance.

The hidden technical assumptions are the following: X should be Polish; the measure m_x should depend measurably on $x \in X$; each m_x should have finite first moment (see section 1.4).

Definition 19 (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.)

Maybe the most fundamental example is the following.

Exercise 20 (Discrete cube).
Let $\{0, 1\}^N$ be the discrete cube equipped with its L^1 metric and uniform probability measure. Compute the coarse Ricci curvature of the 1-step random walk for a pair of points x, y at distance 1.

Next, we did everything so that in Riemannian manifolds, we get the correct value of Ricci curvature up to some scaling (and up to checking that parallel transport between balls is indeed an optimal coupling up

to higher-order terms). See also condition (xii) in [RS05], Theorem 1.5, which uses an infimum of Ricci curvature instead.

Exercise 21 (Riemannian manifolds).
Let (X, d) be a smooth complete N-dimensional Riemannian manifold. Equip it with the ε-step random walk, for ε small enough.

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

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

The ε^2 factor reflects the fact that Riemannian manifolds are locally Euclidean up to second order.

In both these cases, we have computed curvature only for "close" points x, y. This is justified by the following simple yet very useful property.

Exercise 22 (Geodesic spaces).
Suppose that (X, d) is α-geodesic in the sense that for any two points $x, y \in X$, there exists an integer n and a sequence $x_0 = x, x_1, \ldots, x_n = y$ such that $d(x, y) = \sum d(x_i, x_{i+1})$ and $d(x_i, x_{i+1}) \leqslant \alpha$.

Then, if $\kappa(x, y) \geqslant \kappa$ for all pairs of points with $d(x, y) \leqslant \alpha$, then $\kappa(x, y) \geqslant \kappa$ for all pairs of points $x, y \in X$.

For example, a graph is 1-geodesic, and a Riemannian manifold is α-geodesic for any $\alpha > 0$: in both cases, this is the very definition of the distance.

This property is to be kept in mind in the next series of examples.

Exercise 23 ($\mathbb{Z}^N$ 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 $x, y \in \mathbb{Z}^N$, the coarse Ricci curvature along (xy) is 0.

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. For example, the triangular tiling of the plane has 0 curvature, as well as $\mathbb{R}^N$ equipped with an L^p norm.

Our last basic example was $\mathbb{R}^N$ equipped with a Gaussian measure. Following the spirit of Bakry–Émery theory, we use (a discretization of) the natural random process having a Gaussian measure as its invariant distribution, namely the Ornstein–Uhlenbeck process.

Exercise 24 (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 $\mathrm{e}^{-\alpha\delta t}x$, of variance $s^2(1 - \mathrm{e}^{-2\alpha\delta t})/2\alpha \sim s^2\delta t$ for small δt.

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

On Riemannian manifolds this generalizes as follows.

Exercise 25 (Ricci curvature à la Bakry–Émery).
Let X be an N-dimensional Riemannian manifold and F be a tangent vector field. Consider the differential operator

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

associated with the stochastic differential equation

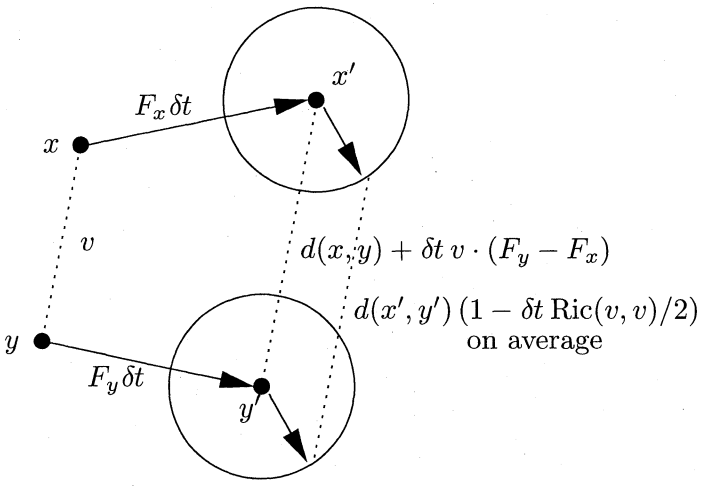
$$\mathrm{d}X_t = F\,\mathrm{d}t + \mathrm{d}B_t$$

where B_t is the Brownian motion in X. The Ricci curvature (in the Bakry–Émery sense) of this operator, applied to a tangent vector v is $\frac{1}{2}\operatorname{Ric}(v, v) - v \cdot \nabla_v F$.

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

$$\kappa(x, y) = \delta t \left(\frac{1}{2}\operatorname{Ric}(v, v) - v \cdot \nabla_v F + O(d(x, y)) + O(\sqrt{\delta t}) \right)$$



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 [BE85], the generator of Brownian motion is $\frac{1}{2}\Delta$ instead of Δ, hence the $\frac{1}{2}$ factor for the Ricci part.

Maybe the reason for the additional $-v \cdot \nabla_v F$ in Ricci curvature à la Bakry–Émery is made clearer in this context: it is simply the quantity by which the flow of F 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. Combining the Markov chain with an isometry of the space has no effect whatsoever on our definition.

Exercise 26 (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 $1/N$.

Exercise 27 (Binomial and Poisson distributions).
Consider the discrete cube $\{0,1\}^N$ equipped with the following continuous-time random walk: during a time interval δt, each digit 0 has a probability $p\,\delta t$ of becoming a 1 and each digit 1 has a probability $q\,\delta t$ of becoming a 0. What is the coarse Ricci curvature of (a small-time discretization of) this random walk?

A particularly interesting case is when $N \to \infty$ and $p = \lambda/N, q = 1$. Then the number of 1's asymptotically follows a Poisson law of parameter λ; the random walk "projected" on $\mathbb{N}$ by considering only the number of 1's tends to the so-called $M/M/\infty$ *process* on $\mathbb{N}$.

The next example relates our definition to a traditional generalization of negative sectional curvature. Although negative Ricci curvature is generally not very useful in Riemannian geometry, it is nice to see that at least the definition is consistent. (This exercise requires good knowledge of δ-hyperbolic geometry.)

Exercise 28 (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 consisting in performing k steps of the simple random walk. Let $x, y \in X$. Then $\kappa(x,y) = -\frac{2k}{d(x,y)}\,(1 - o(1))$ when k and $d(x,y)$ 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. The argument applies to trees or discrete δ-hyperbolic spaces with a uniform lower bound on the exponential growth rate of balls as well.

Exercise 29 (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 $U(S) := -\sum_{x\sim y\in G} S(x)S(y) - h\sum_x S(x)$ 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 [Gri67]. 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 [Dob70] (or to $G(\beta) < 1$ in the notation of [Gri67], Eq. (19)). By comparison, the exact value of the critical β for the Ising model on the regular infinite tree of valency v is $\frac{1}{2}\ln\left(\frac{v}{v-2}\right)$, which shows asymptotic optimality of this criterion. When block dynamics (see [Mar04]) are used instead of single-site updates, positive coarse Ricci curvature of the block dynamics Markov chain is equivalent to the Dobrushin–Shlosman criterion [DS85].

Positive curvature in our sense 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 essentially equivalent to the main results of the literature under the Dobrushin–Shlosman criterion (see e.g. the review [Mar04]), but may be a quick way to prove them. Note that in our setting we do not need the underlying graph to be $\mathbb{Z}^N$.

Exercise 30.
Make precise comparisons between the results obtained on Ising-like models using theorems on positive discrete coarse Ricci curvature, and the results from the literature.

We end this series of examples by asking for new ones.

Problem A (Log-concave measures).
We have seen that coarse Ricci curvature is positive for $\mathbb{R}^N$ equipped with a Gaussian measure, and this generalizes to smooth, uniformly strictly log-concave measures. What happens for a general log-concave measure? The next example would be a convex set (whose boundary has "positive curvature" in an intuitive geometric sense), with the associated process a Brownian motion conditioned not to leave the convex body.

Problem B (Finsler manifolds).
We have seen that coarse Ricci curvature is 0 for $\mathbb{R}^N$ equipped with an L^p norm. Does this give anything interesting in Finsler manifolds? (Compare [Oht] and [OS].)

Problem C (Nilpotent groups).
We have seen that curvature of $\mathbb{Z}^N$ is 0. What happens on discrete or continuous nilpotent groups?

For example, on the discrete Heisenberg group $\langle a, b, c \mid ac = ca, bc = cb, [a, b] = c \rangle$, the natural discrete random walk analogous to the hypoelliptic diffusion operator on the continuous Heisenberg group is the random walk generated by a and b. Since these generators are free up to length 8, clearly coarse Ricci curvature is negative at small scales, but does it tend to 0 at larger and larger scales?

§2.2. Elementary properties

We leave here as exercices a series of simple properties associated with positive coarse Ricci curvature. We say that coarse Ricci curvature is at least κ if for any pair of points $x, y \in X$ we have $\kappa(x, y) \geqslant \kappa$.

The first such result shows that our notion is a direct generalization of one of the results of Bakry and Émery in positive Ricci curvature (part (i) of Theorem 15 above). This was actually suggested in [RS05].

Exercise 31 (Lipschitz norm contraction).
Coarse Ricci curvature is at least κ if and only if the random walk operator M maps 1-Lipschitz functions to $(1 - \kappa)$-Lipschitz functions. (*Hint:* One direction is easy; use Kantorovich duality for the other.)

Statements equivalent to the following proposition also appear in [Dob70] (Theorem 3), in [Dob96] (Proposition 14.3), in the second edition of [Che04] (Theorem 5.22), in [DGW04] (in the proof of Proposition 2.10) and in [Oli].

Exercise 32 (W_1-contraction).
Coarse Ricci curvature is at least κ if and only if the random walk acting on the space of probability measures with finite first moment, equipped with the L^1 transportation distance, is a $(1-\kappa)$-contraction. (*Hint:* First check that this space is stable under the random walk action, then either use Kantorovich duality or the fact from [Vil08] that optimal couplings can be chosen in a measurable way.)

Corollary 33 (Convergence). *Suppose that coarse Ricci curvature is at least $\kappa > 0$. Then there exists a unique invariant probability measure ν. Moreover, for any probability measure μ we have*

$$W_1(\mu * m^{*t}, \nu) \leqslant W_1(\mu, \nu)\,(1-\kappa)^t \leqslant (\operatorname{diam} X)\,(1-\kappa)^t$$

and, for $x \in X$,

$$W_1(m_x^{*t}, \nu) \leqslant \tfrac{W_1(\delta_x, m_x)}{\kappa}\,(1-\kappa)^t.$$

The latter estimate is sometimes very useful because $W_1(\delta_x, m_x)$ (the "jump" of the random walk at x) is often readily accessible.

Exercise 34.
Use this to prove an $O(N \ln N)$ estimate for mixing time of the lazy simple random walk on the discrete cube (in comparison with the $O(N^2)$ from the Poincaré inequality).

Exercise 35 (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 $\kappa + \kappa' - \kappa\kappa'$.

Exercise 36 (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 $\sum \alpha_i \kappa_i$.

Exercise 37 (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 $\inf \alpha_i \kappa_i$. (The "infimum" aspect is clear: if some component mixes very badly or not at all, then so does the whole process.)

We close this section with an open problem. In several examples above, the natural process was a continuous-time one. When the space is finite or compact, or when one has good explicit knowledge of the process (as for Ornstein–Uhlenbeck on $\mathbb{R}^N$), discretization works very well, but this might not be the case in full generality.

Problem D (Continuous-time).
Suppose a continuous-time Markov semigroup $(m_x^t)_{x\in X, t\in\mathbb{R}_+}$ is given. One can define a coarse Ricci curvature in a straightforward manner as

$$\kappa(x,y) := \liminf_{t\to 0^+} \frac{1}{t} \frac{d(x,y) - W_1(m_x^t, m_y^t)}{d(x,y)}$$

but then, in the proofs of the elementary properties above, there arise non-trivial issues of commutation between limits and integrals, especially if the generator of the process is unbounded. Is this definition, combined with some assumption on the process (e.g. non-explosion), enough to get all the properties above in full generality, for both diffusions and jump processes? Given an unbounded generator for the process, is positivity of the $\kappa(x,y)$ above enough to ensure non-explosion? (One could directly use the Lipschitz norm contraction as a definition [RS05, Jou07], but first, this is not a local criterion, and second, it only defines a lower bound on Ricci curvature, not a value at a given point.)

§2.3. Results in positive coarse Ricci curvature

2.3.1. More notation

Before stating the theorems, we need two more definitions which capture coarse analogues of the diffusion constant and dimension of the space. Here, as above, we consider a metric space (X,d) equipped with a random walk $(m_x)_{x\in X}$.

Definition 38 (Diffusion constant). 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

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

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

Definition 39 (Local dimension). Let the *local dimension* at x be

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

and $n := \inf_x n_x$.

Exercise 40.
Consider the ε-step random walk on an N-dimensional Riemannian manifold. Then $\sigma(x) = \varepsilon\sqrt{\frac{N}{N+2}} + o(\varepsilon)$ and $N - 1 \leqslant n_x \leqslant N$ (I do not know the exact value).

About this definition of dimension. Obviously $n_x \geqslant 1$. For the discrete-time Brownian motion on a N-dimensional Riemannian manifold, one has $n_x \approx N$. 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.

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

so that for each $x \in X$ the local dimension of X at x is $n_x = \mathrm{StatDim}(X, d, m_x)$. 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.

2.3.2. Spectral gap

We now give a generalization to the Lichnerowicz theorem (Theorem 12) stating that positive curvature implies a spectral gap. We begin with a study of Lipschitz versus L^2 norms.

Exercise 41 (Variance of Lipschitz functions).
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 $\frac{\sigma^2}{n\kappa(2-\kappa)} \leqslant \frac{\sigma^2}{n\kappa}$. In particular, Lipschitz functions are in $L^2(\nu)$. (*Hint:* Iterate the formula $\operatorname{Var} f = \operatorname{Var} \mathrm{M} f + \int \operatorname{Var}_{m_x} f \, \mathrm{d}\nu(x)$.)

Proposition 42 (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 $L^2(X, \nu)/\{const\}$ is at most $1 - \kappa$.

Compare Theorem 1.9 in [CW94] (Theorem 9.18 in [Che04]).

Proof (sketch). The spectral radius w.r.t. Lipschitz norm of the operator M acting on Lipschitz functions is at most $1 - \kappa$ by the results above. In full generality, since the Lipschitz norm controls the L^2 norm by the above, the spectral radius of M w.r.t. the L^2 norm, restricted to the subspace of Lipschitz functions in L^2, is $\leqslant 1 - \kappa$ as well. Now Lipschitz functions are dense in L^2, and for a bounded, self-adjoint operator it is enough to control the spectral radius on a dense subspace. Q.E.D.

Problem E (Non-reversible spectral gap).
What happens in the non-reversible case? There are different ways to formulate the question: spectral radius, norm of the operator, Poincaré inequality. (Note that a Poincaré inequality with a worse constant and with a "blurred" gradient always holds, cf. the section on log-Sobolev inequality in [Oll09].)

In the reversible case, we get that the spectrum of Δ is included in $[-2; -\kappa]$ together with the two Poincaré inequalities

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

and

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

as a corollary.

Exercise 43.
Run through all the examples above, and compare the spectral gap obtained by coarse Ricci curvature to the actual value of the spectral gap, when known (e.g. on the cube).

Problem F (Sharp Lichnerowicz theorem).
For the ε-step random walk on a Riemannian manifold, the operator $\Delta = \mathrm{M} - \mathrm{Id}$ of the random walk behaves like $\frac{\varepsilon^2}{2(N+2)}$ times the Laplace–Beltrami operator and the coarse Ricci curvature is $\frac{\varepsilon^2}{2(N+2)}$ times ordinary Ricci curvature, so that we get a spectral gap estimate of $\inf \mathrm{Ric}(v)$ for the Laplace–Beltrami operator. On the other hand, the Lichnerowicz theorem has a qualitatively comparable but slightly better spectral gap estimate $\frac{N}{N-1} \inf \mathrm{Ric}(v)$, which is sharp for the sphere. This is because our definition of $\kappa(x, y)$ somehow overlooks that the sectional curvature $K(v, v)$ in the direction of xy is 0. Is there a way to take this into account? (Note that our estimate is sharp for the discrete cube as well as for the Ornstein–Uhlenbeck process, so the phenomenon is rather specific to this situation.)

Problem G (Non-constant curvature).
The estimate above uses the infimum of $\kappa(x, y)$. Is it possible to relax this assumption and, for example, include situations where κ takes "not too many" negative or zero values? Using the coarse Ricci curvature of the iterates m_x^t for some $t \geqslant 2$ should "smoothen out" exceptional values of $\kappa(x, y)$, so that for large t the coarse Ricci curvature of m_x^t should be close to an "average" coarse Ricci curvature of m_x (probably involving large deviations of the average value of $\kappa(x, y)$ along trajectories of the random walk).

This may be interesting e.g. on random objects (graphs...) where locally some negative curvature is bound to occur somewhere.

2.3.3. Concentration of measure

One of our main motivations was to find a common geometric property between three concentrated spaces, namely the sphere, the discrete cube and the Gaussian space. It is very satisfying that the property thus found implies concentration of measure.

We now state the theorem without proof (see [Oll09]). Remember the notation in Section 2.3.1.

The estimated Gaussian variance is $\sigma^2/n\kappa$, in the notation of section 2.3.1. However, concentration is not always Gaussian far away

from the mean, as exemplified by binomial distributions on the cube or $M/M/\infty$ queues. 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; 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 predicted by the Gromov–Lévy theorem (Theorem 14). 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 44 (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

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

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

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

Remark 45. Actually $\sigma(x)^2/n_x\kappa$ 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 $C = 0$. Statements somewhat similar to this latter case appear in [DGW04] (under a strong local assumption) and in [Jou07].

Exercise 46 (Continuous-time situations).
Observe that the theorem above has a well-defined limit when we replace the random walk $m = (m_x)_{x\in X}$ with the lazy random walk m' whose transition probabilities are $m'_x := (1 - \delta t)\,\delta_x + \delta t\, m_x$. The point is that when $\delta t \to 0$, both σ_x^2 and κ scale like δt (and n_x tends to 1). This means that we can apply Theorem 44 to continuous-time examples.

Exercise 47.
Review the various examples and check that the order of magnitude of the Gaussian variance is correct, especially for Riemannian manifolds and for binomial distributions on the cube.

Exercise 48.
Remove the various technical assumptions (Lipschitz $\sigma(x)^2$, bounded σ_∞) and find counter-examples to the resulting statements. Also find an example when the Gaussian regime practically disappears and only the exponential part is visible.

Problem H (Isoperimetric profile and curvature at infinity).
Suppose that $\inf \kappa(x, y) = 0$ but that the same infimum taken on balls of increasing radii around some origin is non-zero. Is there a systematic correspondence between the way curvature decreases to 0 at infinity and the isoperimetric profile? (Compare the section of [Oll09] devoted to the relationship between non-negative coarse Ricci curvature and exponential concentration.) An interesting example is the $M/M/k$ queue.

Problem I (Local assumptions).
The condition $\sigma_\infty < \infty$ can be replaced with a local 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 [DGW04].) However, this is not at all 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). In particular, when taking a continuous-time limit as above, such estimates diverge. So, is there a way to relax the assumption $\sigma_\infty < \infty$, yet keep an estimate based on the local variance σ^2, and can this be done so that the estimate stays bounded when taking a continuous-time limit?

Problem J (Functional inequalities).
The Laplace transform estimate $\mathbb{E}e^{\lambda(f-\mathbb{E}f)} \leqslant e^{D^2\lambda^2/2}$ often used to establish Gaussian concentration for a measure ν is equivalent, by a result of Bobkov and Götze [BG99], to the following inequality: $W_1(\mu,\nu) \leqslant \sqrt{2D^2 \operatorname{Ent}(\mathrm{d}\mu/\mathrm{d}\nu)}$ for any probability measure $\mu \ll \nu$. Is there a way to formulate our results in terms of functional inequalities? As such, the inequality above will fail as concentration can be non-Gaussian far away from the mean (e.g. in the simple example of the binomial distributions on the cube), so in a coarse setting it might be necessary to plug additive terms in the formulation of the inequality to account for what happens at small measures or small scales. Another suggestion by Villani is to use a Talagrand inequality where the L^2 transportation distance is replaced with a quadratic-then-linear transportation cost and use the results in [GL07].

2.3.4. Further results: Convergence of empirical measures, Log-Sobolev inequality, Gromov–Hausdorff convergence

Several more results are proven under these or similar assumptions.

— For example, in his preprint [Jou], Joulin proves that, under a positive curvature assumption, the empirical means of Lipschitz functions are concentrated and satisfy a Poisson-like deviation inequality (together with a control of their deviation from the actual expectation under the invariant measure). The order of magnitude for the variance at time t is, with our notation, $\frac{\sup \sigma(x)^2}{t\kappa^2}$, and the transition from Gaussian to Poisson behavior of the tail is controlled by σ_∞ as above. (See also [DGW04] for related results.)

— A generalization of the results of Bakry–Émery mentioned above (Theorem 15) holds [Oll09]. We use a kind of "blurred Lipschitz constant" defined by $\nabla_\lambda f(x) := \sup_{y,y' \in X} \frac{|f(y)-f(y')|}{d(y,y')} e^{-\lambda d(x,y)-\lambda d(y,y')}$. The larger λ is, the closer this is to the usual Lipschitz norm, but there is a maximal possible value of λ depending on the object: typically for

a manifold one can take $\lambda \to \infty$ and thus recover the usual gradient, whereas in a graph one has to take $\lambda \approx 1$.

With this gradient, one can prove generalizations to points (*ii*) and (*iii*) of Theorem 15. Namely, any positive function $f : X \to \mathbb{R}$ with $\nabla_\lambda f < \infty$ satisfies

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

and moreover, we have the contraction property

$$\nabla_\lambda(\mathrm{M}f) \leqslant (1 - \kappa/2)\, \mathrm{M}(\nabla_\lambda f)$$

(so compared to Theorem 15, we lose a factor 4 or so in the constants for Riemannian manifolds; this is to be expected for a theorem valid both on discrete and continuous settings).

Exercise 49 (Herbst argument for ∇_λ).
Use the Herbst argument to show that a logarithmic Sobolev inequality with the gradient $\nabla_\lambda f$ as above implies Gaussian-then-exponential concentration, where the transition from Gaussian to exponential is controlled by the value of λ. (*Hint:* Show that for $\lambda' \leqslant \lambda$, for any 1-Lipschitz function f we have the chain rule $(\nabla_\lambda \mathrm{e}^{\lambda' f})(x) \leqslant \lambda' \mathrm{e}^{\lambda' f(x)}$. For this use the inequality $\left|\mathrm{e}^a - \mathrm{e}^b\right| \leqslant |a-b| \frac{\mathrm{e}^a + \mathrm{e}^b}{2}$. Then apply the usual Herbst argument.) Does the argument work the other way round?

— Oliveira [Oli] used contraction of transportation distances by Markov chains to substantially improve mixing time estimates of a random walk on the set of orthogonal matrices known as *Kac's random walk*, which consists, at each step, in composing the current matrix with a rotation of random angle in the plane generated by two randomly chosen coordinates. This is consistent, of course, with the positive Ricci curvature of $\mathrm{SO}(N)$ as a Riemannian manifold, but Kac's random walk is more practical than Brownian motion on $\mathrm{SO}(N)$.

— Finally, since the objects used in the definition of coarse Ricci curvature involve only integrals of the distance function under the transition kernel, it is kind of tautological [Oll09] to prove continuity theorems for coarse Ricci curvature in the Gromov–Hausdorff topology (suitably extended to include convergence of the Markov kernel).

2.3.5. A few more problems

Problem K (Sturm–Lott–Villani definition).
What is the relationship (if any) between our notion and the one defined

by Sturm and Lott–Villani [Stu06, LV]? The latter is generally more difficult to work out on concrete examples, and is not so well suited to discrete settings (though see [BS]), but under the stronger $\mathrm{CD}(K,N)$ version, some more theorems are proven, including the Brunn–Minkowski inequality and Bishop–Gromov comparison theorem, together with applications to the Finsler case [Oht, OS].

Problem L (Bishop–Gromov theorem).
Is it possible to generalize more traditional theorems of positive Ricci curvature, i.e. the Bishop–Gromov theorem, or something close to the isoperimetric form of the Gromov–Lévy theorem? It is not clear what a reference constant curvature space would be in this context. Observe for example that, in the discrete cube, the growth of balls is exponential-like for small values of the radius (namely N, $N(N-1)/2$, etc.). Such theorems may be limited to manifold-like spaces for which a reference comparison space exists. Yet in the case of the cube, the isoperimetric behavior of balls still "slows down" in a positive-curvature-like way. A maybe useful definition of the "boundary" of a part A is $W_1(1_A, 1_A * m)$. Also compare Problem R below.

Problem M (Entropy decay).
The logarithmic Sobolev inequality (under the form comparing $\mathrm{Ent}\, f^2$ to $\int \|\nabla f\|^2$, not under the modified form comparing $\mathrm{Ent}\, f$ to $\int \|\nabla f\|^2 / f$) usually implies an exponential decreasing of entropy by the Markov chain. Is there some form of this phenomenon in our setting? (Once more, it is necessary to keep in mind the case of binomial distributions on the cube, for which the modified form of the Sobolev logarithmic inequality was introduced.)

Problem N (Discrete Ricci flow).
Define a "discrete Ricci flow" by letting the distance on X evolve according to coarse Ricci curvature

$$\frac{\mathrm{d}}{\mathrm{d}t} d(x,y) = -\kappa(x,y)\, d(x,y)$$

where $\kappa(x,y)$ is computed using the current value of the distance (and by either keeping the same transition kernel m_x or having it evolve according to some rule). What can be said of the resulting evolution? (Note that if the same transition kernel is kept, then this will only compare to the usual Ricci flow up to a change of time, since, e.g. on a Riemannian sphere, this will amount to using smaller and smaller "diffusion constants" whereas the diffusion constant C in the Ricci flow

$\frac{dg}{dt} = -C\,\mathrm{Ric}$ is taken constant; in particular, the diameter of a sphere will tend exponentially towards 0 instead of linearly.)

Problem O (Up to δ).
The constraint $W_1(m_x, m_y) \leqslant (1-\kappa)\, d(x,y)$ may be quite strong when x and y are too close, even if the measures m_x, m_y have a larger support. In order to eliminate completely the influence of small scales, and in the spirit of δ-hyperbolic spaces, we can define a "positive curvature up to δ" condition. Namely, $\kappa(x,y)$ is the best $\leqslant 1$ constant in the inequality

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

so that positive curvature up to some δ becomes an *open* property in Gromov–Hausdorff topology. Which theorems extend to this setting? Is it possible, in such a situation, to choose a discrete subset $X' \subset X$ and to redefine the random walk on X' in a reasonable way such that it has positive coarse Ricci curvature?

Problem P (Discrete sectional curvature).
A notion equivalent to non-negative sectional curvature for Riemannian manifolds can be obtained by requiring that there be a coupling between m_x and m_y, such that the coupling moves *all* points of m_x by *at most* $d(x,y)$. (This amounts to replacing W_1 with the L^∞ transportation distance in the definition.) Does this have any interesting properties? Is it possible to get an actual value for sectional curvature? (In this definition, the contribution from x and y themselves will generally prevent getting non-zero values.) Is this related to positive sectional curvature in the sense of Alexandrov? (Though the latter cannot be applied to discrete spaces.)

Problem Q (Discrete scalar curvature).
In Riemannian geometry, scalar curvature at x is the average of $\mathrm{Ric}(v)$ over all unit vectors v around x. It controls, in particular, the growth of the volume of balls. Here one can transpose this definition and set $S(x) := \int \kappa(x,y)\, dm_x(y)$ (where maybe a weight depending on $d(x,y)$ should be added). Does it have any interesting properties?

Problem R (L^2 Bonnet–Myers and the dimension parameter).
For an L^2 version of the Bonnet–Myers theorem to hold, it is necessary to make stronger assumptions than positive curvature, namely that for any points x, x' and for any small enough pair of times t, t' one has

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

whereas before one used only the case $t = t'$. (The second term is obtained by considering Gaussian measures of variance t and t' centered at x and x' in $\mathbb{R}^N$.) Then (see the section on strong Bonnet–Myers theorem in [Oll09]) one gets a diameter estimate $\operatorname{diam} X \leqslant \pi\sqrt{\frac{C}{2\kappa}}$ so that C plays the role of $N-1$. Is the constant C somehow related to a "dimension", in particular to the "dimension" n in the Bakry–Émery $CD(K,n)$ condition?

Problem S (Alexandrov spaces).
What happens for spaces with positive sectional curvature in the sense of Alexandrov? Do they have positive Ricci curvature for a reasonable choice of m_x? This seems to be unknown for the Sturm–Lott–Villani definition too. Would it be enough to approximate these spaces by manifolds or use a parallel transport in Alexandrov spaces? (See also Problem P above.)

Problem T (Expanders).
Is there a family of expanders (i.e. a family of graphs of bounded degree, spectral gap bounded away from 0 and diameter tending to ∞) with non-negative Ricci curvature? (Suggested by A. Naor and E. Milman.)

Problem U (Permutation groups).
For the permutation groups, with respect to the random walk generated by transpositions, Ricci curvature is positive but does not allow to recover concentration of measure with the correct order of magnitude. Is this related to results by N. Beresticky about the δ-hyperbolic-like properties of the permutation groups, which thus appear to have a mixture of positive and negative curvature properties?

References

[ABCFGMRS00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer, Sur les Inégalités de Sobolev Logarithmiques, Panor. Synthèses, **10**, Soc. Math. France, 2000.

[BD97] R. Bubley and M. E. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, FOCS 1997, 223–231.

[BE85] D. Bakry and M. Émery, Diffusions hypercontractives, In: Séminaire de Probabilités, XIX, 1983/84, Lecture Notes in Math., **1123**, Springer, Berlin, 1985, pp. 177–206.

[BG99] S. Bobkov and F. Götze, Exponential integrability and transportation cost related to garithmic Sobolev inequalities, J. Funct. Anal., **163** (1999), 1–28.

[Bré99] P. Brémaud, Markov Chains, Texts Appl. Math., **31**, Springer, New York, 1999.

[BS] A.-I. Bonciocat and K.-T. Sturm, Mass transportation and rough curvature bounds for discrete spaces, preprint.

[Che04] M.-F. Chen, From Markov Chains to Non-equilibrium Particle Systems. Second ed., World Scientific, 2004.

[CW94] M.-F. Chen and F.-Y. Wang, Application of coupling method to the first eigenvalue on manifold, Sci. China Ser. A, **37** (1994), 1–14.

[DC92] M. P. do Carmo, Riemannian Geometry, Math. Theory Appl., Birkhäuser, 1992.

[DGW04] H. Djellout, A. Guillin and L. Wu, Transportation cost—information inequalities and applications to random dynamical systems and diffusions, Ann. Prob., **32** (2004), 2702–2732.

[Dob70] R. L. Dobrušin, Definition of a system of random variables by means of conditional distributions (Russian), Teor. Verojatnost. i Primenen., **15** (1970), 469–497; English translation: Prescribing a system of random variables by conditional expectations, Theor. Probability Appl., **15** (1970), 458–486.

[Dob96] R. Dobrushin, Perturbation methods of the theory of Gibbsian fields, In: Lectures on Probability Theory and Statistics, Lectures from the 24th Saint-Flour Summer School, July 7–23, 1994, (ed. P. Bernard), Lecture Notes in Math., **1648**, Springer-Verlag, Berlin, 1996, pp. 1–66.

[DS85] R. L. Dobrushin and S. B. Shlosman, Constructive criterion for the uniqueness of Gibbs field, In: Statistical Physics and Dynamical Systems, Papers from the Second Colloquium and Workshop on Random Fields: Rigorous Results in Statistical Mechanics, (eds. J. Fritz, A. Jaffe and D. Szász), Köszeg, August 26–September 1, 1984, Progr. Phys., **10**, Birkhäuser, Boston, 1985, pp. 347–370.

[DS96] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, Ann. Appl. Probab., **6** (1996), 695–750.

[GL07] N. Gozlan and Ch. Léonard, A large deviation approach to some transportation cost inequalities, Probab. Theory Relat. Fields, **139** (2007), 235–283.

[Gri67] R. B. Griffiths, Correlations in Ising ferromagnets. III, Commun. Math. Phys., **6** (1967), 121–127.

[Gro86] M. Gromov, Isoperimetric inequalities in Riemannian manifolds, In: Asymptotic Theory of Finite Dimensional Normed Spaces, (eds. V. Milman and G. Schechtman), Lecture Notes in Math., **1200**, Springer-Verlag, Berlin, 1986, pp. 114–129.

[Gro99] M. Gromov, Metric Structures for Riemannian and Non-Riemannian Spaces, Progr. Math., **152**, Birkhäuser, 1999.

[Hsu02] E. Hsu, Stochastic Analysis on Manifolds, Grad. Stud. Math., **38**, Amer. Math. Soc., 2002.

[Jou07] A. Joulin, Poisson-type deviation inequalities for curved continuous time Markov chains, Bernoulli, **13** (2007), 782–798.

[Jou] A. Joulin, A new Poisson-type deviation inequality for Markov jump process with positive Wasserstein curvature, preprint.

[Led01] M. Ledoux, The concentration of measure phenomenon, Math. Surveys Monogr., **89**, Amer. Math. Soc., 2001.

[Lév22] P. Lévy, Leçons d'analyse fonctionnelle, Gauthier-Villars, Paris (1922). New edition in 1951 under the title Problèmes concrets d'analyse fonctionnelle.

[LV] J. Lott and C. Villani, Ricci curvature for metric-measure spaces via optimal transport, preprint.

[Mar04] F. Martinelli, Relaxation times of Markov chains in statistical mechanics and combinatorial structures, In: Probability on Discrete Structures, (ed. H. Kesten), Encyclopaedia Math. Sci., **110**, Springer-Verlag, Berlin, 2004, pp. 175–262.

[Mas07] P. Massart, Concentration inequalities and model selection, In: Lectures from the 33rd Summer School on Probability Theory, Saint-Flour, July 6–23, 2003, Lecture Notes in Math., **1896**, Springer-Verlag, 2007.

[Oht] S.-i. Ohta, Finsler interpolation inequalities, preprint.

[Oli] R. Imbuzeiro Oliveira, On the convergence to equilibrium of Kac's random walk on matrices, preprint, arXiv:0705.2253.

[Oll07] Y. Ollivier, Ricci curvature of metric spaces, C. R. Math. Acad. Sci. Paris, **345** (2007), 643–646.

[Oll09] Y. Ollivier, Ricci curvature of Markov chains on metric spaces, J. Funct. Anal., **256** (2009), 810–864.

[Oll] Y. Ollivier, Géométrie et concentration de la mesure, présentation du groupe de travail, http://www.umpa.ens-lyon.fr/~yollivie/publs/geoconc.pdf.

[OS] S.-i. Ohta and K.-T. Sturm, Heat flow on Finsler manifolds, preprint.

[Pet06] P. Petersen, Riemannian Geometry. Second ed., Grad. Texts in Math., **171**, Springer-Verlag, 2006.

[RS05] M.-K. von Renesse and K.-T. Sturm, Transport inequalities, gradient estimates, and Ricci curvature, Comm. Pure Appl. Math., **68** (2005), 923–940.

[Sch01] G. Schechtman, Concentration, results and applications, In: Handbook of the Geometry of Banach Spaces, **2**, North-Holland, Amsterdam, 2003, pp. 1603–1634.

[Stu06] K.-T. Sturm, On the geometry of metric measure spaces, Acta Math., **196** (2006), 65–177.

[Vil03] C. Villani, Topics in Optimal Transportation, Grad. Stud. Math., **58**, Amer. Math. Soc., Providence, RI, 2003.

[Vil08] C. Villani, Optimal Transport. Old and New, Grundlehren Math. Wiss., **338**, Springer-Verlag, 2008.

CNRS
UMPA, École normale supérieure de Lyon
46, allée d'Italie
69364 Lyon Cedex 07
France

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