

# SYNTHETIC THEORY OF RICCI CURVATURE BOUNDS

CÉDRIC VILLANI

ABSTRACT. Synthetic theory of Ricci curvature bounds is reviewed, from the conditions which led to its birth, up to some of its latest developments.

## INTRODUCTION

This is the story of a mathematical theory which was born from the encounter of several fields. Those fields were: Riemannian geometry, gradient flows, information, and optimal transport; and their encounter took place about 15 years ago. A major result of the interaction was the synthetic theory of Ricci curvature bounds, which was formalized around 2005 and now seems to be reaching maturity, after ten years of fast, sustained growth. As the theory is still in rapid evolution, it is often difficult, even for experts, to trace its progress accurately.

In this short survey, intended to serve as written source for my Takagi lectures in June 2015, I shall

- comment on the meaning of “synthetic”;
- review in haste the four fields involved;
- summarize the way in which the interaction takes place;
- describe the fundamentals of the theory;
- review the connections to a few other geometric theories;
- discuss possible future directions of development;
- list a few selected references, many of them landmarks in the field.

Throughout the document, I shall insist on the global picture, at the expense of technical details.

It is a pleasure to thank three close collaborators whose help was immeasurable during the years in which I actively worked on the subject: Felix Otto, John Lott, and Luigi Ambrosio. To this short list I will add Yann Brenier and Wilfrid Gangbo, for giving me opportunities to jump in the field; Luis Caffarelli and Craig Evans, for their early support; Michel Ledoux and Eric Carlen, for the enormous influence which they had on me; Robert McCann, for his pioneering role; Karl-Theodor Sturm, for

his huge impact on the theory and his constant help; and the beautiful international small crowd of bright young mathematicians who have been enriching the subject at a fast pace, and more than once resolved problems which seemed unsurmountable. Final thanks are due to the organizers of the Takagi lectures, for the honor and pleasure brought upon me by this invitation. A preliminary version of this set of notes was written for the Chern Lectures in the Mathematical Center of Tsinghua University, January 2015; it was also the basis of my contribution to the ceremony of the 50th anniversary of the FIM in ETH Zürich, June 2015.

### 1. SYNTHETIC POINT OF VIEW

As in my book [53] I will use the well-known notion of convexity to illustrate the concept of “synthetic” point of view, and compare it with the concept of “analytic” point of view.

If  $\Omega$  is a convex open set in  $\mathbb{R}^n$  there are two natural ways to define the notion of convexity for a function  $\varphi : \Omega \rightarrow \mathbb{R}$ :

- (a)  $\forall x \in \Omega, \quad \nabla^2 \varphi(x) \geq 0$ , where  $\nabla^2 \varphi$  is the Hessian matrix of  $\varphi$ ;
- (b)  $\forall x, y \in \Omega, \quad \forall t \in [0, 1], \quad \varphi((1-t)x + ty) \leq (1-t)\varphi(x) + t\varphi(y)$ .

Definition (a) is local and effective: it can often be checked in practice with reasonable effort; it also comes with an auxiliary quantity, the Hessian of  $\varphi$ , which can be used to refine and quantify the notion of convexity. This is the *analytic* definition, based on a computation. It stands in contrast to Definition (b) which is less precise and almost impossible to check directly.

In spite of its shortcomings, Definition (b) also has enormous advantages. First, it is more general: no need for the Hessian to be well defined, the definition also makes sense for nonsmooth convex functions such as  $\varphi : x \mapsto |x|$ . It is also very useful, as the starting point for so many convexity inequalities and results. It is the *synthetic* definition, relying on qualitative properties (in this case, a geometric property of the graph) rather than computations.

In fact, it is the combination of formulations (a) and (b) which turns out to be so fruitful: an important part of our convexity games involves Definition (a) to establish convexity, and Definition (b) to use it.

Then of course, it is crucial that both formulations be equivalent: such is the case when  $\varphi$  is, say,  $C^2$ . And, as said earlier, when that is not the case, we still can use formulation (b).

Actually, more is true: if (b) holds, then *automatically*  $\varphi$  enjoys some regularity and some version of (a) holds true. This is the content of Alexandrov’s regularity

theorem: if  $\varphi : \Omega \rightarrow \mathbb{R}$  is convex (in sense (b)), then  $\varphi$  is twice almost everywhere differentiable, so  $\nabla^2\varphi$  exists almost everywhere, and it is nonnegative wherever it is defined.

In Euclidean geometry there is also such a distinction between analytic geometry (based on equations on coordinates – linear equations for lines, quadratic equations for parabolas, circles, ellipses, etc.) and synthetic geometry (in the ancient Greek style, with properties of circles, parallel lines, etc.) Both have their rewards, the analytic approach being often more effective and the synthetic one more economical, more to the point, and, well, more synthetic.

## 2. THE FORCES IN PRESENCE

The four theories which follow are all well-established and respectable; and, for a long time, roughly up to the mid-nineties, they lived happily without any connection — except, arguably, for some links between curvature and gradient flow, for instance in positively curved Alexandrov spaces or the theory of Ricci flow.

**2.1. Curvature theory.** Curvature is the basis of non-Euclidean geometry. First introduced by Gauss and further developed by his student Riemann, curvature measures the degree of non-Euclidean behavior in geodesics. It comes in several variants.

**2.1.1. Sectional curvature.** Sectional curvature is the simplest and most fundamental notion of curvature. If  $x$  is a point in a Riemannian manifold and  $u, v$  are two unit orthogonal tangent vectors at  $x$ , then the sectional curvature  $\kappa = \kappa(u, v)$  may be defined as the dominant non-Euclidean correction to the distance between the geodesics  $\gamma_u(t) = \exp_x(tu)$  and  $\gamma_v(t) = \exp_x(tv)$  (which have constant speed and emanate from  $x$  with velocities  $u$  and  $v$  respectively):

$$d(\gamma_u(t), \gamma_v(t)) = \sqrt{2}t \left( 1 - \frac{\kappa}{12}t^2 + O(t^3) \right) \quad \text{as } t \rightarrow 0. \quad (1)$$

Notice the negative sign in front of  $\kappa$ : positive curvature means reduced distances. It can be shown that  $\kappa$  actually depends only on the tangent plane, the *section* generated by  $u$  and  $v$ .

The form of inequality (1) makes it clear that  $\kappa$  is invariant by isometry. Spaces of constant sectional curvatures  $\kappa$  are classified: locally they are Euclidean spaces ( $\kappa = 0$ ),  $n$ -dimensional spheres ( $\kappa > 0$ ), and hyperbolic spaces ( $\kappa < 0$ ).

2.1.2. *Synthetic sectional curvature.* A sectional curvature bound may also be introduced in a synthetic way, by *comparison* to geometric properties in a space of constant curvature. For instance, one may define nonnegative sectional curvature by the property that all (sufficiently small) geodesic triangles have the sum of their angles at least equal to  $\pi$ : with obvious notation,  $\sum \alpha_i \geq \pi$ . Equivalently, one may define nonnegative sectional curvature by the property that for any triangle  $abc$ , if  $m$  is the midpoint of  $(bc)$  and  $a_0b_0c_0$  is an isometric triangle drawn on  $\mathbb{R}^2$ , and  $m_0$  is the midpoint of  $(b_0c_0)$ , then  $d(a, m) \geq d(a_0, m_0)$ ; in short, triangles in positive curvature have longer medians. At first presentation it looks plausible that such a basic property can be used to significantly translate curvature properties into geometric statements; such is indeed the case [7].

2.1.3. *Ricci curvature.* When dealing with surfaces, sectional curvature, then called Gauss curvature, is all that one has to know. In dimension  $n \geq 3$ , further notions of curvature must be defined, and more care is needed.

The Ricci curvature is famous, among others, for its use in general relativity and in Ricci flow. This tensor is defined as follows: if  $u = e_1$  is a unit vector, then introduce  $(e_2, \dots, e_n)$  in such a way that  $(e_1, \dots, e_n)$  is an orthonormal basis of  $T_x M$ , and let

$$\text{Ric}(u) := \sum_{j=2}^n \kappa(e_1, e_j). \quad (2)$$

Not only does this turn out to be independent of the choice of  $(e_2, \dots, e_n)$ , but it is also a quadratic form in  $u$ . This is the Ricci curvature: it tells about *volume distortion* rather than distance distortion.

There is a formula which captures very well the infinitesimal meaning of Ricci curvature: if  $\xi$  is a vector field, and  $\mathcal{J}(t) := \det(d_x \exp_x(t\xi))$ , then

$$\frac{d^2}{dt^2}(\mathcal{J}^{1/n}) + \frac{\text{Ric}(\dot{\gamma}, \dot{\gamma})}{n}(\mathcal{J}^{1/n}) \leq 0. \quad (3)$$

Here  $\dot{\gamma}(t) = \exp_x(t\xi)$  is the geodesic starting from  $x$  with velocity  $\xi(x)$ . This tells us about the infinitesimal rate of change of volume along the flow generated by  $\xi$ . As usual in fluid mechanics, there is also an equivalent Eulerian formulation:

$$\xi \cdot \nabla(\nabla \cdot \xi) - \Delta \frac{|\xi|^2}{2} + \frac{(\nabla \cdot \xi)^2}{n} + \text{Ric}(\xi, \xi) \leq 0. \quad (4)$$

Usually the latter formula is stated for  $\xi = \nabla\psi$ , in the form

$$-\nabla\psi \cdot \nabla\Delta\psi + \Delta \frac{|\nabla\psi|^2}{2} \geq \frac{(\Delta\psi)^2}{n} + \text{Ric}(\nabla\psi, \nabla\psi). \quad (5)$$

Comparison of (1) on the one hand, (3) or (4) on the other hand, shows several important differences in principle:

- While (1) is about distances, inequality (3) on the contrary is in terms of Jacobian determinant, which means volume. This explains why Ricci curvature is the favorite curvature of probabilists. At the same time, probability theory will very often involve changes of measures; to keep the distortion interpretation, one will need to twist the definition of Ricci curvature.

- Dimension appears in (3), and in fact curvature and dimension are woven together in this inequality; while dimension played no role in sectional curvature.

- Formula (3) is an inequality, while (1) was an equality. The inequality is unavoidable, because at some place in the derivation a Cauchy–Schwarz inequality is used, in the form  $\|\nabla\xi\|^2 \geq (\nabla \cdot \xi)^2/n$ . In fact there is the neat *Bochner identity*,

$$-\nabla\psi \cdot \nabla\Delta\psi + \Delta\frac{|\nabla\psi|^2}{2} = \|\nabla^2\psi\|^2 + \text{Ric}(\nabla\psi, \nabla\psi); \quad (6)$$

but it is not easy to use, for instance because  $\nabla^2\psi$  does not have any simple relevant interpretation, while  $\Delta\psi$  does have one, as the divergence of the vector field  $\nabla\psi$ .

Inequality (5) can be seen as a property of the Laplace operator, noting that  $|\nabla\psi|^2 = (\Delta\psi^2 - 2\psi\Delta\psi)/2$ . Conversely, one may replace  $\Delta$  by a diffusion-drift operator, say  $L = \Delta - \nabla V \cdot \nabla$ ; in a measure-theoretical view, this amounts to replace the uniform volume measure by the Boltzmann-type measure  $e^{-V}\text{vol}$ , which is the equilibrium measure for  $L$ . Then it makes sense to define the  $\text{CD}(K, N)$  criterion for the linear diffusion operator  $L$ :

$$-\nabla\psi \cdot \nabla L\psi + L\frac{|\nabla\psi|^2}{2} \geq \frac{(L\psi)^2}{N} + K|\nabla\psi|^2. \quad (7)$$

Morally, this means that the Ricci curvature is bounded below by  $K$  and the dimension is bounded above by  $N$ ; here  $K$  might be any real number, while  $N$  might be any positive number (in practice  $N \geq 1$ ). As a typical example, the Gaussian space  $(\mathbb{R}^n, \gamma)$ , where  $\gamma$  is the standard (unit covariance matrix) Gaussian distribution, can be considered as a  $\text{CD}(1, \infty)$  space, independently of the dimension: this property plays a crucial role in the interplay between statistics and geometry, and is consistent with the well-known fact that a Gaussian distribution is like the infinite-dimensional counterpart of the uniform measure on the sphere. The corresponding  $\text{CD}(K, \infty)$  criterion correspond to

$$\frac{d^2}{dt^2}(\log \mathcal{J}) + K(\log \mathcal{J}) \leq 0. \quad (8)$$

and

$$-\nabla\psi \cdot \nabla L\psi + L \frac{|\nabla\psi|^2}{2} \geq K|\nabla\psi|^2. \quad (9)$$

Inequalities (7) and (9), or their Lagrangian counterparts, are the working heart of Ricci curvature analysis. They are used in countless applications: measure growth control, isoperimetric inequalities, spectral gap estimates, heat kernel bounds, measure concentration theory, control of stochastic processes, etc.

A relatively recent but emblematic such estimate is the *curved Brunn–Minkowski inequality*: if  $X$  and  $Y$  are two compact sets in a manifold  $M$  with nonnegative Ricci curvature, and  $M$  is the set of midpoints of  $X$  and  $Y$ , then

$$\text{vol}[M]^{1/n} \geq \frac{1}{2} \left( \text{vol}[X]^{1/n} + \text{vol}[Y]^{1/n} \right). \quad (10)$$

This captures the idea that in positive curvature, *intermediate points form a large set* (this is in accordance with the fact that the flow has a tendency to reduce distances: to get the target right, one has to start by enlarging the volume more than one would expect). Note the formal relation with (3): the same exponent  $1/n$ , and a convexity inequality, expressed for the infinitesimal volume there, and for the integral volume here. Further note that when the manifold is  $\mathbb{R}^n$ , then (10) reduces to the classical “algebraic” Brunn–Minkowski inequality:

$$\left| \frac{X+Y}{2} \right|^{1/n} \geq \frac{1}{2} (|X|^{1/n} + |Y|^{1/n}).$$

2.1.4. *Scalar curvature.*  $S = \text{tr}(\text{Ric})$  is the scalar curvature, much more tricky to interpret. It does play a fundamental role in a number of problems in analysis. I shall say nothing about it here, and refer to the survey [26] and the treatises [16, 20] for an overview of the theory of curvature.

2.2. **Nonsmooth gradient flows.** A gradient flow is defined by two ingredients: a geometry and a function, which may be called “energy” even though its physical meaning may be very different. Typically the geometry is provided by a Riemannian manifold  $\mathcal{M}$  and the energy is a function  $\mathcal{E} : \mathcal{M} \rightarrow \mathbb{R}$ . The Riemannian structure transforms the differential of  $U$  into a vector field, called the gradient vector field; then the gradient flow is defined by the ordinary differential equation

$$\dot{X} = -\text{grad } \mathcal{E}(X). \quad (11)$$

Gradient flows appear in a number of contexts and are associated with all kinds of dissipative phenomena. In various respects they are the dissipative counterpart of Hamiltonian flows in the conservative world. They are very popular in optimization,

as a way to search for the minimum of  $\mathcal{E}$ ; for instance they are a key tool in modern artificial intelligence theory.

Gradient flows can be defined in finite or infinite dimension. For instance, the heat equation is the gradient flow of the Dirichlet energy

$$\mathcal{E}(u) = \frac{1}{2} \int |\nabla u|^2$$

in the  $L^2$  (Hilbert) geometry.

The need for gradient flows in a nonsmooth context arose long ago. De Giorgi developed it with applications to image processing in mind. In the absence of smoothness, the gradient of  $\mathcal{E}$  may be ill-defined, but this can be often compensated by convexity or semiconvexity properties: as a general rule, the gradient flow of a convex functional is well-defined, nonexpanding, and so on. To define and study nonsmooth gradient flows, two main strategies emerged.

*2.2.1. Time discretization: Generalized minimizing movements.* This formalism was a favorite of the Italian school, starting with De Giorgi. Choose a time step in the form of a small number  $\tau > 0$ . Starting from an initial condition  $X_0$ , define  $X_0^{(\tau)} = X_0$  and for each  $k \in \mathbb{N}$

$$X_k^{(\tau)} := \operatorname{argmin}_X \left[ \mathcal{E}(X) + \frac{d(X_{k-1}^{(\tau)}, X)^2}{2\tau} \right]. \quad (12)$$

Then as  $\tau \rightarrow 0$  and  $k\tau \rightarrow t$ , show that, extracting a subsequence if necessary,

$$X_k^{(\tau)} \simeq Y(k\tau) \simeq Y(t);$$

the limit  $Y$  is a generalized gradient flow.

There is nothing mysterious in equation (12): if  $X_k^{(\tau)} \simeq X_{k-1}^{(\tau)}$ , and the geometry is infinitesimally Hilbert, the approximate Euler–Lagrange equation should be

$$\operatorname{grad} \mathcal{E}(X_k^{(\tau)}) + \left( \frac{X_k^{(\tau)} - X_{k-1}^{(\tau)}}{\tau} \right) = 0,$$

which is a discrete formulation of (11).

Of course if  $\mathcal{E}$  is lower semicontinuous, then the minimization problem in (12) has a solution. The program can be carried out under various semi-continuity assumptions for  $\mathcal{E}$  and  $|\operatorname{grad} \mathcal{E}|$ ; see [3]. One can prove typically the 1/2-Hölder regularity of  $Y(t)$  in the time variable  $t$ , the finiteness of  $\mathcal{E}(t)$  for any  $t > 0$ , etc.

2.2.2. *Evolution Variational Inequalities.* The other approach goes back to B enilan and others: it relies on the variations of geodesics. If  $Z$  is a point in the phase space, then one may differentiate the distance of  $Z$  to the curve  $X(t)$ : whenever that differential is well-defined,

$$\frac{d}{dt} \left( \frac{d(Z, X(t))^2}{2} \right) = \left\langle \frac{dX}{dt}, \frac{d\gamma}{ds} \right\rangle \Big|_{s=1}, \quad (13)$$

where  $\gamma$  is a minimizing geodesic joining  $\gamma(0) = Z$  to  $\gamma(1) = X(t)$ . If  $X(t)$  evolves according to (11), then the right-hand side in (13) coincides with  $-(d/ds)\mathcal{E}(\gamma(s))$ . So the idea is to compare  $(d/dt)d(Z, X(t))^2/2$  with  $-(d/ds)\mathcal{E}(\gamma(s))$ . These are rates of variation along different time evolutions. This may be simplified by using a convexity assumption: indeed, for instance, if  $\mathcal{E}$  is convex, then

$$\frac{d}{ds} \Big|_{s=1} \mathcal{E}(\gamma(s)) \geq \mathcal{E}(\gamma(1)) - \mathcal{E}(\gamma(0)) = \mathcal{E}(X(t)) - \mathcal{E}(Z).$$

Further, even if the distance function is not differentiable, it can always be upper differentiated. Then the evolution inequality becomes

$$\forall Z, \quad \frac{d^+}{dt} \left( \frac{d(Z, X(t))^2}{2} \right) \geq \mathcal{E}(X(t)) - \mathcal{E}(Z). \quad (14)$$

This is in case  $\mathcal{E}$  is convex; there are similar formulations for semiconvex functionals  $\mathcal{E}$ . Many variants have been explored; once again I shall refer to [3] for a complete overview.

2.3. **Information theory.** Information theory was born formally in the fifties, through the efforts of Shannon to quantify the flux of information in a language, or signal. To this aim certain key quantities were introduced; the most famous is the entropy

$$S_\nu(\mu) = - \int \rho \log \rho \, d\nu, \quad \rho = \frac{d\mu}{d\nu}. \quad (15)$$

Often one works with the information  $H_\nu(\mu) = -S_\nu(\mu)$ , which is convex (and unfortunately called ‘‘entropy’’ by many mathematicians). The logarithmic form imposed itself to Shannon from two basic principles:

- the typical ‘‘informational value’’ of an elementary signal should be a function of its frequency  $\rho$  (rare is valuable, so the elementary information should be an increasing function of  $1/\rho$ )
- independent signal sources should add up their informations.



Since independence is associated with multiplication of probabilities, going from multiplication to addition was done through the logarithmic function; then the integration of  $\log(1/\rho)$  gave rise to  $-\int \rho \log \rho$ . Simple as it was, this procedure had led Shannon to recover Boltzmann's celebrated formula, which is associated with a rich history and many important theorems in statistics. In its most general formulation, Boltzmann's formula can be summarized by Sanov's theorem:

$$\mathbb{P} \left[ \frac{1}{N} \sum \delta_{X_i} \simeq \mu \right] \simeq e^{-NH_\nu(\mu)}, \quad (16)$$

where the  $X_i$  are independent random variables distributed according to the law  $\nu$ .

The rate of entropy variation along the heat equation is another cornerstone of information theory:

$$\frac{d}{dt} H_\nu(\mu_t) = -I_\nu(\mu_t), \quad (17)$$

where  $(\mu_t)$  solves the diffusion equation  $\partial_t \mu_t = L\mu_t$ ,  $L\mu = \Delta\mu + \nabla \cdot (\mu \nabla V)$ ,  $\nu = e^{-V} \text{vol}$ , and  $I_\nu$  is the Fisher information:

$$I_\nu(\mu) = \int \frac{|\nabla \rho|^2}{\rho} d\nu, \quad \rho = \frac{d\mu}{d\nu}. \quad (18)$$

This functional is also a key object in statistics, appearing for instance as the asymptotic covariance matrix of the maximum likelihood estimate. For more on the subject I shall refer to the classical book by Cover and Thomas [15].

All kinds of variants exist, in the form of functionals of the type

$$U_\nu(\mu) = \int U \left( \frac{d\mu}{d\nu} \right) d\nu, \quad (19)$$

where  $U$  is a nonlinearity. Typically one assumes  $U(0) = 0$ , and  $U$  convex; there are some physical and mathematical reasons for that. The resulting functionals have variable informational content and come up in a huge number of modelling issues, for instance as energy-type functionals associated with nonlinear diffusion equations.

**2.4. One ring to bring them all and bind them.** The theory which will act as a link between gradient flows, information theory and geometry is the theory of optimal transport.

**2.4.1. Introduction.** Optimal transport theory is one of the most versatile theories developed in analysis over the past decades. Of the four theories which are reviewed here, it is arguably the oldest and the most changing. It takes its roots in two founding papers:

- the 1781 memoir by Monge asking about the most economical way to transport a distribution of mass from a given state to a prescribed one;
- the 1942 paper by Kantorovich investigating the same problem in a probabilistic formulation, with computability and optimization in mind.

Both Monge and Kantorovich were outstanding scientists whose careers were lying at the intersection of science and public policy. The life of Monge was somewhat romanticized by E.T. Bell in his classics, *Men of Mathematics*; while Kantorovich is one of the heroes in the unclassifiable modern novel *Red Plenty* by Francis Spufford.

The basic data in optimal transport consists of a cost function  $c$  in two variables, and two probability measures, say  $\mu_0$  and  $\mu_1$ , defined on a source space  $\mathcal{X}_0$  and a target space  $\mathcal{X}_1$  respectively. Then one defines

- the *Monge variational problem*,

$$\inf_T \left\{ \int c(x, T(x)) \mu_0(dx); \quad T_{\#}\mu_0 = \mu_1 \right\}. \quad (20)$$

Recall that  $T_{\#}\mu[A] = \mu[T^{-1}(A)]$  defines the image measure of  $\mu$  by  $T$ , which is also an abstract way to introduce a change of variables. Here  $T$  is called the transport map. Problem (20) is a highly nonlinear, nonconvex minimization problem, first because  $c$  is nonlinear and in general nonconvex, but also because of the complicated constraint on  $T$ , which in a Euclidean context may involve the Jacobian determinant  $\det(dT)$ .

- the *Kantorovich variational problem*,

$$\inf_{\pi} \left\{ \iint c(x, y) \pi(dx dy); \quad \pi \in \Pi(\mu_0, \mu_1) \right\}. \quad (21)$$

Here  $\Pi(\mu_0, \mu_1)$  is the set of joint probability measures having marginals  $\mu_0$  and  $\mu_1$  respectively, and  $\pi$  is called a *transport plan*. The ansatz  $\pi = (\text{Id}, T)_{\#}\mu_0$  reduces (21) to (20); thus the Kantorovich problem can be seen as a relaxed version of the Monge problem, even though it was in fact introduced independently. In the Kantorovich problem, the infimum is achieved under minimal assumptions, while more stringent conditions are required for the existence of a minimizer in the Monge problem.

Both the Monge and the Kantorovich problem are expressions of the same problem: how to rearrange a given distribution of mass  $\mu_0$ , defined on  $\mathcal{X}_0$ , into another distribution of mass  $\mu_1$ , defined on  $\mathcal{X}_1$ . Under very general assumptions (say if  $\mu_0$

has no atoms and  $c$  is continuous), the values of the optimum in (20) and (21) coincide; this common value is called the optimal transport cost. It is often the case that  $\mathcal{X}_0 = \mathcal{X}_1$ , but this might not always be.

2.4.2. *Duality.* For a long time the most visible part of optimal transport theory was the quest for duality results and their use. Indeed, the Kantorovich variational problem belongs to the theory of linear programming, so it admits a dual variational result, and this is expressed through the Kantorovich duality: under very general assumptions,

$$\min_{\pi \in \Pi(\mu_0, \mu_1)} \iint c(x, y) \pi(dx dy) = \sup_{\psi_1 - \psi_0 \leq c} \left\{ \int \psi_1(y) d\mu_1(y) - \int \psi_0(x) d\mu_0(x) \right\}. \quad (22)$$

In the right-hand side the expression  $\psi_1 - \psi_0 \leq c$  really means

$$\forall x, y, \quad \psi_1(y) - \psi_0(x) \leq c(x, y).$$

There is an economical interpretation of (22) in terms of economy, prices and profit (which, incidentally, made this theorem look like heresy to the eyes of ideologists of communist Russia); then  $\psi_0$  and  $\psi_1$  should be interpreted as prices, and the condition is about the difference of values, between the start and the end of the transport, never exceeding the transport cost.

From there the notion of  $c$ -transform arises naturally: if  $\psi$  is any function on  $\mathcal{X}_0$ , define a “dual” function on  $\mathcal{X}_1$  by the formula

$$\psi^c(y) = \inf_{x \in \mathcal{X}_0} [\psi(x) + c(x, y)];$$

and symmetrically, if  $\psi$  is any function on  $\mathcal{X}_1$ , the dual function on  $\mathcal{X}_0$  will be

$$\psi_c(x) = \sup_{y \in \mathcal{X}_1} [\psi(y) - c(x, y)].$$

Then optimization over the price functions and optimization over the transport map become strongly related issues.

2.4.3. *A turn in the theory: characterizing the optimal map.* A huge revival of optimal transport theory started at the end of the eighties, when three independent contributions of Mather, Brenier and Cullen showed that optimal transport can bring great help in other fields such as dynamical systems and partial differential equations. (It took long before Mather’s contribution was explicitly recast in terms of optimal transport, even though this makes it crystal clear in retrospect [53].) At the same time, Brenier and Cullen showed the importance of solving the Monge problem and understanding the structure of the optimal transport map.

An emblematic result from that period is Brenier’s theorem (independently obtained by Rachev–Rüschendorf): if  $c(x, y) = |x - y|^2$  in  $\mathbb{R}^n$  and  $\mu_0$  is absolutely continuous with respect to Lebesgue measure, then the optimal transport map  $T$  is uniquely defined and characterized by the form

$$T = \nabla\varphi, \quad \varphi \text{ a convex function.} \quad (23)$$

The extension to Riemannian manifolds was achieved by McCann at the end of the nineties: when the cost function is  $c(x, y) = d(x, y)^2$  on a Riemannian manifold  $M$ , then

$$T = \exp(\nabla\psi), \quad \psi \text{ a } c\text{-convex function.} \quad (24)$$

Here  $c$ -convex means that  $\psi = (\psi^c)_c$ , or equivalently that there exists  $\zeta : M \rightarrow \mathbb{R} \cup \{-\infty\}$  such that  $\psi = \zeta_c$ , that is

$$\forall x \quad \psi(x) = \sup_y \left[ c(x, y) - \zeta(y) \right]. \quad (25)$$

This may be thought of as a generalized Legendre transform, adapted to the cost function.

Here are two important remarks on the notion of  $c$ -convexity:

- $c$ -convexity implies semiconvexity, so  $\psi$  is just as regular as a convex function;
- any  $C^2$ -small convex function on a compact Riemannian manifold  $M$  is  $c$ -convex.

In particular,  $c$ -convex functions form a “rich” family.

Another case of importance is when  $c(x, y) = d(x, y)$ ; this was the first cost considered by Monge, and it was also favored by Kantorovich because it leads to a simpler duality formula. It is however very tricky to study, and general conclusions did not appear before the end of the nineties [2, 12, 9, 21]. The main results in the field give a representation of optimal transport for  $c = d$  in a Riemannian manifold:

- the shared mass does not matter, and can be “cancelled out” since it does not need to move: in fact the transportation problem can be recast as a transshipment problem, depending only on the difference  $\mu - \nu$ ;
- nontrivial transport occurs along geodesic rays, and the direction of transport is given by a gradient  $\nabla\Psi$ ;
- on each ray one may rearrange the transport in a number of ways, but there is a distinguished monotone transport.

2.4.4. *Further structure.* Another important concept which has been associated with optimal transport since the time of Kantorovich and which has been much revived recently is that of *transport distance*. For  $p \in [1, \infty)$  let us define  $C_p(\mu_0, \mu_1)$  as the

value of the optimal transport cost between  $\mu_0$  and  $\mu_1$  when the cost function is  $c(x, y) = d(x, y)^p$ ; and

$$W_p(\mu_0, \mu_1) = C_p(\mu_0, \mu_1)^{1/p}. \quad (26)$$

This defines a family of distances on the set of probability measures (with some subtleties at infinity if the space is noncompact), metrizing the topology of weak convergence. The denomination of Wasserstein distance has become standard for these distances, even though it is historically not very accurate. By far the most important cases are again  $p = 1$  and  $p = 2$ . They correspond to the same topology (if the base space is compact), but to distinct geometries. If the base space is not compact, there might be some differences; the space of probability measures with finite  $p$ -moment will be denoted by  $P_p(\mathcal{X})$ , and may be equipped with the distance  $W_p$ .

Then there is also a notion of *transport interpolation*, when the base space itself admits an interpolation, in particular if it is a geodesic space. Let us write  $I_t(x, y)$  for the geodesic interpolation between  $x$  and  $y$  at time  $t \in [0, 1]$  (let us assume that there is such a recipe, for instance defined almost everywhere); then if  $\pi$  is a transport plan we may deduce from it an interpolation at the level of measures:

$$\mu_t = (I_t)_\# \pi, \quad t \in [0, 1]. \quad (27)$$

In the mid-nineties, McCann showed how this procedure could lead to precious insight in certain geometric properties; following his usage, this is called *displacement interpolation*, and was pushed extremely far later.

Both the distance  $W_p$  and the displacement interpolation endow the space of probability measures, say  $P(\mathcal{X})$ , with a geometry of its own, which is deduced from the geometry on  $\mathcal{X}$  but also contains it, through the identification of  $x$  with  $\delta_x$ . A key discovery from the past fifteen years is that *relevant information on the geometry of  $\mathcal{X}$  can be expressed through geometric properties of  $P(\mathcal{X})$* .

2.4.5. *Further reference.* The existence, uniqueness and regularity of optimal transport, the possibility of interpolation, the underlying partial differential equations: all this has turned into a huge field in the past two decades. My book [53] was written to be a reference source on the subject, at a time when it was evolving at fast speed.

### 3. ENCOUNTER OF THE THIRD TYPE

The encounter which led to the synthetic theory of Ricci curvature was one of strong interaction and even hijacking. It modified the geometric picture and, to some extent, the interpretation of Ricci curvature bounds; and this was done from

the intervention of a theory that seemed very remote. In this section I shall say a bit more about the way in which this occurred.

**3.1. Precursors.** At least two precursors can be mentioned. The first one was McCann’s PhD Thesis [37], studying convexity properties of information-theoretical functionals in the geometry of optimal transport. A main result, slightly restated, was that if a convex nonlinearity  $U$  is given with  $U(0) = 0$ , then the functional

$$\mathcal{U} : \mu \longmapsto \int U \left( \frac{d\mu}{dx} \right) dx \quad (28)$$

on  $P(\mathbb{R}^n)$ , is geodesically convex in  $W_2$  if and only if

$$s^n U(s^{-n}) \text{ is a convex function of } s \in \mathbb{R}_+. \quad (29)$$

Condition (29) is saturated when  $U(r) = -r^{1-1/n}$ ; McCann showed how this  $1/n$  is related to the same exponent in the Brunn–Minkowski inequality.

The second precursor, playing a more indirect role, was the discovery by Marton [36] that comparisons between optimal transport distance  $W_p$  and information theoretical functionals like  $H_\nu$  provide a powerful way to encode concentration properties. This field would later be developed by Talagrand and others. One of the neat results proven by Talagrand is that if  $\gamma$  stands for the normalized Gaussian distribution on  $\mathbb{R}^n$ , then

$$\forall \mu \in P_2(\mathbb{R}^n), \quad W_2(\mu, \gamma) \leq \sqrt{2H_\gamma(\mu)}. \quad (30)$$

This “transport-entropy inequality”, called Talagrand inequality, compares two fundamental ways to measure the discrepancy between two probability measures. It is related to the theory of concentration of measure: the right-hand side tells about the initial spreading of  $\mu$  in Gaussian space (for instance if  $\mu$  is the restriction of  $\gamma$  to a subset  $C$ , then the right-hand side depends on the logarithm of the volume of  $C$ ); while the left-hand side tells us about the effort needed to “invade” the whole space if we spread the mass of  $\mu$ . So this is a functional, “probabilistic” way to express the basic mantra of measure concentration, namely that if a set is not too small in probability measure, then its enlargement quickly takes up most of the measure. Pioneered by Lévy and Dvoretzky, concentration theory has turned into a huge field, reviewed by Ledoux [30].

**3.2. Jordan–Kinderlehrer–Otto.** In 1998 Jordan, Kinderlehrer and Otto published a paper of considerable importance. This paper was not solving any known open problem, nor was it technically very difficult; but it was providing a new view on a well-known object. Cheating a bit, one may summarize its contents in one

sentence: *The gradient flow of the information functional  $H_\nu$ , in the geometry of  $W_2$ , is the heat equation.* This theorem, stated by the authors in  $\mathbb{R}^n$ , would later be proven and reproven in insane generality, all the way to metric spaces in a certain sense. It would also be adapted to other functionals and distances. Here is a smooth geometric version that will be sufficient to give a flavor of the result.

Let  $M$  be a compact Riemannian manifold and let  $\nu(dx) = e^{-V(x)} \text{vol}(dx)$  be a reference probability measure on  $M$ . Let  $\mu_0$  be an initial probability measure on  $M$ . Apply the procedure described in Subsection 2.2.1 to define  $(\mu_k^{(\tau)})_{k \in \mathbb{N}}$ ; then  $\mu_k^{(\tau)}$  converges in the limit  $\tau \rightarrow 0$ ,  $k\tau \rightarrow t$ , to  $\mu(t)$ , which is the solution, evaluated at time  $t$ , of the natural diffusion equation with equilibrium state  $\nu$ :

$$\frac{\partial \mu}{\partial t} = \Delta \mu_t + \nabla \cdot (\nabla V \mu_t).$$

In the sequel, I shall say “heat equation” for this process even if, from the physical point of view, it is a drift-diffusion equation.

All of a sudden, with this contribution, optimal transport theory and gradient flows were related. This unorthodox construction of a heat flow on the space of measures is now called the JKO scheme and would later prove to be usable in extreme generality.

**3.3. Otto–Villani.** In 1998 I heard Otto lecturing on his gradient flows formalism, and in particular the way to deduce from it a geometric structure on probability measures; he would later publish this in an important conceptual paper [43]. Shortly after my encounter with Otto, I happened to be reading a survey article by Ledoux on concentration theory, which was a precursor of his monograph [30]. It struck me that there should be a link between these two pieces of mathematics, and it did not take long to find it. This triggered my paper with Otto [44].

The main results in our manuscript were

(a) the proof of a new theorem in concentration theory: a logarithmic Sobolev inequality automatically implies a Talagrand inequality with the same constant. More precisely, assuming compactness for simplicity, the information-theoretical inequality

$$\forall \mu \in P(M), \quad H_\nu(\mu) \leq \frac{I_\nu(\mu)}{K}, \quad (31)$$

where  $K > 0$  is a positive constant, implies the transport-entropy inequality

$$\forall \mu \in P(M), \quad W_2(\mu, \nu) \leq \sqrt{\frac{2H_\nu(\mu)}{K}}. \quad (32)$$

(There are versions for noncompact spaces, and again, this theorem has been generalized to an amazing degree, even to plain metric spaces [25].)

(b) the proof of a partial converse under certain Ricci curvature conditions; this was mainly based on the HWI interpolation inequality, which under a  $CD(0, \infty)$  condition expresses the domination of  $H_\nu$  by  $W_2$  and  $I_\nu$  together.

(c) the formalization of a link between curvature, information theory and optimal transport, relying mainly on an expansion of the formal Riemannian intuition exposed in [43]. The Riemannian structure of the distance  $W_2$  was pointed out, thanks to a formula which was due to Brenier–Benamou and reinterpreted by Otto:

$$\|\dot{\mu}\|_\mu = \inf \left\{ \sqrt{\int |\xi|^2 d\mu}; \quad \dot{\mu} + \nabla \cdot (\mu\xi) = 0 \right\}. \quad (33)$$

(d) the emphasis on the Hamilton–Jacobi equation as the evolution for the potential driving the geodesics of optimal transport:

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mu \nabla \varphi) = 0, \quad \frac{\partial \varphi}{\partial t} + \frac{|\nabla \varphi|^2}{2} = 0. \quad (34)$$

(e) a conjecture according to which  $H_\nu$  is convex in the geometry of  $W_2$  if a Ricci curvature condition  $CD(0, \infty)$  holds true for  $\nu$ .

While the formal setting was there, some technical tools were missing, especially in relation to the nonsmoothness of optimal transport and the Riemannian formalism. Much of what was stated there on formal grounds would later be proven and motivate new developments.

By the way, in my personal bibliography this paper is at the same time one of those which were easiest to write and quickest to be accepted (Paul Malliavin, who handled it himself, accepted it the day after it was submitted); but it is also the most quoted and the one which triggered the largest amount of developments.

**3.4. The aftermath.** Two papers which were published after [44] deserve particular notice: Cordero-Erausquin–McCann–Schmuckenschläger [14] and Sturm–Von Renesse [49]. These contributions

- established certain key technical lemmas (generalization of Alexandrov’s theorem, nonsmooth change of variables formula, estimates on Jacobi fields);
- proved some new geometric inequalities such as the curved Brunn–Minkowski inequality;



- showed that  $\text{CD}(K, \infty)$  can be characterized either by properties of uniform convexity of  $H_\nu$  along optimal transport, or by contraction properties of the heat equation. Variants were obtained for all  $\text{CD}(K, N)$  conditions.

With this the scene was ready for the start of the synthetic theory of Ricci bounds. It would mainly consist in *using the properties of certain nonlinear, integral functionals with respect to the geometry of optimal transport, to define curvature-dimension conditions without any reference to smoothness; and to derive geometric consequences thereof.*

The transport cost would be the quadratic geodesic distance:  $c = d^2$ . As for the nonlinearities to be used, they would belong to specific classes: the displacement class of dimension  $N$ ,  $\mathcal{DC}_N$ , is defined as the set of all continuous convex functions  $U : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that  $U(0) = 0$  and

- for  $N < \infty$ ,  $U(s^{-N})s^N$  is convex in  $s$ ,
- for  $N = \infty$ ,  $U(e^{-s})e^s$  is convex in  $s$ .

So by construction the theory would mix information-theoretical functionals and optimal transport. Other ingredients that would later be developed along with the theory include:

- nonsmooth analysis, in particular analysis in metric spaces; this was in line with research done by the Italian and Nordic schools, as well as Cheeger, Sturm and others;
- Hamilton–Jacobi equations, in relation with the Hopf–Lax formula, in non-Euclidean spaces — for the purpose of the theory, basic results were extended from Euclidean to Riemannian, then to geodesic or even metric spaces;
- localization, which is a way to reduce global geometric inequalities to inequalities on geodesic lines; it takes its roots in works of Payne–Weinberger, Gromov–Milman, Kannan–Lovász–Simonovits.

For me, the story was also deeply woven together with my two books on optimal transport, written on the occasion of research courses, in Georgia Tech and in Saint-Flour respectively. The first one, *Topics in Optimal Transportation* [51], was motivated in the first place by the encounter of the four fields described above, and the desire to give a synthetic overview. It contributed a lot to advertise for it. The second book, *Optimal transport, old and new* [53], was triggered by the start of the synthetic theory of Ricci bounds; this is one of the reasons why it focused much more on generality and fundamentals of the theory, removing all spurious smoothness assumptions whenever possible. At the same time, it was an attempt (the last, certainly) of an exhaustive review of a field which was growing explosively. But it

was also a research project on its own, since a large number of results were rewritten, proved or improved specifically for this book. Let me mention that I am currently slowly working on a revision of this monograph, time seeming ripe because, seven years after its publication, most of the conjectures which were put forward therein seem to have been settled.

#### 4. WEAK CURVATURE-DIMENSION CONDITIONS

In 2004, I encountered Lott in Berkeley, and he suggested that we start working on a synthetic theory of Ricci curvature bounds. At the same time, Sturm was doing the same in Bonn, independently. The resulting papers [34, 47] founded the theory, and thus the geometric spaces defined there are often called LSV for Lott–Sturm–Villani. The precise assumptions vary from paper to paper; some of these variations are superficial, while other ones may be deeper. In this section I shall present some of these definitions, focusing for simplicity on compact spaces, even though the whole theory works at least for locally compact, complete geodesic metric spaces under appropriate moment conditions (and sometimes even the assumption of local compactness can be removed). As for the presentation of the definitions, I shall proceed in an inductive way, from particular to general in some sense.

**4.1. Preliminary definitions.** Recall that

- the metric derivative  $|\dot{\gamma}|$  of a path  $\gamma$ , valued in a metric space  $(\mathcal{X}, d)$ , is defined by

$$|\dot{\gamma}| = \limsup_{s \downarrow 0} \frac{d(\gamma(t), \gamma(t+s))}{s};$$

- the length of a Lipschitz path  $\gamma : [0, 1] \rightarrow \mathcal{X}$  is defined by

$$L(\gamma) = \int_0^1 |\dot{\gamma}| dt = \sup \left\{ \sum_{i=1}^N d(\gamma(t_i), \gamma(t_{i+1})), \quad 0 = t_1 < t_2 < \dots < t_N = 1 \right\};$$

- a Lipschitz path  $\gamma : [0, 1] \rightarrow \mathcal{X}$  is said to be geodesic if  $L(\gamma) = d(\gamma(0), \gamma(1))$ ;
- a metric space  $(\mathcal{X}, d)$  is said to be geodesic if any two points in  $\mathcal{X}$  are joined by at least one geodesic;
- the length is invariant under time reparameterization, and by default a geodesic can be assumed to have constant speed;
- a metric measured space (mms) is a metric space equipped with a Borel measure;
- a nonbranching metric space is one in which any two geodesics which coincide on a nontrivial time interval are in fact equal to each other.

4.2. **Definitions.** If a compact metric space  $(\mathcal{X}, d)$  is given, then the set  $P(\mathcal{X})$  of Borel probability measures on  $\mathcal{X}$ , equipped with the topology of weak convergence, is a compact space. Unless otherwise stated,  $P(\mathcal{X})$  will be further equipped with the distance  $W_2$  derived from the quadratic geodesic transport cost,  $c(x, y) = d(x, y)^2$ . Then the resulting metric space  $P_2(\mathcal{X})$  is itself a compact geodesic space. All this can be adapted to noncompact spaces, with just a bit of care.

**Definition 1.** A compact geodesic metric measured space  $(\mathcal{X}, d, \nu)$  is said to be weakly  $\text{CD}(0, \infty)$  if any two probability measures  $\mu_0, \mu_1$  on  $\mathcal{X}$  can be joined by a constant-speed geodesic  $(\mu_t)_{0 \leq t \leq 1}$ , such that

$$\forall t \in [0, 1] \quad H_\nu(\mu_t) \leq (1 - t) H_\nu(\mu_0) + t H_\nu(\mu_1). \quad (35)$$

This is the simplest definition. It does not matter if the probability measures appearing in it are imposed to be absolutely continuous, or if they are allowed to be singular as well. More importantly, Definition 1 may be reinforced in at least three ways:

- by replacing the requirement on  $H_\nu$  by a requirement on a more general class of functionals, namely all functionals  $U_\nu$  with  $U \in \mathcal{DC}_\infty$ ;
- by requiring inequality (35) to hold along *all* geodesics  $(\mu_t)$ , rather than just one (this is sometimes called “strong displacement convexity property”);
- by requiring the differential convexity inequality  $(d^2/dt^2)H_\nu(\mu_t) \geq 0$  rather than the integral version (35).

The first extension does not seem to have led to a significantly different theory, so it is to some extent a matter of taste. The other two extensions should be handled with more care, in particular because the stability of the resulting definitions is not granted. However, if the space  $(\mathcal{X}, d)$  is nonbranching, it can be shown that all these definitions are equivalent: the nonbranching assumption allows to go back and forth between the local and global versions of these properties, reducing the integral inequalities to inequalities on geodesics.

Whichever choice is made, when applied to a compact Riemannian manifold equipped with its geodesic distance and with a smooth measure  $\nu = e^{-V(x)} \text{vol}(dx)$ , Definition 1 is equivalent to the requirement that  $(M, \nu)$  satisfies the  $\text{CD}(0, \infty)$  criterion. What makes this equivalence possible, following our remarks in Subsection 2.4.3, is

- the possibility of “integrating” the concavity inequality (8);
- the fact that singularities due to nonsmoothness of optimal transport always “modify inequalities in the right direction”;
- the possibility to “test” all directions of displacement via optimal transport.

So one can see a parallel between distribution theory and Definition 1: in both situations, it is about testing a certain property through a certain arbitrary spreading. In distribution theory this is done by integration against an arbitrary test function; while in Definition 1 this is done by considering arbitrary measures  $\mu_0$  and  $\mu_1$ .

Now, how to generalize this to other values of  $K$  and  $N$ ? Let us restrict to values of  $N \geq 1$  and forget that  $N = 1$  has certain peculiarities. Handling  $\text{CD}(0, N)$  rather than  $\text{CD}(0, \infty)$  is rather straightforward: it is all about replacing  $H_\nu$  by its “ $N$ -dimensional” counterpart

$$H_{N,\nu}(\mu) = N \int \rho(1 - \rho^{-1/N}) d\nu = \int U_N(\rho) d\nu, \quad (36)$$

where  $U_N(r) = Nr(1 - r^{-1/N})$ ; and replacing, correlatively, the displacement convexity class  $\mathcal{DC}_\infty$  by  $\mathcal{DC}_N$ .

**Definition 2.** A compact geodesic metric measured space  $(\mathcal{X}, d, \nu)$  is said to be weakly  $\text{CD}(0, N)$ ,  $N < \infty$ , if any two probability measures  $\mu_0, \mu_1$  on  $\mathcal{X}$  can be joined by a constant-speed geodesic  $(\mu_t)_{0 \leq t \leq 1}$  such that

$$\forall t \in [0, 1] \quad H_{N,\nu}(\mu_t) \leq (1 - t) H_{N,\nu}(\mu_0) + t H_{N,\nu}(\mu_1). \quad (37)$$

Again, the various subtleties and equivalences which were discussed above remain true in this case.

So far there is widespread agreement among experts that the definitions are about right. The story is more convoluted for nonzero values of  $K$ . In the synthetic theory of sectional curvature, nonzero curvature bounds are handled through comparison with spheres or hyperbolic planes, and specific coefficients (of sine or hyperbolic nature) appear. When it comes to the synthetic theory of Ricci curvature, among a number of attempts, two main approaches have emerged, due to Sturm and Erbar–Kuwada–Sturm respectively. To present them, I shall focus on the case of positive curvature, which means positive lower bound; but it can all be adapted to negative lower bound as well.

(i) A first recipe is to introduce “distorted” information functionals, of the style

$$U_{\pi,\nu}^\beta(\mu) = \int_{\mathcal{X}} U \left( \frac{\rho(x)}{\beta(x,y)} \right) \beta(x,y) \pi(dy|x) \nu(dx), \quad (38)$$

where  $\pi(dy|x)$  is the disintegration of  $\pi$  with respect to its  $x$ -marginal. Moreover,  $\beta(x,y) > 0$  is a distortion coefficient which can be related to volume distortion along geodesics and can be compared to the distortion in model spaces like  $N$ -dimensional spheres or  $N$ -dimensional hyperbolic spaces. Notice the dependence on the coupling

$\pi$  and the asymmetric role played by  $x$  and  $y$ . Accordingly, in the definitions, one will distinguish between the coupling  $\pi$  and the “reversed” coupling  $\tilde{\pi}$ , in which the variables  $x$  and  $y$  have been swapped. ( $\tilde{\pi} = S_{\#}\pi$ ,  $S(x, y) = (y, x)$ .)

There are two reasonable choices for the comparison distortion coefficients. One, obviously, is to use the real distortion coefficients of the model spaces: they can be defined as the ratio of the actual volume distortion to the corresponding Euclidean volume distortion, both being measured at time  $t$  in an interpolation going from  $x$  to  $y$ . A computation yields the values

$$\beta_t^{(K,N)}(x, y) = \left( \frac{\sin(t\alpha)}{t \sin \alpha} \right)^{N-1}, \quad \alpha = \sqrt{\frac{K}{N-1}} d(x, y). \quad (39)$$

In this formula, the repeated occurrence of  $N - 1$  is a consequence of the important geometrical fact that *curvature is not felt in the direction of geodesic motion*, but only in the  $N - 1$  transversal directions. Important as it is, this is a subtle phenomenon, which is not obvious to capture in computations involving Jacobian determinants or Bochner formula, and requires to explicitly separate the direction of motion from the rest. Thus it also makes sense to consider artificial distortion coefficients in which all  $N$  directions would be treated the same. These are

$$\beta_t^{*(K,N)}(x, y) = \left( \frac{\sin(t\alpha)}{t \sin \alpha} \right)^N, \quad \alpha = \sqrt{\frac{K}{N}} d(x, y). \quad (40)$$

Then we may go from  $\text{CD}(0, N)$  to  $\text{CD}(K, N)$  by adapting (37) and distort the functional  $H_{N,\nu}$  with either  $\beta$  or  $\beta^*$ . (I shall write  $H_{N,\pi,\nu}^\beta$  as a shorthand of  $(U_N)_{\pi,\nu}^\beta$ .) If  $\mathcal{X}$  is a smooth manifold, then *these two choices are equivalent*, which reveals a self-improvement property of distortion coefficients. In fact, at least in the nonbranching case, both resulting inequalities are locally equivalent.

As a final comment, please note that special care has to be given to the case  $N = 1$ , which I skip in this overview; and that it all makes sense also for noninteger values of  $N$ , even if this means abandoning the  $N$ -dimensional spheres and the interpretation in terms of  $N$  “directions”. (To some extent, it is possible to replace spheres by their one-dimensional projections on the real line, and then this gives a family of measure, with a “dimension” parameter  $N$  that can easily be varied continuously.)

(ii) Another recipe to handle nonzero values of  $K$  is to replace the underlying convexity differential inequality  $\ddot{h} \geq 0$  by a more complicated inequality of the form

$$\ddot{h} + \frac{\dot{h}^2}{m} \geq K\sigma, \quad t \in [0, 1]. \quad (41)$$

While inequality (41) is nonlinear, it still satisfies a maximum principle, and thus can be used in comparison principles. Let us write

$$\Psi_t^{K,m}(\psi_0, \psi_1, \sigma) = \text{the solution of (41) at time } t, \quad (42)$$

taking values  $\psi_0$  and  $\psi_1$  at time 0 and 1 respectively.

Then it turns that we can go from  $\text{CD}(0, \infty)$  to  $\text{CD}(K, N)$  by replacing the linear interpolation on the right-hand side of 35 by the nonlinear interpolation defined by (42). Please note that this uses only the *entropy* functional  $H_\nu$ , not the finite-dimensional counterparts  $H_{N,\nu}$ .

All in all, we end up with three choices for the synthetic  $\text{CD}(K, N)$ ; that is, three different notions of weak  $\text{CD}(K, N)$  spaces, which will correspond to slightly different sets of notation.

**Definition 3.** *Let  $(\mathcal{X}, d, \nu)$  be a compact measured geodesic space. It is said to satisfy  $\text{CD}(K, N)$ , or  $\text{CD}^*(K, N)$ , or  $\text{CD}^e(K, N)$ , with  $N \in [1, \infty)$ , if, for any  $\mu_0, \mu_1 \in P(\mathcal{X})$  one can find a constant speed geodesic  $(\mu_t)_{0 \leq t \leq 1}$  in  $(P(\mathcal{X}), W_2)$  such that the following inequality is satisfied:*

- for  $\text{CD}(K, N)$ :  $H_{N,\nu}(\mu_t) \leq (1-t) H_{N,\tilde{\pi},\nu}^{\beta_{1-t}^{(K,N)}}(\mu_0) + t H_{N,\pi,\nu}^{\beta_t^{(K,N)}}(\mu_1)$ ;
- for  $\text{CD}^*(K, N)$ :  $H_{N,\nu}(\mu_t) \leq (1-t) H_{N,\tilde{\pi},\nu}^{\beta_{1-t}^{*(K,N)}}(\mu_0) + t H_{N,\pi,\nu}^{\beta_t^{*(K,N)}}(\mu_1)$ ;
- for  $\text{CD}^e(K, N)$ :  $H_\nu(\mu_t) \leq \Psi_t^{K,N}(H_\nu(\mu_0), H_\nu(\mu_1), W_2(\mu_0, \mu_1)^2)$ .

Again, if  $(\mathcal{X}, d)$  is nonbranching, then this inequality may be replaced by the differential version

$$\frac{d^2}{dt^2} H_\nu(\mu_t) + \frac{\left(\frac{d}{dt} H_\nu(\mu_t)\right)^2}{N} \geq K W_2(\mu_0, \mu_1)^2.$$

Finally, these definitions can be adapted to the case  $N = \infty$ , and the inequalities above have to be replaced, respectively,

- for  $\text{CD}(K, \infty)$  or  $\text{CD}^*(K, \infty)$ , by

$$H_\nu(\mu_t) \leq (1-t) H_{\tilde{\pi},\nu}^{\beta_{1-t}^{(K,\infty)}}(\mu_0) + t H_{\pi,\nu}^{\beta_t^{(K,\infty)}}(\mu_1), \quad (43)$$

where

$$\beta_t^{(K,\infty)}(x, y) = e^{\frac{1}{6}K(1-t^2)d(x,y)^2};$$

- for  $\text{CD}^e(K, \infty)$ , by

$$H_\nu(\mu_t) \leq (1-t) H_\nu(\mu_0) + t H_\nu(\mu_1) - \frac{K t(1-t)}{2} W_2(\mu_0, \mu_1)^2. \quad (44)$$

The latter inequality, expressing the  $K$ -convexity of  $H_\nu$  in the metric  $W_2$ , had long been identified as a plausible  $\text{CD}(K, \infty)$  condition [34].

So we have a choice of defining inequalities when  $K \neq 0$ . How can we compare the resulting definitions? Assuming that  $(\mathcal{X}, d)$  is nonbranching, one can show that

- $\text{CD}(K, N)$  is the most demanding: it implies both  $\text{CD}^*(K, N)$  and  $\text{CD}^e(K, N)$ . It is also the one which naturally leads to sharp dimensional constants.
- $\text{CD}^*(K, N)$  implies  $\text{CD}(K^*, N)$  with  $K^* = K(1 - 1/N)$ . This deterioration of the constant by a factor  $1 - 1/N$  is typical.
- $\text{CD}^*(K, N)$  is equivalent to  $\text{CD}^e(K, N)$ ; and  $\text{CD}^*(K, N)$  is locally equivalent to  $\text{CD}(K, N)$ .
- $\text{CD}^*(K, N)$  can be globalized: that is, if it is true locally (in the neighborhood of any point) then it also holds true globally.
- It is equivalent to ask that  $\mathcal{X}$  satisfies  $\text{CD}^*(N - 1, N)$ , or that the metric cone built over  $\mathcal{X}$  satisfies  $\text{CD}(0, N + 1)$ . (As the archetypal example, think of the flat  $\mathbb{R}^{N+1}$  as the cone built over  $\mathbb{S}^N$ .)
- $\text{CD}^*(K, N)$  also leads to sharp dimensional inequalities such as the sharp version of Brunn–Minkowski’s inequality; but this is based on more tricky lines of reasonings such as localization techniques.

All this suggests that the “correct” definition is not  $\text{CD}$ , but either  $\text{CD}^*$  (more complicated) or  $\text{CD}^e$  (more simple), both being equivalent in a nonbranching context. It remains an open problem whether the weak  $\text{CD}(K, N)$  property is in fact equivalent to the weak  $\text{CD}^*(K, N)$  property, or whether it is a strictly stronger notion (in which case it should probably be dismissed because of the failure of globalization property).

**4.3. Stability.** So much effort has been spent on definitions so far – but finding the right definition was one of the main goals. I already mentioned that all those definitions do extend the classical  $\text{CD}(K, N)$  criterion. But what else can be said of them?

The first major property is that the various notions of weak  $\text{CD}(K, N)$  spaces are all stable under a very weak but very natural notion of convergence of metric-measure spaces, namely the measured Gromov–Hausdorff topology. It means, roughly speaking, pointwise convergence of distances and weak convergence of measures. Let us note that this stability property was a well-known open problem even in the case of smooth manifolds.

Stability is proven easily by taking advantage of two main properties:

- Optimal transport (say with cost  $c = d^2$ ) is stable under Gromov–Hausdorff convergence; for instance, if  $\mathcal{X}^k \rightarrow \mathcal{X}$  in the Gromov–Hausdorff topology, and  $(\mu_t^k)$  is a geodesic path in  $P_2(\mathcal{X}_k)$ , for each  $k$ , then up to extraction of a subsequence it converges weakly to a geodesic path  $(\mu_t)$  in  $P_2(\mathcal{X})$ . In fact, the convergence of  $\mathcal{X}_k$  to  $\mathcal{X}$  implies the convergence of  $P_2(\mathcal{X}_k)$  to  $P_2(\mathcal{X})$ .

- Nonlinear functionals of information type are also stable, and more precisely lower semicontinuous, under weak convergence of measures, for *both* the integrand and the reference measure. In fact, if  $U$  is a convex nonlinearity, and  $(\mu^k)_{k \in \mathbb{N}}, (\nu^k)_{k \in \mathbb{N}}$  are two sequences of probability measures converging weakly to  $\mu$  and  $\nu$  respectively, then

$$U_\nu(\mu) \leq \liminf U_{\nu^k}(\mu^k).$$

Furthermore, these functionals are also stable under push-forward: if  $f$  is any measurable map, then

$$U_{f\#\nu}(f\#\mu) \leq U_\nu(\mu).$$

At least when  $U(r) = r \log r$ , the latter inequality has clear informational content (applying a function to a signal can only reduce the amount of information); under Sanov’s interpretation (of the entropy as a large deviation rate), it coincides with the “contraction property” which is dear to specialists of large deviations.

**4.4. Generality.** Let me write here generically CD for either CD, or  $\text{CD}^*$ , or  $\text{CD}^e$ . Weak  $\text{CD}(K, N)$  spaces are strictly more general than classical  $\text{CD}(K, N)$  spaces; in particular, they may contain singular limits of  $\text{CD}(K, N)$  spaces, presenting cone-type singularities for instance. Such singularities were already well-known in the context of the synthetic theory of sectional curvature.

But, at a more fundamental level, weak  $\text{CD}(K, N)$  spaces also encompass geometries which never occur with sectional curvature bounds. Indeed, non-Euclidean normed spaces all belong to this class; more precisely,  $\mathbb{R}^n$ , equipped with any norm and with the  $n$ -dimensional Lebesgue measure, defines a weak  $\text{CD}(0, n)$  space. This may look innocent but is a major novelty with respect to the sectional approach: indeed, a normed space satisfies sectional curvature inequalities (one-sided or two-sided) *only* if it is a Euclidean space. More generally, weak  $\text{CD}(K, N)$  spaces can handle certain classes of Finsler geometries.

This wealth of spaces can be seen as an advantage or as a drawback by geometers. On the one hand, normed spaces have the same isoperimetric inequalities as Euclidean spaces (with just a bit more generality, this is sometimes called the theory of Wulff isoperimetry); so it is nice to encompass them in a common setting. On the other hand, there are some key properties from Riemannian geometry which



definitely do not hold in normed spaces. A most emblematic one is the *splitting property*, which can be stated only for unbounded geometries. Its classical formulation goes like this: if a measured Riemannian manifold satisfies  $\text{CD}(0, N)$  and contains an infinite line  $\gamma$ , then it can be factored up into the metric product of  $\gamma$  with a submanifold of codimension 1, up to isometry of course. While this property may look rather specific, it plays a crucial role in establishing rectifiability of certain metric spaces, by splitting off successive  $\mathbb{R}$  factors from a tangent space obtained by metric blow-up.

The tension between these two points of view was resolved by the introduction of an extra condition reinforcing the curvature-dimension conditions. This condition, introduced by Ambrosio, Gigli and Savaré [5], is denoted with the letter  $R$  (for Riemannian); so it gives rise to weak  $\text{RCD}$ ,  $\text{RCD}^*$  or  $\text{RCD}^e$  spaces. It consists in imposing, on top of the considered curvature condition, the requirement that

$$W^{1,2}(\mathcal{X}, d) \quad \text{is a Hilbert space.} \quad (45)$$

This extra condition prevents non-Euclidean normed spaces, and, remarkably, narrows the class of  $\text{CD}(K, N)$  spaces down to what seems to be the right level of generality for all “genuinely Riemannian” applications. A few remarks are in order.

- The space  $W^{1,2}$  is the Sobolev space of functions with square-integrable gradient. It is classically defined in a smooth context by

$$\|f\|_{W^{1,2}}^2 = \|f\|_{L^2}^2 + \|\nabla f\|_{L^2}^2;$$

but here one has to be careful in the generalized definition of  $\|\nabla f\|_{L^2}^2$ . It should be understood in the sense devised by Cheeger after a careful study; the construction goes like this:

- an upper gradient for  $f$  is a measurable function  $g$  such that for any Lipschitz curve  $\gamma$ ,

$$f(\gamma(1)) - f(\gamma(0)) \leq \int_0^1 g(\gamma(s)) |\dot{\gamma}_s| ds.$$

- for a Lipschitz function  $f$ , a distinguished upper gradient is the metric gradient

$$|Df| = \limsup_{y \rightarrow x} \frac{|f(y) - f(x)|}{d(x, y)}.$$

- for an  $L^2$  function  $f$ , an  $L^2$ -relaxed gradient is an  $L^2$  function which is almost everywhere bounded below by the weak limit of  $|Df_n|$ , where  $f_n$  is some sequence converging in the weak  $L^2$  sense to  $f$ .

- the Cheeger  $W^{1,2}$  seminorm,  $\|Df\|_*$ , is the  $L^2$  norm of the minimal relaxed gradient (here minimal means in the  $L^2$  sense).
- the Cheeger functional is

$$\text{Ch}_*(f) = \frac{1}{2} \|Df\|_*^2,$$

and it allows to define an  $L^2$ -Sobolev space, whose analysis has been the subject of an important body of work.

- Requiring the Hilbertian nature of  $W^{1,2}$  is obviously a weak way to impose the Euclidean nature of the tangent spaces, so it really has a Riemannian flavor. This condition is spread all throughout  $\mathcal{X}$ , so that it does allow for nonsmooth spaces in which not every tangent cone is a Euclidean space.

- Ambrosio, Gigli and Savaré showed that this “Riemannian” condition can also be expressed by the *linearity of the heat equation*. Let us come back to this later.

- This geometric condition on  $W^{1,2}$  is in spirit a first-order differential condition; so there is no reason to expect it to be stable in the measured Gromov–Hausdorff topology. However, *when it is combined with a  $\text{CD}(K, N)$  condition*, it turns out to be stable. The analogy is that a second-order bound makes it possible to pass to the limit in the first-order conditions. (Similarly, if a sequence of *convex* functions converges pointwise, one may pass to the limit in the subdifferential.)

- The RCD spaces seem to answer all criticisms that one may raise against the large degree of generality of CD spaces; in particular, they do not include normed spaces besides Euclidean space, and, as a major result by Gigli [24], they do enjoy a splitting property.

**4.5. Heat equation.** This story started with a reinterpretation of the heat equation in terms of optimal transport and information. It also came back into the realm of the heat equation when this reinterpretation became a way to construct nonsmooth heat flow with astounding generality. This was apparent in the result of Ambrosio, Gigli and Savaré [4] showing that in a geodesic space,

- the JKO scheme always converges to a gradient flow solution of the Cheeger functional;

- under a weak  $\text{CD}(K, \infty)$  bound, there is a unique, well-defined heat flow, which can be constructed either as the  $L^2$  gradient flow of the Cheeger functional or, equivalently, as the  $W_2$  gradient flow of the Boltzmann information  $H_\nu$ .

These results hold under the even weaker abstract conditions that  $|\nabla^- H_\nu|$  (which under appropriate assumptions is the square root of the Fisher information) is lower

semicontinuous. But expressing them in terms of lower Ricci curvature bounds, which even in the smooth case are recognized as the key estimate for a well-defined heat flow, is extremely satisfactory. At the same time, the authors study the relations with the alternative evolution variational formulation of the heat equation.

The Ambrosio–Gigli–Savaré theorem is the final accomplishment in a series of generalizations of the Jordan–Kinderlehrer–Otto results. A key technical lemma in this field is due to Kuwada [29]: it allows to estimate the speed of variation, in  $W_2$  sense, of any gradient flow of the Cheeger functional; its core can be seen as an estimate on a nonsmooth Hamilton–Jacobi equation.

It should also be noted that the heat flow is in general nonlinear. This is in the order of things, even for smooth non-Riemannian geometries. Actually, the linearity of the heat flow is equivalent to the property that  $W^{1,2}$  be a Hilbert space.

**4.6. Functional inequalities.** Ever since McCann’s proof of the Brunn–Minkowski inequality, the idea to use optimal transport to study or even establish functional inequalities has gone a long way. Some of the main ideas are to replace volumes by integral functionals of probability measures, and surfaces by their infinitesimal derivatives under some evolution. A precursor of this approach was Gromov’s proof of the classical isoperimetric inequality by using a transport map (in that case, not the optimal transport one).

For instance, to prove the curved Brunn–Minkowski inequality in a measured geodesic space  $(\mathcal{X}, d, \nu)$ , pick up arbitrary compact sets  $A_0$  and  $A_1$  with positive measures, define  $\mu_0$  as the restriction of  $\nu$  to  $A_0$ ,  $\mu_1$  as the restriction of  $\nu$  to  $A_1$ , and study the variations of the functionals  $H_{N,\nu}$ : under a weak  $\text{CD}(0, N)$  condition,

$$H_{N,\nu}(\mu_{1/2}) \leq \frac{1}{2} \left( H_{N,\nu}(\mu_0) + H_{N,\nu}(\mu_1) \right).$$

Then the right-hand side is  $-(\nu[A_0]^{1/N} + \nu[A_1]^{1/N})/2$ ; while the left-hand side can be compared, through Jensen’s inequality, to  $-\nu[S_{1/2}]^{1/N}$ , where  $S_{1/2}$  is the support of  $\mu_{1/2}$ . Since this support is in turn included in the set of midpoints of  $A_0$  and  $A_1$ , the desired inequality follows.

This scheme is typical: applying a convexity inequality to well-chosen probability measures and geodesics in  $W_2$  geometry leads to a surprisingly large number of inequalities. Sometimes this goes also through restriction to a subset of geodesics, or rescaling, or local study, or blow-up, etc. Semigroup arguments are also popular in this area, with particular emphasis on the heat flow and Hamilton–Jacobi equations, in a smooth or nonsmooth setting.

One of the first results in the study of nonsmooth  $\text{CD}(K, N)$  spaces was the doubling property: that is, one can bound the measure of a ball  $B(x, 2r)$  by a constant multiple of the measure of the ball  $B(x, r)$ . This is captured by the Bishop–Gromov inequality, stating that the rate of growth of  $\nu[B(x, r)]$ , as a function of  $r$ , is controlled by the rate of growth in the model  $\text{CD}(K, N)$  space. This result is also a key to Gromov’s precompactness theorem, according to which the set of  $\text{CD}(K, N)$  measured manifolds, with an a priori bound on the diameter, is precompact in the measured Gromov–Hausdorff topology.

It was much more tricky to nail down the local Poincaré inequality at a satisfactory level of generality. A metric-measure space  $(\mathcal{X}, d, \nu)$  is said to satisfy a local Poincaré inequality if, for any Lipschitz function  $u$ , any point  $x_0 \in \mathcal{X}$  and any radius  $r > 0$ ,

$$\int_{B_r(x_0)} \left| u(x) - \langle u \rangle_{B_r(x_0)} \right| d\nu(x) \leq Cr \int_{B_{2r}(x_0)} |\nabla u(x)| d\nu(x), \quad (46)$$

where  $\int_B = (\nu[B])^{-1} \int_B$  is the averaged integral over  $B$ , and  $\langle u \rangle_B = \int_B u d\nu$  is the average of the function  $u$  on  $B$ . Several variants of these inequalities exist: one may require that the ball on the right-hand side be the same as the ball on the left-hand side; one may replace the  $L^1$  integrals by  $L^p$  norms, with possibly different values of  $p$  on both sides; etc. However, all these inequalities are related, and they are one of the pillars of analysis in metric spaces, as a way to control global variations by local ones. In the case of  $\text{CD}(K, N)$  spaces, for some time the proof of the local Poincaré inequality was relying on a nonnatural nonbranching assumption, until a surprisingly neat and simple argument was found by Rajala [46] to handle the full generality.

A further coup de théâtre occurred in the field when Cavalletti and Mondino [10], introduced the technique of *localization* in this nonsmooth setting. Localization, which grew up from works by Payne–Weinberger, Gromov–Milman and Kannan–Lovász–Simonovits, goes through the reduction of an  $n$ -dimensional inequality to a collection of one-dimensional inequalities holding on geodesic lines. The main idea is captured by the following decomposition: if a certain function  $f$  with zero mean is given, then one can decompose the space into a part in which  $f$  is essentially zero, and a set of geodesics  $\gamma$  such that the integral of  $f$  on each  $\gamma$  is also zero; one can also disintegrate the measure  $\nu$  according to this set of geodesics, and thus write  $\nu = \int \nu_\gamma \bar{\nu}(d\gamma)$ . Klartag discovered that  $L^1$ -optimal transport could be used as a tool to establish this decomposition; and Cavalletti–Mondino showed that this could be adapted to a nonsmooth setting, using only a nonbranching assumption. It is very striking that this bit relies on the study of  $L^1$ -transport (with cost  $c = d$ ), while

virtually all other applications in geometry rely on  $L^2$ -transport (with cost  $c = d^2$ ).

With these tools, an impressive list of functional inequalities could be derived. Most of them fall in one of the following categories:

- volume control: Brunn–Minkowski, doubling (Bishop–Gromov inequality); and the functional counterpart of Brunn–Minkowski, the so-called Prékopa–Leindler inequalities.
- concentration inequalities, like the Talagrand inequalities.
- isoperimetric and spectral inequalities: Lévy–Gromov isoperimetric inequality, Sobolev and logarithmic Sobolev inequalities,  $L^2$ –Poincaré inequalities, local Poincaré inequalities...
- heat kernel estimates: Li–Yau inequalities and their many variants.

A large part of my book [53] was devoted to presented a synthetic view of all these inequalities and their links, both in smooth and nonsmooth contexts.

**4.7. Summary.** Now that more dust has settled, one may argue that the synthetic theory of Ricci bounds can be developed at three levels of generality, which are all interesting:

(i) *General metric spaces.* In this case we have a choice between  $\text{CD}(K, N)$ ,  $\text{CD}^*(K, N)$  or  $\text{CD}^e(K, N)$  criteria. In such generality, under curvature-dimension conditions one still has access, with possibly slight variations in the constants, to

- Poincaré inequalities, local and global;
- Sobolev and logarithmic Sobolev inequalities;
- Bishop–Gromov and Brunn–Minkowski inequalities;
- Gromov’s precompactness theorem, as well as various natural metrics on the space of bounded  $\text{CD}(K, N)$  spaces (like Sturm’s distance);
- a well-defined heat flow, which is at the same time the  $L^2$ -gradient flow of the Cheeger functional, and the  $W_2$ -gradient flow of the Boltzmann information  $H_\nu$ .

(ii) *Nonbranching  $\text{CD}^*(K, N)$  spaces,* or equivalently nonbranching  $\text{CD}^e(K, N)$  spaces. Examples do include Banach spaces and certain Finsler geometries. Then in addition we have

- globalization: if the curvature bound holds true locally, then it also holds true globally, without any degradation of constants;
- sharp dimensional isoperimetric-type inequalities, including sharp versions of the Brunn–Minkowski, Lévy–Gromov and  $L^2$ -Sobolev inequalities;
- almost rigidity theorems comparable to those existing in the smooth context.

While nonbranching might seem sometimes a stringent assumption (not stable under Gromov–Hausdorff topology), these results remain true if the space is *essentially nonbranching*, that is, if any optimal transport is concentrated on a set of nonbranching geodesics.

(iii)  $\text{RCD}^*(K, N)$  spaces. These spaces are automatically essentially nonbranching, so this category is strictly stronger than the previous one. It enjoys

- splitting theorem in nonnegative curvature (Gigli);
- a distributional Laplace operator;
- geometric estimates on the heat kernel (Garofalo–Mondino);
- a Bochner formula (Erbar–Kuwada–Sturm); with this, essentially any relevant which is known to hold in the smooth  $\text{CD}(K, N)$  setting can be extended to  $\text{RCD}^*(K, N)$  spaces;
- information-theoretical inequalities of Shannon type (which are probabilistic analogues of the Brunn–Minkowski inequality, and hold true in nonnegative curvature);
- the possibility to solve the Monge problem, and prove the existence of an optimal transport, under general conditions (Gigli);
- rectifiability (Mondino–Naber), even though at the moment the dimension of the tangent space might vary from place to place. This is, in spirit, reminiscent of the situation for convex functions: a synthetic curvature (second-order) bound automatically implies a certain regularity.

**4.8. Extensions?** I already mentioned that Definitions 1 to 3 can be generalized, without much difficulty, to complete, locally compact, possibly noncompact spaces.

It is more tricky to consider spaces  $\mathcal{X}$  which are not locally compact, typically infinite-dimensional. Another related extension is in the direction of length spaces which are not necessarily geodesic spaces. There does not seem to be a general theory so far, but some results in those directions are established and considered by Sturm, Ambrosio and their collaborators. As a typical example, according to Sturm the Wiener space can be seen as an infinite-dimensional  $\text{CD}(1, \infty)$  space, just as one would expect or hope.

## 5. CONNECTION WITH OTHER THEORIES

It was clear from the start that the theory of synthetic Ricci curvature bounds is in relation with an array of mathematical theories going way beyond geometry. But even within geometry, it is worth pointing out a few theories which are strongly related to our subject here.

**5.1. Alexandrov spaces.** One of the first questions which comes to mind is the relation between the synthetic theory of Ricci curvature bounds and the synthetic theory of sectional curvature bounds. The latter is defined using comparison of angles and distances between triangles in the space of interest and in a reference model space of constant curvature, as in Subsection 2.1.2. The resulting spaces are called *Cartan–Alexandrov–Toponogov spaces*, or *CAT spaces*, or *Alexandrov spaces*. There is a theory for upper bounds, and a theory for lower bounds; the corresponding spaces are sometimes denoted  $\text{CAT}^-(\kappa)$  (sectional curvature bounded above by  $\kappa$ ) or  $\text{CAT}^+(\kappa)$  spaces (sectional curvature bounded below by  $\kappa$ ), and their properties are very different; there is also a theory of two-sided curvature bounds, which is somehow less interesting.

Trees are typical examples of  $\text{CAT}^-$  spaces, which can be very rough; but  $\text{CAT}^+$  spaces can never be too wild, at least when they have finite dimension, in which case the dimension  $n$  is defined without ambiguity. Decades of works, especially by the Russian and Japanese schools, have shown how to develop analysis on  $\text{CAT}^+$  spaces, including local coordinate systems, Laplace operators, gradient flows, etc.

Since classical Ricci curvature is deduced from sectional curvature, sectional curvature bounds are stronger than Ricci curvature bounds. In particular, if an  $n$ -dimensional Riemannian manifold has sectional curvatures bounded below by  $\kappa$ , then it satisfies the classical  $\text{CD}((n-1)\kappa, n)$  criterion. Thus, a satisfactory synthetic theory of Ricci curvature should keep this implication at the level of nonsmooth sectional curvature bounds: any  $n$ -dimensional  $\text{CAT}^+(\kappa)$  space should also satisfy a weak  $\text{CD}((n-1)\kappa, n)$  inequality. Note that this link is far from obvious, given the different formalisms in which CAT spaces (defined with lengths and angles) and weak  $\text{CD}(K, N)$  spaces (defined with probability measures) are expressed.

The problem stood wide open until it was solved by Petrunin [45], who gave full details in the particular case  $\kappa = 0$ . Since then these links have been confirmed: Alexandrov spaces are automatically  $\text{RCD}^*$  spaces.

**5.2. Finsler geometries.** Finsler geometries are those in which each tangent space is equipped with a norm that may be non-Euclidean. The resulting diversity of geometries is very rich, but also confusing – for instance, there are a number of Finslerian notions of curvature, in contrast with the Riemannian setting where there is a universally accepted and uniquely defined notion of curvature tensor.

While the notion of synthetic sectional curvature bound is useless in Finsler geometries, the notion of synthetic Ricci curvature on the other hand can be used in this field, with interesting results. Under certain assumptions, it coincides with the

so-called *flag curvature*. The study of this connection was pioneered by Ohta and further developed by Sturm and others [40].

**5.3. Diffusion equations.** Curvature-dimension bounds do not only result in convexity-type properties of information functionals along geodesics of optimal transport: they also imply contraction properties along the solutions of the heat equation, or more generally diffusion equations. The latter equations may be of linear or nonlinear type, and involve classical heat equation as well as porous medium equations or fast diffusion equations.

From a geometric perspective, the difference between these two approaches corresponds to emphasizing either geodesics or gradient flows, which are the two most important categories of differential equations. From the point of view of partial differential equations, it corresponds to emphasizing Hamilton–Jacobi equations or diffusion equations, which are two fundamental classes of evolutions.

One of the most neat consequences of curvature-dimension bounds, combining all the fields mentioned in Section 2, is the following contraction estimate: under the condition  $\text{CD}(K, \infty)$ , if any two probability-valued solutions of the heat equation are given, say  $(\mu_t)_{t \geq 0}$  and  $(\tilde{\mu}_t)_{t \geq 0}$ , then

$$W_2(\mu_t, \tilde{\mu}_t) \leq e^{Kt} W_2(\mu_0, \tilde{\mu}_0).$$

In particular, *the heat flow is nonexpanding in nonnegative curvature*.

There are many variants of this estimate, some of which involve dimensionality (replacing the heat equation by the fast diffusion equation  $\partial_t \rho = \Delta \rho^{1-1/n}$ , for instance). It is also possible to develop a synthetic theory of Ricci curvature based on them; but this approach is somehow more convoluted, and seems to lead to weaker results than the approach exposed in Section 4. Furthermore, it only applies to RCD-type spaces, since these contractivity estimates may fail for Finsler geometries for instance. Nevertheless, the interplay with optimal transport here has led to a better understanding of the properties of diffusion flows.

Another related nonlinear diffusion equation which has been used in a geometric context, with spectacular results, is the Ricci flow, say  $\partial_t g = -2 \text{Ric}$ , where  $g$  is the Riemannian metric. As pointed out by McCann and Topping [38], Ricci flow can be understood in a synthetic way as the “most economical” evolution equation which guarantees the nonexpansivity of the heat flow when the metric is allowed to vary. This was the starting point of a series of contributions reinterpreting Ricci flow estimates in terms of optimal transport, and put forward mainly by Lott [33] and Topping [50].



**5.4. Discrete spaces.** Discrete mathematics has recently become more fashionable and richer with the rise of computer science. This can be seen in the development of discrete probability, discrete stochastic processes, discrete analysis, discrete geometry... In this context it is natural to ask whether curvature, playing such a crucial role in continuous geometry, also has a field of application in a discrete setting.

A priori, curvature is a truly continuous notion; of course, in an analytic approach, it is always possible to discretize the usual differential formulas for curvature, but this leads to cumbersome expressions whose use is not very clear. On the other hand, it is not very difficult to discretize the synthetic formulation of Ricci curvature bounds, as expressed in Section 4. One may achieve this discretization in a number of slightly different ways, either through the behavior of information functionals along approximate geodesic measure-valued paths, or through the contraction properties of diffusion processes. Among a number of works, one may consult [41], [42], [18] for this.

As a striking example of result in this direction, one can now give a precise meaning to the statement that the hypercube  $\{-1, 1\}^N$ , equipped with the uniform probability measure, has (discrete) Ricci curvature equal to  $K = 1/(2N)$ . This provides a precise answer to a seemingly absurd question asked by Stroock in a 1998 seminar in Institut Henri Poincaré: *What is the Ricci curvature of the discrete hypercube?*

Of course, deriving curvature estimates in this context is not a final goal; the point is that these bounds may be useful to derive other consequences, expressed in terms of geometry or probability theory. As an example, together with Ollivier we established the following Brunn–Minkowski type inequality: if  $A$  and  $B$  are two nonempty subsets of  $\{0, 1\}^N$ , and  $M$  is the set of midpoints of  $A$  and  $B$ , then

$$\log |M| \geq \frac{1}{2}(\log |A| + \log |B|) + \frac{K}{8}d(A, B)^2.$$

**5.5. MTW curvature.** In the first place, one of the main reasons why nonsmooth analysis made its way in optimal transport theory is the hard reality that optimal transport is sometimes not smooth: even if the measures  $\mu_0$  and  $\mu_1$  enjoy all the regularity that one could dream of (say  $C^\infty$  densities with upper and lower bounds), it may be that optimal transport is plainly discontinuous.

On the other hand, over the past decade, nearly optimal geometric conditions were discovered, which guarantee the continuity of optimal transport under just upper and lower bounds on the densities, and the smoothness of optimal transport under further smoothness estimates on these densities. These conditions were discovered by Ma, Trudinger and Wang [35] and later studied by a number of authors, including a key contribution by Loeper [31]. So far they have been expressed only in a smooth

Riemannian context, and take the form of the nonnegativity of a nonlocal (fourth-order) curvature tensor, called the MTW curvature. This tensor may also be seen as the collection of sectional curvatures of a specific metric, which is obtained as the mixed second-order derivative of the squared distance function.

The resulting curvature condition is tricky and very rigid, in some sense: for instance, a nearly round ellipsoid will satisfy it, while an elongated ellipsoid will not satisfy it, and optimal transport on it is sometimes discontinuous. However, this is the second main direction in which optimal transport has contributed to an unexpected development in the notion of curvature. Indeed, besides the regularity of optimal transport, MTW curvature has proven useful in certain developments of independent interest:

- solving certain problems in theoretical economics;
- providing natural examples and counterexamples for regularity in the theory of fully nonlinear partial differential equations of Monge–Ampère type;
- yielding new geometric estimates on the shape of the important but elusive cut locus in Riemannian geometry.

As a striking example about the last topic, the MTW tensor was one of the key ingredients which allowed Figalli, Rifford and me [19] to prove the following “nearly round sphere” theorem: if the sphere  $\mathbb{S}^n$  is equipped with a Riemannian metric that is nearly round in the  $C^4$  topology, then all injectivity domains of the resulting geometry are strictly convex.

Here, the injectivity domain of a point  $x$  is the maximal domain of the exponential map at  $x$ , which is also the open region in the tangent space made of all geodesics emanating from  $x$  and evaluated before cut time. So our theorem shows that a nearly round sphere, viewed from any point, is strictly convex. The deceiving simplicity of the statement hides a tricky and indirect proof, and one of the first notable stability results proven over the past decades for the notoriously unstable cut locus. It is also, to my knowledge, the first such statement in which positive curvature has a good effect – negative curvature is good for the study of cut locus, because it rules out focalization which is the main cause of headache in the business; but positive curvature leaves plenty of room for focalization.

Let me finally mention that there is, to some extent, a synthetic version of the MTW condition [52]; it can be understood as a sort of nonlocal reinforcement of the  $\text{CAT}^+(0)$  condition.

## 6. CONCLUSIONS, AND WHAT NEXT?

A decade after the birth of the synthetic theory of Ricci bounds, the theory now seems to be ripe. Some of its main results are

- stability results for Ricci curvature bounds under weak notions of convergence (as noted before, this problem was notoriously open, even for smooth geometries);
- new tools to study limits of Riemannian manifolds, with a view to suggest alternative approaches to the elaborate tools developed by Cheeger and Colding [13] on the subject;
- the construction of the heat flow in what is arguably the most general setting that one could dream of;
- new, robust proofs of certain key theorems in geometry, like the local Poincaré inequality, or maybe most strikingly (and most dear to me), the Lévy–Gromov isoperimetric inequality, whose proof used to rely on crazy amounts of regularity;
- important side progress in metric analysis, including a better understanding of Sobolev spaces, Hamilton–Jacobi equations, and gradient flows;
- new approaches to curvature in non-Riemannian contexts, like discrete spaces or Finsler geometries.

For sure, the theory is not complete yet, and some problems do remain open. Here is a short nonexhaustive list:

- Are  $\text{CD}(K, N)$  and  $\text{CD}^*(K, N)$  definitions from Subsection 4 equivalent, at least in a nonbranching setting?
- Are  $\text{RCD}^*(K, N)$  spaces automatically rectifiable with a well-defined dimension?
- Can one push the theory of weak  $\text{CD}(K, N)$  spaces to the point where it will encompass all currently known results about weak limits of  $\text{CD}(K, N)$  manifolds?
- What about a list of significant examples and counterexamples for branching spaces?
- Are there interesting weak  $\text{CD}(K, N)$  spaces which cannot be obtained as limits of Riemannian manifolds?

While these problems all look difficult and interesting, they also look like refinements in comparison of the fundamental difficulties which were solved during the past ten years.

Besides, the introduction of optimal transport and information theory in geometry has led to new points of view, which have already been helpful to study directions of research and establish conclusions. This added heuristic value was already present from the beginning in Otto’s study of nonlinear diffusion equations [43]: the gradient flow interpretation was naturally tuiding him to certain estimates, which would

have been difficult to guess otherwise. It is also striking to note that one of the early applications of optimal transport to Riemannian geometry was the discovery of the curved Brunn–Minkowski inequality, whose shape was not even conjectured before the question naturally arose through optimal transport. It is also an issue of interpretation, namely the understanding of the so-called Ricci O’Neil theorem, which was one of the early motivations of Lott for setting up the program of synthetic Ricci bounds.

Here are a few other examples that I am aware of, in which optimal transport acted as a guide to finding a relevant interpretation:

- the method which I developed with Grünewald, Otto and Westdickenberg to revisit the hydrodynamic limit of particle systems of Ginzburg–Landau type [27];
- new large deviation interpretation of diffusion processes put forward by Peletier and collaborators in a series of works; among other things, their analysis [1] shows that the JKO scheme for the heat equation be interpreted in statistical terms, with Boltzmann’s entropy arising from Sanov’s theorem and the square distance being related to the log central limit theorem;
- the new unconditional estimates of Funano [23] for spectral values of  $-\Delta$  in nonnegative Ricci curvature, showing that successive eigenvalues  $\lambda_k$  satisfy  $\lambda_k/\lambda_1 \leq C^k$  for some universal  $C > 0$  (since then the dependence on  $k$  has been tremendously improved, but the first idea for such a universal result came from optimal transport interpretation);
- the general theorems of stability of diffusion equations by Ambrosio–Savaré–Zambotti[6];
- the already mentioned works [33, 50], reinterpreting some of the Ricci flow theory, including some portions of Perelman’s proof of the Poincaré conjecture, in terms of optimal transport.

This last topic naturally leads to the question of constructing a synthetic notion of Ricci flow. One motivation for this would be to define a Ricci flow on a nonsmooth space such as an Alexandrov space, and thus enrich the toolbox of geometrical analysis in this context. A different motivation would be to deal with the singularities naturally developed by the Ricci flow even in a smooth setting, thus avoiding the surgery procedure which has been used to get rid of these singularities.

Another direction of research which has gained vitality in the past decade is the study of the geometry of the Wasserstein space  $P_2$ : Besides the fact that it is a geodesic space, does it support a truly differentiable structure in some sense? a differential calculus? a parallel transport? a Riemannian calculus? an algebraic

topology? an infinite-dimensional canonical reference measure? a curvature operator? a heat equation? a Laplace operator? a Hamilton–Jacobi equation? a Brownian motion, which would provide a canonical notion of stochastically evolving probability measure? All these issues have been studied, sometimes informally and sometimes rigorously, over the past years. Some of the best contributions in the field have been done by Sturm, Gangbo and their collaborators. To quote just one striking result, Sturm constructed a beautiful “entropic measure” [48] whose definition looks rather fundamental, and at same time quite difficult to use.

A related problem which has been a source of wonder ever since the time of the encounter of optimal transport with gradient flows, is whether one can define *Hamiltonian flows* in optimal transport structure, and what can be gained from that insight. Natural candidates are known: Vlasov equation, 2-dimensional incompressible Euler equation in vorticity formulation, semigeostrophic equations... But issues of existence, well-posedness and asymptotic behavior turn out to be way more tricky, and simultaneously much more demanding in terms of smoothness, for these would-be Hamiltonian flows than for gradient flows. Accordingly, constructions and results are more indirect and partial.

Finally, let me say that I would not take bets about the most dynamic directions of research to come. As I was writing my first book on the subject, *Topics in Optimal Transportation*, I had the vague feeling that the theory was rounding up and approaching a reasonable state of coherence. But this was followed by an explosion of results which took me by surprise. As I was writing my second book, I was more lucid, and realizing that the theory was bound to continue to blow up, but still, the speed at which this occurred left me amazed – only seven years later, if I wanted to write now a comprehensive book with the same amount of detail as *Optimal transport, old and new*, it should be roughly be twice as thick. Thus I don’t think I take much risk by adapting to this context Poincaré’s prediction that *surprising results shall be obtained*.

#### SELECTED BIBLIOGRAPHY

- [1] ADAMS, S., DIRR, N., PELETIER, M.A., AND ZIMMER, J. From a Large-Deviations Principle to the Wasserstein gradient flow: a new micro-macro passage. *Commun. Math. Phys.* 307 (2011), 791–815.
- [2] AMBROSIO, L. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, vol. 1812 of *Lecture Notes in Math.* Springer, Berlin, 2003, pp. 1–52.

- [3] AMBROSIO, L., GIGLI, N., AND SAVARÉ, G. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.
- [4] AMBROSIO, L., GIGLI, N., AND SAVARÉ, G. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.* 195, 2 (2014), 289–391.
- [5] AMBROSIO, L., GIGLI, N., AND SAVARÉ, G. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.* 163 (2014), 1405–1490.
- [6] AMBROSIO, L., SAVARÉ, G., AND ZAMBOTTI, L. Existence and Stability for Fokker–Planck equations with log-concave reference measure. *Probab. Theory Related Fields* 145 (2009), 517–564.
- [7] BURAGO, D., BURAGO, Y., AND IVANOV, S. *A course in metric geometry*, vol. 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. Errata list available online at [www.pdmi.ras.ru/staff/burag](http://www.pdmi.ras.ru/staff/burag)
- [8] BURAGO, Y., GROMOV, M., AND PERELMAN, G. A. D. Aleksandrov spaces with curvatures bounded below. *Uspekhi Mat. Nauk* 47, 2(284) (1992), 3–51, 222.
- [9] CAVALLETTI, F. Monge problem in metric measure spaces with Riemannian curvature-dimension condition. *Nonlinear Anal.* 99 (2014), 493–551.
- [10] CAVALLETTI, F., AND MONDINO, A. Sharp and rigid inequalities in metric-measure spaces with lower Ricci curvature bounds. Preprint, 2015.
- [11] CAVALLETTI, F., AND MONDINO, A. Sharp geometric and functional inequalities in metric measure spaces with lower Ricci curvature bounds. Preprint, 2015.
- [12] CHAMPION, TH., AND DE PASCALE, L. The Monge problem in  $\mathbb{R}^d$ . *Duke Math. J.* 157, 3 (2011), 551–572.
- [13] CHEEGER, J., AND COLDING, T. H. On the structure of spaces with Ricci curvature bounded below. *J. Differential Geom.* 46, 3 (1997), 406–480 (Part I); *J. Differential Geom.* 54, 1 (2000), 13–35 (Part II) and 37–74 (Part III).
- [14] CORDERO-ERAUSQUIN, D., MCCANN, R. J., AND SCHMUCKENSCHLÄGER, M. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* 146, 2 (2001), 219–257.
- [15] COVER, T. M., AND THOMAS, J. A. *Elements of information theory*. John Wiley & Sons Inc., New York, 1991. A Wiley–Interscience Publication.

- [16] DO CARMO, M. P. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.
- [17] ERBAR, M., KUWADA, K., AND STURM, K.-T. On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Invent. Math.*, Published online, Dec 2014. <http://dx.doi.org/10.1007/s00222-014-0563-7>
- [18] ERBAR, M., AND MAAS, J. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.* 206, 3 (2012), 997–1038.
- [19] FIGALLI, A., RIFFORD, L., AND VILLANI, C. Nearly round spheres look convex. *Amer. J. Math.* 134 (2012), 109–139.
- [20] GALLOT, S., HULIN, D., AND LAFONTAINE, J. *Riemannian geometry*, second ed. Universitext. Springer-Verlag, Berlin, 1990.
- [21] EVANS, L. C. AND GANGBO, W. Differential equations methods for the Monge–Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.* 137, no. 653, 1999.
- [22] FIGALLI, A., KIM, Y.H., AND MCCANN, R.J. When is multidimensional screening a convex program? *J. of Economic Theory* 146, 2 (2011), 454–478.
- [23] FUNANO, K. Eigenvalues of Laplacian and multi-way isoperimetric constants on weighted Riemannian manifolds. Preprint, 2014. Available online at <https://sites.google.com/site/keifunanoshomepage/>
- [24] GIGLI, N. The splitting theorem in non-smooth context. Preprint, 2013. Available online at [arXiv:1302.5555v1](https://arxiv.org/abs/1302.5555v1) An overview of this paper is also available on the author's Web page.
- [25] GOZLAN, N. A characterization of dimension free concentration inequalities. *Annals of Probability* 37, 6 (2009), 2480–2498.
- [26] GROMOV, M. Sign and geometric meaning of curvature. *Rend. Sem. Mat. Fis. Milano* 61 (1991), 9–123 (1994).
- [27] GRUNEWALD, N., OTTO, F., VILLANI, C., AND WESTDICKENBERG, M. A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. H. Poincaré Probab. Stat.* 45, 2 (2009), 302–351.
- [28] JORDAN, R., KINDERLEHRER, D., AND OTTO, F. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* 29, 1 (1998), 1–17.
- [29] KUWADA, K. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.* 258, 11 (2010), 3758–3774.

- [30] LEDOUX, M. *The concentration of measure phenomenon*. American Mathematical Society, Providence, RI, 2001.
- [31] LOEPER, G. On the regularity of solutions of optimal transportation problems. *Acta Math.* 202, 2 (2009), 241–283.
- [32] LOEPER, G., AND VILLANI, C. Regularity of optimal transport in curved geometry: the nonfocal case. *Duke Math. J.* 151 (2010), 431–485.
- [33] LOTT, J. Optimal transport and Perelman’s reduced volume. *Calc. Var. Partial Differential Equations* 36 (2009), 49–84.
- [34] LOTT, J., AND VILLANI, C. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Math.* 169, p. 903–991
- [35] MA, X.-N., TRUDINGER, N. S., AND WANG, X.-J. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* 177, 2 (2005), 151–183.
- [36] MARTON, K. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* 6 (1996), 556–571.
- [37] MCCANN, R. J. A convexity principle for interacting gases. *Adv. Math.* 128, 1 (1997), 153–179.
- [38] MCCANN, R.J., AND TOPPING, P.M. Ricci flow, entropy and optimal transportation. *Amer. J. Math.* 132, 3 (2010), 711–730.
- [39] MONDINO, A., AND NABER, A. Structure Theory of Metric-Measure Spaces with Lower Ricci Curvature Bounds I. Preprint, [arXiv:1405.2222](https://arxiv.org/abs/1405.2222).
- [40] OHTA, SH.-I. Optimal transport and Ricci curvature in Finsler geometry. *Adv. Stud. Pure Math.* 57 (2010), 323–342.
- [41] OLLIVIER, Y. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, *Adv. Stud. Pure Math.* 57, Math. Soc. Japan, pp. 343–381, 2010.
- [42] OLLIVIER, Y. A curved Brunn-Minkowski inequality on the discrete hypercube, Or: What is the Ricci curvature of the discrete hypercube? *SIAM J. Discr. Math.* 26, 3 (2012), 983–996.
- [43] OTTO, F. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* 26, 1–2 (2001), 101–174.
- [44] OTTO, F., AND VILLANI, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* 173, 2 (2000), 361–400.
- [45] PETRUNIN, A. Alexandrov meets Lott–Sturm–Villani. *Münster J. Math.* 4 (2011), 53–64.



- [46] RAJALA, T. Local Poincaré inequalities from stable curvature conditions on metric spaces. *Calc. Var. Partial Differential Equations* 44 (2012), 477–494.
- [47] STURM, K.-T. On the geometry of metric measure spaces. *Acta Math.* 196, 1 (2006), 65–131 (Part I) and 133–177 (Part II).
- [48] STURM, K.-T. Entropic Measure on multidimensional spaces. In *Stochastic analysis, random fields and applications, VI. Progress in Probability*. Birkhuser, 2010.
- [49] STURM, K.-T., AND VON RENESSE, M.-K. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Comm. Pure Appl. Math.* 58, 7 (2005), 923–940.
- [50] TOPPING, P.  $\mathcal{L}$ -optimal transportation for Ricci flow. *J. reine und angewandte Mathematik* 636 (2009), 93–122.
- [51] VILLANI, C. *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [52] VILLANI, C. Weak stability of a fourth-order curvature condition arising in optimal transport theory. *J. Funct. Anal.* 255, 9 (2008), 2683–2708.
- [53] VILLANI, C. *Optimal transport, old and new*, vol. 338 of *Grundlehren der mathematischen Wissenschaften*, Vol. 338, Springer, 2009.

These references may be complemented by various lecture notes which can easily be found online; besides the obvious recommendation to my own writings, I should mention the excellent survey papers and presentations done by the Italian school; and also the lecture notes which I edited with Hervé Pajot and Yann Ollivier, *Optimal Transportation – Theory and Applications*, London Mathematical Society Lecture Note Series No. 413, 2004.

UNIVERSITÉ DE LYON & FRENCH ACADEMY OF SCIENCES

INSTITUT HENRI POINCARÉ, 11 RUE PIERRE ET MARIE CURIE, F-75231 PARIS CEDEX 05, FRANCE

villani@math.univ-lyon1.fr