Japan. J. Math. 11, 219–263 (2016)
DOI: 10.1007/s11537-016-1531-3

JAPANESE JOURNAL OF MATHEMATICS

Synthetic theory of Ricci curvature bounds*

Cédric Villani

Received: 1 March 2016 / Revised: 16 July 2016 / Accepted: 27 July 2016
Published online: 29 August 2016

Communicated by: Toshiyuki Kobayashi

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

Keywords and phrases: Ricci curvature, optimal transport, metric geometry, synthetic geometry, information geometry

Mathematics Subject Classification (2010): 53B21, 90C08

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";

---

* This article is based on the 15th Takagi Lectures that the author delivered at Tohoku University on June 27 and 28, 2015.

C. Villani
Université de Lyon Claude Bernard & French Academy of Sciences and
Institut Henri Poincaré, 11 rue Pierre et Marie Curie, F-75231 Paris Cedex 05, France
(e-mail: villani@math.univ-lyon1.fr)

- 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. Further 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; and the annual meeting of VIASM, Hanoi, in August 2015; I am grateful to all these organizations for their support. Final improvements were incorporated for publication in the *Japanese Journal of Mathematics*, with the help of two outstanding anonymous referees.

1. Synthetic point of view

As in my book [73] I will use the familiar notion of convexity to illustrate the concept of "synthetic" point of view, and compare it with the concept of "analytic" point of view.

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

(a) $\forall x \in \Omega, \qquad \nabla^2\varphi(x) \geq 0$, where $\nabla^2\varphi$ is the Hessian matrix of φ;
(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 φ, 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. This 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: many of our convexity games involve 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 φ 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* φ enjoys some regularity and some version of (a) holds true. This is the content of Alexandrov's regularity theorem: if $\varphi : \Omega \to \mathbb{R}$ is convex (in sense (b)), then φ is twice almost everywhere differentiable, so $\nabla^2\varphi$ exists almost everywhere, and 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 ancient Greek style, with properties of circles, parallel lines, etc.). Both have their rewards: the analytic approach is often more effective, while the synthetic one is more economical, more to the point, and, well, more synthetic!

2. The forces in presence

The four theories which are evoked in the sequel are all well-established and respectable; and, for a long time, roughly up to the mid-nineties, they lived happily with hardly any connection between them—except, arguably, for some links between curvature and gradient flow (for instance in the geometry of positively curved Alexandrov spaces or the theory of Ricci flow), and some links between geometry and information theory (as in the theory of information geometry [2]).

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 respective velocities u and v):

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

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

The form of inequality (1) makes it clear that κ is invariant by isometry. Spaces of constant sectional curvatures κ 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 π: that is, 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 [11, 12].

2.1.3. Ricci curvature When dealing with surfaces, sectional curvature, then called Gauss curvature, is the only curvature that one needs to know. But in dimension $n \geq 3$, further notions of curvature must be defined, and more care is needed.

The Ricci curvature is famous, among other things, 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 $(e_2, \ldots, e_n)$ in such a way that $(e_1, \ldots, e_n)$ is an orthonormal basis of T_xM, and let

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

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

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

$$\frac{d^2}{dt^2}(\mathscr{J}^{1/n}) + \frac{\mathrm{Ric}\,(\dot\gamma,\dot\gamma)}{n}(\mathscr{J}^{1/n}) \le 0. \tag{3}$$

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

$$\xi \cdot \nabla(\nabla \cdot \xi) - \Delta \frac{|\xi|^2}{2} + \frac{(\nabla \cdot \xi)^2}{n} + \mathrm{Ric}\,(\xi,\xi) \le 0. \tag{4}$$

Usually the latter formula is stated for $\xi = \nabla\psi$, where ψ is an arbitrary smooth function, in the form

$$-\nabla\psi \cdot \nabla\Delta\psi + \Delta\frac{|\nabla\psi|^2}{2} \ge \frac{(\Delta\psi)^2}{n} + \mathrm{Ric}\,(\nabla\psi,\nabla\psi). \tag{5}$$

As an immediate corollary, if $n \le N$ and Ric $\ge Kg$ (g the Riemannian metric, $N \in [1,\infty)$, $K \in \mathbb{R}$), one has the **curvature-dimension inequality** $\mathrm{CD}(K,N)$:

$$\forall\psi \in C^3(M), \quad -\nabla\psi \cdot \nabla\Delta\psi + \Delta\frac{|\nabla\psi|^2}{2} \ge \frac{(\Delta\psi)^2}{N} + K\,|\nabla\psi|^2. \tag{6}$$

This inequality also works for $\mathrm{CD}(K,\infty)$: just remove the term in $(\Delta\psi)^2$. The Lagrangian reformulations of $\mathrm{CD}(K,N)$ are

$$\frac{d^2}{dt^2}(\mathscr{J}^{1/N}) + \frac{K\,|\dot\gamma|^2}{N}(\mathscr{J}^{1/N}) \le 0 \quad \text{for } N < \infty; \tag{7}$$

$$\frac{d^2}{dt^2}(\log \mathscr{J}) + K|\dot\gamma|^2 \le 0 \quad \text{for } N = \infty. \tag{8}$$

Comparison of (1) on the one hand, (3)–(8) on the other hand, reveals several important differences in spirit:

- While (1) is about distances, inequality (3) 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*,

$$-\nabla\psi\cdot\nabla\Delta\psi + \Delta\frac{|\nabla\psi|^2}{2} = \|\nabla^2\psi\|^2 + \mathrm{Ric}\,(\nabla\psi,\nabla\psi); \qquad (9)$$

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 (6) can be seen as a property of the Laplace operator, noting that $|\nabla\psi|^2 = (\Delta\psi^2 - 2\psi\Delta\psi)/2$. But one may replace Δ 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 $\nu = e^{-V}\mathrm{vol}$, which is the equilibrium measure for L. Then it makes sense to define the $\mathrm{CD}(K,N)$ criterion for the measure ν or the linear diffusion operator L:

$$-\nabla\psi\cdot\nabla L\psi + L\,\frac{|\nabla\psi|^2}{2} \geq \frac{(L\psi)^2}{N} + K|\nabla\psi|^2. \qquad (10)$$

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 $N \geq 1$). Note that inequalities (7) or (8) will still hold true if $\mathcal{J}(t,x)$ is understood as

$$\lim_{r\to 0}\frac{\nu[\exp(t\nabla\psi)(B_r(x))]}{\nu[B_r(x)]} = |\det d_x\exp(t\nabla\psi)|\,\frac{e^{-V(\exp_x t\nabla\psi(x))}}{e^{-V(x)}}.$$

As a typical example, the Gaussian space $(\mathbb{R}^n,\gamma)$, where γ is the standard (unit covariance matrix) Gaussian distribution, can be considered as a $\mathrm{CD}(1,\infty)$ 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.

Inequality (6), or its Lagrangian counterparts, is the working heart of Ricci curvature analysis. It is 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 with nonnegative Ricci curvature, and M is the set of midpoints of X and Y, then

$$\mathrm{vol}\,[M]^{1/n} \geq \frac{1}{2}(\mathrm{vol}\,[X]^{1/n} + \mathrm{vol}\,[Y]^{1/n}). \qquad (11)$$

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 (11) reduces to the classical "algebraic" Brunn–Minkowski inequality:

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

which has a long history in the geometry of Euclidean space.

Another key ingredient in the history of $\mathrm{CD}(K, N)$ spaces is the Bakry–Émery calculus, which is a set of techniques and recipes involving nonlinear changes of functions, commutators, diffusion operators, and differential calculus of first and second order. The beautiful reviews by Ledoux provide a possible entry point to this theory.

2.1.4. Scalar curvature $S = \mathrm{tr}\,(\mathrm{Ric})$ 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 [37] and the treatises [22,30] 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 $\mathscr{M}$ and the energy is a function $\mathscr{E} : \mathscr{M} \to \mathbb{R}$. 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

$$\dot{X} = -\,\mathrm{grad}\,\mathscr{E}(X). \tag{12}$$

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 $\mathscr{E}$; 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

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

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 $\mathscr{E}$ 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 of analysis, 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 $k \in \mathbb{N}$

$$X_k^{(\tau)} := \arg\min_X \Big[\mathscr{E}(X) + \frac{d(X_{k-1}^{(\tau)}, X)^2}{2\tau}\Big]. \tag{13}$$

Then as $\tau \to 0$ and $k\tau \to t$, 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 (13): if $X_k^{(\tau)} \simeq X_{k-1}^{(\tau)}$, and the geometry is infinitesimally Hilbert, the approximate Euler–Lagrange equation should be

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

which is a discrete formulation of (12).

Of course if $\mathscr{E}$ is lower semicontinuous and there is a compactness estimate, then the minimization problem in (13) has a solution. The program can be carried out under various assumptions on $\mathscr{E}$ and $|\operatorname{grad}\mathscr{E}|$; see [4]. One can prove typically the $1/2$-Hölder regularity of $Y(t)$ in the time variable t, the finiteness of $\mathscr{E}(Y(t))$ for any $t > 0$, etc.

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

$$\frac{d}{dt}\Big(\frac{d(Z, X(t))^2}{2}\Big) = \Big\langle \frac{dX}{dt}, \frac{d\gamma}{ds}\Big\rangle\Big|_{s=1}, \tag{14}$$

where $(\gamma(s))_{0\le s\le 1}$ is a minimizing geodesic joining $\gamma(0) = Z$ to $\gamma(1) = X(t)$. If $X(t)$ evolves according to (12), then the right-hand side in (14) coincides with $-(d/ds)\mathscr{E}(\gamma(s))$, so the idea is to compare $(d/dt)d(Z,X(t))^2/2$ with $-(d/ds)\mathscr{E}(\gamma(s))$. These are rates of variation along different time evolutions. This may be simplified by using a convexity assumption: indeed, for instance, if $\mathscr{E}$ is convex, then

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

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

$$\forall Z, \quad -\frac{d^+}{dt}\Big(\frac{d(Z,X(t))^2}{2}\Big) \geq \mathscr{E}(X(t)) - \mathscr{E}(Z). \tag{15}$$

This is in case $\mathscr{E}$ is convex; there are similar formulations for semiconvex functionals $\mathscr{E}$. Many variants have been explored; once again I shall refer to [4] 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

$$S_\nu(\mu) = -\int \rho \log \rho \, d\nu, \quad \rho = \frac{d\mu}{d\nu}. \tag{16}$$

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 ρ (what is rare is valuable, so the elementary information should be an increasing function of $1/\rho$)
- independent signal sources should add up their respective information.

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:

$$\mathbb{P}\Big[\frac{1}{N}\sum \delta_{X_i} \simeq \mu\Big] \simeq e^{-NH_\nu(\mu)}, \tag{17}$$

where the X_i are independent random variables distributed according to the law ν.

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

$$\frac{d}{dt} H_\nu(\mu_t) = -I_\nu(\mu_t), \tag{18}$$

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

$$I_\nu(\mu) = \int \frac{|\nabla\rho|^2}{\rho}\, d\nu, \quad \rho = \frac{d\mu}{d\nu}. \tag{19}$$

The latter 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 [20].

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

$$U_\nu(\mu) = \int U\left(\frac{d\mu}{d\nu}\right) d\nu, \tag{20}$$

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 and geometry is the Monge–Kantorovich 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 narrated and 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 F. Spufford.

The basic data in optimal transport consists of a cost function c in two variables, and two probability measures, say μ_0 and μ_1, defined on a source space $\mathscr{X}_0$ and a target space $\mathscr{X}_1$ respectively. Then one defines

• the *Monge variational problem*,

$$\inf_{T:\mathscr{X}_0\to\mathscr{X}_1}\left\{\int c(x,T(x))\,\mu_0(dx);\ T_\sharp\mu_0=\mu_1\right\}. \tag{21}$$

Recall that $T_\sharp\mu[A]=\mu[T^{-1}(A)]$ defines the image measure of μ by T, which is also an abstract way to introduce a change of variables, so that $\int(\varphi\circ T)\,d\mu=\int\varphi\,d(T_\sharp\mu)$; any such measurable map T is called a *transport map*, and if it achieves the infimum it will be an *optimal transport map*. The minimization problem (21) is highly nonlinear and nonconvex, 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,*

$$\inf_{\pi\in\Pi(\mathscr{X}_0\times\mathscr{X}_1)}\left\{\iint c(x,y)\,\pi(dx\,dy);\ \pi\in\Pi(\mu_0,\mu_1)\right\}. \tag{22}$$

Here $P(\mathscr{X}_0\times\mathscr{X}_1)$ is the set of probability measures on $\mathscr{X}_0\times\mathscr{X}_1$ (joint probability measures), while $\Pi(\mu_0,\mu_1)$ is the set of joint probability measures having marginals μ_0 and μ_1 respectively. Any probability measure $\pi\in\Pi(\mu_0,\mu_1)$ is called a *transport plan*, and if it achieves the minimum in (22) it will be an *optimal transport plan*. The ansatz $\pi=(\mathrm{Id}\,,T)_\sharp\mu_0$ reduces (22) to (21); 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 problems are formalizations of the same question: how to optimally rearrange a given distribution of mass μ_0, defined on $\mathscr{X}_0$, into another distribution of mass μ_1, defined on $\mathscr{X}_1$? Under very general assumptions (say if μ_0 has no atoms and c is continuous), the values of the optima in (21) and (22) coincide; this common value is called the optimal transport cost. It is often the case that $\mathscr{X}_0=\mathscr{X}_1$, but this might not always be.

2.4.2. Duality For a long time the most active 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,

$$\min_{\pi \in \Pi(\mu_0,\mu_1)} \iint c(x,y)\,\pi(dx\,dy) = \sup_{\psi_1 - \psi_0 \leq c} \left\{ \int \psi_1(y)\, d\mu_1(y) - \int \psi_0(x)\, d\mu_0(x) \right\}. \qquad (23)$$

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 (23) in terms of economy, prices and profit (which, incidentally, made this theorem look like heresy to the eyes of ideologists of communist Russia); then ψ_0 and ψ_1 should be interpreted as prices, and the condition is that the difference of values between the start and the end of the transport should never exceed the transport cost.

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

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

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

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

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 [73].) At the same time, Brenier and Cullen showed the relevance of solving the Monge problem and understanding the structure of the optimal transport map, not just the value of the transport cost.

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 μ_0 is absolutely continuous with respect to Lebesgue measure, then the optimal transport map T is uniquely defined and characterized by the form

$$T = \nabla \varphi, \quad \varphi \text{ a convex function.} \qquad (24)$$

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

$$T = \exp(\nabla\psi), \quad \psi \text{ a } c\text{-convex function.} \tag{25}$$

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

$$\forall x, \quad \psi(x) = \sup_y [\zeta(y) - c(x, y)]. \tag{26}$$

The relation between ζ and ψ may be thought of as a generalized Legendre transform, adapted to the cost function. Similarly, c-convexity is a generalization of the usual notion of convexity.

Here are two important remarks on these properties:

- c-convexity implies semiconvexity, so ψ is just as regular as a convex function;
- any C^2-small 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 [3, 17, 13, 25]. 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 may 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 $\nabla\Psi$;
- transport rays do not cross except possibly at their endpoints, and even that is a negligible event when μ, ν are absolutely continuous and have no shared mass.
- 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 much revived recently, is that of *transport distance*. Let $(\mathcal{X}, d)$ be a separable complete metric space. For $p \in [1, \infty)$ let us define

$$C_p(\mu_0, \mu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \iint d(x, y)^p \, \pi(dx\, dy)$$

as the value of the optimal transport cost between μ_0 and μ_1 when the cost function is $c(x, y) = d(x, y)^p$; and

$$W_p(\mu_0, \mu_1) = C_p(\mu_0, \mu_1)^{1/p}. \tag{27}$$

Then W_p defines a distance on the set $P_p(\mathscr{X})$ of probability measures with finite p-moment. The convergence is the weak convergence against continuous functions φ with $\varphi(x) = O(1 + d(x_0, x)^p)$ (x_0 arbitrary); so if $\mathscr{X}$ is bounded, this is just the weak convergence of probability measures, that is the convergence against bounded continuous test functions.

The denomination of Wasserstein distance has become standard for the W_p functionals, even though it is historically not very accurate. By far the most important cases are again $p = 1$ and $p = 2$. If the base space is compact, they correspond to the same topology, but to distinct geometries.

Then there is also a notion of *transport interpolation*, when the base space itself admits an interpolation recipe, 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, defined almost everywhere and measurable); then if π is a transport plan we may deduce from it an interpolation at the level of measures:

$$\mu_t = (I_t)_\# \pi, \quad t \in [0, 1]. \tag{28}$$

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 $P(\mathscr{X})$, with a geometry of its own, which is deduced from the geometry on $\mathscr{X}$ but also contains it: indeed, $x \mapsto \delta_x$ maps $\mathscr{X}$ isometrically into $P_p(\mathscr{X})$. A key discovery from the past fifteen years is that *relevant information on the geometry of $\mathscr{X}$ can be expressed through geometric properties of $P(\mathscr{X})$.*

2.4.5. Further reference The existence, uniqueness and regularity of optimal transport, the properties of transport interpolation (displacement interpolation), the underlying partial differential equations: all this has turned into a huge field. My book [73] 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 should be mentioned. The first one was McCann's PhD Thesis [54], studying convexity properties of information-theoretical functionals in the geometry of optimal transport. A main result in this Thesis (slightly restated here), is that if a convex nonlinearity U is given with $U(0) = 0$, then the functional

$$\mathscr{U} : \mu \longmapsto \int U\left(\frac{d\mu}{dx}\right) dx \tag{29}$$

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

$$s^n\, U(s^{-n}) \text{ is a convex function of } s \in \mathbb{R}_+. \tag{30}$$

Condition (30) 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 [53] that comparisons between optimal transport distance W_p and information theoretical functionals like H_ν 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 γ stands for the normalized Gaussian distribution on $\mathbb{R}^n$, then

$$\forall \mu \in P_2(\mathbb{R}^n), \quad W_2(\mu, \gamma) \leq \sqrt{2H_\gamma(\mu)}. \tag{31}$$

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 μ in Gaussian space (for instance if μ is the restriction of γ to a subset C, then the right-hand side depends on the logarithm of the Gaussian measure of C); while the left-hand side tells us about the effort needed to "invade" the whole space if we spread the mass of μ. 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, then its enlargement quickly takes up most of the measure. Pioneered by Lévy and V. Milman, concentration theory has turned into a huge field, reviewed by Ledoux [46].

3.2. Jordan–Kinderlehrer–Otto

In 1998 Jordan, Kinderlehrer and Otto published a paper [40] 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_ν, in the geometry of W_2, is the heat equation.* This theorem, stated by the authors in $\mathbb{R}^n$ with a measure $\nu(dx) = e^{-V(x)}\,dx$, 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. Below 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 $\nu(dx) = e^{-V(x)}\,\mathrm{vol}\,(dx)$ be a reference probability measure on M. Let μ_0 be an initial probability measure on M. Apply the procedure described in Subsect. 2.2.1 with $d = W_2$ and $\mathscr{E} = H_\nu$; this defines $(\mu_k^{(\tau)})_{k\in\mathbb{N}}$. Then $\mu_k^{(\tau)}$ converges in the limit $\tau \to 0$, $k\tau \to t$, to $\mu(t)$, which is the solution, evaluated at time t, of the natural diffusion equation with equilibrium state ν:

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

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 [61]. Shortly after that encounter, I happened to be reading a survey article by Ledoux on concentration theory, which was a precursor of his monograph [46]. 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 [62], in which the relation to geometry was explored.

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, the information-theoretical inequality

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

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

$$\forall \mu \in P(M), \quad W_2(\mu, \nu) \leq \sqrt{\frac{2H_\nu(\mu)}{K}}. \tag{33}$$

The original proof required some minor assumptions about the behavior of ν at infinity, but soon there were versions for noncompact spaces, and again, this theorem has been generalized to an amazing degree, even to plain metric spaces [36]. Note that it gives a new understanding of Talagrand's original inequality, as a consequence of the logarithmic Sobolev inequality.

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

(c) the formalization of a link between curvature, information theory and optimal transport, relying mainly on an expansion of the formal Riemannian intuition in [61]. 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:

$$\|\dot{\mu}\|_\mu = \inf\left\{\sqrt{\int |\xi|^2 \, d\mu}; \quad \dot{\mu} + \nabla \cdot (\mu \xi) = 0\right\}. \tag{34}$$

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

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mu \nabla \varphi) = 0, \quad \frac{\partial \varphi}{\partial t} + \frac{|\nabla \varphi|^2}{2} = 0. \tag{35}$$

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

While the formal setting was there, some technical tools were missing, especially in relation to the nonsmoothness of optimal transport and the Riemannian formalism. Most 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 easiest to be accepted (Paul Malliavin, who handled it himself, accepted it the day after it was submitted to the *Journal of Functional Analysis*); still, 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 [62] deserve particular notice: Cordero-Erausquin–McCann–Schmuckenschläger [19] and Sturm–von Renesse [69]. 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 $\mathrm{CD}(K, \infty)$ can be characterized either by properties of uniform convexity of H_ν along optimal transport, or by contraction properties of the heat equation. Variants were obtained for all $\mathrm{CD}(K, N)$ conditions.

With this the scene was ready for the development of the synthetic theory of Ricci bounds. It would mainly consist in *using the properties of certain non-linear, 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, $\mathscr{D}C_N$, is defined as the set of all continuous convex functions $U : \mathbb{R}_+ \to \mathbb{R}$ 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 fields 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 of nonsmooth analysis, as well as Cheeger, Sturm and others; finer degrees of understanding were brought by Ambrosio, Gigli and Savaré, in particular the analysis of the nonsmooth Sobolev spaces.
- 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; as shown by Ambrosio–Gigli–Savaré, the Hamilton–Jacobi equation $\partial_t h + \frac{|\nabla h|^2}{2} = 0$ (for almost all times) and the Hopf–Lax formula $h(t, x) = \inf_y \left(h_0(y) + \frac{d(x,y)^2}{2t}\right)$ always describe the same evolution...
- localization, which is a way to reduce global geometric inequalities to inequalities on geodesic lines forming a kind of partition of the geometry; this strategy takes its roots in works of Payne–Weinberger, Gromov–Milman,

Kannan–Lovász–Simonovits, and was connected to smooth optimal transport by Klartag, before Cavalletti and Mondino extended this to nonsmooth settings.

For me, the story was also deeply woven 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* [71], 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 optimal transport. The second book, *Optimal Transport. Old and New* [73], 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, probably) 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 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. The resulting papers [51,67] founded the theory, and thus the geometric spaces defined there are often called LSV for Lott–Sturm–Villani. 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 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. Preliminaries

Recall that

- the metric derivative $|\dot\gamma|$ of a path γ, valued in a metric space $(\mathscr{X}, d)$, is defined by
$$|\dot\gamma_t| = \limsup_{s\downarrow 0} \frac{d(\gamma(t),\gamma(t+s))}{s};$$

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

$$L(\gamma) = \int_0^1 |\dot{\gamma}_t| \, dt$$
$$= \sup \Big\{ \sum_{i=1}^{N-1} d(\gamma(t_i), \gamma(t_{i+1})), \; 0 = t_1 < t_2 < \cdots < t_N = 1 \Big\};$$

- a Lipschitz path $\gamma : [0,1] \to \mathcal{X}$ 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. Weak curvature-dimension properties

If a compact geodesic 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. This leads to the following new definitions:

- the evaluation map e_t is defined by $e_t(\gamma) = \gamma(t)$;
- a geodesic path $(\mu_t)_{0 \le t \le 1}$ is a geodesic in $P_2(\mathcal{X})$ equipped with the distance W_2;
- a dynamical optimal transference plan Π is a probability measure on the set of geodesic curves γ, such that $(\pi_0, \pi_1)_\# \Pi$ is an optimal transference plan; then $\mu_t = (e_t)_\# \Pi$ is a geodesic path in $P_2(\mathcal{X})$.

In the last definition, the optimal transference plan π and the geodesic path (μ_t) are said to be associated.

Definition 1. *A compact geodesic metric measured space $(\mathcal{X}, d, \nu)$ is said to be weakly* CD$(0, \infty)$ *if any two probability measures μ_0, μ_1 on $\mathcal{X}$ can be joined by a constant-speed geodesic $(\mu_t)_{0 \le t \le 1}$, such that*

$$\forall t \in [0,1] \quad H_\nu(\mu_t) \le (1-t) H_\nu(\mu_0) + t H_\nu(\mu_1). \qquad (36)$$

This is the simplest in a family of related definitions. 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_ν by a requirement on a more general class of functionals, namely all functionals U_ν with $U \in \mathscr{D}C_\infty$;
- by requiring inequality (36) to hold along *all* geodesics (μ_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 (36).

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 $(\mathscr{X}, d)$ 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 $\nu = e^{-V(x)}\,\mathrm{vol}\,(dx)$, Definition 1 is equivalent to the requirement that (M, ν) satisfies the CD$(0,\infty)$ criterion; that is, (6) with $K = 0$ and $N = \infty$. What makes this equivalence possible, following our remarks in Subsect. 2.4.3, is

- the possibility of "integrating" the concavity inequality (3);
- 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 μ_0 and μ_1.

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

$$H_{N,\nu}(\mu) = N \int \rho\,(1 - \rho^{-1/N})\,d\nu = \int U_N(\rho)\,d\nu, \tag{37}$$

where $U_N(r) = Nr(1 - r^{-1/N})$; and replacing, correlatively, the displacement convexity class $\mathscr{D}C_\infty$ by $\mathscr{D}C_N$.

Definition 2. *A compact geodesic metric measured space* $(\mathscr{X}, d, \nu)$ *is said to be weakly* CD(0, N)*,* $N < \infty$*, if any two probability measures* μ_0, μ_1 *on* $\mathscr{X}$ *can be joined by a constant-speed geodesic* $(\mu_t)_{0\le t\le 1}$ *such that*

$$\forall t \in [0,1], \quad H_{N,\nu}(\mu_t) \le (1-t)H_{N,\nu}(\mu_0) + tH_{N,\nu}(\mu_1). \tag{38}$$

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 [67] and Erbar–Kuwada–Sturm [23] 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

$$U^{\beta}_{\pi,\nu}(\mu) = \int_{\mathscr{X}} U\Big(\frac{\rho(x)}{\beta(x,y)}\Big)\,\beta(x,y)\,\pi(dy|x)\,\nu(dx), \tag{39}$$

where $\pi(dy|x)$ is the disintegration of π 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 π and the asymmetric role played by x and y. Accordingly, in the definitions, one will distinguish between the coupling π and the "reversed" coupling $\check{\pi}$, in which the variables x and y have been swapped: $\check{\pi} = S_\sharp\pi$, $S(x,y) = (y,x)$. In this formula π should be associated to the geodesic path (μ_t).

There are two reasonable choices for the comparison distortion coefficients. The most natural idea, first explored by Sturm [67], was 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

$$\beta_t^{(K,N)}(x,y) = \Big(\frac{\sin(t\alpha)}{t\sin\alpha}\Big)^{N-1}, \quad \alpha = \sqrt{\frac{K}{N-1}}\,d(x,y). \tag{40}$$

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 gain of one dimension 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 coefficients are

$$\overset{*}{\beta}_t^{(K,N)}(x,y) = \left(\frac{\sin(t\alpha)}{t\sin\alpha}\right)^N, \quad \alpha = \sqrt{\frac{K}{N}}\, d(x,y), \tag{41}$$

and they were considered by Bacher and Sturm [10] in the context of optimal transport, see also Deng and Sturm [21].

Then we may go from $\mathrm{CD}(0,N)$ to $\mathrm{CD}(K,N)$ by adapting (38) and distort the functional $H_{N,\nu}$ with either β or $\overset{*}{\beta}$. (I shall write $H^{\beta}_{N,\pi,\nu}$ as a shorthand of $(U_N)^{\beta}_{\pi,\nu}$.) If $\mathscr{X}$ is a smooth manifold, then *these two choices are equivalent*, which reveals a self-improvement property of distortion coefficients. In fact, at least in nonbranching geometries, 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 non-integer 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 measures, with a "dimension" parameter N that can easily be varied continuously.) Actually it is even possible to consider $N < 0$, but let us skip this!

(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

$$\ddot{h} - \frac{\dot{h}^2}{m} \geq K\sigma, \quad t \in [0,1]. \tag{42}$$

While inequality (42) is nonlinear, it still satisfies a maximum principle, and thus can be used in comparison principles. Let us write

$$\Psi_t^{K,m}(h_0,h_1,\sigma) = \begin{array}{l}\text{the solution of } \ddot{h} - \dot{h}^2/m = K\sigma \\ \text{taking values } h_0 \text{ and } h_1 \text{ at time 0 and 1} \\ \text{respectively, evaluated at time } t\end{array} \tag{43}$$

Then it turns that we can go from $\mathrm{CD}(0,\infty)$ to $\mathrm{CD}(K,N)$ by replacing the linear interpolation on the right-hand side of (36) by the nonlinear interpolation defined by (43). Please note that this uses only the *entropy* functional H_ν, not the finite-dimensional counterparts $H_{N,\nu}$.

All in all, we end up with three choices for the synthetic $\mathrm{CD}(K,N)$; that is, three different notions of weak $\mathrm{CD}(K,N)$ 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 $\mathrm{CD}(K,N)$, or $\mathrm{CD}^*(K,N)$, or $\mathrm{CD}^e(K,N)$, with $N\in[1,\infty)$, if, for any $\mu_0,\mu_1\in P(\mathcal{X})$ one can find a constant speed geodesic $(\mu_t)_{0\le t\le 1}$ in $(P(\mathcal{X}),W_2)$, and an associated optimal transference plan π, such that the following inequality is satisfied:*

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

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

$$\frac{d^2}{dt^2}H_\nu(\mu_t)-\frac{(\frac{d}{dt}H_\nu(\mu_t))^2}{N}\ge K\,W_2(\mu_0,\mu_1)^2.$$

Definition 4. *The formulas in* (3) *can be adapted to the case $N=\infty$, if the defining inequalities there are replaced, respectively,*

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

$$H_\nu(\mu_t)\le(1-t)H_{\pi,\nu}^{\beta_{1-t}^{(K,\infty)}}(\mu_0)+tH_{\check{\pi},\nu}^{\beta_t^{(K,\infty)}}(\mu_1), \tag{44}$$

where

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

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

$$H_\nu(\mu_t)\le(1-t)H_\nu(\mu_0)+tH_\nu(\mu_1)-\frac{Kt(1-t)}{2}W_2(\mu_0,\mu_1)^2. \tag{45}$$

The latter inequality, expressing the K-convexity of H_ν in the metric W_2, had long been identified as a plausible $\mathrm{CD}(K,\infty)$ condition [51].

So we have a choice of defining inequalities when $K\ne 0$. How can we compare the resulting definitions? Assuming that $(\mathcal{X},d)$ is nonbranching, one can show [10,15,21,23,42] that

- $\mathrm{CD}(K,N)$ is the most demanding: it implies both $\mathrm{CD}^*(K,N)$ and $\mathrm{CD}^e(K,N)$. It is also the one which naturally leads to sharp dimensional constants.
- $\mathrm{CD}^*(K,N)$ implies $\mathrm{CD}(K^*,N)$ with $K^*=K(1-1/N)$. This deterioration of the constant by a factor $1-1/N$ is typical.

- $\mathrm{CD}^*(K, N)$ is equivalent to $\mathrm{CD}^e(K, N)$; and $\mathrm{CD}^*(K, N)$ is locally equivalent to $\mathrm{CD}(K, N)$.
- $\mathrm{CD}^*(K, N)$ can be globalized: that is, if it holds locally (in the neighborhood of any point) then it also holds globally.
- It is equivalent to ask that $\mathscr{X}$ satisfies $\mathrm{CD}^*(N-1, N)$, or that the metric cone built over $\mathscr{X}$ satisfies $\mathrm{CD}(0, N+1)$. (As the archetypal example, think of the flat $\mathbb{R}^{N+1}$ as the cone built over $\mathbb{S}^N$.)
- $\mathrm{CD}^*(K, N)$ 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 CD, but either CD^* (more complicated) or CD^e (more simple), both being equivalent in a nonbranching context. It remains an open problem whether the weak $\mathrm{CD}(K, N)$ property is in fact equivalent to the weak $\mathrm{CD}^*(K, N)$ property, or whether it is a strictly stronger notion (in which case it should probably be dismissed because of the failure of globalization property).

Open Problem 5. In a nonbranching metric measure space, does $\mathrm{CD}^*(K, N)$ imply $\mathrm{CD}(K, N)$?

4.3. Measure Contraction Property

The Measure Contraction Property, or MCP, is another inequality involving geodesics and measures. Introduced by Sturm [66] and developed by various authors [18, Part III, Appendix 2] [44,45], it was reformulated in the context of optimal transport by Ohta [57] and independently Sturm [67, Part II] with minor variations. While strictly weaker than the weak $\mathrm{CD}(K, N)$ criteria of the previous subsection, MCP is useful to study certain degenerate geometries, for instance subriemannian spaces.

Definition 6. *A metric-measure space* $(\mathscr{X}, d, \nu)$ *is said to satisfy* $\mathrm{MCP}(K, N)$ *if for any* $x \in \mathscr{X}$ *and any measurable set* $A \in \mathscr{X}$ *with* $\nu[A] > 0$*, there exists a dynamical optimal transference plan* $\Pi = \Pi_{x,A}$ *such that* $(e_0)_\sharp \Pi = \delta_x$*,* $(e_1)_\sharp \Pi = (1_A \nu)/\nu[A]$ *is the restriction of* ν *to* A*, and*

$$\mu_t \geq (e_t)_\sharp (t\,\nu[A]\,\beta_t(\gamma)\,\Pi(d\gamma)).$$

Here $\beta_t(\gamma) = \beta_t(\gamma(0), \gamma(1))$ is the distortion coefficient along γ at time t.

The main advantages of this criterion are that

- it is automatically implied by $\mathrm{CD}(K, N)$;
- when applied to a smooth manifold of dimension N, it is equivalent to the usual $\mathrm{CD}(K, N)$ criterion;

- it applies to very singular spaces, such as the Heisenberg group.

One main disadvantage, though, is that it needs dimension information: when applied to a manifold (M, g) of dimension $n \neq N$ it fails to be equivalent to $\mathrm{CD}(K, N)$. Thus it cannot really earn the title of weak curvature-dimension inequality; still it may be useful at times.

4.4. 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 $\mathrm{CD}(K, N)$ criterion. But what else can be said of them?

The first major property is that the various notions of weak $\mathrm{CD}(K, N)$ spaces are all stable under a very weak but very natural notion of convergence of metric-measure spaces, namely the measured Gromov–Hausdorff topology. This topology 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. It also holds for $\mathrm{MCP}(K, N)$.

Stability of $\mathrm{CD}(K, N)$ is easily proven by taking advantage of two main properties:

- Optimal transport (say with cost $c = d^2$) is stable under Gromov–Hausdorff convergence; for instance, if $\mathscr{X}^k \to \mathscr{X}$ in the Gromov–Hausdorff topology, and (μ_t^k) is a geodesic path in $P_2(\mathscr{X}_k)$, for each k, then up to extraction of a subsequence it converges weakly to a geodesic path (μ_t) in $P_2(\mathscr{X})$. In fact, the convergence of $\mathscr{X}_k$ to $\mathscr{X}$ is equivalent to the convergence of $P_2(\mathscr{X}_k)$ to $P_2(\mathscr{X})$.
- Nonlinear functionals of information type are also stable, and more precisely lower semicontinuous, under weak convergence of *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 (in the sense of probability measures, that is, against $C_b(\mathscr{X})$) to μ and ν 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

$$U_{f_\sharp \nu}(f_\sharp \mu) \leq U_\nu(\mu).$$

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.

4.5. Generality

Let me write here generically CD for either CD, or CD*, or CD^e. Weak CD(K,N) spaces are strictly more general than classical CD(K,N) spaces; in particular, this class contains singular limits of 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; the synthetic notion of Ricci curvature allows for even more complicated singular behavior.

But, at a more fundamental level, weak 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 CD$(0,n)$ space. More generally, weak CD(K,N) spaces can handle certain classes of Finsler geometries.

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.

This wealth of weak CD(K,N) spaces can be seen as an advantage or as a drawback by geometers. On the one hand, normed spaces are quite natural metric spaces, and 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 all 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 CD$(0,N)$ and contains an infinite line γ (a geodesic which is minimizing for all times in $\mathbb{R}$), then it can be factored up into the metric product of γ 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 [6], is denoted with the letter R (for Riemannian); so it gives rise to weak RCD, RCD* or RCD^e spaces. It consists in imposing, on top of the considered curvature condition, the requirement that the Sobolev space $W^{1,2}$ be Hilbert.

Definition 7. *A metric-measure space* $(\mathscr{X},d,\nu)$ *is said to be* RCD(K,N) *(resp.* RCD$^*(K,N)$*,* RCD$^e(K,N)$*) if it is a weak* CD(K,N) *(resp.* CD$^*(K,N)$*,* CD$^e(K,N)$*) space and*

$$W^{1,2}(\mathscr{X},d)\ \text{ is a Hilbert space.} \tag{46}$$

This extra condition excludes non-Euclidean normed spaces, and, remarkably, narrows the class of $\mathrm{CD}(K, N)$ 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 γ,

$$f(\gamma(1)) - f(\gamma(0)) \le \int_0^1 g(\gamma(s))\,|\dot\gamma_s|\,ds.$$

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

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

  - 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 minimality is in the L^2 sense).
  - the Cheeger functional is

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

    and it allows to define an L^2-Sobolev space $W^{1,2}$, whose analysis has been the subject of an important body of work.
- Ambrosio, Gigli and Savaré [5] were motivated and inspired by optimal transport theory to develop the study of $W^{1,2}$; in particular, they proved the equivalence of two natural definitions of Sobolev spaces, one based on "vertical" variations (obtained by adding a small function) and the other ones based on "horizontal" variations (obtained by composing with a perturbation of the identity).
- 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 $\mathscr{X}$, so that it does allow for nonsmooth spaces in which not every tangent cone is a Euclidean space.

- Gigli noticed that this “Riemannian” condition can also be expressed by the *linearity of the heat equation*, and that this would be useful in exploring the stability of this notion [5].
- The 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* $\mathrm{CD}(K, N)$ *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, this class does not include normed spaces besides Euclidean space, and, as a major result by Gigli [35], $\mathrm{RCD}(0, N)$ spaces do enjoy a splitting property.

Finally, let us recall that, on the other side of the spectrum, if one wishes to allow even more generality than $\mathrm{CD}(K, N)$ notions, the MCP property is still available.

4.6. 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 [5] showing that in a geodesic space,

- the JKO scheme (evoked in Sect. 3.2) always converges to a trajectory of the L^2 gradient flow of the Cheeger functional;
- under a weak $\mathrm{CD}(K, \infty)$ 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_ν.

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 a lower Ricci curvature bound, which even in the smooth case is recognized as the key estimate for the well-posedness of the 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 [43]: it allows to bound the speed of variation, in W_2 sense, of any gradient flow of the Cheeger functional (any "solution of the heat equation") by the square root of the Fisher information; this 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 the property that $W^{1,2}$ be a Hilbert space.

4.7. Functional inequalities

Ever since McCann's transport-based 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 $(\mathscr{X}, d, \nu)$, pick up arbitrary compact sets A_0 and A_1 with positive measures, define μ_0 as the restriction of ν to A_0, μ_1 as the restriction of ν to A_1, and study the variations of the functionals $H_{N,\nu}$: under a weak $\mathrm{CD}(0, N)$ condition,

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

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, rescaling, differentiation, local study, blow-up, etc. For instance, Sobolev-type inequalities are captured by time-differentiating functional inequalities such as those which appear in Definition 3. 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 $\mathrm{CD}(K, N)$ 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 corresponding model $\mathrm{CD}(K,N)$ space. This result is also a key to Gromov's precompactness theorem, according to which the set of $\mathrm{CD}(K,N)$ 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. Let us say that a metric-measure space $(\mathscr{X}, d, \nu)$ is said to satisfy a local Poincaré inequality with constant C if, for any Lipschitz function u, any point $x_0 \in \mathscr{X}$ and any radius $r > 0$,

$$\fint_{B_r(x_0)} |u(x) - \langle u\rangle_{B_r(x_0)}|\, d\nu(x) \le C r \fint_{B_{2r}(x_0)} |\nabla u(x)|\, d\nu(x), \tag{47}$$

where $\fint_B = (\nu[B])^{-1} \int_B$ is the averaged integral over B, and $\langle u\rangle_B = \fint_B u\, d\nu$ 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 $\mathrm{CD}(K,N)$ spaces, for some time the proof of the local Poincaré inequality was relying on an unnatural nonbranching assumption, until a surprisingly neat and simple argument, focusing on the median rather than the mean, was found by Rajala [64] to establish the inequality in the highest generality.

Since the doubling and Poincaré inequalities are the two basic cornerstones of the modern analysis of metric spaces, weak $\mathrm{CD}(K,N)$ spaces fit into this context. But then, they also enjoy much more properties related to Sobolev inequalities, concentration of measure, etc.

A coup de théâtre occurred in the field when Cavalletti and Mondino [15] 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 statement: 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 disjoint geodesics γ such that the integral of f on each γ is also zero. Then the idea is to disintegrate the measure ν according to this set of geodesics, and thus write $\nu = \int \nu_\gamma\, \omega(d\gamma)$. Klartag discovered that L^1-optimal transport could be used as a tool to establish this decomposition, since the geodesic lines of transport are essentially noncrossing. Then Cavalletti and Mondino [15,16] 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 $c = d^2$). Both cost functions are reconciled in [15] because in this study it all boils down to particular types of transports which are optimal for both the d and d^2 cost functions.

With these tools, an impressive list of functional inequalities could be derived. Most of them belong 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^p-Poincaré inequalities, local Poincaré inequalities...
- heat kernel estimates: Li–Yau inequalities and their many variants.

A large part of my book [73] was devoted to presenting a synthetic view of many of these inequalities and their links, both in smooth and nonsmooth contexts. But since the writing of this book, the picture has been enriched, most remarkably with the emblematic Lévy–Gromov inequality: in the nonsmooth version of Cavalletti–Mondino, it asserts that a weak $\mathrm{CD}^*(K, N)$ space has an isoperimetric profile which is bounded below by the isoperimetric profile of the reference space. In other words, if $(\mathscr{X}, d, \nu)$ is $\mathrm{CD}^*(K, N)$ with $\nu[\mathscr{X}] = 1$, and σ is the normalized measure on the N-dimensional sphere of radius $\sqrt{(N-1)/K}$, then $\nu[A] = \sigma[B]$ implies $\nu^+[\partial A] \geq \sigma^+[\partial B]$ (here I am abusing notation by writing $\nu^+[\partial A]$ for the Minkowski content $\liminf(\nu[A^\varepsilon] - \nu[A])/\varepsilon$, as $\varepsilon \downarrow 0$).

Another problem which was tackled in weak $\mathrm{CD}(K, N)$ spaces was the solution of the optimal transport problem. It turned out that the main structure results for both the $c = d^2$ and the $c = d$ cost functions can be extended from the smooth setting to the case of essentially nonbranching $\mathrm{CD}^*(K, N)$ spaces. In the case $c = d^2$, one can show the existence of a Monge transport map which is determined by a c-convex potential [32]; while in the case $c = d$ one can prove a decomposition of the space into noncrossing geoedesics, with properties which are very similar to those in the solution of the classical Monge problem [13].

4.8. Rigidity

At the core of the synthetic theory of Ricci curvature lies the intrinsic study of weak $\mathrm{CD}(K, N)$ spaces, which are not a priori assumed to satisfy any regularity property. But it is also natural to ask whether these spaces do satisfy, a posteriori, some regularity. As an analogy, remember that convex functions $\mathbb{R}^n \to \mathbb{R}$ are automatically twice differentiable almost everywhere.

In the context of finite-dimensional Alexandrov spaces, some rather strong regularity results are known: in particular, angles can be defined everywhere; also, apart from a negligible set of singular points, these spaces are topological manifolds, on which a metric tensor g is defined, and even the Christoffel symbols can be constructed as measures.

For weak $\mathrm{CD}(K, N)$ spaces, things are more tricky. Counterexamples by Cheeger–Naber show that, for instance, even the existence of angles is not guaranteed everywhere, contrary to what occurs on Alexandrov spaces. One the one hand, because these spaces enjoy the doubling and Poincaré inequalities, Cheeger's differentiability theory applies to construct some sort of first-order differential calculus, and for instance some analogue of Rademacher's theorem will hold.

Mondino–Naber [56] prove more: an $\mathrm{RCD}(K, N)$ space with N finite is automatically rectifiable (basically: a countable union of Lipschitz images of measurable subsets of Euclidean spaces). In this result the dimension of the space may vary from place to place; it remains an open problem to show that $\mathrm{RCD}(K, N)$ implies the rectifiability with a uniform dimension. One may also speculate that there is a second differentiability structure in the sense of Honda [39].

Related to this problem is Gigli's construction [35,33,34] of a distributional Laplacian on $\mathrm{RCD}(K, N)$ spaces when $N < \infty$.

Another regularity issue is the rigidity of measure: if a metric measure space $(\mathscr{X}, d, \nu)$ is given with some curvature-dimension condition, say $\mathrm{RCD}(K, N)$, does it imply that any other measure ν' which makes it an $\mathrm{RCD}(K, N)$ space is absolutely continuous with respect to ν? Cavalletti–Mondino [14] establish some results in this direction.

4.9. Summary

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

- Bishop–Gromov and thus doubling inequalities;
- Bonnet–Myers diameter bounds;

- Poincaré inequalities, local and global;
- Brunn–Minkowski inequalities;
- Sobolev and logarithmic Sobolev inequalities;
- 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_ν.

Further, Gromov's precompactness theorem adapts to this setting: If $\mathscr{G}(K, N, R)$ stands for the set of, say, $\mathrm{CD}(K, N)$ spaces with diameter bounded by R, with $N < \infty$ and $R < \infty$, then $\mathscr{G}(K, N, R)$ is compact (not just precompact as in the smooth setting). Further, there are various natural metrics on $\mathscr{G}(K, N, R)$, such as Sturm's distance.

Also, the measure contraction property $\mathrm{MCP}(K, N)$ may be invoked but only leads to a smaller set of properties: Bishop–Gromov and Bonnet–Myers. A typical example is the Heisenberg group which does satisfy MCP(0, 5) but no curvature-dimension condition, neither the Brunn–Minkowski inequality [41].

(ii) *Nonbranching* $\mathrm{CD}^*(K, N)$ *spaces*, or equivalently nonbranching $\mathrm{CD}^e(K, N)$ 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;
- the possibility to solve the quadratic Monge problem (Gigli [32]) and, to some extent, the linear Monge problem (Cavalletti [13]);
- 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.

The possibility to encompass Finsler geometries can make the reader enthusiastic, but it will really become striking when more examples will be at hand. Let us record an open problem which seems important:

Open Problem 8. Find simple and general conditions under which the unit sphere of a norm in $\mathbb{R}^{N+1}$ is a $\mathrm{CD}^*(K, N)$ space with $K > 0$.

Note that the Cone property proven by Ketterer [42] for $\mathrm{CD}^*(K, N)$ spaces does not provide a simple answer, because the metric cone of the unit sphere of a non-Euclidean norm will not be the Euclidean $\mathbb{R}^{N+1}$.

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

- splitting theorem in nonnegative curvature (Gigli [35]);
- a distributional Laplace and Hessian-type operator (Gigli [35,33,34]);
- Li–Yau type estimates on the heat kernel (Garofalo–Mondino [31]);
- a Bochner formula (Erbar–Kuwada–Sturm [23]);
- a kind of Bakry–Émery calculus (Ambrosio–Mondino–Savaré [7,8]), which also implies the local-to-global property.

With this, it is likely that essentially any relevant theorem which is known to hold in the smooth $\mathrm{CD}(K, N)$ setting can be extended to $\mathrm{RCD}^*(K, N)$ spaces.

To the above list let us add

- information-theoretical inequalities of Shannon type (which are probabilistic analogues of the Brunn–Minkowski inequality, and hold true in nonnegative curvature) [23];
- good contraction properties for linear and nonlinear heat flows [23,65];
- rectifiability (Mondino–Naber [56]).

4.10. 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 $\mathscr{X}$ 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 $\mathrm{CD}(1, \infty)$ 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 Subsect. 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 $\mathrm{CAT}^{-}(\kappa)$ (sectional curvature bounded above by κ) or $\mathrm{CAT}^{+}(\kappa)$ spaces (sectional curvature bounded below by κ), 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 CAT^{-} spaces, which can be very rough; but 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 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 κ, then it satisfies the classical $\mathrm{CD}((n-1)\kappa, n)$ criterion. Thus, a satisfactory synthetic theory of Ricci curvature should keep this implication at the level of nonsmooth sectional curvature bounds: any n-dimensional $\mathrm{CAT}^{+}(\kappa)$ space should also satisfy a weak $\mathrm{CD}((n-1)\kappa, n)$ inequality. Note that this link is far from obvious, given the different formalisms in which CAT spaces (defined with lengths and angles) and weak $\mathrm{CD}(K, N)$ spaces (defined with probability measures) are expressed.

The problem stood wide open until it was solved by Petrunin [63], who gave full details in the particular case $\kappa = 0$. Since then these links have been confirmed: Alexandrov spaces are automatically 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. The study of this connection was pioneered by Ohta and further developed by Sturm and others [58]. As mentioned in Open Problem 8, finding $\mathrm{CD}(K, N)$ examples within the class of unit balls for non-Euclidean norms would add considerable value to this extension.

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 Sect. 2, is the following contraction estimate: under the condition $\mathrm{RCD}(K,\infty)$, if any two probability-valued solutions of the heat equation are given, say $(\mu_t)_{t\geq 0}$ and $(\widetilde{\mu}_t)_{t\geq 0}$, then

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

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 Sect. 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 $\partial_t g = -2\,\mathrm{Ric}$, where g is the Riemannian metric. As pointed out by McCann and Topping [55], 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 [50] and Topping [70].

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 Sect. 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 many works, one may consult [59], [60], [24] 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 μ_0 and μ_1 enjoy all the regularity that one could dream of (say C^∞ densities with upper and lower bounds), it may be that optimal transport is plainly discontinuous [73, Chapter 12].

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 [52] and later studied by a number of authors, including a key contribution by Loeper [48]. 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, and optimal transport on that elongated ellipsoid 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 [26];
- providing natural examples and counterexamples for regularity in the theory of fully nonlinear partial differential equations of Monge–Ampère type [48, 49];
- 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 [27] 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 [72]; it can be understood as a sort of nonlocal reinforcement of the $\mathrm{CAT}^+(0)$ 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 [18] 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 can always be enriched. Besides those which were already mentioned, here is a short nonexhaustive list of problems that it would be desirable to solve:

Open Problem 9. If $(\mathscr{X}, d, \nu)$ is a metric-measured space, of dimension 2 in some suitable sense, is $\mathrm{CAT}^+(\kappa)$ equivalent to $\mathrm{CD}(\kappa, 2)$ (rather than just implying it)?

Open Problem 10. Can one push the theory of weak $\mathrm{CD}(K, N)$ spaces to the point where it will encompass all currently known results about weak limits of $\mathrm{CD}(K, N)$ manifolds?

Open Problem 11. Enrich the list of significant examples and counterexamples for branching spaces.

Open Problem 12. Are there interesting weak $\mathrm{CD}(K, N)$ spaces which cannot be obtained as limits of Riemannian manifolds?

Another desirable goal would be to simplify the rigidity statements from [42, 15], which are quite intricate.

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 [61]: the gradient flow interpretation was naturally guiding 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 form 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 [38];
- 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 and Shioya [29, 28, 47] for spectral values of $-\Delta$ in nonnegative Ricci curvature, showing that successive eigenvalues λ_k satisfy $\lambda_k/\lambda_1 \le C^k$ 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 [9];
- the already mentioned works [50, 70], reinterpreting some of the Ricci flow theory, including some portions of Perel'man'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, Gigli, Gangbo and their collaborators. To quote just one striking result, Sturm constructed a beautiful “entropic measure” [68] 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 for the former.

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. Now, only seven years later, if I wanted to write 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 do not think I take much risk by adapting to this context Poincaré’s prediction that *surprising results shall be obtained*.

References

[1] S. Adams, N. Dirr, M.A. Peletier and J. Zimmer, From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage, Comm. Math. Phys., **307** (2011), 791–815.

[2] S. Amari and H. Nagaoka, Methods of Information Geometry. Translated from the 1993 Japanese Original by Daishi Harada, Transl. Math. Monogr., **191**, Amer. Math. Soc., Providence, RI; Oxford Univ. Press, Oxford, 2000.

[3] L. Ambrosio, Lecture notes on optimal transport problems, In: Mathematical Aspects of Evolving Interfaces, Funchal, 2000, Lecture Notes in Math., **1812**, Springer-Verlag, 2003, pp. 1–52.

[4] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Mmeasures, Lectures Math. ETH Zurich, Birkhäuser Verlag, Basel, 2005.

[5] L. Ambrosio, N. Gigli and G. Savaré, Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below, Invent. Math., **195** (2014), 289–391.

[6] L. Ambrosio, N. Gigli and G. Savaré, Metric measure spaces with Riemannian Ricci curvature bounded from below, Duke Math. J., **163** (2014), 1405–1490.

[7] L. Ambrosio, N. Gigli and G. Savaré, Bakry–Émery curvature-dimension condition and Riemannian Ricci curvature bounds, Ann. Probab., **43** (2015), 339–404.
[8] L. Ambrosio, A. Mondino and G. Savaré, On the Bakry–Émery condition, the gradient estimates and the local-to-global property of $\mathrm{RCD}^*(K, N)$ metric measure spaces, J. Geom. Anal., **26** (2016), 24–56.
[9] L. Ambrosio, G. Savaré and L. Zambotti, Existence and stability for Fokker–Planck equations with log-concave reference measure, Probab. Theory Related Fields, **145** (2009), 517–564.
[10] K. Bacher and K.-T. Sturm, Localization and tensorization properties of the curvature-dimension condition for metric measure spaces, J. Funct. Anal., **259** (2010), 28–56.
[11] D. Burago, Y. Burago and S. Ivanov, A Course in Metric Geometry, Grad. Stud. Math., **33**, Amer. Math. Soc., Providence, RI, 2001. Errata list available online at http://www.pdmi.ras.ru/staff/burag.
[12] Y. Burago, M. Gromov and G.Ya. Perel'man, A.D. Aleksandrov spaces with curvatures bounded below, Uspekhi Mat. Nauk, **47**, 2(284) (1992), 3–51.
[13] F. Cavalletti, Monge problem in metric measure spaces with Riemannian curvature-dimension condition, Nonlinear Anal., **99** (2014), 136–151.
[14] F. Cavalletti and A. Mondino, Measure rigidity of Ricci curvature lower bounds, Adv. Math., **286** (2016), 430–480.
[15] F. Cavalletti and A. Mondino, Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds, preprint, 2015.
[16] F. Cavalletti and A. Mondino, Sharp geometric and functional inequalities in metric measure spaces with lower Ricci curvature bounds, preprint, 2015.
[17] T. Champion and L. De Pascale, The Monge problem in $\mathbb{R}^d$, Duke Math. J., **157** (2011), 551–572.
[18] J. Cheeger and T.H. Colding, On the structure of spaces with Ricci curvature bounded below. I, J. Differential Geom., **46** (1997), 406–480; II, J. Differential Geom., **54** (2000), 13–35; III, J. Differential Geom., **54** (2000), 37–74.
[19] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger, A Riemannian interpolation inequality à la Borell, Brascamp and Lieb, Invent. Math., **146** (2001), 219–257.
[20] T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley Ser. Telecom., A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1991.
[21] Q. Deng and K.-T. Sturm, Localization and tensorization properties of the curvature-dimension condition for metric measure spaces. II, J. Funct. Anal., **260** (2011), 3718–3725.
[22] M.P. do Carmo, Riemannian Geometry. Translated from the second Portuguese edition by Francis Flaherty, Math. Theory Appl., Birkhäuser Boston Inc., Boston, MA, 1992.
[23] M. Erbar, K. Kuwada and K.-T. Sturm, On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces, Invent. Math., **201** (2015), 993–1071.
[24] M. Erbar and J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy, Arch. Ration. Mech. Anal., **206** (2012), 997–1038.
[25] L.C. Evans and W. Gangbo, Differential Equations Methods for the Monge–Kantorovich Mass Transfer Problem, Mem. Amer. Math. Soc., **137**, no. 653, Amer. Math. Soc., Providence, RI, 1999.
[26] A. Figalli, Y.-H. Kim and R.J. McCann, When is multidimensional screening a convex program?, J. Econom. Theory, **146** (2011), 454–478.
[27] A. Figalli, L. Rifford and C. Villani, Nearly round spheres look convex, Amer. J. Math., **134** (2012), 109–139.
[28] K. Funano, Eigenvalues of Laplacian and multi-way isoperimetric constants on weighted Riemannian manifolds, preprint, 2014. Available online at https://sites.google.com/site/keifunanoshomepage/.

[29] K. Funano and T. Shioya, Concentration, Ricci curvature, and eigenvalues of Laplacian, Geom. Funct. Anal., **23** (2013), 888–936.

[30] S. Gallot, D. Hulin and J. Lafontaine, Riemannian Geometry. Second ed., Universitext, Springer-Verlag, 1990.

[31] N. Garofalo and A. Mondino, Li–Yau and Harnack type inequalities in $\mathrm{RCD}^*(K,N)$ metric measure spaces, Nonlinear Anal., **95** (2014), 721–734.

[32] N. Gigli, Optimal maps in non branching spaces with Ricci curvature bounded from below, Geom. Funct. Anal., **22** (2012), 990–999.

[33] N. Gigli, On the Differential Structure of Metric Measure Spaces and Applications, Mem. Amer. Math. Soc., **236**, no. 1113, Amer. Math. Soc., Providence, RI, 2015.

[34] N. Gigli, Nonsmooth differential geometry—An approach tailored for spaces with Ricci curvature bounded from below, preprint, arXiv:1407.0809.

[35] N. Gigli, The splitting theorem in non-smooth context, preprint, arXiv:1302.5555v1. An overview of the proof has also been published in Anal. Geom. Metr. Spaces, **2** (2014), 169–213.

[36] N. Gozlan, A characterization of dimension free concentration in terms of transportation inequalities, Ann. Probab., **37** (2009), 2480–2498.

[37] M. Gromov, Sign and geometric meaning of curvature, Rend. Sem. Mat. Fis. Milano, **61** (1991), 9–123.

[38] N. Grunewald, F. Otto, C. Villani and M.G. Westdickenberg, A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit, Ann. Inst. Henri Poincaré Probab. Stat., **45** (2009), 302–351.

[39] S. Honda, A weakly second-order differential structure on rectifiable metric measure spaces, Geom. Topol., **18** (2014), 633–668.

[40] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal., **29** (1998), 1–17.

[41] N. Juillet, Geometric inequalities and generalized Ricci bounds in the Heisenberg group, Int. Math. Res. Not. IMRN, **2009**, 2347–2373.

[42] C. Ketterer, Cones over metric measure spaces and the maximal diameter theorem, J. Math. Pures Appl. (9), **103** (2015), 1228–1275.

[43] K. Kuwada, Duality on gradient estimates and Wasserstein controls, J. Funct. Anal., **258** (2010), 3758–3774.

[44] K. Kuwae and T. Shioya, On generalized measure contraction property and energy functionals over Lipschitz maps, Potential Anal., **15** (2001), 105–121.

[45] K. Kuwae and T. Shioya, Sobolev and Dirichlet spaces over maps between metric spaces, J. Reine Angew. Math., **555** (2003), 39–75.

[46] M. Ledoux, The Concentration of Measure Phenomenon, Math. Surveys Monogr., **89**, Amer. Math. Soc., Providence, RI, 2001.

[47] S. Liu, An optimal dimension-free upper bound for eigenvalue ratios, preprint, arXiv:1405.2213.

[48] G. Loeper, On the regularity of solutions of optimal transportation problems, Acta Math., **202** (2009), 241–283.

[49] G. Loeper and C. Villani, Regularity of optimal transport in curved geometry: the nonfocal case, Duke Math. J., **151** (2010), 431–485.

[50] J. Lott, Optimal transport and Perelman's reduced volume, Calc. Var. Partial Differential Equations, **36** (2009), 49–84.

[51] J. Lott and C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math. (2), **169** (2009), 903–991.

[52] X.-N. Ma, N.S. Trudinger and X.-J. Wang, Regularity of potential functions of the optimal transportation problem, Arch. Ration. Mech. Anal., **177** (2005), 151–183.

[53] K. Marton, A measure concentration inequality for contracting Markov chains, Geom. Funct. Anal., **6** (1996), 556–571.

[54] R.J. McCann, A convexity principle for interacting gases, Ph.D. thesis, Princeton Univ., 1994; Adv. Math., **128** (1997), 153–179.

[55] R.J. McCann and P.M. Topping, Ricci flow, entropy and optimal transportation, Amer. J. Math., **132** (2010), 711–730.

[56] A. Mondino and A. Naber, Structure theory of metric-measure spaces with lower Ricci curvature bounds I, preprint, arXiv:1405.2222.

[57] S. Ohta, On the measure contraction property of metric measure spaces, Comment. Math. Helv., **82** (2007), 805–828.

[58] S. Ohta, Optimal transport and Ricci curvature in Finsler geometry, In: Probabilistic Approach to Geometry, Adv. Stud. Pure Math., **57**, Math. Soc. Japan, Tokyo, 2010, pp. 323–342.

[59] Y. Ollivier, A survey of Ricci curvature for metric spaces and Markov chains, In: Probabilistic Approach to Geometry, Adv. Stud. Pure Math., **57**, Math. Soc. Japan, Tokyo, 2010, pp. 343–381.

[60] Y. Ollivier and C. Villani, A curved Brunn–Minkowski inequality on the discrete hypercube, or: What is the Ricci curvature of the discrete hypercube?, SIAM J. Discrete Math., **26** (2012), 983–996.

[61] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations, **26** (2001), 101–174.

[62] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal., **173** (2000), 361–400.

[63] A. Petrunin, Alexandrov meets Lott–Villani–Sturm, Münster J. Math., **4** (2011), 53–64.

[64] T. Rajala, Local Poincaré inequalities from stable curvature conditions on metric spaces, Calc. Var. Partial Differential Equations, **44** (2012), 477–494.

[65] G. Savaré, Self-improvement of the Bakry–Émery condition and Wasserstein contraction of the heat flow in $\mathrm{RCD}(K,\infty)$ metric measure spaces, Discrete Contin. Dyn. Syst., **34** (2014), 1641–1661.

[66] K.-T. Sturm, Diffusion processes and heat kernels on metric spaces, Ann. Probab., **26** (1998), 1–55.

[67] K.-T. Sturm, On the geometry of metric measure spaces. I, Acta Math., **196** (2006), 65–131; II, Acta Math., **196** (2006), 133–177.

[68] K.-T. Sturm, Entropic measure on multidimensional spaces, In: Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab., **63**, Birkhäuser, Basel, 2011.

[69] K.-T. Sturm and M.-K. von Renesse, Transport inequalities, gradient estimates, entropy, and Ricci curvature, Comm. Pure Appl. Math., **58** (2005), 923–940.

[70] P. Topping, $\mathscr{L}$-optimal transportation for Ricci flow, J. Reine Angew. Math., **636** (2009), 93–122.

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

[72] C. Villani, Stability of a 4th-order curvature condition arising in optimal transport theory, J. Funct. Anal., **255** (2008), 2683–2708.

[73] C. Villani, Optimal Transport. Old and New, Grundlehren Math. Wiss., **338**, Springer-Verlag, 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, 2014.