Université de Nice Sophia-Antipolis
Laboratoire J. A. Dieudonne

Mémoire de synthèse pour l'obtention de l'

HABILITATION À DIRIGER DES RECHERCHES

en **Sciences Mathématiques**

présenté par

Nicola Gigli

Geometric and analytic properties of some non smooth structures

Contents

Introduction

This work is to give an overview over the research that I've done up to now. The overall idea behind my interests has been that of understanding analytical and geometrical properties of non smooth spaces both from both a theoretical and a practical point of view. The two main classes of spaces on which I focussed are: the Wasserstein space $(\mathscr{P}_2(M), W_2)$ built over a Riemannian manifold, and abstract metric and metric-measure spaces, in particular those with Ricci curvature bounded below.

Not all of my works are reported/synthesized here, but only the most recent ones ($\sim$ last 2 years). For instance I won't discuss the Hahn-Banach theorem on the Wasserstein space (proved in [8]), and, most importantly, I won't describe the results of the book [3] which I wrote during the PhD together with Ambrosio and Savaré. Still, some results of this book are mentioned, when they deserve as starting point for further analysis.

Concerning the presentation style, I will not give details about the rigorous proof of the various statements, unless the proof itself contains ideas which I believe worth of notice. In other words, I will try to convey only the main concepts and tools behind the theorems, leaving aside technicalities.

The first Chapter is devoted to the study of the geometry of the Wasserstein space, which has been for me a natural continuation of the studies done during the PhD. Indeed, in the book [3] we gave a rigorous description of various concepts introduced by Otto in his seminal work [39], where he described how the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ can be viewed as a sort of Riemannian manifold. We used this point of view to implement a general theory of gradient flows on $(\mathscr{P}_2(\mathbb{R}^d), W_2)$, linking it to the study of certain parabolic PDEs. What I did after the work on the book was concluded, has been to understand how far the interpretation of the Wasserstein space as Riemannian manifold can be pushed.

In the second Chapter I discuss some recent examples of PDEs which can be seen as gradient flows.

Chapters from the third to the last are related, in a very broad sense, to the study of the heat flow on a general metric measure space, and in particular on those having Ricci curvature bounded from below in the sense of Lott-Sturm-Villani ($CD(K,\infty)$ spaces in short), which has been my main research subject in the last year. The results presented are partly obtained by myself only, and partly in collaboration with other authors (Kuwada and Ohta on one side [14] and Ambrosio and Savaré on the other [4], [5]).

More in detail, in the third chapter I study the relation between the Hopf-Lax formula and the Hamilton-Jacobi equation, showing that the former produces solutions to the latter in a purely metric setting. This result has been inspired by a work of Lott and Villani [32], where they proved an analogous statement in the setting of metric measure spaces having doubling and Poincaré. What turned out, is that actually the variational structure of the Hopf-Lax formula allows for arguments independent on the given measure, which therefore lead to more general results. Surprisingly, the study of the Hamilton-Jacobi equation is a key tool for understanding the behavior of the heat flow in a metric-measure setting.

Chapter 4 describes why it is reasonable to define the heat flow on $CD(K,\infty)$ spaces as gradient flow of the relative entropy. Indeed it is well known that this is a good definition on a smooth setting, but on the abstract case it is a priori unclear why such gradient flows exists and is unique. A general stability result for gradient flows is also given, which implies in particular that the heat flow on $CD(K,\infty)$ spaces is stable under measured-Gromov-Hausdorff convergence.

Chapter 5 contains, in a pretty detailed way, a new description of Sobolev spaces over metric measure spaces, where tools and ideas coming from optimal transport as well as the study of the Hamilton-Jacobi equation play a key role. This new approach coincides with the previous well known construction, but the new tools allow for a finer description of the Sobolev space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$, allowing for instance to prove that Lipschitz functions are dense in energy, regardless of any assumption on the metric measure space $(X, \mathsf{d}, \mathfrak{m})$.

Chapter 6 contains the proof of my main research achievement of the last year: in a general $CD(K, \infty)$ space $(X, \mathsf{d}, \mathfrak{m})$ the gradient flow of the relative entropy in $(\mathscr{P}_2(X), W_2)$ produces the same evolution of the gradient flow of the natural Dirichlet energy in L^2. This is in complete accordance with what happens in a smooth setting, where it was already well known that the heat flow can be regarded as any of these two gradient flows.

Finally, in Chapter 7 I present, without any proof, the results of a very recent work where it is proposed a new definition of Ricci curvature bounds on metric measure spaces which rules out Finsler geometries.

1 Structure of the Wasserstein space

One of the major breakthrough in the study of the optimal transport problem has been done by Otto in [39], where he described how the Wasserstein space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ can be viewed as a sort of infinite dimensional Riemannian manifold. The understanding of how far Riemannian calculus on this space can be pushed has been a topic on which I invested a lot of energies. The first results in this direction are those contained in the monograph [3] which I wrote together with Ambrosio and Savaré. In this work, we needed to develop a rigorous first order calculus to give a meaning to notions like 'subdifferential of a (geodesically) convex functional' and develop a general theory to link the study of gradient flows on $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ with that of certain parabolic PDEs. Some of the results in [3] are recalled in the introduction below, as they deserve as starting point for the further investigation that I carried on after my PhD.

Beside the introduction below, this chapter is split into three sections. In the first one I describe the precise structure of the tangent cone of $(\mathscr{P}_2(M), W_2)$ at a general measure μ, explaing in particular under which conditions it is an Hilbert space (published in [12]).

In the second one I present an overview about second order calculus over the 'manifold' $(\mathscr{P}_2(M), W_2)$, i.e. I will introduce the covariant derivative, compute it for some class of vector fields, discuss the problem of existence of parallel transport and introduce the curvature tensor. Much more could be said on this topic (e.g. one can rigorously describe the differential of the exponential map and the Jacobi equation, showing that the former produces solutions to the latter, in accordance with the smooth case), but I preferred to give here just an introduction to the subject, in the attempt of conveying the basic ideas, rather than drilling down as much results as possible. Full development of the second order calculus is contained in the monograph [10]. I remark that the results contained in [12] and [10] answered problems raised by Villani in his book [47], where he asked both for a deeper understanding of the structure of $(\mathscr{P}_2(M), W_2)$ as a manifold and also up to what extent a rigorous second order calculus could have been developed.

In the last section I will present some 'spot' results about the structure of $(\mathscr{P}_2(M), W_2)$, which I think add value to the overall clarification of the geometry of the Wasserstein space.

1.1 Reminders

This introduction is to make a quick overview about what is meant by 'weak Riemannian structure' of the space $(\mathscr{P}_2(M), W_2)$, statements are taken from the book [3] which I wrote with Ambrosio and Savaré during my PhD.

I shall start with some heuristic considerations about geodesics: let $M = \mathbb{R}^d$ and (μ_t) be a constant speed geodesic on $\mathscr{P}_2(\mathbb{R}^d)$ induced by some optimal map T, i.e.:

$$\mu_t = \big((1-t)\mathrm{Id} + tT\big)_\# \mu_0.$$

Then a simple calculation shows that (μ_t) satisfies the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \nabla\cdot(v_t\mu_t) = 0,$$

with $v_t := (T - \mathrm{Id}) \circ ((1-t)\mathrm{Id} + tT)^{-1}$ for every t, in the sense of distributions. Indeed for $\phi \in C_c^\infty(\mathbb{R}^d)$ it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \phi\,\mathrm{d}\mu_t = \frac{\mathrm{d}}{\mathrm{d}t}\int \phi\big((1-t)\mathrm{Id}+tT\big)\,\mathrm{d}\mu_0 = \int \langle \nabla\phi\big((1-t)\mathrm{Id}+tT\big), T-\mathrm{Id}\rangle\,\mathrm{d}\mu_0 = \int \langle \nabla\phi, v_t\rangle\,\mathrm{d}\mu_t.$$

Now, the continuity equation describes the link between the motion of the continuum μ_t and the instantaneous velocity $v_t : \mathbb{R}^d \to \mathbb{R}^d$ of every "atom" of μ_t. It is therefore natural to think at the vector field v_t as the infinitesimal variation of the continuum μ_t.

From this perspective, one might expect that the set of "smooth" curves on $\mathscr{P}_2(\mathbb{R}^d)$ (and more generally on $\mathscr{P}_2(M)$) is somehow linked to the set of solutions of the continuity equation. This is actually the case, as I'm going to recall now.

Since $(\mathscr{P}_2(M), W_2)$ is just a metric space, there is no a priori notion of smoothness for a curve: the best one can do is to speak about *absolutely continuous curves*. Recall that a curve (x_t) with values in a metric space (X, d_X) is said absolutely continuous if there exists a function $f \in L^1(0,1)$ such that

$$\mathsf{d}_X(x_t, x_s) \leq \int_t^s f(r)\,\mathrm{d}r, \qquad \forall t < s \in [0,1]. \tag{1.1}$$

In this case, for a.e. t the *metric derivative* $|\dot{x}_t|$ exists, given by

$$|\dot{x}_t| := \lim_{h \to 0} \frac{\mathsf{d}_X(x_{t+h}, x_t)}{|h|}, \tag{1.2}$$

and $|\dot{x}_t| \in L^1(0,1)$ and is the smallest L^1 function (up to negligible sets) for which inequality (1.1) is satisfied.

The link between absolutely continuous curves in $\mathscr{P}_2(M)$ and the continuity equation is given by the following theorem, where I will write $\|v\|_\mu$ for the norm of the vector field v in $L^2(\mu)$.

Theorem 1.1 (Characterization of absolutely continuous curves) *Let M be a smooth complete Riemannian manifold without boundary. Then the following holds.*
*(**A**) For every absolutely continuous curve $(\mu_t) \subset \mathscr{P}_2(M)$ there exists a Borel family of vector fields v_t on M such that $\|v_t\|_{L^2(\mu_t)} \leq |\dot{\mu}_t|$ for a.e. t and the continuity equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \nabla \cdot (v_t \mu_t) = 0, \tag{1.3}$$

holds in the sense of distributions.
*(**B**) Conversely, if (μ_t, v_t) satisfies the continuity equation* (1.3) *in the sense of distributions and $\int_0^1 \|v_t\|_{L^2(\mu_t)}\,\mathrm{d}t < \infty$, then up to redefining (μ_t) on a negligible set of times, (μ_t) is an absolutely continuous curve on $\mathscr{P}_2(M)$ and $|\dot{\mu}_t| \leq \|v_t\|_{L^2(\mu_t)}$ for a.e. $t \in [0,1]$.*

The first intuition about a connection between the Wasserstein distance and the continuity equation is due to Benamou and Brenier ([17]), and was further investigated by Otto in his paper [39]. The rigorous statement of the theorem as presented here was proved in [3][1]. Let me remark that in this theorem there are no regularity assumptions on the μ_t's.

Starting from Theorem 1.1 a reasonably rich first order calculus on $(\mathscr{P}_2(M), W_2)$ can be developed, I recall below some key statement.

A first consequence is the *Benamou-Brenier formula*, which says that

$$W_2(\mu^0, \mu^1) = \inf\left\{\int_0^1 \|v_t\|_{\mu_t}\,\mathrm{d}t\right\}, \tag{1.4}$$

[1]actually, in [3] the theorem was stated in the case $M = \mathbb{R}^d$, but the generalization to the case of manifolds is straightforward, as this is a first order statement and as such cannot see the presence of curvature. A rigorous proof of the case of Riemannian manifold starting from the Euclidean one follows, for instance, by an application of Nash's embedding theorem (see e.g. [2] for more details).

where the infimum is taken among all weakly continuous distributional solutions of the continuity equation (μ_t, v_t) such that $\mu_0 = \mu^0$ and $\mu_1 = \mu^1$.

The proof of (1.4) follows directly from Theorem 1.1: the inequality $\geq$ is a consequence of (**A**) (considering a geodesic connecting μ_0 to μ_1), and $\leq$ of (**B**).

An interesting feature of equation (1.4) is that it relates the 'static' optimal transport problem, to the 'dynamic' problem of interpolating between measures with absolutely continuous curves.

The formula also suggests that the scalar product in $L^2(\mu)$ should be considered as the metric tensor on $\mathscr{P}_2(M)$ at μ, so that (1.4) tells nothing but the fact that the Wasserstein distance is the Riemannian distance associated to the scalar product in $L^2(\mu)$. Yet, to give a more precise meaning this statement, there is another important remark to make: given an absolutely continuous curve $(\mu_t) \subset \mathscr{P}_2(M)$, in general there is no unique choice of vector field (v_t) such that the continuity equation (1.3) is satisfied. Indeed, if (1.3) holds and w_t is any Borel family of vector fields such that $\nabla \cdot (w_t \mu_t) = 0$ for a.e. t, then the continuity equation is satisfied also with the vector fields $(v_t + w_t)$. We need then to understand whether there is some natural selection principle to associate uniquely a family of vector fields (v_t) to a given absolutely continuous curve, which we will then call the *velocity vector field* of the curve.

There are two possible approaches:

Algebraic approach. The fact that for distributional solutions of the continuity equation the vector field (v_t) acts only on gradients of smooth functions suggests that the v_t's should be taken in the set of gradients as well, or, more rigorously, v_t should belong to

$$\overline{\left\{ \nabla\varphi \ : \ \varphi \in C_c^\infty(M) \right\}}^{L^2(\mu_t)} \tag{1.5}$$

for a.e. $t \in [0, 1]$.

Variational approach. The fact that the continuity equation is linear in v_t and the L^2 norm is strictly convex, implies that there exists a unique, up to a negligible set in time, family of vector fields $v_t \in L^2(\mu_t)$, $t \in [0, 1]$, with minimal norm for a.e. t, among the vector fields compatible with the curve (μ_t) via the continuity equation. It is immediate to verify that v_t is of minimal norm if and only if it belongs to the set

$$\left\{ v \in L^2(\mu_t) \ : \ \int \langle v, w \rangle \, \mathrm{d}\mu_t = 0, \ \forall w \in L^2(\mu_t) \ s.t. \ \nabla \cdot (w\mu_t) = 0 \right\}. \tag{1.6}$$

The important point here is that the sets defined by (1.5) and (1.6) are the same, as it is easy to check. Therefore it is natural to give the following definition.

Definition 1.2 (The tangent space) *Let $\mu \in \mathscr{P}_2(M)$. The tangent space $\mathrm{Tan}_\mu \mathscr{P}_2(M)$ at $\mathscr{P}_2(M)$ in μ is defined as*

$$\begin{aligned}
\mathrm{Tan}_\mu \mathscr{P}_2(M) :=& \overline{\left\{ \nabla\varphi \ : \ \varphi \in C_c^\infty(M) \right\}}^{L^2(\mu)} \\
=& \left\{ v \in L^2(\mu) \ : \ \int \langle v, w \rangle \, \mathrm{d}\mu = 0, \ \forall w \in L^2(\mu) \ s.t. \ \nabla \cdot (w\mu) = 0 \right\}
\end{aligned}$$

Thus we now have a definition of tangent space for every $\mu \in \mathscr{P}_2(M)$ and this tangent space is naturally endowed with a scalar product: the one of $L^2(\mu)$. This fact, Theorem 1.1 and formula (1.4) are the bases of the so-called weak Riemannian structure of $(\mathscr{P}_2(M), W_2)$.

I shall conclude recalling some properties of $(\mathscr{P}_2(M), W_2)$ which resemble those of a Riemannian manifold. For simplicity, I will deal with the case $M = \mathbb{R}^d$ only and I will assume that the measures I'm dealing with are absolutely continuous, so that optimal maps exist. Still, I remark that analogous statements hold without these assumptions.

Thus in the next three propositions (μ_t) is an absolutely continuous curve in $\mathscr{P}_2(\mathbb{R}^d)$ such that $\mu_t \ll \mathcal{L}^d$ for every t, and (v_t) is the unique, up to a negligible set of times, family of vector fields such that the continuity equation holds and $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_2(\mathbb{R}^d)$ for a.e. t.

Proposition 1.3 (v_t can be recovered by infinitesimal displacement) *Let (μ_t) and (v_t) as above. Also, let T_t^s be the optimal transport map from μ_t to μ_s. Then for a.e. $t \in [0,1]$ it holds*

$$v_t = \lim_{s \to t} \frac{T_t^s - \mathrm{Id}}{s - t},$$

the limit being understood in $L^2(\mu_t)$.

Proposition 1.4 ("Displacement tangency") *Let (μ_t) and (v_t) as above. Then for a.e. $t \in [0,1]$ it holds*

$$\lim_{h \to 0} \frac{W_2\big(\mu_{t+h}, (\mathrm{Id} + hv_t)_\# \mu_t\big)}{h} = 0. \tag{1.7}$$

Proposition 1.5 (Derivative of the squared distance) *Let (μ_t) and (v_t) as above and $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. Then for a.e. $t \in [0,1]$ it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{W_2^2(\mu_t, \nu)}{2} = -\int \langle v_t, T_t - \mathrm{Id} \rangle \, \mathrm{d}\mu_t,$$

where T_t is the (unique) optimal transport map from μ_t to ν.

1.2 First order structure

We just discussed the definition of tangent space at a certain measure $\mu \in \mathscr{P}_2(M)$, as 'space of gradients'. As highlighted by Propositions 1.3, 1.4 and 1.5, this notion is coupled with several results which are reminiscent of those valid in a genuine Riemannian context. Still, both from theoretical and practical reasons the space $\mathrm{Tan}_\mu\mathscr{P}_2(M)$ can be sometime inadequate because 'too small': consider for instance the case $\mu = \delta_x$, then it is immediate to check that $\mathrm{Tan}_{\delta_x}\mathscr{P}_2(M) = \mathbb{R}^{\dim M}$. Thus the 'infinite dimensional manifold' $(\mathscr{P}_2(M), W_2)$ has a tangent space which is finite dimensional, which is - at least - bizarre. It is also evident that in this case $\mathrm{Tan}_{\delta_x}\mathscr{P}_2(M)$ does not reasonably describe the structure of $\mathscr{P}_2(M)$ close to δ_x. Another issue is that when studying the subdifferential of geodesically convex functionals on $\mathscr{P}_2(M)$, in order to obtain its weak-strong closure (which is the key tool from where to start any sort of analysis), one soon realizes that at least for singular measures some enlargement of $\mathrm{Tan}_\mu\mathscr{P}_2(M)$ is needed.

It is mainly for this latter reason that in the book [3] we introduced the *geometric tangent space* $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ as the tangent cone built over the geodesic metric space $\mathscr{P}_2(M)$. The definition is the following. Let the set Geod_μ be defined by:

$$\mathrm{Geod}_\mu := \left\{ \begin{array}{l} \text{constant speed geodesics starting from } \mu \\ \text{and defined on some interval of the kind } [0,T] \end{array} \right\} \Big/ \approx,$$

where we say that $(\mu_t) \approx (\mu'_t)$ provided they coincide on some right neighborhood of 0. We then endow Geod_μ with the distance D defined by:

$$D\big((\mu_t),(\mu'_t)\big) := \overline{\lim_{t\downarrow 0}} \frac{W_2(\mu_t,\mu'_t)}{t}. \tag{1.8}$$

The *Geometric Tangent space* $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is then defined as the completion of Geod_μ w.r.t. the distance D.

It is natural to ask the following question: what is the relation between the "space of gradients" $\mathrm{Tan}_\mu\mathscr{P}_2(M)$ and the "space of directions" $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$?

In order to answer this question, start recalling that given $\varphi \in C_c^\infty(M)$, the map $t \mapsto (\exp(t\nabla\varphi))_\#\mu$ is a constant speed geodesic on a right neighborhood of 0. This means that there is a natural map ι_μ from the set $\{\nabla\varphi : \varphi \in C_c^\infty(M)\}$ into Geod_μ, and therefore into $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$, which sends $\nabla\varphi$ into the (equivalence class of the) geodesic $t \mapsto (\exp(t\nabla\varphi))_\#\mu$.

In [12] I proved the following result, which answer to the previous question.

Theorem 1.6 (The tangent space) *Let $\mu \in \mathscr{P}_2(M)$. Then:*

- *the $\overline{\lim}$ in (1.8) is always a limit,*
- *the metric space $(\mathbf{Tan}_\mu(\mathscr{P}_2(M)), D)$ is complete and separable,*
- *the map $\iota_\mu : \{\nabla\varphi\} \to \mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is an injective isometry, where on the source space we put the L^2 distance w.r.t. μ. Thus, ι_μ always extends to a natural isometric embedding of $\mathrm{Tan}_\mu\mathscr{P}_2(M)$ into $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$.*

Furthermore, the following statements are equivalent:

i) the space $(\mathbf{Tan}_\mu(\mathscr{P}_2(M)), D)$ is an Hilbert space,

ii) the map $\iota_\mu : \mathrm{Tan}_\mu\mathscr{P}_2(M) \to \mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is surjective,

iii) the measure μ has the following property. For any measure $\nu \in \mathscr{P}_2(M)$ there exists a unique optimal plan from μ to ν and this plan is induced by a map.

The proof of this result is rather technical; I will omit it, focussing instead on the implications of the second part of the theorem.

Notice that *(iii)* is telling that there is a strict link between the properties of $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$, which is an object related to the local structure of $\mathscr{P}_2(M)$ around μ, and the ones of the transport problem, which is a global problem.

Also, to further drill down the topic one may ask under which conditions *(iii)* is indeed true. Brenier-McCann's theorems ensure that this is the case if $\mu \ll \mathrm{vol}$, and further investigations (see for instance [34]) revealed that it is sufficient to ask that $\mu(E) = 0$ for any Lipschitz hypersurface $E \subset M$. Actually, to find a condition which is both sufficient and necessary for *(iii)* to be true is not hard, as I remarked in [12]. For simplicity I detail it in the case $M = \mathbb{R}^d$.

The point is that we know that optimal plans are always concentrated on the subdifferential of a convex function, and viceversa for any convex function φ there exists at least a plan having μ as first marginal which is concentrated on $\partial^-\varphi$ and is therefore optimal[2]. Therefore it is easy to check that (iii) is true if and only if for any convex function φ its set of points of non differentiability (which is the same as the set of points where the subdifferential is multivalued) is μ-negligible. At this point one only needs to understand how the set of non differentiability points of a convex function is made. This problem of convex analysis has a known answer, due to Zajíček ([48]): E is the set of non differentiability points of a convex function if and only if it can be covered by countably many $c-c$ hypersurface, where a $c-c$ hypersurface is a $d-1$ dimensional surface which can be written as graph of the difference of two convex functions.

Similar arguments apply also to the case of manifolds: one defines $c-c$ hypersurface on a manifold is an hypersurface which read in an appropriate chart is the graph of the difference of two convex function, and the shows that (iii) is true if and only if $\mu(E)=0$ for any $c-c$ hypersurface E.

In summary: a measure μ is well behaved w.r.t. the optimal transport problem if and only if its geometric tangent space is an Hilbert space, and if and only if it gives no mass to $c-c$ hypersurfaces.

1.3 Second order structure

In the monograph [10], partly based on the previous paper [1] written in collaboration with Ambrosio and based on results of my PhD thesis, I investigated the second order structure of the Wasserstein space $(\mathscr{P}_2(M), W_2)$. The point being the following: we know from Otto's work that $(\mathscr{P}_2(M), W_2)$ has a sort of Riemannian structure and from the analysis made in [3] and (partly) discussed also in the previous sections we also know that this interpretation is not just formal, but can be described rigorously. It is therefore natural to question whether it is also possible to describe rigorously the second order structure of the space, if any. This amounts in finding a good definition for objects like Covariant Derivative and Curvature Tensor and in proving existence results for things like Parallel Transport and Jacobi Fields.

I've been not the only one interested in this topics: Lott wrote a paper ([30]) on the subject, too, where he computed at a formal level the Covariant Derivative and the Curvature Tensor on $\mathscr{P}_2(M)$. His point of view was not to reach the highest possible generality, but only in trying to derive the correct formulas assuming everything to be smooth. This approach allowed him to do the far reaching computations he did, but on the other hand left open the question of 'when' his computations where really justified.

His work has been done contemporarily to my PhD thesis, where I presented a 'concrete' construction of the parallel transport along a certain class of curves in the case of $\mathscr{P}_2(\mathbb{R}^d)$, showing also by an example why out of such class it may very well fail to exist. Thus he worked in the more complex case of the Wasserstein space built over a manifold and wrote down the formulas for the relevant second order objects only at a formal level, while I sticked to the technically simpler case of the Wasserstein space over $\mathbb{R}^d$, but giving an existence result. Later one, after becoming aware of Lott's work, I produced the monograph [10] whose results extend both his and mine.

Back to math. I will start describing what has been the overall strategy to develop the

[2]well, actually one should assume that $\partial^-\varphi$ is non empty μ-a.e. and impose some growth condition on it to ensure that the cost of the plan is finite, so that its second marginal belongs to $\mathscr{P}_2(\mathbb{R}^d)$ as well, but these are minor technicalities which I neglect here

second order calculus over $(\mathscr{P}_2(M), W_2)$.

On a typical course of basic Riemannian geometry, one of the first concepts introduced is that of Levi-Civita connection, which identifies the only natural ("natural" here means: "compatible with the Riemannian structure") way of differentiating vector fields on the manifold. It would therefore be natural to set up the discussion on the second order analysis on $\mathscr{P}_2(M)$ by giving the definition of Levi-Civita connection in this setting. However, this cannot be done. The reason is that we don't have a notion of smoothness for vector fields, therefore not only we don't know how to covariantly differentiate vector fields, but we don't know either which are the vector fields regular enough to be differentiated. In a purely Riemannian setting this problem does not appear, as a Riemannian manifold borns as smooth manifold on which we define a scalar product on each tangent space; but the space $\mathscr{P}_2(M)$ does not have a smooth structure (there is no diffeomorphism of a small ball around the origin in $\mathrm{Tan}_\mu \mathscr{P}_2(M)$ onto a neighborhood of μ in $\mathscr{P}_2(M)$). Thus, one has to proceed in a different way, which I describe now:

Regular curves. First of all, we drop the idea of defining a smooth vector field on the whole "manifold". We will rather concentrate on finding an appropriate definition of smoothness for vector fields defined along curves. We will see that to do this, we will need to work with a particular kind of curves, which we will call *regular*, see Definition 1.7.

Smoothness of vector fields. We will then be able to define the smoothness of vector fields defined along regular curves (Definition 1.8). Among others, a notion of smoothness of particular relevance is that of *absolutely continuous* vector fields: for this kind of vector fields we have a natural notion of *total derivative* (not to be confused with the covariant one, see Definition 1.9).

Levi-Civita connection. At this point we have all the ingredients we need to define the covariant derivative and to prove that it is the Levi-Civita connection on $\mathscr{P}_2(M)$ (Definiton 1.10 and discussion thereafter).

Parallel transport. This is the main existence result on this subject: I will prove that along regular curves the parallel transport always exists (Theorem 1.14). I will also discuss a counterexample to the existence of parallel transport along a non-regular geodesic (Example 1.15). This will show that the definition of regular curve is not just operationally needed to provide a definition of smoothness of vector fields, but is actually intrinsically related to the geometry of $\mathscr{P}_2(M)$.

Calculus of derivatives. Using the technical tools developed for the study of the parallel transport, I will explicitly compute the covariant derivatives of basic examples of vector fields.

Curvature. I conclude the discussion by showing how the concepts developed can lead to a rigorous definition of the curvature tensor on $\mathscr{P}_2(M)$.

As said at the beginning of the chapter, more can be said, but it is not the purpose of this memo to give full details. For this reason, I shall also work only in the simpler case $M = \mathbb{R}^d$, on which anyway the theory can be fully appreciated.

A word on notation: I will write $\|v\|_\mu$ and $\langle v, w\rangle_\mu$ for the norm of the vector field v and the scalar product of the vector fields v, w in the space $L^2(\mu)$ (which I will denote by L^2_μ), respectively.

I shall start with the definition of regular curve, which will play a key role in what follows.

Definition 1.7 (Regular curve) *Let (μ_t) be an absolutely continuous curve and let (v_t) be its velocity vector field, that is (v_t) is the unique vector field - up to equality for a.e. t - such* *that* $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_2(\mathbb{R}^d)$ *for a.e.* t *and the continuity equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \nabla \cdot (v_t\mu_t) = 0,$$

holds in the sense of distributions (recall Theorem 1.1 and Definition 1.2). We say that (μ_t) *is regular provided*

$$\int_0^1 \mathrm{Lip}(v_t)\,\mathrm{d}t < \infty. \tag{1.9}$$

If (μ_t) is regular, by the classical Cauchy-Lipschitz theory we know that there exists a unique family of maps $\mathbf{T}(t,s,\cdot) : \mathrm{supp}(\mu_t) \to \mathrm{supp}(\mu_s)$ satisfying

$$\begin{cases} \dfrac{d}{ds}\mathbf{T}(t,s,x) = v_s(\mathbf{T}(t,s,x)), & \forall t \in [0,1],\ x \in \mathrm{supp}(\mu_t),\ a.e.\ s \in [0,1], \\ \mathbf{T}(t,t,x) = x, & \forall t \in [0,1],\ x \in \mathrm{supp}(\mu_t). \end{cases} \tag{1.10}$$

And it is also easy to check that these maps satisfy the additional properties

$$\begin{aligned} \mathbf{T}(r,s,\cdot) \circ \mathbf{T}(t,r,\cdot) &= \mathbf{T}(t,s,\cdot) & \forall t,r,s \in [0,1], \\ \mathbf{T}(t,s,\cdot)_{\#}\mu_t &= \mu_s, & \forall t,s \in [0,1]. \end{aligned}$$

I will call this family of maps the *flow maps* of the curve (μ_t).

The introduction of regular curves is meaningful from two points of view:
Algebraic point of view. The main problem in defining the time-smoothness of a vector field (u_t) defined along a curve (μ_t) is that for different times, the vectors u_t belong to different spaces, as $L^2_{\mu_t}$ vary in time as well. If the support of μ_t is small (think for instance to the case of a measure supported on some submanifold of $\mathbb{R}^d$ which is moving in time), it is also possible that the vector field $u_t \in L^2_{\mu_t}$ is not even defined μ_s-a.e. for $s \neq t$. In this direction, the existence of the flow maps has a key role, indeed the right composition with $\mathbf{T}(t,s,\cdot)$ provides a bijective isometry from $L^2_{\mu_s}$ to $L^2_{\mu_t}$, and therefore we can read the regularity of (u_t) by looking at the regularity of $t \mapsto u_t \circ \mathbf{T}(0,t,\cdot) \in L^2_{\mu_0}$ (see Definition 1.8).
Geometric point of view. A key, and non trivial, consequence of the fact that the v_t's are Lipschitz is the fact that the *angle between tangent spaces* varies smoothly along a regular curve (see Lemma 1.13). As we will see, this will be the key enabler for the proof of existence of the Parallel Transport.

It can be proved that the set of regular curves is dense in the set of absolutely continuous curves on $\mathscr{P}_2(\mathbb{R}^d)$ with respect to uniform convergence plus convergence of length, so that to some extent 'there are many of them'. It should be also noticed that in general geodesics are not automatically regular (to see why, just consider a geodesic from a delta to a measure which is not a delta), however a consequence of the fact that 'optimal maps from intermediate points are Lipschitz' is that if (μ_t) is a geodesic on $[0,1]$, then its restriction to $[\varepsilon, 1-\varepsilon]$ is regular for any $\varepsilon > 0$.

We can now start considering vector fields along regular curves and their time regularity.

Definition 1.8 (Vector fields along a curve and time regularity) *A vector field along a curve* (μ_t) *is a Borel map* $(t,x) \mapsto u_t(x)$ *such that* $u_t \in L^2_{\mu_t}$ *for a.e.* t. *It will be denoted by* (u_t). *If* (μ_t) *is a regular curve,* $\mathbf{T}(t,s,\cdot)$ *are its flow maps and* (u_t) *is a vector field defined along it, we say that* (u_t) *is absolutely continuous provided the map*

$$t \mapsto u_t \circ \mathbf{T}(t_0,t,\cdot) \in L^2_{\mu_{t_0}}$$

is absolutely continuous for every $t_0 \in [0,1]$.

Since $u_t \circ \mathbf{T}(t_1, t, \cdot) = u_t \circ \mathbf{T}(t_0, t, \cdot) \circ \mathbf{T}(t_1, t_0, \cdot)$ and the composition with $\mathbf{T}(t_1, t_0, \cdot)$ provides an isometry from $L^2_{\mu_{t_0}}$ to $L^2_{\mu_{t_1}}$, it is sufficient to check the regularity of $t \mapsto u_t \circ \mathbf{T}(t_0, t, \cdot)$ for *some* $t_0 \in [0,1]$ to be sure that the same regularity holds for every t_0. It should also be noticed that we are not requiring the vector field to be tangent.

Coupled with the definition of absolute continuity, there is the one of total derivative.

Definition 1.9 (Total derivative) *With the same notation as above, assume that (u_t) is an absolutely continuous vector field. Its total derivative is defined as:*

$$\frac{\mathbf{d}}{\mathrm{d}t} u_t := \lim_{h \to 0} \frac{u_{t+h} \circ \mathbf{T}(t, t+h, \cdot) - u_t}{h},$$

where the limit is intended in $L^2_{\mu_t}$.

Notice that the total derivative may fail to be tangent, even if (u_t) is.

It is not hard to check that the total derivative is well defined for a.e. t and that it is an L^1 vector field, in the sense that it holds $\int_0^1 \left\| \frac{\mathbf{d}}{\mathrm{d}t} u_t \right\|_{\mu_t} \mathrm{d}t < \infty$.

An important property is the *Leibniz rule*: for any couple of absolutely continuous vector fields (u_t^1), (u_t^2) along the same regular curve (μ_t) the map $t \mapsto \left\langle u_t^1, u_t^2 \right\rangle_{\mu_t}$ is absolutely continuous and it holds

$$\frac{\mathrm{d}}{\mathrm{d}t} \left\langle u_t^1, u_t^2 \right\rangle_{\mu_t} = \left\langle \frac{\mathbf{d}}{\mathrm{d}t} u_t^1, u_t^2 \right\rangle_{\mu_t} + \left\langle u_t^1, \frac{\mathbf{d}}{\mathrm{d}t} u_t^2 \right\rangle_{\mu_t}, \qquad a.e.\ t. \tag{1.11}$$

This fact follows immediately from the identity

$$\left\langle u_t^1, u_t^2 \right\rangle_{\mu_t} = \left\langle u_t^1 \circ \mathbf{T}(t_0, t, \cdot), u_t^2 \circ \mathbf{T}(t_0, t, \cdot) \right\rangle_{\mu_{t_0}}, \qquad \forall t \in [0,1].$$

Also, it should be noticed that if $(x,t) \mapsto \xi_t(x)$ is a C_c^∞ vector field on $\mathbb{R}^d$, then

$$\frac{\mathbf{d}}{\mathrm{d}t} \xi_t = \partial_t \xi_t + \nabla \xi_t \cdot v_t, \qquad a.e.\ t, \tag{1.12}$$

which shows that the total derivative is nothing but the *convective derivative* well known in fluid dynamics.

We can now introduce the *covariant derivative*. For $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, we denote by $\mathrm{P}_\mu : L^2_\mu \to \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ the orthogonal projection, and put $\mathrm{P}_\mu^\perp := \mathrm{Id} - \mathrm{P}_\mu$.

Definition 1.10 (Covariant derivative) *Let (u_t) be an absolutely continuous and* tangent *vector field along the regular curve (μ_t). Its covariant derivative is defined as*

$$\frac{\mathbf{D}}{\mathrm{d}t} u_t := \mathrm{P}_{\mu_t} \left(\frac{\mathbf{d}}{\mathrm{d}t} u_t \right). \tag{1.13}$$

Clearly, the covariant derivative is well defined for a.e. t and is an L^1 vector field.

In order to call this derivation covariant derivative, we should check that is *compatible with the metric* and *torsion free*. The first fact is a trivial consequence of the Leibniz rule (1.11), indeed if (u_t^1), (u_t^2) are tangent absolutely continuous vector fields we have:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \left\langle u_t^1, u_t^2 \right\rangle_{\mu_t} &= \left\langle \frac{\mathbf{d}}{\mathrm{d}t} u_t^1, u_t^2 \right\rangle_{\mu_t} + \left\langle u_t^1, \frac{\mathbf{d}}{\mathrm{d}t} u_t^2 \right\rangle_{\mu_t} \\ &= \left\langle \mathrm{P}_{\mu_t} \left(\frac{\mathbf{d}}{\mathrm{d}t} u_t^1 \right), u_t^2 \right\rangle_{\mu_t} + \left\langle u_t^1, \mathrm{P}_{\mu_t} \left(\frac{\mathbf{d}}{\mathrm{d}t} u_t^2 \right) \right\rangle_{\mu_t} = \left\langle \frac{\mathbf{D}}{\mathrm{d}t} u_t^1, u_t^2 \right\rangle_{\mu_t} + \left\langle u_t^1, \frac{\mathbf{D}}{\mathrm{d}t} u_t^2 \right\rangle_{\mu_t}. \end{aligned} \tag{1.14}$$

The proof of the torsion free identity is a bit more complicated, as one should at first identify what is the Lie bracket of two vector fields; I will omit it.

Given the definition of covariant derivative, the one of parallel transport follows naturally.

Definition 1.11 (Parallel transport) *Let (μ_t) be a regular curve. A tangent vector field (u_t) along it is a parallel transport if it is absolutely continuous and*

$$\frac{\mathbf{D}}{\mathrm{d}t} u_t = 0, \qquad a.e.\ t.$$

The compatibility with the metric yields that the scalar product of two parallel transports is preserved in time. In particular, this fact and the linearity of the notion of parallel transport give uniqueness of the parallel transport itself, in the sense that for any $u^0 \in \mathrm{Tan}_{\mu_0}\mathscr{P}_2(\mathbb{R}^d)$ there exists at most one parallel transport (u_t) along (μ_t) satisfying $u_0 = u^0$.

Thus the problem is to show the existence. The proof is based on geometrical arguments, and in order to understand them, there is an important analogy which I want to point out. We already know that the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ looks like a Riemannian manifold, but actually it has also stronger similarities with a Riemannian manifold M embedded in some bigger space (say, on some Euclidean space $\mathbb{R}^D$), indeed in both cases:

- we have a natural presence of non tangent vectors: elements of $L^2_\mu \setminus \mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ for $\mathscr{P}_2(\mathbb{R}^d)$, and vectors in $\mathbb{R}^D$ non tangent to the manifold for the embedded case.

- The scalar product in the tangent space can be naturally defined also for non tangent vectors: scalar product in L^2_μ for the space $\mathscr{P}_2(\mathbb{R}^d)$, and the scalar product in $\mathbb{R}^D$ for the embedded case. This means in particular that there are natural orthogonal projections from the set of tangent and non tangent vectors onto the set of tangent vectors: $\mathrm{P}_\mu : L^2_\mu \to \mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ for $\mathscr{P}_2(\mathbb{R}^d)$ and $P_x : \mathbb{R}^D \to T_xM$ for the embedded case.

- The Covariant derivative of a tangent vector field is given by projecting the "time derivative" onto the tangent space. Indeed, for the space $\mathscr{P}_2(\mathbb{R}^d)$ we defined the covariant derivative via the formula (1.13), while for the embedded manifold it holds:

  $$\nabla_{\dot\gamma_t} u_t = P_{\gamma_t}\left(\frac{\mathrm{d}}{\mathrm{d}t} u_t\right), \tag{1.15}$$

  where $t \mapsto \gamma_t$ is a smooth curve and $t \mapsto u_t \in T_{\gamma_t}M$ is a smooth tangent vector field.

Given these analogies, I shall first give a proof of the existence of the parallel transport along a smooth curve in an embedded Riemannian manifold, then I will describe how to reproduce the same arguments in the Wasserstein space. Clearly, for the case of Riemannian manifolds the arguments that I will describe are not the natural way, nor the shortest, of proving existence of the parallel transport; yet they are interesting for the discussion here because they can be adapted to the Wasserstein space.

Example 1.12 (Parallel transport on an embedded manifold) Let M be a given smooth Riemannian manifold embedded on $\mathbb{R}^D$, $t \mapsto \gamma_t \in M$ a smooth curve on $[0,1]$ and $u^0 \in T_{\gamma_0}M$

a given tangent vector. Our goal is to prove the existence of an absolutely continuous vector field $t \mapsto u_t \in T_{\gamma_t}M$ such that $u_0 = u^0$ and

$$P_{\gamma_t}\left(\frac{\mathrm{d}}{\mathrm{d}t}u_t\right) = 0, \qquad a.e.\ t.$$

For any $t, s \in [0,1]$, let $\mathrm{tr}_t^s : T_{\gamma_t}\mathbb{R}^D \to T_{\gamma_s}\mathbb{R}^D$ be the natural translation map which takes a vector with base point γ_t (tangent or not to the manifold) and gives back the translated of this vector with base point γ_s. Notice that an effect of the curvature of the manifold and the chosen embedding on $\mathbb{R}^D$, is that $\mathrm{tr}_t^s(u)$ may be not tangent to M even if u is. Now define $P_t^s : T_{\gamma_t}\mathbb{R}^D \to T_{\gamma_s}M$ by

$$P_t^s(u) := P_{\gamma_s}(\mathrm{tr}_t^s(u)), \qquad \forall u \in T_{\gamma_t}\mathbb{R}^D.$$

An immediate consequence of the smoothness of M and γ are the two inequalities:

$$|\mathrm{tr}_t^s(u) - P_t^s(u)| \le C|u||s-t|, \qquad \forall t,\ s \in [0,1] \text{ and } u \in T_{\gamma_t}M, \tag{1.16a}$$

$$|P_t^s(u)| \le C|u||s-t|, \qquad \forall t,\ s \in [0,1] \text{ and } u \in T_{\gamma_t}^{\perp}M, \tag{1.16b}$$

where $T_{\gamma_t}^{\perp}M$ is the orthogonal complement of $T_{\gamma_t}M$ in $T_{\gamma_t}\mathbb{R}^D$. The existence proof that I will provide is based on these two inequalities only.

The idea is to produce for any partition of $[0,1]$ an approximation of the parallel transport and check that when the partition becomes finer, this approximation converges. More precisely, given a partition $\mathcal{P} = \{0 = t_0 < t_1 \cdots < t_N = 1\}$ of $[0,1]$ and $u \in T_{\gamma_0}(M)$, we define $\mathcal{P}(u) \in T_{\gamma_1}M$ by

$$\mathcal{P}(u) := P_{t_{N-1}}^{t_N}(P_{t_{N-2}}^{t_{N-1}}(\cdots P_{t_0}^{t_1}(u))),$$

and the goal it to show that for any $\varepsilon > 0$ there exists a partition $\mathcal{P}$ such that for any partition $\mathcal{Q}$ finer than $\mathcal{P}$ it holds $|\mathcal{P}(u) - \mathcal{Q}(u)| \le \varepsilon|u|$.

The fact that this procedure approximates the parallel transport can be heuristically deduced by the identity

$$\nabla_{\gamma_t} P_0^t(u)_{|t=0} = 0, \quad \forall u \in T_{\gamma_0}M, \tag{1.17}$$

which tells that the vectors $P_0^t(u)$ are a first order approximation at $t = 0$ of the parallel transport. To prove (1.17) we notice that taking (1.15) into account, (1.17) is equivalent to

$$|P_t^0(\mathrm{tr}_0^t(u) - P_0^t(u))| = o(t), \qquad u \in T_{\gamma(0)}M, \tag{1.18}$$

and thus equation (1.18) follows by applying inequalities (1.16) (note that $\mathrm{tr}_0^t(u) - P_0^t(u) \in T_{\gamma_t}^{\perp}M$):

$$|P_t^0(\mathrm{tr}_0^t(u) - P_0^t(u))| \le Ct|\mathrm{tr}_0^t(u) - P_0^t(u)| \le C^2t^2|u|. \tag{1.19}$$

The comparison of $\mathcal{P}(u)$ and $\mathcal{Q}(u)$ is based on the inequality

$$\left|P_{s_1}^{s_3}(u) - P_{s_2}^{s_3}(P_{s_1}^{s_2}(u))\right| \le C^2|u||s_1 - s_2||s_2 - s_3|, \quad \forall u \in T_{\gamma_{s_1}}M,$$

valid for any $0 \le s_1 \le s_2 \le s_3 \le 1$, whose proof is similar to that of (1.19). Then, an easy induction shows that for any $0 \le s_1 < \cdots < s_N \le 1$ and $u \in T_{\gamma_{s_1}}M$ we have

$$\begin{aligned}
& \left|P_{s_1}^{s_N}(u) - P_{s_{N-1}}^{s_N}(P_{s_{N-2}}^{s_{N-1}}(\cdots(P_{s_1}^{s_2}(u))))\right| \\
\le\ & \left|P_{s_1}^{s_N}(u) - P_{s_{N-1}}^{s_N}(P_{s_1}^{s_{N-1}}(u))\right| + \left|P_{s_{N-1}}^{s_N}(P_{s_1}^{s_{N-1}}(u)) - P_{s_{N-1}}^{s_N}(P_{s_{N-2}}^{s_{N-1}}(\cdots(P_{s_1}^{s_2}(u))))\right| \\
\le\ & C^2|u||s_{N_1} - s_1|!s_N - s_{N-1}| + \left|P_{s_1}^{s_{N-1}}(u) - P_{s_{N-2}}^{s_{N-1}}(\cdots(P_{s_1}^{s_2}(u)))\right| \\
\le\ & \cdots \\
\le\ & C^2|u|\sum_{i=2}^{N-1}|s_1 - s_i||s_i - s_{i+1}| \le C^2|u||s_1 - s_N|^2. \qquad (1.20)
\end{aligned}$$

With this result, it is immediate to check that the limit of $\mathcal{P}(u)$ as $\mathcal{P}$ becomes finer exists: given $\varepsilon > 0$, it is sufficient to pick $\mathcal{P} := \{0 = t_0 < t_1 \cdots < t_N = 1\}$ such that $\sum_i |t_i - t_{i-1}|^2 < \varepsilon/C^2$ to conclude that if $\mathcal{Q}$ is finer than $\mathcal{P}$ it holds $|\mathcal{P}(u) - \mathcal{Q}(u)| \le \varepsilon|u|$ (just repeatedly apply (1.20) to the various partitions induced by $\mathcal{Q}$ on the intervals $[t_i, t_{i+1}]$).

Thus for any $u \in T_{\gamma_0}M$ this process produces a limit vector in $T_{\gamma_1}M$, which means that we built a map T_0^1 from $T_{\gamma_0}M$ to $T_{\gamma_1}M$. The same procedure applied to the restriction of (γ_t) to the interval $[t, s]$ gives a map $T_t^s : T_{\gamma_t}M \to T_{\gamma_s}M$. It is then easy to verify, starting from (1.17), that for any $u \in T_{\gamma_0}M$ the vector field $t \mapsto T_0^t(u)$ is precisely the parallel transport of u along (γ_t). ■

Now back to the Wasserstein case. To follow the analogy with the Riemannian case, keep in mind that the analogous of the translation map tr_t^s is the right composition with $\mathbf{T}(s, t, \cdot)$, and the analogous of the map P_t^s is

$$\mathscr{P}_t^s(u) := \mathrm{P}_{\mu_s}(u \circ \mathbf{T}(s, t, \cdot)),$$

which maps $L^2_{\mu_t}$ onto $\mathrm{Tan}_{\mu_s}\mathscr{P}_2(\mathbb{R}^d)$. We saw that the key to prove the existence of the parallel transport in the embedded Riemannian case are inequalities (1.16). Thus, given that we want to imitate the approach in the Wasserstein setting, we need to produce an analogous of those inequalities. This is the content of the following key lemma, where I will denote by $\mathrm{Tan}_\mu^\perp(\mathscr{P}_2(\mathbb{R}^d))$ the Normal space at μ, i.e. the orthogonal complement of $\mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ in L^2_μ.

Lemma 1.13 (Control of the angles between tangent spaces) *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $T : \mathbb{R}^d \to \mathbb{R}^d$ be any Borel map satisfying $T_\#\mu = \nu$. Then it holds:*

$$\|v \circ T - \mathrm{P}_\mu(v \circ T)\|_\mu \le \|v\|_\nu \,\mathrm{Lip}(T - \mathrm{Id}), \qquad \forall v \in \mathrm{Tan}_\nu\mathscr{P}_2(\mathbb{R}^d),$$

and, if T is invertible, it also holds

$$\|\mathrm{P}_\mu(w \circ T)\|_\mu \le \|w\|_\nu \,\mathrm{Lip}(T^{-1} - \mathrm{Id}), \qquad \forall w \in \mathrm{Tan}_\nu^\perp(\mathscr{P}_2(\mathbb{R}^d)).$$

To see why this lemma is true, assume for simplicity that $T - \mathrm{Id} \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$ and notice that the first inequality is equivalent to

$$\|\nabla\varphi \circ T - \mathrm{P}_\mu(\nabla\varphi \circ T)\|_\mu \le \|\nabla\varphi\|_\nu \,\mathrm{Lip}(T - \mathrm{Id}), \qquad \forall \varphi \in C_c^\infty(\mathbb{R}^d). \qquad (1.21)$$

Thus what we have to do is to find a function whose gradient is not too far from $\nabla\varphi\circ T$ in L^2_μ. Let's consider the function $\varphi\circ T$: since $\varphi\circ T\in C^\infty_c(\mathbb{R}^d)$,we have $\nabla(\varphi\circ T)=\nabla T\cdot(\nabla\varphi)\circ T\in \mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ and thus

$$\begin{aligned}\|\nabla\varphi\circ T-\mathrm{P}_\mu(\nabla\varphi\circ T)\|_\mu &\le \|\nabla\varphi\circ T-\nabla T\cdot(\nabla\varphi)\circ T\|_\mu\\ &=\left(\int |(I-\nabla T(x))\cdot\nabla\varphi(T(x))|^2\,\mathrm{d}\mu(x)\right)^{1/2}\\ &\le\left(\int |\nabla\varphi(T(x))|^2\|\nabla(\mathrm{Id}-T)(x)\|^2_{op}\,\mathrm{d}\mu(x)\right)^{1/2}\\ &\le \|\nabla\varphi\|_\nu\,\mathrm{Lip}(T-\mathrm{Id}),\end{aligned}$$

where I is the identity matrix and $\|\nabla(\mathrm{Id}-T)(x)\|_{op}$ is the operator norm of the linear functional from $\mathbb{R}^d$ to $\mathbb{R}^d$ given by $v\mapsto\nabla(\mathrm{Id}-T)(x)\cdot v$.

The rest of the proof follows by a pretty standard approximation procedure, and the second inequality comes from a duality argument.

From this lemma and the readily checked inequality

$$\mathrm{Lip}\Big(\mathbf{T}(s,t,\cdot)-\mathrm{Id}\Big)\le e^{\left|\int_t^s\mathrm{Lip}(v_r)\,\mathrm{d}r\right|}-1\le C\left|\int_t^s\mathrm{Lip}(v_r)\,\mathrm{d}r\right|,\qquad \forall t,s\in[0,1],$$

where $C:=e^{\int_0^1\mathrm{Lip}(v_r)\,\mathrm{d}r}-1$, it is immediate to verify that it holds:

$$\begin{aligned}\|u\circ\mathbf{T}(s,t,\cdot)-\mathscr{P}_t^s(u)\|_{\mu_s}&\le C\|u\|_{\mu_t}\left|\int_t^s\mathrm{Lip}(v_r)\,\mathrm{d}r\right|,\qquad u\in\mathrm{Tan}_{\mu_t}\mathscr{P}_2(\mathbb{R}^d),\\ \|\mathscr{P}_t^s(u)\|_{\mu_s}&\le C\|u\|_{\mu_t}\left|\int_t^s\mathrm{Lip}(v_r)\,\mathrm{d}r\right|,\qquad u\in\mathrm{Tan}^\perp_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)).\end{aligned}\tag{1.22}$$

These inequalities are perfectly analogous to the (1.16) (well, the only difference is that here the bound on the angle is L^1 in t,s while for the embedded case it was L^∞, but this does not really change anything). Therefore the arguments presented before apply also to this case, and we can derive the existence of the parallel transport along regular curves:

Theorem 1.14 (Parallel transport along regular curves) *Let (μ_t) be a regular curve and $u^0\in\mathrm{Tan}_{\mu_0}\mathscr{P}_2(\mathbb{R}^d)$. Then there exists a parallel transport (u_t) along (μ_t) such that $u_0=u^0$.*

Now we know that the parallel transport exists along regular curves, and we know also that regular curves are dense, it is therefore natural to try to construct the parallel transport along any absolutely continuous curve via some limiting argument. However, this cannot be done, as the following counterexample shows.

Example 1.15 (Non existence of parallel transport along a non regular geodesic) Let $Q=[0,1]\times[0,1]$ be the unit square in $\mathbb{R}^2$ and let T_i, $i=1,2,3,4$, be the four open triangles in which Q is divided by its diagonals. Let $\mu_0:=\chi_Q\mathscr{L}^2$ and define the function $v:Q\to\mathbb{R}^2$ as the gradient of the convex map $\max\{|x|,|y|\}$, as in the figure. Set also $w=v^\perp$, the rotation by $\pi/2$ of v, in Q and $w=0$ out of Q. Notice that $\nabla\cdot(w\mu_0)=0$.

Set $\mu_t:=(\mathrm{Id}+tv)_\#\mu_0$ and observe that, for positive t, the support Q_t of μ_t is made of 4 connected components, each one the translation of one of the sets T_i, and that $\mu_t=\chi_{Q_t}\mathscr{L}^2$.

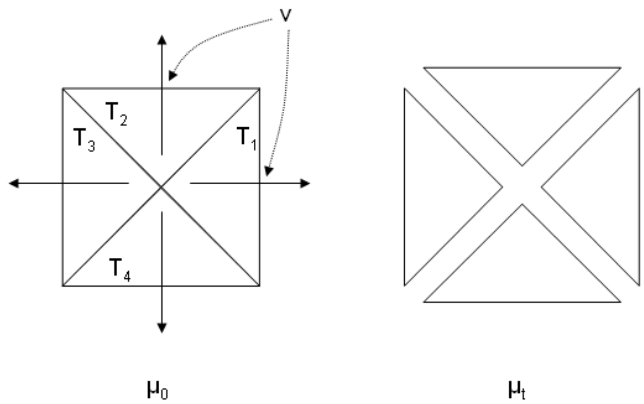

Figure 1: The support of μ_0 is split along the curve

It is immediate to check that (μ_t) is a regular geodesic on $[\varepsilon, 1]$ for every $\varepsilon > 0$, but not on $[0,1]$. Fix $\varepsilon > 0$ and note that, by construction, the flow maps of μ_t in $[\varepsilon, 1]$ are given by

$$\mathbf{T}(t, s, \cdot) = (\mathrm{Id} + sv) \circ (\mathrm{Id} + tv)^{-1}, \quad \forall t, s \in [\varepsilon, 1].$$

Now, set $w_t := w \circ \mathbf{T}(t, 0, \cdot)$ and notice that w_t is tangent at μ_t (because w_t is constant in the connected components of the support of μ_t, so we can define a C_c^∞ function to be affine on each connected component and with gradient given by w_t, and then use the space between the components themselves to rearrange smoothly the function). Since $w_{t+h} \circ \mathbf{T}(t, t+h, \cdot) = w_t$, we have $\frac{\mathbf{d}}{\mathrm{d}t} w_t = 0$ and a fortiori $\frac{\mathbf{D}}{\mathrm{d}t} w_t = 0$. Thus (w_t) is a parallel transport in $[\varepsilon, 1]$. Furthermore, since $\nabla \cdot (w\mu_0) = 0$, we have $w_0 = w \notin \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^2))$. Therefore there is no way to extend w_t to a continuous *tangent* vector field on the whole $[0,1]$. In particular, there is no way to extend the parallel transport up to $t = 0$. ■

To move on with the analysis, I shall show how with the tools just built it is possible to start developing a vector calculus on $(\mathscr{P}_2(\mathbb{R}^d), W_2)$. In particular, I will consider the following problem. Let (u_t) be an absolutely continuous vector field along the regular curve (μ_t) and consider the vector field $(\mathrm{P}_{\mu_t}(u_t))$. From inequalities (1.22) one can check that the latter is absolutely continuous as well. The question is: can we derive a formula which expresses its covariant derivative?

At a formal level, we expect that it holds something like

$$\frac{\mathbf{D}}{\mathrm{d}t} \mathrm{P}_{\mu_t}(u_t) = \mathrm{P}_{\mu_t}\left(\frac{\mathbf{d}}{\mathrm{d}t} u_t\right) + \left(\begin{array}{l} \text{some operator - which we may} \\ \text{think as the covariant derivative} \\ \text{of } \mathrm{P}_{\mu_t} \text{ - applied to } u_t \end{array}\right). \tag{1.23}$$

This heuristic is indeed true, and to turn it into a rigorous statement we will need to introduce the interesting tensor $\mathcal{N}_\mu$.

Start fixing $\varphi \in C_c^\infty(\mathbb{R}^d)$ and apply the Leibniz rule for the total and covariant derivatives ((1.11) and (1.14)), to get that for a.e. $t \in [0,1]$ it holds

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle u_t, \nabla\varphi \rangle_{\mu_t} = \left\langle \frac{\mathbf{d}}{\mathrm{d}t} u_t, \nabla\varphi \right\rangle_{\mu_t} + \left\langle u_t, \frac{\mathbf{d}}{\mathrm{d}t} \nabla\varphi \right\rangle_{\mu_t},$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \mathrm{P}_{\mu_t}(u_t), \nabla\varphi \rangle_{\mu_t} = \left\langle \frac{\mathbf{D}}{\mathrm{d}t} \mathrm{P}_{\mu_t}(u_t), \nabla\varphi \right\rangle_{\mu_t} + \left\langle \mathrm{P}_{\mu_t}(u_t), \frac{\mathbf{D}}{\mathrm{d}t} \nabla\varphi \right\rangle_{\mu_t}.$$

Since $\nabla\varphi \in \mathrm{Tan}_{\mu_t}\mathscr{P}_2(\mathbb{R}^d)$ for any t, it holds $\langle \mathrm{P}_{\mu_t}(u_t), \nabla\varphi\rangle_{\mu_t} = \langle u_t, \nabla\varphi\rangle_{\mu_t}$ for any $t \in [0,1]$, and thus the left hand sides of the previous equations are equal for a.e. t. Recalling formula (1.12) we have $\frac{\mathbf{d}}{\mathrm{d}t}\nabla\varphi = \nabla^2\varphi \cdot v_t$ and $\frac{\mathbf{D}}{\mathrm{d}t}\nabla\varphi = \mathrm{P}_{\mu_t}(\nabla^2\varphi\cdot v_t)$, thus from the equality of the right hand sides we obtain

$$\begin{aligned}\left\langle \frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t), \nabla\varphi\right\rangle_{\mu_t} &= \left\langle \frac{\mathbf{d}}{\mathrm{d}t}u_t, \nabla\varphi\right\rangle_{\mu_t} + \left\langle u_t, \nabla^2\varphi\cdot v_t\right\rangle_{\mu_t} - \left\langle \mathrm{P}_{\mu_t}(u_t), \mathrm{P}_{\mu_t}(\nabla^2\varphi\cdot v_t)\right\rangle_{\mu_t} \\ &= \left\langle \frac{\mathbf{d}}{\mathrm{d}t}u_t, \nabla\varphi\right\rangle_{\mu_t} + \left\langle \mathrm{P}^\perp_{\mu_t}(u_t), \mathrm{P}^\perp_{\mu_t}(\nabla^2\varphi\cdot v_t)\right\rangle_{\mu_t}.\end{aligned} \tag{1.24}$$

This formula characterizes the scalar product of $\frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t)$ with any $\nabla\varphi$ when φ varies on $C^\infty_c(\mathbb{R}^d)$. Since the set $\{\nabla\varphi\}$ is dense in $\mathrm{Tan}_{\mu_t}\mathscr{P}_2(\mathbb{R}^d)$ for any $t \in [0,1]$, the formula actually identifies $\frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t)$.

However, from this expression it is unclear what is the value of $\left\langle \frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t), w\right\rangle_{\mu_t}$ for a general $w \in \mathrm{Tan}_{\mu_t}\mathscr{P}_2(\mathbb{R}^d)$, because some regularity of $\nabla\varphi$ seems required to compute $\nabla^2\varphi\cdot v_t$. In order to better understand what the value of $\frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t)$ is, fix $t \in [0,1]$ and assume for a moment that $v_t \in C^\infty_c(\mathbb{R}^d)$. Then compute the gradient of $x \mapsto \langle \nabla\varphi(x), v_t(x)\rangle$ to obtain

$$\nabla\langle \nabla\varphi, v_t\rangle = \nabla^2\varphi\cdot v_t + \nabla v_t^{\mathrm{t}}\cdot\nabla\varphi,$$

and consider this expression as an equality between vector fields in $L^2_{\mu_t}$. Taking the projection onto the Normal space we derive

$$\mathrm{P}^\perp_{\mu_t}(\nabla^2\varphi\cdot v_t) + \mathrm{P}^\perp_{\mu_t}(\nabla v_t^{\mathrm{t}}\cdot\nabla\varphi) = 0.$$

Plugging the expression for $\mathrm{P}^\perp_{\mu_t}(\nabla^2\varphi\cdot v_t)$ into the formula for the covariant derivative we get

$$\begin{aligned}\left\langle \frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t), \nabla\varphi\right\rangle_{\mu_t} &= \left\langle \frac{\mathbf{d}}{\mathrm{d}t}u_t, \nabla\varphi\right\rangle_{\mu_t} - \left\langle \mathrm{P}^\perp_{\mu_t}(u_t), \mathrm{P}^\perp_{\mu_t}(\nabla v_t^{\mathrm{t}}\cdot\nabla\varphi)\right\rangle_{\mu_t} \\ &= \left\langle \frac{\mathbf{d}}{\mathrm{d}t}u_t, \nabla\varphi\right\rangle_{\mu_t} - \left\langle \nabla v_t\cdot\mathrm{P}^\perp_{\mu_t}(u_t), \nabla\varphi\right\rangle_{\mu_t},\end{aligned}$$

which identifies $\frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t)$ as

$$\frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t) = \mathrm{P}_{\mu_t}\left(\frac{\mathbf{d}}{\mathrm{d}t}u_t - \nabla v_t\cdot\mathrm{P}^\perp_{\mu_t}(u_t)\right), \tag{1.25}$$

consistently with (1.23).

We found this expression assuming that v_t was a smooth vector field, but given that we know that $\frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t)$ exists for a.e. t for an arbitrary regular curve (μ_t), it is realistic to believe that the expression makes sense also for general Lipschitz v_t's. The problem is that the object ∇v_t may very well be not defined μ_t-a.e. for arbitrary μ_t and Lipschitz v_t (Rademacher's theorem is of no help here, because we are not assuming the measures μ_t to be absolutely continuous w.r.t. the Lebesgue measure). To give a meaning to formula (1.25) we need to introduce a new tensor.

Definition 1.16 (The Lipschitz non Lipschitz space) *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. The set $\mathrm{LNL}_\mu \subset [L^2_\mu]^2$ is the set of couples of vector fields (u,v) such that $\min\{\mathrm{Lip}(u), \mathrm{Lip}(v)\} < \infty$, i.e. the set of couples of vectors such that at least one of them is Lipschitz.*

We say that a sequence $(u_n, v_n) \in \mathrm{LnL}_\mu$ *converges to* $(u, v) \in \mathrm{LnL}_\mu$ *provided* $\|u_n - u\|_\mu \to 0$, $\|v_n - v\|_\mu \to 0$ *and*

$$\sup_n \min\{\mathrm{Lip}(u_n), \mathrm{Lip}(v_n)\} < \infty.$$

Then it possible to prove the following theorem:

Theorem 1.17 (The Normal tensor) *Let* $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. *The map*

$$\begin{array}{rcl} \mathcal{N}_\mu(u, v) : [C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)]^2 & \to & \mathrm{Tan}_\mu^\perp(\mathscr{P}_2(\mathbb{R}^d)), \\ (u, v) & \mapsto & \mathrm{P}_\mu^\perp(\nabla u^{\mathrm{t}} \cdot v) \end{array}$$

extends uniquely to a sequentially continuous bilinear and antisymmetric map, still denoted by $\mathcal{N}_\mu$, *from* LnL_μ *in* $\mathrm{Tan}_\mu^\perp(\mathscr{P}_2(\mathbb{R}^d))$ *for which the bound*

$$\|\mathcal{N}_\mu(u, v)\|_\mu \le \min\{\mathrm{Lip}(u)\|v\|_\mu, \mathrm{Lip}(v)\|u\|_\mu\}, \tag{1.26}$$

holds.

I will not present the proof of this theorem, which anyway is based only on some careful approximation procedure. What I believe is more interesting to underline is that the theorem produces an antisymmetric tensor which is well defined as soon as at least one of the vector fields involved is Lipschitz, but not if none of them is. As far as I know, this is a pretty uncommon property.

Definition 1.18 (The operators $\mathcal{O}_v(\cdot)$ and $\mathcal{O}_v^*(\cdot)$) *Let* $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ *and* $v \in L^2_\mu$ *a Lipschitz vector field. Then the operator* $u \mapsto \mathcal{O}_v(u)$ *is defined by*

$$\mathcal{O}_v(u) := \mathcal{N}_\mu(v, u),$$

and the operator $u \mapsto \mathcal{O}_v^*(u)$ *as the adjoint of* $\mathcal{O}_v(\cdot)$, *i.e.*

$$\langle \mathcal{O}_v^*(u), w \rangle_\mu := \langle u, \mathcal{O}_v(w) \rangle_\mu, \qquad \forall w \in L^2_\mu.$$

The bound (1.26) ensures that the operator norm of $\mathcal{O}_v(\cdot)$ and $\mathcal{O}_v^*(\cdot)$ is bounded by $\mathrm{Lip}(v)$. Notice also that if v is smooth we have the representation

$$\begin{aligned} \mathcal{O}_v(u) &= \mathrm{P}_\mu^\perp(\nabla v^{\mathrm{t}} \cdot u), \\ \mathcal{O}_v^*(u) &= \nabla v \cdot \mathrm{P}_\mu^\perp(u). \end{aligned}$$

I should remark that in writing $\mathcal{O}_v(u)$, $\mathcal{O}_v^*(u)$ I'm losing the reference to the base measure μ, which certainly plays a role in the definition; this simplifies the notation and hopefully should create no confusion, as the measure I'm referring to should always be clear from the context.

The introduction of the operators $\mathcal{O}_v(\cdot)$ and $\mathcal{O}_v^*(\cdot)$ allows to give a precise meaning to formula (1.25) for general regular curves. Indeed, starting from (1.25) it is immediate to verify that it holds the formula

$$\frac{\mathbf{D}}{\mathrm{d}t}\mathrm{P}_{\mu_t}(u_t) = \mathrm{P}_{\mu_t}\left(\frac{\mathbf{d}}{\mathrm{d}t}u_t - \mathcal{O}_{v_t}^*(u_t)\right), \tag{1.27}$$

which now makes perfectly sense for a.e. t for any regular curve (μ_t).

Interestingly enough, this sort of differential calculus can be developed quite a lot, for instance one can show that it holds

$$\frac{\mathbf{d}}{\mathrm{d}t}\mathcal{O}_{v_t}(u_t) = \mathcal{O}_{\frac{\mathbf{d}}{\mathrm{d}t}v_t}(u_t) + \mathcal{O}_{v_t}\left(\frac{\mathbf{d}}{\mathrm{d}t}u_t\right) - \mathcal{O}_{v_t}\left(\mathcal{O}_{v_t}(u_t)\right) + \mathrm{P}_{\mu_t}\Big(\mathcal{O}^*_{v_t}\left(\mathcal{O}_{v_t}(u_t)\right)\Big),$$
$$\frac{\mathbf{d}}{\mathrm{d}t}\mathcal{O}^*_{v_t}(u_t) = \mathcal{O}^*_{\frac{\mathbf{d}}{\mathrm{d}t}v_t}(u_t) + \mathcal{O}^*_{v_t}\left(\frac{\mathbf{d}}{\mathrm{d}t}u_t\right) - \mathcal{O}^*_{v_t}\left(\mathcal{O}^*_{v_t}(u_t)\right) + \mathcal{O}^*_{v_t}\left(\mathcal{O}_{v_t}\left(\mathrm{P}_{\mu_t}(u_t)\right)\right),$$

as soon as the velocity vector field (v_t) of the regular curve considered satisfies $\int_0^1 \mathrm{Lip}\left(\frac{\mathbf{d}}{\mathrm{d}t}v_t\right)\,\mathrm{d}t < \infty$, An important feature of these equations is that to express the derivatives of $(\mathcal{O}_{v_t}(u_t))$ and $(\mathcal{O}^*_{v_t}(u_t))$ "no new operators appear". Therefore we can recursively calculate derivatives of any order of the vector fields $(\mathrm{P}_{\mu_t}(u_t))$, $(\mathrm{P}^\perp_{\mu_t}(u_t))$, $\mathcal{O}_{v_t}(u_t)$ and $\mathcal{O}^*_{v_t}(u_t)$, provided - of course - that we make appropriate regularity assumptions on the vector field (u_t) and on the velocity vector field (v_t). An example of result which can be proved following this direction is that the operator $t \mapsto \mathrm{P}_{\mu_t}(\cdot)$ is analytic along (the restriction of) a geodesic.

Also, the understanding of the properties of the tensor $\mathcal{N}_\mu$ allows for a precise description of the curvature tensor of the "manifold" $\mathscr{P}_2(\mathbb{R}^d)$, and I shall conclude this section describing this latter point.

Following the analogy with the Riemannian case, one could be lead to define the curvature tensor in the following way: given three vector fields $\mu \mapsto \nabla\varphi^i_\mu \in \mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$, $i = 1, \ldots, 3$, the curvature tensor $\mathbf{R}$ calculated on them at the measure μ is defined as:

$$\mathbf{R}(\nabla\varphi^1_\mu, \nabla\varphi^2_\mu)(\nabla\varphi^3_\mu) := \boldsymbol{\nabla}_{\nabla\varphi^2_\mu}(\boldsymbol{\nabla}_{\nabla\varphi^1_\mu}\nabla\varphi^3_\mu) - \boldsymbol{\nabla}_{\nabla\varphi^1_\mu}(\boldsymbol{\nabla}_{\nabla\varphi^2_\mu}\nabla\varphi^3_\mu) + \boldsymbol{\nabla}_{[\nabla\varphi^1_\mu, \nabla\varphi^2_\mu]}\nabla\varphi^3_\mu,$$

where the objects like $\boldsymbol{\nabla}_{\nabla\varphi_\mu}(\nabla\psi_\mu)$, are, heuristically speaking, the covariant derivative of the vector field $\mu \mapsto \nabla\psi_\mu$ along the vector field $\mu \mapsto \nabla\varphi_\mu$.

However, in order to give a precise meaning to the above formula, one should be sure, at least, that the derivatives he is taking exist. Such an approach is possible, but heavy: indeed, consider that one should define what are C^1 and C^2 vector fields, and in doing so he cannot just consider derivatives along curves, because he would need to be sure that "the partial derivatives have the right symmetries", otherwise there won't be those cancellations which let the above operator be a tensor.

Instead, we adopt the following strategy:

- First we calculate the curvature tensor for some very specific kind of vector fields, for which we are able to do and justify the calculations. Specifically, we will consider vector fields of the kind $\mu \mapsto \nabla\varphi$, where the function $\varphi \in C^\infty_c(\mathbb{R}^d)$ does not depend on the measure μ.

- Then we prove that the object found is actually a tensor, i.e. that its value depends only on the $\mu-$a.e. value of the considered vector fields, and not on the fact that we obtained the formula assuming that the functions φ's were independent on the measure.

- Finally, we discuss under which conditions the object found is well defined.

Pick $\varphi_1, \varphi_2, \varphi_3, \varphi_4 \in C^\infty_c(\mathbb{R}^d)$ and observe that the curve $t \mapsto (\mathrm{Id} + t\nabla\varphi_2)_\#\mu$ is a regular geodesic on an interval $[-T, T]$ for T sufficiently small. It is then immediate to verify that a vector field of the kind $(\nabla\varphi_3)$ along it is absolutely continuous, with continuous covariant derivative, and that such derivative calculated at $t = 0$ is given by $\mathrm{P}_\mu(\nabla^2\varphi_3 \cdot \nabla\varphi_2)$. Thus we can write:

$$\boldsymbol{\nabla}_{\nabla\varphi_2}\nabla\varphi_3(\mu) := \mathrm{P}_\mu(\nabla^2\varphi_3 \cdot \nabla\varphi_2). \tag{1.28}$$

By the same arguments, the vector field $(\nabla^2\varphi_3 \cdot \nabla\varphi_2)$ can be covariantly differentiated at $t = 0$ along the regular curve $t \mapsto (\mathrm{Id} + t\nabla\varphi_1)_\#\mu$, so that recalling the formula (1.27) we get

$$\boldsymbol{\nabla}_{\nabla\varphi_1}(\boldsymbol{\nabla}_{\nabla\varphi_2}\nabla\varphi_3)(\mu) = \mathrm{P}_\mu\Big(\nabla(\nabla^2\varphi_3 \cdot \nabla\varphi_2) \cdot \nabla\varphi_1 - \mathcal{O}^*_{\nabla\varphi_1}\left(\nabla^2\varphi_3 \cdot \nabla\varphi_2\right)\Big).$$

It is now just a matter of computations starting from the definition

$$\mathbf{R}(\nabla\varphi_1, \nabla\varphi_2)(\nabla\varphi_3) := \boldsymbol{\nabla}_{\nabla\varphi_2}(\boldsymbol{\nabla}_{\nabla\varphi_1}\nabla\varphi_3) - \boldsymbol{\nabla}_{\nabla\varphi_1}(\boldsymbol{\nabla}_{\nabla\varphi_2}\nabla\varphi_3) + \boldsymbol{\nabla}_{[\nabla\varphi_1,\nabla\varphi_2]}\nabla\varphi_3$$

to check that it holds

$$\begin{aligned} \langle\mathbf{R}(\nabla\varphi_1, \nabla\varphi_2)\nabla\varphi_3, \nabla\varphi_4\rangle_\mu = &\langle\mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_3), \mathcal{N}_\mu(\nabla\varphi_2, \nabla\varphi_4)\rangle_\mu \\ &- \langle\mathcal{N}_\mu(\nabla\varphi_2, \nabla\varphi_3), \mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_4)\rangle_\mu \\ &+ 2\,\langle\mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_2), \mathcal{N}_\mu(\nabla\varphi_3, \nabla\varphi_4)\rangle_\mu. \end{aligned} \tag{1.29}$$

The antisymmetry of $\mathcal{N}_\mu$ yields that $\mathbf{R}$ has all the standard symmetry properties and the first Bianchi identity, and with some work it is also possible to check that it satisfies the second Bianchi identity as well. Also, it immediately follows that $\mathbf{R}$ is actually a tensor: indeed the left hand side of equation (1.29) is a tensor w.r.t. the fourth entry, so that the claim follows from the symmetries of the right hand side.

Concerning the domain of definition of the curvature tensor, the following statement holds, whose proof follows from the properties of the normal tensor $\mathcal{N}_\mu$:

Proposition 1.19 *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. Then the curvature tensor, thought as map from $[\{\nabla\varphi\}]^4$ to $\mathbb{R}$ given by* (1.29)*, extends uniquely to a sequentially continuous map on the set of 4-ples of vector fields in L^2_μ in which at least 3 vector fields are Lipschitz, where we say that $(v^1_n, v^2_n, v^3_n, v^4_n)$ is converging to (v^1, v^2, v^3, v^4) if there is convergence in L^2_μ on each coordinate and*

$$\sup_n \mathrm{Lip}(v^i_n) < \infty,$$

for at least 3 indexes i.

Thus, in order for the curvature tensor to be well defined we need at least 3 of the 4 vector fields involved to be Lipschitz. However, for some related notion of curvature the situation simplifies. Of particular relevance is the case of sectional curvature:

Example 1.20 (The sectional curvature) If we evaluate the curvature tensor $\mathbf{R}$ on a 4-ple of vectors of the kind (u, v, u, v) and we recall the antisymmetry of $\mathcal{N}_\mu$ we obtain

$$\langle\mathbf{R}(u, v)u, v\rangle_\mu = 3\,\|\mathcal{N}_\mu(u, v)\|^2_\mu.$$

Thanks to the simplification of the formula, the value of $\langle\mathbf{R}(u, v)u, v\rangle_\mu$ is well defined as soon as either u or v is Lipschitz. That is, $\langle\mathbf{R}(u, v)u, v\rangle_\mu$ is well defined for $(u, v) \in \mathrm{L{\scriptstyle N}L}_\mu$. In analogy with the Riemannian case we can therefore define the sectional curvature $\mathbf{K}(u, v)$ at the measure μ along the directions u, v by

$$\mathbf{K}(u, v) := \frac{\langle\mathbf{R}(u, v)u, v\rangle_\mu}{\|u\|^2_\mu\|v\|^2_\mu - \langle u, v\rangle^2_\mu} = \frac{3\,\|\mathcal{N}_\mu(u, v)\|^2_\mu}{\|u\|^2_\mu\|v\|^2_\mu - \langle u, v\rangle^2_\mu}, \qquad \forall (u, v) \in \mathrm{L{\scriptstyle N}L}_\mu.$$

This expression shows that the sectional curvatures of $\mathscr{P}_2(\mathbb{R}^d)$ are positive, and provides a rigorous proof of the analogous formula found by Otto in [39] and formally computed using O'Neill formula. ■

1.4 Other results on the optimal transport problem

1.4.1 Regularity of Kantorovich potentials on non compact manifolds

It is well known that on a compact Riemannian manifold with cost=distance-squared/2, c-concave functions are Lipschitz and semiconcave, this fact being a simple consequence of the fact that the functions $x \mapsto \mathsf{d}^2(x, y)$ are uniformly Lipschitz and uniformly semiconcave in y coupled with the definition of c-concavity:

$$\varphi \text{ is } c\text{-concave if and only if} \quad \varphi(x) = \inf_y c(x, y) - \psi(y) \quad \text{for some } \psi. \tag{1.30}$$

This regularity of c-concave functions implies that optimal transport maps are a.e. differentiable, indeed McCann showed in [35] that optimal maps T can be written as

$$T(x) = \exp_x(-\nabla\varphi(x)), \tag{1.31}$$

for some c-concave function φ, called Kantorovich potential. In turn, the differentiability of optimal maps is a key tool which allows, for instance, to give a meaning to the change of variable formula

$$\eta(T(x)) = \frac{\rho(x)}{|\det \nabla T|(x)}, \qquad \mathrm{vol} - a.e.\ x$$

where ρ, η are respectively the source and target densities.

When M is not compact, this argument breaks down because there is not anymore uniform Lipschitzianity/semiconcavity of the squared distance functions. This creates problems not only in the change of variable formula, but also in the construction of the optimal map T via the formula (1.31). In the attempt to recover the existence of optimal maps via equation (1.31), some authors (see for instance [23]) passed to a weak formulation with the approximate gradient $\tilde{\nabla}\varphi$ in place of $\nabla\varphi$, showing that the Kantorovich potential has sufficient regularity to still justify the formula.

In a joint work with Figalli [6], we understood that the introduction of such analytic tool is actually unneeded, as regardless of any curvature assumption on the manifold and assumptions on the given measures to transport, the Kantorovich potentials are always locally semiconcave in the 'region of interest'. This shows not only that (1.31) holds, but also that the change of variable formula is true. More precisely, we proved that

Theorem 1.21 *Let φ be a Kantorovich potential as above, set $D := \{\varphi < +\infty\}$ and let Ω be the interior of D. Then φ is locally semiconvex in Ω, $\partial^c\varphi(x)$ is non-empty for any x in Ω, and $\partial^c\varphi$ is locally bounded in Ω. Moreover, $D \setminus \Omega$ can be covered by a countable number of semiconvex surfaces*[3] *of dimension $(n-1)$.*

I remark that semiconvex surfaces are particular cases of $c-c$ hypersurfaces (see the discussion made in Section 1.2), therefore the theorem also ensures that as soon as a measure μ satisfies

[3] actually, in the mentioned paper we proved that $D \setminus \Omega$ is $n-1$-rectifiable. I improved the result to the current statement in [12]

(iii) of Theorem 1.6, it is concentrated on Ω, where the Kantorovich potential is locally semiconcave. Thus this statement perfectly fits into the global picture of the optimal transport problem.

The tools we used in the proof are mainly geometrical and inspired from a discussion made by Villani in [47, Chapter 10], where he introduces the assumptions ($\mathbf{H}\infty_{\mathbf{1}}$) and ($\mathbf{H}\infty_{\mathbf{2}}$), and proves that if the cost function satisfies these assumptions, then a result closely related to ours holds (see [47, Theorem 10.24]).

I shall now sketch the proof of the local semiconcavity result.

Step 1: φ is locally bounded in Ω.
Since φ is defined by a infimum of continuous functions, the fact that φ is locally bounded from above is immediate. Hence we only need to prove the bound from below.

We argue by contradiction, and we assume the existence of a sequence $x_n \to \overline{x} \in \Omega$ such that $\varphi(x_n) \to -\infty$. For every $n \in \mathbb{N}$, let us choose $y_n \in M$ a point such that

$$\varphi(x_n) \geq c(x_n, y_n) - \psi(y_n) - 1. \tag{1.32}$$

In particular, as $c = \frac{\mathsf{d}^2}{2} \geq 0$, we have $\psi(y_n) \to +\infty$. Hence, since

$$\mathbb{R} \ni \varphi(\overline{x}) \leq c(x, y_n) - \psi(y_n),$$

we deduce that $c(x, y_n) \to +\infty$, which further implies $c(x_n, y_n) = \frac{\mathsf{d}(x_n,y_n)^2}{2} \to +\infty$.

Now, let $\gamma_n : [0, \mathsf{d}(x_n, y_n)] \to M$ be a minimizing geodesic parameterized by arc-length connecting x_n to y_n. Since $\mathsf{d}(x_n, y_n) \to +\infty$, any geodesic γ_n is defined at least on an interval $[0, \ell]$, for some $\ell > 0$. Let us define the following set:

$$C_n := \Big\{ x \in M \,:\, \text{there exists } t \in [0, \ell] \text{ s.t. } \mathsf{d}(x, \gamma_n(t)) \leq t/2 \Big\}.$$

We claim that

$$\sup_{C_n} \varphi \to -\infty \qquad \text{as } n \to +\infty. \tag{1.33}$$

Indeed, if $d(x, \gamma_n(t)) \leq t/2$ for some $t \in [0, \ell]$, thanks to the triangle inequality and (1.32) we have

$$\begin{aligned}
\varphi(x) \leq c(x, y_n) - \psi(y_n) &\leq \frac{[\mathsf{d}(\gamma_n(t), y_n) + \mathsf{d}(x, \gamma_n(t))]^2}{2} - \psi(y_n) \\
&\leq \frac{[\mathsf{d}(\gamma_n(t), y_n) + t/2]^2}{2} - \psi(y_n) = \frac{[\mathsf{d}(x_n, y_n) - t/2]^2}{2} - \psi(y_n) \\
&\leq \frac{\mathsf{d}(x_n, y_n)^2}{2} - \mathsf{d}(x_n, y_n)\, \frac{t}{2} + \frac{\ell^2}{8} - \psi(y_n) \\
&\leq \varphi(x_n) + 1 + \frac{\ell^2}{8},
\end{aligned}$$

where in the second line we used the identity $d(\gamma_n(t), y_n) = d(x_n, y_n) - t$. Thus the claim (1.33) is proved.

Now, letting $n \to +\infty$ and using the local compactness of M it is easy to check that (1.33) implies the existence of an open set $A \subset \Omega$ containing $\overline{x}$ in its boundary and on which φ is identically $-\infty$. Which is absurdum by definition of Ω.

Step 2: $\partial^c\varphi(x)$ is non-empty and bounded as x varies in a compact subset of Ω. In particular, φ is locally semiconcave in Ω.

Let $K \subset\subset \Omega$, take $\overline{x} \in K$, and let $y \in M$ be such that

$$\varphi(\overline{x}) \geq c(\overline{x}, y) - \psi(y) - 1.$$

We claim that $\mathsf{d}(\overline{x}, y)$ is uniformly bounded, independently of y satisfying the above inequality. Indeed, assuming without loss of generality $\mathsf{d}(\overline{x}, y) \geq 1$, as in the proof of Step 1 we can consider the point $\gamma(\ell)$ on the (unit speed) geodesic from $\overline{x}$ to y, where $\ell \leq \mathsf{d}(K, \Omega^c)/2$. Then we have

$$\varphi(\gamma(\ell)) \leq c(\gamma(\ell), y) - \psi(y) \leq \frac{\mathsf{d}^2(\overline{x}, y)}{2}(1-\ell) - \psi(y) \leq \varphi(\overline{x}) + 1 - \ell\frac{\mathsf{d}^2(\overline{x}, y)}{2}$$

Since by Step 1 φ is uniformly bounded on the set $\{x \in M : \mathsf{d}(x, K) \leq \ell\} \subset\subset \Omega$, the claim follows.

Thus we just proved that if x varies in a compact subset K of Ω, any sequence of minimizers in (1.30) must lie in a bounded set H. Hence any such sequence must have a convergent subsequence, which proves that $\partial^c\varphi(x)$ is non empty and bounded. It also follows that for $x \in K$ it holds

$$\varphi(x) = \inf_{y \in H} \frac{\mathsf{d}^2(x, y)}{2} - \psi(y),$$

and since $x \mapsto \mathsf{d}^2(x, y)$ is semiconcave on K for any $y \in H$, the local semiconcavity of φ follows as well.

1.4.2 Regularity of the map $\nu \mapsto T_\mu^\nu$

Consider the transport problem in $\mathbb{R}^d$ with cost=distance-squared, let $\mu \ll \mathcal{L}^d$ be a measure with finite second moment. For $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, let T_μ^ν be the optimal transport map from μ to ν.

A consequence of the stability of the set of optimal plans w.r.t. weak convergence, is that the map

$$\mathscr{P}_2(\mathbb{R}^d) \ni \nu \qquad \mapsto \qquad T_\mu^\nu \in L^2(\mathbb{R}^d, \mathbb{R}^d; \mu),$$

is continuous.

The question, loosely formulated, that I want to address now is: does this map have any better regularity? In particular, if we have a Lipschitz curve (ν_t), do we have that $t \mapsto T_\mu^{\nu_t}$ is Lipschitz as well?

The general answer is no. In [11] I produced an example of geodesic (ν_t) such that $T_\mu^{\nu_t}$ is at most $\frac{1}{2}$-Hölder continuous. On the other direction, an argument by Ambrosio shows that $C^{1/2}$ regularity is achievable, so that the example gives the sharp bound. I also mention that a similar question has been investigated also by Loeper [29]. He obtained a result of the following kind: he assumed $\nu_t = (X(t, \cdot))_\#\mu$, with $\mu = \mathcal{L}^d|_\Omega$ for some open set Ω, and $X(t, x) : [0, 1] \times \Omega \to \mathbb{R}^d$ with both X and $\partial_t X$ L^∞ in space and time, and he derived that $t \mapsto T_\mu^{\nu_t} \in L^2(\mathbb{R}^d, \mathbb{R}^d : \mu)$ is of bounded variation.

I turn to the description of the example. Notice that it is built over the example (provided in [3]) which shows that $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ is an Alexandrov space with curvature ≥ 0. Note that with minor modifications one can let all the measures involved be absolutely continuous with C_c^∞ densities.

Let $A := (-2,1)$, $B := (2,1)$, $C := (0,-2)$ and $O := (0,0)$. Since the strict inequality

$$|A-O|^2 + |O-C|^2 = 5+4 < 13+0 = |A-C|^2 + |O-O|^2$$

holds, where $|\cdot|$ is the euclidean norm, we have that for $r>0$ small enough it holds

$$|A-O'|^2 + |O-C'|^2 < |A-C'|^2 + |O-O'|^2, \quad \forall O' \in B_r(O),\ C' \in B_r(C). \tag{1.34}$$

Fix such an r and define the measures

$$\begin{aligned}\mu_0 :=& \tfrac{1}{2}\left(\delta_A + \delta_O\right),\\ \mu_1 :=& \tfrac{1}{2}\left(\delta_B + \delta_O\right),\\ \sigma :=& (2\pi r^2)^{-1}\left(\mathcal{L}^2|_{B_r(O)\cup B_r(C)}\right).\end{aligned}$$

Inequality (1.34) implies that the optimal transport map T_0 from σ to μ_0 satisfies $T_0(B_r(O)) = \{A\}$ and $T_0(B_r(C)) = \{O\}$. Symmetrically, for the optimal transport map T_1 from σ to μ_1 it holds $T_1(B_r(O)) = \{B\}$ and $T_1(B_r(C)) = \{O\}$.

Now observe that since

$$|A-O|^2 + |O-B|^2 = 5+5 < 16+0 = |A-B|^2 + |O-O|^2,$$

there is a unique optimal plan between μ_0 and μ_1 and this plan is induced by the map S, seen from μ_0, given by $S(A) = O$ and $S(O) = B$. Observe that it holds $S(T_0(B_r(O))) \neq T_1(B_r(O))$.

Let $\mu_t := ((1-t)\mathrm{Id} + tS)_\# \mu_0$ and T_t be the optimal transport map from σ to μ_t. Let $D_t := (1-t)A$ and $E_t := tB$, so that $\mathrm{supp}(\mu_t) = \{D_t, E_t\}$.

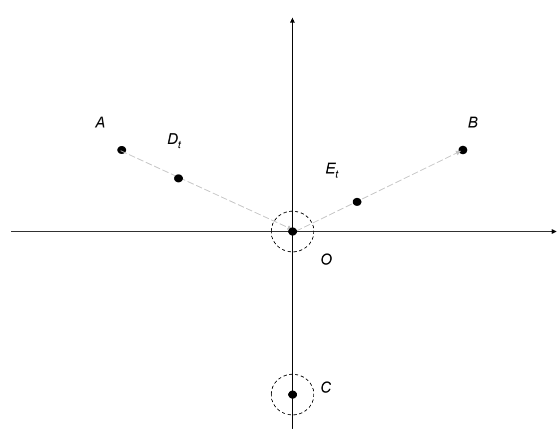

Figure 2: Position of the masses

Here it comes the main idea of the example. We claim that the map $t \to T_t \in L^2(\mathbb{R}^2, \mathbb{R}^2; \sigma)$ is not C^α for $\alpha > 1/2$: we will argue by contradiction. Suppose that for some $\alpha > 1/2$ the map is C^α, let χ be the characteristic function of $B_r(0)$ (i.e. $\chi(B_r(0)) = \{1\}$ and $\chi(\mathbb{R}^2 \setminus B_r(0)) = \{0\}$) and observe that from the inequality

$$\int |T_t - T_s|^2 \chi d\sigma \leq \int |T_t - T_s|^2 d\sigma,$$

we get that 'any regularity of $t \mapsto T_t$ seen as curve in $L^2(\mathbb{R}^2, \mathbb{R}^2; \sigma)$ is inherited by the curve $t \mapsto T_t$ seen as curve with values in $L^2(\mathbb{R}^2, \mathbb{R}^2, 2\chi\sigma)$' (the factor 2 stands just for the renormalization of the mass). In particular the map $t \mapsto T_t \in L^2(\mathbb{R}^2, \mathbb{R}^2; 2\chi\sigma)$ is C^α, too. Therefore defining the measures

$$\nu_t := (T_t)_\#(2\chi\sigma),$$

and using the inequality

$$W^2(\nu_t, \nu_s) \leq \int |T_t - T_s|^2 \, \mathrm{d}(2\chi\sigma),$$

we get that the curve $t \mapsto \nu_t \in (\mathscr{P}_2(\mathbb{R}^d), W_2)$ is C^α. The contradiction comes from the fact that the mass of ν_0 lies entirely on D_0, while the mass of ν_1 is on E_1. To make the contradiction evident, define the function $f : [0,1] \to [0,1]$ as $f(t) := \nu_t(D_t)$ and observe that it holds $f(0) = 1$ and $f(1) = 0$. Now we want to evaluate the distance $W(\nu_t, \nu_s)$: roughly speaking, the best way to move the mass from ν_t to ν_s is to move as much mass as poissible from D_t to D_s, as much mass as possible from E_t to E_s and then 'to adjust the rest'. More precisely, it can be easily checked that the optimal transport plan between ν_t and ν_s is given by

$$\begin{aligned}
&\min\{f(t), f(s)\}\delta_{(D_t, D_s)} + \min\{1 - f(t), 1 - f(s)\}\delta_{(E_t, E_s)} \\
&+ (f(t) - f(s))^+ \delta_{(D_t, E_s)} + (f(s) - f(t))^+ \delta_{(E_t, D_s)},
\end{aligned}$$

as its support is either $\{(D_t, D_s), (E_t, E_s), (D_t, E_s)\}$ or $\{(D_t, D_s), (E_t, E_s), (E_t, D_s)\}$ (depending on whether $f(t) \geq f(s)$ or viceversa, respectively) and both of these sets are cyclically monotone. Therefore we get

$$\begin{aligned}
W_2^2(\nu_t, \nu_s) = &\min\{f(t), f(s)\}|D_t - D_s|^2 + \min\{1 - f(t), 1 - f(s)\}|E_t - E_s|^2 \\
&+ (f(t) - f(s))^+ |D_t - E_s|^2 + (f(s) - f(t))^+ |E_t - D_s|^2.
\end{aligned}$$

Considering only the last two terms of the expression on the right, and choosing $|s - t| < 1/2$ we get the bound

$$W_2(\nu_t, \nu_s) \geq \frac{\sqrt{5}}{2} \sqrt{f(t) - f(s)}.$$

From the fact that $t \mapsto \nu_t \in (\mathscr{P}_2(\mathbb{R}^d), W_2)$ is C^α we get

$$\sqrt{f(t) - f(s)} \leq c|t - s|^\alpha, \quad \forall t, s \text{ s.t. } |s - t| < 1/2,$$

for some constant c. The contradiction follows: indeed the above inequality and the fact that $\alpha > 1/2$ implies that f is constant on $[0,1]$, while we know that $f(0) = 1$ and $f(1) = 0$.

2 PDEs with gradient flow structure

A subject on which I spent many energies since my PhD studies has been the study of gradient flows both from the theoretical point of view, and from the one of applications. The book which I wrote together with L. Ambrosio and G. Savaré contains essentially all of what I've done during the PhD. Here I want to focus on more recent results, where techniques coming from the study of gradient flows help the study of certain PDEs.

The common feature of the three sections below is the fact that it will be given a metric space (X, d) and a functional E on X, then for an initial datum x_0 and a time step τ I will consider the following well known recursive minimization scheme, called implicit Euler scheme or minimizing movements technique. I put $x_0^\tau := x_0$ and for every $n \in \mathbb{N}$ the point $x_{n+1}^\tau \in X$ is chosen among the minimizers of

$$y \qquad \mapsto \qquad E(y) + \frac{\mathsf{d}^2(x_n^\tau, y)}{2\tau}. \tag{2.1}$$

This defines a sequence (x_n^τ) and one can show that in a quite high generality when $\tau \downarrow 0$ these sequence converge, after appropriate time rescaling, to a gradient flow of E (see [3]).

Beside the potential interest of this construction in an abstract setting, the point on which I focus in this chapter is that with appropriate choice of functional/metric space this construction can produce solutions to various PDEs. This, of course, is not a new observation: the added value here is the study of new PDEs via this technique.

2.1 The heat equation with Dirichlet boundary data

In the seminal paper [26] Jordan, Kinderleher and Otto understood that the heat equation on $\mathbb{R}^d$ can be viewed as gradient flow of the relative entropy w.r.t. the Wasserstein distance W_2, so that in particular the scheme described above applied to the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ and the relative entropy functional

$$\mathrm{Ent}(\mu) := \begin{cases} \displaystyle\int_\Omega \rho \log \rho \, \mathrm{d}\mathcal{L}^d, & \text{if } \mu = \rho \mathcal{L}^d, \\ +\infty, & \text{otherwise} \end{cases} \tag{2.2}$$

converges to solutions of the heat equation.

If one replaces the metric space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ with $(\mathscr{P}(\Omega), W_2)$, where $\Omega \subset \mathbb{R}^d$ is a bounded open set, it will still converge to solutions of the heat equation, and the boundary data will be the homogeneous Neumann ones. The fact that these are the boundary conditions arising can be heuristically guessed observing that given that the distance involved is W_2, and given that this distance is well defined only for couples of measure with the same mass, the mass itself must be preserved along the flow.

Motivated by the intent to find an analogous approach to construct solutions of the heat flow subject to Dirichlet boundary condition with a transport-like approach, in a joint paper with A. Figalli ([7]) we introduced a new transportation distance Wb_2 between measures whose main features are:

- It metrizes the weak convergence of positive Borel measures in Ω belonging to the space

$$\mathcal{M}_\Omega := \left\{ \mu \ : \ \int \mathsf{d}^2(x, \partial\Omega) \, \mathrm{d}\mu(x) < \infty \right\}. \tag{2.3}$$

Notice that $\mathcal{M}_\Omega$ contains all positive finite measures on Ω and that the claim made is perfectly analogous to what happens for the common Wasserstein distances, but without any mass constraint.

- The resulting metric space $(\mathcal{M}_\Omega, Wb_2)$ is always geodesic. This is a particularly interesting property compared to what happens in the classical Wasserstein space: indeed the space $(\mathscr{P}(\Omega), W_2)$ is geodesic if and only if Ω is convex. In our case, the convexity of the open set is not required (actually, not even connectedness is needed).
- The natural approach via time discretization and implicit Euler scheme pictured above and applied to the relative entropy functional, leads to weak solution of the heat equation with Dirichlet boundary condition.

As a drawback, the entropy functional does not have the same nice properties it has in the classical Wasserstein space, since it is *not* geodesically convex. Because of this:

- We were not able to prove any kind of contractivity result for the flow.
- Actually, we were not even able to prove uniqueness of the limit of the minimizing movements scheme. (Of course one knows by standard PDE techniques that weak solutions of the heat equation with Dirichlet boundary conditions are unique, therefore a posteriori it is clear that the limit has to be unique - what I'm saying here is that we do not know whether such uniqueness may be deduced a priori via techniques similar, e.g., to those appeared in [3].)

The distance Wb_2 is defined in the following way (the 'b' stands to recall that we have some room to play with the $b$oundary of Ω). Let $\Omega \subset \mathbb{R}^d$ be a bounded open set, and let $\mathcal{M}_\Omega$ be defined by (2.3). We define the distance Wb_2 on $\mathcal{M}_\Omega$ as a result of the following problem:

Problem 1 (A variant of the transportation problem) *Let $\mu, \nu \in \mathcal{M}_\Omega$. The set of admissible couplings* $\mathrm{Adm}_b(\mu,\nu)$ *is defined as the set of positive measures $\boldsymbol{\gamma}$ on $\overline{\Omega} \times \overline{\Omega}$ satisfying*

$$\pi^1_\# \boldsymbol{\gamma}|_\Omega = \mu, \qquad \pi^2_\# \boldsymbol{\gamma}|_\Omega = \nu. \tag{2.4}$$

For any non-negative measure $\boldsymbol{\gamma}$ on $\overline{\Omega} \times \overline{\Omega}$, we define its cost *$C(\boldsymbol{\gamma})$ as*

$$C(\boldsymbol{\gamma}) := \int_{\overline{\Omega} \times \overline{\Omega}} |x-y|^2 \, \mathrm{d}\boldsymbol{\gamma}(x,y).$$

The distance $Wb_2(\mu,\nu)$ is then defined as:

$$Wb_2^2(\mu,\nu) := \inf_{\boldsymbol{\gamma} \in \mathrm{Adm}_b(\mu,\nu)} C(\boldsymbol{\gamma}).$$

The difference between Wb_2 and W_2 relies on the fact that an admissible coupling is a measure on the *closure* of $\Omega \times \Omega$, rather than just on $\Omega \times \Omega$, and that the marginals are required to coincide with the given measures only *inside* Ω. This means that we can use $\partial\Omega$ as an infinite reserve: we can 'take' as mass as we wish from the boundary, or 'give' it back some of the mass, provided we pay the transportation cost. This is why this distance is well defined for measures which do not have the same mass.

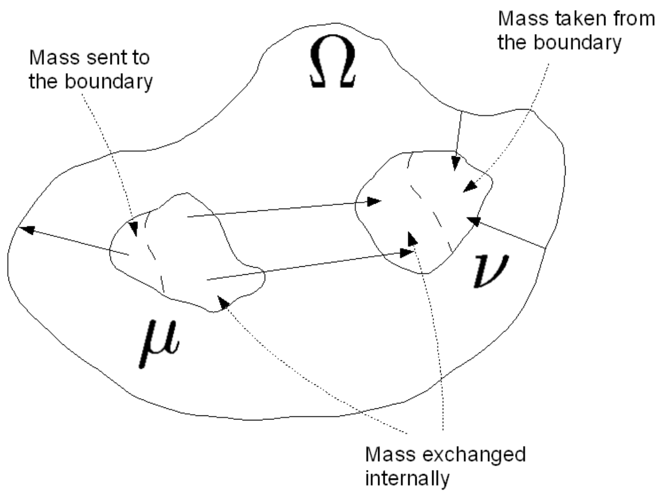

Figure 3: Example of admissible transport plan

Then, what we proved is that the discretization scheme applied to the same relative entropy functional defined in (2.2) produces solutions to the heat equation with Dirichlet boundary conditions:

$$\begin{cases} \partial_t \rho_t = \Delta \rho_t & \text{in } \Omega \times (0, \infty), \\ \rho_t = e^{-1} & \text{in } \partial\Omega \times [0, \infty). \end{cases}$$

The fact that the limit curve is a solution of the heat equation is not surprising, and can be shown more or less with the same ideas used for the standard distance W_2.

It is more interesting to understand why the boundary conditions that appear are the Dirichlet ones. The first thing to notice is that the functional Ent has a unique minimum in $\mathcal{M}_\Omega$, namely the measure $\frac{1}{e}\mathcal{L}_{|\Omega}$, which has constant density $\frac{1}{e}$. This already suggests that whatever the starting measure is, when $t \to \infty$ we should converge to $\mathcal{L}_{|\Omega}$, independently of the mass of the initial datum, and this excludes the Neumann boundary condition and illustrates why one should expect that the boundary condition is the Dirichlet one, with constant value equal to $\frac{1}{e}$. The correct proof of this ansatz passes from the study of the discrete scheme. The idea is based on the following two facts:

- the cost of moving/taking mass from the boundary is small if the point where we want to take/move mass is close to the boundary itself,
- in terms of entropy, it is certainly better to have as most points as possible with density close to $\frac{1}{e}$.

Turning these two properties into quantitative statements is not hard. What comes out is that any minimizer of the discrete scheme has trace 1 on the boundary of Ω, and the estimates passes to the limit, giving the claimed property for the solution of the heat equation.

Finally, I remark that slightly modifying the functional, one can achieve constant Dirichlet boundary data with any positive constant in place of $\frac{1}{e}$: putting $\rho \log \rho - c\rho$ in place of $\rho \log \rho$ inside the integral gives boundary value e^{c-1} (because the minimizer of $z \mapsto z \log z - cz$ is that value). What is not possible to do is to is to produce non constant Dirichlet boundary data without adding a drift term to the heat equation; this is perfectly in line with what happens with the use of the classical distance W_2.

2.2 The Burgers equation

In a joint paper with Otto ([16]) we investigated the relation between (a generalization of) the 1-dimensional Burgers equation

$$\partial_t\theta_t + \partial_x(\theta_t(1-\theta_t)) = 0, \tag{2.5}$$

and the gradient flow theory on a two-phase Wasserstein space (see below for the definition). Such relation was already understood in a previous paper by Otto ([38]), but in our work several proofs have been simplified and a deeper analysis has been carried out.

The point is the following. Consider the set

$$\mathcal{M} := \Big\{\theta : [-1,1] \to [0,1] \ : \ \int \theta \, \mathrm{d}\mathcal{L}^1 = 1\Big\},$$

and endow it with the distance

$$\mathsf{d}^2(\theta,\tilde\theta) := W_2^2(\theta,\tilde\theta) + W_2^2(1-\theta, 1-\tilde\theta),$$

having identified a measure with its density. Then we can see $\mathcal{M}$ as a 'submanifold' of the product manifold $\mathscr{P}_2(\mathbb{R}) \times \mathscr{P}_2(\mathbb{R})$.

On $\mathcal{M}$ we consider the functional $\mathcal{E} : \mathcal{M} \to [-1,1]$ given by

$$\mathcal{E}(\theta) := \int_{-1}^{1} z\,\theta(z)\,\mathrm{d}z.$$

It is then not hard to check that, at least from the formal viewpoint, the gradient flow of $\mathcal{E}$ in $(\mathcal{M},\mathsf{d})$ produces solutions to the Burgers equation (2.5). This fact can be proved rigorously, but it is not the main point here.

What is interesting to observe is that neither at the level of Burgers' equation, nor at the one of gradient flows of $\mathcal{E}$ one can expect a general uniqueness statement (see below); yet, the natural scheme used to build gradient flows for $\mathcal{E}$ selects precisely the 'correct' solution of the equation.

Notice that:

- For the Burgers equation one typically does not have C^1 regularity of the solution, so that one has to deal with distributional solutions. However, they don't grant uniqueness and in order to restore the existence of a unique solution, the notion of *entropy* solution comes into play.

- The functional $\mathcal{E}$ is not semi(geodesically)convex on the 'manifold' $\mathcal{M}$, so that it is very possible that it admits more than one gradient flow for some given initial data. It is unclear whether there is a general principle to select the 'most natural' gradient flow (see the discussion below), but there is a natural scheme to produce gradient flows: the Euler scheme.

Concerning this latter point, observe that given the variational nature of the implicit Euler scheme, it is natural to guess that in a situation where uniqueness is not guaranteed, it selects the gradient flow which 'decreases the energy fastest'. Consider for instance the case of the energy E on $\mathbb{R}_+$ given by $E(x) := -x^{4/3}$. It is a classical fact of ODEs' theory that there are infinite gradient flows for it starting from 0, the two extremal solutions being the stationary curve $x_t^1 \equiv 0$ and the curve which immediately moves $x_t^2 := ct^{3/2}$, with $c = (8/9)^{3/2}$. Of the two, the second one certainly decreases the energy E faster. A straightforward computation shows that this second curve is the one selected by the implicit Euler scheme.

To formalize the concept of 'decreasing the energy fastest', let me observe that given a gradient flow (x_t) of E, the first derivative of the energy along the flow is prescribed by the gradient flow equation, as it holds $\partial_t E(x_t) = \nabla E(x_t) \cdot x_t' = -|\nabla E|^2(x_t)$, however if E is not C^2, as in the case considered, the second derivative is not. Indeed in our case we have $\partial_{tt} E(x_t^1) = 0$ and $\partial_{tt} E(x_t^2) = -2c$. In other words, the derivative of $|\nabla E|^2$ along the flow is not given, and is higher for those flows which decrease the energy fastest. Thus one can try to give the following definition.

Definition 2.1 (Gradient flows which decreases the energy fastest) *Let $E : \mathbb{R}^d \to \mathbb{R}$ be a C^1 functional and $(x_t) \subset \mathbb{R}^d$ a curve. Then we say that (x_t) is a gradient flow which decreases the energy fastest, provided it is a gradient flow for E and for any $t_0 \geq 0$ the following is true. If (y_t) is a gradient flow for E starting from x_{t_0}, then*

$$\frac{\mathrm{d}^+}{\mathrm{d}t} |\nabla E|^2(x_t)_{|t=t_0} \geq \frac{\mathrm{d}^+}{\mathrm{d}t} |\nabla E|^2(y_t)_{|t=0}. \tag{2.6}$$

Still at the level of formal analogies, this definition helps understanding why the gradient flow which decreases the energy fastest for the functional E on $\mathcal{M}$ should correspond to the entropy solution of (2.5). Indeed, some (formal) computations show that the squared slope of $\mathcal{E}$ in $(\mathcal{M}, \mathsf{d})$ is given by

$$|\nabla E|^2(\theta) = \int \theta(1-\theta)\, \mathrm{d}\mathcal{L}^1,$$

and therefore is a strictly concave function of θ. As such, $\eta(z) := -z(1-z)$ is an admissible entropy for (2.5) and we know from the work of Dafermos [21] that the entropy solution (θ_t) is characterized among all weak solutions by the inequality

$$\frac{\mathrm{d}^+}{\mathrm{d}t} \int \eta(\theta_t)_{|t=t_0} \leq \frac{\mathrm{d}^+}{\mathrm{d}t} \int \eta(\tilde{\theta}_t)_{|t=0}, \tag{2.7}$$

where $(\tilde{\theta}_t)$ is any other weak solution coinciding with (θ_t) up to the time $t = t_0$.

The parallelism between (2.6) and (2.7) is what drove our intuition in proving that the minimizing movements scheme for $\mathcal{E}$ on $\mathcal{M}$ produces the entropy solution for (2.5).

Yet, I want to underline that all the discussion that I made here is purely formal, and actually unneeded in the rigorous proof of our result. In particular, the definition of gradient flow which decreases the energy fastest that I gave here, in general does not guarantees existence, not even for C^1 functionals on $\mathbb{R}^2$. And even when there is existence, it is very possible that the implicit Euler scheme does not select it. In my opinion, these two facts (proved by counterexamples in [16]) and the fact that in the 'practical' case of the Burgers equation the formal argument suggests a statement which is actually true, show that there is still something 'behind the scenes' which needs to be undestood.

2.3 The Navier-Stokes equations

In a paper with S. Mosconi [13] we used a variant of the implicit Euler scheme to study the Navier-Stokes equations, and reprove many known basic results about them within a single framework. For this scheme we proved that:

- in any dimension it produces Hopf solutions,
- in dimension 3 it converges to suitable solutions,
- if the initial datum is in H^1 it produces strong solutions in some interval $[0,T)$,
- if the initial datum satisfies the classical smallness conditions, the scheme produces strong solutions in $[0,\infty)$.

In order to describe the approach, I will assume that we are working in the $d-$dimensional flat torus $\mathbb{T}^d=\mathbb{R}^d/\mathbb{Z}^d$, so that we don't have to worry about the boundary conditions. Let $\overline{u}$ be a smooth divergence free vector field on the $\mathbb{T}^d$ and consider the Navier-Stokes equations with initial datum $\overline{u}$:

$$\left\{\begin{array}{rll} \partial_t u_t+(u_t\cdot\nabla)u_t+\nabla p_t &=\Delta u_t, & \text{in } [0,\infty)\times\mathbb{T}^d,\\ \nabla\cdot u_t &=0, & \text{in } \mathbb{T}^d \quad \forall t,\\ u_0 &=\overline{u}, & \text{in } \mathbb{T}^d, \end{array}\right. \tag{2.8}$$

Now given a smooth vector field u, define its flow map $\mathbb{R}\times\mathbb{T}^d\ni(t,x)\mapsto X_t^u(x)\in\mathbb{T}^d$ as the only solution of

$$\left\{\begin{array}{rl} \partial_t X_t^u &=u\circ X_t^u,\\ X_0^u &=\mathrm{Id}. \end{array}\right.$$

Then for a given time step $\tau>0$ define $u_0^\tau:=\overline{u}$ and u_{n+1}^τ as the unique minimizer of

$$v \qquad \mapsto \qquad \frac{1}{2}\int_{\mathbb{T}^d}|\nabla v|^2\,\mathrm{d}\mathcal{L}^d+\frac{\|v\circ X_\tau^{u_n^\tau}-u_n^\tau\|_{L^2}^2}{2\tau}, \tag{2.9}$$

among all L^2 and divergence free vector fields v.

From a theoretical point of view, it is worth noticing the structural difference between this minimization scheme and the one in (2.1): the distance term is perturbed by something depending on the previous minimization step. Hence, to some extent, this fact is saying that the Navier-Stokes equations can be seen as a sort of 'non-autonomous and implicit' gradient flow.

In practice, it is natural to compare (2.9) with the problem of minimizing $L^2\ni g\mapsto \frac{1}{2}\int|\nabla g|^2\,\mathrm{d}\mathcal{L}^d+\frac{\|g-f\|_{L^2}^2}{2t}$ used to build solutions of the heat equations. There two differences:

- we are not minimizing over all L^2 vector fields, but only among divergence free ones (this will produce the pressure term),
- we are perturbing the L^2 distance with the right composition by the flow map $X_\tau^{u_n^\tau}$ (this will produce the non-linear term).

I shall now informally describe why this minimization procedure gives solutions of the Navier-Stokes equations. For notational simplicity, consider the first step in the minimization procedure. It is not hard to check that the unique minimum u_1^τ of (2.9) satisfies

$$\frac{u_1^\tau-\overline{u}\circ X_{-\tau}^{\overline{u}}}{\tau}+\nabla p_1^\tau=\Delta u_1^\tau, \tag{2.10}$$

where p_1^τ is identified, up to additive constants, by

$$\Delta p_1^\tau = \nabla \cdot \left(\frac{\overline{u} \circ X_{-\tau}^{\overline{u}}}{\tau} \right).$$

The point is that (2.10) is a time discretization of (2.8). Indeed, the term

$$\frac{u_1^\tau - \overline{u} \circ X_{-\tau}^{\overline{u}}}{\tau} = \left(\frac{u_1^\tau \circ X_\tau^{\overline{u}} - \overline{u}}{\tau} \right) \circ X_{-\tau}^{\overline{u}},$$

is the time discretization of the convective derivative

$$\partial_t u_t + (u_t \cdot \nabla) u_t = \Big(\partial_t (u_t \circ T_t) \Big) \circ T_t^{-1},$$

where here $[0,\infty) \times \mathbb{T}^d \ni (t,x) \mapsto T_t(x) \in \mathbb{T}^d$ is the flow map (or particle-trajectory map) associated to (u_t), i.e.:

$$\begin{cases} \partial_t T_t &= u_t \circ T_t, \\ T_0 &= \mathrm{Id}. \end{cases}$$

and the pressure term satisfies

$$\Delta p_1^\tau = \nabla \cdot \left(\frac{\overline{u} \circ X_{-\tau}^{\overline{u}}}{\tau} \right) = \nabla \cdot \left(\frac{\overline{u} \circ X_{-\tau}^{\overline{u}} - \overline{u}}{\tau} \right),$$

which is a time discretization of

$$\Delta p_t = \nabla \cdot \big((u_t \cdot \nabla) u_t \big) = \nabla \cdot \big(\partial_t u_t + (u_t \cdot \nabla) u_t \big),$$

the latter being the formula identifying the pressure in (2.8).

Now, to prove that the scheme produces a distributional solution of (2.8), just multiply (2.10) by a smooth ξ and integrate over $\mathbb{T}^d$ to get

$$\left\langle \frac{u_1^\tau - \overline{u}}{\tau}, \xi \right\rangle_{L^2} - \left\langle \overline{u}, \frac{\xi \circ X_\tau^{\overline{u}} - \xi}{\tau} \right\rangle_{L^2} + \langle p_1^\tau, \nabla \cdot \xi \rangle_{L^2} = \langle u_1^\tau, \Delta \xi \rangle_{L^2},$$

having used the fact that $(X_\tau^{\overline{u}})_\# \mathcal{L}^d = \mathcal{L}^d$, which is a consequence of the fact that $\nabla \cdot \overline{u} = 0$. Thus from the identity

$$\xi \circ X_\tau^{\overline{u}} - \xi = \int_0^\tau \partial_t \xi \circ X_t^{\overline{u}} \, \mathrm{d}t = \int_0^\tau \nabla \xi \circ X_t^{\overline{u}} \cdot \overline{u} \circ X_t^{\overline{u}} \, \mathrm{d}t,$$

and a simple iteration, it is not hard to see that the discrete solutions produce approximate distributional solutions. To gain the (discrete) energy inequality, multiply (2.10) by u_1^τ and integrate to get

$$\|u_1^\tau\|_{L^2}^2 - \left\langle u_1^\tau, \overline{u} \circ X_{-\tau}^{\overline{u}} \right\rangle_{L^2} = \tau \langle u_1^\tau, \Delta u_1^\tau \rangle_{L^2}.$$

Now notice that from

$$\begin{aligned} \langle u_1^\tau, \Delta u_1^\tau \rangle_{L^2} &= -\|\nabla u_1^\tau\|_{L^2}^2, \\ \left\langle u_1^\tau, \overline{u} \circ X_{-\tau}^{\overline{u}} \right\rangle_{L^2} &\le \frac{1}{2}\|u_1^\tau\|_{L^2}^2 + \frac{1}{2}\|\overline{u} \circ X_{-\tau}^{\overline{u}}\|_{L^2}^2 = \frac{1}{2}\|u_1^\tau\|_{L^2}^2 + \frac{1}{2}\|\overline{u}\|_{L^2}^2, \end{aligned}$$

one can conclude that $\frac{1}{2}\|u_1^\tau\|_{L^2}^2 + \tau\|\nabla u_1^\tau\|_{L^2}^2 \le \frac{1}{2}\|\overline{u}\|_{L^2}$, which by iteration gives

$$\frac{1}{2}\|u_m^\tau\|_{L^2}^2 + \tau \sum_{i=n+1}^{m} \|\nabla u_i^\tau\|_{L^2}^2 \le \frac{1}{2}\|u_n^\tau\|_{L^2}, \qquad \forall n, m \in \mathbb{N}.$$

Once one has these discrete equations, to pass to the limit and get Hopf solutions is easy, and along similar lines one can prove the other claimed statements concerning suitable solutions and strong solutions.

Let me underline that, as far as I know, this scheme does not produce any new estimate/guess about the Navier-Stokes equation. Still, I believe it is interesting to know both that these equations have a sort of gradient flow structure, and that many known basic results about them can be recovered using a single technique.

3 Hopf-Lax formula and Hamilton-Jacobi equation in metric spaces

It is well known that on $\mathbb{R}^d$ the Hopf-Lax formula

$$Q_t f(x) := \inf_y f(y) + \frac{|x-y|^2}{2t},$$

produces the unique viscosity solution of the Hamilton-Jacobi equation

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t + \frac{|\nabla f_t|^2}{2} = 0,$$

with initial data f.

In this short chapter, whose content is extracted from [4], I will show that a strong relation between the Hopf-Lax formula and the Hamilton-Jacobi equation holds in a purely metric setting.

Let me assume for simplicity that the metric space (X, d) we are dealing with is compact (this is largely unneeded), and fix a continuous function $f : X \to \mathbb{R}$. For $t > 0$ define $F : (0, \infty) \times X^2 \to \mathbb{R}$ by

$$F(t, x, y) := f(y) + \frac{\mathsf{d}^2(x, y)}{2t},$$

and $Q_t f : X \to \mathbb{R}$ by

$$Q_t f(x) := \inf_{y \in X} F(t, x, y), \tag{3.1}$$

with $Q_0 f := f$.

Now denote by $|\nabla g|$ the *local Lipschitz constant of* g defined by

$$|\nabla g|(x) := \varlimsup_{y \to x} \frac{|g(y) - g(x)|}{\mathsf{d}(x, y)},$$

and by $|\nabla^{\pm} g|$ its one sided counterparts, namely the *ascending and descending slopes*, defined respectively by

$$\begin{aligned} |\nabla^+ g|(x) &:= \varlimsup_{y \to x} \frac{(g(y) - g(x))^+}{\mathsf{d}(x, y)}, \\ |\nabla^- g|(x) &:= \varlimsup_{y \to x} \frac{(g(y) - g(x))^-}{\mathsf{d}(x, y)}. \end{aligned} \tag{3.2}$$

If x is isolated, all these object are taken 0 by definition.

Then the following theorem holds.

Theorem 3.1 *With the above notation, the map* $(0, \infty) \times X \ni (t, x) \mapsto Q_t f(x)$ *is locally Lipschitz and for any* $x \in X$ *it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_t f(x) + \frac{|\nabla Q_t f|^2(x)}{2} \leq 0, \tag{3.3}$$

with at most countably many exceptions in $(0, \infty)$.

If (X, d) *is a geodesic space, then equality holds in* (3.3), *with at most countably many exceptions in* $(0, \infty)$ *as well.*

Notice that the statement does not tell anything about viscosity solutions of (3.3), the reason simply being that as of today there is no notion of viscosity solution in a metric setting.

Before describing the approach to the proof of this theorem, I want to underline that the first intuition about the purely metric relation between the Hopf-Lax formula and the Hamilton-Jacobi equation is due to Lott and Villani, who in [31] proved an analogous result under the additional conditions that the space is equipped with a reference doubling measure $\mathfrak{m}$ and that $(X, \mathsf{d}, \mathfrak{m})$ supports a local Poincaré inequality. These assumption allowed them to use the deep results obtained by Cheeger in [19] to somehow mimic (up to several non trivial technicalities) the standard proof available in a smooth setting. What we added in [4] was the fact that the result can be achieved relying only on the metric structure, taking advantage of the variational structure of the Hopf-Lax formula.

The first step in the proof of 3.1 is the introduction of the functions $D^+, D^- : (0,\infty)\times X \to [0,\infty)$:

$$\begin{aligned} D^+(t,x) &:= \sup \mathsf{d}(x,y),\\ D^-(t,x) &:= \inf \mathsf{d}(x,y), \end{aligned}$$

where in both cases y varies among the minimizers of $F(t,x,\cdot)$ (which exist because of the continuity of f and the compactness of X). It is immediate to check that the sup and the inf in the definition of $D^\pm$ are realized. Also, a simple stability argument shows that D^+ is upper semicontinuous and D^- lower semicontinuous.

An argument which goes back to De Giorgi (which is key when studying the problem of gradient flows in metric spaces) shows the following.

Lemma 3.2 (De Giorgi's variational interpolation) *For any $x \in X$ the map $t \mapsto Q_t f(x)$ is locally semiconcave and it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_t f(x) = -\frac{\big(D^\pm(t,x)\big)^2}{2t^2},$$

for any $t \in (0,\infty) \setminus N$, where N is at most countable. In particular, $D^+(t,x) = D^-(t,x)$ for any $t \in (0,\infty)\setminus N$.

This is a known fact whose proof can be found in [3, Theorem 3.1.4].

Now notice that for given $t > 0$ the map $x \mapsto Q_t f(x)$ is certainly Lipschitz, because from the boundedness of X we deduce that the maps $x \mapsto \mathsf{d}^2(x,y)$ are uniformly Lipschitz in y. Hence to conclude the proof of the first part of Theorem 3.1 the only thing we need to check is that

$$\frac{D^+(x,t)}{t} \geq |\nabla Q_t f|(x). \tag{3.4}$$

To prove this, notice that

$$\begin{aligned} Q_t f(x) - Q_t f(y) \leq F(t,x,y_t) - F(t,y,y_t) &= \frac{\mathsf{d}^2(x,y_t) - \mathsf{d}^2(y,y_t)}{2t}\\ &\leq \mathsf{d}(x,y)\frac{\mathsf{d}(x,y_t) + \mathsf{d}(y,y_t)}{2t} \leq \mathsf{d}(x,y)\frac{\mathsf{d}(x,y) + 2D^+(t,y)}{2t}, \end{aligned}$$

where y_t is any minimizer of $F(t,y,\cdot)$. Dividing by $\mathsf{d}(x,y)$ and letting $y \to x$ we get (3.4) for the ascending slope in place of the local Lipschitz constant. To get the same inequality for the descending slope, just reverse the roles of x and y in the previous inequality.

In the case of geodesic spaces, to show that equality holds in (3.3) it is sufficient to show that

$$|\nabla^- Q_t f|(x) = \frac{D^+(t,x)}{t}, \qquad \forall t \in (0,\infty),\ x \in X.$$

To this aim, fix t, x, let $x_t \in \operatorname{argmin} F(t,x,\cdot)$ be such that $\mathsf{d}(x,x_t) = D^+(t,x)$ and let γ be a constant speed geodesic connecting x to x_t. Then it holds

$$\begin{aligned} Q_t f(x) - Q_t f(\gamma_s) \geq F(t,x,x_t) - F(t,\gamma_s,x_t) &= \frac{\mathsf{d}^2(x,x_t) - \mathsf{d}^2(\gamma_s,x_t)}{2t} \\ &= \mathsf{d}^2(x,x_t)\frac{2s - s^2}{2t} = \mathsf{d}(x,\gamma_s) D^+(t,x)\frac{2-s}{2t}, \end{aligned}$$

so that dividing by $\mathsf{d}(x,\gamma_s)$ and letting $s \downarrow 0$ we get the claim.

4 The heat flow on $CD(K,\infty)$ spaces as gradient flow of the entropy

4.1 Introduction

In recent years, a lot of attention has been given to the interplay between optimal transport and analysis in non-smooth setting, the majors breakthrough having been the works of Lott and Villani on one side ([33]) and of Sturm on the other ([43], [44]) where an abstract definition of Ricci curvature bound from below for metric measure spaces has been given.

A definition of particular relevance is that of $CD(K,\infty)$ space (the ∞ stands to say that no upper bound on the dimension is given, the general definition is about $CD(K,N)$ spaces which are spaces with Ricci curvature bounded below by K and dimension bounded above by N).

To give the definition I first recall that given a measure space $(X,\mathfrak{m})$, the relative entropy functional $\mathrm{Ent}_{\mathfrak{m}} : \mathscr{P}(X) \to [0,\infty]$ is defined by

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) := \begin{cases} \displaystyle\int \rho \log(\rho)\,\mathrm{d}\mathfrak{m} & \text{if } \mu = \rho\mathfrak{m}, \\ +\infty & \text{otherwise.} \end{cases}$$

Definition 4.1 (Metric measure spaces with Ricci curvature bounded below) *We say that $(X,\mathsf{d},\mathfrak{m})$ has Ricci curvature bounded below by $K\in\mathbb{R}$ (in short: it is a $CD(K,\infty)$ space) provided the relative entropy functional is K-geodesically convex on $(\mathscr{P}_2(X),W_2)$. In other words, $(X,\mathsf{d},\mathfrak{m})$ is a $CD(K,\infty)$ space provided for any $\mu_0,\mu_1\in\mathscr{P}_2(X)$ there exists a constant speed geodesic (μ_t) connecting them such that*

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t) \le (1-t)\mathrm{Ent}_{\mathfrak{m}}(\mu_0) + t\mathrm{Ent}_{\mathfrak{m}}(\mu_1) - \frac{K}{2}t(1-t)W_2^2(\mu_0,\mu_1).$$

While several analytical and geometric properties of $CD(K,N)$ spaces have been proved in [33], [43], a number of open questions remained open. Among others, Villani posed in [47] the following one: is it well defined an heat flow on $CD(K,\infty)$ spaces? The question is natural because on Riemannian manifolds with Ricci curvature bounded below the heat flow preserves the mass. Also, from the seminal paper [26] we know that the heat flow on $\mathbb{R}^d$ can be viewed as gradient flow of the relative entropy w.r.t. W_2 (later works generalized this fact to other smooth structures like Riemannian manifolds [22] and [46] and Finsler ones [37]), which gives a nice ‘coincidence’ given that $CD(K,\infty)$ spaces are defined exactly asking for a good behavior of the relative entropy w.r.t. the Wasserstein geometry.

It is therefore natural to try to *define* the heat flow on a $CD(K,\infty)$ space as gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ w.r.t. W_2: following this path, what one should do is to prove that the heat flow actually exists and is unique. This is precisely what I did in [9] (more recent results will be detailed in the following chapters). To discuss the point, I shall recall few facts about the general theory (as developed in [3]) of gradient flows of geodesically convex functionals on metric spaces.

Let (Y,d_Y) be a complete and separable metric space (which in the future will be the Wasserstein space built over a $CD(K,\infty)$ space) and $E : Y \to \mathbb{R}\cup\{+\infty\}$ a functional. I will denote by $D(E)$ its domain, i.e. $D(E) := \{E < \infty\}$. E is said K-geodesically convex provided for any $y_0, y_1 \in Y$ there exists a constant speed geodesic (y_t) connecting them such that

$$E(y_t) \leq (1-t)E(y_0) + tE(y_1) - \frac{K}{2}t(1-t)\mathsf{d}_Y^2(y_0, y_1), \qquad \forall t \in [0,1].$$

If E is K-geodesically convex, then the descending slope (recall (3.2)) admits the representation

$$|\nabla^- E|(y) = \sup_{z \neq y} \left(\frac{E(y) - E(z)}{\mathsf{d}_Y(y,z)} - \frac{K^-}{2}\mathsf{d}_Y(y,z) \right)^+. \tag{4.1}$$

A non trivial consequence (see [3, Corollary 2.4.10]) of K-geodesic convexity and lower semicontinuity is that the slope is an upper gradient for E, i.e. for any absolutely continuous curve $(y_t) \subset D(E)$ it holds

$$|E(y_t) - E(y_s)| \leq \int_t^s |\nabla^- E|(y_r)|\dot{y}_r| \, \mathrm{d}r, \qquad \forall t < s,$$

thus an application of Young's inequality shows that it holds

$$E(y_0) \leq E(y_t) + \frac{1}{2}\int_0^t |\dot{y}_r|^2 \, \mathrm{d}r + \frac{1}{2}\int_0^t |\nabla^- E|^2(y_s) \, \mathrm{d}r, \qquad \forall t. \tag{4.2}$$

Notice that by construction the equality holds if and only if for a.e. t the slope of E at y_t is realized along the curve and we have $|\dot{y}_t| = |\nabla^- E|(y_t)$. It is easy to see that if E is a smooth functional on $\mathbb{R}^d$, then this latter two conditions characterize solutions of

$$y_t' = -\nabla E(y_t).$$

This motivates the following definition:

Definition 4.2 (Gradient flow of geodesically convex functionals) *Let $E : Y \to \mathbb{R} \cup \{+\infty\}$ be a K-geodesically convex and lower semicontinuous functional and $\overline{y} \in D(E)$. We say that $[0,\infty) \ni t \mapsto y_t \in E$ is a gradient flow for E starting from $\overline{y}$ provided it is a continuous curve, locally absolutely continuous in $(0,\infty)$, $y_0 = \overline{y}$ and it holds*

$$E(y_0) = E(y_t) + \frac{1}{2}\int_0^t |\dot{y}_r|^2 \, \mathrm{d}r + \frac{1}{2}\int_0^t |\nabla^- E|^2(y_s) \, \mathrm{d}r, \qquad \forall t \geq 0. \tag{4.3}$$

The question is now: do we have in general existence and uniqueness of the gradient flow of a geodesically convex functional? Concerning uniqueness, the answer is *no*, so that in particular one cannot even expect any sort of contractivity of the distance along two flows. This can be easily seen considering the metric space $\mathbb{R}^2$ endowed with the L^∞ norm, and the functional $E(x,y) := x$. In this case any Lipschitz curve $t \mapsto (x_t, y_t)$ satisfying

$$x_t' = -1, \qquad |y_t'| \leq 1, \qquad a.e. \;\; t,$$

is a gradient flow for E starting from $(0,0)$.

Concerning existence, the following general result holds (see [3, Corollary 2.4.12]):

Theorem 4.3 *Let (Y, d_Y) be a complete and separable metric space and $E : Y \to \mathbb{R} \cup \{+\infty\}$ a K-geodesically convex and l.s.c. functional. Assume that there is a topology σ on Y, such that:*

i) d_Y*-bounded sublevels of* E *are* σ*-relatively compact,*

ii) d_Y *and* E *are* σ*-lower semicontinuous on* d_Y*-bounded sets,*

iii) $|\nabla^- E|$ *is* σ*-lower semicontinuous on* d_Y*-bounded sublevels of* E*.*

Then every $y_0 \in D(E)$ *is the starting point of a gradient flow for* E*.*

If we want to apply this theorem to the relative entropy functional on $(\mathscr{P}_2(X), W_2)$, where $(X, \mathsf{d}, \mathfrak{m})$ is a $CD(K,\infty)$ space, we can deduce existence of gradient flows of the entropy only if (X, d) is compact. Indeed under this assumption $(\mathscr{P}(X), W_2)$ is compact as well and we can choose as topology σ on $\mathscr{P}(X)$ the same topology induced by W_2, i.e. the weak topology. The only thing to check is the lower semicontinuity of the slope, which is a simple consequence of formula (4.1) and of the lower semicontinuity of $\mathrm{Ent}_{\mathfrak{m}}$.

When X is not compact, the only reasonable choice for the topology σ in order to be sure that $(i), (ii)$ are satisfied is the weak topology. However, with this choice is not at all clear why (iii) is fulfilled as well (notice that in (4.1) the term containing the distance is at the denominator). Part of my contribution to the topic has been to show that with this choice of σ actually (iii) is true, see below.

4.2 Existence and uniqueness

One of the two results that I proved in [9] was:

Theorem 4.4 *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be a metric measure space, with* (X, d) *complete and separable*[4] *and* $\mathfrak{m} \in \mathscr{P}(X)$*. Assume that it is a* $CD(K,\infty)$ *space. Then for any* $\mu \in D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathscr{P}_2(X)$ *there exists a unique gradient flow of* $\mathrm{Ent}_{\mathfrak{m}}$ *in* $(\mathscr{P}_2(X), W_2)$ *starting from* μ*.*

The proof of this theorem relies on the following lemma, which I believe of independent interest.

Lemma 4.5 *With the same assumption of Theorem 4.4 the squared slope* $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2$ *of the entropy is convex (w.r.t. affine interpolation) and lower semicontinuous w.r.t. weak convergence of measures on bounded sublevels of* $\mathrm{Ent}_{\mathfrak{m}}$*.*

I remark that this lemma has been the first example where the 'horizontal world' interacted with the 'vertical world' in the abstract setting: indeed, notice that the hypothesis of the lemma are related to the W_2-structure only, where the object which plays a role is the displacement interpolation, similarly, the object investigated, namely the squared slope, is defined via the interaction of the entropy functional and the Wasserstein geometry. Yet, the thesis tells that the squared slope is convex w.r.t. the classical affine interpolation of measures, an interpolation which is often unnatural in the context of optimal transport.

Before discussing which are the key ideas behind the proof of Lemma 4.5, I describe why it implies Theorem 4.4. For existence, it is sufficient to apply Theorem 4.3 with the weak topology as topology σ. For uniqueness, the proof is by contradiction: assume that for some

[4] to be precise, the proof that I gave for uniqueness works in general Polish spaces, while the one for existence only in locally compact spaces. The extension from the locally compact case to the general one follows from the tightness of the sublevels of $\mathrm{Ent}_{\mathfrak{m}}$, a remark which I missed in [9]. This observation (and much more) is contained in [4].

$\mu \in D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathscr{P}_2(X)$ there are two different gradient flows (μ_t^1), (μ_t^2) starting from it. Then for any $T > 0$ and $i = 1, 2$ we would have

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) = \mathrm{Ent}_{\mathfrak{m}}(\mu_T^i) + \frac{1}{2}\int_0^T |\dot{\mu}_t^i|^2 \,\mathrm{d}t + \frac{1}{2}\int_0^T |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t^i) \,\mathrm{d}t.$$

Now consider the curve $\mu_t := \frac{\mu_t^1 + \mu_t^2}{2}$. The convexity of W_2^2 w.r.t. affine interpolation easily gives that (μ_t) is absolutely continuous and that $2|\dot{\mu}_t|^2 \leq |\dot{\mu}_t^1|^2 + |\dot{\mu}_t^2|^2$ for a.e. t. By Lemma 4.5 we also get that $2|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t) \leq |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t^1) + |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t^2)$. Finally, the relative entropy is *strictly* convex, hence for any $T > 0$ such that $\mu_T^1 \neq \mu_T^2$ we would have

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) > \mathrm{Ent}_{\mathfrak{m}}(\mu_T) + \frac{1}{2}\int_0^T |\dot{\mu}_t|^2 \,\mathrm{d}t + \frac{1}{2}\int_0^T |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t) \,\mathrm{d}t,$$

which is impossible, because of inequality (4.2).

Notice that this uniqueness proof tells nothing about the contractivity of W_2 along the flow. Actually, it has been proved later by Ohta and Sturm ([45]) that contractivity fails on $\mathbb{R}^d$ equipped with the Lebesgue measure and any norm not coming from a scalar product.

Thus everything boils down in proving Lemma 4.5. In order explain the idea (I will focus on the convexity only, as then the lower semicontinuity follows by pretty standard approximations arguments), it is necessary to introduce the notion of *push forward by a plan.*

Definition 4.6 (Push forward via a plan) *Let $\mu \in \mathscr{P}(X)$ and $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ be such that $\mu \ll \pi^1_\#\boldsymbol{\gamma}$, say $\mu = \rho\pi^1_\#\boldsymbol{\gamma}$. Then the measure $\boldsymbol{\gamma}_\#\mu \in \mathscr{P}(X)$ is defined by*

$$\boldsymbol{\gamma}_\#\mu := \eta\pi^2_\#\boldsymbol{\gamma}, \qquad \eta \text{ being given by} \qquad \eta(y) := \int_X \rho(x) \,\mathrm{d}\boldsymbol{\gamma}_y(x),$$

where $\{\boldsymbol{\gamma}_y\}_y$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. the projection on the second marginal. I also denote by $\boldsymbol{\gamma}_\mu \in \mathscr{P}(X^2)$ the plan defined by $\mathrm{d}\boldsymbol{\gamma}_\mu := \rho \circ \pi^1 \mathrm{d}\boldsymbol{\gamma}$, so that $\boldsymbol{\gamma}_\#\mu = \pi^2\boldsymbol{\gamma}_\mu$.

This construction was firstly used, with a different notation, by Sturm in [43] to prove stability of Ricci curvature bound, and later independently rediscovered by myself and Savaré.

I will say that a plan $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ is of *bounded compression*, provided $c\mathfrak{m} \leq \pi^1_\#\boldsymbol{\gamma}, \pi^2_\#\boldsymbol{\gamma} \leq C\mathfrak{m}$ for some c, C and $\mathsf{d} \in L^\infty(\boldsymbol{\gamma})$. The set of plans with bounded compression is pretty rich, in particular it holds the following approximation lemma:

Lemma 4.7 *Let $(X, \mathsf{d}, \mathfrak{m})$ be a metric measure space and $\mu, \nu \in D(\mathrm{Ent}_{\mathfrak{m}})$. Then there exists a sequence $(\boldsymbol{\gamma}^n)$ of plans with bounded compression such that*

$$\mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}^n_\#\mu) \to \mathrm{Ent}_{\mathfrak{m}}(\nu),$$
$$\int \mathsf{d}^2(x, y) \,\mathrm{d}\boldsymbol{\gamma}_\mu(x, y) \to W_2^2(\mu, \nu).$$

The argument for the proof consists in picking an optimal plan $\boldsymbol{\gamma} \in \mathrm{Opt}(\mu, \nu)$, restricting it to the set $\{(x, y) : \frac{\mathrm{d}\mu}{\mathrm{d}\pi^1_{\boldsymbol{\gamma}}}(x) + \frac{\mathrm{d}\nu}{\mathrm{d}\pi^2_{\boldsymbol{\gamma}}}(y) + \mathsf{d}(x, y) < n\}$ and then adding a small multiple of $(\mathrm{Id}, \mathrm{Id})_\#\mathfrak{m}$.

A trivial relation between the construction of push forward via a plan and the transport problem is the fact that the map

$$\mu \qquad \mapsto \qquad C(\boldsymbol{\gamma}_\mu),$$

is linear on its domain of definition, where by $C(\boldsymbol{\gamma})$ we denote the cost of the plan $\boldsymbol{\gamma}$, given by $C(\boldsymbol{\gamma}) := \int \mathsf{d}^2(x,y)\mathrm{d}\boldsymbol{\gamma}(x,y)$.

What is non trivial, is the following relation between the operation $\boldsymbol{\gamma}_\#$ and the relative entropy functional:

Proposition 4.8 *Let $(X,\mathsf{d},\mathfrak{m})$ be a metric measure space and $\gamma \in \mathscr{P}(X^2)$ with bounded compression. Then the map*

$$\mu \qquad \mapsto \qquad E_{\boldsymbol{\gamma}}(\mu) := \mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\#\mu)$$

is convex (w.r.t. affine interpolation) on $D(\mathrm{Ent}_{\mathfrak{m}})$.

The first proof of a prototype version of this statement was due to Savaré in [41], and later generalized by myself in its current version in [9]. The proof is not hard: it follows by applying Jensen's inequality to the second derivative of the map $t \mapsto E_{\boldsymbol{\gamma}}((1-t)\mu_0 + t\mu_1)$.

Having said this, we are now ready to conclude the proof of Lemma 4.5. Here is where the K-geodesic convexity of the entropy comes into play: from the representation formula (4.1) we have

$$|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu) = \sup_\nu \frac{\Big(\big(\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\nu) - \frac{K^-}{2}W_2^2(\mu,\nu)\big)^+\Big)^2}{W_2^2(\mu,\nu)},$$

so that using the approximation Lemma 4.7 we also get

$$|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu) = \sup_{\boldsymbol{\gamma}} \frac{\Big(\big(E_{\boldsymbol{\gamma}}(\mu) - \frac{K^-}{2}C(\boldsymbol{\gamma}_\mu)\big)^+\Big)^2}{C(\boldsymbol{\gamma}_\mu)},$$

where the sup is taken among all plans with bounded compression. To conclude that this latter expression is convex, it is sufficient to show that for a given plan $\boldsymbol{\gamma}$ with bounded compression the map

$$\mu \qquad \mapsto \qquad \frac{\Big(\big(E_{\boldsymbol{\gamma}}(\mu) - \frac{K^-}{2}C(\boldsymbol{\gamma}_\mu)\big)^+\Big)^2}{C(\boldsymbol{\gamma}_\mu)},$$

is convex. But this is obvious, because from the convexity of $\mu \mapsto E_{\boldsymbol{\gamma}}$ and the linearity of $\mu \mapsto C(\boldsymbol{\gamma}_\mu)$ we deduce the convexity of

$$\mu \qquad \mapsto \qquad \Big(E_{\boldsymbol{\gamma}}(\mu) - \frac{K^-}{2}C(\boldsymbol{\gamma}_\mu)\Big)^+,$$

so that the conclusion follows using again the linearity of $\mu \mapsto C(\boldsymbol{\gamma}_\mu)$ in conjunction with the fact that the map $(a,b) \mapsto \frac{a^2}{b}$ is convex on $[0,\infty)^2$ and increasing in a.

4.3 Stability

If one accepts the definition of heat flow on a $CD(K,\infty)$ space as gradient flow of the relative entropy, as in the approach just described, the following natural question arises. Assume that $(X_n, \mathsf{d}_n, \mathfrak{m}_n)$ is a sequence of $CD(K,\infty)$ spaces converging to a limit space $(X, \mathsf{d}, \mathfrak{m})$ in the measured Gromov-Hausdorff sense, then do the heat flows on the approximating spaces converge to the heat flow in the limit space? As we will see in a moment, the answer is yes. For simplicity I will stick to the compact case.

Let me first point out that if the approximating spaces are Riemannian manifolds with dimension uniformly bounded from above, a consequence of the analysis done by Cheeger and Colding in [20] is that this question has indeed affirmative answer.

The setting under consideration here is, however, more general for two reasons, beside the obvious one that we are considering abstract non-smooth spaces: there is no a priori bound on the dimension, and the heat flow is potentially not linear. In particular, notice that this latter fact prevents the definition and study of the associate Dirichlet form (which is the path followed in [20]).

Let me recall that in proving stability of Ricci curvature bounds Lott and Villani on one side and Sturm on the other proved that if $(X_n, \mathsf{d}_n, \mathfrak{m}_n)$ converge to $(X, \mathsf{d}, \mathfrak{m})$ in the measured-Gromov-Hausdorff sense, then the relative entropies $\mathrm{Ent}_{\mathfrak{m}_n}$ Γ-converge to the relative entropy $\mathrm{Ent}_{\mathfrak{m}}$ in the limit space (up to isometrically embed all the spaces into a common one, which is always possible under Gromov-Hausdorff convergence). Since the notion of K-geodesic convexity is easily seen to be stable under Γ-convergence, stability of Ricci curvature bounds follows.

Therefore the question on stability of the heat flow fits into the more general one: are gradient flows of K-geodesically convex functionals stable under Γ-convergence? The answer is yes, but before turning to the relevant definitions, let me remark that since in general gradient flows of K-geodesically convex functionals are not unique, stability should be understood in the sense of closure, i.e. any limit of gradient flows is a gradient flow.

Recall that given a Polish space (Y, d_Y) and functionals $E_n, E : Y \to \mathbb{R} \cup \{+\infty\}$, $n \in \mathbb{N}$, one says that E_n Γ-converges to E, and writes $\Gamma - \lim_n E_n = E$ provided the following two are true:

$$
\begin{aligned}
E(y) &\leq \inf_{(y_n)} \varliminf_{n\to\infty} E_n(y_n), &\qquad &(\text{the } \Gamma - \varliminf \text{ inequality})\\
E(y) &\geq \inf_{(y_n)} \varlimsup_{n\to\infty} E_n(y_n), &\qquad &(\text{the } \Gamma - \varlimsup \text{ inequality})
\end{aligned}
\tag{4.4}
$$

where in both cases the inf is taken among all sequences $(y_n) \subset Y$ converging to y.

The stability result proven in [9] is then the following.

Theorem 4.9 (Stability of gradient flows) *Let (Y, d_Y) be a compact space and $E_n, E : Y \to \mathbb{R}\cup\{+\infty\}$, $n \in \mathbb{N}$, K-geodesically convex and lower semicontinuous functionals such that $\Gamma - \lim_n E_n = E$. Also, let $y_0 \in D(E)$, $(y_0^n) \subset Y$ be a sequence such that $\lim_{n\to\infty} E_n(y_0^n) = E(y_0)$ and for every n let (y_t^n) be a gradient flow of E_n starting from y_0^n. Then*

- *the sequence of curves (y_t^n) is relatively compact w.r.t. locally uniform convergence,*
- *any limit curve (y_t) is a gradient flow of E starting from z.*

The relative compactness of the sequence is a simple consequence of the definition of gradient flow: indeed notice that from (4.3) we get that for any $T > 0$ it holds

$$\int_0^T |\dot{y}_t^n|^2 \, \mathrm{d}t \leq E_n(y_0^n) - \inf E_n, \qquad \forall n \in \mathbb{N},$$

so that from the fact that $\overline{\lim}_n E_n(y_0^n) = E(y_0) < \infty$, the uniform bound from below and the compactness of Y we get the relative compactness of the set of curves.

Thus the problem is only in proving that any limit curve is actually a gradient flow, and up to pass to a subsequence I can assume that the full sequence of curves (y_t^n) is converging locally uniformly to (y_t). Thanks to inequality (4.2), the goal is to show that

$$E(y_0) \geq E(y_T) + \frac{1}{2}\int_0^T |\dot{y}_t|^2 \, \mathrm{d}t + \frac{1}{2}\int_0^T |\nabla^- E|^2(y_t) \, \mathrm{d}t, \qquad \forall T > 0,$$

knowing the same inequality for the (y_t^n)'s. From the assumptions it trivially follows that $\underline{\lim}_{n\to\infty} E_n(y_T^n) \geq E(y_T)$ and $\underline{\lim}_{n\to\infty} \int_0^T |\dot{y}_t^n|^2 \, \mathrm{d}t \geq \int_0^T |\dot{y}_t|^2 \, \mathrm{d}t$, thus to conclude it is sufficient to show that

$$|\nabla^- E|(z_t) \leq \underline{\lim}_{n\to\infty} |\nabla^- E_n|(z_t^n), \qquad \forall t \geq 0.$$

Here is where the K-geodesic convexity comes into play. Indeed, recall that from formula (4.1) we have

$$|\nabla^- E|(\tilde{z}) = \sup_y \left(\frac{E(\tilde{z}) - E(y)}{\mathsf{d}(\tilde{z}, y)} + \frac{K^-}{2}\tilde{\mathsf{d}}(\tilde{z}, y) \right)^+.$$

Now fix $\tilde{z}, y$ and find a sequence (y_n) realizing the $\Gamma - \overline{\lim}$ inequality for $E(y)$, and let $(\tilde{z}_n)$ be any sequence converging to $\tilde{z}$. The validity of

$$\begin{aligned} \tilde{\mathsf{d}}(\tilde{z}, y) &= \lim_{n\to\infty} \tilde{\mathsf{d}}(\tilde{z}_n, y_n), \\ E(\tilde{z}) &\leq \underline{\lim}_{n\to\infty} E_n(\tilde{z}_n), \\ E(y) &= \lim_{n\to\infty} E_n(y_n), \end{aligned}$$

gives

$$\begin{aligned} \left(\frac{E(\tilde{z}) - E(y)}{\tilde{\mathsf{d}}(\tilde{z}, y)} + \frac{K^-}{2}\tilde{\mathsf{d}}(\tilde{z}, y) \right)^+ &\leq \underline{\lim}_{n\to\infty} \left(\frac{E_n(\tilde{z}) - E_n(y)}{\tilde{\mathsf{d}}(\tilde{z}_n, y_n)} + \frac{K^-}{2}\tilde{\mathsf{d}}(\tilde{z}_n, y_n) \right)^+ \\ &\leq \underline{\lim}_{n\to\infty} |\nabla^- E_n|(\tilde{z}_n), \end{aligned}$$

and the conclusion.

An interesting consequence of the stability of the heat flow is the following result.

Theorem 4.10 *Let F be a Finsler manifold which is compact, smooth, and without boundary. Then F can be realized as Gromov-Hausdorff limit of a sequence of Riemannian manifolds with Ricci curvature uniformly bounded from below if and only if it is Riemannian.*

This result was already known under the additional assumption that the approximating sequence has dimension uniformly bounded from above: this came from the very fine analysis carried on by Cheeger and Colding in [20] that showed that a.e. point in the limit space must have a Euclidean tangent space, which is certainly a statement much stronger than the one of Theorem 4.10. On the other hand, if one cares only about ruling out Finsler geometries at the limit, the stability result provides a much simpler proof, which holds even without the upper bound on the dimension: it is sufficient to recall that the heat flow on a Finsler manifold is linear if and only if the manifold is Riemannian.

5 Sobolev spaces over metric measure spaces

5.1 Introduction

One of the results contained in the recent joint work with Ambrosio and Savaré on the calculus and heat flow over metric measure spaces [4], has been a finer description of the Sobolev spaces over a general metric measure space. Such topic has been deeply investigated in the last decades, I refer to [25] for an overview on the subject and detailed references. Yet, some basic questions remained unanswered, and one of our contributions has been to prove the following density result:

Theorem 5.1 *Let $(X, \mathsf{d}, \mathfrak{m})$ be a metric measure space with (X, d) Polish and $\mathfrak{m}$ locally finite. Then Lipschitz functions are dense in energy in $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$, i.e. for any $f \in W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ there exists a sequence (f_n) of Lipschitz functions converging to f in L^2 and such that $\|f_n\|_{W^{1,2}} \to \|f\|_{W^{1,2}}$.*

Such a result was known under the assumption that the measure is doubling and the space supports a local Poincare' inequality (thanks to the fine analysis done by Cheeger in [19]), but the general case was open. In the following I will deal with the technically simpler case in which $\mathfrak{m}$ is a probability measure.

To see the point, and why this question has been hard to answer to, I shall recall the usual definition of $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$. Although not needed, in what comes next I will always work under the simplifying assumption that $\mathfrak{m}$ is a finite measure: anyway, Theorem 5.1 will follow by some localization argument.

The first thing to understand when trying to propose such a definition, is that it is (necessary and) sufficient to provide a lower semicontinuous map $E : L^2 \to [0, \infty]$ which plays the role of the Dirichlet energy in $\mathbb{R}^d$: the semicontinuity of E ensures by standard arguments that the space

$$W^{1,2}(X, \mathsf{d}, \mathfrak{m}) := \Big\{ f \in L^2 \ : \ E(f) < \infty \Big\},$$

endowed with the norm

$$\|f\|_{W^{1,2}} := \sqrt{\|f\|_{L^2} + E(f)},$$

is a Banach space (this is the best we can hope for: even on $\mathbb{R}^d$ equipped with a norm not coming from a scalar product the space $W^{1,2}$ is not Hilbert).

In order to define E, two - a posteriori equivalents - paths have been followed: one is to proceed by relaxation, the other by looking at the upper gradient property along 'almost all curves'. For the first approach (followed by Cheeger in [19]), one says that G is an upper gradient for $f \in L^2(X, \mathfrak{m})$ provided the inequality

$$\big| f(\gamma_0) - f(\gamma_1) \big| \leq \int_\gamma G,$$

is true for any absolutely continuous curve $\gamma : [0, 1] \to X$, where by $\int_\gamma G$ is meant $\int_0^1 G(\gamma_t)|\dot\gamma_t| \, \mathrm{d}t$. Then the energy of $f \in L^2(X, \mathfrak{m})$ is defined as

$$E(f) := \inf \varliminf_{n \to \infty} \|G_n\|^2_{L^2},$$

where the infimum is taken among all sequences (f_n) converging to f in L^2, and the G_n's are upper gradients for f_n.

For the second approach (followed by Shanmugalingam in [42]) one needs a way to measure how big sets of curves are. The notion which comes into play is that of 2-Modulus, defined by

$$\mathrm{Mod}_2(\Gamma) := \inf\Big\{\|\rho\|_{L^2} \ : \ \int_\gamma \rho \geq 1, \ \ \forall \gamma \in \Gamma\Big\}, \qquad \forall \Gamma \subset AC([0,1],X)$$

Is it possible to show that the 2-Modulus is an outer measure on the space of absolutely continuous curves. Then one may say that $f \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$ provided there exists $\tilde f = f$ $\mathfrak{m}$-a.e. and $G \in L^2(X,\mathfrak{m})$ such that

$$\big|\tilde f(\gamma_0) - \tilde f(\gamma_1)\big| \leq \int_\gamma G, \qquad \forall \gamma \in AC([0,1],X) \setminus \mathcal{N},$$

where $\mathrm{Mod}_2(\mathcal{N}) = 0$. In this case, one defines the energy $\tilde E : L^2 \to [0,\infty]$ by putting

$$\tilde E(f) := \inf \|G\|^2_{L^2},$$

where the infimum is taken among all G's satisfying the previous condition. Thanks to the properties of the 2-modulus, it is possible to show that $\tilde E$ is indeed L^2 lower semicontinuous, so that it leads to a good definition of the Sobolev space. The advantage of this approach is that there is no need of relaxation, so that it is, at least in principle, easier to bound from below the size of upper gradients.

Using a key lemma due to Fuglede, Shanmugalingam proved in [42] that the two approaches are actually the same, so that both the constructions lead to the same Sobolev space and in particular $E = \tilde E$.

It is also important to observe that for functions $f \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$, not only is well defined the 'Dirichlet' energy $\frac{1}{2}E(f)$, but also a pointwise object $|\nabla f|_C$ which provides an integral representation E, in the sense this it satisfies $E(f) = \int |\nabla f|^2_C \,\mathrm{d}\mathfrak{m}$ and for any sequence of functions (f_n) converging to f in L^2 and any choice of upper gradients G_n for the f_n's, it holds $|\nabla f|_C \leq G$ $\mathfrak{m}$-a.e., where G is any weak limit in L^2 of (G_n).

Finally, let me underline that an application of Fuglede's lemma shows that the construction(s) presented here lead to the standard Sobolev space $H^1(\mathbb{R}^d)$ when applied to the Euclidean case. This fact also shows why the notion of 2-modulus is important.

To conclude this introduction, let me underline why Theorem 5.1 is not trivial. The point is that to know that a function f has an upper gradient G in L^2, while ensuring, for instance, that the function is absolutely continuous along 'most' curves, gives no informations at all on its Lipschitz constant. Even if one knows a priori that G is bounded (and typically one does not have such information) the conclusion is non trivial, unless one works in geodesic spaces (or more generally, λ-convex spaces, where the length of the minimal curve joining two points is controlled by λ-times the distance between the points themselves).

5.2 A new relaxation procedure: relaxed gradients

The simplest thing one might try to do in order to ensure weak density of Lipschitz functions in the Sobolev space is to change the relaxation procedure: rather than relaxing upper gradients, one can relax directly the local Lipschitz constant. More in detail, for a given function $f : X \to \mathbb{R}$ one defines its local Lipschitz constant $|\nabla f| : X \to [0,\infty]$ (as in (3.2)) by

$$|\nabla f|(x) := \varlimsup_{y \to x} \frac{|f(x) - f(y)|}{\mathsf{d}(x,y)},$$

where this lim sup is taken by definition 0 if x is isolated.

Then one can define the *Cheeger energy*[5] $\mathsf{Ch}(f) : L^2(X, \mathfrak{m}) \to [0, \infty]$ by

$$\mathsf{Ch}(f) := \inf_{(f_n)} \varliminf_{n\to\infty} \frac{1}{2} \int |\nabla f_n|^2 \, d\mathfrak{m},$$

where the inf is taken among all sequences (f_n) of Lipschitz functions converging to f in L^2.

As before, the lower semicontinuity of Ch ensures that the domain $D(\mathsf{Ch})$ of the Cheeger energy endowed with the norm $\|f\|^2_{W^{1,2}} := \|f\|^2_{L^2} + 2\mathsf{Ch}(f)$ is a Banach space. Also, as before, for functions $f \in D(\mathsf{Ch})$ there exists a pointwise representative $|\nabla f|_*$, which we called *relaxed gradient* in [4], such that $\mathsf{Ch}(f) = \frac{1}{2} \int |\nabla f|_*^2 \, d\mathfrak{m}$. This relaxed gradient has the characterizing property that $|\nabla f|_* \le G$ whenever G is the weak limit of $(|\nabla f_n|)$, the sequence (f_n) being made of Lipschitz functions converging to f in $L^2(X, \mathfrak{m})$.

It is also easy to check that for the relaxed gradient they hold the standard calculus rules available on a metric setting, i.e.

$$\begin{aligned}
|\nabla(\alpha f + \beta g)|_* &\le |\alpha||\nabla f|_* + |\beta||\nabla g|_*, &&\forall f, g \in D(\mathsf{Ch}),\ \alpha, \beta \in \mathbb{R}\\
|\nabla(fg)|_* &\le |f||\nabla g|_* + |g||\nabla f|_*, &&\forall f, g \in D(\mathsf{Ch}) \cap L^\infty,\\
|\nabla(\varphi \circ f)|_* &\le |\varphi' \circ f||\nabla f|_*, &&\forall f \in D(\mathsf{Ch}),\ \varphi \in C^1(\mathbb{R}) \cap \mathrm{Lip}(\mathbb{R}),
\end{aligned}$$

where the last inequality is an equality if φ is non decreasing.

In summary, the Cheeger's energy has all the properties one wishes from an energy defined on a metric space, and the problem is 'only' to show that it coincides with the energy E previously defined. In order to prove this, we need to develop some calculus rule and to call into play optimal transport.

Notice that the only trivial relation which is granted from the construction, is that

$$|\nabla f|_* \ge |\nabla f|_C, \qquad \mathfrak{m} - a.e., \tag{5.1}$$

whenever $f \in D(\mathsf{Ch})$. This can be seen from the fact that for a Lipschitz function f, the local Lipschitz constant is certainly an upper gradient, hence in the relaxation procedure which gives the definition of $|\nabla f|_C$ we are relaxing over a set bigger than the one used to define $|\nabla f|_*$.

From the definition of Ch we can extract a definition of Laplacian in a purely variational way. Indeed, from the definition of Ch we see that not only it is lower semicontinuous on L^2, but also convex. Hence its subdifferential $\partial\mathsf{Ch}(f)$ is a well defined object: given $f \in D(\mathsf{Ch})$ we say that $v \in \partial\mathsf{Ch}(f)$ provided

$$\mathsf{Ch}(f) + \int v(g - f) \, d\mathfrak{m} \le \mathsf{Ch}(g), \qquad \forall g \in L^2(X, \mathfrak{m}).$$

Notice that $\partial\mathsf{Ch}(f)$ is closed and convex, possibly empty.

Then we say that f is in the domain of the Laplacian provided $\partial\mathsf{Ch}(f) \ne \emptyset$, and in this case we define $\Delta f := -v$, where v is the element of minimal norm in the subdifferential of

[5] the name 'Cheeger energy' has been preferred over 'Dirichlet energy' both to acknowledge the important contribution of Cheeger to analysis in metric spaces, and to avoid potential confusion: Ch is typically not a quadratic function, which is the same as to say that $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ can be not an Hilbert space, while the terminology 'Dirichlet energy' strongly reminds Dirichlet forms

Ch at f. It is a one line computation to check that this definition is consistent with the classical one for the space $H^1(\mathbb{R}^d)$. Notice that the Laplacian is a linear operator if and only if $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ is Hilbert. Still, it is always 1-homogeneous.

The standard theory of gradient flow of convex functionals in an Hilbert setting (see for instance [18]) ensures that for any $f \in L^2$ there exists a unique curve $[0,\infty) \ni t \mapsto f_t \in L^2$ which is locally absolutely continuous, converges to f when $t \downarrow 0$, and satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t \in -\partial \mathsf{Ch}(f_t), \qquad a.e.\ t. \tag{5.2}$$

This curve is called *gradient flow* of Ch starting from f (this definition is consistent with the abstract one given in Definition 4.2). It also has the following property:

$$\frac{\mathrm{d}^+}{\mathrm{d}t} f_t = \Delta f_t, \qquad \forall t \geq 0,$$

which justifies the definition of Laplacian. It is also easy to check that the gradient flow is mass preserving (i.e. $\int f_t \,\mathrm{d}\mathfrak{m} = \int f_0 \,\mathrm{d}\mathfrak{m}$ for any $t \geq 0$) and that it obeys the maximum principle (i.e. $f_0 \geq c$ implies $f_t \geq c$ for any $t \geq 0$, and similarly for bounds from above).

Interesting enough, for the Laplacian they hold calculus rules which are strongly reminiscent of those valid in a smooth setting:

$$\begin{aligned} \left| \int g \Delta f \,\mathrm{d}\mathfrak{m} \right| &\leq \int |\nabla f|_* |\nabla g|_* \,\mathrm{d}\mathfrak{m}, \qquad \forall f \in D(\Delta),\ g \in D(\mathsf{Ch}), \\ \int \varphi \circ f \Delta f \,\mathrm{d}\mathfrak{m} &= -\int \varphi' \circ f |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m}, \end{aligned} \tag{5.3}$$

where in the second formula $\varphi : \mathbb{R} \to \mathbb{R}$ is C^1, Lipschitz and non decreasing. The proof of these formulas follows directly from the deifnition, for instance, the first one follows noticing that for any $v \in \partial \mathsf{Ch}(f)$ it holds

$$\frac{1}{2} \int |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m} + \varepsilon \int v g \,\mathrm{d}\mathfrak{m} \leq \frac{1}{2} \int |\nabla f + \varepsilon g|_*^2 \,\mathrm{d}\mathfrak{m},$$

and using the inequality $|\nabla f + \varepsilon g|_* \leq |\nabla f|_* + \varepsilon |\nabla g|_*$.

An immediate consequence of the 'integration by parts' formula (5.3) is that if (f_t) is a gradient flow of Ch and $f_0 \geq c > 0$, then it holds

$$\frac{\mathrm{d}}{\mathrm{d}t} \int f_t \log f_t \,\mathrm{d}\mathfrak{m} = \int (\log f_t + 1) \Delta f_t \,\mathrm{d}\mathfrak{m} = -\int \frac{|\nabla f_t|_*^2}{f_t} \,\mathrm{d}\mathfrak{m}, \qquad a.e.\ t, \tag{5.4}$$

which is consistent with the entropy dissipation formula along the heat flow valid in a smooth setting.

What is totally non trivial, is the relation of the gradient flow of Ch with the Wasserstein distance W_2. Such relation, firstly exploited in [14], will be key both in proving the density of Lipschitz functions in the Sobolev space, and in proving that in $CD(K,\infty)$ spaces the W_2-gradient flow of the entropy coincides with the L^2-gradient flow of Ch. What makes the following lemma non trivial is the fact that puts in relation two different worlds: on one side we have the L^2-gradient flow of Ch, on the other we are relating such flow with the Wasserstein distance, which is a priori a totally different object.

Lemma 5.2 (Key estimate for the Wasserstein velocity) *Let $f_0 \in L^2(X, \mathfrak{m})$ be such that $\int f_0 \, \mathrm{d}\mathfrak{m} = 1$ and $f_0 \geq c > 0$ for some c, and consider the gradient flow (f_t) of Ch starting from f_0. Define the measures $\mu_t := f_t \mathfrak{m}$ (which are probability measures, thanks to the mass preservation and the maximum principle). The the curve $t \mapsto \mu_t \in \mathscr{P}(X)$ is locally absolutely continuous w.r.t. W_2 and for its metric speed it holds the bound*

$$|\dot{\mu}_t|^2 \leq \int \frac{|\nabla f_t|_*^2}{f_t} \, \mathrm{d}\mathfrak{m}, \qquad a.e.\ t. \tag{5.5}$$

Notice that the bound provided is sharp, in the sense that on smooth spaces equality holds. The idea to prove this lemma, which is due to Kuwada and appeared firstly in the technically simpler case of Alexandrov spaces in [14], is to pass to the dual formulation of optimal transport and then to use the properties of the Hopf-Lax semigroup. Indeed, we know that it holds

$$\frac{W_2^2(\mu_t, \mu_s)}{2} = \sup_{\varphi \in \mathrm{Lip}(X)} \int \varphi \, \mathrm{d}\mu_t + \int \varphi^c \, \mathrm{d}\mu_s = \sup_{\psi \in \mathrm{Lip}(X)} \int Q_1(\psi) \, \mathrm{d}\mu_s - \int \psi \, \mathrm{d}\mu_t, \tag{5.6}$$

where $Q_1(\psi)$ is defined via the Hopf-Lax formula as in (3.1). Since $t \mapsto Q_t(\psi)$ is Lipschitz with values in $C_b(X)$, it is easy to check that the following calculations are justified:

$$\begin{aligned}
&\int Q_1(\psi) \, \mathrm{d}\mu_s - \int \psi \, \mathrm{d}\mu_t \\
&\qquad = \int_0^1 \int \frac{\mathrm{d}}{\mathrm{d}t} \big(Q_r(\psi) f_{t+r(s-t)} \big) \, \mathrm{d}\mathfrak{m} \, \mathrm{d}r \\
&\qquad = \iint_0^1 f_{t+r(s-t)} \frac{\mathrm{d}}{\mathrm{d}t} Q_r(\psi) + Q_r(\psi) \frac{\mathrm{d}}{\mathrm{d}t} f_{t+r(s-t)} \, \mathrm{d}r \, \mathrm{d}\mathfrak{m} \\
&\qquad \overset{(3.3)}{\leq} \iint_0^1 -\frac{|\nabla Q_r|^2}{2} f_{t+r(s-t)} + (s-t) Q_r(\psi) \Delta f_{t+r(s-t)} \, \mathrm{d}r \, \mathrm{d}\mathfrak{m} \\
&\qquad \overset{(5.3)}{\leq} \iint_0^1 -\frac{|\nabla Q_r|^2}{2} f_{t+r(s-t)} + (s-t) |\nabla Q_r(\psi)|_* |\nabla f_{t+r(s-t)}|_* \, \mathrm{d}r \, \mathrm{d}\mathfrak{m} \\
&\qquad \leq \iint_0^1 \left(-\frac{|\nabla Q_r|^2}{2} + \frac{|\nabla Q_r \psi|_*^2}{2} \right) f_{t+r(s-t)} + \frac{(s-t)^2 |\nabla f_{t+r(s-t)}|_*^2}{2 f_{t+r(s-t)}} \, \mathrm{d}r \, \mathrm{d}\mathfrak{m} \\
&\qquad \leq \frac{(s-t)^2}{2} \iint_0^1 \frac{|\nabla f_{t+r(s-t)}|_*^2}{f_{t+r(s-t)}} \, \mathrm{d}r \, \mathrm{d}\mathfrak{m},
\end{aligned}$$

where in the last equality we used the fact that for any Lipschitz function g it holds $|\nabla g|_* \leq |\nabla g|$ $\mathfrak{m}$-a.e., which is a trivial consequence of the definition. Thus we found a bound on $\int Q_1(\psi) \, \mathrm{d}\mu_s - \int \psi \, \mathrm{d}\mu_t$ which is independent on ψ. Equation (5.6) then gives

$$\frac{W_2^2(\mu_t, \mu_s)}{2} \leq \frac{(s-t)^2}{2} \iint_0^1 \frac{|\nabla f_{t+r(s-t)}|_*^2}{f_{t+r(s-t)}} \, \mathrm{d}r \, \mathrm{d}\mathfrak{m},$$

from which Lemma 5.2 easily follows.

5.3 A new notion of null set of curves: weak gradients

Here I recall another possible definition of Sobolev space on a metric measure space, inspired to the approach of Shanmugalingam, and in the next section I will show why all these notions of 'norm of gradient' actually coincide. I will denote by $\mathrm{e}_t : C([0,1],X) \to X$, $t \in [0,1]$, the evaluation map defined by

$$\mathrm{e}_t(\gamma) := \gamma_t.$$

Definition 5.3 (Test plans and negligible set of curves) *A Borel probability measure* $\boldsymbol{\pi} \in \mathscr{P}(C([0,1]),X)$ *is a test plan provided* $\iint_0^1 |\dot\gamma_t|^2 \,\mathrm{d}t\mathrm{d}\boldsymbol{\pi}(\gamma) < \infty$ *and there exists* $C > 0$ *such that*

$$(\mathrm{e}_t)_\# \boldsymbol{\pi} \le C\mathfrak{m}, \qquad \forall t \in [0,1].$$

A Borel set $\Gamma \subset AC^2([0,1],X)$ *is said negligible provided* $\boldsymbol{\pi}(\Gamma) = 0$ *for any test plan* $\boldsymbol{\pi}$*. A property which is true for any curve* $\gamma \in AC^2([0,1],X)$ *except a negligible set is said to hold for a.e. curve.*

Notice that if $\mathrm{Mod}_2(\Gamma) = 0$, then Γ is negligible. Indeed for $\boldsymbol{\pi}$ test plan and $\rho \in L^2(X,\mathfrak{m})$ such that $\int_\gamma \rho \ge 1$ for any $\gamma \in \Gamma$ we have

$$\begin{aligned}
\boldsymbol{\pi}(\Gamma) &\le \iint_\gamma \rho \,\mathrm{d}\boldsymbol{\pi}(\boldsymbol{\gamma}) = \iint_0^1 \rho(\gamma_t)|\dot\gamma_t| \,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma) \\
&\le \sqrt{\iint_0^1 \rho^2(\gamma_t)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma)}\sqrt{\iint_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma)} \\
&\le \sqrt{C}\sqrt{\int \rho^2(x)\,\mathrm{d}\mathfrak{m}(x)}\sqrt{\iint_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma)},
\end{aligned}$$

which implies the claim.

Coupled with the notion of negligible set, we can give the definition of functions which are Sobolev along a.e. curve:

Definition 5.4 (Functions Sobolev along a.e. curve) *We say that* $f : X \to \mathbb{R}$ *is Sobolev along a.e. curve provided for a.e. curve the map* $t \mapsto f(\gamma_t)$ *coincides a.e. in* $[0,1]$ *and in* $\{0,1\}$ *with an absolutely continuous map* f_γ*. In this case we say that* $G : X \to [0,\infty]$ *is a weak upper gradient for* f *provided*

$$\big|f(\gamma_0) - f(\gamma_1)\big| \le \int_\gamma G, \qquad a.e.\ \gamma. \tag{5.7}$$

The definition of negligible set of curves is given to ensure basic invariance and lower-semicontinuity properties, so that, for instance, if (f_n) is a sequence of functions Sobolev along a.e. curve which converges $\mathfrak{m}$-a.e. to f, and if G_n is a weak upper gradient and $G_n \to G$ weakly in $L^2(X,\mathfrak{m})$, then f is Sobolev along a.e. curve and G is a weak upper gradient of f.

Also, it is possible to show that for f Sobolev along a.e. curve there exists a minimal $\mathfrak{m}$-a.e. function G satisfying (5.7): we will denote such G by $|\nabla f|_w$ and call it *minimal weak gradient.*

Since we know that a Mod_2 null set is also a negligible set, we obtain immediately from the definition that $|\nabla f|_C \ge |\nabla f|_w$ $\mathfrak{m}$-a.e., which coupled with (5.1) gives the chain of inequalities

$$|\nabla f|_w \le |\nabla f|_C \le |\nabla f|_*, \qquad \mathfrak{m} - a.e.,$$

so that to conclude the proof of Theorem 5.1 we need to show that $|\nabla f|_w \ge |\nabla f|_*$ $\mathfrak{m}$-a.e..

In order to describe the key argument to achieve this inequality, I need to recall the *superposition principle* valid in a generic metric space (proved by Lisini in [28], generalizing the arguments in [3] for the Euclidean setting).

Theorem 5.5 (Superposition principle) *Let $(\mu_t) \subset \mathscr{P}(X)$ be an $AC^2([0,1], \mathscr{P}(X))$ curve w.r.t. the distance W_2. Then there exists $\boldsymbol{\pi} \in \mathscr{P}(C([0,1], X))$ concentrated on $AC^2([0,1], X)$ such that*

$$
\begin{aligned}
(e_t)_\# \boldsymbol{\pi} &= \mu_t, \qquad \forall t \in [0,1], \\
\int |\dot\gamma_t|^2 \, \mathrm{d}\boldsymbol{\pi}(\gamma) &= |\dot\mu_t|^2, \qquad a.e.\ t \in [0,1].
\end{aligned}
\tag{5.8}
$$

Our job is now to prove that $|\nabla f|_w \geq |\nabla f|_*$ $\mathfrak{m}$-a.e., and in order to do so the first thing we need to do is to find a way to bound from below the minimal weak gradient $|\nabla f|_w$. I remark that, to some extent, this should not be an hard task, because the minimal weak gradient is something which is defined in inequality (5.7) as to be greater or equal than something else (as opposed to the relaxed gradient $|\nabla f|_*$ which is defined by relaxation and thus is 'naturally less or equal than something'). The only thing we need to do, roughly said, is to find some test plan in order to be able to check inequality (5.7) along sufficiently many curves.

We will achieve this by looking at the gradient flow of Ch. Indeed, we know that given $f_0 \in L^2(X, \mathfrak{m})$ such that $0 < c \leq f_0 \leq C < \infty$ and $\int f_0 \, \mathrm{d}\mathfrak{m} = 1$, the gradient flow (f_t) satisfies $0 < c \leq f_t \leq C < \infty$ and $\int f_t \, \mathrm{d}\mathfrak{m}$ for any $t \geq 0$. Furthermore, by Lemma 5.2 we also know that the curve $t \mapsto \mu_t := f_t \mathfrak{m}$ is absolutely continuous w.r.t. W_2 and its metric derivative satisfies the bound (5.5). Now we apply the superposition principle given by Theorem 5.5 to get the existence of a plan $\boldsymbol{\pi} \in \mathscr{P}(C([0,1], X))$ satisfying (5.8). The maximum principle ensures that this plan is a test plan. Hence the following calculation is justified:

$$
\begin{aligned}
\int f_0 \log f_0 \, \mathrm{d}\mathfrak{m} - \int f_t \log f_t \, \mathrm{d}\mathfrak{m} &\leq \int \log f_0 (f_0 - f_t) \, \mathrm{d}\mathfrak{m} \\
&= \int \log(f_0 \circ \mathrm{e}_0) - \log(f_0 \circ \mathrm{e}_t) \, \mathrm{d}\boldsymbol{\pi} \\
&\leq \iint_0^1 |\nabla \log(f_0)|_w(\gamma_t) |\dot\gamma_t| \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\pi}(\gamma) \\
&\leq \frac{1}{2} \iint_0^1 |\nabla \log(f_0)|_w^2(\gamma_t) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\pi}(\gamma) + \frac{1}{2} \iint_0^1 |\dot\gamma_t|^2 \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\pi}(\gamma) \\
&= \frac{1}{2} \iint_0^1 \frac{|\nabla f_0|_w^2}{f_0^2} f_t \, \mathrm{d}t \, \mathrm{d}\mathfrak{m} + \frac{1}{2} \iint_0^1 |\dot\gamma_t|^2 \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\pi}(\gamma)
\end{aligned}
\tag{5.9}
$$

5.4 Identification of the two gradients

At this stage, we have all the ingredients needed to conclude that $|\nabla f|_* = |\nabla f|_w$. Indeed from (5.9), (5.8) and (5.5) we get

$$
\lim_{t \downarrow 0} \frac{\int f_0 \log f_0 \, \mathrm{d}\mathfrak{m} - \int f_t \log f_t \, \mathrm{d}\mathfrak{m}}{t} \leq \frac{1}{2} \int \frac{|\nabla f_0|_w^2}{f_0} \, \mathrm{d}\mathfrak{m} + \frac{1}{2} \int \frac{|\nabla f_0|_*^2}{f_0} \, \mathrm{d}\mathfrak{m},
$$

while from (5.4) we know that

$$
\lim_{t \downarrow 0} \frac{\int f_0 \log f_0 \, \mathrm{d}\mathfrak{m} - \int f_t \log f_t \, \mathrm{d}\mathfrak{m}}{t} = \int \frac{|\nabla f_0|_*^2}{f_0} \, \mathrm{d}\mathfrak{m}.
$$

Hence it must hold

$$\int \frac{|\nabla f_0|_w^2}{f_0}\,\mathrm{d}\mathfrak{m} \geq \int \frac{|\nabla f_0|_*^2}{f_0}\,\mathrm{d}\mathfrak{m},$$

which, together with the pointwise estimate $|\nabla f|_w \leq |\nabla f|_*$ grants the conclusion.

6 Two points of view on the heat flow as gradient flow

It well known that on $\mathbb{R}^d$ the gradient flow of the Dirichlet energy in L^2 coincides with the gradient flow of the relative entropy in the Wasserstein space, as both produces solutions to the heat equation. More precisely, given $f \in L^2(\mathbb{R}^d, \mathcal{L}^d)$ such that $\mu := f\mathcal{L}^d \in \mathscr{P}_2(\mathbb{R}^d)$, if we let (f_t) be the gradient flow of the Dirichlet energy starting from f and (μ_t) the one of the relative entropy starting from μ, it holds $\mu_t = f_t\mathcal{L}^d$ for any $t \geq 0$.

The same property is true also on Riemannian manifolds (proved by Erbar in [22], see also Villani's arguments in [46]) and on Finsler manifolds (by Ohta and Sturm in [37]).

In all these situations, the structure of the proof of the identification is the same:

- One studies the gradient flow of the Dirichlet energy in L^2 and writes down the PDE solved by it,
- then he studies also the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ w.r.t. W_2 and writes down the PDE solved by it,
- he realizes the the two PDEs are actually the same: the heat equation.
- Finally, and this is the key step, he calls into play his knowledge of PDEs to assert that solutions to the heat equation uniquely depend on their initial data. Hence the two gradient flows must coincide.

Thus we know that in all the smooth situations the two gradient flows coincide. It is then natural to guess that there must be some deep reasons in order for this equivalence to be true, and that the same should hold in a non smooth context as well. The natural abstract setting where to work on is that of $CD(K, \infty)$ spaces, because, as already recalled in Chapter 4 we know from the study of Riemannian geometry that a sufficient condition in order for the heat flow to preserve the mass is that the Ricci curvature is bounded from below (mass preservation is important if we want to work with the Wasserstein distance).

On $CD(K, \infty)$ spaces we have a natural analogue for the Dirichlet energy: the Cheeger energy Ch (which, as discussed, is well defined on general metric measure spaces), and we also studied its gradient flow in L^2. Also, by the discussion made in Chapter 4 we know that the gradient flow of the relative entropy w.r.t. W_2 exists and is unique.

Thus the question is: do these two gradient flows coincide in this generality?

The difficulty in answering this question relies on the fact that the strategy used for the smooth case cannot be applied, because we don't have any a priori uniqueness result for the solutions of the heat equation (written, for instance, as in (5.2)). And even if such uniqueness result were available, we would still have the problem to show that the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ satisfies the heat equation.

Hence a new approach is needed. Such approach has been developed firstly in [14], and later generalized in [4] , where we proved the following result:

Theorem 6.1 *Let $(X, \mathsf{d}, \mathfrak{m})$ be a $CD(K, \infty)$ space, $f_0 \in L^2(X, \mathfrak{m})$ such that $\mu_0 := f_0\mathfrak{m} \in \mathscr{P}_2(X)$ and define $(f_t) \subset L^2(X, \mathfrak{m})$ as the gradient flow of Ch w.r.t. L^2 starting from f_0 and $(\mu_t) \subset \mathscr{P}_2(X)$ as the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ w.r.t. W_2 starting from μ_0.*

Then $\mu_t = f_t\mathfrak{m}$ for any $t \geq 0$.

The proof of this result has been one of the main research goals that I tried to accomplish in the last 2 years, and the paper [9] where I prove existence, uniqueness and stability of the W_2-gradient flow of the relative entropy should be regarded as the first step in this direction. Also, all the technical tools like the study of the Hamilton-Jacobi equation and the refined analysis of the Sobolev spaces discussed in Chapters 3 and 5 have been developed exactly to achieve this result (although I believe that they are of independent interest).

With all the machinery that we have at disposal now, the proof is not hard. Indeed, we know from Theorem 4.4 that the W_2-gradient flow of the entropy is unique, so that to conclude it is sufficient to show that the curve $t \mapsto \nu_t := f_t\mathfrak{m}$ satisfies the inequality

$$\mathrm{Ent}_{\mathfrak{m}}(\nu_0) \geq \mathrm{Ent}_{\mathfrak{m}}(\nu_T) + \frac{1}{2}\int_0^T |\dot{\nu}_t|^2 \,\mathrm{d}t + \frac{1}{2}\int_0^T |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\nu_t)\,\mathrm{d}t, \qquad \forall T > 0.$$

Thanks to the absolute continuity of $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\nu_t) \in \mathbb{R}$ and of $t \mapsto \nu_t \in \mathscr{P}_2(X)$ (the latter being proved in Lemma 5.2), it is sufficient to show that

$$\begin{aligned}
-\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}_{\mathfrak{m}}(\nu_t) &= \int \frac{|\nabla f_t|_*^2}{f_t}\,\mathrm{d}\mathfrak{m}, \qquad && a.e.\ t,\\
|\dot{\nu}_t|^2 &\leq \int \frac{|\nabla f_t|_*^2}{f_t}\,\mathrm{d}\mathfrak{m}, \qquad && a.e.\ t,\\
|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(f\mathfrak{m}) &\leq \int \frac{|\nabla f|_*^2}{f}\,\mathrm{d}\mathfrak{m}, \qquad && \forall f \in L^2(X,\mathfrak{m}),\ s.t.\ f\mathfrak{m} \in \mathscr{P}_2(X).
\end{aligned}$$

The first equation is proved in 5.4, and the second one in Lemma 5.2, so that to conclude the proof it is sufficient to prove the bound from above on the slope. This is not hard, as Lott and Villani proved in [33] the same statement with the local Lipschitz constant in place of the relaxed gradient, so that to conclude it is enough to use the lower semicontinuity of the slope w.r.t. weak convergence of measures.

7 Riemannian Ricci curvature bounds

We already discussed the definition of spaces with Ricci curvature bounded below, and concerning the appearance of Finsler geometries we saw that:

- Smooth Finsler manifolds are $CD(K,\infty)$ spaces for appropriate K.
- Finsler manifolds cannot arise as limit of Riemannian manifolds with Ricci curvature bounded from below.

In particular, this shows that the class of $CD(K,\infty)$ spaces strictly contains the closure of the class of Riemannian manifolds with uniform bounds from below on the Ricci curvature. Thus, regardless of whether one wants or not to include in the study Finsler geometries, it is natural to ask if there is a more restrictive notion of Ricci bound which rules them out, in order to have a better description of the closure of the class of Riemannian manifolds.

In the recent paper [5] written in collaboration with Ambrosio and Savaré we made a first step in this direction by introducing the notion of spaces with Riemannian Ricci curvature bounded from below (in short $RCD(K,\infty)$ spaces). Up to minor technicalities - which I won't discuss here - what we add to the $CD(K,\infty)$ condition is the linearity of the heat flow, or, which is the same, we ask for the Sobolev space $W^{1,2}$ to be an Hilbert space.

Clearly, Riemannian manifolds with Ricci curvature bounded below by K are $RCD(K,\infty)$ spaces, as they are $CD(K,\infty)$ spaces and the heat flow is linear on them. Also, from the studied made in [40], [49], [36] and [15] we also know that finite dimensional Alexandrov spaces with curvature bounded below are $RCD(K,\infty)$ spaces as well.

The stability of this notion can be deduced, for instance, from the stability result discussed in Section 4.3, which grants that if the approximating sequence of spaces have linear heat flow, then the same is true for the limit space as well.

Hence $RCD(K,\infty)$ spaces have the same basic properties of $CD(K,\infty)$ spaces, which gives to this notion the right of being called a synthetic (or weak) notion of Ricci curvature bound.

The point is then to understand which kind of gains about analysis/geometry of these spaces one gets from adding this linearity condition. A first non trivial consequence is that the heat flow K-contracts the distance W_2, i.e.

$$W_2(\mu_t,\nu_t)\leq e^{-Kt}W_2(\mu_0,\nu_0),\qquad \forall t\geq 0,$$

whenever $(\mu_t),(\nu_t)\subset\mathscr{P}_2(X)$ are two gradient flows of the entropy.

By a duality argument (see [27]) this property implies the Bakry-Emery gradient estimate

$$|\nabla \mathsf{h}_t(f)|_*^2(x)\leq e^{-2Kt}\mathsf{h}_t(|\nabla f|_*^2)(x),\qquad \forall t\geq 0,\ \mathfrak{m}-a.e.\ x,$$

where $\mathsf{h}_t:L^2(X,\mathfrak{m})\to L^2(X,\mathfrak{m})$ is the heat flow seen as gradient flow of Ch. If $(X,\mathsf{d},\mathfrak{m})$ is doubling and supports a local Poincaré inequality, then also the Lipschitz regularity of the heat kernel is deduced (following an argument described in [14]).

Also, notice that the linearity of the heat flow is equivalent to the fact that Ch is a quadratic form, which means that

$$\mathcal{E}(f,g):=\mathsf{Ch}(f+g)-\mathsf{Ch}(f)-\mathsf{Ch}(g),\qquad \forall f,g\in W^{1,2}(X,\mathsf{d},\mathfrak{m}),$$

induces a closed Dirichlet form on $L^2(X,\mathfrak{m})$ (closure follows from the L^2-lower semicontinuity of Ch). Hence it is natural to compare the calculus on $RCD(K,\infty)$ spaces with the abstract one available for Dirichlet forms (see [24]). The picture here is pretty clear and, I believe, satisfactory. Recall that to $f \in D(\mathcal{E})$ one can associate the energy measure $[f]$ defined by

$$[f](\varphi) := -\mathcal{E}(f, f\varphi) + \mathcal{E}(f^2/2, \varphi).$$

Using the calculus tools developed in the previous chapters, we are able to show that under the only assumption that $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ is Hilbert (so no bound on the curvature is needed here), the energy measure coincides with $|\nabla f|_*^2\mathfrak{m}$. Also, if $(X,\mathsf{d},\mathfrak{m})$ is a $RCD(K,\infty)$ space, then we can also show that the distance d coincides with the intrinsic distance $\mathsf{d}_{\mathcal{E}}$ induced by the form, which is defined by

$$\mathsf{d}_{\mathcal{E}}(x,y) := \sup\Big\{|g(x)-g(y)| \ : \ g \in D(\mathcal{E}) \cap C(X),\ [g] \le \mathfrak{m}\Big\}.$$

Taking advantage of these identification and of the locality of $\mathcal{E}$ (which is a consequence of the locality of the notion $|\nabla f|_*$), one can also see that on $RCD(K,\infty)$ spaces it is well defined a continuous Brownian motion.

Finally, for $RCD(K,\infty)$ spaces we have been able to prove tensorization and locality properties which are in line to those available for $CD(K,\infty)$ spaces.

References

Personal publications

[1] L. Ambrosio and N. Gigli, *Constrution of the parallel transport in the Wasserstein space*, Meth. Appl., 15 (2010), pp. 1–30.

[2] ——, *User's guide to optimal transport theory*, Lecture notes of C.I.M.E. summer school, (2011).

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[4] ——, *Calculus and heat flows in metric measure spaces with Ricci curvature bounded from below*, to appear, (2011).

[5] ——, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, to appear, (2011).

[6] A. Figalli and N. Gigli, *Local semiconvexity of Kantorovich potentials on non-compact manifolds*, ESAIM, Control Optim. Calc. Var., (2011).

[7] ——, *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, J. Math. Pures Appl., 94 (August 2010), pp. 107–130.

[8] N. Gigli, *On the weak closure of convex sets of probability measures*, Rend. di Mat., 29 (2009), pp. 133–141.

[9] ——, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations, 39 (2010), pp. 101–120.

[10] ——, *Second order calculus on* $(\mathscr{P}_2(M), W_2)$, Memoirs of the AMS, (2010).

[11] ——, *On Hölder continuity-in-time of the optimal transport map towards measures along a curve*, Proc. Edim. Math. Soc., 54 (2011), pp. 401–409.

[12] ——, *On the inverse implication of Brenier-McCann theorems and the structure of* $(\mathscr{P}_2(M), W_2)$, Meth. Appl., (2011).

[13] ——, *A variational approach to the Navier-Stokes equations*, to appear, (2011).

[14] N. Gigli, K. Kuwada, and S.-I. Ohta, *Heat flow on Alexandrov spaces*, to appear, (2010).

[15] N. Gigli and S.-I. Ohta, *First variation formula in Wasserstein spaces over compact Alexandrov spaces*, Canad. Math. Bull., (2011).

[16] N. Gigli and F. Otto, *Entropic Burgers' equation via a minimizing movement scheme based on the Wasserstein metric*, Calc. Var. and Part. Diff. Eq., (2011).

Miscellaneous

[17] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[18] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[19] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal., 9 (1999), pp. 428–517.

[20] J. Cheeger and T. Colding, *On the structure of spaces with Ricci curvature bounded below I*, J. Diff. Geom., 46 (1997), pp. 37–74.

[21] C. Dafermos, *The entropy rate admissibility criterion for solutions of hyperbolic conservation laws*, J. Diff. Eq., 14 (1973), pp. 202–212.

[22] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Annales de l'Institut Henri Poincaré - Probabilités et Statistiques, 46 (2010), pp. 1–23.

[23] A. Fathi and A. Figalli, *Optimal transportation on non-compact manifolds*, Israel J. Math., 175 (2010), pp. 1–59.

[24] M. Fukushima, *Dirichlet forms and Markov processes*, vol. 23 of North-Holland Mathematical Library, North-Holland Publishing Co., Amsterdam, 1980.

[25] J. Heinonen, *Non smooth calculus*, Bull. of the AMS, 44 (2007), pp. 163–232.

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

[27] K. Kuwada, *Duality on gradient estimates and Wasserstein controls*, Journal of Functional Analysis, 258 (2010), pp. 3758–3774.

[28] S. Lisini, *Characterization of absolutely continuous curves in Wasserstein spaces*, Calc. Var. Partial Differential Equations, 28 (2007), pp. 85–120.

[29] G. Loeper, *On the regularity of the polar factorization for time dependent maps*, Calc. Var. and Part. Diff. Eq., 22 (2004), pp. 343–374.

[30] J. Lott, *Some geometric calculations on Wasserstein space*, Comm. Math. Phys., 277 (2008), pp. 423–437.

[31] J. Lott and C. Villani, *Hamilton-Jacobi semigroup on length spaces and applications*, J. Math. Pures Appl. (9), 88 (2007), pp. 219–229.

[32] ——, *Weak curvature bounds and functional inequalities*, J. Funct. Anal., 245 (2007), pp. 311–333.

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

[34] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J., 80 (1995), pp. 309–323.

[35] ——, *Polar factorization of maps on riemannian manifolds*, Geometric and Functional Analysis, 11 (2001), pp. 589–608.

[36] S.-I. Ohta, *Gradient flows on Wasserstein spaces over compact Alexandrov spaces*, Amer. J. Math., 131 (2009), pp. 475–516.

[37] S.-I. Ohta and K.-T. Sturm, *Heat flow on Finsler manifolds*, Comm. Pure Appl. Math., 62 (2009), pp. 1386–1433.

[38] F. Otto, *Evolution of microstructure in unstable porous media flow: a relaxational approach*, Comm. Pure Appl. Math., 52 (1999), pp. 873–915.

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

[40] A. Petruinin, *Alexandrov meets Lott–Villani–Sturm*, preprint, Available at arXiv:1003.5948, (2009).

[41] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris, 345 (2007), pp. 151–154.

[42] N. Shanmugalingam, *Newtonian spaces: an extension of Sobolev spaces to metric measure spaces*, Rev. Mat. Iberoamericana, 16 (2000), pp. 243–279.

[43] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math., 196 (2006), pp. 65–131.

[44] ——, *On the geometry of metric measure spaces. II*, Acta Math., 196 (2006), pp. 133–177.

[45] K.-T. Sturm and S.-I. Ohta, *Non-contraction of heat flow on Minkowski spaces*, (2010).

[46] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[47] ——, *Optimal transport. Old and new*, vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.

[48] L. Zajíček, *On the differentiability of convex functions in finite and infinite dimensional spaces*, Czechoslovak Math. J., 104 (1979), pp. 340–348.

[49] H.-C. Zhang and X.-P. Zhu, *Ricci curvature on Alexandrov spaces and rigidity theorems*, Comm. Anal. Geom., 18 (2010), pp. 503–553.