Lectures in Mathematics
ETH Zürich
Department of Mathematics
Research Institute of Mathematics

Managing Editor:
Michael Struwe

Luigi Ambrosio
Nicola Gigli
Giuseppe Savaré

Gradient Flows

in Metric Spaces and in the Space of Probability Measures

Birkhäuser Verlag
Basel · Boston · Berlin

Authors:

Luigi Ambrosio
Nicola Gigli
Scuola Normale Superiore
Piazza die Cavalieri 7
I-56126 Pisa
ambrosio@sns.it
n.gigli@sns.it

Giuseppe Savaré
Dipartimento di Matematica
Università di Pavia
Via Ferrata, 1
I-27100 Pavia
savare@imati.cnr.it

2000 Mathematical Subject Classification 28A33, 28A50, 35K55, 35K90, 47H05, 47J35, 49J40, 65M15

A CIP catalogue record for this book is available from the
Library of Congress, Washington D.C., USA

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN 3-7643-2428-7 Birkhäuser Verlag, Basel – Boston – Berlin

Part of Springer Science+Business Media
Printed on acid-free paper produced from chlorine-free pulp. TCF
Printed in Germany
ISBN-10: 3-7643-2428-7
ISBN-13: 978-3-7643-2428-5

9 8 7 6 5 4 3 2 1 www.birkhauser.ch

Contents

Introduction

This book is devoted to a theory of gradient flows in spaces which are not necessarily endowed with a natural linear or differentiable structure. It is made of two parts, the first one concerning gradient flows in metric spaces and the second one devoted to gradient flows in the L^2-Wasserstein space of probability measures[1] on a separable Hilbert space X (we consider the L^p-Wasserstein distance, $p \in (1, \infty)$, as well).

The two parts have some connections, due to the fact that the Wasserstein space of probability measures provides an important model to which the "metric" theory applies, but the book is conceived in such a way that the two parts can be read independently, the first one by the reader more interested to Non-Smooth Analysis and Analysis in Metric Spaces, and the second one by the reader more oriented to the applications in Partial Differential Equations, Measure Theory and Probability.

The occasion for writing this book came with the NachDiplom course taught by the first author in the ETH in Zürich in the fall of 2001. The course covered only part of the material presented here, and then with the contribution of the second and third author (in particular on the error estimates of Part I and on the generalized convexity properties of Part II) the project evolved in the form of the present book. As a result, it should be conceived in part as a textbook, since we try to present as much as possible the material in a self-contained way, and in part as a research book, with new results never appeared elsewhere.

Now we pass to a more detailed description of the content of the book, splitting the presentation in two parts; for the bibliographical notes we mostly refer to each single chapter.

Part I

In Chapter 1 we introduce some basic tools from Analysis in Metric Spaces. The

[1] This distance is also commonly attributed in the literature to Kantorovich-Rubinstein. Actually Prof. V.Bogachev kindly pointed out to us that the correct spelling of the name Wasserstein should be "Vasershtein" [124] and that the attribution to Kantorovich and Rubinstein is much more correct. We kept the attribution to Wassertein and the wrong spelling because this terminology is by now standard in many recent papers on the subject (gradient flows) closely related to our present work

first one is the metric derivative: we show, following the simple argument in [7], that for any metric space $(\mathscr{S}, d)$ and any absolutely continuous map $v : (a,b) \subset \mathbb{R} \to \mathscr{S}$ the limit

$$|v'|(t) := \lim_{h\to 0} \frac{d\,(v(t+h), v(t))}{|h|}$$

exists for $\mathscr{L}^1$-a.e. $t \in (a,b)$ and $d(v(s), v(t)) \le \int_s^t |v'|(r)\,dr$ for any interval $(s,t) \subset (a,b)$. This is a kind of metric version of Rademacher's theorem, see also [12] and the references therein for the extension to maps defined on subsets of $\mathbb{R}^d$.
In Section 1.2 we introduce the notion of upper gradient, a weak concept for the modulus of the gradient, following with some minor variants the approach in [81], [41]. We say that a function $g : \mathscr{S} \to [0, +\infty]$ is a *strong upper gradient* for $\phi : \mathscr{S} \to (-\infty, +\infty]$ if for every absolutely continuous curve $v : (a,b) \to \mathscr{S}$ the function $g \circ v$ is Borel and

$$\big|\phi(v(t)) - \phi(v(s))\big| \le \int_s^t g(v(r))|v'|(r)\,dr \quad \forall\, a < s \le t < b. \tag{1}$$

In particular, if $g \circ v|v'| \in L^1(a,b)$ then $\phi \circ v$ is absolutely continuous and

$$|(\phi \circ v)'(t)| \le g(v(t))|v'|(t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b). \tag{2}$$

We also introduce the concept of weak upper gradient, where we require only that (2) holds with the approximate derivative of $\phi \circ v$, whenever $\phi \circ v$ is a function of (essential) bounded variation. Among all possible choices of upper gradients, the local [52] and global slopes of ϕ are canonical and respectively defined by:

$$|\partial\phi|(v) := \limsup_{w\to v} \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)}, \quad \mathfrak{l}_\phi(v) := \sup_{w\ne v} \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)}. \tag{3}$$

In our setting, $\mathfrak{l}_\phi(\cdot)$ provides the natural "one sided" bounds for difference quotients modeled on the analogous one [41] for Lipschitz functionals, where the positive part of $\phi(v) - \phi(w)$ is replaced by the modulus.

We prove in Theorem 1.2.5 that the function $|\partial\phi|$ is a weak upper gradient for ϕ and that, if ϕ is lower semicontinuous, $\mathfrak{l}_\phi$ is a strong upper gradient for ϕ.
In Section 1.3 we introduce our main object of study, the notion of curve of maximal slope in a general metric setting. The presentation here follows the one in [8], on the basis of the ideas introduced in [52] and further developed in [53], [95]. To illustrate the heuristic ideas behind, let us start with the classical setting of a gradient flow

$$u'(t) = -\nabla\phi\,(u(t)) \tag{4}$$

in a Hilbert space. If we take the modulus in both sides we have the equation $|u'|(t) = |\nabla\phi(u(t))|$ which makes sense in a metric setting, interpreting the left hand side as the metric derivative and the right hand side as an upper gradient of ϕ (for instance the local slope $|\partial\phi|$, as in [8]). However, in passing from (4) to a scalar equation we clearly have a loss of information. This information can be retained by looking at the derivative of the energy:

$$\frac{d}{dt}\phi\left(u(t)\right) = \langle u'(t), \nabla\phi\left(u(t)\right)\rangle = -|u'(t)||\nabla\phi\left(u(t)\right)| = -\frac{1}{2}|u'|^2(t) - \frac{1}{2}|\nabla\phi\left(u(t)\right)|^2.$$

The second equality holds iff u' and $-\nabla\phi(u)$ are parallel and the third equality holds iff $|u'|$ and $|\nabla\phi(u)|$ are equal, so that we can rewrite (4) as

$$\frac{1}{2}|u'|^2(t) + \frac{1}{2}|\nabla\phi\left(u(t)\right)|^2 = -\frac{d}{dt}\phi\left(u(t)\right).$$

Passing to an integral formulation and replacing $|\nabla\phi(u)|$ with $g(u)$, where g is an upper gradient of ϕ, we say that u is a curve of maximal slope with respect to g if

$$\frac{1}{2}\int_s^t \left(|u'|^2(r) + |g\left(u(r)\right)|^2\right) dr \le \phi\left(u(s)\right) - \phi\left(u(t)\right) \tag{5}$$

for $\mathscr{L}^1$-a.e. s, t with $s \le t$. In the case when g is a strong upper gradient, the energy is absolutely continuous in time, the inequality above is an equality and it holds for any s, $t \ge 0$ with $s \le t$.
This concept of curve of maximal slope is very natural, as we will see, also in connection with the problem of the convergence of the implicit Euler scheme. Indeed, we will see that (5) has also a discrete counterpart, see (11) and (3.2.4). A brief comparison between the notion of curves of maximal slope and the more usual notion of gradient flows in Banach spaces is addressed in Section 1.4. We shall see that the metric approach is useful even in a linear framework, e.g. when the Banach space does not satisfy the Radon-Nikodým property (so that there exist absolutely continuous curves which are not a.e. differentiable) and therefore gradient flows cannot be characterized by a differential inclusion.

In Chapter 2 we study the problem of the existence of curves of maximal slope starting from a given initial datum $u_0 \in \mathscr{S}$ and the convergence of (a variational formulation of) the implicit Euler scheme. Given a time step $\tau > 0$ and a discrete initial datum $U_\tau^0 \approx u_0$, we use the classical variational problem

$$U_\tau^n \in \operatorname{argmin}\left\{\phi(v) + \frac{1}{2\tau}d^2(v, U_\tau^{n-1}) : v \in \mathscr{S}\right\} \tag{6}$$

to find, given U_τ^{n-1}, the next value U_τ^n. We consider also the case of a variable time step when τ depends on n as well (see Remark 2.0.3). Also, we have preferred to distinguish the role played by the distance d (which, together with ϕ, governs the direction of the flow) by the role played by an auxiliary topology σ on $\mathscr{S}$, that could be weaker than the one induced by d, ensuring compactness of the sublevel sets of the minimizing functional of (6) (this ensures existence of minimizers in (6)). In this introductory presentation we consider for simplicity the case of a uniform step size τ independent of n and of an energy functional ϕ whose sublevel

sets $\{\phi \le c\}$, $c \in \mathbb{R}$, are compact with respect to the distance topology; we also suppose that $U^0_\tau = u_0, \phi(u_0) < +\infty$. This ensures a compactness property of the discrete trajectories and therefore the existence of limit trajectories as $\tau \downarrow 0$ (the so-called generalized minimizing movements in De Giorgi's terminology, see [51]).
In Section 2.3 we state some general existence results for curves of maximal slope. The first result is stated in Theorem 2.3.1 and it is the more basic one: we show that if the relaxed slope

$$|\partial^- \phi|(u) := \inf \Big\{ \liminf_{n\to\infty} |\partial\phi|(u_n) : \ u_n \to u, \ \sup_n \{d(u_n,u), \phi(u_n)\} < +\infty \Big\} \tag{7}$$

is a weak upper gradient for ϕ, and if ϕ is continuous along bounded sequences in $\mathscr{S}$ on which both ϕ and $|\partial\phi|$ are bounded, then any limit trajectory is a curve of maximal slope with respect to $|\partial^-\phi|(u)$. If $|\partial^-\phi|(u)$ is a strong upper gradient we can drop the continuity assumption on ϕ and obtain in Theorem 2.3.3 that any limit trajectory is a curve of maximal slope with respect to $|\partial^-\phi|(u)$. In particular this leads to the energy identity

$$\frac{1}{2}\int_s^t \Big(|u'|^2(r) + |\partial^-\phi|^2(u(r)) \Big)\, dr = \phi\,(u(s)) - \phi\,(u(t)) \tag{8}$$

for any interval $[s,t] \subset [0,+\infty)$. One can also show strong L^2 convergence of several quantities associated to discrete trajectories to their continuous counterpart, see (2.3.6) and (2.3.7).
In Section 2.4 we consider the case of convex functionals. Here convexity or, more generally, λ-convexity has to be understood (see [84], [97]) in the following sense:

$$\phi(\gamma_t) \le (1-t)\phi(\gamma_0) + t\phi(\gamma_1) - \frac{1}{2}\lambda\, t(1-t) d^2(\gamma_0,\gamma_1) \qquad \forall t \in [0,1] \tag{9}$$

for any constant speed minimal geodesic $\gamma_t : [0,1] \to \mathscr{S}$ (but more general class of interpolating curves could also be considered). We show that for λ-convex functionals with $\lambda \ge 0$ the local and global slopes coincide. Moreover, for any λ-convex functional the local slope $|\partial\phi|$ is a a strong upper gradient and it is lower semicontinuous, therefore the results of the previous section apply and we obtain existence of curves of maximal slope with respect to $|\partial\phi|$ and the energy identity (8). Assuming $\lambda > 0$ we prove some estimates which imply exponential convergence of $u(t)$ to the minimum point of the energy as $t \to +\infty$. At this level of generality an open problem is the uniqueness of curves of maximal slope: this problem is open even in the case when $\mathscr{S}$ is a Banach space. We are able to get uniqueness, together with error estimates for the Euler scheme, only under stronger convexity assumptions (see Chapter 4 and also Section 11.1.2 in Part II, where uniqueness is obtained in the Wasserstein space using its differentiable structure). Finally, we prove in Theorem 2.4.15 a metric counterpart of Brezis' result [28, Theorem 3.2, page. 57], showing that the right metric derivative of $t \mapsto u(t)$ and the right derivative of $t \mapsto \phi(u(t))$ exist at any $t > 0$; in addition the equation

$$\frac{d}{dt_+}\phi(u(t)) = -|\partial\phi|^2(u(t)) = -|u'_+|^2(t) = -|\partial\phi|(u(t))\,|u'_+|(t)$$

holds in a pointwise sense in $(0, +\infty)$.

Chapter 3 is devoted to some proofs of the convergence and regularity theorems stated in the previous chapter. We study in particular the Moreau–Yosida approximation ϕ_τ of ϕ (a natural object of study in connection with (6)), defined by

$$\phi_\tau(u) := \inf\left\{\phi(v) + \frac{1}{2\tau}d^2(v,u) :\ v \in \mathscr{S}\right\} \qquad u \in \mathscr{S},\ \tau > 0. \tag{10}$$

Notice that since $v = u$ is admissible in the variational problem defining ϕ_τ, we have the obvious inequality

$$\frac{1}{2\tau}d^2(u, u_\tau) \le \phi(u) - \phi(u_\tau)$$

for any minimizer u_τ (here we assume that for $\tau > 0$ sufficiently small the infimum is attained). Following an interpolation argument due to De Giorgi this elementary inequality can be improved (see Theorem 3.1.4), getting

$$\frac{d^2(u_\tau, u)}{2\tau} + \int_0^\tau \frac{d^2(u_r, u)}{2r^2}\,dr = \phi(u) - \phi(u_\tau). \tag{11}$$

Combining this identity with the slope estimate (see Lemma 3.1.3)

$$|\partial\phi|(u_\tau) \le \frac{d(u_\tau, u)}{\tau},$$

we obtain the sharper inequality

$$\frac{d^2(u_\tau, u)}{2\tau} + \int_0^\tau \frac{|\partial\phi|^2(u_r)}{2}\,dr \le \phi(u) - \phi(u_\tau).$$

If we interpret $r \mapsto u_r$ as a kind of "variational" interpolation between u and u_τ, and if we apply this estimate repeatedly to all pairs $(u, u_\tau) = (U_\tau^{n-1}, U_\tau^n)$ arising in the Euler scheme, we obtain a discrete analogue of (5). This is the argument underlying the basic convergence Theorem 2.3.1. Notice that this variational interpolation does not coincide (being dependent on ϕ), even in a linear framework, with the standard piecewise linear interpolation.

Chapter 4 addresses the general questions related to the well posedness of curves of maximal slope, i.e. uniqueness, continuous dependence on the initial datum, convergence of the approximation scheme and possibly optimal error estimates, asymptotic behavior. All these properties have been deeply studied for l.s.c. convex functionals ϕ in Hilbert spaces, where it is possible to prove that the Euler scheme (6) converges (with an optimal rate depending on the regularity of u_0) for each choice of initial datum in the closure of the domain of ϕ and generates a contraction semigroup which exhibits a regularizing effect and can be characterized by a system of variational inequalities.

We already mentioned the lackness of a corresponding Banach space theory: if one hopes to reproduce the Hilbertian result in a purely metric framework it is natural to think that the so called "parallelogram rule"

$$\left\|\frac{\gamma_0+\gamma_1}{2}\right\|^2+\left\|\frac{\gamma_0-\gamma_1}{2}\right\|^2=\frac{1}{2}\|\gamma_0\|^2+\frac{1}{2}\|\gamma_1\|^2, \tag{12}$$

which provides a metric characterization of Hilbertian norms, should play a crucial role.

It is well known that (12) is strictly related to the uniform modulus of convexity of the norm: in fact, considering a general convex combination $\gamma_t = (1-t)\gamma_0 + t\gamma_1$ instead of the middle point between γ_0 and γ_1, and evaluating the distance $d(\gamma_t, v) := \|\gamma_t - v\|$ from a generic point v instead of 0, we easily see that (12) can be rephrased as

$$d(\gamma_t,v)^2=(1-t)d(\gamma_0,v)^2+td(\gamma_1,v)^2-t(1-t)d(\gamma_0,\gamma_1)^2 \quad \forall\, t\in[0,1]. \tag{13}$$

It was one of the main contribution of U. Mayer [96] to show that in a general geodesically complete metric space the 2-convexity inequality

$$d(\gamma_t,v)^2\le(1-t)d(\gamma_0,v)^2+td(\gamma_1,v)^2-t(1-t)d(\gamma_0,\gamma_1)^2 \quad \forall\, t\in[0,1]. \tag{14}$$

(where now γ_t is a constant speed minimal geodesic connecting γ_0 to γ_1: cf. (9)) is a sufficient condition to prove a well posedness result by mimicking the celebrated Crandall-Ligget generation result for contraction semigroups associated to m-accretive operators in Banach spaces.

For a Riemannian manifold (14) is equivalent to a global nonpositivity condition on the sectional curvature: Aleksandrov introduced condition (14) for general metric spaces, which are now called NPC (Non Positively Curved) spaces.

Unfortunately, the L^2-Wasserstein space, which provides one of the main motivating example of the present theory, satisfies the opposite (generally strict) inequality, which characterizes Positively Curved space.

Our main result consists in the possibility to choose more freely the family of connecting curves, which do not have to be geodesics any more: we simply suppose that for each triple of points γ_0, γ_1, v there exists a curve γ_t connecting γ_0 to γ_1 and satisfying (14) and (9); we shall see in the second Part of this book that this considerably weaker condition is satisfied by various interesting examples in the L^2-Wasserstein space.

Even if the Crandall-Ligget technique cannot be applied under these more general assumptions, we are able to prove a completely analogous generation result for a regularizing contraction semigroup, together with the optimal error estimate (here $\lambda = 0$) at each point t of the discrete mesh

$$d^2(u(t), U_\tau(t)) \le \tau\Big(\phi(u_0) - \phi_\tau(u_0)\Big) \le \frac{\tau^2}{2}|\partial\phi|^2(u_0).$$

Part II

Chapter 5 contains some preliminary and basic facts about Measure Theory and Probability in a general separable metric space X. In the first section we introduce the narrow convergence and discuss its relation with tightness, lower semicontinuity, and p-uniform integrability; a particular attention is devoted in Section 5.1.2 to the case when X is an Hilbert space and the strong or weak topologies are considered. In the second section we introduce the push-forward operator $\mu \mapsto \boldsymbol{r}_\#\mu$ between measures and discuss its main properties. Section 5.3 is devoted to the disintegration theorem for measures and to the related and classical concept of measure-valued map. The relationships between convergence of maps and narrow convergence of the associated plans, typical in the theory of Young measures (see for instance [128, 129, 23, 123, 20]), are presented in Section 5.4.
Finally, the last section of the chapter contains a discussion on the area formula for maps $f : A \subset \mathbb{R}^d \to \mathbb{R}^d$ under minimal regularity assumptions on f (in the same spirit of [77]), so that the classical formula for the change of density

$$f_\# \left(\rho\mathscr{L}^d\right) = \frac{\rho}{|\mathrm{det}\nabla f|} \circ f^{-1}|_{f(A)}\mathscr{L}^d$$

still makes sense. These results apply in particular to the classical case when f is the gradient of a convex function (this fact was proved first by a different argument in [97]). In the same section we introduce the classical concepts of *approximate continuity* and *approximate differentiability* which will play an important role in establishing the existence and the differentiability of optimal transport maps.

Chapter 6 is entirely devoted to the general results on optimal transportation problems between probability measures μ, ν: in the first section they are studied in a Polish/Radon space X with a cost function $c : X^2 \to [0, +\infty]$. We consider the strong formulation of the problem with transport maps due to Monge, see (6.0.1), and its weak formulation with transport plans

$$\min\left\{\int_{X^2} c(x, y)\, d\boldsymbol{\gamma} :\ \boldsymbol{\gamma} \in \Gamma(\mu, \nu)\right\} \tag{15}$$

due to Kantorovich. Here $\Gamma(\mu, \nu)$ denotes the class of all $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ such that $\pi^1_\#\boldsymbol{\gamma} = \mu$ and $\pi^2_\#\boldsymbol{\gamma} = \nu$ ($\pi^i : X^2 \to X$, $i = 1, 2$ are the canonical projections) and in the following we shall denote by $\Gamma_o(\mu, \nu)$ the class of optimal plans for (15).

In Section 6.1 we discuss the duality formula

$$\min\ (15) = \sup\left\{\int_X \varphi\, d\mu + \int_X \psi\, d\nu :\ \varphi(x)+\psi(y) \le c(x,y)\right\}$$

for the Kantorovich problem and the necessary and sufficient optimality conditions for transport plans. These can be expressed in two basically equivalent ways (under suitable a-priori estimates from above on the cost function): a transport plan $\boldsymbol{\gamma}$ is optimal if and only if its support is c-monotone, i.e.

$$\sum_{i=1}^n c(x_i, y_{\sigma(i)}) \ge \sum_{i=1}^n c(x_i, y_i) \quad \text{for any permutation } \sigma \text{ of } \{1,\ldots,n\}$$

for any choice of $(x_i, y_i) \in \operatorname{supp}\boldsymbol{\gamma}$, $1 \le i \le n$. Alternatively, a transport plan $\boldsymbol{\gamma}$ is optimal if and only if there exist (φ, ψ) such that $\varphi(x) + \psi(y) \le c(x,y)$ for any (x,y) and

$$\varphi(x) + \psi(y) = c(x,y) \qquad \boldsymbol{\gamma}\text{-a.e. in } X \times X. \tag{16}$$

The pair (φ, ψ) can be built in a canonical way, independent of the optimal plan $\boldsymbol{\gamma}$, looking for maximizing pairs in the duality formula (6.1.1). In the presentation of these facts we have been following mostly [14], [71], [112], [126]; see also [61].
Section 6.2 is devoted to the problem of the existence of optimal transport maps $\boldsymbol{t}_\mu^\nu$, under the assumption that X is an Hilbert space and the initial measure μ is absolutely continuous (in the infinite dimensional case we assume that the measure μ vanishes on all Gaussian null sets); we consider mostly the case when the cost function is the p-power, with $p > 1$, of the distance. We include also (see Theorem 6.2.10) an existence result in the case when X is a separable Hilbert space (compare with the results [68, 69, 89] in Wiener spaces, where the cost function $c(x,y)$ is finite only when $x-y$ is in the Cameron-Martin space). The proofs follow the by now standard approach of differentiating with respect to x the relation (16) to obtain that for μ-a.e. x there is a unique y such that (16) holds (the relation $x \mapsto y$ then gives the desired optimal transport map $y = \boldsymbol{t}_\mu^\nu(x)$).

The Wasserstein distances and their geometric properties are the main subjects of Chapter 7. In Section 7.1 we define the p-Wasserstein distance and we recall its basic properties, emphasizing the fact that the space $\mathscr{P}_p(X)$ endowed with this distance is complete and separable but not locally compact when the underlying space X is not compact.
The second section of Chapter 7 deals with the characterization of constant speed geodesics in $\mathscr{P}_p(X)$ (here X is an Hilbert space), parametrized on the unit interval $[0,1]$. Given the endpoints μ_0, μ_1 of the geodesic, we show that there exists an optimal plan $\boldsymbol{\gamma}$ between μ_0 and μ_1 such that

$$\mu_t = \left(t\pi^2 + (1-t)\pi^1\right)_\# \boldsymbol{\gamma} \qquad \forall t \in [0,1]. \tag{17}$$

Conversely, given any optimal plan $\boldsymbol{\gamma}$, the formula above defines a constant speed geodesic. In the case when plans are induced by transport maps, (17) reduces to

$$\mu_t = \left(t\,\boldsymbol{t}_{\mu_0}^{\mu_1} + (1-t)\boldsymbol{i}\right)_\# \mu_0 \qquad \forall t \in [0,1]. \tag{18}$$

We show also in Lemma 7.2.1 that there is a unique transport plan joining a point in the interior of a geodesic to one of the endpoints; in addition this transport plan is induced by a transport map (this does not require any absolute continuity assumption on the endpoints and will provide a useful technical tool to approximate plans with transports).
In Section 7.3 we focus our attention on the L^2-Wasserstein distance: we will prove a semi-concavity inequality for the squared distance function $\psi(t) := \frac{1}{2}W_2^2(\mu_t, \mu)$ from a fixed measure μ along a constant speed minimal geodesic μ_t, $t \in [0, 1]$

$$W_2^2(\mu_t, \mu) \geq tW_2^2(\mu_1, \mu) + (1-t)W_2^2(\mu_0, \mu) - t(1-t)W_2^2(\mu_0, \mu_1) \tag{19}$$

and we discuss its geometric counterpart; we also provide a precise formula to evaluate the time derivative of ψ and we show trough an explicit counterexample that ψ does not satisfy any λ-convexity property, for any $\lambda \in \mathbb{R}$. Conversely, (19) shows that ψ is semi-concave and that $\mathscr{P}_2(X)$ is a Positively curved (PC) metric space.

Chapter 8 plays an important role in the theory developed in this book. In the first section we review some classical results about the continuity/transport equation

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \qquad \text{in } X \times (a, b) \tag{20}$$

in a finite dimensional euclidean space X and the representation formula for its solution by the Characteristics method, when the velocity vector field v_t satisfies a p-summability property with respect to the measures μ_t and a local Lipschitz condition. When this last space-regularity properties does not hold, one can still recover a probabilistic representation result, through Young measures in the space of X-valued time dependent curves: this approach is presented in Section 8.2.
The main result of this chapter, presented in Section 8.3, is that the class of solutions of the transport equation (20) (in the infinite dimensional case the equation can still be interpreted in a weak sense using cylindrical test functions) coincides with the class of absolutely continuous curves μ_t with values in the Wasserstein space. Specifically, given an absolutely continuous curve μ_t one can always find a "velocity field" $v_t \in L^p(\mu_t; X)$ such that (20) holds; in addition, by construction we get that the norm of the velocity field can be estimated by the metric derivative:

$$\|v_t\|_{L^p(\mu_t)} \leq |\mu'|(t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a, b). \tag{21}$$

Conversely, any solution (μ_t, v_t) of (20) with $\int_a^b \|v_t\|_{L^p(\mu_t)}\, dt < +\infty$ induces an absolutely continuous curve μ_t, whose metric derivative can be estimated by $\|v_t\|_{L^p(\mu_t)}$ for $\mathscr{L}^1$-a.e. $t \in (a, b)$. As a consequence of (8.2.1) we see that among all velocity fields v_t which produce the same flow μ_t, there is an optimal one with smallest L^p norm, equal to the metric derivative of μ_t; we view this optimal field as the "tangent" vector field to the curve μ_t. To make this statement more precise, let us consider for instance the case when $p = 2$ and X is finite dimensional: in this case the tangent vector field is characterized, among all possible velocity fields, by the property

$$v_t \in \overline{\{\nabla\varphi : \ \varphi \in C_c^\infty(X)\}}^{L^2(\mu_t;X)} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b). \tag{22}$$

In general one has to consider a duality map j_q between L^q and L^p (since gradients are thought as covectors, and therefore as elements of L^q) and gradients of cylindrical test functions if X is infinite dimensional.
In the next Section 8.4 we investigate the properties of the above defined tangent vector. A first consequence of the characterization of absolutely continuous curves is a result, given in Proposition 8.4.6, concerning the infinitesimal behaviour of the Wasserstein distance along absolutely continuous curves μ_t: given the tangent vector field v_t to the curve, we show that

$$\lim_{h\to 0} \frac{W_p(\mu_{t+h}, (\boldsymbol{i} + hv_t)_\# \mu_t)}{|h|} = 0 \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b). \tag{23}$$

Moreover the optimal transport plans between μ_t and μ_{t+h}, rescaled in a suitable way, converge to the optimal transport plan $(\boldsymbol{i} \times v_t)_\# \mu_t$ associated to v_t (see (8.4.6)). This Proposition shows that the infinitesimal behaviour of the Wasserstein distance is governed by transport maps even in the situations when globally optimal transport maps fail to exist (recall that the existence of optimal transport maps requires assumptions on the initial measure μ).
Another interesting result is a formula for the derivative of the distance from a fixed measure along any absolutely continuous curve μ_t in $\mathscr{P}_p(X)$: one can show for any $p \in (1,\infty)$ that

$$\frac{d}{dt} W_p^p(\mu_t, \bar\mu) = p \int_{X^2} \langle v_t(x_1), x_1 - x_2\rangle |x_1 - x_2|^{p-2} \, d\boldsymbol{\gamma}_t(x_1,x_2) \tag{24}$$

for any optimal plan $\boldsymbol{\gamma}_t$ between μ_t and $\bar\mu$; here v_t is any admissible velocity vector field associated to μ_t through the continuity equation (20). This "generic" differentiability along absolutely continuous curves is sufficient for our purposes, see for instance Theorem 11.1.4 where uniqueness of gradient flows is proved.
Another consequence of the characterization of absolutely continuous curves in $\mathscr{P}_2(X)$ is the variational representation formula

$$W_2^2(\mu_0, \mu_1) = \min\left\{ \int_0^1 \|v_t\|^2_{L^2(\mu_t)} \, dt : \ \frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0 \right\}. \tag{25}$$

Again, these formulas still hold with the necessary adaptations if either $p \in (1, +\infty)$ (in this case we have a kind of Finsler metric) or X is infinite dimensional. We also show that optimal transport maps belong to $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ under quite general conditions.
The characterization (22) of velocity vectors and the additional properties we

listed above, strongly suggest to consider the following "regular" tangent bundle to $\mathscr{P}_2(X)$

$$\mathrm{Tan}_\mu \mathscr{P}_2(X) := \overline{\{\nabla\varphi :\ \varphi \in C_c^\infty(X)\}}^{L^2(\mu;X)} \qquad \forall \mu \in \mathscr{P}_2(X), \tag{26}$$

endowed with the natural L^2 metric. Up to a $\mathscr{L}^1$-negligible set in (a, b), it contains and characterizes all the tangent velocity vectors to absolutely continuous curves. In this way we recover in a general framework the Riemannian interpretation of the Wasserstein distance developed by Otto in [107] (see also [106], [83] and also [38]): indeed, the right hand side in (25) is nothing but the minimal length, computed with respect to the metric tensor, of all absolutely continuous curves connecting μ_0 to μ_1. This formula was independently discovered also in [21], and used for numerical purposes. In the original paper [107], instead, (25) is derived using formally the concept of Riemannian submersion and the family of maps $\phi \mapsto \phi_\# \mu$ (indexed by μ) from Arnold's space of diffeomorphisms into the Wasserstein space. In the last Section 8.5 we compare the "regular" tangent space 26 with the tangent cone obtained by taking the closure in $L^p(\mu; X)$ of all the optimal transport maps and we will prove the remarkable result that these two notions coincide.

In Chapter 9 we study the convexity properties of functionals $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$. Here "convexity" refers to convexity along geodesics (as in [97], [107], where these properties have been first studied), whose characterization has been given in the previous Section 7.2. More generally, as in the metric part of the book, we consider λ-convex functionals as well, and in Section 9.2 we investigate some more general convexity properties in $\mathscr{P}_2(X)$. The motivation comes from the fact, discussed in Part I, that error estimates for the implicit Euler approximation of gradient flows seem to require joint convexity properties of the functional and of the squared distance function. As shown by a formal computation in [107], the function $W_2^2(\cdot, \mu)$ is not 1-convex along classical geodesics μ_t and we have actually the reverse inequality (19) (cf. Corollary 7.3.2). It is then natural to look for different kind of interpolating curves, along which the distance behaves nicely, and for functionals which are convex along this new class of curves.
To this aim, given an absolutely continuous measure μ, we consider the family of "generalized geodesics"

$$\mu_t := \big((1-t)\boldsymbol{t}_\mu^{\mu_0} + t\,\boldsymbol{t}_\mu^{\mu_1}\big)_\#\,\mu \qquad t \in [0, 1],$$

among all possible optimal transport maps $\boldsymbol{t}_\mu^{\mu_0}$, $\boldsymbol{t}_\mu^{\mu_1}$. As usual we get rid of the absolute continuity assumption on μ by considering the family of 3-plans

$$\big\{\boldsymbol{\gamma} \in \mathscr{P}(X^3) :\ (\pi^1, \pi^2)_\#\boldsymbol{\gamma} \in \Gamma_o(\mu, \mu_0),\ (\pi^1, \pi^3)_\#\boldsymbol{\gamma} \in \Gamma_o(\mu, \mu_1)\big\},$$

and the corresponding family of generalized geodesics:

$$\mu_t := \big((1-t)\pi^2 + t\pi^3\big)_\#\,\boldsymbol{\gamma} \qquad t \in [0, 1].$$

We prove in Lemma 9.2.1 the key fact that $W_2^2(\cdot,\mu)$ is 1-convex along these generalized geodesics. Thanks to the theory developed in Part I, the convexity of $W_2^2(\cdot,\mu)$ along the generalized geodesics leads to error estimates for the Euler scheme, provided the energy functional ϕ is λ-convex, for some $\lambda\in\mathbb{R}$, along any curve in this family. It turns out that almost all the known examples of convex functionals along geodesics, which we study in some detail in Section 9.3, satisfy this stronger convexity property; following a terminology introduced by C. Villani, we will consider functionals which are the sum of three different kinds of contribution: the *potential* and the *interaction energy*, induced by convex functions $V, W : X\to(-\infty,+\infty]$

$$\mathcal{V}(\mu)=\int_X V(x)\,d\mu(x),\qquad \mathcal{W}(\mu)=\int_{X^2} W(x-y)\,d\mu\times\mu(x),$$

and finally the *internal energy*

$$\mathcal{F}(\mu):=\int_{\mathbb{R}^d} F\Big(\frac{d\mu}{d\mathscr{L}^d}(x)\Big)\,d\mathscr{L}^d(x), \tag{27}$$

$F:[0,+\infty)\to\mathbb{R}$ being the energy density, which should satisfy an even stronger condition than convexity.
The last Section 9.4 discusses the link between the geodesic convexity of the Relative Entropy functional (without any restriction on the dimension of the space; we also consider a more general class of relative integral functionals, obtained replacing $\mathscr{L}^d$ in (27) by a general probability measure γ in X)

$$\mathcal{H}(\mu|\gamma):=\begin{cases}\displaystyle\int_X \frac{d\mu}{d\gamma}\log\left(\frac{d\mu}{d\gamma}\right)d\gamma & \text{if } \mu\ll\gamma,\\ +\infty & \text{otherwise,}\end{cases} \tag{28}$$

and the "log" concavity of the reference measure γ, a concept which is strictly related to various powerful functional analytic inequalities. The main result here states that $\mathcal{H}(\cdot|\gamma)$ is convex along geodesics in $\mathscr{P}_p(X)$ (here the exponent p can be freely chosen, and also generalized geodesics in $\mathscr{P}_2(X)$ can be considered) if and only if γ is "log" concave, i.e. for every couple of open sets A, $B\subset X$ we have

$$\log\gamma((1-t)A+tB)\ge(1-t)\log\gamma(A)+t\log\gamma(B)\qquad t\in[0,1].$$

When $X=\mathbb{R}^d$ and $\gamma\ll\mathscr{L}^d$, this condition is equivalent to the representation $\gamma=e^{-V}\cdot\mathscr{L}^d$ for some l.s.c. convex potential $V:\mathbb{R}^d\to(-\infty,+\infty]$ whose domain has not empty interior in $\mathbb{R}^d$.

One of the goal of the last two chapters is to establish a theory sufficiently powerful to reproduce in the Wasserstein framework the nice results valid for convex functionals and their gradient flows in Hilbert spaces. In this respect an essential ingredient is the concept of (Fréchet) subdifferential of a l.s.c. functional $\phi:\mathscr{P}_p(X)\to(-\infty,+\infty]$ (see also [37, 38]), which is introduced and systematically

studied in Chapter 10.
In order to motivate the relevant definitions and to suggest a possible guideline for the development of the theory, we start by recalling five main properties satisfied by the Fréchet subdifferential in Hilbert spaces. In Section 10.1 we prove that a natural transposition of the same definitions in the Wasserstein space $\mathscr{P}_2(X)$, when only regular measures belong to the proper domain of ϕ (or even of its metric slope $|\partial\phi|$), is possible and they enjoy completely analogous properties as in the flat case. Since this exposition is easier to follow than the one of Section 10.3 for arbitrary measures, here we briefly sketch the main points.
First of all, the subdifferential $\partial\phi(\mu)$ contains all the vectors $\boldsymbol{\xi} \in L^2(\mu; X)$ such that

$$\phi(\nu) - \phi(\mu) \geq \int_X \langle \boldsymbol{\xi}, \boldsymbol{t}_\mu^\nu - \boldsymbol{i} \rangle \, d\mu + o\left(W_2(\nu, \mu)\right). \tag{29}$$

If μ is a minimizer of ϕ, then $0 \in \partial\phi(\mu)$; more generally, if $\mu_\tau \in \mathscr{P}_2(X)$ minimizes

$$\nu \mapsto \frac{1}{2\tau} W_2^2(\nu, \mu) + \phi(\nu),$$

then the corresponding "Euler" equation reads

$$\frac{\boldsymbol{t}_{\mu_\tau}^\mu - \boldsymbol{i}}{\tau} \in \partial\phi(\mu_\tau).$$

As in the linear case, when ϕ is convex along geodesics, the subdifferential (29) can also be characterized by the global system of variational inequalities

$$\phi(\nu) - \phi(\mu) \geq \int_X \langle \boldsymbol{\xi}, \boldsymbol{t}_\mu^\nu - \boldsymbol{i} \rangle \, d\mu \quad \forall \, \nu \in \mathscr{P}_2(X), \tag{30}$$

and it is "monotone", since

$$\xi_i \in \partial\phi(\mu_i), \quad i = 1, 2 \quad \Longrightarrow \quad \int_X \langle \boldsymbol{\xi}_2(\boldsymbol{t}_{\mu_1}^{\mu_2}(x)) - \boldsymbol{\xi}_1(x), \boldsymbol{t}_{\mu_1}^{\mu_2}(x) - x \rangle \, d\mu_1(x) \geq 0;$$

the fact that $\boldsymbol{\xi}_2$ is evaluated on $\boldsymbol{t}_{\mu_1}^{\mu_2}$ in the above formula should not be surprising, since subdifferentials of ϕ in different measures μ_1, μ_2 belong to different vector ($L^2(\mu_i; X)$) spaces (like in Riemannian geometry), so that they can be added or subtracted only after a composition with a suitable transport map.
Closure properties like

$$\mu_h \to \mu \quad \text{in } \mathscr{P}_2(X), \quad \boldsymbol{\xi}_h \rightharpoonup \boldsymbol{\xi}, \quad \boldsymbol{\xi}_h \in \partial\phi(\mu_h) \quad \Longrightarrow \quad \boldsymbol{\xi} \in \partial\phi(\mu), \tag{31}$$

(here one should intend the weak convergence of the vector fields $\boldsymbol{\xi}_h$, which are defined in the varying spaces $L^2(\mu_h; X)$, according to the notion we introduced in Section 5.4) play a crucial role: they hold for convex functionals and define the class of "regular" functionals. In this class the minimal norm of the subdifferential coincides with the metric slope of the functional

$$|\partial\phi|(\mu) = \min\Big\{ \|\boldsymbol{\xi}\|_{L^2(\mu;X)} : \boldsymbol{\xi} \in \partial\phi(\mu) \Big\},$$

and we can prove the chain rule

$$\frac{d}{dt}\phi(\mu_t) = \int_X \langle \boldsymbol{\xi}, \boldsymbol{v}_t\rangle \, d\mu_t \quad \forall \boldsymbol{\xi} \in \partial\phi(\mu_t),$$

for $\mathscr{L}^1$-a.e. (approximate) differentiability point of $t \mapsto \phi(\mu_t)$ along an absolutely continuous curve μ_t, whose metric velocity is $\boldsymbol{v}_t$.
Section 10.2 is entirely devoted to study the (sub- and super-) differentiability properties of the p-Wasserstein distances: here the assumption that the measures are absolutely continuous w.r.t. the Lebesgue one is too restrictive, and our efforts are mainly devoted to circumvent the difficulty that optimal transport maps do not exist in general. Thus we should deal with plans instead of maps and the results we obtain provide the right way to introduce the concept of subdifferential in full generality, i.e. without restriction to absolutely continuous measures, in the next Section 10.3.
To this aim, we need first to define, for given $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ and $\mu := \pi^1_\#\boldsymbol{\gamma}$, the class of 3-plans

$$\Gamma_o(\boldsymbol{\gamma}, \nu) := \left\{\boldsymbol{\gamma} \in \mathscr{P}(X^3) : \ (\pi^1, \pi^2)_\#\boldsymbol{\mu} = \boldsymbol{\gamma}, \ (\pi^1, \pi^3)_\#\boldsymbol{\mu} \in \Gamma_o(\mu, \nu)\right\}.$$

Notice that in the particular case when $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{\xi})_\#\mu$ is induced by a transport map and μ is absolutely continuous, then $\Gamma_o(\boldsymbol{\gamma}, \nu)$ contains only one element

$$\Gamma_o(\boldsymbol{\gamma}, \nu) = \left\{\left(\boldsymbol{i} \times \boldsymbol{\xi} \times \boldsymbol{t}_\mu^\nu\right)_\#\mu\right\} \tag{32}$$

Thus we say that $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ is a general plan subdifferential in $\boldsymbol{\partial}\phi(\mu)$ if its first marginal is μ, its second marginal has finite q-moment, and the asymptotic inequality (29) can be rephrased as

$$\phi(\nu) - \phi(\mu) - \int_{X^3} \langle x_2, x_3 - x_1\rangle \, d\boldsymbol{\mu}(x_1, x_2, x_3) \geq o\big(W_2(\mu, \nu)\big), \tag{33}$$

for some 3-plan $\boldsymbol{\mu}$ (depending on ν) in $\Gamma_o(\boldsymbol{\gamma}, \nu)$.
When ϕ is convex (a similar characterization also holds for λ-convexity) along geodesics, this asymptotic property can be reformulated by means of a system of variational inequalities, analogous to (30): $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu)$ if and only if

$$\forall \nu \in \mathscr{P}_p(X) \quad \exists \boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma}, \nu) : \qquad \phi(\nu) \geq \phi(\mu) + \int_{X^3} \langle x_2, x_3 - x_1\rangle \, d\boldsymbol{\mu}. \tag{34}$$

If condition (32) holds then conditions (33) and (34) reduce of course to (29) and (30) respectively.
This general concept of subdifferential, whose elements are transport plans rather than tangent vectors (or maps) is useful to establish the typical identities of Convex Analysis: we extend to this more general situation all the main properties we discussed in the linear case and we also show that in the λ-convex case tools of

Γ-convergence theory fit quite well in our approach, by providing flexible closure and approximation results for subdifferentials.
In particular, we prove in Theorem 10.3.10 that, as in the classical Hilbert setting, the minimal norm of the subdifferential (in the present case, the q-moment of its second marginal) coincides with the descending slope:

$$\min\left\{\int_{X^2}|x_2|^q\,d\boldsymbol{\gamma}:\ \boldsymbol{\gamma}\in\boldsymbol{\partial}\phi(\mu)\right\}=|\partial\phi|^q(\mu),\tag{35}$$

and the above minimum is assumed by a unique plan $\boldsymbol{\partial}^\circ\phi(\mu)$, which provides the so called "minimal selection" in $\boldsymbol{\partial}\phi(\mu)$ and enjoys many distinguished properties among all the subdifferentials in $\boldsymbol{\partial}\phi(\mu)$. Notice that this result is more difficult than the analogous property in linear spaces, since the q-moment of (the second marginal of) a plan is linear map, and therefore it is not strictly convex. Besides its intrinsic interest, this result provides a "bridge" between De Giorgi's metric concept of gradient flow, based on the descending slope, and the concepts of gradient flow which use the differentiable structure (we come to this point later on).
The last Section 10.4 collects many examples of subdifferentials for the various functionals considered in Chapter 9; among the others, here we recall Example 10.4.6, where the geometric investigations of Chapter 7 yield the precise expression for the subdifferential of the opposite 2-Wasserstein distance, Example 10.4.8, where we show that even in infinite dimensional Hilbert spaces the Relative Fisher Information coincides with the squared slope of the Relative Entropy $\mathcal{H}(\cdot|\gamma)$, when γ is log-concave, and 10.4.7 where the subdifferential of a general functional resulting from the sum of the potential, interaction, and internal energies

$$\phi(\mu)=\int_{\mathbb{R}^d}V(x)\,d\mu(x)+\int_{\mathbb{R}^{2d}}W(x-y)\,d\mu\times\mu(x,y)+\int_{\mathbb{R}^d}F(d\mu/d\mathscr{L}^d)\,dx,$$

is characterized: under quite general assumptions on V,W,F (which allow for potentials with arbitrary growth and also assuming the value $+\infty$) we will show that the minimal selection $\boldsymbol{\partial}^\circ\phi(\mu)$ is in fact induced by the transport map $\boldsymbol{w}=\partial^\circ\phi(\mu)\in L^q(\mu;\mathbb{R}^d)$ defined by

$$\rho\boldsymbol{w}=\nabla L_F(\rho)+\rho\nabla v+\rho(\nabla W*\rho),\quad \mu=\rho\cdot\mathscr{L}^d,\quad L_F(\rho)=\rho F'(\rho)-F(\rho).$$

In the last Chapter 11 we define gradient flows in $\mathscr{P}_p(X)$, X being a separable Hilbert space, and we combine the main points presented in this book to study these flows under many different points of view.
For the sake of simplicity, in this introduction we consider only the more relevant case $p=2$: a locally absolutely continuous curve $\mu_t:(0,+\infty)\to\mathscr{P}_2(X)$, with $|\mu'|\in L^2_{loc}(0,+\infty)$ is said to be a gradient flow relative to the functional $\phi:\mathscr{P}_2(X)\to(-\infty,+\infty]$ if its velocity vector v_t satisfies

$$-v_t\in\partial\phi(\mu_t),\qquad \text{for } \mathscr{L}^1\text{-a.e. } t\in(0,+\infty).\tag{36}$$

For functionals ϕ satisfying the regularity property (31), in Theorem 11.1.3 we show that this "differential" concept of gradient flow is equivalent to the "metric" concept of curve of maximal slope introduced in Part I, see in particular Section 1.3 in Chapter 1. The equivalence passes through the pointwise identity (35).
When the functional is λ-convex along geodesics, in Theorem 11.1.4 we show that gradient flows are uniquely determined by their initial condition

$$\lim_{t\downarrow 0} \mu_t = \mu_0.$$

The proof of this fact depends on the differentiability properties of the squared Wasserstein distance studied in Section 8.3. When the measures μ_t are absolutely continuous and the functional is λ-convex along geodesics, this condition reduces to the system

$$\begin{cases} \dot\mu_t + \nabla\cdot(v_t\mu_t) = 0 \quad \text{in } X\times(0,+\infty), \\ \phi(\nu) \geq \phi(\mu_t) - \displaystyle\int_X \langle v_t, \boldsymbol{t}_{\mu_t}^{\nu} - \boldsymbol{i}\rangle\, d\mu_t + \lambda W_2^2(\nu,\mu_t) \\ \qquad\qquad \forall \nu\in\mathscr{P}_2(X), \quad \text{for } \mathscr{L}^1\text{-a.e. } t>0. \end{cases} \tag{37}$$

Section 11.1.3 is devoted to a general convergence result (up to extraction of a suitable subsequence) of the Minimizing Movement scheme, following a direct approach, which is intrinsically limited to the case when $p = 2$ and the measures μ_t are absolutely continuous. Apart from these restrictions, the functional ϕ could be quite general, so that only a relaxed version of (36) can be obtained in the limit.
Existence of gradient flows is obtained in Theorem 11.2.1 for initial data $\mu_0 \in \overline{D(\phi)}$ and l.s.c. functionals which are λ-convex along generalized geodesics in $\mathscr{P}_2(X)$: this strong result is one of the main applications of the abstract theory developed in Chapter 4 to the Wasserstein framework and, besides optimal error estimates for the convergence of the Minimizing Movement scheme, it provides many additional informations on the regularity the semigroup properties, the asymptotic behaviour as $t \to +\infty$, the pointwise differential properties, the approximations, and the stability w.r.t. perturbations of the functional of the gradient flows. Applications are then given in Section 11.2.1 to various evolutionary PDE's in finite and infinite dimensions, modeled on the examples discussed in Section 10.4.
In Section 11.3 we consider the wider class of regular functionals in $\mathscr{P}_p(X)$ even for $p \neq 2$ and we prove existence of gradient flows when μ_0 belongs to the domain of ϕ and suitable local compactness properties of the sublevel of ϕ are satisfied. This approach uses basically the compactness/energy arguments of the theory developed in Chapter 2 and the equivalence between gradient flows and curves of maximal slope.

The Appendix collects some auxiliary results: the first two sections are devoted to lower semicontinuity and convergence results for integral functionals on product spaces, when the integrand satisfies only a normal or Carathéodory condition, and one of the marginals of the involved sequence of measures is fixed.

In the last two sections we follow the main ideas of the theory of Positively curved (PC) metric space and we are able to identify the geometric tangent cone $\mathbf{Tan}_\mu \mathscr{P}_2(X)$ to $\mathscr{P}_2(X)$ at a measure μ. In a general metric space this tangent space is obtained by taking the completion in a suitable distance of the abstract set of all the curve which are minimal constant speed geodesics at least in a small neighborhood of their starting point μ.
In our case, by identifying these geodesics with suitable transport plans, we can give an explicit characterization of the tangent space and we will see that, if $\mu \in \mathscr{P}_2^r(X)$, it coincides with the closure in $L^2(\mu; X)$ of the gradients of smooth functions and with the closed cone generated by all optimal transport maps, thus with the tangent space (10.4.1) we introduced in Section 8.4.

Acknowledgements. During the development of this project, that took almost three years, we had many useful conversations with colleagues and friends on the topics treated in this book. In particular we wish to thank Y.Brenier, J.A. Carrillo, L.C. Evans, W. Gangbo, N. Ghoussub, R. Mc Cann, F. Otto, G. Toscani and C. Villani. We also warmly thank the PhD student Stefano Lisini for his careful reading of a large part of this manuscript.

Notation

$\lvert v'\rvert(t)$	Metric derivative of $v:(a,b)\to\mathscr{S}$, see Theorem 1.1.2
$AC^p(a,b;\mathscr{S})$	Absolutely continuous $v:(a,b)\to\mathscr{S}$ with $\lvert v'\rvert\in L^p(a,b)$
$B_r(x)$	Open ball of radius r centered at x in a metric space
$D(\phi)$	Domain of the functional ϕ, see (1.2.1)
$\lvert\partial\phi\rvert(v),\ \mathfrak{l}_\phi(v)$	Local and global slopes of ϕ, see Definition 1.2.4
$\mathrm{Lip}(\phi,A)$	Lipschitz constant of the function ϕ in the set A
$\partial\phi(v)$	Fréchet subdifferential of ϕ in Banach (1.4.7), Hilbert (10.0.1), or Wasserstein spaces, see Definition 10.1.1 and (10.3.12)
$\partial^\circ\phi(\mu)$	Minimal selection map in the subdifferential, see Section 1.4 and (10.1.14)
$\lvert\partial^-\phi\rvert(v)$	Relaxed slope of ϕ, see (2.3.1)
$\Phi(\tau,u;v)$	Quadratic perturbation of ϕ by $d^2(u,\cdot)/2\tau$, see (2.0.3b)
$J_\tau[u]$	Resolvent operator, see (2.0.5)
$\overline{U}_\tau(t)$	Piecewise constant interpolation of U^n_τ, see (2.0.7)
$MM(\Phi;u_0)$	Minimizing movement of ϕ, see Definition 2.0.6
$GMM(\Phi;u_0)$	Generalized minimizing movement of ϕ, see Definition 2.0.6
$\phi_\tau(u)$	Moreau–Yosida approximation of ϕ, see Definition 3.1.1
$\tilde{U}_\tau(t)$	De Giorgi's interpolation of U^n_τ, see (3.2.1)
$\mathscr{B}(X)$	Borel sets in a separable metric space X
$C^0_b(X)$	Space of continuous and bounded real functions defined on X
$C^\infty_c(\mathbb{R}^d)$	Space of smooth real functions with compact support in $\mathbb{R}^d$
$\mathscr{P}(X)$	Probability measures in a separable metric space X
$\mathscr{P}_p(X)$	Probability measures with finite p-th moment, see (5.1.22)
$\mathscr{P}_{pq}(X\times X)$	Probability measures with finite p,q-th moments, see (10.3.2)
$L^p(\mu;X)$	L^p space of μ-measurable X-valued maps, see (5.4.3)
X_ϖ	The Hilbert space X endowed with a weaker (normed) topology, see Section 5.1.2
$\tilde f,\ \tilde\nabla f$	Approximate limit and differential of a function f, see Definition 5.5.1
$\mathrm{supp}\,\mu$	Support of μ, see (5.0.1)
$\mathrm{span}\,C$	Linear envelope generated by a subset C of a vector space
$\boldsymbol{r}_\#\mu$	Push-forward of μ through $\boldsymbol{r}$, see (5.2.1)
π^i, $\pi^{i,j}$	Projection operators on a product space $\boldsymbol{X}$, see (5.2.9)
$\Gamma(\mu^1,\mu^2)$	2-plans with given marginals μ^1, μ^2
$\Gamma_o(\mu^1,\mu^2)$	Optimal 2-plans with given marginals μ^1, μ^2
$\boldsymbol{i}$	Identity map
$\boldsymbol{t}^\nu_\mu$	Optimal transport map between μ and ν, see (7.1.4)
$W_p(\mu,\nu)$	p-th Wasserstein distance between μ and ν
$W_{\boldsymbol{\mu}}(\mu,\nu)$	Pseudo-Wasserstein distance induced by $\boldsymbol{\mu}$, see (7.3.2)
$W_{p,\boldsymbol{\mu}}(\mu,\nu)$	Pseudo pth-Wasserstein distance induced $\boldsymbol{\mu}$, see (10.2.9)
$\pi^{i\to j}_t,\ \pi^{i\to j,k}_t$	Interpolated projections, see (7.2.2)
j_p	Duality map between L^p and $L^{p'}$, see (8.3.1)

$\Pi_d(X)$	d-dimensional projections on a Hilbert space X, see Definition 5.1.11
$\mathrm{Cyl}(X)$	Cylindrical test functions on a Hilbert space X, see Definition 5.1.11
$\bar{\boldsymbol{\gamma}}(x)$	Barycentric projection of a plan $\boldsymbol{\gamma}$ in $\mathscr{P}(X \times X)$, see (5.4.9)
$\mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$	Tangent bundle to $\mathscr{P}_p(X)$, see Definition 8.4.1
$\Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)$	3-plans $\boldsymbol{\gamma}$ such that $\pi^{1,3}_{\#}\boldsymbol{\gamma} \in \Gamma_o(\pi^1_{\#}\boldsymbol{\mu}^{1\,2}, \mu^3)$
$\boldsymbol{\partial}\phi(\mu)$	Extended Fréchet subdifferential of ϕ at μ, see Definitions 10.3.1
$\boldsymbol{\partial}^\circ\phi(\mu)$	Minimal selection plan in the subdifferential, see Theorem 10.3.11

Part I

Gradient Flow in Metric Spaces

Chapter 1

Curves and Gradients in Metric Spaces

As we briefly discussed in the introduction, the notion of gradient flows in a metric space $\mathscr{S}$ relies on two elementary but basic concepts: the metric derivative of an absolutely continuous curve with values in $\mathscr{S}$ and the upper gradients of a functional defined in $\mathscr{S}$. The related definitions are presented in the next two sections (a more detailed treatment of this topic can be found for instance in [15]); the last one deals with curves of maximal slope.
When $\mathscr{S}$ is a Banach space and its distance is induced by the norm, one can expect that curves of maximal slope could also be characterized as solutions of (doubly, if $\mathscr{S}$ is not Hilbertian) nonlinear (sub)differential inclusions: this aspect is discussed in the last part of this chapter.
Throughout this chapter (and in the following ones of this first part)

$$(\mathscr{S}, d) \quad \text{will be a given } \textit{complete metric space}; \tag{1.0.1}$$

we will denote by (a, b) a generic open (possibly unbounded) interval of $\mathbb{R}$.

1.1 Absolutely continuous curves and metric derivative

Definition 1.1.1 (Absolutely continuous curves). *Let $(\mathscr{S}, d)$ be a complete metric space and let $v : (a, b) \to \mathscr{S}$ be a curve; we say that v belongs to $AC^p(a, b; \mathscr{S})$, for $p \in [1, +\infty]$, if there exists $m \in L^p(a, b)$ such that*

$$d(v(s), v(t)) \leq \int_s^t m(r)\, dr \qquad \forall\, a < s \leq t < b. \tag{1.1.1}$$

In the case $p = 1$ we are dealing with absolutely continuous *curves and we will denote the corresponding space simply with $AC(a, b; \mathscr{S})$.*

We recall also that a map $\varphi : (a, b) \to \mathbb{R}$ is said to have *finite pointwise variation* if

$$\sup\left\{\sum_{i=1}^{n-1} |\varphi(t_{i+1}) - \varphi(t_i)| : \ a < t_1 < \cdots < t_n < b\right\} < +\infty. \tag{1.1.2}$$

It is well known that any bounded monotone function has finite pointwise variation and that any function with finite pointwise variation can be written as the difference of two bounded monotone functions.

Any curve in $AC^p(a, b; \mathscr{S})$ is uniformly continuous; if $a > -\infty$ (resp. $b < +\infty$) we will denote by $v(a+)$ (resp. $v(b-)$) the right (resp. left) limit of v, which exists since $\mathscr{S}$ is complete. The above limit exist even in the case $a = -\infty$ (resp. $b = +\infty$) if $v \in AC(a, b; \mathscr{S})$. Among all the possible choices of m in (1.1.1) there exists a minimal one, which is provided by the following theorem (see [7, 8, 15]).

Theorem 1.1.2 (Metric derivative). *Let $p \in [1, +\infty]$. Then for any curve v in $AC^p(a, b; \mathscr{S})$ the limit*

$$|v'|(t) := \lim_{s \to t} \frac{d(v(s), v(t))}{|s - t|} \tag{1.1.3}$$

exists for $\mathscr{L}^1$-a.e. $t \in (a, b)$. Moreover the function $t \mapsto |v'|(t)$ belongs to $L^p(a, b)$, it is an admissible integrand for the right hand side of (1.1.1), *and it is minimal in the following sense:*

$$\begin{gathered} |v'|(t) \le m(t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (a, b), \\ \textit{for each function } m \textit{ satisfying } (1.1.1). \end{gathered} \tag{1.1.4}$$

Proof. Let $(y_n) \subset \mathscr{S}$ be dense in $v((a, b))$ and let $\mathsf{d}_n(t) := d(y_n, v(t))$. Since all functions d_n are absolutely continuous in (a, b) the function

$$\mathsf{d}(t) := \sup_{n \in \mathbb{N}} |\mathsf{d}_n'(t)|$$

is well defined $\mathscr{L}^1$-a.e. in (a, b). Let $t \in (a, b)$ be a point where all functions d_n are differentiable and notice that

$$\liminf_{s \to t} \frac{d(v(s), v(t))}{|s - t|} \ge \sup_{n \in \mathbb{N}} \liminf_{s \to t} \frac{|\mathsf{d}_n(s) - \mathsf{d}_n(t)|}{|s - t|} = \mathsf{d}(t).$$

This inequality together with (1.1.1) shows that $\mathsf{d} \le m$ $\mathscr{L}^1$-a.e., therefore $\mathsf{d} \in L^p(a, b)$. On the other hand the definition of d gives

$$d(v(s), v(t)) = \sup_{n \in \mathbb{N}} |\mathsf{d}_n(s) - \mathsf{d}_n(t)| \le \int_s^t \mathsf{d}(r)\, dr \qquad \forall s,\, t \in (a, b),\ s \le t,$$

and therefore

$$\limsup_{s \to t} \frac{d(v(s), v(t))}{|s - t|} \le \mathsf{d}(t)$$

at any Lebesgue point t of d. □

In the next remark we deal with the case when the target space is a dual Banach space, see for instance [12].

Remark 1.1.3 (Derivative in Banach spaces). Suppose that $\mathscr{S} = \mathscr{B}$ is a *reflexive Banach space* (respectively: a *dual Banach space*): then a curve v belongs to $AC^p(a,b;\mathscr{S})$ if and only if it is differentiable (resp. weakly*-differentiable) at $\mathscr{L}^1$-a.e. point $t \in (a,b)$, its derivative v' belongs to $L^p(a,b;\mathscr{B})$ (resp. to $L^p_{w^*}(a,b;\mathscr{B})$) and

$$v(t) - v(s) = \int_s^t v'(r)\,dr \quad \forall\, a < s \le t < b. \tag{1.1.5}$$

In this case,

$$\|v'(t)\|_{\mathscr{B}} = |v'|(t) \quad \mathscr{L}^1\text{-a.e. in } (a,b). \tag{1.1.6}$$

Lemma 1.1.4 (Lipschitz and arc-length reparametrizations). *Let v be a curve in $AC(a,b;\mathscr{S})$ with length $L := \int_a^b |v'|(t)\,dt$.*
(a) *For every $\varepsilon > 0$ there exists a strictly increasing absolutely continuous map*

$$\mathsf{s}_\varepsilon : (a,b) \to (0, L_\varepsilon) \quad \textit{with } \mathsf{s}_\varepsilon(a+) = 0,\ \mathsf{s}_\varepsilon(b-) = L_\varepsilon := L + \varepsilon(b-a), \tag{1.1.7}$$

and a Lipschitz curve $\hat v_\varepsilon : (0, L_\varepsilon) \to \mathscr{S}$ such that

$$v = \hat v_\varepsilon \circ \mathsf{s}_\varepsilon, \quad |\hat v_\varepsilon'| \circ \mathsf{s}_\varepsilon = \frac{|v'|}{\varepsilon + |v'|} \in L^\infty(a,b). \tag{1.1.8}$$

The map s_ε admits a Lipschitz continuous inverse $\mathsf{t}_\varepsilon : (0, L_\varepsilon) \to (a,b)$ with Lipschitz constant less than ε^{-1}, and $\hat v_\varepsilon = v \circ \mathsf{t}_\varepsilon$.
(b) *There exists an increasing absolutely continuous map*

$$\mathsf{s} : (a,b) \to [0,L] \quad \textit{with } \mathsf{s}(a+) = 0,\ \mathsf{s}(b-) = L, \tag{1.1.9}$$

and a Lipschitz curve $\hat v : [0,L] \to \mathscr{S}$ such that

$$v = \hat v \circ \mathsf{s}, \quad |\hat v'| = 1 \quad \mathscr{L}^1\textit{-a.e. in } [0,L]. \tag{1.1.10}$$

Proof. Let us first consider the case (a) with $\varepsilon > 0$; we simply define

$$\mathsf{s}_\varepsilon(t) := \int_a^t \big(\varepsilon + |v'|(\theta)\big)\,d\theta, \quad t \in (a,b); \tag{1.1.11}$$

s_ε is strictly increasing with $\mathsf{s}_\varepsilon' \ge \varepsilon$, $\mathsf{s}_\varepsilon\big((a,b)\big) = (0, L_\varepsilon)$, its inverse map t_ε : $(0, L_\varepsilon) \to (a,b)$ satisfies a Lipschitz condition with constant $\le \varepsilon^{-1}$, and

$$\mathsf{t}_\varepsilon' \circ \mathsf{s}_\varepsilon = \frac{1}{\varepsilon + |v'|} \quad \mathscr{L}^1\text{-a.e. in } (a,b).$$

Setting $\hat v^\varepsilon := v \circ \mathsf{t}_\varepsilon$, for every choice of $t_i = \mathsf{t}_\varepsilon(s_i)$ with $0 < s_1 < s_2 < L_\varepsilon$ we have

$$\begin{aligned} d(\hat v_\varepsilon(s_1), \hat v_\varepsilon(s_2)) = d(v(t_1), v(t_2)) &\le \int_{t_1}^{t_2} |v'|(t)\,dt \\ &\le \mathsf{s}_\varepsilon(t_2) - \mathsf{s}_\varepsilon(t_1) - \varepsilon(t_2 - t_1) = s_2 - s_1 - \varepsilon(t_2 - t_1), \end{aligned} \tag{1.1.12}$$