Lectures in Mathematics ETH Zürich

Luigi Ambrosio
Nicola Gigli
Giuseppe Savaré

Gradient Flows

in Metric Spaces and in the Space of Probability Measures

Birkhäuser

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 $d(v(s), v(t)) \leq \int_s^t |v'|(r)\,dr$ 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

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

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

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

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:

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

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

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

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"

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

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

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

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

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

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)
$\pi^i,\ \pi^{i,j}$	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*

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

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:*

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

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. $\square$

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 $\mathsf{t}_\varepsilon :$ $(0, L_\varepsilon) \to (a, b)$ 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

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

so that $\hat{v}_\varepsilon$ is 1-Lipschitz and can be extended to $[0, L_\varepsilon]$ since $\hat{v}_\varepsilon(0+) = v(a+)$ and $\hat{v}_\varepsilon(L_\varepsilon-) = v$; dividing the above inequality by $s_2 - s_1$ and passing to the limit as $s_2 \to s_1$ we get the bound

$$|\hat{v}'_\varepsilon| \circ \mathsf{s}_\varepsilon \le 1 - \frac{\varepsilon}{\varepsilon + |v'|} = \frac{|v'|}{\varepsilon + |v'|} \quad \mathscr{L}^1\text{-a.e. in } (a,b). \tag{1.1.13}$$

On the other hand,

$$\begin{aligned} d(v(t_2), v(t_1)) = d(\hat{v}_\varepsilon(s_2), \hat{v}_\varepsilon(s_1)) &\le \int_{s_1}^{s_2} |\hat{v}'_\varepsilon|(s)\, ds \\ = \int_{t_1}^{t_2} |\hat{v}'_\varepsilon|(\mathsf{s}_\varepsilon(t))\, \mathsf{s}'_\varepsilon(t)\, dt &\le \int_{t_1}^{t_2} \big(|\hat{v}'_\varepsilon| \circ \mathsf{s}_\varepsilon\big) \big(\varepsilon + |v'|\big)\, dt. \end{aligned} \tag{1.1.14}$$

By (1.1.4) we obtain

$$|v'| \le \big(|\hat{v}'_\varepsilon| \circ \mathsf{s}_\varepsilon\big) \big(\varepsilon + |v'|\big) \quad \mathscr{L}^1\text{-a.e. in } (a,b),$$

which, combined with the converse inequality (1.1.13), yields (1.1.8).

(b) We define $\mathsf{s} := \mathsf{s}_0$ for $\varepsilon = 0$ by (1.1.11) and we consider the left continuous, increasing map

$$\mathsf{t}(s) := \min\big\{t \in [a,b] : \mathsf{s}(t) = s\big\}, \quad s \in [0, L],$$

which satisfies $\mathsf{s}(\mathsf{t}(s)) = s$ in $[0, L]$. Moreover, still denoting by v its continuous extension to the closed interval $[a,b]$, we observe that

$$\mathsf{t}(\mathsf{s}(t)) \le t, \quad v(\mathsf{t}(\mathsf{s}(t)) = v(t) \quad \forall\, t \in [a,b], \tag{1.1.15}$$

since

$$d(v(\mathsf{t}(\mathsf{s}(t)), v(t)) = \int_{\mathsf{t}(\mathsf{s}(t))}^{t} |v'|(\theta)\, d\theta = \mathsf{s}(t) - \mathsf{s}(t) = 0.$$

Defining $\hat{v} := v \circ \mathsf{t}$ as above, (1.1.12) (with $\varepsilon = 0$) shows that $\hat{v}$ is 1-Lipschitz and (1.1.15) yields $v = \hat{v} \circ \mathsf{s}$. Finally, (1.1.14) shows that $|\hat{v}'| \circ \mathsf{s} = 1$ $\mathscr{L}^1$-a.e. in (a,b). □

1.2 Upper gradients

In this section we define a kind of "modulus of the gradient" for real valued functions defined on metric spaces, following essentially the approach of [81, 41].

Let $\phi : \mathscr{S} \to (-\infty, +\infty]$ be an extended real functional, with proper effective domain

$$D(\phi) := \{v \in \mathscr{S} : \phi(v) < +\infty\} \ne \varnothing. \tag{1.2.1}$$

If $\mathscr{S}$ is a vector space and ϕ is differentiable, then $|\nabla\phi|$ has the following natural variational characterization:

$$g \geq |\nabla\phi| \quad \Leftrightarrow \quad \begin{array}{c} \big|(\phi\circ v)'\big| \leq g(v)|v'| \\ \text{for every regular curve } v:(a,b)\to\mathscr{S}. \end{array} \tag{1.2.2}$$

We want to define a notion of "upper gradient" g for ϕ modeled on (1.2.2). A first possibility is to use an integral formulation of (1.2.2) along absolutely continuous curves.

Definition 1.2.1 (Strong upper gradients, [81, 41]). *A function $g : \mathscr{S} \to [0, +\infty]$ is a* strong upper gradient *for ϕ if for every absolutely continuous curve $v \in AC(a, b; \mathscr{S})$ the function $g \circ v$ is Borel and*

$$\big|\phi(v(t)) - \phi(v(s))\big| \leq \int_s^t g(v(r))|v'|(r)\,dr \quad \forall\, a < s \leq t < b. \tag{1.2.3}$$

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

$$|(\phi\circ v)'(t)| \leq g(v(t))|v'|(t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (a,b). \tag{1.2.4}$$

We also introduce a weaker notion, based on a pointwise formulation:

Definition 1.2.2 (Weak upper gradients). *A function $g : \mathscr{S} \to [0, +\infty]$ is a* weak upper gradient *for ϕ if every curve $v \in AC(a, b; \mathscr{S})$ such that*

(i) *$g \circ v|v'| \in L^1(a, b)$;*

(ii) *$\phi \circ v$ is $\mathscr{L}^1$-a.e. equal in (a, b) to a function φ with finite pointwise variation in (a, b);*

we have

$$|\varphi'(t)| \leq g(v(t))|v'|(t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (a,b). \tag{1.2.5}$$

In this case, if $\phi \circ v \in AC(a, b)$ then $\varphi = \phi \circ v$ and (1.2.3) holds.

Remark 1.2.3 (Approximate derivative). Condition (ii) of Definition 1.2.2 is equivalent to say that $\phi \circ v$ has *essential bounded variation* in (a, b). Accordingly, condition (1.2.5) could be stated without any reference to φ by replacing $\varphi'(t)$ with the *approximate derivative* of $\phi \circ v$ (see Definition 5.5.1).

Among all the possible choices for an upper gradient of ϕ, we recall the definition of the local and global slopes (see also [41], [52]):

Definition 1.2.4 (Slopes). *The* local and global slopes *of ϕ at $v \in D(\phi)$ are defined by*

$$|\partial\phi|(v) := \limsup_{w\to v} \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)}, \quad \mathfrak{l}_\phi(v) := \sup_{w\neq v} \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)}. \tag{1.2.6}$$

Theorem 1.2.5 (Slopes are upper gradients). *The function $|\partial\phi|$ is a weak upper gradient for ϕ. If ϕ is d-lower semicontinuous then $\mathfrak{l}_\phi$ is a strong upper gradient for ϕ.*

Proof. In order to show that $|\partial\phi|$ is a weak upper gradient we consider an absolutely continuous curve $v : (a,b) \to \mathscr{S}$ satisfying the assumptions of Definition 1.2.2; we introduce the set

$$A := \big\{ t \in (a,b) : \ \phi(v(t)) = \varphi(t), \ \varphi \text{ is differentiable at } t, \ \exists\, |v'|(t) \big\}$$

and we observe that $(a,b) \setminus A$ is $\mathscr{L}^1$-negligible.

If the derivative of φ vanishes at $t \in A$ then (1.2.5) is surely satisfied, therefore it is not restrictive to consider points $t \in A$ such that $\varphi'(t) \neq 0$. In order to fix the ideas, let us suppose that $t \in A$ and $\varphi'(t) > 0$; since $d(v(s), v(t)) \neq 0$ when $s \in A \setminus \{t\}$ belongs to a suitable neighborhood of t we have

$$\begin{aligned} |\varphi'(t)| = \varphi'(t) &= \lim_{s\uparrow t, s\in A} \frac{\phi(v(t)) - \phi(v(s))}{t-s} = \lim_{s\uparrow t, s\in A} \frac{\phi(v(t)) - \phi(v(s))}{d(v(s),v(t))} \frac{d(v(s),v(t))}{t-s} \\ &\le \limsup_{s\uparrow t, s\in A} \frac{\phi(v(t)) - \phi(v(s))}{d(v(s),v(t))} \lim_{s\uparrow t, s\in A} \frac{d(v(s),v(t))}{t-s} \le |\partial\phi|(v(t))\, |v'|(t). \end{aligned}$$

In order to check the second part of the Theorem, we notice first that $v \mapsto \mathfrak{l}_\phi(v)$ is lower semicontinuous in $\mathscr{S}$. Indeed, if $w \neq v$ and $v_h \to v$ then $w \neq v_h$ for h large enough and therefore

$$\liminf_{h\to\infty} \mathfrak{l}_\phi(v_h) \ge \liminf_{h\to\infty} \frac{\big(\phi(v_h) - \phi(w)\big)^+}{d(v_h, w)} \ge \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)}.$$

By taking the supremum w.r.t. w the lower semicontinuity follows.

Let now v be a curve in $AC(a,b;\mathscr{S})$ satisfying $\mathfrak{l}_\phi(v)|v'| \in L^1(a,b)$ and notice that $\mathfrak{l}_\phi(v)$ is lower semicontinuous, therefore Borel. We apply Lemma 1.1.4 with $\varepsilon = 0$, and for the increasing and absolutely continuous map $\mathsf{s} := \mathsf{s}_0 : [a,b] \to [0,L]$ defined by (1.1.11) we set

$$\hat v(s) := v(\mathsf{t}(s)), \quad \varphi(s) := \phi(\hat v(s)), \quad g(s) := \mathfrak{l}_\phi(\hat v(s)) \qquad s \in (0,L)$$

and we observe that for each couple $s_1,\, s_2 \in (0,L)$ we have $(\varphi(s_1) - \varphi(s_2))^+ \le g(s_1)|s_2 - s_1|$, hence

$$|\varphi(s_1) - \varphi(s_2)| \le \max[g(s_1), g(s_2)]\, |s_2 - s_1|. \tag{1.2.7}$$

The 1-dimensional change of variables formula gives

$$\int_0^L g(s)\, ds = \int_a^b \mathfrak{l}_\phi(v(t))|v'|(t)\, dt < +\infty, \tag{1.2.8}$$

therefore $g \in L^1(0,L)$ and (1.2.7) shows that φ belongs to the metric Sobolev space $W^{1,1}_m(0,L)$ in the sense of Hajłasz [80]. By a difference quotients argument this condition implies (see Lemma 1.2.6 below and [15]) that φ belongs to the conventional Sobolev space $W^{1,1}(0,L)$ and we simply have to check that φ coincides with its continuous representative. Since $\hat v$ is a Lipschitz map we immediately see that φ is lower semicontinuous in $(0,L)$: therefore continuity follows if we show that

$$\limsup_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \varphi(s+r)\,dr \le \varphi(s) \quad \text{for all } s\in(0,L). \tag{1.2.9}$$

Invoking (1.2.7) we get

$$\limsup_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \big(\varphi(s+r)-\varphi(s)\big)\,dr \le \limsup_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \big(\varphi(s+r)-\varphi(s)\big)^+\,dr$$
$$\le \limsup_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} g(s+r)\,|r|\,dr \le \limsup_{\varepsilon\downarrow 0} \frac{1}{2}\int_{-\varepsilon}^{\varepsilon} g(s+r)\,dr = 0.$$

Since $\phi(v(t)) = \phi(\hat v(\mathsf{s}(t))) = \varphi(\mathsf{s}(t))$, we obtain the absolute continuity of $\phi\circ v$; using the inequality $\mathfrak{l}_\phi(v) \ge |\partial\phi|(v)$ and the the fact that $|\partial\phi|$ is an upper gradient we conclude. □

Lemma 1.2.6. *Let φ, $g \in L^1(a,b)$ with $g \ge 0$ and assume that there exists a $\mathscr{L}^1$-negligible set $N \subset (a,b)$ such that*

$$|\varphi(s)-\varphi(t)| \le (g(s)+g(t))\,|s-t| \qquad \forall s,\, t\in(a,b)\setminus N.$$

Then $\varphi \in W^{1,1}(a,b)$ and $|\varphi'| \le 2g$ $\mathscr{L}^1$-a.e. in (a,b).

Proof. For every $\zeta \in C^\infty_c(a,b)$ we have

$$\begin{aligned} T(\zeta) &:= \int_a^b \varphi(t)\,\zeta'(t)\,dt = \lim_{h\to 0}\int_a^b \varphi(t)\frac{\zeta(t+h)-\zeta(t)}{h}\,dt \\ &= \lim_{h\to 0}\int_a^b \frac{\varphi(t-h)-\varphi(t)}{h}\zeta(t)\,dt \le \limsup_{h\to 0}\int_a^b \big(g(t-h)+g(t)\big)|\zeta(t)|\,dt \\ &= 2\int_a^b g(t)|\zeta(t)|\,dt \le 2\|g\|_{L^1(a,b)}\sup_{[a,b]}|\zeta|. \end{aligned}$$

We obtain from Riesz representation theorem that T can be represented by a signed measure λ in (a,b) having total variation less that $2\|g\|_{L^1(a,b)}$. Then, the inequality

$$\left|\int_a^b \zeta(t)\,d\lambda\right| \le 2\int_a^b |\zeta(t)||g(t)|\,dt \qquad \forall \zeta\in C^\infty_c(a,b)$$

immediately gives that $|\lambda| \le 2|g|\mathscr{L}^1$. □

1.3 Curves of maximal slope

The notion of curves of maximal slope have been introduced (in a slight different form) in [52] and further developed in [53, 95]. Our presentation essentially follows the ideas of [8], combining them with the "upper gradient" point of view.

In order to motivate the main Definition 1.3.2 of this section, let us initially consider the finite dimensional case of the Euclidean space $\mathscr{S} := \mathbb{R}^d$ with scalar product $\langle\cdot,\cdot\rangle$ and norm $|\cdot|$. The gradient $\nabla\phi$ of a smooth real functional $\phi : \mathscr{S} \to \mathbb{R}$ can be defined taking the derivative of ϕ along regular curves, i.e.

$$\boldsymbol{g} = \nabla\phi \quad \Leftrightarrow \quad \begin{array}{c} \big(\phi\circ v\big)' = \langle \boldsymbol{g}(v), v'\rangle \\ \text{for every regular curve } v:(0,+\infty)\to\mathscr{S}, \end{array} \tag{1.3.1}$$

and its modulus $|\nabla\phi|$ has the natural variational characterization (1.2.2). In this case, a steepest descent curve u for ϕ, i.e. a solution of the equation

$$u'(t) = -\nabla\phi(u(t)) \qquad t > 0, \tag{1.3.2}$$

can be characterized by the following two scalar conditions in $(0,+\infty)$

$$\big(\phi\circ u\big)' = -|\nabla\phi(u)|\,|u'|, \tag{1.3.3a}$$

$$|u'| = |\nabla\phi(u)|; \tag{1.3.3b}$$

in fact, (1.3.3a) forces the direction of the velocity u' to be opposite to the gradient one, whereas the modulus of u' is determined by (1.3.3b). (1.3.3a,b) are also equivalent, via Young inequality, to the single equation

$$\big(\phi\circ u\big)' = -\frac{1}{2}|u'|^2 - \frac{1}{2}|\nabla\phi(u)|^2 \quad \text{in } (0,+\infty). \tag{1.3.3c}$$

It is interesting to note that we can impose (1.3.3a,b) or (1.3.3c) as a system of differential inequalities in the couple (u, g), the first one saying that the function g is an upper bound for the modulus of the gradient (an "upper gradient", as we have seen in the previous section)

$$\big|\big(\phi\circ v\big)'\big| \le g(v)|v'| \quad \text{for every regular curve } v:(0,+\infty)\to\mathscr{S}, \tag{1.3.4a}$$

the second one imposing that the functional ϕ decreases along u as much as possible compatibly with (1.3.4a), i.e.

$$\big(\phi\circ u\big)' \le -g(u)|u'| \quad \text{in } (0,+\infty), \tag{1.3.4b}$$

and the last one prescribing the dependence of $|u'|$ on $g(u)$

$$|u'| = g(u) \quad \text{in } (0,+\infty), \tag{1.3.4c}$$

or even in a single formula

$$\big(\phi \circ u\big)' \le -\frac{1}{2}|u'|^2 - \frac{1}{2}g(u)^2 \quad \text{in } (0,+\infty). \tag{1.3.4d}$$

Whereas equations (1.3.1), (1.3.2) make sense only in a Hilbert-Riemannian framework, the formulation (1.3.4a,b,c,d) is of purely metric nature and can be extended to more general metric spaces $(\mathscr{S}, d)$, provided we understand $|u'|$ as the metric derivative of u. Of course, the concept of upper gradient provides only an upper estimate for the modulus of $\nabla\phi$ in the regular case, but it is enough to define steepest descent curves, i.e. curves which realize the minimal selection of $\frac{d}{dt}\phi(u(t))$ compatible with (1.2.4).

Remark 1.3.1 (**p, q variants).** Instead of (1.3.2) we can consider more general nonlinear coupling between time derivative and gradient, which naturally appears when a non euclidean distance in $\mathscr{S}$ is considered: in the last section of the present chapter we will briefly discuss the case of a Banach space.
In the easier Euclidean setting, the simplest generalization leads to an equation of the type

$$j(u'(t)) = -\nabla\phi(u(t)) \quad t > 0, \quad \text{with} \quad j(v) = \alpha(|v|)\frac{v}{|v|} \tag{1.3.5}$$

for a continuous, strictly increasing and surjective map $\alpha : [0,+\infty) \to [0,+\infty)$. In this case, the velocity u' still takes the opposite direction of $\nabla\phi(u)$ yielding (1.3.3a), but equation (1.3.3b) for its modulus is substituted by the monotone condition

$$\alpha(|u'|) = |\nabla\phi(u)|. \tag{1.3.6}$$

Introducing the strictly convex primitive function ψ of α and its conjugate ψ^*

$$\psi(z) := \int_0^z \alpha(r)\,dr, \quad \psi^*(z^*) := \max_{x\in[0,+\infty)} z^*x - \psi(x), \quad z, z^* \in [0,+\infty), \tag{1.3.7}$$

(1.3.5) is therefore equivalent to

$$\big(\phi \circ u\big)' \le -\psi(|u'|) - \psi^*(|\nabla\phi(u)|) \quad \text{in } (0,+\infty), \tag{1.3.8}$$

which, in the metric framework, could be relaxed to

$$\big(\phi \circ u\big)' \le -\psi(|u'|) - \psi^*(g((u)) \quad \text{in } (0,+\infty), \tag{1.3.9}$$

for an upper gradient g satisfying (1.3.4a).
Even if many results could be extended to this general situation, for the sake of simplicity in the present book we will consider only a p, q-setting, where $p, q \in (1,+\infty)$ are conjugate exponent $p^{-1} + q^{-1} = 1$, corresponding to the choices

$$\alpha(z) := z^{p-1}, \quad \psi(z) = \frac{1}{p}z^p, \quad \psi^*(z^*) = \frac{1}{q}(z^*)^q,$$

and to the equation

$$j_p(u'(t)) = -\nabla\phi(u(t)), \quad j_p(v) := \begin{cases} |v|^{p-2}v & \text{for } v \neq 0, \\ 0 & \text{if } v = 0. \end{cases} \tag{1.3.10}$$

Thus the idea is that (1.3.3a) is still imposed and (1.3.3b) is substituted by

$$|u'|^{p-1} = |\nabla\phi(u)| \quad \text{or, equivalently,} \quad |u'| = |\nabla\phi(u)|^{q-1} \tag{1.3.11}$$

and therefore, taking into account the strict convexity of $|\cdot|^p$, in the purely metric framework we end up with the inequality

$$\big(\phi \circ u\big)' \leq -\frac{1}{p}|u'|^p - \frac{1}{q}g(u)^q \quad \text{in } (0, +\infty). \tag{1.3.12}$$

Recalling (1.3.4a), (1.3.4d), and (1.3.12), we introduce the following definition:

Definition 1.3.2 (Curves of maximal slope). *We say that a locally absolutely continuous map $u : (a, b) \to \mathscr{S}$ is a p-*curve of maximal slope*, $p \in (1, +\infty)$ (we will often omit to mention p in the quadratic case), for the functional ϕ with respect to its upper gradient g, if $\phi \circ u$ is $\mathscr{L}^1$-a.e. equal to a non-increasing map φ and*

$$\varphi'(t) \leq -\frac{1}{p}|u'|^p(t) - \frac{1}{q}g^q(u(t)) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (a, b). \tag{1.3.13}$$

Remark 1.3.3. Observe that (1.2.5) and (1.3.13) yield

$$|u'|^p(t) = g^q(u(t)) = -\varphi'(t) \quad \mathscr{L}^1\text{-a.e. in } (a, b), \tag{1.3.14}$$

in particular $u \in AC^p_{\rm loc}(a, b; \mathscr{S})$ and $g \circ u \in L^q_{\rm loc}(a, b)$. If u is a curve of maximal slope for ϕ with respect to a strong upper gradient g, then $\phi(u(t)) \equiv \varphi(t)$ is a locally absolutely continuous map in (a, b) and the energy identity

$$\frac{1}{p}\int_s^t |u'|^p(r)\,dr + \frac{1}{q}\int_s^t g^q(r)\,dr = \phi(u(s)) - \phi(u(t)) \tag{1.3.15}$$

holds in each interval $[s, t] \subset (a, b)$.

1.4 Curves of maximal slope in Hilbert and Banach spaces

We conclude this chapter dedicated to slopes and upper gradients by giving a closer look to the case when

$$\mathscr{S} = \mathscr{B} \text{ is a Banach space with norm } \|\cdot\|; \tag{1.4.1}$$

we denote by $\langle \cdot, \cdot \rangle$ the duality between $\mathscr{B}$ and its dual $\mathscr{B}'$ and by $\| \cdot \|_*$ the dual norm in $\mathscr{B}'$.
Let us first consider a C^1 functional $\phi : \mathscr{B} \to \mathbb{R}$: the chain rule (1.3.1) characterizes the Frèchet differential $D\phi : \mathscr{B} \to \mathscr{B}'$, which is defined by

$$\boldsymbol{g} = D\phi(v) \quad \Leftrightarrow \quad \lim_{w \to v} \frac{\phi(w) - \phi(v) - \langle \boldsymbol{g}, w - v \rangle}{\|w - v\|} = 0 \quad \forall\, v \in \mathscr{B}.$$

Since the metric derivative $|v'|$ of a regular curve v coincides with the norm of the velocity vector $\|v'\|$, it is easy to show that upper gradients involve the dual norm of $D\phi(v)$: by (1.2.2) g is an upper gradient for ϕ iff

$$g \geq \|D\phi(v)\|_* \quad \forall\, v \in \mathscr{B}. \tag{1.4.2}$$

In this case, the steepest descent conditions (1.3.3a), (1.3.4b) become

$$\langle D\phi(u), u' \rangle = \big(\phi \circ u\big)' \leq -\|u'\|\, g(u) \leq -\|u'\|\, \|D\phi(u)\|_*, \tag{1.4.3}$$

whereas (1.3.3b) could take the more general p, q form (1.3.11) (but see also (1.3.6))

$$\|u'\|^{p-1} = \|D\phi(u)\|_*. \tag{1.4.4}$$

Combining (1.4.3) and (1.4.4) we end up with the doubly nonlinear differential inclusion

$$\mathfrak{J}_p(u'(t)) \ni -D\phi(u(t)) \quad t > 0, \tag{1.4.5}$$

where $\mathfrak{J}_p : \mathscr{B} \to 2^{\mathscr{B}'}$ is the p-duality map defined by

$$\xi \in \mathfrak{J}_p(v) \quad \Leftrightarrow \quad \langle \xi, v \rangle = \|v\|^p = \|\xi\|_*^q = \|v\|\, \|\xi\|_*, \tag{1.4.6}$$

which is single valued if the norm $\| \cdot \|$ of $\mathscr{B}$ is differentiable.

We want now to extend the previous considerations to a non-smooth setting. Recall that the *Fréchet subdifferential* $\partial\phi(v) \subset \mathscr{B}'$ of a functional $\phi : \mathscr{B} \to (-\infty, +\infty]$ at a point $v \in D(\phi)$ is defined by

$$\xi \in \partial\phi(v) \quad \Longleftrightarrow \quad \liminf_{w \to v} \frac{\phi(w) - (\phi(v) + \langle \xi, w - v \rangle)}{\|w - v\|_{\mathscr{B}}} \geq 0. \tag{1.4.7}$$

As usual, $D(\partial\phi)$ denotes the subset of $\mathscr{B}$ given by all the elements $v \in D(\phi)$ such that $\partial\phi(v) \neq \varnothing$; $\partial\phi(v)$ is a (strongly) closed convex set and we will suppose that

$$\partial\phi(v) \quad \text{is weakly}^* \text{ closed} \quad \forall\, v \in D(\partial\phi); \tag{1.4.8}$$

(1.4.8) is surely satisfied if e.g. $\mathscr{B}$ is reflexive or ϕ is convex (see the next Proposition 1.4.4). $\partial^\circ\phi(v)$ is the subset of elements of minimal (dual) norm in $\partial\phi(v)$, which reduces to a single point if the dual norm of $\mathscr{B}$ is strictly convex. Notice that

$$|\partial\phi|(v) = \limsup_{w \to 0} \frac{\phi(v) - \phi(v + w)}{\|w\|} \leq \limsup_{w \to 0} \langle \xi, \frac{w}{\|w\|} \rangle \leq \|\xi\|_* \quad \forall\, \xi \in \partial\phi(v).$$

Therefore, if we extend the function $v \mapsto \|\partial^\circ\phi(v)\|_*$ to $+\infty$ outside of $D(\partial\phi)$ we have

$$|\partial\phi|(v) \leq \|\partial^\circ\phi(v)\|_* \quad \forall\, v \in \mathscr{B}, \tag{1.4.9}$$

and we obtain from Theorem 1.2.5 that

$$\text{the map } \ v \mapsto \|\partial^\circ\phi(v)\|_* \quad \text{is a weak upper gradient for } \phi. \tag{1.4.10}$$

In the next proposition we characterize the ($\mathscr{L}^1$-a.e. differentiable) curves of maximal slope with respect to the upper gradient (1.4.10) as the solution of a suitable doubly nonlinear differential inclusion: in the case when $\mathscr{S}$ is a reflexive Banach space and ϕ is convex, these kind of evolution equations have been studied in [43, 42]; we refer to these contributions and to [127] for many examples of partial differential equations which can be studied by this abstract approach.

Proposition 1.4.1 (Doubly nonlinear differential inclusions). *Let us consider a proper l.s.c. functional $\phi : \mathscr{B} \to (-\infty, +\infty]$ satisfying* (1.4.8) *and a curve $u \in AC^p(a, b; \mathscr{B})$ which is differentiable at $\mathscr{L}^1$-a.e. point of (a, b) (see Remark* 1.1.3*). If u is a p-curve of maximal slope for ϕ with respect to the weak upper gradient* (1.4.10)*, then*

$$\mathfrak{J}_p(u'(t)) \supset -\partial^\circ\phi(u(t)) \neq \varnothing \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a, b); \tag{1.4.11}$$

in particular, if the norm of $\mathscr{B}$ is differentiable, we have

$$\mathfrak{J}_p(u'(t)) = -\partial^\circ\phi(u(t)) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a, b). \tag{1.4.12}$$

Conversely, if u satisfies (1.4.11) *and $\phi \circ u$ is ($\mathscr{L}^1$-a.e. equal to) a non increasing function, then u is a p-curve of maximal slope.*

Proof. Let us suppose that u is a p-curve of maximal slope for ϕ with respect to the upper gradient (1.4.10) and let φ be a non increasing map $\mathscr{L}^1$-a.e. equal to $\phi \circ u$ satisfying (1.3.13).
Then we can find a $\mathscr{L}^1$-negligible subset $N \subset (a, b)$ such that for every $t \in (a, b)\setminus N$ u and φ are differentiable at t, $\phi(u(t)) = \varphi(t)$, the inequality of (1.3.13) holds, and Definition (1.4.13) yields the chain rule

$$\varphi'(t) = \langle \xi, u'(t)\rangle \quad \forall\, \xi \in \partial^\circ\phi(u(t)). \tag{1.4.13}$$

It follows that for $t \in (a, b) \setminus N$

$$\langle \xi, u'(t)\rangle = \varphi'(t) \leq -\frac{1}{p}\|u'(t)\|^p - \frac{1}{q}\|\xi\|_*^q \quad \forall\, \xi \in \partial^\circ\phi(u(t)), \tag{1.4.14}$$

which yields (1.4.11). When the norm of $\mathscr{B}$ is differentiable the duality map $\mathfrak{J}_p$ is single-valued and the dual norm $\|\cdot\|_*$ is strictly convex, so that $\partial^\circ\phi$ contains at most one element: therefore (1.4.11) reduces to (1.4.12).
The converse implication follows by the same argument, since (1.4.11) and the chain rule (1.4.13) yields (1.3.13). □

Corollary 1.4.2 (Gradient flows in Hilbert spaces). *If $\mathscr{S} = \mathscr{B} = \mathscr{B}'$ is an Hilbert space, usually identified with its dual through the Risz isomorphism $\mathfrak{J}_2$, any 2-curve of maximal slope $u \in AC^2_{loc}(a,b;\mathscr{B})$ with respect to $\|\partial^\circ\phi(v)\|$ satisfies the gradient flow equation*

$$u'(t) = -\partial^\circ\phi(u(t)) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b). \tag{1.4.15}$$

Remark 1.4.3 (Non reflexive Banach spaces). The previous Proposition 1.4.1 strongly depends on the $\mathscr{L}^1$-a.e. differentiability of the considered curve and we have seen in Remark 1.1.3 that absolutely continuous curves enjoy this property if the underlying Banach space $\mathscr{B}$ satisfies the Radon-Nikodým property, e.g. if it is reflexive. One of the advantage of the purely metric formulation (1.3.13) is that it does not require any vector differentiability property of those curves and therefore it can be stated in any Banach space.

The next section will provide general existence and approximation results for curves of maximal slope with respect to the upper gradient $|\partial\phi|$: it is therefore important to know if $\|\partial^\circ\phi(v)\|_* = |\partial\phi|(v)$. In the following Proposition we deal with the case when ϕ is convex and l.s.c., proving in particular that $\|\partial^\circ\phi(v)\|_*$ is *a strong upper gradient* and coincides with $|\partial\phi|(v)$ and $\mathfrak{l}_\phi(v)$.

Proposition 1.4.4 (Slope and subdifferential of convex functions). *Let $\mathscr{B}$ be a Banach space and let $\phi : \mathscr{B} \to (-\infty, +\infty]$ be convex and l.s.c. Then*

$$\xi \in \partial\phi(v) \quad \Longleftrightarrow \quad \phi(w) - (\phi(v) + \langle \xi, w - v\rangle) \geq 0 \quad \forall\, w \in \mathscr{B} \tag{1.4.16}$$

for any $v \in D(\phi)$, the graph of $\partial\phi$ in $\mathscr{B} \times \mathscr{B}'$ is strongly-weakly* closed (in particular (1.4.8) *holds), with*

$$\xi_n \in \partial\phi(v_n),\ v_n \to v,\ \xi_n \rightharpoonup^* \xi \quad \Longrightarrow \quad \xi \in \partial\phi(v),\ \phi(v_n) \to \phi(v), \tag{1.4.17}$$

and

$$|\partial\phi|(v) = \min\Big\{\|\xi\|_* : \xi \in \partial\phi(v)\Big\} = \|\partial^\circ\phi(v)\|_* \quad \forall\, v \in \mathscr{B}. \tag{1.4.18}$$

Moreover

$$|\partial\phi|(v) = \mathfrak{l}_\phi(v) \quad \forall\, v \in \mathscr{B}, \tag{1.4.19}$$

so that, by Theorem 1.2.5, *$|\partial\phi|(v)$ is a strong upper gradient.*

Proof. The equivalence (1.4.16) and the identity (1.4.19) are simple consequence of the monotonicity of difference quotients of convex functions.

For every $w \in \mathscr{B}$ the map $(v,\xi) \mapsto \phi(w) - \phi(v) - \langle \xi, w - v\rangle$ is upper-semicontinuous with respect to the strong-weak*-topology in the product $\mathscr{B} \times \mathscr{B}'$; thus by (1.4.16) the graph of $\partial\phi$ is closed in this topology; this shows the first implication of (1.4.17). the second one follows from (1.4.16), which yields

$$|\phi(v) - \phi(v_n)| \leq \|v_n - v\|\big(\|\xi_n\|_* + \|\xi\|_*\big).$$

The inequality

$$\frac{\phi(v)-\phi(v+w)}{\|w\|} \le \langle \xi, \frac{w}{\|w\|} \rangle \qquad \forall w \in \mathscr{B} \setminus \{0\}$$

yields that $\mathfrak{l}_\phi(v)$ can be estimated from above by $\|\xi\|_{\mathscr{B}'}$ for any $\xi \in \partial\phi$. Assuming that $\mathfrak{l}_\phi(v)$ is finite, to conclude the proof we need only to show the existence of $\xi \in \partial\phi(v)$ such that $\|\xi\|_{\mathscr{B}'} \le \mathfrak{l}_\phi(v)$. By definition we know that

$$-\mathfrak{l}_\phi(v)\|w\| \le \phi(v+w) - \phi(v) \quad \forall\, w \in \mathscr{B}, \tag{1.4.20}$$

i.e. the convex epigraph

$$\big\{(w,r) \in \mathscr{B} \times \mathbb{R} : r \ge \phi(v+w) - \phi(v)\big\}$$

of the function $w \mapsto \phi(v+w) - \phi(v)$ is disjoint from the open convex hypograph in $\mathscr{B} \times \mathbb{R}$

$$\big\{(w,r) \in \mathscr{B} \times \mathbb{R} : r < -\mathfrak{l}_\phi(v)\|w\|\big\}$$

Therefore we can apply a geometric version of Hahn-Banach theorem to obtain $\xi \in \mathscr{B}'$, $\alpha \in \mathbb{R}$ such that

$$-\mathfrak{l}_\phi(v)\|w\| \le \langle \xi, w\rangle + \alpha \le \phi(v+w) - \phi(v) \quad \forall\, w \in \mathscr{B}.$$

Taking $w = 0$ we get $\alpha = 0$; the first inequality shows that $\|\xi\|_{\mathscr{B}'} \le \mathfrak{l}_\phi(v)$ and the second one, according to (1.4.16), means that $\xi \in \partial\phi(v)$. □

The above results can be easily extended to C^1 perturbations of convex functions.

Corollary 1.4.5 (C^1-perturbations of convex functions). *Let us suppose that $\phi : \mathscr{B} \to (-\infty, +\infty]$ admits the decomposition $\phi = \phi_1 + \phi_2$, where ϕ_1 is a proper, l.s.c., and convex functional, whereas $\phi_2 : \mathscr{B} \to \mathbb{R}$ is of class C^1. Then $\partial\phi = \partial\phi_1 + D\phi_2$ satisfies* (1.4.17) *and* (1.4.18)*, and $|\partial\phi|(v)$ is a strong upper gradient for ϕ.*

Proof. The sum rule $\partial\phi = \partial\phi_1 + D\phi_2$ follows directly from Definition (1.4.7) and the differentiability of ϕ_2.
In order to check the closure property (1.4.17), we observe that if $\xi_n \in \partial\phi(v_n)$ and $(v_n, \xi_n) \to (v, \xi)$ in the strong-weak* topology of $\mathscr{B} \times \mathscr{B}'$ then

$$\xi_n - D\phi_2(v_n) \in \partial\phi_1(v_n), \quad \xi_n - D\phi_2(v_n) \stackrel{*}{\rightharpoonup} \xi - D\phi_2(v) \in \partial\phi_2(v),$$

since $D\phi_2$ is continuous and ϕ_1 is convex: we obtain $\xi \in \partial\phi(v)$ and $\phi_1(v_n) \to \phi_1(v)$ which yield (1.4.17) being ϕ_2 continuous.

Finally, since we can add to ϕ_1 and subtract to ϕ_2 an arbitrary linear and continuous functional, in order to prove (1.4.18) it is not restrictive to suppose that $D\phi_2(v) = 0$; it follows that

$$
\begin{aligned}
|\partial\phi|(v) &= \limsup_{w\to v} \frac{\big(\phi(w)-\phi(v)\big)^+}{\|w-v\|} \\
&\geq \limsup_{w\to v} \frac{\big(\phi_1(w)-\phi_1(v)\big)^+}{\|w-v\|} - \limsup_{w\to v} \frac{|\phi_2(w)-\phi_2(w)|}{\|w-v\|} \\
&= |\partial\phi_1|(v) = \|\partial^\circ\phi_1(v)\|_* = \|\partial^\circ\phi(v)\|_*.
\end{aligned}
$$

Combining this inequality with the opposite one (1.4.9), we conclude. □

Let us rephrase the last conclusion of the previous Corollary, which is quite interesting in the case $\mathscr{B}$ does not satisfy the Radon-Nikodým property.

Remark 1.4.6 ("Upper" chain rule for (even non reflexive) Banach spaces). If $\phi : \mathscr{B} \to (-\infty, +\infty]$ is lower semicontinuou convex function (or a C^1 perturbation as in Corollary 1.4.5), v is a curve in $AC(a,b;\mathscr{B})$ with $\|\partial^\circ\phi\|_x * |v'| \in L^1(a,b)$, then $\phi\circ v$ is absolutely continuous in (a,b); if $\mathscr{B}$ has the Radon-Nikodým property, then

$$\frac{d}{dt}\phi\circ v(t) = \langle \partial^\circ\phi(v(t)), v'(t)\rangle \quad \text{for } \mathscr{L}^1\text{-a.e. } t \text{ in } (a,b);$$

for general Banach spaces, one can always write the upper estimate

$$\Big|\frac{d}{dt}\phi\circ v(t)\Big| \leq \|\partial^\circ\phi(v(t))\|_*|v'|(t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \text{ in } (a,b). \tag{1.4.21}$$

In the next chapter we will see how the last two proposition can be extended to a general class of functions defined on metric spaces and satisfying suitable geometric convexity conditions.

Chapter 2

Existence of Curves of Maximal Slope and their Variational Approximation

The main object of our investigation is the solution of the following Cauchy problem in the complete metric space $(\mathscr{S}, d)$:

Problem 2.0.1. *Given a functional* $\phi : \mathscr{S} \to (-\infty, +\infty]$ *and an initial datum* $u_0 \in D(\phi)$*, find a (p-)curve* u *of maximal slope in* $(0, +\infty)$ *for* ϕ *such that* $u(0+) = u_0$*.*

To keep our presentation simpler we focus our attention to the case $p = 2$; at the end of each section we will add some comments on the validity of the main statements in the general case $p \in (1, +\infty)$.

The interest of studying Problem 2.0.1 in such an abstract framework relies also on the approximation scheme which can be used to construct a curve of maximal slope. In the Euclidean setting one of the simplest ways to solve numerically (1.3.2) is provided by the Implicit Euler Method: for a given sequence $\boldsymbol{\tau} := \{\tau_n\}_{n=1}^{+\infty}$ of (strictly positive) time steps with $|\boldsymbol{\tau}| := \sup_n \tau_n < +\infty$ associated to the partition of the time interval $(0, +\infty)$

$$\begin{gathered} \mathcal{P}_{\boldsymbol{\tau}} := \{0 = t^0_{\boldsymbol{\tau}} < t^1_{\boldsymbol{\tau}} < \cdots < t^n_{\boldsymbol{\tau}} < \cdots\}, \quad I^n_{\boldsymbol{\tau}} := (t^{n-1}_{\boldsymbol{\tau}}, t^n_{\boldsymbol{\tau}}], \\ \tau_n = t^n_{\boldsymbol{\tau}} - t^{n-1}_{\boldsymbol{\tau}}, \quad \lim_{n\to\infty} t^n_{\boldsymbol{\tau}} = \sum_{k=1}^{+\infty} \tau_k = +\infty, \end{gathered} \tag{2.0.1}$$

one should find an approximate solution $U^n_{\boldsymbol{\tau}} \approx u(t^n_{\boldsymbol{\tau}}),\ n = 1, \ldots$, by solving iteratively the equation in the unknown $U^n_{\boldsymbol{\tau}}$ starting from an initial value $U^0_{\boldsymbol{\tau}} \approx u_0$

$$\frac{U^n_{\boldsymbol{\tau}} - U^{n-1}_{\boldsymbol{\tau}}}{\tau_n} = -\nabla\phi(U^n_{\boldsymbol{\tau}}) \quad n = 1, \ldots \tag{2.0.2}$$

Since (2.0.2) is the Euler equation associated to the functional in the variable V

$$\Phi(\tau_n, U_{\boldsymbol{\tau}}^{n-1}; V) := \frac{1}{2\tau_n}|V - U_{\boldsymbol{\tau}}^{n-1}|^2 + \phi(V) \qquad V \in \mathscr{S}, \tag{2.0.3a}$$

one can restrict the set of solutions of (1.3.14) to the minimum points of (2.0.3a), which can also be settled in a general metric context, simply replacing the modulus by the distance

$$\Phi(\tau_n, U_{\boldsymbol{\tau}}^{n-1}; V) := \frac{1}{2\tau_n}d^2(V, U_{\boldsymbol{\tau}}^{n-1}) + \phi(V) \qquad V \in \mathscr{S}. \tag{2.0.3b}$$

We thus end up with the recursive scheme

$$\begin{cases} U_{\boldsymbol{\tau}}^0 \text{ is given; whenever } U_{\boldsymbol{\tau}}^1, \dots, U_{\boldsymbol{\tau}}^{n-1} \text{ are known,} \\ \\ \text{find } U_{\boldsymbol{\tau}}^n \in \mathscr{S}: \quad \Phi(\tau_n, U_{\boldsymbol{\tau}}^{n-1}; U_{\boldsymbol{\tau}}^n) \le \Phi(\tau_n, U_{\boldsymbol{\tau}}^{n-1}; V) \quad \forall V \in \mathscr{S}. \end{cases} \tag{2.0.4}$$

The (multivalued) operator which provides all the solution $U_{\boldsymbol{\tau}}^n$ of (2.0.4) for a given $U_{\boldsymbol{\tau}}^{n-1}$ is sometimes called *resolvent operator*: for a general $\tau > 0$ and $U \in \mathscr{S}$ it is defined by

$$\begin{gathered} J_\tau[U] := \operatorname{argmin} \Phi(\tau, U; \cdot), \text{ i.e.} \\ U_\tau \in J_\tau[U] \iff \Phi(\tau, U; U_\tau) \le \Phi(\tau, U; V) \quad \forall V \in \mathscr{S}. \end{gathered} \tag{2.0.5}$$

Thus a sequence $\{U_{\boldsymbol{\tau}}^n\}_{n=0}^{+\infty}$ solves the recursive scheme (2.0.4) if and only if

$$U_{\boldsymbol{\tau}}^n \in J_{\tau_n}[U_{\boldsymbol{\tau}}^{n-1}] \quad \forall n \ge 1. \tag{2.0.6}$$

Definition 2.0.2 (Discrete solution). *Let us suppose that for a choice of* $\boldsymbol{\tau}$ *and* $U_{\boldsymbol{\tau}}^0 \in \mathscr{S}$ *a sequence* $\{U_{\boldsymbol{\tau}}^n\}_{n=1}^{+\infty}$ *solving* (2.0.4) *exists, so that we can interpolate the discrete values by the piecewise constant function* $\overline{U}_{\boldsymbol{\tau}}$, *defined by*

$$\overline{U}_{\boldsymbol{\tau}}(0) = U_{\boldsymbol{\tau}}^0, \qquad \overline{U}_{\boldsymbol{\tau}}(t) \equiv U_{\boldsymbol{\tau}}^n \quad \text{if } t \in (t_{\boldsymbol{\tau}}^{n-1}, t_{\boldsymbol{\tau}}^n], \quad n \ge 1. \tag{2.0.7}$$

We call $\overline{U}_{\boldsymbol{\tau}}$ *a "discrete solution" corresponding to the partition* $\mathcal{P}_{\boldsymbol{\tau}}$.

Remark 2.0.3 (Uniform partitions). From a theoretical point of view, the simpler choice of uniform partitions of time step $\tau > 0$

$$\mathcal{P}_\tau := \{0, \tau, 2\tau, \cdots\}, \qquad t_{\boldsymbol{\tau}}^n := n\tau, \qquad \tau_n := \tau = |\boldsymbol{\tau}|, \tag{2.0.8}$$

would be sufficient to state all the following existence results; in this case for $U_{\boldsymbol{\tau}}^0 := u_0$ we get

$$\overline{U}_\tau(t) \in (J_{t/n})^n[u_0] \qquad \text{with } \tau := t/n. \tag{2.0.9}$$

On the other hand, we will also address the related issue of deriving optimal error estimates for this kind of approximation scheme and in this case the possibility to choose freely the time steps is a crucial feature from the numerical point of view. The reader which is not interested in such a numerical issue can simply reformulate all the following theorems in terms of the uniform choice (2.0.8).

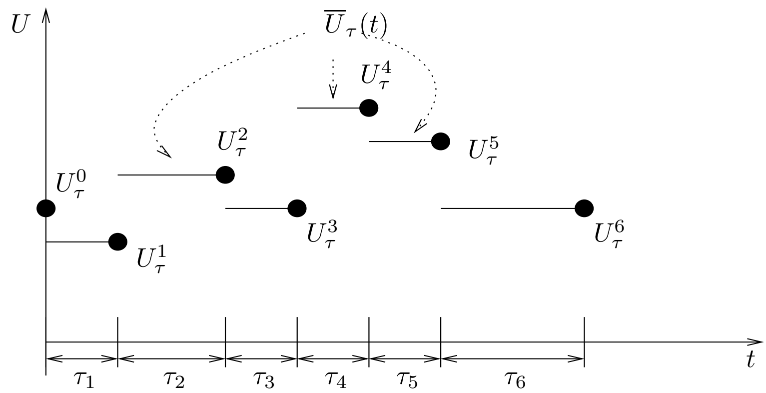


Figure 2.1: Partition of the time interval and piecewise constant interpolant.

We can thus state our main approximation problem:

Problem 2.0.4. *Find conditions on* $(u_0, U^0_{\boldsymbol\tau}, \phi, \mathscr{S})$ *which ensure that the minimization algorithm* (2.0.4) *is solvable and, up to a possible extraction of a subsequence* $(\boldsymbol\tau_k)$ *of admissible partitions with* $|\boldsymbol\tau_k| \downarrow 0$*, the curves* $\overline{U}_{\boldsymbol\tau_k}$ *converge to a solution of Problem* 2.0.1 *with respect to a suitable topology* σ *on* $\mathscr{S}$*.*

Remark 2.0.5 (The choice of the topology σ). Since the simplest choice would be to study the convergence of the scheme with respect to the topology induced by the distance d on $\mathscr{S}$, one may wonder about the opportunity to introduce another topology σ on $\mathscr{S}$. On the other hand, many examples (e.g. in the framework of reflexive Banach spaces) show the importance to deal with an auxiliary weaker topology, which allows for more flexibility to derive compactness properties. The idea here is to distinguish between the role played by the distance d (which is an essential ingredient of the approximation scheme through the functional Φ and of the definition of gradient flow through the notions of metric derivative (1.1.3) and upper gradient) and the convergence properties of the approximation scheme (which a priori can be studied with respect to different topologies).
From now on we adopt the convention to write

$$u_n \stackrel{\sigma}{\rightharpoonup} u \text{ for the convergence w.r.t. } \sigma, \quad u_n \stackrel{d}{\to} u \text{ for the convergence w.r.t. } d.$$

Besides the natural Hilbertian setting (see e.g. [28, 115] and a more detailed list of references in [102]), Problem 2.0.4 has been considered by many authors in various particular contexts [93, 127, 6, 94, 98, 83, 107]; in [51] (see also [8]) E. De Giorgi proposed a general approach to this kind of problems, suggesting that the iteration scheme itself could be used to define and select an appropriate notion of "gradient flow" in a non Euclidean setting; similar ideas [45] occur in the definition of the so called "mild solutions" for nonlinear evolution equations in Banach spaces. Here we borrow and we adapt from [51] an important definition.

Definition 2.0.6 (Minimizing movements). *For a given functional* Φ *defined as in* (2.0.3b) *and an initial datum* $u_0 \in \mathscr{S}$ *we say that a curve* $u : [0,+\infty) \to \mathscr{S}$ *is a* mimimizing movement *for* Φ *starting from* u_0 *if for every partition* $\boldsymbol{\tau}$ *(with sufficiently small* $|\boldsymbol{\tau}|$*) there exists a discrete solution* $\overline{U}_{\boldsymbol{\tau}}$ *defined as in* (2.0.4), (2.0.7) *such that*

$$\lim_{|\boldsymbol{\tau}|\downarrow 0} \phi(U^0_{\boldsymbol{\tau}}) = \phi(u_0), \quad \limsup_{|\boldsymbol{\tau}|\downarrow 0} d(U^0_{\boldsymbol{\tau}}, u_0) < +\infty, \\ \overline{U}_{\boldsymbol{\tau}}(t) \stackrel{\sigma}{\rightharpoonup} u(t) \quad \forall\, t \in [0,+\infty). \tag{2.0.10}$$

We denote by $MM(\Phi; u_0)$ *the collection of all the minimizing movements for* Φ *starting from* u_0.
Analogously, we say that a curve $u : [0,+\infty) \to \mathscr{S}$ *is a* generalized mimimizing movement *for* Φ *starting from* u_0 *if there exists a sequence of partitions* $\boldsymbol{\tau}_k$ *with* $|\boldsymbol{\tau}_k| \downarrow 0$ *and a corresponding sequence of discrete solutions* $\overline{U}_{\boldsymbol{\tau}_k}$ *defined as in* (2.0.4), (2.0.7) *such that*

$$\lim_{k\to\infty} \phi(U^0_{\boldsymbol{\tau}_k}) = \phi(u_0), \quad \limsup_{k\to\infty} d(U^0_{\boldsymbol{\tau}_k}, u_0) < +\infty, \\ \overline{U}_{\boldsymbol{\tau}_k}(t) \stackrel{\sigma}{\rightharpoonup} u(t) \quad \forall\, t \in [0,+\infty). \tag{2.0.11}$$

We denote by $GMM(\Phi; u_0)$ *the collection of all the generalized minimizing movements for* Φ *starting from* u_0.

The easiest question introduced by Problem 2.0.4, i.e. the existence of discrete solutions corresponding to given partitions $\boldsymbol{\tau}$ of $(0,+\infty)$, can be easily approached by the direct method of the Calculus of Variations, which ensures the existence of a minimum for (2.0.3b) under suitable lower semicontinuity, coercivity, and compactness assumptions: in the next (sub)section we fix the main topological properties we will deal with.

Remark 2.0.7 (The p-scheme). When we want to approximate p-curves of maximal slope, we simply change the Definition (2.0.3b) of Φ as

$$\Phi(\tau_n, V; U^n_{\boldsymbol{\tau}}) := \frac{1}{p\,\tau_n^{p-1}} d^p(V, U^{n-1}_{\boldsymbol{\tau}}) + \phi(V) \qquad V \in \mathscr{S}. \tag{2.0.12}$$

2.1 Main topological assumptions

As usual, we are dealing with a complete metric space $(\mathscr{S}, d)$; in the sequel we are supposing that

$$\sigma \text{ is an Hausdorff topology on } \mathscr{S} \text{ compatible with } d,$$

in the sense that σ is weaker than the topology induced by d and d is sequentially σ-lower semicontinuous:

$$(u_n, v_n) \stackrel{\sigma}{\rightharpoonup} (u, v) \quad \Rightarrow \quad \liminf_{n\to\infty} d(u_n, v_n) \ge d(u, v). \tag{2.1.1}$$

Here are the various kind of assumptions on the proper (see (1.2.1)) functional $\phi : \mathscr{S} \to (-\infty, +\infty]$ we are dealing with:
2.1a: Lower semicontinuity. We suppose that ϕ is sequentially σ-lower semicontinuous on d-bounded sets

$$\sup_{n,m} d(u_n, u_m) < +\infty, \quad u_n \stackrel{\sigma}{\rightharpoonup} u \quad \Rightarrow \quad \liminf_{n\to\infty} \phi(u_n) \geq \phi(u). \tag{2.1.2a}$$

2.1b: Coercivity. There exist $\tau_* > 0$ and $u_* \in \mathscr{S}$ such that

$$\phi_{\tau_*}(u_*) := \inf_{v\in\mathscr{S}} \Phi(\tau_*, u_*; v) > -\infty. \tag{2.1.2b}$$

2.1c: Compactness. Every d-bounded set contained in a sublevel of ϕ is relatively σ-sequentially compact: i.e.,

$$\begin{gathered}\text{if } (u_n) \subset \mathscr{S} \text{ with } \sup_n \phi(u_n) < +\infty, \quad \sup_{n,m} d(u_n, u_m) < +\infty, \\ \text{then } (u_n) \text{ admits a } \sigma\text{-convergent subsequence.}\end{gathered} \tag{2.1.2c}$$

Remark 2.1.1 (The case when σ is induced by d). Of course, the choice

"$\sigma :=$ the topology induced by the distance d"

is always admissible: in this case Assumption 2.1a simply means that

$$\phi \text{ is } d\text{-lower semicontinuous,} \tag{2.1.3a}$$

and 2.1c says that

$$d\text{-bounded subsets of a sublevel of } \phi \text{ are relatively compact in } \mathscr{S}. \tag{2.1.3b}$$

In particular, if

$$\text{the sublevels } \Big\{v \in \mathscr{S} : \phi(v) \leq c\Big\} \text{ are (strongly) compact} \tag{2.1.3c}$$

then all the previous assumptions hold: this is the simplest situation which is covered by this framework.
Weakly lower semicontinuous functionals in reflexive (or dual) Banach spaces provides another example which fits in this setting.

In the following

we will always assume that ϕ is lower semicontinuous and coercive,
i.e. that the first two properties 2.1a,b hold;

we will see that, in some circumstances, compactness is not necessary, since the structure of the minimization algorithm and stronger convexity assumptions on d and ϕ will directly provide convergence estimates with respect to the distance d.

Remark 2.1.2. We did not try to present a minimal set of assumptions: e.g., compactness 2.1c implies that σ is weaker than d on the sublevels of ϕ, (2.1.1) could be imposed only on the sublevel of ϕ,... On the other hand, as we already said, we will not always assume 2.1c, therefore some redundancy at this initial level simplifies the exposition in the sequel.

2.2 Solvability of the discrete problem and compactness of discrete trajectories

Observe that (2.1.2b) is surely satisfied by all τ_* and $u_* \in \mathscr{S}$ if ϕ is bounded from below, i.e.

$$\inf_{\mathscr{S}} \phi > -\infty; \quad \text{in this case } \phi_{\tau_*}(u_*) \geq \inf_{\mathscr{S}} \phi \quad \forall\, u_* \in \mathscr{S},\ \tau_* > 0. \tag{2.2.1}$$

Taking into account (2.1.2b) it is natural to define

$$\tau_*(\phi) := \sup\Big\{\tau_* > 0 : \phi_{\tau_*}(u_*) > -\infty \quad \text{for some } u_* \in \mathscr{S}\Big\}. \tag{2.2.2}$$

Lemma 2.2.1. *If $\phi_{\tau_*}(u_*) > -\infty$ as in Assumption 2.1b and $\tau < \tau_* \leq \tau_*(\phi)$, then*

$$\phi_\tau(u) \geq \phi_{\tau_*}(u_*) - \frac{1}{\tau_* - \tau} d^2(u_*, u) > -\infty \quad \forall\, u \in \mathscr{S}, \tag{2.2.3}$$

$$d^2(v,u) \leq \frac{4\tau\tau_*}{\tau_* - \tau}\Big(\Phi(\tau,u;v) - \phi_{\tau_*}(u_*) + \frac{1}{\tau_* - \tau} d^2(u_*,u)\Big) \quad \forall\, u,\, v \in \mathscr{S}. \tag{2.2.4}$$

In particular, the sublevels of $\Phi(\tau,u;\cdot)$ are bounded in $\mathscr{S}$.

Proof. Invoking the Cauchy-type inequality

$$d^2(v,u_*) \leq (1+\varepsilon) d^2(v,u) + (1+\varepsilon^{-1}) d^2(u_*,u) \quad \forall\, \varepsilon > 0,\ u, v \in \mathscr{S}, \tag{2.2.5}$$

we get for $\varepsilon := (\tau_* - \tau)/(\tau_* + \tau)$

$$\frac{1}{2\tau_*} d^2(v,u_*) \leq \frac{1}{\tau + \tau_*} d^2(v,u) + \frac{1}{\tau_* - \tau} d^2(u_*,u),$$

so that (2.1.2b) yields for each $u, v \in \mathscr{S}$ and $\tau < \tau_*$

$$\begin{aligned}
\Phi(\tau,u;v) &= \frac{\tau_* - \tau}{2\tau(\tau+\tau_*)} d^2(v,u) + \frac{1}{\tau+\tau_*} d^2(v,u) + \phi(v) \\
&\geq \frac{\tau_* - \tau}{4\tau\tau_*} d^2(v,u) + \phi_{\tau_*}(u_*) - \frac{1}{\tau_* - \tau} d^2(u_*,u) \qquad (2.2.6)\\
&\geq \phi_{\tau_*}(u_*) - \frac{1}{\tau_* - \tau} d^2(u_*,u). \qquad (2.2.7)
\end{aligned}$$

We obtain (2.2.3) by taking the infimum w.r.t. v in (2.2.7); (2.2.4) follows directly from (2.2.6). □

Corollary 2.2.2 (Existence of the discrete solutions). *If the topological assumptions of Section 2.1 are verified, then for every $\tau < \tau_*(\phi)$ and $u \in \mathscr{S}$ the functional $\Phi(\tau,u;\cdot)$ admits a minimum in $\mathscr{S}$; in particular for every choice of $U^0_{\boldsymbol{\tau}} \in \mathscr{S}$ and of a partition $\mathcal{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \tau_*(\phi)$, there exists at least one discrete solution $\overline{U}_{\boldsymbol{\tau}}$.*

Proof. (2.2.4) shows that both $d(v,u)$ and $\phi(v)$ remain bounded on the sublevels of $\Phi(\tau,u;v)$. Lower semicontinuity and compactness yield

$$\text{the sublevels } \Big\{v \in \mathscr{S} :\ \Phi(\tau,u;v) \le c\Big\} \text{ are } \sigma\text{-sequentially compact.}$$

The existence of a minimum for (2.0.4) then follows by a well known compactness and lower semicontinuity argument. □

The following preliminary result provides compactness for the family of discrete solutions:

Proposition 2.2.3 (Compactness). *Let us suppose that all the assumptions of Section* 2.1 *hold and let* Λ *be a family of partitions with* $\inf_{\boldsymbol{\tau}\in\Lambda}|\boldsymbol{\tau}| = 0$. *If the corresponding family of initial data* $\{U^0_{\boldsymbol{\tau}}\}_{\boldsymbol{\tau}\in\Lambda}$ *satisfies*

$$\sup_{\boldsymbol{\tau}\in\Lambda} \phi(U^0_{\boldsymbol{\tau}}) < +\infty, \qquad \sup_{\boldsymbol{\tau}\in\Lambda} d(U^0_{\boldsymbol{\tau}}, u_0) < +\infty, \tag{2.2.8}$$

then there exist a sequence $(\boldsymbol{\tau}_k) \subset \Lambda$ *with* $|\boldsymbol{\tau}_k| \downarrow 0$ *and a limit curve* $u \in AC^2_{\rm loc}([0,+\infty);\mathscr{S})$ *such that*

$$\overline{U}_{\boldsymbol{\tau}_k}(t) \stackrel{\sigma}{\rightharpoonup} u(t) \qquad \forall\, t \in [0,+\infty). \tag{2.2.9}$$

In particular, if $U^0_{\boldsymbol{\tau}_k} \stackrel{\sigma}{\rightharpoonup} u_0$ *and* $\phi(U^0_{\boldsymbol{\tau}_k}) \to \phi(u_0)$ *as* $k \to \infty$, *then* $u(0+) = u_0$ *and* $u \in GMM(\Phi;u_0)$, *which is therefore a nonempty set.*

We prove this proposition in the next Section 3.

Remark 2.2.4 (p-estimates). In the general case $p \neq 2$, Lemma 2.2.1 and Proposition 2.2.3 still hold (with different constants) simply replacing 2 with p: thus the limiting curve belongs to $AC^p_{\rm loc}([0,+\infty);\mathscr{S})$.

2.3 Generalized minimizing movements and curves of maximal slope

In this section we present two different sets of general conditions which provide a general answer for Problems 2.0.1, 2.0.4, and a direct connection between curves of maximal slope and generalized minimizing movements. They are both related to some kind of lower semicontinuity property of the local slope of ϕ, which can be well expressed by its *relaxed slope*, i.e. (a slight modification of) the *sequentially* σ*-lower semicontinuous envelope* of $|\partial\phi|$:

$$|\partial^-\phi|(u) := \inf\Big\{\liminf_{n\to\infty} |\partial\phi|(u_n) : u_n \stackrel{\sigma}{\rightharpoonup} u,\ \sup_n\{d(u_n,u),\phi(u_n)\} < +\infty\Big\}. \tag{2.3.1}$$

This following result holds (up to considering the appropriate p-scheme, cf. Remark 2.0.7) for every $p \in (1, +\infty)$. Notice that the compactness assumption 2.1c (which was a crucial ingredient in Proposition 2.2.3) is not needed here: Theorems 2.3.1 and 2.3.3 hold whenever one knows that a curve u belongs to $GMM(\Phi; u_0)$; besides Proposition 2.2.3, they can be combined with any convergence results for the variational Euler scheme (2.0.4), e.g. with the results of Chapter 4.

Theorem 2.3.1 ($GMM(\phi; u_0)$ are curves of maximal slope I). *Let us assume that ϕ is lower semicontinuous and coercive according to Assumptions* 2.1a,b*; if*

$$v \in \mathscr{S} \mapsto |\partial^- \phi|(v) \quad \textit{is a } \text{weak upper gradient } \textit{for } \phi, \tag{2.3.2}$$

and ϕ satisfies the continuity condition

$$\sup_{n\in\mathbb{N}} \Big\{ |\partial\phi|(v_n), d(v_n, v_0), \phi(v_n) \Big\} < +\infty, \quad v_n \overset{\sigma}{\rightharpoonup} v \quad \Rightarrow \quad \phi(v_n) \to \phi(v), \tag{2.3.3}$$

then every curve $u \in GMM(\Phi; u_0)$ with $u_0 \in D(\phi)$ is a curve of maximal slope for ϕ w.r.t. $|\partial^- \phi|$.

Remark 2.3.2. Observe that, in view of Theorem 1.2.5, (2.3.2) is always satisfied if $|\partial\phi|$ is σ-sequentially lower semicontinuous, i.e. $|\partial^-\phi| = |\partial\phi|$.

In order to state our strongest result we define a piecewise constant function $|U'_{\boldsymbol{\tau}}|$ on $(0, +\infty)$, relative to the partition $\mathcal{P}_{\boldsymbol{\tau}}$, by

$$|U'_{\boldsymbol{\tau}}|(t) = \frac{d(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}})}{t^n_{\boldsymbol{\tau}} - t^{n-1}_{\boldsymbol{\tau}}} \quad \text{if } t \in (t^{n-1}_{\boldsymbol{\tau}}, t^n_{\boldsymbol{\tau}}). \tag{2.3.4}$$

Our notation is justified by the fact that $|U'_{\boldsymbol{\tau}}|$ is really the modulus of the derivative of the piecewise affine interpolant of $U^n_{\boldsymbol{\tau}}$ when $\mathscr{S}$ in an Hilbert space and d is induced by its scalar product.

Theorem 2.3.3 ($GMM(\Phi; u_0)$ are curves of maximal slope II (energy identity)). *Suppose that the lower semicontinuity and coercivity assumptions* 2.1a,b *hold; if*

$$v \in \mathscr{S} \mapsto |\partial^- \phi|(v) \textit{ is a } \text{strong upper gradient } \textit{for } \phi,$$

then every curve $u \in GMM(\Phi; u_0)$ with $u_0 \in D(\phi)$ is a curve of maximal slope for ϕ w.r.t. $|\partial^-\phi|$ and in particular u satisfies the energy identity

$$\tfrac{1}{2} \int_0^T |u'|^2(t)\, dt + \tfrac{1}{2} \int_0^T |\partial^- \phi|^2(u(t))\, dt + \phi(u(T)) = \phi(u_0) \qquad \forall T > 0. \tag{2.3.5}$$

Moreover, if $\{\overline{U}_{\boldsymbol{\tau}_k}\}_{k\in\mathbb{N}}$ is a sequence of discrete solutions satisfying (2.2.8) *and* (2.2.9)*, we have*

$$\lim_{n\to\infty} \phi(\overline{U}_{\tau_n}(t)) = \phi(u(t)) \quad \forall\, t \in [0, +\infty), \tag{2.3.6}$$

$$\lim_{n\to\infty} |\partial\phi|(\overline{U}_{\tau_n}) = |\partial^-\phi|(u) \quad \textit{in } L^2_{\text{loc}}([0, +\infty)), \tag{2.3.7}$$

$$\lim_{n\to\infty} |U'_{\tau_n}| = |u'| \quad \textit{in } L^2_{\text{loc}}([0, +\infty)). \tag{2.3.8}$$

In the case $p \neq 2$ the energy identity reads

$$\frac{1}{p}\int_0^T |u'|^p(t)\,dt + \frac{1}{q}\int_0^T |\partial^-\phi|^q(u(t))\,dt + \phi(u(T)) = \phi(u_0) \qquad \forall T > 0, \quad (2.3.9)$$

and the limiting relations (2.3.7), (2.3.8) *should be intended in* $L^q_{\rm loc}([0,+\infty))$, $L^p_{\rm loc}([0,+\infty))$ *respectively.*

Remark 2.3.4. Whenever the functional ϕ satisfies the topological assumptions of Section 2.1, the previous theorems 2.3.1 and 2.3.3 can be applied by following at least two different strategies:

(i) One can try to show that the slope $|\partial\phi|$ is σ-lower semicontinuous (i.e. $|\partial^-\phi| = |\partial\phi|$). In the case when $\mathscr{S}$ is a Banach space and (1.4.18) holds, as in Section 1.4, this property usually corresponds to a strong-weak* closure of the graph of the Fréchet subdifferential of ϕ, as we will see in the next example 2.3.5. Once the σ-lower semicontinuity of $|\partial\phi|$ is proved, then one has to check the continuity property (2.3.3) (for Theorem 2.3.1) or that $|\partial\phi|$ is a *strong* upper gradient for ϕ (for Theorem 2.3.3).

(ii) The second possibility, when the slope $|\partial\phi|$ is not lower semicontinuous, is to prove directly that the relaxed slope is an upper gradient for ϕ, i.e. it satisfies a sort of chain rule. This approach is quite useful to dealing with gradient flows of non regular perturbations of convex functional in Hilbert spaces and has been applied to some evolution equations arising in quasi-stationary phase field problems [114].

We postpone the proofs and more detailed statements of Theorems 2.3.1 and 2.3.3 to the next chapter and we conclude the present section by an important application to the Banach case.

Example 2.3.5 (Doubly nonlinear evolution equations in Banach spaces). Let us consider the Banach space setting $\mathscr{S} = \mathscr{B}$ introduced in Section 1.4 and let us suppose that $\mathscr{B}$ satisfies the Radon-Nikodym property, so that absolutely continuous curves in $\mathscr{B}$ are $\mathscr{L}^1$-a.e. differentiable.
We want to apply the previous metric results (following the first strategy of Remark 2.3.4) to find solutions of the doubly nonlinear differential inclusion (1.4.11) for a functional $\phi : \mathscr{B} \to (-\infty, +\infty]$. Two properties seem crucial: the first one establishes a link between the metric slope and the Fréchet subdifferential of ϕ

$$|\partial\phi|(v) = \min\Big\{\|\xi\|_* : \xi \in \partial\phi(v)\Big\} = \|\partial^\circ\phi(v)\|_* \quad \forall\, v \in \mathscr{B}, \qquad (2.3.10a)$$

and the second one is σ-weak* closure of the graph of $\partial\phi$ in $\mathscr{B} \times \mathscr{B}'$

$$\xi_n \in \partial\phi(v_n),\ v_n \overset{\sigma}{\rightharpoonup} v,\ \xi_n \rightharpoonup^* \xi,\ \sup_n \phi(v_n) < +\infty \implies \xi \in \partial\phi(v). \quad (2.3.10b)$$

The following result is immediate:

Lemma 2.3.6 (Lower semicontinuity of $|\partial\phi|$). *Let us suppose that the functional $\phi : \mathscr{B} \to (-\infty, +\infty]$ satisfies* (2.3.10a) *and* (2.3.10b)*. Then $|\partial^-\phi| = |\partial\phi|$.*

Proof. If $(v_n) \subset \mathscr{B}$ converges to v in the topology σ and $|\partial\phi|(v_n) \to \rho < +\infty$ as $n \to \infty$, by (2.3.10a) there exists $\xi_n \in \partial\phi(v_n)$ such that $\|\xi_n\|_* = |\partial\phi|(v_n)$ is uniformly bounded; up to an extraction of a suitable subsequence, we can suppose that $\xi_n \rightharpoonup^* \xi$ and, by (2.3.10b), $\xi \in \partial\phi(v)$. Since the dual norm is weakly* lower semicontinuous, a further application of (2.3.10a) yields

$$|\partial\phi|(v) \le \|\xi\|_* \le \liminf_{n\to\infty} \|\xi_n\|_* \le \rho \quad \square$$

The following theorem is a variant of a result of [95]: notice that a perturbation of class C^1 of the functional ϕ corresponds to a C^0 perturbation of its subdifferential and no general existence results are known for differential equations associated to C^0 vector fields in infinite dimensional (even Hilbertian) vector spaces.

Theorem 2.3.7 (Existence for C^1 perturbation of convex functionals). *Let $\phi : \mathscr{B} \to (-\infty, +\infty]$ be a lower semicontinuous functional satisfying the coercivity assumption* (2.1.2b) *and the (strong) compactness assumption* (2.1.3b)*. If $\phi = \phi_1 + \phi_2$ admits the decomposition of Corollary* 1.4.5 *with respect to a convex l.s.c. function ϕ_1 and a C^1 function ϕ_2, then for every $u_0 \in D(\phi)$ $GMM(\Phi; u_0)$ is non empty and its elements u are solutions of the doubly nonlinear differential inclusion*

$$\mathfrak{J}_p(u'(t)) \supset -\partial^\circ\phi(u(t)) \quad t > 0; \quad u(0+) = u_0, \tag{2.3.11}$$

that satisfy the energy identity

$$\int_0^T |u'(t)|^p \, dt + \phi(u(T)) = \phi(u_0) \quad \forall T > 0. \tag{2.3.12}$$

Proof. In this case σ is the *strong* topology of $\mathscr{B}$, as in Remark 2.1.1. We can combine Proposition 2.2.3 ($GMM(\Phi; u_0)$ is non empty), Corollary 1.4.5, Lemma 2.3.6, and Theorem 2.3.3 (every $u \in GMM(\Phi; u_0)$ is a curve of maximal slope for $|\partial\phi|$ and satisfies the energy identity (2.3.9)), Corollary 1.4.5 and Proposition 1.4.1 (curves of maximal slope for $|\partial\phi|$ solve (2.3.11)). $\square$

In fact, condition (2.3.10b) is almost enough to prove the existence of solutions to (2.3.11):

Theorem 2.3.8 (Existence under the closure condition (2.3.10b)**).** *Let $\phi : \mathscr{B} \to (-\infty, +\infty]$ be a functional satisfying all the Assumptions of Section* 2.1*. If ϕ satisfies* (2.3.10b) *and* (2.3.3)*, then it also satisfies* (2.3.10a)*. In particular, for every $u_0 \in D(\phi)$ $GMM(\Phi; u_0)$ is non empty and its elements u are curves of maximal slope which solve the doubly nonlinear differential inclusion* (2.3.11)*.*

The proof of this theorem relies on the implication (2.3.10b) $\Rightarrow$ (2.3.10a) for a functional ϕ which satisfies the assumptions of Section 2.1: we will prove the above implication in the next chapter.

We conclude this section by showing an example where the above theorem can be applied by choosing an auxiliary topology σ weaker than the strong one; the following situation is typical for evolution equations in unbounded domains.

Remark 2.3.9 (An example where σ is weaker than the strong topology of $\mathscr{B}$). Let us consider the Banach space $\mathscr{B} := L^p(\mathbb{R}^d)$, $1 < p < +\infty$, and let $F : \mathbb{R} \to [0, +\infty)$ be a nonnegative C^1 convex function satisfying $F(0) = 0$. We consider the functional

$$\phi(v) := \begin{cases} \int_{\mathbb{R}^d} \Big(\frac{1}{2}|\nabla v(x)|^2 + F(v(x))\Big)\,dx & \text{if } \nabla v \in L^2(\mathbb{R}^d),\ F(v) \in L^1(\mathbb{R}^d), \\ +\infty & \text{otherwise.} \end{cases} \tag{2.3.13}$$

Since F' is increasing, it is not difficult to check that $\xi \in L^q(\mathbb{R}^d)$ belongs to $\partial\phi(v)$ if and only if

$$\Delta v \in L^q(\mathbb{R}^d), \quad F'(v) \in L^q(\mathbb{R}^d), \quad \xi = -\Delta v + F'(v); \tag{2.3.14}$$

the curve u of maximal slope associated to ϕ should be a solution of the Cauchy problem

$$\begin{cases} |\partial_t u|^{p-2}\partial_t u - \Delta u + F'(u) = 0 & \text{in } \mathbb{R}^d \times (0, +\infty), \\ \qquad\qquad u(\cdot, 0+) = u_0(\cdot) & \text{in } \mathbb{R}^d. \end{cases} \tag{2.3.15}$$

Moreover, if v satisfies (2.3.14), it enjoys the a priori estimates

$$\|F'(v)\|_{L^q(\mathbb{R}^d)} \le \|\xi\|_{L^q(\mathbb{R}^d)}, \quad \|\Delta v\|_{L^q(\mathbb{R}^d)} \le 2\|\xi\|_{L^q(\mathbb{R}^d)}.$$

Since ϕ is a convex functional in $L^p(\mathbb{R}^d)$, we know that $|\partial\phi|$ is lower semicontinuous w.r.t. the strong $L^p(\mathbb{R}^d)$ topology, but the sublevel sets

$$\Big\{v \in L^p(\mathbb{R}^d) : \|v\|_{L^p(\mathbb{R}^d)} + \phi(v) \le c\Big\}$$

are not compact in $L^p(\mathbb{R}^d)$.
Let us show that ϕ satisfies (2.3.10b) with respect to the weak L^p-topology σ. If $\xi_n = -\Delta v_n + F'(v_n) \rightharpoonup \xi$ in $L^q(\mathbb{R}^d)$ $\sup_n \phi(v_n) < +\infty$, and $v_n \rightharpoonup v$ in $L^p(\mathbb{R}^d)$, the a priori bounds and Rellich compactness theorem yields

$$\xi = -\Delta v + \eta, \quad F'(v_n) \rightharpoonup \eta \text{ in } L^q(\mathbb{R}^d), \quad v_n \to v \text{ in } L^2_{\rm loc}(\mathbb{R}^d)$$

Up to extracting a further subsequence, we can assume that v_n converges to v $\mathscr{L}^d$-a.e. in $\mathbb{R}^d$ so that $\eta = F'(v)$ and therefore $\xi \in \partial\phi(v)$.

2.4 The (geodesically) convex case

In this section we will consider a notion of convexity along classes of curves in the metric space $\mathscr{S}$: a particular attention is devoted to functionals ϕ which are convex along the geodesics of the metric space $\mathscr{S}$. Let us first introduce the relevant definitions.

Definition 2.4.1 (λ-convexity along curves). *A functional $\phi : \mathscr{S} \to (-\infty, +\infty]$ is called* convex *on a curve $\gamma : t \in [0,1] \mapsto \gamma_t \in \mathscr{S}$ if*

$$\phi(\gamma_t) \le (1-t)\phi(\gamma_0) + t\phi(\gamma_1) \qquad \forall t \in [0,1]. \tag{2.4.1}$$

More generally, we say that ϕ is λ-convex on γ for some $\lambda \in \mathbb{R}$ if

$$\phi(\gamma_t) \le (1-t)\phi(\gamma_0) + t\phi(\gamma_1) - \frac{1}{2}\lambda\, t(1-t)d^2(\gamma_0,\gamma_1) \qquad \forall t \in [0,1]. \tag{2.4.2}$$

Notice that we require that the usual convexity inequality holds with respect to the initial and final point of the curve γ; of course, if $\phi\circ\gamma$ is a real convex function in $[0,1]$ then (2.4.1) surely holds. Among all the possible curves connecting points in $\mathscr{S}$, we are interested to the so called *geodesics*, i.e. to length minimizing curves.

Definition 2.4.2 (Constant speed geodesics). *A curve $\gamma : [0,1] \to \mathscr{S}$ is a (constant speed) geodesic if*

$$d(\gamma_s,\gamma_t) = d(\gamma_0,\gamma_1)(t-s) \quad \forall\, 0 \le s \le t \le 1. \tag{2.4.3}$$

Definition 2.4.3 (λ-geodesically convex functionals). *We say that a functional $\phi : \mathscr{S} \to (-\infty,+\infty]$ is λ-geodesically convex if for any $v_0, v_1 \in D(\phi)$ there exists a constant speed geodesic γ with $\gamma_0 = v_0, \gamma_1 = v_1$ such that ϕ is λ-convex on γ.*

Remark 2.4.4 (Euclidean case). In Euclidean spaces, the largest λ such that ϕ is λ-convex along segments (canonically parametrized on $[0,1]$) is, for smooth functions ϕ, the infimum w.r.t. x of the smallest eigenvalue of $\nabla^2\phi(x)$. In this case ϕ is λ-convex if and only if $v \mapsto \phi(v) - \frac{1}{2}\lambda|v|^2$ is convex. In particular the map $v \mapsto \frac{1}{2}|v-w|^2$ is 1-convex, as the following elementary identity, depending on $t \in \mathbb{R}$, shows

$$|(1-t)v_0 + tv_1 - w|^2 = (1-t)|v_0 - w|^2 + t|v_1 - w|^2 - t(1-t)|v_0 - v_1|^2. \tag{2.4.4}$$

It is not difficult to show that (2.4.4) forces the norm $|\cdot|$ to be induced by a scalar product: in fact, choosing $w = 0$, $t = 1/2$ we see that $|\cdot|$ satisfies the parallelogram rule (12). It is interesting that the same conclusion holds if (2.4.4) is replaced by the corresponding 1-convexity inequality for $t \in [0,1]$.

λ-convexity along geodesics is the easiest assumption of geometric type, which allows for a simple application of the theory presented in the previous Section 2.3, as we shall see in a moment. This property results from the necessity to join two points v_0, $v_1 \in D(\phi)$ by a curve along which both the distance (the curve should be a geodesic) and the functional (λ-convexity) behave nicely. From the "Minimizing Movement" point of view, its importance is clear, since the distance and the functional are the two components of the family of variational functionals $v \mapsto \Phi(\tau, w; v)$ defined by (2.0.3b).

We can easily check that if $\phi : \mathscr{S} \to (-\infty,+\infty]$ is a λ-geodesically convex functional, then for every couple of points v_0, $v_1 \in D(\phi)$ the functional $v \mapsto$

$\Phi(\tau, v_0; v)$ is $(\tau^{-1}+\lambda)$-convex along a suitable geodesic γ connecting v_0 to v_1. In fact in this case we have

$$\begin{aligned}
&\Phi(\tau, v_0;\gamma_t) = \frac{1}{2\tau}d^2(v_0,\gamma_t)+\phi(\gamma_t) = \frac{1}{2\tau}t^2d^2(v_0,v_1)+\phi(\gamma_t)\\
&\le \frac{1}{2\tau}t^2d^2(v_0,v_1)+(1-t)\phi(v_0)+t\phi(v_1)-\frac{1}{2}\lambda t(1-t)d^2(v_0,v_1)\\
&= (1-t)\phi(v_0)+t\Big(\frac{1}{2\tau}d^2(v_0,v_1)+\phi(v_1)\Big)-\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)t(1-t)d^2(v_0,v_1)\\
&= (1-t)\Phi(\tau,v_0;v_0)+t\Phi(\tau,v_0;v_1)-\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)t(1-t)d^2(v_0,v_1). \qquad (2.4.5)
\end{aligned}$$

On the other hand, one can ask if it is possible to formulate a more general assumption directly on Φ: it is interesting that the results we are presenting in this section hold even if there exists an arbitrary curve γ (not necessarily a geodesic) along which Φ is $(\tau^{-1}+\lambda)$-convex, for every $\tau > 0$ such that $\tau^{-1}+\lambda > 0$ (i.e. $0 < \tau < \frac{1}{\lambda^-}$, where $\frac{1}{\lambda^-} = +\infty$ if $\lambda \ge 0$).

Assumption 2.4.5 (Convexity of Φ). *For a given $\lambda \in \mathbb{R}$ we suppose that for any $v_0, v_1 \in D(\phi)$ there exists a curve γ with $\gamma_0 = v_0, \gamma_1 = v_1$ such that*

$$v \mapsto \Phi(\tau, v_0; v) \quad \textit{is } (\tau^{-1}+\lambda)\textit{-convex on } \gamma, \quad \forall\, 0<\tau<\frac{1}{\lambda^-}, \qquad (2.4.6)$$

where $\Phi(\tau, v_0; v) = \frac{1}{2\tau}d^2(v_0, v)+\phi(v)$ is the functional introduced in (2.0.3b).

Any function ϕ satisfying the previous Assumption 2.4.5 for some $\lambda \in \mathbb{R}$ trivially satisfies the same condition for all $\lambda' < \lambda$.
(2.4.6) is equivalent, for every $v_0, v_1 \in D(\phi)$, to the existence of points v_t, $t \in [0,1]$, such that

$$\begin{aligned}
&\frac{1}{2\tau}d^2(v_0,v_t)+\phi(v_t) \qquad (2.4.7a)\\
&\le (1-t)\phi(v_0)+t\phi(v_1)+\frac{1}{2\tau}td^2(v_0,v_1)-\frac{1}{2}\Big(\frac{1}{\tau}+\lambda\Big)t(1-t)d^2(v_0,v_1)\\
&= (1-t)\phi(v_0)+t\phi(v_1)+\frac{t}{2\tau}\Big(t-\lambda\tau(1-t)\Big)d^2(v_0,v_1) \qquad (2.4.7b)
\end{aligned}$$

Neglecting the first term in the left-hand side (2.4.7a) and dividing by t we also get

$$\frac{\phi(v_t)-\phi(v_0)}{t} \le \phi(v_1)-\phi(v_0)+\frac{1}{2\tau}\Big(t-\lambda\tau(1-t)\Big)d^2(v_0,v_1). \qquad (2.4.7c)$$

Remark 2.4.6 (v_t are independent of τ). For the sake of simplicity we supposed that that the points v_t in (2.4.7a) are independent of τ, even if many of the following results still hold in the case when v_t are allowed to depend on τ and ϕ is coercive (2.1.2b). Here we make explicit two useful consequences of the fact that

v_t is independent of τ: first of all, multiplying the inequality (2.4.7a,b) by τ and passing to the limit as $\tau \to 0^+$ we get

$$d(v_0, v_t) \le t d(v_0, v_1). \tag{2.4.8}$$

When $\lambda \ge 0$, we can pass to the limit as $\tau \to +\infty$ in (2.4.7a,b) showing that ϕ is λ-convex along the curve v_t. Thus in this case (2.4.7c) becomes

$$\frac{\phi(v_t) - \phi(v_0)}{t} \le \phi(v_1) - \phi(v_0) - \frac{1}{2}\lambda(1-t)d^2(v_0, v_1). \tag{2.4.9}$$

Remark 2.4.7 (p-modulus of convexity). In the case of p-curves of maximal slope, $p \ne 2$, one should consider a related notion of modulus of convexity depending on the p-power of the distance. Here we do not exploit this variant and we will often assume directly (2.4.7a,b) with $\lambda = 0$.

λ-convexity assumption provides a useful information about the value of $\tau_*(\phi)$ as defined by (2.2.2) and on the existence of a minimum point for ϕ.

Lemma 2.4.8 (Coercivity for convex functionals). *Assume that Assumption* 2.4.5 *holds for some* $\lambda \in \mathbb{R}$ *(with* $\lambda = 0$ *if* $p \ne 2$*). If*

$$\exists\, u_* \in D(\phi),\ r_* > 0 :\ m_* := \inf\Big\{\phi(v) : v \in \mathscr{S},\ d(v, u_*) \le r_*\Big\} > -\infty, \tag{2.4.10}$$

(e.g. if either the coercivity 2.1b *or the lower semicontinuity and compactness* 2.1a,c *assumptions hold), then*

$$\tau_*(\phi) \ge \frac{1}{\lambda^-}, \quad \textit{in particular } \tau_*(\phi) = +\infty \textit{ if } \lambda \ge 0. \tag{2.4.11}$$

If $\lambda > 0$ *then* ϕ *is bounded from below and if it is lower semicontinuous then it has a unique minimum point* $\bar u$*:*

$$\exists!\, \bar u \in \mathscr{S} :\ \phi(\bar u) = \min_{\mathscr{S}} \phi > -\infty. \tag{2.4.12}$$

Proof. Let u_*, r_*, m_* as in (2.4.10), $0 < \tau < \frac{1}{\lambda^-}$. If $v \in D(\phi)$ and $d(v, u_*) > r_*$ we apply the convexity property (2.4.7c) with $v_0 := u_*,\ v_1 := v$ and $t = r^*/d(u_*, v)$ to find $v_* := v_t \in D(\phi)$ satisfying

$$\phi(v) - \frac{\lambda}{2} d^2(v, u_*) \ge \phi(u_*) + c_* d(v, u_*), \quad c_* := \frac{\phi(v_*) - \phi(u_*) - \frac{1}{2}(\tau^{-1} + \lambda) r_*^2}{r^*}$$

By (2.4.8) $dv_*u \le r_*$ and therefore $\phi(v_*) \ge m_*$ by (2.4.10); applying Young inequality we get

$$\phi(v) - \frac{\lambda - \varepsilon}{2} d^2(v, u_*) \ge \phi(u_*) - \frac{1}{2\varepsilon} c_*^2 \quad \forall\, v \in D(\phi),\ \varepsilon > 0. \tag{2.4.13}$$

(2.4.13) shows that $\tau_*(\phi) \geq -\lambda^{-1}$ if $\lambda < 0$, and that ϕ is bounded from below if $\lambda > 0$. If the lower semicontinuity-compactness Assumptions 2.1a,c hold, then the existence of a minimum $\bar{u}$ for ϕ follows directly from (2.4.13). We can also prove the existence (and the uniqueness) of the minimum by a completeness argument, thus assuming the ϕ is simply lower semicontinuous with respect to the distance d and avoiding compactness: just take a minimizing sequence (v_n) with

$$\phi(v_n) \leq \inf_{\mathscr{S}} \phi + \omega_n, \qquad \lim_{n\to\infty} \omega_n = 0,$$

and apply the λ-convexity property of ϕ stated by Remark 2.4.6 along a curve $\gamma_{n,m}$ connecting v_n to v_m. Choosing $t = 1/2$ we obtain

$$\begin{aligned}\frac{\lambda}{8} d^2(v_n, v_m) &\leq \frac{1}{2}\phi(v_m) + \frac{1}{2}\phi(v_n) - \phi(\gamma_{n,m}(1/2)) \\ &\leq \frac{1}{2}\omega_m + \frac{1}{2}\omega_n \to 0 \quad \text{as } n, m \to \infty.\end{aligned}$$

□

The next result, though simple, provides the crucial estimate for λ-convex functions; the reader could compare (2.4.14) with the classification introduced in [95, page 293]: following the notation of that paper, it is not difficult to check that λ-convex functions belongs to the class $\mathcal{K}(\mathscr{S}; 0, 2)$.

Theorem 2.4.9. *If the convexity Assumption* 2.4.5 *holds for some* $\lambda \in \mathbb{R}$, *then the local slope* $|\partial\phi|$ *admits the representation*

$$|\partial\phi|(v) = \sup_{w\neq v} \left(\frac{\phi(v) - \phi(w)}{d(v,w)} + \frac{1}{2}\lambda d(v,w) \right)^+ \qquad \forall\, v \in D(\phi). \tag{2.4.14}$$

In particular, when $\lambda \geq 0$ *the local slope coincides with the global one, i.e.*

$$|\partial\phi|(v) = \sup_{w\neq v} \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)} = \mathfrak{l}_\phi(v) \qquad \forall\, v \in D(\phi). \tag{2.4.15}$$

Proof. First of all we observe that

$$\begin{aligned}|\partial\phi|(v) &= \limsup_{w\to v} \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)} = \limsup_{w\to v} \left(\frac{\phi(v) - \phi(w)}{d(v,w)} + \tfrac{1}{2}\lambda\, d(v,w) \right)^+ \\ &\leq \sup_{w\neq v} \left(\frac{\phi(v) - \phi(w)}{d(v,w)} + \tfrac{1}{2}\lambda d(v,w) \right)^+ .\end{aligned}$$

In order to prove the opposite inequality it is not restrictive to suppose

$$v \in D(\phi), \quad w \neq v \quad \text{with} \quad \phi(v) - \phi(w) + \tfrac{1}{2}\lambda d^2(v,w) > 0; \tag{2.4.16}$$

applying (2.4.7c) with $v_0 = v$ and $v_1 = w$ to get v_t satisfying for every $0 < \tau < \frac{1}{\lambda^-}$

$$\frac{\phi(v) - \phi(v_t)}{d(v, v_t)} \geq \left(\frac{\phi(v) - \phi(w)}{d(v,w)} + \frac{1}{2\tau}\big(\lambda\tau(1-t) - t\big) d(v,w) \right) \frac{t\, d(v,w)}{d(v, v_t)}.$$

Since $d(v, v_t) \le t d(v, w)$, the sign inequality of (2.4.16) yields

$$|\partial\phi|(v) \ge \limsup_{t\downarrow 0} \frac{\phi(v) - \phi(v_t)}{d(v, v_t)} \ge \frac{\phi(v) - \phi(w)}{d(v, w)} + \tfrac{1}{2}\lambda d(v, w) \quad \forall\, w \in \mathscr{S}.$$

Then (2.4.14) follows easily by taking the supremum with respect to w. □

Recalling the Definition (1.2.6) of global slope $\mathfrak{l}_\phi$, from (2.4.14) we easily get

$$|\partial\phi|(v) \le \mathfrak{l}_\phi(v) \le |\partial\phi|(v) + \frac{\lambda^-}{2} \operatorname{diam} \mathscr{S} \quad \forall\, v \in \mathscr{S}. \tag{2.4.17}$$

The following corollary is an immediate consequence of the above upper bound.

Corollary 2.4.10. *Suppose that $\phi : \mathscr{S} \to (-\infty, +\infty]$ satisfies the convexity assumption* 2.4.5 *for some $\lambda \in \mathbb{R}$ and it is d-lower semicontinuous. Then $|\partial\phi|$ is a strong upper gradient for ϕ and it is d-lower semicontinuous.*

Proof. In the case when $\lambda \ge 0$ or $\mathscr{S}$ is bounded, we can simply apply Theorem 1.2.5. In the case $\lambda < 0$ and $\operatorname{diam} \mathscr{S} = +\infty$, recalling that $|\partial\phi|$ is a weak upper gradient, we should check that for any curve $z \in AC(a, b; \mathscr{S})$ with $|\partial\phi|(z)|z'| \in L^1(a, b)$ the function $\phi \circ z$ is absolutely continuous.
It is not restrictive to assume that (a, b) is a bounded interval and the curve z is extended by continuity to $[a, b]$. We simply introduce the compact metric space $\mathscr{S}_0 := z([a, b])$ with the metric induced by $\mathscr{S}$ and we consider the related global slope of ϕ, denoted by $\mathfrak{l}^0_\phi(\cdot)$; (2.4.14) yields

$$\mathfrak{l}^0_\phi(v) = \sup_{w \in \mathscr{S}_0 \setminus \{v\}} \frac{(\phi(v) - \phi(w))^+}{d(v, w)} \le |\partial\phi|(v) - \frac{1}{2}\lambda \operatorname{diam} \mathscr{S}_0 \qquad \forall\, v \in \mathscr{S}_0.$$

In particular $\mathfrak{l}_\phi(z)|z'| \in L^1(a, b)$ and therefore Theorem 1.2.5 yields the desired absolute continuity of $\phi \circ z$.

In order to prove the lower semicontinuity of $|\partial\phi|$ we argue as in the proof of Theorem 1.2.5, where we proved that $\mathfrak{l}_\phi$ is d-lower semicontinuous, and use (2.4.14). □

We can now state two existence results for curves of maximal slope, the first one assuming that there is compactness with respect to the topology induced by d and the second one assuming that there is compactness with respect to σ. Both of them hold even in the case of p-curves (with $\lambda = 0$ for $p \ne 2$) and combine Proposition 2.2.3 and Theorem 2.3.3 following the first strategy of Remark 2.3.4.

Corollary 2.4.11 (Existence of curves of maximal slope I). *Suppose that $\phi : \mathscr{S} \to (-\infty, +\infty]$ satisfies the convexity Assumption* 2.4.5 *for some $\lambda \in \mathbb{R}$ (with $\lambda = 0$ for $p \ne 2$), and the lower semicontinuity-compactness Assumptions* 2.1a,c *for the topology σ* induced by the distance *d, as in* (2.1.3a,b) *of Remark* 2.1.1*. Then every $u_0 \in D(\phi)$ is the initial point of a curve of maximal slope for ϕ with respect to (the strong upper gradient) $|\partial\phi|$ and the conclusions of Theorem* 2.3.3 *hold.*

Proof. By Lemma 2.2.1, ϕ also satisfies the coercivity Assumption 2.1b. Proposition 2.2.3 shows that $GMM(\Phi; u_0)$ is not empty; moreover, the above corollaries yield that $|\partial^- \phi| = |\partial \phi|$ is a strong upper gradient for ϕ, since the compactness assumption of Section 2.1 holds for the topology induced by the distance. Then we can apply Theorem 2.3.3. □

Corollary 2.4.12 (Existence of curves of maximal slope II). *Suppose that the functional $\phi : \mathscr{S} \to (-\infty, +\infty]$ satisfies the convexity Assumption* 2.4.5 *for some $\lambda \in \mathbb{R}$ and the lower semicontinuity-compactness Assumptions* 2.1a,c. *If the map $v \mapsto |\partial\phi|(v)$ is σ-sequentially lower semicontinuous on d-bounded subsets of sublevels of ϕ, then every $u_0 \in D(\phi)$ is the starting point of a curve of maximal slope for ϕ with respect to $|\partial\phi|$ and the conclusions of Theorem* 2.3.3 *hold.*

Proof. Again we observe that Corollary 2.4.10 and our assumption yield that $|\partial^- \phi| = |\partial\phi|$ is a strong upper gradient for ϕ. Invoking Proposition 2.2.3 and Theorem 2.3.3 again we conclude. □

2-curves of maximal slopes of λ-convex functionals with $\lambda > 0$ exhibit exponential convergence to the minimum point of the functional (which exists under the weak condition (2.4.10) of Lemma 2.4.8), with exponential convergence to 0 of the energy. The crucial estimates are stated in the following lemma:

Lemma 2.4.13. *Assume that $\phi : \mathscr{S} \to (-\infty, +\infty]$ is a d-lower semicontinuous functional satisfying the convexity Assumption* 2.4.5 *with $\lambda > 0$. Then*

$$\phi(u) - \inf_{\mathscr{S}} \phi \le \frac{1}{2\lambda} |\partial\phi|^2(u) \quad \forall u \in D(\phi). \tag{2.4.18}$$

Moreover, if $\bar{u} \in D(\phi)$ is the (unique) minimizer for ϕ, then

$$\frac{\lambda}{2} d^2(u, \bar{u}) \le \phi(u) - \phi(\bar{u}) \le \frac{1}{2\lambda} |\partial\phi|^2(u) \quad \forall u \in D(\phi). \tag{2.4.19}$$

Proof. (2.4.18) is an immediate consequence of Young inequality and (2.4.14), which for every $v \in D(\phi)$ with $\phi(v) \le \phi(u)$ yields

$$\phi(u) - \phi(v) \le |\partial\phi|(u)\, d(u, v) - \frac{\lambda}{2} d(u, v)^2 \le \frac{1}{2\lambda} |\partial\phi|^2(u). \tag{2.4.20}$$

On the other hand, if $\bar{u}$ is a minimum for ϕ, we can apply (2.4.9) with $v_0 := \bar{u}$ and $v_1 := u$: since $\phi(v_t) \ge \phi(\bar{u})$ we obtain

$$\frac{\lambda}{2}(1 - t) d^2(u, \bar{u}) \le \phi(u) - \phi(\bar{u});$$

taking the limit as $t \downarrow 0$ we conclude. □

Theorem 2.4.14. *Assume that $\phi : \mathscr{S} \to (-\infty, +\infty]$ is a d-lower semicontinuous functional satisfying the convexity Assumption 2.4.5 with $\lambda > 0$ and having a minimum point $\bar{u}$. Then any curve of maximal slope u w.r.t. to $|\partial\phi|$ satisfies for every $t \geq t_0 > 0$*

$$\tfrac{1}{2}\lambda d^2(u(t), \bar{u}) \leq \phi(u(t)) - \phi(\bar{u}) \leq (\phi(u(t_0)) - \phi(\bar{u}))e^{-2\lambda(t-t_0)}. \tag{2.4.21}$$

Proof. Since the time derivative of (the absolutely continuous map, since $|\partial\phi|$ is a strong upper gradient) $\Delta(t) := \phi(u(t)) - \phi(\bar{u})$ is $-|\partial\phi|^2(u(t))$, we obtain the differential inequality $\Delta'(t) \leq -2\lambda\Delta(t)$, whence the second inequality in (2.4.21) follows; the first one is simply (2.4.19). □

Even in a metric framework, if a curve of maximal slope is a Generalized Minimizing Movement, it exhibits a sort of regularizing effect allowing for a finer description of the differential equation at each point of the interval, if we consider right derivatives. It is interesting to compare the next theorem with Brezis' result [28, Theorem 3.2, page. 57].

Theorem 2.4.15. *Let us suppose that $\phi : \mathscr{S} \to (-\infty, +\infty]$ is d-lower semicontinuous, and it satisfies* (2.4.10) *and the convexity Assumption* 2.4.5 *for some $\lambda \in \mathbb{R}$. If $u_0 \in D(\phi)$ then each element $u \in GMM(u_0; \Phi)$ is locally Lipschitz in $(0, +\infty)$ and satisfies the following properties:*

(i) *The right metric derivative*

$$|u'_+|(t) := \lim_{s \downarrow t} \frac{d(u(s), u(t))}{s - t} \tag{2.4.22}$$

exists and $u(t) \in D(|\partial\phi|)$ for all $t > 0$.

(ii) *The map $t \mapsto e^{-2\lambda^- t}\phi(u(t))$ is convex; the map $t \mapsto e^{\lambda t}|\partial\phi|(u(t))$ is non-increasing, right continuous, and satisfies*

$$\tfrac{T}{2}|\partial\phi|^2(u(T)) \leq e^{2\lambda^- T}\Big(\phi(u_0) - \phi_T(u_0)\Big), \tag{2.4.23}$$

$$T|\partial\phi|^2(u(T)) \leq (1 + 2\lambda^+ T)e^{-2\lambda T}\Big(\phi(u_0) - \inf_{\mathscr{S}} \phi\Big), \tag{2.4.24}$$

where $\phi_T(u_0)$ is the Moreau-Yosida approximation of ϕ defined as in (2.1.2b)

$$\phi_T(u_0) := \inf_{v \in \mathscr{S}} \Phi(T, u_0; v) = \inf_{v \in \mathscr{S}} \frac{1}{2T} d^2(v, u_0) + \phi(v). \tag{2.4.25}$$

(iii) *The equation*

$$\frac{d}{dt_+}\phi(u(t)) = -|\partial\phi|^2(u(t)) = -|u'_+|^2(t) = -|\partial\phi|(u(t))\,|u'_+|(t) \tag{2.4.26}$$

is satisfied at every point of $(0, +\infty)$.

Remark 2.4.16. The statements of the above Theorem hold up to $t = 0$ if $u_0 \in D(|\partial\phi|)$.

Remark 2.4.17. In the case $p \neq 2$ and $\lambda = 0$ the previous Theorem still holds, provided (2.4.23), (2.4.24), and (2.4.26) are properly reformulated:

$$\tfrac{T}{q}|\partial\phi|^q(u(T)) \leq \Big(\phi(u_0) - \phi_T(u_0)\Big), \quad T|\partial\phi|^q(u(T)) \leq \Big(\phi(u_0) - \inf_{\mathscr{S}} \phi\Big), \tag{2.4.27}$$

$$\frac{d}{dt_+}\phi(u(t)) = -|\partial\phi|^q(u(t)) = -|u'_+|^p(t) = -|\partial\phi|(u(t))\,|u'_+|(t). \tag{2.4.28}$$

Chapter 3

Proofs of the Convergence Theorems

We divide the proof of the main convergence theorems in four steps: first of all, we study a single minimization problem of the scheme (2.0.4); stability estimates are then derived for discrete solutions which yield Proposition 2.2.3 by a compactness argument. Finally, convergence is obtained by combining the a priori energy estimates with the gradient properties of the relaxed slope. We will conclude this section with the proof of Theorem 2.4.15.

3.1 Moreau-Yosida approximation

In this section we will study a single minimization problem (2.0.4), which is strictly related to the Moreau-Yosida approximation of the functional ϕ. The convergence of the scheme will be addressed in the next subsections.

Definition 3.1.1 (Moreau-Yosida approximation). *Let us suppose that $\phi : \mathscr{S} \to (-\infty, +\infty]$ is a d-lower semicontinuous and coercive functional; for $\tau > 0$ the* Moreau-Yosida approximation *ϕ_τ of ϕ is defined as*

$$\phi_\tau(u) := \inf_{v \in \mathscr{S}} \Phi(\tau, u; v) = \inf_{v \in \mathscr{S}} \Big\{ \frac{1}{2\tau} d^2(v, u) + \phi(v) \Big\}. \tag{3.1.1}$$

We also set

$$\begin{gathered} J_\tau[u] := \operatorname{argmin} \Phi(\tau, u; \cdot), \ \ i.e. \\ u_\tau \in J_\tau[u] \ \Leftrightarrow \ \Phi(\tau, u; u_\tau) \le \Phi(\tau, u; v) \quad \forall\, v \in \mathscr{S}, \end{gathered} \tag{3.1.2}$$

and, if $J_\tau[u] \neq \emptyset$,

$$\mathtt{d}^+_\tau(u) := \sup_{u_\tau \in J_\tau[u]} d(u_\tau, u), \quad \mathtt{d}^-_\tau(u) := \inf_{u_\tau \in J_\tau[u]} d(u_\tau, u). \tag{3.1.3}$$

For the sake of simplicity, in this section we will often suppose that

$$J_\tau[u] \neq \emptyset \quad \forall\, u \in \mathscr{S},\ 0 < \tau < \tau_*(\phi), \tag{3.1.4}$$

even if many results hold without this assumption. The following properties are well known:

Lemma 3.1.2 (Monotonicity and continuity of $\phi_\tau(u)$). *The map $(\tau, u) \mapsto \phi_\tau(u)$ is continuous in $(0, \tau_*(\phi)) \times \mathscr{S}$.*
If $0 < \tau_0 < \tau_1$ and $u_{\tau_i} \in J_{\tau_i}[u]$ then

$$\begin{aligned} &\phi(u) \geq \phi_{\tau_0}(u) \geq \phi_{\tau_1}(u), \quad d(u_{\tau_0}, u) \leq d(u_{\tau_1}, u),\\ &\phi(u) \geq \phi(u_{\tau_0}) \geq \phi(u_{\tau_1}), \quad \mathtt{d}^+_{\tau_0}(u) \leq \mathtt{d}^-_{\tau_1}(u) \leq \mathtt{d}^+_{\tau_1}(u). \end{aligned} \tag{3.1.5}$$

In particular, if ϕ satisfies (3.1.4), *it holds*

$$\lim_{\tau\downarrow 0} \phi_\tau(u) = \lim_{\tau\downarrow 0} \inf_{u_\tau \in J_\tau[u]} \phi(u_\tau) = \phi(u), \quad \text{if } u \in \overline{D(\phi)} \text{ then } \lim_{\tau\downarrow 0} \mathtt{d}^+_\tau(u) = 0, \tag{3.1.6}$$

and there exists an (at most) countable set $\mathscr{N}_u \subset (0, \tau_*(\phi))$ such that

$$\mathtt{d}^-_\tau(u) = \mathtt{d}^+_\tau(u) \quad \forall\, \tau \in (0, \tau_*(\phi)) \setminus \mathscr{N}_u. \tag{3.1.7}$$

Proof. To prove the continuity of ϕ_τ let us consider sequences $(\tau_n, u_n) \subset (0, \tau_*(\phi)) \times \mathscr{S}$ convergent to (τ, u) and a corresponding sequence $(v_n) \subset D(\phi)$ such that

$$\lim_{n\to\infty} \Big(\Phi(\tau_n, u_n; v_n) - \phi_{\tau_n}(u_n)\Big) = 0.$$

We easily obtain

$$\limsup_{n\to\infty} \Phi(\tau_n, u_n; v_n) = \limsup_{n\to\infty} \phi_{\tau_n}(u_n) \leq \limsup_{n\to\infty} \Phi(\tau_n, u_n; v) = \Phi(\tau, u; v) \quad \forall\, v \in \mathscr{S}. \tag{3.1.8}$$

Taking the infimum w.r.t. v we get $\limsup_n \phi_{\tau_n}(u_n) \leq \phi_\tau(u)$. By (2.2.4) we deduce that (v_n) is a bounded sequence in $\mathscr{S}$; therefore

$$\begin{aligned} \liminf_{n\to\infty} \phi_{\tau_n}(u_n) &= \liminf_{n\to\infty} \Phi(\tau_n, u_n; v_n) \geq \liminf_{n\to\infty} \frac{1}{2\tau_n}\Big(d(v_n, u) - d(u_n, u)\Big)^2 + \phi(v_n)\\ &\geq \liminf_{n\to\infty} \frac{1}{2\tau_n} d^2(v_n, u) - \frac{1}{\tau_n} d(v_n, u) d(u_n, u) + \phi(v_n) \geq \phi_\tau(u), \end{aligned}$$

and this inequality proves the continuity of ϕ_τ.

The first inequality of (3.1.5) follows easily from the analogous monotonicity property of $\tau \mapsto \Phi(\tau, u; v)$ for each $u, v \in \mathscr{S}$. The second one follows from

$$\begin{aligned} \frac{1}{2\tau_0} d^2(u_{\tau_0}, u) + \phi(u_{\tau_0}) &\leq \frac{1}{2\tau_0} d^2(u_{\tau_1}, u) + \phi(u_{\tau_1})\\ &= \big(\frac{1}{2\tau_0} - \frac{1}{2\tau_1}\big) d^2(u_{\tau_1}, u) + \frac{1}{2\tau_1} d^2(u_{\tau_1}, u) + \phi(u_{\tau_1})\\ &\leq \big(\frac{1}{2\tau_0} - \frac{1}{2\tau_1}\big) d^2(u_{\tau_1}, u) + \frac{1}{2\tau_1} d^2(u_{\tau_0}, u) + \phi(u_{\tau_0}) \end{aligned}$$

i.e.

$$\Big(\frac{1}{2\tau_0}-\frac{1}{2\tau_1}\Big)d^2(u_{\tau_0},u)\le\Big(\frac{1}{2\tau_0}-\frac{1}{2\tau_1}\Big)d^2(u_{\tau_1},u).$$

The third inequality follows by combining the preceding one with

$$\frac{1}{2\tau_1}d^2(u_{\tau_1},u)+\phi(u_{\tau_1})\le\frac{1}{2\tau_1}d^2(u_{\tau_0},u)+\phi(u_{\tau_0}).$$

The first limit in (3.1.6) is a simple consequence of the monotonicity property (3.1.5) and of the d-lower semicontinuity of ϕ. In order to prove the last one, observe that

$$d^2(u_\tau,u)\le-2\tau\phi(u_\tau)+d^2(v,u)+2\tau\phi(v)\quad\forall\,v\in D(\phi),\ u_\tau\in J_\tau[u].$$

We take the supremum with respect to $u_\tau\in J_\tau[u]$ and we recall (2.2.3); passing to the limit as $\tau\downarrow 0$ we get

$$\limsup_{\tau\downarrow 0}(\mathtt{d}_\tau^+(u))^2\le d^2(v,u)\quad\forall\,v\in D(\phi).$$

Since $u\in\overline{D(\phi)}$ we conclude. □

The second lemma provides a very useful pointwise estimate on the local slope of ϕ.

Lemma 3.1.3 (Slope estimate). *If $u_\tau\in J_\tau[u]$, then $u_\tau\in D(|\partial\phi|)$ and*

$$|\partial\phi|(u_\tau)\le\frac{d(u_\tau,u)}{\tau}. \tag{3.1.9}$$

In particular $D(|\partial\phi|)$ is d-dense in $D(\phi)$.

Proof. Starting from (3.1.2) we easily get

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

for every $v\in D(\phi)$. Dividing the equation by $d(v,u_\tau)$ we get

$$\limsup_{v\to u_\tau}\frac{\big(\phi(u_\tau)-\phi(v)\big)^+}{d(v,u_\tau)}\le\limsup_{v\to u_\tau}\frac{1}{2\tau}\Big(d(v,u)+d(u_\tau,u)\Big)=\frac{d(u,u_\tau)}{\tau}.$$ □

The next estimate will play a crucial role in the subsequent convergence proofs; we observe that for any open interval (τ_0,τ_1) with $0<\tau_0<\tau_1<\tau_*(\phi)$, (3.1.5) yields (recall (1.1.2))

$$\text{the maps } \tau\mapsto\tau^{-1}\mathtt{d}_\tau^\pm(u) \text{ have finite pointwise variation in } (\tau_0,\tau_1). \tag{3.1.10}$$

Theorem 3.1.4 (Derivative of $\phi_\tau(u)$). *Assume that* (3.1.4) *holds so that for* $\tau \in (0,\tau_*(\phi))$ *the infimum in* (3.1.1) *is attained. For every* $u \in \mathscr{S}$ *the map* $\tau \mapsto \phi_\tau(u)$ *is locally Lipschitz in the open interval* $(0,\tau_*(\phi))$ *and*

$$\frac{d}{d\tau}\phi_\tau(u) = -\frac{\big(\mathtt{d}^\pm_\tau(u)\big)^2}{2\tau^2}, \quad \textit{for every } \tau \in (0,\tau_*(\phi)) \setminus \mathscr{N}_u, \tag{3.1.11}$$

where $\mathscr{N}_u$ *is the (at most countable) set introduced in* (3.1.7).
In particular, if $u_0 \in D(\phi)$, *for every* $\tau \in (0,\tau_*(\phi))$ *we have*

$$\frac{d^2(u_\tau,u)}{2\tau} + \int_0^\tau \frac{\big(\mathtt{d}^\pm_r(u)\big)^2}{2r^2}\,dr = \phi(u) - \phi(u_\tau) \quad \forall\, u_\tau \in J_\tau[u]. \tag{3.1.12}$$

Proof. We simply observe that for every $\tau_0, \tau_1 \in (0,\tau_*(\phi))$ and $u_{\tau_1} \in J_{\tau_1}[u]$

$$\begin{aligned}\phi_{\tau_0}(u) - \phi_{\tau_1}(u) &\le \Phi(\tau_0,u;u_{\tau_1}) - \Phi(\tau_1,u;u_{\tau_1})\\ &= \frac{1}{2\tau_0}d^2(u_{\tau_1},u) - \frac{1}{2\tau_1}d^2(u_{\tau_1},u) = \frac{\tau_1-\tau_0}{2\tau_1\tau_0}d^2(u_{\tau_1},u),\end{aligned} \tag{3.1.13}$$

and, changing sign to each term of the inequality and interchanging τ_0 with τ_1

$$\phi_{\tau_0}(u) - \phi_{\tau_1}(u) \ge \frac{\tau_1-\tau_0}{2\tau_1\tau_0}d^2(u_{\tau_0},u), \tag{3.1.14}$$

so that, being the map $\tau \mapsto \phi_\tau(u)$ is non increasing,

$$0 \le \frac{\big(\mathtt{d}^+_{\tau_0}(u)\big)^2}{2\tau_1\tau_0} \le \frac{\phi_{\tau_0}(u) - \phi_{\tau_1}(u)}{\tau_1-\tau_0} \le \frac{\big(\mathtt{d}^-_{\tau_1}(u)\big)^2}{2\tau_1\tau_0} \quad \text{if } 0<\tau_0<\tau_1. \tag{3.1.15}$$

(3.1.15) shows that $\tau \mapsto \phi_\tau(u)$ is locally Lipschitz in $(0,\tau_*(\phi))$. Passing to the limit as $\tau_1 \downarrow \tau, \tau_0 \uparrow \tau$ we obtain (3.1.11). Integrating (3.1.11) from $0<\tau_0$ to $\tau>\tau_0$ we obtain

$$\phi_\tau(u) + \int_{\tau_0}^\tau \frac{\big(\mathtt{d}^\pm_r(u)\big)^2}{2r}\,dr = \phi_{\tau_0}(u);$$

if $\phi(u_0) < +\infty$, we can pass to the limit in the previous identity as $\tau_0 \downarrow 0$: recalling (3.1.6) we get (3.1.12). □

The next result provides a sort of duality characterization of the local slope (1.2.6) in terms of the Moreau-Yosida approximation of ϕ

Lemma 3.1.5 (Duality formula for the local slope). *We have*

$$\frac{1}{2}|\partial\phi|^2(u) = \limsup_{\tau\to 0}\frac{\phi(u)-\phi_\tau(u)}{\tau}. \tag{3.1.16}$$

Moreover, if the infimum of (3.1.1) *is attained at* u_τ *according to* (3.1.4), *there exists a sequence* $\tau_n \downarrow 0$ *such that*

$$|\partial\phi|^2(u) = \lim_{n\to\infty}\frac{d^2(u_{\tau_n},u)}{\tau_n^2} = \lim_{n\to\infty}\frac{\phi(u)-\phi(u_{\tau_n})}{\tau_n} \ge \liminf_{\tau\downarrow 0}|\partial\phi|^2(u_\tau). \tag{3.1.17}$$

Proof. It is not restrictive to suppose $|\partial\phi|(u) > 0$. We use the elementary identity

$$\frac{1}{2}x^2 = \sup_{y>x} xy - \frac{1}{2}y^2 \quad \forall\, x \in (0, +\infty)$$

obtaining for each $0 < a^{-1} < |\partial\phi|(u)$, $\varepsilon > 0$

$$\begin{aligned}
\frac{1}{2}|\partial\phi|^2(u) &= \limsup_{v\to u} \frac{1}{2}\left(\frac{\big(\phi(u)-\phi(v)\big)^+}{d(u,v)}\right)^2 \\
&= \limsup_{v\to u} \sup_{y>a^{-1}} \left(\frac{\big(\phi(u)-\phi(v)\big)^+}{d(u,v)} y - \frac{1}{2}y^2\right) \\
&= \limsup_{v\to u} \sup_{0<\tau<a\,d(u,v)} \left(\frac{\big(\phi(u)-\phi(v)\big)^+}{d(u,v)} \frac{d(u,v)}{\tau} - \frac{1}{2}\frac{d^2(u,v)}{\tau^2}\right) \\
&\le \sup_{v\in\mathscr{S}\setminus\{u\}} \sup_{0<\tau<\varepsilon} \left(\frac{\big(\phi(u)-\phi(v)\big)^+}{d(u,v)} \frac{d(u,v)}{\tau} - \frac{1}{2}\frac{d^2(u,v)}{\tau^2}\right) \\
&= \sup_{0<\tau<\varepsilon} \sup_{v\in\mathscr{S}\setminus\{u\}} \left(\frac{\big(\phi(u)-\phi(v)\big)^+}{d(u,v)} \frac{d(u,v)}{\tau} - \frac{1}{2}\frac{d^2(u,v)}{\tau^2}\right) \\
&= \sup_{0<\tau<\varepsilon} \frac{\phi(u)-\phi_\tau(u)}{\tau},
\end{aligned}$$

where we used the fact that

$$\begin{aligned}
\sup_{v\in\mathscr{S}\setminus\{u\}} \left(\big(\phi(u)-\phi(v)\big)^+ - \frac{1}{2\tau}d^2(u,v)\right) &= \sup_{v\in\mathscr{S}\setminus\{u\}} \left(\phi(u)-\phi(v) - \frac{1}{2\tau}d^2(u,v)\right) \\
&= \phi(u) - \phi_\tau(u).
\end{aligned}$$

Passing to the limit as $\varepsilon \downarrow 0$ we get

$$\frac{1}{2}|\partial\phi|^2(u) \le \limsup_{\tau\downarrow 0} \frac{\phi(u)-\phi_\tau(u)}{\tau}. \tag{3.1.18}$$

On the other hand, supposing that the infimum in (3.1.1) is attained (otherwise, we argue by approximation), we have

$$\begin{aligned}
\limsup_{\tau\downarrow 0} \frac{\phi(u)-\phi_\tau(u)}{\tau} &= \limsup_{\tau\downarrow 0} \left(\frac{\phi(u)-\phi(u_\tau)}{\tau} - \frac{d^2(u,u_\tau)}{2\tau^2}\right) \\
&\le \limsup_{\tau\downarrow 0} \left(|\partial\phi|(u)\frac{d(u,u_\tau)}{\tau} - \frac{d^2(u,u_\tau)}{2\tau^2}\right) \le \frac{1}{2}|\partial\phi|^2(u),
\end{aligned} \tag{3.1.19}$$

which proves (3.1.16). Combining (3.1.18) and (3.1.19), we get

$$\liminf_{\tau\downarrow 0}\left||\partial\phi|(u)-\frac{d(u,u_\tau)}{\tau}\right|^2\le 0,$$

which yields (3.1.17), via (3.1.9). □

Now we show that convexity Assumption 2.4.5 leads to stronger estimates.

Theorem 3.1.6 (Slope estimates for convex functionals). *Let us suppose that ϕ satisfies the convexity Assumption 2.4.5 for some $\lambda\in\mathbb{R}$.*

(i) *If $u\in D(|\partial\phi|)$, $1+\lambda\tau>0$ and $u_\tau\in J_\tau[u]$, then*

$$\begin{aligned}(1+\lambda\tau)|\partial\phi|^2(u_\tau)\le(1+\lambda\tau)\frac{d^2(u_\tau,u)}{\tau^2}&\le 2\frac{\phi(u)-\phi_\tau(u)}{\tau}\\ &\le\frac{1}{1+\lambda\tau}\,|\partial\phi|^2(u).\end{aligned}\tag{3.1.20}$$

The last inequality holds even though $J_\tau[u]=\emptyset$.

(ii) *If $u\in D(\phi)$, $u_\tau\in J_\tau[u]$, and $\lambda\ge 0$ then*

$$\phi(u_\tau)-\inf_{\mathscr{S}}\phi\le\frac{1}{(1+\lambda\tau)^2}\Big(\phi(u)-\inf_{\mathscr{S}}\phi\Big).\tag{3.1.21}$$

(iii) *If $\lambda\ge 0$ then*

$$\sup_{\tau>0}\frac{\phi(u)-\phi_\tau(u)}{\tau}=\frac12|\partial\phi|^2(u)\quad\forall\,u\in D(\phi).\tag{3.1.22}$$

Proof. (i) The first inequality has already been proved in (3.1.9); in order to prove the second one, we apply (2.4.7b) with $v_0=u$, $v_1=u_\tau$ to find a point γ_t such that

$$\begin{aligned}\frac{1}{2\tau}d^2(u,u_\tau)+\phi(u_\tau)&\le\frac{1}{2\tau}d^2(u,\gamma_t)+\phi(\gamma_t)\\ &\le\frac{t}{2\tau}\Big(t-\lambda\tau\,(1-t)\Big)d^2(u,u_\tau)+(1-t)\phi(u)+t\phi(u_\tau).\end{aligned}$$

Since the right hand quadratic function has a minimum for $t=1$, taking the left derivative we obtain

$$\Big(\frac{\lambda}{2}+\frac{1}{\tau}\Big)d^2(u,u_\tau)+\phi(u_\tau)-\phi(u)\le 0,$$

or, equivalently,

$$\frac12(1+\lambda\tau)\frac{d^2(u,u_\tau)}{\tau^2}\le\frac{\phi(u)-\phi(u_\tau)}{\tau}-\frac{d^2(u,u_\tau)}{2\tau^2}=\frac{\phi(u)-\phi_\tau(u)}{\tau}.$$

The last inequality of (3.1.20) is a simple consequence of Theorem 2.4.9: we write

$$\frac{\phi(u)-\phi_\tau(u)}{\tau}=\frac{\phi(u)-\phi(u_\tau)}{\tau}-\frac{d^2(u_\tau,u)}{2\tau^2}\le|\partial\phi|(u)\frac{d(u_\tau,u)}{\tau}-(1+\lambda\tau)\frac{d^2(u_\tau,u)}{2\tau^2}$$

and we apply the Young inequality, observing that $1-\lambda\tau>0$. If $J_\tau[u]=\emptyset$ we repeat the argument above with a minimizing sequence.
(ii) Starting from (3.1.20) we easily get

$$\frac{\phi(u)-\phi(u_\tau)}{\tau}\ge(1+\tfrac{1}{2}\lambda\tau)|\partial\phi|^2(u_\tau)$$

and, recalling (2.4.18) and (3.1.9), we obtain

$$\big(\phi(u)-\inf_{\mathscr S}\phi\big)-\big(\phi(u_\tau)-\inf_{\mathscr S}\phi\big)=\phi(u)-\phi(u_\tau)\ge 2\lambda\tau(1+\tfrac{1}{2}\lambda\tau)\big(\phi(u_\tau)-\inf_{\mathscr S}\phi\big) \tag{3.1.23}$$

which gives (3.1.21).
(iii) follows by (3.1.20) and (3.1.16).

Remark 3.1.7 (p-estimates). It is easy to check that Lemma 3.1.2, 3.1.3, 3.1.5 still hold in the general p-case, with the estimates

$$|\partial\phi|^q(u_\tau)\le\frac{d^p(u_\tau,u)}{\tau^p},\qquad \frac{1}{q}|\partial\phi|^q(u)=\limsup_{\tau\to0}\frac{\phi(u)-\phi_\tau(u)}{\tau}, \tag{3.1.24}$$

$$|\partial\phi|^q(u)=\lim_{n\to\infty}\frac{d^p(u_{\tau_n},u)}{\tau_n^p}=\lim_{n\to\infty}\frac{\phi(u)-\phi(u_{\tau_n})}{\tau_n}\ge\liminf_{\tau\downarrow0}|\partial\phi|^q(u_\tau). \tag{3.1.25}$$

(3.1.11) becomes

$$\frac{d}{d\tau}\phi_\tau(u)=-\frac{\big(\mathtt{d}_\tau^\pm(u)\big)^p}{q\,\tau^p}\qquad\forall\,\tau\in(0,\tau_*(\phi))\setminus\mathscr N_u, \tag{3.1.26}$$

and therefore (3.1.12) reads

$$\tau\frac{d^p(u_\tau,u)}{p\,\tau^p}+\int_0^\tau\frac{\big(\mathtt{d}_r^\pm(u)\big)^p}{q\,r^p}\,dr=\phi(u)-\phi(u_\tau)\qquad\forall\,\tau\in(0,\tau_*(\phi)). \tag{3.1.27}$$

Finally, for $\lambda=0$ the estimates of Theorem 3.1.6 easily extend to

$$|\partial\phi|^q(u_\tau)\le\frac{d^p(u_\tau,u)}{\tau^p}\le q\frac{\phi(u)-\phi_\tau(u)}{\tau}\le|\partial\phi|^q(u), \tag{3.1.28}$$

$$\sup_{\tau>0}\frac{\phi(u)-\phi_\tau(u)}{\tau}=\frac{1}{q}|\partial\phi|^q(u). \tag{3.1.29}$$

3.2 A priori estimates for the discrete solutions

In order to obtain a sharp energy estimate for discrete solutions giving the differential inequality (1.3.13) as $|\boldsymbol{\tau}| \to 0$, we follow a variational interpolation argument due to De Giorgi, which is based on the identity (3.1.12) in Theorem 3.1.4. As in the previous section, we will assume (3.1.4), so that at least one solution of the variational scheme (2.0.4) exists if $|\boldsymbol{\tau}| < \tau_*(\phi)$.

Definition 3.2.1 (De Giorgi variational interpolation). *Let $\{U^n_{\boldsymbol{\tau}}\}_{n=0}^{+\infty}$ be a solution of the variational scheme* (2.0.4)*; we will denote by $\tilde{U}_{\boldsymbol{\tau}} : [0, +\infty) \to \mathscr{S}$ any interpolation of the discrete values satisfying*

$$\tilde{U}_{\boldsymbol{\tau}}(t) = \tilde{U}_{\boldsymbol{\tau}}(t^{n-1}_{\boldsymbol{\tau}} + \delta) \in J_\delta[U^{n-1}_{\boldsymbol{\tau}}] \quad \textit{if } t = t^{n-1}_{\boldsymbol{\tau}} + \delta \in (t^{n-1}_{\boldsymbol{\tau}}, t^n_{\boldsymbol{\tau}}]. \tag{3.2.1}$$

We also introduce the real valued function $G_{\boldsymbol{\tau}}$ defined by

$$G_{\boldsymbol{\tau}}(t) := \frac{\mathsf{d}^+_\delta(U^{n-1}_{\boldsymbol{\tau}})}{\delta} \geq \frac{d(\tilde{U}_{\boldsymbol{\tau}}(t), U^{n-1}_{\boldsymbol{\tau}})}{t - t^{n-1}_{\boldsymbol{\tau}}} \qquad \text{if } t = t^{n-1}_{\boldsymbol{\tau}} + \delta \in (t^{n-1}_{\boldsymbol{\tau}}, t^n_{\boldsymbol{\tau}}]. \tag{3.2.2}$$

Observe that $G_{\boldsymbol{\tau}}$ is a Borel map thanks to (3.1.10), and (3.1.9) yields

$$|\partial\phi|(\tilde{U}_{\boldsymbol{\tau}}(t)) \leq G_{\boldsymbol{\tau}}(t) \qquad \forall t \in (0, +\infty). \tag{3.2.3}$$

Lemma 3.2.2 (A priori estimates). *Let $|\boldsymbol{\tau}| \in (0, \tau_*)$, let $\{U^n_{\boldsymbol{\tau}}\}_{n=0}^{+\infty}$ be a sequence solving the variational scheme* (2.0.4)*, and let $|U'_{\boldsymbol{\tau}}|$, $G_{\boldsymbol{\tau}}$ be respectively defined by* (2.3.4)*,* (3.2.2)*. Then for each couple of integers $1 \leq i \leq j$ we have*

$$\frac{1}{2}\int_{t^i_{\boldsymbol{\tau}}}^{t^j_{\boldsymbol{\tau}}} |U'_{\boldsymbol{\tau}}|^2(t)\,dt + \frac{1}{2}\int_{t^i_{\boldsymbol{\tau}}}^{t^j_{\boldsymbol{\tau}}} G^2_{\boldsymbol{\tau}}(t)\,dt + \phi(U^j_{\boldsymbol{\tau}}) = \phi(U^i_{\boldsymbol{\tau}}). \tag{3.2.4}$$

Moreover, for any $u_* \in \mathscr{S}$, S, $T > 0$, there exists a constant $C = C(u_*, \tau_*(\phi), S, T)$ such that if

$$\phi(U^0_{\boldsymbol{\tau}}) \leq S, \quad d^2(U^0_{\boldsymbol{\tau}}, u_*) \leq S, \quad t^N_{\boldsymbol{\tau}} \leq T, \quad |\boldsymbol{\tau}| \leq \tau_*(\phi)/8, \tag{3.2.5}$$

we have for $1 \leq n \leq N$

$$d^2(U^n_{\boldsymbol{\tau}}, u_*) \leq C, \quad \sum_{j=1}^n \frac{d^2(U^j_{\boldsymbol{\tau}}, U^{j-1}_{\boldsymbol{\tau}})}{2\tau_j} \leq \phi(U^0_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}) \leq C, \tag{3.2.6}$$

$$d^2(\tilde{U}_{\boldsymbol{\tau}}(t), \overline{U}_\tau(t)) \leq C|\boldsymbol{\tau}| \quad \forall\, t \in [0, T]. \tag{3.2.7}$$

Proof. Starting from (3.1.12) for $u := U^{n-1}_{\boldsymbol{\tau}}$, $u_\tau := U^n_{\boldsymbol{\tau}}$ and observing that for $r \in (0, \tau_n)$ $u_r = \tilde{U}_{\boldsymbol{\tau}}(t^{n-1}_{\boldsymbol{\tau}} + r)$, we get

$$\frac{d^2(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}})}{2\tau_n} + \frac{1}{2}\int_{I^n_{\boldsymbol{\tau}}} \frac{d^2(\tilde{U}_{\boldsymbol{\tau}}(t), U^{n-1}_{\boldsymbol{\tau}})}{(t - t^{n-1}_{\boldsymbol{\tau}})^2}\,dt \leq \phi(U^{n-1}_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}).$$

Recalling the definition of $|U'_{\boldsymbol{\tau}}|$ and $G_{\boldsymbol{\tau}}$ we rewrite the inequality as

$$\frac{1}{2}\int_{I^n_{\boldsymbol{\tau}}} |U'_{\boldsymbol{\tau}}|^2\,dt + \frac{1}{2}\int_{I^n_{\boldsymbol{\tau}}} |G_{\boldsymbol{\tau}}(t)|^2\,dt \le \phi(U^{n-1}_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}).$$

Summing up from $n = i+1$ to $n = j$ we obtain (3.2.4). The same argument, neglecting the (nonnegative) integral terms gives

$$\frac{1}{2}\sum_{n=1}^{N} \tau_n \left(\frac{d(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}})}{\tau_n}\right)^2 \le \phi(U^0_{\boldsymbol{\tau}}) - \phi(U^N_{\boldsymbol{\tau}}). \tag{3.2.8}$$

Now we observe that for every $\varepsilon > 0$ and $\tau_* < \tau_*(\phi)$ we have

$$\begin{aligned}
\frac{1}{2}d^2(U^n_{\boldsymbol{\tau}}, u_*) - \frac{1}{2}d^2(U^0_{\boldsymbol{\tau}}, u_*) &= \sum_{j=1}^{n}\Big(\frac{1}{2}d^2(U^j_{\boldsymbol{\tau}}, u_*) - \frac{1}{2}d^2(U^{j-1}_{\boldsymbol{\tau}}, u_*)\Big)\\
&\le \sum_{j=1}^{n} d(U^j_{\boldsymbol{\tau}}, U^{j-1}_{\boldsymbol{\tau}})d(U^j_{\boldsymbol{\tau}}, u_*)\\
&\le \varepsilon\sum_{j=1}^{n}\frac{d^2(U^j_{\boldsymbol{\tau}}, U^{j-1}_{\boldsymbol{\tau}})}{2\tau_j} + \frac{1}{2\varepsilon}\sum_{j=1}^{n}\tau_j d^2(U^j_{\boldsymbol{\tau}}, u_*)\\
&\le \varepsilon\phi(U^0_{\boldsymbol{\tau}}) - \varepsilon\phi_{\tau_*}(u_*) + \frac{\varepsilon}{2\tau_*}d^2(U^n_{\boldsymbol{\tau}}, u_*) + \frac{1}{2\varepsilon}\sum_{j=1}^{n}\tau_j d^2(U^j_{\boldsymbol{\tau}}, u_*).
\end{aligned}$$

Choosing $\varepsilon := \tau_*/2$ we get

$$\begin{aligned}
d^2(U^n_{\boldsymbol{\tau}}, u_*) &\le 2\Big(d^2(U^0_{\boldsymbol{\tau}}, u_*) + \tau_*\phi(U^0_{\boldsymbol{\tau}}) - \tau_*\phi_{\tau_*}(u_*)\Big) + \frac{4}{\tau_*}\sum_{j=1}^{n}\tau_j d^2(U^j_{\boldsymbol{\tau}}, u_*)\\
&\le 2\big(S + \tau_* S - \tau_*\phi_{\tau_*}(u_*)\big) + \tfrac{4}{\tau_*}\sum_{j=1}^{n}\tau_j d^2(U^j_{\boldsymbol{\tau}}, u_*),
\end{aligned} \tag{3.2.9}$$

where we used the obvious bound

$$\phi(U^n_{\boldsymbol{\tau}}) \ge \phi_{\tau_*}(u_*) - \frac{1}{2\tau_*}d^2(U^n_{\boldsymbol{\tau}}, u_*). \tag{3.2.10}$$

By applying the Gronwall lemma 3.2.4 below with $a_n := d^2(U^n_{\boldsymbol{\tau}}, u_*)$, $A := 2\big(S + \tau_* S - \tau_*\phi_{\tau_*}(u_*)\big)$, and $\alpha := 4/\tau_*$, we get

$$a_n \le Be^{\alpha_{\boldsymbol{\tau}}\, t^{n-1}_{\boldsymbol{\tau}}} \le Be^{\alpha_{\boldsymbol{\tau}}\, T}, \quad \alpha_{\boldsymbol{\tau}} := \frac{\alpha}{1-\alpha|\boldsymbol{\tau}|} = \frac{4}{\tau_* - 4|\boldsymbol{\tau}|}, \quad B := \frac{A}{1-\alpha|\boldsymbol{\tau}|},$$

provided $\alpha|\boldsymbol{\tau}| < 1$. Applying this estimate to (3.2.9) and choosing, e.g. $\tau_* := 3\tau_*(\phi)/4$, we obtain the first inequality of (3.2.6). The second inequality follows then from (3.2.8) and (3.2.10) for $n := N$. Finally, (3.2.7) follows by (3.1.5) and (3.2.6) since for $t \in I^j_{\boldsymbol\tau}$, $j \le n$,

$$d^2(\tilde U_{\boldsymbol\tau}(t), \overline U_{\tau}(t)) \le 2d^2(U^{j-1}_{\boldsymbol\tau}, U^j_{\boldsymbol\tau}) + 2d^2(\tilde U_{\boldsymbol\tau}(t), U^{j-1}_{\boldsymbol\tau})$$
$$\le 4d^2(U^{j-1}_{\boldsymbol\tau}, U^j_{\boldsymbol\tau}) \le 4|\boldsymbol\tau| \sum_{j=1}^n \frac{d^2(U^{j-1}_{\boldsymbol\tau}, U^j_{\boldsymbol\tau})}{\tau_j}.$$

□

Remark 3.2.3 (Easier estimates when ϕ is bounded from below). Observe that when ϕ is bounded from below then (3.2.6) and (3.2.7) become considerably easier, since they are a trivial consequence of (3.2.8):

$$\frac{1}{2}d^2(U^n_{\boldsymbol\tau}, U^0_*) \le \frac{1}{2T}\sum_{j=1}^n \frac{d^2(U^j_{\boldsymbol\tau}, U^{j-1}_{\boldsymbol\tau})}{\tau_j} \le \frac{1}{T}\Big(\phi(U^0_{\boldsymbol\tau}) - \inf_{\mathscr S}\phi\Big), \tag{3.2.11}$$

$$d^2(\tilde U_{\boldsymbol\tau}(t), \overline U_{\tau}(t)) \le 4|\boldsymbol\tau|\big(\phi(U^0_{\boldsymbol\tau}) - \inf_{\mathscr S}\phi\big). \tag{3.2.12}$$

Lemma 3.2.4 (A discrete version of Gronwall Lemma). *Let A, $\alpha \in [0,+\infty)$ and, for $n \ge 1$, let a_n, $\tau_n \in [0,+\infty)$ be satisfying*

$$a_n \le A + \alpha\sum_{j=1}^n \tau_j a_j \quad \forall n \ge 1, \qquad m := \sup_{n\in\mathbb N} \alpha\tau_n < 1. \tag{3.2.13}$$

Then, setting $\beta = \alpha/(1-m)$, $B := A/(1-m)$ and $\tau_0 = 0$, we have

$$a_n \le Be^{\beta\sum\limits_{i=0}^{n-1}\tau_i} \qquad \forall n \ge 1. \tag{3.2.14}$$

Proof. Let $t^j := \sum_{i=1}^j \tau_i$ for $j \ge 1$. First of all, we observe that (3.2.13) gives

$$a_n \le B + \beta\sum_{j=1}^{n-1}\tau_j a_j \quad \forall n \ge 2, \qquad a_1 \le B. \tag{3.2.15}$$

We argue by induction: observe that for $n = 1$ (3.2.14) reduces to $a_1 \le B$. Supposing that (3.2.14) holds for $1 \le n \le k$, and observing that $e^{\beta t^{j-1}} \le e^{\beta t}$ for any $t \in (t^{j-1}, t^j]$, we get

$$\begin{aligned} a_{k+1} &\le B + \beta\sum_{j=1}^k \tau_j a_j \le B + B\beta\sum_{j=1}^k \tau_j e^{\beta t^{j-1}} \le B + B\beta\sum_{j=1}^k \int_{t^{j-1}}^{t^j} e^{\beta t}\,dt \\ &= B + B\beta\int_0^{t^k} e^{\beta t}\,dt = B + B\beta\frac{e^{\beta t^k}-1}{\beta} = Be^{\beta t^k}. \end{aligned}$$

□

Remark 3.2.5. Lemma 3.2.2 still holds for $p \neq 2$ with the obvious variants

$$|\partial\phi|^q(\tilde{U}_{\boldsymbol{\tau}}(t)) \le G^p_{\boldsymbol{\tau}}(t), \tag{3.2.16}$$

$$\frac{1}{p}\int_{t^i_{\boldsymbol{\tau}}}^{t^j_{\boldsymbol{\tau}}} |U'_{\boldsymbol{\tau}}(t)|^p\,dt + \frac{1}{q}\int_{t^i_{\boldsymbol{\tau}}}^{t^j_{\boldsymbol{\tau}}} G^p_{\boldsymbol{\tau}}(t)\,dt + \phi(U^j_{\boldsymbol{\tau}}) \le \phi(U^i_{\boldsymbol{\tau}}), \tag{3.2.17}$$

$$d^p(U^n_{\boldsymbol{\tau}}, u_*) \le C, \quad \sum_{j=1}^n \tau_j \frac{d^p(U^j_{\boldsymbol{\tau}}, U^{j-1}_{\boldsymbol{\tau}})}{p\,\tau_j^p} \le \phi(U^0_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}) \le C, \tag{3.2.18}$$

$$d^p(\tilde{U}_{\boldsymbol{\tau}}(t), \overline{U}_{\tau}(t)) \le C|\boldsymbol{\tau}| \quad \forall\, t \in [0,T]. \tag{3.2.19}$$

3.3 A compactness argument

The following result combines the ideas of Ascoli-Arzelà and Aubin-Lions compactness Theorems: weak compactness (w.r.t. σ) and strong equicontinuity (w.r.t. d) yield pointwise convergence up to subsequences.

Proposition 3.3.1 (A refined version of Ascoli-Arzelà theorem). *Let $T > 0$, let $K \subset \mathscr{S}$ be a sequentially compact set w.r.t. σ, and let $u_n : [0,T] \to \mathscr{S}$ be curves such that*

$$u_n(t) \in K \quad \forall\, n \in \mathbb{N},\ t \in [0,T], \tag{3.3.1}$$

$$\limsup_{n\to\infty} d(u_n(s), u_n(t)) \le \omega(s,t) \quad \forall\, s,\, t \in [0,T], \tag{3.3.2}$$

for a (symmetric) function $\omega : [0,T] \times [0,T] \to [0,+\infty)$, such that

$$\lim_{(s,t)\to(r,r)} \omega(s,t) = 0 \quad \forall\, r \in [0,T] \setminus \mathscr{C}, \tag{3.3.3}$$

where $\mathscr{C}$ is an (at most) countable subset of $[0,T]$. Then there exist an increasing subsequence $k \mapsto n(k)$ and a limit curve $u : [0,T] \to \mathscr{S}$ such that

$$u_{n(k)}(t) \stackrel{\sigma}{\rightharpoonup} u(t) \quad \forall\, t \in [0,T], \quad u \text{ is } d\text{-continuous in } [0,T] \setminus \mathscr{C}. \tag{3.3.4}$$

Remark 3.3.2 (The case when ω is induced by a finite measure). An important case where the previous theorem can be applied is provided by a (symmetric) function ω of the form

$$\omega(s,t) = \mu([s,t]) \quad \forall\, 0 \le s \le t \le T, \tag{3.3.5}$$

where μ is a non negative and finite measure on $[0,T]$; in this case $\mathscr{C}$ is set of the atoms of μ.

Proof. Being K sequentially compact, by a standard diagonal argument we can find a subsequence $k \mapsto n(k)$ and a function $u : (\mathbb{Q} \cap [0,T]) \cup \mathscr{C} \to K$ such that

$$u_{n(k)}(t) \stackrel{\sigma}{\rightharpoonup} u(t), \quad d(u(s), u(t)) \le \omega(s,t) \quad \forall\, s,t \in \big(\mathbb{Q} \cap [0,T]\big) \cup \mathscr{C}; \tag{3.3.6}$$

the distance inequality in (3.3.6) follows from (2.1.1). Since K is d-complete, thanks to (3.3.6) we can extend u to $[0,T]$ by continuity. Therefore we can uniquely define a curve (still denoted by) $u : [0,T] \to \mathscr{S}$ which is continuous at each point of $[0,T] \setminus \mathscr{C}$.

In order to prove that $u_{n(k)}(t) \stackrel{\sigma}{\rightharpoonup} u(t)$ also for all $t \in [0,T] \setminus \mathscr{C}$ it is sufficient to show that every converging subsequence of $u_{n(k)}(t)$ converges to $u(t)$: if $u_{n(k)'}(t) \stackrel{\sigma}{\rightharpoonup} v$ is such a subsequence, we have

$$d(u(s), v) \le \liminf_{k\to\infty} d(u_{n(k)'}(s), u_{n(k)'}(t)) \le \omega(s,t) \quad \forall\, s \in [0,T] \cap \mathbb{Q}.$$

If $t \notin \mathscr{C}$ we can let $s \uparrow t$ to obtain that $v = u(t)$. □

An immediate application of Proposition 3.3.1 and Remark 3.3.2 is a classical result, due to Helly, on compactness of monotone functions w.r.t. pointwise convergence.

Lemma 3.3.3 (Helly). *Suppose that (φ_n) are non increasing functions defined in $[0,T]$ with values in $[-\infty,+\infty]$. Then there exist a subsequence $k \mapsto n(k)$ and a non increasing map $\varphi : [0,T] \to [-\infty,+\infty]$ such that $\lim_{k\to\infty} \varphi_{n(k)}(t) = \varphi(t)$ for any $t \in [0,T]$.*

Proof. It is not restrictive to assume, up to a left composition, that all functions φ_n have their values in $[0,1]$. Denoting by μ_n the derivatives in the sense of distributions of φ_n, it suffices to extract a subsequence such that $\mu_{n(k)}$ narrowly converge in $[0,T]$ to a finite and non negative measure μ in $[0,T]$. Then the assumptions of Remark 3.3.2 are fulfilled because

$$\limsup_{k\to\infty} |\varphi_{n(k)}(s) - \varphi_{n(k)}(t)| \le \limsup_{k\to\infty} \mu_{n(k)}([s,t]) \le \mu([s,t])$$

whenever $0 \le s \le t \le T$. □

Corollary 3.3.4. *Let us fix $p \in (1,+\infty)$ and let be given a family Λ of admissible partitions of $(0,+\infty)$ with*

$$\inf_{\boldsymbol{\tau}\in\Lambda} |\boldsymbol{\tau}| = 0, \quad \sup_{\boldsymbol{\tau}\in\Lambda} |\boldsymbol{\tau}| < \tau_*(\phi),$$

and a corresponding family of initial data $\{U^0_{\boldsymbol{\tau}}\}_{\boldsymbol{\tau}\in\Lambda}$ satisfying

$$U^0_{\boldsymbol{\tau}} \stackrel{\sigma}{\rightharpoonup} u_0, \quad \phi(U^0_{\boldsymbol{\tau}}) \to \phi(u_0) \ \textit{as } |\boldsymbol{\tau}| \downarrow 0, \quad \sup_{\boldsymbol{\tau}\in\Lambda} d^p(U^0_{\boldsymbol{\tau}}, u_0) < +\infty. \tag{3.3.7}$$

Then there exist a sequence $(\boldsymbol{\tau}_n) \subset \Lambda$ with $\lim_n |\boldsymbol{\tau}_n| = 0$, a limit curve u which belongs to $AC^p_{\rm loc}([0,+\infty);\mathscr{S})$, a non-increasing function $\varphi : [0,+\infty) \to \mathbb{R}$, and a function $A \in L^p_{\rm loc}([0,+\infty))$ such that

$$\overline{U}_{\boldsymbol{\tau}_n}(t) \stackrel{\sigma}{\rightharpoonup} u(t), \quad \tilde{U}_{\boldsymbol{\tau}_n}(t) \stackrel{\sigma}{\rightharpoonup} u(t) \quad \textit{as } n \to \infty \quad \forall\, t \ge 0, \tag{3.3.8}$$

$$\varphi(t) := \lim_{n\to\infty} \phi(\overline{U}_{\boldsymbol{\tau}_n}(t)) \geq \phi(u(t)) \quad \forall\, t \geq 0, \quad \phi(u(0)) = \phi(u_0), \tag{3.3.9}$$

$$|U'_{\boldsymbol{\tau}_n}| \rightharpoonup A \textit{ in } L^p_{\rm loc}([0,+\infty)), \quad A(t) \geq |u'|(t) \textit{ for } \mathscr{L}^1\textit{-a.e. } t \in (0,+\infty), \tag{3.3.10}$$

$$\liminf_{n\to\infty} G_{\tau_n}(t) \geq |\partial^-\phi|(u(t)) \quad \forall\, t \geq 0. \tag{3.3.11}$$

Proof. As usual we limit to consider the case $p = 2$, the modifications to deal with the general case being obvious. Condition (3.3.7) ensures that for any $T > 0$ the constant C defined by Lemma 3.2.2 remains uniformly bounded with respect to $\boldsymbol{\tau} \in \Lambda$. Therefore the estimate (3.2.6) and the Assumptions of section 2.1 show that the curves $\overline{U}_{\boldsymbol{\tau}} : [0,T] \to \mathscr{S}$, $\boldsymbol{\tau} \in \Lambda$, take their values in a σ-sequentially compact set. We can find a sequence $(\boldsymbol{\tau}_n) \subset \Lambda$ with $|\boldsymbol{\tau}_n| \downarrow 0$ such that $|U'_{\boldsymbol{\tau}_n}|$ weakly converge in $L^2(0,+\infty)$ to some function A and, by Lemma 3.3.3, the limit in (3.3.9) exists.

For fixed $0 \leq s < t$ let us define

$$s(n) := \max\big\{r \in \mathcal{P}_{\boldsymbol{\tau}_n} :\ r < s\big\}, \qquad t(n) := \min\big\{r \in \mathcal{P}_{\boldsymbol{\tau}_n} :\ t < r\big\},$$

so that

$$s(n) \leq s \leq t \leq t(n), \quad \lim_{n\to\infty} s(n) = s, \quad \lim_{n\to\infty} t(n) = t.$$

We have

$$d(\overline{U}_{\boldsymbol{\tau}_n}(s), \overline{U}_{\boldsymbol{\tau}_n}(t)) \leq \int_{s(n)}^{t(n)} |U'_{\boldsymbol{\tau}_n}|(r)\, dr \tag{3.3.12}$$

and therefore

$$\limsup_{n\to\infty} d(\overline{U}_{\boldsymbol{\tau}_n}(s), \overline{U}_{\boldsymbol{\tau}_n}(t)) \leq \int_s^t A(r)\, dr. \tag{3.3.13}$$

Applying Proposition 3.3.1 and (3.2.7), possibly extracting one more subsequence, we can find $u \in AC^2_{\rm loc}([0,+\infty);\mathscr{S})$ such that (3.3.8) and (3.3.9) hold true. Moreover, the limit inequality $d(u(s),u(t)) \leq \int_s^t A(r)\, dr$ immediately gives that $|u'| \leq A$ $\mathscr{L}^1$-a.e. in $(0,+\infty)$. Finally, (3.3.11) follows from Fatou's Lemma and the estimate (3.2.3), which yields

$$|\partial^-\phi|(u(t)) \leq \liminf_{n\to\infty} |\partial\phi|(\tilde{U}_{\boldsymbol{\tau}_n}(t)) \leq \liminf_{n\to\infty} G_{\boldsymbol{\tau}_n}(t).$$

3.4 Conclusion of the proofs of the convergence theorems

Proof of Theorem 2.3.3. Combining (3.3.10), (3.3.11), (3.3.9) and using eventually (3.2.4) we easily get

$$\begin{aligned}
&\frac{1}{2}\int_0^t |u'|^2(s)\,ds + \frac{1}{2}\int_0^t |\partial^-\phi|^2(u(s))\,ds + \phi(u(t)) \\
\le\; &\frac{1}{2}\int_0^t A^2(s)\,ds + \frac{1}{2}\int_0^t \liminf_{n\to\infty} G^2_{\boldsymbol{\tau}_n}(s)\,ds + \lim_{n\to\infty}\phi(\overline{U}_{\boldsymbol{\tau}_n}(t)) \\
\le\; &\liminf_{n\to\infty} \frac{1}{2}\int_0^t \left|U'_{\boldsymbol{\tau}_n}(s)\right|^2 ds + \frac{1}{2}\int_0^t G^2_{\boldsymbol{\tau}_n}(s)\,ds + \phi(\overline{U}_{\boldsymbol{\tau}_n}(t)) \le \phi(u_0).
\end{aligned} \tag{3.4.1}$$

On the other hand, since $|\partial^-\phi|$ is a strong upper gradient for ϕ we have

$$\phi(u_0) \le \phi(u(t)) + \int_0^t |\partial^-\phi|(u(s))|u'|(s)\,ds, \tag{3.4.2}$$

and therefore

$$\begin{aligned}
|u'|(t) &= |\partial^-\phi|(u(t)) \quad \text{for } \mathscr{L}^1\text{-a.e. } t\in(0,+\infty),\\
\phi(u_0)-\phi(u(t)) &= \int_0^t |\partial^-\phi|(u(s))|u'|(s)\,ds.
\end{aligned} \tag{3.4.3}$$

It follows that $t\mapsto\phi(u(t))$ is locally absolutely continuous and

$$\frac{d}{dt}\phi(u(t)) = -|\partial^-\phi|(u(t))|u'|(t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t\in(0,+\infty). \qquad \square$$

Proof of Theorem 2.3.1. Observe that Corollary 3.3.4 still holds under the assumptions of Theorem 2.3.1. Denoting by $\varphi(t)$ the limit in (3.3.9) and arguing as in (3.4.1) we get

$$\frac{1}{2}\int_s^t |u'|^2(r)\,dr + \frac{1}{2}\int_s^t |\partial^-\phi|^2(u(r))\,dr \le \varphi(s)-\varphi(t) \tag{3.4.4}$$

for $0\le s\le t$, and

$$\int_0^t \liminf_{n\to\infty} G^2_{\boldsymbol{\tau}_n}(r)\,dr < +\infty \quad \forall\, t>0. \tag{3.4.5}$$

Therefore

$$-\varphi'(t) \ge \frac{1}{2}|u'|^2(t) + \frac{1}{2}|\partial^-\phi|^2(u(t)) \quad \text{for } \mathscr{L}^1\text{-a.e. } t\in(0,+\infty). \tag{3.4.6}$$

Moreover, (3.2.3) and (3.4.5) yield

$$\liminf_{n\to\infty} |\partial\phi|(\tilde{U}_{\boldsymbol{\tau}_n}(t)) \le \liminf_{n\to\infty} G_{\boldsymbol{\tau}_n}(t) < +\infty \quad \text{for } \mathscr{L}^1\text{-a.e. } t\in(0,+\infty),$$

so that (3.3.8) and the continuity assumption (2.3.3) give

$$\varphi(t) = \phi(u(t)) \quad \text{for } \mathscr{L}^1\text{-a.e. } t\in(0,+\infty). \tag{3.4.7}$$

We can conclude that u is a curve of maximal slope for ϕ with respect to is (weak) upper gradient $|\partial^-\phi|$. $\square$

Proof of Theorem 2.4.15. We start with a simple consequence of (3.1.20) and (3.1.21).

Lemma 3.4.1. *Let us suppose that* ϕ *satisfies Assumption* 2.4.5 *with* $\lambda|\boldsymbol{\tau}| > -1$, *and let us set*

$$\lambda_{\boldsymbol{\tau}} := \frac{\log(1+\lambda|\boldsymbol{\tau}|)}{|\boldsymbol{\tau}|} = \inf_{n>0} \frac{\log(1+\lambda\tau_n)}{\tau_n} \le \lambda. \tag{3.4.8}$$

The sequences

$$n \mapsto e^{\lambda_{\boldsymbol{\tau}} t^n_{\boldsymbol{\tau}}} |\partial\phi|(U^n_{\boldsymbol{\tau}}), \quad n \mapsto e^{2\lambda^+_{\boldsymbol{\tau}} t^n_{\boldsymbol{\tau}}} \Big(\phi(U^n_{\boldsymbol{\tau}}) - \inf_{\mathscr{S}} \phi\Big) \quad \textit{are non increasing.} \tag{3.4.9}$$

Proof. From (3.1.20) we get for every $n > 1$

$$\begin{aligned} e^{\lambda_{\boldsymbol{\tau}} t^n_{\boldsymbol{\tau}}} |\partial\phi|(U^n_{\boldsymbol{\tau}}) &\le e^{\lambda_{\boldsymbol{\tau}} t^n_{\boldsymbol{\tau}}} \frac{1}{1+\lambda\tau_n} |\partial\phi|(U^{n-1}_{\boldsymbol{\tau}}) \le e^{\lambda_{\boldsymbol{\tau}} t^n_{\boldsymbol{\tau}}} e^{-\lambda_{\boldsymbol{\tau}}\tau_n} |\partial\phi|(U^{n-1}_{\boldsymbol{\tau}}) \\ &= e^{\lambda_{\boldsymbol{\tau}} t^{n-1}_{\boldsymbol{\tau}}} |\partial\phi|(U^{n-1}_{\boldsymbol{\tau}}), \end{aligned}$$

where we used the inequality

$$\frac{1}{1+\lambda\tau_n} \le e^{-\lambda_{\boldsymbol{\tau}}\tau_n}, \tag{3.4.10}$$

which follows directly from Definition (3.4.8) (whose last inequality is a consequence of the concavity of the map $t \mapsto \log(1+\lambda t)$). The second property of (3.4.9) follows by an analogous argument and from (3.1.21). □

Setting

$$\bar G_{\boldsymbol{\tau}}(t) := e^{\lambda_{\boldsymbol{\tau}} t^n_{\boldsymbol{\tau}}} |\partial\phi|(U^n_{\boldsymbol{\tau}}), \quad \text{if } t \in (t^{n-1}_{\boldsymbol{\tau}}, t^n_{\boldsymbol{\tau}}],$$

$\bar G_{\boldsymbol{\tau}}$ is a non-increasing function: taking into account the $L^2_{\rm loc}$ convergence of slopes given by (2.3.7) and Helly's Theorem 3.3.3, we can suppose, up to extracting a suitable subsequence, that for a non-increasing function $\bar G$ we have

$$\lim_{k\to\infty} \bar G_{\boldsymbol{\tau}_k}(t) = \bar G(t) \quad \forall t \ge 0, \qquad \bar G(t) = e^{\lambda t} |\partial\phi|(u(t)) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,+\infty).$$

On the other hand, we know that the map $t \mapsto e^{\lambda t}|\partial\phi|(u(t))$ is lower semicontinuous and therefore it coincides with the right continuous representative of $\bar G$. Since $|u'|(t) = |\partial\phi|(u(t))$ for $\mathscr{L}^1$-a.e. $t \in (0,+\infty)$, we deduce that $|u'|$ admits a right continuous representative and it is essentially bounded in each interval $(\delta, 1/\delta)$ for $0 < \delta < 1$. From the inequality

$$d(u(t+h), u(t)) \le \int_t^{t+h} |\partial\phi|(u(r))\, dr \qquad \forall\, t \ge 0,$$

we get

$$\limsup_{h\downarrow 0} \frac{d(u(t+h), u(t))}{h} \le |\partial\phi|(u(t)) \qquad \forall\, t \ge 0. \tag{3.4.11}$$

Since for any μ, $\bar\phi \in \mathbb{R}$ and $\mathscr{L}^1$-a.e. $t > 0$

$$\frac{d}{dt} e^{2\mu t}\big(\phi(u(t)) - \bar\phi\big) = -e^{2\mu t}|\partial\phi|^2(u(t)) + 2\mu e^{2\mu t}\big(\phi(u(t)) - \bar\phi\big), \tag{3.4.12}$$

choosing $\mu := \min(\lambda, 0)$ and $\bar\phi \le \inf_{t\in[0,T]} \phi(u(t))$, we find that the map $t \mapsto e^{2\mu t}\big(\phi(u(t)) - \bar\phi\big)$ is convex in $[0,T]$. It follows that the map $t \mapsto \phi(u(t))$ is right differentiable at each point $t \in [0,+\infty)$ and

$$\frac{d}{dt_+}\phi(u(t)) = -|\partial\phi|^2(u(t)) \quad \forall\, t \ge 0\,. \tag{3.4.13}$$

On the other hand, it is easy to check that

$$\frac{d}{dt_+}\phi(u(t)) \ge -|\partial\phi|(u(t)) \cdot \liminf_{h\downarrow 0} \frac{d(u(t+h), u(t))}{h}. \tag{3.4.14}$$

Combining (3.4.14) with (3.4.11) and (3.4.13) we get

$$|u'_+|(t) = \lim_{h\downarrow 0} \frac{d(u(t+h), u(t))}{h} = |\partial\phi|(u(t)) \quad \text{at each point } t \in (0,+\infty).$$

Inequality (2.4.23) follows by the energy inequality, since for $\lambda \le 0$

$$\begin{aligned}\tfrac{T}{2} e^{2\lambda T}|\partial\phi|^2(u(T)) &\le \tfrac12 \int_0^T e^{2\lambda t}|\partial\phi|^2(u(t))\,dt \le \phi(u_0) - \phi(u(T)) - \tfrac12 \int_0^T |u'|^2(t)\,dt \\ &\le \phi(u_0) - \phi(u(T)) - \tfrac{1}{2T} d^2(u_0, u(T)) \le \phi(u_0) - \phi_T(u_0).\end{aligned}$$

Finally (2.4.24) for $\lambda \ge 0$ follows by an integration of (3.4.12), choosing $\mu := \lambda$, $\bar\phi := \inf_{\mathscr{S}} \phi$, and taking into account (2.4.21):

$$\begin{aligned}e^{2\lambda T}\big(\phi(u(T)) - \bar\phi\big) + Te^{2\lambda T}|\partial\phi|^2(u(T)) &\le \phi(u_0) - \bar\phi + 2\lambda \int_0^T e^{2\lambda t}\big(\phi(u(t)) - \bar\phi\big)\,dt \\ &\le (1 + 2\lambda T)(\phi(u_0) - \bar\phi).\end{aligned}$$

□

We conclude this section with a discrete analogous to Theorem 2.4.14, whose proof follows directly from (3.4.9) and Lemma 2.4.13.

Corollary 3.4.2. *Assume that $\phi : \mathscr{S} \to (-\infty, +\infty]$ is a d-lower semicontinuous functional satisfying Assumption 2.4.5 with $\lambda > 0$ and let $\overline{U}_{\boldsymbol\tau}$ be a discrete solution. Then*

$$\tfrac12 \lambda d^2(U^n_{\boldsymbol\tau}, \bar u) \le \phi(U^n_{\boldsymbol\tau}) - \phi(\bar u) \le (\phi(U^0_{\boldsymbol\tau}) - \phi(\bar u))e^{-2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}. \tag{3.4.15}$$

Chapter 4

Uniqueness, Generation of Contraction Semigroups, Error Estimates

In all this section we consider the "quadratic" approximation scheme (2.0.3b), (2.0.4) for 2-curves of maximal slope and we identify the "weak" topology σ with the "strong" one induced by the distance d as in Remark 2.1.1: thus we are assuming that

$$\begin{gathered} p=2, \quad (\mathscr{S},d) \text{ is a complete metric space and} \\ \phi:\mathscr{S}\to(-\infty,+\infty] \text{ is a proper, coercive (2.4.10), l.s.c. functional,} \end{gathered} \tag{4.0.1}$$

but *we are not imposing any compactness assumptions on the sublevels of* ϕ. Existence, uniqueness and semigroup properties for minimizing movement $u \in MM(\Phi; u_0)$ (and not simply the generalized ones, recall Definition 2.0.6) are well known in the case of lower semicontinuous *convex functionals in Hilbert spaces* [28]. In this framework the resolvent operator in $J_\tau[\cdot]$ (3.1.2) is single valued and *non expansive*, i.e.

$$d(J_\tau[u],J_\tau[v]) \le d(u,v) \quad \forall\, u,v\in\mathscr{S},\ \tau>0; \tag{4.0.2}$$

this property is a key ingredient, as in the celebrated Crandall-Ligget generation Theorem [46], to prove the uniform convergence of the exponential formula (cf. (2.0.9))

$$u(t)=\lim_{n\to\infty}(J_{t/n})^n[u_0], \qquad d\Big(u(t),(J_{t/n})^n[u_0]\Big)\le\frac{2|\partial\phi|(u_0)\,t}{\sqrt{n}}, \tag{4.0.3}$$

and therefore to define a contraction semigroup on $\overline{D(\phi)}$. Being generated by a convex functional, this semigroup exhibits a nice regularizing effect [27], since

$u(t) \in D(|\partial\phi|)$ whenever $t > 0$ even if the starting value u_0 simply belongs to $\overline{D(\phi)}$. Moreover the function u can be characterized as the unique solution of the *evolution variational inequality*

$$\langle \tfrac{d}{dt} u(t), u(t) - v\rangle + \phi(u(t)) \le \phi(v) \quad \forall\, v \in D(\phi), \tag{4.0.4}$$

$\langle \cdot, \cdot \rangle$ being the scalar product in $\mathscr{S}$.

More recently, optimal *a priori* and *a posteriori* error estimates have also been derived [18, 115, 102]: the original $O(\tau^{1/2}) = O(1/\sqrt{n})$ order of convergence established by Crandall and Ligget for $u_0 \in D(|\partial\phi|)$ and a uniform partition (2.0.8), has been improved to

$$d\Big(u(t), (J_{t/n})^n[u_0]\Big) \le \frac{|\partial\phi|(u_0)\, t}{n\sqrt{2}} \tag{4.0.5}$$

and extended to the general scheme (2.0.4), (2.0.7)

$$d^2(\overline{U}_{\boldsymbol{\tau}}(t), u(t)) \le |\boldsymbol{\tau}| \Big(\phi(u_0) - \inf_{\mathscr{S}} \phi\Big), \quad d^2(\overline{U}_{\boldsymbol{\tau}}(t), u(t)) \le |\boldsymbol{\tau}|^2\, \frac{|\partial\phi|^2(u_0)}{2}, \tag{4.0.6}$$

thus establishing an optimal error estimate of the same order $O(|\boldsymbol{\tau}|)$ of the Euler method in a smooth and finite dimensional setting.

Similar results for gradient flows of convex functionals in general (non Hilbertian) Banach spaces are still completely open: at least heuristically, this fact suggests that some structural property of the distance should play a crucial role, besides the convexity of the functional ϕ.

A first step in this direction has been obtained by U. Mayer [96] (see also [85]), who considered gradient flows of geodesically convex functionals on *non-positively curved metric spaces*: these are length spaces (i.e. each couple of points v_0, v_1 can be connected through a minimal geodesic) where the distance maps

$$v \mapsto \tfrac{1}{2} d^2(v, w) \text{ are 1-convex along geodesics} \quad \forall\, w \in \mathscr{S}. \tag{4.0.7}$$

This property was introduced by Aleksandrov on the basis of the analogous inequality satisfied in Euclidean spaces (2.4.4) and in Riemannian manifolds of non-positive sectional curvature [84, §2.3]; it allows to prove (4.0.2), and to obtain the generation formula (4.0.3) by following the same Crandall-Liggett arguments. Observe that Mayer's assumptions yield in particular that the variational functional defined by (2.0.3b)

$$\begin{aligned} &v \mapsto \Phi(\tau, w; v) = \frac{1}{2\tau} d^2(v, w) + \phi(v) \\ &\text{is } (\tau^{-1} + \lambda)\text{-}\textit{convex along geodesics} \quad \forall\, w \in \mathscr{S}. \end{aligned} \tag{4.0.8}$$

These assumptions, though quite general, do not cover the case of the metric space of probability measures endowed with the L^2-Wasserstein distance: we will show in Section 7.3 that, in fact, the distance of this space satisfies the opposite inequality, thus providing a positively curved space, as formally suggested also by [107]. Example 7.3.3 will also show that the squared L^2-Wasserstein distance does not satisfy any λ-convexity properties, even for negative choice of $\lambda \in \mathbb{R}$.

Our idea is to concentrate our attention directly on the functional $\Phi(\tau, w; \cdot)$ and to allow more flexibility in the choice of the connecting curves, along which it has to satisfies the convexity assumption (4.0.8): we formalize this requirement in the following assumption:

Assumption 4.0.1 **($(\tau^{-1} + \lambda)$-convexity of $\Phi(\tau, u; \cdot)$).** *We suppose that for every choice of w, v_0, and v_1 in $D(\phi)$ there exists a curve $\gamma = \gamma_t$, $t \in [0, 1]$, with $\gamma_0 = v_0, \gamma_1 = v_1$ such that*

$$v \mapsto \Phi(\tau, w; v) \text{ is } \Big(\frac{1}{\tau} + \lambda\Big)\text{-convex on } \gamma \text{ for each } 0 < \tau < \frac{1}{\lambda^-}, \tag{4.0.9}$$

i.e. the map $\Phi(\tau, w; \gamma_t)$ satisfies the inequality

$$\Phi(\tau, w; \gamma_t) \le (1-t)\Phi(\tau, w; v_0) + t\Phi(\tau, w; v_1) - \frac{1+\lambda\tau}{2\tau} t(1-t) d^2(v_0, v_1). \tag{4.0.10}$$

Remark 4.0.2. Of course, Assumption 4.0.1 covers the case of a (geodesically) λ-convex functional on a nonpositively curved metric space considered by [96], in particular the case of a (geodesically) λ-convex functional in a Riemannian manifold of nonpositive sectional curvature or in a Hilbert space.

Remark 4.0.3. Assumption 4.0.1 is *stronger* than 2.4.5, since this last one is a particular case of (4.0.1) when the "base point" w coincides with v_0.

We collect the main results in this case

Theorem 4.0.4 (Generation and main properties of the evolution semigroup). *Let us assume that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold for some $\lambda \in \mathbb{R}$.*

i) Convergence and exponential formula: *for each $u_0 \in \overline{D(\phi)}$ there exists a unique element $u = S[u_0]$ in $MM(\Phi; u_0)$ which therefore can be expressed through the exponential formula*

$$u(t) = S[u_0](t) = \lim_{n\to\infty} (J_{t/n})^n [u_0]. \tag{4.0.11}$$

ii) Regularizing effect: *u is a locally Lipschitz curve of maximal slope with $u(t) \in D(|\partial\phi|) \subset D(\phi)$ for $t > 0$; in particular, if $\lambda \ge 0$, the following a priori bounds hold:*

$$\begin{aligned} \phi(u(t)) \le \phi_t(u_0) \le \phi(v) + \frac{1}{2t} d^2(v, u_0) \quad & \forall\, v \in D(\phi), \\ |\partial\phi|^2(u(t)) \le |\partial\phi|^2(v) + \frac{1}{t^2} d^2(v, u_0) \quad & \forall\, v \in D(|\partial\phi|). \end{aligned} \tag{4.0.12}$$

iii) Uniqueness and evolution variational inequalities: *u is the unique solution of the evolution variational inequality*

$$\frac{1}{2}\frac{d}{dt}d^2(u(t),v)+\frac{1}{2}\lambda d^2(u(t),v)+\phi(u(t))\le\phi(v)\quad \mathscr{L}^1\text{-a.e. } t>0,\ \forall v\in D(\phi), \tag{4.0.13}$$

among all the locally absolutely continuous curves such that $\lim_{t\downarrow 0} u(t)=u_0$ *in $\mathscr{S}$.*

iv) Contraction semigroup: *The map $t\mapsto S[u_0](t)$ is a λ-contracting semigroup i.e.*

$$d\big(S[u_0](t),S[v_0](t)\big)\le e^{-\lambda t}d(u_0,v_0)\quad \forall\, u_0,v_0\in\overline{D(\phi)}. \tag{4.0.14}$$

v) Optimal a priori estimate: *if $u_0\in D(\phi)$ and $\lambda=0$ then*

$$d^2\Big(S[u_0](t),(J_{t/n})^n[u_0]\Big)\le\frac{t}{n}\Big(\phi(u_0)-\phi_{t/n}(u_0)\Big)\le\frac{t^2}{2n^2}|\partial\phi|^2(u_0). \tag{4.0.15}$$

Remark 4.0.5. Let us collect some comments about this result:

(a) The regularizing effect provided by (4.0.12) is stronger than the analogous property proved in Theorem 2.4.15 for λ-convex function, since in this case we simply need $u_0\in\overline{D(\phi)}$ instead of $u_0\in D(\phi)$. Inequality (4.0.12) also implies a faster decay of $|\partial\phi|(u(t))$ as $t\uparrow+\infty$.

(b) Since for differentiable curves u in a Hilbert space $\mathscr{S}=\mathscr{H}$

$$\langle\frac{d}{dt}u(t),u(t)-v\rangle=\frac{1}{2}\frac{d}{dt}|u(t)-v|^2=\frac{1}{2}\frac{d}{dt}d^2(u(t),v)\quad\forall\, v\in\mathscr{H},$$

the variational inequality formulation (4.0.13) is formally equivalent to (4.0.4) (in the case $\lambda=0$), but it does not require neither the existence of the pointwise derivative of u nor a vectorial structure. A similar idea was introduced by P. Bénilan [22] for the definition of the integral solutions of evolution equations governed by m-accretive operators in Banach spaces. The integral formulation corresponds to condier (4.0.13) in the weaker distributional sense:

$$\frac{1}{2}d^2(u(t),v)-\frac{1}{2}d^2(u(s),v)\le\int_s^t\Big(\phi(v)-\phi(u(r))-\frac{\lambda}{2}d^2(u(r),v)\Big)\,dr, \tag{4.0.16}$$

for every $v\in D(\phi)$ and $0<s<t$; in this way, one can simply require that u is a continuous curve with $\phi\circ u\in L^1_{loc}(0,+\infty)$, thus avoiding any *a priori* regularity assumption on the evolution curve. It would not be difficult to show that there exists at most one integral solution with prescribed initial datum and that this formulation is equivalent to (4.0.13).

(c) The semigroup S satisfies the contracting property (4.0.14) (e.g. for $\lambda = 0$) even if at the discrete level the resolvent operator does not satisfy in general the analogous property (4.0.2).

(d) In the case $\lambda > 0$ (4.0.14) provides another estimates of the exponential decay of the solution u to the unique minimum point $\bar{u}$ of ϕ (cf. (2.4.12)), as already discussed in Theorem 2.4.14, i.e.

$$d(u(t), \bar{u}) \le e^{-\lambda t} d(u_0, \bar{u}) \quad \forall\, t \ge 0. \tag{4.0.17}$$

(e) The estimates (4.0.15) are exactly the same of the Hilbert framework: in fact the first one is even slightly better than the previously known results, since it exhibits an order of convergence $o(\sqrt{1/n})$ instead of $O(\sqrt{1/n})$ for $u_0 \in D(\phi)$ and it shows that the error is related to the speed of convergence of the Moreau-Yosida approximation ϕ_τ to ϕ as $\tau \downarrow 0$. Starting from this formula, it would not be difficult to relate the order of convergence to the regularity of u_0, measured in suitable (nonlinear) interpolation classes between $D(\phi)$ and $D(|\partial\phi|)$ (see e.g. [29], [19]).

In the limiting case $\lambda = 0$ the exponential decay does not occur, in general, but we can still prove some weaker results on the asymptotic behaviour of u, which are easy consequences of (4.0.12) and of (4.0.13).

Corollary 4.0.6. *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold with* $\lambda = 0$, *and that* $\bar{u}$ *is a minimum point for* ϕ. *Then the solution* $u = S[u_0]$ *provided by Theorem* 4.0.4 *satisfies*

$$\begin{gathered} |\partial\phi|(u(t)) \le \frac{d(u_0, \bar{u})}{t}, \qquad \phi(u(t)) - \phi(\bar{u}) \le \frac{d^2(u_0, \bar{u})}{2t}, \\ \text{the map} \quad t \mapsto d(u(t), \bar{u}) \quad \text{is not increasing.} \end{gathered} \tag{4.0.18}$$

In particular, if the sublevels of ϕ *are compact, then* $u(t) \stackrel{d}{\to} u_\infty$ *as* $t \to \infty$ *and* u_∞ *is a minimum point for* ϕ.

General a priori and a posteriori error estimates. (4.0.15) is a particular case of the general error estimates which can also be proved for non uniform partitions; quite surprisingly, they reproduce exactly the same structure of the Hilbertian setting and can be derived by a preliminary *a posteriori error analysis* (we refer to [102] for a detailed account of the various contributions to the subject of the *a priori and a posteriori* error estimates in the Hilbert case).

As we have already seen in (4.0.15), for each estimate the order of convergence depends on the regularity of the initial datum: the best one is obtained if $u_0 \in D(|\partial\phi|)$, whereas an intermediate order $O(\sqrt{|\boldsymbol{\tau}|})$ can be proved if $u_0 \in D(\phi)$; simple linear examples show that these bounds are optimal.

We first present the most interesting result for $\lambda = 0$ and then we will show how the various constants are affected by different values of λ.

Theorem 4.0.7 (The case $\lambda = 0$). *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold with* $\lambda = 0$*, let* $u \in MM(\Phi; u_0)$ *be the unique solution of the equation* (4.0.13) *and let* $\overline{U}_{\boldsymbol{\tau}}$ *be a discrete solution associated to the partition* $\mathcal{P}_{\boldsymbol{\tau}}$ (2.0.1)*. If* $u_0 \in D(\phi)$ *and* $T = t^N_{\boldsymbol{\tau}} \in \mathcal{P}_{\boldsymbol{\tau}}$

$$d^2(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \le d^2(U^0_{\boldsymbol{\tau}}, u_0) + \sum_{n=1}^{N-1} \tau_n^2 \mathscr{E}^n_{\boldsymbol{\tau}}, \tag{4.0.19}$$

where

$$\mathscr{E}^n_{\boldsymbol{\tau}} := \frac{\phi(U^{n-1}_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}})}{\tau_n} - \frac{d^2(U^{n-1}_{\boldsymbol{\tau}}, U^n_{\boldsymbol{\tau}})}{\tau_n^2} \tag{4.0.20}$$

and

$$\sum_{n=1}^{N} \tau_n^2 \mathscr{E}^n_{\boldsymbol{\tau}} \le |\boldsymbol{\tau}| \Big(\phi(U^0_{\boldsymbol{\tau}}) - \phi_T(U^0_{\boldsymbol{\tau}})\Big); \tag{4.0.21}$$

if $U^0_{\boldsymbol{\tau}} \equiv u_0$ *we have*

$$d^2(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \le |\boldsymbol{\tau}|\Big(\phi(u_0) - \phi_T(u_0)\Big) \le |\boldsymbol{\tau}|\Big(\phi(u_0) - \inf_{\mathscr{S}} \phi\Big) \quad \forall T > 0. \tag{4.0.22}$$

If $U^0_{\boldsymbol{\tau}} \in D(|\partial\phi|)$ *we have*

$$\sum_{n=1}^{N} \tau_n^2 \mathscr{E}^n_{\boldsymbol{\tau}} \le \frac{|\boldsymbol{\tau}|^2}{2} |\partial\phi|^2(U^0_{\boldsymbol{\tau}}); \tag{4.0.23}$$

if $U^0_{\boldsymbol{\tau}} \equiv u_0$ *we have*

$$d^2(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \le \frac{|\boldsymbol{\tau}|^2}{2} |\partial\phi|^2(u_0) \quad \forall T > 0. \tag{4.0.24}$$

Remark 4.0.8. (4.0.21) is slightly worse than (4.0.15), which in the case of a uniform mesh and $u_0 \in D(\phi)$ provides an $o(\sqrt{|\boldsymbol{\tau}|})$ estimates, instead of $O(\sqrt{|\boldsymbol{\tau}|})$: this fact depends on a finer cancellation effect which seems to be related to the choice of uniform step sizes.

In the case $\lambda \neq 0$ the error $d(\overline{U}_{\boldsymbol{\tau}}(T), u(T))$ should be affected by an exponential factor $e^{-\lambda T}$, corresponding to (4.0.14) or $e^{-\lambda_{\boldsymbol{\tau}} T}$, where

$$\lambda_{\boldsymbol{\tau}} := \frac{\log(1 + \lambda|\boldsymbol{\tau}|)}{|\boldsymbol{\tau}|} \quad \text{as for the discrete bounds of Lemma 3.4.1;} \tag{4.0.25}$$

the involved constants could also be perturbed by the presence of λ: here the main technical difficulty is to obtain estimates which exhibit the right coefficient of the exponential grow (or decay) and constants which reduce to the optimal ones (4.0.22), (4.0.24) when $\lambda = 0$.

We limit us to detail the *a priori* bounds of the error: we adopt the convention to denote by $c = c(\lambda, |\boldsymbol{\tau}|, T)$ the constants which depend only on the parameters λ, $|\boldsymbol{\tau}|$, T, exhibit at most a polynomial (in fact linear or quadratic) growth with respect to T, and are asymptotic to 1 as $\lambda \to 0$.

Theorem 4.0.9 (The case $\lambda < 0$). *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *holds for* $\lambda < 0$, *let* $u \in MM(\Phi; u_0)$ *be the unique solution of the equation* (4.0.13) *and let* $\overline{U}_{\boldsymbol{\tau}}$ *be the discrete solution associated to the partition* $\mathcal{P}_{\boldsymbol{\tau}}$ *in* (2.0.1) *with* $|\boldsymbol{\tau}| < (-\lambda)^{-1}$. *If* $U^0_{\boldsymbol{\tau}} = u_0 \in D(\phi)$ *we have*

$$d^2(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \leq c\,|\boldsymbol{\tau}|\Big(\phi(u_0) - \inf_{\mathscr{S}} \phi\Big) e^{-2\lambda T}, \qquad c := \Big(1 + \sqrt{\tfrac{4}{3}|\lambda|\,|\boldsymbol{\tau}|}\Big)^2. \tag{4.0.26}$$

If $U^0_{\boldsymbol{\tau}} = u_0 \in D(|\partial\phi|)$, $\lambda_{\boldsymbol{\tau}}$ *is defined as in* (4.0.25), *and* $T_{\boldsymbol{\tau}} = \min\big\{t^k_{\boldsymbol{\tau}} \in \mathcal{P}_{\boldsymbol{\tau}} : t^k_{\boldsymbol{\tau}} \geq T\big\}$, *we have*

$$d(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \leq c\,\frac{|\boldsymbol{\tau}|}{\sqrt{2}}|\partial\phi|(u_0)\, e^{-\lambda_{\boldsymbol{\tau}} T}, \qquad c := \frac{1 + 2|\lambda|\, T_{\boldsymbol{\tau}}}{1 + \lambda|\boldsymbol{\tau}|}. \tag{4.0.27}$$

We recall that in the case $\lambda > 0$ the function ϕ is bounded from below.

Theorem 4.0.10 (The case $\lambda > 0$). *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold for* $\lambda > 0$, *let* $u \in MM(\Phi; u_0)$ *be the unique solution of the equation* (4.0.13), *let* $\overline{U}_{\boldsymbol{\tau}}$ *be a discrete solution associated to the partition* $\mathcal{P}_{\boldsymbol{\tau}}$ (2.0.1), *and let* $\lambda_{\boldsymbol{\tau}}$ *be defined as in* (4.0.25). *If* $U^0_{\boldsymbol{\tau}} = u_0 \in D(\phi)$ *and* $T_{\boldsymbol{\tau}} \in \mathcal{P}_{\boldsymbol{\tau}}$ *is defined as in the above Theorem, we have*

$$\begin{aligned} d^2(\overline{U}_{\boldsymbol{\tau}}(T), u(T))) &\leq c\,|\boldsymbol{\tau}|\Big(\phi(u_0) - \inf_{\mathscr{S}} \phi\Big) e^{-2\lambda_{\boldsymbol{\tau}} T}, \\ c &:= (1 + \lambda|\boldsymbol{\tau}|)\big(1 + \sqrt{2\lambda T_{\boldsymbol{\tau}}}\big)^4. \end{aligned} \tag{4.0.28}$$

If $U^0_{\boldsymbol{\tau}} = u_0 \in D(|\partial\phi|)$ *we have*

$$d^2(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \leq c\,\frac{|\boldsymbol{\tau}|^2}{2}|\partial\phi|^2(u_0)\, e^{-2\lambda_{\boldsymbol{\tau}} T}, \qquad c := 1 + 2\lambda\, T_{\boldsymbol{\tau}}. \tag{4.0.29}$$

We split the proof of the previous theorems in many steps:

4.1.1: discrete variational inequalities. First of all we derive the variational evolution inequalities (4.1.3), which are the discrete counterparts of (4.0.13). They provide a crucial property satisfied by the discrete solutions and are a simple consequence of the convexity assumption 4.0.1; all the subsequent estimates can be deduced from this fundamental point.

4.1.2: Cauchy-type estimates. Here we introduce a general way to pass from a discrete variational inequality to a continuous one, though affected by a perturbation term; the main technical difficulty is the lackness of an underlying linear structure, which prevents an easy interpolation of the discrete values in the ambient space $\mathscr{S}$. We circumvent this fact by considering affine interpolations of the values of the functions instead of trying to interpolate their arguments (see also [101] for a similar approach). Once continuous versions of the evolution variational inequalities are at our disposal, it will not be difficult to derive Cauchy-type estimates, by also applying a Gronwall lemma in the case $\lambda \neq 0$.

4.2: convergence. This section is devoted to control the perturbation terms in the previously derived estimates, in order to prove the convergence of the scheme. We first consider the easier case $u_0 \in D(\phi)$ and then we extend the results to a general $u_0 \in \overline{D(\phi)}$.

4.3: regularizing effect and semigroup generation. Here we show that the unique element $u \in MM(\Phi; u_0)$ exhibits the regularizing effect (4.0.12) and then derives the differential characterization (4.0.13) which also yield the λ-contracting semigroup property (4.0.14).

4.4: optimal error estimates. Finally, we refine the error estimates which have been derived in the first section, and we prove Theorems 4.0.7, 4.0.9, 4.0.10, and the related estimate (4.0.15). For the ease of the reader, the main ideas are first presented in the case $\lambda = 0$; the more technical results for $\lambda \neq 0$ are discussed in Section 4.4.2

4.1 Cauchy-type estimates for discrete solutions

4.1.1 Discrete variational inequalities

Let us first state an auxiliary lemma:

Lemma 4.1.1. *Let us suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold for some* $\lambda \in \mathbb{R}$, *and let* $0 < \tau < \frac{1}{\lambda^-}$. *If* $u \in \overline{D(\phi)}$ *and* (v_n) *is a sequence in* $D(\phi)$ *satisfying*

$$\limsup_{n \to \infty} \Phi(\tau, u; v_n) \leq \phi_\tau(u), \tag{4.1.1}$$

then (v_n) *converges to* $v \in D(\phi)$ *and* $v = u_\tau = J_\tau[u]$ *is the unique element of* $J_\tau[u]$.

Proof. Being $u \in \overline{D(\phi)}$, we can find a sequence $(u_n) \subset D(\phi)$ converging to u such that

$$\limsup_{n \to \infty} \Phi(\tau, u_n; v_n) = \limsup_{n \to \infty} \Phi(\tau, u; v_n) \leq \phi_\tau(u).$$

We argue as in the proof of Lemma 2.4.8: observe that, being ϕ_τ continuous (cf. Lemma 3.1.2) and $\phi_\tau(u) < +\infty$, (4.1.1) yields

$$\begin{aligned}\Phi(\tau, u_n; v_n) &= \phi_\tau(u_n) + \Big(\phi_\tau(u) - \phi_\tau(u_n)\Big) + \Big(\Phi(\tau, u_n; v_n) - \phi_\tau(u)\Big) \\ &= \phi_\tau(u_n) + \omega_n \quad \text{with} \quad \limsup_{n \to \infty} \omega_n \leq 0.\end{aligned}$$

We apply the convexity property (4.0.10) with $w := u_n, v_0 := v_n, v_1 := v_m$ at $t = 1/2$ to find $v_{n,m}$ such that

$$\phi_\tau(u_n) \leq \Phi(\tau, u_n; v_{n,m}) \leq \phi_\tau(u_n) + \frac{\omega_n + \omega_m}{2} - \frac{1 + \lambda\tau}{8\tau} d^2(v_n, v_m).$$

Since $1 + \lambda\tau > 0$ this implies that

$$\limsup_{n,m\to\infty} d^2(v_n, v_m) \le \frac{4\tau}{1+\lambda\tau} \limsup_{n,m\to\infty} \big(\omega_n + \omega_m\big) = 0,$$

therefore (v_n) is a Cauchy sequence and the lower semicontinuity of ϕ gives that $\Phi(\tau, u; v) = \phi_\tau(u)$, i.e. $v \in J_\tau[u]$. The same argument also shows that v is the unique element of $J_\tau[u]$. □

The following result is a significant improvement of Theorem 3.1.6:

Theorem 4.1.2 (Variational inequalities for u_τ). *Let us suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *holds for some* $\lambda \in \mathbb{R}$.

(i) *If* $u \in \overline{D(\phi)}$ *and* $\lambda\tau > -1$ *then the minimum problem* (2.0.5) *has a unique solution* $u_\tau = J_\tau[u]$. *The map* $u \in \overline{D(\phi)} \mapsto J_\tau[u]$ *is continuous.*

(ii) *If* $u \in \overline{D(\phi)}$ *and* $u_\tau = J_\tau[u]$, *for each* $v \in D(\phi)$ *we have*

$$\frac{1}{2\tau} d^2(u_\tau, v) - \frac{1}{2\tau} d^2(u, v) + \frac{1}{2}\lambda d^2(u_\tau, v) \le \phi(v) - \phi_\tau(u). \tag{4.1.2}$$

Proof. (i) In order to show the existence of a minimum point $u_\tau \in J_\tau[u]$ we simply apply the previous Lemma 4.1.1 by choosing an arbitrary minimizing sequence, thus satisfying (4.1.1).

The continuity of J_τ follows by the same argument; simply take a sequence $(u_n) \subset \overline{D(\phi)}$ converging to u and observe that $v_n := J_\tau[u_n]$ is bounded in $\mathscr{S}$ and satisfies

$$\limsup_{n\to\infty} \Phi(\tau, u; v_n) = \limsup_{n\to\infty} \Phi(\tau, u_n; v_n) = \lim_{n\to\infty} \phi_\tau(u_n) = \phi_\tau(u).$$

(ii) Since the map J_τ is continuous, by a standard approximation argument we can suppose $u \in D(\phi)$. We apply (4.0.10) again with $w := u$, $v_0 := u_\tau$ and $v_1 := v$, obtaining a family $v_t \in D(\phi)$, $t \in (0, 1)$, such that

$$\Phi(\tau, u; u_\tau) \le \Phi(\tau, u; v_t) \le (1-t)\Phi(\tau, u; u_\tau) + t\Phi(\tau, u; v) - \frac{1+\lambda\tau}{2\tau} t(1-t) d^2(u_\tau, v).$$

Subtracting $\Phi(\tau, u; u_\tau)$ by each term of the inequality, dividing by t, and passing to the limit as $t \downarrow 0$ we get

$$0 \le -\Phi(\tau, u; u_\tau) + \Phi(\tau, u; v) - \frac{1+\lambda\tau}{2\tau} d^2(u_\tau, v)$$

which is equivalent to (4.1.2) since $\phi_\tau(u) = \Phi(\tau, u; u_\tau)$. □

Corollary 4.1.3 (Variational inequalities for discrete solutions). *Under the same assumptions of the previous Lemma, every discrete solution* $\{U^n_{\boldsymbol\tau}\}_{n=0}^{+\infty}$ *with* $U^0_{\boldsymbol\tau} \in \overline{D(\phi)}$ *satisfies*

$$\begin{aligned}\frac{1}{2\tau_n}\Big(d^2(U^n_{\boldsymbol\tau},V)-d^2(U^{n-1}_{\boldsymbol\tau},V)\Big)+\frac12\lambda\,d^2(U^n_{\boldsymbol\tau},V)&\\ \le \phi(V)-\phi(U^n_{\boldsymbol\tau})-\frac{1}{2\tau_n}d^2(U^n_{\boldsymbol\tau},U^{n-1}_{\boldsymbol\tau})\quad \forall\,V\in D(\phi),\ n\ge 1.&\end{aligned}\tag{4.1.3}$$

4.1.2 Piecewise affine interpolation and comparison results

Now we formalize a general way to write a discrete difference inequality as a continuous one: first of all, let us introduce the "delayed" piecewise constant function $\underline{U}_{\boldsymbol\tau}$

$$\underline{U}_{\boldsymbol\tau}(t)\equiv U^{n-1}_{\boldsymbol\tau}\quad \text{if } t\in(t^{n-1}_{\boldsymbol\tau},t^n_{\boldsymbol\tau}],$$

and the interpolating functions

$$\ell_{\boldsymbol\tau}(t):=\frac{t-t^{n-1}_{\boldsymbol\tau}}{\tau_n},\quad 1-\ell_{\boldsymbol\tau}(t)=\frac{t^n_{\boldsymbol\tau}-t}{\tau_n}\quad \text{if } t\in(t^{n-1}_{\boldsymbol\tau},t^n_{\boldsymbol\tau}].\tag{4.1.4}$$

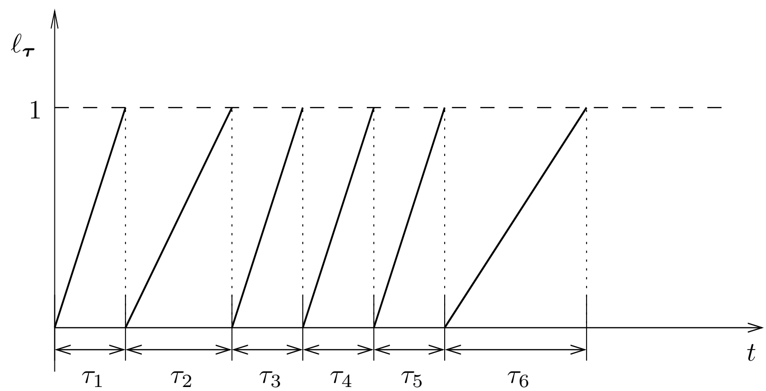


Figure 4.1: The interpolating functions $\ell_{\boldsymbol\tau}$.

If $\zeta:\mathscr{S}\to(-\infty,+\infty]$ is a function which is finite on the discrete solution $\{U^n_{\boldsymbol\tau}\}_{n=0}^{+\infty}$, we can define its *affine interpolation* as

$$\begin{aligned}\zeta_{\boldsymbol\tau}(t):=&(1-\ell_{\boldsymbol\tau}(t))\zeta(\underline{U}_{\boldsymbol\tau}(t))+\ell_{\boldsymbol\tau}(t)\zeta(\overline{U}_{\boldsymbol\tau}(t))\\ =&(1-\ell_{\boldsymbol\tau}(t))\zeta(U^{n-1}_{\boldsymbol\tau})+\ell_{\boldsymbol\tau}(t)\zeta(U^n_{\boldsymbol\tau})\quad \text{if } t\in(t^{n-1}_{\boldsymbol\tau},t^n_{\boldsymbol\tau}].\end{aligned}\tag{4.1.5}$$

In other words, $\zeta_{\boldsymbol\tau}$ is the continuous piecewise affine function which interpolates the values $\zeta(U^n_{\boldsymbol\tau})$ at the nodes $t^n_{\boldsymbol\tau}$ of the partition $\mathcal{P}_{\boldsymbol\tau}$. In this way, for $V\in\mathscr{S}$, we can consider the functions

$$d^2_{\boldsymbol{\tau}}(t;V) := (1-\ell_{\boldsymbol{\tau}}(t))d^2(U^{n-1}_{\boldsymbol{\tau}},V) + \ell_{\boldsymbol{\tau}}(t)d^2(U^n_{\boldsymbol{\tau}},V) \qquad t\in(t^{n-1}_{\boldsymbol{\tau}},t^n_{\boldsymbol{\tau}}], \tag{4.1.6}$$

$$\varphi_{\boldsymbol{\tau}}(t) := (1-\ell_{\boldsymbol{\tau}}(t))\phi(U^{n-1}_{\boldsymbol{\tau}}) + \ell_{\boldsymbol{\tau}}(t)\phi(U^n_{\boldsymbol{\tau}}) \qquad t\in(t^{n-1}_{\boldsymbol{\tau}},t^n_{\boldsymbol{\tau}}]. \tag{4.1.7}$$

The main idea here is to "interpolate a function" instead of evaluating it on a (more difficult) interpolation of the arguments (see also [101] for another application of this technique); of course, for convex functional in Euclidean space these two approaches are slightly different but in our metric framework the first one is particularly convenient.

Finally, to every discrete solution $\{U^n_{\boldsymbol{\tau}}\}_{n=0}^{+\infty}\subset D(\phi)$ defined as before we associate the "squared discrete derivative"

$$D^n_{\boldsymbol{\tau}} := \frac{d^2(U^{n-1}_{\boldsymbol{\tau}},U^n_{\boldsymbol{\tau}})}{\tau_n^2}, \quad n=1,\cdots, \tag{4.1.8}$$

and the residual function $\mathscr{R}_{\boldsymbol{\tau}}$, defined for $t\in(t^{n-1}_{\boldsymbol{\tau}},t^n_{\boldsymbol{\tau}}]$ by

$$\mathscr{R}_{\boldsymbol{\tau}}(t) := 2(1-\ell_{\boldsymbol{\tau}}(t))\Big(\phi(U^{n-1}_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}) - \frac{\tau_n}{2}D^n_{\boldsymbol{\tau}}\Big) - \ell_{\boldsymbol{\tau}}(t)\tau_n D^n_{\boldsymbol{\tau}} \tag{4.1.9a}$$

$$= 2(1-\ell_{\boldsymbol{\tau}}(t))\tau_n\mathscr{E}^n_{\boldsymbol{\tau}} + \big(1-2\ell_{\boldsymbol{\tau}}(t)\big)\tau_n D^n_{\boldsymbol{\tau}}. \tag{4.1.9b}$$

Observe that (3.1.20) yields

$$\begin{aligned}(1+\lambda\tau_n)|\partial\phi|^2(U^n_{\boldsymbol{\tau}}) &\le (1+\lambda\tau_n)D^n_{\boldsymbol{\tau}} \le \frac{2}{\tau_n}\Big(\phi(U^{n-1}_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}) - \frac{\tau_n}{2}D^n_{\boldsymbol{\tau}}\Big)\\ &\le \frac{1}{1+\lambda\tau_n}|\partial\phi|^2(U^{n-1}_{\boldsymbol{\tau}}) \le \frac{1}{1+\lambda\tau_n}D^{n-1}_{\boldsymbol{\tau}},\end{aligned} \tag{4.1.10}$$

so that, if $U^{n-1}_{\boldsymbol{\tau}}\in D(|\partial\phi|)$ then (4.1.9a) yields

$$\mathscr{R}_{\boldsymbol{\tau}}(t) \le \tau_n\frac{1-\ell_{\boldsymbol{\tau}}(t)}{1+\lambda\tau_n}|\partial\phi|^2(U^{n-1}_{\boldsymbol{\tau}}) - \ell_{\boldsymbol{\tau}}(t)\tau_n D^n_{\boldsymbol{\tau}} \qquad t\in(t^{n-1}_{\boldsymbol{\tau}},t^n_{\boldsymbol{\tau}}]. \tag{4.1.11}$$

Theorem 4.1.4. *Let us suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold for* $\lambda\in\mathbb{R}$, *and* $U^0_{\boldsymbol{\tau}}\in D(\phi)$. *The interpolated functions* $d_{\boldsymbol{\tau}}$, $\varphi_{\boldsymbol{\tau}}$ *defined as in* (4.1.6), (4.1.7) *starting from the discrete solution* $\{U^n_{\boldsymbol{\tau}}\}_{n=0}^{+\infty}$ *satisfy the following system of variational inequalities almost everywhere in* $(0,+\infty)$*:*

$$\frac12\frac{d}{dt}d^2_{\boldsymbol{\tau}}(t;V) + \frac{\lambda}{2}d^2(\overline{U}_{\boldsymbol{\tau}}(t),V) + \varphi_{\boldsymbol{\tau}}(t) - \phi(V) \le \frac12\mathscr{R}_{\boldsymbol{\tau}}(t) \quad \forall V\in D(\phi). \tag{4.1.12}$$

Proof. If $t \in (t_{\boldsymbol{\tau}}^{n-1}, t_{\boldsymbol{\tau}}^n]$, using (4.1.3) we obtain

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}d^2_{\boldsymbol{\tau}}(t;V) + \frac{1}{2}\lambda d^2(\overline{U}_{\boldsymbol{\tau}}(t), V) + \varphi_{\boldsymbol{\tau}}(t) - \phi(V)\\
&= \frac{1}{2\tau_n}\Big(d^2(U^n_{\boldsymbol{\tau}}, V) - d^2(U^{n-1}_{\boldsymbol{\tau}}, V)\Big) + \frac{1}{2}\lambda d^2(U^n_{\boldsymbol{\tau}}, V) + \varphi_{\boldsymbol{\tau}}(t) - \phi(V)\\
&\le -\frac{1}{2\tau_n}d^2(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}}) + \phi(V) - \phi(U^n_{\boldsymbol{\tau}}) + \varphi_{\boldsymbol{\tau}}(t) - \phi(V)\\
&= -\frac{1}{2\tau_n}d^2(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}}) + (1 - \ell_{\boldsymbol{\tau}}(t))\Big(\phi(U^{n-1}_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}})\Big)\\
&= (1 - \ell_{\boldsymbol{\tau}}(t))\Big(\phi(U^{n-1}_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}) - \frac{1}{2\tau_n}d^2(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}})\Big) - \ell_{\boldsymbol{\tau}}(t)\frac{1}{2\tau_n}d^2(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}}).
\end{aligned}$$

Recalling the Definition (4.1.9a) of $\mathscr{R}_{\boldsymbol{\tau}}(t)$ we conclude. □

Comparison between discrete solutions for $\lambda = 0$. In the next Corollary we are finally able to compare two discrete solutions.

Corollary 4.1.5 (Comparison for $\lambda = 0$). *Under the same assumptions of Theorem 4.1.4, let us suppose that $\lambda = 0$ and let $\{U^m_{\boldsymbol{\eta}}\}_{m=0}^{+\infty}$, $U^0_{\boldsymbol{\eta}} \in D(\phi)$, be another discrete solution associated to the admissible partition*

$$\mathcal{P}_{\boldsymbol{\eta}} := \Big\{0 = t^0_{\boldsymbol{\eta}} < t^1_{\boldsymbol{\eta}} < \ldots < t^m_{\boldsymbol{\eta}}, \ldots\Big\}, \quad \eta_m = t^m_{\boldsymbol{\eta}} - t^{m-1}_{\boldsymbol{\eta}}. \tag{4.1.13}$$

The continuous and piecewise affine function

$$d^2_{\boldsymbol{\tau\eta}}(t,s) := (1 - \ell_{\boldsymbol{\eta}}(s))d^2_{\boldsymbol{\tau}}(t; \underline{U}_{\boldsymbol{\eta}}(s)) + \ell_{\boldsymbol{\eta}}(s)d^2_{\boldsymbol{\tau}}(t; \overline{U}_{\boldsymbol{\eta}}(s)) \quad t, s \ge 0 \tag{4.1.14}$$

satisfies the differential inequality

$$\frac{d}{dt}d^2_{\boldsymbol{\tau\eta}}(t,t) \le \mathscr{R}_{\boldsymbol{\tau}}(t) + \mathscr{R}_{\boldsymbol{\eta}}(t) \quad \forall\, t \in (0, +\infty) \setminus (\mathcal{P}_{\boldsymbol{\tau}} \cup \mathcal{P}_{\boldsymbol{\eta}}) \tag{4.1.15}$$

and therefore the integral bound

$$d^2_{\boldsymbol{\tau\eta}}(T,T) \le d^2(U^0_{\boldsymbol{\tau}}, U^0_{\boldsymbol{\eta}}) + \int_0^T \Big(\mathscr{R}_{\boldsymbol{\tau}}(t) + \mathscr{R}_{\boldsymbol{\eta}}(t)\Big)\, dt. \tag{4.1.16}$$

Proof. Defining the function $\varphi_{\boldsymbol{\eta}}(s)$ as in (4.1.5) by

$$\varphi_{\boldsymbol{\eta}}(s) := (1 - \ell_{\boldsymbol{\eta}}(s))\phi(\underline{U}_{\boldsymbol{\eta}}(s)) + \ell_{\boldsymbol{\eta}}(s)\phi(\overline{U}_{\boldsymbol{\eta}}(s)), \tag{4.1.17}$$

a convex combination of (4.1.12) for $V := \underline{U}_{\boldsymbol{\eta}}(s)$ and $V := \overline{U}_{\boldsymbol{\eta}}(s)$ yields

$$\frac{1}{2}\frac{\partial}{\partial t}d^2_{\boldsymbol{\tau\eta}}(t,s) + \varphi_{\boldsymbol{\tau}}(t) - \varphi_{\boldsymbol{\eta}}(s) \le \frac{1}{2}\mathscr{R}_{\boldsymbol{\tau}}(t) \quad \forall\, s > 0,\ t \in (0, +\infty) \setminus \mathcal{P}_{\boldsymbol{\tau}}.$$

Analogously, writing (4.1.12) for the function $d^2_{\boldsymbol\eta}$ defined as in (4.1.6)

$$d^2_{\boldsymbol\eta}(s;V) := (1-\ell_{\boldsymbol\eta}(s))d^2(\underline{U}_{\boldsymbol\eta}(s),V) + \ell_{\boldsymbol\eta}(s)d^2(\overline{U}_{\boldsymbol\eta}(s),V),$$

and reversing the roles of $\boldsymbol\eta$ and $\boldsymbol\tau$ we obtain

$$\frac{1}{2}\frac{\partial}{\partial s}d^2_{\boldsymbol\eta\boldsymbol\tau}(s,t) + \varphi_{\boldsymbol\eta}(s) - \varphi_{\boldsymbol\tau}(t) \le \frac{1}{2}\mathscr{R}_{\boldsymbol\eta}(s) \quad \forall\, t>0,\ s\in(0,+\infty)\setminus\mathcal{P}_{\boldsymbol\eta},$$

where

$$d^2_{\boldsymbol\eta\boldsymbol\tau}(s,t) := (1-\ell_{\boldsymbol\tau}(t))d^2_{\boldsymbol\eta}(s;U^{n-1}_{\boldsymbol\tau}) + \ell_{\boldsymbol\tau}(t)d^2_{\boldsymbol\eta}(s;U^n_{\boldsymbol\tau}) \quad \text{for } t\in(t^{n-1}_{\boldsymbol\tau},t^n_{\boldsymbol\tau}]. \tag{4.1.18}$$

Summing up the two contributions we find

$$\frac{\partial}{\partial t}d^2_{\boldsymbol\tau\boldsymbol\eta}(t,s) + \frac{\partial}{\partial s}d^2_{\boldsymbol\eta\boldsymbol\tau}(s,t) \le \mathscr{R}_{\boldsymbol\tau}(t) + \mathscr{R}_{\boldsymbol\eta}(s) \quad \forall\, s,t\in(0,+\infty)\setminus(\mathcal{P}_{\boldsymbol\tau}\cup\mathcal{P}_{\boldsymbol\eta}).$$

Finally, by the symmetry property

$$d^2_{\boldsymbol\tau\boldsymbol\eta}(t,s) = d^2_{\boldsymbol\eta\boldsymbol\tau}(s,t), \tag{4.1.19}$$

evaluating the previous inequality for $s=t$ we end up with (4.1.15). □

Comparison between discrete solutions for $\lambda \neq 0$. If $\lambda\neq 0$ we need to rewrite (4.1.12) in a more convenient form; let us first observe that the concavity of the square root provides the inequalities for $V\in\mathscr{S}$

$$(1-\ell_{\boldsymbol\tau}(t))d(\underline{U}_{\boldsymbol\tau}(t),V) + \ell_{\boldsymbol\tau}(t)d(\overline{U}_{\boldsymbol\tau}(t),V) \le d_{\boldsymbol\tau}(t,V) \quad \forall\, t>0, \tag{4.1.20}$$

$$(1-\ell_{\boldsymbol\eta}(s))d_{\boldsymbol\tau}(t,\underline{U}_{\boldsymbol\eta}(s)) + \ell_{\boldsymbol\eta}(s)d_{\boldsymbol\tau}(t,\overline{U}_{\boldsymbol\eta}(s)) \le d_{\boldsymbol\tau\boldsymbol\eta}(t,s) \quad \forall\, t,s>0. \tag{4.1.21}$$

Lemma 4.1.6. *Under the same assumptions of Theorem* 4.1.4, *for a discrete solution* $\{U^n_{\boldsymbol\tau}\}_{n=0}^{+\infty}$ *with* $U^0_{\boldsymbol\tau}\in D(\phi)$ *let us define*

$$\mathscr{D}_{\boldsymbol\tau}(t) := \big(1-\ell_{\boldsymbol\tau}(t)\big)d(\overline{U}_{\boldsymbol\tau}(t),\underline{U}_{\boldsymbol\tau}(t)) = \tau_n\big(1-\ell_{\boldsymbol\tau}(t)\big)\sqrt{D^n_{\boldsymbol\tau}}, \quad t\in(t^{n-1}_{\boldsymbol\tau},t^n_{\boldsymbol\tau}]. \tag{4.1.22}$$

Then for every element $V\in D(\phi)$ *the interpolated functions* $d_{\boldsymbol\tau},\varphi_{\boldsymbol\tau}$ *defined by* (4.1.6) *and* (4.1.7) *satisfy the following system of variational inequalities almost everywhere in* $(0,+\infty)$*:*

$$\frac{d}{dt}\frac{1}{2}d^2_{\boldsymbol\tau}(t;V) + \frac{\lambda}{2}d^2_{\boldsymbol\tau}(t;V) - |\lambda|\mathscr{D}_{\boldsymbol\tau}(t)d_{\boldsymbol\tau}(t;V) + \varphi_{\boldsymbol\tau}(t) - \phi(V) \le \frac{1}{2}\mathscr{R}_{\boldsymbol\tau}(t) + \frac{\lambda^-}{2}\mathscr{D}^2_{\boldsymbol\tau}(t), \tag{4.1.23}$$

where $\lambda^- = \max(-\lambda,0)$.

Proof. If $\lambda\ge 0$ the inequality (4.1.23) is an immediate consequence of (4.1.12) and

$$-2d_{\boldsymbol\tau}(t;V)\mathscr{D}_{\boldsymbol\tau}(t) \le d^2(\overline{U}_{\boldsymbol\tau}(t),V) - d^2_{\boldsymbol\tau}(t;V)$$

which, in turn, follows by the triangle inequality. If $\lambda < 0$ it follows by (4.1.12) and

$$d^2(\overline{U}_{\boldsymbol{\tau}}(t),V) - d^2_{\boldsymbol{\tau}}(t;V) \le 2d_{\boldsymbol{\tau}}(t;V)\mathscr{D}_{\boldsymbol{\tau}}(t) + \mathscr{D}^2_{\boldsymbol{\tau}}(t). \tag{4.1.24}$$

Let us prove (4.1.24). Suppose $t \in (t^{n-1}_{\boldsymbol{\tau}}, t^n_{\boldsymbol{\tau}}]$ and $d^2(\overline{U}_{\boldsymbol{\tau}}(t),V) \ge d^2_{\boldsymbol{\tau}}(t;V)$, otherwise (4.1.24) is obvious; the elementary identity $a^2 - b^2 = 2b(a-b) + (a-b)^2$ yields

$$\begin{aligned} d^2(\overline{U}_{\boldsymbol{\tau}}(t),V) - d^2_{\boldsymbol{\tau}}(t;V) &= 2d_{\boldsymbol{\tau}}(t;V)\big(d(\overline{U}_{\boldsymbol{\tau}}(t),V) - d_{\boldsymbol{\tau}}(t;V)\big) \\ &+ \big(d(\overline{U}_{\boldsymbol{\tau}}(t),V) - d_{\boldsymbol{\tau}}(t;V)\big)^2. \end{aligned}$$

On the other hand the concavity inequality (4.1.20) gives

$$\begin{aligned} d(\overline{U}_{\boldsymbol{\tau}}(t),V) - d_{\boldsymbol{\tau}}(t;V) &\le d(\overline{U}_{\boldsymbol{\tau}}(t),V) - (1-\ell_{\boldsymbol{\tau}}(t))d(\underline{U}_{\boldsymbol{\tau}}(t),V) \\ &- \ell_{\boldsymbol{\tau}}(t)d(\overline{U}_{\boldsymbol{\tau}}(t),V) \le \mathscr{D}_{\boldsymbol{\tau}}(t). \end{aligned}$$

These two inequalities imply (4.1.24). □

Corollary 4.1.7 (Comparison for $\lambda \neq 0$). *Under the same assumption of the previous Lemma, let $\mathcal{P}_{\boldsymbol{\tau}}, \mathcal{P}_{\boldsymbol{\eta}}$ be two admissible partitions; the "error" function $d_{\boldsymbol{\tau\eta}}(t,s)$ defined by* (4.1.14) *satisfies the differential inequality*

$$\begin{aligned} \frac{d}{dt}d^2_{\boldsymbol{\tau\eta}}(t,t) + 2\lambda d^2_{\boldsymbol{\tau\eta}}(t,t) \le & 2|\lambda|\big(\mathscr{D}_{\boldsymbol{\tau}}(t) + \mathscr{D}_{\boldsymbol{\eta}}(t)\big)d_{\boldsymbol{\tau\eta}}(t,t) \\ & + (\mathscr{R}_{\boldsymbol{\tau}}(t) + \mathscr{R}_{\boldsymbol{\eta}}(t)) + \lambda^-\big(\mathscr{D}^2_{\boldsymbol{\tau}}(t) + \mathscr{D}^2_{\boldsymbol{\eta}}(t)\big), \end{aligned} \tag{4.1.25}$$

and therefore the Gronwall-like estimate

$$\begin{aligned} e^{\lambda T}d_{\boldsymbol{\tau\eta}}(T,T) \le & \left(d^2(U^0_{\boldsymbol{\tau}}, V^0_{\boldsymbol{\eta}}) + \mathsf{R}_{\boldsymbol{\tau}}(T) + \mathsf{R}_{\boldsymbol{\eta}}(T) + \int_0^T e^{2\lambda t}\lambda^-\big(\mathscr{D}^2_{\boldsymbol{\tau}}(t) + \mathscr{D}^2_{\boldsymbol{\eta}}(t)\big)\,dt\right)^{1/2} \\ & + 2\int_0^T |\lambda|e^{\lambda t}\big(\mathscr{D}_{\boldsymbol{\tau}}(t) + \mathscr{D}_{\boldsymbol{\eta}}(t)\big)\,dt, \end{aligned} \tag{4.1.26}$$

where $\mathsf{R}_{\boldsymbol{\tau}}$ (and analogously $\mathsf{R}_{\boldsymbol{\eta}}$) are defined by

$$\mathsf{R}_{\boldsymbol{\tau}}(T) := \sup_{t\in[0,T]}\int_0^t e^{2\lambda r}\mathscr{R}_{\boldsymbol{\tau}}(r)\,dr \le \int_0^T e^{2\lambda r}\Big(\mathscr{R}_{\boldsymbol{\tau}}(r)\Big)^+\,dr \quad \forall T > 0. \tag{4.1.27}$$

Proof. Starting from the inequality (4.1.23) we easily obtain (4.1.25) by arguing as in Corollary 4.1.5 and by using (4.1.21). Inequality (4.1.26) is a direct consequence of (4.1.25) and of the following version of the Gronwall Lemma [18]. □

Lemma 4.1.8 (A version of Gronwall Lemma). *Let $x : [0,+\infty) \to \mathbb{R}$ be a locally absolutely continuous function, let $a,\ b \in L^1_{\rm loc}([0,+\infty))$ be given functions satisfying, for $\lambda \in \mathbb{R}$,*

$$\frac{d}{dt}x^2(t) + 2\lambda x^2(t) \le a(t) + 2b(t)x(t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t > 0. \tag{4.1.28}$$

Then for every $T > 0$ we have

$$e^{\lambda T}|x(T)| \le \Big(x^2(0) + \sup_{t\in[0,T]}\int_0^t e^{2\lambda s}a(s)\,ds\Big)^{1/2} + 2\int_0^T e^{\lambda t}|b(t)|\,dt. \tag{4.1.29}$$

Proof. Multiplying (4.1.28) by $e^{2\lambda t}$ we obtain

$$\frac{d}{dt}\big(e^{\lambda t}x(t)\big)^2 \le e^{2\lambda t}a(t) + 2e^{\lambda t}b(t)\big(e^{\lambda t}x(t)\big) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0, \tag{4.1.30}$$

therefore it is sufficient to prove (4.1.29) for $\lambda = 0$.

Introducing the functions

$$\begin{aligned} X(T) &:= \sup_{t\in(0,T)}|x(t)|, \quad A(T) := \sup_{t\in(0,T)}\int_0^t a(s)\,ds,\\ B(T) &:= \int_0^T |b(s)|\,ds, \end{aligned} \tag{4.1.31}$$

and integrating the equation we obtain

$$x^2(t) \le x^2(0) + \int_0^t a(s)\,ds + 2B(t)X(t) \quad \forall\, t > 0. \tag{4.1.32}$$

Therefore, taking the supremum w.r.t. $t \in [0,T]$ we get

$$X^2(T) \le x^2(0) + A(T) + 2B(T)X(T), \tag{4.1.33}$$

and adding $B^2(T)$ to both sides gives

$$X(T) \le B(T) + \sqrt{B^2(T) + x^2(0) + A(T)} \le 2B(T) + \sqrt{x^2(0) + A(T)}.$$

Recalling (4.1.31) we obtain (4.1.29). □

4.2 Convergence of discrete solutions

4.2.1 Convergence when the initial datum $u_0 \in D(\phi)$

The previous Corollaries 4.1.5, 4.1.7 show the importance to obtain a priori bounds of the integral of $\mathscr{R}_{\boldsymbol{\tau}}, \mathscr{D}_{\boldsymbol{\tau}}$, and $\mathscr{D}^2_{\boldsymbol{\tau}}$. In this section we mainly focus our attention on the convergence of the discrete solutions, by quickly deriving rough estimates of these integrals and we postpone a finer analysis of the error to Section 4.4. It is not restrictive to assume $\lambda \le 0$.

Lemma 4.2.1. *Let us suppose that the convexity Assumption* 4.0.1 *holds with* $\lambda \le 0$, *let* $\mathscr{R}_{\boldsymbol\tau}, \mathscr{D}_{\boldsymbol\tau}$ *be the residual terms associated to a discrete solution* $\{U^n_{\boldsymbol\tau}\}_{n=0}^{+\infty}$ *defined as in* (4.1.9a), (4.1.22), *and let us choose* T *in the interval* $I^N_{\boldsymbol\tau} = (t^{N-1}_{\boldsymbol\tau}, t^N_{\boldsymbol\tau}]$. *Then*

$$\int_0^T e^{2\lambda t}\Big([\mathscr{R}_{\boldsymbol\tau}(t)]^+ - \lambda \mathscr{D}^2_{\boldsymbol\tau}(t)\Big)\,dt \le |\boldsymbol\tau|\Big(\phi(U^0_{\boldsymbol\tau}) - \phi(U^N_{\boldsymbol\tau})\Big), \tag{4.2.1}$$

$$\Big(\int_0^T |\lambda| e^{\lambda t}\mathscr{D}_{\boldsymbol\tau}(t)\,dt\Big)^2 \le \frac{1}{2}\int_0^T |\lambda| e^{2\lambda t}\mathscr{D}^2_{\boldsymbol\tau}(t)\,dt \tag{4.2.2}$$

$$\le \frac{|\lambda|\,|\boldsymbol\tau|^2}{3}\Big(\phi(U^0_{\boldsymbol\tau}) - \phi(U^N_{\boldsymbol\tau})\Big). \tag{4.2.3}$$

Proof. First of all we observe that

$$\int_{I^n_{\boldsymbol\tau}} \big[\mathscr{R}_{\boldsymbol\tau}(t)\big]^+\,dt \le \tau_n\Big(\phi(U^{n-1}_{\boldsymbol\tau}) - \phi(U^n_{\boldsymbol\tau}) - \frac{d^2(U^n_{\boldsymbol\tau}, U^{n-1}_{\boldsymbol\tau})}{2\tau_n}\Big), \tag{4.2.4}$$

which is a direct consequence of (4.1.9a) and

$$\phi(U^{n-1}_{\boldsymbol\tau}) - \phi(U^n_{\boldsymbol\tau}) - \frac{d^2(U^n_{\boldsymbol\tau}, U^{n-1}_{\boldsymbol\tau})}{2\tau_n} \ge 0, \qquad \int_{I^n_{\boldsymbol\tau}} (1-\ell_{\boldsymbol\tau}(t))\,dt = \int_{I^n_{\boldsymbol\tau}} \ell_{\boldsymbol\tau}(t)\,dt = \frac{1}{2}.$$

Since

$$\int_{I^n_{\boldsymbol\tau}} (1-\ell_{\boldsymbol\tau}(t))^2\,dt = \frac{1}{3}\tau_n,$$

and

$$\int_{I^n_{\boldsymbol\tau}} |\lambda|\mathscr{D}^2_{\boldsymbol\tau}(t)\,dt \le \frac{1}{3}|\lambda|\tau_n d^2(U^n_{\boldsymbol\tau}, U^{n-1}_{\boldsymbol\tau}) \le \frac{1}{3}d^2(U^n_{\boldsymbol\tau}, U^{n-1}_{\boldsymbol\tau}), \tag{4.2.5}$$

from (4.2.4) we get

$$\int_{I^n_{\boldsymbol\tau}} e^{2\lambda t}\Big[\big(\mathscr{R}_{\boldsymbol\tau}(t)\big)^+ + |\lambda|\mathscr{D}^2_{\boldsymbol\tau}(t)\Big]\,dt \le \tau_n\big(\phi(U^{n-1}_{\boldsymbol\tau}) - \phi(U^n_{\boldsymbol\tau})\big) \tag{4.2.6}$$

which yields (4.2.1). Starting from (4.2.5) and recalling (3.2.8) we obtain

$$\int_0^T |\lambda|\mathscr{D}^2_{\boldsymbol\tau}(t)\,dt \le \frac{2}{3}|\lambda|\,|\boldsymbol\tau|^2(\phi(U^0_{\boldsymbol\tau}) - \phi(U^N_{\boldsymbol\tau})),$$

so that

$$\Big(\int_0^T |\lambda| e^{\lambda t}\mathscr{D}_{\boldsymbol\tau}(t)\,dt\Big)^2 \le \int_0^T |\lambda| e^{2\lambda t}\,dt \int_0^T |\lambda|\mathscr{D}^2_{\boldsymbol\tau}(t)\,dt \le \frac{|\lambda|\,|\boldsymbol\tau|^2}{3}\Big(\phi(U^0_{\boldsymbol\tau}) - \phi(U^N_{\boldsymbol\tau})\Big),$$

which yields (4.2.2) and (4.2.3). □

Theorem 4.2.2. *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold for* $\lambda \in \mathbb{R}$ *and*

$$\lim_{|\boldsymbol{\tau}|\downarrow 0} d(U^0_{\boldsymbol{\tau}}, u_0) = 0, \quad \sup_{\boldsymbol{\tau}} \phi(U^0_{\boldsymbol{\tau}}) = S < +\infty. \tag{4.2.7}$$

Then the family $\{\overline{U}_{\boldsymbol{\tau}}\}_{\boldsymbol{\tau}}$ *of the discrete solutions generated by* $U^0_{\boldsymbol{\tau}}$ *is convergent to a function* u *as* $|\boldsymbol{\tau}| \downarrow 0$, *uniformly in each bounded interval* $[0,T]$*; in particular* u *is the unique element of* $MM(\Phi; u_0)$.

Proof. We fix a time $t \in [0,T]$ and we prove that $\{\overline{U}_{\boldsymbol{\tau}}(t)\}_{\boldsymbol{\tau}}$ is a Cauchy family as $|\boldsymbol{\tau}|$ goes to 0. We already know from the a priori estimates of Lemma 3.2.2 that there exists a constant C dependent on S, T, λ but independent of $\boldsymbol{\tau}$ such that

$$d^2(\overline{U}_{\boldsymbol{\tau}}(t), \underline{U}_{\boldsymbol{\tau}}(t)) \le C\,|\boldsymbol{\tau}|, \quad \phi(U^0_{\boldsymbol{\tau}}) - \phi(U^n_{\boldsymbol{\tau}}) \le C \quad 1 \le n \le N, \tag{4.2.8}$$

for the integer N such that the interval $I^N_{\boldsymbol{\tau}}$ contains T. Moreover, choosing two partitions $\mathcal{P}_{\boldsymbol{\tau}}, \mathcal{P}_{\boldsymbol{\eta}}$ as in Corollary 4.1.7, by (4.1.14) we have

$$\begin{aligned} d^2(\overline{U}_{\boldsymbol{\tau}}(t), \overline{U}_{\boldsymbol{\eta}}(t)) &\le 3d^2_{\boldsymbol{\tau\eta}}(t,t) + 3d^2(\underline{U}_{\boldsymbol{\tau}}(t), \overline{U}_{\boldsymbol{\tau}}(t)) + 3d^2(\underline{U}_{\boldsymbol{\eta}}(t), \overline{U}_{\boldsymbol{\eta}}(t)) \\ &\le 3d^2_{\boldsymbol{\tau\eta}}(t,t) + 3C(|\boldsymbol{\tau}| + |\boldsymbol{\eta}|), \end{aligned}$$

therefore we simply have to show that $\lim_{|\boldsymbol{\tau}|,|\boldsymbol{\eta}|\downarrow 0} d_{\boldsymbol{\tau\eta}}(t,t) = 0$. By (4.1.26), (4.2.1), and (4.2.3) we obtain

$$e^{2\lambda t} d^2_{\boldsymbol{\tau\eta}}(t,t) \le 2d^2(U^0_{\boldsymbol{\tau}}, U^0_{\boldsymbol{\eta}}) + 2C\big(|\boldsymbol{\tau}| + |\boldsymbol{\eta}|\big) + 2|\lambda| C(|\boldsymbol{\tau}|^2 + |\boldsymbol{\eta}|^2), \tag{4.2.9}$$

and this conclude the proof of the convergence; since the constant C in the bound (4.2.9) is independent of t, the convergence is also uniform in $[0,T]$.
Finally, it is easy to check that the limit does not depend on the particular family of initial data $(U^0_{\boldsymbol{\tau}})$ satisfying (4.2.7): if $(V^0_{\boldsymbol{\tau}})$ is another sequence approximating u_0, we can apply the same convergence result to a third family $(W^0_{\boldsymbol{\tau}})$ which coincides with the previous ones along two different subsequences of step sizes $\boldsymbol{\tau}_n$, $\boldsymbol{\tau}'_n$ with $|\boldsymbol{\tau}_n|, |\boldsymbol{\tau}'_n| \downarrow 0$ as $n \to \infty$. □

Corollary 4.2.3. *Under the same assumption of the previous Theorem, let* $u = MM(\Phi; u_0)$ *and let* $\overline{U}_{\boldsymbol{\tau}}$ *be the discrete solution associated to the partition* $\mathcal{P}_{\boldsymbol{\tau}}$. *Then if* $T \in \mathcal{P}_{\boldsymbol{\tau}}$ *and* $\lambda = 0$ *we have*

$$d^2(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \le d^2(U^0_{\boldsymbol{\tau}}, u_0) + \int_0^T \mathscr{R}_{\boldsymbol{\tau}}(t)\,dt, \tag{4.2.10}$$

whereas for $\lambda \ne 0$ *we have*

$$\begin{aligned} e^{\lambda T} d(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \le &\left(d^2(U^0_{\boldsymbol{\tau}}, u_0) + \mathsf{R}_{\boldsymbol{\tau}}(T) + \int_0^T e^{2\lambda t} \lambda^- \mathscr{D}^2_{\boldsymbol{\tau}}(t)\,dt \right)^{1/2} \\ &+ 2\int_0^T |\lambda| e^{\lambda t} \mathscr{D}_{\boldsymbol{\tau}}(t)\,dt, \end{aligned} \tag{4.2.11}$$

where $\mathsf{R}_{\boldsymbol{\tau}}$ *is defined by* (4.1.27).

Proof. We simply pass to the limit as $|\boldsymbol{\eta}| \downarrow 0$ in (4.1.16) or (4.1.26), observing that the integrals of $\big(\mathscr{R}_{\boldsymbol{\eta}}\big)^+, \mathscr{D}_{\boldsymbol{\eta}}, \mathscr{D}^2_{\boldsymbol{\eta}}$ are infinitesimal by the estimates of Lemma 4.2.1; on the other hand, by (4.2.8) we have for $T \in \mathcal{P}_{\boldsymbol{\tau}}$

$$\lim_{|\boldsymbol{\eta}|\downarrow 0} d_{\boldsymbol{\tau\eta}}(T,T) = d_{\boldsymbol{\tau}}(T, u(T)), \quad \text{and} \quad d_{\boldsymbol{\tau}}(T,u(T)) = d(\overline{U}_{\boldsymbol{\tau}}(T), u(T)). \qquad \square$$

4.2.2 Convergence when the initial datum $u_0 \in \overline{D(\phi)}$.

Now we conclude the proof of (4.0.11) in the statement of Theorem 4.0.4 when the starting point belongs to the closure in $\mathscr{S}$ of the proper domain of ϕ: in this case, it is more difficult to exhibit an explicit order of convergence for the approximate solutions and we have to take care of the loss of regularity of the initial datum.

Let us start with a comparison result between two discrete solutions related to the same partition $\mathcal{P}_{\boldsymbol{\tau}}$:

Lemma 4.2.4. *Let $\overline{U}_{\boldsymbol{\tau}}, \overline{V}_{\boldsymbol{\tau}}$ be discrete solutions associated to the same choice of step size $\boldsymbol{\tau}$ and to the initial values $U^0_{\boldsymbol{\tau}} \in \overline{D(\phi)}, V^0_{\boldsymbol{\tau}} \in D(\phi)$ respectively. If $T \in I^N_{\boldsymbol{\tau}} = (t^{N-1}_{\boldsymbol{\tau}}, t^N_{\boldsymbol{\tau}}]$, and $\lambda_{\boldsymbol{\tau}}$ is defined in (4.0.25), then for $-1 < \lambda|\boldsymbol{\tau}| \leq 0$ we have*

$$\begin{aligned} e^{2\lambda_{\boldsymbol{\tau}}(T+|\boldsymbol{\tau}|)} d^2(\overline{U}_{\boldsymbol{\tau}}(T), \overline{V}_{\boldsymbol{\tau}}(T)) &\leq e^{2\lambda_{\boldsymbol{\tau}} t^N_{\boldsymbol{\tau}}} d^2(U^N_{\boldsymbol{\tau}}, V^N_{\boldsymbol{\tau}}) \\ &\leq d^2(U^0_{\boldsymbol{\tau}}, V^0_{\boldsymbol{\tau}}) + 2|\boldsymbol{\tau}|\Big(\phi(V^0_{\boldsymbol{\tau}}) - \phi(V^N_{\boldsymbol{\tau}})\Big). \end{aligned} \tag{4.2.12}$$

Proof. Choosing $V := V^{n-1}_{\boldsymbol{\tau}}$ in (4.1.3) and multiplying the inequality by $2\tau_n$ we obtain

$$\begin{aligned} d^2(U^n_{\boldsymbol{\tau}}, V^{n-1}_{\boldsymbol{\tau}}) - d^2(U^{n-1}_{\boldsymbol{\tau}}, V^{n-1}_{\boldsymbol{\tau}}) &\leq 2\tau_n\phi(V^{n-1}_{\boldsymbol{\tau}}) - 2\tau_n\phi(U^n_{\boldsymbol{\tau}}) - d^2(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}}) \\ &- \lambda\tau_n d^2(U^n_{\boldsymbol{\tau}}, V^{n-1}_{\boldsymbol{\tau}}). \end{aligned}$$

Analogously, we choose $V := U^n_{\boldsymbol{\tau}}$ in the discrete inequality (4.1.3) written for the discrete solution $\{V^n_{\boldsymbol{\tau}}\}_{n=0}^{+\infty}$ obtaining

$$(1+\lambda\tau_n)d^2(V^n_{\boldsymbol{\tau}}, U^n_{\boldsymbol{\tau}}) - d^2(V^{n-1}_{\boldsymbol{\tau}}, U^n_{\boldsymbol{\tau}}) \leq 2\tau_n\phi(U^n_{\boldsymbol{\tau}}) - 2\tau_n\phi(V^n_{\boldsymbol{\tau}}) - d^2(V^n_{\boldsymbol{\tau}}, V^{n-1}_{\boldsymbol{\tau}}).$$

Recalling the elementary inequality $(a+b)^2 \leq \varepsilon^{-1}a^2 + (1-\varepsilon)^{-1}b^2$, $0 < \varepsilon < 1$, choosing $\varepsilon := -\lambda\tau_n$ we get

$$-\lambda\tau_n d^2(U^n_{\boldsymbol{\tau}}, V^{n-1}_{\boldsymbol{\tau}}) \leq d^2(U^n_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}}) - \frac{\lambda\tau_n}{1+\lambda\tau_n} d^2(U^{n-1}_{\boldsymbol{\tau}}, V^{n-1}_{\boldsymbol{\tau}});$$

summing up the previous inequalities we obtain

$$(1+\lambda\tau_n)d^2(V^n_{\boldsymbol{\tau}}, U^n_{\boldsymbol{\tau}}) - \frac{1}{1+\lambda\tau_n} d^2(U^{n-1}_{\boldsymbol{\tau}}, V^{n-1}_{\boldsymbol{\tau}}) \leq 2\tau_n\Big(\phi(V^{n-1}_{\boldsymbol{\tau}}) - \phi(V^n_{\boldsymbol{\tau}})\Big).$$

Multiplying the inequality by $e^{\lambda_{\boldsymbol{\tau}}(2t^{n-1}_{\boldsymbol{\tau}}+\tau_n)} < 1$ and recalling that $\phi(V^{n-1}_{\boldsymbol{\tau}}) \geq \phi(V^n_{\boldsymbol{\tau}})$, we get by (3.4.10)

$$e^{2\lambda_{\boldsymbol{\tau}} t^n_{\boldsymbol{\tau}}} d^2(V^n_{\boldsymbol{\tau}}, U^n_{\boldsymbol{\tau}}) \leq e^{2\lambda_{\boldsymbol{\tau}} t^{n-1}_{\boldsymbol{\tau}}} d^2(V^{n-1}_{\boldsymbol{\tau}}, U^{n-1}_{\boldsymbol{\tau}}) + 2\tau_n\Big(\phi(V^{n-1}_{\boldsymbol{\tau}}) - \phi(V^n_{\boldsymbol{\tau}})\Big).$$

Summing these inequalities from $n = 1$ to N we get (4.2.12). □

The following Corollary extends the previous Theorem 4.2.2 and concludes the *proof* of the convergence part of Theorem 4.0.4:

Corollary 4.2.5. *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold for* $\lambda \in \mathbb{R}$ *and*

$$U^0_{\boldsymbol{\tau}} \in \overline{D(\phi)}, \qquad \lim_{|\boldsymbol{\tau}|\downarrow 0} d(U^0_{\boldsymbol{\tau}}, u_0) = 0. \tag{4.2.13}$$

The family $\{\overline{U}_{\boldsymbol{\tau}}\}_{\boldsymbol{\tau}}$ *of the discrete solutions generated by* $U^0_{\boldsymbol{\tau}}$ *is convergent to the function* $u = S[u_0]$ *as* $|\boldsymbol{\tau}| \downarrow 0$ *defined by Corollary* 4.3.3*, uniformly in each bounded interval* $[0, T]$*; in particular* u *is the unique element of* $MM(\Phi; u_0)$.

Proof. It is not restrictive to assume $\lambda \le 0$. Let $\overline{U}_{\boldsymbol{\tau}}, \overline{U}_{\boldsymbol{\eta}}$ be two discrete solutions corresponding to the admissible partitions $\mathcal{P}_{\boldsymbol{\tau}}, \mathcal{P}_{\boldsymbol{\eta}}$, let us choose an arbitrary initial datum $v_0 \in D(\phi)$, and let us introduce the correspondent discrete solutions $\overline{V}_{\boldsymbol{\tau}}, \overline{V}_{\boldsymbol{\eta}}$ associated to the same partitions $\mathcal{P}_{\boldsymbol{\tau}}, \mathcal{P}_{\boldsymbol{\eta}}$ with $V^0_{\boldsymbol{\tau}} = V^0_{\boldsymbol{\eta}} = v_0$.

Applying the previous Lemma 4.2.4 we get

$$\begin{aligned} d(\overline{U}_{\boldsymbol{\tau}}(t), \overline{U}_{\boldsymbol{\eta}}(t)) &\le d(\overline{U}_{\boldsymbol{\tau}}(t), \overline{V}_{\boldsymbol{\tau}}(t)) + d(\overline{V}_{\boldsymbol{\tau}}(t), \overline{V}_{\boldsymbol{\eta}}(t)) + d(\overline{V}_{\boldsymbol{\eta}}(t), \overline{U}_{\boldsymbol{\eta}}(t)) \\ &\le e^{-\lambda_{\boldsymbol{\tau}}(t+|\boldsymbol{\tau}|)} \Big[d^2(v_0, U^0_{\boldsymbol{\tau}}) + 2|\boldsymbol{\tau}|[\phi(v_0) - \phi(\overline{V}_{\boldsymbol{\tau}}(t))] \Big]^{1/2} \\ &\quad + e^{-\lambda_{\boldsymbol{\eta}}(t+|\boldsymbol{\eta}|)} \Big[d^2(v_0, U^0_{\boldsymbol{\eta}}) + 2|\boldsymbol{\tau}|[\phi(v_0) - \phi(\overline{V}_{\boldsymbol{\eta}}(t))] \Big]^{1/2} + d(\overline{V}_{\boldsymbol{\tau}}(t), \overline{V}_{\boldsymbol{\eta}}(t)). \end{aligned}$$

Since $v_0 \in D(\phi)$, passing to the limit as $|\boldsymbol{\tau}|, |\boldsymbol{\eta}| \downarrow 0$ and applying Theorem 4.2.2, we get

$$\limsup_{|\boldsymbol{\tau}|,|\boldsymbol{\eta}|\downarrow 0} d(\overline{U}_{\boldsymbol{\tau}}(t), \overline{U}_{\boldsymbol{\eta}}(t)) \le 2e^{-\lambda t} d(u_0, v_0) \quad \forall\, v_0 \in D(\phi).$$

Since $u_0 \in \overline{D(\phi)}$, taking the infimum with respect to v_0 we conclude. □

4.3 Regularizing effect, uniqueness and the semigroup property

The λ-contractivity property is an immediate consequence of Lemma 4.2.4:

Proposition 4.3.1. *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold,* $\lambda \in \mathbb{R}$*. If* $u_0, v_0 \in \overline{D(\phi)}$ *and* $u = MM(u_0; \Phi), v = MM(v_0; \Phi)$*, then*

$$d(u(t), v(t)) \le e^{-\lambda t} d(u_0, v_0). \tag{4.3.1}$$

Proof. If $v_0 \in D(\phi)$, we can simply pass to the limit as $|\boldsymbol{\tau}| \downarrow 0$ in (4.2.12), choosing e.g. $U^0_{\boldsymbol{\tau}} = u_0, V^0_{\boldsymbol{\tau}} = v_0$.

When $v_0 \in \overline{D(\phi)} \setminus D(\phi)$, we consider an auxiliary initial datum $w_0 \in D(\phi)$ and the Minimizing Movement $w = MM(w_0;\Phi)$, obtaining by the triangular inequality

$$d(u(t),v(t)) \le d(u(t),w(t)) + d(w(t),v(t)) \le e^{-\lambda t}\big(d(u_0,w_0)+d(w_0,v_0)\big).$$

(4.3.1) follows now by taking the infimum of the right hand member of the previous inequality w.r.t. $w_0 \in D(\phi)$. □

Theorem 4.3.2. *Suppose that* (4.0.1) *and the convexity Assumption* 4.0.1 *hold,* $\lambda \in \mathbb{R}$. *If* $u \in MM(u_0;\Phi)$ *then* u *satisfies* (4.0.13). *In particular, setting*

$$\iota(T) := \int_0^T e^{\lambda t}\,dt = \frac{e^{\lambda T}-1}{T}, \tag{4.3.2}$$

we have

$$\phi(u(T)) \le \frac{1}{\iota(T)}\int_0^T \phi(u(t))\,e^{\lambda t}dt \le \phi_{\iota(T)}(u_0), \tag{4.3.3}$$

and, if $\lambda \ge 0$,

$$\begin{aligned} |\partial\phi|(u(T)) &\le \frac{1}{T} d(u_0,u(T)),\\ |\partial\phi|^2(u(T)) &\le |\partial\phi|^2(V) + \frac{1}{T^2} d^2(V,u_0) \quad \forall\, V \in D(|\partial\phi|). \end{aligned} \tag{4.3.4}$$

Proof. By a simple approximation argument via the λ-contraction property of Proposition 4.3.1 and the lower semicontinuity of ϕ, it is not restrictive to assume $u_0 \in D(\phi)$. In this case, we already know from Theorem 2.4.15 that u is locally Lipschitz in $(0,+\infty)$. Keeping the same notation of Section 4.1.2, observe that

$$\lim_{|\boldsymbol{\tau}|\downarrow 0} d_{\boldsymbol{\tau}}(t,V) = d(u(t),V), \quad \lim_{|\boldsymbol{\tau}|\downarrow 0} \varphi_{\boldsymbol{\tau}}(t) = \phi(u(t)) \quad \forall\, t \ge 0,\ V \in \mathscr{S}.$$

Integrating (4.1.12) from S to T and passing to the limit as $|\boldsymbol{\tau}| \downarrow 0$ gives

$$\frac{1}{2}d^2(u(T),V) - \frac{1}{2}d^2(u(S),V) + \int_S^T \Big(\phi(u(t)) + \frac{\lambda}{2}d^2(u(t),V)\Big)\,dt \le (T-S)\phi(V) \tag{4.3.5}$$

which easily yields (4.0.13). Moreover, multiplying (4.0.13) by $e^{\lambda t}$ and integrating from 0 to T, since $t \mapsto \phi(u(t))$ is decreasing we have

$$\iota(T)\phi(u(T)) \le \int_0^T \phi(u(t))\,e^{\lambda t}dt \le \iota(T)\phi(V) + \frac{1}{2}d^2(u_0,V) - \frac{e^{\lambda T}}{2}d^2(u(T),V)$$

for any $V \in D(\phi)$. Taking the infimum w.r.t. V we get (4.3.3). Finally, if $\lambda = 0$, multiplying (2.4.26) by t and integrating in time we get

$$\begin{aligned}\frac{T^2}{2}|\partial\phi|^2(u(T)) &\leq \int_0^T t|\partial\phi|^2(u(t))\,dt \leq -\int_0^T t\big(\phi(u(t))\big)'\,dt \\ &= \int_0^T \phi(u(t))\,dt - T\phi(u(T)) \\ &\leq T\phi(V) + \frac{1}{2}d^2(u_0,V) - T\phi(u(T)) - \frac{1}{2}d^2(u(T),V).\end{aligned}$$

Choosing $V := u(T)$ yields the first estimate of (4.3.4); on the other hand, if $V \in D(|\partial\phi|)$ the right hand side of the last formula can be bounded by

$$T|\partial\phi|(V)d(V,u(T)) - \frac{1}{2}d^2(u(T),V) + \frac{1}{2}d^2(u_0,V) \leq \frac{T^2}{2}|\partial\phi|^2(V) + \frac{1}{2}d^2(u_0,V),$$

which gives the second inequality of (4.3.4). □

Corollary 4.3.3. *The λ-contractive map $u_0 \mapsto S[u_0](t)$, $S[u_0]$ being the Minimizing movement $MM(u_0;\Phi)$, provides the unique solution of the evolution variational inequality* (4.0.13)*, and it satisfies the semigroup property $S[u_0](t+s) = S[S[u_0](t)](s)$ for every choice of $t, s \geq 0$.*

Proof. Let us first observe that if u is a continuous solution of the system (4.0.13), then an integration from $t-h$ to t gives for every $v \in D(\phi)$

$$\frac{1}{2}d^2(u(t),v) + \frac{1}{2}d^2(u(t-h),v) + \int_{t-h}^t \Big(\frac{\lambda}{2}d^2(u(r),v) + \phi(u(r))\Big)\,dr \leq h\phi(v).$$

Dividing by h and passing to the limit as $h \downarrow 0$, the lower semicontinuity of ϕ and Fatou's Lemma yield

$$\begin{aligned}\limsup_{h\downarrow 0} h^{-1}\Big(\frac{1}{2}d^2(u(t),v) - \frac{1}{2}d^2(u(t-h),v)\Big)& \\ + \frac{\lambda}{2}d^2(u(t),v) + \phi(u(t)) \leq \phi(v) \quad \forall\, t > 0.&\end{aligned} \tag{4.3.6}$$

By the same argument we also get the analogous pointwise estimate for the right derivative

$$\begin{aligned}\limsup_{h\downarrow 0} h^{-1}\Big(\frac{1}{2}d^2(u(t+h),v) - \frac{1}{2}d^2(u(t),v)\Big)& \\ + \frac{\lambda}{2}d^2(u(t),v) + \phi(u(t)) \leq \phi(v) \quad \forall\, t > 0.&\end{aligned} \tag{4.3.7}$$

Let now $u, w \in AC_{\rm loc}(0,+\infty;\mathscr{S})$ be two curves valued in $D(\phi)$ which satisfy the system (4.0.13) and take (by continuity as $t \downarrow 0$) the initial values $u_0, w_0 \in \overline{D(\phi)}$.

Choosing $v := w(t)$ in (4.3.6), $v := u(t)$ in the analogous inequality (4.3.7) written for the function w, and applying the next lemma we find that

$$\frac{d}{dt}d^2(u(t),w(t)) + 2\lambda d^2(u(t),w(t)) \le 0 \quad \text{for } \mathscr{L}^1\text{-a.e. } t>0,$$

i.e.

$$\frac{d}{dt}e^{2\lambda t}d^2(u(t),w(t)) \le 0, \quad d^2(u(t),w(t)) \le e^{-2\lambda t}d^2(u_0,w_0) \quad \forall\, t>0.$$

In particular, if $u_0 = w_0$ the functions u, w coincides and therefore the system (4.0.13) has at most one solution for a given initial datum u_0.

Since the curve $u(t) := S[u_0](t)$, defined as the value at t of $u \in MM(u_0;\Phi)$ for $u_0 \in D(\phi)$, solves (4.0.13), we obtain that u is the unique solution of (4.0.13). The semigroup property follows easily be the uniqueness for solutions of (4.0.13). □

The following elementary lemma is stated just for convenience for functions in the unit interval $(0,1)$.

Lemma 4.3.4. *Let $d(s,t) : (0,1)^2 \to \mathbb{R}$ be a map satisfying*

$$|d(s,t) - d(s',t)| \le |v(s) - v(s')|, \quad |d(s,t) - d(s,t')| \le |v(t) - v(t')|$$

for any s, t, s', $t' \in (0,1)$, for some locally absolutely continuous map $v : (0,1) \to \mathbb{R}$ and let $\delta(t) := d(t,t)$. Then δ is locally absolutely continuous in $(0,1)$ and

$$\frac{d}{dt}\delta(t) \le \limsup_{h\downarrow 0}\frac{d(t,t)-d(t-h,t)}{h} + \limsup_{h\downarrow 0}\frac{d(t,t+h)-d(t,t)}{h} \qquad \mathscr{L}^1\text{-a.e. in } (0,1)$$

Proof. Since $|\delta(s) - \delta(t)| \le 2|v(s) - v(t)|$ the function δ is locally absolutely continuous. We fix a nonnegative function $\zeta \in C_c^\infty(0,1)$ and $h > 0$ such that $\pm h + \operatorname{supp}\zeta \subset (0,1)$. We have then

$$\begin{aligned} -\int_0^1 \delta(t)\frac{\zeta(t+h)-\zeta(t)}{h}\,dt &= \int_0^1 \zeta(t)\frac{d(t,t)-d(t-h,t-h)}{h}\,dt \\ = \int_0^1 \zeta(t)\frac{d(t,t)-d(t-h,t)}{h}\,dt &+ \int_0^1 \zeta(t+h)\frac{d(t,t+h)-d(t,t)}{h}\,dt, \end{aligned}$$

where the last equality follows by adding and subtracting $d(t-h,t)$ and then making a change of variables in the last integral. Since

$$h^{-1}\,|d(t,t) - d(t-h,t)| \le h^{-1}\,|v(t) - v(t-h)| \to |v'(t)| \quad \text{in } L^1_{\rm loc}(0,1) \text{ as } h\downarrow 0$$

and an analogous inequality holds for the other difference quotient, we can apply (an extended version of) Fatou's Lemma and pass to the (superior) limit in the integrals as $h \downarrow 0$; denoting by a and b the two upper derivatives in the statement of the Lemma we get $-\int \delta\zeta'\,dt \le \int (a+b)\zeta\,dt$, whence the inequality between distributions follows. □

4.4 Optimal error estimates

4.4.1 The case $\lambda = 0$

In this section we mainly focus our attention on the case $\lambda = 0$ and we postpone the analysis of the other situation to Section 4.4.2.

Lemma 4.4.1. *Let us suppose that the convexity Assumption* 4.0.1 *holds for* $\lambda = 0$, *let* $\mathscr{R}_{\boldsymbol\tau}, \mathscr{E}^n_{\boldsymbol\tau}$ *be defined as in* (4.1.9a) *and* (4.0.20), *let* $I^n_{\boldsymbol\tau} := (t^{n-1}_{\boldsymbol\tau}, t^n_{\boldsymbol\tau}]$, *and let us define*

$$\mathscr{I}_{\boldsymbol\tau}(T) := \int_{t^{N-1}_{\boldsymbol\tau}}^T \mathscr{R}_{\boldsymbol\tau}(t)\,dt \quad \textit{for } T \in I^N_{\boldsymbol\tau} = (t^{N-1}_{\boldsymbol\tau}, t^N_{\boldsymbol\tau}]. \tag{4.4.1}$$

Then

$$\int_{I^n_{\boldsymbol\tau}} \mathscr{R}_{\boldsymbol\tau}(t)\,dt = \tau_n^2 \mathscr{E}^n_{\boldsymbol\tau}, \tag{4.4.2}$$

$$\mathscr{I}_{\boldsymbol\tau}(T) \le \tau_N\Big(\phi(U^{N-1}_{\boldsymbol\tau}) - \phi(U^N_{\boldsymbol\tau}) - \frac{1}{2}\tau_N D^N_{\boldsymbol\tau}\Big) \le \frac{1}{2}\tau_N^2 |\partial\phi|^2(U^{N-1}_{\boldsymbol\tau}), \tag{4.4.3}$$

$$\mathscr{E}^n_{\boldsymbol\tau} \le \frac{1}{2}\Big(|\partial\phi|^2(U^{n-1}_{\boldsymbol\tau}) - D^n_{\boldsymbol\tau}\Big) \le \frac{1}{2}\Big(|\partial\phi|^2(U^{n-1}_{\boldsymbol\tau}) - |\partial\phi|^2(U^n_{\boldsymbol\tau})\Big), \tag{4.4.4}$$

$$\int_0^T \mathscr{R}_{\boldsymbol\tau}(t)\,dt \le \sum_{n=1}^{N-1} \tau_n^2 \mathscr{E}^n_{\boldsymbol\tau} + \mathscr{I}_{\boldsymbol\tau}(T). \tag{4.4.5}$$

Proof. (4.4.2) follows directly from (4.1.9b) since

$$\int_{I^n_{\boldsymbol\tau}} (1 - \ell_{\boldsymbol\tau}(t))\,dt = \int_{I^n_{\boldsymbol\tau}} \ell_{\boldsymbol\tau}(t)\,dt = \frac{1}{2}, \quad \int_{I^n_{\boldsymbol\tau}} (1 - 2\ell_{\boldsymbol\tau}(t))\,dt = 0. \tag{4.4.6}$$

(4.2.4) and (4.1.10) yield (4.4.3) and (4.4.4); finally, (4.4.7) is a direct consequence of (4.4.2) and (4.4.1). □

Corollary 4.4.2. *Under the same assumption of the previous lemma, let us suppose that* $\lambda = 0$ *and* $U^0_{\boldsymbol\tau} \in D(\phi)$; *then we have*

$$\sum_{n=1}^{N-1} \tau_n^2 \mathscr{E}^n_{\boldsymbol\tau} + \mathscr{I}_{\boldsymbol\tau}(T) \le |\boldsymbol\tau|\Big\{\phi(U^0_{\boldsymbol\tau}) - \phi_T(U^0_{\boldsymbol\tau})\Big\} \le |\boldsymbol\tau|\Big\{\phi(U^0_{\boldsymbol\tau}) - \inf_{\mathscr{S}} \phi\Big\}, \tag{4.4.7}$$

and, if $U^0_{\boldsymbol\tau} \in D(|\partial\phi|)$,

$$\sum_{n=1}^{N-1} \tau_n^2 \mathscr{E}^n_{\boldsymbol\tau} + \mathscr{I}_{\boldsymbol\tau}(T) \le \frac{1}{2}|\boldsymbol\tau|^2\,|\partial\phi|^2(U^0_{\boldsymbol\tau}). \tag{4.4.8}$$

Moreover, when the partition $\mathcal{P}_{\boldsymbol\tau}$ *is* uniform *(i.e.* $\tau_n \equiv \tau = |\boldsymbol\tau|$ *is independent of* n, *cf. Remark* 2.0.3*), then the following sharper estimate holds, too:*

$$\int_0^T \mathscr{R}_{\boldsymbol\tau}(t)\,dt \le \sum_{n=1}^{N-1} \tau_n^2 \mathscr{E}^n_{\boldsymbol\tau} + \mathscr{I}_{\boldsymbol\tau}(T) \le \tau\Big\{\phi(U^0_{\boldsymbol\tau}) - \phi_\tau(U^0_{\boldsymbol\tau})\Big\} \le \frac{\tau^2}{2}\,|\partial\phi|^2(U^0_{\boldsymbol\tau}). \tag{4.4.9}$$

Proof. Since $\mathscr{E}^n_{\boldsymbol\tau} \geq 0$ by (4.1.10), we easily have

$$\begin{aligned}\sum_{n=1}^{N-1} \tau_n^2 \mathscr{E}^n_{\boldsymbol\tau} &\leq |\boldsymbol\tau| \sum_{n=1}^{N-1} \Big(\big(\phi(U^{n-1}_{\boldsymbol\tau}) - \phi(U^n_{\boldsymbol\tau})\big) - \tau_n D^n_{\boldsymbol\tau}\Big)\\ &\leq |\boldsymbol\tau| \sum_{n=1}^{N-1} \Big(\phi(U^{n-1}_{\boldsymbol\tau}) - \phi(U^n_{\boldsymbol\tau})\Big) - |\boldsymbol\tau| \sum_{n=1}^{N-1} \tau_n D^n_{\boldsymbol\tau}\\ &= |\boldsymbol\tau| \Big\{\phi(U^0_{\boldsymbol\tau}) - \phi(U^{N-1}_{\boldsymbol\tau}) - |\boldsymbol\tau| \sum_{n=1}^{N-1} \tau_n D^n_{\boldsymbol\tau}\Big\}.\end{aligned}$$

Summing up the contribution of $\mathscr{I}_{\boldsymbol\tau}(T)$ and recalling that

$$\sum_{n=1}^{N} \tau_n D^n_{\boldsymbol\tau} = \sum_{n=1}^{N} \frac{d^2(U^n_{\boldsymbol\tau}, U^{n-1}_{\boldsymbol\tau})}{\tau_n} \geq \frac{1}{T} d^2(U^0_{\boldsymbol\tau}, U^N_{\boldsymbol\tau}), \tag{4.4.10}$$

we obtain (4.4.7).

Since $n \mapsto |\partial\phi|^2(U^n_{\boldsymbol\tau})$ is decreasing, too, if $U^0_{\boldsymbol\tau} \in D(|\partial\phi|)$ then (4.4.4) yields

$$\begin{aligned}\sum_{n=1}^{N} \tau_n^2 \mathscr{E}^n_{\boldsymbol\tau} &\leq \frac{|\boldsymbol\tau|^2}{2} \sum_{n=1}^{N-1} \Big(|\partial\phi|^2(U^{n-1}_{\boldsymbol\tau}) - |\partial\phi|^2(U^{n-1}_{\boldsymbol\tau})\Big)\\ &\leq \frac{|\boldsymbol\tau|^2}{2} \Big(|\partial\phi|^2(U^0_{\boldsymbol\tau}) - |\partial\phi|^2(U^{N-1}_{\boldsymbol\tau})\Big) \leq \frac{|\boldsymbol\tau|^2}{2} |\partial\phi|^2(U^0_{\boldsymbol\tau}) - \mathscr{I}_{\boldsymbol\tau}(T),\end{aligned}$$

which proves (4.4.8). When $\tau_n \equiv \tau$ we can use a different estimate for $\mathscr{E}^n_{\boldsymbol\tau}$ which comes from (4.1.10)

$$\begin{aligned}\mathscr{E}^n_{\boldsymbol\tau} &\leq \tau^{-1}\Big(\big(\phi(U^{n-1}_{\boldsymbol\tau}) - \phi_\tau(U^{n-1}_{\boldsymbol\tau})\big) - \frac{1}{2}\tau_n D^n_{\boldsymbol\tau}\Big)\\ &\leq \tau^{-1}\Big(\big(\phi(U^{n-1}_{\boldsymbol\tau}) - \phi_\tau(U^{n-1}_{\boldsymbol\tau})\big) - \big(\phi(U^n_{\boldsymbol\tau}) - \phi_\tau(U^n_{\boldsymbol\tau})\big)\Big),\end{aligned} \tag{4.4.11}$$

thus obtaining

$$\begin{aligned}\sum_{n=1}^{N-1} \tau^2 \mathscr{E}^n_{\boldsymbol\tau} &\leq \tau \sum_{n=1}^{N-1} \Big(\big(\phi(U^{n-1}_{\boldsymbol\tau}) - \phi_\tau(U^{n-1}_{\boldsymbol\tau})\big) - \big(\phi(U^n_{\boldsymbol\tau}) - \phi_\tau(U^n_{\boldsymbol\tau})\big)\Big)\\ &\leq \tau\big(\phi(U^0_{\boldsymbol\tau}) - \phi_\tau(U^0_{\boldsymbol\tau})\big) - \tau\big(\phi(U^{N-1}_{\boldsymbol\tau}) - \phi_\tau(U^{N-1}_{\boldsymbol\tau})\big)\Big)\\ &\leq \tau\big(\phi(U^0_{\boldsymbol\tau}) - \phi_\tau(U^0_{\boldsymbol\tau})\big) - \mathscr{I}_{\boldsymbol\tau}(T),\end{aligned}$$

which proves (4.4.9). □

Corollary 4.4.3. *Suppose that the convexity Assumption* 4.0.1 *holds with* $\lambda \geq 0$. *Then the estimate* (4.0.15) *of Theorem* 4.0.4 *and all the estimates of Theorem* 4.0.7 *hold.*

Proof. We simply apply (4.2.10) and the results of the previous corollary. Observe that when $T = t^N_{\boldsymbol\tau} \in \mathcal{P}_{\boldsymbol\tau}$ then $\mathscr{I}_{\boldsymbol\tau}(T) = 0$, so that we have (4.0.19) without any correction term. □

4.4.2 The case $\lambda \neq 0$

First of all, let us observe that the first estimate (4.0.26) of Theorem 4.0.9 follows directly from Corollary 4.2.3 and (4.2.1), (4.2.3).

In order to get the other error bounds, we need refined estimates of the integral terms in the right-hand side of (4.2.11). Since $\lambda_{\boldsymbol{\tau}} \leq \lambda$, by replacing λ by $\lambda_{\boldsymbol{\tau}}$ in the left-hand side of the differential inequality (4.1.25), we easily get bounds analogous to (4.1.26) and (4.2.11) where the coefficient $\lambda_{\boldsymbol{\tau}}$ occours in each exponential term, thus obtaining for $U^0_{\boldsymbol{\tau}} = u_0$

$$\begin{aligned} e^{\lambda_{\boldsymbol{\tau}} T} d(\overline{U}_{\boldsymbol{\tau}}(T), u(T)) \leq \left(\mathsf{R}_{\boldsymbol{\tau}}(T) + \lambda^- \int_0^T e^{2\lambda_{\boldsymbol{\tau}} t} \mathscr{D}^2_{\boldsymbol{\tau}}(t)\, dt \right)^{1/2} \\ + 2 \int_0^T |\lambda| e^{\lambda_{\boldsymbol{\tau}} t} \mathscr{D}_{\boldsymbol{\tau}}(t)\, dt. \end{aligned} \tag{4.4.12}$$

Let us observe that if $T \in (t^{N-1}_{\boldsymbol{\tau}}, t^N_{\boldsymbol{\tau}}]$ for some $N \in \mathbb{N}$,

$$\mathsf{R}_{\boldsymbol{\tau}}(T) = \sup_{t \in [0,T]} \int_0^t e^{2\lambda_{\boldsymbol{\tau}} t} \mathscr{R}_{\boldsymbol{\tau}}(r)\, dr \tag{4.4.13a}$$

$$\leq \sup_{1 \leq M \leq N} \Big(\int_0^{t^{M-1}_{\boldsymbol{\tau}}} e^{2\lambda_{\boldsymbol{\tau}} t} \mathscr{R}_{\boldsymbol{\tau}}(r)\, dr + \int_{I^M_{\boldsymbol{\tau}}} e^{2\lambda_{\boldsymbol{\tau}} t} \big[\mathscr{R}_{\boldsymbol{\tau}}(r)\big]^+ dr \Big) \tag{4.4.13b}$$

$$\leq \sup_{1 \leq M \leq N} \Big(\sum_{n=1}^{M-1} \int_{I^n_{\boldsymbol{\tau}}} e^{2\lambda_{\boldsymbol{\tau}} t} \mathscr{R}_{\boldsymbol{\tau}}(r)\, dr + \int_{I^M_{\boldsymbol{\tau}}} e^{2\lambda_{\boldsymbol{\tau}} t} \big[\mathscr{R}_{\boldsymbol{\tau}}(r)\big]^+ dr \Big), \tag{4.4.13c}$$

and, recalling (4.1.11), the integral of the positive part of $\mathscr{R}_{\boldsymbol{\tau}}$ can be bounded by

$$\int_{I^M_{\boldsymbol{\tau}}} e^{2\lambda_{\boldsymbol{\tau}} t} \big[\mathscr{R}_{\boldsymbol{\tau}}(r)\big]^+ dr \leq \tau_M^2 \frac{\max\big[e^{2\lambda_{\boldsymbol{\tau}} t^{M-1}_{\boldsymbol{\tau}}}, e^{2\lambda_{\boldsymbol{\tau}} t^M_{\boldsymbol{\tau}}}\big]}{2(1+\lambda\tau_M)} |\partial\phi|^2(U^{M-1}_{\boldsymbol{\tau}}). \tag{4.4.14}$$

The next two lemmas provide the estimates of the other integral in the right-hand side of (4.4.13b) and of the integrals involving $\mathscr{D}_{\boldsymbol{\tau}}$ in (4.4.12). Combining these results with (4.4.12) we complete the proof of Theorems 4.0.9 and 4.0.10.

Proposition 4.4.4. *Suppose that $\lambda < 0$ and $U^0_{\boldsymbol{\tau}} \in D(\partial\phi)$; then for $T > 0$ we have*

$$\mathsf{R}_{\boldsymbol{\tau}}(T) \leq \frac{|\boldsymbol{\tau}|^2}{2(1+\lambda|\boldsymbol{\tau}|)} |\partial\phi|^2(U^0_{\boldsymbol{\tau}}), \tag{4.4.15}$$

and, recalling that $T_{\boldsymbol{\tau}} := \min\big\{t^k_{\boldsymbol{\tau}} \in \mathcal{P}_{\boldsymbol{\tau}} : t^k_{\boldsymbol{\tau}} \geq T\big\}$,

$$\begin{aligned} |\lambda| \int_0^T e^{2\lambda_{\boldsymbol{\tau}} t} \mathscr{D}^2_{\boldsymbol{\tau}}(t)\, dt &\leq |\boldsymbol{\tau}|^2 \frac{|\lambda|\, T_{\boldsymbol{\tau}}}{3(1+\lambda|\boldsymbol{\tau}|)^2} |\partial\phi|^2(U^0_{\boldsymbol{\tau}}), \\ 2|\lambda| \int_0^T e^{\lambda_{\boldsymbol{\tau}} t} \mathscr{D}_{\boldsymbol{\tau}}(t)\, dt &\leq |\boldsymbol{\tau}| \frac{|\lambda|\, T_{\boldsymbol{\tau}}}{1+\lambda|\boldsymbol{\tau}|} |\partial\phi|(U^0_{\boldsymbol{\tau}}). \end{aligned} \tag{4.4.16}$$

Proof. Let us suppose that $T \in I^N_{\boldsymbol\tau} = (t^{N-1}_{\boldsymbol\tau}, t^N_{\boldsymbol\tau}]$ so that $T_{\boldsymbol\tau} = t^N_{\boldsymbol\tau}$, and $1 \le M \le N$. Since

$$\int_{I^n_{\boldsymbol\tau}} e^{2\lambda_{\boldsymbol\tau} t}(1-\ell_{\boldsymbol\tau}(t))\,dt \le \frac{1}{2}\tau_n e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}, \tag{4.4.17}$$

$$\int_{I^n_{\boldsymbol\tau}} e^{2\lambda t}\ell_{\boldsymbol\tau}(t)\,dt \ge \frac{1}{2}\tau_n e^{\lambda(t^{n-1}_{\boldsymbol\tau}+t^n_{\boldsymbol\tau})} = \frac{1}{2}\tau_n e^{2\lambda t^{n-1}_{\boldsymbol\tau}}\, e^{\lambda\tau_n}, \tag{4.4.18}$$

recalling (4.1.11) and (3.4.10) we get

$$\begin{aligned}\int_{I^n_{\boldsymbol\tau}} e^{2\lambda_{\boldsymbol\tau} t}\mathscr{R}_{\boldsymbol\tau}(t)\,dt &\le \frac{\tau_n^2}{2(1+\lambda\tau_n)} e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}\Big\{|\partial\phi|^2(U^{n-1}_{\boldsymbol\tau}) - (1+\lambda\tau_n)e^{\lambda_{\boldsymbol\tau}\tau_n}|\partial\phi|^2(U^n_{\boldsymbol\tau})\Big\}\\ &\le \frac{\tau_n^2}{2(1+\lambda\tau_n)}\Big(e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}|\partial\phi|^2(U^{n-1}_{\boldsymbol\tau}) - e^{2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}|\partial\phi|^2(U^n_{\boldsymbol\tau})\Big).\end{aligned}$$

Since the map $n \mapsto e^{2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}|\partial\phi|^2(U^n_{\boldsymbol\tau})$ is decreasing, we get

$$\sum_{n=1}^{M-1}\int_{I^n_{\boldsymbol\tau}} e^{2\lambda_{\boldsymbol\tau} t}\mathscr{R}_{\boldsymbol\tau}(t)\,dt \le \frac{|\boldsymbol\tau|^2}{2(1+\lambda|\boldsymbol\tau|)}\Big(|\partial\phi|^2(U^0_{\boldsymbol\tau}) - e^{2\lambda_{\boldsymbol\tau} t^{M-1}_{\boldsymbol\tau}}|\partial\phi|^2(U^{M-1}_{\boldsymbol\tau})\Big).$$

Taking into account (4.4.14) we obtain (4.4.15). Finally, we easily have

$$\begin{aligned}|\lambda|\int_{I^n_{\boldsymbol\tau}} e^{2\lambda_{\boldsymbol\tau} t}\mathscr{D}^2_{\boldsymbol\tau}(t)\,dt &\le \frac{|\lambda|\tau_n}{3}e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}} d^2(U^n_{\boldsymbol\tau}, U^{n-1}_{\boldsymbol\tau})\\ &\le \frac{|\lambda|\tau_n^3}{3(1+\lambda\tau_n)^2}e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}|\partial\phi|^2(U^{n-1}_{\boldsymbol\tau}) \le \frac{|\lambda||\boldsymbol\tau|^2\tau_n}{3(1+\lambda|\boldsymbol\tau|)}|\partial\phi|^2(U^0_{\boldsymbol\tau}),\end{aligned}$$

and

$$\begin{aligned}2|\lambda|\int_{I^n_{\boldsymbol\tau}} e^{\lambda_{\boldsymbol\tau} t}\mathscr{D}_{\boldsymbol\tau}(t)\,dt &\le |\lambda|\tau_n e^{\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}} d(U^n_{\boldsymbol\tau}, U^{n-1}_{\boldsymbol\tau})\\ &\le \frac{|\lambda|\tau_n^2}{1+\lambda\tau_n}e^{\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}|\partial\phi|(U^{n-1}_{\boldsymbol\tau}) \le \frac{|\lambda||\boldsymbol\tau|\tau_n}{1+\lambda|\boldsymbol\tau|}|\partial\phi|(U^0_{\boldsymbol\tau}).\end{aligned}$$

Summing up all the contribution from $n = 1$ to N we obtain (4.4.16). □

Proposition 4.4.5. *Assume that* $\lambda > 0$, $\inf_{\mathscr{S}}\phi = 0$, $U^0_{\boldsymbol\tau} \in D(\phi)$, *and* $T_{\boldsymbol\tau}$ *is defined as in the above proposition. We have*

$$\mathsf{R}_{\boldsymbol\tau}(T) \le \int_0^T e^{2\lambda_{\boldsymbol\tau} t}\Big(\mathscr{R}_{\boldsymbol\tau}(t)\Big)^+\,dt \le |\boldsymbol\tau|(1+\lambda|\boldsymbol\tau|)(1+\lambda T_{\boldsymbol\tau})\phi(U^0_{\boldsymbol\tau}), \tag{4.4.19}$$

$$\int_0^T e^{\lambda_{\boldsymbol\tau} t}\mathscr{D}_{\boldsymbol\tau}(t)\,dt \le |\boldsymbol\tau|\Big(2T_{\boldsymbol\tau}(1+\lambda T_{\boldsymbol\tau})\phi(U^0_{\boldsymbol\tau})\Big)^{1/2}. \tag{4.4.20}$$

Moreover, if $U^0_{\boldsymbol\tau} \in D(|\partial\phi|)$ *then*

$$\mathsf{R}_{\boldsymbol\tau}(T) \le \frac{1}{2}|\boldsymbol\tau|^2(1+\lambda T_{\boldsymbol\tau})\,|\partial\phi|^2(U^0_{\boldsymbol\tau}), \tag{4.4.21}$$

$$2\int_0^T e^{\lambda_{\boldsymbol\tau} t}\mathscr{D}_{\boldsymbol\tau}(t)\,dt \le T_{\boldsymbol\tau}|\boldsymbol\tau||\partial\phi|(U^0_{\boldsymbol\tau}). \tag{4.4.22}$$

Proof. As before suppose that $T \in I^N_{\boldsymbol\tau} = (t^{N-1}_{\boldsymbol\tau}, t^N_{\boldsymbol\tau}]$. Since Lemma 2.4.13 yields

$$D^n_{\boldsymbol\tau} \ge |\partial\phi|^2(U^n_{\boldsymbol\tau}) \ge 2\lambda\phi(U^n_{\boldsymbol\tau}),$$

by (4.1.9a) and recalling (3.4.10) and (3.4.9), we get

$$\begin{aligned}
\int_{I^n_{\boldsymbol\tau}} e^{2\lambda_{\boldsymbol\tau} t}\Big(\mathscr{R}_{\boldsymbol\tau}(t)\Big)^+ dt &\le \tau_n e^{2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}\Big(\phi(U^{n-1}_{\boldsymbol\tau}) - (1+\lambda\tau_n)\phi(U^n_{\boldsymbol\tau})\Big)\\
&\le \tau_n e^{2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}(1+\lambda\tau_n)\Big(\frac{\lambda\tau_n}{(1+\lambda\tau_n)^2}\phi(U^{n-1}_{\boldsymbol\tau}) + \frac{1}{(1+\lambda\tau_n)^2}\phi(U^{n-1}_{\boldsymbol\tau}) - \phi(U^n_{\boldsymbol\tau})\Big)\\
&\le |\boldsymbol\tau|(1+\lambda|\boldsymbol\tau|)\Big(\lambda\tau_n e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}\phi(U^{n-1}_{\boldsymbol\tau}) + e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}\phi(U^{n-1}_{\boldsymbol\tau}) - e^{2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}\phi(U^n_{\boldsymbol\tau})\Big)\\
&\le |\boldsymbol\tau|(1+\lambda|\boldsymbol\tau|)\Big(\lambda\tau_n\phi(U^0_{\boldsymbol\tau}) + e^{2\lambda_{\boldsymbol\tau} t^{n-1}_{\boldsymbol\tau}}\phi(U^{n-1}_{\boldsymbol\tau}) - e^{2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}\phi(U^n_{\boldsymbol\tau})\Big).
\end{aligned}$$

Summing up for $n = 1$ to N we obtain

$$\int_0^T e^{2\lambda_{\boldsymbol\tau} t}\mathscr{R}_{\boldsymbol\tau}(t)\,dt \le |\boldsymbol\tau|(1+\lambda|\boldsymbol\tau|)(1+\lambda T_{\boldsymbol\tau})\phi(U^0_{\boldsymbol\tau}).$$

Moreover,

$$2\int_0^T e^{\lambda_{\boldsymbol\tau} t}\mathscr{D}_{\boldsymbol\tau}(t)\,dt \le \sum_{n=1}^N \tau_n^2 e^{\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}\sqrt{D^n_{\boldsymbol\tau}} \le \sqrt{2T_{\boldsymbol\tau}}\,|\boldsymbol\tau|\Big(\sum_{n=1}^N \tau_n e^{2\lambda_{\boldsymbol\tau} t^n_{\boldsymbol\tau}}\frac{D^n_{\boldsymbol\tau}}{2}\Big)^{1/2}$$

and

$$(1+\lambda\tau_n)\tau_n\frac{D^n_{\boldsymbol\tau}}{2} \le \Big(\phi(U^{n-1}_{\boldsymbol\tau}) - (1+\lambda\tau_n)\phi(U^n_{\boldsymbol\tau})\Big). \tag{4.4.23}$$

Arguing as before, we find

$$2\int_0^T e^{\lambda_{\boldsymbol\tau} t}\mathscr{D}_{\boldsymbol\tau}(t)\,dt \le |\boldsymbol\tau|\Big(2T_{\boldsymbol\tau}(1+\lambda T_{\boldsymbol\tau})\phi(U^0_{\boldsymbol\tau})\Big)^{1/2}.$$

Finally, if $U^0_{\boldsymbol\tau} \in D(|\partial\phi|)$, we first observe that

$$\int_{I^n_{\boldsymbol\tau}} e^{2\lambda_{\boldsymbol\tau} t}(1 - 2\ell_{\boldsymbol\tau}(t))\,dt \le 0, \tag{4.4.24}$$

so that by (4.1.9b) we have

$$\int_{I_{\boldsymbol{\tau}}^n} e^{2\lambda_{\boldsymbol{\tau}} t} \mathscr{R}_{\boldsymbol{\tau}}(t)\, dt \le \tau_n^2 e^{2\lambda_{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^n} \mathscr{E}_{\boldsymbol{\tau}}^n. \tag{4.4.25}$$

Since

$$\begin{aligned} 2\mathscr{E}_{\boldsymbol{\tau}}^n &\le \frac{1}{1+\lambda\tau_n}|\partial\phi|^2(U_{\boldsymbol{\tau}}^{n-1}) - |\partial\phi|^2(U_{\boldsymbol{\tau}}^n) \\ &\le \frac{\lambda\tau_n}{(1+\lambda\tau_n)^2}|\partial\phi|^2(U_{\boldsymbol{\tau}}^{n-1}) + \frac{1}{(1+\lambda\tau_n)^2}|\partial\phi|^2(U_{\boldsymbol{\tau}}^{n-1}) - |\partial\phi|^2(U_{\boldsymbol{\tau}}^n) \end{aligned}$$

we obtain

$$\begin{aligned} \int_{I_{\boldsymbol{\tau}}^n} e^{2\lambda_{\boldsymbol{\tau}} t} \mathscr{R}_{\boldsymbol{\tau}}(t)\, dt \quad &\le \quad \frac{1}{2}\tau_n^2\Big(e^{2\lambda_{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^{n-1}}|\partial\phi|^2(U_{\boldsymbol{\tau}}^{n-1}) - e^{2\lambda_{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^n}|\partial\phi|^2(U_{\boldsymbol{\tau}}^n)\Big) \\ &+ \quad \frac{\lambda}{2}\tau_n^3 e^{2\lambda_{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^{n-1}}|\partial\phi|^2(U_{\boldsymbol{\tau}}^{n-1}). \end{aligned}$$

Summing up from $n = 1$ to $M - 1$ and adding the contribution of the integral in the last interval $I_{\boldsymbol{\tau}}^M$ as in (4.4.14), by a repeated application of (4.1.10) we find

$$\begin{aligned} &\sum_{n=1}^{M-1} \int_{I_{\boldsymbol{\tau}}^n} e^{2\lambda_{\boldsymbol{\tau}} t}\mathscr{R}_{\boldsymbol{\tau}}(t)\, dt + \int_{I_{\boldsymbol{\tau}}^M} e^{2\lambda_{\boldsymbol{\tau}} t}\Big(\mathscr{R}_{\boldsymbol{\tau}}(t)\Big)^+ dt \\ &\quad\le \frac{|\boldsymbol{\tau}|^2}{2}\Big(|\partial\phi|^2(U_{\boldsymbol{\tau}}^0) - e^{2\lambda_{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^{M-1}}|\partial\phi|^2(U_{\boldsymbol{\tau}}^{M-1})\Big) + \frac{\lambda|\boldsymbol{\tau}|^2 t_{\boldsymbol{\tau}}^{M-1}}{2}|\partial\phi|^2(U_{\boldsymbol{\tau}}^0) \\ &\qquad + \tau_M^2\frac{e^{2\lambda_{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^{M-1}}}{2}|\partial\phi|^2(U_{\boldsymbol{\tau}}^{M-1}) + \tau_M^2\frac{\lambda\tau_M e^{2\lambda_{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^M}}{(1+\lambda_{\boldsymbol{\tau}}\tau_M)^2}|\partial\phi|(U_{\boldsymbol{\tau}}^{M-1}) \\ &\quad\le \frac{|\boldsymbol{\tau}|^2}{2}|\partial\phi|^2(U_{\boldsymbol{\tau}}^0)(1+\lambda t_{\boldsymbol{\tau}}^M), \end{aligned}$$

which yields (4.4.21). Analogously,

$$2\int_0^T e^{2\lambda_{\boldsymbol{\tau}} t}\mathscr{D}_{\boldsymbol{\tau}}(t)\, dt \le T_{\boldsymbol{\tau}}|\partial\phi|(U_{\boldsymbol{\tau}}^0),$$

which concludes the proof. □

Part II

Gradient Flow in the Space of Probability Measures

Chapter 5

Preliminary Results on Measure Theory

In this chapter we introduce, mostly without proofs, some basic measure-theoretic tools needed in the next chapters. We decided to present the most significant result in the quite general framework of *separable metric spaces* in view of possible applications to infinite dimensional Hilbert (or Banach) spaces, thus avoiding any local compactness assumption (we refer to the treatises [109, 59, 60, 117, 55] for comprehensive presentations of this subject).

At this preliminary level, the existence of an equivalent complete metric (Polish spaces) only enters in the *compact inner regularity* (5.1.9) or *tightness* (5.1.8) of every Borel measure (it is a consequence of Ulam's Theorem [60, 7.1.4], a particular case of the converse implication in Prokhorov Theorem 5.1.3), which in particular appears in the so called disintegration theorem 5.3.1 and its consequences; this inner approximation condition is satisfied by a wider class of even non complete metric spaces (the so called *Radon spaces* [117, page 117]) and it will be sufficient for our aims. Since weak topologies in Hilbert-Banach spaces are not metrizable, it will also be useful (see Lemma 5.1.12) to deal with auxiliary non complete metrics, still satisfying (5.1.9).

Even if the presentation looks more abstract and the assumptions very weak with respect to the more usual finite dimensional Euclidean setting of the standard theory for evolutionary PDE's, this approach is sufficiently powerful to provide all the crucial results and allows for a great flexibility.

Let X be a separable metric space. We denote by $\mathscr{B}(X)$ the family of the Borel subsets of X, by $\mathscr{P}(X)$ the family of all Borel probability measures on X. The *support* $\operatorname{supp}\mu \subset X$ of $\mu \in \mathscr{P}(X)$ is the closed set defined by

$$\operatorname{supp}\mu := \Big\{x \in X : \mu(U) > 0 \quad \text{for each neighborhood } U \text{ of } x\Big\}. \tag{5.0.1}$$

When $\boldsymbol{X} = X_1 \times \ldots \times X_k$ is a product space, we will often use *bold* letters to indicate Borel measures $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$. Recall that for separable metric spaces $X_1, \ldots, X_k$ the Borel σ-algebra coincides with the product one

$$\mathscr{B}(\boldsymbol{X}) = \mathscr{B}(X_1) \times \mathscr{B}(X_2) \times \cdots \times \mathscr{B}(X_k). \tag{5.0.2}$$

5.1 Narrow convergence, tightness, and uniform integrability

Conformally to the probabilistic terminology, we say that a sequence $(\mu_n) \subset \mathscr{P}(X)$ is *narrowly* convergent to $\mu \in \mathscr{P}(X)$ as $n \to \infty$ if

$$\lim_{n\to\infty} \int_X f(x)\, d\mu_n(x) = \int_X f(x)\, d\mu(x) \tag{5.1.1}$$

for every function $f \in C^0_b(X)$, the space of continuous and bounded real functions defined on X.

Of course, it is sufficient to check (5.1.1) on any subset $\mathscr{C}$ of bounded continuous functions whose linear envelope $\operatorname{span}\mathscr{C}$ is uniformly dense (i.e. dense in the uniform topology induced by the "sup" norm) in $C^0_b(X)$. Even better, let us suppose that $\mathscr{C}_0 \subset C^0_b(X)$ satisfies the approximation properties

$$\int_X f(x)\, d\mu(x) = \sup\Big\{ \int_X h(x)\, d\mu(x) : \ h \in \mathscr{C}_0,\, h \le f \Big\} \tag{5.1.2a}$$

$$= \inf\Big\{ \int_X h(x)\, d\mu(x) : \ h \in \mathscr{C}_0,\, h \ge f \Big\}, \tag{5.1.2b}$$

for every $f \in \mathscr{C}$; then if (5.1.1) holds for every $f \in \mathscr{C}_0$, then it holds for every continuous and bounded function f. In fact for every $f \in \mathscr{C}$ we easily have

$$\begin{aligned}\liminf_{n\to\infty} \int_X f(x)\, d\mu_n(x) &\ge \sup_{h\in\mathscr{C}_0, h\le f} \ \liminf_{n\to\infty} \int_X h(x)\, d\mu_n(x) \\ &= \sup_{h\in\mathscr{C}_0, h\le f} \int_X h(x)\, d\mu(x) = \int_X f(x)\, d\mu(x),\end{aligned} \tag{5.1.3}$$

and the opposite inequality for the "lim sup" can be obtained in a similar way starting from (5.1.2b). Thus every $f \in \mathscr{C}$ satisfies (5.1.1), and we get the same property for every $f \in C^0_b(X)$ since $\operatorname{span}\mathscr{C}$ is uniformly dense in $C^0_b(X)$.

If d is any metric for X, the subset of d-uniformly (or d-Lipschitz) continuous and bounded real functions provides an important example [119, Th. 3.1.5] satisfying (5.1.2a,b). For, we can pointwise approximate a continuous and bounded function f from below with an increasing sequence of bounded Lipschitz functions f_k (they are particular examples of the Moreau-Yosida approximations for the exponent $p = 1$, see Section 3.1)

$$f_k(x) := \inf_y f(y) + kd(x,y), \quad \text{with} \quad \begin{cases} \inf f \le f_k(x) \le f(x) \le \sup f, \\ f(x) = \lim_{k\to\infty} f_k(x) = \sup_{k\in\mathbb{N}} f_k(x), \end{cases} \tag{5.1.4}$$

thus obtaining (5.1.2a) by Fatou's lemma; changing f to $-f$ we obtain (5.1.2b).

A slight refinement of this argument provides a *countable* set of d-Lipschitz functions satisfying (5.1.2a,b) for every function $f \in C^0_b(X)$: we simply choose a countable dense set $D \subset X$ and we consider the countable family of functions $h : X \to \mathbb{R}$ of the type

$$\begin{gathered} h(x) = \big(q_1 + q_2 d(x,y)\big) \wedge k \\ \text{for some } q_1, q_2, k \in \mathbb{Q}, \quad q_2,\ k \in (0,1),\ y \in D. \end{gathered} \tag{5.1.5a}$$

We denote by $\mathscr{C}_1$ the collection generated from this set by taking the infimum of a finite number of functions, thus satisfying

$$\sup_{x\in X} |h(x)| < 1, \quad \mathrm{Lip}(h,X) < 1 \quad \forall\, h \in \mathscr{C}_1; \tag{5.1.5b}$$

finally we set

$$\mathscr{C}_0 = \big\{\lambda h : h \in \mathscr{C}_1, \lambda \in \mathbb{Q}\big\}. \tag{5.1.5c}$$

As showed by the next remark, the above constructions are useful, since in general $C^0_b(X)$ (endowed with the uniform topology) is not separable, unless X is compact.

Remark 5.1.1 (Narrow convergence is induced by a distance). It is well known that narrow convergence is induced by a distance on $\mathscr{P}(X)$: an admissible choice is obtained by ordinating each element of $\mathscr{C}_1$ in a sequence (f_k) and setting

$$\delta(\mu,\nu) := \sum_{k=1}^{\infty} 2^{-k} \Big| \int_X f_k\, d\mu - \int_X f_k\, d\nu \Big|. \tag{5.1.6}$$

If d is a complete bounded metric for X we could also choose any p-Wasserstein distance on $\mathscr{P}(X)$ (see Chap. 7 and Remark 7.1.7). In particular, the family of all converging sequences is sufficient to characterize the narrow topology and we do not have to distinguish between compact and sequentially compact subsets.

Remark 5.1.2 (Narrow topology coincides with the weak* topology of $\big(C^0_b(X)\big)'$). $\mathscr{P}(X)$ can be identified with a convex subset of the unitary ball of the dual space $\big(C^0_b(X)\big)'$: by definition, narrow convergence is induced by the weak* topology of $\big(C^0_b(X)\big)'$. This identification is useful to characterize the closed convex hull in $\mathscr{P}(X)$ of a given set $\mathcal{K} \subset \mathscr{P}(X)$: Hahn-Banach theorem shows that

$$\mu \in \overline{\mathrm{Conv}}\,(\mathcal{K}) \quad \Longleftrightarrow \quad \int_X f\, d\mu \le \sup_{\nu\in\mathcal{K}} \int_X f\, d\nu \quad \forall\, f \in C^0_b(X). \tag{5.1.7}$$

For instance we can prove the separability of $\mathscr{P}(X)$ by choosing $\mathcal{K} := \{\delta_x : x \in D\}$, where D is a countable dense subset of X: by (5.1.7) we easily check that $\mathscr{P}(X) = \overline{\mathrm{Conv}\,\mathcal{K}}$ and therefore the subset of all the convex combinations with rational coefficients of δ-measures concentrated in D is narrowly dense in $\mathscr{P}(X)$.

The following theorem provides a useful characterization of relatively compact sets with respect to the narrow topology.

Theorem 5.1.3 (Prokhorov, [55, III-59]). *If a set $\mathcal{K} \subset \mathscr{P}(X)$ is* tight, *i.e.*

$$\forall\,\varepsilon > 0 \quad \exists\, K_\varepsilon \text{ compact in } X \text{ such that} \quad \mu(X \setminus K_\varepsilon) \le \varepsilon \quad \forall\,\mu \in \mathcal{K}, \tag{5.1.8}$$

then $\mathcal{K}$ is relatively compact in $\mathscr{P}(X)$. Conversely, if there exists an equivalent complete *metric for X, i.e. X is a so called* Polish space, *then every relatively compact subset of $\mathscr{P}(X)$ is tight.*

Observe in particular that in a Polish space X each measure $\mu \in \mathscr{P}(X)$ is tight; moreover, compact inner approximation holds for every Borel set:

$$\forall\, B \in \mathscr{B}(X),\ \varepsilon > 0 \quad \exists\, K_\varepsilon \Subset B: \quad \mu(B \setminus K_\varepsilon) \le \varepsilon. \tag{5.1.9}$$

In fact, this approximation property holds for a more general class of spaces, the so-called *Radon spaces* [117].

Definition 5.1.4 (Radon spaces). *A separable metric space X is a* Radon space *if every Borel probability measure $\mu \in \mathscr{P}(X)$ satisfies* (5.1.9).

When the elements of $\mathcal{K} \subset X$ are ordinated in a sequence (μ_n) of tight measures (which is always the case if X is a Radon space), then the tightness condition (5.1.8) can also be reformulated as

$$\inf_{K \Subset X} \limsup_{n\to\infty} \mu_n(X \setminus K) = 0, \tag{5.1.10a}$$

or, equivalently since $\mu_n(X) \equiv 1$,

$$\sup_{K \Subset X} \liminf_{n\to\infty} \mu_n(K) = 1. \tag{5.1.10b}$$

An interesting result by Le Cam [91], [60, 11.5.3], shows that

$$\begin{gathered}\text{in a (metric, separable) Radon space } X,\\ \textit{every narrowly converging sequence } (\mu_n) \subset \mathscr{P}(X) \textit{ is tight.}\end{gathered} \tag{5.1.11}$$

Remark 5.1.5 (An integral condition for tightness). It is easy to check that (5.1.8) is equivalent to the following condition: there exists a function $\varphi : X \to [0, +\infty]$, whose sublevels $\{x \in X : \varphi(x) \le c\}$ are compact in X, such that

$$\sup_{\mu \in \mathcal{K}} \int_X \varphi(x)\, d\mu(x) < +\infty. \tag{5.1.12}$$

For, if $\{\varepsilon_n\}_{n=0}^{\infty}$ is a sequence with $\sum_{n=0}^{+\infty} \varepsilon_n < +\infty$ and $K_n := K_{\varepsilon_n}$ is an (increasing) sequence of compact sets satisfying (5.1.8), the function

$$\varphi(x) := \inf\Big\{n \ge 0 : x \in K_n\Big\} = \sum_{n=0}^{+\infty} \chi_{X\setminus K_n}(x), \tag{5.1.13}$$

satisfies (5.1.12). Conversely, if $\mathcal{K}$ satisfies (5.1.12), Chebichev inequality shows that (5.1.8) is satisfied by the family of sublevels of φ.

We conclude this part by a well known result comparing narrow convergence with convergence in the sense of distributions when $X = \mathbb{R}^d$.

Remark 5.1.6 (Narrow and distributional convergence in $X = \mathbb{R}^d$). For $n \in \mathbb{N}$ let μ_n, μ be Borel probability measures in the euclidean space $X = \mathbb{R}^d$ such that

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f(x)\, d\mu_n(x) = \int_{\mathbb{R}^d} f(x)\, d\mu(x) \quad \forall\, f \in C_c^\infty(\mathbb{R}^d). \tag{5.1.14}$$

Then the sequence (μ_n) is tight and it narrowly converges to μ as $n \to \infty$.
For, if $\zeta \in C_c^\infty(\mathbb{R}^d)$ satisfies

$$0 \le \zeta \le 1, \quad \zeta(x) = 1 \quad \text{if } |x| \le 1/2, \quad \zeta(x) = 0 \quad \text{if } |x| \ge 1,$$

and $\zeta_k(x) := \zeta(x/k)$, we have

$$\liminf_{n\to\infty} \mu_n(\overline{B_k(0)}) \ge \lim_{n\to\infty} \int_{\mathbb{R}^d} \zeta_k(x)\, d\mu_n(x) = \int_{\mathbb{R}^d} \zeta_k(x)\, d\mu(x);$$

since Lebesgue dominated convergence theorem yields

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} \zeta_k(x)\, d\mu(x) = 1,$$

choosing k sufficiently big we can verify the tightness condition (5.1.10b). By Prokhorov theorem the sequence (μ_n) has at least one narrowly convergence subsequence: a standard approximation result by convolution shows that any narrow limit point of the sequence (μ_n) should coincide with μ, which is therefore the narrow limit of the whole sequence (recall that the narrow topology is metrizable, see Remark 5.1.1).

5.1.1 Unbounded and l.s.c. integrands

When one needs to pass to the limit in expressions like (5.1.1) w.r.t. *unbounded or lower semicontinuous* functions f, the following two properties are quite useful. The first one is a lower semicontinuity property:

$$\liminf_{n\to\infty} \int_X g(x)\, d\mu_n(x) \ge \int_X g(x)\, d\mu(x) \tag{5.1.15}$$

for every sequence $(\mu_n) \subset \mathscr{P}(X)$ narrowly convergent to μ and any l.s.c. function $g : X \to (-\infty, +\infty]$ bounded from below: it follows by the same approximation argument of (5.1.3), by truncating the Moreau-Yosida approximations (5.1.4); in this case l.s.c. functions satisfy only the approximation property (5.1.2a), where e.g. $\mathscr{C}_0$ is given by (5.1.5a,b,c).
Changing g in $-g$ one gets the corresponding "lim sup" inequality for upper semicontinuous functions bounded from above. In particular, choosing as g the characteristic functions of open and closed subset of X, we obtain

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G) \quad \forall\, G \text{ open in } X, \tag{5.1.16}$$

$$\limsup_{n\to\infty} \mu_n(F) \leq \mu(F) \quad \forall\, F \text{ closed in } X. \tag{5.1.17}$$

The statement of the second property requires the following definitions: we say that a Borel function $g : X \to [0, +\infty]$ is *uniformly integrable* w.r.t. a given set $\mathcal{K} \subset \mathscr{P}(X)$ if

$$\lim_{k\to\infty} \int_{\{x:g(x)\geq k\}} g(x)\, d\mu(x) = 0 \quad \text{uniformly w.r.t. } \mu \in \mathcal{K}. \tag{5.1.18}$$

If d is a given metric for X, in the particular case of $g(x) := d(x, \bar{x})^p$, for some (and thus any) $\bar{x} \in X$ and a given $p > 0$, i.e. if

$$\lim_{k\to\infty} \int_{X\setminus B_k(\bar{x})} d^p(\bar{x}, x)\, d\mu(x) = 0 \quad \text{uniformly w.r.t. } \mu \in \mathcal{K}, \tag{5.1.19}$$

we say that the set $\mathcal{K} \subset \mathscr{P}(X)$ *has uniformly integrable p-moments*. Notice that if

$$0 < p < p_1 \quad \text{and} \quad \sup_{\mu\in\mathcal{K}} \int_X d(x, \bar{x})^{p_1}\, d\mu(x) < +\infty, \tag{5.1.20}$$

then $\mathcal{K}$ has uniformly integrable p-moments. In the case when $X = \mathbb{R}^d$ with the usual Euclidean distance, any family $\mathcal{K} \subset \mathscr{P}(\mathbb{R}^d)$ satisfying (5.1.20) is tight. The following lemma provides a characterization of p-uniformly integrable families, extending the validity of (5.1.1) to unbounded but with p-growth functions, i.e. functions $f : X \to \mathbb{R}$ such that

$$|f(x)| \leq A + B\, d^p(\bar{x}, x) \qquad \forall x \in X, \tag{5.1.21}$$

for some A, $B \geq 0$ and $\bar{x} \in X$. We denote by $\mathscr{P}_p(X)$ the subset

$$\mathscr{P}_p(X) := \Big\{\mu \in \mathscr{P}(X) : \int_X d(x, \bar{x})^p\, d\mu(x) < +\infty \quad \text{for some } \bar{x} \in X\Big\}. \tag{5.1.22}$$

Lemma 5.1.7. *Let (μ_n) be a sequence in $\mathscr{P}(X)$ narrowly convergent to $\mu \in \mathscr{P}(X)$. If $f : X \to \mathbb{R}$ is continuous, $g : X \to (-\infty, +\infty]$ is lower semicontinuous, and*

$|f|, g^-$ *are uniformly integrable w.r.t. the set* $\{\mu_n\}_{n\in\mathbb{N}}$*, then*

$$\liminf_{n\to\infty} \int_X g(x)\, d\mu_n(x) \geq \int_X g(x)\, d\mu(x) > -\infty, \tag{5.1.23a}$$

$$\lim_{n\to\infty} \int_X f(x)\, d\mu_n(x) = \int_X f(x)\, d\mu(x). \tag{5.1.23b}$$

Conversely, if $f : X \to [0, +\infty)$ *is continuous,* μ_n*-integrable, and*

$$\limsup_{n\to\infty} \int_X f(x)\, d\mu_n(x) \leq \int_X f(x)\, d\mu(x) < +\infty, \tag{5.1.24}$$

then f *is uniformly integrable w.r.t.* $\{\mu_n\}_{n\in\mathbb{N}}$*.*
In particular, a family $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathscr{P}(X)$ *has uniformly integrable p-moments iff* (5.1.1) *holds for every continuous function* $f : X \to \mathbb{R}$ *with p-growth.*

Proof. If μ_n narrowly converges to μ as $n \to \infty$ and g is lower semicontinuous, (5.1.15) yields

$$\liminf_{n\to\infty} \int_X g_k\, d\mu_n \geq \int_X g_k\, d\mu \qquad \forall k \in \mathbb{N},$$

where $g_k := g \vee (-k)$, $k \geq 0$. On the other hand, since g^- is uniformly integrable w.r.t. $\{\mu_n\}_{n\in\mathbb{N}}$ and $g_k \geq g$, (5.1.18) gives

$$\sup_{n\in\mathbb{N}} \left(\int_X g_k\, d\mu_n - \int_X g\, d\mu_n \right) \leq \sup_{n\in\mathbb{N}} \int_{\{x: g^-(x)\geq k\}} g^-\, d\mu_n \to 0$$

as $k \to \infty$. Using these two facts we obtain (5.1.23a). As usual, (5.1.23b) follows by applying (5.1.23a) to $g := f$ and $g := -f$.

Conversely, let $f : X \to [0, +\infty)$ be a continuous function satisfying (5.1.24) and let

$$f^k(x) := f(x) \wedge k, \ \forall\, x \in X, \quad F^k := \big\{x \in X : f(x) \geq k\big\};$$

since f^k is continuous and bounded and F^k is a closed subset of X, recalling (5.1.17) and (5.1.15) we have for any $\epsilon > 0$

$$\begin{aligned}\limsup_{n\to\infty} \int_{\{x: f(x)\geq k\}} f\, d\mu_n &= \limsup_{n\to\infty} \Big(\int_X \big(f - f^k\big)\, d\mu_n + k\mu_n(F^k) \Big) \\ &\leq \int_X \big(f - f^k\big)\, d\mu + k\mu(F^k) = \int_{F^k} f\, d\mu < \varepsilon\end{aligned}$$

for k sufficiently large. Since f is uniformly integrable for finite subsets of $\{\mu_n\}_{n\in\mathbb{N}}$, this easily leads to the uniform integrability of f. $\square$

There exists an interesting link between narrow convergence of probability measures and Kuratowski convergence of their supports:

Proposition 5.1.8. *If $(\mu_n) \subset \mathscr{P}(X)$ is a sequence narrowly converging to $\mu \in \mathscr{P}(X)$ then* $\operatorname{supp}\mu \subset K\!-\!\liminf_{n\to\infty} \operatorname{supp}\mu_n$, *i.e.*

$$\forall\, x \in \operatorname{supp}\mu \quad \exists\, x_n \in \operatorname{supp}\mu_n : \quad \lim_{n\to\infty} x_n = x. \tag{5.1.25}$$

Proof. Let $x \in \operatorname{supp}\mu$ and let $B_{1/k}(x)$ be the open ball of center x and radius $1/k$ with respect to the distance d on X. By (5.1.16) we obtain

$$\liminf_{n\to\infty} \mu_n(B_{1/k}(x)) \geq \mu(B_{1/k}(x)) > 0;$$

thus the strictly increasing sequence

$$j_0 := 0, \quad j_k := \min\Big\{n \in \mathbb{N} : n > j_{k-1}, \quad \operatorname{supp}\mu_m \cap B_{1/k}(x) \neq \varnothing \quad \forall\, m \geq n\Big\}$$

is well defined. For $j_k \leq n < j_{k+1}$ pick a point $x_n \in \operatorname{supp}\mu_n \cap B_{1/k}(x)$: clearly the sequence (x_n) satisfies (5.1.25). □

Corollary 5.1.9 (Convergence of Dirac masses). *A sequence $(x_n) \subset X$ is convergent in X iff the sequence (δ_{x_n}) is narrowly convergent in $\mathscr{P}(X)$; in this case, the limit measure μ is δ_x, x being the limit of the sequence (x_n).*

Proposition 5.1.10. *Let $(\mu_n) \subset \mathscr{P}(X)$ be a sequence narrowly converging to $\mu \in \mathscr{P}(X)$ and let $f, g : X \to (-\infty, +\infty]$ be Borel functions such that $|f|, g^-$ are uniformly integrable with respect to $\{\mu_n\}_{n\in\mathbb{N}}$. If for any $\varepsilon > 0$ there exists a closed set $A \subset X$ such that*

$$f|_A \text{ is continuous}, \quad g|_A \text{ is l.s.c.}, \quad \text{and} \quad \limsup_{n\to\infty} \mu_n(X \setminus A) < \varepsilon, \tag{5.1.26}$$

then (5.1.1) *and* (5.1.15) *hold.*

Proof. As usual we can limit us to consider the l.s.c. case; using the uniform integrability of g^- with respect to $\{\mu_n\}_{n\in\mathbb{N}}$, a truncation argument, and arguing as in the first part of the proof of Lemma 5.1.7, we reduce immediately ourselves to the case when g is bounded from below by a constant $-M \leq 0$. Let $\varepsilon > 0, k \in \mathbb{N}$ be fixed and let $A \subset X$ be a closed set such that (5.1.26) holds. We consider the truncated functions $g^k(x) := g(x) \wedge k$ for $x \in X$, and the lower semicontinuous $\tilde g^k$

$$\tilde g^k(x) = \begin{cases} g^k(x) & \text{if } x \in A, \\ k & \text{if } x \in X \setminus A, \end{cases}$$

which extends $g^k|_A$ to X. We obtain

$$\begin{aligned}\liminf_{n\to\infty} \int_X g\, d\mu_n \geq \liminf_{n\to\infty} \int_X g^k\, d\mu_n &\geq \liminf_{n\to\infty} \Big(\int_X \tilde g^k\, d\mu_n + \int_{X\setminus A} \big(g^k - \tilde g^k\big)\, d\mu_n\Big) \\ &\geq \liminf_{n\to\infty} \int_X \tilde g^k\, d\mu_n - (M+k) \limsup_{n\to\infty} \mu_n(X \setminus A) \\ &\geq \int_X \tilde g^k(x)\, d\mu - \varepsilon(M+k) \geq \int_X g^k(x)\, d\mu - (k+M)\varepsilon.\end{aligned}$$

Passing to the limit, first as $\varepsilon \downarrow 0$ and then as $k \uparrow \infty$ we obtain (5.1.15). □

5.1.2 Hilbert spaces and weak topologies

Let X be a separable, infinite dimensional, Hilbert space, with norm $|\cdot|$ and scalar product $\langle \cdot, \cdot \rangle$; in many circumstances it would be useful to rephrase the results of the previous section with respect to the weak topology $\sigma(X, X')$ of X. Unfortunately, the weak topology is not induced by a distance on X, thus the previous statements are not immediately applicable.

We can circumvent this difficulty by the following simple trick: we introduce a new continuous norm $\|\cdot\|_\varpi$, inducing a topology ϖ globally weaker than $\sigma(X, X')$, but coinciding with $\sigma(X, X')$ on bounded sets (with respect to the original stronger norm $|\cdot|$). In particular bounded sets of X are relatively compact w.r.t. ϖ and Borel sets with respect to the three topologies coincide.

For instance, if $\{e_n\}_{n=1}^{+\infty}$ is an orthonormal basis of X, an admissible choice is

$$\|x\|_\varpi^2 := \sum_{n=1}^{\infty} \frac{1}{n^2} \langle x, e_n \rangle^2. \tag{5.1.27}$$

In fact, if $(x_k) \subset X$ is a bounded sequence, we can extract a subsequence, still denoted by x_k, weakly converging to x in X; since $\langle x_k - x, e_n \rangle \to 0$ as $k \to \infty$ for each $n \geq 1$, Lebesgue dominated convergence theorem yields

$$\lim_{k\to\infty} \|x_k - x\|_\varpi^2 = \lim_{k\to\infty} \sum_{n=1}^{\infty} \frac{1}{n^2} \langle x_k - x, e_n \rangle^2 = 0.$$

We denote by X_ϖ the new pre-Hilbertian topological vector space. We will also introduce the space of smooth cylindrical functions $\mathrm{Cyl}(X)$: observe that for finite dimensional spaces, X_ϖ is homeomorphic to X and $\mathrm{Cyl}(X) = C_c^\infty(X)$.

Definition 5.1.11 (Finite dimensional projection and smooth cylindrical functions). *We denote by $\Pi_d(X)$ the space of all maps $\pi : X \to \mathbb{R}^d$ of the form*

$$\pi(x) = (\langle x, e_1 \rangle, \langle x, e_2 \rangle, \ldots, \langle x, e_d \rangle) \qquad x \in X, \tag{5.1.28}$$

where $\{e_1, \ldots, e_d\}$ is any orthonormal family of vectors in X. The adjoint map

$$\pi^* : y \in \mathbb{R}^d \to \sum_{k=1}^{d} y_k e_k \in \mathrm{span}(e_1, \ldots, e_k) \subset X \tag{5.1.29}$$

is a linear isometry of $\mathbb{R}^d$ onto $\mathrm{span}(e_1, \ldots, e_d)$ *so that*

$$\begin{aligned} \pi \circ \pi^* \quad & \textit{is the identity in } \mathbb{R}^d \textit{ and} \\ \hat{\pi} := \pi^* \circ \pi \quad & \textit{is the orthogonal projection of } X \textit{ onto } \mathrm{span}(e_1, \ldots, e_d). \end{aligned} \tag{5.1.30}$$

We denote by $\mathrm{Cyl}(X)$ *the functions $\varphi = \psi \circ \pi$ with $\pi \in \Pi_d(X)$ and $\psi \in C_c^\infty(\mathbb{R}^d)$.*

Notice that any $\varphi = \psi \circ \pi \in \mathrm{Cyl}(X)$ is a Lipschitz function, everywhere differentiable in the Fréchet sense, and that φ is also continuous with respect to the weak topology of X and to X_ϖ (if the corresponding orthogonal systems coincide). Moreover $\nabla\varphi = \pi^* \circ \nabla\psi \circ \pi$.

The following properties are immediate:

Lemma 5.1.12. *Let X be a separable Hilbert space and let X_ϖ be the pre-Hilbertian vector space whose norm is defined by* (5.1.27).

(a) *If K is weakly compact in X then K is strongly compact in X_ϖ.*

(b) *If*

$$g : X \to (-\infty, +\infty] \quad \text{is weakly l.s.c.} \quad \text{and} \quad \lim_{|x|\to\infty} g(x) = +\infty, \tag{5.1.31}$$

then it is lower semicontinuous in X_ϖ with compact sublevels.

(c) *Let us denote by $\overline{B_R} := \{x \in X : |x| \le R\}$ the centered closed balls w.r.t. the strong norm; if $\mathcal{K} \subset \mathscr{P}(X)$ satisfies the* weak tightness condition

$$\forall\, \varepsilon > 0 \quad \exists R_\varepsilon > 0 \text{ such that } \mu(X \setminus \overline{B_{R_\varepsilon}}) \le \varepsilon \quad \forall\, \mu \in \mathcal{K}, \tag{5.1.32}$$

then $\mathcal{K}$ is tight in $\mathscr{P}(X_\varpi)$ and therefore relatively compact in $\mathscr{P}(X_\varpi)$.

(d) *If the sequence $(\mu_n) \subset \mathscr{P}(X)$ is narrowly converging to μ in $\mathscr{P}(X_\varpi)$ and it is weakly tight according to* (5.1.32)*, then for every Borel functions $f, g : X \to (-\infty, +\infty]$ such that $g^-, |f|$ are uniformly integrable and f (resp. g) is weakly continuous (resp. l.s.c.) on bounded sets of X, we have*

$$\liminf_{n\to\infty} \int_X g(x)\, d\mu_n(x) \ge \int_X g(x)\, d\mu(x), \tag{5.1.33a}$$

$$\lim_{n\to\infty} \int_X f(x)\, d\mu_n(x) = \int_X f(x)\, d\mu(x). \tag{5.1.33b}$$

(e) *$\mathcal{K} \subset \mathscr{P}(X)$ is weakly tight according to* (5.1.32) *iff there exists a Borel function $h : X \to [0, +\infty]$ such that $h(x) \to +\infty$ as $|x| \to \infty$ and*

$$\sup_{\mu\in\mathcal{K}} \int_X h(x)\, d\mu(x) < +\infty. \tag{5.1.34}$$

(f) *If the sequence $(\mu_n) \subset \mathscr{P}(X)$ is weakly tight according to* (5.1.32)*, then it narrowly converges to μ in $\mathscr{P}(X_\varpi)$ iff*

$$\lim_{n\to\infty} \int_X \varphi(x)\, d\mu_n(x) = \int_X \varphi(x)\, d\mu(x) \quad \forall\, \varphi \in \mathrm{Cyl}(X). \tag{5.1.35}$$

Proof. (a) and (b) are trivial and (c) is a direct consequence of the fact that bounded and closed convex sets are compact in X_ϖ. Since on bounded subsets of X the topology of X_ϖ coincides with the weak one, (d) follows from Proposition 5.1.10.

One implication in (e) follows directly from Chebichev inequality. The other one can be proved arguing as in Remark 5.1.5.
Finally, one implication in (f) is a consequence of (5.1.33b) of (d), since (smooth) cylindrical functions are bounded and weakly continuous. In order to prove the converse implication, we can simply check that any two narrow limit point μ^1, μ^2 of the sequence (μ_n) in $\mathscr{P}(X_\varpi)$ should coincide. For, let $f \in C^0_b(X)$ and π_d be the map (5.1.28), so that $\hat\pi_d := \pi^*_d \circ \pi_d$ is the orthogonal projection of X onto $X_d = \mathrm{span}(e_1, \cdots, e_d)$. We set $\psi_d := f \circ \pi^*_d \in C^0_b(\mathbb{R}^d)$, $\varphi_d := \psi_d \circ \pi_d = f \circ \hat\pi_d \in \mathrm{Cyl}(X)$; by (5.1.35) we know

$$\int_X \varphi(x)\, d\mu^1(x) = \int_X \varphi(x)\, d\mu^2(x) \quad \forall\, \varphi \in \mathrm{Cyl}(X); \tag{5.1.36}$$

a standard approximation argument for bounded continuous functions defined in $\mathbb{R}^d$ by smooth functions in $C^\infty_c(\mathbb{R}^d)$ as in Remark 5.1.6 yields (5.1.36) for $\varphi := \varphi_d$ and $d \in \mathbb{N}$; therefore

$$\int_X f(\hat\pi_d(x))\, d\mu^1(x) = \int_X f(\hat\pi_d(x))\, d\mu^2(x) \quad \forall\, d \in \mathbb{N}.$$

Passing to the limit as $d \to \infty$, since $\hat\pi_d(x) \to x$ for every $x \in X$, Lebesgue dominated convergence theorem yields

$$\int_X f(x)\, d\mu^1(x) = \int_X f(x)\, d\mu^2(x).$$

Since f is an arbitrary function in $C^0_b(X)$ we obtain $\mu^1 = \mu^2$. □

In the following theorem we will show that narrow convergence in $\mathscr{P}(X_\varpi)$ and convergence of the p-moment $\int_X |x|^p\, d\mu_h(x)$ (but more general integrands are allowed) yields convergence in $\mathscr{P}(X)$, thus obtaining the measure-theoretic version of the fact that weak convergence and convergence of the norms in X imply strong convergence. We will show a different proof of this fact at the end of Section 7.1.

Theorem 5.1.13. *Let $j : [0, +\infty) \to [0, +\infty)$ be a continuous, strictly increasing and surjective map, and let μ_n, $\mu \in \mathscr{P}(X)$ be satisfying*

$$\mu_n \to \mu \quad \text{in } \mathscr{P}(X_\varpi), \qquad \lim_{n\to\infty} \int_X j(|x|)\, d\mu_n(x) = \int_X j(|x|)\, d\mu < +\infty. \tag{5.1.37}$$

Then μ_n converge to μ in $\mathscr{P}(X)$.

Proof. Observe that the family $\{\mu_n\}_{n\in\mathbb{N}}$ is weakly tight, according to (5.1.32). We consider the vector space $\mathscr{H}$ of continuous functions $h : X \to \mathbb{R}$ satisfying the growth condition (compare with (5.1.21))

$$\exists\, A, B \geq 0 : \quad |h(x)| \leq A + Bj(|x|) \quad \forall\, x \in X, \tag{5.1.38}$$

and
$$\lim_{n\to\infty}\int_X h(x)\,d\mu_n(x) = \int_X h(x)\,d\mu(x). \tag{5.1.39}$$
Observe that $\mathscr{H}$ is closed with respect to uniform convergence of functions and contains the constants and the function $j(|\cdot|)$.
By the monotonicity argument outlined at the beginning of Section 5.1 we need only to check that the infimum of a finite number of functions of the form
$$x \mapsto \big(q_1 + q_2|x-y|\big) \wedge k, \quad q_1 \in \mathbb{R}, \quad q_2, k \geq 0, \quad y \in X, \tag{5.1.40}$$
belongs to $\mathscr{H}$. To this aim, let us consider the convex cone $\mathscr{A} \subset \mathscr{H}$ of strongly continuous functions which satisfy (5.1.38), (5.1.39), and are weakly lower semicontinuous. Notice that, truncated affine functions of the type
$$x \mapsto (-l)\vee\big(a+\langle x,y\rangle\big)\wedge m, \quad \text{for } l,m \geq 0,\ a \in \mathbb{R},\ y \in X \quad \text{belongs to } \mathscr{A}, \tag{5.1.41}$$
since they are bounded, weakly continuous, and condition (5.1.39) follows by (d) of Lemma 5.1.12.
Let us first prove that $\mathscr{A}$ is a lattice.

Claim 1. *If f, $g \in C^0(X)$ satisfy* (5.1.38), *are weakly lower semicontinuous, and $f + g \in \mathscr{A}$, then both f, $g \in \mathscr{A}$.*

Indeed, by (5.1.33a) we have
$$\begin{aligned}\int_X (f+g)\,d\mu &= \lim_{n\to\infty}\int_X (f+g)\,d\mu_n \geq \limsup_{n\to\infty}\int_X f\,d\mu_n + \liminf_{n\to\infty}\int_X g\,d\mu_n\\ &\geq \int_X f\,d\mu + \int_X g\,d\mu = \int_X (f+g)\,d\mu,\end{aligned}$$
which yields
$$\limsup_{n\to\infty}\int_X f\,d\mu_n + \liminf_{n\to\infty}\int_X g\,d\mu_n = \int_X f\,d\mu + \int_X g\,d\mu; \tag{5.1.42}$$
since by (5.1.33a)
$$\limsup_{n\to\infty}\int_X f\,d\mu_n \geq \int_X f\,d\mu, \quad \liminf_{n\to\infty}\int_X g\,d\mu_n \geq \int_X g\,d\mu,$$
(5.1.42) yields
$$\limsup_{n\to\infty}\int_X f\,d\mu_n = \int_X f\,d\mu, \quad \liminf_{n\to\infty}\int_X g\,d\mu_n = \int_X g\,d\mu;$$
inverting the role of f and g we obtain $f, g \in \mathscr{A}$.

Claim 1 immediately implies that $\mathscr{A}$ is a lattice, as
$$f,g \in \mathscr{A} \quad\Rightarrow\quad f+g = (f\wedge g)+(f\vee g) \in \mathscr{A} \quad\Rightarrow\quad f\wedge g,\ f\vee g \in \mathscr{A}.$$

Since

$$\Big(q_1 + q_2|x-y|\Big) \wedge k = \Big(q_1 + \big(q_2|x-y|\big) \wedge (k-q_1)\Big) \wedge k,$$

it remains to show that

$$\text{all functions } x \mapsto |x-y| \wedge k, \text{ for } y \in X,\ k \geq 0, \text{ belong to } \mathscr{A}. \tag{5.1.43}$$

To this aim, we need a further claim.

Claim 2. If $f \in \mathscr{A}$ and $\theta : \mathbb{R} \to \mathbb{R}$ is a uniformly continuous, bounded, increasing function, then $\theta \circ f \in \mathscr{A}$.

Indeed, since θ can be uniformly approximated by a sequence of Lipschitz continuous increasing maps, it is not restrictive to assume that θ is Lipschitz, bounded, and its Lipschitz constant is less than 1; in this case also $x \mapsto x - \theta(x)$ is Lipschitz and increasing, thus $\theta \circ f$ and $f - f \circ \theta$ are still weakly lower semicontinuous they satisfies the growth condition (5.1.38) and and their sum is $f \in \mathscr{A}$: we can apply Claim 1.

Let us consider (5.1.43) in the case $y = 0$ first: we fix $R > 0$ and we consider the continuous increasing function θ_R which vanishes in $(-\infty, 0)$ and satisfies

$$\theta_R(s) := \big(j^{-1}(s)\big)^2 \wedge R^2,\ s \geq 0, \quad \text{so that} \quad r^2 \wedge R^2 = \theta_R(j(r)) \quad \forall\, r \geq 0.$$

By Claim 2, we deduce that the map f_R defined by $f_R(x) := |x|^2 \wedge R^2$ belongs to $\mathscr{A}$.

Now, for fixed $k, l, m > 0$ and $y \in X$, we set

$$g_{l,m}(x) := (-l) \vee \Big(-2\langle x, y\rangle + |y|^2\Big) \wedge m, \quad g_{R,l.m,k} := \Big(\big(f_R + g_{l,m}\big) \vee 0\Big)^{1/2} \wedge k,$$

and we know by the lattice property, the previous claim, and (5.1.41) that $g_{R,l,m,k} \in \mathscr{A}$. Choosing now $R \geq l + k^2$ and $m \geq k$ the expression of $g_{R,l,m,k}$ simplifies to

$$g_{R,l,m,k}(x) = \tilde{g}_{l,k}(x) := \bigg(\Big(|x|^2 + \big(-2\langle x, y\rangle + |y|^2\big) \vee (-l)\Big) \vee 0\bigg)^{1/2} \wedge k,$$

which belongs to $\mathscr{A}$, is decreasing with respect to l, and satisfies

$$\lim_{l \to \infty} \tilde{g}_{l,k}(x) = \inf_{l \in \mathbb{N}} \tilde{g}_{l,k}(x) = |x-y| \wedge k \quad \forall\, x \in X.$$

It follows that

$$\limsup_{n \to \infty} \int_X \big(|x-y| \wedge k\big)\, d\mu_n(x) \leq \limsup_{n \to \infty} \int_X \tilde{g}_{l,k}(x)\, d\mu_n(x) = \int_X \tilde{g}_{l,k}(x)\, d\mu(x);$$

passing to the limit as $l \to +\infty$, and recalling that the corresponding "lim inf" inequality is provided by (5.1.33a) of Lemma 5.1.12, we obtain (5.1.43). □

5.2 Transport of measures

If $X_1,\ X_2$ are separable metric spaces, $\mu \in \mathscr{P}(X_1)$, and $\boldsymbol{r} : X_1 \to X_2$ is a Borel (or, more generally, μ-measurable) map, we denote by $\boldsymbol{r}_\#\mu \in \mathscr{P}(X_2)$ the *push-forward of μ through $\boldsymbol{r}$*, defined by

$$\boldsymbol{r}_\#\mu(B) := \mu(\boldsymbol{r}^{-1}(B)) \quad \forall\, B \in \mathscr{B}(X_2). \tag{5.2.1}$$

More generally we have

$$\int_{X_1} f(\boldsymbol{r}(x))\, d\mu(x) = \int_{X_2} f(y)\, d\,\boldsymbol{r}_\#\mu(y) \tag{5.2.2}$$

for every bounded (or $\boldsymbol{r}_\#\mu$-integrable) Borel function $f : X_2 \to \mathbb{R}$. It is easy to check that

$$\nu \ll \mu \quad \Longrightarrow \quad \boldsymbol{r}_\#\nu \ll \boldsymbol{r}_\#\mu \qquad \forall\, \mu, \nu \in \mathscr{P}(X_1). \tag{5.2.3}$$

In the following we will extensively use the following composition rule

$$(\boldsymbol{r} \circ \boldsymbol{s})_\#\mu = \boldsymbol{r}_\#(\boldsymbol{s}_\#\mu) \quad \text{where } \boldsymbol{s} : X_1 \to X_2,\ \boldsymbol{r} : X_2 \to X_3,\ \mu \in \mathscr{P}(X_1). \tag{5.2.4}$$

Furthermore, if $\boldsymbol{r} : X_1 \to X_2$ is a continuous map, then

$$\boldsymbol{r}_\# : \mathscr{P}(X_1) \to \mathscr{P}(X_2) \text{ is continuous w.r.t. the narrow convergence} \tag{5.2.5}$$

and

$$\boldsymbol{r}\big(\operatorname{supp}\mu\big) \subset \operatorname{supp}\boldsymbol{r}_\#\mu = \overline{\boldsymbol{r}\big(\operatorname{supp}\mu\big)}. \tag{5.2.6}$$

Lemma 5.2.1. *Let $\boldsymbol{r}_n : X_1 \to X_2$ be Borel maps uniformly converging to $\boldsymbol{r}$ on compact subsets of X_1 and let $(\mu_n) \subset \mathscr{P}(X_1)$ be a tight sequence narrowly converging to μ. If $\boldsymbol{r}$ is continuous, then $(\boldsymbol{r}_n)_\#\mu_n$ narrowly converge to $\boldsymbol{r}_\#\mu$.*

Proof. Let f be a bounded continuous function in X_2. We will prove the lim inf inequality

$$\liminf_{n\to\infty} \int_{X_2} f\, d(\boldsymbol{r}_n)_\#\mu_n \geq \int_{X_2} f\, d\boldsymbol{r}_\#\mu,$$

as the lim sup simply follows replacing f by $-f$. To this aim, possibly adding to f a constant, we can assume that $f \geq 0$. For any compact set $K \subset X_1$ the uniform convergence of $\boldsymbol{r}_n$ to $\boldsymbol{r}$ on K gives the uniform convergence of $f \circ \boldsymbol{r}_n$ to $f \circ \boldsymbol{r}$ on K, therefore (5.1.15) gives

$$\begin{aligned}
\liminf_{n\to\infty} \int_{X_1} f \circ \boldsymbol{r}_n\, d\mu_n &\geq \liminf_{n\to\infty} \int_K f \circ \boldsymbol{r}_n\, d\mu_n = \liminf_{n\to\infty} \int_K f \circ \boldsymbol{r}\, d\mu_n \\
&\geq \big(-\sup f\big) \sup_n \mu_n(X_1 \setminus K) + \liminf_{n\to\infty} \int_{X_1} f \circ \boldsymbol{r}\, d\mu_n \\
&\geq \big(-\sup f\big) \sup_n \mu_n(X_1 \setminus K) + \int_{X_1} f \circ \boldsymbol{r}\, d\mu.
\end{aligned}$$

Since $\{\mu_n\}_{n\in\mathbb{N}}$ is tight, we can find an increasing sequence of compact set K_m such that $\lim_m \sup_n \mu_n(X_1 \setminus K_m) = 0$. Putting $K = K_m$ in the inequality above and letting $m \uparrow +\infty$ the proof is achieved. □

Lemma 5.2.2 (Tightness criterion). *Let X, X_1, $X_2, \ldots, X_N$ be separable metric spaces and let $\boldsymbol{r}^i : X \to X_i$ be continuous maps such that the product map*

$$\boldsymbol{r} := \boldsymbol{r}^1 \times \boldsymbol{r}^2 \times \ldots \times \boldsymbol{r}^N : X \to X_1 \times \ldots \times X_N \quad \textit{is proper.} \tag{5.2.7}$$

Let $\mathcal{K} \subset \mathscr{P}(X)$ be such that $\mathcal{K}_i := \boldsymbol{r}^i_\#(\mathcal{K})$ is tight in $\mathscr{P}(X_i)$ for $i = 1, \ldots, N$. Then also $\mathcal{K}$ is tight in $\mathscr{P}(X)$.

Proof. For every $\mu \in \mathscr{P}(X)$ we denote by μ_i the measure $\mu_i := \boldsymbol{r}^i_\# \mu$. By definition, for each $\varepsilon > 0$ there exist compact sets $K_i \subset X_i$ such that $\mu_i(X_i \setminus K_i) \le \varepsilon/N$ for any $\mu \in \mathcal{K}$; it follows that $\mu(X \setminus (\boldsymbol{r}^i)^{-1}(K_i)) \le \varepsilon/N$ and

$$\mu\left(X \setminus \bigcap_{i=1}^N (\boldsymbol{r}^i)^{-1}(K_i)\right) \le \sum_{i=1}^N \mu\left(X \setminus (\boldsymbol{r}^i)^{-1}(K_i)\right) \le \varepsilon \qquad \forall \mu \in \mathcal{K}. \tag{5.2.8}$$

On the other hand $\cap_{i=1}^N (\boldsymbol{r}^i)^{-1}(K_i) = \boldsymbol{r}^{-1}(K_1 \times K_2 \times \ldots \times K_N)$, which is compact by (5.2.7). □

For an integer $N \ge 2$ and $i,\ j = 1, \ldots, N$, we denote by π^i, $\pi^{i,j}$ the projection operators defined on the product space $\boldsymbol{X} := X_1 \times \ldots \times X_N$ respectively defined by

$$\pi^i : (x_1, \ldots, x_N) \mapsto x_i \in X_i, \quad \pi^{i,j} : (x_1, \ldots, x_N) \mapsto (x_i, x_j) \in X_i \times X_j. \tag{5.2.9}$$

If $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$, the *marginals* of $\boldsymbol{\mu}$ are the probability measures

$$\mu^i := \pi^i_\# \boldsymbol{\mu} \in \mathscr{P}(X_i), \quad \boldsymbol{\mu}^{i\,j} := \pi^{i,j}_\# \boldsymbol{\mu} \in \mathscr{P}(X_i \times X_j). \tag{5.2.10}$$

If $\mu^i \in \mathscr{P}(X_i)$, $i = 1, \ldots, N$, the class of *multiple plans* with marginals μ^i is defined by

$$\Gamma(\mu^1, \ldots, \mu^N) := \Big\{\boldsymbol{\mu} \in \mathscr{P}(X_1 \times \ldots \times X_N) : \pi^i_\# \boldsymbol{\mu} = \mu^i,\ i = 1, \ldots, N\Big\}. \tag{5.2.11}$$

In the case $N = 2$ a measure $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ is also called *transport plan* between μ^1 and μ^2. Notice also that

$$\Gamma(\mu^1, \mu^2) = \{\mu^1 \times \mu^2\} \quad \text{if either } \mu^1 \text{ or } \mu^2 \text{ is a Dirac mass.} \tag{5.2.12}$$

We will mostly consider multiple plans with $N = 2$ or $N = 3$. To each couple of measures $\mu^1 \in \mathscr{P}(X_1), \mu^2 = \boldsymbol{r}_\# \mu^1 \in \mathscr{P}(X_2)$ linked by a Borel transport map $\boldsymbol{r} : X_1 \to X_2$ we can associate the transport plan

$$\boldsymbol{\mu} := (\boldsymbol{i} \times \boldsymbol{r})_\# \mu^1 \in \Gamma(\mu^1, \mu^2), \quad \boldsymbol{i} \text{ being the identity map on } X_1. \tag{5.2.13}$$

If $\boldsymbol{\mu}$ is representable as in (5.2.13) then we say that $\boldsymbol{\mu}$ is *induced* by $\boldsymbol{r}$. Each transport plan $\boldsymbol{\mu}$ concentrated on a $\boldsymbol{\mu}$-measurable graph in $X_1 \times X_2$ admits the representation (5.2.13) for some μ^1-measurable map $\boldsymbol{r}$, which therefore transports

μ^1 to μ^2 (see, e.g., [9]; the same result holds for Borel graphs and maps if X_1, X_2 are Polish spaces [117, p. 107])

We define also the *inverse* $\boldsymbol{\mu}^{-1} \in \mathscr{P}(X_2 \times X_1)$ of a transport plan $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2)$ by $i_\#\boldsymbol{\mu}$, where $i(x_1, x_2) = (x_2, x_1)$.

Remark 5.2.3. By Lemma 5.2.2, if $X_1, X_2, \cdots, X_N$ are Radon spaces (i.e. each measure $\mu^i \in \mathscr{P}(X_i)$ is tight), $\Gamma(\mu^1, \ldots, \mu^N)$ is compact in $\mathscr{P}(X)$ and not empty, since it contains at least $\mu^1 \times \ldots \times \mu^N$. If for some Borel functions $g_i : X_i \to [0, +\infty]$

$$\int_{X_i} g_i(x_i)\, d\mu^i(x_i) < +\infty \quad i = 1, \ldots, N, \tag{5.2.14}$$

then it is easy to check that $g(x) := \sum_{i=1}^N g_i(x_i)$ defined in the product space $\boldsymbol{X} = X_1 \times X_2 \times \cdots \times X_N$ is uniformly integrable with respect to $\Gamma(\mu^1, \ldots, \mu^N)$.

When X is a separable Hilbert space as in Section 5.1.2, the following result provides a sufficient condition for the convergence of the integrals $\int_{X^2} \langle x_1, x_1 \rangle\, d\mu_h$ even in the case when the measures μ_h do not converge narrowly with respect to the strong topology.

Lemma 5.2.4. *Let $(\boldsymbol{\mu}_n) \subset \mathscr{P}(X \times X)$ be a sequence narrowly converging to $\boldsymbol{\mu}$ in $\mathscr{P}(X \times X_\varpi)$, with*

$$\sup_n \int_{X^2} |x_1|^p + |x_2|^q \, d\boldsymbol{\mu}_n(x_1, x_2) < +\infty, \quad p,\, q \in (1, \infty),\ p^{-1} + q^{-1} = 1. \tag{5.2.15}$$

If either $\pi^1_\#\boldsymbol{\mu}_n$ have uniformly integrable p-moments or $\pi^2_\#\boldsymbol{\mu}_n$ have uniformly integrable q-moments, then

$$\lim_{n\to\infty} \int_{X\times X} \langle x_1, x_2 \rangle \, d\boldsymbol{\mu}_n = \int_{X\times X} \langle x_1, x_2 \rangle \, d\boldsymbol{\mu}.$$

Proof. We assume to fix the ideas that $\pi^2_\#\boldsymbol{\mu}_n$ have uniformly integrable q-moments and we show that the function $(x_1, x_2) \mapsto g(x_1, x_2) := |x_1| \cdot |x_2|$ is uniformly integrable. For any $k, m \in \mathbb{N}$ we have

$$g(x_1, x_2) \geq k, \quad |x_2| \leq m \quad \Rightarrow \quad |x_1| \geq k/m$$

and therefore

$$\int_{\{g \geq k\}} g\, d\boldsymbol{\mu}_n \leq m \int_{\{|x_1| \geq k/m\}} |x_1|\, d\pi^1_\#\boldsymbol{\mu}_n + C\Big(\int_{\{|x_2| \geq m\}} |x_2|^q\, d\pi^2_\#\boldsymbol{\mu}_n\Big)^{1/q}$$

where $C^p := \sup_n \int_X |x_1|^p \, d\boldsymbol{\mu}_n$. Taking the supremum w.r.t. n and the lim sup as $k \to \infty$, since $\pi^1_\#\boldsymbol{\mu}_n$ has uniformly integrable 1-moments by (5.1.20) we have

$$\limsup_{k\to\infty} \sup_n \int_{\{g \geq k\}} g\, d\boldsymbol{\mu}_n \leq \sup_n C\Big(\int_{\{|x_2| \geq m\}} |x_2|^q\, d\pi^2_\#\boldsymbol{\mu}_n\Big)^{1/q}$$

Letting $m \to \infty$ we conclude.

In the finite dimensional case (or even if $\boldsymbol{\mu}_n \to \boldsymbol{\mu}$ in $\mathscr{P}(X \times X)$) we conclude immediately, since the map $(x_1, x_2) \mapsto \langle x_1, x_2 \rangle$ is continuous in $X \times X$.

In the infinite dimensional case, let $\overline{B_R}$ be the centered closed ball of radius R in X which is compact in X_ϖ. The map $(x_1, x_2) \mapsto \langle x_1, x_2 \rangle$ is continuous in each closed set $X \times B_R$ with respect to the $X \times X_\varpi$ topology and (5.2.15) yields

$$\limsup_{n,R\to\infty} \boldsymbol{\mu}_n\big(X^2 \setminus (X \times B_R)\big) = 0.$$

Therefore we conclude by invoking Proposition 5.1.10. □

5.3 Measure-valued maps and disintegration theorem

Let X, Y be separable metric spaces and let $x \in X \mapsto \mu_x \in \mathscr{P}(Y)$ be a measure-valued map. We say that μ_x is a Borel map if $x \mapsto \mu_x(B)$ is a Borel map for any Borel set $B \subset Y$, or equivalently if this property holds for any open set $A \subset Y$. By the monotone class theorem we have also that

$$x \in X \mapsto \int_Y f(x,y)\,d\mu_x(y) \text{ is Borel} \tag{5.3.1}$$

for every bounded (or nonnegative) Borel function $f : X \times Y \to \mathbb{R}$.

By (5.3.1) the formula

$$\boldsymbol{\mu}(f) = \int_X \left(\int_Y f(x,y)\,d\mu_x(y) \right) d\nu(x)$$

defines for any $\nu \in \mathscr{P}(X)$ a unique measure $\boldsymbol{\mu} \in \mathscr{P}(X \times Y)$, that will be denoted by $\int_X \mu_x\,d\nu(x)$. Actually any $\boldsymbol{\mu} \in \mathscr{P}(X \times Y)$ whose first marginal is ν can be represented in this way. This is implied by the so-called *disintegration theorem* (related to the existence of conditional probability measures in Probability), see for instance [55, III-70].

Theorem 5.3.1 (Disintegration). *Let $\boldsymbol{X}$, X be Radon separable metric spaces, $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$, let $\pi : \boldsymbol{X} \to X$ be a Borel map and let $\nu = \pi_\# \boldsymbol{\mu} \in \mathscr{P}(X)$. Then there exists a ν-a.e. uniquely determined Borel family of probability measures $\{\mu_x\}_{x\in X} \subset \mathscr{P}(\boldsymbol{X})$ such that*

$$\mu_x(\boldsymbol{X} \setminus \pi^{-1}(x)) = 0 \quad \textit{for } \nu\textit{-a.e. } x \in X \tag{5.3.2}$$

and

$$\int_{\boldsymbol{X}} f(\boldsymbol{x})\,d\boldsymbol{\mu}(\boldsymbol{x}) = \int_X \left(\int_{\pi^{-1}(x)} f(\boldsymbol{x})\,d\mu_x(\boldsymbol{x}) \right) d\nu(x) \tag{5.3.3}$$

for every Borel map $f : \boldsymbol{X} \to [0, +\infty]$.

In particular, when $\boldsymbol{X} := X_1 \times X_2, X := X_1$, $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2)$, $\nu = \mu^1 = \pi^1_\# \boldsymbol{\mu}$, we can canonically identify each fiber $(\pi^1)^{-1}(x_1)$ with X_2 and find a Borel family of probability measures $\{\mu_{x_1}\}_{x_1 \in X_1} \subset \mathscr{P}(X_2)$ (which is μ^1-a.e. uniquely determined) such that $\boldsymbol{\mu} := \int_{X_1} \mu_{x_1}\, d\mu^1(x_1)$.

As an application of the disintegration theorem we can prove existence, and in some cases uniqueness, of multiple plans with given marginals.

Lemma 5.3.2. *Let X_1, X_2, X_3 be Radon separable metric spaces and let $\boldsymbol{\gamma}^{1\,2} \in \mathscr{P}(X_1 \times X_2)$, $\boldsymbol{\gamma}^{1\,3} \in \mathscr{P}(X_1 \times X_3)$ such that $\pi^1_\# \boldsymbol{\gamma}^{1\,2} = \pi^1_\# \boldsymbol{\gamma}^{1\,3} = \mu^1$. Then there exists $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2 \times X_3)$ such that*

$$\pi^{1,2}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,2}, \qquad \pi^{1,3}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,3}. \tag{5.3.4}$$

Moreover, if $\boldsymbol{\gamma}^{1\,2} = \int \gamma^{1\,2}_{x_1}\, d\mu^1$, $\boldsymbol{\gamma}^{1\,3} = \int \gamma^{1\,3}_{x_1}\, d\mu^1$ and $\boldsymbol{\mu} = \int \boldsymbol{\mu}_{x_1}\, d\mu^1$ are the disintegrations of $\boldsymbol{\gamma}^{1\,2}$, $\boldsymbol{\gamma}^{1\,3}$ and $\boldsymbol{\mu}$ with respect to μ^1, (5.3.4) *is equivalent to*

$$\boldsymbol{\mu}_{x_1} \in \Gamma(\gamma^{1\,2}_{x_1}, \gamma^{1\,3}_{x_1}) \subset \mathscr{P}(X_2 \times X_3) \quad \textit{for } \mu^1\textit{-a.e. } x_1 \in X_1. \tag{5.3.5}$$

In particular (5.2.12) *implies that the measure $\boldsymbol{\mu}$ is unique if either $\boldsymbol{\gamma}^{1\,2}$ or $\boldsymbol{\gamma}^{1\,3}$ are induced by a transport. We denote by $\Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ the subset of plans $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2 \times X_3)$ satisfying* (5.3.4).

Proof. With the notation introduced in the statement of the theorem, the measure $\boldsymbol{\mu}$ whose disintegration w.r.t. x_1 is

$$\int_{X_1} \gamma^{1\,2}_{x_1} \times \gamma^{1\,3}_{x_1}\, d\mu^1(x_1)$$

has the required properties.

Now we prove the equivalence between (5.3.4) and (5.3.5). If $\boldsymbol{\mu}$ satisfies $\pi^{1,2}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,2}$ and $\pi^{1,3}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^{1\,3}$, then

$$\boldsymbol{\gamma}^{1\,2} = \pi^{1,2}_\# \boldsymbol{\mu} = \int_{X_1} \pi^2_\# \boldsymbol{\mu}_{x_1}\, d\mu^1(x_1)$$

and the uniqueness of the disintegration gives $\pi^2_\# \boldsymbol{\mu}_{x_1} = \gamma^{1\,2}_{x_1}$ for μ^1-a.e. $x_1 \in X_1$. A similar argument gives that $\pi^3_\# \boldsymbol{\mu}_{x_1} = \gamma^{1\,3}_{x_1}$ for μ^1-a.e. $x_1 \in X_1$.

Conversely, let us suppose that $\boldsymbol{\mu}$ satisfies (5.3.5) and let $f : X_1 \times X_2 \to \mathbb{R}$

be a bounded Borel function; the computation

$$\begin{aligned}\int_{X_1\times X_2} f(x_1,x_2)\,d\pi^{1,2}_\#\boldsymbol{\mu} &= \int_{X_1\times X_2\times X_3} f(x_1,x_2)\,d\boldsymbol{\mu}(x_1,x_2,x_3)\\ &= \int_{X_1}\Big(\int_{X_2\times X_3} f(x_1,x_2)\,d\boldsymbol{\mu}_{x_1}(x_2,x_3)\Big)\,d\mu^1(x_1)\\ &= \int_{X_1}\Big(\int_{X_2} f(x_1,x_2)\,d\pi^2_\#\boldsymbol{\mu}_{x_1}(x_2)\Big)\,d\mu^1(x_1)\\ &= \int_{X_1}\Big(\int_{X_2} f(x_1,x_2)\,d\gamma^{1\,2}_{x_1}(x_2)\Big)\,d\mu^1(x_1)\\ &= \int_{X_1\times X_2} f(x_1,x_2)\,d\boldsymbol{\gamma}^{1\,2}(x_1,x_2)\end{aligned}$$

shows that $\pi^{1,2}_\#\boldsymbol{\mu}=\boldsymbol{\gamma}^{1\,2}$. A similar argument proves that $\pi^{1,3}_\#\boldsymbol{\mu}=\boldsymbol{\gamma}^{1\,3}$. □

Remark 5.3.3 (Composition of plans). An analogous situation occurs when $\boldsymbol{\gamma}^{1\,2}\in\mathscr{P}(X_1\times X_2)$ and $\boldsymbol{\gamma}^{2\,3}\in\mathscr{P}(X_2\times X_3)$. In this case we say that

$$\boldsymbol{\mu}\in\Gamma^2(\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3})\quad\text{if}\quad \pi^{1,2}_\#\boldsymbol{\mu}=\boldsymbol{\gamma}^{1\,2},\quad \pi^{2,3}_\#\boldsymbol{\mu}=\boldsymbol{\gamma}^{2\,3}. \tag{5.3.6}$$

Of course, $\Gamma^2(\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3})$ is not empty iff $\pi^2_\#\boldsymbol{\gamma}^{1\,2}=\pi^1_\#\boldsymbol{\gamma}^{2\,3}$. In this case, the measure $\pi^{1,3}_\#\boldsymbol{\mu}$, with $\boldsymbol{\mu}\in\Gamma^2(\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3})$ constructed as in the proof of Lemma 5.3.2, belongs by construction to $\Gamma(\mu^1,\mu^3)$; it will be called *composition* of $\boldsymbol{\gamma}^{2\,3}$ and $\boldsymbol{\gamma}^{1\,2}$ and denoted by $\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}$. We have then

$$\int_{X_1\times X_3} f(x_1,x_3)\,d(\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}) = \int_{X_2}\Big(\int_{X_1\times X_3} f(x_1,x_3)\,d\gamma^{1\,2}_{x_2}\times\gamma^{2\,3}_{x_2}\Big)\,d\mu^2(x_2) \tag{5.3.7}$$

for any bounded Borel function $f:X_1\times X_3\to\mathbb{R}$. The name is justified since in the case $\boldsymbol{\gamma}^{1\,2},\boldsymbol{\gamma}^{2\,3}$ are induced by the transports $\boldsymbol{r}^{1\,2},\boldsymbol{r}^{2\,3}$, then the plan $\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}$ is induced by the composition map $\boldsymbol{r}^{2\,3}\circ\boldsymbol{r}^{1\,2}$: this fact can be easily checked starting from (5.3.7)

$$\begin{aligned}\int_{X_1\times X_3} f(x_1,x_3)\,d(\boldsymbol{\gamma}^{2\,3}\circ\boldsymbol{\gamma}^{1\,2}) &= \int_{X_2}\Big(\int_{X_1} f(x_1,\boldsymbol{r}^{2\,3}(x_2))\,d\gamma^{1\,2}_{x_2}(x_1)\Big)\,d\mu^2(x_2)\\ &= \int_{X_1\times X_2} f(x_1,\boldsymbol{r}^{2\,3}(x_2))\,d\boldsymbol{\gamma}^{1\,2}(x_1,x_2)\\ &= \int_{X_1} f(x_1,\boldsymbol{r}^{2\,3}(\boldsymbol{r}^{1\,2}(x_1)))\,d\mu^1(x_1).\end{aligned}$$

Notice that by (5.2.12) this construction is canonical only if either $(\boldsymbol{\gamma}^{1\,2})^{-1}$ or $\boldsymbol{\gamma}^{2\,3}$ are induced by a transport.

In the proof of the completeness of the Wasserstein distance we will also need the following useful extensions of Lemma 5.3.2 to a countable product of Radon spaces.

Lemma 5.3.4. *Let X_i, $i \in \mathbb{N}$, be a sequence of Radon separable metric spaces, $\mu^i \in \mathscr{P}(X_i)$ and $\boldsymbol{\alpha}^{i(i+1)} \in \Gamma(\mu^i, \mu^{i+1}), \boldsymbol{\beta}^{1\,i} \in \Gamma(\mu^1, \mu^i)$. Let $\boldsymbol{X}_\infty := \Pi_{i\in\mathbb{N}} X_i$, with the canonical product topology. Then there exist $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathscr{P}(\boldsymbol{X}_\infty)$ such that*

$$\pi^{i,i+1}_\# \boldsymbol{\mu} = \boldsymbol{\alpha}^{i\,(i+1)}, \quad \pi^{1,i}_\# \boldsymbol{\nu} = \boldsymbol{\beta}^{1\,i} \quad \forall\, i \in \mathbb{N}. \tag{5.3.8}$$

Proof. Let $\boldsymbol{X}_n := \Pi_{i=1}^n X_i = \boldsymbol{X}_{n-1} \times X_n$ and let $\pi^{\boldsymbol{n}} : \boldsymbol{X}_m \to \boldsymbol{X}_n$, $m \geq n$, be the projection onto the first n coordinates. In order to show the existence of $\boldsymbol{\mu}$, we set $\boldsymbol{\mu}^2 := \boldsymbol{\alpha}^{1\,2}$ and we apply recursively Lemma 5.3.2 and Remark 5.3.3 with $\boldsymbol{\mu}^n \in \mathscr{P}(\boldsymbol{X}_{n-1} \times X_n), \boldsymbol{\alpha}^{n(n+1)} \in \mathscr{P}(X_n \times X_{n+1})$, $n \geq 2$, to obtain a sequence $\boldsymbol{\mu}^{n+1} \in \mathscr{P}(\boldsymbol{X}_{n+1})$ satisfying

$$\pi^{\boldsymbol{n}}_\# \boldsymbol{\mu}^{n+1} = \boldsymbol{\mu}^n, \quad \pi^{n,n+1}_\# \boldsymbol{\mu}^{n+1} = \boldsymbol{\alpha}^{n(n+1)}.$$

Kolmogorov's Theorem [55, §51] provides a measure $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X}_\infty)$ such that $\pi^{\boldsymbol{n}}_\# \boldsymbol{\mu} = \boldsymbol{\mu}^n$ and therefore

$$\pi^{n-1,n}_\# \boldsymbol{\mu} = \pi^{n-1,n}_\# \big(\pi^{\boldsymbol{n}}_\# \boldsymbol{\mu}\big) = \pi^{n-1,n}_\# \boldsymbol{\mu}^n = \boldsymbol{\alpha}^{(n-1)\,n}.$$

The existence of $\boldsymbol{\nu}$ can be proved by a similar argument, by setting $\boldsymbol{\nu}^2 := \boldsymbol{\beta}^{1\,2}$ and by applying recursively Lemma 5.3.2 to $\boldsymbol{\nu}^n \in \mathscr{P}(X_1 \times \boldsymbol{X}_{n-1}), \boldsymbol{\beta}^{1(n+1)} \in \mathscr{P}(X_1 \times X_{n+1})$, $n \geq 2$: we can find a sequence $\boldsymbol{\nu}^{n+1} \in \mathscr{P}(\boldsymbol{X}_{n+1})$ satisfying

$$\pi^{\boldsymbol{n}}_\# \boldsymbol{\nu}^{n+1} = \boldsymbol{\nu}^n, \quad \pi^{1,n+1}_\# \boldsymbol{\nu}^{n+1} = \boldsymbol{\beta}^{1(n+1)}.$$

Kolmogorov's Theorem [55, §51] provides a measure $\boldsymbol{\nu} \in \mathscr{P}(\boldsymbol{X}_\infty)$ such that $\pi^{\boldsymbol{n}}_\# \boldsymbol{\nu} = \boldsymbol{\nu}^n$ and therefore

$$\pi^{1,n}_\# \boldsymbol{\nu} = \pi^{1,n}_\# \big(\pi^{\boldsymbol{n}}_\# \boldsymbol{\nu}\big) = \pi^{1,n}_\# \boldsymbol{\nu}^n = \boldsymbol{\beta}^{1\,n}$$ □

5.4 Convergence of plans and convergence of maps

In this section we investigate the relation between the convergence of maps and the convergence of the associated plans.

Let us first recall that if $X, Y_1, \ldots, Y_k$ are separable metric spaces with $\boldsymbol{Y} := Y_1 \times \ldots \times Y_k$, $\mu \in \mathscr{P}(X)$, and $\boldsymbol{r}_i : X \to Y_i$, $i = 1, \cdots, k$, then the product map

$$\begin{gathered} \boldsymbol{r} := (\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_k) : X \to \boldsymbol{Y} \quad \text{is Borel (μ-measurable) iff} \\ \text{each map } \boldsymbol{r}_i : X \to Y_i \text{ is Borel (resp. μ-measurable).} \end{gathered} \tag{5.4.1}$$

In particular, if $\boldsymbol{r}, \boldsymbol{s} : X \to Y$ are μ-measurable, then their distance $d_Y(\boldsymbol{r}(\cdot), \boldsymbol{s}(\cdot))$ is a μ-measurable real map.

We can thus define the convergence in measure of a sequence of μ-measurable maps $\boldsymbol{r}_n : X \to Y$ to a μ-measurable map $\boldsymbol{r}$ by asking that

$$\lim_{n\to\infty} \mu\big(\{x \in X : d_Y(\boldsymbol{r}_n(x), \boldsymbol{r}(x)) > \varepsilon\}\big) = 0 \quad \forall\, \varepsilon > 0. \tag{5.4.2}$$

We can also introduce the L^p spaces (see e.g. [7])

$$L^p(\mu; Y) := \Big\{\boldsymbol{r} : X \to Y \ \mu\text{-measurable} : \int_X d_Y^p(\boldsymbol{r}(x), \bar y)\, d\mu(x) < +\infty \\ \text{for some (and thus any) } \bar y \in Y\Big\}. \tag{5.4.3}$$

with the distance

$$\boldsymbol{d}(\boldsymbol{r}, \boldsymbol{s})_{L^p(\mu;Y)} := \Big(\int_X d_Y^p(\boldsymbol{r}(x), \boldsymbol{s}(y))\, d\mu(x)\Big)^{1/p}; \tag{5.4.4}$$

it is easy to check that $L^p(\mu; Y)$ is complete iff Y is complete. When Y is a (separable) Hilbert space and $p \geq 1$, then the above distance is induced by the norm

$$\|\boldsymbol{r}\|_{L^p(\mu;Y)} := \Big(\int_X |\boldsymbol{r}(x)|_Y^p\, d\mu(x)\Big)^{1/p}; \tag{5.4.5}$$

for $\boldsymbol{r} \in L^1(\mu; Y)$ the (vector valued) integral $\int_X \boldsymbol{r}(x)\, d\mu(x) \in Y$ of $\boldsymbol{r}$ is well defined and satisfies

$$\int_X \langle y, \boldsymbol{r}(x)\rangle\, d\mu(x) = \langle y, \int_X \boldsymbol{r}(x)\, d\mu(x)\rangle \quad \forall\, y \in Y, \tag{5.4.6}$$

$$\phi\Big(\int_X \boldsymbol{r}(x)\, d\mu(x)\Big) \leq \int_X \phi(\boldsymbol{r}(x))\, d\mu(x) \tag{5.4.7}$$

for every proper, convex and l.s.c. function $\phi : Y \to (-\infty, +\infty]$ (Jensen's inequality).

In the following lemma we consider first the case when the reference measure μ is fixed, and show the equivalence between narrow convergence of the plans $(\boldsymbol{i} \times \boldsymbol{r}_n)_\# \mu$ and convergence in measure and in $L^p(\mu; Y)$ of $\boldsymbol{r}_n$, when the limiting plan is induced by a transport $\boldsymbol{r}$.

Lemma 5.4.1 (Narrow convergence of plans and convergence in measure). *Let $\mu \in \mathscr{P}(X)$ and let $\boldsymbol{r}_n, \boldsymbol{r} : X \to Y$ be Borel maps. Then the plans $(\boldsymbol{i} \times \boldsymbol{r}_n)_\# \mu$ narrowly convergence to $(\boldsymbol{i} \times \boldsymbol{r})_\# \mu$ in $\mathscr{P}(X \times Y)$ as $n \to \infty$ if and only if $\boldsymbol{r}_n$ converge in measure to $\boldsymbol{r}$.*
Moreover, the measures $(\boldsymbol{r}_n)_\# \mu$ have uniformly integrable p-moments iff $\boldsymbol{r}_n$ converges to $\boldsymbol{r}$ in $L^p(\mu; Y)$.

Proof. Since for every Borel map $\boldsymbol{s} : X \to Y$

$$\int_{X\times Y} \varphi(x, y)\, d(\boldsymbol{i} \times \boldsymbol{s})_\# \mu = \int_X \varphi(x, \boldsymbol{s}(x))\, d\mu(x) \qquad \forall \varphi \in C_b^0(X \times Y)$$

and convergence in measure is stable by composition with continuous functions, it is immediate to check that convergence in measure of the maps implies narrow convergence of the plans.
The converse implication can be obtained as follows: fix $\epsilon > 0$, a continuous function ψ_ϵ with $0 \le \psi_\varepsilon \le 1$, $\psi_\varepsilon(0) = 0$ and $\psi_\varepsilon(t) = 1$ whenever $|t| > \epsilon$ and a continuous function $\tilde{\boldsymbol{r}}$ such that $\mu(\{\boldsymbol{r} \neq \tilde{\boldsymbol{r}}\}) < \epsilon$. Then, using the test function $\varphi_\epsilon(x, y) = \psi_\epsilon(d_Y(y, \tilde{\boldsymbol{r}}(x)))$ we obtain

$$\begin{aligned}\limsup_{n\to\infty} \mu(\{d_Y(\boldsymbol{r}_n, \tilde{\boldsymbol{r}}) > \epsilon\}) &\le \limsup_{n\to\infty} \int_{X\times Y} \varphi_\epsilon \, d(\boldsymbol{i} \times \boldsymbol{r}_n)_\# \mu = \int_{X\times Y} \varphi_\epsilon \, d(\boldsymbol{i} \times \boldsymbol{r})_\# \mu \\ &= \int_X \psi_\epsilon(d_Y(\boldsymbol{r}(x), \tilde{\boldsymbol{r}}(x))) \, d\mu(x) \le \epsilon.\end{aligned}$$

Taking into account our choice of $\tilde{\boldsymbol{r}}$ we obtain $\limsup\limits_{n\to\infty} \mu(\{d_Y(\boldsymbol{r}_n, \boldsymbol{r}) > \epsilon\}) \le 2\epsilon$.

The second part of the lemma follows easily by Vitali dominated convergence theorem and the identities

$$\begin{aligned}\lim_{n\to\infty} \int_X d_Y^p(\boldsymbol{r}_n(x), \bar{y}) \, d\mu(x) &= \lim_{n\to\infty} \int_Y d_Y^p(y, \bar{y}) \, d\big((\boldsymbol{r}_n)_\# \mu\big)(y) \\ = \int_X d_Y^p(\boldsymbol{r}(x), \bar{y}) \, d\mu(x) &= \int_Y d_Y^p(y, \bar{y}) \, d\big(\boldsymbol{r}_\# \mu\big)(y),\end{aligned} \tag{5.4.8}$$

which hold either if $\boldsymbol{r}_n$ converges to $\boldsymbol{r}$ in $L^p(\mu; Y)$ or if the family $(\boldsymbol{r}_n)_\# \mu$, $n \in \mathbb{N}$, has uniformly integrable p-moments. □

In the rest of this section we assume that X is a separable Hilbert space as in Section 5.1.2.

Definition 5.4.2 (Barycentric projection). *The barycentric projection $\bar{\boldsymbol{\gamma}} : X \to X$ of a plan $\boldsymbol{\gamma} \in \mathscr{P}(X \times X)$, which admits the disintegration $\boldsymbol{\gamma} = \int_X \gamma_{x_1} \, d\mu(x_1)$ with respect to its first marginal $\mu = \pi^1_\# \boldsymbol{\gamma}$, is defined as*

$$\bar{\boldsymbol{\gamma}}(x_1) := \int_X x_2 \, d\gamma_{x_1}(x_2) \quad \textit{for } \mu\textit{-a.e. } x_1 \in X \tag{5.4.9}$$

provided γ_{x_1} has finite first moment for μ-a.e. x_1.

Assume that we are given maps $\boldsymbol{v}_n \in L^p(\mu_n; X)$: here we have to be careful in the meaning of the convergence of vectors $\boldsymbol{v}_n$, which belong to different L^p-spaces. Two approaches seem natural:

(i) we can consider the narrow limit in $\mathscr{P}(X_\varpi)$ of the X-valued measures $\boldsymbol{\nu}_n := \boldsymbol{v}_n \mu_n$ (component by component);

(ii) we can consider the limit $\boldsymbol{\gamma}$ of the associated plans $\boldsymbol{\gamma}_n := (\boldsymbol{i} \times \boldsymbol{v}_n)_\# \mu_n$ in $\mathscr{P}_2(X_\varpi \times X_\varpi)$, recovering a limit vector v by taking the barycenter of $\boldsymbol{\gamma}$.

In fact, these two approaches yields equivalent notions: we formalize the point (i) in the following definition, and then we see that it coincides with (ii).

Definition 5.4.3. *Let $(\boldsymbol{e}_j)$ be an orthonormal basis of X, let $(\mu_n) \subset \mathscr{P}(X)$ be narrowly converging to μ in $\mathscr{P}(X_\varpi)$ and let $\boldsymbol{v}_n \in L^1(\mu_n; X)$. We say that $\boldsymbol{v}_n$ weakly converge to $\boldsymbol{v} \in L^1(\mu; X)$ if*

$$\lim_{n\to\infty} \int_X \zeta(x)\langle \boldsymbol{e}_j, \boldsymbol{v}_n(x)\rangle \, d\mu_n(x) = \int_X \zeta(x)\langle \boldsymbol{e}_j, \boldsymbol{v}(x)\rangle \, d\mu(x) \tag{5.4.10}$$

for every $\zeta \in \mathrm{Cyl}(X)$ and any $j \in \mathbb{N}$. We say that $\boldsymbol{v}_n$ converges strongly *to $\boldsymbol{v}$ in L^p, $p > 1$, if* (5.4.10) *holds and*

$$\limsup_{n\to\infty} \|\boldsymbol{v}_n\|_{L^p(\mu_n;X)} \le \|\boldsymbol{v}\|_{L^p(\mu;X)}. \tag{5.4.11}$$

It is easy to check that the limit $\boldsymbol{v}$, if it exists, is unique.

Theorem 5.4.4. *Let $p > 1$, let $(\mu_n) \subset \mathscr{P}(X)$ be narrowly converging to μ in $\mathscr{P}(X_\varpi)$ and let $\boldsymbol{v}_n \in L^p(\mu_n; X)$ be such that*

$$\sup_{n\in\mathbb{N}} \int_X |\boldsymbol{v}_n(x)|^p \, d\mu_n(x) < +\infty. \tag{5.4.12}$$

Then the following statements hold:

(i) *The family of plans $\boldsymbol{\gamma}_n := (\boldsymbol{i} \times \boldsymbol{v}_n)_\# \mu_n$ has limit points in $\mathscr{P}(X_\varpi \times X_\varpi)$ as $n \to \infty$ and the sequence $(\boldsymbol{v}_n)$ has weak limit points as $n \to \infty$.*

(ii) *$\boldsymbol{v}_n$ weakly converge to $\boldsymbol{v} \in L^p(\mu; X)$ according to Definition 5.4.3 if and only if $\boldsymbol{v}$ is the barycenter of any limit point of the sequence of plans $\boldsymbol{\gamma}_n$ in $\mathscr{P}(X_\varpi \times X_\varpi)$; in this case*

$$\liminf_{n\to\infty} \int_X g(\boldsymbol{v}_n(x)) \, d\mu_n(x) \ge \int_X g(\boldsymbol{v}(x)) \, d\mu(x), \tag{5.4.13}$$

for every convex *and l.s.c. function $g : X \to (-\infty, +\infty]$.*

(iii) *If $\boldsymbol{v}_n$ strongly converge to $\boldsymbol{v}$ in L^p then $\boldsymbol{\gamma}_n$ narrowly converge to $(\boldsymbol{i} \times \boldsymbol{v})_\# \mu$ in $\mathscr{P}(X_\varpi \times X)$ and*

$$\lim_{n\to\infty} \|v_n\|^p_{L^p(\mu_n;X)} = \lim_{n\to\infty} \int_{X^2} |x_2|^p \, d\boldsymbol{\gamma}_n = \|v\|^p_{L^p(\mu;X)}. \tag{5.4.14}$$

If, in addition, μ_n narrowly converge to μ in $\mathscr{P}(X)$ then $\boldsymbol{\gamma}_n$ narrowly converge to $(\boldsymbol{i} \times \boldsymbol{v})_\# \mu$ in $\mathscr{P}(X \times X)$. Finally, if μ_n has uniformly integrable p-moments, then

$$\lim_{n\to\infty} \int_X f(x, v_n(x)) \, d\mu_n(x) = \int_X f(x, v(x)) \, d\mu(x), \tag{5.4.15}$$

for every continuous function $f : X \times X \to \mathbb{R}$ with p-growth according to (5.1.21).

Proof. (i) Observe that Lemma 5.2.2 ensures that the sequence $(\boldsymbol{\gamma}_n)$ is relatively compact in $\mathscr{P}(X_\varpi \times X_\varpi)$, since (see also Lemma 5.1.12) $\pi^1_\# \boldsymbol{\gamma}_n = \mu_n \to \mu$ in $\mathscr{P}(X_\varpi)$ and $\pi^2_\# \boldsymbol{\gamma}_n$ is relatively compact in $\mathscr{P}(X_\varpi)$ by (5.4.12).

(ii) For every $j \in \mathbb{N}$ and any $\varphi \in \mathrm{Cyl}(X)$ we have

$$\int_X \varphi(x)\langle \boldsymbol{e}_j, \boldsymbol{v}_n(x)\rangle \, d\mu_n(x) = \int_{X\times X} \varphi(x_1)\langle \boldsymbol{e}_j, x_2\rangle \, d\boldsymbol{\gamma}_n(x_1,x_2).$$

Since $|x_2|$ is uniformly integrable w.r.t. $(\boldsymbol{\gamma}_n)$, Proposition 5.1.10 yields

$$\begin{aligned}\lim_{k\to\infty} \int_{X\times X} \varphi(x_1)\langle \boldsymbol{e}_j, x_2\rangle \, d\boldsymbol{\gamma}_{n_k}(x_1,x_2) &= \int_{X\times X} \varphi(x_1)\langle \boldsymbol{e}_j, x_2\rangle \, d\boldsymbol{\gamma}(x_1,x_2)\\ &= \int_X \varphi(x_1)\langle \boldsymbol{e}_j, \bar{\boldsymbol{\gamma}}(x_1)\rangle \, d\mu(x_1)\end{aligned}$$

for every subsequence $(\boldsymbol{\gamma}_{n_k})$ converging to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_\varpi \times X_\varpi)$. Therefore, (5.4.10) holds if and only if $\boldsymbol{v} = \bar{\boldsymbol{\gamma}}$ for every limit point $\boldsymbol{\gamma}$.
(5.4.13) follows by Jensen's inequality and (5.1.33a), being g weakly lower semicontinuous.

(iii) If $\boldsymbol{\gamma}$ is a limit point of $\boldsymbol{\gamma}_n$ as in (ii), taking into account that $\boldsymbol{v} = \bar{\boldsymbol{\gamma}}$ we have

$$\int_{X\times X} |x_2|^p \, d\boldsymbol{\gamma} \le \liminf_{n\to\infty} \int_{X\times X} |x_2|^p \, d\boldsymbol{\gamma}_n = \int_X |\bar{\boldsymbol{\gamma}}|^p \, d\mu.$$

Hence, by disintegrating $\boldsymbol{\gamma}$ with respect to x_1 we get

$$\int_X \left(\int_X |x_2|^p \, d\boldsymbol{\gamma}_{x_1}\right)^p - |\bar{\boldsymbol{\gamma}}(x_1)|^p \, d\mu(x_1) = 0$$

and so Jensen's inequality gives that $\boldsymbol{\gamma}_{x_1} = \delta_{\boldsymbol{v}(x_1)}$ for μ-a.e. x_1, i.e. $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{v})_\# \mu$. This proves the narrow convergence of $\boldsymbol{\gamma}_n$ to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_\varpi \times X_\varpi)$ and (5.4.14). By applying Theorem 5.1.13 we obtain that the second marginals of $\boldsymbol{\gamma}_n$ are also converging in the stronger narrow topology of $\mathscr{P}(X)$. Lemma 5.2.2 yields that the sequence $\boldsymbol{\gamma}_n$ is tight in $\mathscr{P}(X_\varpi \times X)$ and therefore converges to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_\varpi \times X)$. The last part of the statement follows again by Lemma 5.2.2 and Lemma 5.1.7. □

5.5 Approximate differentiability and area formula in Euclidean spaces

Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a function. Then, denoting by $\Sigma = D(\nabla f)$ the Borel set where f is differentiable, there is a sequence of sets $\Sigma_n \uparrow \Sigma$ such that $f|_{\Sigma_n}$ is a Lipschitz function for any n (see [65, 3.1.8]). Therefore the well-known area formula for Lipschitz maps (see for instance [64, 65]) extends to this general class of maps and reads as follows:

$$\int_\Sigma h(x)|\mathrm{det}\nabla f|(x)\, dx = \int_{\mathbb{R}^d} \sum_{x\in\Sigma\cap f^{-1}(y)} h(x)\, dy \tag{5.5.1}$$

for any Borel function $h : \mathbb{R}^d \to [0, +\infty]$. Actually, these results hold more generally for the *approximately differentiable maps*, whose definition and main properties are recalled below.

Definition 5.5.1 (Approximate limit and approximate differential). *Let $\Omega \subset \mathbb{R}^d$ be an open set and $f : \Omega \to \mathbb{R}^m$. We say that f has an approximate limit (respectively, approximate differential) at $x \in \Omega$ if there exists a function $g : \Omega \to \mathbb{R}^m$ continuous (resp. differentiable) at x such that the set $\{f \neq g\}$ has density 0 at x. In this case the approximate limit (resp. approximate differential) will be denoted by $\tilde{f}(x)$ (resp. $\tilde{\nabla} f(x)$).*

It is immediate to check that the definition above is well posed, i.e. it does not depend on the choice of g. An equivalent and more traditional (see [65]) definition of approximate limit goes as follows: we say that $z \in \mathbb{R}^m$ is the approximate limit of f at x if all sets

$$\{y : \ |f(y) - z| > \epsilon\} \qquad \epsilon > 0$$

have density 0 at x. Analogously, a linear map $L : \mathbb{R}^d \to \mathbb{R}^m$ is said to be the approximate differential of f at x if f has an approximate limit at x and all sets

$$\left\{ y : \ \frac{|f(y) - \tilde{f}(x) - L(y - x)|}{|y - x|} > \epsilon \right\} \qquad \epsilon > 0$$

have density 0 at x.
The latter definitions have the advantage of being more intrinsic and do not rely on an auxiliary function g. We have chosen the former definitions because they are more practical, as we will see, for our purposes. For instance, a property that immediately follows by the definition, and that will be used very often in the sequel, is the *locality* principle: if f has approximate limit $\tilde{f}(x)$ (resp. approximate differential $\tilde{\nabla} f(x)$) for any $x \in B$, with B Borel, then g has approximate limit (resp. approximate differential) equal to $\tilde{f}(x)$ (resp. $\tilde{\nabla} f(x)$) for $\mathscr{L}^d$-a.e. $x \in B$, and precisely at all points x where the coincidence set $B \cap \{f = g\}$ has density 1.

Remark 5.5.2. Recall that if $f : \Omega \to \mathbb{R}^m$ is $\mathscr{L}^d$-measurable, then it has approximate limit $\tilde{f}(x)$ at $\mathscr{L}^d$-a.e. $x \in \Omega$ and $f(x) = \tilde{f}(x)$ $\mathscr{L}^d$-a.e.. In particular every Lebesgue measurable set B has density 1 at $\mathscr{L}^d$-a.e. point of B.

Denoting by Σ_f the Borel set (see for instance [7]) of points where f is approximately differentiable, it is still true by [65, 3.1.8] that there exists a sequence of sets $\Sigma_n \uparrow \Sigma_f$ such that $\tilde{f}|_{\Sigma_n}$ is a Lipschitz function for any n. By Mc Shane theorem we can extend $\tilde{f}|_{\Sigma_n}$ to Lipschitz functions g_n defined on the whole of $\mathbb{R}^d$. In the case $m = d$, by applying the area formula to g_n on Σ_n and noticing that (by definition) $\nabla g_n = \tilde{\nabla} f$ $\mathscr{L}^d$-a.e. on Σ_n we obtain

$$\int_{\Sigma_f} h(x) |\det \tilde{\nabla} f|(x)\, dx = \int_{\mathbb{R}^d} \sum_{x \in \Sigma_f \cap \tilde{f}^{-1}(y)} h(x)\, dy \qquad (5.5.2)$$

for any Borel function $h : \mathbb{R}^d \to [0, +\infty]$.

This formula leads to a simple rule for computing the density of the push-forward of measures absolutely continuous w.r.t. $\mathscr{L}^d$.

Lemma 5.5.3 (Density of the push-forward). *Let $\rho \in L^1(\mathbb{R}^d)$ be a nonnegative function and assume that there exists a Borel set $\Sigma \subset \Sigma_f$ such that $\tilde{f}|_\Sigma$ is injective and the difference $\{\rho > 0\} \setminus \Sigma$ is $\mathscr{L}^d$-negligible. Then $f_\#\big(\rho\mathscr{L}^d\big) \ll \mathscr{L}^d$ if and only if $|\det\tilde{\nabla} f| > 0$ $\mathscr{L}^d$-a.e. on Σ and in this case*

$$f_\#\big(\rho\mathscr{L}^d\big) = \frac{\rho}{|\det\tilde{\nabla} f|} \circ \tilde{f}^{-1}|_{f(\Sigma)}\mathscr{L}^d.$$

Proof. If $|\det\tilde{\nabla} f| > 0$ $\mathscr{L}^d$-a.e. on Σ we can put $h = \rho\chi_{\tilde{f}^{-1}(B)\cap\Sigma}/|\det\tilde{\nabla} f|$ in (5.5.2), with $B \in \mathscr{B}(\mathbb{R}^d)$, to obtain

$$\int_{\tilde{f}^{-1}(B)} \rho\,dx = \int_{\tilde{f}^{-1}(B)\cap\Sigma} \rho\,dx = \int_{B\cap\tilde{f}(\Sigma)} \frac{\rho(\tilde{f}^{-1}(y))}{|\det\tilde{\nabla} f(\tilde{f}^{-1}(y))|}\,dy.$$

Conversely, if there is a Borel set $B \subset \Sigma$ with $\mathscr{L}^d(B) > 0$ and $|\det\tilde{\nabla} f| = 0$ on B the area formula gives $\mathscr{L}^d(\tilde{f}(B)) = 0$. On the other hand

$$f_\#(\rho\mathscr{L}^d)(\tilde{f}(B)) = \int_{f^{-1}(\tilde{f}(B))} \rho\,dx > 0$$

because at $\mathscr{L}^d$-a.e. $x \in B$ we have $f(x) = \tilde{f}(x)$ and $\rho(x) > 0$. Hence $f_\#(\rho\mathscr{L}^d)$ is not absolutely continuous with respect to $\mathscr{L}^d$. □

By applying the area formula again we obtain the rule for computing integrals of the densities:

$$\int_{\mathbb{R}^d} F\left(\frac{f_\#(\rho\mathscr{L}^d)}{\mathscr{L}^d}\right) dx = \int_{\mathbb{R}^d} F\left(\frac{\rho}{|\det\tilde{\nabla} f|}\right) |\det\tilde{\nabla} f|\,dx \tag{5.5.3}$$

for any Borel function $F : \mathbb{R} \to [0, +\infty]$ with $F(0) = 0$. Notice that in this formula the set Σ does not appear anymore (due to the fact that $F(0) = 0$ and $\rho = 0$ out of Σ), so it holds provided f is approximately differentiable $\rho\mathscr{L}^d$-a.e., it is $\rho\mathscr{L}^d$-essentially injective (i.e. there exists a Borel set Σ such that $\tilde{f}|_\Sigma$ is injective and $\rho = 0$ $\mathscr{L}^d$-a.e. out of Σ) and $|\det\tilde{\nabla} f| > 0$ $\rho\mathscr{L}^d$-a.e.

We will apply mostly these formulas when f is the gradient of a convex function g (corresponding to optimal transport map for the quadratic cost function), or is an optimal transport map. In the former case actually approximate differentiability is not needed thanks to the following result (see for instance [4, 64]).

Theorem 5.5.4 (Aleksandrov). *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then ∇g is differentiable $\mathscr{L}^d$-a.e. in its domain, its gradient $\nabla^2 g(x)$ is a symmetric matrix for $\mathscr{L}^d$-a.e. $x \in \mathbb{R}^d$, and g has second order Taylor expansion*

$$g(y) = g(x) + \langle\nabla g(x), y - x\rangle + \frac{1}{2}\langle\nabla^2 g(x), y - x\rangle + o(|y - x|^2) \quad \text{as } y \to x. \tag{5.5.4}$$

Notice that ∇g is also monotone

$$\langle \nabla g(x_1) - \nabla g(x_2), x_1 - x_2 \rangle \geq 0 \qquad x_1,\, x_2 \in D(\nabla g),$$

and that the above inequality is strict if g is *strictly* convex: in this case, it is immediate to check that ∇g is injective on $D(\nabla g)$, and that $|\det \nabla^2 g| > 0$ on the differentiability set of ∇g if g is *uniformly* convex.

Chapter 6

The Optimal Transportation Problem

Let X, Y be separable metric spaces such that any Borel probability measure in X, Y is tight (5.1.9), i.e. Radon spaces, according to Definition 5.1.4, and let $c : X \times Y \to [0, +\infty]$ be a Borel cost function. Given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ the optimal transport problem, in Monge's formulation, is given by

$$\inf \left\{ \int_X c(x, \boldsymbol{t}(x))\, d\mu(x) : \ \boldsymbol{t}_\# \mu = \nu \right\}. \tag{6.0.1}$$

This problem can be ill posed because sometimes there is no transport map $\boldsymbol{t}$ such that $\boldsymbol{t}_\#\mu = \nu$ (this happens for instance when μ is a Dirac mass and ν is not a Dirac mass). Kantorovich's formulation

$$\min \left\{ \int_{X \times Y} c(x, y)\, d\boldsymbol{\gamma}(x, y) : \ \boldsymbol{\gamma} \in \Gamma(\mu, \nu) \right\} \tag{6.0.2}$$

circumvents this problem (as $\mu \times \nu \in \Gamma(\mu, \nu)$). The existence of an optimal transport plan, when c is l.s.c., is provided by (5.1.15) and by the tightness of $\Gamma(\mu, \nu)$ (this property is equivalent to the tightness of μ, ν, a property always guaranteed in Radon spaces).

The problem (6.0.2) is truly a weak formulation of (6.0.1) in the following sense: if c is bounded and continuous, and if μ has no atom, then the "min" in (6.0.2) is equal to the "inf" in (6.0.1), see [70], [9]. This result can also be extended to unbounded cost functions, under the assumption (6.1.8), see [111].

In some special situations one can directly show the existence of optimal transport maps without any assumption on the cost function (besides positivity and lower semicontinuity).

Theorem 6.0.1 (Birkhoff theorem). *Let C be the convex set of all doubly stochastic $N \times N$ matrices, i.e. those matrices M whose entries M_{ij} satisfy*

$$\sum_{i=1}^{N} M_{ij} = \frac{1}{N} \quad \forall j = 1, \dots, N, \qquad \sum_{j=1}^{N} M_{ij} = \frac{1}{N} \quad \forall i = 1, \dots, N.$$

Then, the extreme points of C are permutation matrices, i.e. those matrices of the form

$$M_{ij} = \frac{1}{N}\delta_{i\sigma(j)} \quad \text{for some permutation } \sigma \text{ of } \{1, \dots, N\}.$$

In particular, if μ (resp. ν) can be represented as the sum of N Dirac masses in distinct points x_i (resp. distinct points y_j) with weight $1/N$, then the minimum in (6.0.2) *is always provided by a transport map.*

Proof. For a proof the first statement see, for instance, the simple argument at the end of the introduction of [126].
The convex set $\Gamma(\mu, \nu)$ can be canonically identified with C, writing $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}(\{x_i\} \times \{y_j\})$, and transport maps correspond to permutation matrices. Since the energy functional is linear on $\Gamma(\mu, \nu)$, the minimum is surely attained on a extreme point of $\Gamma(\mu, \nu)$ and therefore on a transport map. □

Another special occasion occurs when $X = Y = \mathbb{R}$. In this case we can use the distribution function

$$F_\mu(t) := \mu\left((-\infty, t)\right) \qquad t \in \mathbb{R}$$

to characterize optimal transport maps and to give a simple formula for the minimum value in (6.0.2). We need to define also an inverse of F_μ, by the formula (notice that a priori F_μ need not be continuous or strictly increasing)

$$F_\mu^{-1}(s) := \sup\{x \in \mathbb{R} : \ F_\mu(x) \le s\} \qquad s \in [0, 1].$$

Theorem 6.0.2 (Optimal transportation in $\mathbb{R}$). *Let μ, $\nu \in \mathscr{P}_p(\mathbb{R})$ and let $c(x, y) = h(x - y)$, with $h \ge 0$ convex and with p growth.*

(i) *If μ has no atom, i.e. F_μ is continuous, then $F_\nu^{-1} \circ F_\mu$ is an optimal transport map. It is the unique optimal transport map if h is strictly convex.*

(ii) *We have*

$$\min\left\{\int_{\mathbb{R}^2} c(x, y)\, d\boldsymbol{\gamma} : \ \boldsymbol{\gamma} \in \Gamma(\mu, \nu)\right\} = \int_0^1 c\left(F_\mu^{-1}(s), F_\nu^{-1}(s)\right) ds. \tag{6.0.3}$$

Proof. For the proof of the first statement see for instance [126], [71].
(ii) In this proof we use the following two elementary properties of the distribution function when μ has no atom: first, $F_{\mu\#}\mu = \chi_{(0,1)}\mathscr{L}^1$ (this fact can be checked in an elementary way on intervals and we omit the argument), second

$F_\mu^{-1} \circ F_\mu(x) = x$ for μ-a.e. x. The second property simply follows by the observation that the (maximal) open intervals in which F_μ is constant correspond, by the very definition of F_μ, to intervals where μ has no mass. Using statement (i) we have then

$$\begin{aligned}\int_{\mathbb{R}} c\left(x, F_\nu^{-1} \circ F_\mu(x)\right) d\mu(x) &= \int_{\mathbb{R}} c\left(F_\mu^{-1} \circ F_\mu(x), F_\nu^{-1} \circ F_\mu(x)\right) d\mu(x) \\ &= \int_0^1 c\left(F_\mu^{-1}(s), F_\nu^{-1}(s)\right) ds,\end{aligned}$$

in the case when μ has no atom. The general case can be achieved through a simple approximation. □

6.1 Optimality conditions

In this section we discuss the optimality conditions in the variational problem (6.0.2), assuming always that $c : X \times Y \to [0, +\infty]$ is a proper l.s.c. function.

Theorem 6.1.1 (Duality formula). *The minimum of the Kantorovich problem* (6.0.2) *is equal to*

$$\sup\left\{\int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y)\right\} \tag{6.1.1}$$

where the supremum runs among all pairs $(\varphi, \psi) \in C_b^0(X) \times C_b^0(Y)$ *such that* $\varphi(x) + \psi(y) \leq c(x, y)$.

Proof. This identity is well-known if c is bounded and continuous, see for instance [92, 112, 126]. A possible strategy is to show first that the support of any optimal plan is a c-monotone set, according to Definition 6.1.3 below, and than use this fact to build a maximizing pair (we will give this construction in Theorem 6.1.4 below, under more general assumptions on c).

In the general case it suffices to approximate c from below by an increasing sequence of bounded continuous functions c_h, defined for instance by (compare with (5.1.4))

$$c_h(x, y) := \inf_{(x', y') \in X \times Y} \left\{c(x', y') \wedge h + h d_X(x, x') + h d_Y(y, y')\right\},$$

noticing that a simple compactness argument gives

$$\min\left\{\int_{X \times Y} c_h\, d\boldsymbol{\gamma} : \ \boldsymbol{\gamma} \in \Gamma(\mu, \nu)\right\} \quad \uparrow \quad \min\left\{\int_{X \times Y} c\, d\boldsymbol{\gamma} : \ \boldsymbol{\gamma} \in \Gamma(\mu, \nu)\right\}$$

and that any pair (φ, ψ) such that $\varphi + \psi \leq c_h$ is admissible in (6.1.1). □

We recall briefly the definitions of c-transform, c-concavity and c-cyclical monotonicity, referring to the papers [56], [71] and to the book [112] for a more detailed analysis.

Definition 6.1.2 (c-transform, c-concavity). (1) *For $u : X \to \overline{\mathbb{R}}$, the* c-transform *$u^c : Y \to \overline{\mathbb{R}}$ is defined by*

$$u^c(y) := \inf_{x \in X} c(x, y) - u(x)$$

with the convention that the sum is $+\infty$ whenever $c(x, y) = +\infty$ and $u(x) = +\infty$. Analogously, for $v : Y \to \overline{\mathbb{R}}$, the c-transform *$v^c : X \to \overline{\mathbb{R}}$ is defined by*

$$v^c(x) := \inf_{y \in Y} c(x, y) - v(y)$$

with the same convention when an indetermination of the sum is present.
(2) *We say that $u : X \to \overline{\mathbb{R}}$ is* c-concave *if $u = v^c$ for some v; equivalently, u is c-concave if there is some family $\{(y_i, t_i)\}_{i \in I} \subset Y \times \overline{\mathbb{R}}$ such that*

$$u(x) = \inf_{i \in I} c(x, y_i) + t_i \qquad \forall x \in X. \tag{6.1.2}$$

An analogous definition can be given for functions $v : Y \to \overline{\mathbb{R}}$.

It is not hard to show that $u^{cc} \geq u$ and that equality holds if and only if u is c-concave. Analogously, $v^{cc} \geq v$ and equality holds if and only if v is c-concave.

Let us also introduce the concept of c-monotonicity.

Definition 6.1.3 (c-monotonicity). *We say that $\Gamma \subset X \times Y$ is* c-monotone *if*

$$\sum_{i=1}^{n} c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^{n} c(x_i, y_i)$$

whenever $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ and σ is a permutation of $\{1, \ldots, n\}$.

With these definitions we can prove the following result concerning necessary and sufficient optimality conditions and the existence of maximizing pairs (φ, ψ) in (6.1.1). The proof is taken from [14], see also [126], [71], [112] for similar earlier results (note however that conditions (6.1.3) and (6.1.4) do not apply to the cost functions considered in [68, 69, 89], in a infinite-dimensional framework).

Theorem 6.1.4 (Necessary and sufficient optimality conditions).
(Necessity) If $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$ is optimal and $\int_{X \times Y} c \, d\boldsymbol{\gamma} < +\infty$, then $\boldsymbol{\gamma}$ is concentrated on a c-monotone Borel subset of $X \times Y$. Moreover, if c is continuous, then $\operatorname{supp} \boldsymbol{\gamma}$ *is c-monotone.*
(Sufficiency) Assume that c is real-valued, $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$ is concentrated on a c-monotone Borel subset of $X \times Y$, and

$$\mu\left(\left\{x \in X : \int_Y c(x, y) \, d\nu(y) < +\infty\right\}\right) > 0, \tag{6.1.3}$$

$$\nu\left(\left\{y \in Y: \int_X c(x,y)\,d\mu(x) < +\infty\right\}\right) > 0. \tag{6.1.4}$$

Then $\boldsymbol{\gamma}$ *is optimal,* $\int_{X\times Y} c\,d\boldsymbol{\gamma} < +\infty$ *and there exists a maximizing pair* (φ,ψ) *in* (6.1.1) *with* φ *c-concave and* $\psi = \varphi^c$.

Proof. Let (φ_n,ψ_n) be a maximizing sequence in (6.1.1) and let $c_n = c - \varphi_n - \psi_n$. Since

$$\int_{X\times Y} c_n\,d\boldsymbol{\gamma} = \int_{X\times Y} c\,d\boldsymbol{\gamma} - \int_X \varphi_n\,d\mu - \int_Y \psi_n\,d\nu \to 0$$

and $c_n \geq 0$ we can find a subsequence $c_{n(k)}$ and a Borel set Γ on which $\boldsymbol{\gamma}$ is concentrated and c is finite, such that $c_{n(k)} \to 0$ on Γ. If $\{(x_i,y_i)\}_{1\le i\le p} \subset \Gamma$ and σ is a permutation of $\{1,\ldots,p\}$ we get

$$\begin{aligned}\sum_{i=1}^p c(x_i,y_{\sigma(i)}) &\geq \sum_{i=1}^p \varphi_{n(k)}(x_i) + \psi_{n(k)}(y_{\sigma(i)})\\ &= \sum_{i=1}^p \varphi_{n(k)}(x_i) + \psi_{n(k)}(y_i) = \sum_{i=1}^p c(x_i,y_i) - c_{n(k)}(x_i,y_i)\end{aligned}$$

for any k. Letting $k \to \infty$ the c-monotonicity of Γ follows.

Now we show the converse implication, assuming that (6.1.3) and (6.1.4) hold. We denote by Γ a Borel and c-monotone set on which $\boldsymbol{\gamma}$ is concentrated; without loss of generality we can assume that $\Gamma = \cup_k \Gamma_k$ with Γ_k compact and $c|_{\Gamma_k}$ continuous. We choose continuous functions c_l such that $c_l \uparrow c$ and split the proof in several steps.

Step 1. There exists a c-concave Borel function $\varphi : X \to [-\infty,+\infty)$ such that $\varphi(x) > -\infty$ for μ-a.e. $x \in X$ and

$$\varphi(x') \leq \varphi(x) + c(x',y) - c(x,y) \qquad \forall x' \in X,\ \ (x,y) \in \Gamma. \tag{6.1.5}$$

To this aim, we use the explicit construction given in the generalized Rockafellar theorem in [116], setting

$$\begin{aligned}\varphi(x) := \inf\{&c(x,y_p) - c(x_p,y_p) + c(x_p,y_{p-1}) - c(x_{p-1},y_{p-1})\\ &+ \cdots + c(x_1,y_0) - c(x_0,y_0)\}\end{aligned}$$

where $(x_0,y_0) \in \Gamma_1$ is fixed and the infimum runs among all integers p and collections $\{(x_i,y_i)\}_{1\le i\le p} \subset \Gamma$.

It can be easily checked that

$$\varphi = \lim_{p\to\infty}\lim_{m\to\infty}\lim_{l\to\infty} \varphi_{p,m,l},$$

where

$$\begin{aligned}\varphi_{p,m,l}(x) := \inf\{&c_l(x,y_p) - c(x_p,y_p) + c_l(x_p,y_{p-1}) - c(x_{p-1},y_{p-1})\\ &+ \cdots + c_l(x_1,y_0) - c(x_0,y_0)\}\end{aligned}$$

and the infimum is made among all collections $\{(x_i, y_i)\}_{1\le i\le p} \subset \Gamma_m$. As all functions $\varphi_{p,m,l}$ are upper semicontinuous we obtain that φ is a Borel function.

Arguing as in [116] it is straightforward to check that $\varphi(x_0) = 0$ and that (6.1.5) holds. Choosing $x' = x_0$ we obtain that $\varphi > -\infty$ on $\pi_X(\Gamma)$ (here we use the assumption that c is real-valued). But since $\boldsymbol{\gamma}$ is concentrated on Γ the Borel set $\pi_X(\Gamma)$ has full measure with respect to $\mu = \pi_{X\#}\boldsymbol{\gamma}$, hence $\varphi \in \mathbb{R}$ μ-a.e.

Step 2. Now we show that $\psi := \varphi^c$ is ν-measurable, real-valued ν-a.e. and that

$$\varphi + \psi = c \quad \text{on } \Gamma. \tag{6.1.6}$$

It suffices to study ψ on $\pi_Y(\Gamma)$: indeed, as γ is concentrated on Γ, the Borel set $\pi_Y(\Gamma)$ has full measure with respect to $\nu = \pi_{Y\#}\gamma$. For $y \in \pi_Y(\Gamma)$ we notice that (6.1.5) gives

$$\psi(y) = c(x,y) - \varphi(x) \in \mathbb{R} \qquad \forall x \in \Gamma_y := \{x :\ (x,y) \in \Gamma\}.$$

In order to show that ψ is ν-measurable we use the disintegration $\boldsymbol{\gamma} = \gamma_y \times \nu$ of $\boldsymbol{\gamma}$ with respect to y and notice that the probability measure γ_y is concentrated on Γ_y for ν-a.e. y, therefore

$$\psi(y) = \int_X c(x,y) - \varphi(x)\, d\gamma_y(x) \qquad \text{for } \nu\text{-a.e. } y.$$

Since $y \mapsto \gamma_y$ is a Borel measure-valued map we obtain that ψ is ν-measurable.

Step 3. We show that φ^+ and ψ^+ are integrable with respect to μ and ν respectively (here we use (6.1.3) and (6.1.4)). By (6.1.3) we can choose x in such a way that $\int_Y c(x,y)\, d\nu(y)$ is finite and $\varphi(x) \in \mathbb{R}$, so that by integrating on Y the inequality $\psi^+ \le c(x,\cdot) + \varphi^-(x)$ we obtain that $\psi^+ \in L^1(Y,\nu)$. The argument for φ^+ uses (6.1.4) and is similar.

Step 4. Conclusion. The semi-integrability of φ and ψ gives the null-Lagrangian identity

$$\int_{X\times Y} (\varphi + \psi)\, d\tilde{\boldsymbol{\gamma}} = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \in \mathbb{R} \cup \{-\infty\} \qquad \forall \tilde{\boldsymbol{\gamma}} \in \Gamma(\mu,\nu),$$

so that choosing $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$ we obtain from (6.1.6) that $\int_{X\times Y} c\, d\boldsymbol{\gamma} < +\infty$ and $\varphi \in L^1(X,\mu)$, $\psi \in L^1(Y,\nu)$. Moreover, for any $\tilde{\gamma} \in \Gamma(\mu,\nu)$ we get

$$\begin{aligned}
\int_{X\times Y} c\, d\tilde{\boldsymbol{\gamma}} \ &\ge\ \int_{X\times Y} (\varphi + \psi)\, d\tilde{\boldsymbol{\gamma}} = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \\
&=\ \int_{X\times Y} (\varphi + \psi)\, d\boldsymbol{\gamma} = \int_\Gamma (\varphi + \psi)\, d\boldsymbol{\gamma} = \int_{X\times Y} c\, d\boldsymbol{\gamma}.
\end{aligned}$$

This chain of inequalities gives that γ is optimal and, at the same time, that (φ,ψ) is optimal in (6.1.1). □

We say that a Borel function $\varphi \in L^1(X,\mu)$ is a *maximal Kantorovich potential* if (φ,φ^c) is a maximizing pair in (6.1.1). In many applications it is useful to write the optimality conditions using a maximal Kantorovich potential, instead of the cyclical monotonicity.

Theorem 6.1.5. *Let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, assume that* (6.1.3) *and* (6.1.4) *hold, that c is real-valued and that the* sup *in* (6.1.1) *is finite. Then there exists a maximizing pair (φ,φ^c) in* (6.1.1) *and if $\boldsymbol{\gamma} \in \Gamma(\mu,\nu)$ is optimal then*

$$\varphi(x) + \varphi^c(y) = c(x,y) \qquad \boldsymbol{\gamma}\text{-a.e. in } X \times Y. \tag{6.1.7}$$

Moreover, if there exists a Borel potential $\varphi \in L^1(X,\mu)$ such that (6.1.7) *holds, then $\boldsymbol{\gamma}$ is optimal.*

Proof. The existence of a maximizing pair is a direct consequence of the sufficiency part of the previous theorem, choosing an optimal $\boldsymbol{\gamma}$ and (by the necessity part of the statement) a c-monotone set on which $\boldsymbol{\gamma}$ is concentrated.

If $\boldsymbol{\gamma}$ is optimal then

$$\int_{X\times Y} (c - \varphi - \varphi^c)\, d\boldsymbol{\gamma} = \int_{X\times Y} c\, d\boldsymbol{\gamma} - \int_X \varphi\, d\mu - \int_Y \varphi^c\, d\nu = 0.$$

As the integrand is nonnegative, it must vanish $\boldsymbol{\gamma}$-a.e. The converse implication is analogous. ☐

Remark 6.1.6. The assumptions (6.1.3), (6.1.4) are implied by

$$\int_{X\times Y} c(x,y)\, d\mu \times \nu(x,y) < +\infty. \tag{6.1.8}$$

In turn, (6.1.8) is implied by the condition

$$c(x,y) \le a(x) + b(y) \quad \text{with} \quad a \in L^1(\mu),\ b \in L^1(\nu).$$

6.2 Optimal transport maps and their regularity

In this section we go back to the original Monge problem (6.0.1), finding natural conditions on c and μ ensuring the existence of optimal transport maps.

Definition 6.2.1 (Gaussian measures and Gaussian null sets). *Let X be a separable Banach space with dual X', and let $\mu \in \mathscr{P}(X)$. We say that μ is a nondegenerate Gaussian (probability) measure in X if for any $L \in X'$ the image measure $L_\#\mu \in \mathscr{P}(\mathbb{R})$ has a Gaussian distribution, i.e. there exist $m = m(L) \in \mathbb{R}$ and $\sigma = \sigma(L) > 0$ such that*

$$\mu\left(\{x \in X:\ a < L(x) < b\}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{-|t-m|^2/2\sigma^2}\, dt \quad \forall\, (a,b) \subset \mathbb{R}.$$

We say that $B \in \mathscr{B}(X)$ is a Gaussian null set if $\mu(B) = 0$ for any nondegenerate Gaussian measure μ in X.

We refer to [25] for the general theory of Gaussian measures. Here we use Gaussian measures only to define the σ-ideal of Gaussian null sets. Starting from Definition 6.2.1 and recalling (5.2.4), it is easy to check that if μ is a (nondegenerate) Gaussian measure in X and Y is another (separable) Banach space, then

$$\begin{aligned} &\pi_{\#}\mu \quad \text{is a (nondegenerate) Gaussian measure in } Y \\ &\text{for every continuous (surjective) linear map} \quad \pi : X \to Y. \end{aligned} \tag{6.2.1}$$

One can also check that in the case $X = \mathbb{R}^d$ nondegenerate Gaussian measures are absolutely continuous with respect to $\mathscr{L}^d$, with density given by

$$\frac{1}{\sqrt{(2\pi)^d \det A}} e^{-\frac{1}{2}\langle A^{-1}(x-m),(x-m)\rangle}$$

for some $m \in \mathbb{R}^d$ and some positive definite symmetric matrix A. Therefore Gaussian null sets coincide with $\mathscr{L}^d$-negligible sets. See also [47] for the equivalence between Gaussian null sets and null sets in the sense of Aronszajn, a concept that involves only the Lebesgue measure on the real line.

Definition 6.2.2 (Regular measures). *We say that $\mu \in \mathscr{P}(X)$ is regular if $\mu(B) = 0$ for any Gaussian null set B. We denote by $\mathscr{P}^r(X)$ the class of regular measures.*

By definition of Gaussian null sets, all Gaussian measures are regular. By the above remarks on Gaussian null sets, in the finite dimensional case $X = \mathbb{R}^d$ the class $\mathscr{P}^r(X)$ reduces to the standard family of measures absolutely continuous with respect to $\mathscr{L}^d$.

We recall the following classical infinite-dimensional version of Rademacher's theorem (see for instance Theorem 5.11.1 in [25]).

Theorem 6.2.3 (Differentiability of Lipschitz functions). *Let X be a separable Hilbert space and let $\phi : X \to \mathbb{R}$ be a locally Lipschitz function. Then the set of points where ϕ is not Gateaux differentiable is a Gaussian null set.*

Theorem 6.2.4 (Optimal transport maps in $\mathbb{R}^d$). *Assume that $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, $c(x, y) = h(x - y)$ with $h : \mathbb{R}^d \to [0, +\infty)$ strictly convex, and the minimum in* (6.0.2) *finite.*
If μ, ν satisfy (6.1.3), (6.1.4), *and $\mu \in \mathscr{P}^r(\mathbb{R}^d)$, then the Kantorovich problem* (6.0.2) *has a unique solution $\boldsymbol{\mu}$ and this solution is induced by an* optimal transport, *i.e. there exists a Borel map $\boldsymbol{r} : \mathbb{R}^d \to \mathbb{R}^d$ such that the representation* (5.2.13) *holds. We have also*

$$\boldsymbol{r}(x) = x - (\partial h)^{-1}\left(\tilde\nabla\varphi(x)\right) \qquad \textit{for } \mu\textit{-a.e. } x, \tag{6.2.2}$$

for any c-concave and maximal Kantorovich potential φ (recall that $\tilde\nabla$ stands for the approximate differential).

Proof. By the necessity part in Theorem 6.1.4 we have the existence of an optimal plan, concentrated on a c-monotone subset of $\mathbb{R}^d \times \mathbb{R}^d$. By the sufficiency part we obtain the existence of a c-concave maximal Kantorovich potential φ. Theorem 6.1.5 gives that for μ-a.e. x there exists y such that $\varphi(x) + \varphi^c(y) = c(x, y)$. We have to show that y is unique and given by (6.2.2). To this aim, for any $R > 0$ we define

$$\varphi_R(x) := \inf_{z \in B_R(0)} c(x, z) - \varphi^c(z) \qquad x \in \mathbb{R}^d.$$

Notice that all functions φ_R are locally Lipschitz in $\mathbb{R}^d$ for R large enough (as soon as there is some z with $|z| < R$ and $\varphi^c(z) > -\infty$) and therefore differentiable $\mathscr{L}^d$-a.e. Moreover, the above mentioned existence of y for μ-a.e. x implies that the decreasing family of sets $\{\varphi < \varphi_R\}$ has a μ-negligible intersection, i.e. μ-a.e. x belongs to $\{\varphi = \varphi_R\}$ for R large enough.

It follows that for μ-a.e. x the following two conditions are satisfied: x is a point of density 1 of $\{\varphi = \varphi_R\}$ for some R (recall Remark 5.5.2 and φ_R is differentiable at x. By the very definition of approximate differential, φ is approximately differentiable at x and $\tilde{\nabla}\varphi(x) = \nabla\varphi_R(x)$. If $\varphi(x) + \varphi^c(y) = h(x-y)$, since $x' \mapsto h(x'-y) - \varphi(x')$ attains its minimum (equal to $\varphi^c(y)$) at x, by differentiation of both sides we get

$$\tilde{\nabla}\varphi(x) \in \partial h(x - y).$$

This immediately gives that y is unique and given by (6.2.2). □

In the following remark we point out some extensions of the previous existence result and we recall some cases when the approximate differential in (6.2.2) is indeed a classical differential.

Remark 6.2.5. a) Classical differential. As the proof shows, the approximate differential is actually a classical differential if ν has a bounded support. Under a technical condition on the level sets of h at infinity (this condition includes the model case $h(z) = |z|^p$, $p > 1$) the differential is still classical even when ν has an unbounded support, see [71].

b) More general initial measures. It has been shown in [71] that for $h \in C^{1,1}_{\rm loc}(\mathbb{R}^d)$ and ν with bounded support the same properties hold if μ satisfies the more general condition

$$\mu(B) = 0 \quad \text{whenever } B \in \mathscr{B}(\mathbb{R}^d) \text{ and } \mathscr{H}^{d-1}(B) < +\infty. \tag{6.2.3}$$

The proof is based on a refinement of Rademacher theorem, valid for convex or semi-convex functions, see for instance [4].

c) The case when h is not strictly convex. Here the difficulty arises from the fact that $(\partial h)^{-1}$ is not single-valued in general, so the first variation argument of the previous proofs does not produce anymore a unique y, for given x. This problem, even when $h(z) = \|z\|$ for some norm $\|\cdot\|$ in $\mathbb{R}^d$, is not yet completely understood, see the discussions in [13]. Only the case when $\|\cdot\|$ is the Euclidean norm (or, more generally, a C^2 and uniformly convex norm) has been settled (see [121], [62], [33], [122], [9], [14]). See also [13] for an existence result in the case when the norm $\|\cdot\|$ is crystalline (i.e. its unit sphere is contained in finitely many hyperplanes).

6.2.1 Approximate differentiability of the optimal transport map

In many applications it is useful to know that the optimal transport map is differentiable, at least in the approximate sense. The following theorem answers to this question and shows, adapting to a non-smooth setting an argument in [103], that the differential of the optimal transport map is diagonalizable and has nonnegative eigenvalues. Notice that our assumption on the cost includes the model case $c(x,y) = |x-y|^p$, $p > 1$. In the proof of the theorem we will use a weak version of the second order Taylor expansion, but still sufficient to have a maximum principle.

Definition 6.2.6 (Approximate second order expansion). *Let $\Omega \subset \mathbb{R}^d$ be an open set and $\varphi : \Omega \to \mathbb{R}$. We say that φ has an approximate second order expansion at $x \in \Omega$ if*

$$\lim_{y \to x,\, y \in E} \frac{\varphi(y) - a - \langle b, y-x \rangle - \langle A(y-x), (y-x) \rangle}{|y-x|^2} = 0 \tag{6.2.4}$$

for some $a \in \mathbb{R}$, $b \in \mathbb{R}^d$ and some symmetric matrix A, with E having density 1 *at x.*

It is immediate to check that $a = \tilde{\varphi}(x)$, $b = \tilde{\nabla}\varphi(x)$ and that A is uniquely determined: we will denote it by $\tilde{\nabla}^2\varphi(x)$. Moreover, if φ has a minimum at x then $b = 0$ and $A \geq 0$.

Theorem 6.2.7 (Approximate differentiability of the transport map). *Assume that $\mu \in \mathscr{P}^r(\mathbb{R}^d)$, $\nu \in \mathscr{P}(\mathbb{R}^d)$ and let $c(x,y) = h(x-y)$ with $h : \mathbb{R}^d \to [0,+\infty)$ strictly convex with superlinear growth, $h \in C^1(\mathbb{R}^d) \cap C^2(\mathbb{R}^d \setminus \{0\})$, and $\nabla^2 h$ is positive definite in $\mathbb{R}^d \setminus \{0\}$. If the minimum in* (6.0.2) *is finite, then for μ-a.e. $x \in \mathbb{R}^d$ the optimal transport map $\boldsymbol{r}$ is approximately differentiable at x and $\tilde{\nabla}\boldsymbol{r}(x)$ is diagonalizable with nonnegative eigenvalues.*

Proof. Let φ be a maximal Kantorovich potential and let $N = \{\boldsymbol{r}(x) \neq x\}$. Clearly it suffices to show that the claimed properties are true μ-a.e. on N (as outside of N the approximate differential of $\boldsymbol{r}$ is the identity). We consider the countable family of triplets of balls (B, B', B'') centered at a rational point of $\mathbb{R}^d$, with $\overline{B} \subset B'$, $\overline{B}' \subset B''$ and with rational radii, the family of sets

$$N_{B,B',B''} := \{x \in B : \ \boldsymbol{r}(x) \in B'' \setminus B'\},$$

and the family of functions

$$\varphi_{B,B',B''}(x) := \min_{y \in B'' \setminus B'} h(x-y) - \varphi^c(y) \qquad x \in B.$$

Notice that $\varphi_{B,B',B''} = \varphi$ μ-a.e. on $N_{B,B',B''}$, as the minimum of $y \mapsto h(x-y) - \varphi^c(y)$ is achieved at $y = \boldsymbol{r}(x) \in B'' \setminus B'$ for μ-a.e. x.

Let $C = C(B, B', B'')$ be the Lipschitz constant $\mathrm{Lip}(\nabla h, B - (B'' \setminus B'))$ of ∇h in the set $B - (B'' \setminus B')$; it follows that all maps

$$x \mapsto h(x-y) - \varphi^c(y) - \frac{C}{2}|x|^2, \qquad y \in B'' \setminus B',$$

are concave in B, and therefore $\varphi_{B,B',B''} - C|x|^2/2$ is concave in B as well. By Alexandrov's differentiability theorem (see 5.5.4) we obtain that $\varphi_{B,B',B''}$ are twice differentiable and have a classical second order Taylor expansion for $\mathscr{L}^d$-a.e. $x \in B$.

Clearly the set N is contained in the union of all sets $N_{B,B',B''}$, therefore, by Remark 5.5.2, $\mathscr{L}^d$-a.e. $x \in N$ is a point of density 1 for one of the sets $N_{B,B',B''}$ and $\varphi_{B,B',B''}$ is twice differentiable at x. By Definition 6.2.6 we obtain that φ is twice differentiable in the approximate sense at x and (6.2.4) holds with $a = \varphi(x)$, $b = \tilde\nabla\varphi(x) = \nabla\varphi_{B,B',B''}$ and $A = \tilde\nabla^2\varphi(x) = \nabla^2\varphi_{B,B',B''}/2$. Since

$$\boldsymbol{r}(x) = x - (\partial h)^{-1}(\tilde\nabla\varphi(x)) = x - \nabla h^*(\tilde\nabla\varphi(x)),$$

we obtain that $\boldsymbol{r}$ is approximately differentiable μ-a.e. on N.

Since h has a superlinear growth at infinity, the gradient map $\nabla h : \mathbb{R}^d \to \mathbb{R}^d$ is a bijection and its inverse is ∇h^*, where h^* is the conjugate of h. Therefore ∇h^* is differentiable on $\mathbb{R}^d \setminus \{\nabla h(0)\}$.

Fix now a point x where the above properties hold and set $y = \boldsymbol{r}(x)$. Since $x' \mapsto h(x'-y) - \varphi(x')$ achieves its minimum, equal to $-\varphi^c(y)$, at x, we get

$$\nabla^2 h(x-y) \geq \tilde\nabla^2\varphi(x).$$

On the other hand, the identity $\nabla h(\nabla h^*(p)) = p$ gives

$$\nabla^2 h\left(\nabla h^*(p)\right) = \left[\nabla^2 h^*(p)\right]^{-1}.$$

Using the identity above with $p = \tilde\nabla\varphi(x) \neq \nabla h(0)$ we obtain

$$\left[\nabla^2 h^*(\tilde\nabla\varphi(x))\right]^{-1} \geq \tilde\nabla^2\varphi(x).$$

By Lemma 6.2.8 below with $A := \nabla^2 h^*(\tilde\nabla\varphi(x))$ and $B := -\tilde\nabla^2\varphi(x)$ we obtain that $\tilde\nabla\boldsymbol{r}(x) = \boldsymbol{i} + AB$ is diagonalizable and it has nonnegative eigenvalues. □

Again, under more restrictive assumptions (e.g. the supports of the two measures are compact and $\mathrm{dist}\,(\mathrm{supp}\,\mu, \mathrm{supp}\,\nu) > 0$) one can show that the optimal transport map $\boldsymbol{r}$ is μ-a.e. differentiable in a classical sense. As discussed in Section 5.5, approximate differentiability is however sufficient to establish an area formula and the rule for the computation of the density of $\boldsymbol{r}_\#(\rho\mathscr{L}^d)$.

The following elementary lemma is also taken from [103].

Lemma 6.2.8. *Let A, B be symmetric matrices with A positive definite. If $-B \le A^{-1}$ then $\boldsymbol{i} + AB$ is diagonalizable and has nonnegative eigenvalues.*

Proof. Let C be a positive definite symmetric matrix such that $C^2 = A$. Since

$$\boldsymbol{i} + AB = C(\boldsymbol{i} + CBC)C^{-1}$$

and since $\boldsymbol{i} + CBC$ is symmetric we obtain that $\boldsymbol{i} + AB$ is diagonalizable. In order to show that the eigenvalues are nonnegative we estimate:

$$\begin{aligned}\langle(\boldsymbol{i} + CBC)\xi, \xi\rangle &= |\xi|^2 + \langle C\xi, BC\xi\rangle \ge |\xi|^2 - \langle C\xi, A^{-1}C\xi\rangle \\ &= |\xi|^2 - \langle \xi, CA^{-1}C\xi\rangle = 0 \quad \square\end{aligned}$$

□

In the following theorem we establish, under more restrictive assumptions on $\boldsymbol{r}$ or h, some properties of the distributional derivative of $\boldsymbol{r}$ and the nonnegativity of the distributional divergence of $\boldsymbol{r}$ (or, better, of a canonical extension of $\boldsymbol{r}$ to the whole of $\mathbb{R}^d$: recall that $\boldsymbol{r}$ is a priori defined only μ-a.e.).

Theorem 6.2.9 (Distributional derivative of $\boldsymbol{r}$). *Let μ, $\nu \in \mathscr{P}^r(\mathbb{R}^d)$, with $\operatorname{supp}\nu$ bounded, let $c(x, y) = h(x - y)$ with $h : \mathbb{R}^d \to [0, +\infty)$ strictly convex and with superlinear growth and assume that the minimum in (6.0.2) is finite. Let $\boldsymbol{r}$ be the optimal transport map between μ and ν. Then*

(i) *If $h \in C^2(\mathbb{R}^d)$ is locally uniformly convex then $\boldsymbol{r}$ has a canonical $BV_{\rm loc}$ extension to $\mathbb{R}^d$ satisfying $D \cdot \boldsymbol{r} \ge 0$.*

(ii) *If $h \in C^2(\mathbb{R}^d \setminus \{0\})$ and $\nabla h(0) = 0$ we can find equi-bounded maps $\boldsymbol{r}_k \in BV_{\rm loc}(\mathbb{R}^d)$ satisfying $D \cdot \boldsymbol{r}_k \ge 0$ such that $\mu(\{\boldsymbol{r}_k \ne \boldsymbol{r}\}) \to 0$ as $k \to \infty$.*

Proof. (i) By the argument used in the proof of Theorem 6.2.4 we know that there exists a c-concave potential φ of the form

$$\varphi(x) = \inf_{y \in \operatorname{supp}\nu} h(x - y) - \psi(y) \tag{6.2.5a}$$

with $\psi = -\infty$ on $\mathbb{R}^d \setminus \operatorname{supp}\nu$, such that

$$\boldsymbol{r}(x) = x - (\nabla h)^{-1}(\tilde{\nabla}\varphi) \quad \mu\text{-a.e. in } \mathbb{R}^d. \tag{6.2.5b}$$

We take as an extension of $\boldsymbol{r}$ the right hand side in the previous identity (6.2.5b), for φ given by (6.2.5a). Notice that, on any ball B, all functions

$$x \mapsto h(x - y) - \psi(y) - C|x|^2 \quad \text{for } y \in \operatorname{supp}\nu, \quad \psi(y) > -\infty,$$

are concave for C large enough (depending on B and $\operatorname{supp}\nu$), so that $\varphi - C|x|^2$ is concave in B as well. This proves that φ is locally Lipschitz and locally BV in $\mathbb{R}^d$ and therefore, since the inverse of ∇h is locally Lipschitz in $\mathbb{R}^d$ as well (by the local uniform convexity assumption on h and the superlinear growth condition), also $\boldsymbol{r}$ is locally BV.

Let us show that $\boldsymbol{r}(x) \in \operatorname{supp}\nu$ and that $x' \mapsto \varphi(x') - h(x'-y)$ attains its maximum at x when $y = \boldsymbol{r}(x)$ for $\mathscr{L}^d$-a.e. $x \in \mathbb{R}^d$. Indeed, fix x where φ is differentiable and let $\bar{y} \in \operatorname{supp}\nu$ be a minimizer of $y \mapsto h(x-y)-\psi(y)$ (without loss of generality we can assume that ψ is upper semicontinuous: being $\operatorname{supp}\nu$ compact and $\psi(y) < +\infty$ for every $y \in X$, a minimizer exists). Then $\varphi(x')-h(x'-\bar{y})$ attains its maximum at x since (6.2.5a) yields

$$\varphi(x') - h(x'-\bar{y}) \leq h(x'-\bar{y}) - \psi(\bar{y}) - h(x'-\bar{y}) = -\psi(\bar{y}) = \varphi(x) - h(x-\bar{y}),$$

and a differentiation yields $\bar{y} = \boldsymbol{r}(x)$.

It remains to show that $D\cdot\boldsymbol{r} \geq 0$. Since $\max_{\operatorname{supp}\nu} h(x-\cdot)$ is locally bounded we can find a strictly positive function $\rho \in L^1(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} \max_{y\in\operatorname{supp}\nu} h(x-y)\rho(x)\,dx < +\infty. \tag{6.2.6}$$

Let $\bar{\mu} = \rho\mathscr{L}^d$, and notice that the minimality property above shows that the graph of $\boldsymbol{r}$ is (essentially, excluding points x where φ is not differentiable) c-monotone: indeed for any choice of differentiability points $x_1,\ldots,x_n$ of and for any permutation σ of $\{1,\ldots,n\}$ we have

$$\sum_{i=1}^n \varphi(x_{\sigma(i)}) - h(x_{\sigma(i)} - \boldsymbol{r}(x_i)) \leq \sum_{i=1}^n \varphi(x_i) - h(x_i - \boldsymbol{r}(x_i)).$$

Removing from both sides $\sum_i \varphi(x_i)$ we obtain the c-monotonicity inequality.

Therefore, since by (6.2.6) the cost associated to $\boldsymbol{r}$ is finite, Theorem 6.1.4 gives that $\boldsymbol{r}$ is an optimal map between $\bar{\mu}$ and $\boldsymbol{r}_{\#}\bar{\mu}$.

This optimality property of the extended map $\boldsymbol{r}$ shows that it suffices to prove that $D\cdot\boldsymbol{r} \geq 0$ only when $\operatorname{supp}\nu$ is made by finitely many points: the general case can be achieved by approximation, using the fact that optimality relative to $\bar{\mu}$ is stable in the limit and yields $L^p(\bar{\mu})$ convergence of the maps (see Lemma 5.4.1) and then, up to subsequences, $\mathscr{L}^d$-a.e. convergence, due to the fact that $\rho > 0$ $\mathscr{L}^d$-a.e. Under the assumption that $\operatorname{supp}\nu$ is finite the function $\boldsymbol{r}$ takes only finitely many values $\{y_1,\ldots,y_m\}$ and the distributional divergence is given by

$$D\cdot\boldsymbol{r} = \langle \boldsymbol{r}^+ - \boldsymbol{r}^-, \boldsymbol{n}\rangle \chi_S \mathscr{H}^{d-1},$$

where $\boldsymbol{r}^\pm$ are the approximate one sided limits on the approximate jump set S of $\boldsymbol{r}$ and $\boldsymbol{n}$ is the approximate normal to the jump set. For a given Borel choice of $\boldsymbol{n}$, let us consider the sets

$$S_{ij} := \Big\{x \in S : \boldsymbol{r}^-(x) = y_i,\ \boldsymbol{r}^+(x) = y_j\Big\} \quad 1 \leq i,j \leq m,\ i \neq j, \quad S = \bigcup_{i\neq j} S_{ij}.$$

Since each neighborhood of $x \in S_{ij}$ contains points $x^\pm$ such that $\boldsymbol{r}(x^\pm) = \boldsymbol{r}^\pm(x)$ is the unique minimizer of $y \mapsto h(x^\pm - y) - \psi(y)$ in $\{y_1,\cdots,y_m\}$, S_{ij} is contained in ∂E_{ij}, with

$$E_{ij} := \{x \in \mathbb{R}^d : \ h(x - y_i) - \psi(y_i) < h(x - y_j) - \psi(y_j)\} \qquad 1 \le i \ne j \le m$$

and the classical inner normal to E_{ij} is parallel (with the same direction) to the nonvanishing vector $\nabla h(x - y_j) - \nabla h(x - y_i)$. Therefore it suffices to check the inequality

$$\langle y_i - y_j, \nabla h(x - y_j) - \nabla h(x - y_i)\rangle \ge 0.$$

This is a direct consequence of the monotonicity of ∇h:

$$\langle (x - y_j) - (x - y_i), \nabla h(x - y_j) - \nabla h(x - y_i)\rangle \ge 0.$$

(ii) Let $h_k \ge h$ be in $C^2(\mathbb{R}^d)$ and locally uniformly convex, with the property that for any $z \in \mathbb{R}^d$ we have $h_k(z) = h(z)$ and $(\nabla h_k)^{-1}(z) = (\nabla h)^{-1}(z)$ for k large enough (the proof of the existence of this approximation, a regularization of h near the origin, is left to the reader) and let φ, ψ as in the proof of (i). We define

$$\varphi_k(x) := \inf_{y \in \operatorname{supp} \nu} h_k(x - y) - \psi(y)$$

so that $\varphi_k \ge \varphi$. Since the infimum in the problem defining φ is attained (by $y = \boldsymbol{r}(x)$) for μ-a.e. x, it follows that $\varphi_k(x) = \varphi(x)$ for μ-a.e. x for k large enough (precisely, such that $h_k(x - \boldsymbol{r}(x)) = h(x - \boldsymbol{r}(x))$, so that $\mu(\{\varphi_k \ne \varphi\}) \to 0$ as $k \to \infty$. Setting

$$\boldsymbol{r}_k := \boldsymbol{i} - (\nabla h_k)^{-1}(\tilde{\nabla}\varphi_k)$$

we know, by the c-monotonicity argument seen in the proof of statement (i), that $\boldsymbol{r}_k$ are optimal transport maps relative to the costs $h_k(x - y)$, that $\boldsymbol{r}_k \in \operatorname{supp} \nu$ μ-a.e. and that $D \cdot \boldsymbol{r}_k \ge 0$. Since the approximate differentials coincide at points of density 1 of the coincidence set we have $\mu(\{\tilde{\nabla}\varphi_k \ne \tilde{\nabla}\varphi\}) \to 0$ as $k \to \infty$ and therefore $\mu(\{\boldsymbol{r}_k \ne \boldsymbol{r}\}) \to 0$ as $h \to \infty$. □

6.2.2 The infinite dimensional case

In the infinite dimensional case we consider for simplicity only the case when $c(x, y) = |x - y|^p/p$, $p > 1$; when ν has a bounded support we are still able to recover, by the same argument used in the finite dimensional case, a differential characterization of the optimal transport map.

We denote by $\mathscr{P}_p^r(X)$ the intersection of $\mathscr{P}_p(X)$ (see (5.1.22)) with $\mathscr{P}^r(X)$.

Theorem 6.2.10 (Optimal transport maps in Hilbert spaces). *Assume that X is a separable Hilbert space, let $\mu \in \mathscr{P}_p^r(X)$, $\nu \in \mathscr{P}_p(X)$ and let $c(x, y) = |x-y|^p/p$ for $p \in (1, +\infty)$, $q^{-1} + p^{-1} = 1$. Then the Kantorovich problem* (6.0.2) *has a unique solution $\boldsymbol{\mu}$ and this solution is induced by an* optimal transport*, i.e. there exists* *a Borel map* $\boldsymbol{r} \in L^p(X, \mu; X)$ *such that the representation* (5.2.13) *holds. If* ν *has a bounded support we have also*

$$\boldsymbol{r}(x) = x - |\nabla\varphi(x)|^{q-2}\nabla\varphi(x) \qquad \text{for } \mu\text{-a.e. } x, \tag{6.2.7}$$

for some locally Lipschitz, c-concave and maximal Kantorovich potential φ *(here* $\nabla\varphi$ *denotes the Gateaux differential of* φ*).*

Proof. Let us assume first that $\operatorname{supp}\nu$ is bounded. We first define a canonical Kantorovich potential, taking into account the boundedness assumption on $\operatorname{supp}\nu$, as follows. Let ϕ be any maximal Kantorovich potential and define

$$\varphi(x) := \inf_{y \in \operatorname{supp}\nu} c(x, y) - \phi^c(y) \qquad x \in X. \tag{6.2.8}$$

Notice that the optimality conditions on ϕ ensure that for μ-a.e. x the infimum above is attained. By construction φ is a locally Lipschitz function and it is still a maximal Kantorovich potential. Indeed, $\varphi = \phi$ μ-a.e. and since φ is the c-transform of the function ψ equal to ϕ^c on $\operatorname{supp}\nu$ and equal to $-\infty$ otherwise we have $\varphi^c = (\psi^c)^c \geq \psi = \phi^c$ on $\operatorname{supp}\nu$.

As in the proof of Theorem 6.2.4 it can be shown that for μ-a.e. x there is only one y such that $\varphi(x) + \varphi^c(y) = c(x, y)$, and that y is given by (6.2.7); the only difference is that we have to consider Theorem 6.2.3 instead of the classical Rademacher theorem.

In the general case when $\operatorname{supp}\nu$ is possibly unbounded we can still prove existence and uniqueness of an optimal transport map as follows. Let $\boldsymbol{\gamma} \in \Gamma_o(\mu, \nu)$, let $\boldsymbol{\gamma}_n = \chi_{B_n}(y)\boldsymbol{\gamma}$ where $B_n := B_n(0)$ is the centered open ball of radius n, and let μ_n, ν_n be the marginals of $\boldsymbol{\gamma}_n$ (in particular $\nu_n = \chi_{B_n}\nu$ and μ_n is absolutely continuous with respect to μ, therefore still regular). By Theorem 6.1.5 we know that $\operatorname{supp}\boldsymbol{\gamma}$ is $|\cdot|^p$-monotone, and therefore $\operatorname{supp}\boldsymbol{\gamma}_n$ is $|\cdot|^p$-monotone as well. By applying Theorem 6.1.5 again and the first part of the present proof, we obtain that $\boldsymbol{\gamma}_n$ is an optimal plan, induced by a unique transport map $\boldsymbol{r}_n$. The inequality

$$(\boldsymbol{i} \times \boldsymbol{r}_n)_\# \mu_n = \boldsymbol{\gamma}_n \leq \boldsymbol{\gamma}_m = (\boldsymbol{i} \times \boldsymbol{r}_m)_\# \mu_m$$

immediately gives (for instance by disintegration of both sides with respect to x)

$$\boldsymbol{r}_n = \boldsymbol{r}_m \quad \mu_n\text{-a.e. whenever } n < m.$$

Therefore the map $\boldsymbol{r}$ such that $\boldsymbol{r} = \boldsymbol{r}_n$ μ_n-a.e. for any n is well defined, and passing to the limit as $n \to \infty$ in the identity $\boldsymbol{\gamma}_n = (\boldsymbol{i} \times \boldsymbol{r})_\# \mu_n$ we obtain $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{r})_\# \mu$. This proves that $\boldsymbol{r}$ is an optimal transport map, and that any optimal plan is induced by an optimal transport map.

If there were two different optimal transport maps $\boldsymbol{r}$, $\boldsymbol{r}'$, then we could build an optimal transport plan

$$\boldsymbol{\gamma} := \frac{1}{2}\int_X \delta_{\boldsymbol{r}(x)} + \delta_{\boldsymbol{r}'(x)}\, d\mu(x)$$

which is not induced by any transport map. This contradiction proves the uniqueness of $\boldsymbol{r}$. □

Remark 6.2.11 (Essential injectivity of the transport map). Notice also that if ν is regular as well, under the assumption of Theorem 6.2.4 or Theorem 6.2.10, then the optimal transport map $\boldsymbol{r}$ between μ and ν is *μ-essentially injective* (i.e. its restriction to a set with full μ-measure is injective). This follows by the fact that, denoting by $\boldsymbol{s}$ the optimal transport map between ν and μ, the uniqueness of optimal plans gives $(\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu = [(\boldsymbol{s} \times \boldsymbol{i})_{\#}\nu]^{-1}$, which leads to $\boldsymbol{s} \circ \boldsymbol{r} = \boldsymbol{i}$ μ-a.e. and to the essential injectivity of $\boldsymbol{r}$.

In the case when $p = 2$ and $\mu, \nu \in \mathscr{P}_2^r(\mathbb{R}^d)$ we can actually prove *strict monotonicity* of the optimal transport map.

Proposition 6.2.12 (Strict monotonicity of $\boldsymbol{r}$). *Let $\mu, \nu \in \mathscr{P}_2^r(\mathbb{R}^d)$, and let $\boldsymbol{r}$ be the unique optimal transport map relative to the cost $c(x,y) = |x-y|^2/2$. Then $\nabla \boldsymbol{r} > 0$ μ-a.e. and there exists a μ-negligible set $N \subset \mathbb{R}^d$ such that*

$$\langle \boldsymbol{r}(x_1) - \boldsymbol{r}(x_2), x_1 - x_2 \rangle > 0 \qquad \forall x_1,\, x_2 \in \mathbb{R}^d \setminus N. \tag{6.2.9}$$

Proof. Let φ be a c-concave maximal Kantorovich potential. The c-concavity of φ and its construction ensure that $\varphi < +\infty$ globally, that $\varphi > -\infty$ μ-a.e. and that $\varphi - |x|^2/2$ is concave. In particular, denoting by C the interior of the convex hull of $\{\varphi \in \mathbb{R}\}$, we have that φ is finite on C and μ is concentrated on C. We have also that the optimal transport map $\boldsymbol{r}$ can be represented as $\nabla\phi$ with $\phi = |x|^2/2 - \varphi$ convex. Recalling that, by Alexandrov's theorem 5.5.4 convex functions are twice differentiable $\mathscr{L}^d$-a.e. in the classical sense, we can apply Lemma 5.5.3 to obtain that $\nabla \boldsymbol{r} > 0$ μ-a.e. in C, due to the fact that $\boldsymbol{r}_{\#}\mu \ll \mathscr{L}^d$.

Let now N be the μ-negligible set of points $x \in C$ where either ϕ is not twice differentiable or $\nabla^2\phi$ has some zero eigenvalue. The monotonicity inequality then gives (with $x_t = (1-t)x + ty$)

$$\langle \nabla\phi(y) - \nabla\phi(x), y - x \rangle \geq \lim_{t \downarrow 0} \frac{1}{t^2} \langle \nabla\phi(x_t) - \nabla\psi(x), x_t - x \rangle > 0$$

for any $x,\, y \in C \setminus N$. □

6.2.3 The quadratic case $p = 2$

In the case of $c(x,y) := \frac{1}{2}|x-y|^2$ in a Hilbert space X, the theory developed in the previous sections presents some more interesting features and stronger links with classical convex analysis.

Here we quote the most relevant aspects.

- *A function $u : X \to \bar{\mathbb{R}}$ is c-concave iff $u - \frac{1}{2}|\cdot|^2$ is u.s.c. and concave, i.e. $\tilde{u}(x) := \frac{1}{2}|x|^2 - u(x)$ is l.s.c. and convex.*
  For, from the representation of (6.1.2) we get

$$u(x) - \frac{1}{2}|x|^2 = \inf_{i \in I} t_i + \frac{1}{2}|y_i|^2 - \langle x, y_i \rangle.$$

This means that $u(x) - |x|^2/2$ is the infimum of a family of linear continuous functional on X.

- *If $v = u^c$ is the c-transform of u then $\tilde v = \tilde u^*$, the Legendre-Fenchel-Moreau conjugate functional defined as*

$$\tilde u^*(y) := \sup_{x\in X} \langle x, y\rangle - \tilde u(x).$$

We simply have

$$\tilde v(y) = \frac{1}{2}|y|^2 - u^c(y) = \sup_{x\in X} \frac{1}{2}|y|^2 - \frac{1}{2}|x-y|^2 + u(x)$$
$$= \sup_{x\in X} \langle x, y\rangle - \big(\frac{1}{2}|x|^2 - u(x)\big) = \sup_{x\in X} \langle x, y\rangle - \tilde u(x).$$

- *A subset Γ of X^2 is c-monotone according to Definition* 6.1.3 *iff it is* cyclically monotone*, i.e. for every cyclical choice of points $(x_1^k, x_2^k) \in \Gamma$, $k = 0, \ldots, N$, with $(x_1^0, x_2^0) = (x_1^N, x_2^N)$, we have*

$$\sum_{k=1}^{N} \langle x_1^k - x_1^{k-1}, x_2^k\rangle \geq 0. \tag{6.2.10}$$

In particular, by Rockafellar theorem, c-monotone sets are always contained in the graph of the subdifferential

$$\{(x, y) : \ y \in \partial\varphi(x)\}$$

of a convex l.s.c function φ. Conversely, any subset of such a graph is c-monotone.

- *Suppose that μ, $\nu \in \mathscr{P}_2(X)$ and $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$. Then the following properties are equivalent:*
  - *$\boldsymbol{\gamma}$ is optimal;*
  - supp $\boldsymbol{\gamma}$ *is cyclically monotone;*
  - *there exists a convex, l.s.c. potential $\tilde\varphi \in L^1(X, \mu)$ such that*

$$\langle x, y\rangle = \tilde\varphi(x) + \tilde\varphi^*(y) \quad \boldsymbol{\gamma}\text{-a.e. in } X^2. \tag{6.2.11}$$

Equivalently, we can also state (6.2.11) by saying that $y \in \partial\varphi(x)$ for $\boldsymbol{\gamma}$-a.e. $(x, y) \in X^2$. In particular, if $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu$ then there exists a l.s.c. convex functional φ such that $\boldsymbol{r}(x) \in \partial\varphi(x)$ for μ-a.e. $x \in X$.

- *Suppose that $X = \mathbb{R}^d$ and $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$, $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. Then there exists a unique optimal transport plan and this plan is induced by a transport map $\boldsymbol{r}$. If $\nu \in \mathscr{P}_2^r(\mathbb{R}^d)$ as well, then $\boldsymbol{r}$ is μ-essentially injective and fulfills* (6.2.9).

Chapter 7

The Wasserstein Distance and its Behaviour along Geodesics

In this chapter we will introduce the p-th Wasserstein distance $W_p(\mu,\nu)$ between two measures $\mu,\nu \in \mathscr{P}_p(X)$. The first section is devoted to its preliminary properties, in connection with the optimal transportation problems studied in the previous chapter and with narrow convergence: the main topological results are valid in general metric spaces.

In the last two sections we will focus our attention to the case when X is an Hilbert space: we will characterize the (minimal, constant speed) geodesics with respect to the Wasserstein distance and, for $p=2$ and a given $\nu \in \mathscr{P}_2(X)$, we will study the behaviour of the map $\mu \mapsto W_2^2(\mu,\nu)$ along geodesics: in particular, we will give a precise formula for its derivative along geodesics and and we will prove its semi-concavity, an important geometric property which is related to a metric version of suitable curvature inequalities.

7.1 The Wasserstein distance

Let X be a separable metric space satisfying the Radon property (5.1.9) and $p \geq 1$. The (p-th) *Wasserstein distance* between two probability measures μ^1, $\mu^2 \in \mathscr{P}_p(X)$ is defined by

$$\begin{aligned} W_p^p(\mu^1,\mu^2) &:= \min\Big\{ \int_{X^2} d(x_1,x_2)^p \, d\boldsymbol{\mu}(x_1,x_2) : \ \boldsymbol{\mu} \in \Gamma(\mu^1,\mu^2) \Big\} \\ &= \min\Big\{ \boldsymbol{d}(x_1,x_2)^p_{L^p(\boldsymbol{\mu};X)} : \ \boldsymbol{\mu} \in \Gamma(\mu^1,\mu^2) \Big\}. \end{aligned} \tag{7.1.1}$$

Using Remark 5.3.3 we can show that the function defined above is indeed a distance. Indeed, if $\mu^i \in \mathscr{P}_p(X)$ for $i=1,2,3$, $\boldsymbol{\gamma}^{1\,2}$ is optimal between μ^1 and μ^2 and $\boldsymbol{\gamma}^{2\,3}$ is optimal between μ^2 and μ^3 we can find $\boldsymbol{\gamma} \in \mathscr{P}(X^3)$ such that

$\pi^{1\,2}_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}^{1\,2}$ and $\pi^{2\,3}_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}^{2\,3}$. The plan $\boldsymbol{\gamma}^{1\,3} := \pi^{1\,3}_{\#}\boldsymbol{\gamma}$ belongs to $\Gamma(\mu^1,\mu^3)$ and since

$$W_p(\mu^1,\mu^2) = \boldsymbol{d}(x_1,x_2)_{L^p(\boldsymbol{\gamma};X)}, \quad W_p(\mu^2,\mu^3) = \boldsymbol{d}(x_1,x_2)_{L^p(\boldsymbol{\gamma};X)}$$

and

$$\boldsymbol{d}(x_1,x_3)_{L^p(\boldsymbol{\gamma}^{1\,3};X)} = \boldsymbol{d}(x_1,x_3)_{L^p(\boldsymbol{\gamma};X)},$$

we immediately get $W_p(\mu^1,\mu^3) \le W_p(\mu^1,\mu^2) + W_p(\mu^2,\mu^3)$ from the standard triangle inequality of the L^p distance.

In the particular case when $p = 1$ and μ and ν have a bounded support we can use the duality formula (6.1.1) and the fact that c-concavity coincides with 1-Lipschitz continuity and $\varphi^c = -\varphi$ for the cost $c(x,y) = d(x,y)$ to obtain

$$W_1(\mu,\nu) = \sup\left\{\int \varphi\, d(\mu-\nu) : \ \varphi : X \to \mathbb{R} \ \text{1-Lipschitz}\right\}. \tag{7.1.2}$$

We denote by $\Gamma_o(\mu^1,\mu^2) \subset \Gamma(\mu^1,\mu^2)$ (which also depends on p, even if we omit to indicate explicitly this dependence) the convex and narrowly compact set of *optimal plans* where the minimum is attained, i.e.

$$\boldsymbol{\gamma} \in \Gamma_o(\mu^1,\mu^2) \quad \Longleftrightarrow \quad \int_{X^2} d(x_1,x_2)^p\, d\boldsymbol{\gamma}(x_1,x_2) = W_p^p(\mu^1,\mu^2). \tag{7.1.3}$$

When $\Gamma_o(\mu^1,\mu^2)$ contains a unique plan $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu^1$ induced by a transport map $\boldsymbol{r}$ as in (5.2.13), we will also denote $\boldsymbol{r}$ by $\boldsymbol{t}_{\mu^1}^{\mu^2}$; therefore $\boldsymbol{t}_{\mu^1}^{\mu^2}$ is characterized by

$$\boldsymbol{t}_{\mu^1}^{\mu^2} : X \to X, \quad \big(\boldsymbol{t}_{\mu^1}^{\mu^2}\big)_{\#}\mu^1 = \mu^2, \quad \Gamma_o(\mu^1,\mu^2) = \big\{\big(\boldsymbol{i} \times \boldsymbol{t}_{\mu^1}^{\mu^2}\big)_{\#}\mu^1\big\}, \tag{7.1.4}$$

it is the unique (strict) minimizer of the optimal transportation problem in the original Monge's formulation (6.0.1), and satisfies

$$\int_X d\big(x, \boldsymbol{t}_{\mu^1}^{\mu^2}(x)\big)^p\, d\mu^1(x) = W_p^p(\mu^1,\mu^2). \tag{7.1.5}$$

Given $\boldsymbol{\mu}$-measurable maps $\boldsymbol{r}, \boldsymbol{s} : \boldsymbol{X} \to X$, a very useful inequality giving an estimate from above of the Wasserstein distance is

$$W_p(\boldsymbol{r}_{\#}\boldsymbol{\mu}, \boldsymbol{s}_{\#}\boldsymbol{\mu}) \le \boldsymbol{d}(\boldsymbol{r},\boldsymbol{s})_{L^p(\boldsymbol{\mu};X)}. \tag{7.1.6}$$

It holds because $\boldsymbol{\gamma} = (\boldsymbol{r},\boldsymbol{s})_{\#}\boldsymbol{\mu} \in \Gamma(\boldsymbol{r}_{\#}\boldsymbol{\mu}, \boldsymbol{s}_{\#}\boldsymbol{\mu})$ and $\int d(x_1,x_2)^p\, d\boldsymbol{\gamma} = \boldsymbol{d}(\boldsymbol{r},\boldsymbol{s})^p_{L^p(\boldsymbol{\mu};X)}$.

From Theorem 6.1.4 we derive that $\boldsymbol{\mu}$ is optimal iff its support is $d(\cdot,\cdot)^p$-*monotone* according to Definition 6.1.3, i.e.

$$\sum_{k=1}^N d\big(x_1^k, x_2^k\big)^p \le \sum_{k=1}^N d\big(x_1^k, x_2^{\sigma(k)}\big)^p \tag{7.1.7}$$

for every choice of $(x_1^k, x_2^k) \in \operatorname{supp} \boldsymbol{\mu}$, $k = 1, \ldots, N$, and for every permutation $\sigma : \{1, \ldots, N\} \to \{1, \ldots, N\}$ (actually Theorem 6.1.4 shows only that $\boldsymbol{\mu}$ has to be concentrated on a c-monotone set, but since in this case the cost is continuous the c-monotonicity holds, by a density argument, for the whole support of $\boldsymbol{\mu}$).

Remark 7.1.1. It is not difficult to check that supports of optimal plans satisfy the slightly stronger property

$$\overline{\bigcup_{\boldsymbol{\gamma} \in \Gamma_o(\mu^1, \mu^2)} \operatorname{supp} \boldsymbol{\gamma}} \quad \text{is } d(\cdot, \cdot)^p - \text{monotone.} \tag{7.1.8}$$

For, we take a sequence $(\boldsymbol{\gamma}_n)$ narrowly dense in $\Gamma_o(\mu^1, \mu^2)$ and we consider the new plan $\bar{\boldsymbol{\gamma}} := \sum_n 2^{-n} \boldsymbol{\gamma}_n$. The plan $\bar{\boldsymbol{\gamma}}$ is optimal, too, and its support coincides with (7.1.8).

Remark 7.1.2 (Cyclical monotonicity in the case when X is Hilbert). When $p = 2$ and X is a (pre-)Hilbert space, condition (7.1.7) is equivalent to the classical *cyclical monotonicity* of $\operatorname{supp} \boldsymbol{\mu}$, i.e. for every cyclical choice of points $(x_1^k, x_2^k) \in \operatorname{supp} \boldsymbol{\mu}$, $k = 0, \ldots, N$, with $(x_1^0, x_2^0) = (x_1^N, x_2^N)$, we have

$$\sum_{k=1}^{N} \langle x_1^k - x_1^{k-1}, x_2^k \rangle \geq 0. \tag{7.1.9}$$

In particular, if $\boldsymbol{r} = \nabla \phi$ for some convex C^1 function ϕ then $\boldsymbol{r}$ is a 2-optimal transport map for every measure $\mu \in \mathscr{P}_2(X)$ such that $\int |\boldsymbol{r}|^2 \, d\mu < +\infty$.

A useful application of the necessary and sufficient optimality conditions is given by the following stability of optimality with respect to narrow convergence.

Proposition 7.1.3 (Stability of optimality and narrow lower semicontinuity). *Let $(\mu_n^1), (\mu_n^2) \subset \mathscr{P}_p(X)$ be two sequences narrowly converging to μ^1, μ^2 respectively, and let $\boldsymbol{\mu}_n \in \Gamma_o(\mu_n^1, \mu_n^2)$ be a sequence of optimal plans with $\int_{X^2} d(x_1, x_2)^p \, d\boldsymbol{\mu}_n$ bounded.*
Then $(\boldsymbol{\mu}_n)$ is narrowly relatively compact in $\mathscr{P}(X^2)$ and any narrow limit point $\boldsymbol{\mu}$ belongs to $\Gamma_o(\mu^1, \mu^2)$, with

$$\begin{aligned} W_p(\mu^1, \mu^2) &= \int_{X^2} d(x_1, x_2)^p \, d\boldsymbol{\mu}(x_1, x_2) \\ &\leq \liminf_{n \to \infty} \int_{X^2} d(x_1, x_2)^p \, d\boldsymbol{\mu}_n(x_1, x_2) = \liminf_{n \to \infty} W_p(\mu_n^1, \mu_n^2). \end{aligned} \tag{7.1.10}$$

Proof. The relative compactness of the sequence $(\boldsymbol{\mu}_n)$ is a consequence of Lemma 5.2.2 and the "lim inf" inequality in (7.1.10) is a direct consequence of (5.1.15), which in particular yields $\int_{X^2} d(x_1, x_2)^p \, d\boldsymbol{\mu} < +\infty$.
Using proposition 5.1.8 it is immediate to check by approximation that the support of $\boldsymbol{\mu}$ is $d(\cdot, \cdot)^p$-monotone. □

When X is a Hilbert space, the Wasserstein distance is lower semicontinuous w.r.t. the weaker narrow convergence in $\mathscr{P}(X_\varpi)$:

Lemma 7.1.4 (Weak narrow lower semicontinuity of W_p in Hilbert spaces). *Let X be a (separable) Hilbert space and let (μ_n^1), $(\mu_n^2) \subset \mathscr{P}_p(X)$ be two weakly tight sequences (according to (5.1.32)) narrowly converging to μ^1, μ^2 in $\mathscr{P}(X_\varpi)$. Then*

$$W_p(\mu^1, \mu^2) \le \liminf_{n\to\infty} W_p(\mu_n^1, \mu_n^2). \tag{7.1.11}$$

Proof. The map $(x_1, x_2) \mapsto |x_1 - x_2|^p$ is weakly l.s.c. in $X \times X$: we simply argue as in the previous proof and we apply Lemma 5.1.12(c). Notice that in this case the first line of (7.1.10) is an inequality "$\le$", since we do not know that the limit plan μ is optimal any more; nevertheless, the inequality is sufficient to obtain (7.1.11). $\square$

Proposition 7.1.5 (Convergence, compactness and completeness). *$\mathscr{P}_p(X)$ endowed with the p-Wasserstein distance is a separable metric space which is complete if X is complete. A set $\mathcal{K} \subset \mathscr{P}_p(X)$ is relatively compact iff it is p-uniformly integrable and tight. In particular, for a given sequence $(\mu_n) \subset \mathscr{P}_p(X)$ we have*

$$\lim_{n\to\infty} W_p(\mu_n, \mu) = 0 \iff \begin{cases} \mu_n \text{ narrowly converge to } \mu, \\ (\mu_n) \text{ has uniformly integrable } p\text{-moments.} \end{cases} \tag{7.1.12}$$

Proof. Let us first prove the completeness of $\mathscr{P}_p(X)$, by assuming that X is complete. It suffices to show that any sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathscr{P}_p(X)$ such that

$$\sum_{n=1}^{\infty} W_p(\mu^n, \mu^{n+1}) < +\infty$$

is converging. We choose $\boldsymbol{\alpha}^{n\,(n+1)} \in \Gamma_o(\mu^n, \mu^{n+1})$ and use Lemma 5.3.4 to find $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$, with $\boldsymbol{X} = X^{\mathbb{N}}$, satisfying (5.3.8). It follows that

$$\sum_{n=1}^{\infty} \boldsymbol{d}(\pi^n, \pi^{n+1})_{L^p(\boldsymbol{\mu};X)} < +\infty.$$

Therefore, (π^n) is a Cauchy sequence in $L^p(\boldsymbol{\mu}; X)$, which is a complete metric space, and admits a limit map $\pi^\infty \in L^p(\boldsymbol{\mu}; X)$. Setting $\mu_\infty := \pi^\infty_\# \boldsymbol{\mu} \in \mathscr{P}_p(X)$, we easily find

$$\begin{aligned} \limsup_{n\to\infty} W_p(\mu^n, \mu^\infty) &\le \limsup_{n\to\infty} \boldsymbol{d}(\pi^n, \pi^\infty)_{L^p(\boldsymbol{\mu};X)} \\ &\le \limsup_{n\to\infty} \sum_{j=n}^{\infty} \boldsymbol{d}(\pi^{j+1}, \pi^j)_{L^p(\boldsymbol{\mu};X)} = 0. \end{aligned}$$

We will prove now the equivalence (7.1.12) (a different argument in locally compact spaces, based on the duality formula (7.1.2), is available for instance in [126]).

First we suppose that $W_p(\mu_n,\mu) \to 0$. Arguing as before, we can choose optimal plans $\boldsymbol{\beta}^{1\,n} \in \Gamma_o(\mu,\mu_n,)$ and use Lemma 5.3.4 (with $\mu_1 := \mu$) to find $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$ satisfying (5.3.8). It follows that

$$\lim_{n\to\infty} \boldsymbol{d}(\pi^n,\pi^1)_{L^p(\boldsymbol{X},\boldsymbol{\mu};X)} = 0,$$

and therefore, for every continuous real function f with p-growth the Vitali dominated convergence theorem gives

$$\lim_{n\to\infty}\int_X f(x)\,d\mu_n(x) = \lim_{n\to\infty}\int_{\boldsymbol{X}} f(\pi^n(\boldsymbol{x}))\,d\boldsymbol{\mu}(\boldsymbol{x}) = \int_{\boldsymbol{X}} f(\pi^1(\boldsymbol{x}))\,d\boldsymbol{\mu}(\boldsymbol{x})$$
$$= \int_X f(x)\,d\mu(x).$$

By lemma 5.1.7 we obtain the narrow convergence and the uniform p-integrability of the sequence (μ_n).

Conversely, let us suppose that the sequence (μ_n) has uniformly integrable p-moments and it is narrowly converging to μ; in particular, by (5.4.7), the set $\{\mu,\mu_n : n\in\mathbb{N}\}$ is tight. As before, let us choose $\boldsymbol{\alpha}^{1\,n} \in \Gamma_o(\mu,\mu_n)$: it easy to check that the sequence $(\boldsymbol{\alpha}^{1\,n})$ is p-uniformly integrable and tight in $\mathscr{P}(X\times X)$ (see Lemma 5.2.2): a subsequence $k\mapsto n_k$ exists such that $\boldsymbol{\alpha}^{1\,n_k} \to \boldsymbol{\alpha}$ narrowly, with $\boldsymbol{\alpha} \in \Gamma_o(\mu,\mu)$ by Proposition 7.1.3. Applying Lemma 5.1.7 we get

$$\lim_{k\to\infty} W_p^p(\mu,\mu_{n_k}) = \lim_{k\to\infty}\int_{X\times X} |x_1-x_2|^p\,d\boldsymbol{\alpha}^{1\,n_k}(x_1,x_2)$$
$$= \int_{X\times X} |x_1-x_2|^p\,d\boldsymbol{\alpha}(x_1,x_2) = 0.$$

Since the limit is independent of the subsequence n_k we get the convergence of μ_n with respect to the Wasserstein distance. Using (7.1.12) it is now immediate to check that convex combinations of Dirac masses with centers in a countable dense subset of X and with rational coefficients are dense in $\mathscr{P}_p(X)$, therefore $\mathscr{P}_p(X)$ is separable. □

It is interesting to note that in the previous proof of the equivalence between narrow and Wasserstein topology (on sets with uniformly integrable p-moments), one implication (the topology induced by the Wasserstein distance is stronger than the narrow one) could be directly deduced from (7.1.2) via the approximation arguments discussed in Section 5.1, thus avoiding Lemma 5.3.4; this implication is therefore considerably easier than the converse one, which relies on the stability property 7.1.3 and therefore on the main characterization results of Chapter 6 for optimal transportation problems. However the argument via Lemma 5.3.2 seems to be necessary to get completeness, at least in infinite dimensions.

Remark 7.1.6 (Limit of the optimal plan). As a byproduct of the previous proof, we obtain that if $\mu_n \to \mu$ in $\mathscr{P}_p(X)$ and $\boldsymbol{\mu}_n \in \Gamma_o(\mu,\mu_n)$, then

$$\boldsymbol{\mu}_n \to (\boldsymbol{i}\times\boldsymbol{i})_{\#}\mu \quad \text{in } \mathscr{P}_p(X\times X). \tag{7.1.13}$$

Remark 7.1.7 ($\mathscr{P}(X)$ is a Polish space if X is Polish). By taking an equivalent *bounded* metric on X, all the Wasserstein distances induce the topology of narrow convergence between probability measures: as we already noticed in Remark 5.1.1, the narrow topology $\mathscr{P}(X)$ is metrizable; moreover, if X is a Polish space, then $\mathscr{P}(X)$ is a Polish space, too.

Remark 7.1.8 (Relative compactness of $\mathscr{P}_p(X)$-bounded sets). When X is infinite dimensional Hilbert space, bounded subset in $\mathscr{P}_p(X)$ are not relatively compact in $\mathscr{P}(X)$ any more, but they are relatively compact in $\mathscr{P}(X_\varpi)$.

Remark 7.1.9 ($\mathscr{P}_p(X)$ is locally compact only if X is compact). If X is not compact, the space $\mathscr{P}_p(X)$ is not locally compact, not even in the case when $X = \mathbb{R}^d$ is finite dimensional. Indeed, assume that for some $\epsilon > 0$ and $x_0 \in X$ the closed ball in $\mathscr{P}_p(X)$

$$\mathcal{B}_\varepsilon := \Big\{\mu \in \mathscr{P}_p(X) : W_p(\mu, \delta_{x_0}) \le \varepsilon\Big\} = \Big\{\mu \in \mathscr{P}_p(X) : \int_X d(x, x_0)^p \, d\mu(x) \le \varepsilon^p\Big\}$$

is compact and let us prove that an arbitrary sequence $(x_n) \in X$ admits a convergent subsequence. It is not restrictive to assume $\liminf_{n\to\infty} d(x_n, x_0) > 0$ (otherwise (x_n) admits a subsequence converging to x_0), and therefore $\inf_{n\in\mathbb{N}} d(x_n, x_0) = \delta > 0$. We consider the real numbers

$$m_n = \frac{(\delta \wedge \varepsilon)^p}{d(x_n, x_0)^p} \le 1, \quad \text{so that} \quad m_n d(x_n, x_0)^p = (\delta \wedge \varepsilon)^p;$$

the sequence of measures $\mu_n := (1 - m_n)\delta_{x_0} + m_n\delta_{x_n}$ belongs to $\mathcal{B}_\varepsilon$ since $W_p(\mu_n, \delta_{x_0}) = \varepsilon \wedge \delta$ and therefore admits a subsequence $(\mu_{n'})$ converging to some $\mu \ne \delta_{x_0}$ in $\mathscr{P}_p(X)$.
Since (m_n) is bounded, too, it is not restrictive to assume that $m_{n'} \to m \in [0, 1]$ which should be strictly positive, being $\mu \ne \delta_{x_0}$. By Proposition 5.1.8 (see also Corollary 5.1.9) it follows that μ takes the form $(1 - m)\delta_{x_0} + m\delta_x$ for some $x \in X$, and therefore $x_{n'} \to x$.

Lemma 7.1.10 (Approximation by convolution). *Let $\mu \in \mathscr{P}_p(\mathbb{R}^d)$ and let $(\rho_\varepsilon) \subset C^\infty(\mathbb{R}^d)$ be a family of nonnegative mollifiers such that*

$$\rho_\varepsilon(x) := \varepsilon^{-d}\rho(x/\varepsilon), \; \int_{\mathbb{R}^d} \rho(x)\, dx = 1, \; \mathsf{m}_p^p(\rho) := \int_{\mathbb{R}^d} |x|^p \rho(x)\, dx < +\infty. \quad (7.1.14)$$

Then if $\mu_\varepsilon := \mu * \rho_\varepsilon$

$$W_p(\mu, \mu_\varepsilon) \le \varepsilon \mathsf{m}_p(\rho), \quad (7.1.15)$$

and therefore μ_ε converges to μ in $\mathscr{P}_p(\mathbb{R}^d)$ as $\varepsilon \downarrow 0$.

Proof. We introduce the family of plans $\boldsymbol{\gamma}_\varepsilon := \int \rho_\varepsilon(\cdot - x)\mathscr{L}^d \, d\mu(x)$ defined by

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x, y)\, d\boldsymbol{\gamma}_\varepsilon(x, y) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x, y)\rho_\varepsilon(y - x)\, dy\, d\mu(x)$$

which obviously satisfy $\boldsymbol{\gamma}_\varepsilon \in \Gamma(\mu, \mu_\varepsilon)$. Therefore

$$W_p^p(\mu, \mu_\varepsilon) \le \iint_{(\mathbb{R}^d)^2} |x-y|^p \, d\boldsymbol{\gamma}_\varepsilon(x,y) = \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |x-y|^p \rho_\varepsilon(y-x) \, dy \Big) \, d\mu(x)$$
$$= \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |z|^p \rho_\varepsilon(z) \, dz \Big) \, d\mu(x) = \int_{\mathbb{R}^d} |\varepsilon z|^p \rho(z) \, dz = \varepsilon^p \int_{\mathbb{R}^d} |z|^p \rho(z) \, dz \qquad \square$$

Remark 7.1.11. Combining Proposition 5.1.13 with $j(r) := r^p$, $1 < p < +\infty$, and Lemma 5.1.7 we get the following useful characterization of the convergence in $\mathscr{P}_p(X)$, which is particularly interesting when X is infinite dimensional Hilbert space:

$$\lim_{n\to\infty} W_p(\mu_n, \mu) = 0 \iff \begin{cases} \mu_n \text{ narrowly converge to } \mu \text{ in } \mathscr{P}(X_\varpi), \\ \lim_{n\to\infty} \int_X |x|^p \, d\mu_n(x) = \int_X |x|^p \, d\mu(x). \end{cases} \tag{7.1.16}$$

Since we have at our disposal new powerful results (which are consequences of the theory presented in Chapter 6) we conclude this section by showing a simpler proof of (7.1.16), which could be extended to the case of uniformly convex Banach spaces.

Proof. Let us consider the (Radon, separable) metric space X_ϖ with the distance induced by the norm $\|\cdot\|_\varpi$; since $\|\cdot\|_\varpi^p \le |\cdot|^p$, (7.1.16) and Lemma 5.1.7 show that $\|\cdot\|_\varpi^p$ is uniformly integrable w.r.t. the sequence (μ_n). Applying (7.1.12) of Proposition 7.1.5 in X_ϖ (this characterization does not require the completeness of the metric space), we obtain that μ_n converges to μ in the p-Wasserstein distance of $\mathscr{P}_p(X_\varpi)$. It follows by Remark 7.1.6 that any sequence of plans $\boldsymbol{\mu}_n \in \Gamma(\mu_n, \mu)$, optimal in $\mathscr{P}_p(X_\varpi)$, satisfies

$$\boldsymbol{\mu}_n \to (\boldsymbol{i} \times \boldsymbol{i})_\# \mu \quad \text{in } \mathscr{P}_p(X_\varpi \times X_\varpi) \quad \text{as } n \to \infty. \tag{7.1.17}$$

We suppose $p \ge 2$ and we integrate with respect to $\boldsymbol{\mu}_n$ the inequality (c_p is a strictly positive constant, $j_p(x_1) = |x_1|^{p-2} x_1$)

$$c_p |x_1 - x_2|^p \le \frac{1}{p}|x_2|^p - \frac{1}{p}|x_1|^p - \langle j_p(x_1), x_2 - x_1 \rangle \quad \forall\, x_1, x_2 \in X,$$

which we will prove in Lemma 10.2.1; we obtain

$$c_p W_p^p(\mu, \mu_n) \le \int_{X\times X} c_p |x_1 - x_2|^p \, d\boldsymbol{\mu}_n(x_1, x_2) \tag{7.1.18a}$$
$$\le \int_{X\times X} \Big(\frac{1}{p}|x_2|^p - \frac{1}{p}|x_1|^p - \langle j_p(x_1), x_2 - x_1 \rangle \Big) \, d\boldsymbol{\mu}_n(x_1, x_2)$$
$$= \frac{1}{p} \int_X |x_2|^p \, d\mu_n(x_2) - \frac{1}{p} \int_X |x_1|^p \, d\mu(x_1) - \int_{X\times X} \langle y_1, y_2 \rangle \, d\hat{\boldsymbol{\mu}}_n(y_1, y_2), \tag{7.1.18b}$$

where

$$\hat{\boldsymbol{\mu}}_n := \big(j_p \circ \pi^1, \pi^2 - \pi^1\big)_{\#}\boldsymbol{\mu}_n.$$

Since the first marginal of $\hat{\boldsymbol{\mu}}_n$ is fixed in $\mathscr{P}_q(X)$, it is easy to check by Lemma 5.2.1 that

$$\hat{\boldsymbol{\mu}}_n \to \big((j_p)_{\#}\mu\big) \times \delta_0 \quad \text{in } \mathscr{P}(X \times X_{\varpi}) \quad \text{as } n \to \infty,$$

and that $(\boldsymbol{\mu}_n)$ satisfies the assumptions of Lemma 5.2.4; therefore, passing to the limit as $n \to \infty$ in (7.1.18a,b), the convergence of the moments (7.1.16) and Lemma 5.2.4 yield $W_p(\mu, \mu_n) \to 0$.
The case $p < 2$ follows by the same argument and inequality (10.2.5). □

7.2 Interpolation and geodesics

In this section we are assuming that X is a separable Hilbert space and $p > 1$, and we show that constant speed geodesics in $\mathscr{P}_p(X)$ coincide with a suitable class of interpolations obtained from optimal transport plans. Recall that a curve $\mu_t \in \mathscr{P}_p(X)$, $t \in [0,1]$, is a constant speed geodesic (see also (2.4.3)) if

$$W_p(\mu_s, \mu_t) = (t-s)W_p(\mu_0, \mu_1) \quad \forall\, 0 \le s \le t \le 1. \tag{7.2.1}$$

If $\boldsymbol{\mu} \in \mathscr{P}(X^N)$, $N \ge 2$, $1 \le i,j,k \le N$, and $t \in [0,1]$ we set

$$\pi_t^{i\to j} := (1-t)\pi^i + t\pi^j : X^N \to X, \tag{7.2.2}$$

$$\pi_t^{i\to j,k} := (1-t)\pi^{i,k} + t\pi^{j,k} : X^N \to X^2, \tag{7.2.3}$$

$$\mu_t^{i\to j} := (\pi_t^{i\to j})_{\#}\boldsymbol{\mu} \in \mathscr{P}(X), \tag{7.2.4}$$

$$\boldsymbol{\mu}_t^{i\to j,k} := (\pi_t^{i\to j,k})_{\#}\boldsymbol{\mu} \in \mathscr{P}(X^2). \tag{7.2.5}$$

It is well known that $\Gamma_o(\mu^1, \mu^2)$ can contain in general more than one element. In the following lemma we show that along a geodesic the optimal plans to the extreme points μ_0, μ_1 are unique if we consider μ_t, $t \in (0,1)$, as the initial measure.

Lemma 7.2.1. *Let $(\mu_t)_{t\in[0,1]}$ be a constant speed geodesic in $\mathscr{P}_p(X)$ and let $t \in (0,1)$. Then $\Gamma_o(\mu_t, \mu_1)$ (resp. $\Gamma_o(\mu_0, \mu_t)$) contains a unique plan $\boldsymbol{\mu}^{t\,1}$ (resp. $\boldsymbol{\mu}^{0\,t}$) and this plan (resp. $(\boldsymbol{\mu}^{0\,t})^{-1}$) is induced by a transport. Moreover, $\boldsymbol{\mu} = \boldsymbol{\mu}^{t\,1} \circ \boldsymbol{\mu}^{0\,t} \in \Gamma_o(\mu_0, \mu_1)$ and*

$$\boldsymbol{\mu}^{0\,t} = (\pi_t^{1,1\to 2})_{\#}\boldsymbol{\mu}, \quad \boldsymbol{\mu}^{t\,1} = (\pi_t^{1\to 2,2})_{\#}\boldsymbol{\mu}. \tag{7.2.6}$$

Proof. For $t \in (0,1)$ let $\boldsymbol{\gamma}$ (resp. $\boldsymbol{\eta}$) be optimal transport plans between μ_0 and μ_t (resp. μ_t and μ_1). In order to clarify the structure of the proof it is convenient to view μ_0, μ_t, μ_1 as measures in $\mathscr{P}(X_1)$, $\mathscr{P}(X_2)$, $\mathscr{P}(X_3)$, where X_i are distinct copies of X. Then, we can define

$$\boldsymbol{\lambda} := \int_{X_2} \gamma_{x_2} \times \eta_{x_2}\, d\mu_t(x_2) \in \Gamma(\mu_0, \mu_t, \mu_1)$$

where $\gamma = \int_{X_2} \gamma_{x_2}\,d\mu_t$ and $\eta = \int_{X_2} \eta_{x_2}\,d\mu_t$ are the disintegrations of γ and η with respect to the common variable x_2. Then, since (recall the composition of plans in Remark 5.3.3)

$$\boldsymbol{\mu} = \boldsymbol{\eta} \circ \boldsymbol{\gamma} = \pi^{1,3}_{\#}\boldsymbol{\lambda} \in \Gamma(\mu_0, \mu_1)$$

we get

$$\begin{aligned} W_p(\mu_0,\mu_1) &\leq \|x_1 - x_3\|_{L^p(\boldsymbol{\mu};X)} \leq \|x_1 - x_2\|_{L^p(\boldsymbol{\lambda};X)} + \|x_2 - x_3\|_{L^p(\boldsymbol{\lambda};X)} \\ &= \|x_1 - x_2\|_{L^p(\boldsymbol{\gamma};X)} + \|x_2 - x_3\|_{L^p(\boldsymbol{\eta};X)} = W_p(\mu_0,\mu_1). \end{aligned}$$

This proves that $\boldsymbol{\mu}$ is optimal; moreover, since all inequalities are equalities and the L^p-norm is strictly convex, we get that there exists $\alpha > 0$ such that $x_2 - x_1 = \alpha(x_3 - x_1)$ for $\boldsymbol{\lambda}$-a.e. triple (x_1, x_2, x_3). Using the fact that $W_p(\mu_t, \mu_0) = tW_p(\mu_0, \mu_1)$ we obtain $\alpha = t$ and therefore

$$x_2 - x_1 = t(x_3 - x_1) \quad \boldsymbol{\lambda}\text{-a.e. in } X_1 \times X_2 \times X_3.$$

Denoting by $z(x_2)$ the barycenter of γ_{x_2}, the linearity of this relation w.r.t. x_1 yields

$$x_2 - z(x_2) = t(x_3 - z(x_2)) \quad \boldsymbol{\eta}\text{-a.e. in } X_2 \times X_3.$$

Hence $\boldsymbol{\eta}$ is induced by the transport $\boldsymbol{r}_t(x_2) = x_2/t - z(x_2)(1-t)/t$. Since z depends on $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ have been chosen independently, this proves that $\boldsymbol{\eta}$ is unique, so that $\boldsymbol{\eta} = \boldsymbol{\mu}^{t\,1}$, the measure defined in (7.2.6). Inverting the order of μ_0 and μ_1, we obtain the other identity. □

Theorem 7.2.2 (Characterization of constant speed geodesics). *If $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$ then the curve $t \mapsto \mu_t := \mu_t^{1\to 2}$ is a constant speed geodesic connecting μ^1 to μ^2. Conversely, any constant speed geodesic $\mu_t : [0,1] \to \mathscr{P}_p(X)$ connecting μ^1 to μ^2 has this representation for a suitable $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$, which can be constructed from any point μ_t, $0 < t < 1$, as in the previous Lemma.*

Proof. By (7.1.6) we get

$$W_p(\mu_t, \mu_s) \leq (t-s)W_p(\mu^1, \mu^2) \quad \forall s,\, t \in (0,1),\, s \leq t. \tag{7.2.7}$$

If there is a strict inequality for some $s < t$ we immediately derive a contradiction by applying the triangle inequality with the points μ_0, μ_s, μ_t and μ_1. Therefore equality holds and μ_t is a constant speed geodesic.

Let μ_t be a constant speed geodesic and for a fixed $t \in (0,1)$ let $\boldsymbol{\mu} := \boldsymbol{\mu}^{t\,1} \circ \boldsymbol{\mu}^{0\,t}$ be as in Lemma 7.2.1. Since $\boldsymbol{\mu}^{0\,t} = (\pi_t^{1,1\to 2})_{\#}\boldsymbol{\mu}$ is the unique element of $\Gamma_o(\mu_0, \mu_t)$ and the curve $s \mapsto \mu_{ts}$, $s \in [0,1]$ is a constant speed geodesic, we get

$$\mu_{st} = (\pi_s^{1\to 2})_{\#}\boldsymbol{\mu}^{0\,t} = (\pi_s^{1\to 2} \circ \pi_t^{1,1\to 2})_{\#}\boldsymbol{\mu} = (\pi_{st}^{1\to 2})_{\#}\boldsymbol{\mu}.$$

Inverting μ_0 with μ_1 we conclude. □

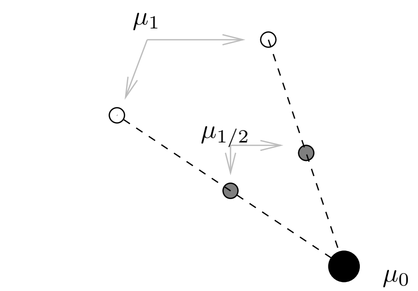


Figure 7.1: An example of geodesic: the mass of μ^0 splits into two parts

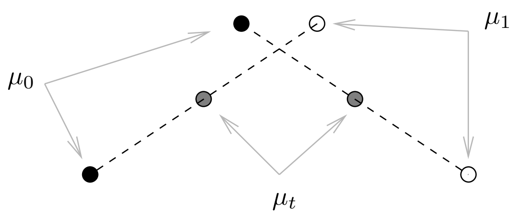


Figure 7.2: Another example of geodesic: the trajectories may intersect

In the case $X = \mathbb{R}$, using the explicit representation (6.0.3) for the Wasserstein distance in terms of the inverses of distribution functions, we get

$$F^{-1}_{\mu_t^{1\to 2}} = (1-t)F^{-1}_{\mu^1} + tF^{-1}_{\mu^2} \quad \mathscr{L}^1\text{-a.e. in } (0,1). \tag{7.2.8}$$

for any geodesic $\mu_t^{1\to 2}$ induced by $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$.

7.3 The curvature properties of $\mathscr{P}_2(X)$

In this section we consider the particular case $p = 2$ and we establish some finer geometric properties of $\mathscr{P}_2(X)$.

In particular we will prove in Theorem 7.3.2 the *semiconcavity inequality* of the Wasserstein distance from a fixed measure μ^3 along the constant speed geodesics $\mu_t^{1\to 2}$ connecting μ^1 to μ^2:

$$W_2^2(\mu_t^{1\to 2}, \mu^3) \geq (1-t)W_2^2(\mu^1, \mu^3) + tW_2^2(\mu^2, \mu^3) - t(1-t)W_2^2(\mu^1, \mu^2). \tag{7.3.1}$$

According to Aleksandrov's metric notion of curvature (see [5] and Section 12.3 in the Appendix), this inequality can be interpreted by saying that the Wasserstein space is a positively curved metric space (in short, a *PC*-space). This was already pointed out by a formal computation in [107], showing also that generically the inequality is strict (see Example 7.3.3). See also Section 12.3 in the Appendix, where we recall some basic facts of the theory of positively curved metric spaces.

For $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3) \subset \mathscr{P}_2(X^3)$ and $i, j, k \in \{1, 2, 3\}$, $t \in [0, 1]$ we set

$$W_{\boldsymbol{\mu}}^2(\mu_t^{i\to j}, \mu^k) := \int_{X^3} |(1-t)x_i + tx_j - x_k|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3). \tag{7.3.2}$$

By (7.1.6) we get

$$W_2^2(\mu_t^{i\to j}, \mu^k) \le W_{\boldsymbol{\mu}}^2(\mu_t^{i\to j}, \mu^k). \tag{7.3.3}$$

Moreover, the Hilbertian identity

$$|(1-t)a + tb - c|^2 = (1-t)|a-c|^2 + t|b-c|^2 - t(1-t)|b-a|^2$$

gives

$$W_{\boldsymbol{\mu}}^2(\mu_t^{1\to 2}, \mu^3) = (1-t)W_{\boldsymbol{\mu}}^2(\mu^1, \mu^3) + tW_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) - t(1-t)W_{\boldsymbol{\mu}}^2(\mu^1, \mu^2), \tag{7.3.4}$$

and the related differential identities

$$\frac{d}{dt}W_{\boldsymbol{\mu}}^2(\mu_t^{1\to 2}, \mu^3) = W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) - W_{\boldsymbol{\mu}}^2(\mu^1, \mu^3) + (2t-1)W_{\boldsymbol{\mu}}^2(\mu^1, \mu^2) \tag{7.3.5}$$

$$= \frac{1}{1-t}\Big(W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) - W_{\boldsymbol{\mu}}^2(\mu_t^{1\to 2}, \mu^2) - W_{\boldsymbol{\mu}}^2(\mu_t^{1\to 2}, \mu^3)\Big) \tag{7.3.6}$$

$$= \frac{1}{t}\Big(W_{\boldsymbol{\mu}}^2(\mu_t^{1\to 2}, \mu^1) + W_{\boldsymbol{\mu}}^2(\mu_t^{1\to 2}, \mu^3) - W_{\boldsymbol{\mu}}^2(\mu^1, \mu^3)\Big). \tag{7.3.7}$$

Proposition 7.3.1. *Let $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2)$, $t \in (0, 1)$ and $\boldsymbol{\mu}^{t\,3} \in \Gamma_o(\mu_t^{1\to 2}, \mu^3)$. Then there exists a plan*

$$\boldsymbol{\mu}_t \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3) \quad \text{such that} \quad (\pi_t^{1\to 2,3})_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{t\,3}, \tag{7.3.8}$$

and this plan is unique if $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$. For each plan $\boldsymbol{\mu}_t$ satisfying (7.3.8) *we have*

$$W_2^2(\mu_t^{1\to 2}, \mu^3) = (1-t)W_{\boldsymbol{\mu}_t}^2(\mu^1, \mu^3) + tW_{\boldsymbol{\mu}_t}^2(\mu^2, \mu^3) - t(1-t)W_{\boldsymbol{\mu}_t}^2(\mu^1, \mu^2). \tag{7.3.9}$$

Proof. Let $\Sigma_t : X^2 \to X^2$ and $\Lambda_t : X^3 \to X^3$ be the homeomorphisms defined by

$$\Sigma_t(x_1, x_2) := ((1-t)x_1 + tx_2, x_2), \quad \Lambda_t(x_1, x_2, x_3) = ((1-t)x_1 + tx_2, x_2, x_3)$$

and notice that $\boldsymbol{\mu}$ has the required properties if and only if $\boldsymbol{\nu} := \Lambda_{t\#}\boldsymbol{\mu}$ satisfies

$$\pi_\#^{1,2}\boldsymbol{\nu} = \Sigma_{t\#}\boldsymbol{\mu}^{1\,2}, \quad \pi_\#^{1,3}\boldsymbol{\nu} = \boldsymbol{\mu}^{t\,3}. \tag{7.3.10}$$

Then, Lemma 5.3.2 says that there exists a plan $\boldsymbol{\nu}$ fulfilling (7.3.10) and, since Λ_t is invertible, this proves the existence of $\boldsymbol{\mu}$. When $\boldsymbol{\mu}^{1\,2}$ is optimal, since $\Sigma_{t\#}\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu_t^{1\to 2}, \mu^2)$, we infer from Lemma 7.2.1 that $\Sigma_{t\#}\boldsymbol{\mu}^{1\,2}$ is unique and induced by a transport map and therefore $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ are uniquely determined. □

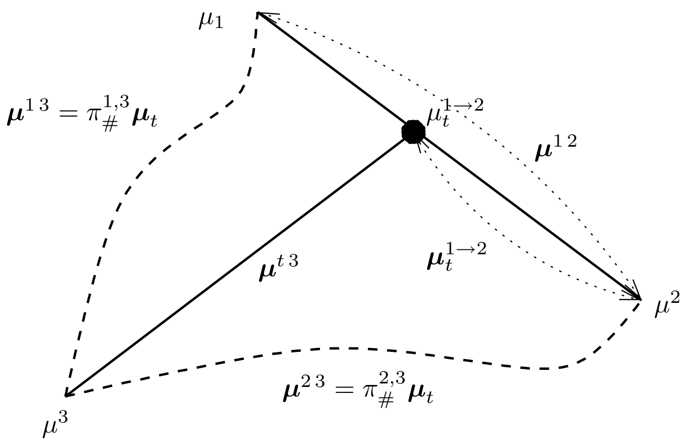


Figure 7.3: $\boldsymbol{\mu}^{1\,2}$ and $\boldsymbol{\mu}^{t\,3}$ are given optimal plans; $\boldsymbol{\mu}^{2\,3}$ and $\boldsymbol{\mu}^{1\,3}$ are not optimal, in general

Theorem 7.3.2 ($\mathscr{P}_2(X)$ is a *PC*-space). *For each choice of μ^1, μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1,\mu^2)$ we have*

$$W_2^2(\mu_t^{1\to 2},\mu^3) \geq (1-t)W_2^2(\mu^1,\mu^3) + tW_2^2(\mu^2,\mu^3) - t(1-t)W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2) \quad (7.3.11)$$

and the map $t \mapsto W_2^2(\mu_t^{1\to 2},\mu^3) - t^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2)$ is concave in $[0,1]$. In particular, choosing $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1,\mu^2)$ (see Figure 7.3) we have

$$W_2^2(\mu_t^{1\to 2},\mu^3) \geq (1-t)W_2^2(\mu^1,\mu^3) + tW_2^2(\mu^2,\mu^3) - t(1-t)W_2^2(\mu^1,\mu^2) \quad (7.3.12)$$

and therefore $\mathscr{P}_2(X)$ is a PC-space.

Proof. (7.3.11) is a direct consequence of (7.3.9) and (7.3.3). In order to prove the concavity property we choose λ, t_1, $t_2 \in [0,1]$, $t := (1-\lambda)t_1 + \lambda t_2$, and we have only to develop the obvious calculations:

$$\begin{aligned}
&W_2^2(\mu_t^{1\to 2},\mu^3) - t^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2) = W_2^2(\mu_\lambda^{t_1\to t_2},\mu^3) - t^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2)\\
&\geq (1-\lambda)W_2^2(\mu^{t_1},\mu^3) + \lambda W_2^2(\mu^{t_2},\mu^3) - \Big(\lambda(1-\lambda)(t_2-t_1)^2 + t^2\Big)W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2)\\
&= (1-\lambda)\big[W_2^2(\mu_{t_1}^{1\to 2},\mu^3) - t_1^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2)\big] + \lambda\big[W_2^2(\mu_{t_2}^{1\to 2},\mu^3) - t_2^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2)\big]\,.
\end{aligned}$$

In the case $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1,\mu^2)$ is sufficient to note that $W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2) = W_2^2(\mu^1,\mu^2)$. □

Example 7.3.3 (Strict positivity of the sectional curvature). The following example shows that in general the inequality (7.3.1) is strict. Let

$$\mu^1 := \frac{1}{2}\left(\delta_{(1,1)} + \delta_{(5,3)}\right), \quad \mu^2 := \frac{1}{2}\left(\delta_{(-1,1)} + \delta_{(-5,3)}\right), \quad \mu^3 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(0,-4)}\right).$$

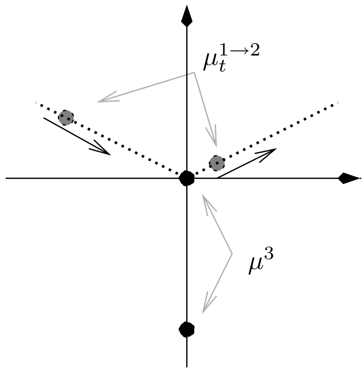


Figure 7.4: μ^3 is the sum of deltas on black dots, $\mu_t^{1\to 2}$ is moving along the dotted lines

Then, it is immediate to check that $W_2^2(\mu^1,\mu^2)=40$, $W_2^2(\mu^1,\mu^3)=30$, and $W_2^2(\mu^2,\mu^3)=30$. On the other hand, the unique constant speed geodesic joining μ^1 to μ^2 is given by

$$\mu_t := \frac{1}{2}\left(\delta_{(1-6t,1+2t)} + \delta_{(5-6t,3-2t)}\right)$$

and a simple computation gives

$$24 = W_2^2(\mu_{1/2},\mu^3) > \frac{30}{2} + \frac{30}{2} - \frac{40}{4}.$$

Formula (7.3.11) is useful to evaluate the directional derivative of the Wasserstein distance. If $\boldsymbol{\mu}^{1\,2}\in\Gamma(\mu^1,\mu^2)$, general properties of concave maps ensures that for each point $t\in[0,1)$ there exists the right derivative

$$\frac{d}{dt+}W_2^2(\mu_t^{1\to 2},\mu^3) := \lim_{t'\downarrow t}\frac{W_2^2(\mu_{t'}^{1\to 2},\mu^3) - W_2^2(\mu_t^{1\to 2},\mu^3)}{t'-t}$$

and, for $t\in(0,1]$, the left derivative

$$\frac{d}{dt-}W_2^2(\mu_t^{1\to 2},\mu^3) := \lim_{t'\uparrow t}\frac{W_2^2(\mu_t^{1\to 2},\mu^3) - W_2^2(\mu_{t'}^{1\to 2},\mu^3)}{t-t'}$$

satisfying

$$\frac{d}{dt+}W_2^2(\mu_t^{1\to 2},\mu^3) \le \frac{d}{dt-}W_2^2(\mu_t^{1\to 2},\mu^3) \quad \forall\, t\in(0,1)$$

and, for a (at most) countable subset $\mathscr{N}\subset(0,1)$

$$\frac{d}{dt+}W_2^2(\mu_t^{1\to 2},\mu^3) = \frac{d}{dt-}W_2^2(\mu_t^{1\to 2},\mu^3) \quad \forall\, t\in(0,1)\setminus\mathscr{N}. \tag{7.3.13}$$

Corollary 7.3.4. *Let* μ^1, μ^2, $\mu^3 \in \mathscr{P}_2(X)$, $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1,\mu^2)$, $t \in [0,1]$, *and*

$\boldsymbol{\mu}_t \in \Gamma(\boldsymbol{\mu}^{1\,2},\mu^3)$ *such that* $(\pi_t^{1\to 2,3})_\#\boldsymbol{\mu} \in \Gamma_o(\mu_t^{1\to 2},\mu^3)$ *as in Proposition* 7.3.1.

Then

$$\begin{aligned}\frac{d}{dt_+}W_2^2(\mu_t^{1\to 2},\mu^3) &\le W^2_{\boldsymbol{\mu}_t}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}_t}(\mu^1,\mu^3) + (2t-1)W^2_{\boldsymbol{\mu}_t}(\mu^1,\mu^2)\\ &= \frac{1}{1-t}\Big(W^2_{\boldsymbol{\mu}_t}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}_t}(\mu_t^{1\to 2},\mu^2) - W_2^2(\mu_t^{1\to 2},\mu^3)\Big)\\ &= \frac{1}{t}\Big(W^2_{\boldsymbol{\mu}_t}(\mu_t^{1\to 2},\mu^1) + W_2^2(\mu_t^{1\to 2},\mu^3) - W^2_{\boldsymbol{\mu}_t}(\mu^1,\mu^3)\Big)\\ &\le \frac{d}{dt_-}W_2^2(\mu_t^{1\to 2},\mu^3).\end{aligned} \tag{7.3.14}$$

In particular, equality holds in the previous formula whenever t *belongs to the set of differentiability of the distance, i.e.* $t \in (0,1)\setminus\mathscr{N}$.

Proof. We simply observe that

$$W_2^2(\mu_{t'}^{1\to 2},\mu^3) \le W^2_{\boldsymbol{\mu}_t}(\mu_{t'}^{1\to 2},\mu^3) \quad \text{if } t'\neq t, \qquad W_2^2(\mu_t^{1\to 2},\mu^3) = W^2_{\boldsymbol{\mu}_t}(\mu_t^{1\to 2},\mu^3),$$

and we apply (7.3.9) and (7.3.5), (7.3.6), (7.3.7) to evaluate the right and left derivatives. □

We conclude this section by a precise characterization of the right derivative (7.3.14) at time $t=0$; we need to introduce some more definitions.

Definition 7.3.5 (A new class of multiple plans). *Let* $\boldsymbol{\mu}^{1\,2} \in \mathscr{P}_2(X^2)$ *and* $\mu^3 \in \mathscr{P}_2(X)$. *We say that* $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\mu}^{1\,2},\mu^3)$ *belongs to* $\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)$ *if* $\pi^{1,3}_\#\boldsymbol{\mu} \in \Gamma_o(\mu^1,\mu^3)$.

Proposition 7.3.6. *Let* $\boldsymbol{\mu}^{1\,2}\in\Gamma(\mu^1,\mu^2)$, $\mu^3\in\mathscr{P}_2(X)$. *Then for every* $\boldsymbol{\mu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)$ *such that*

$$\int_{X^3}|x_2-x_3|^2\,d\boldsymbol{\mu} = \min\Big\{\int_{X^3}|x_2-x_3|^2\,d\boldsymbol{\nu} : \boldsymbol{\nu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)\Big\} \tag{7.3.15}$$

we have

$$\begin{aligned}\frac{d}{dt_+}W_2^2(\mu_t^{1\to 2},\mu^3)_{|t=0} &= \Big(W^2_{\boldsymbol{\mu}}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}}(\mu^1,\mu^2) - W_2^2(\mu^1,\mu^3)\Big)\\ &= -2\int_{X^3}\langle x_2-x_1, x_3-x_1\rangle\,d\boldsymbol{\mu}.\end{aligned} \tag{7.3.16}$$

Proof. We already know by (7.3.14) that

$$\frac{d}{dt_+}W_2^2(\mu_t^{1\to 2},\mu^3)_{|t=0} \le \Big(W^2_{\boldsymbol{\mu}}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}}(\mu^1,\mu^2) - W_2^2(\mu^1,\mu^3)\Big)$$

so that we simply have to prove the opposite inequality. Let $\mathscr{N}$ be the negligible set defined by (7.3.13); thanks to (7.3.14) and to the semiconcavity of the squared distance map, we have

$$\begin{aligned}\frac{d}{dt+}W_2^2(\mu_t^{1\to 2},\mu^3) &= \lim_{t\downarrow 0, t\notin\mathscr{N}} \frac{d}{dt+}W_2^2(\mu_t^{1\to 2},\mu^3)\\ &= \lim_{t\downarrow 0, t\notin\mathscr{N}} \frac{1}{1-t}\Big(W^2_{\boldsymbol{\mu}_t}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}^{1\,2}}(\mu_t^{1\to 2},\mu^2) - W_2^2(\mu_t^{1\to 2},\mu^3)\Big)\\ &\geq \Big(W^2_{\boldsymbol{\mu}_0}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2) - W_2^2(\mu^1,\mu^3)\Big),\end{aligned}$$

where $\boldsymbol{\mu}_0$ is any narrow accumulation point of $\boldsymbol{\mu}_t$ as $t\downarrow 0$. By Proposition 7.1.3 $\pi^{1\,2}_\#\boldsymbol{\mu}_0 = \boldsymbol{\mu}^{1\,2}$, $\pi^{1\,3}_\#\boldsymbol{\mu}_0 \in \Gamma_o(\mu^1,\mu^3)$. Invoking (7.3.14) again, we conclude. □

Since the integrals of $|x_1-x_2|^2$ and of $|x_1-x_3|^2$ do not depend on the choice of $\boldsymbol{\nu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)$, we can reformulate (7.3.16) as

$$\frac{d}{dt_+}W_2^2(\mu_t^{1\to 2},\mu^3)_{|t=0} = \min_{\boldsymbol{\nu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)} -2\int_{X^3}\langle x_2-x_1, x_3-x_1\rangle\, d\boldsymbol{\nu}. \tag{7.3.17}$$

Chapter 8

Absolutely Continuous Curves in $\mathscr{P}_p(X)$ and the Continuity Equation

In this chapter we endow $\mathscr{P}_p(X)$, when X is a separable Hilbert space, with a kind of differential structure, consistent with the metric structure introduced in the previous chapter. Our starting point is the analysis of absolutely continuous curves $\mu_t : (a,b) \to \mathscr{P}_p(X)$ and of their metric derivative $|\mu'|(t)$: recall that these concepts depend only on the metric structure of $\mathscr{P}_p(X)$, by Definition 1.1.1 and (1.1.3). We show in Theorem 8.3.1 that for $p > 1$ this class of curves coincides with (distributional, in the duality with smooth cylindrical test functions) solutions of the continuity equation

$$\frac{\partial}{\partial t}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \qquad \text{in } X \times (a,b).$$

More precisely, given an absolutely continuous curve μ_t, one can find a Borel time-dependent velocity field $v_t : X \to X$ such that $\|v_t\|_{L^p(\mu_t)} \le |\mu'|(t)$ for $\mathscr{L}^1$-a.e. $t \in (a,b)$ and the continuity equation holds. Conversely, if μ_t solve the continuity equation for some Borel velocity field v_t with $\int_a^b \|v_t\|_{L^p(\mu_t)}\,dt < +\infty$, then μ_t is an absolutely continuous curve and $\|v_t\|_{L^p(\mu_t)} \ge |\mu'|(t)$ for $\mathscr{L}^1$-a.e. $t \in (a,b)$.

As a consequence of Theorem 8.3.1 we see that among all velocity fields v_t which produce the same flow μ_t, there is a unique optimal one with smallest $L^p(\mu_t; X)$-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, one can show that the minimality of the L^p norm of v_t is characterized by the property

$$v_t \in \overline{\{j_q(\nabla\varphi) : \ \varphi \in \mathrm{Cyl}(X)\}}^{L^p(\mu_t;X)} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b), \tag{8.0.1}$$

where q is the conjugate exponent of p and $j_q : L^q(\mu; X) \to L^p(\mu; X)$ is the duality map, i.e. $j_q(v) = |v|^{q-2}v$ (here gradients are thought as covectors, and therefore as elements of L^q).

The characterization (8.0.1) of tangent vectors strongly suggests, in the case $p = 2$, to consider the following tangent to $\mathscr{P}_2(X)$

$$\mathrm{Tan}_\mu \mathscr{P}_2(X) := \overline{\{\nabla\varphi :\ \varphi \in \mathrm{Cyl}(X)\}}^{L^2(\mu;X)} \qquad \forall \mu \in \mathscr{P}_2(X), \tag{8.0.2}$$

endowed with the natural L^2 metric. Moreover, as a consequence of the characterization of absolutely continuous curves in $\mathscr{P}_2(X)$, we recover the Benamou–Brenier (see [21], where the formula was introduced for numerical purposes) formula for the Wasserstein distance:

$$W_2^2(\mu_0, \mu_1) = \min \left\{ \int_0^1 \|v_t\|^2_{L^2(\mu_t;X)}\, dt :\ \frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0 \right\}. \tag{8.0.3}$$

Indeed, for any admissible curve we use the inequality between L^2 norm of v_t and metric derivative to obtain:

$$\int_0^1 \|v_t\|^2_{L^2(\mu_t;X)}\, dt \geq \int_0^1 |\mu'|^2(t)\, dt \geq W_2^2(\mu_0, \mu_1).$$

Conversely, since we know that $\mathscr{P}_2(X)$ is a length space, we can use a geodesic μ_t and its tangent vector field v_t to obtain equality in (8.0.3). Similar arguments work in the case $p > 1$ as well, with the only drawback that a priori the L^p closure of $j_q(\nabla\varphi)$ is not a vector space in general, so we are able only to define a tangent cone. We also show that optimal transport maps belong to $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ under quite general conditions.

In this way we recover in a more general framework the *Riemannian interpretation* of the Wasserstein distance developed by Otto in [107] (see also [106], [83]) and used to study the long time behaviour of the porous medium equation. In the original paper [107], (8.0.3) is derived in the case $X = \mathbb{R}^d$ using formally the concept of Riemannian submersion and the family of maps $\phi \mapsto \phi_\#\mu$ (indexed by $\mu \ll \mathscr{L}^d$) from Arnold's space of diffeomorphisms into the Wasserstein space. In Otto's formalism tangent vectors are rather thought as $\boldsymbol{s} = \frac{d}{dt}\mu_t$ and these vectors are identified, via the continuity equation, with $-D\cdot(v_{\boldsymbol{s}}\mu_t)$. Moreover $v_{\boldsymbol{s}}$ is chosen to be the gradient of a function $\psi_{\boldsymbol{s}}$, so that $D\cdot(\nabla\psi_{\boldsymbol{s}}\mu_t) = -\boldsymbol{s}$. Then the metric tensor is induced by the identification $\boldsymbol{s} \mapsto \nabla\phi_{\boldsymbol{s}}$ as follows:

$$\langle \boldsymbol{s}, \boldsymbol{s}' \rangle_{\mu_t} := \int_{\mathbb{R}^d} \langle \nabla\psi_{\boldsymbol{s}}, \nabla\psi_{\boldsymbol{s}'} \rangle\, d\mu_t.$$

As noticed in [107], both the identification between tangent vectors and gradients and the scalar product depend on μ_t, and these facts lead to a non trivial geometry of the Wasserstein space. We prefer instead to consider directly v_t as the tangent vectors, allowing them to be not necessarily gradients: this leads to (8.0.2).

Another 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

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

Moreover the optimal transport plans between μ_t and μ_{t+h}, rescaled in a suitable way, converge to the 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 regularity assumptions on the initial measure μ). As a consequence, we will obtain in Theorem 8.4.7 a formula for the derivative of the map $t\mapsto W_p^p(\mu_t,\nu)$.

8.1 The continuity equation in $\mathbb{R}^d$

In this section we collect some results on the continuity equation

$$\partial_t\mu_t + \nabla\cdot(v_t\mu_t) = 0 \qquad \text{in } \mathbb{R}^d\times(0,T), \tag{8.1.1}$$

which we will need in the sequel. Here μ_t is a Borel family of probability measures on $\mathbb{R}^d$ defined for t in the open interval $I := (0,T)$, $v : (x,t)\mapsto v_t(x)\in\mathbb{R}^d$ is a Borel velocity field such that

$$\int_0^T\int_{\mathbb{R}^d} |v_t(x)|\,d\mu_t(x)\,dt < +\infty, \tag{8.1.2}$$

and we suppose that (8.1.1) holds in the sense of distributions, i.e.

$$\begin{gathered}\int_0^T\int_{\mathbb{R}^d}\Big(\partial_t\varphi(x,t) + \langle v_t(x),\nabla_x\varphi(x,t)\rangle\Big)\,d\mu_t(x)\,dt = 0,\\ \forall\,\varphi\in C_c^\infty(\mathbb{R}^d\times(0,T)).\end{gathered} \tag{8.1.3}$$

Remark 8.1.1 (More general test functions). By a simple regularization argument via convolution, it is easy to show that (8.1.3) holds if $\varphi\in C_c^1\big(\mathbb{R}^d\times(0,T)\big)$ as well. Moreover, under condition (8.1.2), we can also consider bounded test functions φ, with bounded gradient, whose support has a compact projection in $(0,T)$ (that is, the support in x need not be compact): it suffices to approximate φ by $\varphi\chi_R$ where $\chi_R\in C_c^\infty(\mathbb{R}^d)$, $0\le\chi_R\le 1$, $|\nabla\chi_R|\le 2$ and $\chi_R = 1$ on $B_R(0)$. This more general choice of the test functions is consistent with the infinite-dimensional case, where cylindrical test functions will be considered, see Definition 5.1.11 and (8.3.8).

First of all we recall some (technical) preliminaries.

Lemma 8.1.2 (Continuous representative). *Let μ_t be a Borel family of probability measures satisfying* (8.1.3) *for a Borel vector field v_t satisfying* (8.1.2). *Then there exists a* narrowly continuous *curve $t \in [0,T] \mapsto \tilde\mu_t \in \mathscr{P}(\mathbb{R}^d)$ such that $\mu_t = \tilde\mu_t$ for $\mathscr{L}^1$-a.e. $t \in (0,T)$. Moreover, if $\varphi \in C^1_c(\mathbb{R}^d \times [0,T])$ and $t_1 \le t_2 \in [0,T]$ we have*

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(x,t_2)\,d\tilde\mu_{t_2}(x) - \int_{\mathbb{R}^d} \varphi(x,t_1)\,d\tilde\mu_{t_1}(x)\\ = \int_{t_1}^{t_2}\int_{\mathbb{R}^d} \Big(\partial_t\varphi + \langle \nabla\varphi, v_t\rangle\Big)\,d\mu_t(x)\,dt.\end{aligned} \tag{8.1.4}$$

Proof. Let us take $\varphi(x,t) = \eta(t)\zeta(x)$, $\eta \in C^\infty_c(0,T)$ and $\zeta \in C^\infty_c(\mathbb{R}^d)$; we have

$$-\int_0^T \eta'(t)\Big(\int_{\mathbb{R}^d} \zeta(x)\,d\mu_t(x)\Big)\,dt = \int_0^T \eta(t)\Big(\int_{\mathbb{R}^d} \langle\nabla\zeta(x), v_t(x)\rangle\,d\mu_t(x)\Big)\,dt,$$

so that the map

$$t \mapsto \mu_t(\zeta) = \int_{\mathbb{R}^d} \zeta(x)\,d\mu_t(x)$$

belongs to $W^{1,1}(0,T)$ with distributional derivative

$$\dot\mu_t(\zeta) = \int_{\mathbb{R}^d} \langle\nabla\zeta(x), v_t(x)\rangle\,d\mu_t(x) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,T) \tag{8.1.5}$$

with

$$|\dot\mu_t(\zeta)| \le V(t)\sup_{\mathbb{R}^d}|\nabla\zeta|, \quad V(t) := \int_{\mathbb{R}^d} |v_t(x)|\,d\mu_t(x), \quad V \in L^1(0,T). \tag{8.1.6}$$

If L_ζ is the set of its Lebesgue points, we know that $\mathscr{L}^1((0,T)\setminus L_\zeta) = 0$. Let us now take a countable set Z which is dense in $C^1_c(\mathbb{R}^d)$ with respect the usual C^1 norm $\|\zeta\|_{C^1} = \sup_{\mathbb{R}^d}(|\zeta|, |\nabla\zeta|)$ and let us set $L_Z := \cap_{\zeta\in Z} L_\zeta$. The restriction of the curve μ to L_Z provides a uniformly continuous family of bounded functionals on $C^1_c(\mathbb{R}^d)$, since (8.1.6) shows

$$|\mu_t(\zeta) - \mu_s(\zeta)| \le \|\zeta\|_{C^1}\int_s^t V(\lambda)\,d\lambda \quad \forall\, s,t \in L_Z.$$

Therefore, it can be extended in a unique way to a continuous curve $\{\tilde\mu_t\}_{t\in[0,T]}$ in $[C^1_c(\mathbb{R}^d)]'$. If we show that $\{\mu_t\}_{t\in L_Z}$ is also tight, the extension provides a continuous curve in $\mathscr{P}(\mathbb{R}^d)$.

For, let us consider nonnegative, smooth functions $\zeta_k : \mathbb{R}^d \to [0,1]$, $k \in \mathbb{N}$, such that

$$\zeta_k(x) = 1 \quad \text{if } |x| \le k, \quad \zeta_k(x) = 0 \quad \text{if } |x| \ge k+1, \quad |\nabla\zeta_k(x)| \le 2.$$

It is not restrictive to suppose that $\zeta_k \in Z$. Applying the previous formula (8.1.5), for $t,\, s \in L_Z$ we have

$$|\mu_t(\zeta_k) - \mu_s(\zeta_k)| \le a_k := 2\int_0^T \int_{k<|x|<k+1} |v_\lambda(x)|\, d\mu_\lambda(x)\, d\lambda,$$

with $\sum_{k=1}^{+\infty} a_k < +\infty$. For a fixed $s \in L_Z$ and $\varepsilon > 0$, being μ_s tight, we can find $k \in \mathbb{N}$ such that $\mu_s(\zeta_k) > 1 - \varepsilon/2$ and $a_k < \varepsilon/2$. It follows that

$$\mu_t(\overline{B_{k+1}(0)}) \ge \mu_t(\zeta_k) \ge 1 - \varepsilon \quad \forall t \in L_Z.$$

Now we show (8.1.4). Let us choose $\varphi \in C^1_c(\mathbb{R}^d \times [0,T])$ and set $\varphi_\varepsilon(x,t) = \eta_\varepsilon(t)\varphi(x,t)$, where $\eta_\varepsilon \in C^\infty_c(t_1,t_2)$ such that

$$0 \le \eta_\varepsilon(t) \le 1, \quad \lim_{\varepsilon\downarrow 0} \eta_\varepsilon(t) = \chi_{(t_1,t_2)}(t)\ \forall t \in [0,T], \quad \lim_{\varepsilon\downarrow 0} \eta'_\varepsilon = \delta_{t_1} - \delta_{t_2}$$

in the duality with continuous functions in $[0,T]$. We get

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t(\eta_\varepsilon\varphi) + \langle \nabla_x(\eta_\varepsilon\varphi), v_t\rangle\Big)\, d\mu_t(x)\, dt \\ &= \int_0^T \eta_\varepsilon(t) \int_{\mathbb{R}^d} \Big(\partial_t\varphi(x,t) + \langle v_t(x), \nabla_x\varphi(x,t)\rangle\Big)\, d\mu_t(x)\, dt \\ &\quad + \int_0^T \eta'_\varepsilon(t) \int_{\mathbb{R}^d} \varphi(x,t)\, d\tilde\mu_t(x)\, dt. \end{aligned}$$

Passing to the limit as ε vanishes and invoking the continuity of $\tilde\mu_t$, we get (8.1.4). □

Lemma 8.1.3 (Time rescaling). *Let* $\mathsf{t} : s \in [0,T'] \to \mathsf{t}(s) \in [0,T]$ *be a strictly increasing absolutely continuous map with absolutely continuous inverse* $\mathsf{s} := \mathsf{t}^{-1}$. *Then* (μ_t, v_t) *is a distributional solution of* (8.1.1) *if and only if*

$$\hat\mu := \mu\circ\mathsf{t},\ \hat v := \mathsf{t}' v\circ\mathsf{t}, \qquad \textit{is a distributional solution of } (8.1.1) \textit{ on } (0,T').$$

Proof. By an elementary smoothing argument we can assume that s is continuously differentiable and $\mathsf{s}' > 0$. We choose $\hat\varphi \in C^1_c(\mathbb{R}^d \times (0,T'))$ and let us set $\varphi(x,t) := \hat\varphi(x,\mathsf{s}(t))$; since $\varphi \in C^1_c(\mathbb{R}^d \times (0,T))$ we have

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} \big(\mathsf{s}'(t)\partial_s\hat\varphi(x,\mathsf{s}(t)) + \langle \nabla\hat\varphi(x,\mathsf{s}(t)), \hat v_t(x)\rangle\big)\, d\mu_t(x)\, dt \\ &= \int_0^T \mathsf{s}'(t) \int_{\mathbb{R}^d} \Big(\partial_s\hat\varphi(x,\mathsf{s}(t)) + \langle \nabla_x\hat\varphi(x,\mathsf{s}(t)), \frac{v_t(x)}{\mathsf{s}'(t)}\rangle\Big)\, d\mu_t(x)\, dt \\ &= \int_0^{T'} \int_{\mathbb{R}^d} \Big(\partial_s\hat\varphi(x,s) + \langle \nabla_x\hat\varphi(x,s), \mathsf{t}'(s) v_{\mathsf{t}(s)}(x)\rangle\Big)\, d\hat\mu_s(x)\, ds. \end{aligned}$$

□

When the velocity field v_t is more regular, the classical method of characteristics provides an explicit solution of (8.1.1).

First we recall an elementary result of the theory of ordinary differential equations.

Lemma 8.1.4 (The characteristic system of ODE). *Let v_t be a Borel vector field such that for every compact set $B \subset \mathbb{R}^d$*

$$\int_0^T \Big(\sup_B |v_t| + \mathrm{Lip}(v_t, B)\Big)\, dt < +\infty. \tag{8.1.7}$$

Then, for every $x \in \mathbb{R}^d$ and $s \in [0,T]$ the ODE

$$X_s(x,s) = x, \quad \frac{d}{dt} X_t(x,s) = v_t(X_t(x,s)), \tag{8.1.8}$$

admits a unique maximal solution defined in an interval $I(x,s)$ relatively open in $[0,T]$ and containing s as (relatively) internal point.
Furthermore, if $t \mapsto |X_t(x,s)|$ is bounded in the interior of $I(x,s)$ then $I(x,s) = [0,T]$; finally, if v satisfies the global bounds analogous to (8.1.7)

$$S := \int_0^T \Big(\sup_{\mathbb{R}^d} |v_t| + \mathrm{Lip}(v_t, \mathbb{R}^d)\Big)\, dt < +\infty, \tag{8.1.9}$$

then the flow map X satisfies

$$\int_0^T \sup_{x\in\mathbb{R}^d} |\partial_t X_t(x,s)|\, dt \le S, \qquad \sup_{t,s\in[0,T]} \mathrm{Lip}(X_t(\cdot,s), \mathbb{R}^d) \le e^S. \tag{8.1.10}$$

For simplicity, we set $X_t(x) := X_t(x,0)$ in the particular case $s = 0$ and we denote by $\tau(x) := \sup I(x,0)$ the length of the maximal time domain of the characteristics leaving from x at $t = 0$.

Remark 8.1.5 (The characteristics method for backward first order linear PDE's). Characteristics provide a useful representation formula for classical solutions of the backward equation (formally adjoint to (8.1.1))

$$\partial_t\varphi + \langle v_t, \nabla\varphi\rangle = \psi \quad \text{in } \mathbb{R}^d \times (0,T), \quad \varphi(x,T) = \varphi_T(x) \quad x \in \mathbb{R}^d, \tag{8.1.11}$$

when, e.g., $\psi \in C^1_b(\mathbb{R}^d \times (0,T)), \varphi_T \in C^1_b(\mathbb{R}^d)$ and v satisfies the global bounds (8.1.9), so that maximal solutions are always defined in $[0,T]$. A direct calculation shows that

$$\varphi(x,t) := \varphi_T(X_T(x,t)) - \int_t^T \psi(X_s(x,t), s)\, ds \tag{8.1.12}$$

solve (8.1.11). For $X_s(X_t(x,0),t) = X_s(x,0)$ yields

$$\varphi(X_t(x,0),t) = \varphi_T(X_T(x,0)) - \int_t^T \psi(X_s(x,0), s)\, ds,$$

and differentiating both sides with respect to t we obtain

$$\left[\frac{\partial\varphi}{\partial t} + \langle v_t, \nabla\varphi\rangle\right](X_t(x,0),t) = \psi(X_t(x,0),t).$$

Since x (and then $X_t(x,0)$) is arbitrary we conclude that (8.1.18) is fulfilled.

Now we use characteristics to prove the existence, the uniqueness, and a representation formula of the solution of the continuity equation, under suitable assumption on v.

Lemma 8.1.6. *Let v_t be a Borel velocity field satisfying* (8.1.7), (8.1.2), *let $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$, and let X_t be the maximal solution of the ODE* (8.1.8) *(corresponding to $s=0$). Suppose that for some $\bar t \in (0,T]$*

$$\tau(x) > \bar t \quad \textit{for } \mu_0\textit{-a.e. } x \in \mathbb{R}^d. \tag{8.1.13}$$

Then $t \mapsto \mu_t := (X_t)_\#\mu_0$ is a continuous solution of (8.1.1) *in $[0,\bar t]$.*

Proof. The continuity of μ_t follows easily since $\lim_{s\to t} X_s(x) = X_t(x)$ for μ_0-a.e. $x \in \mathbb{R}^d$: thus for every continuous and bounded function $\zeta : \mathbb{R}^d \to \mathbb{R}$ the dominated convergence theorem yields

$$\lim_{s\to t}\int_{\mathbb{R}^d}\zeta\,d\mu_s = \lim_{s\to t}\int_{\mathbb{R}^d}\zeta(X_s(x))\,d\mu_0(x) = \int_{\mathbb{R}^d}\zeta(X_t(x))\,d\mu_0(x) = \int_{\mathbb{R}^d}\zeta\,d\mu_t.$$

For any $\varphi \in C_c^\infty(\mathbb{R}^d \times (0,\bar t))$ and for μ_0-a.e. $x \in \mathbb{R}^d$ the maps $t \mapsto \phi_t(x) := \varphi(X_t(x),t)$ are absolutely continuous in $(0,\bar t)$, with

$$\dot\phi_t(x) = \partial_t\varphi(X_t(x),t) + \langle\nabla\varphi(X_t(x),t), v_t(X_t(x))\rangle = \Lambda(\cdot,t)\circ X_t,$$

where $\Lambda(x,t) := \partial_t\varphi(x,t) + \langle\nabla\varphi(x,t), v_t(x)\rangle$. We thus have

$$\begin{aligned}\int_0^T\int_{\mathbb{R}^d}|\dot\phi_t(x)|\,d\mu_0(x)\,dt &= \int_0^T\int_{\mathbb{R}^d}|\Lambda(X_t(x),t)|\,d\mu_0(x)\,dt\\ &= \int_0^T\int_{\mathbb{R}^d}|\Lambda(x,t)|\,d\mu_t(x)\,dt\\ &\le \mathrm{Lip}(\varphi)\Big(T + \int_0^T\int_{\mathbb{R}^d}|v_t(x)|\,d\mu_t(x)\,dt\Big) < +\infty\end{aligned}$$

and therefore

$$\begin{aligned}0 &= \int_{\mathbb{R}^d}\varphi(x,\bar t)\,d\mu_{\bar t}(x) - \int_{\mathbb{R}^d}\varphi(x,0)\,d\mu_0(x) = \int_{\mathbb{R}^d}\Big(\varphi(X_{\bar t}(x),\bar t) - \varphi(x,0)\Big)\,d\mu_0(x)\\ &= \int_{\mathbb{R}^d}\Big(\int_0^{\bar t}\dot\phi_t(x)\,dt\Big)\,d\mu_0(x) = \int_0^{\bar t}\int_{\mathbb{R}^d}\big(\partial_t\varphi + \langle\nabla\varphi, v_t\rangle\big)\,d\mu_t\,dt,\end{aligned}$$

by a simple application of Fubini's theorem. □

We want to prove that, under reasonable assumptions, in fact *any* solution of (8.1.1) can be represented as in Lemma 8.1.6. The first step is a uniqueness theorem for the continuity equation under minimal regularity assumptions on the velocity field. Notice that the only global information on v_t is (8.1.14). The proof, based on a classical duality argument (see for instance [57, 9]), could be much simplified by the assumption that the velocity field is globally bounded, but we prefer to keep here a version of the lemma stronger than the one actually needed in the proof of Theorem 8.3.1.

Proposition 8.1.7 (Uniqueness and comparison for the continuity equation). *Let σ_t be a narrowly continuous family of signed measures solving $\partial_t\sigma_t + \nabla\cdot(v_t\sigma_t) = 0$ in $\mathbb{R}^d \times (0,T)$, with $\sigma_0 \le 0$,*

$$\int_0^T \int_{\mathbb{R}^d} |v_t|\, d|\sigma_t| dt < +\infty, \tag{8.1.14}$$

and

$$\int_0^T \left(|\sigma_t|(B) + \sup_B |v_t| + \mathrm{Lip}(v_t, B) \right) dt < +\infty$$

for any bounded closed set $B \subset \mathbb{R}^d$. Then $\sigma_t \le 0$ for any $t \in [0,T]$.

Proof. Fix $\psi \in C_c^\infty(\mathbb{R}^d \times (0,T))$ with $0 \le \psi \le 1$, $R > 0$, and a smooth cut-off function

$$\begin{gathered}\chi_R(\cdot) = \chi(\cdot/R) \in C_c^\infty(\mathbb{R}^d) \quad \text{such that } 0 \le \chi_R \le 1,\ |\nabla\chi_R| \le 2/R,\\ \chi_R \equiv 1 \text{ on } B_R(0), \text{ and } \chi_R \equiv 0 \text{ on } \mathbb{R}^d \setminus B_{2R}(0).\end{gathered} \tag{8.1.15}$$

We define w_t so that $w_t = v_t$ on $B_{2R}(0) \times (0,T)$, $w_t = 0$ if $t \notin [0,T]$ and

$$\sup_{\mathbb{R}^d} |w_t| + \mathrm{Lip}(w_t, \mathbb{R}^d) \le \sup_{B_{2R}(0)} |v_t| + \mathrm{Lip}(v_t, B_{2R}(0)) \quad \forall\, t \in [0,T]. \tag{8.1.16}$$

Let w_t^ε be obtained from w_t by a double mollification with respect to the space and time variables: notice that w_t^ε satisfy

$$\sup_{\varepsilon\in(0,1)} \int_0^T \left(\sup_{\mathbb{R}^d} |w_t^\varepsilon| + \mathrm{Lip}(w_t^\varepsilon, \mathbb{R}^d) \right) dt < +\infty. \tag{8.1.17}$$

We now build, by the method of characteristics described in Remark 8.1.5, a smooth solution $\varphi^\varepsilon : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ of the PDE

$$\frac{\partial\varphi^\varepsilon}{\partial t} + \langle w_t^\varepsilon, \nabla\varphi^\varepsilon \rangle = \psi \quad \text{in } \mathbb{R}^d \times (0,T), \quad \varphi^\varepsilon(x,T) = 0 \quad x \in \mathbb{R}^d. \tag{8.1.18}$$

Combining the representation formula (8.1.12), the uniform bound (8.1.17), and the estimate (8.1.10), it is easy to check that $0 \ge \varphi^\varepsilon \ge -T$ and $|\nabla\varphi^\varepsilon|$ is uniformly bounded with respect to ε, t and x.

We insert now the test function $\varphi^\varepsilon\chi_R$ in the continuity equation and take into account that $\sigma_0 \le 0$ and $\varphi^\varepsilon \le 0$ to obtain

$$\begin{aligned} 0 \quad &\ge \quad -\int_{\mathbb{R}^d} \varphi^\varepsilon \chi_R \, d\sigma_0 = \int_0^T \int_{\mathbb{R}^d} \chi_R \frac{\partial \varphi^\varepsilon}{\partial t} + \langle v_t, \chi_R \nabla\varphi^\varepsilon + \varphi^\varepsilon \nabla\chi_R\rangle \, d\sigma_t dt \\ &= \quad \int_0^T \int_{\mathbb{R}^d} \chi_R(\psi + \langle v_t - w_t^\varepsilon, \nabla\varphi^\varepsilon\rangle) \, d\sigma_t dt + \int_0^T \int_{\mathbb{R}^d} \varphi^\varepsilon \langle \nabla\chi_R, v_t\rangle \, d\sigma_t dt \\ &\ge \quad \int_0^T \int_{\mathbb{R}^d} \chi_R(\psi + \langle v_t - w_t^\varepsilon, \nabla\varphi^\varepsilon\rangle) \, d\sigma_t dt - \int_0^T \int_{\mathbb{R}^d} |\nabla\chi_R||v_t| \, d|\sigma_t| \, dt. \end{aligned}$$

Letting $\varepsilon \downarrow 0$ and using the uniform bound on $|\nabla\varphi^\varepsilon|$ and the fact that $w_t = v_t$ on $\operatorname{supp}\chi_R \times [0,T]$, we get

$$\int_0^T \int_{\mathbb{R}^d} \chi_R \psi \, d\sigma_t \, dt \le \int_0^T \int_{\mathbb{R}^d} |\nabla\chi_R||v_t| \, d|\sigma_t| \, dt \le \frac{2}{R} \int_0^T \int_{R \le |x| \le 2R} |v_t| \, d|\sigma_t| \, dt.$$

Eventually letting $R \to \infty$ we obtain that $\int_0^T \int_{\mathbb{R}^d} \psi \, d\sigma_t dt \le 0$. Since ψ is arbitrary the proof is achieved. □

Proposition 8.1.8 (Representation formula for the continuity equation). *Let μ_t, $t \in [0,T]$, be a narrowly continuous family of Borel probability measures solving the continuity equation* (8.1.1) *w.r.t. a Borel vector field v_t satisfying* (8.1.7) *and* (8.1.2)*. Then for μ_0-a.e. $x \in \mathbb{R}^d$ the characteristic system* (8.1.8) *admits a globally defined solution $X_t(x)$ in $[0,T]$ and*

$$\mu_t = (X_t)_\# \mu_0 \quad \forall\, t \in [0,T]. \tag{8.1.19}$$

Moreover, if

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p \, d\mu_t(x) \, dt < +\infty \quad \textit{for some } p > 1, \tag{8.1.20}$$

then the velocity field v_t is the time derivative of X_t in the L^p-sense

$$\lim_{h\downarrow 0} \int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} - v_t(X_t(x)) \right|^p d\mu_0(x) \, dt = 0, \tag{8.1.21}$$

$$\lim_{h\to 0} \frac{X_{t+h}(x,t) - x}{h} = v_t(x) \quad \textit{in } L^p(\mu_t;\mathbb{R}^d) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (0,T). \tag{8.1.22}$$

Proof. Let $E_s = \{\tau > s\}$ and let us use the fact that, proved in Lemma 8.1.6, that $t \mapsto X_{t\#}(\chi_{E_s}\mu_0)$ is the solution of (8.1.1) in $[0,s]$. By Proposition 8.1.7 we get also

$$X_{t\#}(\chi_{E_s}\mu_0) \le \mu_t \qquad \text{whenever } 0 \le t \le s.$$

Using the previous inequality with $s = t$ we can estimate:

$$\begin{aligned}
\int_{\mathbb{R}^d} \sup_{(0,\tau(x))} |X_t(x) - x| \, d\mu_0(x) &\leq \int_{\mathbb{R}^d} \int_0^{\tau(x)} |\dot{X}_t(x)| \, d\mu_0(x) \\
&= \int_{\mathbb{R}^d} \int_0^{\tau(x)} |v_t(X_t(x))| \, d\mu_0(x) \\
&= \int_0^T \int_{E_t} |v_t(X_t(x))| \, d\mu_0(x) \, dt \\
&\leq \int_0^T \int_{\mathbb{R}^d} |v_t| \, d\mu_t \, dt.
\end{aligned}$$

It follows that $X_t(x)$ is bounded on $(0, \tau(x))$ for μ_0-a.e. $x \in \mathbb{R}^d$ and therefore X_t is globally defined in $[0, T]$ for μ_0-a.e. in $\mathbb{R}^d$. Applying Lemma 8.1.6 and Proposition 8.1.7 we obtain (8.1.19).

Now we observe that the differential quotient $D_h(x,t) := h^{-1}(X_{t+h}(x) - X_t(x))$ can be bounded in $L^p(\mu_0 \times \mathscr{L}^1)$ by

$$\begin{aligned}
&\int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} \right|^p d\mu_0(x) \, dt \\
&= \int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{1}{h} \int_0^h v_{t+s}(X_{t+s}(x)) \, ds \right|^p d\mu_0(x) \, dt \\
&\leq \int_0^{T-h} \int_{\mathbb{R}^d} \frac{1}{h} \int_0^h |v_{t+s}(X_{t+s}(x))|^p \, ds \, d\mu_0(x) \, dt \\
&\leq \int_0^T \int_{\mathbb{R}^d} |v_t(X_t(x))|^p \, d\mu_0(x) \, dt < +\infty.
\end{aligned}$$

Since we already know that D_h is pointwise converging to $v_t \circ X_t$ $\mu_0 \times \mathscr{L}^1$-a.e. in $\mathbb{R}^d \times (0,T)$, we obtain the strong convergence in $L^p(\mu_0 \times \mathscr{L}^1)$, i.e. (8.1.21).

Finally, we can consider $t \mapsto X_t(\cdot)$ and $t \mapsto v_t(X_t(\cdot)$ as maps from $(0,T)$ to $L^p(\mu_0; \mathbb{R}^d)$; (8.1.21) is then equivalent to

$$\lim_{h \downarrow 0} \int_0^{T-h} \left\| \frac{X_{t+h} - X_t}{h} - v_t(X_t) \right\|_{L^p(\mu_0;\mathbb{R}^d)}^p dt = 0,$$

and it shows that $t \mapsto X_t(\cdot)$ belongs to $AC^p(0,T; L^p(\mu_0; \mathbb{R}^d))$. General results for absolutely continuous maps in reflexive Banach spaces (see 1.1.3) yield that X_t is differentiable $\mathscr{L}^1$-a.e. in $(0,T)$, so that

$$\lim_{h \to 0} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} - v_t(X_t(x)) \right|^p d\mu_0(x) = 0 \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,T).$$

Since $X_{t+h}(x) = X_h(X_t(x), t)$, we obtain (8.1.22). □

Now we state an approximation result for general solution of (8.1.1) with more regular ones, satisfying the conditions of the previous Proposition 8.1.8.

Lemma 8.1.9 (Approximation by regular curves). *Let $p \geq 1$ and let μ_t be a time-continuous solution of* (8.1.1) *w.r.t. a velocity field satisfying the p-integrability condition*

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p \, d\mu_t(x) \, dt < +\infty. \tag{8.1.23}$$

Let $(\rho_\varepsilon) \subset C^\infty(\mathbb{R}^d)$ be a family of strictly positive *mollifiers in the x variable, (e.g. $\rho_\varepsilon(x) = (2\pi\varepsilon)^{-d/2} \exp(-|x|^2/2\varepsilon)$), and set*

$$\mu_t^\varepsilon := \mu_t * \rho_\varepsilon, \quad E_t^\varepsilon := (v_t \mu_t) * \rho_\varepsilon, \quad v_t^\varepsilon := \frac{E_t^\varepsilon}{\mu_t^\varepsilon}. \tag{8.1.24}$$

Then μ_t^ε is a continuous solution of (8.1.1) *w.r.t. v_t^ε, which satisfy the local regularity assumptions* (8.1.7) *and the uniform integrability bounds*

$$\int_{\mathbb{R}^d} |v_t^\varepsilon(x)|^p \, d\mu_t^\varepsilon(x) \leq \int_{\mathbb{R}^d} |v_t(x)|^p \, d\mu_t(x) \quad \forall \, t \in (0, T). \tag{8.1.25}$$

Moreover, $E_t^\varepsilon \to v_t \mu_t$ narrowly and

$$\lim_{\varepsilon \downarrow 0} \|v_t^\varepsilon\|_{L^p(\mu_t^\varepsilon;\mathbb{R}^d)} = \|v_t\|_{L^p(\mu_t;\mathbb{R}^d)} \qquad \forall t \in (0, T). \tag{8.1.26}$$

Proof. With a slight abuse of notation, we are denoting the measure μ_t^ε and its density w.r.t. $\mathscr{L}^d$ by the same symbol. Notice first that $|E^\varepsilon|(t, \cdot)$ and its spatial gradient are uniformly bounded in space by the product of $\|v_t\|_{L^1(\mu_t)}$ with a constant depending on ε, and the first quantity is integrable in time. Analogously, $|\mu_t^\varepsilon|(t, \cdot)$ and its spatial gradient are uniformly bounded in space by a constant depending on ε. Therefore, as $v_t^\varepsilon = E_t^\varepsilon / \mu_t^\varepsilon$, the local regularity assumptions (8.1.7) is fulfilled if

$$\inf_{|x| \leq R, \, t \in [0,T]} \mu_t^\varepsilon(x) > 0 \qquad \text{for any } \varepsilon > 0, \ R > 0.$$

This property is immediate, since μ_t^ε are continuous w.r.t. t and equi-continuous w.r.t. x, and therefore continuous in both variables.

Lemma 8.1.10 shows that (8.1.25) holds. Notice also that μ_t^ε solve the continuity equation

$$\partial_t \mu_t^\varepsilon + \nabla \cdot (v_t^\varepsilon \mu_t^\varepsilon) = 0 \qquad \text{in } \mathbb{R}^d \times (0, T), \tag{8.1.27}$$

because, by construction, $\nabla \cdot (v_t^\varepsilon \, \mu_t^\varepsilon) = \nabla \cdot ((v_t \mu_t) * \rho_\varepsilon) = (\nabla \cdot (v_t \mu_t)) * \rho_\varepsilon$. Finally, general lower semicontinuity results on integral functionals defined on measures of the form

$$(E, \mu) \mapsto \int_{\mathbb{R}^d} \left| \frac{E}{\mu} \right|^p d\mu$$

(see for instance Theorem 2.34 and Example 2.36 in [11]) provide (8.1.26). □

Lemma 8.1.10. *Let $p \geq 1$, $\mu \in \mathscr{P}(\mathbb{R}^d)$ and let E be a $\mathbb{R}^m$-valued measure in $\mathbb{R}^d$ with finite total variation and absolutely continuous with respect to μ. Then*

$$\int_{\mathbb{R}^d} \left|\frac{E * \rho}{\mu * \rho}\right|^p \mu * \rho \, dx \leq \int_{\mathbb{R}^d} \left|\frac{E}{\mu}\right|^p d\mu$$

for any convolution kernel ρ.

Proof. We use Jensen inequality in the following form: if $\Phi : \mathbb{R}^{m+1} \to [0, +\infty]$ is convex, l.s.c. and positively 1-homogeneous, then

$$\Phi\left(\int_{\mathbb{R}^d} \psi(x)\, d\theta(x)\right) \leq \int_{\mathbb{R}^d} \Phi(\psi(x))\, d\theta(x)$$

for any Borel map $\psi : \mathbb{R}^d \to \mathbb{R}^{m+1}$ and any positive and finite measure θ in $\mathbb{R}^d$ (by rescaling θ to be a probability measure and looking at the image measure $\psi_\#\theta$ the formula reduces to the standard Jensen inequality). Fix $x \in \mathbb{R}^d$ and apply the inequality above with $\psi := (E/\mu, 1)$, $\theta := \rho(x - \cdot)\mu$ and

$$\Phi(z,t) := \begin{cases} \dfrac{|z|^p}{t^{p-1}} & \text{if } t > 0 \\ 0 & \text{if } (z,t) = (0,0) \\ +\infty & \text{if either } t < 0 \text{ or } t = 0, z \neq 0, \end{cases}$$

to obtain

$$\begin{aligned} \left|\frac{E * \rho(x)}{\mu * \rho(x)}\right|^p \mu * \rho(x) &= \Phi\left(\int_{\mathbb{R}^d} \frac{E}{\mu}(y)\rho(x-y)\, d\mu(y), \int \rho(x-y) d\mu(y)\right) \\ &\leq \int_{\mathbb{R}^d} \Phi(\frac{E}{\mu}(y), 1)\rho(x-y)\, d\mu(y) \\ &= \int_{\mathbb{R}^d} \left|\frac{E}{\mu}\right|^p (y)\rho(x-y)\, d\mu(y). \end{aligned}$$

An integration with respect to x leads to the desired inequality. □

8.2 A probabilistic representation of solutions of the continuity equation

In this section we extend Proposition 8.1.8 to the case when the vector field fails to satisfy (8.1.7) and is in particular not Lipschitz w.r.t. x. Of course in this situation we have to take into account that characteristics are not unique, and we do that by considering suitable probability measures in the space Γ_T of continuous maps from $[0, T]$ into $\mathbb{R}^d$, endowed with the sup norm. The results presented here are not used in the rest of the book, but we believe that they can have an independent interest. Indeed, this kind of notion plays an important role in the uniqueness and stability of Lagrangian flows in [10] and provides an alternative way to the approach of [57].

Our basic representation formula for solutions $\mu_t^{\boldsymbol{\eta}}$ of the continuity equation (8.1.1) is given by

$$\int_{\mathbb{R}^d} \varphi \, d\mu_t^{\boldsymbol{\eta}} := \int_{\mathbb{R}^d \times \Gamma_T} \varphi(\gamma(t)) \, d\boldsymbol{\eta}(x, \gamma) \qquad \forall \varphi \in C_b^0(\mathbb{R}^d), \;\; t \in [0, T] \tag{8.2.1}$$

where $\boldsymbol{\eta}$ is a probability measure in $\mathbb{R}^d \times \Gamma_T$. In the case when $\boldsymbol{\eta}$ is the push forward under $x \mapsto (x, X_\cdot(x))$ of μ_0 (here we are considering X as a function mapping $x \in \mathbb{R}^d$ into the solution curve $t \mapsto X_t(x)$ in Γ_T) we see that the measures $\mu_t^{\boldsymbol{\eta}}$ implicitly defined by (8.2.1) simply reduce to the standard ones considered in Proposition 8.1.8, i.e. $\mu_t^{\boldsymbol{\eta}} = X_t(\cdot)_\# \mu_0$.

By introducing the evaluation maps

$$\mathsf{e}_t : (x, \gamma) \in \mathbb{R}^d \times \Gamma_T \mapsto \gamma(t) \in \mathbb{R}^d, \quad \text{for } t \in [0, T], \tag{8.2.2}$$

(8.2.1) can also be written as

$$\mu_t^{\boldsymbol{\eta}} = (\mathsf{e}_t)_\# \boldsymbol{\eta}. \tag{8.2.3}$$

Theorem 8.2.1 (Probabilistic representation). *Let $\mu_t : [0, T] \to \mathscr{P}(\mathbb{R}^d)$ be a narrowly continuous solution of the continuity equation* (8.1.1) *for a suitable Borel vector field $v(t, x) = v_t(x)$ satisfying* (8.1.20) *for some $p > 1$. Then there exists a probability measure $\boldsymbol{\eta}$ in $\mathbb{R}^d \times \Gamma_T$ such that*

(i) *$\boldsymbol{\eta}$ is concentrated on the set of pairs (x, γ) such that $\gamma \in AC^p(0, T; \mathbb{R}^d)$ is a solution of the ODE $\dot\gamma(t) = v_t(\gamma(t))$ for $\mathscr{L}^1$-a.e. $t \in (0, T)$, with $\gamma(0) = x$;*

(ii) *$\mu_t = \mu_t^{\boldsymbol{\eta}}$ for any $t \in [0, T]$, with $\mu_t^{\boldsymbol{\eta}}$ defined as in* (8.2.1).

Conversely, any $\boldsymbol{\eta}$ satisfying (i) and

$$\int_0^T \int_{\mathbb{R}^d \times \Gamma_T} |v_t(\gamma(t))| \, d\boldsymbol{\eta}(x, \gamma) \, dt < +\infty, \tag{8.2.4}$$

induces via (8.2.1) *a solution of the continuity equation, with $\mu_0 = \gamma(0)_\# \boldsymbol{\eta}$.*

Proof. We first prove the converse implication, since its proof is much simpler. Indeed, notice that due to assumption (i) the set F of all (t, x, γ) such that either $\dot\gamma(t)$ does not exist or it is different from $v_t(\gamma(t))$ is $\mathscr{L}^1 \times \boldsymbol{\eta}$-negligible. As a consequence, we have

$$\dot\gamma(t) = v_t(\gamma(t)) \quad \boldsymbol{\eta}\text{-a.e., for } \mathscr{L}^1\text{-a.e. } t \in (0, T).$$

It is immediate to check using (8.2.1) that $t \mapsto \mu_t^{\boldsymbol{\eta}}$ is narrowly continuous. Now we check that $t \mapsto \int \zeta \, d\mu_t^{\boldsymbol{\eta}}$ is absolutely continuous for $\zeta \in C^1(\mathbb{R}^d)$ bounded and with a bounded gradient. Indeed, for $s < t$ in I we have

$$
\begin{aligned}
\left|\int_{\mathbb{R}^d} \zeta\, d\mu_s^{\boldsymbol{\eta}} - \int_{\mathbb{R}^d} \zeta\, d\mu_t^{\boldsymbol{\eta}}\right| &\le \int_s^t \int_{\mathbb{R}^d\times\Gamma_T} |\langle \nabla\zeta(\gamma(\tau)), \dot\gamma(\tau)\rangle|\, d\boldsymbol{\eta}\, d\tau \\
&\le \|\nabla\zeta\|_\infty \int_s^t \int_{\mathbb{R}^d\times\Gamma_T} |v_\tau(\gamma(\tau))|\, d\boldsymbol{\eta}\, d\tau.
\end{aligned}
$$

By (8.2.4) this inequality immediately gives the absolute continuity of the map. We have also

$$
\begin{aligned}
\frac{d}{dt}\int_{\mathbb{R}^d} \zeta\, d\mu_t^{\boldsymbol{\eta}} &= \frac{d}{dt}\int_{\mathbb{R}^d\times\Gamma_T} \zeta(\gamma(t))\, d\boldsymbol{\eta} \\
&= \int_{\mathbb{R}^d\times\Gamma_T} \langle \nabla\zeta(\gamma), \dot\gamma(t))\rangle\, d\boldsymbol{\eta} = \int_{\mathbb{R}^d} \langle \nabla\zeta, v_t\rangle\, d\mu_t^{\boldsymbol{\eta}}
\end{aligned}
$$

for $\mathscr{L}^1$-a.e. $t \in (0, T)$. Since this pointwise derivative is also a distributional one, this proves that (8.1.4) holds for test function φ of the form $\zeta(x)\psi(t)$ and therefore for all test functions.

Conversely, let μ_t, v_t be given as in the statement of the theorem and let us apply the regularization Lemma 8.1.9, finding approximations μ_t^ε, v_t^ε satisfying the continuity equation, the uniform integrability condition (8.1.2) and the local regularity assumptions (8.1.7). Therefore, we can apply Proposition 8.1.8, obtaining the representation formula $\mu_t^\varepsilon = (X_t^\varepsilon)_\# \mu_0^\varepsilon$, where X_t^ε is the maximal solution of the ODE $\dot X_t^\varepsilon = v_t^\varepsilon(X_t^\varepsilon)$ with the initial condition $X_0^\varepsilon = x$ (see Lemma 8.1.4). Thinking X^ε as a map from $\mathbb{R}^d$ to Γ_T, we thus define

$$
\boldsymbol{\eta}^\varepsilon := (\boldsymbol{i} \times X_\cdot^\varepsilon)_\# \mu_0^\varepsilon \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T).
$$

Now we claim that the family $\boldsymbol{\eta}^\varepsilon$ is tight as $\varepsilon \downarrow 0$ and that any limit point $\boldsymbol{\eta}$ fulfills (i) and (ii). The tightness of the family can be obtained from Lemma 5.2.2, by choosing the maps $\boldsymbol{r}^1, \boldsymbol{r}^2$ defined in $\mathbb{R}^d \times \Gamma_T$

$$
\boldsymbol{r}^1 : (x, \gamma) \mapsto x \in \mathbb{R}^d, \quad \boldsymbol{r}^2 : (x, \gamma) \mapsto \gamma - x \in \Gamma_T, \tag{8.2.5}
$$

and noticing that $\boldsymbol{r} : \boldsymbol{r}^1 \times \boldsymbol{r}^2 : \mathbb{R}^d \times \Gamma_T \to \mathbb{R}^d \times \Gamma_T$ is proper, the family $\boldsymbol{r}^1_\# \boldsymbol{\eta}^\varepsilon$ is given by the first marginals μ_0^ε which are tight (indeed, they narrowly converge to μ^0), while $\boldsymbol{\beta}^\varepsilon := \boldsymbol{r}^2_\# \boldsymbol{\eta}^\varepsilon$ satisfy

$$
\begin{aligned}
\int_{\Gamma_T} \int_0^T |\dot\gamma|^p\, dt\, d\boldsymbol{\beta}^\varepsilon &= \int_{\mathbb{R}^d} \int_0^T |\dot X_t^\varepsilon(x)|^p\, dt\, d\mu_0^\varepsilon(x) \\
&= \int_{\mathbb{R}^d} \int_0^T |v_t^\varepsilon(X_t^\varepsilon)|^p\, dt\, d\mu_0^\varepsilon(x) = \int_0^T \int_{\mathbb{R}^d} |v_t^\varepsilon(x)|^p\, d\mu_t^\varepsilon(x)\, dt \\
&\le \int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p\, d\mu_t(x)\, dt.
\end{aligned}
$$

Since for $p > 1$ the functional $\gamma \mapsto \int_0^T |\dot\gamma|^p\,dt$ (set to $+\infty$ if $\gamma \notin AC^p((0,T);\mathbb{R}^d)$ or $\gamma(0) \neq 0$) has compact sublevel sets in Γ_T, also $\boldsymbol\beta^\varepsilon$ is tight, due to Remark 5.1.5.

Let now $\boldsymbol\eta$ be a narrow limit point of $\boldsymbol\eta^\varepsilon$, along some infinitesimal sequence ε_i. Since

$$\int_{\mathbb{R}^d} \varphi\,d\mu_t^{\boldsymbol\eta^{\varepsilon_i}} = \int_{\mathbb{R}^d\times\Gamma_T} \varphi(\gamma(t))\,d\boldsymbol\eta^{\varepsilon_i} = \int_{\mathbb{R}^d} \varphi(X_t^{\varepsilon_i})\,d\mu_0^{\varepsilon_i} = \int_{\mathbb{R}^d} \varphi\,d\mu_t^{\varepsilon_i}$$

for any $\varphi \in C_b^0(\mathbb{R}^d)$, we can pass to the limit as $i \to \infty$ to obtain that $\mu_t^{\boldsymbol\eta} = \mu_t$, so that condition (ii) holds.

Finally we check condition (i). Let $w(t,x) = w_t(x)$ be a bounded uniformly continuous function, and let us prove the estimate

$$\int_{\mathbb{R}^d\times\Gamma_T} \left|\gamma(t) - x - \int_0^t w_\tau(\gamma(\tau))\,d\tau\right|^p d\boldsymbol\eta(x,\gamma) \leq (2T)^{p-1}\int_0^T\int_{\mathbb{R}^d} |v_\tau - w_\tau|^p\,d\mu_\tau\,d\tau. \tag{8.2.6}$$

Indeed, we have

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\Gamma_T} \left|\gamma(t) - x - \int_0^t w_\tau(\gamma(\tau))\,d\tau\right|^p d\boldsymbol\eta^\varepsilon(x,\gamma)\\
=\ &\int_{\mathbb{R}^d} \left|X_t^\varepsilon(x) - x - \int_0^t w_\tau(X_\tau^\varepsilon(x))\,d\tau\right|^p d\mu^0(x)\\
=\ &\int_{\mathbb{R}^d} \left|\int_0^t (v_\tau^\varepsilon - w_\tau)(X_\tau^\varepsilon(x))\,d\tau\right|^p d\mu^0(x) \leq t^{p-1}\int_0^t\int_{\mathbb{R}^d} |v_\tau^\varepsilon - w_\tau|^p\,d\mu_t^\varepsilon\,d\tau\\
\leq\ &(2t)^{p-1}\int_0^t\int_{\mathbb{R}^d} |v_\tau^\varepsilon - w_\tau^\varepsilon|^p\,d\mu_t^\varepsilon\,d\tau + (2t)^{p-1}\int_0^t\int_{\mathbb{R}^d} |w_\tau^\varepsilon - w_\tau|^p\,d\mu_t^\varepsilon\,d\tau\\
\leq\ &(2T)^{p-1}\int_0^T\int_{\mathbb{R}^d} |v_\tau - w_\tau|^p\,d\mu_\tau\,d\tau + (2T)^{p-1}\int_0^T \sup_{x\in\mathbb{R}^d} |w_\tau^\varepsilon(x) - w_\tau(x)|^p\,d\tau,
\end{aligned}$$

where in the last two inequalities we have added and subtracted $w_\tau^\varepsilon := w_\tau * \rho_\varepsilon$ and then used Lemma 8.1.10. Setting $\varepsilon = \varepsilon_i$ and passing to the limit as $i \to \infty$ we recover (8.2.6), since the function under the integral is a continuous and nonnegative test function in $\mathbb{R}^d \times \Gamma_T$.

Now let $\mu := \int_0^T \mu_t\,d\mathscr{L}^1(t)$ the Borel measure on $\mathbb{R}^d \times (0,T)$ whose disintegration with respect to $\mathscr{L}^1$ is $\{\mu_t\}_{t\in(0,T)}$ and let $w^n \in C_c^0(\mathbb{R}^d \times (0,T);\mathbb{R}^d)$ be continuous functions with compact support converging to v in $L^p(\mu;\mathbb{R}^d)$. Using the fact that $\mu_t = \mu_t^{\boldsymbol\eta}$ we have

$$\int_{\mathbb{R}^d\times\Gamma_T}\int_0^T |w_\tau^n(\gamma(\tau)) - v_\tau(\gamma(\tau))|^p\,d\tau\,d\boldsymbol\eta = \int_0^T\int_{\mathbb{R}^d} |w_\tau^n - v_\tau|^p\,d\mu_\tau\,d\tau \to 0,$$

as $n \to \infty$ so that, using the triangular inequality in $L^p(\boldsymbol\eta)$, we can pass to the limit as $n \to \infty$ in (8.2.6) with $w = w^n$ to obtain

$$\int_{\mathbb{R}^d\times\Gamma_T} \left|\gamma(t) - x - \int_0^t v_\tau(\gamma(\tau))\,d\tau\right|^p d\boldsymbol\eta(x,\gamma) = 0 \quad \forall\, t \in [0,T], \tag{8.2.7}$$

and therefore

$$\gamma(t) - x - \int_0^t v_\tau(\gamma(\tau))\,d\tau = 0 \qquad \text{for } \boldsymbol{\eta}\text{-a.e. } (x,\gamma)$$

for any $t \in [0,T]$. Choosing all t's in $(0,T)\cap\mathbb{Q}$ we obtain an exceptional $\boldsymbol{\eta}$-negligible set that does not depend on t and use the continuity of γ to show that the identity is fulfilled for any $t \in [0,T]$. □

Notice that due to condition (i) the measure $\boldsymbol{\eta}$ in the previous theorem can also be identified with a measure $\boldsymbol{\sigma}$ in Γ_T whose projection on $\mathbb{R}^d$ via the map $\mathsf{e}_0 : \gamma \mapsto \gamma(0)$ is μ_0 and whose corresponding disintegration $\boldsymbol{\sigma} = \int_{\mathbb{R}^d} \boldsymbol{\sigma}_x\,d\mu_0(x)$ is made by probability measures $\boldsymbol{\sigma}_x$ concentrated on solutions of the ODE starting from x at $t = 0$. In this case (8.2.1) takes the simpler equivalent form

$$\int_{\mathbb{R}^d} \varphi\,d\mu_t^{\boldsymbol{\sigma}} := \int_{\Gamma_T} \varphi(\gamma(t))\,d\boldsymbol{\sigma}(\gamma) \qquad \forall \varphi \in C_b^0(\mathbb{R}^d),\ \ t \in [0,T]. \tag{8.2.8}$$

Finally we notice that the results of this section could be easily be extended to the case when $\mathbb{R}^d$ is replaced by a separable Hilbert space, using a finite dimensional projection argument (see in particular the last part of the proof of Theorem 8.3.1).

8.3 Absolutely continuous curves in $\mathscr{P}_p(X)$

In this section we show that the continuity equation characterizes the class of absolutely continuous curves in $\mathscr{P}_p(X)$, with $p > 1$ and X separable Hilbert space (see [9] for a discussion of the degenerate case $p = 1$ when $X = \mathbb{R}^d$).

Let us first recall that the map $j_p : L^p(\mu; X) \to L^q(\mu; X)$ defined by (here $q = p'$ is the conjugate exponent of p)

$$v \mapsto j_p(v) := \begin{cases} |v|^{p-2}v & \text{if } v \neq 0, \\ 0 & \text{if } v = 0, \end{cases} \tag{8.3.1}$$

provides the differential of the convex functional

$$v \in L^p(\mu; X) \mapsto \frac{1}{p}\int_X |v(x)|^p\,d\mu(x), \tag{8.3.2}$$

for every measure $\mu \in \mathscr{P}(X)$; in particular it satisfies

$$\|j_p(v)\|^q_{L^q(\mu,X)} = \|v\|^p_{L^p(\mu,X)} = \int_X \langle j_p(v), v\rangle\,d\mu(x), \tag{8.3.3}$$

$$w = j_p(v) \quad \Longleftrightarrow \quad v = j_q(w), \tag{8.3.4}$$

$$\frac{1}{p}\|v\|^p_{L^p(\mu;X)} - \frac{1}{p}\|w\|^p_{L^p(\mu;X)} \geq \langle j_p(w), v-w\rangle \quad \forall\, v, w \in L^p(\mu;X). \tag{8.3.5}$$

Recall that the space of smooth cylindrical functions $\mathrm{Cyl}(X)$ has been introduced in Definition 5.1.11; the space $\mathrm{Cyl}(X\times I)$, $I=(a,b)$ being an open interval, is defined analogously considering functions $\psi\in C^\infty_c(\mathbb{R}^d\times I)$ and functions $\varphi(x,t)=\psi(\pi(x),t)$.

Theorem 8.3.1 (Absolutely continuous curves and the continuity equation). *Let I be an open interval in $\mathbb{R}$, let $\mu_t : I\to\mathscr{P}_p(X)$ be an absolutely continuous curve and let $|\mu'|\in L^1(I)$ be its metric derivative, given by Theorem 1.1.2. Then there exists a Borel vector field $v:(x,t)\mapsto v_t(x)$ such that*

$$v_t\in L^p(\mu_t;X), \qquad \|v_t\|_{L^p(\mu_t;X)}\leq |\mu'|(t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t\in I, \tag{8.3.6}$$

and the continuity equation

$$\partial_t\,\mu_t+\nabla\cdot(v_t\mu_t)=0 \qquad \text{in } X\times I \tag{8.3.7}$$

holds in the sense of distributions, i.e.

$$\int_I\int_X \Big(\partial_t\varphi(x,t)+\langle v_t(x),\nabla_x\varphi(x,t)\rangle\Big)\,d\mu_t(x)\,dt=0 \quad \forall\,\varphi\in\mathrm{Cyl}(X\times I). \tag{8.3.8}$$

Moreover, for $\mathscr{L}^1$-a.e. $t\in I$ $j_p(v_t)$ belongs to the closure in $L^q(\mu_t,X)$ of the subspace generated by the gradients $\nabla\varphi$ with $\varphi\in\mathrm{Cyl}(X)$.
Conversely, if a narrowly continuous curve $\mu_t: I\to\mathscr{P}_p(X)$ satisfies the continuity equation for some Borel velocity field v_t with $\|v_t\|_{L^p(\mu_t;X)}\in L^1(I)$ then $\mu_t: I\to\mathscr{P}_p(X)$ is absolutely continuous and $|\mu'|(t)\leq\|v_t\|_{L^p(\mu_t;X)}$ for $\mathscr{L}^1$-a.e. $t\in I$.

Proof. Taking into account Lemma 1.1.4 and Lemma 8.1.3, we will assume with no loss of generality that $|\mu'|\in L^\infty(I)$ in the proof of the first statement. To fix the ideas, we also assume that $I=(0,1)$.

First of all we show that for every $\varphi\in\mathrm{Cyl}(X)$ the function $t\mapsto\mu_t(\varphi)$ is absolutely continuous, and its derivative can be estimated with the metric derivative of μ_t. Indeed, for s, $t\in I$ we have, for $\boldsymbol{\mu}_{st}\in\Gamma_o(\mu_s,\mu_t)$ and using the Hölder inequality,

$$|\mu_t(\varphi)-\mu_s(\varphi)|=\left|\int_{X\times X}\big(\varphi(y)-\varphi(x)\big)\,d\boldsymbol{\mu}_{st}\right|\leq \mathrm{Lip}(\varphi)W_p(\mu_s,\mu_t),$$

whence the absolute continuity follows. In order to estimate more precisely the derivative of $\mu_t(\varphi)$ we introduce the upper semicontinuous and bounded map

$$H(x,y):=\begin{cases}|\nabla\varphi(x)| & \text{if } x=y,\\[2mm] \dfrac{|\varphi(x)-\varphi(y)|}{|x-y|} & \text{if } x\neq y,\end{cases}$$

and notice that, setting $\boldsymbol{\mu}_h = \boldsymbol{\mu}_{(s+h)s}$, we have

$$\begin{aligned}\frac{|\mu_{s+h}(\varphi) - \mu_s(\varphi)|}{|h|} &\leq \frac{1}{|h|}\int_{X\times X} |x-y| H(x,y)\, d\boldsymbol{\mu}_h \\ &\leq \frac{W_p(\mu_{s+h},\mu_s)}{|h|}\left(\int_{X\times X} H^q(x,y)\, d\boldsymbol{\mu}_h\right)^{1/q},\end{aligned}$$

where q is the conjugate exponent of p. If t is a point where $s \mapsto \mu_s$ is metrically differentiable, using the fact that $\boldsymbol{\mu}_h \to (x,x)_\# \mu_t$ narrowly (because their marginals are narrowly converging, any limit point belongs to $\Gamma_o(\mu_t,\mu_t)$ and is concentrated on the diagonal of $X \times X$) we obtain

$$\limsup_{h\to 0} \frac{|\mu_{t+h}(\varphi) - \mu_t(\varphi)|}{|h|} \leq |\mu'|(t)\left(\int_X |H|^q(x,x)\, d\mu_t\right)^{1/q} = |\mu'|(t)\|\nabla\varphi\|_{L^q(\mu_t;X)}. \tag{8.3.9}$$

Set $Q = X \times I$ and let $\mu = \int \mu_t\, dt \in \mathscr{P}(Q)$ be the measure whose disintegration is $\{\mu_t\}_{t\in I}$. For any $\varphi \in \mathrm{Cyl}(Q)$ we have

$$\begin{aligned}\int_Q \partial_s\varphi(x,s)\, d\mu(x,s) &= \lim_{h\downarrow 0}\int_Q \frac{\varphi(x,s)-\varphi(x,s-h)}{h}\, d\mu(x,s) \\ &= \lim_{h\downarrow 0}\int_I \frac{1}{h}\Big(\int_X \varphi(x,s)\, d\mu_s(x) - \int_X \varphi(x,s)\, d\mu_{s+h}(x)\Big)\, ds.\end{aligned}$$

Taking into account (8.3.9), Fatou's Lemma yields

$$\begin{aligned}\left|\int_Q \partial_s\varphi(x,s)\, d\mu(x,s)\right| &\leq \int_J |\mu'|(s)\Big(\int_X |\nabla\varphi(x,s)|^q\, d\mu_s(x)\Big)^{1/q} ds \\ &\leq \Big(\int_J |\mu'|^p(s)\, ds\Big)^{1/p}\Big(\int_Q |\nabla\varphi(x,s)|^q\, d\mu(x,s)\Big)^{1/q},\end{aligned} \tag{8.3.10}$$

where $J \subset I$ is any interval such that $\operatorname{supp}\varphi \subset J \times X$. If $\mathscr{V}$ denotes the closure in $L^q(\mu;X)$ of the subspace $V := \Big\{\nabla\varphi, \quad \varphi \in \mathrm{Cyl}(Q)\Big\}$, the previous formula says that the linear functional $L : V \to \mathbb{R}$ defined by

$$L(\nabla\varphi) := -\int_Q \partial_s\varphi(x,s)\, d\mu(x,s)$$

can be uniquely extended to a bounded functional on $\mathscr{V}$. Therefore the minimum problem

$$\min\left\{\frac{1}{q}\int_Q |w(x,s)|^q\, d\mu(x,s) - L(w) :\ w \in \mathscr{V}\right\} \tag{8.3.11}$$

admits a unique solution $w \in \mathscr{V}$ such that $v := j_q(w)$ satisfies

$$\int_Q \langle v(x,s), \nabla\varphi(x,s)\rangle\, d\mu(x,s) = \langle L, \nabla\varphi\rangle \qquad \forall \varphi \in \mathrm{Cyl}(Q). \tag{8.3.12}$$

Setting $v_t(x) = v(x,t)$ and using the definition of L we obtain (8.3.8). Moreover, choosing a sequence $(\nabla\varphi_n) \subset V$ converging to w in $L^q(\mu; X)$, it is easy to show that for $\mathscr{L}^1$-a.e. $t \in I$ there exists a subsequence $n(i)$ (possibly depending on t) such that $\nabla\varphi_{n(i)}(\cdot, t) \in \mathrm{Cyl}(X)$ converge in $L^q(\mu_t; X)$ to $w(\cdot, t) = j_p(v(\cdot, t))$.

Finally, choosing an interval $J \subset I$ and $\eta \in C_c^\infty(J)$ with $0 \le \eta \le 1$, (8.3.12) and (8.3.10) yield

$$
\begin{aligned}
&\int_Q \eta(s)|v(x,s)|^p\,d\mu(x,s) = \int_Q \eta\langle v, w\rangle\,d\mu = \lim_{n\to\infty}\int_Q \eta\langle v, \nabla\varphi_n\rangle\,d\mu \\
&= \lim_{n\to\infty}\langle L, \nabla(\eta\varphi_n)\rangle \le \Big(\int_J |\mu'|^p(s)\,ds\Big)^{1/p} \lim_{n\to\infty}\Big(\int_{X\times J}|\nabla\varphi_n|^q\,d\mu\Big)^{1/q} \\
&= \Big(\int_J |\mu'|^p(s)\,ds\Big)^{1/p}\Big(\int_{X\times J}|w|^q\,d\mu\Big)^{1/q} = \Big(\int_J |\mu'|^p(s)\,ds\Big)^{1/p}\Big(\int_{X\times J}|v|^p\,d\mu\Big)^{1/q}.
\end{aligned}
$$

Taking a sequence of smooth approximations of the characteristic function of J we obtain

$$\int_J\int_X |v_s(x)|^p\,d\mu_s(x)\,ds \le \int_J |\mu'|^p(s)\,ds, \tag{8.3.13}$$

and therefore

$$\|v_t\|_{L^p(\mu_t,X)} \le |\mu'|(t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in I.$$

Now we show the converse implication, assuming first that $X = \mathbb{R}^d$. We apply the regularization Lemma 8.1.9, finding approximations μ_t^ε, v_t^ε satisfying the continuity equation, the uniform integrability condition (8.1.2) and the local regularity assumptions (8.1.7). Therefore, we can apply Proposition 8.1.8, obtaining the representation formula $\mu_t^\varepsilon = (T_t^\varepsilon)_\#\mu_0^\varepsilon$, where T_t^ε is the maximal solution of the ODE $\dot T_t^\varepsilon = v_t^\varepsilon(T_t^\varepsilon)$ with the initial condition $T_0^\varepsilon = x$ (see Lemma 8.1.4).

Now, taking into account Lemma 8.1.10, we estimate

$$
\begin{aligned}
\int_{\mathbb{R}^d} |T_{t_2}^\varepsilon(x) - T_{t_1}^\varepsilon(x)|^p\,d\mu_0^\varepsilon \;&\le\; (t_2-t_1)^{p-1}\int_{\mathbb{R}^d}\int_{t_1}^{t_2} |\dot T_t^\varepsilon(x)|^p\,dt\,d\mu_0^\varepsilon \qquad (8.3.14)\\
&=\; (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d} |v_t^\varepsilon(x)|^p\,d\mu_t^\varepsilon\,dt \\
&\le\; (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d} |v_t|^p\,d\mu_t dt,
\end{aligned}
$$

therefore the transport plan $\boldsymbol{\gamma}^\varepsilon := (T_{t_1}^\varepsilon \times T_{t_2}^\varepsilon)_\#\mu_0^\varepsilon$ satisfies

$$W_p^p(\mu_{t_1}^\varepsilon, \mu_{t_2}^\varepsilon) \le \int_{\mathbb{R}^{2d}} |x-y|^p\,d\boldsymbol{\gamma}^\varepsilon \le (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d}|v_t|^p\,d\mu_t\,dt.$$

Since for every $t \in I$ μ_t^ε converges narrowly to μ_t as $\varepsilon \to 0$, Lemma 7.1.3 shows that for any limit point $\boldsymbol{\gamma}$ of $\boldsymbol{\gamma}^\varepsilon$ we have

$$W_p^p(\mu_{t_1}, \mu_{t_2}) \le \int_{\mathbb{R}^{2d}} |x-y|^p\,d\boldsymbol{\gamma} \le (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d}|v_t|^p\,d\mu_t dt.$$

Since t_1 and t_2 are arbitrary this implies that μ_t is absolutely continuous and that its metric derivative is less than $\|v_t\|_{L^p(\mu_t;X)}$ for $\mathscr{L}^1$-a.e. $t \in I$.

We conclude the proof considering the general infinite-dimensional case and following a typical reduction argument, by projecting measures on finite dimensional subspaces. Let $\pi^d : X \to \mathbb{R}^d$ be the canonical maps, given by (5.1.28) for an orthonormal basis (e_n) of X, let $\mu_t^d := \pi^d_\# \mu_t \in \mathscr{P}(\mathbb{R}^d)$, and let $\{\mu_{ty}\}_{y\in\mathbb{R}^d}$ be the disintegration of μ_t with respect to μ_t^d as in Theorem 5.3.1. Notice that considering test functions $\varphi = \psi \circ \pi^d$ in (8.1.3), with $\nabla\varphi = (\pi^d)^* \circ \nabla\psi \circ \pi^d$, gives

$$
\begin{aligned}
\frac{d}{dt}\int_X \varphi\, d\mu_t(x) &= \int_X \langle \pi^d(v_t), \nabla\psi \circ \pi^d\rangle\, d\mu_t(x)\\
&= \int_{\mathbb{R}^d} \Big(\int_{(\pi^d)^{-1}(y)} \langle \pi^d(v_t), \nabla\psi\circ\pi^d\rangle\, d\mu_{ty}(x)\Big)\, d\mu_t^d(y)\\
&= \int_{\mathbb{R}^d} \langle \int_{(\pi^d)^{-1}(y)} \pi^d\big(v_t(x)\big)\, d\mu_{ty}(x), \nabla\psi(y)\rangle\, d\mu_t^d(y) = \int_{\mathbb{R}^d} \langle v_t^d(y), \nabla\psi(y)\rangle\, d\mu_t^d(y),
\end{aligned}
$$

with $v_t^d(y) := \int_{(\pi^d)^{-1}(y)} \pi^d\big(v_t^d(x)\big)\, d\mu_{ty}(x)$, and therefore

$$
\partial_t\, \mu_t^d + \nabla \cdot (v_t^d \mu_t^d) = 0 \qquad \text{in } \mathbb{R}^d \times I.
$$

Notice also that, by similar calculations,

$$
\begin{aligned}
\left| \int_{\mathbb{R}^d} \langle v_t^d(y), \chi(y)\rangle\, d\mu_t^d(y) \right| &= \left| \int_X \langle \pi^d(v_t(x)), \chi(\pi^d(x))\rangle\, d\mu_t \right|\\
&\le \|v_t\|_{L^p(\mu_t;X)} \|\chi\|_{L^q(\mu_t^d;\mathbb{R}^d)}
\end{aligned}
$$

for any $\chi \in L^\infty(\mu_t^d;\mathbb{R}^d)$, hence $\|v_t^d\|_{L^p(\mu_t^d;\mathbb{R}^d)} \le \|v_t\|_{L^p(\mu_t;X)}$. Therefore $t \mapsto \mu_t^d$ is an absolutely continuous curve in $\mathscr{P}_p(\mathbb{R}^d)$ and

$$
W_p(\mu_{t_1}^d, \mu_{t_2}^d) \le \int_{t_1}^{t_2} \|v_t^d\|_{L^p(\mu_t^d;\mathbb{R}^d)}\, dt \le \int_{t_1}^{t_2} \|v_t\|_{L^p(\mu_t;X)}\, dt \quad \forall t_1,\, t_2 \in I,\ t_1 \le t_2.
$$

Let now

$$
\hat\mu_t^d = (\pi^d)^*_\# \mu_t^d = \hat\pi^d_\# \mu_t,
$$

be the image of the measures μ_t^d under the isometries $(\pi^d)^* : y \mapsto \sum_1^d y_i e_i$. Passing to the limit as $d \to \infty$ and using the narrow convergence of $\hat\mu_t^d$ to μ_t and (7.1.11) we obtain

$$
W_p(\mu_{t_1}, \mu_{t_2}) \le \int_{t_1}^{t_2} \|v_t\|_{L^p(\mu_t,X)}\, dt \qquad \forall t_1,\, t_2 \in I,\ t_1 \le t_2.
$$

This proves that μ_t is absolutely continuous and that its metric derivative can be estimated with $\|v_t\|_{L^p(\mu_t;X)}$. □

In the case when the measures are constant in time, by combining the previous finite dimensional projection argument and the smoothing technique of Lemma 8.1.9, one obtains an important approximation property. Let us first collect some preliminary useful properties of orthogonal projections of measures and vector fields, some of which we already proved in the last part of the above proof.

Lemma 8.3.2 (Finite dimensional projection of vector fields). *Let* $\mu \in \mathscr{P}_p(X)$, $v \in L^p(\mu; X)$, *and let* $\{e_n\}_{n=1}^{\infty}$ *be a complete orthonormal system of* X, *with the associated canonical maps* $\pi^d, (\pi^d)^*, \hat{\pi}^d$ *given by* (5.1.28), (5.1.29), *and* (5.1.30). *We consider the finite dimensional subspaces* $X^d := \mathrm{span}(e_1, \ldots, e_d)$, *the measures* $\hat{\mu}^d := \hat{\pi}^d_{\#}\mu$, *the disintegration* $\{\mu_x\}_{x \in X^d}$ *of* μ *w.r.t.* $\hat{\mu}^d$ *given by Theorem* 5.3.1, *and the vector field*

$$\hat{v}^d(x) := \int_{(\hat{\pi}^d)^{-1}(x)} \hat{\pi}^d(v(y))\, d\mu_x(y) \quad \text{for } \hat{\mu}^d\text{-a.e. } x \in X^d. \tag{8.3.15}$$

The following properties hold:

(i) $\mathrm{supp}\, \hat{\mu}^d \subset X^d$, $\quad \hat{\mu}^d \to \mu$ *in* $\mathscr{P}_p(X)$ *as* $d \to \infty$. *If* μ *is regular then also* $\hat{\mu}^d|_{X^d}$ *is regular;*

(ii) $\hat{v}^d \in L^p(\hat{\mu}^d; X^d)$ *with*

$$\|\hat{v}^d\|_{L^p(\hat{\mu}^d; X^d)} \le \|v\|_{L^p(\mu; X)}; \tag{8.3.16}$$

(iii) $\hat{v}^d$ *is characterized by the following identity*

$$\int_X \langle \boldsymbol{\zeta}(x), \hat{v}^d(x)\rangle\, d\hat{\mu}^d(x) = \int_X \langle \hat{\pi}^d \boldsymbol{\zeta}(\hat{\pi}^d(x)), v(x)\rangle\, d\mu(x), \tag{8.3.17}$$

for every bounded Borel vector field $\boldsymbol{\zeta} : X \to X$;

(iv) *If* $\nabla \cdot (v\mu) = 0$ *(in the duality with smooth cylindrical maps), then also* $\nabla \cdot (\hat{v}^d \hat{\mu}^d) = 0$;

(v) *for every continuous function* $f : X \times X \to \mathbb{R}$ *with* p*-growth according to* (5.1.21) *we have*

$$\lim_{d\to\infty} \int_{X\times X} f(x, \hat{v}^d(x))\, d\hat{\mu}^d(x) = \int_{X\times X} f(x, v(x))\, d\mu(x). \tag{8.3.18}$$

In particular, $\hat{v}^d\hat{\mu}^d \to v\mu$ *in the duality with* $C^0_b(X; X)$ *and*

$$\lim_{d\to\infty} \|\hat{v}^d\|_{L^p(\mu^d; X)} = \|v\|_{L^p(\mu; X)}. \tag{8.3.19}$$

Proof. (i) is immediate and we have seen in the previous proof that (ii) is a direct consequence of (iii); in order to check this point we simply use the Definition

(8.3.15) of $\hat v^d$ obtaining

$$\begin{aligned}\int_X \langle \boldsymbol\zeta(x), \hat v^d(x)\rangle\, d\hat\mu^d(x) &= \int_X \langle \boldsymbol\zeta(x), \int_{(\hat\pi^d)^{-1}(x)} \hat\pi^d v(y)\, d\mu_x(y)\rangle\, d\hat\mu^d(x)\\ &= \int_X \int_{(\hat\pi^d)^{-1}(x)} \langle \boldsymbol\zeta(\hat\pi^d(y)), \hat\pi^d v(y)\rangle\, d\mu_x(y)\, d\hat\mu^d(x)\\ &= \int_X \langle \boldsymbol\zeta(\hat\pi^d(x)), \hat\pi^d v(x)\rangle\, d\mu(x) = \int_X \langle \hat\pi^d \boldsymbol\zeta(x), v(x)\rangle\, d\mu(x).\end{aligned}$$

(iv) follows by (iii) simply choosing $\boldsymbol\zeta := \nabla(\hat\chi^d_R\varphi)$, for $\varphi \in \mathrm{Cyl}(X)$ and $\hat\chi^d_R := \chi_R \circ \pi^d$ as in (8.1.15), and observing that

$$\hat\pi^d\big(\nabla(\hat\chi^d_R(\hat\pi^d)\varphi(\hat\pi^d))\big) = \nabla\big((\hat\chi^d_R\varphi)\circ\hat\pi^d\big), \quad (\hat\chi^d_R\varphi)\circ\hat\pi^d \in \mathrm{Cyl}(X).$$

Therefore we get

$$\int_X \langle \nabla\varphi, \hat v^d\rangle\, d\hat\mu^d = \lim_{R\uparrow+\infty} \int_X \langle \nabla(\hat\chi^d_R\varphi), \hat v^d\rangle\, d\hat\mu^d = \lim_{R\uparrow+\infty} \int_X \langle \nabla\big((\hat\chi^d_r\varphi)\circ\hat\pi^d\big), v\rangle\, d\mu = 0.$$

Finally, (8.3.17) easily yields

$$\lim_{d\to\infty} \int_X \langle \boldsymbol\zeta, \hat v^d\rangle\, d\hat\mu^d = \int_X \langle \boldsymbol\zeta, v\rangle\, d\mu \quad \forall\, \boldsymbol\zeta \in C^0_b(X;X); \tag{8.3.20}$$

taking into account of (8.3.16), of Definition 5.4.3, and of Theorem 5.4.4, we conclude. □

Proposition 8.3.3 (Approximation by regular measures). *For any $\mu \in \mathscr{P}_p(X)$, any $v \in L^p(\mu;X)$ such that $\nabla\cdot(v\mu) = 0$ (in the duality with smooth cylindrical functions), and any complete orthonormal system $\{e_n\}_{n\ge1}$, there exist measures $\mu_h \in \mathscr{P}_p(X)$ and vectors $v_h \in L^p(\mu_h;X)$, $h\in\mathbb{N}$, such that*

i. $\operatorname{supp}\mu_h \subset X_h := \mathrm{span}(e_1,\ldots,e_h)$ *(in the finite dimensional case we simply set $X_h = X$),*

ii. $\mu_h|_{X_h} \in \mathscr{P}^r_p(X_h)$,

iii. $v_h(x) \in X_h(x) \quad \forall\, x \in X, \qquad \nabla\cdot(v_h\mu_h) = 0,$

iv. $\mu_h \to \mu$ *in $\mathscr{P}_p(X)$ as $h\to\infty$,*

v. *for every continuous function $f : X\times X \to \mathbb{R}$ with p-growth according to* (5.1.21) *we have*

$$\lim_{h\to\infty} \int_{X\times X} f(x, v_h(x))\, d\mu_h(x) = \int_{X\times X} f(x, v(x))\, dx. \tag{8.3.21}$$

In particular, $v_h\mu_h \to v\mu$ in the duality with $C^0_b(X;X)$ and

$$\lim_{h\to\infty} \|v_h\|_{L^p(\mu_h;X)} = \|v\|_{L^p(\mu;X)}.$$

Proof. To each finite dimensional measure and vector field provided by Lemma 8.3.2 we apply the smoothing argument of Lemma 8.1.9; the proof is achieved by a simple diagonal argument. □

8.4 The tangent bundle to $\mathscr{P}_p(X)$

Notice that the continuity equation (8.3.7) involves only the action of v_t on $\nabla\varphi$ with $\varphi \in \mathrm{Cyl}(X)$. Moreover, Theorem 8.3.1 shows that the minimal norm among all possible velocity fields v_t is the metric derivative and that $j_p(v_t)$ belongs to the L^q closure of gradients of functions in $\mathrm{Cyl}(X)$. These facts suggest a "canonical" choice of v_t and the following definition of tangent bundle to $\mathscr{P}_p(X)$.

Definition 8.4.1 (Tangent bundle). *Let $\mu \in \mathscr{P}_p(X)$. We define*

$$\mathrm{Tan}_\mu \mathscr{P}_p(X) := \overline{\{j_q(\nabla\varphi) : \ \varphi \in \mathrm{Cyl}(X)\}}^{L^p(\mu;X)},$$

where $j_q : L^q(\mu;X) \to L^p(\mu;X)$ is the duality map defined in (8.3.1) .

Notice also that $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ can be equivalently defined as the image under j_q of the L^q closure of gradients of smooth cylindrical functions in X. The choice of $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ is motivated by the following variational selection principle (nonlinear in the case $p \neq 2$):

Lemma 8.4.2 (Variational selection of the tangent vectors). *A vector $v \in L^p(\mu;X)$ belongs to the tangent cone $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ iff*

$$\|v+w\|_{L^p(\mu;X)} \geq \|v\|_{L^p(\mu;X)} \quad \forall\, w \in L^p(\mu;X) \text{ such that } \nabla\cdot(w\mu) = 0. \tag{8.4.1}$$

In particular, for every $v \in L^p(\mu;X)$ there exists a unique $\Pi(v) \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$ in the equivalence class of v modulo divergence-free vector fields, $\Pi(v)$ is the element of minimal L^p-norm in this class, and

$$\int_X \langle j_p(v), w - \Pi(w)\rangle \, d\mu(x) = 0 \quad \forall\, v \in \mathrm{Tan}_\mu \mathscr{P}_p(X), \ w \in L^p(\mu;X). \tag{8.4.2}$$

Proof. By the convexity of the L^p norm, (8.4.1) holds iff

$$\int_X \langle j_p(v), w\rangle \, d\mu = 0 \quad \text{for any } w \in L^p(\mu;X) \text{ s.t. } \nabla\cdot(w\mu) = 0 \tag{8.4.3}$$

(here the divergence is understood making the duality with smooth cylindrical test functions) and this is true iff $j_p(v)$ belongs to the L^q closure of $\{\nabla\phi : \ \phi \in \mathrm{Cyl}(X)\}$. Therefore $v = j_q(j_p(v))$ belongs to $\mathrm{Tan}_\mu \mathscr{P}_p(X)$. (8.4.2) follows from (8.4.3) since $w - \Pi(w)$ is divergence free. □

Observe that the projection Π is linear and $\mathrm{Tan}_\mu\mathscr{P}_p(X)$ is a vector space only in the Hilbertian case $p = q = 2$.

The remarks above lead also to the following characterization of divergence-free vector fields:

Proposition 8.4.3. *Let $w \in L^p(\mu; X)$. Then $\nabla \cdot (w\mu) = 0$ iff $\|v - w\|_{L^p(\mu;X)} \geq \|v\|_{L^p(\mu;X)}$ for any $v \in \mathrm{Tan}_\mu\mathscr{P}_p(X)$. Moreover equality holds for some v iff $w = 0$.*

Proof. We already proved that $\nabla \cdot (w\mu) = 0$ implies $\|v - w\|_{L^p(\mu;X)} \geq \|v\|_{L^p(\mu;X)}$ for any $v \in \mathrm{Tan}_\mu\mathscr{P}_p(X)$. Let us prove now the opposite implication. Indeed, being $\mathrm{Tan}_\mu\mathscr{P}_p(X)$ a cone, a differentiation yields

$$\int_X \langle j_p(v), w\rangle \, d\mu = 0 \qquad \forall v \in \mathrm{Tan}_\mu\mathscr{P}_p(X),$$

and choosing $v = j_q(\nabla\varphi)$, with $\varphi \in \mathrm{Cyl}(X)$, we obtain $\int_X \langle \nabla\varphi, w\rangle \, d\mu = 0$ for any $\varphi \in \mathrm{Cyl}(X)$.

We give now an elementary proof of the fact that if equality holds for some v, then $w = 0$. If equality holds for some v the convexity of the L^p norm gives $\|v + tw\|_{L^p(\mu;X)} = \|v\|_{L^p(\mu;X)}$ for any $t \in [0, 1]$, and differentiation with respect to t gives

$$\int_X |v + tw|^{p-2}\langle v + tw, w\rangle \, d\mu = 0 \qquad \forall t \in (0, 1).$$

Differentiating once more (and using the monotone convergence theorem and the convexity of the map $t \mapsto |a + tb|^p$) we eventually obtain

$$\int_X |v + tw|^{p-2}\left[|w|^2 + (p-2)\frac{(\langle v + tw, w\rangle)^2}{|v + tw|^2}\right] d\mu = 0 \quad \forall t \in (0, 1).$$

Since the integrand is nonnegative it immediately follows that $w = 0$. □

In the particular case $p = 2$ the map j_2 is the identity and (8.4.3) gives

$$\mathrm{Tan}^\perp_\mu\mathscr{P}_2(X) = \left\{v \in L^2(\mu, X) : \ \nabla \cdot (v\mu) = 0\right\}. \tag{8.4.4}$$

Remark 8.4.4 (Cotangent space, duality, and quotients). Since tangent vectors acts naturally only on gradient vector fields, one could also define the *cotangent space* as

$$\mathrm{CoTan}_\mu\mathscr{P}_p(X) := \overline{\{\nabla\varphi : \ \varphi \in \mathrm{Cyl}(X)\}}^{L^q(\mu;X)}, \tag{8.4.5}$$

and therefore the tangent space by duality. If $\sim$ denotes the equivalence relation which identifies two vector fields in $L^p(\mu; X)$ if their difference is divergence free, the tangent space could be identified with the quotient space $L^p(\mu; X)/\sim$. Definition 8.4.1 and the related lemma 8.4.2 simply operates a canonical (though nonlinear) selection of an element $\Pi(v)$ in the class of v by using the duality map between the Cotangent and the Tangent space. This distinction becomes superfluous in the Hilbertian case $p = q = 2$, since in that case the tangent and the cotangent spaces turn out to be the same, by the usual identification via the Riesz isomorphism.

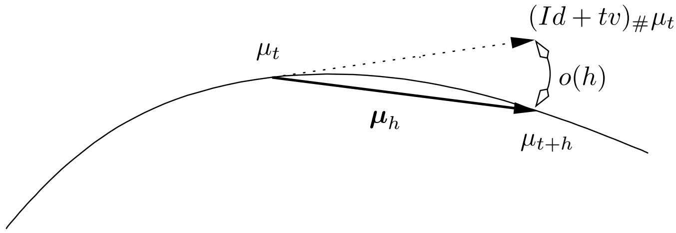
The following two propositions show that the notion of tangent space is consistent with the metric structure, with the continuity equation, and with optimal transport maps (if any).

Proposition 8.4.5 (Tangent vector to a.c. curves). *Let $\mu_t : I \to \mathscr{P}_p(X)$ be an absolutely continuous curve and let $v_t \in L^p(\mu_t; X)$ be such that* (8.3.7) *holds. Then v_t satisfies* (8.3.6) *as well if and only if $v_t = \Pi(v_t) \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ for $\mathscr{L}^1$-a.e. $t \in I$. The vector v_t is uniquely determined $\mathscr{L}^1$-a.e. in I by* (8.3.6) *and* (8.3.7).

Proof. The uniqueness of v_t is a straightforward consequence of the linearity with respect to the velocity field of the continuity equation and of the strict convexity of the L^p norm.

In the proof of Theorem 8.3.1 we built vector fields $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ satisfying (8.3.6) and (8.3.7). By uniqueness, it follows that conditions (8.3.6) and (8.3.7) imply $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ for $\mathscr{L}^1$-a.e. t. □

In the following proposition we recover the tangent vector field to a curve through the infinitesimal behaviour of optimal transport maps, or plans, along the curve. Notice that in the limit we recover a plan $(\boldsymbol{i}\times v_t)_\#\mu_t$ associated to a *classical* transport even in the situation when μ_t are not necessarily absolutely continuous. It is for this reason that we don't need, at least for differential calculus along absolutely continuous curves, the more general notions of tangent space, made by plans instead of maps, discussed in the Appendix.



Proposition 8.4.6 (Optimal plans along a.c. curves). *Let $\mu_t : I \to \mathscr{P}_p(X)$ be an absolutely continuous curve and let $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ be characterized by Proposition* 8.4.5*. Then, for $\mathscr{L}^1$-a.e. $t \in I$ the following property holds: for any choice of $\boldsymbol{\mu}_h \in \Gamma_o(\mu_t, \mu_{t+h})$ we have*

$$\lim_{h\to 0}\big(\pi^1, \frac{1}{h}(\pi^2 - \pi^1)\big)_\#\boldsymbol{\mu}_h = (\boldsymbol{i}\times v_t)_\#\mu_t \qquad \textit{in } \mathscr{P}_p(X\times X) \tag{8.4.6}$$

and

$$\lim_{h\to 0}\frac{W_p(\mu_{t+h}, (\boldsymbol{i}+hv_t)_\#\mu_t)}{|h|} = 0. \tag{8.4.7}$$

In particular, for $\mathscr{L}^1$-a.e. $t \in I$ such that $\mu_t \in \mathscr{P}_p^r(X)$ we have

$$\lim_{h\to 0} \frac{1}{h}(\boldsymbol{t}_{\mu_t}^{\mu_{t+h}} - \boldsymbol{i}) = v_t \qquad \text{in } L^p(\mu_t; X), \tag{8.4.8}$$

where $\boldsymbol{t}_{\mu_t}^{\mu_{t+h}}$ is the unique optimal transport map between μ_t and μ_{t+h}.

Proof. Let $\mathscr{D}_d \subset C_c^\infty(\mathbb{R}^d)$ be a countable set with the following property: for any integer $R > 0$ and any $\psi \in C_c^\infty(\mathbb{R}^d)$ with $\operatorname{supp}\psi \subset B_R$ there exist $(\varphi_n) \subset \mathscr{D}_d$ with $\operatorname{supp}\varphi_n \subset B_R$ and $\varphi_n \to \varphi$ in $C^1(\mathbb{R}^d)$. Let also $\Pi_d \subset \Pi_d(X)$ be a a countable set with the following property: for any $\pi \in \Pi_d(X)$ there exist $\pi_n \in \Pi_d$ such that $\pi_n \to \pi$ uniformly on bounded sets of X (the existence of Π_d follows easily by the separability of X).

We fix $t \in I$ such that $W_p(\mu_{t+h}, \mu_t)/|h| \to |\mu'|(t) = \|v_t\|_{L^p(\mu_t)}$ and

$$\lim_{h\to 0} \frac{\mu_{t+h}(\varphi) - \mu_t(\varphi)}{h} = \int_{\mathbb{R}^d} \langle \nabla\varphi, v_t\rangle \, d\mu_t \qquad \forall \varphi = \psi\circ\pi,\ \psi \in \mathscr{D}_d,\ \pi \in \Pi_d. \tag{8.4.9}$$

Since $\mathscr{D}_d$ and Π_d are countable, the metric differentiation theorem implies that both conditions are fulfilled for $\mathscr{L}^1$-a.e. $t \in I$. Let $\boldsymbol{\mu}_h \in \Gamma_o(\mu_t, \mu_{t+h})$, set

$$\boldsymbol{\nu}_h := \left(\pi^1, \frac{1}{h}(\pi^2 - \pi^1)\right)_{\#} \boldsymbol{\mu}_h,$$

and fix φ as in (8.4.9) and a limit point $\boldsymbol{\nu}_0 = \int \nu_{0x}\, d\mu_t(x)$ of $\boldsymbol{\nu}_h$ as $h \to 0$ (w.r.t. the narrow convergence). We use the identity

$$\frac{\mu_{t+h}(\varphi) - \mu_t(\varphi)}{h} = \frac{1}{h}\int_{X\times X} \varphi(y) - \varphi(x)\, d\boldsymbol{\mu}_h$$
$$= \frac{1}{h}\int_{X\times X} \varphi(x + h(y-x)) - \varphi(x)\, d\boldsymbol{\nu}_h = \int_{X\times X} \langle \nabla\varphi(x), y - x\rangle + \omega_{x,y}(h)\, d\boldsymbol{\nu}_h$$

with $\omega_{x,y}(h)$ bounded and infinitesimal as $h \to 0$, to obtain

$$\int_X \langle \nabla\varphi, v_t\rangle \, d\mu_t = \int_X \int_X \langle y, \nabla\varphi(x)\rangle \, d\nu_{0x}(y)\, d\mu_t(x).$$

Denoting by $\tilde{v}_t(x) = \int_X y \, d\nu_{0x}(y)$ the first moment of ν_{0x}, by a density argument it follows that

$$\nabla \cdot ((\tilde{v}_t - v_t)\mu_t) = 0. \tag{8.4.10}$$

We now claim that

$$\int_X \int_X |y|^p \, d\nu_{0x}(y) d\mu_t(x) \le [|\mu'|(t)]^p. \tag{8.4.11}$$

Indeed

$$\begin{aligned}\int_X \int_X |y|^p \, d\nu_{0x}(y) d\mu_t(x) &\le \liminf_{h\to 0} \int_{X\times X} |y|^p \, d\boldsymbol{\nu}_h \\ &= \liminf_{h\to 0} \frac{1}{h^p} \int_{X\times X} |y-x|^p d\boldsymbol{\mu}_h \\ &= \liminf_{h\to 0} \frac{W_p^p(\mu_{t+h},\mu_t)}{h^p} = |\mu'|^p(t).\end{aligned}$$

From (8.4.11) we obtain that

$$\|\tilde v_t\|^p_{L^p(\mu_t;X)} \le \int_X \int_X |y|^p \, d\nu_{0x} d\mu_t(x) \le [|\mu'|(t)]^p = \|v_t\|^p_{L^p(\mu_t;X)}.$$

Therefore Proposition 8.4.3 entails that $\tilde v_t = v_t$. Moreover, the first inequality above is strict if ν_{0x} is not a Dirac mass in a set of μ_t-positive measure. Therefore ν_{0x} is a Dirac mass for μ_t-a.e. x and $\boldsymbol{\nu}_0 = (\boldsymbol{i} \times v_t)_\# \mu_t$. This proves the narrow convergence of the measures in (8.4.6). Together with convergence of moments, this gives convergence in the Wasserstein metric.

Now we show (8.4.7). Let $\boldsymbol{\mu}_h = \int_X \mu_{hx} \, d\mu_t(x)$ and let us estimate the distance between μ_{t+h} and $(\boldsymbol{i} + h v_t)_\# \mu_t$ with $\pi^{2,3}_\# \left(\int \delta_{x+hv_t} \times \nu_{hx} \, d\mu_t(x)\right)$. We have then

$$\begin{aligned}\frac{W_p^p(\mu_{t+h},(\boldsymbol{i}+hv_t)_\#\mu_t)}{h^p} &\le \int_{X\times X} \frac{1}{h^p} |x + h v_t(x) - y|^p \, d\boldsymbol{\mu}_h \\ &= \int_{X\times X} |v_t(x) - y|^p \, d\boldsymbol{\nu}_h = o(1)\end{aligned}$$

because of (8.4.6).

In the case when $\mu_t \in \mathscr{P}_p^r(X)$, the identity

$$\left(\pi^1, \frac{1}{h}(\pi^2 - \pi^1)\right)_\# \boldsymbol{\mu}_h = \left(\boldsymbol{i} \times \frac{1}{h}(\boldsymbol{t}_{\mu_t}^{\mu_{t+h}} - \boldsymbol{i})\right)_\# \mu_t$$

and the weak convergence at the level of plans give that $\frac{1}{h}(\boldsymbol{t}_{\mu_t}^{\mu_{t+h}} - \boldsymbol{i})\mu_t$ narrowly converge to $v_t \mu_t$. On the other hand our choice of t ensures that the L^p norms converge to the L^p norm of the limit, therefore the convergence of the densities of these measures w.r.t. μ_t is strong in L^p. □

As an application of (8.4.7) we are now able to show the $\mathscr{L}^1$-a.e. differentiability of $t \mapsto W_p(\mu_t, \sigma)$ along absolutely continuous curves μ_t. Recall that for constant speed geodesics more precise results hold, see Chapter 7.

Theorem 8.4.7 (Generic differentiability of $W_p(\mu_t,\sigma)$). *Let $\mu_t : I \to \mathscr{P}_p(X)$ be an absolutely continuous curve, let $\sigma \in \mathscr{P}_p(X)$ and let $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ be its* *tangent vector field, characterized by Proposition* 8.4.5. *Then*

$$\frac{d}{dt}W_p^p(\mu_t,\sigma)=\int_{X^2}p|x_1-x_2|^{p-2}\langle x_1-x_2,v_t(x_1)\rangle\,d\boldsymbol{\gamma}\qquad\forall\boldsymbol{\gamma}\in\Gamma_o(\mu_t,\sigma)\quad(8.4.12)$$

for $\mathscr{L}^1$*-a.e.* $t\in I$.

Proof. We show that the stated property is true at any t where (8.4.7) holds and the derivative of $t\mapsto W_p(\mu_t,\sigma)$ exists (recall that this map is absolutely continuous). Due to (8.4.7), we know that the limit

$$L:=\lim_{h\to0}\frac{W_p^p((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma)-W_p^p(\mu_t,\sigma)}{h}$$

exists and coincides with $\frac{d}{dt}W_p^p(\mu_t,\sigma)$, and we have to show that it is equal to the left hand side in (8.4.12). Choosing any $\boldsymbol{\gamma}\in\Gamma_o(\mu_t,\sigma)$ we can use the plan $\boldsymbol{\eta}:=(\pi^1+hv_t\circ\pi^1,\pi^2)_\#\boldsymbol{\gamma}\in\Gamma((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma)$ to estimate from above $W_p^p((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma)$ as follows:

$$\begin{aligned}W_p^p((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma)&\le\int_{X^2}|x_1-x_2|^p\,d\boldsymbol{\eta}=\int_{X^2}|x_1+hv_t(x_1)-x_2|^p\,d\boldsymbol{\gamma}\\&=W_p^p(\mu_t,\sigma)+h\int_{X^2}p\langle\frac{x_1-x_2}{|x_1-x_2|^{2-p}},v_t(x_1)\rangle\,d\boldsymbol{\gamma}+o(h).\end{aligned}$$

Dividing both sides by h and taking limits as $h\downarrow0$ or $h\uparrow0$ we obtain

$$L\le\int_{X^2}p|x_1-x_2|^{p-2}\langle x_1-x_2,v_t(x_1)\rangle\,d\boldsymbol{\gamma}\le L.\qquad\square$$

The argument in the previous proof leads to the so-called superdifferentiability property of the Wasserstein distance, a theme that we will explore more in detail in Chapter 10 (see in particular Theorem 10.2.2).

Remark 8.4.8 (Derivative formula with an arbitrary velocity vector field). In fact, Proposition 8.5.4 will show that formula (8.4.12) holds for *every* Borel velocity vector field v_t satisfying the continuity equation in the distribution sense (8.3.8) and the L^p-estimate $\|v_t\|_{L^p(\mu_t;X)}\in L^1(I)$.

8.5 Tangent space and optimal maps

In this section we compare the tangent space arising from the closure of gradients of smooth cylindrical function with the tangent space built using optimal maps; the latter one is also compared in the Appendix with the geometric tangent space made with plans (see Theorem 12.4.4).

Proposition 8.4.6 suggests another possible definition of tangent cone to a measure in $\mathscr{P}_p(X)$ (see also Section 12.4 in the Appendix): for any $\mu\in\mathscr{P}_p(X)$ we define

$$\mathrm{Tan}_\mu^r\mathscr{P}_p(X):=\overline{\left\{\lambda(\boldsymbol{r}-\boldsymbol{i}):(\boldsymbol{i}\times\boldsymbol{r})_\#\mu\in\Gamma_o(\mu,\boldsymbol{r}_\#\mu),\ \lambda>0\right\}}^{L^p(\mu;X)}.\qquad(8.5.1)$$

The main result of this section shows that the two notions in fact coincide.

Theorem 8.5.1. *For any* $p \in (1, +\infty)$ *and any* $\mu \in \mathscr{P}_p(X)$ *we have*

$$\mathrm{Tan}_\mu \mathscr{P}_p(X) = \mathrm{Tan}^r_\mu \mathscr{P}_p(X).$$

We split the (not elementary) proof of this result in various steps, which are of independent interest.

The first step provides an inclusion between the tangent cones when the base measure μ is regular.

Proposition 8.5.2 (Optimal displacement maps are tangent). *If* $p \in (1, +\infty)$ *and* $\mu \in \mathscr{P}^r_p(X)$, *then* $\mathrm{Tan}^r_\mu \mathscr{P}_p(X) \subset \mathrm{Tan}_\mu \mathscr{P}_p(X)$, *i.e. for every measure* $\sigma \in \mathscr{P}_p(X)$, *if* $\boldsymbol{t}_\mu^\sigma$ *is the unique optimal transport map between* μ *and* σ *given by Theorem* 6.2.4 *and Theorem* 6.2.10, *we have* $\boldsymbol{t}_\mu^\sigma - \boldsymbol{i} \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$.

Proof. Assume first that $\mathrm{supp}\,\sigma$ is contained in $\overline{B}_R(0)$ for some $R > 0$. Theorem 6.2.4 ensures the representation $\boldsymbol{t}_\mu^\sigma - \boldsymbol{i} = j_q(\nabla\varphi)$, where φ is a locally Lipschitz and $|\cdot|^p$-concave map whose gradient $\nabla\phi = j_p(\boldsymbol{t}_\mu^\sigma - \boldsymbol{i})$ has $(p-1)$-growth (according to (5.1.21)), since $\boldsymbol{t}_\mu^\sigma$ takes its values in a bounded set.

We consider the Euclidean case $X = \mathbb{R}^d$ first and the mollified functions φ_ε. A truncation argument enabling an approximation by gradients with compact support gives that $j_q(\nabla\varphi_\varepsilon)$ belong to $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ (notice also that $\nabla\varphi_\varepsilon$ have still $(p-1)$-growth, uniformly with respect to ϵ). Due to the absolute continuity of μ it is immediate to check using the dominated convergence theorem that $j_q(\nabla\varphi_\varepsilon)$ converge to $j_q(\nabla\varphi)$ in $L^p(\mu;\mathbb{R}^d)$, therefore $j_q(\nabla\varphi) \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$ as well.

In the case when X is an infinite dimensional, separable Hilbert case we argue as follows. Let $\pi^d, (\pi^d)^*, \hat\pi^d$ be the canonical maps given by (5.1.28), (5.1.29), and (5.1.30) for an orthonormal basis $\{e_n\}_{n\ge 1}$ of X. We set

$$\mu^d := \pi^d_\#\mu,\ \nu^d := \pi^d_\#\nu \in \mathscr{P}(\mathbb{R}^d), \qquad \hat\mu^d := \hat\pi^d_\#\mu,\ \hat\nu^d := \hat\pi^d_\#\nu \in \mathscr{P}(X),$$

observing that, by (6.2.1) and (5.2.3), μ^d is absolutely continuous with respect to the d-dimensional Lebesgue measure. Therefore there exists an optimal transportation map $\boldsymbol{r}^d \in L^p(\mu^d;\mathbb{R}^d)$ defined on $\mathbb{R}^d$ such that $\boldsymbol{r}^d_\#\mu^d = \nu^d$ and $\boldsymbol{r}^d - \boldsymbol{i} = j_q(\nabla\psi^d)$ in $\mathbb{R}^d$ for some locally Lipschitz and $|\cdot|^p$-concave map $\psi^d : \mathbb{R}^d \to \mathbb{R}$. By the previous approximation argument, setting $\varphi^d := \psi^d \circ \pi^d$ and

$$\begin{aligned}\hat{\boldsymbol{r}}^d :=& (\pi^d)^* \circ (\boldsymbol{r}^d \circ \pi^d) = (\pi^d)^* \circ \big(j_q(\nabla\psi^d \circ \pi^d) + \pi^d\big)\\ =& j_q\big((\pi^d)^* \circ \nabla\psi^d \circ \pi^d\big) + (\pi^d)^* \circ \pi^d = j_q\big(\nabla\varphi^d\big) + \hat\pi^d\end{aligned}$$

(here we used the commutation property $j_q \circ (\pi^d)^* = (\pi^d)^* \circ j_q$), we get $\hat{\boldsymbol{r}}^d - \hat\pi^d \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$; moreover, being $(\pi^d)^*$ an isometry, it is immediate to check that $\hat{\boldsymbol{r}}^d$ is an optimal map pushing $\hat\mu^d$ on $\hat\nu^d$.

Letting $d \to +\infty$, since

$$\lim_{d\uparrow+\infty} \|\hat\pi^d - \boldsymbol{i}\|_{L^p(\mu;X)} = 0,$$

we conclude by applying the following Lemma.

Finally, when σ has not a bounded support, we can approximate σ in $\mathscr{P}_p(X)$ by measures σ_n with bounded support and we can apply again the following lemma. The details are left to the reader. □

Lemma 8.5.3. *Let μ, $\nu \in \mathscr{P}_p(X)$ such that $\Gamma_o(\mu,\nu) = \{(\boldsymbol{i}\times\boldsymbol{r})_\#\mu\}$ contains only an optimal transportation map $\boldsymbol{r} \in L^p(\mu;X)$, let $\boldsymbol{t}_n \in L^p(\mu;X)$ be a family of maps converging to the identity in $L^p(\mu;X)$ with $\mu_n := (\boldsymbol{t}_n)_\#\mu$, and let $\nu_n \in \mathscr{P}_p(X)$ be converging to ν as $n \to \infty$. Suppose that $\boldsymbol{r}_n \in L^p(\mu_n;X)$ is an optimal transport map from μ_n to ν_n. Then*

$$\lim_{n\to\infty} \|\boldsymbol{r}_n \circ \boldsymbol{t}_n - \boldsymbol{r}\|_{L^p(\mu,X)} = 0. \tag{8.5.2}$$

Proof. Let $\varphi : X \times X \to \mathbb{R}$ any continuous function with p-growth. Since $W_p^p(\mu_n,\mu) \to 0$, $W_p^p(\nu_n,\nu) \to 0$ as $n \to \infty$, by applying Proposition 7.1.3 and Lemma 5.1.7 we get

$$\begin{aligned} \lim_{n\to\infty} \int_X \varphi(\boldsymbol{t}_n(x), \boldsymbol{r}_n(\boldsymbol{t}_n(x)))\, d\mu(x) &= \lim_{n\to\infty} \int_X \varphi(y, \boldsymbol{r}_n(y))\, d\mu_n(y) \\ &= \int_X \varphi(y, \boldsymbol{r}(y))\, d\mu(y). \end{aligned} \tag{8.5.3}$$

Choosing $\varphi(x_1,x_2) := |x_2|^p$ we get that $\boldsymbol{r}_n \circ \boldsymbol{t}_n$ is bounded in $L^p(\mu;X)$ and its norm converges to the norm of $\boldsymbol{r}$; therefore we can assume that $\boldsymbol{r}_n \circ \boldsymbol{t}_n$ is weakly convergent to some map $\boldsymbol{s} \in L^p(\mu;X)$ and we should prove that $\boldsymbol{s} = \boldsymbol{r}$. Thus we choose $\varphi(x_1,x_2) := \zeta(x_1)\langle x_2, z\rangle$ with ζ continuous and bounded and $z \in X$: (8.5.3) yields

$$\lim_{n\to\infty} \int_X \zeta(\boldsymbol{t}_n(x))\langle z, \boldsymbol{r}_n(\boldsymbol{t}_n(x))\rangle\, d\mu(x) = \int_X \zeta(x)\langle z, \boldsymbol{r}(x)\rangle\, d\mu(x),$$

whereas weak convergence provides

$$\begin{aligned} \lim_{n\to\infty} \int_X \zeta(\boldsymbol{t}_n(x))\langle z, \boldsymbol{r}_n(\boldsymbol{t}_n(x))\rangle\, d\mu(x) &= \lim_{n\to\infty} \int_X \zeta(x)\langle z, \boldsymbol{r}_n(\boldsymbol{t}_n(x))\rangle\, d\mu(x) \\ &= \int_X \zeta(x)\langle z, \boldsymbol{s}(x)\rangle\, d\mu(x). \end{aligned}$$

It follows that $\langle z, \boldsymbol{s}(x)\rangle = \langle z, \boldsymbol{r}(x)\rangle$ for μ-a.e. $x \in X$, $\forall\, z \in X$, and therefore being X separable $\boldsymbol{s} = \boldsymbol{r}$ μ-a.e. in X. □

Proposition 8.5.4. *Let μ, $\nu \in \mathscr{P}_p(X)$ and let $\boldsymbol{\gamma} \in \Gamma_o(\mu,\nu)$. For every divergence-free vector field $w \in L^p(\mu;X)$ we have*

$$\int_{X\times X} \langle j_p(x_2 - x_1), w(x_1)\rangle\, d\boldsymbol{\gamma}(x_1,x_2) = 0. \tag{8.5.4}$$

In particular, if $\boldsymbol{r}$ *is an optimal transport map between* μ *and* $\nu = \boldsymbol{r}_{\#}\mu$ *we have*

$$\int_X \langle j_p(\boldsymbol{r}(x) - x), w(x)\rangle \, d\mu(x) = 0 \quad \forall w \in L^p(\mu; X) \ \text{s.t.} \ \nabla \cdot \big(w\mu\big) = 0. \tag{8.5.5}$$

Recalling (8.4.3) *we get that* $\boldsymbol{r} - \boldsymbol{i} \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$.

Proof. We can assume (possibly replacing $\boldsymbol{\gamma}$ by $(\pi_t^{1,1\to 2})_{\#}\boldsymbol{\gamma}$ with t close to 1) that $\boldsymbol{\gamma}$ is the unique optimal transport plan between μ and ν (see Lemma 7.2.1).

By the approximation result stated in Proposition 8.3.3 we can find finite dimensional subspaces X_h, measures $\mu_h \in \mathscr{P}_p(X)$ with support in X_h and regular restriction to X_h converging to μ in $\mathscr{P}_p(X)$, and vectors $w_h \in L^p(\mu_h; X_h)$ such that $\nabla \cdot (w_h \mu_h) = 0$, $(\boldsymbol{i} \times w_h)_{\#}\mu_h \to (\boldsymbol{i} \times w)_{\#}\mu$ in $\mathscr{P}_p(X^2)$. Denoting by $\boldsymbol{t}_h$ the unique optimal transport map between μ_h and $\nu_h := \hat{\pi}^h_{\#}\nu$ (as usual, $\hat{\pi}^h$ is the orthogonal projection of X onto X_h and we identify μ_h and ν_h with their restriction to X_h), we know by Proposition 8.5.2 that $\boldsymbol{t}_h - \boldsymbol{i} \in \mathrm{Tan}^r_{\mu_h}\mathscr{P}_p(X_h)$, and therefore

$$\int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), w_h\rangle \, d\mu_h = 0 \qquad \forall h \in \mathbb{N}.$$

Moreover, the uniqueness of $\boldsymbol{\gamma}$ yields that the transport plans $(\boldsymbol{i} \times \boldsymbol{t}_h)_{\#}\mu_h$ narrowly converge in $\mathscr{P}(X \times X)$ to $\boldsymbol{\gamma}$. Since the marginals of the plans converge in $\mathscr{P}_p(X)$ we have also that the plans are uniformly p-integrable, therefore

$$\begin{aligned}\lim_{h\to\infty} \int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), \tilde{w}\rangle \, d\mu_h &= \lim_{h\to\infty} \int_{X\times X} \langle j_p(x_2 - x_1), \tilde{w}(x_1)\rangle \, d(\boldsymbol{i} \times \boldsymbol{t}_h)_{\#}\mu_h \\ &= \int_{X\times X} \langle j_p(x_2 - x_1), \tilde{w}(x_1)\rangle \, d\boldsymbol{\gamma}\end{aligned}$$

for any continuous function $\tilde{w}$ with linear growth. By Proposition 8.3.3 again (with $f(x_1, x_2) = |x_2 - \tilde{w}(x_1)|^p$) we know that

$$\lim_{\tilde{w} \in C^0_b(X),\ \tilde{w} \to w \text{ in } L^p(\mu;X)} \ \limsup_{h\to\infty} \int_X |w_h - \tilde{w}|^p \, d\mu_h = 0. \tag{8.5.6}$$

Since

$$0 = \int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), \tilde{w}\rangle \, d\mu + \int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), w_h - \tilde{w}\rangle \, d\mu_h \qquad \text{for any } \tilde{w} \in C^0_b(X),$$

passing to the limit as $h \to \infty$ and using Hölder inequality we obtain

$$\left| \int_X \langle j_p(x_2 - x_1), \tilde{w}(x_1)\rangle \, d\boldsymbol{\gamma} \right| \le \sup_h \|\boldsymbol{t}_h - \boldsymbol{i}\|^{1/q}_{L^p(\mu_h;X)} \limsup_{h\to\infty} \|w_h - \tilde{w}\|_{L^p(\mu_h;X)}.$$

Taking into account (8.5.6) we conclude that $\int_X \langle j_p(x_2 - x_1), w(x_1)\rangle \, d\boldsymbol{\gamma} = 0$. □

The above proposition shows that for general measures $\mu \in \mathscr{P}_p(X)$

$$\mathrm{Tan}^r_\mu \mathscr{P}_p(X) \subset \mathrm{Tan}_\mu \mathscr{P}_p(X). \tag{8.5.7}$$

Now we want to prove the opposite inclusion: let us first mention that the case $p = 2$ is particularly simple.

Corollary 8.5.5. *For any $\mu \in \mathscr{P}_2(X)$ we have* $\mathrm{Tan}_\mu \mathscr{P}_2(X) = \mathrm{Tan}^r_\mu \mathscr{P}_2(X)$.

Proof. We should only check the inclusion $\subset$: if $\varphi \in \mathrm{Cyl}(X)$ it is always possible to choose $\lambda > 0$ such that $x \mapsto \frac{1}{2}|x|^2 + \lambda^{-1}\phi(x)$ is convex. Therefore $\boldsymbol{r} := \boldsymbol{i} + \lambda^{-1}\nabla\varphi$ is cyclically monotone, thus an optimal map between μ and $\boldsymbol{r}_\# \mu$; by (8.5.1) we obtain that $\nabla\phi = \lambda(\boldsymbol{r} - \boldsymbol{i})$ belongs to $\mathrm{Tan}^r_\mu \mathscr{P}_2(X)$. □

In the general case $p \in (1, +\infty)$ the desired inclusion follows by the following characterization:

Proposition 8.5.6. *Let $\mu \in \mathscr{P}_p(X)$, $v \in L^p(\mu; X)$, and $\mu_\varepsilon := (\boldsymbol{i} + \varepsilon v)_\# \mu$ for $\varepsilon > 0$. If $v \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$ then*

$$\lim_{\varepsilon \downarrow 0} \frac{W_p(\mu, \mu_\varepsilon)}{\varepsilon} = \|v\|_{L^p(\mu;X)}, \tag{8.5.8}$$

and denoting by $\boldsymbol{\gamma}_\varepsilon \in \Gamma_o(\mu, \mu_\varepsilon)$ a family of optimal plans, we have

$$\lim_{\varepsilon \downarrow 0} \int_{X \times X} \left| \frac{x_2 - x_1 - \varepsilon v(x_1)}{\varepsilon} \right|^p d\boldsymbol{\gamma}_\varepsilon(x_1, x_2) = 0. \tag{8.5.9}$$

Proof. Let us consider the rescaled plans

$$\boldsymbol{\mu}_\varepsilon := \big(\pi^1, \varepsilon^{-1}(\pi^2 - \pi^1)\big)_\# \boldsymbol{\gamma}_\varepsilon \quad \text{for } \boldsymbol{\gamma}_\varepsilon \in \Gamma_o(\mu, \mu_\varepsilon), \tag{8.5.10}$$

observing that

$$\pi^1_\# \boldsymbol{\mu}_\varepsilon = \mu, \quad \int_{X^2} |x_2|^p \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2) = \frac{W_p^p(\mu, \mu_\varepsilon)}{\varepsilon^p} \le \|v\|^p_{L^p(\mu;X)}, \tag{8.5.11}$$

$$\int_{X \times X} \left| \frac{x_2 - x_1 - \varepsilon v(x_1)}{\varepsilon} \right|^p d\boldsymbol{\gamma}_\varepsilon(x_1, x_2) = \int_{X \times X} |x_2 - v(x_1)|^p \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2).$$

For every vanishing sequence $\varepsilon_k \to 0$ we can find a subsequence (still denoted by ε_k) and a limit plan $\boldsymbol{\mu}$ such that $\boldsymbol{\mu}_{\varepsilon_k}$ is narrowly converging to $\boldsymbol{\mu}$ in $\mathscr{P}(X \times X_\varpi)$. In particular, for every smooth cylindrical function $\zeta \in \mathrm{Cyl}(X)$ we have

$$\begin{aligned}
\varepsilon^{-1} \int_X \Big(\zeta(x + \varepsilon v(x)) - \zeta(x)\Big) \, d\mu(x) &= \varepsilon^{-1} \Big(\int_X \zeta(x_2) \, d\mu_\varepsilon(x_2) - \int_X \zeta(x_1) \, d\mu(x_1) \Big) \\
= \int_{X \times X} \frac{\zeta(x_2) - \zeta(x_1)}{\varepsilon} \, d\boldsymbol{\gamma}_\varepsilon(x_1, x_2) &= \int_{X \times X} \frac{\zeta(x_1 + \varepsilon x_2) - \zeta(x_1)}{\varepsilon} \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2) \\
= \int_0^1 \int_{X \times X} \langle \nabla\zeta(x_1 + \varepsilon t x_2), x_2 \rangle \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2) \, dt
\end{aligned} \tag{8.5.12}$$

and

$$\varepsilon^{-1}\int_X \Big(\zeta(x+\varepsilon v(x))-\zeta(x)\Big)\,d\mu(x)=\int_0^1\int_X \langle\nabla\zeta(x+t\varepsilon v(x)),v(x)\rangle\,d\mu(x)\,dt. \tag{8.5.13}$$

Choosing $\varepsilon=\varepsilon_k$ in (8.5.12) and in (8.5.13) and passing to the limit as $k\to\infty$, a repeated application of Lebesgue dominated convergence theorem yields

$$\begin{aligned}\int_X \langle\nabla\zeta(x),v(x)\rangle\,d\mu(x) \qquad\qquad (8.5.14)\\ =\lim_{k\to\infty}\int_0^1\int_X \langle\nabla\zeta(x+t\varepsilon_k v(x)),v(x)\rangle\,d\mu(x)\,dt\\ =\lim_{k\to\infty}\int_0^1\int_{X\times X} \langle\nabla\zeta(x_1+t\varepsilon_k x_2),x_2\rangle\,d\boldsymbol{\mu}_{\varepsilon_k}(x_1,x_2)\,dt\\ =\int_{X\times X}\langle\nabla\zeta(x_1),x_2\rangle\,d\boldsymbol{\mu}(x_1,x_2). \qquad\qquad (8.5.15)\end{aligned}$$

It follows that the limit plan $\boldsymbol{\mu}$ satisfies

$$\int_{X\times X}\langle\nabla\zeta(x_1),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}(x_1,x_2)=0\quad\forall\,\zeta\in \mathrm{Cyl}(X), \tag{8.5.16}$$

and the same relation holds if $\nabla\zeta$ is replaced by any element ξ of the "cotangent space" $\mathrm{CoTan}_\mu\mathscr{P}_p(X)$ (i.e. the closure in $L^q(\mu;X)$ of the gradient vector fields) introduced by (8.4.5).

If $v\in\mathrm{Tan}_\mu\mathscr{P}_p(X)$ and $p\geq 2$, by the p-inequality (10.2.4), we can find a suitable vanishing subsequence $\varepsilon_k\to 0$ and a limit plan $\boldsymbol{\mu}$ such that

$$\begin{aligned}0&\leq c_p\limsup_{\varepsilon\to 0}\int_{X\times X}|x_2-v(x_1)|^p\,d\boldsymbol{\mu}_\varepsilon(x_1,x_2)\\ &\leq\limsup_{\varepsilon\to 0}\int_{X\times X}|x_2|^p-|v(x_1)|^p-p\langle j_p(v(x_1)),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}_\varepsilon(x_1,x_2)\\ &=\lim_{k\to\infty}\frac{W_p^p(\mu,\mu_{\varepsilon_k})}{\varepsilon_k^p}-\|v\|^p_{L^p(\mu;X)}-\int_{X\times X}p\langle j_p(v(x_1)),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}_{\varepsilon_k}(x_1,x_2)\\ &\leq-\int_{X\times X}p\langle j_p(v(x_1)),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}(x_1,x_2)=0\end{aligned}$$

by (8.5.11) and (8.5.16), since $v\in\mathrm{Tan}_\mu\mathscr{P}_p(X)$ is equivalent to $j_p(v)\in\mathrm{CoTan}_\mu\mathscr{P}_p(X)$. The case $p<2$ is completely analogous. □

When μ is regular, the opposite inclusion

$$\mathrm{Tan}_\mu\mathscr{P}_p(X)\subset\mathrm{Tan}^r_\mu\mathscr{P}_p(X),$$

which completes the proof of Theorem 8.5.1, follows easily from the previous proposition: keeping the same notation, we know that $\boldsymbol{\gamma}_\varepsilon$ is induced by an optimal transport map $\boldsymbol{r}_\varepsilon$ so that $\varepsilon^{-1}(\boldsymbol{r}_\varepsilon - \boldsymbol{i}) \in \mathrm{Tan}^r_\mu \mathscr{P}_p(X)$ and (8.5.9) yields

$$\lim_{\varepsilon \to 0} \int_X \left| \frac{\boldsymbol{r}_\varepsilon(x) - x}{\varepsilon} - v(x) \right|^p d\mu(x) = 0. \tag{8.5.17}$$

Therefore v belongs to $\mathrm{Tan}^r_\mu \mathscr{P}_p(X)$.

In the general case, by disintegrating $\boldsymbol{\gamma}_\varepsilon$ with respect to the first variable x_1, a measurable selection theorem [39] allows us to select $\boldsymbol{r}_\varepsilon(x_1)$ such that $\boldsymbol{r}_\varepsilon(x_1) \in \mathrm{supp}\,(\boldsymbol{\gamma}_\varepsilon)_{x_1}$ and

$$\left| \frac{\boldsymbol{r}_\varepsilon(x_1) - x_1}{\varepsilon} - v(x_1) \right|^p \leq 2 \int_X \left| \frac{\boldsymbol{r}_\varepsilon(y) - y}{\varepsilon} - v(y) \right|^p d(\boldsymbol{\gamma}_\varepsilon)_{x_1}(y).$$

Then, since the graph of $\boldsymbol{r}_\varepsilon$ is contained in the support of $\boldsymbol{\gamma}_\varepsilon$, we obtain that $\boldsymbol{r}_\varepsilon$ is $|\cdot|^p$-monotone (so that $\varepsilon^{-1}(\boldsymbol{r}_\varepsilon - \boldsymbol{i}) \in \mathrm{Tan}^r_\mu \mathscr{P}_p(X)$) and (8.5.17) still holds.

Chapter 9

Convex Functionals in $\mathscr{P}_p(X)$

The importance of geodesically convex functionals in Wasserstein spaces was firstly pointed out by McCann [97], who introduced the three basic examples we will discuss in detail in 9.3.1, 9.3.4, 9.3.6. His original motivation was to prove the uniqueness of the minimizer of an energy functional which results from the sum of the above three contributions.

Applications of this idea have been given to (im)prove many deep functional (Brunn-Minkowski, Gaussian, (logarithmic) Sobolev, Isoperimetric, etc.) inequalities: we refer to Villani's book [126, Chap. 6] (see also the survey [72]) for a detailed account on this topic. Connections with evolution equations have also been exploited [103, 107, 108, 1, 38], mainly to study the asymptotic decay of the solution to the equilibrium.

From our point of view, convexity is a crucial tool to study the well posedness and the basic regularity properties of gradient flows, as we showed in Chapters 2 and 4. Thus in this chapter we discuss the basic notions and properties related to this concept: the first part of Section 9.1 is devoted to fixing the notion of convexity along geodesics in $\mathscr{P}_p(X)$, avoiding any unnecessary restriction to regular measures; a useful tool for the subsequent developments is the stability of convexity with respect to Γ-convergence, a well known property in the more usual linear theory.

Unfortunately, Example 9.1.5 shows that the squared 2-Wasserstein distance is not convex along geodesics in $\mathscr{P}_2(X)$: this fact and the theory of Chapter 4 motivate the investigation (of convexity properties) along different interpolating curves, along which the squared 2-Wasserstein distance exhibits a nicer behavior; the second part of Section 9.1 discusses this question and introduces the notion of generalized geodesics. Lemma 9.2.7, though simple, provides a crucial link with the metric theory of Chapter 4.

Section 9.3 discusses in great generality the main examples of geodesically convex functionals, showing that they all satisfy also the stronger convexity along generalized geodesics. The last example is related to the semiconcavity properties of the squared 2-Wasserstein distance, discussed in Theorem 7.3.2.

In the last section we give a closer look to the convexity properties of general Relative Entropy functionals, showing that they are strictly related to the log-concavity of the reference measures. Here we use the full generality of the theory, proving all the significant results even in infinite dimensional Hilbert spaces.

9.1 λ-geodesically convex functionals in $\mathscr{P}_p(X)$

In McCann's approach, functionals are naturally defined on $\mathscr{P}_2^r(\mathbb{R}^d)$ so that for each couple of measures μ^1, $\mu^2 \in \mathscr{P}_2^r(\mathbb{R}^d)$ a unique *optimal transport* map $\boldsymbol{t} = \boldsymbol{t}_{\mu^1}^{\mu^2}$ (see (7.1.4)) always exists: in his terminology, a functional $\phi : \mathscr{P}_2^r(\mathbb{R}^d) \to (-\infty, +\infty]$ is *displacement convex* if

$$
\begin{aligned}
&\text{setting} \quad \mu_t^{1\to 2} := \big(\boldsymbol{i} + t(\boldsymbol{t} - \boldsymbol{i})\big)_{\#}\mu^1, \quad \boldsymbol{t} = \boldsymbol{t}_{\mu^1}^{\mu^2},\\
&\text{the map} \quad t \in [0,1] \mapsto \phi\big(\mu_t^{1\to 2}\big) \quad \text{is convex,} \quad \forall\, \mu^1,\, \mu^2 \in \mathscr{P}_2^r(\mathbb{R}^d).
\end{aligned}
\tag{9.1.1}
$$

In Section 7.2 we have seen that the curve $\mu_t^{1\to 2}$ is the constant speed geodesic connecting μ^1 to μ^2; therefore the following definition seems natural, when we consider functionals whose domain contains general probability measures.

Definition 9.1.1 (λ-convexity along geodesics). *Let X be a separable Hilbert space and let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$. Given $\lambda \in \mathbb{R}$, we say that ϕ is* λ-geodesically convex *in $\mathscr{P}_p(X)$ if for every couple μ^1, $\mu^2 \in \mathscr{P}_p(X)$* there exists *an optimal transfer plan $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$ such that*

$$
\phi(\mu_t^{1\to 2}) \le (1-t)\phi(\mu^1) + t\phi(\mu^2) - \frac{\lambda}{2} t(1-t) W_p^2(\mu^1, \mu^2) \quad \forall\, t \in [0,1], \tag{9.1.2}
$$

where $\mu_t^{1\to 2} = (\pi_t^{1\to 2})_{\#}\boldsymbol{\mu} = \big((1-t)\pi^1 + t\pi^2\big)_{\#}\boldsymbol{\mu}$ is defined as in (7.2.2), π^1, π^2 being the projections onto the first and the second coordinate in X^2, respectively.

Notice that this notion of convexity *depends* on the summability exponent p.

Remark 9.1.2 (The map $t \mapsto \phi(\mu_t^{1\to 2})$ is λ-convex). Actually this definition of λ-convexity expressed through (9.1.2) implies that

$$
\text{the map } t \in [0,1] \mapsto \phi(\mu_t^{1\to 2}) \text{ is } \lambda W_p^2(\mu^1, \mu^2)\text{-convex}, \tag{9.1.3}
$$

thus recovering an (apparently) stronger and more traditional form.
This equivalence follows easily by the fact, proved in Section 7.2, that for $t_1 < t_2$ in $[0,1]$ with $\{t_1, t_2\} \neq \{0,1\}$ the plan $\big(\pi_{t_1}^{1\to 2} \times \pi_{t_2}^{1\to 2}\big)_{\#}\boldsymbol{\mu}$ is the *unique* element of $\Gamma_o(\mu_{t_1}^{1\to 2}, \mu_{t_2}^{1\to 2})$.

Notice that in Definition 9.1.1 we *do not* require (9.1.2) along *all* the optimal plans of $\Gamma_o(\mu^1, \mu^2)$. One of the advantage of this technical point is provided by the following proposition, which will be useful to check convexity in many examples.

Proposition 9.1.3 (Convexity criterion). *Let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be a l.s.c. map such that for any $\mu \in D(\phi)$ there exists $(\mu_h) \subset \mathscr{P}^r_p(X)$ converging to μ in $\mathscr{P}_p(X)$ with $\phi(\mu_h) \to \phi(\mu)$.*
Then ϕ is λ-geodesically convex iff for each $\mu \in D(\phi) \cap \mathscr{P}^r_p(X)$ and for each μ-essentially injective map $\boldsymbol{r} \in L^p(\mu; X)$ whose graph is $|\cdot|^p$-cyclically monotone the map $t \mapsto \phi\left(((1-t)\boldsymbol{i} + t\boldsymbol{r})_{\#}\mu\right)$ is λ-convex in $[0,1]$.

Proof. If $\mu^1 \in \mathscr{P}^r_p(X)$ and $\boldsymbol{r} \in L^p(\mu^1; X)$ is $|\cdot|^p$-cyclically monotone, then $\big((1-t)\boldsymbol{i} + t\boldsymbol{r}\big)_{\#}\mu^1$ is the unique geodesic joining μ^1 to $\mu^2 := \boldsymbol{r}_{\#}\mu^1$. This shows the necessity of the condition.

In order to show its sufficiency, we notice that if μ^1, $\mu^2 \in \mathscr{P}^r_p(X)$ then a unique optimal map $\boldsymbol{t}_{\mu^1}^{\mu^2}$ exists, it belongs to $L^p(\mu^1; X)$ and it is μ^1-essentially injective (by Remark 6.2.11). Therefore the convexity inequality (9.1.2) holds when the initial and final measure are regular. The general case can be recovered through a standard approximation and compactness argument, as in the proof of the next lemma. □

The following natural Γ-convergence result is well known for convex functionals in linear spaces, see for instance Chapter 11 in [50].

Lemma 9.1.4 (Convexity and Γ-convergence). *Let $\phi_h : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be λ-geodesically convex functionals which $\Gamma\big(\mathscr{P}_p(X)\big)$-converge to ϕ as $n \to \infty$, i.e.*

$$\mu_h \to \mu \text{ in } \mathscr{P}_p(X) \quad \Rightarrow \quad \liminf_{h\to\infty} \phi_h(\mu_h) \geq \phi(\mu), \tag{9.1.4}$$

$$\forall\, \mu \in \mathscr{P}_p(X) \quad \exists \mu_h \to \mu \text{ in } \mathscr{P}_p(X): \quad \lim_{h\to\infty} \phi_h(\mu_h) = \phi(\mu). \tag{9.1.5}$$

Then ϕ is λ-geodesically convex.
The same result holds for the $\Gamma\big(\mathscr{P}(X)\big)$-convergence if $\lambda \geq 0$, i.e. if we replace convergence in $\mathscr{P}_p(X)$ with narrow convergence *in $\mathscr{P}(X)$ (thus without assuming the convergence of the p-moments of μ_h) in* (9.1.4), (9.1.5).

Proof. Let us fix $\mu^1, \mu^2 \in D(\phi)$; by (9.1.5) we can find sequences μ^1_h, μ^2_h converging to μ^1, μ^2 in $\mathscr{P}_p(X)$ such that

$$\lim_{n\to\infty} \phi_h(\mu^1_h) = \phi(\mu^1), \quad \lim_{n\to\infty} \phi_h(\mu^2_h) = \phi(\mu^2).$$

Let $\boldsymbol{\mu}_h \in \Gamma_o(\mu^1_h, \mu^2_h)$ be an optimal plan such that (5.1.19) holds for ϕ_h; by Lemma 5.2.2 the sequence $(\boldsymbol{\mu}_h)$ is tight (resp. uniformly p-integrable), because the sequences of their marginals are tight (resp. uniformly p-integrable). Therefore, by Proposition 7.1.5 we can extract a suitable subsequence (still denoted by $\boldsymbol{\mu}_h$) converging to $\boldsymbol{\mu}$ in $\mathscr{P}_p(X \times X)$: we want to show that ϕ is λ-convex along the interpolation $\mu^{1\to 2}_t$ induced by $\boldsymbol{\mu}$.

Since $(\mu_h)_t^{1\to 2} \to \mu_t^{1\to 2}$ in $\mathscr{P}_p(X)$ as $h \to \infty$, (9.1.4) yields easily

$$\begin{aligned}\phi(\mu_t^{1\to 2}) &\le \liminf_{h\to\infty} \phi_h\big((\mu_h)_t^{1\to 2}\big)\\ &\le \liminf_{h\to\infty} \Big((1-t)\phi(\mu_h^1) + t\phi(\mu_h^2) - \frac{\lambda}{2}t(1-t)W_p^2(\mu_h^1,\mu_h^2)\Big)\\ &= (1-t)\phi(\mu^1) + t\phi(\mu^2) - \frac{\lambda}{2}t(1-t)W_p^2(\mu^1,\mu^2). \end{aligned}\tag{9.1.6}$$

In the case of narrow convergence, we can follow the same argument; (9.1.6) becomes an inequality, thanks to (7.1.11), if $\lambda \ge 0$. □

λ-convexity of functionals along geodesics is the simplest condition which allows us to apply the theory developed in Section 2.4. The semigroup generation results of Chapter 4 involve the stronger 1-convexity property of the distance function $W_2^2(\mu^1,\cdot)$ from an arbitrary base point μ^1.

In the 1*-dimensional case* we already know by Theorem 6.0.2 and (7.2.8) that $\mathscr{P}_2(\mathbb{R}^1)$ is isometrically isomorphic to a closed convex subset of an Hilbert space: precisely the space of nondecreasing functions in $(0,1)$ (the inverses of distribution functions), viewed as a subset of $L^2(0,1)$. Thus the 2-Wasserstein distance in $\mathbb{R}$ satisfies the generalized parallelogram rule

$$\begin{aligned} W_2^2(\mu^1,\mu_t^{2\to 3}) = (1-t)W_2^2(\mu^1,\mu^2) + tW_2^2(\mu^1,\mu^3) - t(1-t)W_2^2(\mu^2,\mu^3)\\ \forall\, t\in[0,1],\quad \mu^1,\mu^2,\mu^3 \in \mathscr{P}_2(\mathbb{R}^1). \end{aligned}\tag{9.1.7}$$

If the space X has dimension ≥ 2 the following example shows that there is no constant λ such that $W_2^2(\cdot,\mu^1)$ is λ-convex along geodesics. We will see in the next subsection how to circumvent this difficulty.

Example 9.1.5 (The distance function is not λ-convex along geodesics). Let

$$\mu^2 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(2,1)}\right), \qquad \mu^3 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(-2,1)}\right).$$

Using for instance Theorem 6.0.1 it is easy to check that the unique optimal map $\boldsymbol{r}$ pushing μ^2 to μ^3 maps $(0,0)$ in $(-2,1)$ and $(2,1)$ in $(0,0)$, therefore there is a unique constant speed geodesic joining the two measures, given by

$$\mu_t^{2\to 3} := \frac{1}{2}\left(\delta_{(-2t,t)} + \delta_{(2-2t,1-t)}\right) \qquad t\in[0,1].$$

Choosing $\mu^1 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(0,-2)}\right)$, there are two maps $\boldsymbol{r}_t$, $\boldsymbol{s}_t$ pushing μ^1 to $\mu_t^{2\to 3}$, given by

$$\begin{aligned} \boldsymbol{r}_t(0,0) = (-2t,t),\quad \boldsymbol{r}_t(0,-2) = (2-2t,1-t),\\ \boldsymbol{s}_t(0,0) = (2-2t,1-t),\quad \boldsymbol{s}_t(0,-2) = (-2t,t). \end{aligned}$$

Therefore

$$W_2^2(\mu_t^{2\to 3},\mu^1) = \min\left\{5t^2 - 7t + \frac{13}{2}, 5t^2 - 3t + \frac{9}{2}\right\}$$

has a concave cusp at $t = 1/2$ and therefore is not λ-convex along the geodesic $\mu_t^{2\to 3}$ for any $\lambda \in \mathbb{R}$.

9.2 Convexity along generalized geodesics

In dimension greater than 1, Example 9.1.5 shows that the squared Wasserstein distance functional $\mu \mapsto W_2^2(\mu^1, \mu)$ is not 1-convex along geodesics (in fact, Theorem 7.3.2 shows that it satisfies the opposite inequality).

On the other hand, the theory developed in Chapter 4 indicates that 1-convexity of the squared distance is a quite essential property and that we can exploit the flexibilty in the choice of the connecting curve, along which 1-convexity should be checked. Therefore, here we are looking for such kind of curves (in the case of the "Hilbertian-like" 2-Wasserstein distance) and for the related concept of convexity for functionals.

Let us first suppose that the reference measure μ^1 is regular, i.e. $\mu^1 \in \mathscr{P}_2^r(X)$ and let μ^2, μ^3 be given in $\mathscr{P}_2(X)$; we can find two optimal transport maps $\boldsymbol{t}^2 = \boldsymbol{t}_{\mu^1}^{\mu^2}$, $\boldsymbol{t}^3 = \boldsymbol{t}_{\mu^1}^{\mu^3}$ as in (7.1.4) such that

$$W_2^2(\mu^1, \mu^i) = \int_X |\boldsymbol{t}^i(x) - x|^2 \, d\mu^1(x), \quad i = 2, 3. \tag{9.2.1}$$

Equation (9.2.1) reduces the evaluation of the Wasserstein distance to an integral with respect to the fixed measure μ^1: it is therefore quite natural to interpolate between μ^2 and μ^3 by using $\boldsymbol{t}^2$ and $\boldsymbol{t}^3$, i.e. setting

$$\mu_t^{2\to 3} = (\boldsymbol{t}_t^{2\to 3})_\# \mu^1 \quad \text{where} \quad \boldsymbol{t}_t^{2\to 3} := (1-t)\boldsymbol{t}^2 + t\,\boldsymbol{t}^3, \quad t \in [0,1]. \tag{9.2.2}$$

Since $\boldsymbol{t}_t^{2\to 3}$ is obviously cyclically monotone, we have

$$W_2^2(\mu^1, \mu_t^{2\to 3}) = \int_X |\boldsymbol{t}_t^{2\to 3}(x) - x|^2 \, d\mu^1(x) = \int_X |(1-t)\boldsymbol{t}^2(x) + t\boldsymbol{t}^3(x) - x|^2 \, d\mu^1(x),$$

and therefore an easy calculation shows

$$\begin{aligned} W_2^2(\mu^1, \mu_t^{2\to 3}) &= (1-t)\int_X |\boldsymbol{t}^2(x) - x|^2 \, d\mu^1(x) + t\int_X |\boldsymbol{t}^3(x) - x|^2 \, d\mu^1(x) \\ &- t(1-t)\int_X |\boldsymbol{t}^2(x) - \boldsymbol{t}^3(x)|^2 \, d\mu^1(x) \qquad (9.2.3) \\ &\le (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) - t(1-t)W_2^2(\mu^2, \mu^3), \end{aligned}$$

since

$$\int_X |\boldsymbol{t}^2(x) - \boldsymbol{t}^3(x)|^2 \, d\mu^1(x) \ge W_2^2(\mu^2, \mu^3).$$

This calculation shows that $\frac{1}{2}W_2^2(\mu^1, \cdot)$ is 1-convex along the new interpolating curve $\mu_t^{2\to 3}$ given by (9.2.2).

When μ^1 is not regular, we have to substitute the optimal maps $\boldsymbol{t}_{\mu}^{\mu^2}, \boldsymbol{t}_{\mu^1}^{\mu^3}$ with optimal plans $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$, $\boldsymbol{\mu}^{1\,3} \in \Gamma_o(\mu^1, \mu^3)$: in order to interpolate between them, we shall also introduce a 3-plan

$$\begin{gathered}\boldsymbol{\mu} \in \mathscr{P}_2(X^3) \quad \text{such that} \quad \pi^{1\,2}_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{1\,2},\ \pi^{1\,3}_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{1\,3} \quad \text{and we set} \\ \mu_t^{2\to3} := (\pi_t^{2\to3})_\# \boldsymbol{\mu}, \quad \text{where} \quad \pi_t^{2\to3} := (1-t)\pi^2 + t\pi^3.\end{gathered} \tag{9.2.4}$$

Recalling that in (7.3.2) we set

$$W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) := \int_{X^3} |x_3 - x_2|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) \geq W_2^2(\mu^2, \mu^3), \tag{9.2.5}$$

we have

Lemma 9.2.1. *Let μ^1, μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and let*

$$\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3) \quad \textit{such that} \quad \boldsymbol{\mu}^{1,i} = \pi^{1\,i}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^i), \quad i = 2, 3. \tag{9.2.6}$$

Then, defining $\mu_t^{2\to3}$ as in (9.2.4), we get

$$W_2^2(\mu^1, \mu_t^{2\to3}) = \int_{X^3} |(1-t)x_2 + tx_3 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) \tag{9.2.7a}$$

$$= (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) - t(1-t)W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) \tag{9.2.7b}$$

$$\leq (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) - t(1-t)W_2^2(\mu^2, \mu^3). \tag{9.2.7c}$$

The inequality (9.2.7c) implies that $\frac{1}{2}W_2^2(\mu^1, \cdot)$ is 1-convex along the curve $\mu_t^{2\to3}$.

Proof. We argue as for (9.2.3), by introducing the transfer plan

$$\boldsymbol{\mu}_t^{1,2\to3} := \big((1-t)\pi^{1,2} + t\pi^{1,3}\big)_\# \boldsymbol{\mu} \in \Gamma(\mu^1, \mu_t^{2\to3});$$

by the definition of the Wasserstein distance and the Hilbertian identity (12.3.3) it is immediate to see that

$$W_2^2(\mu^1, \mu_t^{2\to3}) \leq \int_{X\times X} |y_1 - y_2|^2 \, d\boldsymbol{\mu}_t^{1,2\to3}(y_1, y_2) \tag{9.2.8}$$

$$\begin{aligned}&= \int_{X^3} |(1-t)x_2 + tx_3 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) \\ &= \int_{X^3} \Big((1-t)|x_2 - x_1|^2 + t|x_3 - x_1|^2 - t(1-t)|x_2 - x_3|^2\Big) \, d\boldsymbol{\mu}(x_1, x_2, x_3).\end{aligned} \tag{9.2.9}$$

(9.2.9) yields (9.2.7b) since by (9.2.6) we have

$$\int_{X^3} |x_2 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) = \int_{X^2} |x_2 - x_1|^2 \, d\boldsymbol{\mu}^{1\,2}(x_1, x_2) = W_2^2(\mu^1, \mu^2),$$

$$\int_{X^3} |x_3 - x_1|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3) = \int_{X^2} |x_3 - x_1|^2 \, d\boldsymbol{\mu}^{1\,3}(x_1, x_3) = W_2^2(\mu^1, \mu^3);$$

(9.2.7c) follows directly from the inequality (9.2.5).

Moreover, it is possible to see that (9.2.8) is in fact an equality, i.e. $\boldsymbol{\mu}_t^{1,2\to 3} \in \Gamma_o(\mu^1, \mu_t^{2\to 3})$, by checking that the support of $\boldsymbol{\mu}_t^{1,2\to 3}$ is cyclically monotone; by the density property (5.2.6), we can simply check that $\pi_t^{1,2\to 3}\big(\operatorname{supp}\boldsymbol{\mu}\big)$ is cyclically monotone. We choose points $(a_i, b_i) \in \pi_t^{1,2\to 3}\big(\operatorname{supp}\boldsymbol{\mu}\big)$, $i = 1, \ldots, N$ and set $(a_0, b_0) := (a_N, b_N)$; we thus find points b_i', b_i'' such that

$$(a_i, b_i') \in \operatorname{supp}\boldsymbol{\mu}^{1,2}, \quad (a_i, b_i'') \in \operatorname{supp}\boldsymbol{\mu}^{1,3}, \quad b_i = (1-t)b_i' + tb_i''.$$

Therefore the cyclical monotonicity of $\operatorname{supp}\boldsymbol{\mu}^{1,i}$ gives

$$\begin{aligned}\sum_{i=1}^N \langle a_i - a_{i-1}, b_i\rangle &= \sum_{i=1}^N \langle a_i - a_{i-1}, (1-t)b_i' + tb_i''\rangle \\ &= (1-t)\sum_{i=1}^N \langle a_i - a_{i-1}, b_i'\rangle + t\sum_{i=1}^N \langle a_i - a_{i-1}, b_i''\rangle \geq 0. \quad \square\end{aligned}$$

Taking account of Lemma 9.2.1, we introduce the following definitions.

Definition 9.2.2 (Generalized geodesics). *A "generalized geodesic" joining μ^2 to μ^3 (with base μ^1) is a curve of the type*

$$\mu_t^{2\to 3} = (\pi_t^{2\to 3})_\# \boldsymbol{\mu} \qquad t \in [0,1],$$

where

$$\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3) \quad \textit{and} \quad \pi^{1,2}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2), \quad \pi^{1,3}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^3). \tag{9.2.10}$$

Remark 9.2.3. Remember that if $\mu^1 \in \mathscr{P}_2^r(X)$ then by Lemma 5.3.2 and Theorem 6.2.10 there exists a unique generalized geodesic connecting μ^2 to μ^3 with base μ^1, since there exists a unique plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ satisfying the optimality condition $\pi^{1,i}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^i)$, $i = 2, 3$. In fact, denoting by $\boldsymbol{t}^i$ the optimal maps $\boldsymbol{t}_{\mu^1}^{\mu^i}$ pushing μ^1 to μ^i, $i = 2, 3$, $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu} := (\boldsymbol{i} \times \boldsymbol{t}^2 \times \boldsymbol{t}^3)_\# \mu^1. \tag{9.2.11}$$

We thus recover the expression $\mu_t^{2\to 3} = \big((1-t)\boldsymbol{t}^2 + t\,\boldsymbol{t}^3\big)_\# \mu^1$ given by (9.2.2).

Definition 9.2.4 (Convexity along generalized geodesics). *Given $\lambda \in \mathbb{R}$, we say that ϕ is λ-convex along generalized geodesics if for* any *μ^1, μ^2, $\mu^3 \in D(\phi)$ there exists a generalized geodesic $\mu_t^{2\to 3}$ induced by a plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ satisfying (9.2.10) such that*

$$\phi(\mu_t^{2\to 3}) \leq (1-t)\phi(\mu^2) + t\phi(\mu^3) - \frac{\lambda}{2} t(1-t) W^2_{\boldsymbol{\mu}}(\mu^2, \mu^3) \quad \forall t \in [0,1], \tag{9.2.12}$$

where $W^2_{\boldsymbol{\mu}}(\cdot,\cdot)$ is defined in (9.2.5).

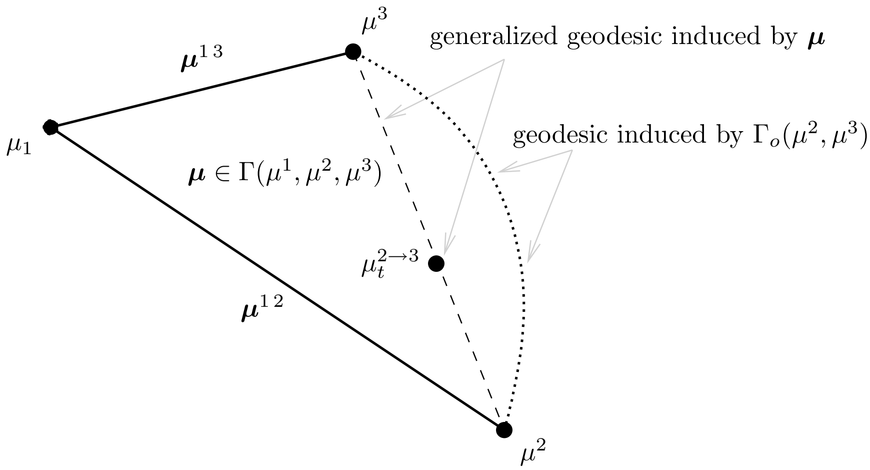


Figure 9.1: Generalized geodesics

Remark 9.2.5 (The case of optimal transport maps). If ϕ is convex along *any* interpolating curve $\mu_t^{2\to 3}$ induced by $\boldsymbol{\mu} \in \Gamma(\mu^2, \mu^3)$, then ϕ is trivially convex along generalized geodesics.

Remark 9.2.6. When $\lambda \neq 0$ Definition 9.2.4 slightly differs from the analogous metric Definition 2.4.1 in the modulus of convexity, since

$$W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) \geq W_2^2(\mu^2, \mu^3). \tag{9.2.13}$$

In particular, when $\lambda > 0$ this condition is stronger than 2.4.1, whereas for $\lambda < 0$ (9.2.12) is weaker. The next lemma motivates this choice.

Lemma 9.2.7 ($(\tau^{-1} + \lambda)$-convexity of $\Phi(\tau, \mu^1; \cdot)$). *Suppose that* $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ *is a proper functional which is* λ*-convex along generalized geodesics for some* $\lambda \in \mathbb{R}$. *Then for each* $\mu^1 \in D(\phi)$ *and* $0 < \tau < \frac{1}{\lambda^-}$ *the functional*

$$\Phi(\tau, \mu^1; \mu) := \frac{1}{2\tau} W_2^2(\mu^1, \mu) + \phi(\mu) \quad \textit{satisfies the convexity Assumption } 4.0.1.$$

Proof. We consider a plan $\boldsymbol{\mu}$ satisfying (9.2.10) and we combine (9.2.7b) and (9.2.12) and use (9.2.13) to obtain

$$\begin{aligned}\Phi(\tau, \mu^1; \mu_t^{2\to 3}) &\leq (1-t)\Phi(\tau, \mu^1; \mu^2) + t\Phi(\tau, \mu^1; \mu^3) - \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big) W_{\boldsymbol{\mu}}^2(\mu^2, \mu^3) \\ &\leq (1-t)\Phi(\tau, \mu^1; \mu^2) + t\Phi(\tau, \mu^1; \mu^3) - \frac{1}{2}\Big(\frac{1}{\tau} + \lambda\Big) W_2^2(\mu^2, \mu^3)\end{aligned}$$

whenever $\tau^{-1} > -\lambda$. □

Remark 9.2.8 (Comparison between the two notions of convexity). If ϕ is λ-convex on generalized geodesics then it is also λ-geodesically convex according to Definition 9.1.1: it is sufficient to notice if we choose $\mu^1 = \mu^3$, then any $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ such that $\pi^{1,3}_\#\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^3)$ is of the form of the form

$$\boldsymbol{\mu} = \int_{X^2} \delta_{x_1}(x_3)\, d\boldsymbol{\mu}^{1\,2}(x_1, x_2) \quad \text{where} \quad \boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2).$$

Therefore, if we impose also that $\boldsymbol{\mu}^{1\,2} = \pi^{1,2}_\#\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$, then $\mu^{2\to 3}_t$ is the canonical geodesic interpolation $\big(t\pi^1 + (1-t)\pi^2\big)_\# \boldsymbol{\mu}^{1\,2}$.
We already know by Example 9.1.5 that $\frac{1}{2}W_2(\cdot, \mu^1)$ *is not λ-convex* along geodesics, and therefore *is not λ-convex* along generalized geodesics. On the other hand, if we choose generalized geodesics with base point μ^1 as in (9.2.10), then $\frac{1}{2}W_2^2(\cdot, \mu^1)$ is *indeed* $1-$convex along these curves by Lemma 9.2.1. As Lemma 9.2.7 shows, this property is the key point to apply the theory of Chapter 4.

For λ-convex functionals on generalized geodesics we present now two properties which are analogous to the ones stated in Lemma 9.1.4 and Proposition 9.1.3. We omit the proofs, which are similar to the previous ones.

Lemma 9.2.9 (Convexity along generalized geodesics and Γ-convergence). *Let $\phi_h : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be λ-convex on generalized geodesics. If ϕ_h $\Gamma\big(\mathscr{P}_2(X)\big)$-converge to ϕ as $h \to \infty$ as in* (9.1.4), (9.1.5), *then ϕ is λ-convex on generalized geodesics. If $\lambda \geq 0$ the same result holds for $\Gamma\big(\mathscr{P}(X)\big)$-convergence, i.e. Γ-convergence with respect to the* narrow topology *of $\mathscr{P}(X)$.*

Proposition 9.2.10 (A criterion for convexity along generalized geodesics). *Let $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be a l.s.c. map such that for any $\mu \in D(\phi)$ there exist $(\mu_h) \subset \mathscr{P}^r_2(X)$ converging to μ with $\phi(\mu_h) \to \phi(\mu)$.*

Then ϕ is λ-convex on generalized geodesics iff for every $\mu \in \mathscr{P}^r_2(X)$ and for every couple of μ-essentially injective maps $\boldsymbol{r}^0$, $\boldsymbol{r}^1 \in L^2(\mu; X)$ whose graph is cyclically monotone we have

$$\begin{aligned}\phi\Big(\big((1-t)\boldsymbol{r}^0 + t\boldsymbol{r}^1\big)_\#\mu\Big) &\leq (1-t)\phi\big(\boldsymbol{r}^0_\#\mu\big) + t\phi\big(\boldsymbol{r}^1_\#\mu\big) \\ &\quad - \frac{\lambda}{2}t(1-t)\int_X |\boldsymbol{r}^0(x) - \boldsymbol{r}^1(x)|^2\, d\mu(x) \quad \forall\, t \in [0,1].\end{aligned} \tag{9.2.14}$$

9.3 Examples of convex functionals in $\mathscr{P}_p(X)$

In this section we introduce the main classes of geodesically convex functionals.

Example 9.3.1 (Potential energy). Let $V : X \to (-\infty, +\infty]$ be a proper, lower semicontinuous function whose negative part has a p-growth, i.e.

$$V(x) \geq -A - B|x|^p \quad \forall\, x \in X \quad \text{for some} \quad A, B \in \mathbb{R}. \tag{9.3.1}$$

In $\mathscr{P}_p(X)$ we define

$$\mathcal{V}(\mu) := \int_X V(x)\, d\mu(x). \tag{9.3.2}$$

Evaluating $\mathcal{V}$ on Dirac's masses we check that $\mathcal{V}$ is proper; since V^- is uniformly integrable w.r.t. any sequence (μ_n) converging in $\mathscr{P}_p(X)$ (see Proposition 7.1.5), Lemma 5.1.7 shows that $\mathcal{V}$ is lower semicontinuous in $\mathscr{P}_p(X)$. If V is bounded from below we have even, thanks to (5.1.15), lower semicontinuity w.r.t. narrow convergence.

Recall that for functionals defined on a Hilbert space, λ-convexity means

$$V((1-t)x_1 + tx_2) \le (1-t)V(x_1) + tV(x_2) - \frac{\lambda}{2}t(1-t)|x_1 - x_2|^2 \quad \forall\, x_1, x_2 \in X. \tag{9.3.3}$$

Proposition 9.3.2 (Convexity of $\mathcal{V}$). *If V is λ-convex then for every μ^1, $\mu^2 \in D(\mathcal{V})$ and $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ we have*

$$\mathcal{V}(\mu_t^{1\to 2}) \le (1-t)\mathcal{V}(\mu^1) + t\mathcal{V}(\mu^2) - \frac{\lambda}{2}t(1-t)\int_{X^2} |x_1 - x_2|^2\, d\boldsymbol{\mu}(x_1, x_2). \tag{9.3.4}$$

In particular:

(i) *If $\boldsymbol{p = 2}$ then the functional $\mathcal{V}$ is λ-convex on generalized geodesics, according to Definition* 9.2.4 *(in fact it is λ-convex along any interpolating curve, cf. Remark* 9.2.5*).*

(ii) *If* $(\boldsymbol{p \le 2,\ \lambda \ge 0})$ *or* $(\boldsymbol{p \ge 2,\ \lambda \le 0})$ *then $\mathcal{V}$ is λ-geodesically convex in $\mathscr{P}_p(X)$.*

Proof. Since V is bounded from below by a continuous affine functional (if $\lambda \ge 0$) or by a quadratic function (if $\lambda < 0$) its negative part satisfies (9.3.1) for the corresponding values of p considered in this lemma; therefore Definition (9.3.2) makes sense.

Integrating (9.3.3) along any admissible transport plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ with $\mu^1, \mu^2 \in D(\mathcal{V})$ we obtain (9.3.4), since

$$\begin{aligned}
\mathcal{V}(\mu_t^{1\to 2}) &= \int_{X^2} V((1-t)x_1 + tx_2)\, d\boldsymbol{\mu}(x_1, x_2) \\
&\le \int_{X^2} \Big((1-t)V(x_1) + tV(x_2) - \frac{\lambda}{2}t(1-t)|x_1 - x_2|^2\Big)\, d\boldsymbol{\mu}(x_1, x_2) \\
&= (1-t)\mathcal{V}(\mu^1) + t\mathcal{V}(\mu^2) - \frac{\lambda}{2}t(1-t)\int_{X^2} |x_1 - x_2|^2\, d\boldsymbol{\mu}(x_1, x_2).
\end{aligned}$$

When $p = 2$ we obtain (9.2.12). When $p \neq 2$ we choose $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$: for $p > 2$ we use the inequality

$$\int_{X^2} |x_1 - x_2|^2\, d\boldsymbol{\mu}(x_1, x_2) \le \Big(\int_{X^2} |x_1 - x_2|^p\, d\boldsymbol{\mu}(x_1, x_2)\Big)^{2/p} = W_p^2(\mu^1, \mu^2),$$

whereas, for $p < 2$, we use the reverse one

$$\int_{X^2} |x_1 - x_2|^2 \, d\boldsymbol{\mu}(x_1, x_2) \geq \Big(\int_{X^2} |x_1 - x_2|^p \, d\boldsymbol{\mu}(x_1, x_2) \Big)^{2/p} = W_p^2(\mu^1, \mu^2).$$

□

Remark 9.3.3. Since $\mathcal{V}(\delta_x) = V(x)$, it is easy to check that the conditions on V are also necessary for the validity of the previous proposition.

Example 9.3.4 (Interaction energy). Let us fix an integer $k > 1$ and let us consider a lower semicontinuous function $W : X^k \to (-\infty, +\infty]$, whose negative part satisfies the usual p-growth condition. Denoting by $\mu^{\times k}$ the measure $\mu \times \mu \times \cdots \times \mu$ on X^k, we set

$$\mathcal{W}_k(\mu) := \int_{X^k} W(x_1, x_2, \ldots, x_k) \, d\mu^{\times k}(x_1, x_2, \ldots, x_k). \tag{9.3.5}$$

If

$$\exists\, x \in X : W(x, x, \ldots, x) < +\infty, \tag{9.3.6}$$

then $\mathcal{W}_k$ is proper; its lower semicontinuity follows from the fact that

$$\mu_n \to \mu \quad \text{in } \mathscr{P}_p(X) \quad \Longrightarrow \quad \mu_n^{\times k} \to \mu^{\times k} \quad \text{in } \mathscr{P}_p(X^k). \tag{9.3.7}$$

Here the typical example is $k = 2$ and $W(x_1, x_2) := \tilde{W}(x_1 - x_2)$ for some $\tilde{W} : X \to (-\infty, +\infty]$ with $\tilde{W}(0) < +\infty$.

Proposition 9.3.5 (Convexity of $\mathcal{W}$). *If W is convex then the functional $\mathcal{W}_k$ is convex along any interpolating curve $\mu_t^{1\to 2}$, $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$, in $\mathscr{P}_p(X)$ (cf. Remark* 9.2.5*).*

Proof. Observe that $\mathcal{W}_k$ is the restriction to the subset

$$\mathscr{P}_p^{\times}(X^k) := \Big\{ \mu^{\times k} : \mu \in \mathscr{P}_p(X) \Big\}$$

of the potential energy functional $\mathcal{W}$ on $\mathscr{P}_p(X^k)$ given by

$$\mathcal{W}(\boldsymbol{\mu}) := \int_{X^k} W(x_1, \ldots, x_k) \, d\boldsymbol{\mu}(x_1, \ldots, x_k).$$

We consider the linear permutation of coordinates $P : (X^2)^k \to (X^k)^2$ defined by

$$P\Big((x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\Big) := \Big((x_1, \ldots x_k), (y_1, \ldots y_k)\Big).$$

If $\boldsymbol{\mu} \in \Gamma(\mu_1, \mu_2)$ then it is easy to check that $P_\# \boldsymbol{\mu}^{\times k} \in \Gamma(\mu_1^{\times k}, \mu_2^{\times k}) \subset \mathscr{P}((X^k)^2)$ and

$$(\pi_t^{1\to 2})_\# P_\# (\boldsymbol{\mu}^{\times k}) = P_\# \Big((\pi_t^{1\to 2})_\# \boldsymbol{\mu} \Big)^{\times k}.$$

Therefore all the convexity properties for $\mathcal{W}_k$ follow from the corresponding ones of $\mathcal{W}$. □

In the next example we limit us to consider the finite dimensional case $X := \mathbb{R}^d$, since the Lebesgue measure $\mathscr{L}^d$ will play a distinguished role.

Example 9.3.6 (Internal energy). Let $F : [0, +\infty) \to (-\infty, +\infty]$ be a proper, lower semicontinuous convex function such that

$$F(0) = 0, \quad \liminf_{s \downarrow 0} \frac{F(s)}{s^\alpha} > -\infty \quad \text{for some } \alpha > \frac{d}{d+p}. \tag{9.3.8}$$

We consider the functional $\mathcal{F} : \mathscr{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ defined by

$$\mathcal{F}(\mu) := \begin{cases} \int_{\mathbb{R}^d} F(\rho(x))\, d\mathscr{L}^d(x) & \text{if } \mu = \rho \cdot \mathscr{L}^d \in \mathscr{P}_p^r(\mathbb{R}^d), \\ +\infty & \text{otherwise,} \end{cases} \tag{9.3.9}$$

and its relaxed envelope $\mathcal{F}^*$ defined as

$$\mathcal{F}^*(\mu) := \inf \Big\{ \liminf_{n \to +\infty} \mathcal{F}(\mu_n) : \mu_n \to \mu \quad \text{in } \mathscr{P}_p(\mathbb{R}^d) \Big\}. \tag{9.3.10}$$

Remark 9.3.7 (The meaning of condition (9.3.8)). Condition (9.3.8) simply guarantees that the negative part of $F(\mu)$ is integrable in $\mathbb{R}^d$. For, let us observe that there exist nonnegative constants c_1, c_2 such that the negative part of F satisfies

$$F^-(s) \le c_1 s + c_2 s^\alpha \quad \forall\, s \in [0, +\infty),$$

and it is not restrictive to suppose $\alpha \le 1$. Since $\mu = \rho\, \mathscr{L}^d \in \mathscr{P}_p(\mathbb{R}^d)$ and $\frac{\alpha p}{1-\alpha} > d$ we have

$$\begin{aligned} &\int_{\mathbb{R}^d} \rho^\alpha(x)\, d\mathscr{L}^d(x) = \int_{\mathbb{R}^d} \rho^\alpha(x)(1+|x|)^{\alpha p}(1+|x|)^{-\alpha p}\, d\mathscr{L}^d(x) \\ &\le \Big(\int_{\mathbb{R}^d} \rho(x)(1+|x|)^p\, d\mathscr{L}^d(x) \Big)^\alpha \Big(\int_{\mathbb{R}^d} (1+|x|)^{-\alpha p/(1-\alpha)}\, d\mathscr{L}^d(x) \Big)^{1-\alpha} < +\infty \end{aligned}$$

and therefore $F^-(\rho) \in L^1(\mathbb{R}^d)$.

Remark 9.3.8 (Lower semicontinuity of $\mathcal{F}$). General results on integral functionals [11] show that [79, 31] $\mathcal{F}^* = \mathcal{F}$ on $\mathscr{P}_p^r(\mathbb{R}^d)$ and that $\mathcal{F}^* = \mathcal{F}$ on the whole of $\mathscr{P}_p(\mathbb{R}^d)$ if F has a superlinear growth at infinity.

Proposition 9.3.9 (Convexity of $\mathcal{F}$). *If*

$$\textit{the map} \quad s \mapsto s^d F(s^{-d}) \quad \textit{is convex and non increasing in } (0, +\infty), \tag{9.3.11}$$

then the functionals $\mathcal{F}$, $\mathcal{F}^*$ are convex along (generalized, if $p = 2$) geodesics in $\mathscr{P}_p(\mathbb{R}^d)$.

Proof. By Proposition 9.1.3 we can limit us to check the geodesic convexity of $\mathcal{F}$: thus we consider two regular measures $\mu^i = \rho^i \mathscr{L}^d \in D(\mathcal{F}) \subset \mathscr{P}_p^r(\mathbb{R}^d)$, $i = 1, 2$, and the optimal transport map $\boldsymbol{r}$ for the p-Wasserstein distance W_p such that

$\boldsymbol{r}_{\#}\mu^1 = \mu^2$. Setting $\boldsymbol{r}_t := (1-t)\boldsymbol{i} + t\boldsymbol{r}$, by Theorem 7.2.2 we know that $\boldsymbol{r}_t$ is an optimal transport map between μ^1 and $\mu_t := r_{t\#}\mu^1$ for any $t \in [0,1]$, and Lemma 7.2.1 (for $t \in [0,1)$) and the assumption $\mu^2 \in \mathscr{P}^r_p(\mathbb{R}^d)$ (for $t = 1$) show that $(\boldsymbol{i} \times \boldsymbol{r}_t)_{\#}\mu^1 = (\boldsymbol{s}_t \times \boldsymbol{i})_{\#}\mu_t$ for some optimal transport map $\boldsymbol{s}_t$, therefore $\boldsymbol{s}_t \circ \boldsymbol{r}_t = \boldsymbol{i}$ μ^1-a.e. in $\mathbb{R}^d$. This proves that $\boldsymbol{r}_t$ is μ^1-essentially injective for any $t \in [0,1]$.

By Theorem 6.2.7 we know that $\boldsymbol{r}$ is approximately differentiable μ^1-a.e. and $\tilde{\nabla}\boldsymbol{r}$ is diagonalizable with nonnegative eigenvalues; since μ^2 is regular, by Lemma 5.5.3 $\det \tilde{\nabla}\boldsymbol{r}(x) > 0$ for μ^1-a.e. $x \in \mathbb{R}^d$. Therefore $\tilde{\nabla}\boldsymbol{r}_t$ is diagonalizable, too, with strictly positive eigenvalues: applying Lemma 5.5.3 again we get $\mu_t^{1\to 2} := (\boldsymbol{r}_t)_{\#}\mu^1 \in \mathscr{P}^r_p(\mathbb{R}^d)$ and

$$\mu_t^{1\to 2} = \rho_t \mathscr{L}^d \quad \text{with} \quad \rho_t(\boldsymbol{r}_t(x)) = \frac{\rho^1(x)}{\det \tilde{\nabla}\boldsymbol{r}_t(x)} \quad \text{for } \mu^1\text{-a.e. } x \in \mathbb{R}^d.$$

By (5.5.3) it follows that

$$\mathcal{F}(\mu_t) = \int_{\mathbb{R}^d} F(\rho_t(y))\,dy = \int_{\mathbb{R}^d} F\Big(\frac{\rho(x)}{\det \tilde{\nabla}\boldsymbol{r}_t(x)}\Big) \det \tilde{\nabla}\boldsymbol{r}_t(x)\,dx.$$

Since for a diagonalizable map D with nonnegative eigenvalues

$$t \mapsto \det((1-t)I + tD)^{1/d} \quad \text{is concave in } [0,1], \tag{9.3.12}$$

the integrand above may be seen as the composition of the convex and non-increasing map $s \mapsto s^d F(\rho(x)/s^d)$ and of the concave map in (9.3.12), so that the resulting map is convex in $[0,1]$ for μ^1-a.e. $x \in \mathbb{R}^d$. Thus we have

$$F\Big(\frac{\rho^1(x)}{\det \tilde{\nabla}\boldsymbol{r}_t(x)}\Big) \det \tilde{\nabla}\boldsymbol{r}_t(x) \le (1-t)F(\rho^1(x)) + tF(\rho^2(x))$$

and the thesis follows by integrating this inequality in $\mathbb{R}^d$.

In order to check the convexity along generalized geodesics in the case $p = 2$, we apply Proposition 9.2.10: we have to choose $\mu \in \mathscr{P}^r_2(X)$ and two optimal transport maps $\boldsymbol{r}^0, \boldsymbol{r}^1 \in L^2(\mu; X)$, setting $\boldsymbol{r}^t := (1-t)\boldsymbol{r}^0 + t\boldsymbol{r}^1$. We know that $\boldsymbol{r}^0$, $\boldsymbol{r}^1$ are approximately differentiable, μ-essentially injective, and that $\tilde{\nabla}\boldsymbol{r}^0$, $\tilde{\nabla}\boldsymbol{r}^1$ are *symmetric* (since $p = 2$) and strictly positive definite for μ-a.e. $x \in \mathbb{R}^d$; moreover, by applying (6.2.9) to $\boldsymbol{r}^0$ and $\boldsymbol{r}^1$ we get

$$\langle \boldsymbol{r}^t(x) - \boldsymbol{r}^t(y), x - y\rangle = (1-t)\langle \boldsymbol{r}^0(x) - \boldsymbol{r}^0(y), x - y\rangle + t\langle \boldsymbol{r}^1(x) - \boldsymbol{r}^1(y), x - y\rangle > 0$$

for x, $y \in \mathbb{R}^d \setminus N$, for a suitable μ-negligible subset N of $\mathbb{R}^d$. It follows that $\boldsymbol{r}^t$ are μ-essentially injective as well and we can argue as before by exploiting the symmetry of $\tilde{\nabla}\boldsymbol{r}^0$, $\tilde{\nabla}\boldsymbol{r}^1$, obtaining

$$\mathcal{F}(\mu^t) \le (1-t)\mathcal{F}(\mu_0) + t\mathcal{F}(\mu_1) \quad \text{for } \mu^t := (\boldsymbol{r}^t)_{\#}\mu. \qquad \square$$

In order to express (9.3.11) in a different way, we introduce the function

$$L_F(z) := zF'(z) - F(z) \quad \text{which satisfies} \ - L_F(e^{-z})e^z = \frac{d}{dz}F(e^{-z})e^z; \quad (9.3.13)$$

denoting by $\hat F$ the modified function $F(e^{-z})e^z$ we have the simple relation

$$\begin{gathered}\hat L_F(z) = -\frac{d}{dz}\hat F(z), \quad \widehat{L^2_F}(z) = -\frac{d}{dz}\hat L_F(z) = \frac{d^2}{dz^2}\hat F(z), \quad \text{where}\\ L^2_F(z) := L_{L_F}(z) = zL'_F(z) - L_F(z).\end{gathered} \quad (9.3.14)$$

The nonincreasing part of condition (9.3.11) is equivalent to say that

$$L_F(z) \geq 0 \quad \forall\, z \in (0,+\infty), \quad (9.3.15)$$

and it is in fact implied by the convexity of F. A simple computation in the case $F \in C^2(0,+\infty)$ shows

$$\frac{d^2}{ds^2}F(s^{-d})s^d = \frac{d^2}{ds^2}\hat F(d\cdot\log s) = \hat L^2_F(d\cdot\log s)\frac{d^2}{s^2} + \hat L_F(d\cdot\log s)\frac{d}{s^2},$$

and therefore

$$(9.3.11) \text{ is equivalent to } L^2_F(z) \geq -\frac{1}{d}L_F(z) \quad \forall\, z \in (0,+\infty), \quad (9.3.16)$$

i.e.

$$zL'_F(z) \geq \big(1-\frac{1}{d}\big)L_F(z), \quad \text{the map } z \mapsto z^{1/d-1}L_F(z) \text{ is non increasing.} \quad (9.3.17)$$

Observe that the bigger is the dimension d, the stronger are the above conditions, which always imply the convexity of F.

Remark 9.3.10 (A "dimension free" condition). The weakest condition on F yielding the geodesic convexity of $\mathcal{F}$ in *any dimension* is therefore

$$L^2_F(z) = zL'_F(z) - L_F(z) \geq 0 \quad \forall\, z \in (0,+\infty). \quad (9.3.18)$$

Taking into account (9.3.14), this is also equivalent to ask that

$$\textit{the map} \quad s \mapsto F(e^{-s})e^s \quad \textit{is convex and non increasing in } (0,+\infty). \quad (9.3.19)$$

Among the functionals F satisfying (9.3.11) we quote:

$$\text{the entropy functional:} \quad F(s) = s\log s, \quad (9.3.20)$$

$$\text{the power functional:} \quad F(s) = \frac{1}{m-1}s^m \quad \text{for } m \geq 1-\frac{1}{d}. \quad (9.3.21)$$

Observe that (9.3.20) and (9.3.21) with $m > 1$ also satisfy (9.3.19) and $\mathcal{F} = \mathcal{F}^*$, by Remark 9.3.8; on the other hands, if $m < 1$, $\mathcal{F}^*$ is given by [79, 31]

$$\mathcal{F}^*(\mu) := \frac{1}{m-1}\int_{\mathbb{R}^d} F(\rho(x))\,d\mathscr{L}^d(x) \quad \text{with } \mu = \rho\cdot\mathscr{L}^d + \mu_s,\ \mu_s \perp \mathscr{L}^d. \tag{9.3.22}$$

In this case the functional takes only account of the density of the absolutely continuous part of μ w.r.t. $\mathscr{L}^d$ and the domain of $\mathcal{F}^*$ is the whole $\mathscr{P}_p(\mathbb{R}^d)$, which strictly contains $\mathscr{P}_p^r(\mathbb{R}^d)$.

Example 9.3.11 (The opposite Wasserstein distance). In the separable Hilbert space X let us fix a base measure $\mu^1 \in \mathscr{P}_2(X)$ and let us consider the functional

$$\phi(\mu) := -\frac{1}{2}W_2^2(\mu^1,\mu). \tag{9.3.23}$$

Proposition 9.3.12. *For each couple μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and each transfer plan $\boldsymbol{\mu}^{2\,3} \in \Gamma(\mu^2,\mu^3)$ we have*

$$\begin{aligned} W_2^2(\mu^1,\mu_t^{2\to3}) \ge{}& (1-t)W_2^2(\mu^1,\mu^2) + tW_2^2(\mu^1,\mu^3)\\ &- t(1-t)\int_{X^2} |x_2-x_3|^2\,d\boldsymbol{\mu}^{2\,3}(x_2,x_3) \quad \forall\, t\in[0,1]. \end{aligned} \tag{9.3.24}$$

In particular, by Remark 9.2.5, the map $\phi : \mu \mapsto -\frac{1}{2}W_2^2(\mu^1,\mu)$ is (-1)-convex along generalized geodesics.

Proof. We argue as in Theorem 7.3.2: by Proposition 7.3.1, for μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and $\boldsymbol{\mu}^{2\,3} \in \Gamma(\mu^2,\mu^3)$ we can find a plan $\boldsymbol{\mu} \in \Gamma(\mu^1,\mu^2,\mu^3)$ such that

$$(\pi_t^{1,2\to3})_\#\boldsymbol{\mu} \in \Gamma_o(\mu^1,\mu_t^{2\to3}), \quad (\pi^{2,3})_\#\boldsymbol{\mu} = \boldsymbol{\mu}^{2\,3}. \tag{9.3.25}$$

Therefore

$$\begin{aligned} W_2^2(\mu^1,\mu_t^{2\to3}) &= \int_{X^3} |(1-t)x_2 + tx_3 - x_1|^2\,d\boldsymbol{\mu}(x_1,x_2,x_3)\\ &= \int_{X^3}\Big((1-t)|x_2-x_1|^2 + t|x_3-x_1|^2 - t(1-t)|x_2-x_3|^2\Big)\,d\boldsymbol{\mu}(x_1,x_2,x_3)\\ &\ge (1-t)W_2^2(\mu^1,\mu^2) + tW_2^2(\mu^1,\mu^3) - t(1-t)\int_{X^2}|x_2-x_3|^2\,d\boldsymbol{\mu}^{2\,3}(x_2,x_3). \end{aligned}$$

□

9.4 Relative entropy and convex functionals of measures

In this section we study in detail the case of relative entropies, which extend even to infinite dimensional spaces the example (9.3.20) discussed in 9.3.6: for more details and developments we refer to [67].

Definition 9.4.1 (Relative entropy). *Let γ, μ be Borel probability measures on a separable Hilbert space X; the* relative entropy *of μ w.r.t. γ is*

$$\mathcal{H}(\mu|\gamma) := \begin{cases} \displaystyle\int_X \frac{d\mu}{d\gamma} \log\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise.} \end{cases} \tag{9.4.1}$$

As in Example 9.3.6 we introduce the nonnegative, l.s.c., (extended) real, (strictly) convex function

$$H(s) := \begin{cases} s(\log s - 1) + 1 & \text{if } s > 0, \\ 1 & \text{if } s = 0, \\ +\infty & \text{if } s < 0, \end{cases} \tag{9.4.2}$$

and we observe that

$$\mathcal{H}(\mu|\gamma) = \int_X H\Big(\frac{d\mu}{d\gamma}\Big)\, d\gamma \ge 0; \quad \mathcal{H}(\mu|\gamma) = 0 \quad \Leftrightarrow \quad \mu = \gamma. \tag{9.4.3}$$

Remark 9.4.2 (Changing γ). Let γ be a Borel measure on X and let $V : X \to (-\infty, +\infty]$ a Borel map such that

$$V^+ \quad \text{has } p\text{-growth (5.1.21)}, \quad \tilde\gamma := e^{-V} \cdot \gamma \quad \text{is a probability measure.} \tag{9.4.4}$$

Then for measures in $\mathscr{P}_p(X)$ the relative entropy w.r.t. γ is well defined by the formula

$$\mathcal{H}(\mu|\gamma) := \mathcal{H}(\mu|\tilde\gamma) - \int_X V(x)\, d\mu(x) \in (-\infty, +\infty] \quad \forall\, \mu \in \mathscr{P}_p(X). \tag{9.4.5}$$

In particular, when $X = \mathbb{R}^d$ and γ is the d-dimensional Lebesgue measure, we find the standard entropy functional introduced in (9.3.20).

More generally, we can consider a

$$\begin{gathered} \text{proper, l.s.c., convex function } F : [0, +\infty) \to [0, +\infty] \\ \text{with superlinear growth} \end{gathered} \tag{9.4.6}$$

and the related functional

$$\mathcal{F}(\mu|\gamma) := \begin{cases} \displaystyle\int_X F\Big(\frac{d\mu}{d\gamma}\Big)\, d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise.} \end{cases} \tag{9.4.7}$$

Lemma 9.4.3 (Joint lower semicontinuity). *Let γ^n, $\mu^n \in \mathscr{P}(X)$ be two sequences narrowly converging to γ, μ in $\mathscr{P}(X_\varpi)$. Then*

$$\liminf_{n\to\infty} \mathcal{H}(\mu^n|\gamma^n) \ge \mathcal{H}(\mu|\gamma), \quad \liminf_{n\to\infty} \mathcal{F}(\mu^n|\gamma^n) \ge \mathcal{F}(\mu|\gamma). \tag{9.4.8}$$

The proof of this lemma follows easily from the next representation formula; before stating it, we need to introduce the conjugate function of F

$$F^*(s^*) := \sup_{s\geq 0} s\cdot s^* - F(s) < +\infty \quad \forall\, s^* \in \mathbb{R}, \tag{9.4.9}$$

so that

$$F(s) = \sup_{s^*\in\mathbb{R}} s^*\cdot s - F^*(s^*); \tag{9.4.10}$$

if $s_0 \geq 0$ is a minimizer of F then

$$F^*(s^*) \geq s^* s_0 - F(s_0), \quad s\geq s_0 \quad \Rightarrow \quad F(s) = \sup_{s^*\geq 0} s^*\cdot s - F^*(s^*). \tag{9.4.11}$$

In the case of the entropy functional, we have $H^*(s^*) = e^{s^*} - 1$.

Lemma 9.4.4 (Duality formula). *For any γ, $\mu \in \mathscr{P}(X)$ we have*

$$\mathcal{F}(\mu|\gamma) = \sup\Big\{ \int_X S^*(x)\, d\mu(x) - \int_X F^*(S^*(x))\, d\gamma(x) : S^* \in C^0_b(X_\varpi)\Big\}. \tag{9.4.12}$$

Proof. This lemma is a particular case of more general results on convex integrals of measures, well known in the case of a finite dimensional space X, see for instance §2.6 of [11]. We present here a brief sketch of the proof for a general Hilbert space; up to an addition of a constant, we can always assume $F^*(0) = -\min_{s\geq 0} F(s) = -F(s_0) = 0$.

Let us denote by $\mathcal{F}'(\mu|\gamma)$ the right hand side of (9.4.12). It is obvious that $\mathcal{F}'(\mu|\gamma) \leq \mathcal{H}(\mu|\gamma)$, so that we have to prove only the converse inequality.

First of all we show that $\mathcal{F}'(\mu|\gamma) < +\infty$ yields that $\mu \ll \gamma$. For let us fix s^*, $\varepsilon > 0$ and a Borel set A with $\gamma(A) \leq \varepsilon/2$. Since μ, γ are tight measures (recall that $\mathscr{B}(X) = \mathscr{B}(X_\varpi)$, compact subset of X are compact in X_ϖ, too, and X_ϖ is a separable metric space) we can find a compact set $K \subset A$, an open set (in X_ϖ) $G \supset A$ and a continuous function $\zeta : X_\varpi \to [0, s^*]$ such that

$$\mu(G\setminus K) \leq \varepsilon, \quad \gamma(G) \leq \varepsilon, \quad \zeta(x) = s^* \quad \text{on } K, \quad \zeta(x) = 0 \quad \text{on } X\setminus G.$$

Since F^* is increasing (by Definition (9.4.9)) and $F^*(0) = 0$, we have

$$\begin{aligned} s^*\mu(K) - F^*(s^*)\varepsilon &\leq \int_K \zeta(x)\, d\mu(x) - \int_G F^*(\zeta(x))\, d\gamma(x) \\ &\leq \int_X \zeta(x)\, d\mu(x) - \int_X F^*(\zeta(x))\, d\gamma(x) \leq \mathcal{F}'(\mu|\gamma) \end{aligned}$$

Taking the supremum w.r.t. $K\subset A$ and $s^* \geq 0$, and using (9.4.11) we get

$$\varepsilon F\big(\mu(A)/\varepsilon\big) \leq \mathcal{F}'(\mu|\gamma) \quad \text{if } \mu(A) \geq \varepsilon s_0.$$

Since $F(s)$ has a superlinear growth as $s \to +\infty$, we conclude that $\mu(A) \to 0$ as $\varepsilon \downarrow 0$.

Now we can suppose that $\mu = \rho \cdot \gamma$ for some Borel function $\rho \in L^1(\gamma)$, so that

$$\mathcal{F}'(\mu|\gamma) = \sup\Big\{ \int_X \big(S^*(x)\rho(x) - F^*(S^*(x))\big)\, d\gamma(x) : S^* \in C^0_b(X_\varpi)\Big\}$$

and, for a suitable dense countable set $C = \{s^*_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$

$$\begin{aligned}\mathcal{F}(\mu|\gamma) &= \int_X \sup_{s^*\in C} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x)\\ &= \lim_{k\to\infty} \int_X \sup_{s^*\in C_k} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x)\end{aligned}$$

where $C_k = \{s^*_1, \cdots, s^*_k\}$. Our thesis follows if we show that for every k

$$\int_X \max_{s^*\in C_k} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x) \le \mathcal{F}'(\mu|\gamma). \tag{9.4.13}$$

For we call

$$A_j = \Big\{x \in X : s^*_j\rho(x) - F^*(s^*_j) \ge s^*_i\rho(x) - F^*(s^*_i) \quad \forall\, i \in \{1, \ldots, k\}\Big\},$$

and

$$A'_1 = A_1, \quad A'_{j+1} = A_{j+1} \setminus \Big(\bigcup_{i=1}^{j} A_i\Big).$$

Since γ is Radon, we find compact sets $K_j \subset A'_j$, X_ϖ-open sets $G_j \supset A_j$ with $G_j \cap K_i = \varnothing$ if $i \ne j$, and X_ϖ-continuous functions ζ_j such that

$$\sum_{j=1}^{k} \gamma(G_j \setminus K_j) + \mu(G_j \setminus K_j) \le \varepsilon, \quad \zeta_j \equiv s^*_j \text{ on } K_j, \quad \zeta_j \equiv 0 \text{ on } X \setminus G_j.$$

Denoting by $\zeta := \sum_{j=1}^k \zeta_j$, $M := \sum_{j=1}^k |s^*_j|$, since the negative part of $F^*(s^*)$ is bounded above by $|s^*|s_0$ we have

$$\begin{aligned}\int_X \max_{s^*\in C_k} \big(s^*\rho(x) - F^*(s^*)\big)\, d\gamma(x) &= \sum_{j=1}^{k} \int_{A'_j} \big(s^*_j\rho(x) - F^*(s^*_j)\big)\, d\gamma(x)\\ \le \sum_{j=1}^{k} \int_{K_j} \big(s^*_j\rho(x) - F^*(s^*_j)\big)\, d\gamma(x) + \varepsilon(M + Ms_0)\\ = \sum_{j=1}^{k} \int_{K_j} \big(\zeta(x)\rho(x) - F^*(\zeta(x))\big)\, d\gamma(x) + \varepsilon(M + Ms_0)\\ \le \int_X \big(\zeta(x)\rho(x) - F^*(\zeta(x))\big)\, d\gamma(x) + \varepsilon(M + Ms_0 + M + F^*(M)).\end{aligned}$$

Passing to the limit as $\varepsilon \downarrow 0$ we get (9.4.13). □

Lemma 9.4.5 (Entropy and marginals). *Let $\pi : X \to X$ be a Borel map. For every couple of probability measures $\gamma, \mu \in \mathscr{P}(X)$ we have*

$$\mathcal{H}(\pi_\#\mu|\pi_\#\gamma) \le \mathcal{H}(\mu|\gamma), \quad \mathcal{F}(\pi_\#\mu|\pi_\#\gamma) \le \mathcal{F}(\mu|\gamma). \tag{9.4.14}$$

Proof. It is not restrictive to assume that $\mu \ll \gamma$: we denote by ρ a Borel map γ-a.e. equal to the density $\frac{d\mu}{d\gamma}$; applying the disintegration theorem we can find a Borel family of probability measures γ_x in X such that $\gamma = \int_X \gamma_x \, d\,\pi_\#\gamma(x)$ and $\gamma_x(X \setminus \pi^{-1}(x)) = 0$ for $\pi_\#\gamma$-a.e. x.

It follows that μ and $\pi_\#\mu$ admit the representation

$$\mu = \int_X \rho\gamma_x \, d\,\pi_\#\gamma(x) \quad \text{and} \quad \pi_\#\mu = \tilde\rho \cdot \pi_\#\gamma \quad \text{with} \quad \tilde\rho(x) := \int_{\pi^{-1}(x)} \rho(y)\, d\gamma_x(y)$$

since for each Borel set $A \subset X$ one has

$$\int_{\pi^{-1}(A)} d\mu(x) = \int_A \Big(\int_{\pi^{-1}(x)} \rho(y)\, d\gamma_x(y) \Big)\, d\,\pi_\#\gamma(x).$$

Jensen inequality yields

$$F(\tilde\rho(x)) \le \int_{\pi^{-1}(x)} F(\rho(y))\, d\gamma_x(y),$$

and therefore

$$\begin{aligned}\mathcal{F}(\pi_\#\mu|\pi_\#\gamma) &= \int_X F(\tilde\rho(x))\, d\,\pi_\#\gamma(x) \le \int_X \Big(\int_{\pi^{-1}(x)} F(\rho(y))\, d\gamma_x(y) \Big)\, d\,\pi_\#\gamma(x) \\ &\le \int_X F(\rho(x))\, d\gamma(x) = \mathcal{F}(\mu|\gamma).\end{aligned}$$

□

Corollary 9.4.6. *Let $\pi^k : X \to X$ be Borel maps such that*

$$\lim_{k\to\infty} \pi^k(x) = x \quad \forall\, x \in X.$$

For every γ, $\mu \in \mathscr{P}(X)$, setting $\gamma^k := \pi^k_\#\gamma$, $\mu^k := \pi^k_\#\mu$, we have

$$\lim_{k\to\infty} \mathcal{H}(\mu^k|\gamma^k) = \mathcal{H}(\mu|\gamma), \quad \lim_{k\to\infty} \mathcal{F}(\mu^k|\gamma^k) = \mathcal{F}(\mu|\gamma). \tag{9.4.15}$$

Proof. Lebesgue's dominated convergence theorem shows that γ^k, μ^k narrowly converge to γ, μ respectively. Combining Lemma 9.4.3 and 9.4.5 we conclude. □

9.4.1 Log-concavity and displacement convexity

We want to characterize the probability measures γ inducing a geodesically convex relative entropy functional $\mathcal{H}(\cdot|\gamma)$ in $\mathscr{P}_p(X)$. The following lemma provides the first crucial property; the argument is strictly related to the proof of the Brunn-Minkowski inequality for the Lebesgue measure, obtained via optimal transportation inequalities [126]. See also [25] for the link between log-concavity and representation formulae like (9.4.23).

Lemma 9.4.7 (γ is log-concave if $\mathcal{H}(\cdot|\gamma)$ is displacement convex). *Suppose that for each couple of probability measures $\mu^1, \mu^2 \in \mathscr{P}(X)$ with bounded support, there exists $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ such that $\mathcal{H}(\cdot|\gamma)$ is convex along the interpolating curve $\mu_t^{1\to 2} = \big((1-t)\pi^1 + t\pi^2\big)_\# \boldsymbol{\mu}$, $t \in [0,1]$. Then for each couple of open sets A, $B \subset X$ and $t \in [0,1]$ we have*

$$\log \gamma((1-t)A + tB) \geq (1-t)\log\gamma(A) + t\log\gamma(B). \tag{9.4.16}$$

Proof. We can obviously assume that $\gamma(A) > 0$, $\gamma(B) > 0$ in (9.4.16); we consider

$$\mu^1 := \gamma(\cdot|A) = \frac{1}{\gamma(A)}\chi_A \cdot \gamma, \quad \mu^2 := \gamma(\cdot|B) = \frac{1}{\gamma(B)}\chi_B \cdot \gamma,$$

observing that

$$\mathcal{H}(\mu^1|\gamma) = -\log\gamma(A), \quad \mathcal{H}(\mu^2|\gamma) = -\log\gamma(B). \tag{9.4.17}$$

If $\mu_t^{1\to 2}$ is induced by a transfer plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ along which the relative entropy is displacement convex, we have

$$\mathcal{H}(\mu_t^{1\to 2}|\gamma) \leq (1-t)\mathcal{H}(\mu^1|\gamma) + t\mathcal{H}(\mu^2|\gamma) = -(1-t)\log\gamma(A) - t\log\gamma(B).$$

On the other hand the measure $\mu_t^{1\to 2}$ is concentrated on $(1-t)A + tB = \pi_t^{1\to 2}(A \times B)$ and the next lemma shows that

$$-\log\gamma((1-t)A + tB) \leq \mathcal{H}(\mu_t^{1\to 2}|\gamma).$$

□

Lemma 9.4.8 (Relative entropy of concentrated measures). *Let $\gamma, \mu \in \mathscr{P}(X)$; if μ is concentrated on a Borel set A, i.e. $\mu(X \setminus A) = 0$, then*

$$\mathcal{H}(\mu|\gamma) \geq -\log\gamma(A). \tag{9.4.18}$$

Proof. It is not restrictive to assume $\mu \ll \gamma$ and $\gamma(A) > 0$; denoting by γ_A the probability measure $\gamma(\cdot|A) := \gamma(A)^{-1}\chi_A \cdot \gamma$, we have

$$\begin{aligned}
\mathcal{H}(\mu|\gamma) &= \int_X \log\Big(\frac{d\mu}{d\gamma}\Big)\,d\mu = \int_A \log\Big(\frac{d\mu}{d\gamma_A}\cdot\frac{1}{\gamma(A)}\Big)\,d\mu \\
&= \int_A \log\Big(\frac{d\mu}{d\gamma_A}\Big)\,d\mu - \int_A \log\big(\gamma(A)\big)\,d\mu = \mathcal{H}(\mu|\gamma_A) - \log\big(\gamma(A)\big) \\
&\geq -\log\big(\gamma(A)\big).
\end{aligned}$$

□

The previous results justifies the following definition:

Definition 9.4.9 (log-**concavity of a measure**). *We say that a Borel probability measure $\gamma \in \mathscr{P}(X)$ on X is* log-concave *if for every couple of open sets A, $B \subset X$ we have*

$$\log \gamma((1-t)A + tB) \geq (1-t)\log \gamma(A) + t \log \gamma(B). \tag{9.4.19}$$

In Definition 9.4.9 and also in the previous theorem we confined ourselves to pairs of open sets, to avoid the non trivial issue of the measurability of $(1-t)A+tB$ when A and B are only Borel (in fact, it is an open set whenever A and B are open). Observe that a log-concave measure γ in particular satisfies

$$\log \gamma(B_r((1-t)x_0 + tx_1)) \geq (1-t)\log \gamma(B_r(x_0)) + t \log \gamma(B_r(x_1)), \tag{9.4.20}$$

for every couple of points $x_0, x_1 \in X$, $r > 0$, $t \in [0,1]$.

We want to show that in fact log concavity is equivalent to the geodesic convexity of the Relative Entropy functional $\mathcal{H}(\cdot|\gamma)$.

Let us first recall some elementary properties of convex sets in $\mathbb{R}^d$. Let $C \subset \mathbb{R}^d$ be a convex set; the *affine dimension* $\dim C$ of C is the linear dimension of its affine envelope

$$\operatorname{aff} C = \Big\{(1-t)x_0 + tx_1 : x_0, x_1 \in C,\ t \in \mathbb{R}\Big\}, \tag{9.4.21}$$

which is an affine subspace of $\mathbb{R}^d$. We denote by $\operatorname{int} C$ the relative interior of C as a subset of $\operatorname{aff} C$: it is possible to show that

$$\operatorname{int} C \neq \varnothing, \quad \overline{\operatorname{int} C} = \overline{C}, \quad \mathscr{H}^k(\overline{C} \setminus \operatorname{int} C) = 0 \quad \text{if } k = \dim C. \tag{9.4.22}$$

Theorem 9.4.10. *Let us suppose that $X = \mathbb{R}^d$ is finite dimensional and $\gamma \in \mathscr{P}(X)$ satisfies the* log-concavity *assumptions on balls* (9.4.20)*. Then $\operatorname{supp}\gamma$ is convex and there exists a convex l.s.c. function $V : X \to (\infty, +\infty]$ such that*

$$\gamma = e^{-V} \cdot \mathscr{H}^k|_{\operatorname{aff}(\operatorname{supp}\gamma)}, \qquad \textit{where } k = \dim(\operatorname{supp}\gamma). \tag{9.4.23}$$

Conversely, if γ admits the representation (9.4.23) *then γ is* log-concave *and the relative entropy functional $\mathcal{H}(\cdot|\gamma)$ is convex along any (generalized, if $p = 2$) geodesic of $\mathscr{P}_p(X)$.*

Proof. Let us suppose that γ satisfies the log-concave inequality on balls and let k be the dimension of $\operatorname{aff}(\operatorname{supp}\gamma)$. Observe that the measure γ satisfies the same inequality (9.4.20) for the balls of $\operatorname{aff}(\operatorname{supp}\gamma)$: up to an isometric change of coordinates it is not restrictive to assume that $k = d$ and $\operatorname{aff}(\operatorname{supp}\gamma) = \mathbb{R}^d$.

Let us now introduce the set

$$D := \Big\{x \in \mathbb{R}^d :\ \liminf_{r \downarrow 0} \frac{\gamma(B_r(x))}{r^d} > 0\Big\}. \tag{9.4.24}$$

Since (9.4.20) yields

$$\frac{\gamma(B_r(x_t))}{r^k} \geq \left(\frac{\gamma(B_r(x_0))}{r^k}\right)^{1-t} \left(\frac{\gamma(B_r(x_1))}{r^k}\right)^{t} \quad t \in (0,1), \tag{9.4.25}$$

it is immediate to check that D is a convex subset of $\mathbb{R}^d$ with $D \subset \operatorname{supp}\gamma$.

General results on derivation of Radon measures in $\mathbb{R}^d$ (see for instance Theorem 2.56 in [11]) show that

$$\limsup_{r\downarrow 0} \frac{\gamma(B_r(x))}{r^d} < +\infty \quad \text{for } \mathscr{L}^d\text{-a.e. } x \in \mathbb{R}^d \tag{9.4.26}$$

and

$$\limsup_{r\downarrow 0} \frac{r^d}{\gamma(B_r(x))} < +\infty \quad \text{for } \gamma\text{-a.e. } x \in \mathbb{R}^d. \tag{9.4.27}$$

Using (9.4.27) we see that actually γ is concentrated on D (so that $\operatorname{supp}\gamma \subset \overline{D}$) and therefore, being d the dimension of $\operatorname{aff}(\operatorname{supp}\gamma)$, it follows that d is also the dimension of $\operatorname{aff}(D)$.

If a point $\bar{x} \in \mathbb{R}^d$ exists such that

$$\limsup_{r\downarrow 0} \frac{\gamma(B_r(\bar{x}))}{r^d} = +\infty,$$

then (9.4.25) forces every point of $\operatorname{int}(D)$ to verify the same property, but this would be in contradiction with (9.4.26), since we know that $\operatorname{int}(D)$ has strictly positive $\mathscr{L}^d$-measure. Therefore

$$\limsup_{r\downarrow 0} \frac{\gamma(B_r(x))}{r^d} < +\infty \quad \text{for all } x \in \mathbb{R}^d \tag{9.4.28}$$

and we obtain that $\gamma \ll \mathscr{L}^d$, again by the theory of derivation of Radon measures in $\mathbb{R}^d$. In the sequel we denote by ρ the density of γ w.r.t. $\mathscr{L}^d$ and notice that by Lebesgue differentiation theorem $\rho > 0$ $\mathscr{L}^d$-a.e. in D and $\rho = 0$ $\mathscr{L}^d$-a.e. in $\mathbb{R}^d \setminus D$.

By (9.4.20) the maps

$$V_r(x) = -\log\left(\frac{\gamma(B_r(x))}{\omega_d r^d}\right)$$

are convex on $\mathbb{R}^d$, and (9.4.28) gives that the family $V_r(x)$ is bounded as $r \downarrow 0$ for any $x \in D$. Using the pointwise boundedness of V_r on D and the convexity of V_r it is easy to show that V_r are locally equi-bounded (hence locally equi-continuous) on $\operatorname{int}(D)$ as $r \downarrow 0$. Let W be a limit point of V_r, with respect to the local uniform convergence, as $r \downarrow 0$: W is convex on $\operatorname{int}(D)$ and Lebegue differentiation theorem shows that

$$\exists \lim_{r\downarrow 0} V_r(x) = -\log\rho(x) = W(x) \quad \text{for } \mathscr{L}^d\text{-a.e. } \ x \in \operatorname{int}(D), \tag{9.4.29}$$

so that $\gamma = \rho\mathscr{L}^d = e^{-W}\chi_{\mathrm{int}(D)}\mathscr{L}^d$. In order to get a globally defined convex and l.s.c function V we extend W with the $+\infty$ value out of $\mathrm{int}(D)$ and define V to be its convex and l.s.c. envelope. It turns out that V coincides with W on $\mathrm{int}(D)$, so that still the representation $\gamma = e^{-V}\mathscr{L}^d$ holds.

Conversely, let us suppose that γ admits the representation (9.4.23) for a given convex l.s.c. function V and let μ^1, $\mu^2 \in \mathscr{P}_p(X)$; if their relative entropies are finite then they are absolutely continuous w.r.t. γ and therefore their supports are contained in $\mathrm{aff}(\mathrm{supp}\,\gamma)$. It follows that the support of any optimal plan $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$ in $\mathscr{P}_p(X)$ is contained in $\mathrm{aff}(\mathrm{supp}\,\gamma) \times \mathrm{aff}(\mathrm{supp}\,\gamma)$: up to a linear isometric change of coordinates, it is not restrictive to suppose $\mathrm{aff}(\mathrm{supp}\,\gamma) = \mathbb{R}^d$, μ^1, $\mu^2 \in \mathscr{P}_p(\mathbb{R}^d)$, $\gamma = e^{-V} \cdot \mathscr{L}^d \in \mathscr{P}(\mathbb{R}^d)$.

In this case we introduce the density ρ^i of μ^i w.r.t. $\mathscr{L}^d$ observing that

$$\frac{d\mu^i}{d\gamma} = \rho^i e^V \quad i = 1, 2,$$

where we adopted the convention $0 \cdot (+\infty) = 0$ (recall that $\rho^i(x) = 0$ for $\mathscr{L}^d$-a.e. $x \in \mathbb{R}^d \setminus D(V)$). Therefore the entropy functional can be written as

$$\mathcal{H}(\mu^i|\gamma) = \int_{\mathbb{R}^d} \rho^i(x) \log \rho^i(x)\, dx + \int_{\mathbb{R}^d} V(x)\, d\mu^i(x), \tag{9.4.30}$$

i.e. the sum of two geodesically convex functionals, as we proved discussing Examples 9.3.1 and Examples 9.3.6. Lemma 9.4.7 yields the log-concavity of γ; the case of generalized geodesics in $\mathscr{P}_2(X)$ is completely analogous. □

The previous theorem shows that in finite dimensions log-concavity of γ is equivalent to the convexity of $\mathcal{H}(\mu|\gamma)$ along (even generalized, if $p = 2$) geodesics of anyone of the Wasserstein spaces $\mathscr{P}_p(X)$: the link between these two concepts is provided by the representation formula (9.4.23).

When X is an infinite dimensional Hilbert space, (9.4.23) is no more true in general, but the equivalence between log-concavity and geodesic convexity of the relative entropy still holds. In particular all Gaussian measures, defined in Definition 6.2.1, induce a geodesically convex relative entropy functional (see condition (5) in the statement below).

Theorem 9.4.11. *Let X be a separable Hilbert space and let $\gamma \in \mathscr{P}(X)$. The following properties are equivalent:*

(1) *$\mathcal{H}(\cdot|\gamma)$ is geodesically convex in $\mathscr{P}_p(X)$ for every $p \in (1, +\infty)$.*

(2) *$\mathcal{H}(\cdot|\gamma)$ is convex along generalized geodesics in $\mathscr{P}_2(X)$.*

(3) *For every couple of measures μ^1, $\mu^2 \in \mathscr{P}(X)$ with bounded support there exists a connecting plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2)$ along with $\mathcal{H}(\cdot|\gamma)$ is displacement convex.*

(4) *γ is log-concave.*

(5) *For every finite dimensional orthogonal projection $\pi : X \to X$, $\pi_\# \gamma$ is representable as in* (9.4.23) *for a suitable convex and l.s.c. function V.*

Proof. The implications (1) ⇒ (3) and (2) ⇒ (3) are trivial, and (3) ⇒ (4) follows by Lemma 9.4.7.
Now we show that (4) ⇒ (5), using Theorem 9.4.10: if A, B are (relatively) open subsets of $\pi(X)$ and $t \in [0,1]$ we should prove that

$$\log\Big(\pi_\#\gamma\big((1-t)A+tB\big)\Big) \ge (1-t)\log\Big(\pi_\#\gamma\big(A\big)\Big) + t\log\Big(\pi_\#\gamma\big(B\big)\Big). \quad (9.4.31)$$

By definition $\pi_\#\gamma\big(A\big) = \gamma\big(\pi^{-1}A\big)$, $\pi_\#\gamma\big(B\big) = \gamma\big(\pi^{-1}B\big)$, and it is immediate to check that

$$\pi_\#\gamma\big((1-t)A+tB\big) = \gamma\big((1-t)\pi^{-1}A + t\pi^{-1}B\big)$$

since $\pi^{-1}\big((1-t)A+tB\big) = (1-t)\pi^{-1}A + t\pi^{-1}B$. Thus (9.4.31) follows by the log-concavity of γ applied to the open sets $\pi^{-1}A$, $\pi^{-1}B$.
(5) ⇒ (1): we choose a sequence π^h of finite dimensional orthogonal projections on X such that $\pi^h(x) \to x$ for any $x \in X$ as $h \to \infty$, set $\gamma^h := \pi^h_\#\gamma$ and

$$\phi^h(\mu) := \mathcal{H}(\mu|\gamma^h), \quad \phi(\mu) := \mathcal{H}(\mu|\gamma) \quad \forall\, \mu \in \mathscr{P}(X).$$

Since each functional ϕ^h is geodesically convex in $\mathscr{P}_p(X)$, by Theorem 9.4.10, the thesis follows by Lemma 9.1.4 if we show that ϕ is the Γ-limit of ϕ^h as $h \to \infty$: thus we have to check conditions (9.1.4) and (9.1.5).

(9.1.4) follows immediately by Lemma 9.4.3; in order to check (9.1.5) we simply choose $\mu^h := \pi^h_\#\mu$ and we apply Corollary 9.4.6.
The implications (5) ⇒ (2) follows by the same approximation argument, invoking Lemma 9.2.9. □

If γ is log-concave and F satisfies (9.3.19), then all the integral functionals $\mathcal{F}(\cdot|\gamma)$ introduced in (9.4.7) are geodesically convex in $\mathscr{P}_p(X)$ and convex along generalized geodesics in $\mathscr{P}_2(X)$.

Theorem 9.4.12 (Geodesical convexity for relative integral functionals). *Suppose that γ is* log*-concave and $F : [0,+\infty) \to [0,+\infty]$ satisfies conditions* (9.4.6) *and* (9.3.19)*. Then the integral functional $\mathcal{F}(\cdot|\gamma)$ is geodesically convex in $\mathscr{P}_p(X)$ and convex along generalized geodesics in $\mathscr{P}_2(X)$.*

Proof. The same approximation argument of the proof of the previous theorem shows that it is sufficient to consider the final dimensional case $X := \mathbb{R}^d$. Arguing as in the final part of the proof of Theorem 9.4.10 we can assume that $\gamma := e^{-V}\mathscr{L}^d$ for a convex l.s.c. function $V : \mathbb{R}^d \to (-\infty,+\infty]$ whose domain has not empty interior. For every couple of measure $\mu^1, \mu^2 \in D(\mathcal{F}(\cdot|\gamma))$ we have

$$\mu^i = \rho^i e^V \cdot \gamma, \quad \mathcal{F}(\mu^i|\gamma) = \int_{\mathbb{R}^d} F(\rho^i(x)e^{V(x)})e^{-V(x)}\,dx \quad i = 1,2. \qquad (9.4.32)$$

As in Proposition 9.3.9, we denote by $\boldsymbol{r}$ the optimal transport map for the p-Wasserstein distance pushing μ^1 to μ^2 and we set $\boldsymbol{r}^t := (1-t)\boldsymbol{i}+t\boldsymbol{r}$, $\mu_t := (\boldsymbol{r}^t)_{\#}\mu^1$; arguing as in that proposition, we get

$$\mathcal{F}(\mu_t|\gamma) = \int_{\mathbb{R}^d} F\Big(\frac{\rho(x)e^{V(\boldsymbol{r}_t(x))}}{\det \tilde{\nabla}\boldsymbol{r}^t(x)}\Big) \det \tilde{\nabla}\boldsymbol{r}^t(x)e^{-V(\boldsymbol{r}_t(x))}\,dx, \tag{9.4.33}$$

and the integrand above may be seen as the composition of the convex and non-increasing map $s \mapsto F(\rho(x)e^{-s})e^s$ with the concave curve

$$t \mapsto -V(\boldsymbol{r}_t(x)) + \log(\det \tilde{\nabla}\boldsymbol{r}_t(x)),$$

since $D(x) := \tilde{\nabla}\boldsymbol{r}(x)$ is a diagonalizable map with nonnegative eigenvalues and

$$t \mapsto \log\det\big((1-t)I + tD(x)\big) \quad \text{is concave in } [0,1].$$

The case of convexity along generalized geodesics in $\mathscr{P}_2(\mathbb{R}^d)$ follows by the same argument, recalling the final part of the proof of Proposition 9.3.9 once again. □

Chapter 10

Metric Slope and Subdifferential Calculus in $\mathscr{P}_p(X)$

As we have seen in Section 1.4, in the classical theory of subdifferential calculus for proper, lower semicontinuous functionals $\phi : X \to (-\infty, +\infty]$ defined in a Hilbert space X, the *Fréchet Subdifferential* $\partial\phi : X \to 2^X$ of ϕ is a multivalued operator defined as

$$\xi \in \partial\phi(v) \quad \Longleftrightarrow \quad v \in D(\phi), \quad \liminf_{w \to v} \frac{\phi(w) - \phi(v) - \langle \xi, w - v\rangle}{|w - v|} \geq 0, \qquad (10.0.1)$$

which we will also write in the equivalent form for $v \in D(\phi)$

$$\xi \in \partial\phi(v) \quad \Longleftrightarrow \quad \phi(w) \geq \phi(v) + \langle \xi, w - v\rangle + o\big(|w - v|\big) \quad \text{as } w \to v. \qquad (10.0.2)$$

As usual in multivalued analysis, the proper domain $D(\partial\phi) \subset D(\phi)$ is defined as the set of all $v \in X$ such that $\partial\phi(v) \neq \varnothing$; we will use this convention for all the multivalued operators we will introduce.

The Fréchet subdifferential occurs quite naturally in the Euler equations for minima of (smooth perturbation of) ϕ:

A. Euler equation for quadratic perturbations. If v_τ is a minimizer of

$$w \mapsto \Phi(\tau, v; w) := \phi(w) + \frac{1}{2\tau}|w - v|^2 \quad \text{for some} \quad \tau > 0, \ v \in X \qquad (10.0.3)$$

then

$$v_\tau \in D(\partial\phi) \quad \text{and} \quad -\frac{v_\tau - v}{\tau} \in \partial\phi(v_\tau). \qquad (10.0.4)$$

For **λ-convex functionals** (recall Definition 2.4.1 and Remark 2.4.4) the Fréchet subdifferential enjoys at least two other simple but fundamental properties, which play a crucial role in the corresponding variational theory of evolution equations:

B. Characterization by variational inequalities and monotonicity. If ϕ is λ-convex, then

$$\xi \in \partial\phi(v) \quad \Longleftrightarrow \quad \phi(w) \geq \phi(v) + \langle \xi, w - v \rangle + \frac{\lambda}{2}|w - v|^2 \quad \forall\, w \in D(\phi); \tag{10.0.5}$$

in particular,

$$\xi_i \in \partial\phi(v_i) \quad \Longrightarrow \quad \langle \xi_1 - \xi_2, v_1 - v_2 \rangle \geq \lambda |v_1 - v_2|^2 \quad \forall\, v_1, v_2 \in D(\partial\phi). \tag{10.0.6}$$

C. Convexity and strong-weak closure. [28, Chap. II, Ex. 2.3.4, Prop. 2.5] If ϕ is λ-convex, then $\partial\phi(v)$ is closed and convex, and for every sequences (v_n), $(\xi_n) \in X$ we have

$$\xi_n \in \partial\phi(v_n), \quad v_n \to v,\ \xi_n \rightharpoonup \xi \quad \Longrightarrow \quad \xi \in \partial\phi(v), \quad \phi(v_n) \to \phi(v). \tag{10.0.7}$$

Modeled on the last property **C** and following a terminology introduced by F.H. Clarke, see e.g. [113, Chap. 8], we say that a functional ϕ is *regular* if

$$\left.\begin{matrix} \xi_n \in \partial\phi(v_n), & \varphi_n = \phi(v_n) \\ v_n \to v, \quad \xi_n \rightharpoonup \xi, & \varphi_n \to \varphi \end{matrix}\right\} \quad \Longrightarrow \quad \xi \in \partial\phi(v), \quad \varphi = \phi(v). \tag{10.0.8}$$

D. Minimal selection and slope. (cf. Proposition 1.4.4) If ϕ is regular (in particular if ϕ is λ-convex) for every $v \in D(\phi)$ the *metric slope*

$$|\partial\phi|(v) = \limsup_{w \to v} \frac{(\phi(v) - \phi(w))^+}{|w - v|} \tag{10.0.9}$$

is finite if and only if $\partial\phi(v) \neq \emptyset$ and

$$|\partial\phi|(v) = \min\Big\{ |\xi| : \xi \in \partial\phi(v) \Big\}. \tag{10.0.10}$$

E. Chain rule. If $v : (a, b) \to D(\phi)$ is a curve in X then

$$\frac{d}{dt}\phi(v(t)) = \langle \xi, v'(t) \rangle \quad \forall\, \xi \in \partial\phi(v(t)), \tag{10.0.11}$$

at each point t where v and $\phi \circ v$ are differentiable and $\partial\phi(v(t)) \neq \emptyset$. In particular (see [28, Chap. III, Lemma 3.3] and Remark 1.4.6) if ϕ is also λ-convex, $v \in AC(a, b; X)$ (see Remark 1.1.3), and

$$\int_a^b |\partial\phi|(v(t))|v'(t)|\, dt < +\infty, \tag{10.0.12}$$

then $\phi \circ v$ is absolutely continuous in (a, b) and (10.0.11) holds for $\mathscr{L}^1$-a.e. $t \in (a, b)$.

The aim of this chapter is to extend the notion of Fréchet subdifferentiability and these properties to the Wasserstein framework (see also [38] for related results). In the next section we shall consider the simpler case of regular measures in $\mathscr{P}_2(X)$, where the theory exhibit an evident formal analogy with the Euclidean one.

After a detailed analysis of the differentiability properties of the Wasserstein distance map $W_p(\cdot,\mu)$ from a fixed reference measure $\mu \in \mathscr{P}_p(X)$ that we will carry out in Section 10.2, in the third section we will attack the case of general measures in $\mathscr{P}_p(X)$. Examples are provided in the last section of this chapter.

10.1 Subdifferential calculus in $\mathscr{P}_2^r(X)$: the regular case

In this section we focus our attention to functionals ϕ defined on $\mathscr{P}_2(X)$ (i.e. here $p = 2$) and we present the main definitions and results on subdifferentiability in the (considerably) simplifying assumption that each measure μ in $D(|\partial\phi|)$ can be pushed on every $\nu \in D(\phi)$ by a unique optimal transport map, which we denoted by $\boldsymbol{t}_\mu^\nu$ in (7.1.4). To ensure this property we are supposing that

$$\begin{gathered}\phi : \mathscr{P}_2(X) \to (-\infty, +\infty] \text{ is proper and lower semicontinuous,}\\ \text{with } D(|\partial\phi|) \subset \mathscr{P}_2^r(X);\end{gathered} \tag{10.1.1a}$$

we further simplify some technical point by assuming that for some $\tau_* > 0$ the functional

$$\begin{gathered}\nu \mapsto \Phi(\tau,\mu;\nu) = \frac{1}{2\tau} W_2^2(\mu,\nu) + \phi(\nu) \quad \text{admits at least}\\ \text{a minimum point } \mu_\tau, \text{ for all } \tau \in (0,\tau_*) \text{ and } \mu \in \mathscr{P}_2(X).\end{gathered} \tag{10.1.1b}$$

Notice that $D(\phi) \subset \mathscr{P}_2^r(X)$ is a sufficient but not necessary condition for (10.1.1a), see the example of the internal energy functional discussed in Theorem 10.4.13.

The formal mechanism for translating statements from the euclidean framework to the Wasserstein formalism is simple: if $\mu \leftrightarrow v$ is the reference point, scalar products $\langle\cdot,\cdot\rangle$ have to be intended in the reference Hilbert space $L^2(\mu;X)$ (which contains the tangent space $\mathrm{Tan}_\mu\mathscr{P}_2(X)$) and displacement vectors $w - v$ corresponds to transport maps $\boldsymbol{t}_\mu^\nu - \boldsymbol{i}$. According to these two natural rules, the transposition of (10.0.1) yields

Definition 10.1.1 (Fréchet subdifferential). *Let $\phi : \mathscr{P}_2(X) \to (-\infty,+\infty]$ be a functional satisfying* (10.1.1a) *and let $\mu \in D(|\partial\phi|)$. We say that $\boldsymbol{\xi} \in L^2(\mu;X)$ belongs to the Fréchet subdifferential $\partial\phi(\mu)$ if*

$$\liminf_{\nu\to\mu} \frac{\phi(\nu) - \phi(\mu) - \int_X \langle \boldsymbol{\xi}(x), \boldsymbol{t}_\mu^\nu(x) - x\rangle\, d\mu(x)}{W_2(\mu,\nu)} \geq 0, \tag{10.1.2}$$

or, with equivalent simpler notation,

$$\phi(\nu) - \phi(\mu) \geq \int_X \langle \boldsymbol{\xi}(x), \boldsymbol{t}_\mu^\nu(x) - x\rangle\, d\mu(x) + o\big(W_2(\mu,\nu)\big). \tag{10.1.3}$$

When $\boldsymbol{\xi} \in \partial\phi(\mu)$ *also satisfies*

$$\phi(\boldsymbol{t}_{\#}\mu) - \phi(\mu) \geq \int_X \langle \boldsymbol{\xi}(x), \boldsymbol{t}(x) - x\rangle \, d\mu(x) + o\big(\|\boldsymbol{t} - \boldsymbol{i}\|_{L^2(\mu;X)}\big), \tag{10.1.4}$$

then we will say that $\boldsymbol{\xi}$ *is a* strong *subdifferential.*

It is obvious that $\partial\phi(\mu)$ is a closed convex subset of $L^2(\mu; X)$; in fact, we can also impose that it is contained in the tangent space $\mathrm{Tan}_\mu \mathscr{P}_2(X)$, since the vector $\boldsymbol{\xi}$ in (10.1.3) acts only on tangent vectors (see (8.5.1) and Theorem 8.5.1).

A. Euler equation for quadratic perturbations. When we want to minimize the perturbed functional (10.1.1b) we get a result completely analogous to the euclidean one (10.0.4):

Lemma 10.1.2. *Let* ϕ *be satisfying* (10.1.1a,b) *and let* μ_τ *be a minimizer of* (10.1.1b)*; then* $\mu_\tau \in D(|\partial\phi|)$ *and*

$$\frac{1}{\tau}\big(\boldsymbol{t}_{\mu_\tau}^{\mu} - \boldsymbol{i}\big) \in \partial\phi(\mu_\tau) \quad \textit{is a strong subdifferential.} \tag{10.1.5}$$

Proof. The minimality of μ_τ gives

$$\begin{aligned}
\phi(\nu) - \phi(\mu_\tau) &= \Phi(\tau, \mu; \nu) - \Phi(\tau, \mu; \mu_\tau) + \frac{1}{2\tau}\Big(W_2^2(\mu_\tau, \mu) - W_2^2(\nu, \mu)\Big) \\
&\geq \frac{1}{2\tau}\Big(W_2^2(\mu_\tau, \mu) - W_2^2(\nu, \mu)\Big) \quad \forall \nu \in \mathscr{P}_2(X).
\end{aligned}$$

Now we observe that if $\nu = \boldsymbol{t}_{\#}\mu_\tau$

$$W_2^2(\mu_\tau, \mu) = \int_X |\boldsymbol{t}_{\mu_\tau}^{\mu}(x) - x|^2 \, d\mu_\tau(x), \quad W_2^2(\nu, \mu) \leq \int_X |\boldsymbol{t}(x) - \boldsymbol{t}_{\mu_\tau}^{\mu}(x)|^2 \, d\mu_\tau(x),$$

and therefore the elementary identity $\frac{1}{2}|a|^2 - \frac{1}{2}|b|^2 = \langle a, a - b\rangle - \frac{1}{2}|a - b|^2$ yields

$$\begin{aligned}
\phi(\nu) - \phi(\mu_\tau) &\geq \frac{1}{2\tau}\int_X \Big(|\boldsymbol{t}_{\mu_\tau}^{\mu}(x) - x|^2 - |\boldsymbol{t}_{\mu_\tau}^{\mu}(x) - \boldsymbol{t}(x)|^2\Big) \, d\mu_\tau(x) \\
&= \int_X \Big(\frac{1}{\tau}\langle \boldsymbol{t}_{\mu_\tau}^{\mu}(x) - x, \boldsymbol{t}(x) - x\rangle - \frac{1}{2\tau}|\boldsymbol{t}(x) - x|^2\Big) \, d\mu_\tau(x) \\
&= \int_X \frac{1}{\tau}\langle \boldsymbol{t}_{\mu_\tau}^{\mu}(x) - x, \boldsymbol{t}(x) - x\rangle \, d\mu_\tau(x) - \frac{1}{2\tau}\|\boldsymbol{t} - \boldsymbol{i}\|^2_{L^2(\mu_\tau;X)}.
\end{aligned}$$

We deduce $\frac{1}{\tau}\big(\boldsymbol{t}_{\mu_\tau}^{\mu} - \boldsymbol{i}\big) \in \partial\phi(\mu_\tau)$ and the strong subdifferentiability condition. □

The above result, though simple, is very useful and usually provides the first crucial information when one looks for the differential properties of discrete solutions of the variational scheme (2.0.4). The nice argument which combine the minimality of μ_τ and the possibility to use any "test" transport map $\boldsymbol{t}$ to estimate $W_2^2(\boldsymbol{t}_{\#}\nu, \mu)$ was originally introduced by F. Otto.

10.1.1 The case of λ-convex functionals along geodesics

Let us now focus our attention to the case of a λ-convex functional:

$$\phi \quad \text{is } \lambda\text{-convex on geodesics, according to Definition 9.1.1.} \tag{10.1.6}$$

B. Characterization by Variational inequalities and monotonicity. *Suppose that ϕ satisfies* (10.1.1a) *and* (10.1.6). *Then a vector $\boldsymbol{\xi} \in L^2(\mu; X)$ belongs to the Fréchet subdifferential of ϕ at μ iff*

$$\phi(\nu) - \phi(\mu) \geq \int_X \langle \boldsymbol{\xi}(x), \boldsymbol{t}_\mu^\nu(x) - x \rangle \, d\mu(x) + \frac{\lambda}{2} W_2^2(\mu, \nu) \quad \forall \, \nu \in D(\phi). \tag{10.1.7}$$

In particular if $\boldsymbol{\xi}_i \in \partial\phi(\mu_i)$, $i = 1, 2$, and $\boldsymbol{t} = \boldsymbol{t}_{\mu_1}^{\mu_2}$ is the optimal transport map, then

$$\int_X \langle \boldsymbol{\xi}_2(\boldsymbol{t}(x)) - \boldsymbol{\xi}_1(x), \boldsymbol{t}(x) - x \rangle \, d\mu_1(x) \geq \lambda W_2^2(\mu_1, \mu_2). \tag{10.1.8}$$

Proof. One implication is trivial. To prove the other one, suppose that $\boldsymbol{\xi} \in \partial\phi(\mu)$ and $\nu \in D(\phi)$; for $t \in [0, 1]$ we set $\mu_t := (\boldsymbol{i} + t(\boldsymbol{t}_\mu^\nu - \boldsymbol{i}))_{\#}\mu$ and we recall that the λ-convexity yields

$$\frac{\phi(\mu_t) - \phi(\mu)}{t} \leq \phi(\nu) - \phi(\mu) - \frac{\lambda}{2}(1 - t) W_2^2(\mu, \nu).$$

On the other hand, since $W_2(\mu, \mu_t) = t W_2(\mu, \nu)$, Fréchet differentiability yields

$$\begin{aligned} \liminf_{t \downarrow 0} \frac{\phi(\mu_t) - \phi(\mu)}{t} &\geq \liminf_{t \to 0^+} \frac{1}{t} \int_X \langle \boldsymbol{\xi}(x), \boldsymbol{t}_\mu^{\mu_t}(x) - x \rangle \, d\mu(x) \\ &\geq \int_X \langle \boldsymbol{\xi}(x), \boldsymbol{t}_\mu^\nu(x) - x \rangle \, d\mu(x), \end{aligned}$$

since $\boldsymbol{t}_\mu^{\mu_t}(x) = x + t(\boldsymbol{t}_\mu^\nu(x) - x)$. □

C. Convexity and strong-weak closure. The next step is to show the closure of the graph of $\partial\phi$: here one has to be careful in the meaning of the convergence of vectors $\boldsymbol{\xi}_n \in L^2(\mu_n; X)$, which belongs to different L^2-spaces, and we will adopt Definition 5.4.3, see also Theorem 5.4.4 for the main properties of this convergence.

Lemma 10.1.3 (Closure of the subdifferential). *Let ϕ be a λ-convex functional satisfying* (10.1.1a), *let (μ_n) be converging to $\mu \in D(\phi)$ in $\mathscr{P}_2(X)$ and let $\boldsymbol{\xi}_n \in \partial\phi(\mu_n)$ be satisfying*

$$\sup_n \int_X |\boldsymbol{\xi}_n(x)|^2 \, d\mu_n(x) < +\infty, \tag{10.1.9}$$

and converging to $\boldsymbol{\xi}$ according to Definition 5.4.3. *Then $\boldsymbol{\xi} \in \partial\phi(\mu)$.*

Proof. Let us fix $\nu \in D(\phi)$ and the related optimal transport map $\boldsymbol{t}_{\mu_n}^{\nu}$, and let $\boldsymbol{\mu}_n = (\boldsymbol{i} \times \boldsymbol{\xi}_n \times \boldsymbol{t}_{\mu_n}^{\nu})_{\#}\mu_n$. Observe that the sequence $\boldsymbol{\mu}_n$ is relatively compact in $\mathscr{P}(X \times X_{\varpi} \times X)$, by Lemma 5.2.2 (the tightness of second marginals follows by (10.1.9) and Lemma 5.1.12) and

$$\pi_{\#}^{1,3}\boldsymbol{\mu}_n \in \Gamma_o(\mu_n, \nu), \quad \boldsymbol{\gamma}_n = \pi_{\#}^{1,2}\boldsymbol{\mu}_n \quad \text{is as in Theorem 5.4.4.} \tag{10.1.10}$$

By (10.1.7) we know that

$$\phi(\nu) \geq \phi(\mu_n) + \int_{X\times X\times X} \langle x_2, x_3 - x_1\rangle \, d\boldsymbol{\mu}_n(x_1, x_2, x_3) + \frac{\lambda}{2}W_2^2(\mu_n, \nu). \tag{10.1.11}$$

If $\boldsymbol{\mu}$ is any limit point of $\boldsymbol{\mu}_n$ in $\mathscr{P}(X \times X_{\varpi} \times X)$, applying Lemma 5.2.4 and the lower semicontinuity of ϕ (recall that $|x_1|^2$ and $|x_3|^2$ are uniformly integrable w.r.t. $\boldsymbol{\mu}_n$, by the convergence of μ_n in $\mathscr{P}_2(X)$ and the fact that the third marginal of $\boldsymbol{\mu}_n$ is ν) we get

$$\phi(\nu) \geq \phi(\mu) + \int_{X\times X\times X} \langle x_2, x_3 - x_1\rangle \, d\boldsymbol{\mu}(x_1, x_2, x_3) + \frac{\lambda}{2}W_2^2(\mu, \nu). \tag{10.1.12}$$

On the other hand, $\pi_{\#}^{1,3}\boldsymbol{\mu} \in \Gamma_o(\mu, \nu)$ and (10.1.12) easily yields $\mu \in D(|\partial\phi|)$; by (10.1.1a) we know that $\pi_{\#}^{1,3}\boldsymbol{\mu}$ is induced by the unique optimal transport map $\boldsymbol{t}_{\mu}^{\nu}$. Invoking Lemma 5.3.2 we get

$$\begin{aligned}\phi(\nu) &\geq \phi(\mu) + \int_{X\times X} \langle x_2, \boldsymbol{t}_{\mu}^{\nu}(x_1) - x_1\rangle \, d\boldsymbol{\gamma}(x_1, x_2) + \frac{\lambda}{2}W_2^2(\mu, \nu) \\ &= \phi(\mu) + \int_X \langle \bar{\boldsymbol{\gamma}}(x_1), \boldsymbol{t}_{\mu}^{\nu}(x_1) - x_1\rangle \, d\mu(x_1) + \frac{\lambda}{2}W_2^2(\mu, \nu),\end{aligned}$$

with $\boldsymbol{\gamma} = \pi_{\#}^{1,2}\boldsymbol{\mu}$. Since Theorem 5.4.4 yields $\boldsymbol{\xi} = \bar{\boldsymbol{\gamma}}$, we conclude. □

10.1.2 Regular functionals

Definition 10.1.4. *A functional* $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ *satisfying* (10.1.1a) *is* regular *if whenever the* strong *subdifferential* $\boldsymbol{\xi}_n \in \partial\phi(\mu_n)$, $\varphi_n = \phi(\mu_n)$ *satisfy*

$$\begin{cases} \mu_n \to \mu \text{ in } \mathscr{P}_2(X), \quad \varphi_n \to \varphi, \quad \sup_n \|\boldsymbol{\xi}_n\|_{L^2(\mu_n;X)} < +\infty \\ \qquad \boldsymbol{\xi}_n \to \boldsymbol{\xi} \text{ weakly, according to Definition 5.4.3,} \end{cases} \tag{10.1.13}$$

then $\boldsymbol{\xi} \in \partial\phi(\mu)$ *and* $\varphi = \phi(\mu)$.

We just proved that λ-convex functionals are indeed regular.

D. Minimal selection and slope.

Lemma 10.1.5. *Let ϕ be a regular functional satisfying* (10.1.1a,b)*. $\mu \in D(|\partial\phi|)$ if and only if $\partial\phi(\mu)$ is not empty and*

$$|\partial\phi|(\mu) = \min\Big\{\|\boldsymbol{\xi}\|_{L^2(\mu;X)} : \boldsymbol{\xi} \in \partial\phi(\mu)\Big\}. \tag{10.1.14}$$

By the convexity of $\partial\phi(\mu)$ there exists a unique vector $\boldsymbol{\xi} \in \partial\phi(\mu)$ which attains the minimum in (10.1.14)*: we will denote it by $\partial^\circ\phi(\mu)$.*

Proof. It is clear from the very definition of Fréchet subdifferential that

$$|\partial\phi|(\mu) \le \|\boldsymbol{\xi}\|_{L^2(\mu;X)} \quad \forall\, \boldsymbol{\xi} \in \partial\phi(\mu);$$

thus we should prove that if $|\partial\phi|(\mu) < +\infty$ there exists $\boldsymbol{\xi} \in \partial\phi(\mu)$ such that $\|\boldsymbol{\xi}\|_{L^2(\mu;X)} \le |\partial\phi|(\mu)$. We argue by approximation: for $\mu \in D(|\partial\phi|)$ and $\tau \in (0, \tau_*)$, let μ_τ be a minimizer of (10.1.1b); by Lemma 10.1.2 and 3.1.5 we know that

$$\boldsymbol{\xi}_\tau = \frac{1}{\tau}\big(\boldsymbol{t}_{\mu_\tau}^{\mu} - \boldsymbol{i}\big) \in \partial\phi(\mu_\tau), \qquad \int_X |\boldsymbol{\xi}_\tau(x)|^2 \, d\mu_\tau(x) = \frac{W_2^2(\mu, \mu_\tau)}{\tau^2},$$

$\boldsymbol{\xi}_\tau$ is a strong subdifferential, and for a suitable vanishing subsequence $\tau_n \to 0$

$$\lim_{n\to\infty} \int_X |\boldsymbol{\xi}_{\tau_n}(x)|^2 \, d\mu_{\tau_n}(x) = |\partial\phi|^2(\mu). \tag{10.1.15}$$

By Theorem 5.4.4(c) we know that $\boldsymbol{\xi}_\tau$ has some limit point $\boldsymbol{\xi} \in L^2(\mu; X)$ as $\tau \downarrow 0$, according to Definition 5.4.3. By (10.1.13) we conclude. □

E. Chain rule. *Let $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be a regular functional satisfying* (10.1.1a,b)*, and let $\mu : (a, b) \mapsto \mu_t \in D(\phi) \subset \mathscr{P}_2(X)$ be an absolutely continuous curve with tangent velocity vector $\boldsymbol{v}_t$. Let $\Lambda \subset (a, b)$ be the set of points $t \in (a, b)$ such that*

(a) $|\partial\phi|(\mu_t) < +\infty$*;*

(b) *$\phi \circ \mu$ is approximately differentiable at t (recall Definition* 5.5.1*);*

(c) *condition* (8.4.6) *of Proposition* 8.4.6 *holds.*

Then

$$\frac{\tilde{d}}{dt}\phi(\mu_t) = \int_X \langle \boldsymbol{\xi}_t(x), \boldsymbol{v}_t(x)\rangle \, d\mu_t(x) \quad \forall\, \boldsymbol{\xi}_t \in \partial\phi(\mu_t), \quad \forall t \in \Lambda. \tag{10.1.16}$$

Moreover, if ϕ is λ-convex (10.1.6) *and*

$$\int_a^b |\partial\phi|(\mu_t)|\mu'|(t)\, dt < +\infty, \tag{10.1.17}$$

then the map $t \mapsto \phi(\mu_t)$ *is absolutely continuous, and* $(a,b) \setminus \Lambda$ *is* $\mathscr{L}^1$*-negligible.*

Proof. Let $\bar t \in \Lambda$; observing that

$$\boldsymbol{v}_h := \frac{1}{h}\big(\boldsymbol{t}_{\mu_{\bar t}}^{\mu_{\bar t+h}} - \boldsymbol{i}\big) \to v_{\bar t} \quad \text{in } L^2(\mu_{\bar t}; X), \tag{10.1.18}$$

we have

$$\phi(\mu_{\bar t+h}) - \phi(\mu_{\bar t}) \geq h \int_X \langle \boldsymbol{v}_h(x), \boldsymbol{\xi}_{\bar t}(x)\rangle \, d\mu_{\bar t}(x) + o(h). \tag{10.1.19}$$

Dividing by h and taking the right and left limits as $h \to 0$ we obtain that the left and right approximate derivatives $\tilde d/dt_{\pm}\phi(\mu_t)$ satisfy

$$\frac{\tilde d}{dt_+}\phi(\mu_t)|_{t=\bar t} \geq \int_X \langle v_{\bar t}(x), \boldsymbol{\xi}_{\bar t}(x)\rangle \, d\mu_{\bar t}(x), \quad \frac{\tilde d}{dt_-}\phi(\mu_t)|_{t=\bar t} \leq \int_X \langle v_{\bar t}(x), \boldsymbol{\xi}_{\bar t}(x)\rangle \, d\mu_{\bar t}(x)$$

and therefore we find (10.1.16).

In the λ-convex case, since $|\partial\phi|$ is a strong upper gradient (see Definition 1.2.1 and Corollary 2.4.10), we already know that $t \mapsto \phi(\mu_t)$ is absolutely continuous in (a,b) and thus the conditions (a, b, c) hold $\mathscr{L}^1$-a.e. in (a,b). □

10.2 Differentiability properties of the p-Wasserstein distance

In this section we present a careful analysis of the differentiability properties of the Wasserstein distance function $\nu \mapsto W_p^p(\mu,\nu)$ from a fixed measure $\mu \in \mathscr{P}_p(X)$.

This important example, which is a basic ingredient of the *Minimizing Movement* approach developed in Chapter 2, will provide some basic tools for dealing with more general functionals (as in step **A** of the previous section) and will suggest the right way to define their Fréchet subdifferential in terms of plans.

The main ingredient, a super-differentiability result which is essential to the developments of the next Section 10.3, is provided by Theorem 10.2.2. The remaining part is devoted to study the (more delicate) sub-differentiability properties of W_p, which are interesting by themselves, even if they do not play a crucial role in the sequel.

First of all, we recall a useful property of the differential $j_p(x) = |x|^{p-2}x$ of the function $p^{-1}|x|^p$ in X, $p \in (1,\infty)$; we first introduce the strictly positive and continuous function

$$h_p(t_0) := \int_0^1 (1-t)|t-t_0|^{p-2}\, dt = \frac{(1-t_0)^p - t_0^p + p t_0^{p-1}}{p(p-1)}, \tag{10.2.1}$$

and the positive constants

$$c_p := \min_{t_0 \in [0,1]} h_p(t_0), \quad C_p := \max_{t_0\in[0,1]} h_p(t_0), \tag{10.2.2}$$

observing that

$$c_p \geq \frac{2^{1-p}}{p(p-1)} \quad \text{for } p \geq 2, \quad C_p \leq \frac{2^{2-p}}{p-1} \quad \text{for } p \leq 2. \tag{10.2.3}$$

Lemma 10.2.1. *If $p \geq 2$ then for each couple of points x_1, x_2 in the Hilbert space X we have*

$$\begin{aligned} c_p|x_1 - x_2|^p &\leq \frac{1}{p}|x_2|^p - \frac{1}{p}|x_1|^p - \langle j_p(x_1), x_2 - x_1\rangle \\ &\leq \frac{(p-1)}{2}|x_2 - x_1|^2 \max\big(|x_2|, |x_1|\big)^{p-2}. \end{aligned} \tag{10.2.4}$$

Analogously, if $p \leq 2$ we have

$$\begin{aligned} &\frac{p-1}{2}|x_2 - x_1|^2 \min\big(|x_2|, |x_1|\big)^{p-2} \\ &\leq \frac{1}{p}|x_2|^p - \frac{1}{p}|x_1|^p - \langle j_p(x_1), x_2 - x_1\rangle \leq C_p|x_2 - x_1|^p. \end{aligned} \tag{10.2.5}$$

Proof. Let us denote by x_t, $t \in (0,1)$, the segment $x_t := (1-t)x_1 + tx_2$; it is not restrictive to suppose that $x_t \neq 0$ for each value of t. Therefore the convex map $t \mapsto p^{-1}|x_t|^p$ is of class C^2 and denoting by $g(t)$ its (nonnegative) second derivative we have

$$\frac{1}{p}|x_2|^p - \frac{1}{p}|x_1|^p - \langle j_p(x_1), x_2 - x_1\rangle = \int_0^1 (1-t)g(t)\,dt. \tag{10.2.6}$$

A direct calculation shows

$$\begin{aligned} \frac{d}{dt}p^{-1}|x|^p &= |x_t|^{p-2}\langle x_t, x_2 - x_1\rangle, \\ g(t) = \frac{d^2}{dt^2}p^{-1}|x|^p &= |x_t|^{p-2}|x_2 - x_1|^2 + (p-2)|x_t|^{p-4}\big(\langle x_t, x_2 - x_1\rangle\big)^2, \end{aligned}$$

and therefore

$$\begin{aligned} |x_t|^{p-2}|x_2 - x_1|^2 \leq g(t) &\leq (p-1)|x_t|^{p-2}|x_2 - x_1|^2 \quad \text{if } p \geq 2, \\ (p-1)|x_t|^{p-2}|x_2 - x_1|^2 \leq g(t) &\leq |x_t|^{p-2}|x_2 - x_1|^2 \quad \text{if } p \leq 2. \end{aligned}$$

The second inequality of (10.2.4) and the first one of (10.2.5) follow easily by (10.2.6). In order to prove the other inequalities, let us denote by $t_0 \in [0,1]$ the value corresponding to the point of minimal norm along the segment x_t. It is easy to check that

$$|x_t| \geq |x_t - x_{t_0}| = |x_2 - x_1|\,|t - t_0|;$$

taking into account of (10.2.2) we conclude. □

We can apply the above result to establishing a sort of *super-differentiability* property of the Wasserstein distance; to clarify our notation, we will call μ^2 the reference measure, and we are studying the map

$$\psi : \mu \mapsto \psi(\mu) := \frac{1}{p} W_p^p(\mu, \mu^2) \quad \text{near a given measure} \quad \mu^1 \in \mathscr{P}_p(X). \tag{10.2.7}$$

Some other notation will be useful: for a given plan $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2) \subset \mathscr{P}_p(X \times X)$ and $\mu^3 \in \mathscr{P}_p(X)$ we set

$$\Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3) := \Big\{ \boldsymbol{\mu} \in \mathscr{P}_p(X \times X \times X) : \pi^{1,2}_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{1\,2}, \quad \pi^3_\# \boldsymbol{\mu} = \mu^3 \Big\}, \tag{10.2.8}$$

which is a subset of $\Gamma(\mu^1, \mu^2, \mu^3)$; a "3-plan" $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ induces the "pseudo-distances"

$$W^p_{p,\boldsymbol{\mu}}(\mu^i, \mu^j) := \int_{X^3} |x_i - x_j|^p \, d\boldsymbol{\mu}(x_1, x_2, x_3) \qquad i, j \in \{1, 2, 3\}, \tag{10.2.9}$$

some of which reduce to the Wasserstein ones, if $\pi^{i,j}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^i, \mu^j)$. In particular, we will often consider

$$\Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3) := \Big\{ \boldsymbol{\mu} \in \mathscr{P}_p(X^3) : \pi^{1,2}_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{1\,2}, \quad \pi^{1,3}_\# \boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^3) \Big\}, \tag{10.2.10}$$

observing that for $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$ and $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)$ we have

$$W_{p,\boldsymbol{\mu}}(\mu^1, \mu^2) = W_p(\mu^1, \mu^2) \quad \text{and} \quad W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3) = W_p(\mu^1, \mu^3). \tag{10.2.11}$$

Theorem 10.2.2 (Super-differentiability of W_p). *Let us fix μ^1, $\mu^2 \in \mathscr{P}_p(X)$, $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$, and let ψ be defined as in (10.2.7). Then for every $\mu^3 \in \mathscr{P}_p(X)$ and $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3)$ we have*

$$\psi(\mu^3) - \psi(\mu^1) + \int_{X^3} \langle j_p(x_2 - x_1), x_3 - x_1 \rangle \, d\boldsymbol{\mu} \le o\big(W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3)\big) \tag{10.2.12}$$

where for $p \ge 2$

$$o\big(W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3)\big) = (p-1) W^2_{p,\boldsymbol{\mu}}(\mu^1, \mu^3) \Big(W_p(\mu^1, \mu^2) + W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3) \Big)^{p-2} \tag{10.2.13}$$

and for $p \le 2$

$$o\big(W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3)\big) = \frac{2^{2-p}}{p-1} W^p_{p,\boldsymbol{\mu}}(\mu^1, \mu^3). \tag{10.2.14}$$

In particular

$$\limsup_{\substack{W_{p,\boldsymbol{\mu}}(\mu^3, \mu^1) \to 0 \\ \boldsymbol{\mu} \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3)}} \frac{\psi(\mu^3) - \psi(\mu^1) + \int_{X^3} \langle j_p(x_2 - x_1), x_3 - x_1 \rangle \, d\boldsymbol{\mu}}{W_{p,\boldsymbol{\mu}}(\mu^3, \mu^1)} \le 0. \tag{10.2.15}$$

If we restrict $\boldsymbol{\mu}$ to belong to $\Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)$, then we can replace $W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3)$ with $W_p(\mu^1, \mu^3)$ in (10.2.12), (10.2.13), (10.2.14), (10.2.15).

Proof. Let us check (10.2.12) for $p \geq 2$, the other case being even easier: since $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3)$ and $\boldsymbol{\mu}^{1\,2}$ is optimal, we have by (10.2.4)

$$
\begin{aligned}
\frac{1}{p} W_p^p(\mu^3, \mu^2) - \frac{1}{p} W_p^p(\mu^1, \mu^2) + \int_{X^3} \langle j_p(x_2 - x_1), x_3 - x_1 \rangle \, d\boldsymbol{\mu} \\
\leq \int_{X^3} \Big(\frac{1}{p}|x_3 - x_2|^p - \frac{1}{p}|x_2 - x_1|^p - \langle j_p(x_2 - x_1), x_3 - x_1 \rangle \Big) \, d\boldsymbol{\mu} \\
\leq (p-1) \int_{X^3} |x_3 - x_1|^2 \max\big(|x_2 - x_1|, |x_3 - x_2|\big)^{p-2} \, d\boldsymbol{\mu} \\
\leq (p-1) \Big(\int_{X^3} |x_3 - x_1|^p \, d\boldsymbol{\mu} \Big)^{2/p} \Big(\int_{X^3} \big(|x_2 - x_1| + |x_3 - x_1|\big)^p \, d\boldsymbol{\mu} \Big)^{(p-2)/p} \\
\leq (p-1) W_{p,\boldsymbol{\mu}}^2(\mu^1, \mu^3) \Big(W_p(\mu^1, \mu^2) + W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3) \Big)^{p-2}.
\end{aligned}
$$

□

Remark 10.2.3 (Super-differentiability). Recalling that, at least in the case $p = 2$, the function $-\psi$ is (-1)-convex along geodesics, it is not surprising that we proved a *super*-differentiability result for ψ, i.e. a *sub*-differentiability property for $-\psi$. The converse property requires a more refined argument and it does not hold in general: we will discuss this property in the next theorem.

Remark 10.2.4 (The regular case). Let us suppose that $\mu^1 \in \mathscr{P}_p^r(X)$ in the previous statement; then $\Gamma_o(\mu^1, \mu^2)$ contains the unique plan $\boldsymbol{\mu}^{1\,2} = \big(\boldsymbol{i} \times \boldsymbol{t}_{\mu^1}^{\mu^2}\big)_\# \mu^1$ and $\Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)$ contains the unique plan $\boldsymbol{\mu} = \big(\boldsymbol{i} \times \boldsymbol{t}_{\mu^1}^{\mu^2} \times \boldsymbol{t}_{\mu^1}^{\mu^3}\big)_\# \mu^1$; therefore (10.2.12) becomes (up to a change of sign)

$$
\psi(\mu^3) - \psi(\mu^1) + \int_X \langle j_p\big(\boldsymbol{t}_{\mu^1}^{\mu^2}(x_1) - x_1\big), \boldsymbol{t}_{\mu^1}^{\mu^3}(x_1) - x_1 \rangle \, d\mu^1(x_1) \leq o\big(W_p(\mu^1, \mu^3)\big). \tag{10.2.16}
$$

Recalling Definition 10.1.1 we could say that

$$
j_p\big(\boldsymbol{t}_{\mu^1}^{\mu^2} - \boldsymbol{i}\big) \in \partial(-\psi)(\mu^1), \tag{10.2.17}
$$

which is formally analogous to the euclidean formula

$$
j_p(x_2 - x_1) \in \partial(-\psi)(x_1) \quad \text{where} \quad \psi(x_1) := \frac{1}{p}|x_1 - x_2|^p. \tag{10.2.18}
$$

In the case of a general measure μ^1, (10.2.17) suggests an extended notion for the subdifferential of $-\psi$: anticipating the definition of the next section, we will say that the rescaled plans

$$
\boldsymbol{\gamma} := \big(x_1, j_p(x_2 - x_1)\big)_\# \boldsymbol{\mu}^{1\,2} \quad \text{for } \boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2) \tag{10.2.19}
$$

will belong to the extended subdifferential $\boldsymbol{\partial}(-\psi)(\mu^1)$.

Remark 10.2.5. One may wonder about the use of general plans $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3)$ instead of $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)$ in (10.2.12),$\cdots$,(10.2.15): this choice corresponds to consider more general perturbation of μ^1 than those obtained by optimal transports: this is the reason why these perturbations have to be measured in the pseudo-distance $W_{p,\boldsymbol{\mu}}(\cdot,\cdot)$ instead of $W_p(\cdot,\cdot)$.

This choice will reveal more flexible and useful when one considers the Euler equation for minima of the Yosida approximation of a given functional, as we will discuss in the point **A** of the next section and in some of the examples of Section 10.4.

When $\Gamma_o(\mu^1, \mu^2)$ contains a unique element *induced by a transport*, then we can prove a corresponding sub-differentiability (and therefore differentiability) of the Wasserstein distance.

Theorem 10.2.6 (Sub-differentiability of W_p). *Let us fix μ^1, $\mu^2 \in \mathscr{P}_p(X)$ and let us suppose that $\Gamma_o(\mu^1, \mu^2)$ contains a unique element $\boldsymbol{\mu}^{1\,2} = \big(\boldsymbol{i} \times \boldsymbol{r}\big)_{\#}\mu^1$. Then the distance function $\mu \mapsto \psi(\mu) = \frac{1}{p} W_p^p(\mu, \mu^2)$ satisfies*

$$\liminf_{\substack{W_p(\mu^1,\mu^3)\to 0\\ \boldsymbol{\mu}\in\Gamma_o(\mu^1,\mu^3)}} \frac{\psi(\mu^3) - \psi(\mu^1) + \int_{X^2} \langle j_p(\boldsymbol{r}(x_1) - x_1), x_3 - x_1\rangle \, d\boldsymbol{\mu}}{W_p(\mu^3, \mu^1)} \geq 0. \tag{10.2.20}$$

Proof. Being $\boldsymbol{\mu}^{1\,2}$ induced by a transport $\boldsymbol{r}$, any element $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3)$ is of the form $\boldsymbol{\mu} = \big(x_1, \boldsymbol{r}(x_1), x_2\big)_{\#}\boldsymbol{\mu}^{1\,3}$ for some $\boldsymbol{\mu}^{1\,3} \in \Gamma_o(\mu^1, \mu^3)$. In particular we can rewrite (10.2.20) in the form

$$\liminf_{\substack{W_p(\mu^1,\mu^3)\to 0\\ \boldsymbol{\mu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)}} \frac{\psi(\mu^3) - \psi(\mu^1) + \int_{X^3} \langle j_p(x_2 - x_1), x_3 - x_1\rangle \, d\boldsymbol{\mu}}{W_p(\mu^3, \mu^1)} \geq 0. \tag{10.2.21}$$

Let us choose a sequence (μ^3_n) converging to μ^1 in $\mathscr{P}_p(X)$ and a corresponding sequence of plans $\boldsymbol{\mu}_n \in \Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3_n)$ such that

$$\liminf_{\substack{\mu^3\to\mu^1\\ \boldsymbol{\mu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)}} \frac{\psi(\mu^3) - \psi(\mu^1) + \int_{X^3} \langle j_p(x_2 - x_1), x_3 - x_1\rangle \, d\boldsymbol{\mu}}{W_p(\mu^3, \mu^1)} =$$
$$\lim_{n\to\infty} \frac{\psi(\mu^3_n) - \psi(\mu^1) + \int_{X^3} \langle j_p(x_2 - x_1), x_3 - x_1\rangle \, d\boldsymbol{\mu}_n}{W_p(\mu^3_n, \mu^1)}.$$

Choosing $\boldsymbol{\beta}_n$ such that

$$\pi^{1,3}_{\#}\boldsymbol{\beta}_n = \pi^{1,3}_{\#}\boldsymbol{\mu}_n \in \Gamma_o(\mu^1, \mu^3_n), \quad \pi^{2,3}_{\#}\boldsymbol{\beta}_n \in \Gamma_o(\mu^2, \mu^3_n),$$

we observe that $\pi^{1,2}_{\#}\boldsymbol{\beta}_n \in \Gamma(\mu^1, \mu^2)$, so that

$$W_p^p(\mu^3_n, \mu^2) - W_p^p(\mu^1, \mu^2) \geq \int_{X^3} \Big(|x_3 - x_2|^p - |x_1 - x_2|^p\Big) \, d\boldsymbol{\beta}_n.$$

Now we denote by $\lambda_n := W_p(\mu^3_n, \mu^1)$ and we rescale $\boldsymbol{\mu}_n$, $\boldsymbol{\beta}_n$ so that

$$\boldsymbol{\mu}_n = \big(\pi^1, \pi^2, \pi^1 + \lambda_n \pi^3\big)_{\#} \hat{\boldsymbol{\mu}}_n, \quad \boldsymbol{\beta}_n = \big(\pi^1, \pi^2, \pi^1 + \lambda_n \pi^3\big)_{\#} \hat{\boldsymbol{\beta}}_n$$

obtaining

$$\begin{aligned}\frac{1}{p} W_p^p(\mu^3_n, \mu^2) - \frac{1}{p} W_p^p(\mu^1, \mu^2) &\geq \frac{1}{p} \int_{X^3} \Big(|x_2 - x_1 - \lambda_n x_3|^p - |x_2 - x_1|^p \Big)\, d\hat{\boldsymbol{\beta}}_n \\ &\geq -\lambda_n \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\beta}}_n,\end{aligned}$$

$$\int_{X^3} \langle j_p(x_2 - x_1), x_3 - x_1 \rangle \, d\boldsymbol{\mu}_n = \lambda_n \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\mu}}_n.$$

It follows that the "lim inf" of (10.2.21) is bounded from below by

$$\limsup_{n\to\infty} \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\mu}}_n - \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\beta}}_n.$$

Let us extract subsequences (still denoted by $\hat{\boldsymbol{\mu}}_n$, $\hat{\boldsymbol{\beta}}_n$) narrowly converging in $\mathscr{P}(X \times X \times X_\varpi)$ to $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\beta}} \in \mathscr{P}_p(X^3)$: by construction $\pi^{1,3}_{\#} \hat{\boldsymbol{\mu}} = \pi^{1,3}_{\#} \hat{\boldsymbol{\beta}}$ and, applying the next Lemma 10.2.8, we get

$$\pi^{1,2}_{\#} \hat{\boldsymbol{\beta}}_n = \pi^{1,2}_{\#} \boldsymbol{\beta}_n \to \boldsymbol{\mu}^{1\,2} \quad \text{in } \mathscr{P}_p(X \times X),$$

so that $\pi^{1,2}_{\#} \hat{\boldsymbol{\mu}} = \pi^{1,2}_{\#} \hat{\boldsymbol{\beta}} = \boldsymbol{\mu}^{1\,2}$; therefore, since $\boldsymbol{\mu}^{1\,2}$ is induced by a transport map, Lemma 5.3.2 gives that $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\beta}}$. By Lemma 5.2.4 we conclude that

$$\begin{aligned}&\limsup_{n\to\infty} \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\mu}}_n - \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\beta}}_n \\ &= \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\mu}} - \int_{X^3} \langle j_p(x_2 - x_1), x_3 \rangle \, d\hat{\boldsymbol{\beta}} = 0.\end{aligned}$$

□

As a corollary of the super-differentiability property (10.2.16) and of the sub-differentiability property (10.2.20) we obtain a differentiability property of the Wasserstein distance at regular measures.

Corollary 10.2.7 (Differentiability of W_p at regular measures). *If $\mu^1 \in \mathscr{P}^r_p(X)$ then for every $\mu^2 \in \mathscr{P}_p(X)$ the distance function $\psi : \mu \mapsto \frac{1}{p} W_p^p(\mu, \mu^2)$ satisfies*

$$\lim_{\mu^3 \to \mu^1} \frac{\psi(\mu^3) - \psi(\mu^1) + \int_X \langle j_p(\boldsymbol{t}_{\mu^1}^{\mu^2}(x) - x), \boldsymbol{t}_{\mu^1}^{\mu^3}(x) - x \rangle \, d\mu^1(x)}{W_p(\mu^3, \mu^1)} = 0. \qquad (10.2.22)$$

We prove now a result we used in the proof of Proposition 10.2.6.

Lemma 10.2.8 (Continuity of optimal plans). *Let $\mu^1, \mu^2 \in \mathscr{P}_p(X)$ and assume that $\Gamma_o(\mu^1,\mu^2) = \{\boldsymbol{\mu}^{1\,2}\}$. Then, for any choice of $\boldsymbol{\beta} \in \Gamma(\mu^1,\mu^2,\mu^3) \subset \mathscr{P}_p(X^3)$ with $\pi^{2,3}_\#\boldsymbol{\beta} \in \Gamma_o(\mu^2,\mu^3)$ we have*

$$\pi^{1,2}_\#\boldsymbol{\beta} \to \boldsymbol{\mu}^{1\,2} \quad \text{in } \mathscr{P}_p(X \times X) \quad \text{as } W_{p,\boldsymbol{\beta}}(\mu^1,\mu^3) \to 0.$$

Proof. Notice that the triangular inequality yields

$$\begin{aligned} W_{p,\boldsymbol{\beta}}(\mu^1,\mu^2) &\le W_{p,\boldsymbol{\beta}}(\mu^1,\mu^3) + W_p(\mu^2,\mu^3) \\ &\le W_{p,\boldsymbol{\beta}}(\mu^1,\mu^3) + W_p(\mu^1,\mu^2) + W_p(\mu^1,\mu^3) \\ &\le 2W_{p,\boldsymbol{\beta}}(\mu^1,\mu^3) + W_p(\mu^1,\mu^2) \end{aligned}$$

therefore any limit point of $\pi^{1,2}_\#\boldsymbol{\beta}$ (belonging to the $\mathscr{P}_p(X)$-compact set $\Gamma(\mu^1,\mu^2)$) as $W_{p,\boldsymbol{\beta}}(\mu^1,\mu^3) \to 0$ belongs to $\Gamma_o(\mu^1,\mu^2)$. Since $\Gamma_o(\mu^1,\mu^2) = \{\boldsymbol{\mu}^{1\,2}\}$ this proves the convergence of $\pi^{1,2}_\#\boldsymbol{\beta}$ to $\boldsymbol{\mu}^{1\,2}$. □

10.3 Subdifferential calculus in $\mathscr{P}_p(X)$: the general case

When one tries to extend the results of the previous Section 10.1 to functionals which should be "differentiable" on general (thus possibly not regular) probability measures, one realizes immediately that vector transport fields are no more sufficient to describe a satisfactory notion of subdifferential, even for convex functionals. There are at least two main reasons for that:

- *Minima μ_τ of quadratic perturbations* (10.1.1b) *cannot be pushed to the reference measure μ by a transport map:* thus the starting point (10.1.5) of point **A** is no more valid, in general. Notice that this property is essential to prove the existence of a minimal selection in $\partial\phi(\mu)$ when the metric slope $|\partial\phi|(\mu)$ is finite (point **D**).
- *The reference measure μ cannot be pushed to general "testing" measures ν by a transport $\boldsymbol{t}_\mu^\nu$*: thus the formal identification of the Euclidean difference vector $w - v$ with the displacement map $\boldsymbol{t}_\mu^\nu - \boldsymbol{i}$ is no longer available. Notice that this was an essential ingredient in Definition (10.1.2) and in the subsequent points **B, C, E.**

The above remarks suggest that *rescaled plans* with assigned first marginal μ should be used instead of vector fields to describe a useful notion of subdifferential. Of course, dealing with plans is less intuitive and notation becomes more complex; moreover, if for vector fields $\boldsymbol{t}$, $\boldsymbol{s} \in L^2(\mu;X)$ the scalar product $\langle \boldsymbol{t}, \boldsymbol{s}\rangle_{L^2(\mu;X)}$ is unambiguously defined, things become subtler when one tries to find an analogous coupling for two plans $\boldsymbol{\gamma}$, $\boldsymbol{\sigma}$ whose first marginals is μ.

Nevertheless, reasoning in terms of plans allow to recover all the main properties for a subdifferential theory, which we detailed at the beginning of the present chapter both in the Euclidean and in the $\mathscr{P}_2^r(X)$-case.

In this section we are thus considering a functional

$$\phi : \mathscr{P}_p(X) \to (-\infty, +\infty], \text{ proper and lower semicontinuous} \tag{10.3.1a}$$

such that

$$\begin{gathered}\nu \mapsto \Phi(\tau, \mu; \nu) = \frac{1}{p\tau^{p-1}} W_p^p(\mu, \nu) + \phi(\nu) \quad \text{admits at least} \\ \text{a minimum point } \mu_\tau, \text{ for all } \tau \in (0, \tau_*) \text{ and } \mu \in \mathscr{P}_p(X).\end{gathered} \tag{10.3.1b}$$

This condition is surely satisfied if ϕ is bounded below and lower semicontinuous w.r.t. narrow convergence of $\mathscr{P}(X_\varpi)$ on W_p-bounded sets. If $p = 2$ and ϕ is λ-convex along generalized geodesics, then lower semicontinuity w.r.t. W_2 is sufficient, thanks to Theorem 4.1.2.

In order to deal with the case $p \neq 2$ we introduce the set

$$\mathscr{P}_{pq}(X \times X) := \Big\{ \boldsymbol{\mu} \in \mathscr{P}(X \times X) : |\boldsymbol{\mu}|_{1,p} + |\boldsymbol{\mu}|_{2,q} < +\infty \Big\} \tag{10.3.2}$$

where, for $\boldsymbol{\mu} \in \mathscr{P}(X \times X)$, we defined

$$|\boldsymbol{\mu}|_{j,p}^p := \int_{X \times X} |x_j|^p \, d\boldsymbol{\mu}(x_1, x_2), \quad j = 1, 2, \ p > 1. \tag{10.3.3}$$

Recalling (7.1.12), we will say that a sequence $(\boldsymbol{\mu}_n) \subset \mathscr{P}_{pq}(X \times X)$ converges to $\boldsymbol{\mu}$ in $\mathscr{P}_{pq}(X \times X)$ as $n \to \infty$ if

$$\begin{gathered}\boldsymbol{\mu}_n \text{ narrowly converge to } \boldsymbol{\mu} \text{ in } \mathscr{P}(X \times X) \text{ and} \\ |\boldsymbol{\mu}_n|_{1,p} \to |\boldsymbol{\mu}|_{1,p}, \quad |\boldsymbol{\mu}_n|_{2,q} \to |\boldsymbol{\mu}|_{2,q} \quad \text{as } n \to \infty.\end{gathered} \tag{10.3.4}$$

By applying Theorem 5.1.13 it is easy to check that we can replace the first condition in (10.3.4) by the weaker one

$$\boldsymbol{\mu}_n \text{ narrowly converge to } \boldsymbol{\mu} \text{ in } \mathscr{P}(X_\varpi \times X_\varpi). \tag{10.3.5}$$

The above notion of convergence (10.3.4) is induced by a distance: e.g. we can take the sum of a distance inducing the narrow convergence in $\mathscr{P}(X \times X)$ and the p, q Wasserstein distances between the first and the second marginals of a given couple of plans $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathscr{P}_{pq}(X \times X)$. When $p = q$ this distance is equivalent to the p-Wasserstein distance in $\mathscr{P}_p(X^2)$.

Definition 10.3.1 (Extended Fréchet subdifferential). *Let $q = p' = \frac{p}{p-1}$, let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be a functional satisfying* (10.3.1a) *and let $\mu^1 \in D(\phi)$. We say that $\boldsymbol{\gamma} \in \mathscr{P}_{pq}(X \times X)$, belongs to the (extended) Fréchet subdifferential $\boldsymbol{\partial}\phi(\mu^1)$ if*

(i) $\pi^1_\# \boldsymbol{\gamma} = \mu^1$;

$$\text{(ii)}\quad \phi(\mu^3) - \phi(\mu^1) \ge \inf_{\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma},\mu^3)} \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu} + o\big(W_p(\mu^1, \mu^3)\big). \tag{10.3.6}$$

We say that $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu^1)$ *is also a* strong *Fréchet subdifferential if for every* $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\gamma}, \mu^3)$ *it satisfies the stronger condition*

$$\phi(\mu^3) - \phi(\mu^1) \ge \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu} + o\big(W_{p,\boldsymbol{\mu}}(\mu^1, \mu^3)\big). \tag{10.3.7}$$

Remark 10.3.2 (First variation along vector fields). If $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu^1)$ is a *strong* subdifferential, we can choose μ^3 and $\boldsymbol{\mu}$ of the type

$$\mu^3 := (x_1 + \varepsilon \boldsymbol{r}(x_1))_\# \mu^1, \quad \boldsymbol{\mu} = (x_1, x_2, x_1 + \varepsilon \boldsymbol{r}(x_1))_\# \boldsymbol{\gamma} \tag{10.3.8}$$

for an arbitrary vector field $\boldsymbol{r} \in L^p(\mu^1; X)$.

In this case $W_{p,\boldsymbol{\mu}}(\mu^3, \mu) = \varepsilon \|\boldsymbol{r}\|_{L^p(\mu;X)}$, and therefore we get a lower bound for the directional derivative of ϕ along $\boldsymbol{r}$:

$$\liminf_{\varepsilon \downarrow 0} \frac{\phi\big((\boldsymbol{i} + \varepsilon \boldsymbol{r})_\# \mu^1\big) - \phi(\mu^1)}{\varepsilon} \ge \int_{X^2} \langle x_2, \boldsymbol{r}(x_1) \rangle \, d\boldsymbol{\gamma}(x_1, x_2) \tag{10.3.9}$$

$$= \int_X \langle \bar{\boldsymbol{\gamma}}(x_1), \boldsymbol{r}(x_1) \rangle \, d\mu^1(x_1). \tag{10.3.10}$$

Observe that this property is stronger than the corresponding one (10.3.6) satisfied by a generic element of the Fréchet subdifferential of ϕ, since we are free to take variations along arbitrary vector fields, whereas (10.3.6) forces us to use only p-optimal transports. We will see that each minimizer of (10.3.1b) is a point of *strong* subdifferentiability: this is particularly useful in the case of functionals which are not geodesically convex (see e.g. [75]).
On the other hand, requiring (10.3.7) for general subdifferentials would induce a too strong notion, which would not satisfy in general the closure property of the next Lemma 10.3.8 even for λ-convex functionals. (10.3.9) will be related to extra properties of μ^1, and an important example will be provided by minimizers of (10.3.1b), as we will discuss in Lemma 10.3.4.

Remark 10.3.3 (Consistency). Suppose that $p = 2$ and that $\mu^1 \in \mathscr{P}_2^r(X)$; then $\boldsymbol{\xi} \in L^2(\mu^1; X)$ belongs to the Fréchet subdifferential $\partial\phi(\mu^1)$, according to Definition 10.1.1, if and only if

$$\boldsymbol{\gamma} = \big(\boldsymbol{i} \times \boldsymbol{\xi}\big)_\# \mu^1 \in \boldsymbol{\partial}\phi(\mu^1). \tag{10.3.11}$$

In fact, $\Gamma_o(\boldsymbol{\gamma}, \mu^3)$ contains the unique element $\boldsymbol{\mu} = \big(\boldsymbol{i} \times \boldsymbol{\xi} \times \boldsymbol{t}_{\mu^1}^{\mu^3}\big)_\# \mu^1$, so that the integral in (10.3.6) becomes

$$\inf_{\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma},\mu^3)} \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu} = \int_X \langle \boldsymbol{\xi}(x), \boldsymbol{t}_{\mu^1}^{\mu^3}(x) - x \rangle \, d\mu^1(x),$$

as in (10.1.3).

Moreover, if $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu^1)$ according to Definition 10.3.1, then its barycentric projection $\bar{\boldsymbol{\gamma}} \in L^2(\mu^1; X)$ according to (5.4.9) belongs to $\partial\phi(\mu^1)$. This property follows easily from the relation for $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma}, \mu^3)$

$$\int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu}(x_1, x_2, x_3) = \int_{X^2} \langle x_2, \boldsymbol{t}_{\mu^1}^{\mu^3}(x_1) - x_1 \rangle \, d\boldsymbol{\gamma}(x_1, x_2)$$
$$= \int_X \langle \bar{\boldsymbol{\gamma}}(x_1), \boldsymbol{t}_{\mu^1}^{\mu^3}(x_1) - x_1 \rangle \, d\mu^1(x_1).$$

Motivated by the above remark, we will introduce the shorter notation

$$\boldsymbol{\xi} \in \partial\phi(\mu^1) \quad \Longleftrightarrow \quad \boldsymbol{\xi} \in L^q(\mu^1; X), \quad (\boldsymbol{i} \times \boldsymbol{\xi})_{\#}\mu^1 \in \boldsymbol{\partial}\phi(\mu^1), \tag{10.3.12}$$

observing that, by Lemma 5.3.2, $\boldsymbol{\xi} \in L^q(\mu^1; X)$ belongs to $\partial\phi(\mu^1)$ if

$$\phi(\mu^2) - \phi(\mu^1) \geq \inf_{\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)} \int_{X^2} \langle \boldsymbol{\xi}(x_1), x_2 - x_1 \rangle \, d\boldsymbol{\mu} + o\big(W_p(\mu^1, \mu^2)\big), \tag{10.3.13}$$

and, also for general $p \in (1, +\infty)$,

$$\text{if } \mu \in \mathscr{P}_p^r(X), \ \boldsymbol{\gamma} \in \mathscr{P}_{pq}(X \times X), \quad (\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu) \quad \Longleftrightarrow \quad \bar{\boldsymbol{\gamma}} \in \partial\phi(\mu)). \tag{10.3.14}$$

With the notion introduced in Definition 10.3.1 we can now revisit the five properties **A,B,C,D,E** discussed at the beginning of this chapter. The starting point is an easy consequence of Theorem 10.2.2.

A. Euler equation for the Moreau-Yosida approximations.

Lemma 10.3.4. *Let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be satisfying* (10.3.1a) *and let μ_τ be a minimizer of* (10.3.1b)*; if $\hat{\boldsymbol{\gamma}}_\tau \in \Gamma_o(\mu_\tau, \mu)$, then the rescaled plans*

$$\boldsymbol{\gamma}_\tau := (\boldsymbol{\rho}_\tau)_{\#}\hat{\boldsymbol{\gamma}}_\tau \quad \textit{with} \quad \boldsymbol{\rho}_\tau(x_1, x_2) := \Big(x_1, j_p\big(\frac{x_2 - x_1}{\tau}\big)\Big) \tag{10.3.15}$$

and the associated plans $\boldsymbol{\mu}_\tau \in \Gamma(\boldsymbol{\gamma}_\tau, \mu^3)$ satisfy

$$\phi(\mu^3) - \phi(\mu_\tau) - \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu}_\tau \geq o(W_{p,\boldsymbol{\mu}_\tau}(\mu^3, \mu_\tau)). \tag{10.3.16}$$

In particular, restricting (10.3.16) *to plans $\boldsymbol{\mu}_\tau \in \Gamma_o(\boldsymbol{\gamma}_\tau, \mu^3)$, we get*

$$\boldsymbol{\gamma}_\tau \in \boldsymbol{\partial}\phi(\mu_\tau) \tag{10.3.17}$$

and $\boldsymbol{\gamma}_\tau$ is also a strong *subdifferential, according to* (10.3.7).

Proof. If μ_τ is a minimizer of (10.3.1b), Theorem 10.2.2 yields for $\mu^1 := \mu_\tau$, $\mu^2 := \mu$, and for every $\mu^3 \in \mathscr{P}_p(X)$ and $\hat{\boldsymbol{\mu}}_\tau \in \Gamma(\hat{\boldsymbol{\gamma}}_\tau, \mu^3)$

$$\phi(\mu^3) - \phi(\mu_\tau) \geq -\Big(\frac{1}{p\tau^{p-1}} W_p^p(\mu^3, \mu) - \frac{1}{p\tau^{p-1}} W_p^p(\mu_\tau, \mu)\Big)$$
$$\geq \int_{X^3} \langle j_p\big(\frac{x_2 - x_1}{\tau}\big), x_3 - x_1 \rangle \, d\hat{\boldsymbol{\mu}}_\tau - o(W_{p,\hat{\boldsymbol{\mu}}_\tau}(\mu^3, \mu_\tau)),$$

which is exactly (10.3.16), after the appropriate rescaling (10.3.15). □

Observe that in the previous corollary $\boldsymbol{\gamma}_\tau$ belongs to the set of all the rescaled optimal plans, whose first marginal is μ_τ; moreover

$$\int_{X^2} |x_2|^q d\boldsymbol{\gamma}_\tau = \int_{X^2} \left|\frac{x_2 - x_1}{\tau}\right|^p d\hat{\boldsymbol{\gamma}}_\tau = \Big(\frac{W_p^p(\mu_\tau, \mu)}{\tau^p}\Big) < +\infty. \tag{10.3.18}$$

Remark 10.3.5 ($\partial\phi(\mu_\tau)$ is not empty). We shall see at the end of Section 10.4.6 that, at least for $p = 2$, there exists a rescaled plan $\boldsymbol{\gamma}_\tau^\circ = (\boldsymbol{\rho}_\tau)_\# \hat{\boldsymbol{\gamma}}_\tau^\circ$ for some $\hat{\boldsymbol{\gamma}}_\tau^\circ \in \Gamma_o(\mu_\tau, \mu)$ whose barycenter $\bar{\boldsymbol{\gamma}}_\tau^\circ$ is a strong subdifferential in $\partial\phi(\mu_\tau)$: it is characterized by the minimum condition

$$\tau \|\bar{\boldsymbol{\gamma}}_\tau^\circ\|_{L^2(\mu_\tau;X)} = \min\Big\{ \|\bar{\boldsymbol{\gamma}}_\tau - \boldsymbol{i}\|_{L^2(\mu_\tau;X)} : \hat{\boldsymbol{\gamma}}_\tau \in \Gamma_o(\mu_\tau, \mu)\Big\}. \tag{10.3.19}$$

In particular $\partial\phi(\mu_\tau)$ is not empty.

10.3.1 The case of λ-convex functionals along geodesics

As in Section 10.1, we turn now our attention to λ-(geodesically) convex functionals. We already recalled in Section 9.1 what "convexity" here means.

B. Characterization by Variational inequalities and monotonicity

Theorem 10.3.6. *Let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be a proper, lower semicontinuous and λ-convex functional. A plan $\boldsymbol{\gamma} \in \mathscr{P}_{pq}(X \times X)$ belongs to $\boldsymbol{\partial}\phi(\mu^1)$ if and only if*

(i) $\pi^1_\# \boldsymbol{\gamma} = \mu^1$;

(ii) *for any $\mu^3 \in \mathscr{P}_p(X)$ there exists $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma}, \mu^3)$ satisfying*

$$\phi(\mu^3) - \phi(\mu^1) \geq \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu} + \frac{\lambda}{2} W_p^2(\mu^1, \mu^3). \tag{10.3.20}$$

Moreover, (10.3.20) *holds for every plan $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma}, \mu^3)$ such that ϕ is λ-convex along $\pi^{1,2}_\# \boldsymbol{\mu}$. For every couple of subdifferentials $\boldsymbol{\gamma}^i \in \boldsymbol{\partial}\phi(\mu^i)$, $i = 1, 2$, there exists a plan $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) \subset \mathscr{P}(X^2 \times X^2)$ such that $\pi^{1,3}_\# \boldsymbol{\gamma} \in \Gamma_o(\mu^1, \mu^2)$ and*

$$\int_{X^4} \langle x_2 - x_4, x_1 - x_3 \rangle \, d\boldsymbol{\mu} \geq \lambda W_p^2(\mu^1, \mu^3). \tag{10.3.21}$$

Proof. (10.3.20) directly yields (10.3.6); conversely, if (10.3.6) holds, we fix $\mu^3 \in D(\phi)$ and we apply (10.3.6) to the measures $\mu_t^{1\to 3}$ induced by some plan $\boldsymbol{\mu}^{1\,3} \in \Gamma_o(\mu^1, \mu^3)$ along which ϕ is λ-convex. Thus we find plans $\hat{\boldsymbol{\mu}}_t \in \Gamma_o(\boldsymbol{\gamma}, \mu_t^{1\to 3})$ such that

$$\phi(\mu_t^{1\to 3}) - \phi(\mu^1) \geq \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\hat{\boldsymbol{\mu}}_t + o(t) = t \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu}_t + o(t),$$

where $\boldsymbol{\mu}_t \in \Gamma_o(\boldsymbol{\gamma}, \mu^3)$ is defined by the relation $\hat{\boldsymbol{\mu}}_t := (\pi_t^{1,2,1\to 3})_\# \boldsymbol{\mu}_t$. On the other hand, the λ-convexity of ϕ yields for

$$\begin{aligned}\phi(\mu^3) - \phi(\mu^1) &\geq \frac{\phi(\mu_t^{1\to 3}) - \phi(\mu^1) + \frac{\lambda}{2} t(1-t) W_p^2(\mu^1, \mu^3)}{t} \\ &\geq \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu}_t + \frac{\lambda}{2}(1-t) W_p^2(\mu^1, \mu^3) + o(1).\end{aligned}$$

Passing to the limit as $t \downarrow 0$ we get (10.3.20) from Lemma 5.2.4, where $\boldsymbol{\mu}$ is any limit point of $\boldsymbol{\mu}_t$ in $\mathscr{P}(X \times X_\varpi \times X)$ as $t \downarrow 0$ (recall Lemma 5.2.2).

(10.3.21) follows by the same argument, simply inverting the role of μ^3 (which now is called μ^2) and μ^1 : notice that for a given optimal plan $\boldsymbol{\mu}^{1\,3} \in \Gamma_o(\mu^1, \mu^2)$ along which ϕ is λ-convex, we can always find a 4-plan $\boldsymbol{\mu}$ such that

$$\pi^{1,2}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^1, \quad \pi^{3,4}_\# \boldsymbol{\mu} = \boldsymbol{\gamma}^2, \quad \pi^{1,3}_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{1\,3}. \quad \square$$

Remark 10.3.7. The proof shows that if $\boldsymbol{\mu}^{1\,3} \in \Gamma_o(\mu^1, \mu^3)$ is an optimal plan along which ϕ is λ-convex, we can always choose $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma}, \mu^3)$ in (10.3.20) such that $\pi^{1,3}_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{1\,3}$.

C. Convexity, strong-weak closure, and Γ-convergence. The following lemma extends Lemma 10.1.3 to the more general setting of subdifferential plans; we also take account of a varying family of functionals ϕ_h which are $\Gamma(\mathscr{P}_p(X))$-convergent to ϕ as $n \to \infty$, as in (9.1.4), (9.1.5).

Lemma 10.3.8 (Closure of the subdifferential). *Let $\phi_h : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be λ-geodesically functionals which $\Gamma(\mathscr{P}_p(X))$-converge to ϕ as $h \to \infty$. If*

$$\begin{aligned}&\boldsymbol{\gamma}_h \in \boldsymbol{\partial}\phi_h(\mu_h), \quad \mu_h \to \mu \quad \textit{in } \mathscr{P}_p(X), \quad \mu \in D(\phi) \\ &\sup_h |\boldsymbol{\gamma}_h|_{2,q} < +\infty, \quad \boldsymbol{\gamma}_h \to \boldsymbol{\gamma} \quad \textit{in } \mathscr{P}(X \times X_\varpi),\end{aligned} \tag{10.3.22}$$

then $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu)$.

Proof. By (9.1.5) for a given $\mu^3 \in D(\phi)$ we can find a sequence μ_h^3 converging to μ^3 in $\mathscr{P}_p(X)$ such that $\phi_h(\mu_h^3) \to \phi(\mu^3)$ as $h \to \infty$. Theorem 10.3.6 yields plans $\boldsymbol{\mu}_h \in \Gamma_o(\boldsymbol{\gamma}_h, \mu_h^3)$ such that

$$\phi_h(\mu_h^3) - \phi_h(\mu_h) \geq \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu}_h + \frac{\lambda}{2} W_p^2(\mu_h, \mu_h^3).$$

Let $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma}, \mu^3)$ be a limit point in $\mathscr{P}(X \times X_\varpi \times X)$ of $\boldsymbol{\mu}_h$ (its existence follows by Lemma 5.2.2 together with Lemma 5.1.12). We wish to pass to the limit in this inequality. To this aim, notice that the upper limit of the first side is less than $\phi(\mu^3) - \phi(\mu^1)$, thanks to (9.1.4), therefore it suffices to show that

$$\lim_{h \to \infty} \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu}_h = \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\mu}. \tag{10.3.23}$$

Since

$$\int_{X^3} \langle x_2, x_3 - x_1\rangle \, d\boldsymbol{\mu}_h = \int_{X^3} \langle x_2, x_3\rangle \, d\pi^{2,3}_{\#}\boldsymbol{\mu}_h - \int_{X^3} \langle x_2, x_1\rangle \, d\boldsymbol{\mu}_h$$

and the same decomposition can be done for $\boldsymbol{\mu}$, we may apply Lemma 5.2.4 first to $\pi^{2,3}_{\#}\boldsymbol{\mu}_h$ (whose second marginal is μ_h^3) and then to the bounded sequence $\boldsymbol{\mu}_h$ (whose first marginals is μ_h): since the these marginals are converging in $\mathscr{P}_p(X)$ and therefore have uniformly integrable p-moments, we obtain (10.3.23). □

As a consequence one obtains also that for lower semicontinuous functionals the graph of $\boldsymbol{\partial}\phi(\mu)$ is closed w.r.t. narrow convergence in $\mathscr{P}(X \times X_\varpi)$ along pq-bounded sequences.

10.3.2 Regular functionals

We can introduce a property analogous to 10.1.4 even in the case of the extended subdifferential.

Definition 10.3.9 (Regular functionals). *A functional $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ satisfying* (10.3.1a) *is* regular *if whenever the* strong *subdifferentials $\boldsymbol{\gamma}_n \in \boldsymbol{\partial}\phi(\mu_n)$, $\varphi_n = \phi(\mu_n)$ satisfy*

$$\begin{aligned} \varphi_n \to \varphi \in \mathbb{R}, \quad & \mu_n \to \mu \quad \textit{in } \mathscr{P}_p(X), \\ \sup_n |\boldsymbol{\gamma}_n|_{2,q} < +\infty, \quad & \boldsymbol{\gamma}_n \to \boldsymbol{\gamma} \quad \textit{in } \mathscr{P}(X \times X_\varpi), \end{aligned} \tag{10.3.24}$$

then $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu)$ and $\varphi = \phi(\mu)$.

D. Minimal selection and slope

Theorem 10.3.10 (Metric slope and subdifferential). *Let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be a regular functional satisfying* (10.3.1a,b)*. Then $\mu \in D(|\partial\phi|)$ if and only if $\boldsymbol{\partial}\phi(\mu)$ is not empty and we have*

$$|\partial\phi|(\mu) = \min\Big\{ |\boldsymbol{\gamma}|_{2,q} : \boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu) \Big\} \qquad \forall \mu \in D(|\partial\phi|) = D(\boldsymbol{\partial}\phi). \tag{10.3.25}$$

Moreover, if μ_τ is a minimizer of (10.3.1b) *and $\boldsymbol{\gamma}_\tau$ is defined as in Lemma* 10.3.4*, then there exists a vanishing sequence $\tau_n \to 0$ such that*

$$|\partial\phi|^q(\mu) = \lim_{n\to\infty} \frac{W_p^p(\mu, \mu_{\tau_n})}{{\tau_n}^p} = \lim_{n\to\infty} \frac{\phi(\mu) - \phi(\mu_{\tau_n})}{\tau_n} = \lim_{n\to\infty} |\boldsymbol{\gamma}_{\tau_n}|^q_{2,q}. \tag{10.3.26}$$

Finally if τ_n is a vanishing sequence satisfying (10.3.26)*, then any limit point $\boldsymbol{\gamma}$ of the (relatively compact) family $(\boldsymbol{\gamma}_{\tau_n})$ in $\mathscr{P}(X \times X_\varpi)$ is a minimizer of* (10.3.25) *and it is also a limit point in the topology of $\mathscr{P}_{pq}(X \times X)$. When $p = 2$ the same result holds for the sequence of strong subdifferentials $\bar{\boldsymbol{\gamma}}^\circ_{\tau_n} \in \partial\phi(\mu_{\tau_n})$, provided by Remark* 10.3.5.

Proof. Let us first prove that

$$|\partial\phi|(\mu) \le |\boldsymbol{\gamma}|_{2,q} \quad \forall\, \boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu). \tag{10.3.27}$$

This follows directly from (10.3.6), since if $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu)$ for each $\mu^3 \in D(\phi)$ and $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\gamma}, \mu^3)$ satisfying (10.3.6) we get the estimate

$$\begin{aligned}\phi(\mu) - \phi(\mu^3) &\le \Big(\int |x_2|^q \, d\boldsymbol{\mu} \Big)^{1/q} \Big(\int |x_3 - x_1|^p \, d\boldsymbol{\mu} \Big)^{1/p} + o\big(W_p(\mu, \mu^3)\big) \\ &= |\boldsymbol{\gamma}|_{2,q} W_p(\mu, \mu^3) + o\big(W_p(\mu, \mu^3)\big),\end{aligned}$$

which is independent on the choice of $\boldsymbol{\mu}$. Dividing by $W_p(\mu, \mu^3)$ and passing to the limit as $\mu^3 \to \mu$ we get (10.3.27).

Conversely, let $\mu \in D(|\partial\phi|)$ and let us denote by μ_τ a minimizer of (10.3.1b). If $\boldsymbol{\gamma}_\tau \in \boldsymbol{\partial}\phi(\mu_\tau)$ is defined as in Lemma 10.3.4, we know by Remark 3.1.7 and (10.3.18) that $\boldsymbol{\gamma}_\tau$ is a *strong* subdifferential and (10.3.26) holds for a suitable vanishing subsequence $\tau_n \to 0$. Since $\mu_\tau \to \mu$ in $\mathscr{P}_p(X)$ as $\tau \downarrow 0$, the regularity of ϕ ensures that any limit point $\boldsymbol{\gamma}$ in $\mathscr{P}(X \times X_\varpi)$ of the family $\boldsymbol{\gamma}_{\tau_n}$ as $n \to \infty$ belongs to $\boldsymbol{\partial}\phi(\mu)$. By the lower semicontinuity of the map $\boldsymbol{\gamma} \mapsto |\boldsymbol{\gamma}|_{2,q}$ with respect to narrow convergence in $\mathscr{P}(X \times X_\varpi)$, we obtain $|\boldsymbol{\gamma}|_{2,q} \le |\partial\phi|(\mu)$ which, combined with (10.3.27) yields that $\boldsymbol{\gamma}$ is a minimizer of (10.3.25). Applying Theorem 5.1.13 to the second marginal of $\boldsymbol{\gamma}_{\tau_n}$ we conclude.
The argument for $\bar{\boldsymbol{\gamma}}^\circ_{\tau_n}$ is completely analogous. □

When one considers vectors instead of plans, it is easy to show that there exists a unique selection of minimal norm in $\partial\phi(\mu)$ by an argument of strict convexity of the norm. In the case of plans, this result is no more obvious, since the map $\boldsymbol{\gamma} \mapsto |\boldsymbol{\gamma}|_{2,q}$ is linear along convex combination. One can try to circumvent this difficulty by considering convex *interpolation* of plans, but it is not clear if $\boldsymbol{\partial}\phi(\mu)$ is stable under this kind of interpolation. On the other hand, strong subdifferentials are closed under interpolation, so that suitably combining interpolation and approximation, we can prove that the minimal selection is unique even for plan subdifferentials.

Theorem 10.3.11 (Minimal selection). *Let ϕ be regular functional satisfying* (10.3.1a,b), *and let $\mu \in D(\boldsymbol{\partial}\phi)$. There exists a unique plan $\boldsymbol{\gamma}_0 \in \boldsymbol{\partial}\phi(\mu)$ which attains the minimum*

$$|\boldsymbol{\gamma}_0|_{2,q} = \min\Big\{ |\boldsymbol{\gamma}|_{2,q} : \ \boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu) \Big\} = |\partial\phi|(\mu). \tag{10.3.28}$$

Consequently $\boldsymbol{\gamma}_0$ is the unique narrow limit point in $\mathscr{P}(X \times X_\varpi)$ and in $\mathscr{P}_{pq}(X \times X)$ of any family (of strong *subdifferentials, according to* (10.3.7)*) $\boldsymbol{\gamma}_{\tau_n}$ (when $p = 2$ we can also choose the barycenters $\bar{\boldsymbol{\gamma}}^\circ_{\tau_n}$ as in Remark* 10.3.5*) satisfying the asymptotic property* (10.3.26) *of the previous theorem, and we will denote it by the symbol $\boldsymbol{\partial}^\circ\phi(\mu)$.*

Proof. Let $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu)$ be attaining the minimum in (10.3.25) (the existence of minimizers is a direct consequence of the regularity of ϕ and a compactness argument in $\mathscr{P}(X \times X_\varpi)$ based on Lemma 5.2.2 and Lemma 5.1.12), let μ_τ be a minimizer of (10.3.1b); by the definition of subdifferential, we can find plans $\hat{\boldsymbol{\mu}}_\tau \in \Gamma_o(\boldsymbol{\gamma}, \mu_\tau)$ such that

$$\phi(\mu_\tau) - \phi(\mu) \geq \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\hat{\boldsymbol{\mu}}_\tau + o\big(W_p(\mu, \mu_\tau)\big). \tag{10.3.29}$$

We rescale $\hat{\boldsymbol{\mu}}_\tau$ as

$$\boldsymbol{\mu}_\tau = \big(\pi^1, \pi^2, \tau^{-1}(\pi^1 - \pi^3)\big)_\# \hat{\boldsymbol{\mu}}_\tau$$

and we consider a limit point $\boldsymbol{\mu} \in \mathscr{P}_p(X^3)$ of $\boldsymbol{\mu}_{\tau_n}$ in $\mathscr{P}(X \times X \times X_\varpi)$, $\tau_n \to 0$ being a vanishing sequence satisfying (10.3.26).

By the previous theorem we know that

$$\lim_{n\to\infty} \frac{\phi(\mu) - \phi(\mu_{\tau_n})}{\tau_n} = |\partial\phi|^q(\mu) = |\boldsymbol{\gamma}|_{2,q}^q = \int_{X^3} |x_2|^q \, d\boldsymbol{\mu} = \int_{X^3} |x_3|^p \, d\boldsymbol{\mu},$$

and by (10.3.29) and Lemma 5.2.4

$$\lim_{n\to\infty} \frac{\phi(\mu) - \phi(\mu_{\tau_n})}{\tau_n} \leq \int_{X^3} \langle x_2, x_3 \rangle \, d\boldsymbol{\mu},$$

so that

$$\int_{X^3} \Big(\frac{1}{q}|x_2|^q + \frac{1}{p}|x_3|^p - \langle x_2, x_3 \rangle\Big) \, d\boldsymbol{\mu} \leq 0, \tag{10.3.30}$$

i.e. $x_2 = j_p(x_3)$ for $\boldsymbol{\mu}$-a.e. $(x_1, x_2, x_3) \in X^3$. It follows that all the sequence $\boldsymbol{\mu}_{\tau_n}$ is converging to $\boldsymbol{\mu}$ and $\big(\pi^1, j_p \circ \pi^3\big)_\# \boldsymbol{\mu}_{\tau_n}$ is converging to $\boldsymbol{\gamma}$ in $\mathscr{P}(X \times X)$. We observe that

$$\boldsymbol{\gamma}_\tau := \big(\pi^1 - \tau\pi^3, j_p \circ \pi^3\big)_\# \boldsymbol{\mu}_\tau \in \boldsymbol{\partial}\phi(\mu_\tau) \quad \text{is a rescaled optimal plan as in (10.3.15),}$$

and $\boldsymbol{\gamma}_{\tau_n}$ has the same limit points in $\mathscr{P}_p(X \times X_\varpi)$ of $\big(\pi^1, j_p \circ \pi^3\big)_\# \boldsymbol{\mu}_{\tau_n}$ by Lemma 5.2.1; therefore $\boldsymbol{\gamma}_{\tau_n}$ converges to $\boldsymbol{\gamma}$ in $\mathscr{P}(X \times X)$.

Let us now suppose that $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \boldsymbol{\partial}\phi(\mu)$ attain the minimum in (10.2.13). We thus find two families $\boldsymbol{\gamma}_{i,\tau_n} \in \boldsymbol{\partial}\phi(\mu_{\tau_n})$ such that

$$\boldsymbol{\gamma}_{i,\tau_n} \to \boldsymbol{\gamma}_i, \quad |\boldsymbol{\gamma}_{i,\tau_n}|_{2,q} \to |\boldsymbol{\gamma}_i|_{2,q} = |\partial\phi|(\mu) \quad \text{as } n \to \infty.$$

Being $\boldsymbol{\gamma}_{i,\tau}$ *strong* subdifferentials, the next lemma shows that for every 3-plan $\boldsymbol{\nu}_{\tau_n}$ such that $\pi^{1,2}_\# \boldsymbol{\nu}_{\tau_n} = \boldsymbol{\gamma}_{1,\tau_n}$, $\pi^{1,3}_\# \boldsymbol{\nu}_{\tau_n} = \boldsymbol{\gamma}_{2,\tau_n}$ the interpolated plan

$$\boldsymbol{\gamma}_{1/2,\tau_n} := (\pi^{1,2\to3}_{1/2})_\# \boldsymbol{\nu}_{\tau_n}$$

still is a strong subdifferential in $\boldsymbol{\partial}\phi(\mu)$. Being $\{\boldsymbol{\nu}_{\tau_n}\}_{n\in\mathbb{N}}$ a tight family, possibly extracting a vanishing subsequence $\tau_{n'}$ so that $\boldsymbol{\nu}_{\tau_{n'}} \to \boldsymbol{\nu}$ in $\mathscr{P}(X \times X_\varpi \times X_\varpi)$, we know that

$$\pi^{1,2}_\#\boldsymbol{\nu} = \boldsymbol{\gamma}_1,\ \pi^{1,3}_\#\boldsymbol{\nu} = \boldsymbol{\gamma}_2, \quad (\pi^{1,2\to 3}_{1/2})_\#\boldsymbol{\nu} \in \boldsymbol{\partial}\phi(\mu).$$

The uniform convexity of the L^p-norm and the minimality of $\boldsymbol{\gamma}_i$ yield $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$. □

Lemma 10.3.12 (Interpolation of strong subdifferentials). *If $\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3} \in \mathscr{P}_{pq}(X \times X)$ belong to the* strong *subdifferential of a functional ϕ at $\mu \in \mathscr{P}_p(X)$, then for every $\boldsymbol{\gamma} \in \mathscr{P}_p(X^3)$ and $t \in [0,1]$ we have*

$$\pi^{1,i}_\#\boldsymbol{\gamma} = \boldsymbol{\gamma}^{1\,i} \quad i = 2,3 \quad \Longrightarrow \quad \boldsymbol{\gamma}^{1,2\to 3}_t = (\pi^{1,2\to 3}_t)_\#\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu), \tag{10.3.31}$$

$\boldsymbol{\gamma}^{1,2\to 3}_t$ being also a strong *subdifferential.*

Proof. For $\mu^4 \in D(\phi)$ and $\boldsymbol{\mu}_t \in \Gamma(\boldsymbol{\gamma}^{1,2\to 3}_t, \mu^4)$, arguing as in Proposition 7.3.1, it is not difficult to construct a new plan $\boldsymbol{\mu} \in \mathscr{P}(X^4)$ such that

$$\pi^{1,2,3}_\#\boldsymbol{\mu} = \boldsymbol{\gamma}, \quad \big(\pi^1, \pi^{2\to 3}_t, \pi^4\big)_\#\boldsymbol{\mu} = \boldsymbol{\mu}_t.$$

Since

$$\pi^{1,2,4}_\#\boldsymbol{\mu} \in \Gamma(\boldsymbol{\gamma}^{1\,2}, \mu^4), \quad \pi^{1,3,4}_\#\boldsymbol{\mu} \in \Gamma(\boldsymbol{\gamma}^{1\,3}, \mu^4), \quad (\pi^{1,2\to 3,4}_t)_\#\boldsymbol{\mu} \in \Gamma(\boldsymbol{\gamma}^{1,2\to 3}_t, \mu^4),$$

applying (10.3.7) we get

$$\phi(\mu^4) - \phi(\mu) \geq \int_{X^4} \langle x_2, x_4 - x_1\rangle \, d\boldsymbol{\mu} + o\big(W_{p,\boldsymbol{\mu}}(\mu, \mu^4)\big),$$
$$\phi(\mu^4) - \phi(\mu) \geq \int_{X^4} \langle x_3, x_4 - x_1\rangle \, d\boldsymbol{\mu} + o\big(W_{p,\boldsymbol{\mu}}(\mu, \mu^4)\big),$$

so that

$$\begin{aligned}\phi(\mu^4) - \phi(\mu) &\geq \int_{X^4} \langle (1-t)x_2 + tx_3, x_4 - x_1\rangle \, d\boldsymbol{\mu} + o\big(W_{p,\boldsymbol{\mu}}(\mu, \mu^4)\big)\\ &= \int_{X^3} \langle y, x_4 - x_1\rangle \, d\boldsymbol{\mu}_t(x_1, y, x_4) + o\big(W_{p,\boldsymbol{\mu}_t}(\mu, \mu^4)\big).\end{aligned}$$

□

Remark 10.3.13 (The distinguished role of the minimal selection $\boldsymbol{\partial}^\circ\phi(\mu)$). The above theorem is particularly useful in combination with Remark 10.3.2: in many examples it shows that the minimal selection $\boldsymbol{\partial}^\circ\phi(\mu)$ enjoys both the variational inequalities characterization (10.3.20) along optimal transports and a directional derivative inequality like (10.3.9) along general smooth vector fields.
This last property is not a consequence of general abstract conditions (like convexity for (10.3.20)), but it can directly checked by approximating $\boldsymbol{\partial}^\circ\phi(\mu)$ as in Theorem 10.3.11 and showing that the differential properties provided by (10.3.9) on the approximating sequence pass to the limit if the vector field $\boldsymbol{r}$ is sufficiently regular.

Remark 10.3.14 (A refined convergence result for λ-convex functionals). If the regular functional ϕ is λ-geodesically convex in $\mathscr{P}_p(X)$ according to Definition 9.1.1,and satisfies *(10.3.1a,b)*, the whole rescaled family $(\boldsymbol{\gamma}_\tau)_{0<\tau<1/\lambda^-}$ considered in Theorem 10.3.10 and 10.3.11 satisfies

$$|\partial\phi|^q(\mu) = \lim_{\tau\downarrow 0} \frac{W_p^p(\mu,\mu_\tau)}{\tau^p} = \lim_{\tau\downarrow 0} \frac{\phi(\mu)-\phi(\mu_\tau)}{\tau} = \lim_{\tau\downarrow 0} |\boldsymbol{\gamma}_\tau|^q_{2,q} \tag{10.3.32}$$

and therefore is converging to $\boldsymbol{\gamma} = \boldsymbol{\partial}^\circ\phi(\mu)$ in $\mathscr{P}_p(X\times X)$ as $\tau\downarrow 0$: it is sufficient to apply the estimates of Theorem 3.1.6 and Remark 3.1.7.

Combining Theorem 10.3.11 with (10.3.14) we show that for measures $\mu \in \mathscr{P}_p^r(X)\cap D(|\partial\phi|)$ the minimal selection $\boldsymbol{\partial}^\circ\phi(\mu)$ is induced by a (unique) transport map in the cotangent space of μ, that we call $\partial^\circ\phi(\mu)$.

Corollary 10.3.15 ($\boldsymbol{\partial}^\circ\phi(\mu) = \partial^\circ\phi(\mu)$ if μ is regular). *If ϕ is a regular functional and $\mu \in \mathscr{P}_p^r(X)\cap D(|\partial\phi|)$ then*

$$\{(\boldsymbol{i}\times\boldsymbol{\xi})_\#\mu\} = \boldsymbol{\partial}^\circ\phi(\mu) \tag{10.3.33}$$

for some map $\boldsymbol{\xi}$ with $j_q(\boldsymbol{\xi}) \in \mathrm{Tan}^r_\mu\mathscr{P}_p(X) = \mathrm{Tan}_\mu\mathscr{P}_p(X)$. We denote this vector by $\partial^\circ\phi(\mu)$.

Proof. If $\boldsymbol{\gamma}\in\boldsymbol{\partial}^\circ\phi(\mu)$ then by (10.3.14) the plan $(\boldsymbol{i}\times\bar{\boldsymbol{\gamma}})_\#\mu$ belongs to $\boldsymbol{\partial}\phi(\mu)$ and therefore

$$\int_{X^2} |x_2|^q\,d\boldsymbol{\gamma}(x_1,x_2) = \int_X \Big(\int_X |x_2|^q\,d\boldsymbol{\gamma}_{x_1}(x_2)\Big)\,d\mu(x_1) \geq \int_X |\bar{\boldsymbol{\gamma}}(x_1)|^q\,d\mu(x_1).$$

The minimality of $\boldsymbol{\gamma}$ and the usual strict convexity argument yield $\boldsymbol{\gamma}_{x_1} = \delta_{\bar{\boldsymbol{\gamma}}(x_1)}$ for μ-a.e. $x_1\in X$, i.e. $\boldsymbol{\gamma} = (\boldsymbol{i}\times\bar{\boldsymbol{\gamma}})_\#\mu$.

To show that $j_q(\bar{\boldsymbol{\gamma}})$ belongs to $\mathrm{Tan}^r_\mu\mathscr{P}_p(X)$ we observe that by the regularity of μ the rescaled plans $\boldsymbol{\gamma}_\tau$ introduced in Lemma 10.3.4 and Theorem 10.3.10 are given by

$$\boldsymbol{\gamma}_\tau = \Big(\boldsymbol{t}_\mu^{\mu_\tau}\times j_p\big(\frac{\boldsymbol{i}-\boldsymbol{t}_\mu^{\mu_\tau}}{\tau}\big)\Big)_\#\mu.$$

Since, by Theorem 10.3.11, $\boldsymbol{\gamma}_\tau$ narrowly converge in $\mathscr{P}(X\times X_\varpi)$ to $\boldsymbol{\gamma}$ as $\tau\downarrow 0$, choosing test functions of the form $\varphi(x_1)x_2$ with φ Lipschitz, we easily get

$$j_p\big(\frac{\boldsymbol{i}-\boldsymbol{t}_\mu^{\mu_\tau}}{\tau}\big) \to \bar{\boldsymbol{\gamma}} \ \text{ in the duality with Lipschitz functions.}$$

On the other hand, since $|\boldsymbol{\gamma}_\tau|_{2,q}\to|\boldsymbol{\gamma}|_{2,q}$ we have also

$$\lim_{\tau\downarrow 0}\Big\| j_p\big(\frac{\boldsymbol{i}-\boldsymbol{t}_\mu^{\mu_\tau}}{\tau}\big)\Big\|_{L^q(\mu;X)} = \|\bar{\boldsymbol{\gamma}}\|_{L^q(\mu;X)}$$

and therefore the two informations together give that $j_p\big(\tau^{-1}(\boldsymbol{i}-\boldsymbol{t}_\mu^{\mu_\tau})\big)\to\bar{\boldsymbol{\gamma}}$ in $L^q(\mu;X)$. By applying the duality map j_q we obtain that $\tau^{-1}(\boldsymbol{i}-\boldsymbol{t}_\mu^{\mu_\tau})\to j_q(\bar{\boldsymbol{\gamma}})$ in $L^p(\mu;X)$, so that $j_q(\bar{\boldsymbol{\gamma}})\in\mathrm{Tan}^r_\mu\mathscr{P}_p(X)$. □

Let us now consider a sequence of functionals (ϕ_h) which is $\Gamma(\mathscr{P}_p(X))$-converging to ϕ; if ϕ_h are λ-geodesically convex, Lemma 10.3.8 shows that limits of Fréchet subdifferentials of ϕ_h are Fréchet subdifferentials of ϕ. In many situations a converse approximation results would be useful, too; in other words, it would interesting to know if a given plan $\boldsymbol{\gamma} \in \boldsymbol{\partial}\phi(\mu)$ can be approximated by a sequence of plans $\boldsymbol{\gamma}_h \in \boldsymbol{\partial}\phi_h(\mu_h)$.

If $\boldsymbol{\gamma}$ is the minimal selection $\boldsymbol{\partial}^\circ\phi(\mu)$ and we reinforce a little bit the convergence assumption on ϕ_h, this approximation can always be performed, and we can also find an approximating sequence $\boldsymbol{\gamma}_h$ of *strong* subdifferentials. We can thus reproduce in the Wasserstein setting the same result which in the Euclidean case follows from the convergence of (ϕ_h) in the sense of Mosco (i.e. with different topologies in the lim inf inequality (9.1.4) and the lim sup inequality (9.1.5)).

Lemma 10.3.16. *Suppose that $\phi_h : \mathscr{P}_p(X) \to (-\infty, +\infty]$ is a sequence of functionals which satisfy* (10.3.1b) *(for some $\tau_* > 0$ independent of h) and the equicoercivity-like condition*

$$\inf_{\substack{\nu\in\mathscr{P}_p(X)\\ h\in\mathbb{N}}} \left\{\phi_h(\nu) + \frac{1}{p\tau_*^{p-1}} W_p^p(\bar\mu, \nu)\right\} > -\infty \tag{10.3.34}$$

for some $\bar\mu \in \mathscr{P}_p(X)$. Assume that ϕ is a proper regular functional which is the limit of ϕ_h in the sense that

$$\left.\begin{array}{l}\mu_h \to \mu \text{ in } \mathscr{P}(X_\varpi),\\ \sup_{h\in\mathbb{N}} \int_X |x|^p\, d\mu_h < +\infty\end{array}\right\} \implies \liminf_{h\to\infty} \phi_h(\mu_h) \ge \phi(\mu), \tag{10.3.35}$$

$$\forall\, \mu \in \mathscr{P}_p(X) \quad \exists \mu_h \to \mu \text{ in } \mathscr{P}_p(X): \quad \lim_{h\to\infty} \phi_h(\mu_h) = \phi(\mu). \tag{10.3.36}$$

Then ϕ satisfies (10.3.1a,b) *and for every $\mu \in D(\boldsymbol{\partial}\phi)$ there exist a sequence (μ_h) converging to μ in $\mathscr{P}_p(X)$ and* strong *subdifferentials $\boldsymbol{\gamma}_h \in \boldsymbol{\partial}\phi_h(\mu_h)$ such that*

$$\boldsymbol{\gamma}_h \to \boldsymbol{\partial}^\circ\phi(\mu) \quad \text{in } \mathscr{P}_{pq}(X\times X) \quad \text{as } h\to\infty. \tag{10.3.37}$$

The proof is based on the next typical Γ-convergence lemma, ensuring convergence of minimizers to minimizers and convergence of the extremal values.

Lemma 10.3.17. *Under the same assumptions of Lemma* 10.3.16*, for a given sequence $(\mu^h) \subset \mathscr{P}_p(X)$ converging to μ in $\mathscr{P}_p(X)$ and $\tau \in (0, \tau_*)$ such that* (10.3.34) *holds, let us consider sequences μ_τ^h, $\boldsymbol{\gamma}_\tau^h$ such that*

$$\mu_\tau^h \text{ is a minimum for } \quad \nu \mapsto \phi_h(\nu) + \frac{1}{p\tau^{p-1}} W_p^p(\mu^h, \nu), \tag{10.3.38}$$

and $\boldsymbol{\gamma}_\tau^h \in \boldsymbol{\partial}\phi(\mu_\tau^h)$ is obtained by rescaling an optimal plan $\hat{\boldsymbol{\gamma}}_\tau^h \in \Gamma_o(\mu_\tau^h, \mu^h)$ as in Lemma 10.3.4*:*

$$\boldsymbol{\gamma}_\tau^h := (\boldsymbol{\rho}_\tau)_\# \hat{\boldsymbol{\gamma}}_\tau^h \quad \text{with} \quad \boldsymbol{\rho}_\tau(x_1, x_2) := \Big(x_1, j_p\big(\frac{x_2 - x_1}{\tau}\big)\Big). \tag{10.3.39}$$

Then the families $\{\mu_\tau^h\}_{h\in\mathbb{N}}$, $\{\boldsymbol{\gamma}_\tau^h\}_{h\in\mathbb{N}}$, *and* $\{\hat{\boldsymbol{\gamma}}_\tau^h\}_{h\in\mathbb{N}}$ *are relatively compact in* $\mathscr{P}_p(X)$, $\mathscr{P}_p(X\times X)$, *and in* $\mathscr{P}_{pq}(X\times X)$ *respectively, Furthermore, for any* $\boldsymbol{\gamma}_\tau = \lim_i \boldsymbol{\gamma}_\tau^{h_i}$, *the measure* $\mu_\tau := \pi^1_\#\boldsymbol{\gamma}_\tau$ *minimizes* (10.3.1b) *and*

$$\lim_{i\to\infty}\phi_{h_i}(\mu_\tau^{h_i}) = \phi(\mu_\tau), \qquad \lim_{i\to\infty} W_p(\mu_\tau^{h_i},\mu^{h_i}) = W_p(\mu_\tau,\mu). \tag{10.3.40}$$

Proof. (10.3.34) and the estimate (2.2.4) (see also Remark 2.2.4) yield that μ_τ^h is bounded in $\mathscr{P}_p(X)$ and therefore the sequence $\hat{\boldsymbol{\gamma}}_\tau^h$ is narrowly relatively compact in $\mathscr{P}(X_\varpi\times X)$. Let $\hat{\boldsymbol{\gamma}}_\tau$ be a limit point of $\hat{\boldsymbol{\gamma}}_\tau^{h_i}$ as $i\to\infty$ and let $\mu_\tau := \pi^1_\#\hat{\boldsymbol{\gamma}}_\tau = \pi^1_\#\boldsymbol{\gamma}_\tau$; we choose $\nu\in D(\phi)$ and a sequence (ν^h) converging to ν in $\mathscr{P}_p(X)$ such that $\phi_h(\nu^h)\to\phi(\nu)$ as in (10.3.36). Passing to the limit in the inequality

$$\phi_h(\mu_\tau^h) + \frac{1}{p\tau^{p-1}}W_p^p(\mu_\tau^h,\mu^h) \le \phi_h(\nu^h) + \frac{1}{p\tau^{p-1}}W_p^p(\nu^h,\mu^h)$$

with $n = h_i$, using (10.3.35) and the lower semicontinuity of the Wasserstein distance w.r.t. narrow convergence in $\mathscr{P}_p(X_\varpi)$ (see Lemma 7.1.4), we get

$$\begin{aligned}\phi(\mu_\tau) + \frac{1}{p\tau^{p-1}}W_p^p(\mu_\tau,\mu) &\le \limsup_{h\to\infty}\phi_h(\mu_\tau^h) + \frac{1}{p\tau^{p-1}}W_p^p(\mu_\tau^h,\mu^h)\\ &\le \phi(\nu) + \frac{1}{p\tau^{p-1}}W_p^p(\nu,\mu).\end{aligned} \tag{10.3.41}$$

This shows that μ_τ minimizes (10.3.1b). Choosing $\nu=\mu_\tau$, the same argument provides convergence in energy, i.e.

$$\lim_{i\to\infty}\phi_{h_i}(\mu_\tau^{h_i}) + \frac{W_p^p(\mu_\tau^{h_i},\mu)}{p\tau^{p-1}} = \phi(\mu_\tau) + \frac{W_p^p(\mu_\tau,\mu)}{p\tau^{p-1}}.$$

But since the two terms are separately lower semicontinuous we obtain (10.3.40).

By applying (7.1.16), we obtain that the second marginals of the plans $\big(\pi^2,\pi^1-\pi^2\big)_\#\hat{\boldsymbol{\gamma}}_\tau^{h_i}$ are narrowly converging in $\mathscr{P}(X)$ and have uniformly integrable p-moments; since the first marginals (i.e. μ^h) are also converging in $\mathscr{P}_p(X)$, we obtain

$$\big(\pi^2,\pi^1-\pi^2\big)_\#\hat{\boldsymbol{\gamma}}_\tau^{h_i} \to \big(\pi^2,\pi^1-\pi^2\big)_\#\hat{\boldsymbol{\gamma}}_\tau \quad \text{in } \mathscr{P}_p(X\times X).$$

It follows that $\hat{\boldsymbol{\gamma}}_\tau^{h_i}\to\hat{\boldsymbol{\gamma}}_\tau$ in $\mathscr{P}_p(X\times X)$ and, as a consequence, and $\boldsymbol{\gamma}_\tau^{h_i}\to\boldsymbol{\gamma}_\tau$ in $\mathscr{P}_{pq}(X\times X)$ and $\mu_\tau^{h_i}\to\mu_\tau$ in $\mathscr{P}_p(X)$. □

Proof of Lemma 10.3.16. Let d be a distance in $\mathscr{P}_{pq}(X\times X)$ inducing the convergence (10.3.4); by the same construction of Proposition 5.1.8, Lemma 10.3.16 is equivalent to check that any open ball centered at $\boldsymbol{\gamma} = \boldsymbol{\partial}^\circ\phi(\mu)$ contains strong subdifferentials $\boldsymbol{\gamma}_h\in\boldsymbol{\partial}\phi(\mu^h)$ for sufficiently large h.

We argue by contradiction: thus we suppose that $\varepsilon>0$ and a sequence $h_i\to\infty$ exist such that

$$\boldsymbol{\gamma}_{h_i} \text{ is a strong subdifferential in } \boldsymbol{\partial}\phi_{h_i}(\mu_{h_i}) \quad\Longrightarrow\quad d(\boldsymbol{\gamma}_{h_i},\boldsymbol{\gamma})>\varepsilon. \tag{10.3.42}$$

We perform a diagonal argument (first keep τ fixed and let $h_i \to \infty$, then let $\tau \downarrow 0$): by the previous lemma we know that for any $\tau \in (0, \tau_*)$ the family $\{\boldsymbol{\gamma}_\tau^{h_i}\}_{i\in\mathbb{N}}$ (defined as in (10.3.38) and (10.3.39)) has a limit point $\boldsymbol{\gamma}_\tau$ in $\mathscr{P}_{pq}(X \times X)$. We take a sequence $\tau_n \to 0$ such that (10.3.26) is fulfilled by $\mu_{\tau_n} = \pi^1_\# \boldsymbol{\gamma}_{\tau_n}$ and by Theorem 10.3.11 we can find $\bar n \in \mathbb{N}$ such that $d(\boldsymbol{\gamma}_{\tau_{\bar n}}, \boldsymbol{\gamma}) < \varepsilon/2$. Since $\boldsymbol{\gamma}_{\tau_{\bar n}}$ is a limit point of $(\boldsymbol{\gamma}^{h_i}_{\tau_{\bar n}})$ in $\mathscr{P}_{pq}(X\times X)$ as $i \to \infty$, we get a contradiction with (10.3.42). $\square$

E. Chain rule. We conclude this section by proving a chain rule for functionals along absolutely continuous curves.

Proposition 10.3.18 (Chain rule). *Let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be a regular functional satisfying* (10.3.1a,b), *and let $\mu : (a,b) \mapsto \mu_t \in D(\phi) \subset \mathscr{P}_p(X)$ be an absolutely continuous curve with tangent velocity vector $\boldsymbol{v}_t$. Let $\Lambda \subset (a,b)$ be the set of points $t \in (a,b)$ such that*

(a) *$|\partial\phi|(\mu_t) < +\infty$;*

(b) *$\phi \circ \mu$ is approximately differentiable at t;*

(c) *condition* (8.4.6) *of Proposition* 8.4.6 *holds.*

Then

$$\frac{\widetilde{d}}{dt}\phi(\mu_t) = \int \langle x_2, \boldsymbol{v}_t(x_1)\rangle \, d\boldsymbol{\gamma}_t(x_1,x_2) \quad \forall\, \boldsymbol{\gamma}_t \in \boldsymbol{\partial}\phi(\mu_t),\ \forall\, t \in \Lambda. \tag{10.3.43}$$

Moreover, if ϕ is λ-convex and (10.1.17) *holds, then the map $t \mapsto \phi(\mu_t)$ is absolutely continuous and $(a,b)\setminus\Lambda$ is $\mathscr{L}^1$-negligible.*

Proof. We have simply to evaluate the time derivative of $\phi\circ\mu$ at a point $\bar t \in \Lambda$. We take $\boldsymbol{\gamma}_{\bar t} \in \boldsymbol{\partial}\phi(\mu_{\bar t})$ and $\hat{\boldsymbol{\mu}}_h \in \Gamma_o(\boldsymbol{\gamma}_{\bar t}, \mu_{\bar t + h})$ so that

$$\begin{aligned} \phi(\mu_{\bar t+h}) - \phi(\mu_{\bar t}) &\geq \int_{X^3} \langle x_2, x_3 - x_1\rangle \, d\hat{\boldsymbol{\mu}}_h + o(h) \\ &= h\int_{X^3} \langle x_2, x_3\rangle \, d\boldsymbol{\mu}_h + o(h), \end{aligned} \tag{10.3.44}$$

with $\boldsymbol{\mu}_h := \big(\pi^1, \pi^2, h^{-1}(\pi^3 - \pi^1)\big)_\# \hat{\boldsymbol{\mu}}_h$. We know by (8.4.6) that

$$\lim_{h\to 0} \pi^{1,3}_\# \boldsymbol{\mu}_h = \big(\boldsymbol{i} \times \boldsymbol{v}_{\bar t}\big)_\# \mu_{\bar t}, \quad \text{in } \mathscr{P}_p(X\times X)$$

and therefore, since $\pi^{1,2}_\# \boldsymbol{\mu}_h = \boldsymbol{\gamma}_{\bar t}$, we infer from Lemma 5.3.2 that

$$\lim_{h\to 0} \boldsymbol{\mu}_h = \big(x_1, x_2, \boldsymbol{v}_{\bar t}(x_1)\big)_\# \boldsymbol{\gamma}_{\bar t} \qquad \text{in } \mathscr{P}_p(X^3).$$

Therefore, dividing by h and passing to the limit in (10.3.44) we obtain that the approximate derivatives $\tilde d/dt_\pm \phi(\mu_t)$ satisfy

$$\frac{\tilde d}{dt_+}\phi(\mu_t)\Big|_{t=\bar t} \geq \int \langle x_2, \boldsymbol{v}_{\bar t}(x_1)\rangle \, d\boldsymbol{\gamma}_{\bar t}, \quad \frac{\tilde d}{dt_-}\phi(\mu_t)\Big|_{t=\bar t} \leq \int \langle x_2, \boldsymbol{v}_{\bar t}(x_1)\rangle \, d\boldsymbol{\gamma}_{\bar t}$$

and therefore we find (10.3.43).

In the convex case, since $|\partial\phi|$ is a strong upper gradient, we already know that $\phi(\mu_t)$ is absolutely continuous and thus $(a,b)\setminus\Lambda$ is $\mathscr{L}^1$-negligible. □

10.4 Example of subdifferentials

In this section we consider in the detail the subdifferential of the convex functionals presented in Chapter 9 (potential energy, interaction energy, internal energy, negative Wasserstein distance), with a particular attention to the characterization of the elements with minimal norm.

We start by considering a general, but *smooth*, situation.

10.4.1 Variational integrals: the smooth case

In order to clarify the underlying structure of many examples and the link between the notion of Wasserstein subdifferential and the standard variational calculus for integral functionals, we first consider the case of a variational integral of the type

$$\mathscr{F}(\mu) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} F(x,\rho(x),\nabla\rho(x))\,dx & \text{if } \mu=\rho\cdot\mathscr{L}^d \text{ with } \rho\in C^1(\mathbb{R}^d) \\ +\infty & \text{otherwise.}\end{cases} \tag{10.4.1}$$

Since we are not claiming any generality and we are only interested in the form of the subdifferential, we will assume enough regularity to justify all the computations; therefore, we suppose that $F:\mathbb{R}^d\times[0,+\infty)\times\mathbb{R}^d\to[0,+\infty)$ is a C^2 function with $F(x,0,p)=0$ for every x, $p\in\mathbb{R}^d$ and we consider the case of a smooth and strictly positive density ρ: as usual, we denote by $(x,z,p)\in\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d$ the variables of F and by $\delta\mathscr{F}/\delta\rho$ the first variation density

$$\frac{\delta\mathscr{F}}{\delta\rho}(x) := F_z(x,\rho(x),\nabla\rho(x)) - \nabla\cdot F_p(x,\rho(x),\nabla\rho(x)). \tag{10.4.2}$$

Lemma 10.4.1. *If $\mu=\rho\cdot\mathscr{L}^d\in\mathscr{P}_p^r(\mathbb{R}^d)$ with $\rho\in C^1(\mathbb{R}^d)$ satisfies $\mathscr{F}(\mu)<+\infty$ and $\boldsymbol{w}\in L^q(\mu;\mathbb{R}^d)$ belongs to the strong subdifferential of $\mathscr{F}$ at μ, then*

$$\boldsymbol{w}(x) = \nabla\frac{\delta\mathscr{F}}{\delta\rho}(x) \quad \textit{for } \mu\textit{-a.e. } x\in\mathbb{R}^d, \tag{10.4.3}$$

and for every vector field $\boldsymbol{\xi}\in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d}\boldsymbol{w}(x)\cdot\boldsymbol{\xi}(x)\,d\mu(x) = -\int_{\mathbb{R}^d}\frac{\delta\mathscr{F}}{\delta\rho}(x)\,\nabla\cdot\big(\rho(x)\boldsymbol{\xi}(x)\big)\,dx. \tag{10.4.4}$$

The same result holds if $p=2$, $\rho\in C^2_c(\mathbb{R}^d)$ and $\boldsymbol{w}\in\partial\mathscr{F}(\mu)\cap\mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$.

Proof. We take a smooth vector field $\boldsymbol{\xi} \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$ and we set for $\varepsilon \in \mathbb{R}$ sufficiently small $\mu_\varepsilon := (\boldsymbol{i}+\varepsilon\boldsymbol{\xi})_\#\mu$. If $\boldsymbol{w}$ is a strong subdifferential, by (10.3.10) we know that

$$\limsup_{\varepsilon\uparrow 0}\frac{\mathscr{F}(\mu_\varepsilon)-\mathscr{F}(\mu)}{\varepsilon}\le\int_{\mathbb{R}^d}\boldsymbol{w}(x)\cdot\boldsymbol{\xi}(x)\,d\mu(x)\le\liminf_{\varepsilon\downarrow 0}\frac{\mathscr{F}(\mu_\varepsilon)-\mathscr{F}(\mu)}{\varepsilon};\tag{10.4.5}$$

on the other hand, by Lemma 5.5.3 we know that $\mu_\varepsilon=\rho_\varepsilon\mathscr{L}^d$ with

$$\rho_\varepsilon(y)=\frac{\rho}{\det(I+\varepsilon\nabla\boldsymbol{\xi})}\circ\big(\boldsymbol{i}+\varepsilon\boldsymbol{\xi}\big)^{-1}(y)\quad\forall\,y\in\mathbb{R}^d.\tag{10.4.6}$$

The map $(x,\varepsilon)\mapsto\rho_\varepsilon(x)$ is of class C^2 with $\rho_\varepsilon(x)=\rho(x)$ outside a compact set and

$$\rho_\varepsilon(x)_{|\varepsilon=0}=\rho(x),\qquad\frac{\partial\rho_\varepsilon(x)}{\partial\varepsilon}_{|\varepsilon=0}=-\nabla\cdot\big(\rho(x)\boldsymbol{\xi}(x)\big).\tag{10.4.7}$$

Standard variational formulae (see e.g. [76, Vol. I, 1.2.1]) yield

$$\lim_{\varepsilon\to 0}\frac{\mathscr{F}(\mu_\varepsilon)-\mathscr{F}(\mu)}{\varepsilon}=-\int_{\mathbb{R}^d}\frac{\delta\mathscr{F}}{\delta\rho}(x)\,\nabla\cdot\big(\rho(x)\boldsymbol{\xi}(x)\big)\,dx,\tag{10.4.8}$$

which shows (10.4.4).

Let us now suppose that $p=2$ and $\boldsymbol{w}\in\partial\mathscr{F}(\mu)\cap\mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$; then (10.4.8) holds whenever $\boldsymbol{i}+\varepsilon\boldsymbol{\xi}$ is, an optimal transport map for $|\varepsilon|$ small enough, and in particular for gradient vector fields $\boldsymbol{\xi}=\nabla\zeta$ with $\zeta\in C_c^\infty(\mathbb{R}^d)$. Since $\mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ is the closure in $L^2(\mu;\mathbb{R}^d)$ of the space of such gradients, we have

$$\int_{\mathbb{R}^d}\boldsymbol{w}(x)\cdot\boldsymbol{\xi}(x)\,d\mu(x)=-\int_{\mathbb{R}^d}\nabla\frac{\delta\mathscr{F}}{\delta\rho}(x)\cdot\boldsymbol{\xi}(x)\,d\mu(x)\quad\forall\,\boldsymbol{\xi}\in\mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d).\tag{10.4.9}$$

We obtain (10.4.3) noticing that $\delta\mathscr{F}/\delta\rho\in\mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$, by the assumption that $\rho\in C_c^2(\mathbb{R}^d)$. □

10.4.2 The potential energy

Let $V:X\to(-\infty,+\infty]$ be a proper, l.s.c. and λ-convex functional (here it is sufficient to consider the case $\lambda\le 0$) and let $\mathcal{V}$ be the functional defined by (9.3.2) on $\mathscr{P}_p(X)$ (here $p\ge 2$ if $\lambda<0$). We denote by $\mathrm{graph}\,\partial V$ the graph of the Fréchet subdifferential of V in $X\times X$, i.e. the subset of the couples $(x_1,x_2)\in X\times X$ satisfying

$$V(x_3)\ge V(x_1)+\langle x_2,x_3-x_1\rangle+\frac{\lambda}{2}|x_1-x_2|^2\quad\forall\,x_3\in X.\tag{10.4.10}$$

As usual, $\partial^o V(x)$ denotes the element of minimal norm in $\partial V(x)$.

Proposition 10.4.2. *Let* $\boldsymbol{\gamma} \in \mathscr{P}_{pq}(X \times X)$ *with* $\mu = \pi^1_\# \boldsymbol{\gamma}$.

(i) $\boldsymbol{\gamma}$ *is a* strong *subdifferential of* $\mathcal{V}$ *at* μ *if and only if it satisfies* $\operatorname{supp} \boldsymbol{\gamma} \subset \operatorname{graph} \partial V$.

(ii) $\boldsymbol{\gamma} = \boldsymbol{\partial}^\circ \mathcal{V}(\mu)$ *if and only if* $\boldsymbol{\gamma}$ *is induced by the transport* $\boldsymbol{\xi} = \partial^\circ V$, *i.e.* $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{\xi})_\# \mu$ *where* $\boldsymbol{\xi}(x) = \partial^\circ V(x)$ *for* μ*-a.e.* $x \in X$*; in particular*

$$|\partial \phi|^q(\mu) = \int_X |\partial V|^q(x)\, d\mu(x) = \int_X |\partial^\circ V(x)|^q\, d\mu(x). \tag{10.4.11}$$

Proof. We suppose $\lambda = 0$ and $p = 2$: the proof of the general case can be obtained by obvious modifications.

(i) If $\boldsymbol{\gamma} \in \mathscr{P}_{pq}(X \times X)$ with $\operatorname{supp} \boldsymbol{\gamma} \subset \operatorname{graph} \partial V$ then for every $\mu^3 \in D(\mathcal{V})$ and $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\gamma}, \mu^3)$ (10.4.10) holds $\boldsymbol{\mu}$-a.e. in X^3 and therefore

$$\mathcal{V}(\mu^3) - \mathcal{V}(\mu) - \int_{X^3} \langle x_2, x_3 - x_1 \rangle\, d\boldsymbol{\mu} = \int_{X^3} \Big(V(x_3) - V(x_2) - \langle x_2, x_3 - x_1 \rangle \Big)\, d\boldsymbol{\mu} \geq 0.$$

Conversely, suppose that (10.3.7) holds: then choosing $\mu_3 \in D(\mathcal{V})$, $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\gamma}, \mu_3)$, $\boldsymbol{\mu}_t := (x_1, x_2, (1-t)x_1 + t x_3)_\# \boldsymbol{\mu}$, $\mu_t^{1 \to 3} := \pi^3_\# \boldsymbol{\mu}_t$, and recalling that $\mathcal{V}$ is convex along any interpolating plan, we have

$$\mathcal{V}(\mu^3) - \mathcal{V}(\mu^1) \geq \liminf_{t \downarrow 0} \frac{\mathcal{V}(\mu_t^{1\to 3}) - \mathcal{V}(\mu^1)}{t} \geq \int_{X^3} \langle x_2, x_3 - x_1 \rangle\, d\boldsymbol{\mu}. \tag{10.4.12}$$

If $\operatorname{supp} \boldsymbol{\gamma}$ is not a subset of $\operatorname{graph} \partial V$, by the lower semicontinuity of V we can find $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3 \in X$ and $\rho > 0$ such that

$$V(\hat{x}_3) < V(x_1) + \langle x_2, \hat{x}_3 - x_1 \rangle \quad \forall\, (x_1, x_2) \in R := B_\rho(\hat{x}_1) \times B_\rho(\hat{x}_2)$$

and $\boldsymbol{\gamma}(R) > 0$; thus, integrating the above relation in R with respect to $\boldsymbol{\gamma}$ yields

$$\int_R \Big(V(\hat{x}_3) - V(x_1) - \langle x_2, \hat{x}_3 - x_1 \rangle \Big)\, d\boldsymbol{\gamma}(x_1, x_2) < 0. \tag{10.4.13}$$

We introduce the map $\boldsymbol{r}$ equal to x_1 on $X^2 \setminus R$ and equal to $\hat{x}_3$ on R and we set $\mu^3 := \boldsymbol{r}_\# \boldsymbol{\gamma}$, $\boldsymbol{\mu} := (x_1, x_2, \boldsymbol{r}(x_1, x_2))_\# \boldsymbol{\gamma} \in \Gamma(\boldsymbol{\gamma}, \mu^3)$. Applying (10.4.12) we get

$$\begin{aligned}
0 &\leq \mathcal{V}(\mu^3) - \mathcal{V}(\mu^1) - \int_{X^3} \langle x_2, x_3 - x_1 \rangle\, d\boldsymbol{\mu}(x_1, x_2, x_3) \\
&= \int_{X^2} \Big(V(\boldsymbol{r}(x_1, x_2)) - V(x_1) - \langle x_2, \boldsymbol{r}(x_1, x_2) - x_1 \rangle \Big)\, d\boldsymbol{\gamma}(x_1, x_2) \\
&= \int_R \Big(V(\hat{x}_3) - V(x_1) - \langle x_2, \hat{x}_3 - x_1 \rangle \Big)\, d\boldsymbol{\gamma}(x_1, x_2),
\end{aligned}$$

which contradicts (10.4.13).

(ii) We now show (10.4.11), which in particular characterizes $\boldsymbol{\partial}^\circ\mathcal{V}(\mu)$. For $\tau > 0$ we call $\boldsymbol{r}_\tau$ the resolvent map which provides for any x the unique solution of the strictly convex minimization problem

$$\min_{y\in X} \frac{1}{2\tau}|y-x|^2 + V(y).$$

We set $\mu_\tau := (\boldsymbol{r}_\tau)_\#\mu$ and we check that μ_τ is the minimizer of (10.3.1b): for every $\nu \in D(\mathcal{V})$ and $\boldsymbol{\gamma} \in \Gamma_o(\mu,\nu)$ we have

$$\begin{aligned}\mathcal{V}(\nu) + \frac{1}{2\tau}W_2^2(\nu,\mu) &= \int_{X^2} V(x_2) + \frac{1}{2\tau}|x_2-x_1|^2\,d\boldsymbol{\gamma}\\ &\geq \int_X V(\boldsymbol{r}_\tau(x_1)) + \frac{1}{2\tau}|\boldsymbol{r}_\tau(x_1)-x_1|^2\,d\mu\\ &\geq \mathcal{V}(\mu_\tau) + \frac{1}{2\tau}W_2^2(\mu,\mu_\tau).\end{aligned}$$

Recalling that $|\partial\mathcal{V}|^2(\mu) = \lim_{\tau\downarrow 0} W_2^2(\mu,\mu_\tau)/\tau^2$ and that

$$\frac{|x-\boldsymbol{r}_\tau(x)|^2}{\tau^2} \leq |\partial^\circ V(x)|^2, \quad \lim_{\tau\downarrow 0}\frac{|x-\boldsymbol{r}_\tau(x)|^2}{\tau^2} = |\partial^\circ V(x)|^2,$$

we obtain

$$|\partial\mathcal{V}|^2(\mu) = \lim_{\tau\downarrow 0}\int_X \frac{|x-\boldsymbol{r}_\tau(x)|^2}{\tau^2}\,d\mu(x) = \int_X |\partial^\circ V(x)|^2\,d\mu(x). \qquad \square$$

Remark 10.4.3. It would not be difficult to show that if $V \in C^1(X)$ is a functional with bounded Fréchet derivatives, then $\mathcal{V}$ is regular and $\boldsymbol{\xi} = \partial^\circ\mathcal{V}(\mu)$ iff $\boldsymbol{\xi}(x) = \nabla V(x)$.

10.4.3 The internal energy

Let $\mathcal{F}$ be the functional

$$\mathcal{F}(\mu) := \begin{cases}\displaystyle\int_{\mathbb{R}^d} F(\rho(x))\,d\mathscr{L}^d(x) & \text{if } \mu = \rho\cdot\mathscr{L}^d \in \mathscr{P}_p^r(\mathbb{R}^d),\\ +\infty & \text{otherwise,}\end{cases} \tag{10.4.14}$$

for a convex differentiable function satisfying

$$F(0) = 0, \quad \liminf_{s\downarrow 0}\frac{F(s)}{s^\alpha} > -\infty \quad \text{for some } \alpha > \frac{d}{d+p} \tag{10.4.15}$$

as in Example 9.3.6. Recall that if F has superlinear growth at infinity then the functional $\mathcal{F}$ is l.s.c. with respect to the narrow convergence (indeed, under this growth condition the lower semicontinuity can be checked w.r.t. to the stronger weak L^1 convergence, by Dunford-Pettis theorem, and lower semicontinuity w.r.t. weak L^1 convergence is a direct consequence of the convexity on $\mathcal{F}$). More generally, in the case when F has a (sub-)linear growth, i.e.

$$\lim_{z\to+\infty}\frac{F(z)}{z}=\sup_{z>0}\frac{F(z)}{z}=\theta<+\infty, \tag{10.4.16}$$

we consider the lower semicontinuous envelope of $\mathcal{F}$, given by

$$\mathcal{F}^*(\mu)=\int_{\mathbb{R}^d}F(\rho)\,dx+\theta\mu_s(\mathbb{R}^d),\quad \mu=\rho\cdot\mathscr{L}^d+\mu_s,\quad \mu_s\perp\mathscr{L}^d. \tag{10.4.17}$$

We set $L_F(z)=zF'(z)-F(z):[0,+\infty)\to[0,+\infty)$ and we observe that L_F is strictly related to the *convex* function

$$G(z,s):=sF(z/s),\quad z\in[0,+\infty),\ s\in(0,+\infty), \tag{10.4.18}$$

since

$$\frac{\partial}{\partial s}G(z,s)=-\frac{z}{s}F'(z/s)+F(z/s)=-L_F(z/s). \tag{10.4.19}$$

In particular (recall that $F(0)=0$, by (10.4.15))

$$G(z,s)\le F(z)\quad\text{for } s\ge 1,\qquad \frac{F(z)-G(z,s)}{s-1}\uparrow L_F(z)\quad\text{as } s\downarrow 1. \tag{10.4.20}$$

We will also suppose that F satisfies condition (9.3.11), i.e.

$$\text{the map}\quad s\mapsto s^dF(s^{-d})\quad\text{is convex and non increasing in }(0,+\infty), \tag{10.4.21}$$

yielding the geodesic convexity of $\mathcal{F}$, $\mathcal{F}^*$.

The following lemma shows the existence of the directional derivative of $\mathcal{F}$ (or $\mathcal{F}^*$) along a suitable class of directions including all optimal transport maps.

Lemma 10.4.4 (Directional derivative of $\mathcal{F}^*$). *Suppose that $F:[0,+\infty)\to\mathbb{R}$ is a convex differentiable function satisfying* (10.4.21) *and* (10.4.15)*. Let $\mu=\rho\mathscr{L}^d+\mu_s\in D(\mathcal{F}^*)$, $\boldsymbol{r}\in L^p(\mu;\mathbb{R}^d)$ and $\bar t>0$ be such that*

(i) *$\boldsymbol{r}$ is approximately differentiable $\rho\mathscr{L}^d$-a.e. and $\boldsymbol{r}_t:=(1-t)\boldsymbol{i}+t\boldsymbol{r}$ is $\rho\mathscr{L}^d$-injective with $|\det\tilde\nabla\boldsymbol{r}_t(x)|>0$ $\rho\mathscr{L}^d$-a.e., for any $t\in[0,\bar t]$;*

(ii) *$\tilde\nabla\boldsymbol{r}_{\bar t}$ is diagonalizable with positive eigenvalues;*

(iii) *$\mu_s\perp\mathscr{L}^d$ and $(\boldsymbol{r}_t)_\#\mu_s\perp\mathscr{L}^d$ for any $t\in[0,\bar t]$;*

(iv) *$\mathcal{F}^*((\boldsymbol{r}_{\bar t})_\#\mu)<+\infty$.*

Then the map $t\mapsto t^{-1}\big(\mathcal{F}^*((\boldsymbol{r}_t)_\#\mu)-\mathcal{F}^*(\mu)\big)$ is nondecreasing in $[0,\bar t]$ and

$$+\infty>\lim_{t\downarrow 0}\frac{\mathcal{F}^*((\boldsymbol{r}_t)_\#\mu)-\mathcal{F}^*(\mu)}{t}=-\int_{\mathbb{R}^d}L_F(\rho)\mathrm{tr}\,\tilde\nabla(\boldsymbol{r}-\boldsymbol{i})\,dx. \tag{10.4.22}$$

The identity above still holds when assumptions (ii) *on $\boldsymbol{r}$ is replaced by* (ii′) $\|\tilde\nabla(\boldsymbol{r}-\boldsymbol{i})\|_{L^\infty(\rho\mathscr{L}^d;\mathbb{R}^{d\times d})}<+\infty$ *(in particular if* $\boldsymbol{r}-\boldsymbol{i}\in C^\infty_c(\mathbb{R}^d;\mathbb{R}^d)$*),*

and F *satisfies in addition the "doubling" condition*

$$\exists\, C>0: \quad F(z+w)\le C\big(1+F(z)+F(w)\big) \quad \forall\, z,\, w. \tag{10.4.23}$$

Proof. By assumptions (i) and (ii), taking into account Lemma 5.5.3 and the representation (10.4.17) of $\mathcal{F}^*$ we have

$$\begin{aligned}\mathcal{F}^*((\boldsymbol{r}_t)_\#\mu)-\mathcal{F}^*(\mu) &= \int_{\mathbb{R}^d} F\left(\frac{\rho(x)}{\det\tilde\nabla\boldsymbol{r}_t(x)}\right)\det\tilde\nabla\boldsymbol{r}_t(x)\,dx-\int_{\mathbb{R}^d}F(\rho(x))\,dx\\ &= \int_{\mathbb{R}^d}\Big(G(\rho(x),\det\tilde\nabla\boldsymbol{r}_t(x))-F(\rho(x))\Big)\,dx\end{aligned}$$

for any $t\in(0,\bar t]$. By the argument given in Proposition 9.3.9, (iii), assumption (10.4.21) together with (9.3.12) imply that the function

$$\frac{G(\rho(x),\det\tilde\nabla\boldsymbol{r}_t)-F(\rho(x))}{t} \qquad t\in(0,\bar t] \tag{10.4.24}$$

is nondecreasing w.r.t. t and bounded above by an integrable function (take $t=\bar t$ and apply (iv)). Therefore the monotone convergence theorem gives

$$\lim_{t\downarrow 0}\frac{\mathcal{F}^*((\boldsymbol{r}_t)_\#\mu)-\mathcal{F}^*(\mu)}{t}=\int_{\mathbb{R}^d}\frac{d}{dt}G(\rho(x),\det\tilde\nabla\boldsymbol{r}_t(x))\big|_{t=0}\,dx$$

and the expansion $\det\tilde\nabla\boldsymbol{r}_t=1+t\,\mathrm{tr}\,\tilde\nabla(\boldsymbol{r}-\boldsymbol{i})+o(t)$ together with (10.4.19) give the result.

In the case when (ii′) holds, the argument is analogous but, since condition (ii) fails, we cannot rely anymore on the monotonicity of the function in (10.4.24). However, using the inequalities

$$F(w)-F(0)\le wF'(w)\le F(2w)-F(w)$$

and the doubling condition we easily see that the derivative w.r.t. s of the function $G(z,s)$ can be bounded by $C(1+F^+(z))$ for $|s-1|\le 1/2$. Therefore we can use the dominated convergence theorem instead of the monotone convergence theorem to pass to the limit. □

The next technical lemma shows that we can "integrate by parts" in (10.4.22) preserving the inequality, if $L_F(\rho)$ is locally in $W^{1,1}$.

Lemma 10.4.5 (A "weak" integration by parts formula). *Under the same assumptions of Lemma* 10.4.4*, let us suppose that*

(i) $\mathrm{supp}\,\mu\subset\overline{\Omega}$, Ω *being a convex open subset of* $\mathbb{R}^d$ *(not necessarily bounded);*

(ii) $L_F(\rho)\in W^{1,1}_{\mathrm{loc}}(\Omega)$*;*

(iii) $K = \mathrm{supp}(\boldsymbol{r}_{\bar t})_{\#}\mu$ *is a compact subset of* Ω *for some* $\bar t \in (0,1]$*;*

(iv) $\boldsymbol{r}$ *satisfies the property* (ii) *of Theorem* 6.2.9*, i.e. there exists a sequence* $\boldsymbol{r}_h$ *of functions in* $BV_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ *such that* $\boldsymbol{r}_h(x) \in K$ *for* $\rho\mathscr{L}^d$*-a.e.* $x \in \mathbb{R}^d$ *and* $\lim_{h\to\infty} \mu\big(\{\boldsymbol{r}_h \neq \boldsymbol{r}\}\big) = 0$.

Then we can find an increasing family of nonnegative Lipschitz functions $\chi_k : \mathbb{R}^d \to [0,1]$ *with compact support in* Ω *such that* $\chi_k \uparrow \chi_\Omega$ *and*

$$-\int_{\mathbb{R}^d} L_F(\rho(x))\mathrm{tr}\,\tilde\nabla(\boldsymbol{r}-\boldsymbol{i})\,dx \geq \limsup_{k\to\infty}\int_{\mathbb{R}^d}\langle \nabla L_F(\rho), \boldsymbol{r}-\boldsymbol{i}\rangle \chi_k\,dx. \tag{10.4.25}$$

Proof. Possibly replacing $\boldsymbol{r}$ by $\boldsymbol{r}_{\bar t}$, we can assume that $\bar t = 1$ in (iii). Let us first recall that by Calderon-Zygmund theorem (see for instance [11]) for every vector field $\boldsymbol{s} \in BV_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ the approximate divergence $\mathrm{tr}\,(\tilde\nabla \boldsymbol{s})$ is the absolutely continuous part of the distributional divergence $D\cdot\boldsymbol{s}$; therefore we have

$$\int_{\mathbb{R}^d} v\,\mathrm{tr}\,(\tilde\nabla\boldsymbol{s})\,dx \leq -\int_{\mathbb{R}^d}\langle\nabla v,\boldsymbol{s}\rangle\,dx, \tag{10.4.26}$$

provided $D\cdot\boldsymbol{s}\geq 0$ and $v \in C^\infty_c(\mathbb{R}^d)$ is nonnegative. If $\boldsymbol{s}$ is bounded, by approximation the same inequality remains true for every nonnegative function $v \in W^{1,1}(\mathbb{R}^d)$. For every Lipschitz function $\eta : \mathbb{R}^d \to [0,1]$ with compact support in Ω and each function $\boldsymbol{s} = \boldsymbol{r}_h$, choosing $v := \eta L_F(\rho) \in W^{1,1}(\mathbb{R}^d)$ we get

$$\int_{\mathbb{R}^d}\big(\eta L_F(\rho)\big)\,\mathrm{tr}\,(\tilde\nabla\boldsymbol{r}_h)\,dx \leq -\int_{\mathbb{R}^d}\langle\nabla\big(\eta L_F(\rho)\big),\boldsymbol{r}_h\rangle\,dx; \tag{10.4.27}$$

taking into account that $\mu(\{\tilde\nabla\boldsymbol{r}\neq\tilde\nabla\boldsymbol{r}_h\})\to 0$ (because the approximate differentials coincide at points of density one of the coincidence set) and that $L_F(\rho)$ vanishes where ρ vanishes, we recover the inequality

$$\int_{\mathbb{R}^d}\big(\eta L_F(\rho)\big)\,\mathrm{tr}\,(\tilde\nabla\boldsymbol{r})\,dx \leq -\int_{\mathbb{R}^d}\langle\nabla\big(\eta L_F(\rho)\big),\boldsymbol{r}\rangle\,dx. \tag{10.4.28}$$

On the other hand, a standard integration by parts yields

$$\int_{\Omega}\big(\eta L_F(\rho)\big)\,\mathrm{tr}\,(\nabla\boldsymbol{i})\,dx = -\int_{\Omega}\langle\nabla\big(\eta L_F(\rho)\big),\boldsymbol{i}\rangle\,dx; \tag{10.4.29}$$

summing up with (10.4.28) and inverting the sign we find

$$-\int_{\mathbb{R}^d}\big(\eta L_F(\rho)\big)\,\mathrm{tr}\,(\tilde\nabla(\boldsymbol{r}-\boldsymbol{i}))\,dx \geq \int_{\mathbb{R}^d}\langle\nabla\big(\eta L_F(\rho)\big),\boldsymbol{r}-\boldsymbol{i}\rangle\,dx. \tag{10.4.30}$$

Now we choose carefully the test function η. We consider an increasing family bounded open convex sets Ω_k such that

$$\overline{\Omega_k} \subset\subset \Omega, \quad \Omega = \bigcup_{k=1}^{\infty}\Omega_k$$

and for each convex set Ω_k we consider the function

$$\chi_k(x) := k\, d(x, \mathbb{R}^d \setminus \Omega_k) \wedge 1. \tag{10.4.31}$$

χ_k is an increasing family of nonnegative Lipschitz functions which take their values in $[0,1]$ and satisfy $\chi_k(x) \equiv 1$ if $d(x, \mathbb{R}^d \setminus \Omega_k) \geq \frac{1}{k}$; in particular, $\chi_k \equiv 1$ in K for k sufficiently large. Moreover χ_k is concave in Ω_k, since the distance function $d(\cdot, \mathbb{R}^d \setminus \Omega_k)$ is concave. Choosing $\eta := \chi_k$ in (10.4.30) we get

$$\begin{aligned} -\int_{\mathbb{R}^d} \big(\chi_k L_F(\rho)\big) \operatorname{tr}(\tilde{\nabla}(\boldsymbol{r} - \boldsymbol{i}))\, dx &\geq \int_{\mathbb{R}^d} \langle \nabla L_F(\rho), \boldsymbol{r} - \boldsymbol{i}\rangle\, \chi_k\, dx \\ &\quad + \int_{\Omega_k} \langle \nabla \chi_k, \boldsymbol{r} - \boldsymbol{i}\rangle\, L_F(\rho)\, dx \qquad (10.4.32) \\ &\geq \int_{\mathbb{R}^d} \langle \nabla L_F(\rho), \boldsymbol{r} - \boldsymbol{i}\rangle\, \chi_k\, dx \end{aligned}$$

since the integrand of (10.4.32) is nonnegative: in fact, for $\mathscr{L}^d$-a.e. $x \in \Omega_k$ where $L_F(\rho(x))$ is strictly positive, the concavity of χ_k and $\boldsymbol{r}(x) \in K$ yields

$$\langle \nabla \chi_k(x), \boldsymbol{r}(x) - \boldsymbol{i}(x)\rangle \geq \chi_k(\boldsymbol{r}(x)) - \chi_k(x) = 1 - \chi_k(x) \geq 0.$$

Passing to the limit as $k \to \infty$ in the previous integral inequality, we obtain (10.4.25) (recall that the function in the left hand side of (10.4.25) is semiintegrable by (10.4.22)). □

In the following two theorems we characterize $\partial^\circ \mathcal{F}^*(\mu)$ and give (under the doubling condition, but see Remark 10.4.7) a formula for the slope of the functional, showing that $\nabla L_F(\rho)/\rho$ is the minimal selection in the subdifferential. Since $\mathcal{F}^* = \mathcal{F}$ in the superlinear case, we consider the functional $\mathcal{F}^*$ only.

Theorem 10.4.6 (Slope and subdifferential of $\mathcal{F}^*$). *Suppose that $F : [0, +\infty) \to \mathbb{R}$ is a convex differentiable function satisfying* (10.4.15), (10.4.21) *and* (10.4.23). *Assume that $\mathcal{F}^*$ has finite slope at $\mu \in \mathscr{P}_p(\mathbb{R}^d)$ with $\mu = \rho \cdot \mathscr{L}^d + \mu_s$ and $\mu_s \perp \mathscr{L}^d$. Then the following statements hold:*

(a) *$L_F(\rho) \in W^{1,1}(\mathbb{R}^d)$ and $\nabla L_F(\rho) = \boldsymbol{w}\rho$ for some function $\boldsymbol{w} \in L^q(\rho\mathscr{L}^d; \mathbb{R}^d)$. Moreover*

$$\Big(\int_{\mathbb{R}^d} |\boldsymbol{w}(x)|^q \rho(x)\, dx\Big)^{1/q} \leq |\partial \mathcal{F}^*|(\mu) < +\infty. \tag{10.4.33}$$

(b) *If $\mu \in \mathscr{P}_p^r(\mathbb{R}^d)$ then equality holds in* (10.4.33) *and $\boldsymbol{w} = \partial^\circ \mathcal{F}^*(\mu)$.*

Conversely, if $L_F(\rho) \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ and $\nabla L_F(\rho) = \boldsymbol{w}\rho$ for some $\boldsymbol{w} \in L^q(\mu; \mathbb{R}^d)$, then $\mathcal{F}^*$ has a finite slope at $\mu = \rho\mathscr{L}^d$ and $\boldsymbol{w} = \partial^\circ \mathcal{F}^*(\mu)$.

Proof. (a) We apply first (10.4.22) with $\boldsymbol{r} = 0$ $\rho\mathscr{L}^d$-a.e. and $\boldsymbol{r} = \boldsymbol{i}$ on a $\mathscr{L}^d$-negligible set on which μ^s is concentrated and take into account that

$$W_p(\mu, ((1-t)\boldsymbol{i} + t\boldsymbol{r})_{\#}\mu) \le t\|\boldsymbol{i}\|_{L^p(\rho\mathscr{L}^d;\mathbb{R}^d)}$$

to obtain

$$d\int_{\mathbb{R}^d} L_F(\rho)\,dx \le |\partial\mathcal{F}^*|(\mu)\|\boldsymbol{i}\|_{L^p(\rho\mathscr{L}^d;\mathbb{R}^d)},$$

so that $L_F(\rho) \in L^1(\mathbb{R}^d)$. Next, we apply (10.4.22) with $\boldsymbol{r}-\boldsymbol{i}$ equal to a $C_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$ function $\boldsymbol{t}$ $\rho\mathscr{L}^d$-a.e. (notice that condition (i) holds with $\bar{t} < \sup|\nabla\boldsymbol{t}|$) and equal to 0 on a $\mathscr{L}^d$-negligible set on which μ_s is concentrated, and use again the inequality $W_p(\mu, ((1-t)\boldsymbol{i} + t\boldsymbol{r})_{\#}\mu) \le t\|\boldsymbol{r}-\boldsymbol{i}\|_{L^p(\rho\mathscr{L}^d)}$ to obtain

$$\int_{\mathbb{R}^d} L_F(\rho)\mathrm{tr}\,(\nabla\boldsymbol{t})\,dx \le |\partial\mathcal{F}^*|(\mu)\|\boldsymbol{t}\|_{L^p(\rho\mathscr{L}^d)} \le |\partial\mathcal{F}^*|(\mu)\sup_{\mathbb{R}^d}|\boldsymbol{t}|,$$

having used also the fact that the approximate differential $\tilde{\nabla}(\boldsymbol{r}-\boldsymbol{i})$ (by definition) coincides with the classical differential $\nabla\boldsymbol{t}$ $\rho\mathscr{L}^d$-a.e. As $\boldsymbol{t}$ is arbitrary, Riesz theorem gives that $L_F(\rho)$ is a function of bounded variation (i.e. its distributional derivative $DL_F(\rho)$ is a finite $\mathbb{R}^d$-valued measure in $\mathbb{R}^d$), so that we can rewrite the inequality as

$$\left|\sum_{i=1}^d \int_{\mathbb{R}^d} \boldsymbol{t}_i\,dD_iL_F(\rho)\right| \le |\partial\mathcal{F}|(\mu)\|\boldsymbol{t}\|_{L^p(\rho\mathscr{L}^d;\mathbb{R}^d)}.$$

By L^p duality theory there exists $\boldsymbol{w} \in L^q(\rho\mathscr{L}^d;\mathbb{R}^d)$ with $\|\boldsymbol{w}\|_q \le |\partial\mathcal{F}|(\mu)$ such that

$$\sum_{i=1}^d \int_{\mathbb{R}^d} \boldsymbol{t}_i\,dD_iL_F(\rho) = \int_{\mathbb{R}^d} \langle\boldsymbol{w},\boldsymbol{t}\rangle\,d\rho\mathscr{L}^d \qquad \forall\boldsymbol{t} \in C_c^\infty(\mathbb{R}^d,\mathbb{R}^d).$$

Therefore $L_F(\rho) \in W^{1,1}(\mathbb{R}^d)$ and $\nabla L_F(\rho) = \boldsymbol{w}\rho$. This leads to the inequality $\le$ in (10.4.33).

(b) Assume now that $\mu \in \mathscr{P}_p^r(\mathbb{R}^d)$. In order to show that equality holds in (10.4.33) we will prove that $(\boldsymbol{i}\times\boldsymbol{w})_{\#}\mu$ belongs to $\boldsymbol{\partial}\mathcal{F}^*(\mu)$. We have to show that (10.1.7) holds for any $\nu \in \mathscr{P}_p(\mathbb{R}^d)$ and, by approximation, we can assume that $\nu \in \mathscr{P}_p^r(\mathbb{R}^d)$ and that $\mathcal{F}^*(\nu)$ is finite. Using the doubling condition it is also easy to find a sequence of measures ν_h with compact support converging to ν in $\mathscr{P}_p(X)$ and such that $\mathcal{F}^*(\nu_h)$ converges to $\mathcal{F}^*(\nu)$, hence we can also assume that $\mathrm{supp}\,\nu$ is compact. Setting $\boldsymbol{r} = \boldsymbol{t}_\mu^\nu$, by Theorem 6.2.7 and the argument in the beginning of Proposition 9.3.9, we know that all the conditions of Lemma 10.4.4 are fulfilled. Theorem 6.2.9 shows that also Lemma 10.4.5 holds; therefore, by applying (10.4.22), the geodesic convexity of $\mathcal{F}^*$, and (10.4.25) we obtain

$$\begin{aligned}\mathcal{F}^*(\nu) - \mathcal{F}^*(\mu) &\ge \limsup_{h\to\infty}\int_{\mathbb{R}^d}\langle\nabla L_F(\rho),(\boldsymbol{r}-\boldsymbol{i})\rangle\chi_h\,dx\\ &= \limsup_{h\to\infty}\int_{\mathbb{R}^d}\langle\boldsymbol{w},(\boldsymbol{r}-\boldsymbol{i})\rangle\chi_h\rho\,dx = \int_{\mathbb{R}^d}\langle\boldsymbol{w},\boldsymbol{r}-\boldsymbol{i}\rangle\,d\mu,\end{aligned}$$

proving that $\boldsymbol{w} \in \partial\mathcal{F}^*(\mu)$.

Finally, we notice that our proof that $\boldsymbol{w} = \nabla L_F(\rho)/\rho \in \partial\mathcal{F}^*(\mu)$ does not use the finiteness of slope, but only the assumption $\boldsymbol{w} \in L^q(\mu;\mathbb{R}^d)$, therefore these conditions imply that the subdifferential is not empty and that the slope is finite. □

Remark 10.4.7 (The non-doubling case). The doubling assumption seems to play an essential role in the previous proof, as it allows to differentiate the energy functionals along smooth and compactly supported directions. Notice also that the proof that the assumptions $L_F(\rho) \in W^{1,1}(\mathbb{R}^d)$ and $\nabla L_F(\rho) = \boldsymbol{w}\rho$ with $w \in L^q$ imply that $\boldsymbol{w}$ is in the subdifferential does not use the doubling condition. In the non-doubling case the characterization of the minimal subdifferential could still be obtained through a monotone approximation argument of F by doubling functions F_n (which yields indeed the Γ-convergence of the corresponding energy functionals) based on Lemma 10.3.16. This argument is explained in detail in a more relevant case for the applications, the entropy functional in infinite dimensions (see Theorem 10.4.17: in this case the approximating functionals are the entropies with respect to finite dimensional projections of the reference measure).

Let us now consider a particular class of functions F with sublinear growth: assuming that $\theta - F(z)/z \to 0$ sufficiently slowly as $z \to \infty$, we prove that finiteness of slope implies absolute continuity of the measure. This assumptions covers all power functions $-t^m$ with $m > 1 - 1/d$ (leaving open only the case $m = 1 - 1/d$, where still (10.4.21) holds).

Theorem 10.4.8 (Finiteness of slope implies regularity). *Let us suppose that F is a convex differentiable function in $[0,+\infty)$ satisfying* (10.4.15), (10.4.21), *and*

$$\lim_{z\to+\infty} z^{1/d}\Big(\theta - \frac{F(z)}{z}\Big) = +\infty. \tag{10.4.34}$$

If the metric slope $|\partial\mathcal{F}^*|(\mu)$ is finite at $\mu \in \mathscr{P}_p(X)$ then $\mu \in \mathscr{P}^r_p(X)$.

Proof. Let $\mu \in D(\mathcal{F}^*)$ be fixed, assume that $|\partial\mathcal{F}^*|(\mu) < \infty$, and write $\mu = \rho\mathscr{L}^d + \mu^s$, with μ^s singular with respect to $\mathscr{L}^d$. We call E a Borel $\mathscr{L}^d$-negligible set on which μ^s is concentrated, i.e.

$$E \in \mathscr{B}(\mathbb{R}^d), \quad \mu^s(\mathbb{R}^d \setminus E) = 0, \quad \mathscr{L}^d(E) = 0.$$

We now claim that μ is absolutely continuous. If not, let $Q = [0,1]^d$ and let $\boldsymbol{r}$ be the optimal transport map between $\mu_0 := \chi_Q\mathscr{L}^d$ and μ. By Theorem 6.2.7 we know that $\boldsymbol{r}$ is approximately differentiable and that $\tilde\nabla\boldsymbol{r}$ is diagonalizable with nonnegative eigenvalues μ_0-a.e.; the argument in the beginning of Proposition 9.3.9 shows that $(1-t)\boldsymbol{i} + t\boldsymbol{r}$ is μ_0-essentially injective for any $t \in [0,1)$.

We define $Q_1 := \boldsymbol{r}^{-1}(E)$ and $Q_2 = Q \setminus Q_1$. Since E is $\mathscr{L}^d$-negligible the area formula (5.5.2) gives $\det\tilde\nabla\boldsymbol{r} = 0$ $\mathscr{L}^d$-a.e. on Q_1. Notice also that $\mathscr{L}^d(Q_1)$ is the total mass of μ^s.

We define $\boldsymbol{r}_t := t\boldsymbol{i}+(1-t)\boldsymbol{r}$, $\mu_t := (\boldsymbol{r}_t)_\#\mu_0$ and $J_t := \det\tilde\nabla\boldsymbol{r}_t$. The concavity of the map $t\mapsto J_t^{1/d}$ and the fact that $J_0(x)\ge 0,\ J_1(x)=1$ $\mathscr{L}^d$-a.e. in Q yield

$$J_t(x)\ge t^d \quad \text{for } \mathscr{L}^d\text{-a.e. } x\in Q. \tag{10.4.35}$$

Therefore from (5.5.3) we get $\mu_t=(\rho_t^1+\rho_t^2)\mathscr{L}^d$ with

$$\rho_t^i=\frac{1}{J_t\circ\boldsymbol{r}_t^{-1}}|_{\boldsymbol{r}_t(Q_i)}.$$

Moreover, the μ_0-essential injectivity of $\boldsymbol{r}_t$ and the fact that $\mu_t\ll\mathscr{L}^d$ imply that $\boldsymbol{r}_t(Q_1)\cap\boldsymbol{r}_t(Q_2)$ is $\mathscr{L}^d$-negligible, so that we have the decomposition

$$\mathcal{F}^*(\mu_t)=\mathcal{F}(\mu_t)=\int_{Q_1}F\Big(\frac{1}{J_t(y)}\Big)J_t(y)\,dy+\int_{\boldsymbol{r}_t(Q_2)}F\Big(\frac{1}{J_t(\boldsymbol{r}_t^{-1}(x))}\Big)\,dx. \tag{10.4.36}$$

On the other hand, $\boldsymbol{r}|_{Q_2}$ is the optimal transport map between $\chi_{Q_2}\mathscr{L}^d$ and $\rho\mathscr{L}^d$ and ρ_t^2 is the value of the unique constant speed geodesic at time t, see Chapter 7 (here we apply the interpolation theory to pairs of measures whose common total mass is not necessarily 1).

Since $\rho\mathscr{L}^d$ is regular as well we can find the optimal transport map $\boldsymbol{s}$ between $\rho\mathscr{L}^d$ and $\chi_{Q_2}\mathscr{L}^d$ and setting $\boldsymbol{s}_t=((1-t)\boldsymbol{i}+t\boldsymbol{s})$, the uniqueness of geodesic interpolation gives $\rho_t^2\mathscr{L}^d=(\boldsymbol{s}_t)_\#\big(\rho\mathscr{L}^d\big)$, hence

$$\int_{\boldsymbol{r}_t(Q_2)}F\Big(\frac{1}{J_t(\boldsymbol{r}_t^{-1}(x))}\Big)\,dx=\mathcal{F}(\rho_t^2)=\mathcal{F}^*((\boldsymbol{s}_t)_\#\big(\rho\mathscr{L}^d\big))=\int_{\mathbb{R}^d}F\Big(\frac{\rho(x)}{\check J_t(x)}\Big)\check J_t(x)\,dx,$$

with $\check J_t:=\det\tilde\nabla\hat{\boldsymbol{s}}_t$ and

$$\begin{aligned}\frac{\mathcal{F}^*(\mu)-\mathcal{F}^*(\mu_t)}{t}&=t^{-1}\Big(\int_{\mathbb{R}^d}F(\rho(x))\,dx-\int_{\mathbb{R}^d}F\Big(\frac{\rho(x)}{\check J_t(x)}\Big)\check J_t(x)\,dx\Big)\\&\quad+t^{-1}\int_{Q_1}\Big(\theta-F\Big(\frac{1}{J_t(y)}\Big)J_t(y)\Big)\,dy.\end{aligned}$$

From Lemma 10.4.5 and (10.4.22) we get

$$\begin{aligned}-\infty<A:=&-\int_{\mathbb{R}^d}\langle\nabla L_F(\rho(x)),\boldsymbol{s}(x)-x\rangle\,dx\\&\le\lim_{t\downarrow0}t^{-1}\Big(\int_{\mathbb{R}^d}F(\rho(x))\,dx-\int_{\mathbb{R}^d}F\Big(\frac{\rho(x)}{\check J_t(x)}\Big)\check J_t(x)\,dx\Big).\end{aligned}$$

Passing to the limit as $t\downarrow 0$ and using the identity $W_p(\mu_t,\mu_0)=tW_p(\mu_0,\mu)$ we get

$$\limsup_{t\downarrow0}\int_{Q_1}t^{-1}\Big(\theta-F\Big(\frac{1}{J_t(y)}\Big)J_t(y)\Big)\,dy\le|\partial\mathcal{F}|(\mu)W_p(\mu_0,\mu)-A<+\infty.$$

We observe that if $\det \tilde\nabla \boldsymbol{r}(y) = 0$ we have $\lim_{t\downarrow 0} J_t(y) = 0$ and by (10.4.35) we get

$$\liminf_{t\downarrow 0} t^{-1}\Big(\theta - F\Big(\frac{1}{J_t(y)}\Big)J_t(y)\Big) \geq \liminf_{t\downarrow 0} J_t^{-1/d}(y)\Big(\theta - F\Big(\frac{1}{J_t(y)}\Big)J_t(y)\Big)$$
$$= \liminf_{z\uparrow+\infty} z^{1/d}\Big(\theta - F(z)/z\Big) = +\infty.$$

Since $\det \tilde\nabla \boldsymbol{r}(y) = 0$ for $\mathscr{L}^d$-a.e. $y \in Q_1$, Fatou's Lemma yields

$$\liminf_{t\downarrow 0} \int_{Q_1} t^{-1}\Big(\theta - F\Big(\frac{1}{J_t(y)}\Big)J_t(y)\Big)\,dy = +\infty$$

whenever $\mathscr{L}^d(Q_1) = \mu^s(\mathbb{R}^d) > 0$. □

10.4.4 The relative internal energy

In this section we briefly discuss the modifications which should be apported to the previous results, when one consider a relative energy functional as in Section 9.4.

We thus consider a log-concave probability measure $\gamma = e^{-V}\cdot\mathscr{L}^d \in \mathscr{P}(\mathbb{R}^d)$ induced by a convex l.s.c. potential

$$V : \mathbb{R}^d \to (-\infty,+\infty], \quad \text{with} \quad \Omega = \operatorname{int} D(V) \neq \emptyset. \tag{10.4.37}$$

We are also assuming that the energy density

$$\begin{gathered} F : [0,+\infty) \to [0,+\infty] \quad \text{is convex and l.s.c.,} \\ \text{it satisfies the doubling property (10.4.23),} \\ \text{and the geodesic convexity condition (9.3.19),} \end{gathered} \tag{10.4.38}$$

which yield that the map $s \mapsto \hat F(s) := F(e^{-s})e^s$ is convex and non increasing in $\mathbb{R}$. The functional

$$\mathcal{F}(\mu|\gamma) := \int_{\mathbb{R}^d} F(\sigma)\,d\gamma = \int_\Omega F\Big(\rho/e^{-V}\Big)e^{-V}\,dx, \qquad \mu = \sigma\cdot\gamma = \rho\mathscr{L}^d \tag{10.4.39}$$

is therefore geodesically convex in $\mathscr{P}_p(\mathbb{R}^d)$, by Theorem 9.4.12. It is easy to check that whenever $\hat F$ is not constant (case which corresponds to a linear F and a constant functional $\mathcal{F}$), F has a superlinear growth and therefore $\mathcal{F}$ is lower semicontinuous in $\mathscr{P}_p(X)$.

As already observed in Remark 10.4.7, the doubling property (10.4.38) could be avoided; here we are assuming it for the sake of simplicity.

Theorem 10.4.9 (Subdifferential of $\mathcal{F}(\cdot|\gamma)$). *The functional $\mathcal{F}(\cdot|\gamma)$ has finite slope at $\mu = \sigma\cdot\gamma \in D(\mathcal{F})$ if and only if $L_F(\sigma) \in W^{1,1}_{\rm loc}(\Omega)$ and $\nabla L_F(\sigma) = \sigma\boldsymbol{w}$ for some function $\boldsymbol{w} \in L^q(\mu;\mathbb{R}^d)$. In this case*

$$\Big(\int_{\mathbb{R}^d} |\boldsymbol{w}(x)|^q\,d\mu(x)\Big)^{1/q} = |\partial\mathcal{F}|(\mu), \tag{10.4.40}$$

and $\boldsymbol{w} = \partial^\circ\mathcal{F}(\mu)$.

Proof. We argue as in Theorem 10.4.6: in the present case the directional derivative formula (10.4.22) becomes

$$
\begin{aligned}
+\infty > \lim_{t\downarrow 0} & \frac{\mathcal{F}((\boldsymbol{r}_t)_\#\mu|\gamma) - \mathcal{F}(\mu|\gamma)}{t} \\
&= -\int_{\mathbb{R}^d} L_F(\rho/e^{-V})\Big(e^{-V}\operatorname{tr}\tilde{\nabla}(\boldsymbol{r}-\boldsymbol{i}) - e^{-V}\langle\nabla V, \boldsymbol{r}-\boldsymbol{i}\rangle\Big)\,dx \qquad (10.4.41)\\
&= -\int_{\mathbb{R}^d} L_F(\sigma)\operatorname{tr}\tilde{\nabla}\Big(e^{-V}(\boldsymbol{r}-\boldsymbol{i})\Big)\,dx
\end{aligned}
$$

for every vector field $\boldsymbol{r}$ satisfying the assumptions of Lemma 10.4.4 and $\mathcal{F}(\boldsymbol{r}_\#\mu|\gamma) < +\infty$. Choosing as before $\boldsymbol{r} = \boldsymbol{i} + e^V\boldsymbol{t}$, $\boldsymbol{t} \in C^\infty_c(\Omega;\mathbb{R}^d)$, since V is bounded in each compact subset of Ω, we get

$$\int_\Omega L_F(\sigma)\operatorname{tr}\nabla\boldsymbol{t}\,dx \le |\partial\mathcal{F}|(\mu)\sup_{\mathbb{R}^d}|e^V\boldsymbol{t}|,$$

so that $L_F(\sigma) \in BV_{\rm loc}(\Omega)$. Choosing now $\boldsymbol{r} = \boldsymbol{i} + \boldsymbol{t}$ with $\boldsymbol{t} \in C^\infty_c(\Omega;\mathbb{R}^d)$ we get

$$\left|\sum_{i=1}^d \int_\Omega \boldsymbol{t}_i\,dD_iL_F(\sigma)\,d\gamma\right| \le |\partial\mathcal{F}|(\mu)\|\boldsymbol{t}\|_{L^p(\mu;\mathbb{R}^d)}$$

so that there exists $\boldsymbol{w} \in L^p(\mu;\mathbb{R}^d)$ such that

$$\sum_{i=1}^d \int_\Omega \boldsymbol{t}_i\,dD_iL_F(\sigma)\,d\gamma = \int_{\mathbb{R}^d}\langle\boldsymbol{w},\boldsymbol{t}\rangle\,d\mu = \int_{\mathbb{R}^d}\langle\rho\boldsymbol{w},\boldsymbol{t}\rangle e^{-V}\,dx \quad \forall\,\boldsymbol{t} \in C^\infty_c(\Omega;\mathbb{R}^d),$$

thus showing that $L_F(\sigma) \in W^{1,1}_{\rm loc}(\Omega)$ and $\nabla L_F(\sigma) = \rho e^{-V}\boldsymbol{w} = \sigma\boldsymbol{w}$.

Conversely, if $L_F(\sigma) \in W^{1,1}_{\rm loc}(\Omega)$ with $\nabla L_F(\sigma) = \sigma\boldsymbol{w}$ and $\boldsymbol{w} \in L^q(\mu;\mathbb{R}^d)$, arguing as in Lemma 10.4.5 we have for every measure $\nu = \boldsymbol{r}_\#\mu$ with compact support in Ω

$$
\begin{aligned}
\mathcal{F}(\nu|\gamma) - \mathcal{F}(\mu|\gamma) &\ge \limsup_{k\to\infty} -\int_\Omega L_F(\sigma)\operatorname{tr}\tilde{\nabla}\big(e^{-V}(\boldsymbol{r}-\boldsymbol{i})\big)\chi_k\,dx\\
&\ge \limsup_{k\to\infty}\int_\Omega \langle\chi_k\nabla L_F(\sigma) + L_F(\sigma)\nabla\chi_k, \boldsymbol{r}-\boldsymbol{i}\rangle\,d\gamma\\
&\ge \limsup_{k\to\infty}\int_\Omega \langle\nabla L_F(\sigma), \boldsymbol{r}-\boldsymbol{i}\rangle\chi_k\,d\gamma\\
&\ge \limsup_{k\to\infty}\int_\Omega \langle\boldsymbol{w}, \boldsymbol{r}-\boldsymbol{i}\rangle\chi_k\,d\mu = \int_\Omega\langle\boldsymbol{w},\boldsymbol{r}-\boldsymbol{i}\rangle\,d\mu,
\end{aligned}
$$

which shows through a density argument that $\boldsymbol{w} \in \partial\mathcal{F}(\mu)$. □

10.4.5 The interaction energy

In this section we consider the interaction energy functional $\mathcal{W} : \mathscr{P}_p(X) \to [0, +\infty]$ defined by

$$\mathcal{W}(\mu) := \frac{1}{2} \int_{X \times X} W(x-y)\, d\mu \times \mu(x,y).$$

Without loss of generality we shall assume that $W : X \to [0, +\infty)$ is an even function; our main assumption, besides the convexity of W, is the doubling condition

$$\exists\, C_W > 0 : \quad W(x+y) \le C_W\big(1 + W(x) + W(y)\big) \quad \forall\, x,\, y \in X. \tag{10.4.42}$$

Let us first state a preliminary result: we are denoting by $\bar\mu$ the barycenter of the measure μ:

$$\bar\mu := \int_X x\, d\mu(x). \tag{10.4.43}$$

Lemma 10.4.10. *Assume that $W : X \to [0, +\infty)$ is convex, Gateaux differentiable, even, and satisfies the doubling condition* (10.4.42). *Then for any $\mu \in D(\mathcal{W})$ we have*

$$\int_X W(x)\, d\mu(x) \le C_W\big(1 + \mathcal{W}(\mu) + W(\bar\mu)\big) < +\infty, \tag{10.4.44}$$

$$\int_{X \times X} |\nabla W(x-y)|\, d\mu \times \mu(x,y) \le C_W\big(1 + S_W + \mathcal{W}(\mu)\big) < +\infty, \tag{10.4.45}$$

where $S_W := \sup_{|y| \le 1} W(y)$. In particular $\boldsymbol{w} := (\nabla W) * \mu$ is well defined for μ-a.e. $x \in X$, it belongs to $L^1(\mu; X)$, and it satisfies

$$\begin{aligned} \int_{X^2 \times X} &\langle \nabla W(x_1 - x_2), y_1 - x_1\rangle\, d\boldsymbol{\gamma}(x_1, y_1)\, d\mu(x_2) \\ &= \int_{X^2} \langle \boldsymbol{w}(x_1), y_1 - x_1\rangle\, d\boldsymbol{\gamma}(x_1, y_1), \end{aligned} \tag{10.4.46}$$

for every plan $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$ with $\nu \in D(\mathcal{W})$. In particular, choosing $\boldsymbol{\gamma} := (\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu$, we have

$$\int_{X \times X} \langle \nabla W(x-y), \boldsymbol{r}(x)\rangle\, d\mu \times \mu(x,y) = \int_X \langle \boldsymbol{w}(x), \boldsymbol{r}(x)\rangle\, d\mu(x) \tag{10.4.47}$$

for every vector field $\boldsymbol{r} \in L^\infty(\mu; X)$ and for $\boldsymbol{r} := \lambda \boldsymbol{i}$, $\lambda \in \mathbb{R}$.

Proof. By Jensen inequality we have

$$W(x - \bar\mu) \le \int_X W(x-y)\, d\mu(y) \quad \forall\, x \in X, \tag{10.4.48}$$

so that a further integration yields

$$\int_X W(x - \bar\mu)\, d\mu(x) \le \mathcal{W}(\mu); \tag{10.4.49}$$

(10.4.44) follows directly from (10.4.49) and the doubling condition (10.4.42), since $W(x) \le C_W\big(1 + W(x - \bar\mu) + W(\bar\mu)\big)$.

Combining the doubling condition and the convexity of W we also get

$$\begin{aligned}|\nabla W(x)| = \sup_{|y|\le 1} \langle \nabla W(x), y\rangle &\le \sup_{|y|\le 1} W(x+y) - W(x)\\ &\le C_W\big(1 + W(x) + \sup_{|y|\le 1} W(y)\big),\end{aligned} \tag{10.4.50}$$

which yields (10.4.45).

If now $\nu \in D(\mathcal{W})$ and $\boldsymbol{\gamma} \in \Gamma(\mu,\nu)$, then the positive part of the map $(x_1, y_1, x_2) \mapsto \langle \nabla W(x_1 - x_2), y_1 - x_1\rangle$ belongs to $L^1(\boldsymbol{\gamma} \times \mu)$ since convexity yields

$$\langle \nabla W(x_1 - x_2), y_1 - x_1\rangle \le W(y_1 - x_2) - W(x_1 - x_2),$$

and the right hand side of this inequality is integrable:

$$\int_{X^3} W(y_1 - x_2)\, d\boldsymbol{\gamma}\times\mu = \int_{X^2} W(y_1 - x_2)\, d\nu\times\mu \le C\big(1 + \mathcal{W}(\nu) + \mathcal{W}(\mu) + W(\bar\nu - \bar\mu)\big),$$

$$\int_{X^3} W(x_1 - x_2)\, d\boldsymbol{\gamma}\times\mu = \int_{X^2} W(x_1 - x_2)\, d\mu\times\mu = \mathcal{W}(\mu).$$

Therefore we can apply Fubini-Tonelli theorem to obtain

$$\begin{aligned}&\int_{X^3} \langle \nabla W(x_1 - x_2), y_1 - x_1\rangle\, d\boldsymbol{\gamma}\times\mu(x_1, y_1, x_2)\\ &= \int_{X^2} \Big(\int_X \langle \nabla W(x_1 - x_2), y_1 - x_1\rangle\, d\mu(x_2)\Big)\, d\boldsymbol{\gamma}(x_1, y_1)\\ &= \int_{X^2} \langle\Big(\int_X \nabla W(x_1 - x_2)\, d\mu(x_2)\Big), y_1 - x_1\rangle\, d\boldsymbol{\gamma}(x_1, y_1)\\ &= \int_{X^2} \langle \boldsymbol{w}(x_1), y_1 - x_1\rangle\, d\boldsymbol{\gamma}(x_1, y_1),\end{aligned}$$

which yields (10.4.46). □

Theorem 10.4.11 (Minimal subdifferential of $\mathcal{W}$). *Assume that $W : X \to [0, +\infty)$ is convex, Gateaux differentiable, even, and satisfies the doubling condition (10.4.42). Then $\mu \in \mathscr{P}_p(X)$ belongs to $D(|\partial\mathcal{W}|)$ if and only if $\boldsymbol{w} = (\nabla W) * \rho \in L^q(\mu; X)$. In this case $\boldsymbol{w} = \partial^\circ \mathcal{W}(\mu)$.*

Proof. As we did for the internal energy functional, we start by computing the directional derivative of $\mathcal{W}$ along a direction induced by a transport map $\boldsymbol{r} = \boldsymbol{i} + \boldsymbol{t}$, with $\boldsymbol{t}$ bounded and with a compact support (by the growth condition on W, this ensures that $\mathcal{W}(\boldsymbol{r}_{\#}\mu) < +\infty$). Since the map

$$t \mapsto \frac{W\left((x-y) + t(\boldsymbol{t}(x) - \boldsymbol{t}(y))\right) - W(x-y)}{t}$$

is nondecreasing w.r.t. t, the monotone convergence theorem and (10.4.47) give (taking into account that ∇W is an odd function)

$$\begin{aligned}+\infty \quad &> \quad \lim_{t\downarrow 0}\frac{\mathcal{W}((\boldsymbol{i}+t\boldsymbol{t})_{\#}\mu-\mathcal{W}(\mu)}{t}\\ &= \quad \frac{1}{2}\int_{X\times X}\langle\nabla W(x-y),(\boldsymbol{t}(x)-\boldsymbol{t}(y))\rangle\,d\mu\times\mu=\int_X\langle\boldsymbol{w},\boldsymbol{t}\rangle\,d\mu.\end{aligned}$$

On the other hand, since $|\partial\mathcal{W}|(\mu)<+\infty$, using the inequality $W_p((\boldsymbol{i}+t\boldsymbol{t})_{\#}\mu,\mu)\leq\|\boldsymbol{t}\|_{L^p(\mu;X)}$ we get

$$\int_X\langle\boldsymbol{w},\boldsymbol{t}\rangle\,d\mu\geq-|\partial\mathcal{W}|(\mu)\|\boldsymbol{t}\|_{L^p(\mu)};$$

changing the sign of $\boldsymbol{t}$ we obtain

$$\left|\int_X\langle\boldsymbol{w},\boldsymbol{t}\rangle\,d\mu\right|\leq|\partial\mathcal{W}|(\mu)\|\boldsymbol{t}\|_{L^p(\mu)},$$

and this proves that $\boldsymbol{w}\in L^q(\mu;X)$ and that $\|\boldsymbol{w}\|_{L^q}\leq|\partial\mathcal{W}|(\mu)$.

Now we prove that if $\boldsymbol{w}=(\nabla W)*\mu\in L^q(\mu;X)$, then it belongs to $\partial\mathcal{W}(\mu)$. Let us consider a test measure $\nu\in D(\mathcal{W})$, a plan $\boldsymbol{\gamma}\in\Gamma(\mu,\nu)$, and the directional derivative of $\mathcal{W}$ along the direction induced by $\boldsymbol{\gamma}$. Since the map

$$t\mapsto\frac{W\left((1-t)(x_1-x_2)+t(y_1-y_2)\right)-W(x_1-x_2)}{t}$$

is nondecreasing w.r.t. t, the monotone convergence theorem, the fact that ∇W is an odd function, and (10.4.47) give

$$\begin{aligned}\mathcal{W}(\nu)-\mathcal{W}(\mu) \quad &\geq \quad \lim_{t\downarrow 0}\frac{\mathcal{W}(((1-t)\pi^1+t\pi^2)_{\#}\boldsymbol{\gamma}-\mathcal{W}(\mu)}{t}\\ &= \quad \frac{1}{2}\int_{X^2\times X^2}\langle\nabla W(x_1-x_2),(y_1-x_1)-(y_2-x_2)\rangle\,d\boldsymbol{\gamma}\times\boldsymbol{\gamma}\\ &= \quad \int_{X^2}\langle\boldsymbol{w}(x_1),y_1-x_1\rangle\,d\boldsymbol{\gamma}(x_1,y_1),\end{aligned}$$

and this proves that $(\boldsymbol{i}\times\boldsymbol{w})_{\#}\mu\in\boldsymbol{\partial}\mathcal{W}(\mu)$. □

10.4.6 The opposite Wasserstein distance

In this section we compute the (metric) slope of the function $\psi(\cdot):=-\frac{1}{2}W_2^2(\cdot,\mu^2)$, i.e. the limit

$$\frac{1}{2}\limsup_{\nu\to\mu}\frac{W_2^2(\nu,\mu^2)-W_2^2(\mu,\mu^2)}{W_2(\nu,\mu)}=|\partial\psi|(\mu); \tag{10.4.51}$$

observe that the triangle inequality shows that the "lim sup" above is always less than $W_2(\mu,\mu^2)$; however this inequality is always strict when optimal plans are not induced by transports, as the following theorem shows; the right formula for the slope involves the minimal L^2 norm of the barycentric projection of the optimal plans and gives that the minimal selection is always induced by a map.

Theorem 10.4.12 (Minimal subdifferential of the opposite Wasserstein distance). *Let $\psi(\mu) = -\frac{1}{2}W_2^2(\mu, \mu^2)$. Then*

$$|\partial\psi|^2(\mu) = \min\left\{\int_X |\bar{\boldsymbol{\gamma}} - \boldsymbol{i}|^2\, d\mu : \ \boldsymbol{\gamma} \in \Gamma_o(\mu, \mu^2)\right\} \qquad \forall \mu \in \mathscr{P}_2(X), \tag{10.4.52}$$

and $\partial^\circ\psi(\mu) = \bar{\boldsymbol{\gamma}} - \boldsymbol{i}$ is a strong *subdifferential, where $\boldsymbol{\gamma}$ is the unique minimizing plan above.*
Moreover $\mu \mapsto |\partial\psi|(\mu)$ is lower semicontinuous with respect to narrow convergence in $\mathscr{P}(X)$, along sequences bounded in $\mathscr{P}_2(X)$.

Proof. Notice first that the minimum is uniquely attained because of the convexity of $\Gamma_o(\mu, \mu^2)$, the linearity of the barycentric projection, and the strict convexity of the L^2 norm.

We show first that for any $\boldsymbol{\gamma} \in \Gamma_o(\mu, \mu^2)$ the plan $\boldsymbol{\eta} = (\boldsymbol{i} \times (\bar{\boldsymbol{\gamma}} - \boldsymbol{i}))_\# \mu$ belongs to $\boldsymbol{\partial}\psi(\mu)$, i.e. $\bar{\boldsymbol{\gamma}} - \boldsymbol{i} \in \partial\psi(\mu)$, proving the inequality $\le$ in (10.4.52). For every $\nu \in \mathscr{P}_2(X)$ and $\boldsymbol{\nu} \in \Gamma(\mu, \nu)$ it would be sufficient to show that (recall that ψ is -1 convex)

$$\psi(\nu) \ge \psi(\mu) + \int_{X^2} \langle \bar{\boldsymbol{\gamma}}(x_1) - x_1, x_2 - x_1\rangle\, d\boldsymbol{\nu}(x_1, x_2) - \frac{1}{2}W_{\boldsymbol{\nu}}^2(\mu, \nu). \tag{10.4.53}$$

Let $\boldsymbol{\beta} \in \Gamma(\mu, \mu^2, \nu)$ be the 3-plan determined by the condition

$$\boldsymbol{\beta}_{x_1} = \gamma_{x_1} \times \nu_{x_1} \quad \text{for } \mu\text{-a.e. } x_1.$$

Since $\pi^{1,2}_\# \boldsymbol{\beta} = \boldsymbol{\gamma} \in \Gamma_o(\mu, \mu^2)$ and $\pi^{1,3}_\# \boldsymbol{\beta} = \boldsymbol{\nu} \in \Gamma(\mu, \nu)$, we have the inequalities

$$\begin{aligned}
\psi(\nu) - \psi(\mu) + \frac{1}{2}W_{\boldsymbol{\nu}}^2(\mu, \nu) &= -\frac{1}{2}W_2^2(\nu, \mu^2) + \frac{1}{2}W_2^2(\mu, \mu^2) + \frac{1}{2}W_{\boldsymbol{\nu}}^2(\mu, \nu) \\
&\ge -\frac{1}{2}\|x_2 - x_3\|^2_{L^2(\boldsymbol{\beta}; X)} + \frac{1}{2}\|x_2 - x_1\|^2_{L^2(\boldsymbol{\beta}; X)} + \frac{1}{2}\|x_3 - x_1\|^2_{L^2(\boldsymbol{\beta}; X)} \\
&= \int_{X^3} \langle x_2 - x_1, x_3 - x_1\rangle\, d\boldsymbol{\beta}(x_1, x_2, x_3) \\
&= \int_X \Big(\int_{X^2} \langle x_2 - x_1, x_3 - x_1\rangle\, d\gamma_{x_1}(x_2) \times d\nu_{x_1}(x_3)\Big)\, d\mu(x_1) \\
&= \int_X \Big(\int_X \langle \bar{\boldsymbol{\gamma}}(x_1) - x_1, x_3 - x_1\rangle\ d\nu_{x_1}(x_3)\Big)\, d\mu(x_1) \\
&= \int_{X^2} \langle \bar{\boldsymbol{\gamma}}(x_1) - x_1, x_2 - x_1\rangle\, d\boldsymbol{\nu}(x_1, x_2),
\end{aligned}$$

proving (10.4.53).

In order to show that equality holds in (10.4.52) we notice that

$$\left(\int_{X^2} |x_2|^2 \, d\boldsymbol{\gamma}\right)^{1/2} \geq \|\bar{\boldsymbol{\gamma}}\|_{L^2(\mu;X)} \geq \|P_\mu(\bar{\boldsymbol{\gamma}})\|_{L^2(\mu;X)} \qquad \forall \boldsymbol{\gamma} \in \Gamma(\mu, \mu^2),$$

where $P_\mu : L^2(\mu; X) \to \mathrm{Tan}_\mu \mathscr{P}_2(X)$ is the orthogonal projection. Therefore, taking into account that the slope is equal to the minimal norm in the subdifferential, it would be sufficient to show that any $\boldsymbol{\gamma} \in \boldsymbol{\partial}\psi(\mu)$ satisfies $P_\mu(\bar{\boldsymbol{\gamma}}) = \bar{\boldsymbol{\eta}} - \boldsymbol{i}$ for some $\boldsymbol{\eta} \in \Gamma_o(\mu, \mu^2)$. Denoting by $\langle \cdot, \cdot \rangle$ the scalar product in $L^2(\mu; X)$, by applying Hahn-Banach theorem in $\mathrm{Tan}_\mu \mathscr{P}_2(X)$ it suffices to show the inequality

$$\langle P_\mu(\bar{\boldsymbol{\gamma}}), \boldsymbol{v} \rangle \leq \max_{\boldsymbol{w} \in K} \langle \boldsymbol{w}, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{v} \in \mathrm{Tan}_\mu \mathscr{P}_2(X), \tag{10.4.54}$$

where $K \subset L^2(\mu; X)$ is the bounded, closed, and convex set defined by

$$K := \left\{ \bar{\boldsymbol{\eta}} - \boldsymbol{i} : \ \boldsymbol{\eta} \in \Gamma_o(\mu, \mu^2) \right\}.$$

Notice indeed that $K \subset \mathrm{Tan}_\mu \mathscr{P}_2(X)$ by Theorem 12.4.4 and Theorem 8.5.5.

By a density argument it suffices to check (10.4.54) when v has the property that $\boldsymbol{i} + \varepsilon \boldsymbol{v}$ is an optimal map, and the unique one, for some $\varepsilon > 0$: indeed, it suffices to recall that for any optimal transport map $\boldsymbol{r}$ the interpolated maps $\boldsymbol{r}_t = (1-t)\boldsymbol{i} + t\boldsymbol{r}$ are the unique optimal transport maps (see Lemma 7.2.1) for any $t \in [0, 1)$, so that the property above is fulfilled with $\varepsilon = t$ and $\boldsymbol{v} = \boldsymbol{r} - \boldsymbol{i}$. Then we use the fact that the positive cone induced by these vectors is dense in $\mathrm{Tan}_\mu \mathscr{P}_2(X)$, by Theorem 8.5.5. By homogeneity, we assume that $\varepsilon = 1$. Under these assumptions on $\boldsymbol{v}$, for $t \in [0, 1]$ we set

$$\mu_t := (\boldsymbol{i} + t\boldsymbol{v})_\# \mu \quad \text{and} \quad \boldsymbol{\alpha}_t := (\pi^1, \pi^2, (\boldsymbol{i} + tv) \circ \pi^1)_\# \boldsymbol{\gamma},$$

noticing that Lemma 5.3.2 gives that $\boldsymbol{\alpha}_t$ is the unique 3-plan such that $\pi^{1,2}_\# \boldsymbol{\alpha}_t = \boldsymbol{\gamma}$ and $\pi^{1,3}_\# \boldsymbol{\alpha}_t = \mu_t$. As a consequence, since $\boldsymbol{\gamma} \in \boldsymbol{\partial}\psi(\mu)$, the inequality

$$\psi(\mu_t) \geq \psi(\mu) + \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\alpha}_t - \frac{1}{2} W_2^2(\mu, \mu_t) = t \int_X \langle \bar{\boldsymbol{\gamma}}, \boldsymbol{v} \rangle \, d\mu - \frac{1}{2} W_2^2(\mu, \mu_t)$$

must hold. Since $W_2^2(\mu, \mu_t) = t^2 W_2^2(\mu, \mu_1)$, dividing both sides by t and passing to the limit as $t \downarrow 0$ we obtain

$$\liminf_{t \downarrow 0} \frac{W_2^2(\mu, \mu^2) - W_2^2(\mu_t, \mu^2)}{t} \geq 2 \int_X \langle \bar{\boldsymbol{\gamma}}, \boldsymbol{v} \rangle \, d\mu = 2 \int_X \langle P_\mu(\bar{\boldsymbol{\gamma}}), \boldsymbol{v} \rangle \, d\mu.$$

On the other hand, Proposition 7.3.6 and (7.3.17) give that the derivative on the left is equal to

$$\max_{\boldsymbol{\eta} \in \Gamma_o(\mu, \mu^2)} 2 \int_{X^2} \langle \boldsymbol{v}(x_1), x_2 - x_1 \rangle \, d\boldsymbol{\eta} = \max_{\boldsymbol{\eta} \in \Gamma_o(\mu, \mu^2)} 2 \int_{X^2} \langle \boldsymbol{v}(x_1), \bar{\boldsymbol{\eta}}(x_1) - x_1 \rangle \, d\mu,$$

so that, recalling the definition of K, (10.4.53) holds.

Now we show the stated lower semicontinuity of the slope. Let $(\mu_n) \subset \mathscr{P}_2(X)$ be a bounded sequence narrowly converging in $\mathscr{P}_2(X)$ to μ and let $\boldsymbol{\gamma}_n \in \Gamma_o(\mu_n, \mu^2)$. We assume that $|\partial\psi|(\mu_n)$ converges to L and we have to show that $L \geq |\partial\psi|(\mu)$. Since the marginals of $\boldsymbol{\gamma}_n$ are converging in $\mathscr{P}_2(X)$ (and therefore tight), we can assume, possibly extracting a subsequence, that $\boldsymbol{\gamma}_n$ narrowly converge in $\mathscr{P}(X \times X)$ to some plan γ. By Proposition 7.1.3 we obtain that $\gamma \in \Gamma_o(\mu, \mu^2)$. Using test functions of the form $\langle x_2 - x_1, \varphi(x_1)\rangle$, with $\varphi \in C^0_b(X;X)$, we immediately obtain that $(\bar{\boldsymbol{\gamma}}_n - \boldsymbol{i})\mu_n$ narrowly converge in the duality with $C^0_b(X;X)$ to $(\bar{\gamma} - \boldsymbol{i})\mu$.

Now we claim that $\|\bar{\boldsymbol{\gamma}} - \boldsymbol{i}\|_{L^2(\mu)} \leq L$. Indeed, for any $\varphi \in C^0_b(X;X)$ we can pass to the limit as $n \to \infty$ in the inequality

$$\left|\int_X \langle \varphi, \bar{\boldsymbol{\gamma}}_n - \boldsymbol{i}\rangle \, d\mu_n\right| \leq \|\bar{\gamma}_n - \boldsymbol{i}\|_{L^2(\mu_n;X)} \|\varphi\|_{L^2(\mu_n;X)}$$

to obtain

$$\left|\int_X \langle \varphi, \bar{\boldsymbol{\gamma}} - \boldsymbol{i}\rangle \, d\mu\right| \leq L \|\varphi\|_{L^2(\mu;X)},$$

whence the stated inequality follows. Using the inequality $|\partial\psi|(\mu) \leq \|\bar{\gamma} - \boldsymbol{i}\|_{L^2(\mu)}$ we obtain that $|\partial\psi|(\mu) \leq L$. □

We conclude this section by proving Remark 10.3.5: setting now $\psi(\cdot) := -\frac{1}{2}W_2^2(\cdot, \mu)$, we simply observe that if μ_τ is a minimizer of (10.3.1b), then

$$\boldsymbol{\gamma}_\tau \in \boldsymbol{\partial}\psi(\mu_\tau) \quad \Longrightarrow \quad \boldsymbol{\gamma}_\tau \in \boldsymbol{\partial}\phi(\mu_\tau), \tag{10.4.55}$$

since

$$\phi(\nu) - \phi(\mu_\tau) \geq \psi(\nu) - \psi(\mu_\tau) \quad \forall\, \nu \in \mathscr{P}_2(X).$$

10.4.7 The sum of internal, potential and interaction energy

In this section we consider, as in [38], the functional $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ given by the sum of internal, potential and interaction energy:

$$\phi(\mu) := \int_{\mathbb{R}^d} F(\rho)\, dx + \int_{\mathbb{R}^d} V \, d\mu + \frac{1}{2}\int_{\mathbb{R}^d \times R^d} W \, d\mu \times \mu \quad \text{if } \mu = \rho \mathscr{L}^d, \tag{10.4.56}$$

setting $\phi(\mu) = +\infty$ if $\mu \in \mathscr{P}_p(\mathbb{R}^d) \setminus \mathscr{P}_p^r(\mathbb{R}^d)$. Recalling the "doubling condition" stated in (10.4.23), we make the following assumptions on F, V and W:

- **(F)** $F : [0, +\infty) \to \mathbb{R}$ is a doubling, convex differentiable function with superlinear growth satisfying (10.4.15) (i.e. the bounds on F^-) and (10.4.21) (yielding the geodesic convexity of the internal energy).
- **(V)** $V : \mathbb{R}^d \to (-\infty, +\infty]$ is a l.s.c. λ-convex function with proper domain $D(V)$ with nonempty interior $\Omega \subset \mathbb{R}^d$;

(W) $W : \mathbb{R}^d \to [0,+\infty)$ is a convex, differentiable, even function satisfying the doubling condition (10.4.42).

Notice that we have assumed that F has a superlinear growth only for simplicity: also the case of (sub-)linear growth could be considered, proving along the lines of Theorem 10.4.8 that finiteness of the slope implies regularity of the measure. Also the doubling assumptions could be relaxed, see Remark 10.4.7. Finally, the finiteness of ϕ yields

$$\operatorname{supp}\mu \subset \overline{\Omega} = \overline{D(V)}, \quad \mu(\partial\Omega) = 0, \tag{10.4.57}$$

so that its density ρ w.r.t. $\mathscr{L}^d$ can be considered as a function of $L^1(\Omega)$.

The same monotonicity argument used in the proof of Lemma 10.4.4 gives

$$+\infty > \lim_{t\downarrow 0} \frac{\int_{\mathbb{R}^d} V\, d((1-t)\boldsymbol{i} + t\boldsymbol{r}))_{\#}\mu - \int_{\mathbb{R}^d} V\, d\mu}{t} = \int_{\mathbb{R}^d} \langle \nabla V, \boldsymbol{r} - \boldsymbol{i}\rangle\, d\mu, \tag{10.4.58}$$

whenever both $\int_{\mathbb{R}^d} V\, d\mu < +\infty$ and $\int_{\mathbb{R}^d} V\, d\boldsymbol{r}_{\#}\mu < +\infty$.

Analogously, denoting by $\mathcal{W}$ the interaction energy functional induced by $W/2$, arguing as in the first part of Theorem 10.4.11 we have

$$+\infty > \lim_{t\downarrow 0} \frac{\mathcal{W}(((1-t)\boldsymbol{i} + t\boldsymbol{r})_{\#}\mu) - \mathcal{W}(\mu)}{t} = \int_{\mathbb{R}^d} \langle (\nabla W) * \mu), \boldsymbol{r} - \boldsymbol{i}\rangle\, d\mu, \tag{10.4.59}$$

whenever $\mathcal{W}(\mu) + \mathcal{W}(\boldsymbol{r}_{\#}\mu) < +\infty$. The growth condition on W ensures that $\mu \in D(\mathcal{W})$ implies $\boldsymbol{r}_{\#}\mu \in D(\mathcal{W})$ if either $\boldsymbol{r} - \boldsymbol{i}$ is bounded or $\boldsymbol{r} = 2\boldsymbol{i}$ (here we use the doubling condition).

We have the following characterization of the minimal selection in the subdifferential $\partial^\circ\phi(\mu)$:

Theorem 10.4.13 (Minimal subdifferential of ϕ). *A measure $\mu = \rho\mathscr{L}^d \in D(\phi) \subset \mathscr{P}_p(\mathbb{R}^d)$ belongs to $D(|\partial\phi|)$ if and only if $L_F(\rho) \in W^{1,1}_{\rm loc}(\Omega)$ and*

$$\rho\boldsymbol{w} = \nabla L_F(\rho) + \rho\nabla V + \rho(\nabla W) * \rho \quad \textit{for some} \quad \boldsymbol{w} \in L^q(\mu; \mathbb{R}^d). \tag{10.4.60}$$

In this case the vector $\boldsymbol{w}$ defined μ-a.e. by (10.4.60) *is the minimal selection in $\partial\phi(\mu)$, i.e. $\boldsymbol{w} = \partial^\circ\phi(\mu)$.*

Proof. We argue exactly as in the proof of Theorem 10.4.6, computing the Gateaux derivative of ϕ in several directions $\boldsymbol{r}$, using Lemma 10.4.4 for the internal energy and (10.4.58), (10.4.59) respectively for the potential and interaction energy.

Choosing $\boldsymbol{r} = \boldsymbol{i} + \boldsymbol{t}$, with $\boldsymbol{t} \in C^\infty_c(\Omega; \mathbb{R}^d)$, we obtain

$$-\int_{\mathbb{R}^d} L_F(\rho)\nabla\cdot\boldsymbol{t}\, dx + \int_{\mathbb{R}^d} \langle \nabla V, \boldsymbol{t}\rangle\, d\mu + \int_{\mathbb{R}^d} \langle (\nabla W) * \rho, \boldsymbol{t}\rangle\, d\mu \geq -|\partial\phi|(\mu)\|\boldsymbol{t}\|_{L^p(\mu)}. \tag{10.4.61}$$

Since V is locally Lipschitz in Ω and $\nabla W * \rho$ is locally bounded, following the same argument of Theorem 10.4.6, we obtain from (10.4.61) first that $L_F(\rho) \in BV_{\mathrm{loc}}(\mathbb{R}^d)$ and then that $L_F(\rho) \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$, with

$$\nabla L_F(\rho) + \rho\nabla V + \rho(\nabla W) * \rho = \boldsymbol{w}\rho \quad \text{for some } \boldsymbol{w} \in L^q(\mu;\mathbb{R}^d) \tag{10.4.62}$$

with $\|\boldsymbol{w}\|_{L^q} \le |\partial\phi|(\mu)$.

In order to show that the vector $\boldsymbol{w}$ is in the subdifferential (and then, by the previous estimate, it is the minimal selection) we choose eventually a test measure $\nu \in D(\phi)$ with compact support contained in Ω and the associated optimal transport map $\boldsymbol{r} = \boldsymbol{t}_\mu^\nu$; Lemma 10.4.4, (10.4.58), (10.4.59), and Lemma 10.4.5 yield

$$\begin{aligned}
\phi(\nu) - \phi(\mu) &\ge \frac{d}{dt}\phi\left(((1-t)\boldsymbol{i} + t\boldsymbol{r})_{\#}\mu\right)\Big|_{t=0^+} \\
&= -\int_\Omega L_F(\rho)\tilde{\nabla}\cdot(\boldsymbol{r}-\boldsymbol{i})\,dx + \int_\Omega \langle \nabla V, \boldsymbol{r}-\boldsymbol{i}\rangle\,d\mu + \int_\Omega \langle(\nabla W)*\rho, \boldsymbol{r}-\boldsymbol{i}\rangle\,d\mu \\
&\ge \limsup_{h\to\infty}\int_\Omega \langle \nabla L_F(\rho), \boldsymbol{r}-\boldsymbol{i}\rangle\chi_h\,dx + \int_\Omega \langle \nabla V + (\nabla W)*\rho, \boldsymbol{r}-\boldsymbol{i}\rangle\,d\mu \\
&= \limsup_{h\to\infty}\int_\Omega \langle \nabla L_F(\rho) + \rho\nabla V + \rho(\nabla W)*\rho, \boldsymbol{r}-\boldsymbol{i}\rangle\chi_h\,dx \\
&= \int_\Omega \langle \rho\boldsymbol{w}, \boldsymbol{r}-\boldsymbol{i}\rangle\,dx = \int_\Omega \langle \boldsymbol{w}, \boldsymbol{r}-\boldsymbol{i}\rangle\,d\mu.
\end{aligned}$$

Finally, we notice that the proof that $\boldsymbol{w}$ belongs to the subdifferential did not use the finiteness of slope, but only the assumption (previously derived by the finiteness of slope) that $L_F(\rho) \in W^{1,1}_{\mathrm{loc}}(\Omega)$, (10.4.60), and $\phi(\mu) < +\infty$; therefore these conditions imply that the subdifferential is not empty, hence the slope is finite and the vector $\boldsymbol{w}$ is the minimal selection in $\partial\phi(\mu)$. □

We know that for general λ-convex functionals the metric slope is l.s.c. with respect to convergence in $\mathscr{P}_p(\mathbb{R}^d)$. In the case of the functional ϕ of (10.4.56) the slope is also lower semicontinuous w.r.t. the narrow convergence.

Proposition 10.4.14 (Narrow lower semicontinuity of $|\partial\phi|$). *Let us suppose that assumptions* **(F,V,W)** *are satisfied; if $(\mu_n) \subset \mathscr{P}_p(\mathbb{R}^d)$ is a bounded sequence narrowly converging to μ in $\mathscr{P}(\mathbb{R}^d)$ with $\sup_n \phi(\mu_n) < +\infty$, $\boldsymbol{w}_n \in \partial^\circ\phi(\mu_n)$ have bounded $L^q(\mu_n;\mathbb{R}^d)$ norms and are weakly converging to $\boldsymbol{w} \in L^q(\mu;\mathbb{R}^d)$ in the sense of Definition* 5.4.3*, then $\boldsymbol{w} \in \partial^\circ\phi(\mu)$. We have also*

$$\liminf_{n\to\infty} |\partial\phi|(\mu_n) \ge |\partial\phi|(\mu). \tag{10.4.63}$$

Proof. Observe that thanks to Theorem 5.4.4

$$+\infty > \liminf_{n\to\infty} |\partial\phi|^q(\mu_n) = \liminf_{n\to\infty}\int_X |\boldsymbol{w}_n(x)|^q\,d\mu_n(x) \ge \int_X |\boldsymbol{w}(x)|^q\,d\mu(x).$$

Let now ρ_n be such that $\mu_n = \rho_n \mathscr{L}^d$; since the p-moment of μ_n is bounded and the negative part of V has a linear growth, we know that

$$\sup_n \int_{\mathbb{R}^d} F(\rho_n(x))\,dx < +\infty, \quad \sup_n \int_{\mathbb{R}^d} \Big(V^+(x) + |x|^p\Big)\,\rho_n(x)\,dx < +\infty.$$

Thanks to the superlinear growth of F, we deduce that ρ_n weakly converge to ρ in $L^1(\mathbb{R}^d)$, ρ being the Lebesgue density of μ.

Since $L_F(\rho) \le F(2\rho) - 2F(\rho)$, the doubling condition shows that $L_F(\rho_n)$ is bounded in $L^1(\Omega)$; since ∇V is locally bounded, we know that

$$\rho_n \nabla V \to \rho \nabla V, \quad \text{weakly in } L^1_{\rm loc}(\Omega); \tag{10.4.64}$$

(10.4.50) and Lemma 5.1.7 show that

$$\int_{\Omega^2} \langle \nabla W(x-y), \boldsymbol{t}(y)\rangle \rho_n(x)\rho_n(y)\,dx\,dy \to \int_{\Omega^2} \langle \nabla W(x-y), \boldsymbol{t}(y)\rangle \rho(x)\rho(y)\,dx\,dy$$

for every vector field $\boldsymbol{t} \in L^\infty(\Omega)$, so that $\rho_n(\nabla W)*\rho_n$ weakly converge to $\rho(\nabla W)*\rho$ in $L^1(\Omega)$.

We thus deduce that $L_F(\rho_n)$ is bounded in $BV_{\rm loc}(\Omega)$; we can extract a further subsequence such that

$$L_F(\rho_n) \to L \quad \text{in } L^1_{\rm loc}(\mathbb{R}^d) \quad \text{and pointwise } \mathscr{L}^d\text{-a.e.} \tag{10.4.65}$$

A standard truncation argument and the fact that L_F is a monotone function yield $L(x) = L_F(\rho(x))$ for $\mathscr{L}^d$-a.e. $x \in \Omega$, and therefore $\nabla L_F(\rho_n) \to \nabla L_F(\rho)$ in the sense of distributions.

Combining all the above results, we get

$$\rho \boldsymbol{w} = \nabla L_F(\rho) + \rho \nabla V + \rho(\nabla W) * \rho \quad \text{for } \boldsymbol{w} \in L^q(\mu; \mathbb{R}^d), \tag{10.4.66}$$

so that $\boldsymbol{w} = \partial^\circ \phi(\mu)$. □

An interesting particular case of the above result is provided by the relative entropy functional: let us choose $p = 2$, $W \equiv 0$ and

$$F(s) := s \log s, \quad \gamma := \frac{1}{Z} e^{-V} \cdot \mathscr{L}^d = e^{-(V(x) + \log Z)} \cdot \mathscr{L}^d,$$

with $Z > 0$ chosen so that $\gamma(\mathbb{R}^d) = 1$. Recalling Remark 9.4.2, the functional ϕ can also be written as

$$\phi(\mu) = \mathcal{H}(\mu|\gamma) - \log Z. \tag{10.4.67}$$

Since in this case $L_F(\rho) = \rho$, a vector $\boldsymbol{w} \in L^2(\mu; \mathbb{R}^d)$ is the minimal selection $\partial^\circ \phi(\mu)$ if and only if

$$-\int_{\mathbb{R}^d} \nabla \cdot \boldsymbol{\zeta}(x)\,d\mu(x) = \int_{\mathbb{R}^d} \langle \boldsymbol{w}(x), \boldsymbol{\zeta}(x)\rangle\,d\mu(x) - \int_{\mathbb{R}^d} \langle \nabla V(x), \boldsymbol{\zeta}(x)\rangle\,d\mu(x), \tag{10.4.68}$$

for every test function $\boldsymbol{\zeta} \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^d)$; (10.4.68) can also be written in terms of $\sigma = \frac{d\mu}{d\gamma}$ as

$$-\int_{\mathbb{R}^d} \sigma \nabla \cdot (e^{-V(x)}\boldsymbol{\zeta}(x))\,dx = \int_{\mathbb{R}^d} \langle \sigma \boldsymbol{w}(x), e^{-V(x)}\boldsymbol{\zeta}(x)\rangle\,dx, \tag{10.4.69}$$

which shows that $\sigma\boldsymbol{w} = \nabla\sigma$.

10.4.8 Relative entropy and Fisher information in infinite dimensions

Let X be an infinite dimensional (separable) Hilbert space, and let $Q : X \to X$ be a bounded, positive definite, symmetric linear operator of trace class.

We introduce the orthonormal system $(\boldsymbol{e}_n)$ of eigenvectors of Q, i.e. satisfying

$$Q\boldsymbol{e}_n = \lambda_n \boldsymbol{e}_n, \quad \lambda_n > 0, \quad \sum_{n=1}^{\infty} \lambda_n = \operatorname{tr} Q < +\infty. \tag{10.4.70}$$

We denote by X_d the finite dimensional subspace generated by the first d eigenvectors, by $\hat\pi_d$ the orthogonal projection of X onto X_d, and by $Q_d : X \to X_d$ the linear operator defined by $Q_d := \hat\pi_d \circ Q$, $Q_d x = \sum_{j=1}^d \lambda_j \langle x, \boldsymbol{e}_j\rangle \boldsymbol{e}_j$.

Let γ be the centered Gaussian measure with covariance operator Q^{-1}: γ is determined by its finite dimensional projections $\gamma_d := (\hat\pi_d)_\#\gamma$, which are given by

$$\gamma_d := \frac{1}{\sqrt{(2\pi)^d \det Q_d}} e^{-\frac{1}{2}\langle Q_d^{-1}x,x\rangle} \cdot \mathscr{H}^d|_{X_d}. \tag{10.4.71}$$

Notice that

$$\det Q_d := \prod_{j=1}^d \lambda_j \quad \text{and} \quad \langle Q_d^{-1}x, x\rangle = \sum_{j=1}^d \lambda_j^{-1} \langle x, \boldsymbol{e}_j\rangle^2. \tag{10.4.72}$$

In Section 9.4 we studied the properties of the relative entropy functional

$$\phi(\mu) := \mathcal{H}(\mu|\gamma), \tag{10.4.73}$$

which is a geodesically convex functional in $\mathscr{P}_2(X)$.

Let us recall the standard definition of generalized partial derivatives for functions in $L^1(\gamma)$ [25, Def. 5.2.7]:

Definition 10.4.15 (Partial and logarithmic derivatives). *Let $\rho : X \to \mathbb{R}$ be a Borel function with $\int_X |x||\rho|\,d\gamma < +\infty$. The function ρ has generalized partial derivative $\sigma_j := \partial_{\boldsymbol{e}_j}\rho \in L^1(\gamma)$ along $\boldsymbol{e}_j$ if for any smooth cylindrical function $\zeta \in \mathrm{Cyl}(X)$ one has the "integration by parts" formula*

$$-\int_X \rho(x)\partial_{\boldsymbol{e}_j}\zeta(x)\,d\gamma(x) = \int_X \sigma_j(x)\zeta(x)\,d\gamma(x) - \lambda_j^{-1}\int_X \langle x, \boldsymbol{e}_j\rangle \zeta(x)\rho(x)\,d\gamma(x). \tag{10.4.74}$$

We say that $\mu = \rho \cdot \gamma \in \mathscr{P}_2^r(X)$ *has* w_j *as logarithmic partial derivative along* $\boldsymbol{e}_j$ *if* $\partial_{\boldsymbol{e}_j}\rho \in L^1(\gamma)$ *and* $w_j = \frac{\partial_{\boldsymbol{e}_j}\rho}{\rho} \in L^2(\mu)$, *i.e.*

$$-\int_X \partial_{\boldsymbol{e}_j}\zeta(x)\,d\mu(x) = \int_X w_j(x)\zeta(x)\,d\mu(x) - \lambda_j^{-1}\int_X \langle x, \boldsymbol{e}_j\rangle \zeta(x)\,d\mu(x). \quad (10.4.75)$$

By a standard smoothing argument, as in the finite dimensional case, one can check that the integral identities above holds for any cylindrical bounded function ζ of class C^1 with a bounded gradient.

Definition 10.4.16 (Logarithmic gradient and Fisher information functional). *If* μ *has logarithmic partial derivatives* $w_j = \frac{\partial_{\boldsymbol{e}_j}\rho}{\rho} \in L^2(\mu)$ *for every* $j \in \mathbb{N}$ *and*

$$\sum_{j=1}^{\infty}\int_X |w_j(x)|^2\,d\mu(x) < +\infty, \quad (10.4.76)$$

we define the "logarithmic gradient" of μ *as follows:*

$$\boldsymbol{w}(x) = \frac{\nabla\rho}{\rho} := \sum_{j=1}^{\infty} w_j(x)\boldsymbol{e}_j \in L^2(\mu; X). \quad (10.4.77)$$

The Fisher information functional $\mathscr{I}(\mu|\gamma)$ *is defined as* $\|\boldsymbol{w}\|^2_{L^2(\mu;X)}$.

The following theorem shows that the Fisher information functional is indeed the minimal slope of the entropy functional even in the infinite-dimensional case. The proof requires the validity of the statement in the finite dimensional case and an approximation based on the Γ-convergence of the finite-dimensional entropy functionals.

Theorem 10.4.17. *A measure* $\mu = \rho \cdot \gamma \in \mathscr{P}_2^r(X)$ *with finite relative entropy* $\phi(\mu) = \mathcal{H}(\mu|\gamma)$ *belongs to* $D(|\partial\phi|)$ *if and only if* μ *has a logarithmic gradient* $\boldsymbol{w} = \frac{\nabla\rho}{\rho} \in L^2(\mu; X)$ *according to Definition* 10.4.16. *In this case* $\boldsymbol{w} = \partial^\circ\phi(\mu)$ *and*

$$|\partial\phi|^2(\mu) = \int_X \frac{|\nabla\rho(x)|^2}{\rho(x)^2}\,d\mu(x) = \int_X \frac{|\nabla\rho(x)|^2}{\rho(x)}\,d\gamma(x) = \mathscr{I}(\mu|\gamma). \quad (10.4.78)$$

Proof. Let us suppose that all the components w_j of $\boldsymbol{w} \in L^2(\mu; X)$ satisfy (10.4.75). We fix an integer d and we consider the orthogonal projection $\hat{\pi}_d$ of X onto X_d. We consider cylindrical functions of the form $\boldsymbol{\zeta}(x) = \boldsymbol{\psi}(\hat{\pi}_d(x))$ for $\boldsymbol{\psi} : X_d \to X$ bounded, of class C^1 and with a bounded gradient. If we introduce the measure $\mu_d := (\hat{\pi}_d)_\#\mu$, we can disintegrate μ w.r.t. μ_d as $\mu = \int_{X_d}\mu_x\,d\mu_d(x)$, with $\mu_x \in \mathscr{P}(X)$ concentrated on $\hat{\pi}_d^{-1}(x)$ and we can define the vector field

$$\boldsymbol{w}_d(x) := \int_X \boldsymbol{w}(y)\,d\mu_x(y), \quad (10.4.79)$$

which satisfies

$$\int_X \langle \boldsymbol{w}(x), \boldsymbol{\psi}(\hat{\pi}_d(x))\rangle \, d\mu(x) = \int_{X_d} \langle \boldsymbol{w}_d(y), \boldsymbol{\psi}(y)\rangle \, d\mu_d(y). \tag{10.4.80}$$

Choosing $\boldsymbol{\psi}$ of the form $\sum_{j=1}^d \psi_j \boldsymbol{e}_j$, using (10.4.75) and the previous identity, we obtain

$$-\int_{X_d} \sum_{j=1}^d \frac{\partial \psi_j}{\partial \boldsymbol{e}_j}(x)\, d\mu_d(x) = \int_{X_d} \langle \boldsymbol{w}_d(x), \boldsymbol{\psi}(x)\rangle \, d\mu_d(x) - \int_{X_d} \langle Q^{-1}x, \boldsymbol{\psi}(x)\rangle \, d\mu_d(x). \tag{10.4.81}$$

Since (by Jensen inequality) $\boldsymbol{w}_d \in L^2(\mu_d; \mathbb{R}^d)$ and

$$\int_{X_d} |\boldsymbol{w}_d(x)|^2 \, d\mu_d(x) \le \int_X |\boldsymbol{w}(x)|^2 \, d\mu(x), \tag{10.4.82}$$

from (10.4.68) (stated for $\mathbb{R}^d$ but still true for X_d) we obtain that $\boldsymbol{w}_d \in \partial^\circ \phi_d(\mu_d)$ where $\phi_d(\mu) := \mathcal{H}(\mu|\gamma_d)$. Lemma 9.4.3 and Lemma 9.4.5 show that ϕ_d is $\Gamma(\mathscr{P}_2(X))$ converging to ϕ as $d \to +\infty$. Since $\mu_d \to \mu$ in $\mathscr{P}_p(X)$ and $\boldsymbol{w}_d$ are easily seen to be converging to $\boldsymbol{w}$ according to Definition 5.4.3 (by compactness, see Theorem 5.4.4(a), one needs only to check condition (10.4.47) on cylindrical test functions $\boldsymbol{\zeta}$), we can apply Lemma 10.3.8 which shows that $\boldsymbol{w} \in \partial\phi(\mu)$ and therefore

$$|\partial\phi|^2(\mu) \le \mathscr{I}(\mu|\gamma).$$

In order to prove the opposite implication, let us now suppose that $\boldsymbol{w} = \partial^\circ \phi(\mu)$: applying Lemma 10.3.16 to the sequence of functionals $\phi_d(\mu) = \mathcal{H}(\mu|\gamma_d)$, we find two sequences $\nu_d \to \mu$ in $\mathscr{P}_2(X)$ and $\boldsymbol{w}_d \in \partial^\circ \phi_d(\nu_d)$ converging to $\boldsymbol{w}$ according to Definition 5.4.3, i.e. the plans $(\boldsymbol{i} \times \boldsymbol{w}_d)_\# \nu_d \in \boldsymbol{\partial}^\circ \phi_d(\nu_d)$ narrowly converge in $\mathscr{P}(X \times X_\varpi)$ to $(\boldsymbol{i} \times \boldsymbol{w})_\# \mu \in \boldsymbol{\partial}^\circ \phi(\mu)$ (actually the lemma provides the stronger convergence in $\mathscr{P}_2(X \times X)$, not needed here). By the finite dimensional result, we know that

$$-\int_X \partial_{\boldsymbol{e}_j} \zeta(x)\, d\nu_d(x) = \int_X \langle \boldsymbol{w}_d(x), \boldsymbol{e}_j\rangle \zeta(x)\, d\nu_d(x) - \lambda_j^{-1} \int_X \langle x, \boldsymbol{e}_j\rangle \zeta(x)\, d\nu_d(x)$$

for every $j = 1, \ldots, d$ and $\zeta \in \mathrm{Cyl}(X)$. Keeping j and ζ fixed, we can pass to the limit as $d \to \infty$ to obtain (10.4.75) (the convergence of the rightmost integral follows by Lemma 5.1.7). □

Chapter 11

Gradient Flows and Curves of Maximal Slope in $\mathscr{P}_p(X)$

In this chapter we state some of the main results of the paper, concerning existence, uniqueness, approximation, and qualitative properties of gradient flows μ_t generated by a proper, l.s.c. functional ϕ in $\mathscr{P}_p(X)$, X being a separable Hilbert space. Taking into account the first part of this book and the (sub)differential theory developed in the previous chapter, there are at least four possible approaches to gradient flows which can be adapted to the framework of Wasserstein spaces:

1. **The "Minimizing Movement" approximation.** We can simply consider any limit curve of the variational approximation scheme we introduced at the beginning of Chapter 2 (see Definition 2.0.6), i.e. a "Generalized minimizing movement" $GMM(\Phi;\mu_0)$ in the terminology suggested by E. De Giorgi. In the context of $\mathscr{P}_2(\mathbb{R}^d)$ this procedure has been first used in [83, 104, 105, 103, 106] and subsequently it has been applied in many different contexts, e.g. by [82, 99, 107, 73, 74, 78, 66, 35, 36, 1, 75, 63].

2. **Curves of Maximal Slope.** We can look for absolutely continuous curves $\mu_t \in AC^p_{\rm loc}((0,+\infty);\mathscr{P}_p(X))$ which satisfy the differential form of the Energy inequality

$$\frac{d}{dt}\phi(\mu_t) \le -\frac{1}{p}|\mu'|^p(t) - \frac{1}{q}|\partial\phi|^q(\mu_t) \le -|\partial\phi|(\mu_t)\cdot|\mu'|(t) \tag{11.0.1}$$

for $\mathscr{L}^1$-a.e. $t \in (0,+\infty)$. Notice that in the present case of $\mathscr{P}_p(X)$, we established in Chapter 8 a precise description of absolutely continuous curve (in terms of the continuity equation) and of the metric velocity (in terms of the $L^p(\mu_t;X)$-norm of the related velocity vector field); moreover, in Chapter 10 we have shown an equivalent differential characterization of the slope $|\partial\phi|$ in terms of the $L^q(\mu_t;X)$-norm of the Fréchet subdifferential of ϕ.

3. The pointwise differential formulation. Since we have at our disposal a notion of tangent space and the related concepts of velocity vector field v_t and (sub)differential $\partial\phi(\mu_t)$, we can reproduce the simple definition of gradient flow modeled on smooth Riemannian manifold, i.e.

$$v_t \in -\partial\phi(\mu_t), \tag{11.0.2}$$

trying to adapt it to the case $p \neq 2$ and to extended plan subdifferentials.

4. Systems of Evolution Variational Inequalities (E.V.I.). When $p = 2$, in the case of λ-convex functionals along geodesics in $\mathscr{P}_2(X)$, we can try to find solutions of the family of "metric" variational inequalities

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t,\nu) \le \phi(\nu) - \phi(\mu_t) - \frac{\lambda}{2}W_2^2(\mu_t,\nu) \quad \forall\,\nu \in D(\phi). \tag{11.0.3}$$

This formulation provides the best kind of solutions, for which in particular one can prove not only uniqueness, but also error estimates. On the other hand it imposes severe restrictions on the space ($p = 2$) and on the functional (λ-convexity along generalized geodesics).

In any case, the variational approximation scheme is the basic tool for proving existence of gradient flows: at the highest level of generality, when the functional ϕ does not satisfy any convexity or regularity assumption, one can only hope to prove the existence of a limit curve which will satisfy a sort of "relaxed" differential equation: we will present the basic steps of the convergence argument at the end of the next section, in a simplified situation.

As we will see in the next section, when ϕ satisfies more restrictive regularity assumptions, one can show that the first three notions essentially coincide; if ϕ is also λ-convex and $p = 2$, they are also equivalent to the most restrictive fourth one.

It is then possible to prove the convergence of the discrete solutions to a curve of maximal slope (or to a solution of the E.V.I. system) by applying the general theorems of Chapter 2 (respectively, of Chapter 4): we will devote the last two sections to present a brief account of these metric approaches.

11.1 The gradient flow equation and its metric formulations

Definition 11.1.1 (Gradient flows). *We say that a map $\mu_t \in AC^p_{\rm loc}((0,+\infty);\mathscr{P}_p(X))$ is a solution of the gradient flow equation*

$$j_p(v_t) \in -\partial\phi(\mu_t) \quad t > 0, \tag{11.1.1}$$

if denoting by $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ its velocity vector field, its dual vector field $j_p(v_t)$ belongs to the (reduced) subdifferential (10.3.12) *of ϕ at μ_t for $\mathscr{L}^1$-a.e. $t > 0$.*

The above definition is equivalent to the requirement that there exists a Borel vector field v_t such that $v_t \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$ for $\mathscr{L}^1$-a.e. $t > 0$, $\|v_t\|_{L^p(\mu_t)} \in L^p_{\mathrm{loc}}(0, +\infty)$, the continuity equation

$$\partial_t \mu_t + \nabla \cdot \big(v_t \mu_t\big) = 0 \quad \text{in } X \times (0, +\infty) \tag{11.1.2}$$

holds in the sense of distributions according to (8.3.8), and finally

$$j_p(v_t) \in -\partial\phi(\mu_t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0. \tag{11.1.3}$$

The last inclusion is also equivalent (see (10.3.12) and Definition 10.3.1 for the definition of $\partial\phi$) to

$$\big(\boldsymbol{i} \times j_p(-v_t)\big)_{\#}\mu_t \in \boldsymbol{\partial}\phi(\mu_t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0. \tag{11.1.4}$$

Observe that in the case $p = 2$ (11.1.1) simplifies to

$$v_t \in -\partial\phi(\mu_t), \quad \text{or, equivalently,} \quad \big(\boldsymbol{i} \times (-v_t)\big)_{\#}\mu_t \in \boldsymbol{\partial}\phi(\mu_t), \tag{11.1.5}$$

for $\mathscr{L}^1$-a.e. $t > 0$. Before studying the question of existence of solutions to (11.1.1), which we will postpone to the next sections, we want to discuss some preliminary issues.

First of all we mention the basic (but formal, at this level) example, which provides one of the main motivations to study this kind of gradient flows.

Example 11.1.2 (Gradient flows and evolutionary PDE's of diffusion type). In the space-time open cylinder $\mathbb{R}^d \times (0, +\infty)$ we look for *nonnegative solutions* $\rho : \mathbb{R}^d \times (0, +\infty)$ of a parabolic equation of the type

$$\partial_t \rho - \nabla \cdot \Big(\rho \nabla \big(\frac{\delta \mathscr{F}}{\delta \rho}\big)\Big) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty), \tag{11.1.6}$$

where

$$\frac{\delta \mathscr{F}(\rho)}{\delta \rho} = -\nabla \cdot F_p(x, \rho, \nabla\rho) + F_z(x, \rho, \nabla\rho).$$

is the first variation of a typical integral functional as in (10.4.1)

$$\mathscr{F}(\rho) = \int_{\mathbb{R}^d} F(x, \rho(x), \nabla\rho(x))\, dx \tag{11.1.7}$$

associated to a (smooth) Lagrangian $F = F(x, z, p) : \mathbb{R}^d \times [0, +\infty) \times \mathbb{R}^d \to \mathbb{R}$.

Observe that (11.1.6) has the following structure:

$$\partial_t \rho + \nabla \cdot (\rho \boldsymbol{v}) = 0 \qquad \textit{(continuity equation)}, \tag{11.1.8a}$$

$$\rho \boldsymbol{v} = \rho \nabla \psi \qquad \textit{(gradient condition)}, \tag{11.1.8b}$$

$$\psi = -\frac{\delta \mathscr{F}(\rho)}{\delta \rho} \qquad \textit{(nonlinear relation)}. \tag{11.1.8c}$$

Observe that in the case when F depends only on $z = \rho$ then we have

$$\frac{\delta \mathscr{F}(\rho)}{\delta \rho} = F_z(\rho), \quad \rho \nabla F_z(x, \rho) = \nabla L_F(\rho), \quad L_F(z) := zF'(z) - F(z). \tag{11.1.9}$$

Since we look for *nonnegative solutions* having (constant, by (11.1.8a), normalized) *finite mass*

$$\rho(x,t) \geq 0, \quad \int_{\mathbb{R}^d} \rho(x,t)\, dx = 1 \quad \forall t \geq 0, \tag{11.1.10}$$

and *finite quadratic momentum*

$$\int_{\mathbb{R}^d} |x|^2 \rho(x,t)\, dx < +\infty \quad \forall\, t \geq 0, \tag{11.1.11}$$

recalling Example 10.4.1, we can

$$\text{identify } \rho \text{ with the measures } \mu_t := \rho(\cdot, t) \cdot \mathscr{L}^d, \tag{11.1.12}$$

and we consider $\mathscr{F}$ as a functional defined in $\mathscr{P}_2(\mathbb{R}^d)$. Then any smooth positive function ρ is a solution of the system (11.1.8a,b,c) if and only if μ is a solution in $\mathscr{P}_2(\mathbb{R}^d)$ of the Gradient Flow equation (11.1.1) for the functional $\mathscr{F}$.

Observe that (11.1.8a) coincides with (11.1.2), the gradient constraint (11.1.8b) corresponds to the tangent condition $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_2(\mathbb{R}^d)$ of (11.1.3), and the nonlinear coupling $\psi = -\delta\mathscr{F}(\rho)/\delta\rho$ is equivalent to the differential inclusion $v_t \in -\partial\mathscr{F}(\mu_t)$ of (11.1.3).

At this level of generality the equivalence between the system (11.1.8a,b,c) and the evolution equation (11.1.1) is known only for smooth solution (which, by the way, may not exist); nevertheless, the point of view of gradient flow in the Wasserstein spaces, which was introduced by F. Otto in a series of pioneering and enlightening papers [104, 83, 106, 107], still presents some interesting features, whose role should be discussed in each concrete case:

a) The gradient flow formulation (11.1.1) suggests a general variational scheme (the Minimizing Movement approach, which we discussed in the first part of this book and which we will apply in the next sections) to approximate the solution of (11.1.8a,b,c): proving its convergence is interesting both from the theoretical (cf. the papers quoted at the beginning of the chapter) and the numerical point of view [88].

b) The variational scheme exhibits solutions which are *a priori* nonnegative, even if the equation does not satisfies any maximum principle as in the fourth order case [105, 75].

c) Working in Wasserstein spaces allows for weak assumptions on the data: initial values which are general measures (as for fundamental solutions, in the linear cases) fit quite naturally in this framework.

d) The gradient flow structure suggests new contraction and energy estimates, which may be useful to study the asymptotic behaviour of solutions to (11.1.8a,b,c) [107, 17, 34, 38, 2, 118, 58], or to prove uniqueness under weak assumptions on the data.

e) The interplay with the theory of Optimal Transportation provides a novel point of view to get new functional inequalities with sharp constants [108, 125, 3, 44, 16, 54].

f) The variational structure provides an important tool in the study of the dependence of solutions from perturbation of the functional.

g) The setting in space of measures is particularly well suited when one considers evolution equations in infinite dimensions and tries to "pass to the limit" as the dimension d goes to ∞.

11.1.1 Gradient flows and curves of maximal slope

Our first step is to compare solutions to (11.1.1) with the curves of Maximal Slope we introduced in 1.3.2: we are thus discussing the equivalence of the second and of the third formulation introduced at the beginning of this chapter.

As usual, we are at least assuming that

$$\phi : \mathscr{P}_p(X) \to (-\infty, +\infty], \text{ proper and lower semicontinuous,} \tag{11.1.13a}$$

is such that

$$\begin{aligned} &\nu \mapsto \Phi(\tau, \mu; \nu) = \frac{1}{p\tau^{p-1}} W_p^p(\mu, \nu) + \phi(\nu) \quad \text{admits at least} \\ &\text{a minimum point } \mu_\tau, \text{ for all } \tau \in (0, \tau_*) \text{ and } \mu \in \mathscr{P}_p(X). \end{aligned} \tag{11.1.13b}$$

Theorem 11.1.3 (Curves of maximal slope coincide with gradient flows). *Let* $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ *be a* regular functional*, according to Definition* 10.3.9 *satisfying* (11.1.13a,b)*. Then* $\mu_t : (0, +\infty) \to \mathscr{P}_p(X)$ *is a p-curve of maximal slope w.r.t.* $|\partial\phi|$ *(according to Definition* 1.3.2*) iff* μ_t *is a gradient flow and* $t \mapsto \phi(\mu_t)$ *is* $\mathscr{L}^1$*-a.e. equal to a function of bounded variation. In this case the tangent vector field* v_t *to* μ_t *satisfies the* minimal selection principle

$$v_t = -\partial^\circ \phi(\mu_t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t > 0. \tag{11.1.14}$$

Proof. Assume first that μ_t is a p-curve of maximal slope w.r.t. $|\partial\phi|$. We know that there exists a function of (locally) bounded variation $\varphi : (0, +\infty) \to \mathbb{R}$ such that $\phi(\mu_t) = \varphi(t)$ $\mathscr{L}^1$-a.e. in $(0, +\infty)$ and

$$\frac{d}{dt}\varphi(t) = -\frac{1}{p}|\mu'|^p(t) - \frac{1}{q}|\partial\phi|^p(t) \qquad \mathscr{L}^1\text{-a.e. in } (0, +\infty). \tag{11.1.15}$$

Indeed, the inequality $\leq$ follows by the definition of curve of maximal slope, while the opposite inequality follows by the fact, proved in Theorem 1.2.5, that $|\partial\phi|$ is a weak upper gradient of ϕ.

Being ϕ regular, $\boldsymbol{\partial}\phi(\mu_t) \neq \emptyset$ for $\mathscr{L}^1$-a.e. $t > 0$; thus the chain rule 10.3.18 shows that

$$\frac{d}{dt}\varphi(t) = \int_{X^2} \langle x_2, v_t(x_1)\rangle \, d\boldsymbol{\gamma}_t \quad \forall \boldsymbol{\gamma}_t \in \partial\phi(\mu_t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0. \tag{11.1.16}$$

Choosing in particular $\boldsymbol{\gamma}_t = \boldsymbol{\partial}^\circ\phi(\mu_t)$, since the equalities

$$|\mu'|^p(t) = \int_{X^2} |v_t(x_1)|^p \, d\boldsymbol{\gamma}_t, \qquad |\partial\phi|^q(t) = \int_{X^2} |x_2|^q \, d\boldsymbol{\gamma}_t,$$

hold for $\mathscr{L}^1$-a.e. $t > 0$, we get

$$\int_{X^2} \Big(\frac{1}{p}|v_t(x_1)|^p + \frac{1}{q}|x_2|^q + \langle x_2, v_t(x_1)\rangle\Big) \, d\boldsymbol{\gamma}_t(x_1, x_2) = 0. \tag{11.1.17}$$

It follows that

$$x_2 = -j_p(v_t(x_1)) \quad \text{for } \boldsymbol{\gamma}_t \text{ a.e. } (x_1, x_2),$$

i.e. $(\boldsymbol{i} \times j_p(-v_t))_{\#}\mu_t = \boldsymbol{\partial}^\circ\phi(\mu_t)$ or, equivalently, $j_p(v_t) = -\partial^\circ\phi(\mu_t)$.

Conversely, if μ_t is a gradient flow in the sense of (11.1.1) and $\phi \circ \mu$ is a function of (essential) bounded variation, by applying the chain rule once more, we easily get that μ_t is a p-curve of maximal slope w.r.t. $|\partial\phi|$. □

One of the most interesting aspects of the previous characterization is to force $\boldsymbol{\partial}^\circ\phi(\mu_t)$ to be concentrated on the graph of the transport map $-j_p(v_t)$ for $\mathscr{L}^1$-a.e. $t > 0$, even if the measures μ_t do not satisfy any regularity assumption.

11.1.2 Gradient flows for λ-convex functionals

If the functional ϕ is λ-convex along geodesics, for flows $\mu_t : (0, +\infty) \to \mathscr{P}_p^r(X)$ with $\|v_t\|^p_{L^p(\mu_t)}$ locally integrable, Definition (11.1.1) reduces to the system

$$\begin{cases} \partial_t\mu_t + \nabla\cdot(\mu_t\, v_t) = 0 \quad \text{in } X \times (0, +\infty), \\[2ex] -\displaystyle\int_X \langle j_p(v_t), \boldsymbol{t}_{\mu_t}^\sigma - \boldsymbol{i}\rangle \, d\mu_t \leq \phi(\sigma) - \phi(\mu_t) - \frac{\lambda}{2}W_p^2(\sigma, \mu_t) \quad \forall \sigma \in D(\phi), \end{cases} \tag{11.1.18}$$

where the first equation is understood in a weak sense, in the duality with cylindrical functions in $X \times (0, +\infty)$, and the second one holds for $\mathscr{L}^1$-a.e. $t > 0$.

Notice that in the case $p = 2$ it is not necessary to assume that v_t is tangent in (11.1.18): indeed, projecting the velocity field onto the tangent space leaves the continuity equation unchanged (by (8.4.4)) and does not affect the subdifferential inequality (by Proposition 8.5.2). However, whenever the norm of the velocity field has to be minimized, it is natural to assume that the vector field is tangent.

The case $p = 2$ is particularly distinguished, since the left hand side of (11.1.18) is also equal to the time derivative of $\frac{1}{2}W_2^2(\mu_t,\nu)$; this simple fact is the crucial ingredient of the following uniqueness result, which is known in the case $p = 2$ only. This is not very surprising, as even in the "flat" L^p spaces, $p \in (1,+\infty)\setminus\{2\}$, uniqueness of gradient flows is not known.

Theorem 11.1.4 (Uniqueness of gradient flows in the case $p = 2$ and E.V.I.). *Let $p = 2$ and let $\phi : \mathscr{P}_2(X) \to (-\infty,+\infty]$ be a l.s.c. λ-geodesically convex functional. If $\mu_t^i : (0,+\infty) \to \mathscr{P}_2(X)$, $i = 1,2$, are gradient flows satisfying $\mu_t^i \to \mu^i$ as $t \downarrow 0$ in $\mathscr{P}_2(X)$, then*

$$W_2(\mu_t^1,\mu_t^2) \le e^{-\lambda t} W_2(\mu^1,\mu^2) \quad \forall\, t > 0. \tag{11.1.19}$$

In particular, for any $\mu_0 \in \mathscr{P}_2(X)$ there is at most one gradient flow μ_t satisfying the initial Cauchy condition $\mu_t \to \mu_0$ as $t \downarrow 0$ and it is also characterized by the system of "Evolution Variational Inequalities"

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t,\sigma) + \frac{\lambda}{2}W_2^2(\mu_t,\sigma) \le \phi(\sigma) - \phi(\mu_t) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t > 0,\ \forall\, \sigma \in D(\phi). \tag{11.1.20}$$

Proof. Let $\sigma \in D(\phi)$ and let μ_t be a gradient flow satisfying $\mu_t \to \mu_0$ as $t \downarrow 0$. Denoting by v_t the velocity vector of μ_t, and applying the definition of subdifferential we obtain the existence of $\boldsymbol{\gamma}_t \in \Gamma_o(\mu_t,\sigma)$ such that

$$\phi(\sigma) \ge \phi(\mu_t) + \int_{X^2} \langle v_t(x_2), x_1 - x_2\rangle \, d\boldsymbol{\gamma}_t + \frac{\lambda}{2}W_2^2(\mu_t,\sigma). \tag{11.1.21}$$

On the other hand the differentiability of W_2^2 stated in Lemma 8.4.7 gives

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t,\sigma) = \int_{X^2} \langle v_t(x_1), x_1 - x_2\rangle \, d\boldsymbol{\gamma}_t \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,+\infty),$$

and therefore (11.1.20).

Conversely, if μ_t is an absolutely continuous curve satisfying (11.1.20), it is immediate to check that for every *countable subset* $\Sigma \subset D(\phi)$ we can find a $\mathscr{L}^1$-negligible set $\mathscr{N} \subset (0,+\infty)$ such that the velocity vector v_t satisfies (11.1.21) for every $\sigma \in \Sigma$ and $t \in (0,+\infty)\setminus\mathscr{N}$. We can choose now a countable set Σ which is dense in $D(\phi)$ with respect to the distance $W_2(\mu,\nu) + |\phi(\mu)-\phi(\nu)|$ (see Proposition 7.1.5): by a density argument based on Proposition 7.1.3 we conclude that $-v_t \in \partial\phi(\mu_t)$ for $t \in (0,+\infty)\setminus\mathscr{N}$.

Finally, if μ_t^1, μ_t^2 are two gradient flows satisfying the initial Cauchy condition $\mu_t^i \to \mu^i$ as $t \downarrow 0$, $i = 1,2$, it is easy to check that we can apply Lemma 4.3.4 with the choices $d(s,t) := W_2^2(\mu_s^1,\mu_t^2)$, $\delta(t) := d(t,t)$, thus obtaining $\delta' \le -2\lambda\delta$. Since $\delta(0_+) = W_2^2(\mu^1,\mu^2)$ we obtain (11.1.19). □

11.1.3 The convergence of the "Minimizing Movement" scheme

The existence of solutions to (11.1.1) will be obtained as limit of the metric variational scheme we discussed in Chapter 2: let us briefly recall some notation we will extensively use.

The discrete equation Here we consider (for simplicity: see Remark (2.0.1)) a uniform partition $\mathcal{P}_\tau$ of $(0,+\infty)$ by intervals I_τ^n of size $\tau > 0$

$$\mathcal{P}_\tau := \big\{0 < t_\tau^1 = \tau < t_\tau^2 = 2\tau < \cdots < t_\tau^n = n\tau < \cdots\big\}, \quad I_\tau^n := ((n-1)\tau, n\tau];$$

for a given family of initial values M_τ^0 such that

$$M_\tau^0 \to \mu_0 \quad \text{in } \mathscr{P}_p(X), \quad \phi(M_\tau^0) \to \phi(\mu_0) \quad \text{as } \tau \downarrow 0, \tag{11.1.22}$$

assuming that (11.1.13b) is satisfied, for every $\tau \in (0,\tau_*)$ a corresponding family of sequences $(M_\tau^n)_{n\in\mathbb{N}}$ recursively defined as

$$M_\tau^n \quad \text{minimizes} \quad \mu \mapsto \Phi(\tau, M_\tau^{n-1};\mu) \tag{11.1.23}$$

always exists. We call "discrete solution" the piecewise constant interpolant

$$\overline{M}_\tau(t) := M_\tau^n \quad \text{if } t \in ((n-1)\tau, n\tau], \tag{11.1.24}$$

and we say that a curve μ is a Generalized Minimizing Movement of $GMM(\Phi;\mu_0)$ if there exists a sequence $(\tau_k) \downarrow 0$ such that

$$\overline{M}_{\tau_k}(t) \to \mu_t \quad \text{narrowly in } \mathscr{P}(X_\varpi) \text{ for every } t > 0, \quad \text{as } k \to \infty. \tag{11.1.25}$$

It follows from Proposition 2.2.3 that if $\mu_0 \in D(\phi)$ then a generalized Minimizing Movement always exists and it is an absolutely continuous curve $\mu \in AC^p_{\rm loc}\big([0,+\infty);\mathscr{P}_p(X)\big)$.

The main problem is to characterize the equation satisfied by its tangent velocity vector v_t, or, equivalently, to pass to the limit in the "discrete gradient flow" equation satisfied by the discrete solution $\overline{M}_\tau$.

In order to clarify this point, let us first suppose for simplicity, as we did in Section 10.1, that

$$D(|\partial\phi|) \subset \mathscr{P}_p^r(X). \tag{11.1.26}$$

If $\boldsymbol{t}_\tau^n$ is the optimal transport map pushing M_τ^n to M_τ^{n-1}, it is natural to define the discrete velocity vector $\boldsymbol{V}_\tau^n$ as $(\boldsymbol{i} - \boldsymbol{t}_\tau^n)/\tau$. By Lemma 10.1.2 and Theorem 10.4.12

$$-j_p(\boldsymbol{V}_\tau^n) = j_p\Big(\frac{\boldsymbol{t}_\tau^n - \boldsymbol{i}}{\tau}\Big) \in \partial\phi(M_\tau^n), \tag{11.1.27}$$

which can be considered as an Euler implicit discretization of (11.1.1). By introducing the piecewise constant interpolant

$$\overline{\boldsymbol{V}}_\tau(t) := \boldsymbol{V}_\tau^n \quad \text{if } t \in ((n-1)\tau, n\tau], \tag{11.1.28}$$

the identity (11.1.27) reads

$$-j_p(\overline{\boldsymbol{V}}_\tau(t)) \in \partial\phi(\overline{M}_\tau(t)) \quad \text{for } t > 0. \tag{11.1.29}$$

It is not difficult to show that, up to subsequences, $\overline{\boldsymbol{V}}_\tau \rightharpoonup \boldsymbol{v}$ in the distribution sense in $X \times (0, +\infty)$, for some vector field $\boldsymbol{v}$ satisfying

$$\partial_t \mu_t + \nabla \cdot (\mu_t \boldsymbol{v}_t) = 0 \quad \text{in } X \times (0, +\infty), \quad \|\boldsymbol{v}_t\|_{L^p(\mu_t;X)} \in L^p_{\rm loc}(0, +\infty). \tag{11.1.30}$$

A we already said, the main difficulty is to show that the (doubly, if $p \neq 2$) nonlinear equation (11.1.29) is preserved in the limit.

We shall see three different kinds of arguments which give some insights for this problem and reflect different properties of the functional.

1. The first one is a direct development of the compactness method: passing to the limit in the discrete equation satisfied at each step by the approximating sequence M^n_τ, one tries to write a relaxed form of the limit differential equation, assuming only narrow convergences of weak type. It may happen that under suitable closure and convexity assumptions on the sections of the subdifferential, which should be checked in each particular situation, this relaxed version coincides with the stronger one, and therefore one gets an effective solution to (11.1.1).

   In general, however, even under some simplifying assumptions ($p = 2$, all the measures are regular), the results of this direct approach are not completely satisfactory: it could be considered as a first basic step, which should be common to each attempt to apply the Wasserstein formalism for studying a gradient flow.

   In order to clarify the basic arguments of this preliminary strategy, we will try to explain it at the end of this section (in a simplified setting, to keep the presentation easier) without invoking all the abstract results of the first part of this book.

2. The second approach (see Section 11.3) involves the *regularity of the functional* according to Definition 10.3.9, and works for every $p > 1$; in particular it can applied to λ-convex functionals.

   In this case, thanks to Theorem 11.1.3, the gradient flow equation is equivalent to the maximal slope condition, which is of purely metric nature. We can then apply the abstract theory we presented in Chapter 2 and therefore we can prove that any limit curve μ of (11.1.28) is a solution to (11.1.1).

   The key ingredient, which allows to pass to the limit in the "doubly nonlinear" differential inclusion (without any restrictions on the regularity of μ_t) and to gain a better insight on the limit than the previous simpler method, is the refined discrete energy estimate (3.2.4) (related to De Giorgi's variational interpolation (3.2.1)) and the lower semicontinuity of the slope, which follows from the regularity of the functional.

3. The last approach, presented in Section 11.2, is based on the general estimates of Theorem 4.0.4 of Chapter 4: it can be performed only in $\mathscr{P}_2(X)$ and imposes on the functionals the strongest condition of λ-convexity along generalized geodesics.

   Despite the strong convexity requirements on ϕ, which are nevertheless satisfied by all the examples of Section 10.4 in $\mathscr{P}_2(X)$, this approach has many nice features:

   - it does not require compactness assumptions of the sublevels of ϕ in $\mathscr{P}_2(X)$: the convergence of the "Minimizing movement" scheme is proved by a Cauchy-type estimate.
   - The gradient flow equation (11.1.1) is satisfied in the limit, since Theorem 4.0.4 provides directly the system of evolution variational inequalities (11.1.20).
   - It provides our strongest results in terms of regularity, asymptotic behaviour, and error estimates for the continuous solution, which can be directly derived from the general metric setting.
   - It allows for general initial data μ_0 which belong to the *closure* of the domain of ϕ: in particular, one can often directly consider a sort of "nonlinear fundamental solution" for initial values which are concentrated in one point.
   - The Γ-convergence of functionals, in the sense of Lemma 10.3.16, induces the uniform convergence of the corresponding gradient flows.

Let us now present a brief sketch of the first approach: as in Section 10.1 we are assuming $p = 2$ and (11.1.26). We also introduce a limiting version of the subdifferential, modeled on the analogous one introduced by [90, 100] in linear spaces (see also the monograph [113] and [114] for applications to gradient flows in Hilbert spaces).

Definition 11.1.5 (Limiting subdifferentials). *For $\mu \in D(\phi)$, we say that a vector $\boldsymbol{\xi} \in L^2(\mu; X)$ belongs to the* limiting subdifferential $\partial_\ell\phi(\mu)$ of ϕ at μ *if there exist two sequences $\mu_k \in D(\partial\phi)$, $\boldsymbol{\xi}_k \in \partial\phi(\mu_k)$, $\boldsymbol{\xi}_k$ being* strong subdifferentials, *such that*

$$\mu_k \to \mu \quad \textit{narrowly in } \mathscr{P}(X_\varpi), \quad \boldsymbol{\xi}_k \to \boldsymbol{\xi} \quad \textit{weakly, as in Definition 5.4.3,}$$
$$\sup_k \Big(\phi(\mu_k), \int_X \big(|x|^2 + |\boldsymbol{\xi}_k(x)|^2\big)\, d\mu_k(x)\Big) < +\infty. \tag{11.1.31}$$

The following result has a simpler counterpart in the flat framework of Hilbert spaces:

Theorem 11.1.6 (Relaxed gradient flow). *Let us suppose that $p = 2$, the proper and coercive functional $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ is l.s.c. with respect to the narrow*

convergence of $\mathscr{P}(X_\varpi)$*, and it satisfies* $D(|\partial\phi|) \subset \mathscr{P}_2^r(X)$*.*
If $\mu_0 \in D(\phi)$*, then each sequence of discrete solutions* $(\overline{M}_{\tau_h})$ *with vanishing time steps admits a converging subsequence (still denoted by* $\overline{M}_{\tau_h}$*) such that:*

(i) $\overline{M}_{\tau_h}(t) \to \mu_t$ *in* $\mathscr{P}(X_\varpi) \quad \forall\, t \in [0,+\infty)$*;*

(ii) *for any* $T > 0$ *we have* $\overline{\boldsymbol{V}}_{\tau_h} \to \boldsymbol{v} \in L^2(\mu; X)$ *weakly in* $X \times (0,T)$*, as in Definition* 5.4.3*, where* $T^{-1}\mu \in \mathscr{P}(X \times (0,T))$ *is the measure* $T^{-1}\int_0^T \mu_t \, d\mathscr{L}^1(t)$*.*

The map $t \mapsto \mu_t$ *belongs to* $AC^2_{\mathrm{loc}}([0,+\infty); \mathscr{P}_2(X))$*, it satisfies the* continuity equation

$$\partial_t \mu + \nabla \cdot (\mu \boldsymbol{v}) = 0, \tag{11.1.32}$$

and $\boldsymbol{v}$ *satisfies the* relaxed gradient flow

$$-\boldsymbol{v}_t \in \overline{\mathrm{Conv}}\, \partial_\ell \phi(\mu_t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0. \tag{11.1.33}$$

In other words, the limit vector field $-\boldsymbol{v}_t$ belongs to the *closed convex hull* in $L^2(\mu_t; X)$ of the limiting subdifferential $\partial_\ell \phi(\mu_t)$. Before presenting the proof of this theorem, let us show two easy corollaries and two related motivating applications: the proof of the first statement follows from Lemma 10.1.3, providing the inclusion of the limiting subdifferential into the standard subdifferential for coercive and λ-convex functionals.

Corollary 11.1.7 (λ-convex functionals in $\mathscr{P}_2^r(X)$). *Under the same assumptions of the previous theorem, suppose that* ϕ *is a* λ*-geodesically convex functional satisfying* (11.1.26)*, whose sublevels are locally compact in* $\mathscr{P}_2(X)$*; then* μ_t *is a solution of the gradient flow equation* (11.1.5)*.*

Corollary 11.1.8 (Single–valued limiting subdifferential). *Under the same assumptions of the above theorem, suppose that* $\partial_\ell \phi(\mu)$ *contains at most one vector. Then* μ_t *is a solution of*

$$\partial_t \mu + \nabla \cdot (\mu \boldsymbol{v}) = 0, \quad -\boldsymbol{v}_t = \partial_\ell \phi(\mu_t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0. \tag{11.1.34}$$

Example 11.1.9 (Diffusion equations without geodesic convexity). Let $F : [0,+\infty) \to \mathbb{R}$ be a convex, doubling, differentiable functional with superlinear growth satisfying (9.3.8) and let $V : \mathbb{R}^d \to (-\infty, +\infty]$ be a l.s.c. potential. which is bounded from below and locally Lipschitz in the (nonempty) interior Ω of its proper domain $D(V)$, with $\mathscr{L}^d(\partial\Omega) = 0$. Even if the related functional

$$\phi(\mu) := \int_\Omega \Big(F(\rho) + \rho(x) V(x) \Big)\, dx, \quad \mu = \rho \cdot \mathscr{L}^d, \tag{11.1.35}$$

is not λ-geodesically convex in $\mathscr{P}_2(\mathbb{R}^d)$ (since we do not ask for (10.4.21) and V could not satisfy any λ-convexity property), it is not difficult to check that ϕ satisfies the assumptions of Theorem 11.1.6 and

$$\boldsymbol{w} \in \partial_\ell \phi(\mu) \quad \Rightarrow \quad L_F(\rho) \in W^{1,1}_{\mathrm{loc}}(\Omega), \quad \nabla L_F(\rho) = \rho(\boldsymbol{w} - \nabla V). \tag{11.1.36}$$

For, taking directional derivatives of ϕ along smooth transport vector fields $\boldsymbol{t} \in C_c^\infty(\Omega;\mathbb{R}^d)$ as in Section 10.4.1 (see also Lemma 10.4.4), we easily find that any *strong* subdifferential $\boldsymbol{w}_n \in \partial\phi(\mu_n)$ satisfies (11.1.36) and the same argument of Proposition 10.4.14 shows that this relation holds for the limit $\boldsymbol{w} \in \partial_\ell\phi(\mu)$.

Therefore, *for every $\mu_0 = \rho_0\mathscr{L}^d \in D(\phi)$ there exists a solution $\mu_t = \rho_t\mathscr{L}^d$ of the equation*

$$\begin{cases} \partial_t\rho_t + \nabla\cdot(\rho_t\boldsymbol{v}_t) = 0 \quad \textit{in } \mathscr{D}'(\mathbb{R}^d\times(0,+\infty)), \\ -\rho_t\boldsymbol{v}_t = \nabla L_F(\rho_t) + \rho_t\nabla V \quad \textit{in } \Omega, \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t>0. \end{cases} \tag{11.1.37}$$

satisfying the initial condition $\rho_t \to \rho_0$ as $t\downarrow 0$ weakly in $L^1(\mathbb{R}^d)$ and having (locally) finite energy

$$\int_0^T\int_\Omega |\boldsymbol{v}_t(x)|^2\,d\mu_t(x)\,dt = \int_0^T\int_\Omega \frac{|\nabla L_F(\rho_t)+\rho_t\nabla V|^2}{\rho_t}\,dx\,dt < +\infty. \tag{11.1.38}$$

Notice that, even if Ω is bounded, the first equation of (11.1.37) is still imposed in $\mathbb{R}^d$ (in the distribution sense), and therefore it provides a weak formulation of the Neumann boundary condition

$$\partial_{\boldsymbol{n}}(L_F(\rho_t)+\rho_t V) = 0 \quad \text{on } \partial\Omega\times(0,+\infty). \tag{11.1.39}$$

The main difference with the results we are going to show in the case of λ-geodesically convex functionals is that we do not know if a solution of (11.1.38), (11.1.39) is indeed a gradient flow in the variational sense or the differential sense.

Example 11.1.10 (The Quantum drift-diffusion equation as gradient flow of the Fisher information). Let us consider the Fisher information functional (relative to the Lebesgue measure), introduced in 10.4.16

$$\mathscr{I}(\mu) = \mathscr{I}(\mu|\mathscr{L}^d) := \int_{\mathbb{R}^d}\frac{|\nabla\rho(x)|^2}{\rho(x)}\,dx = 4\int_{\mathbb{R}^d}\big|\nabla\sqrt{\rho(x)}\big|^2\,dx \\ \text{if } \mu=\rho\cdot\mathscr{L}^d \text{ with } \sqrt{\rho}\in W^{1,2}(\mathbb{R}^d),\ +\infty \text{ otherwise.} \tag{11.1.40}$$

It is an integral functional as in (11.1.7) corresponding to the (non smooth) Lagrangian

$$F(x,z,p) := \frac{|p|^2}{z}, \quad \text{with} \quad \frac{\delta\mathscr{I}(\rho)}{\delta\rho} = -4\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}} = -4\frac{\sqrt{\rho}\,\Delta\sqrt{\rho}}{\rho}. \tag{11.1.41}$$

It is not known if $\mathscr{I}$ enjoys some λ-convexity or regularity property, but it is still possible to prove [75] that if $\boldsymbol{w}\in L^2(\mu;\mathbb{R}^d)$, $\mu=\rho\cdot\mathscr{L}^d\in\mathscr{P}_2(\mathbb{R}^d)$, then

$$\boldsymbol{w}\in\partial_\ell\mathscr{I}(\mu) \quad\Longrightarrow\quad \begin{cases} \sqrt{\rho}\in W^{2,2}(\mathbb{R}^d), \quad \sqrt{\rho}\,\Delta\sqrt{\rho}\in W^{1,1}(\mathbb{R}^d), \\ \rho\boldsymbol{w} = -4\Big(\nabla\big(\sqrt{\rho}\,\Delta\sqrt{\rho}\big) - 2\nabla\sqrt{\rho}\,\Delta\sqrt{\rho}\Big), \end{cases} \tag{11.1.42}$$

which is an indirect way to write as in (10.2.1)

$$\boldsymbol{w} = \nabla\Big(-4\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\Big) = \nabla\Big(\frac{\delta\mathscr{I}(\rho)}{\delta\rho}\Big).$$

Thanks to (11.1.42) and to Corollary 11.1.8, *for every* $\mu_0 = \rho_0\mathscr{L}^d \in D(\mathscr{I})$ *there exists a solution* $\mu_t = \rho_t\mathscr{L}^d$ *of the Quantum drift-diffusion equation*

$$\partial_t\rho_t + 4\nabla\cdot\Big(\rho_t\nabla\frac{\Delta\sqrt{\rho_t}}{\sqrt{\rho_t}}\Big) = 0 \quad \textit{in } \mathbb{R}^d\times(0,+\infty) \tag{11.1.43}$$

in the sense of (11.1.34) *and* (11.1.42), *and satisfying the initial condition* $\rho_t \to \rho_0$ *as* $t\downarrow 0$ *weakly in* $L^1(\mathbb{R}^d)$.
We refer to [24, 86] for a different approach to the above equation, and to [32, 87, 75] for further investigations and results.

Proof of Theorem 11.1.6. For the sake of simplicity, here we present the proof of Corollary 11.1.8 for a *nonnegative* functional ϕ. The estimates for the general case can be obtained as in Lemma 3.2.2 by means of the discrete Gronwall Lemma 3.2.4, whereas the relaxed inclusion (11.1.33) will follow from Lemma 12.2.2 in the Appendix.
Step 1: a priori estimates. We easily have

$$\frac{\tau}{2}\Big(\frac{W_2(M^n_\tau, M^{n-1}_\tau)}{\tau}\Big)^2 + \phi(M^n_\tau) \le \phi(M^{n-1}_\tau), \tag{11.1.44}$$

which yields

$$\phi(M^n_\tau) \le \phi(M^0_\tau) \quad \forall\, n\in\mathbb{N}, \qquad \sum_{n=1}^{+\infty}\frac{\tau}{2}\Big(\frac{W_2(M^n_\tau, M^{n-1}_\tau)}{\tau}\Big)^2 \le \phi(M^0_\tau); \tag{11.1.45}$$

in terms of $\overline{M}_\tau$ it means that

$$\sup_{t\ge 0}\phi(\overline{M}_\tau(t)) \le \phi(M^0_\tau) \quad \forall\,\tau>0. \tag{11.1.46}$$

From the last inequality of (11.1.45) we get for $0\le m\le n$

$$\begin{aligned} W_2(M^n_\tau, M^m_\tau) &\le \tau\sum_{k=m+1}^{n}\frac{W_2(M^k_\tau, M^{k-1}_\tau)}{\tau} \\ &\le \left(\tau\sum_{k=1}^{n}\frac{W_2^2(M^k_\tau, M^{k-1}_\tau)}{\tau}^2\right)^{1/2}\Big((m-n)\tau\Big)^{1/2} \\ &\le \Big(2\phi(M^0_\tau)\Big)^{1/2}\Big((m-n)\tau\Big)^{1/2}. \end{aligned} \tag{11.1.47}$$

Step 2: compactness and limit trajectory μ_t. (11.1.47) and (11.1.22) show that in each bounded interval $(0,T)$ the values $\{\overline{M}_\tau(t)\}_{\tau>0}$ are bounded in $\mathscr{P}_2(X)$, thus belong to a fixed compact set for the narrow topology of $\mathscr{P}(X_\varpi)$.

By connecting every pair of consecutive discrete values M_τ^{n-1}, M_τ^n with a constant speed geodesic parametrized in the interval $[t_\tau^{n-1}, t_\tau^n]$, we obtain by (11.1.47) a family of Lipschitz curves $\hat{M}_\tau$ satisfying

$$\begin{aligned} W_2(\hat{M}_\tau(t), \hat{M}_\tau(s)) \le C(t-s)^{1/2}, \\ W_2(\hat{M}_\tau(t), \overline{M}_\tau(t)) \le C\sqrt{\tau} \quad \forall\, t,s \in [0,T], \end{aligned} \tag{11.1.48}$$

where C is a constant independent of τ. Since the curves $\hat{M}_\tau$ are uniformly equicontinuous w.r.t. the 2-Wasserstein distance, which induces a stronger convergence than the narrow one of $\mathscr{P}(X_\varpi)$, Ascoli-Arzelà Theorem yields the relative compactness of the family $\{\hat{M}_{\tau_h}\}_{h\in\mathbb{N}}$ in $C^0([0,T];\mathscr{P}(X_\varpi))$ for each bounded interval $[0,T]$; we can therefore extract a vanishing subsequence (still denoted by τ_h) such that statement (i) holds.

Step 3: space-time measures and construction of $\boldsymbol{v}$. Recall that $\boldsymbol{t}_\tau^n$ is the optimal transport map pushing M_τ^n to M_τ^{n-1}, and that the discrete velocity vector $\boldsymbol{V}_\tau^n$ is defined by $(\boldsymbol{i} - \boldsymbol{t}_\tau^n)/\tau$. Let us introduce the discrete rescaled optimal plans

$$\boldsymbol{\gamma}_\tau^n := (\boldsymbol{i} \times \boldsymbol{V}_\tau^n)_\# M_\tau^n \tag{11.1.49}$$

and the piecewise constant interpolants

$$\overline{\boldsymbol{\gamma}}_\tau(t) := \boldsymbol{\gamma}_\tau^n, \qquad \overline{\boldsymbol{t}}_\tau(t) := \boldsymbol{t}_\tau^n \quad \text{if } t \in ((n-1)\tau, n\tau]. \tag{11.1.50}$$

For every bounded time interval $I_T := (0,T]$, denoting by $X_T := X \times I_T$, we can canonically identify $T^{-1}\overline{M}_\tau$, $T^{-1}\mu$ to elements of $\mathscr{P}(X_T)$ and $T^{-1}\overline{\boldsymbol{\gamma}}_\tau$ to an element in $\mathscr{P}(X_T \times X)$, simply by integrating with respect to the (normalized) Lebesgue measure $T^{-1}\mathscr{L}^1$ in I_T. Therefore $\overline{\boldsymbol{V}}_\tau$ is a vector field in $L^2(\overline{M}_\tau; X)$ and $\overline{\boldsymbol{\gamma}}_\tau$ is related to $\overline{M}_\tau$ by

$$\overline{\boldsymbol{\gamma}}_\tau = (\boldsymbol{i}_T \times \overline{\boldsymbol{V}}_\tau)_\# \overline{M}_\tau, \quad \boldsymbol{i}_T(x,t) := x \text{ being the projection of } X_T \text{ onto } X. \tag{11.1.51}$$

Then (11.1.45) yields

$$\begin{aligned} \int_0^T \int_X |\overline{\boldsymbol{V}}_\tau(x,t)|^2 \, d\big(\overline{M}_\tau(t)\big)(x)\, dt &= \int_{X_T} |\overline{\boldsymbol{V}}_\tau(x,t)|^2 \, d\overline{M}_\tau(x,t) \\ &= \int_{X_T \times X} |x_2|^2 \, d\overline{\boldsymbol{\gamma}}_\tau \le 2\phi(M_\tau^0). \end{aligned} \tag{11.1.52}$$

Hence, by Lemma 5.1.12(e), the family $\overline{\boldsymbol{\gamma}}_\tau$ is tight w.r.t. the narrow convergence of $\mathscr{P}(X_\varpi \times I_T \times X_\varpi)$. Denoting by $\boldsymbol{\gamma}$ the narrow limit (up to the extraction of a further subsequence, not relabeled) of $\boldsymbol{\gamma}_{\tau_h}$ in $\mathscr{P}(X_\varpi \times I_T \times X_\varpi)$, passing to the limit as $h \to \infty$ in the identity $\pi^{1,2}_\# \overline{\boldsymbol{\gamma}}_{\tau_h} = \overline{M}_{\tau_h}$ we obtain

$$\pi^{1,2}_\# \boldsymbol{\gamma} = \mu. \tag{11.1.53}$$

Therefore, we define

$$\boldsymbol{v}(x,t) := \int_X x_2 \boldsymbol{\gamma}_{x_1,t}(x_2) \tag{11.1.54}$$

where $\boldsymbol{\gamma}_{x_1,t}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. μ. Then, Theorem 5.4.4 and (11.1.52) give

$$\int_{X_T} |\boldsymbol{v}(x,t)|^2 \, d\mu(x,t) \le \liminf_{h\to\infty} \int_{X_T} |\overline{\boldsymbol{V}}_{\tau_h}(x,t)|^2 \, d\overline{M}_{\tau_h}(x,t) \le 2\phi(\mu_0). \tag{11.1.55}$$

Step 4: the limits μ, $\boldsymbol{v}$ satisfy the continuity equation (11.1.32).
The following argument was introduced, in a simpler setting, in [83]. Let us first observe that for every smooth cylindrical function $\psi \in \mathrm{Cyl}(X)$ we have

$$\begin{aligned}
&\int_X \psi(x)\, d\big(\overline{M}_\tau(t)\big)(x) - \int_X \psi(x)\, d\big(\overline{M}_\tau(t-\tau)\big)(x) \\
&\quad = \int_X \big(\psi(x) - \psi(\overline{\boldsymbol{t}}_\tau(x,t))\big)\, d\big(\overline{M}_\tau(t)\big)(x) \\
&\quad = \int_X \langle \nabla\psi(x), x - \overline{\boldsymbol{t}}_\tau(x,t)\rangle\, d\big(\overline{M}_\tau(t)\big)(x) + \varepsilon(\tau,\psi,t) \\
&\quad = \tau \int_{X\times X} \langle \nabla\psi(x_1), x_2\rangle\, d\big(\overline{\boldsymbol{\gamma}}_\tau(t)\big)(x_1,x_2) + \varepsilon(\tau,\psi,t),
\end{aligned}$$

where, for a suitable constant C_ψ depending only on the second derivatives of ψ

$$\begin{aligned}
|\varepsilon(\tau,\psi,t)| &= \left| \int_X \Big(\psi(x) - \psi(\overline{\boldsymbol{t}}_\tau(x,t)) - \nabla\psi(x)\cdot(x - \overline{\boldsymbol{t}}_\tau(x,t))\Big)\, d\big(\overline{M}_\tau(t)\big)(x) \right| \\
&\le C_\psi \int_X |x - \overline{\boldsymbol{t}}_\tau(x,t)|^2\, d\big(\overline{M}_\tau(t)\big)(x) = C_\psi \tau^2 \int_X \big|\overline{\boldsymbol{V}}_\tau(x,t)\big|^2\, d\big(\overline{M}_\tau(t)\big)(x).
\end{aligned}$$

Choosing now $\phi \in \mathrm{Cyl}(X_T)$, applying the estimate above with $\psi(\cdot) = \phi(t,\cdot)$ and taking into account (11.1.52), we have

$$\begin{aligned}
-\int_{X_T} \partial_t\phi(x,t)\, d\mu(x,t) &= \lim_{h\to\infty} -\int_{X_T} \partial_t\phi(x,t)\, d\overline{M}_{\tau_h}(x,t) = \\
&= \lim_{h\to\infty} -\tau_h^{-1} \int_{X_T} \big(\phi(x,t+\tau_h) - \phi(x,t)\big)\, d\overline{M}_{\tau_h}(x,t) \\
&= \lim_{h\to\infty} \int_{X_T\times X} \langle \nabla\phi(x_1,t), x_2\rangle\, d\boldsymbol{\gamma}_{\tau_h}(x_1,t,x_2) + \tau_h^{-1} \int_0^T \varepsilon(\tau_h, \phi(t,\cdot), t)\, dt \\
&= \int_{X_T\times X} \langle \nabla\phi(x_1,t), x_2\rangle\, d\boldsymbol{\gamma}(x_1,t,x_2) = \int_{X_T} \langle \nabla\phi(x_1,t), \boldsymbol{v}(x_1,t)\rangle\, d\mu(x_1,t).
\end{aligned}$$

Step 5: the limits μ, $\boldsymbol{v}$ satisfy the relaxed equation $-\boldsymbol{v}_t \in \partial_\ell\phi(\mu_t)$. By (11.1.55) and Fatou's Lemma, there exists a Borel set $I_0 \subset (0,T)$ with $\mathscr{L}^1\big((0,T)\setminus I_0\big) = 0$ such that

$$\liminf_{h\to\infty} \int_X \big|\overline{\boldsymbol{V}}_{\tau_h}(x,t)\big|^2 \, d\big(\overline{M}_{\tau_h}(t)\big)(x) < +\infty \quad \forall\, t \in I_0.$$

Since $-\overline{\boldsymbol{V}}_{\tau_h}(t,\cdot)$ is a strong subdifferential for every $h \in \mathbb{N}$, $t > 0$, the definition of limiting differential and the compactness Theorem 5.4.4 show that for any $t \in I_0$ μ_t belongs to the domain of the limiting subdifferential; since $\partial_\ell\phi(\mu_t)$ contains at most one vector, there exists a unique vector $-\tilde{\boldsymbol{v}}(t) \in \partial_\ell\phi(\mu(t))$ for any $t \in I_0$. We have to show that $\tilde{\boldsymbol{v}}(t) = \boldsymbol{v}(t)$ $\mathscr{L}^1$-a.e. in $(0,T)$.

The basic point here is that if $t \in I_0$, $\varepsilon > 0$, $\zeta \in \mathrm{Cyl}(X)$, and $\boldsymbol{e} \in X$, then

$$\begin{aligned}\liminf_{h\to\infty} \int_X &\Big(\zeta(x)\langle \boldsymbol{e}, \overline{\boldsymbol{V}}_{\tau_h}(x,t)\rangle + \varepsilon|\overline{\boldsymbol{V}}_{\tau_h}(x,t)|^2\Big)\, d\big(\overline{M}_{\tau_h}(t)\big)(x) \\ &\geq \int_X \zeta(x)\langle \boldsymbol{e}, \tilde{\boldsymbol{v}}(x,t)\rangle \, d\mu_t(x).\end{aligned} \tag{11.1.56}$$

For, if the left hand side is finite, by extracting a further subsequence we can assume thanks to Theorem 5.4.4 that $\overline{\boldsymbol{V}}_{\tau_h}(t,\cdot)$ is weakly converging in the sense of Definition 5.4.3 and its limit is $\tilde{\boldsymbol{v}}(t,\cdot)$, since this vector is the unique element of $\partial_\ell\phi(\mu_t)$.

Integrating (11.1.56) in time, against a test function $\eta \in C_0^\infty(0,T)$ with values in $[0,1]$, and choosing $\boldsymbol{e}$ among the vectors $\{\boldsymbol{e}_j\}_{j\in\mathbb{N}}$ of an orthonormal basis of X we have

$$\begin{aligned}\liminf_{n\to\infty} &\int_0^T \eta(t)\Big(\int_X \zeta(x)\langle \boldsymbol{e}_j, \overline{\boldsymbol{V}}_{\tau_h}\rangle\, d\big(\overline{M}_{\tau_h}(t)\big)(x)\Big)\, dt + 2\varepsilon\phi(\mu_0) \\ &\geq \liminf_{n\to\infty} \int_0^T \eta(t)\Big(\int_X \zeta(x)\langle \boldsymbol{e}_j, \overline{\boldsymbol{V}}_{\tau_h}\rangle + \varepsilon|\overline{\boldsymbol{V}}_{\tau_h}|^2 d\big(\overline{M}_{\tau_h}(t)\big)(x)\Big)\, dt \\ &\geq \int_0^T \eta(t)\Big(\liminf_{h\to\infty} \int_X \zeta(x)\langle \boldsymbol{e}_j, \overline{\boldsymbol{V}}_{\tau_h}\rangle + \varepsilon|\overline{\boldsymbol{V}}_{\tau_h}|^2 d\big(\overline{M}_{\tau_h}(t)\big)(x)\Big)\, dt \\ &\geq \int_0^T \eta(t)\Big(\int_X \zeta(x)\langle \boldsymbol{e}_j, \tilde{\boldsymbol{v}}\rangle\, d\mu_t(x)\Big)\, dt = \int_{X_T} \eta(t)\zeta(x)\langle \boldsymbol{e}_j, \tilde{\boldsymbol{v}}\rangle\, d\mu.\end{aligned}$$

On the other hand, the narrow convergence of $\overline{\boldsymbol{\gamma}}_{\tau_h}$ to $\boldsymbol{\gamma}$, Lemma 5.1.7 and the definition of $\boldsymbol{v}$ yield

$$\begin{aligned}&\lim_{h\to\infty} \int_{X_T} \eta(t)\zeta(x)\langle \boldsymbol{e}_j, \overline{\boldsymbol{V}}_{\tau_h}\rangle\, d\overline{M}_{\tau_h}(x,t) = \lim_{h\to\infty} \int_{X_T\times X} \eta(t)\zeta(x_1)\langle \boldsymbol{e}_j, x_2\rangle\, d\overline{\boldsymbol{\gamma}}_{\tau_h} \\ &= \int_{X_T\times X} \eta(t)\zeta(x_1)\langle \boldsymbol{e}_j, x_2\rangle\, d\overline{\boldsymbol{\gamma}} = \int_{X_T} \eta(t)\zeta(x)\langle \boldsymbol{e}_j, \boldsymbol{v}\rangle\, d\mu(x,t).\end{aligned}$$

Letting $\varepsilon \downarrow 0$ and changing ζ with $-\zeta$ we eventually get

$$\int_{X_T} \eta(t)\zeta(x)\langle \boldsymbol{e}_j, \boldsymbol{v}\rangle\, d\mu = \int_{X_T} \eta(t)\zeta(x)\langle \boldsymbol{e}_j, \tilde{\boldsymbol{v}}\rangle\, d\mu,$$

Since η, ζ and j are arbitrary this proves that $\boldsymbol{v} = \tilde{\boldsymbol{v}}$. □

11.2 Gradient flows in $\mathscr{P}_2(X)$ for λ-convex functionals along generalized geodesics

In this section we are considering the case of a

$$\begin{gathered}\text{proper, l.s.c. and coercive functional } \phi : \mathscr{P}_2(X) \to (-\infty, \infty], \\ \text{which is } \lambda\text{-}\textit{convex along generalized geodesics}, \end{gathered} \tag{11.2.1}$$

according to Definition 9.2.4; as usual, X is *a separable Hilbert space*. Thus we are assuming that for every choice of $\mu^1, \mu^2, \mu^3 \in D(\phi)$ there exists a 3-plan $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3)$ such that

$$\begin{cases} \pi^{1,2}_{\#}\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2), \quad \pi^{1,3}_{\#}\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^3), \\ \phi(\mu_t^{2\to 3}) \le (1-t)\phi(\mu^2) + t\phi(\mu^3) - \dfrac{\lambda}{2} t(1-t) \displaystyle\int_{X^3} |x_2 - x_3|^2 \, d\boldsymbol{\mu}, \end{cases} \tag{11.2.1a}$$

where $\mu_t^{2\to 3}$ is the interpolation between μ^2 and μ^3 induced by $\boldsymbol{\mu}$. By Lemma 2.4.8, for λ-convex functionals the coercivity assumption can equivalently formulated as

$$\exists\, r_* > 0 : \quad \inf\Big\{\phi(\mu) : \mu \in \mathscr{P}_2(X), \quad \int_X |x|^2 \, d\mu(x) \le r_*\Big\} > -\infty. \tag{11.2.1b}$$

We already observed in Lemma 9.2.7 that (11.2.1a,b) entails the main convexity assumption 4.0.1 of Chapter 4, whereas (4.0.1) corresponds to (11.2.1). By Theorem 4.1.2 the above conditions imply (11.1.13a,b), which are also the minimal assumptions we adopted to develop the subdifferential theory of Chapter 10.

The following theorem reproduces in the Wasserstein setting the metric results of Chapters 2 and 4.

Theorem 11.2.1 (Existence and main properties of gradient flows). *Let us suppose that* $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ *satisfies* (11.2.1) *and let* $\mu_0 \in \overline{D(\phi)}$.

Convergence. *The discrete solution* $\overline{M}_\tau$ *of* (11.1.24) *converges locally uniformly to a locally Lipschitz curve* $\mu := S[\mu_0]$ *in* $\mathscr{P}_2(X)$ *which is the unique gradient flow of* ϕ *with* $\mu(0+) = \mu_0$.

λ-contractive semigroup. *The map* $t \mapsto S[\mu_0](t)$ *is a* λ*-contracting semigroup on* $\overline{D(\phi)}$, *i.e.*

$$W_2(S[\mu_0](t), S[\nu_0](t)) \le e^{-\lambda t} W_2(\mu_0, \nu_0). \tag{11.2.2}$$

Regularizing effect. $\mu_t \in D(\boldsymbol{\partial}\phi) \subset D(\phi)$ for every $t > 0$ *and the map* $t \mapsto e^{\lambda t}|\partial\phi|(\mu_t)$ *is non increasing. For* $\lambda \geq 0$ *it satisfies the regularization estimates*

$$\begin{aligned} \phi(\mu_t) &\leq \frac{1}{2t} W_2^2(\mu_0, \nu) + \phi(\nu) \quad \forall \nu \in D(\phi), \\ |\partial\phi|^2(\mu_t) &\leq |\partial\phi|^2(\nu) + \frac{1}{t^2} W_2^2(\mu_0, \nu) \quad \forall \nu \in D(|\partial\phi|). \end{aligned} \tag{11.2.3}$$

Steepest descent and Evolution variational inequalities. $\mu = S[\mu_0]$ *is a curve of maximal slope, it satisfies the system of evolution variational inequalities* (11.1.20), *and the energy identity*

$$\int_a^b \int_X |v_t(x)|^2 \, d\mu_t(x) \, dt + \phi(\mu_b) = \phi(\mu_a) \quad \forall\, 0 \leq a < b < +\infty. \tag{11.2.4}$$

Asymptotic behaviour. *If* $\lambda > 0$ *then* ϕ *admits a unique minimum point* $\overline{\mu}$ *and*

$$\begin{aligned} W_2(\mu(t), \overline{\mu}) &\leq W_2(\mu(t_0), \overline{\mu}) e^{-\lambda(t-t_0)}, \\ \phi(\mu(t)) - \phi(\overline{\mu}) &\leq \Big(\phi(\mu(t_0)) - \phi(\overline{\mu})\Big) e^{-2\lambda(t-t_0)} \\ |\partial\phi|(\mu(t)) &\leq |\partial\phi|(\mu(t_0)) e^{-\lambda(t-t_0)}. \end{aligned} \tag{11.2.5}$$

If $\lambda = 0$ *and* $\overline{\mu}$ *is a minimum point of* ϕ *then we have*

$$\begin{gathered} |\partial\phi|(\mu(t)) \leq \frac{W_2(\mu_0, \overline{\mu})}{t}, \quad \phi(\mu(t)) - \phi(\overline{\mu}) \leq \frac{W_2^2(\mu_0, \overline{\mu})}{2t}, \\ \textit{the map} \quad t \mapsto W_2(\mu(t), \overline{\mu}) \quad \textit{is not increasing.} \end{gathered} \tag{11.2.6}$$

Right limits and precise pointwise formulation of the equation. *For every* $t, h > 0$ *and* $\hat{\boldsymbol{\mu}}_{t,h} \in \Gamma_o(\mu_t, \mu_{t+h})$ *the* right limit

$$\boldsymbol{\mu}_{t,0} := \lim_{h \downarrow 0} \Big(\pi^1, \frac{\pi^1 - \pi^2}{h}\Big)_{\#} \hat{\boldsymbol{\mu}}_{t,h} \quad \textit{exists in } \mathscr{P}_2(X \times X) \tag{11.2.7}$$

and satisfies

$$\boldsymbol{\mu}_{t,0} = \boldsymbol{\partial}^\circ \phi(\mu_t) \quad \forall t > 0, \tag{11.2.8}$$

$$\frac{d}{dt_+} \phi(\mu(t)) = -\int_{X^2} |x_2|^2 \, d\boldsymbol{\mu}_{t,0} = -|\partial\phi|^2(\mu(t)) = -|\mu'|^2(t) \quad \forall t > 0. \tag{11.2.9}$$

Moreover, (11.2.7), (11.2.8), *and* (11.2.9) *hold at* $t = 0$ *iff* $\mu_0 \in D(\boldsymbol{\partial}\phi) = D(|\partial\phi|)$.

Optimal error estimate. *If* $\lambda \geq 0$ *and* $\mu_0 \in D(\phi)$*, for every* $t = k\tau \in \mathcal{P}_\tau$ *we have*

$$W_2^2(\mu(t), \overline{M}_\tau(t)) \leq \tau\big(\phi(\mu_0) - \phi_\tau(\mu_0)\big) \leq \frac{\tau^2}{2} |\partial\phi|^2(\mu_0). \tag{11.2.10}$$

Stability. *Suppose that* ϕ_h, ϕ *are* λ*-convex functionals along generalized geodesics and satisfy the assumption of Lemma!10.3.16, and let* μ_h, $\mu : (0,+\infty) \to \mathscr{P}_2(X)$ *be the corresponding gradient flows satisfying the initial Cauchy conditions* $\lim_{t\downarrow 0} \mu_h(t) = \mu_{h,0}$, $\lim_{t\downarrow 0} \mu(t) = \mu_0$ *in* $\mathscr{P}_2(X)$*. If*

$$\mu_{h,0} \to \mu_0 \quad \textit{in } \mathscr{P}_2(X) \textit{ as } h \to \infty, \quad \sup_{h\in\mathbb{N}} \phi_h(\mu_{h,0}) < +\infty \tag{11.2.11}$$

then $\mu_h(t)$ *converge to* $\mu(t)$ *in* $\mathscr{P}_2(X)$*, locally uniformly in* $[0,+\infty)$*.*

Proof. We already observed that we can apply Theorem 4.0.4; the *convergence* of the variational scheme, the λ*-contraction property* of the induced semigroup, the *regularizing estimates* (11.1.47), the formulation by evolution variational inequalities (11.1.20), and the optimal error estimates (11.2.10) follow directly from that statement.

Theorem 11.1.4 shows that the limit curve μ satisfies the gradient flow equation (11.1.3) and it is therefore a curve of maximal slope, by Theorem 11.1.3.

The *energy identity* (11.2.4) is then a direct consequence of the metric Theorem 2.3.3 or of the Chain Rule 10.3.18.

Theorem 2.4.15 shows that the map $t \mapsto e^{\lambda t}|\partial\phi|(\mu(t))$ is not increasing; this proves the third formula of (11.2.5). The first one is a simple consequence of (11.2.2), since a minimum point provides a constant solution to the gradient flow equation. The second formula in (11.2.5) follows from Theorem 2.4.14, whereas (11.2.6) corresponds to Corollary 4.0.6.

Let us consider now the *right limit properties* (11.2.7), (11.2.8), and (11.2.9). We already know that $\boldsymbol{\partial}\phi(\mu(t))$ is not empty for $t > 0$: we set $\boldsymbol{\gamma}_t = \boldsymbol{\partial}^\circ\phi(\mu(t))$; Theorem 2.4.15 and Theorem 10.3.11 yield

$$-\frac{d}{dt_+}\phi(\mu(t)) = \int_{X^2} |x_2|^2 \, d\boldsymbol{\gamma}_t = \lim_{h\downarrow 0} \frac{1}{h^2} \int_{X^2} |x_2 - x_1|^2 \, d\hat{\boldsymbol{\mu}}_{t,h}. \tag{11.2.12}$$

As in Proposition 10.3.18 we consider 3-plans $\boldsymbol{\gamma}_{t,h}$ such that

$$(\pi^{1,2})_\#\hat{\boldsymbol{\gamma}}_{t,h} = \boldsymbol{\gamma}_t, \quad (\pi^{1,3})_\#\hat{\boldsymbol{\gamma}}_{t,h} = \hat{\boldsymbol{\mu}}_{t,h},$$

and we define $\boldsymbol{\gamma}_{t,h} := \big(\pi^1, \pi^2, h^{-1}(\pi^1 - \pi^3)\big)_\#\hat{\boldsymbol{\gamma}}_{t,h}$; arguing as in (10.3.44) we get

$$\frac{d}{dt_+}\phi(\mu(t)) \ge \limsup_{h\downarrow 0} -\int_{X^3} \langle x_2, x_3\rangle \, d\boldsymbol{\gamma}_{t,h},$$

while (11.2.12) gives

$$-\frac{d}{dt_+}\phi(\mu(t)) = \int_{X^3} |x_2|^2 \, d\boldsymbol{\gamma}_{t,h} = \lim_{h\downarrow 0} \int_{X^3} |x_3|^2 \, d\boldsymbol{\gamma}_{t,h}.$$

Combining these inequalities we get

$$\limsup_{h\downarrow 0}\int_{X^3}|x_2-x_3|^2\,d\boldsymbol{\gamma}_{t,h}$$
$$\le \limsup_{h\downarrow 0}\Big(\int_{X^3}|x_2|^2\,d\boldsymbol{\gamma}_{t,h}-2\int_{X^3}\langle x_2,x_3\rangle\,d\boldsymbol{\gamma}_{t,h}+\int_{X^3}|x_2|^2\,d\boldsymbol{\gamma}_{t,h}\Big)=0,$$

which shows that

$$\pi^{1,3}_{\#}\boldsymbol{\gamma}_{t,h}=\big(\pi^1,h^{-1}(\pi^1-\pi^2)\big)_{\#}\hat{\boldsymbol{\mu}}_{t,h}\to\boldsymbol{\gamma}_t \quad \text{in } \mathscr{P}_2(X\times X) \quad \text{as } h\downarrow 0. \tag{11.2.13}$$

(11.2.9) then follows from (11.2.12).

Finally, we are proving the last stability property in a fixed time interval $[0,T]$. If $\overline{M}_{h,\tau}$ is the piecewise constant discrete solution of the gradient flow of ϕ_h associated to a fixed time step $\tau>0$, Lemma 10.3.17, the uniqueness of the minimizers M^n_τ given by Lemma 4.1.1, and a simple induction argument show that

$$\overline{M}_{h,\tau}(t)\to\overline{M}_\tau(t) \quad \text{in } \mathscr{P}_2(X) \quad \text{uniformly in } [0,T],\ \forall\tau\in(0,1/\lambda^-) \tag{11.2.14}$$

for the discrete solution $M_\tau(t)$ relative to ϕ. On the other hand, the optimal *a priori error estimates* and the bound on the initial energy show that

$$\sup_{t\in[0,T]}W_2(\overline{M}_{h,\tau}(t),\mu_h(t))\le C\sqrt{\tau}, \qquad \sup_{t\in[0,T]}W_2(\overline{M}_\tau(t),\mu(t))\le C\sqrt{\tau}$$

for a constant C independent of τ. The triangle inequality then proves the uniform convergence of μ_h to μ in $[0,T]$. □

11.2.1 Applications to Evolution PDE's

Here we illustrate some Evolution PDE's arising from the examples of λ-convex functionals given in Chapter 9, whose (minimal) subdifferential has been computed in Chapter 10.

Example 11.2.2 (The linear transport equation for λ-convex potentials). Let $V: X\to(-\infty,+\infty]$ be a proper, l.s.c. and λ-convex potential. We are looking for curves $t\mapsto\mu_t\in\mathscr{P}_2(X)$ which solve the evolution equation

$$\frac{\partial}{\partial t}\mu_t+\nabla\cdot(\mu_t\boldsymbol{v}_t)=0, \quad \text{with } -\boldsymbol{v}_t(x)\in\partial V(x) \text{ for } \mu_t\text{-a.e. } x\in X, \tag{11.2.15}$$

which is the gradient flow in $\mathscr{P}_2(X)$ of the potential energy functional discussed in Example 9.3.1:

$$\mathcal{V}(\mu):=\int_X V(x)\,d\mu(x). \tag{11.2.16}$$

If V is differentiable, (11.2.15) can also be written as

$$\frac{\partial}{\partial t}\mu_t(x) = \nabla \cdot (\mu_t(x)\nabla V(x)) \quad \text{in the distribution sense.} \tag{11.2.17}$$

In the statement of the following theorem we denote by T the λ-contractive semigroup on $\overline{D(V)} \subset X$ induced by the differential inclusion

$$\frac{d}{dt}T_t(x) \in -\partial V(T_t(x)), \quad T_0(x) = x \quad \forall\, x \in \overline{D(V)}. \tag{11.2.18}$$

Recall also that, according to Brezis theorem, $\frac{d}{dt}T_t(x)$ equals $-\partial^\circ V(T_t(x))$ for $\mathscr{L}^1$-a.e. $t > 0$.

Theorem 11.2.3. *For every $\mu_0 \in \mathscr{P}_2(X)$ with* $\operatorname{supp}\mu_0 \subset \overline{D(V)}$ *there exists a unique solution $(\mu_t, \boldsymbol{v})$ of* (11.2.15) *satisfying*

$$\lim_{t\downarrow 0}\mu_t = \mu_0, \qquad \int_X |\boldsymbol{v}_t(x)|^2\, d\mu_t(x) \in L^1_{\mathrm{loc}}(0,+\infty), \tag{11.2.19}$$

and this solution satisfies all the properties stated in Theorem 11.2.1*. In particular, for every $t > 0$ we have the representation formulas $\mu_t = (T_t)_\#\mu_0$ and*

$$\boldsymbol{v}_t(x) = -\partial^\circ V(x) \quad \textit{for } \mu_t\textit{-a.e. } x \in X. \tag{11.2.20}$$

Proof. Proposition 9.3.2 shows that the functional $\mathcal{V}$ satisfies (11.2.1). In order to show that $\mu_0 \in \overline{D(\mathcal{V})}$ we observe that, being $\operatorname{supp}\mu_0 \subset \overline{D(V)}$, we can find a sequence $(\nu_h) \subset D(\mathcal{V})$ of convex combination of Dirac masses

$$\nu_n := \sum_{k=1}^{K_n} \alpha_{n,k}\delta_{x_{n,k}}, \quad \alpha_{n,k} \ge 0, \quad \sum_{k=1}^{K_n}\alpha_{n,k} = 1, \quad x_{n,k} \in D(V), \tag{11.2.21}$$

such that $\nu_n \to \mu_0$ in $\mathscr{P}_2(X)$.

Therefore, we can apply Theorem 11.2.1 and the subdifferential characterization in Proposition 10.4.2 to get (11.2.15) and more precisely (11.2.20).

It is then immediate to check directly that if we choose $\mu_0 = \nu_n$ then

$$\nu_{n,t} := \sum_{k=1}^{K_n} \alpha_{n,k}\delta_{T_t(x_{n,k})} = (T_t)_\#\nu_n \tag{11.2.22}$$

solves (11.2.15) (see also Section 8.1, where the connection between characteristics and solutions of the continuity equation is studied in detail), whereas (11.2.19) follows by the energy identity

$$\int_a^b |\partial^\circ V(T_t(x))|^2\, dt + \phi(T_b(x)) = \phi(T_a(x)) \quad \forall\, x \in \overline{D(V)}.$$

We thus have $\mu_t = \nu_{n,t} = (T_t)_\#\mu_0$ for every initial datum which is a convex combination of Dirac masses in $D(V)$. A standard approximation argument via (11.2.2) yields the representation formula $\mu_t = (T_t)_\#\mu_0$ for every admissible initial measure μ_0. □

Example 11.2.4 (Nonlinear diffusion equations). Let us consider a convex differentiable function $F : [0,+\infty) \to \mathbb{R}$ which satisfies (10.4.15) with $p = 2$, (10.4.21) and (10.4.23): F is the density of the *internal energy functional* $\mathcal{F}$ defined in (10.4.14).

Setting $L_F(z) := zF'(z) - F(z)$, we are looking for *nonnegative* solution of the evolution equation

$$\frac{\partial}{\partial t}\rho_t - \Delta(L_F(\rho_t)) = 0 \quad \text{in } \mathbb{R}^d \times (0,+\infty), \tag{11.2.23}$$

satisfying the (normalized) mass conservation

$$\rho_t \in L^1(\mathbb{R}^d), \qquad \int_{\mathbb{R}^d} \rho_t(x)\,dx = 1 \quad \forall\, t > 0, \tag{11.2.24}$$

the finiteness of the quadratic moment

$$\sup_{t\in(0,T)} \int_{\mathbb{R}^d} |x|^2 \rho_t(x)\,dx < +\infty \quad \forall\, T > 0, \tag{11.2.25}$$

the integrability condition $L_F(\rho) \in L^1_{\rm loc}(\mathbb{R}^d \times (0,+\infty))$, and the initial Cauchy condition

$$\lim_{t\downarrow 0} \rho_t \cdot \mathscr{L}^d = \mu_0 \text{ in } \mathscr{P}_2(\mathbb{R}^d). \tag{11.2.26}$$

Therefore (11.2.23) has the usual distributional meaning

$$\int_0^{+\infty} \int_{\mathbb{R}^d} \Big(- \rho_t(x) \frac{\partial}{\partial t}\zeta(x,t) + L_F(\rho_t(x))\Delta\zeta(x,t)\Big)\,dx\,dt = 0$$

for any $\zeta \in \mathscr{D}(\mathbb{R}^d \times (0,+\infty))$.

Theorem 11.2.5. *Suppose that either F has a superlinear growth or F satisfies* (10.4.34)*. Then for every $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a unique solution*

$$\rho \in AC^2_{\rm loc}((0,+\infty); \mathscr{P}_2(\mathbb{R}^d))$$

of the above equation among those satisfying

$$L_F(\rho) \in L^1_{\rm loc}((0,+\infty); W^{1,1}_{\rm loc}(\mathbb{R}^d)), \qquad \int_{\mathbb{R}^d} \frac{|\nabla L_F(\rho)|^2}{\rho}\,dx \in L^1_{\rm loc}(0,+\infty). \tag{11.2.27}$$

It is the unique gradient flow in $\mathscr{P}_2(\mathbb{R}^d)$ of the (relaxed) functional $\mathcal{F}^*$ defined in (10.4.17)*, which is convex along generalized geodesics. In particular it satisfies all the properties of Theorem* 11.2.1 *for $\lambda = 0$.*

Proof. The proof is a simple combination of Theorem 11.2.1 and of the results of Section 10.4.3 for the functional $\mathcal{F}^*$, noticing that the domain of $\mathcal{F}$ is dense in $\mathscr{P}_2(\mathbb{R}^d)$.

Observe that, even if F has a sublinear growth and μ_0 is not regular (e.g. a Dirac mass), the regularizing effect of the Wasserstein semigroup and Theorem 10.4.8 show that, because of (10.4.34), $\mu_t := S[\mu_0](t)$ is absolutely continuous w.r.t. the Lebesgue measure $\mathscr{L}^d$ for all $t > 0$: its density ρ_t w.r.t. $\mathscr{L}^d$ is therefore well defined and provides a solution of (11.2.23) in the above precised meaning. □

Remark 11.2.6. Equation (11.2.23) is a very classical problem: it has been studied by many authors from different points of view, which is impossibile to recall in detail here.
We only mention that in the case of homogeneous Dirichlet boundary conditions in a bounded domain, H. Brézis showed that the equation is the gradient flow (see [28]) of the convex functional (since L_F is monotone)

$$\psi(\rho) := \int_{\mathbb{R}^d} G_F(\rho)\,dx, \quad \text{where} \quad G_F(\rho) := \int_0^\rho L_F(r)\,dr,$$

in the space $H^{-1}(\Omega)$. We refer to the paper of Otto [107] for a detailed comparison of the two notions of solutions and for a physical justification of the interest of the Wasserstein approach. Notice that here we allow for more general initial data (an arbitrary probability measure), whereas in the H^{-1} formulation Dirac masses are not allowed (but see [110, 40]).

Example 11.2.7 (Drift diffusion equations with non local terms). Let us consider, as in [37, 38], a functional ϕ which is the sum of internal, potential and interaction energy:

$$\phi(\mu) := \int_{\mathbb{R}^d} F(\rho)\,dx + \int_{\mathbb{R}^d} V\,d\mu + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W\,d\mu\times\mu \quad \text{if } \mu = \rho\mathscr{L}^d.$$

Here F, V, W satisfy the assumptions considered in Section 10.4.7; as usual we set $\phi(\mu) = +\infty$ if $\mu \in \mathscr{P}_2(\mathbb{R}^d) \setminus \mathscr{P}_2^r(\mathbb{R}^d)$. The gradient flow of ϕ in $\mathscr{P}_2(\mathbb{R}^d)$ leads to the equation

$$\partial_t\rho_t - \nabla\cdot\Big(\nabla L_F(\rho_t) + \rho_t\nabla V + \rho_t(\nabla W)\star\rho_t\Big) = 0, \tag{11.2.28}$$

coupled with conditions (11.2.24), (11.2.25), (11.2.26).

Theorem 11.2.8. *For every $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a unique distributional solution $\mu_t = \rho_t\mathscr{L}^d$ of* (11.2.28) *among those satisfying $\rho_t\mathscr{L}^d \to \mu_0$ as $t \downarrow 0$, $L_F(\rho_t) \in L^1_{\rm loc}((0,+\infty); W^{1,1}_{\rm loc}(\mathbb{R}^d))$, and*

$$\left\| \frac{\nabla L_F(\rho_t)}{\rho_t} + \nabla V + \big(\nabla W \star \rho_t\big) \right\|_{L^2(\mu_t;\mathbb{R}^d)} \in L^2_{\rm loc}(0,+\infty). \tag{11.2.29}$$

Furthermore, this solution is the gradient flow in $\mathscr{P}_2(\mathbb{R}^d)$ of the functional ϕ and therefore satisfies all the properties stated in Theorem 11.2.1.

Proof. The existence of ρ_t follows by Theorem 11.2.1 and by the characterization, given in Section 10.4.7, of the (minimal) subdifferential of ϕ. The same characterization proves that any ρ_t as in the statement of the theorem is a gradient flow; therefore the uniqueness Theorem 11.1.4 can be applied. □

When F, $V = 0$ we find a model equation for the evolution of granular flows (see e.g. [34]); when $W = 0$ and F is the entropy functional, we find the Fokker–Planck equation with an arbitrary λ-convex potential: it is interesting to compare our result with [26, 48]. Notice that we can also consider evolution equations in convex (bounded or unbounded) domains $\Omega \subset \mathbb{R}^d$ with homogeneous Neumann boundary conditions, simply by setting $V(x) \equiv +\infty$ for $x \in \mathbb{R}^d \setminus \overline{\Omega}$.

Example 11.2.9 (Gradient flow of $-W^2/2$ and geodesics). For a fixed reference measure $\sigma \in \mathscr{P}_2(X)$, X being a separable Hilbert space, let us now consider the functional $\phi(\mu) := -\frac{1}{2}W_2^2(\mu, \sigma)$, as in Theorem 10.4.12. Being ϕ (-1)-convex along generalized geodesics, we can apply Theorem 11.2.1 to show that ϕ generates an evolution semigroup on $\mathscr{P}_2(X)$.

When $\Gamma_o(\sigma, \mu_0)$ contains a plan $\boldsymbol{\gamma}$ such that

$$\big(\pi^1, \pi^1 + T(\pi^2 - \pi^1)\big)_\# \boldsymbol{\gamma} \quad \text{is } \textit{optimal} \text{ for some } T > 1, \tag{11.2.30}$$

then the semigroup moves μ_0 along the geodesics induced by $\boldsymbol{\gamma}$. Lemma 7.2.1 shows that in this case $\boldsymbol{\gamma}$ admits the representation $\boldsymbol{\gamma} = (\boldsymbol{r} \times \boldsymbol{i})_\# \mu_0$ for some transport map $\boldsymbol{r}$ and $\boldsymbol{\gamma}$ is the unique element of $\Gamma_o(\sigma, \mu_0)$.

Theorem 11.2.10. *Let be given two measures σ, $\mu_0 \in \mathscr{P}_2(X)$ and suppose that $\boldsymbol{\gamma} \in \Gamma_o(\sigma, \mu_0)$ satisfies* (11.2.30)*, i.e. the constant speed geodesic*

$$\gamma(s) := \big((1-s)\pi^1 + s\pi^2\big)_\# \boldsymbol{\gamma}$$

can be extended to an interval $[0, T]$, with $T > 1$. Then the formula

$$t \to \mu(t) := \gamma(e^t), \qquad \textit{for } 0 \le t \le \log(T), \tag{11.2.31}$$

gives the gradient flow of $\mu \mapsto -\frac{1}{2}W_2^2(\mu, \sigma)$ starting from μ_0.

Proof. Lemma 7.2.1 shows that $\big((1-e^t)\pi^1 + e^t\pi^2, \pi^1\big)_\# \boldsymbol{\gamma}$ is the unique optimal plan in $\Gamma_o(\mu(t), \sigma)$; therefore

$$W_2^2(\mu(t), \mu(\bar t)) = |e^t - e^{\bar t}|^2 W_2^2(\mu_0, \sigma), \quad W_2^2(\mu(t), \sigma) = e^{2t} W_2^2(\mu_0, \sigma), \tag{11.2.32}$$

so that

$$\frac{d}{dt}\phi(\mu(t)) = -e^{2t} W_2^2(\mu_0, \sigma) = -|\mu'|^2(t). \tag{11.2.33}$$

On the other hand, the characterization of $|\partial\phi|$ given in (10.4.52) gives

$$|\partial\phi|^2(\mu(t)) = e^{2t} W_2^2(\mu_0, \sigma).$$

This shows that $\mu(t)$ is a curve of maximal slope; combining Theorem 11.1.3 with the uniqueness Theorem 11.1.4, we conclude. □

Example 11.2.11 (Fokker–Planck equation in infinite dimension). Let X be a separable Hilbert space and let γ be a reference probability measure which satisfies the log-concavity assumption (9.4.19). The relative entropy functional defined as in (9.4.1)

$$\phi(\mu) := \mathcal{H}(\mu|\gamma) \tag{11.2.34}$$

is then convex along generalized geodesics, according to Theorem 9.4.11; since it is nonnegative and l.s.c. in $\mathscr{P}_2(X)$, its gradient flow generates a contraction semigroup on $\overline{D(\phi)}$, which satisfies all the properties stated in Theorem 11.2.1.

Is is not difficult to check that

$$\overline{D(\phi)} = \Big\{\mu \in \mathscr{P}_2(X) : \operatorname{supp}\mu \subset \operatorname{supp}\gamma\Big\}. \tag{11.2.35}$$

For, $D(\phi)$ contains all the measures of the type

$$\mu_{x_0,\rho} := \frac{1}{\gamma(B_\rho(x_0))}\chi_{B_\rho(x_0)} \cdot \gamma \quad \text{with} \quad x_0 \in \operatorname{supp}\gamma,\ \rho > 0,$$

and their convex combinations, so that

$$\sum_i \alpha_i \delta_{x_i} \in \overline{D(\phi)} \quad \text{if } x_i \in \operatorname{supp}\gamma, \quad \alpha_i \geq 0, \quad \sum_i \alpha_i = 1.$$

Then, Remark 5.1.2 shows (11.2.35).

Let us now consider the particular case of a Gaussian measure γ induced by a bounded, positive definite, symmetric operator Q of trace class, which was considered in Section 10.4.4. Keeping the same notation, from the characterization given in that section of the minimal subdifferential of the relative entropy in terms of the Fisher information functional we obtain the following result (we refer to [49] and to the references therein for a more detailed analysis of this kind of equations).

Theorem 11.2.12. *For every $\mu_0 \in \mathscr{P}_2(X)$ there exists a unique solution $\mu_t = \rho_t \cdot \gamma$, $t > 0$, of the equation*

$$\partial_t \mu_t - \nabla \cdot \big(\gamma \nabla \rho_t\big) = 0, \quad \lim_{t\downarrow 0} \mu_t = \mu_0, \tag{11.2.36}$$

in the distributional sense according to (8.3.8)*, among those satisfying the local integrability of the Fisher information*

$$\int_X \frac{|\nabla \rho_t|^2}{\rho_t}\, d\gamma \in L^1_{\rm loc}(0, +\infty). \tag{11.2.37}$$

Here

$$\nabla \rho_t = \rho_t \big(\frac{\nabla \rho_t}{\rho_t}\big) = \rho_t \partial^\circ \phi(\mu)$$

is defined in terms of the "logarithmic gradient" of μ according to Definition 10.4.16*. μ_t is the gradient flow of the Relative Entropy functional* (11.2.34) *and satisfies all the properties stated in Theorem* 11.2.1*.*

11.3 Gradient flows in $\mathscr{P}_p(X)$ for regular functionals

In this section we are considering the case of a

$$\text{proper, coercive (2.1.2b), and l.s.c. functional } \phi : \mathscr{P}_p(X) \to (-\infty, +\infty], \tag{11.3.1}$$

which will satisfy a suitable regularity assumption like in Definition 10.3.9.

Our main examples will still concern λ-geodesically convex functionals: the main difference with respect to the previous section is thus provided by the Wasserstein distance in $\mathscr{P}_p(X)$, $p \neq 2$, which does not exhibit a sufficiently nice behaviour along generalized geodesics. Thus, even if the functionals would be convex along these more general interpolating curves, the abstract machinery of Chapter 4 could not be applied.

Here we refer instead to the metric theory developed in Chapter 2; that more general convergence proof uses the identity between curves of maximal slopes and gradient flows, we established in Theorem 11.1.3. The crucial assumptions of that approach result from a combination of the *lower semicontinuity of the metric slope* $|\partial\phi|(\cdot)$ and the local compactness of the sublevels of ϕ. It follows that the choice of the right topology becomes crucial.

For $m \in (0, +\infty)$ let us denote by Σ_m the sets

$$\Sigma_m := \Big\{\mu \in \mathscr{P}_p(X) : \phi(\mu) \leq m, \quad \int_X |x|^p \, d\mu(x) \leq m\Big\}. \tag{11.3.2}$$

The sets Σ_m are bounded in $\mathscr{P}_p(X)$ and therefore relatively compact in $\mathscr{P}(X_\varpi)$, by Lemma 5.1.12(e): we are assuming that

Assumption 11.3.1 (Weak lower semicontinuity). *ϕ and $|\partial\phi|$ are lower semicontinuous on Σ_m w.r.t. the narrow convergence of $\mathscr{P}(X_\varpi)$. Moreover, if $\mu_n, \mu \in \Sigma_m$, $\mu_n \to \mu$ in $\mathscr{P}(X_\varpi)$ and $\sup_n |\partial\phi|(\mu_n) < +\infty$, then $\phi(\mu_n) \to \phi(\mu)$.*

Observe that Assumption 11.3.1 is surely satisfied if

$$\begin{gathered}\phi \text{ is a } \textit{regular functional} \text{ according to Definition 10.3.9} \\ \text{and } \Sigma_m \text{ are compact in } \mathscr{P}_p(X).\end{gathered} \tag{11.3.3}$$

In particular, the assumption is satisfied if

$$\phi \text{ is a } \lambda\text{-convex functional and } \Sigma_m \text{ are compact in } \mathscr{P}_p(X), \tag{11.3.4}$$

due to the fact that for λ-convex functionals $|\partial\phi|$ is always lower semicontinuous w.r.t. the narrow convergence of $\mathscr{P}(X)$.

When ϕ is λ-convex but Σ_m are not compact, then one has to check directly on the particular form of $|\partial\phi|$ the lower semicontinuity property with respect to the narrow convergence in $\mathscr{P}(X_\varpi)$.

Theorem 11.3.2 (Existence of gradient flows). *Let $\phi : \mathscr{P}_p(X) \to (-\infty, +\infty]$ be a proper and l.s.c. functional satisfying Assumption 11.3.1. For every initial datum $\mu_0 \in D(\phi)$ each sequence of discrete solutions $\bar{M}_{\tau_k}$ of the variational scheme admits a (not relabeled) subsequence such that:*

(i) *$\overline{M}_{\tau_k}(t)$ narrowly converges in $\mathscr{P}(X_\varpi)$ to μ_t locally uniformly in $[0, +\infty)$, with $\mu \in AC^p_{\rm loc}([0, +\infty); \mathscr{P}_p(X))$.*

(ii) *μ is a solution of the gradient flow equation*

$$j_p(v_t) = -\partial^\circ \phi(\mu_t), \quad \|v_t\|_{L^p(\mu_t;X)} = |\mu'|(t), \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0, \tag{11.3.5}$$

with $\mu_t \to \mu_0$ as $t \downarrow 0$, where v_t is the tangent vector to the curve μ_t.

(iii) *The energy inequality*

$$\int_a^b \int_X |v_t(x)|^p \, d\mu_t(x) \, dt + \phi(\mu_b) \le \phi(\mu_a), \tag{11.3.6}$$

holds for every $b \in [0, +\infty)$ and $a \in [0, b) \setminus \mathscr{N}$, $\mathscr{N}$ being a $\mathscr{L}^1$-negligible subset of $(0, +\infty)$.

Moreover, **if ϕ is also λ-convex along geodesics**, *then we have:*

Energy identity.

$$\int_a^b \int_X |v_t(x)|^p \, d\mu_t(x) \, dt + \phi(\mu_b) = \phi(\mu_a) \quad \forall \, 0 \le a < b < +\infty. \tag{11.3.7}$$

Regularizing effect. *μ_t is locally Lipschitz in $(0, +\infty)$ (in $[0, +\infty)$ if $\mu_0 \in D(|\partial\phi|)$), $\mu_t \in D(\boldsymbol{\partial}\phi)$ for every $t > 0$, $t \mapsto e^{\lambda t}|\partial\phi|(\mu_t)$ is right continuous, and satisfies the bounds (2.4.27), (2.4.28).*

Right limits and pointwise formulation of the equation. *For every t, $h > 0$ and $\hat{\boldsymbol{\mu}}_{t,h} \in \Gamma_o(\mu_t, \mu_{t+h})$ the* right limit

$$\boldsymbol{\mu}_{t,0} := \lim_{h \downarrow 0} \Big(\pi^1, j_p\big(\frac{\pi^1 - \pi^2}{h}\big)\Big)_{\#} \hat{\boldsymbol{\mu}}_{t,h} \quad \textit{exists in } \mathscr{P}_{pq}(X \times X) \tag{11.3.8}$$

and it satisfies

$$\boldsymbol{\mu}_{t,0} = \boldsymbol{\partial}^\circ \phi(\mu_t) \tag{11.3.9}$$

and

$$\frac{d}{dt_+}\phi(\mu_t) = -\int_{X^2} |x_2|^q \, d\boldsymbol{\mu}_{t,0} = -|\partial\phi|^q(\mu_t) = -|\mu'|^p(t) \tag{11.3.10}$$

for any $t > 0$. Finally, (11.3.8), (11.3.9), and (11.3.10) hold at $t = 0$ iff $\mu_0 \in D(\boldsymbol{\partial}\phi) = D(|\partial\phi|)$.

Proof. The first part of the statement is a simple transposition of Theorem 2.3.1: σ is the topology of narrow convergence in $\mathscr{P}(X_\varpi)$.

The second part follows from Theorems 2.3.3, 2.4.12, and 2.4.15; the finer pointwise properties can be proved by the same argument of Theorem 11.2.1. □

Example 11.3.3. Let us consider a functional ϕ which is the sum of internal, potential and interaction energy

$$\phi(\mu) := \int_{\mathbb{R}^d} F(\rho)\,dx + \int_{\mathbb{R}^d} V\,d\mu + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W\,d\mu\times\mu \quad \text{if } \mu = \rho\mathscr{L}^d,$$

where F, V, W satisfy the assumptions considered in Section 10.4.7; as usual we set $\phi(\mu) = +\infty$ if $\mu \in \mathscr{P}_p(\mathbb{R}^d) \setminus \mathscr{P}_p^r(\mathbb{R}^d)$. The gradient flow of ϕ in $\mathscr{P}_p(\mathbb{R}^d)$ leads to the equation

$$\partial_t\rho - \nabla\cdot\Big(\rho j_q\Big(\frac{\nabla L_F(\rho)}{\rho} + \nabla V + (\nabla W)\star\rho\Big)\Big) = 0, \tag{11.3.11}$$

coupled with conditions analogous to (11.2.24), (11.2.25), (11.2.26) for arbitrary $p \in (1, +\infty)$. Arguing as in the proof of Theorem 11.2.8, but replacing 2 by a general exponent $p \in (1, +\infty)$, we obtain the following existence result.

Theorem 11.3.4. *For every $\mu_0 \in D(\phi) \subset \mathscr{P}_p(\mathbb{R}^d)$ there exists a distributional solution $\mu_t = \rho_t\mathscr{L}^d$ of* (11.3.11) *with $L_F(\rho_t) \in L^1_{\rm loc}((0,+\infty); W^{1,1}_{\rm loc}(\mathbb{R}^d))$, $\rho_t\mathscr{L}^d \to \mu_0$ as $t \downarrow 0$ and*

$$\Big\|\frac{\nabla L_F(\rho_t)}{\rho_t} + \nabla V + (\nabla W)\star\rho_t\Big\|_{L^q(\mu_t;X)} \in L^\infty_{\rm loc}(0,+\infty). \tag{11.3.12}$$

Moreover, $t \mapsto \phi(\mu_t)$ satisfies the energy identity and all the other properties stated in Theorem 11.3.2.

Chapter 12

Appendix

12.1 Carathéodory and normal integrands

In this section we recall some standard facts about integrands depending on two variables, measurable w.r.t. the first one, and more regular w.r.t. the second one.

Definition 12.1.1 (Carathéodory and normal integrands). *Let X_1, X_2 be Polish spaces, let $\mu \in \mathscr{P}(X_1)$ and let $\mathscr{L}$ be the Σ-algebra of μ-measurable subsets of X_1. We say that a $\mathscr{L} \times \mathscr{B}(X_2)$-measurable function $f : X_1 \times X_2 \to \mathbb{R}$ is a* Carathéodory *integrand if $x_2 \mapsto f(x_1, x_2)$ is continuous for μ-a.e. $x_1 \in X_1$.*
We say that a $\mathscr{L} \times \mathscr{B}(X_2)$-measurable function $f : X_1 \times X_2 \to [0, +\infty]$ is a normal integrand if $x_2 \mapsto f(x_1, x_2)$ is lower semicontinuous for μ-a.e. $x_1 \in X_1$.

In order to check that a given function f is a Carathéodory integrand the following remark will often be useful.

Remark 12.1.2. Suppose that a function $f : X_1 \times X_2 \to \mathbb{R}$ satisfies

$$\begin{aligned} x_2 \mapsto f(x_1, x_2) \quad & \text{is continuous for } \mu\text{-a.e. } x_1 \in X_1, \\ x_1 \mapsto f(x_1, x_2) \quad & \text{is } \mathscr{L}\text{-measurable for each } x_2 \in X_2. \end{aligned} \tag{12.1.1}$$

Then f is a Carathéodory integrand. Indeed we can approximate f by the $\mathscr{L} \times \mathscr{B}(X_2)$-measurable functions

$$f_\varepsilon(x_1, x_2) := \sum_i f_\varepsilon(x_1, y_i) \chi_{V_i^\varepsilon}(x_2),$$

where $\{V_i^\varepsilon\}$ is a partition of X_2 into (at most) countably many Borel sets with diameter less than ε and $y_i \in V_i^\varepsilon$. By the first condition in (12.1.1) the functions f_ε pointwise converge to u out of a set $N \times X_2$ with $\mu(N) = 0$. Therefore f is $\mathscr{L} \times \mathscr{B}(X_2)$-measurable.

For the proof of the following theorem, we refer to [23, Thm. 1, Cor. 1, Thm. 2($(d) \Rightarrow (a)$)].

Theorem 12.1.3 (Scorza–Dragoni). *Let X_1, X_2 be Polish spaces and let $\mu \in \mathscr{P}(X_1)$; if f is defined in $X_1 \times X_2$ with values in $\mathbb{R}$ (resp. in $[0, +\infty]$) is a Carathéodory (resp. normal) integrand, then for every $\varepsilon > 0$ there exists a continuous (resp. l.s.c. and bounded above by f) function f_ε such that*

$$\mu\left(\{x_1 \in X_1 : \ f(x_1, x_2) \neq f_\varepsilon(x_1, x_2) \quad \text{for some } x_2 \in X_2\}\right) \leq \varepsilon. \tag{12.1.2}$$

12.2 Weak convergence of plans and disintegrations

In this section we examine more closely the relation between narrow convergence and disintegration for families of plans $\boldsymbol{\gamma}^n \in \mathscr{P}(X_1 \times X_2)$ whose first marginal is independent of n.

In the sequel we assume that X_1 and X_2 are Polish spaces, and $\mu_1 \in \mathscr{P}(X_1)$. We start by stating natural continuity and lower semicontinuity properties with respect to narrow convergence of Carathéodory and normal integrands.

Theorem 12.2.1. *Let $\boldsymbol{\gamma}^n \in \mathscr{P}(X_1 \times X_2)$ narrowly converging to $\boldsymbol{\gamma}$ and such that $\pi^1_\# \boldsymbol{\gamma}^n = \mu_1$. Then for every normal integrand f we have*

$$\liminf_{n \to \infty} \int_{X_1 \times X_2} f(x_1, x_2) \, d\boldsymbol{\gamma}^n(x_1, x_2) \geq \int_{X_1 \times X_2} f(x_1, x_2) \, d\boldsymbol{\gamma}(x_1, x_2), \tag{12.2.1}$$

and for every bounded Carathéodory integrand we have

$$\lim_{n \to \infty} \int_{X_1 \times X_2} f(x_1, x_2) \, d\boldsymbol{\gamma}^n(x_1, x_2) = \int_{X_1 \times X_2} f(x_1, x_2) \, d\boldsymbol{\gamma}(x_1, x_2). \tag{12.2.2}$$

Proof. We simply apply Lemma 5.1.10 and the Scorza–Dragoni approximation theorem of the previous section. □

If $\boldsymbol{\gamma}^n$ narrowly converge to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_1 \times X_2)$ and $\pi^1_\# \boldsymbol{\gamma}^n$ is independent of n, the following result provides a finer description of the limit $\boldsymbol{\gamma}$.

Lemma 12.2.2. *Let X_1, X_2 be Polish spaces and let $\boldsymbol{\gamma}^n \in \mathscr{P}(X_1 \times X_2)$ narrowly converging to $\boldsymbol{\gamma}$ and such that $\pi^1_\# \boldsymbol{\gamma}^n = \mu_1$ is independent of n. If $\{\gamma^n_{x_1}\}_{x_1 \in X_1}$, $\{\gamma_{x_1}\}_{x_1 \in X_1}$ are the disintegrations of $\boldsymbol{\gamma}^n$, $\boldsymbol{\gamma}$ w.r.t. μ_1 and $G_{x_1} \subset \mathscr{P}(X_2)$ is the subset of all the narrow accumulation points of $(\gamma^n_{x_1})_{n \in \mathbb{N}}$, then we have*

$$\gamma_{x_1} \subset \overline{\operatorname{conv} G_{x_1}} \quad \text{for } \mu_1\text{-a.e. } x_1 \in X_1. \tag{12.2.3}$$

In particular

$$\operatorname{supp} \gamma_{x_1} \subset \overline{\bigcup_{\gamma \in G_{x_1}} \operatorname{supp} \gamma} \quad \text{for } \mu_1\text{-a.e. } x_1 \in X_1. \tag{12.2.4}$$

Proof. Taking into account Remark 5.1.5 we can find a function $\varphi : X_2 \to [0, +\infty]$ with compact sublevels, such that

$$\int_{X_1\times X_2} \varphi(x_2)\, d\boldsymbol{\gamma}(x_1, x_2) \le \sup_{n\in\mathbb{N}} \int_{X_1\times X_2} \varphi(x_2)\, d\boldsymbol{\gamma}^n(x_1, x_2) = S < +\infty. \tag{12.2.5}$$

In particular, for any open set $A \subset X_1$ and any continuous and bounded function $f : X_2 \to \mathbb{R}$ we have

$$\begin{aligned}\int_{A\times X_2} f(x_2)\, d\boldsymbol{\gamma}(x_1, x_2) + \varepsilon S &\ge \lim_{n\to+\infty} \int_{A\times X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\boldsymbol{\gamma}^n(x_1, x_2)\\ &\ge \int_A \Big(\inf_{\varepsilon>0} \liminf_{n\to\infty} \int_{X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\gamma^n_{x_1}(x_2)\Big)\, d\mu^1(x_1)\end{aligned} \tag{12.2.6}$$

Passing to the limit as $\varepsilon \downarrow 0$ and observing that A is arbitrary, we get

$$\int_{X_2} f(x_2)\, d\gamma_{x_1}(x_2) \ge \inf_{\varepsilon>0} \liminf_{n\to\infty} \int_{X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\gamma^n_{x_1}(x_2) \quad \text{for } \mu\text{-a.e. } x_1$$

and it is not difficult to show using Prokhorov theorem that

$$\liminf_{n\to\infty} \int_{X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\gamma^n_{x_1}(x_2) \ge \inf_{\gamma\in G_{x_1}} \int_{X_2} f(x_2)\, d\gamma(x_2) \tag{12.2.7}$$

and

$$\int_{X_2} f(x_2)\, d\gamma_{x_1}(x_2) \ge \inf_{\gamma\in G_{x_1}} \int_{X_2} f(x_2)\, d\gamma(x_2) \tag{12.2.8}$$

for μ^1-a.e. $x_1 \in X_1$. Choosing f in a countable set $\mathscr{C}_0$ satisfying (5.1.2a,b) we can find a μ^1-negligible subset $N \subset X_1$ such that (12.2.8) holds for each $f \in \mathscr{C}$ and $x_1 \in X_1 \setminus N$. In fact the approximation property (5.1.2a,b) shows that (12.2.8) holds for each function $f \in C^0_b(X_2)$ and therefore Hahn–Banach theorem yields $\gamma_{x_1} \in \overline{\operatorname{conv}}\, G_{x_1}$ for $x_1 \in X_1 \setminus N$. □

We conclude this section with an useful convergence result:

Lemma 12.2.3. *Let X_1 be a Polish space, let X_2 be a separable Hilbert space, and let $f : X_2 \to [0, +\infty]$ be a l.s.c. strictly convex function. Suppose that $(\boldsymbol{\gamma}_n) \subset \mathscr{P}(X_1 \times X_2)$ narrowly converges to $\boldsymbol{\gamma} = \int_{X_1} \boldsymbol{\gamma}_{x_1}\, d\mu_1(x_1)$, with $\mu_1 = \pi^1_\#\boldsymbol{\gamma}$; if the barycenter of $\boldsymbol{\gamma}$ $\bar{\boldsymbol{\gamma}}(x_1) = \int_{X_2} x_2\, d\boldsymbol{\gamma}_{x_1}(x_2)$ exists and satisfies*

$$\liminf_{n\to\infty} \int_{X_1\times X_2} f(x_2)\, d\boldsymbol{\gamma}_n(x_1, x_2) = \int_{X_1} f(\bar{\boldsymbol{\gamma}}_{x_1})\, d\mu_1(x_1) \in \mathbb{R} \tag{12.2.9}$$

then $\boldsymbol{\gamma} = (\boldsymbol{i} \times \bar{\boldsymbol{\gamma}})_\#\mu_1$. The same result holds if $\pi^1_\#\boldsymbol{\gamma}^n = \mu_1$ and $f : X_1 \times X_2 \to [0, +\infty]$ is a normal integrand such that $f(x_1, \cdot)$ is strictly convex for μ_1-a.e. $x_1 \in X_1$; in this case the barycenters $\bar{\boldsymbol{\gamma}}_n$ converge to $\bar{\boldsymbol{\gamma}}$ in μ_1-measure.

Proof. Equality (12.2.9) yields

$$\int_{X_1}\Big(\int_{X_2} f(x_2)\,d\boldsymbol{\gamma}_{x_1}(x_2)\Big)\,d\mu_1(x_1) = \int_{X_1\times X_2} f(x_2)\,d\boldsymbol{\gamma}(x_1,x_2)$$
$$\le \liminf_{n\to+\infty}\int_{X_1\times X_2} f(x_2)\,d\boldsymbol{\gamma}_n(x_1,x_2)$$
$$\le \int_{X_1} f(\bar\gamma(x_1))\,d\gamma_1(x_1),$$

so that Jensen inequality yields

$$\int_{X_2} f(x_2)\,d\boldsymbol{\gamma}_{x_1}(x_2) = f(\bar\gamma(x_1)) \quad \text{for } \mu_1\text{-a.e. } x_1\in X_1$$

and the strict convexity of f yields $\boldsymbol{\gamma}_{x_1} = \delta_{\bar{\boldsymbol{\gamma}}(x_1)}$. The second part of the statement can be proved in an analogous way. □

12.3 PC metric spaces and their geometric tangent cone

In this section we review some basic general facts about *positively curved* (in short PC) spaces in the sense of Aleksandrov [5, 30, 120], and we recall the related notion of tangent cone; in the last section we will discuss its relationships with the tangent space we introduced in Section 8.4 for the Wasserstein space $\mathscr{P}_2(X)$.

Let $(\mathscr{S}, d)$ be a metric space; a *constant speed geodesic* $\boldsymbol{x}^{1\to2} : t\in[0,T]\mapsto x_t\in\mathscr{S}$ connecting x^1 to x^2 is a curve satisfying

$$x_0 = x^1,\quad x_T = x^2,\quad d(x_t,x_s) = \frac{t-s}{T}d(x^1,x^2)\quad \forall\, 0\le s\le t\le T. \tag{12.3.1}$$

In particular we are dealing with geodesics of minimal length whose metric derivative $|\boldsymbol{x}'|(t)$ is constant on $[0,T]$ and equal to $T^{-1}d(x^1,x^2)$.

We say that $\mathscr{S}$ is *geodesically complete* (or *length space*) if each couple of points can be connected by a constant speed geodesic.

Definition 12.3.1 (*PC*-spaces). *A geodesically complete metric space $(\mathscr{S},d)$ is* positively curved *(a PC-space) if for every $x^0\in\mathscr{S}$ and every constant speed geodesic $\boldsymbol{x}^{1\to2} : t\in[0,1]\mapsto x_t^{1\to2}$ connecting x^1 to x^2 it holds*

$$d^2(x_t^{1\to2},x^0) \ge (1-t)d^2(x^1,x^0) + td^2(x^2,x^0) - t(1-t)d^2(x^1,x^2). \tag{12.3.2}$$

Observe that in an Hilbert space X (12.3.2) is in fact an identity, since for $x_t^{1\to2} = (1-t)x^1 + tx^2$ we have

$$|x_t^{1\to2} - x^0|^2 = (1-t)|x^1-x^0|^2 + t|x^2-x^0|^2 - t(1-t)|x^1-x^2|^2. \tag{12.3.3}$$

Therefore condition (12.3.2) can be considered as a sort of comparison property for triangles: let us exploit this fact.

Definition 12.3.2 (Triangles). *A* triangle $\boldsymbol{x}$ *in* $\mathscr{S}$ *is a triple* $\boldsymbol{x}=(\boldsymbol{x}^{1\to 2},\boldsymbol{x}^{2\to 3},\boldsymbol{x}^{3\to 1})$ *of constant speed geodesics connecting (with obvious notation) three points* x^1, x^2, x^3 *in* $\mathscr{S}$. *We denote by* $\triangle=\triangle(\boldsymbol{x})\subset\mathscr{S}$ *the image of the curves* $\boldsymbol{x}^{1\to 2},\boldsymbol{x}^{2\to 3},\boldsymbol{x}^{3\to 1}$.

To each triangle $\boldsymbol{x}$ in $\mathscr{S}$ we can consider a corresponding reference triangle (unique, up to isometric transformation) $\hat{\boldsymbol{x}}=(\hat{\boldsymbol{x}}^{1\to 2},\hat{\boldsymbol{x}}^{2\to 3},\hat{\boldsymbol{x}}^{3\to 1})$ in $\mathbb{R}^2$ connecting the points $\hat{x}^1,\hat{x}^2,\hat{x}^3\in\mathbb{R}^2$ such that

$$|\hat{x}^i-\hat{x}^j|=d(x^i,x^j)\quad i,j=1,2,3. \tag{12.3.4}$$

Two points $x\in\triangle$, $\hat{x}\in\hat{\triangle}$ are correspondent if

$$x=\boldsymbol{x}_t^{i\to j},\quad \hat{x}=\hat{\boldsymbol{x}}_t^{i\to j}\quad \text{for some } t\in[0,1],\ i,j\in\{1,2,3\}.$$

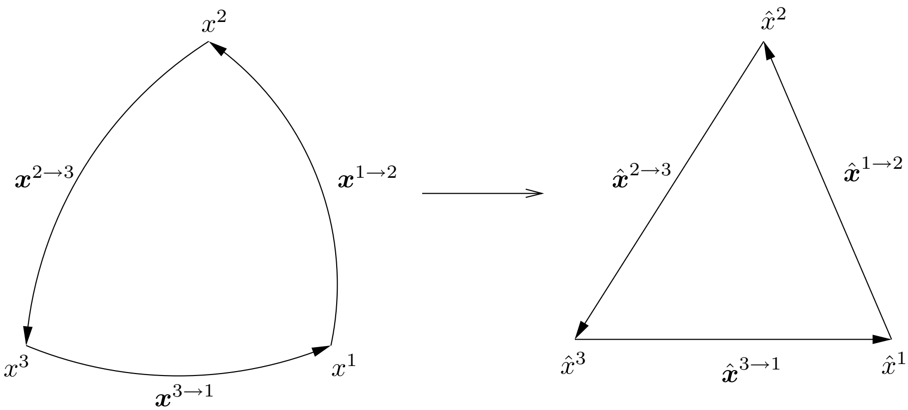


Figure 12.1: on the left the triangle on the PC-space and on the right its euclidean reference.

Proposition 12.3.3 (Triangle comparison). *If* $\mathscr{S}$ *is a PC-space and* $\triangle\subset\mathscr{S},\hat{\triangle}\subset\mathbb{R}^2$ *are two corresponding triangles, then for each couples of correspondent points* $x,y\in\triangle,\hat{x},\hat{y}\in\hat{\triangle}$ *we have*

$$d(x,y)\geq|\hat{x}-\hat{y}|. \tag{12.3.5}$$

Proof. When x or y is a vertex of the triangle, then (12.3.5) is just (12.3.2): thus we have to examine the case (up to permutation of the indexes) $x=x_t^{1\to 2}$, $y=x_s^{1\to 3}$, $t,s\in(0,1)$. Denoting by $\boldsymbol{x}^{1\to t}$ the rescaled geodesic connecting x^1 to $x=x_t^{1\to 2}$ and by introducing a new geodesic $\boldsymbol{x}^{t\to 3}$ connecting x to x^3, we can consider the new triangle $\boldsymbol{x}'=(\boldsymbol{x}^{1\to t},\boldsymbol{x}^{t\to 3},\boldsymbol{x}^{3\to 1})$ connecting x^1,x,x^3. The corresponding euclidean reference $\hat{\boldsymbol{x}}'$ can be constructed keeping fixed $\hat{x}^1$ and $\hat{x}^3$ (and therefore $\hat{y}=\hat{x}_s^{1\to 3}$) and introducing a new point $\hat{x}'$, which in general will be different from $\hat{x}$, such that $|\hat{x}'-\hat{x}^1|=d(x,x^1)$, $|\hat{x}'-\hat{x}^3|=d(x,x^3)$. Applying (12.3.2) we obtain

$$|\hat{x}'-\hat{x}^3|=d(x,x^3)\geq|\hat{x}-\hat{x}^3|$$

and applying the identity (12.3.3) we get

$$\begin{aligned}|\hat{x}' - \hat{y}|^2 &= (1-s)|\hat{x}' - \hat{x}^1|^2 + s|\hat{x}' - \hat{x}^3|^2 - s(1-s)|\hat{x}^3 - \hat{x}^1|^2 \\ &\geq (1-s)|\hat{x} - \hat{x}^1|^2 + s|\hat{x} - \hat{x}^3|^2 - s(1-s)|\hat{x} - \hat{x}^1|^2 = |\hat{x} - \hat{y}|^2\end{aligned}$$

therefore, applying (12.3.2) again to the triangles $\boldsymbol{x}', \hat{\boldsymbol{x}}'$ we obtain

$$d(x,y) \geq |\hat{x}' - \hat{y}'| = |\hat{x}' - \hat{y}| \geq |\hat{x} - \hat{y}|. \qquad \square$$

In a Hilbert space X the angle $\angle(\hat{\boldsymbol{x}}^{1\to 2}, \hat{\boldsymbol{x}}^{1\to 3}) \in [0,\pi]$ between the two segments joining $\hat{x}^1$ to $\hat{x}^2$ and $\hat{x}^1$ to $\hat{x}^3$ can be easily computed by the formula

$$\cos(\angle(\hat{\boldsymbol{x}}^{1\to 2}, \hat{\boldsymbol{x}}^{1\to 3})) = \frac{\langle \hat{x}^2 - \hat{x}^1, \hat{x}^3 - \hat{x}^1\rangle}{|\hat{x}^2 - \hat{x}^1|\,|\hat{x}^3 - \hat{x}^1|} = \alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3), \tag{12.3.6}$$

where

$$\alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3) = \frac{|\hat{x}^2 - \hat{x}^1|^2 + |\hat{x}^3 - \hat{x}^1|^2 - |\hat{x}^3 - \hat{x}^2|^2}{2|\hat{x}^2 - \hat{x}^1|\,|\hat{x}^3 - \hat{x}^1|}. \tag{12.3.7}$$

In particular, if $\hat{x}^{1\to 2}_t := (1-t)\hat{x}^1 + t\hat{x}^2$ and $\hat{x}^{1\to 3}_s := (1-s)\hat{x}^1 + s\hat{x}^3$, we have

$$\alpha(\hat{x}^1; \hat{x}^{1\to 2}_t, \hat{x}^{1\to 3}_s) = \alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3) \quad \forall\, t,\, s \in (0,1]. \tag{12.3.8}$$

Taking into account of (12.3.7), in the case of a general PC-space, it is natural to introduce the function

$$\alpha(x^1; x^2, x^3) := \frac{d(x^2,x^1)^2 + d(x^3,x^1)^2 - d(x^3,x^2)^2}{2d(x^2,x^1)\,d(x^3,x^1)}, \quad x^1 \neq x^2, x^3 \tag{12.3.9}$$

and we have the following monotonicity result.

Lemma 12.3.4 (Angle between geodesics). *Let $(\mathscr{S}, d)$ be a PC-space and let $\boldsymbol{x}^{1\to 2}$, $\boldsymbol{x}^{1\to 3}$ be constant speed geodesics starting from x^1; then the function*

$$t, s \in (0,1] \mapsto \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s) \quad \textit{is nondecreasing in } s,t. \tag{12.3.10}$$

The angle $\angle(\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{1\to 3}) \in [0,\pi]$ *between $\boldsymbol{x}^{1\to 2}$ and $\boldsymbol{x}^{1\to 3}$ is thus defined by the formula*

$$\cos(\angle(\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{1\to 3})) := \inf_{s,t} \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s) = \lim_{s,t\downarrow 0} \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s). \tag{12.3.11}$$

Proof. It is sufficient to prove that $\alpha(x^1; x^2, x^3) \geq \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s)$ for $s,t \in (0,1]$; if $\hat{\boldsymbol{x}}$ is a corresponding reference triangle with vertexes $\hat{x}^1, \hat{x}^2, \hat{x}^3$, we easily have by Proposition 12.3.3 and (12.3.8)

$$\alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s) \leq \alpha(\hat{x}^1; \hat{x}^{1\to 2}_t, \hat{x}^{1\to 3}_s) = \alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3) = \alpha(x^1; x^2, x^3) \qquad \square$$

Remark 12.3.5. Notice that the separate limit as $t \downarrow 0$ is given by

$$\lim_{t\downarrow 0} \alpha(x^1; x_t^{1\to 2}, x_s^{1\to 3}) = \lim_{t\downarrow 0} \frac{t^2 d^2(x^1,x^2) + d^2(x^1, x_s^{1\to 3}) - d^2(x_t^{1\to 2}, x_s^{1\to 3})}{2ts\, d(x^1,x^2)\, d(x^1,x^3)}$$
$$= -\big(2sd(x^1,x^2)\, d(x^1,x^3)\big)^{-1} \frac{d}{dt}\Big(d^2(x_t^{1\to 2}, x_s^{1\to 3})\Big)\big|_{t=0+}$$

and therefore

$$\cos\big(\angle(\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{1\to 3})\big) = -\big(2d(x^1,x^2)\, d(x^1,x^3)\big)^{-1} \frac{\partial^2}{\partial s \partial t}\Big(d^2(x_t^{1\to 2}, x_s^{1\to 3})\Big)\big|_{t,s=0+}$$

For a fixed $x \in \mathscr{S}$ let us denote by $G(x)$ the set of all constant speed geodesics $\boldsymbol{x}$ starting from x and parametrized in some interval $[0, T_{\boldsymbol{x}}]$; recall that the metric velocity of $\boldsymbol{x}$ is $|\boldsymbol{x}'| = d(\boldsymbol{x}(t), x)/t$, $t \in (0,T]$. We set

$$\|\boldsymbol{x}\|_x := |\boldsymbol{x}'|, \quad \langle \boldsymbol{x}, \boldsymbol{y}\rangle_x := \|\boldsymbol{x}\|_x \, \|\boldsymbol{y}\|_x \cos(\angle(\boldsymbol{x},\boldsymbol{y})),$$
$$d_x^2(\boldsymbol{x},\boldsymbol{y}) := \|\boldsymbol{x}\|_x^2 + \|\boldsymbol{y}\|_x^2 - 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle_x. \tag{12.3.12}$$

If $\boldsymbol{x} \in G(x)$ and $\lambda > 0$ we denote by $\lambda\boldsymbol{x}$ the geodesic

$$(\lambda \boldsymbol{x})_t := \boldsymbol{x}_{\lambda t}, \quad T_{\lambda \boldsymbol{x}} = \lambda^{-1} T_{\boldsymbol{x}}, \tag{12.3.13}$$

and we observe that for each $\boldsymbol{x}, \boldsymbol{y} \in G(x)$, $\lambda > 0$, it holds

$$\|\lambda \boldsymbol{x}\|_x = \lambda \|\boldsymbol{x}\|_x, \quad \langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle_x = \langle \boldsymbol{x}, \lambda\boldsymbol{y}\rangle_x = \lambda \langle \boldsymbol{x}, \boldsymbol{y}\rangle_x \tag{12.3.14}$$

Observe that the restriction of a geodesic is still a geodesic; we say that $\boldsymbol{x} \sim \boldsymbol{y}$ if there exist $\varepsilon > 0$ such that $\boldsymbol{x}|_{[0,\varepsilon]} = \boldsymbol{y}|_{[0,\varepsilon]}$.

Theorem 12.3.6 (An abstract notion of Tangent cone). *If $\boldsymbol{x}$, $\boldsymbol{y} : [0,T] \to \mathscr{S}$ are two geodesics starting from x we have*

$$d_x(\boldsymbol{x},\boldsymbol{y}) = \lim_{t\downarrow 0} \frac{d(\boldsymbol{x}_t, \boldsymbol{y}_t)}{t} = \sup_{t\in(0,T]} \frac{d(\boldsymbol{x}_t,\boldsymbol{y}_t)}{t}. \tag{12.3.15}$$

In particular, the function d_x defined by (12.3.12) *is a distance on the quotient space $G(x)/\sim$. The completion of $G(x)/\sim$ is called the* tangent cone $\mathbf{Tan}_x\mathscr{S}$ *at the point x.*

Proof. (12.3.15) follows by a simple computation since for each $s > 0$ (12.3.11) yields

$$\cos\big(\angle(\boldsymbol{x},\boldsymbol{y})\big) = \lim_{t\downarrow 0} \frac{d^2(\boldsymbol{x}_{ts}, x) + d^2(\boldsymbol{y}_{ts}, x) - d^2(\boldsymbol{x}_{ts}, \boldsymbol{y}_{ts})}{2d(\boldsymbol{x}_{ts}, x) d(\boldsymbol{y}_{ts}, x)}$$
$$= \frac{d^2(\boldsymbol{x}_s, x) + d^2(\boldsymbol{y}_s, x)}{2d(\boldsymbol{x}_s, x) d(\boldsymbol{y}_s, x)} - \lim_{t\downarrow 0} \frac{d^2(\boldsymbol{x}_{ts}, \boldsymbol{y}_{ts})}{2t^2 d(\boldsymbol{x}_s, x) d(\boldsymbol{y}_s, x)}$$

and therefore from (12.3.12) we have

$$d_x^2(\boldsymbol{x},\boldsymbol{y}) = \frac{d^2(\boldsymbol{x}_s,x)+d^2(\boldsymbol{y}_s,x)}{s^2} - 2\frac{d(\boldsymbol{x}_s,x)d(\boldsymbol{y}_s,x)}{s^2}\cos(\angle(\boldsymbol{x},\boldsymbol{y}))$$
$$= \lim_{t\downarrow 0}\frac{d^2(\boldsymbol{x}_{ts},\boldsymbol{y}_{ts})}{2t^2s^2}.$$

□

Remark 12.3.7 (The tangent cone as Gromov-Hausdorff blow up of pointed spaces). In the finite dimensional case $\mathbf{Tan}_x\mathscr{S}$ can also be characterized as the Gromov-Hasudorff limit of the sequence of pointed metric spaces $(\mathscr{S}, x, n\cdot d)$ as $n\to\infty$. [30, 7.8.1]

12.4 The geometric tangent spaces in $\mathscr{P}_2(X)$

Taking into account of the abstract definition of Tangent cone 12.3.6 for *PC*-spaces and the fact proved in Section 7.3 that $\mathscr{P}_2(X)$ is a *PC*-space, we want an explicit representation of the abstract tangent space $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ induced by the 2-Wasserstein distance.

First of all we want to determine a precise expression for the angle between two geodesics. Observe that an optimal plan $\boldsymbol{\mu}\in\Gamma_o(\mu^1,\mu^2)$ is associated to the geodesic $\boldsymbol{\mu}^{1\to2}$ with $\mu_t^{1\to2} = (\pi_t^{1\to2})_\#\boldsymbol{\mu}$ whose velocity is equal to the distance between the end points $|\mu'|^2 = \int|x_2-x_1|^2\,d\boldsymbol{\mu}$. If we want to represent each constant speed geodesics, it is convenient to introduce the new "velocity" plans

$$\boldsymbol{\gamma}_\lambda := \big(\pi^1,\lambda(\pi^2-\pi^1)\big)_\#\boldsymbol{\mu}, \tag{12.4.1}$$

that can be used to provide a natural parametrizations for the rescaled geodesic $(\lambda\cdot\boldsymbol{\mu}^{1\to2})_t := \boldsymbol{\mu}_{\lambda t}^{1\to2}$ as follows:

$$\mu_{\lambda t}^{1\to2} = \big((1-\lambda t)\pi^1+\lambda t\pi^2\big)_\#\boldsymbol{\mu} = (\pi^1+t\pi^2)_\#\boldsymbol{\gamma}_\lambda \qquad t\in\big[0,\lambda^{-1}\big]. \tag{12.4.2}$$

Therefore we can identify constant speed geodesics parametrized in some interval $[0,\lambda^{-1}]$ with transport plans $\boldsymbol{\gamma}$ of the type

$$\boldsymbol{\gamma} = \big(\pi^1,\lambda(\pi^2-\pi^1)\big)_\#\boldsymbol{\mu} \quad \text{for some } \textit{optimal plan } \boldsymbol{\mu}\in\mathscr{P}_2(X),$$

and therefore we set

$$\boldsymbol{G}(\mu) := \Big\{\boldsymbol{\gamma}\in\mathscr{P}_2(X^2):\ \big(\pi^1,\pi^1+\varepsilon\pi^2\big)_\#\boldsymbol{\gamma} \text{ is optimal, for some } \varepsilon>0\Big\}. \tag{12.4.3}$$

It easy to check that there is a one-to-one correspondence between $\boldsymbol{G}(\mu)$ and the quotient $G(\mu)/\sim$ introduced in the previous section: for, to each plan $\boldsymbol{\gamma}\in\boldsymbol{G}(\mu)$ we associate the (equivalence class of the) geodesic

$$\mu_t := (\pi^1+t\pi^2)_\#\boldsymbol{\gamma}, \quad 0\le t\le\varepsilon, \tag{12.4.4}$$

where $\varepsilon > 0$ is chosen as in (12.4.3). Conversely, if μ_t, $t \in [0,T]$, is a curve such that $\mu|_{[0,\varepsilon]}$ is a (minimal, constant speed) geodesic, then for every $\lambda^{-1} \in (0,\varepsilon]$ there exists a unique optimal plan $\boldsymbol{\mu}_\lambda \in \Gamma_o(\mu_0, \mu_{\lambda^{-1}})$ such that

$$\mu_t = \big(\pi^1 + \lambda t(\pi^2 - \pi^1)\big)_\# \boldsymbol{\mu}_\lambda \quad t \in [0, \lambda^{-1}];$$

by Theorem 7.2.2

$$0 < \lambda_1^{-1} < \lambda_2^{-1} \le \varepsilon \quad \Longrightarrow \quad \boldsymbol{\mu}_{\lambda_1} = \big(\pi^1, \pi^1 + \lambda_2/\lambda_1(\pi^2 - \pi^1)\big)_\# \boldsymbol{\mu}_{\lambda_2},$$

so that

$$\boldsymbol{\gamma} = \big(\pi^1, \lambda(\pi^2 - \pi^1)\big)_\# \boldsymbol{\mu}_\lambda \quad \text{is independent of } \lambda, \text{ belongs to } \boldsymbol{G}(\mu), \tag{12.4.5}$$

and represents μ_t through (12.4.4).
Motivated by the above discussion, we introduce the following definition:

Definition 12.4.1 (Exponential map in $\mathscr{P}_2(X)$). *For $\mu \in \mathscr{P}(X)$ and $\boldsymbol{\gamma} \in \boldsymbol{G}(\mu)$ we define*

$$\lambda \cdot \boldsymbol{\gamma} := \big(\pi^1, \lambda\pi^2\big)_\# \boldsymbol{\gamma}, \quad \exp_\mu(\boldsymbol{\gamma}) := \big(\pi^1 + \pi^2\big)_\# \boldsymbol{\gamma}. \tag{12.4.6}$$

The notation is justified by the fact that the curve

$$t \mapsto \exp_\mu(t \cdot \boldsymbol{\gamma}) \quad \text{is a constant speed geodesic in some interval } [0,\varepsilon] \tag{12.4.7}$$

whenever $\boldsymbol{\gamma} \in \boldsymbol{G}(\mu)$.
For $\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3} \in \mathscr{P}_2(X^2)$ with $\pi^1_\# \boldsymbol{\gamma}^{1\,i} = \mu$, $i = 2, 3$, we set

$$\|\boldsymbol{\gamma}^{1\,2}\|^2_\mu := \int_{X^2} |x_2|^2 \, d\boldsymbol{\gamma}^{1\,2}(x_1, x_2), \tag{12.4.8}$$

$$\langle \boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3} \rangle_\mu = \max\Big\{ \int_{X^3} \langle x_2, x_3 \rangle \, d\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) \Big\}, \tag{12.4.9}$$

$$W^2_\mu(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) = \min\Big\{ \int_{X^3} |x_2 - x_3|^2 \, d\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) \Big\}, \tag{12.4.10}$$

where $\Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ is the family of all 3-plans in $\boldsymbol{\gamma} \in \mathscr{P}(X^3)$ such that $\pi^{1,2}_\# \boldsymbol{\gamma} = \boldsymbol{\gamma}^{1\,2}$ and $\pi^{1,3}_\# \boldsymbol{\gamma} = \boldsymbol{\gamma}^{1\,3}$.

Proposition 12.4.2. *Suppose that $\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}$ belongs to $\boldsymbol{G}(\mu)$ so that they can be identified with the constant speed geodesics $\boldsymbol{\mu}^{1\to 2}, \boldsymbol{\mu}^{1\to 3}$ through* (12.4.4)*. Then the previous definitions coincide with the corresponding quantities introduced in* (12.3.12) *for general PC-metric spaces.*

Proof. The first identity of (12.4.8) is immediate. In order to prove the second one we apply Proposition 7.3.6, by taking into account Remark 12.3.5: thus we have

$$\langle \boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3} \rangle_\mu = \lim_{s \downarrow 0} 2s^{-1} \int_{X^3} \langle x_2 - x_1, x_3 - x_1 \rangle \, d\boldsymbol{\mu}_s,$$

where $\mu_s^{1\to 3} = \exp_\mu(s\boldsymbol{\gamma}^{1\,3})$ and $\boldsymbol{\mu}_s \in \Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu_s^{1\to 3})$ is chosen among the minimizers of (7.3.15). It is easy to check that we can choose

$$\boldsymbol{\mu}_s = \big(\pi^1, \pi^1 + \pi^2, \pi^1 + s\pi^2\big)_{\#}\boldsymbol{\gamma},$$

where $\boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ realizes the maximum in (12.4.9) (or equivalently the minumum of (12.4.10)) and therefore

$$\begin{aligned}\lim_{s\downarrow 0} s^{-1}\int_{X^3} \langle x_2 - x_1, x_3 - x_1\rangle\, d\boldsymbol{\mu}_s &= \lim_{s\downarrow 0} s^{-1}\int_{X^3} \langle x_2, x_1 + sx_3 - x_1\rangle\, d\boldsymbol{\gamma}\\ &= \int_{X^3} \langle x_2, x_3\rangle\, d\boldsymbol{\gamma}.\end{aligned}$$

The last formula of (12.4.8) follows now directly by the definition (12.3.12). □

If either $\boldsymbol{\gamma}^{1\,2}$ or $\boldsymbol{\gamma}^{1\,3}$ are induced by a transport map $\boldsymbol{t}$, e.g. $\boldsymbol{\gamma}^{1\,2} = \big(\boldsymbol{i}\times\boldsymbol{t}\big)_{\#}\mu$, then the previous formulae are considerably simpler, since

$$\|\boldsymbol{\gamma}^{1\,2}\|_\mu^2 := \int_{X^2} |\boldsymbol{t}(x_1)|^2\, d\mu(x_1) = \|\boldsymbol{t}\|^2_{L^2(\mu;X)}, \tag{12.4.11}$$

$$\langle \boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}\rangle_\mu = \int_{X^2} \langle \boldsymbol{t}(x_1), x_3\rangle\, d\boldsymbol{\gamma}^{1\,3}(x_1, x_3), \tag{12.4.12}$$

$$W_\mu^2(\boldsymbol{\gamma}^{1,2}, \boldsymbol{\gamma}^{1\,3}) = \int_{X^2} |\boldsymbol{t}(x_1) - x_3|^2\, d\boldsymbol{\gamma}^{1\,3}(x_1, x_3). \tag{12.4.13}$$

Finally, if also $\boldsymbol{\gamma}^{1\,3} = \big(\boldsymbol{i}\times\boldsymbol{s}\big)_{\#}\mu$, then (12.4.12) and (12.4.13) become

$$\langle \boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}\rangle_\mu = \int_X \langle \boldsymbol{t}(x_1), \boldsymbol{s}(x_1)\rangle\, d\mu(x_1) = (\boldsymbol{t}, \boldsymbol{s})_{L^2(\mu;X)}, \tag{12.4.14}$$

$$W_\mu^2(\boldsymbol{\gamma}^{1,2}, \boldsymbol{\gamma}^{1\,3}) = \int_X |\boldsymbol{t}(x_1) - \boldsymbol{s}(x_1)|^2\, d\mu(x_1) = \|\boldsymbol{t} - \boldsymbol{s}\|^2_{L^2(\mu;X)}. \tag{12.4.15}$$

These results lead to the following definition.

Definition 12.4.3 (Geometric tangent cone). *The* geometric tangent cone $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ *to* $\mathscr{P}_2(X)$ *at* μ *is the* closure *of* $\boldsymbol{G}(\mu)$ *in* $\mathscr{P}_2(X^2)$ *with respect to the distance* $W_\mu(\cdot,\cdot)$.

In Section 8.4 we already introduced a notion of tangent space $\mathrm{Tan}_\mu\mathscr{P}_2(X)$ and we showed in Theorem 8.5.1 its equivalent characterization in terms of optimal transport maps

$$\mathrm{Tan}_\mu\mathscr{P}_2(X) = \overline{\big\{\lambda(\boldsymbol{r} - \boldsymbol{i}) : \quad (\boldsymbol{i}\times\boldsymbol{r})_{\#}\mu \in \Gamma_o(\mu, \boldsymbol{r}_{\#}\mu),\ \lambda > 0\big\}}^{L^2(\mu;X)}. \tag{12.4.16}$$

In order to compare these two notions, let us recall the Definition 5.4.2 of *barycentric projection* $\bar{\boldsymbol{\gamma}}$ of a plan $\boldsymbol{\gamma} \in \mathscr{P}_2(X^2)$ with $\pi^1_{\#}\boldsymbol{\gamma} = \mu$:

$$\boldsymbol{t} := \bar{\boldsymbol{\gamma}} \quad\Leftrightarrow\quad \boldsymbol{t}(x_1) = \int_X x_2\, d\boldsymbol{\gamma}_{x_1}(x_2), \quad \boldsymbol{t} \in L^2(\mu; X), \tag{12.4.17}$$

which is a nonexpansive map from $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ to $L^2(\mu;X)$. Indeed choosing $\boldsymbol{\gamma}\in\Gamma^1(\boldsymbol{\gamma}^1,\boldsymbol{\gamma}^2)$ and denoting by $\boldsymbol{\gamma}^1_{x_1}$ and $\boldsymbol{\gamma}^2_{x_1}$ the disintegrations of $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$ w.r.t. μ we have

$$\int_X|\bar{\boldsymbol{\gamma}}^1-\bar{\boldsymbol{\gamma}}^2|^2\,d\mu=\int_X\left|\int_{X^2}(x_2-x_3)\,d\boldsymbol{\gamma}_{x_1}\right|^2\,d\mu\le\int_{X^3}|x_2-x_3|^2\,d\boldsymbol{\gamma},$$

so that

$$\|\bar{\boldsymbol{\gamma}}^1-\bar{\boldsymbol{\gamma}}^2\|_{L^2(\mu;X)}\le W_\mu(\boldsymbol{\gamma}^1,\boldsymbol{\gamma}^2). \tag{12.4.18}$$

We have the following result:

Theorem 12.4.4. *For every $\mu\in\mathscr{P}_2(X)$ the reduced tangent space is the image of $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ through the barycentric projection. Moreover, if $\mu\in\mathscr{P}_2^r(X)$, then the barycentric projection is an isometric one-to-one correpondence between $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ and $\mathrm{Tan}_\mu\mathscr{P}_2(X)$.*

Proof. Let us first prove that $\bar{\boldsymbol{\gamma}}\in\mathrm{Tan}_\mu\mathscr{P}_2(X)$ for any $\boldsymbol{\gamma}\in\mathbf{Tan}_\mu\mathscr{P}_2(X)$. By the continuity of the barycentric projection and the identity $\overline{(\pi^1,\pi^1+\varepsilon\pi^2)_\#\boldsymbol{\gamma}}=\boldsymbol{i}+\varepsilon\bar{\boldsymbol{\gamma}}$, it suffices to show that $(\bar{\boldsymbol{\mu}}-\boldsymbol{i})\in\mathrm{Tan}_\mu\mathscr{P}_2(X)$ for any optimal plan $\boldsymbol{\mu}$ whose first marginal is μ. We know that $\operatorname{supp}\boldsymbol{\mu}$ is contained in the graph of the subdifferential of a convex and l.s.c. function $\psi:X\to(-\infty,+\infty]$, i.e.

$$y\in\partial\psi(x)\qquad\text{for any }(x,y)\in\operatorname{supp}\boldsymbol{\gamma}.$$

Since $\partial\psi(x)$ is a closed convex subset of X for every $x\in D(\partial\psi)$, we obtain that $\bar{\boldsymbol{\mu}}(x)=\int_X y\,d\boldsymbol{\mu}_{x_1}(y)\in\partial^-\psi(x)$ for μ-a.e. x; therefore $\bar{\boldsymbol{\mu}}$ is an optimal transport map and $(\bar{\boldsymbol{\mu}}-\boldsymbol{i})\in\mathrm{Tan}_\mu\mathscr{P}_2(X)$.

In order to show that the barycentric projection is onto it suffices to prove that the map $I:\mathrm{Tan}_\mu\mathscr{P}_2(X)\mapsto\mathscr{P}(X\times X)$ defined by $I(\boldsymbol{v}):=(\boldsymbol{i}\times\boldsymbol{v})_\#\mu$ takes its values in $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ and to notice that it satisfies $\overline{I(\boldsymbol{v})}=\boldsymbol{v}$. Since the unique plan in $\Gamma^1\left(I(\boldsymbol{v}),I(\boldsymbol{v}')\right)$ is $(\boldsymbol{i}\times\boldsymbol{v}\times\boldsymbol{v}')_\#\mu$, we have

$$W_\mu^2\left(I(\boldsymbol{v}),I(\boldsymbol{v}')\right)=\int_X|\boldsymbol{v}-\boldsymbol{v}'|^2\,d\mu,$$

so that our thesis follows if $I(\boldsymbol{v})\in\boldsymbol{G}(\mu)$ for every $\boldsymbol{v}$ in the dense subset of $\mathrm{Tan}_\mu\mathscr{P}_2(X)$ introduced in (12.4.16): this last property follows trivially by the definition of $\boldsymbol{G}(\mu)$ (12.4.3). Finally in the case when μ is regular all optimal transport plans in $\boldsymbol{G}(\mu)$ are induced by transports: therefore I is onto and it is the inverse of the barycentric projection. □

Remark 12.4.5 (The exponential map and its inverse). Observe that the exponential map is a contraction since

$$W_2(\exp_\mu(\boldsymbol{\mu}),\exp_\mu(\boldsymbol{\sigma}))\le W_\mu(\boldsymbol{\mu},\boldsymbol{\sigma}), \tag{12.4.19}$$

but in general, it is not injective, even if it is restricted to the tangent space. Nevertheless it admits a natural (multivalued) right inverse defined by

$$\exp_\mu^{-1}(\nu) := \Big\{ \boldsymbol{\mu} \in \boldsymbol{G}(\mu) : \big(\pi^1, \pi^1 + \pi^2\big)_{\#}\boldsymbol{\mu} \in \Gamma_o(\mu, \nu) \Big\}. \tag{12.4.20}$$

We conclude this section with an explicit representation of the distance W_μ defined by (12.4.10).

Proposition 12.4.6. *Let $\boldsymbol{\gamma}^{1\,2}$, $\boldsymbol{\gamma}^{1\,3}$ be two plans in $\mathscr{P}_2(X^2)$ with the same first marginal μ. Then $\boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ realizes the minimum in* (12.4.8) *if and only if its disintegration w.r.t. μ satisfies*

$$\boldsymbol{\gamma}_{x_1} \in \Gamma_o(\boldsymbol{\gamma}^{1\,2}_{x_1}, \boldsymbol{\gamma}^{1\,3}_{x_1}) \quad \textit{for } \mu\textit{-a.e. } x_1 \in X. \tag{12.4.21}$$

Moreover

$$W^2_\mu(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) = \int_X W_2^2(\boldsymbol{\gamma}^{1\,2}_{x_1}, \boldsymbol{\gamma}^{1\,3}_{x_1})\, d\mu(x_1). \tag{12.4.22}$$

Proof. For any $\boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ we clearly have

$$\int_{X^3} |x_2 - x_3|^2\, d\boldsymbol{\gamma} = \int_X \int_{X^2} |x_2 - x_3|^2\, d\boldsymbol{\gamma}_{x_1}\, d\mu(x_1) \geq \int_X W_2^2(\mu^{1,\,2}_{x_1}, \mu^{1,\,3}_{x_1})\, d\mu(x_1).$$

Equality and the necessary and sufficient condition for optimality follows immediately by Lemma 5.3.2 and by the next measurable selection result. □

Lemma 12.4.7. *Suppose that $(\mu^2_{x_1})_{x_1 \in X_1}$, $(\mu^3_{x_1})_{x_1 \in X_1}$ are Borel families of measures in $\mathscr{P}_p(X)$ defined in a Polish space X_1.*

The map

$$x_1 \mapsto W_p^p(\mu^2_{x_1}, \mu^3_{x_1}) \quad \textit{is Borel} \tag{12.4.23}$$

and there exists a Borel family $\boldsymbol{\gamma}_{x_1} \in \mathscr{P}_p(X \times X)$ such that $\boldsymbol{\gamma}_{x_1} \in \Gamma_o(\mu^2_{x_1}, \mu^3_{x_1})$.

Proof. We show first that $x \mapsto \sigma_x$ is a Borel map between X_1 and $\mathscr{P}_p(X)$ whenever $x \mapsto \sigma_x$ is Borel in the sense used in Section 5.3. Indeed by assumption $x \mapsto \sigma_x(A)$ is a Borel map for any open set $A \subset X$ and since

$$\int_X f\, d\sigma_x = \int_0^\infty \sigma_x(\{f > t\})\, dt - \int_{-\infty}^0 \sigma_x(\{f < t\})\, dt$$

and the integral can be approximated by Riemann sums, we have also that $x \mapsto \int_X f\, d\sigma_x$ is Borel for any $f \in C^0_b(X)$.

Let δ be the distance inducing the narrow convergence on $\mathscr{P}(X)$ introduced in (5.1.6). It follows that $x \mapsto \delta(\sigma_x, \sigma)$ is Borel for any $\sigma \in \mathscr{P}(X)$. By (7.1.12) it follows that the distance $\tilde{W}$ defined by

$$\tilde{W}^p(\mu, \sigma) := \delta^p(\mu, \sigma) + \left| \int |x|^p\, d\mu - \int |x|^p\, d\sigma \right|$$

induces the p-Wasserstein topology on $\mathscr{P}_p(X)$; we deduce that $x \mapsto \tilde{W}(\sigma_x, \sigma)$ is Borel for any $\sigma \in \mathscr{P}_p(X)$, therefore $x \mapsto \sigma_x$ is Borel, seen as a function with values in $\mathscr{P}_p(X)$.

In order to prove the second part of the statement, let us observe that the multivalued map $\mu^2, \mu^3 \in \mathscr{P}_p(X) \mapsto \Gamma_o(\mu^2, \mu^3) \subset \mathscr{P}_p(X \times X)$ is upper semicontinuous thanks to Proposition 7.1.3. In particular for each open set $G \subset \mathscr{P}_p(X \times X)$ the set

$$\Big\{(\mu^2, \mu^3) : \Gamma_o(\mu^2, \mu^3) \cap G \neq \varnothing\Big\}$$

is open in $\mathscr{P}_p(X) \times \mathscr{P}_p(X)$. Therefore classical measurable selection theorems (see for instance Theorem III.23 in [39]) give the thesis. $\square$

Bibliography

[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations, to appear, (2002).

[2] ——, *Asymptotic behavior for doubly degenerate parabolic equations*, C. R. Math. Acad. Sci. Paris, 337 (2003), pp. 331–336.

[3] M. Agueh, N. Ghoussoub, and X. Kang, *The optimal evolution of the free energy of interacting gases and its applications*, C. R. Math. Acad. Sci. Paris, 337 (2003), pp. 173–178.

[4] G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in* $\mathbf{R}^n$, Math. Z., 230 (1999), pp. 259–316.

[5] A. D. Aleksandrov, *A theorem on triangles in a metric space and some of its applications*, in Trudy Mat. Inst. Steklov., v 38, Trudy Mat. Inst. Steklov., v 38, Izdat. Akad. Nauk SSSR, Moscow, 1951, pp. 5–23.

[6] F. Almgren, J. E. Taylor, and L. Wang, *Curvature-driven flows: a variational approach*, SIAM J. Control Optim., 31 (1993), pp. 387–438.

[7] L. Ambrosio, *Metric space valued functions of bounded variation*, Ann. Sc. Norm. Sup. Pisa, 17 (1990), pp. 439–478.

[8] ——, *Minimizing movements*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5), 19 (1995), pp. 191–246.

[9] ——, *Lecture notes on optimal transport problem*, in Mathematical aspects of evolving interfaces, CIME summer school in Madeira (Pt), P. Colli and J. Rodrigues, eds., vol. 1812, Springer, 2003, pp. 1–52.

[10] ——, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math., 158 (2004), pp. 227–260.

[11] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, Clarendon Press, Oxford, 2000.

[12] L. Ambrosio and B. Kirchheim, *Rectifiable sets in metric and Banach spaces*, Math. Ann., 318 (2000), pp. 527–555.

[13] L. Ambrosio, B. Kirchheim, and A. Pratelli, *Existence of optimal transport maps for crystalline norms*, Duke Mathematical Journal, to appear, (2003).

[14] L. Ambrosio and A. Pratelli, *Existence and stability results in the L^1 theory of optimal transportation*, in Optimal transportation and applications, Lecture Notes in Mathematics, L. Caffarelli and S. Salsa, eds., vol. 1813, Springer, 2003, pp. 123–160.

[15] L. Ambrosio and P. Tilli, *Selected Topics on "Analysis in Metric Spaces"*, Scuola Normale Superiore, Pisa, 2000.

[16] A. Arnold and J.Dolbeault, *Refined convex Sobolev inequalities*, Tech. Report 431, Ceremade, 2004.

[17] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, *On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations*, Comm. Partial Differential Equations, 26 (2001), pp. 43–100.

[18] C. Baiocchi, *Discretization of evolution variational inequalities*, in Partial differential equations and the calculus of variations, Vol. I, F. Colombini, A. Marino, L. Modica, and S. Spagnolo, eds., Birkhäuser Boston, Boston, MA, 1989, pp. 59–92.

[19] C. Baiocchi and G. Savaré, *Singular perturbation and interpolation*, Math. Models Methods Appl. Sci., 4 (1994), pp. 557–570.

[20] E. J. Balder, *A general approach to lower semicontinuity and lower closure in optimal control theory*, SIAM J. Control Optim., 22 (1984), pp. 570–598.

[21] 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.

[22] P. Bénilan, *Solutions intégrales d'équations d'évolution dans un espace de Banach*, C. R. Acad. Sci. Paris Sér. A-B, 274 (1972), pp. A47–A50.

[23] H. Berliocchi and J.-M. Lasry, *Intégrandes normales et mesures paramétrées en calcul des variations*, Bull. Soc. Math. France, 101 (1973), pp. 129–184.

[24] P. M. Bleher, J. L. Lebowitz, and E. R. Speer, *Existence and positivity of solutions of a fourth-order nonlinear PDE describing interface fluctuations*, Comm. Pure Appl. Math., 47 (1994), pp. 923–942.

[25] V. I. Bogachev, *Gaussian measures*, vol. 62 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1998.

[26] V. I. Bogachev, G. Da Prato, and M. Röckner, *Existence of solutions to weak parabolic equations for measures*, Proc. London Math. Soc. (3), 88 (2004), pp. 753–774.

[27] H. Brezis, *Propriétés régularisantes de certains semi-groupes non linéaires*, Israel J. Math., 9 (1971), pp. 513–534.

[28] 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).

[29] H. Brézis, *Interpolation classes for monotone operators*, in Partial Differential Equations and Related Topics, J. A. Goldstein, ed., vol. 446 of Lecture Notes in Mathematics, Springer, Berlin, 1975, pp. 65–74.

[30] Y. Burago, M. Gromov, and G. Perel′man, *A. D. Aleksandrov spaces with curvatures bounded below*, Uspekhi Mat. Nauk, 47 (1992), pp. 3–51, 222.

[31] G. Buttazzo, *Semicontinuity, relaxation and integral representation in the calculus of variations*, vol. 207 of Pitman Research Notes in Mathematics Series, Longman Scientific & Technical, Harlow, 1989.

[32] M. Cáceres, J. Carrillo, and G.Toscani, *Long-time behavior for a nonlinear fourth order parabolic equation,*, Trans. Amer. Math. Soc., (2003).

[33] L. A. Caffarelli, M. Feldman, and R. J. McCann, *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs*, J. Amer. Math. Soc., 15 (2002), pp. 1–26 (electronic).

[34] E. Caglioti and C. Villani, *Homogeneous cooling states are not always good approximations to granular flows*, Arch. Ration. Mech. Anal., 163 (2002), pp. 329–343.

[35] E. A. Carlen and W. Gangbo, *Constrained steepest descent in the 2-Wasserstein metric*, Ann. of Math. (2), 157 (2003), pp. 807–846.

[36] ——, *Solution of a model Boltzmann equation via steepest descent in the 2-Wasserstein metric*, Arch. Ration. Mech. Anal., 172 (2004), pp. 21–64.

[37] J. A. Carrillo, R. J. McCann, and C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Rev. Mat. Iberoamericana, 19 (2003), pp. 971–1018.

[38] J. A. Carrillo, R. J. McCann, and C. Villani, *Contractions in the 2-Wasserstein space and thermalization of granular media*, To appear, (2004).

[39] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions*, Springer-Verlag, Berlin, 1977. Lecture Notes in Mathematics, Vol. 580.

[40] E. Chasseigne and J. L. Vazquez, *Theory of extended solutions for fast-diffusion equations in optimal classes of data. Radiation from singularities*, Arch. Ration. Mech. Anal., 164 (2002), pp. 133–187.

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

[42] P. Colli, *On some doubly nonlinear evolution equations in Banach spaces*, Japan J. Indust. Appl. Math., 9 (1992), pp. 181–203.

[43] P. Colli and A. Visintin, *On a class of doubly nonlinear evolution equations*, Comm. Partial Differential Equations, 15 (1990), pp. 737–756.

[44] D. Cordero-Erausquin, B. Nazaret, and C. Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities*, Adv. Math., 182 (2004), pp. 307–332.

[45] M. G. Crandall, *Nonlinear semigroups and evolution governed by accretive operators*, in Nonlinear Functional Analysis and its Applications, F. E. Browder, ed., vol. 45 of Proceedings of Symposia in Pure Mathematics, American Mathematical Society, Providence, 1986, pp. 305–338.

[46] M. G. Crandall and T. M. Liggett, *Generation of semi-groups of nonlinear transformations on general Banach spaces*, Amer. J. Math., 93 (1971), pp. 265–298.

[47] M. Csörnyei, *Aronszajn null and Gaussian null sets coincide*, Israel J. Math., 111 (1999), pp. 191–201.

[48] G. Da Prato and A. Lunardi, *Elliptic operators with unbounded drift coefficients and Neumann boundary condition*, J. Differential Equations, 198 (2004), pp. 35–52.

[49] G. Da Prato and J. Zabczyk, *Second order partial differential equations in Hilbert spaces*, vol. 293 of London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 2002.

[50] G. Dal Maso, *An Introduction to* Γ*-Convergence*, vol. 8 of Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser, Boston, 1993.

[51] E. De Giorgi, *New problems on minimizing movements*, in Boundary Value Problems for PDE and Applications, C. Baiocchi and J. L. Lions, eds., Masson, 1993, pp. 81–98.

[52] E. De Giorgi, A. Marino, and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8), 68 (1980), pp. 180–187.

[53] M. Degiovanni, A. Marino, and M. Tosques, *Evolution equations with lack of convexity*, Nonlinear Anal., 9 (1985), pp. 1401–1443.

[54] M. Del Pino, J. Dolbeault, and I. Gentil, *Nonlinear diffusions, hypercontractivity and the optimal* L^p*-Euclidean logarithmic Sobolev inequality*, J. Math. Anal. Appl., 293 (2004), pp. 375–388.

[55] C. Dellacherie and P.-A. Meyer, *Probabilities and potential*, vol. 29 of North-Holland Mathematics Studies, North-Holland Publishing Co., Amsterdam, 1978.

[56] H. Dietrich, *Zur c-Konvexität und c-Subdifferenzierbarkeit von Funktionalen*, Optimization, 19 (1988), pp. 355–371.

[57] R. J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math., 98 (1989), pp. 511–547.

[58] J. Dolbeault, D. Kinderlehrer, and M. Kowalczyk, *Remarks about the Flashing Rachet*, Tech. Report 406, Ceremade, 2004.

[59] R. M. Dudley, *Probabilities and metrics*, Matematisk Institut, Aarhus Universitet, Aarhus, 1976. Convergence of laws on metric spaces, with a view to statistical testing, Lecture Notes Series, No. 45.

[60] ——, *Real analysis and probability*, Wadswoth & Brooks/Cole, Pacific Grove, California, 1989.

[61] L. C. Evans, *Partial differential equations and Monge-Kantorovich mass transfer*, in Current developments in mathematics, 1997 (Cambridge, MA), Int. Press, Boston, MA, 1999, pp. 65–126.

[62] L. C. Evans and W. Gangbo, *Differential equations methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., 137 (1999), pp. viii+66.

[63] L. C. Evans, W. Gangbo, and O. Savin, *Diffeomorphisms and nonlinear heat flows*, (2004).

[64] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

[65] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.

[66] J. Feng and M. Katsoulakis, *A Hamilton Jacobi theory for controlled gradient flows in infinite dimensions,*, tech. report, 2003.

[67] B. Ferrario, G. Savaré, and L. Tubaro, *Fokker–Planck equation in infinite dimensional spaces.* In preparation.

[68] D. Feyel and A. S. Üstünel, *Measure transport on Wiener space and the Girsanov theorem*, C. R. Math. Acad. Sci. Paris, 334 (2002), pp. 1025–1028.

[69] D. Feyel and A. S. Üstünel, *Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space*, Probab. Theory Related Fields, 128 (2004), pp. 347–385.

[70] W. Gangbo, *The Monge mass transfer problem and its applications*, in Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997), vol. 226 of Contemp. Math., Amer. Math. Soc., Providence, RI, 1999, pp. 79–104.

[71] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Math., 177 (1996), pp. 113–161.

[72] R. Gardner, *The Brunn-Minkowski inequality*, Bull. Amer. Math. Soc., 39 (2002), pp. 355–405.

[73] L. Giacomelli and F. Otto, *Variational formulation for the lubrication approximation of the Hele-Shaw flow*, Calc. Var. Partial Differential Equations, 13 (2001), pp. 377–403.

[74] ——, *Rigorous lubrication approximation*, Interfaces Free Bound., 5 (2003), pp. 483–529.

[75] U. Gianazza, G. Toscani, and G. Savaré, *A fourth order parabolic equation and the Wasserstein distance*, tech. report, IMATI-CNR, Pavia, 2004. to appear.

[76] M. Giaquinta and S. Hildebrandt, *Calculus of Variations I*, vol. 310 of Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 1996.

[77] M. Giaquinta, G. Modica, and J. Souček, *Area and the area formula*, in Proceedings of the Second International Conference on Partial Differential Equations (Italian) (Milan, 1992), vol. 62, 1992, pp. 53–87 (1994).

[78] K. Glasner, *A diffuse interface approach to Hele-Shaw flow*, Nonlinearity, 16 (2003), pp. 49–66.

[79] C. Goffman and J. Serrin, *Sublinear functions of measures and variational integrals*, Duke Math. J., 31 (1964), pp. 159–178.

[80] P. Hajłasz, *Sobolev spaces on an arbitrary metric space*, Potential Anal., 5 (1996), pp. 403–415.

[81] J. Heinonen and P. Koskela, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math., 181 (1998), pp. 1–61.

[82] C. Huang and R. Jordan, *Variational formulations for Vlasov-Poisson-Fokker-Planck systems*, Math. Methods Appl. Sci., 23 (2000), pp. 803–843.

[83] 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).

[84] J. Jost, *Nonpositive curvature: geometric and analytic aspects*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1997.

[85] ———, *Nonlinear Dirichlet forms*, in New directions in Dirichlet forms, vol. 8 of AMS/IP Stud. Adv. Math., Amer. Math. Soc., Providence, RI, 1998, pp. 1–47.

[86] A. Jüngel and R. Pinnau, *Global nonnegative solutions of a nonlinear fourth-order parabolic equation for quantum systems*, SIAM J. Math. Anal., 32 (2000), pp. 760–777 (electronic).

[87] A. Jüngel and G. Toscani, *Decay rates of solutions to a nonlinear fourth–order parabolic equation*, Z. Angew. Math. Phys., 54 (2003), pp. 377–386.

[88] D. Kinderlehrer and N. J. Walkington, *Approximation of parabolic equations using the Wasserstein metric*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 837–852.

[89] A. V. Kolesnikov, *Convexity inequalities and optimal transport of infinite dimensional measures*, (2004).

[90] A. J. Kruger and B. Š. Morduhovič, *Extremal points and the Euler equation in nonsmooth optimization problems*, Dokl. Akad. Nauk BSSR, 24 (1980), pp. 684–687, 763.

[91] L. LeCam, *Convergence in distribution of stochastic processes*, Univ. Calif. Publ. Statist., 2 (1957), pp. 207–236.

[92] V. L. Levin and S. T. Rachev, *New duality theorems for marginal problems with some applications in stochastics*, vol. 1412 of Lecture Notes in Math., Springer, Berlin, 1989.

[93] S. Luckhaus, *Solutions for the two-phase Stefan problem with the Gibbs-Thomson Law for the melting temperature*, Euro. Jnl. of Applied Mathematics, 1 (1990), pp. 101–111.

[94] S. Luckhaus and T. Sturzenhecker, *Implicit time discretization for the mean curvature flow equation*, Calc. Var. Partial Differential Equations, 3 (1995), pp. 253–271.

[95] A. Marino, C. Saccon, and M. Tosques, *Curves of maximal slope and parabolic variational inequalities on nonconvex constraints*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), 16 (1989), pp. 281–330.

[96] U. F. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Comm. Anal. Geom., 6 (1998), pp. 199–253.

[97] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.

[98] A. Mielke, F. Theil, and V. I. Levitas, *A variational formulation of rate-independent phase transformations using an extremum principle*, Arch. Ration. Mech. Anal., 162 (2002), pp. 137–177.

[99] T. Mikami, *Dynamical systems in the variational formulation of the Fokker-Planck equation by the Wasserstein metric*, Appl. Math. Optim., 42 (2000), pp. 203–227.

[100] B. S. Mordukhovich, *Nonsmooth analysis with nonconvex generalized differentials and conjugate mappings*, Dokl. Akad. Nauk BSSR, 28 (1984), pp. 976–979.

[101] R. Nochetto and G. Savaré, *Nonlinear evolution governed by accretive operators: Error control and applications*, to appear, (2003).

[102] R. H. Nochetto, G. Savaré, and C. Verdi, *A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations*, Comm. Pure Appl. Math., 53 (2000), pp. 525–589.

[103] F. Otto, *Doubly degenerate diffusion equations as steepest descent*, Manuscript, (1996).

[104] ——, *Dynamics of labyrinthine pattern formation in magnetic fluids: a mean-field theory*, Arch. Rational Mech. Anal., 141 (1998), pp. 63–103.

[105] ——, *Lubrication approximation with prescribed nonzero contact angle*, Comm. Partial Differential Equations, 23 (1998), pp. 2077–2164.

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

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

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

[109] K. R. Parthasarathy, *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967.

[110] M. Pierre, *Uniqueness of the solutions of* $u_t - \Delta\varphi(u) = 0$ *with initial datum a measure*, Nonlinear Anal., 6 (1982), pp. 175–187.

[111] A. Pratelli, *On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation*, To appear, (2004).

[112] S. T. Rachev and L. Rüschendorf, *Mass transportation problems. Vol. I*, Probability and its Applications, Springer-Verlag, New York, 1998. Theory.

[113] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, Springer-Verlag, Berlin, 1998.

[114] R. Rossi and G. Savaré, *Gradient flows of non convex functionals in Hilbert spaces and applications*, tech. report, IMATI-CNR, Pavia, 2004.

[115] J. Rulla, *Error analysis for implicit approximations to solutions to Cauchy problems*, SIAM J. Numer. Anal., 33 (1996), pp. 68–87.

[116] L. Rüschendorf, *On c-optimal random variables*, Statist. Probab. Lett., 27 (1996), pp. 267–270.

[117] L. Schwartz, *Radon measures on arbitrary topological spaces and cylindrical measures*, Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, London, 1973. Tata Institute of Fundamental Research Studies in Mathematics, No. 6.

[118] C. Sparber, J. A. Carrillo, J. Dolbeault, and P. A. Markowich, *On the long-time behavior of the quantum Fokker-Planck equation*, Monatsh. Math., 141 (2004), pp. 237–257.

[119] D. W. Stroock, *Probability theory, an analytic view*, Cambridge University Press, Cambridge, 1993.

[120] K. T. Sturm, *Metric spaces of lower bounded curvature*, Exposition. Math., 17 (1999), pp. 35–47.

[121] V. N. Sudakov, *Geometric problems in the theory of infinite-dimensional probability distributions*, Proc. Steklov Inst. Math., (1979), pp. i–v, 1–178. Cover to cover translation of Trudy Mat. Inst. Steklov **141** (1976).

[122] N. S. Trudinger and X.-J. Wang, *On the Monge mass transfer problem*, Calc. Var. Partial Differential Equations, 13 (2001), pp. 19–31.

[123] M. Valadier, *Young measures*, in Methods of nonconvex analysis (Varenna, 1989), Springer, Berlin, 1990, pp. 152–188.

[124] L. N. Vasershtein, *Markov processes over denumerable products of spaces describing large system of automata*, Problemy Peredači Informacii, 5 (1969), pp. 64–72.

[125] C. Villani, *Optimal transportation, dissipative PDE's and functional inequalities*, in Optimal transportation and applications (Martina Franca, 2001), vol. 1813 of Lecture Notes in Math., Springer, Berlin, 2003, pp. 53–89.

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

[127] A. Visintin, *Models of Phase Transitions*, vol. 28 of Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser, Boston, 1996.

[128] L. C. Young, *Generalized cuves and the existence of an attained absolute minimum in the calculus of variations*, Comptes Rendus de la Soccieté des Sciences et des Lettres de Varsovie (classe III), 30 (1937), pp. 212–234.

[129] ——, *Lectures on the calculus of variations and optimal control theory*, Foreword by Wendell H. Fleming, W. B. Saunders Co., Philadelphia, 1969.

Index