TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

Gradient Flows in General Metric and Wasserstein Spaces

Ausgeführt am Institut für
Diskrete Mathematik und Geometrie
der Technischen Universität Wien

unter der Anleitung von
Univ.Prof. Dipl.-Ing. Dr.techn. Monika Ludwig
und
Senior Lecturer Dipl.-Ing. Dr.techn. Gabriel Maresch
als verantwortlich mitwirkenden Universitätsassistenten

durch

Dominik Leopold Forkert, BSc
Grüntal 4
3412 Kierling

Datum

Unterschrift

Preface

In this thesis we present *gradient flows* on a topological space X. In case of Euclidean space $\mathbb{R}^n$, a gradient flow admits a simple characterization via the differential equation $v' = -\nabla F(v)$ for a given potential $F : \mathbb{R}^n \to \mathbb{R}$. It is well known that under certain regularity assumptions on F for every initial value v_0, there exists a unique solution v of this equation. The family of integral curves $(v(t))_{v(0)\in\mathbb{R}^n}$ is then called a gradient flow with respect to F.

It is not to difficult to generalize this concept to Banach spaces. Indeed, one can formulate the characterizing differential equation by means of the Fréchet derivative. On the other hand, the situation severely complicates if X merely is a metric space. A priori the notion of a derivative or a gradient makes no sense in such a setting. However, one may start with $X = \mathbb{R}^n$ and assume that the potential F is convex. One obtains the following equivalent characterization of a gradient flow:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|v(t)-x|^2 \le F(x) - F\big(v(t)\big) \qquad \forall x \in \mathbb{R}^n,\ \forall t \in \mathbb{R}^+. \tag{1}$$

The inequality stated above is a special case of the so called *evolution variational inequality (EVI)*. Actually, it is straightforward to interpret the expressions in (1) in a general metric setting. Thus, one arrives at

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}d^2(v(t),x) \le F(x) - F\big(v(t)\big) \qquad \forall x \in X,\ \forall t \in \mathbb{R}^+. \tag{2}$$

We want to study properties of curves v, satisfying (2); in particular, we deal with uniqueness and existence results. Hence, inequality (2) (in a slightly more general form) plays a major role throughout most parts of this paper.

The first chapter is devoted to gradient flows on general metric spaces. We start with the development of tools, required for the formulation of weaker variants of the EVI characterization. To this aim, we introduce the *metric differential* due to Kirchheim [34], and the notion of *slopes* which can be seen as metric equivalent of the usual gradient.

We present three variants of gradient flows in metric spaces, including the aforementioned EVI. Although all three definitions agree on Euclidean space, it turns out, that they are not equivalent on arbitrary metric space. We show that EVI is strongest. Moreover, gradient flows in the EVI sense possess a strong contraction property, which implies that there exists at most one gradient flow with respect to a given functional.

Nevertheless, an existence theory – involving deep variational interpolation arguments – may not easily be archived at this level of generality. Therefore, we content ourselves by presenting only principal definitions of the so-called *minimizing movements scheme* which provides a variational interpolation technique for gradient flows in the metric setting. We conclude this chapter with a brief overview of convergence results of the minimizing movements.

In the second chapter we approach the theory of *optimal* transport. We focus on the *Kantorovich problem* which is a relaxation of the *Monge problem*. Both problems arise in various guises and are formulated in a measure theoretic context. However, in its simplest form, the Kantorovich problem reduces to *linear programming*. This allows the very intuitive interpretation of transports by means of e.g. delivering goods from factories to stores along the most cost effective routes.

We show that the Kantorovich problem (unlike the Monge problem) admits an optimal solution under very general assumptions. As a result, one may utilize the theory of optimal transportation to define a family of metrics on a certain subset in the class of probability measures on a metric space. These metrics are known as *Wasserstein distances* – a distorted transliteration of Vaserstein who used one of these metrics in the paper [45]. Although the denotation *Kantorovich distances*, mainly used in the Russian literature, seems to be more appropriate.

The resulting family of metric spaces equipped with the aforementioned distances – called *Wasserstein spaces* – inherits to a great extend the topological and geometrical structure of the underlying metric space. Therefore, it seems reasonable to investigate gradient flows in these spaces. The last part of this paper is devoted to the existence of EVI gradient flows in a Wasserstein space over $\mathbb{R}^n$. The proof of the existence result provided here relies on rather strong assumptions and provides only a suboptimal convergence rate of the corresponding minimizing movements scheme. Nevertheless, the class of admissible functionals is sufficiently large for comprehensive application within the theory of partial differential equations. In a concluding example we show the main ideas applied to the heat equation.

Two appendices provide the required background in measure theory as well as an elementary introduction to flows.

Bibliographical Notes

The main reference for gradient flows in metric spaces is the landmark monograph [3] by Ambrosio, Gigli, Savaré. So far, it provides the most comprehensive treatise of the theory in a general metric setting. Most of the content of Sections 1–2 is taken from this book.

The introduction to gradient flows in Euclidean space in Section 1.3 may be found in a more condensed form in the lecture notes [26] by Daneri, Savaré. In the second part of this section as well as in Section 1.4 we follow the lines of Ambrosio, Gigli [2] – with the exception of Proposition 1.4.1, where we elaborate a sketch by Daneri, Savaré [26], incorporating a *doubling of variables* argument by Villani (cf. Appendix A, Chapter 23 in [47]).

The notion of *minimizing movements* in Section 1.5 goes back to De Giorgi [27]. In the context of Wasserstein spaces it has been first used by Jordan, Kinderlehrer, Otto in their seminal paper [31]. Mayer obtained first existence results of gradient flows in *non-positively curved (NPC) metric spaces* in [39]. The existence theorem by Ambrosio, Gigli, Savaré [3], found in Section 1.5, is a generalization of the former, which allows application in the Wasserstein case. A similar result, yielding suboptimal convergence estimates by a simpler argument, has been obtained by Clément, Desch [21].

Thorough references for the theory of *optimal transport* which is discussed in the second part of this text, are the extensive monographs [46], [47] by Villani. However, only *Brennier's theorem* (see Remark 2.1.6) is taken from the former, and Proposition 2.2.4 is taken the latter monograph. Otherwise, Sections 2.1–2.2 follow the lines of the more recent survey article [13] by Bogachev, Kolesnikov – with two exceptions: The simple proof of the *Kantorovich-Rubinstein theorem* is due to D. Edwards [29], and the proof of Proposition 2.2.6 follows an elementary argument by Bolley [14] (also found in [47]).

Example 2.3.2 in Section 2.3 is taken from [2]. The proof of the main existence result in this section follows a sketch in [4] by Ambrosio, Savaré; in particular, the last step of this proof uses an argument by Clément, Desch [20]. A similar existence result, yielding the same suboptimal convergence estimates by a different way, can be found in the recent paper [23] by Craig.

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisors Monika Ludwig and Gabriel Maresch for their continuous support – specifically in moments of being short of time. It should not go unmentioned that it was Gabriel who introduced me to the fascinating subject of optimal transport and its plenty of applications. Thank You!

Finally, big thanks go out to my parents who encouraged me over the last years and without whom all this would not have been possible.

Table of Contents

1 Gradient Flows in General Metric Spaces

1.1 Absolutely Continuous Curves and the Metric Derivative

We will start with a definition of absolutely continuous curves taking values in a complete metric space (which may lack the additive structure of a vector space).

Notation

In this chapter, by (a,b) we denote a possibly unbounded interval on $\mathbb{R}$ and by (X,d) an arbitrary complete metric space.

1.1.1 **Definition** (Absolutely continuous curves) Given a curve $v : (a,b) \to X$, we say that v belongs to $AC^p((a,b),X)$, $1 \le p \le \infty$, if there exists a function $m \in L^p((a,b),\mathbb{R})$ such that

$$d(v(s),v(t)) \le \int_s^t m \,\mathrm{d}\lambda \qquad \forall (s,t) : a < s \le t < b. \tag{1.1}$$

We say that v belongs to $AC^p_{\text{loc}}((a,b),X)$, $1 \le p \le \infty$, if for every $t \in (a,b)$ there exists an interval $(\acute{a},\acute{b}) \in \mathfrak{U}(x)$ such that $v|_U \in AC^p\big((\acute{a},\acute{b}),X\big)$.

In the case $p = 1$ we say that v is ***absolutely continuous*** or ***locally absolutely continuous*** and simply write $AC((a,b),X)$ or $AC_{\text{loc}}((a,b),X)$ for the corresponding space, instead of $AC^1((a,b),X)$ or $AC^1_{\text{loc}}((a,b),X)$, respectively.

The following lemma is well known from measure theory (see a.e. Theorem 2.5.7 in [12]) and will be particularly useful in the proofs of the next results.

1.1.2 **Lemma** (Absolute continuity of the integral) *Let $(X,\mathcal{A},\mu)$ be a measure space and consider a function $v \in L^p(X,\mu,\mathbb{R})$ for $p \in [1,+\infty]$. Then, for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\int_D |v| \,\mathrm{d}\lambda < \varepsilon \qquad \forall D \in \mathcal{A} : \lambda(D) < \delta.$$

Proof For every $\varepsilon > 0$, there exists a simple function $\phi(x) = \sum_{k=1}^n a_k \mathbb{1}_{A_k}(x)$ with measurable sets $A_k \subseteq X$ and numbers $a_k \in \mathbb{R}$ for all $k \le n$, such that $\|v - \phi\|_p < \varepsilon$.

As a result, by choosing $D \in \mathcal{A}$ with $\lambda(D) < \min\left\{\frac{\varepsilon}{\|\phi\|_\infty}, 1\right\}$ and applying Hölder's inequality with $q \in [1,\infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we obtain

$$\int_D |v| \,\mathrm{d}\lambda \le \int_D |v - \phi| \,\mathrm{d}\lambda + \int_D |\phi| \,\mathrm{d}\lambda < \lambda(D)\,\|v - \phi\|_p + \lambda(D)\,\|\phi\|_\infty < 2\varepsilon.$$

■

It is yet not clear whether the definition given above is equivalent to Definition A.2.1 for real-valued curves in Appendix A. The following simple proposition gives answer to this question.

1.1.3 **Proposition** *Let a real-valued curve $v : (a,b) \to \mathbb{R}$ be given. Then v belongs to $AC((a,b),\mathbb{R})$ iff for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\sum_{i=1}^n |v(b_i) - v(a_i)| < \varepsilon \tag{1.2}$$

for every finite collection of pairwise disjoint intervals $(a_i,b_i) \subseteq (a,b)$ with $\sum_{i=1}^n |b_i - a_i| < \delta$.

Proof Assume that v satisfies (1.2) for every finite collection of pairwise disjoint intervals. Due to Theorem A.2.4, this is equivalent to the existence of a function $m \in L^1((a,b),\mathbb{R})$ such that

$$v(a) - v(x) = \int_a^x m \, \mathrm{d}\lambda \qquad \forall x \in [a, b].$$

Hence, we have

$$|v(t) - v(s)| = \Big|\int_s^t m \, \mathrm{d}\lambda\Big| \leq \int_s^t |m| \, \mathrm{d}\lambda < \infty \qquad \forall (s, t) \subseteq (a, b).$$

Conversely, assume $v \in AC((a, b), \mathbb{R})$ and let $\varepsilon > 0$ be given. According to Definition 1.1.1, there exists $m \in L^1((a, b), \mathbb{R})$ such that (1.1) holds. Note that $m(x) \geq 0$ λ-a.e. in (a, b).
Due to Lemma 1.1.2, we can choose $\delta > 0$ such that

$$\sum_{i=1}^n \int_{a_i}^{b_i} m \, \mathrm{d}\lambda = \int_{\bigcup_{i=1}^n (a_i, b_i)} m \, \mathrm{d}\lambda < \varepsilon$$

for every finite collection of pairwise disjoint intervals $(a_i, b_i) \subset (a, b)$ with

$$\lambda\Big(\bigcup_{i=1}^n (a_i, b_i)\Big) = \sum_{i=1}^n |b_i - a_i| < \delta.$$

We conclude

$$\sum_{i=1}^n |v(b_i) - v(a_i)| \leq \sum_{i=1}^n \int_{a_i}^{b_i} m \, \mathrm{d}\lambda < \varepsilon.$$

■

1.1.4 Facts Consider a curve $v \in AC^p((a, b), X)$.

(i) Note that for every $\varepsilon > 0$, there exits $\delta > 0$ such that

$$d\,(v(s), v(t)) \leq \int_s^t m \, \mathrm{d}\lambda = \int_s^t |m| \, \mathrm{d}\lambda < \varepsilon \qquad \forall s, t \in (a, b) : |s - t| < \delta,$$

due to Lemma 1.1.2. Hence, v is uniformly continuous on (a, b).

(ii) Consider the case when the interval (a, b) is bounded. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in (a, b) such that $\lim_{n\to\infty} a_n = a$. Since v is uniformly continuous on (a, b), $(v(a_n))_{n\in\mathbb{N}}$ inherits the Cauchy property from $(a_n)_{n\in\mathbb{N}}$. Hence, the limit $\lim_{n\to\infty} v(a_n) \in X$ exists due to the completeness of (X, d).
We conclude that the right-sided limit $\lim_{x\searrow a} v(x)$, and similarly the left-sided limit $\lim_{x\nearrow b} v(x)$ exist in (X, d).

(iii) Assume, that $1 < p \leq \infty$. By applying Hölder's inequality with $1 \leq q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we obtain the estimate

$$d(v(s), v(t)) \leq \int_s^t m \, \mathrm{d}\lambda = \int_s^t |m| \, \mathrm{d}\lambda \leq \lambda((s, t)) \, \|m\|_p = \|m\|_p \cdot |s - t| \qquad \forall s, t \in (a, b).$$

Hence, v is Lipschitz continuous with a Lipschitz constant $C_{Lip} = \|m\|_p$.

Although metric spaces lack the linear structure of vector spaces, it is possible to define a certain generalization of a derivative of functions taking values in arbitrary metric spaces.

1.1.5 Definition (Metric differential) A function $f : (a,b) \to X$ is said to be ***metrically differentiable*** at a point $t \in (a,b)$ if the limit

$$|f'|(t) := \lim_{s\to t} \frac{d(f(s),f(t))}{|s-t|} \tag{1.3}$$

does exist. Then $|f'|(t) \in \mathbb{R}$ is called the ***metric differential*** or ***metric derivative*** of f at t.

1.1.6 Example Consider a function $f : (a,b) \to Y$ where $(Y, \|\cdot\|_Y)$ is a Banach space. Then f is metrically differentiable at a point t if f is Fréchet differentiable at t, since

$$\|\mathrm{d}f(t)\|_Y = \left\|\lim_{s\to t} \frac{f(s)-f(t)}{s-t}\right\|_Y = \lim_{s\to t} \frac{\|(f(s)-f(t)\|_Y}{|s-t|} = |f'|(t).$$

◀

Concerning the following theorem, recall that the *limit inferior* of a function $\phi : X \to \mathbb{R} \cup \{+\infty\}$ at a point $x \in X$ is defined as $\liminf_{y\to x} \phi(y) := \sup_{U\in\mathfrak{U}(x)} \inf \phi(U)$, where $\mathfrak{U}(x)$ denotes the neighborhood filter of x.

1.1.7 Theorem *For any curve $v \in AC^p((a,b),X)$ the metric differential $|v'|(t)$ exists λ-a.e. in (a,b) and satisfies the following properties:*

(MD1) *The function $|v'|$ belongs to $L^p((a,b),\mathbb{R})$.*

(MD2) *$|v'|$ is an admissible integrand for the right-hand side of* (1.1.1).

(MD3) *The metric differential is minimal in the following sense that $|v'|(t) \le m(t)$ λ-a.e. in (a,b), for each function $m \in L^p((a,b),\mathbb{R})$ satisfying* (1.1.1).

Proof As $\mathbb{R}$ is separable, there exists a countable set $(y_n)_{n\in\mathbb{N}}$ which is dense in the interval (a,b). Furthermore, $(v(y_n))_{n\in\mathbb{N}}$ is also dense in $\operatorname{ran} v$, due to continuity of v.
Define functions $d\!\!\!/_n(t) := d(y_n, v(t))$ for every $n \in \mathbb{N}$. Since we have

$$|d\!\!\!/_n(s) - d\!\!\!/_n(t)| = |d(y_n, v(s)) - d(y_n, v(t))| \le d(v(s), v(t)) \le \int_s^t m \,\mathrm{d}\lambda \qquad \forall s,t \in (a,b),$$

each $d\!\!\!/_n$ is absolutely continuous in (a,b). Note that each $d\!\!\!/_n$ is also an absolutely continuous real-valued function in the sense of Definition A.2.1, due to Proposition 1.1.3. Henceforth, we can apply Fact A.2.5.i to assure that each derivative $d\!\!\!/_n'(t)$ is well-defined λ-a.e. in (a,b).
As the union of countable many λ-null sets in (a,b) is again a λ-null set, the function

$$\begin{aligned} d\!\!\!/ : (a,b) &\longrightarrow \mathbb{R}_0^+ \cup \{\infty\} \\ t &\longmapsto \sup_{n\in\mathbb{N}} |d\!\!\!/_n'(t)| \end{aligned}$$

is finite λ-a.e. in (a,b).
Now, let $t \in (a,b)$ be a point where all $d\!\!\!/_n$ are differentiable. Notice that, since $d(v(s),v(t)) \ge |d\!\!\!/_n(s) - d\!\!\!/(t)|$ for all $n \in \mathbb{N}$, we have

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

Next, consider a function $m \in L^p((a,b),X)$ which satisfies (1.1) and choose $c \in (a,t)$. Together with Theorem A.2.6, the inequality above shows

$$\bar{d}(t) \leq \liminf_{s\to t} \frac{d(v(s),v(t))}{|s-t|} \leq \liminf_{s\to t} \frac{1}{|s-t|} \int_s^t m\, \mathrm{d}\lambda \leq \tag{1.4}$$

$$\leq \lim_{s\to t} \frac{1}{s-t} \Big(\int_c^t m\, \mathrm{d}\lambda - \int_c^s m\, \mathrm{d}\lambda \Big) = \frac{\mathrm{d}}{\mathrm{d}t} \int_c^t m\, \mathrm{d}\lambda = m(t) \qquad \lambda\text{-a.e. in } (a,b).$$

Hence, we have $\bar{d}(t) \leq m(t)$ λ-a.e. in (a,b), which shows (MD2). Moreover, it is clear that $\bar{d}$ is λ-measurable on (a,b) and we infer $\bar{d} \in L^p((a,b),\mathbb{R})$.

On the other hand, for every $t \in (a,b)$ there exists a subsequence $(y_{n_k})_{k\in\mathbb{N}}$ of $(y_n)_{n\in\mathbb{N}}$ such that $\lim_{k\to\infty} y_{n_k} = t$. Therefore, we obtain

$$d(v(s),v(t)) = \lim_{k\to\infty} \big| d(v(y_{n_k}),v(s)) - d(v(y_{n_k}),v(t)) \big| \qquad \forall s \in (a,b).$$

Together with the reverse triangle inequality $d(v(s),v(t)) \geq |d(v(z),v(s)) - d(v(z),v(t))|$ for all $s,t,z \in (a,b)$, this yields

$$d(v(s),v(t)) = \lim_{k\to\infty} \big| d(v(y_{n_k}),v(s)) - d(v(y_{n_k}),v(t)) \big| = \tag{1.5}$$

$$= \sup_{n\in\mathbb{N}} |d(v(y_n),v(s)) - d(v(y_n),v(t))| = \sup_{n\in\mathbb{N}} |\bar{d}_n(s) - \bar{d}_n(t)| \qquad \forall s,t \in (a,b).$$

Moreover, Theorem A.2.4 and Fact A.2.5.i show that

$$\sup_{n\in\mathbb{N}} |\bar{d}_n(s) - \bar{d}_n(t)| \leq \sup_{n\in\mathbb{N}} \Big| \int_s^t \bar{d}_n'\, \mathrm{d}\lambda \Big| \leq \int_s^t \bar{d}\, \mathrm{d}\lambda \qquad \forall s,t \in (a,b) : s \leq t.$$

Thus, with Theorem A.2.6 and any $c \in (a,t)$ we obtain

$$\limsup_{s\to t} \frac{d(v(s),v(t))}{|s-t|} \leq \limsup_{s\to t} \frac{1}{|s-t|} \int_s^t \bar{d}\, \mathrm{d}\lambda \leq \tag{1.6}$$

$$\leq \lim_{s\to t} \frac{1}{s-t} \Big(\int_c^t \bar{d}\, \mathrm{d}\lambda - \int_c^s \bar{d}\, \mathrm{d}\lambda \Big) = \frac{\mathrm{d}}{\mathrm{d}t} \int_c^t \bar{d}\, \mathrm{d}\lambda = \bar{d}(t) \qquad \lambda\text{-a.e. in } (a,b).$$

Both (1.4) and (1.6) imply that

$$\lim_{s\to t} \frac{d(v(s),v(t))}{|s-t|} = \bar{d}(t) \qquad \lambda\text{-a.e. in } (a,b).$$

Therefore, we have shown (MD1), as well as (MD3). ■

1.1.8 Lemma (Lipschitz and arc-length reparametrization) *Let $v \in AC((a,b),X)$ be an absolutely continuous curve with length $L := \int_a^b |v'|\, \mathrm{d}\lambda$.*

(i) *For every $\varepsilon > 0$ and $L_\varepsilon := L + \varepsilon(b-a)$ there exists a strictly increasing, absolutely continuous map*

$$\varsigma_\varepsilon : (a,b) \to (0,L_\varepsilon) \quad \textit{with} \quad \lim_{t\searrow a} \varsigma_\varepsilon(t) = 0 \quad \textit{and} \quad \lim_{t\nearrow b} \varsigma_\varepsilon(t) = L_\varepsilon,$$

and a Lipschitz curve

$$\hat{v}_\varepsilon : (0,L_\varepsilon) \to X, \quad \textit{such that} \quad v = \hat{v}_\varepsilon \quad \textit{and} \quad |\hat{v}_\varepsilon'| \circ \varsigma_\varepsilon = \frac{|v'|}{\varepsilon + |v'|} \in L^\infty((a,b),\mathbb{R}). \tag{1.7}$$

Moreover, the map ς_ε admits a Lipschitz continuous inverse $\tau_\varepsilon : (0,L_\varepsilon) \to (a,b)$ with a Lipschitz constant ε^{-1} such that $\hat{v}_\varepsilon = v \circ \tau_\varepsilon$.

(ii) *There exists an increasing, absolutely continuous map*

$$\varsigma : (a,b) \to [0,L] \quad \textit{with} \quad \lim_{t\searrow a} \varsigma(t) = 0 \quad \textit{and} \quad \lim_{t\nearrow b} \varsigma(t) = L,$$

and a Lipschitz curve

$$\hat{v} : [0,L] \to X, \quad \textit{such that} \quad v = \hat{v} \circ \varsigma \quad \textit{and} \quad |\hat{v}'| = 1 \ \lambda\textit{-a.e. in } [0,L]. \tag{1.8}$$

Proof First, we will show (i) and consider $\varepsilon > 0$. Define the map $\varsigma_\varepsilon : (a,b) \to \mathbb{R}$ with

$$\varsigma_\varepsilon(t) := \int_a^t \varepsilon + |v'| \, \mathrm{d}\lambda. \tag{1.9}$$

Since $(\varepsilon + |v'|) \in L^1((a,b),\mathbb{R})$, we infer by Theorem A.2.4 that ς_ε is absolutely continuous on (a,b). Thus, we have $\operatorname{ran} \varsigma_\varepsilon = (0, L_\varepsilon)$ and the estimate $\varsigma_\varepsilon' = \varepsilon + |v'| \geq \varepsilon$ λ-a.e. in (a,b) holds. Moreover, ς_ε is strictly increasing by monotonicity of the integral.
As a result, ς_ε is a bijection and admits an inverse map $\tau_\varepsilon : (0,L_\varepsilon) \to (a,b)$. Via the estimate

$$|\varsigma_\varepsilon(t) - \varsigma_\varepsilon(s)| = \int_s^t \varepsilon + |v'| \, \mathrm{d}\lambda \geq \varepsilon\, |t-s| \qquad \forall s,t \in (a,b) : s \leq t$$

we obtain

$$|\varsigma_\varepsilon \circ \tau_\varepsilon(x) - \varsigma_\varepsilon \circ \tau_\varepsilon(y)| = |x-y| \geq \varepsilon\, |\tau_\varepsilon(x) - \tau_\varepsilon(y)| \qquad \forall x,y \in (0,L_\varepsilon).$$

Hence, τ_ε is Lipschitz continuous with ε^{-1} as a Lipschitz constant. Therefore, τ_ε has a finite derivative λ-a.e. in $(0,L_\varepsilon)$, due to Example A.2.11 and Fact A.2.5.i.
Consider a point $t \in (a,b)$ such that both ς_ε and τ_ε are differentiable at t and $\varsigma_\varepsilon(t)$, respectively. Applying the chain rule of differentiation to the expression $t = \tau_\varepsilon \circ \varsigma_\varepsilon(t)$ gives us

$$1 = \tau_\varepsilon' \circ \varsigma_\varepsilon(x) \cdot \varsigma_\varepsilon'(x) = \tau_\varepsilon' \circ \varsigma_\varepsilon(x) \cdot (\varepsilon + |v'|(x)) \,.$$

This establishes

$$\tau_\varepsilon' \circ \varsigma_\varepsilon = \frac{1}{\varepsilon + |v'|} \qquad \lambda\text{-a.e. in } (a,b). \tag{1.10}$$

Now define the mapping $\hat{v}_\varepsilon : (0,L_\varepsilon) \to X$ via $\hat{v}_\varepsilon := v \circ \tau_\varepsilon$. For every pair of points $s = \tau_\varepsilon(x)$, $t = \tau_\varepsilon(y)$ in (a,b) with $0 < x < y < L_\varepsilon$ we have $s < t$ since τ_ε inherits the strict monotonicity from ς_ε. Thus,

$$d\left(\hat{v}_\varepsilon(x), \hat{v}_\varepsilon(y)\right) = d\left(v(s), v(t)\right) \leq \int_s^t |v'| \, \mathrm{d}\lambda \leq \tag{1.11.a}$$

$$\leq \varsigma_\varepsilon(t) - \varsigma_\varepsilon(s) - \varepsilon(t-s) = y - x - \varepsilon(t-s) < y - x. \tag{1.11.b}$$

Consequently, $\hat{v}_\varepsilon$ is Lipschitz continuous with Lipschitz constant 1. Moreover, $\hat{v}_\varepsilon$ can be extended to $[0,L_\varepsilon]$ as $\lim_{x\searrow 0} \hat{v}_\varepsilon(x) = \lim_{t\searrow a} v(t)$ and $\lim_{x\nearrow L_\varepsilon} \hat{v}_\varepsilon(x) = \lim_{t\nearrow b} v(t)$, whereas the one-sided limits of v exist, due to Fact 1.1.4.ii.
Regarding (1.11) we obtain

$$\frac{d\left(\hat{v}_\varepsilon(y), \hat{v}_\varepsilon(x)\right)}{y-x} \leq 1 - \varepsilon\frac{t-s}{y-x} = 1 - \varepsilon\frac{\tau_\varepsilon(y) - \tau_\varepsilon(x)}{y-x}.$$

Passing to the limit $(y \to x)$ in the equation above, we get $|\hat{v}_\varepsilon'| \leq 1 - \varepsilon\tau_\varepsilon'$. Moreover, by virtue of (1.10), we gain the estimate

$$|\hat{v}_\varepsilon'| \circ \varsigma_\varepsilon \leq 1 - \frac{\varepsilon}{\varepsilon + |v'|} = \frac{|v'|}{\varepsilon + |v'|} \qquad \lambda\text{-a.e. in } (a,b).$$

Conversely, due to *change of variables* (see Corollary A.2.8 in Appendix A) as well as the relations $x = \varsigma_\varepsilon(s)$ and $y = \varsigma_\varepsilon(t)$, one has

$$d\left(v(s), v(t)\right) = d\left(\hat{v}_\varepsilon(x), \hat{v}_\varepsilon(y)\right) \leq \int_{\varsigma_\varepsilon(s)}^{\varsigma_\varepsilon(t)} |\hat{v}_\varepsilon'| \,\mathrm{d}\lambda = \int_x^y |\hat{v}_\varepsilon'| \circ \varsigma_\varepsilon \cdot \varsigma_\varepsilon' \,\mathrm{d}\lambda = \int_s^t |\hat{v}_\varepsilon'| \circ \varsigma_\varepsilon \cdot (\varepsilon + |v'|) \,\mathrm{d}\lambda \tag{1.12}$$

Henceforth, by choosing $c \in (a,t)$, we infer

$$|v'|(t) = \lim_{s \to t} \frac{d\left(v(s), v(t)\right)}{|s-t|} \leq \lim_{s \to t} \frac{1}{t-s} \Big(\int_c^t |\hat{v}_\varepsilon'| \circ \varsigma_\varepsilon \cdot (\varepsilon + |v'|) \,\mathrm{d}\lambda - \tag{1.13.a}$$

$$- \int_c^s |\hat{v}_\varepsilon'| \circ \varsigma_\varepsilon \cdot (\varepsilon + |v'|) \,\mathrm{d}\lambda \Big) = |\hat{v}_\varepsilon'| \circ \varsigma_\varepsilon(t) \cdot (\varepsilon + |v'|(t)) \tag{1.13.b}$$

for λ-a.e. $t \in (a,b)$. Together, (1.11) and (1.13) yield the second equality in (1.7).

Now we will show (ii): Analogously to (1.9), we define the map $\varsigma : (a,b) \to [0,L]$ with

$$\varsigma(t) := \int_a^t |v'| \,\mathrm{d}\lambda. \tag{1.14}$$

Clearly, ς is an increasing absolutely continuous map, according to Theorem A.2.4. Moreover, ς is onto by definition.
Consider the *quantile function* of ς, i.e. $\tau : [0,L] \to [a,b]$ with

$$\tau(x) := \inf\left\{s \in [a,b] : x \leq \varsigma(s)\right\} = \min\left\{s \in [a,b] : x = \varsigma(s)\right\}. \tag{1.15}$$

Note that τ is a left-continuous, increasing map. However, τ needs not to be continuous as we may not assume that ς is strictly increasing.
Moreover, the quantile function τ satisfies the property that $\varsigma \circ \tau(x) = x$ for every $x \in [0,L]$. On the other hand, one has $\tau \circ \varsigma(t) \leq t$ for all $t \in [a,b]$.
Recall from part (i) of the proof that v can be extended to the closed interval $[a,b]$. Thus, according to the definition of ς, we have

$$d\left(v \circ \tau \circ \varsigma(t), v(t)\right) \leq \int_{\tau \circ \varsigma(t)}^t |v'| \,\mathrm{d}\lambda = \varsigma(t) - \varsigma \circ \tau \circ \varsigma(t) = \varsigma(t) - \varsigma(t) = 0.$$

Hence, we obtain $v \circ \tau \circ \varsigma = v$ on $[a,b]$. Setting $\hat{v} := v \circ \tau$, this establishes the first equation in (1.8).
Furthermore, due to the estimate

$$d\left(\hat{v}(x), \hat{v}(y)\right) \leq \int_{\tau(x)}^{\tau(y)} |v'| \,\mathrm{d}\lambda = \varsigma \circ \tau(y) - \varsigma \circ \tau(x) = y - x \qquad \forall x, y \in [0,L] : x \leq y, \tag{1.16}$$

one infers that $\hat{v}$ is Lipschitz continuous with 1 as a Lipschitz constant.
Multiplying the inequality in (1.16) with $|x-y|^{-1}$ and taking the limit $(y \to x)$ yields

$$|\hat{v}'| = \lim_{y\to x} \frac{d\left(\hat{v}(x), \hat{v}(y)\right)}{|x-y|} \leq 1 \qquad \lambda\text{-a.e. in } [0, L]. \tag{1.17}$$

Conversely, in a similar fashion to (1.12) and (1.13) one obtains the estimate

$$|v'|(t) = \lim_{s\to t} \frac{d\left(v(s), v(t)\right)}{|s-t|} \leq \lim_{s\nearrow t} \frac{1}{t-s} \int_{\varsigma(s)}^{\varsigma(t)} |\hat{v}'| \,\mathrm{d}\lambda = \lim_{s\nearrow t} \frac{1}{t-s} \int_{s}^{t} |\hat{v}'|\circ\varsigma\cdot|v'| \,\mathrm{d}\lambda = |\hat{v}'|\circ\varsigma(t)\cdot|v'|(t) \tag{1.18}$$

for λ-a.e. $x \in [0, L]$. Since $\operatorname{ran}\varsigma = [0, L] = \operatorname{dom}|v'|$, the inequalities (1.17) and (1.18) combined yield the second equation in (1.8). ■

1.2 Weak and Strong Upper Gradients

In this section we consider functionals which are defined on a metric space X. Lacking linear structure, one cannot hope for defining the gradient of such a functional in a point in terms of a linear mapping.

However, the following example will give rise to a definition which generalizes the *norm of a gradient*.

1.2.1 **Example** Let $(Y, \|\cdot\|_Y)$ be a Banach space and consider a Fréchet differentiable function $\phi : Y \to \mathbb{R}$. Then for any $g : Y \to \mathbb{R}$ one has:

$$\|\nabla\phi\| \leq g \quad \textit{iff} \quad \forall v \in C^\infty((a,b), Y) : |(\phi\circ v)'| \leq g\circ v \cdot \|v'\|_Y . \tag{1.19}$$

Recall that the Fréchet derivative $\nabla\phi(y)$ of ϕ at a point $y \in Y$ is a bounded linear functional from Y to $\mathbb{R}$, i.e. $\nabla\phi : Y \to B(Y, \mathbb{R})$. Hence,

$$\|\nabla\phi(y)\| = \sup_{\|x\|_Y=1} |\nabla\phi(y)x|$$

denotes the operator norm on the space $B(Y, \mathbb{R})$.

Proof of (1.19) Due to the chain rule, we have $(\phi\circ v)'(t) = \nabla\phi\left(v(t)\right) v'(t) \in B(\mathbb{R}, \mathbb{R})$ at any point $t \in (a, b)$. As a result, the sub-multiplicity of the operator norm gives us the estimate

$$|(\phi\circ v)'(t)| \leq \|\nabla\phi\left(v(t)\right)\| \cdot \|v'(t)\|_Y \qquad \forall t \in (a, b). \tag{1.20}$$

Now assume that we have $\|\nabla\phi\| \leq g$ in every point of Y. Then one easily obtains the right statement in (1.19) by means of (1.20).

To prove the converse implication, consider any two points $x, y \in Y$ with $\|x\|_Y = 1$. Define a smooth curve $v : (a, b) \to Y$ via $v(t) := tx + y$. Clearly, we have $v'(t) = x$ for every $t \in (a, b)$. Without loss of generality, we may assume $0 \in (a, b)$. Hence,

$$g(y) = g\circ v(0) \cdot \|v'(0)\|_Y \geq |(\phi\circ v)'(0)| = |\nabla\phi\left(v(0)\right) v'(0)| = |\nabla\phi(y)x| .$$

Since the estimate above remains true for all $x \in Y$ with $\|x\|_Y = 1$, we obtain $g(y) \geq \|\nabla\phi(y)\|$ as desired. ◀

Note that the expression on the right side of (1.19) still makes sense if we assume that Y is merely a complete metric space and interpret $\|v'\|_Y$ as the metric differential $|v'|$ of the curve v (cf. also Example 1.1.6).

For practical reasons we will consider only functionals which take values on the extended real line:

Notation

In this section we denote by $\phi : X \to \mathbb{R} \cup \{+\infty\}$ an extended real functional with *proper* effective domain, i.e. the *effective domain* $\operatorname{dom}\phi := \{x \in X : \phi(x) < +\infty\}$ of ϕ is nonempty.

First, we introduce the following notion, based on a pointwise λ-a.e. formulation of the right-hand side in (1.19).

1.2.2 **Definition** A function $g : X \to [0, +\infty]$ is called a ***weak upper gradient*** for ϕ if every curve $v \in AC((a,b), X)$ with the properties

(WUG1) $g \circ v \cdot |v'| \in L^1((a,b), \mathbb{R})$,

(WUG2) $\phi \circ v$ is of essential bounded variation on (a,b), .i.e. there exists a function $\varphi : (a,b) \to \mathbb{R} \cup \{+\infty\}$ of bounded variation such that $\phi \circ v(t) = \varphi(t)$ λ-a.e. in (a,b),

one has $|\varphi'(t)| \le g \circ v(t) \cdot |v'|(t)$ λ-a.e. in (a,b).

There is also a stronger integral formulation of the right-hand side in (1.19).

1.2.3 **Definition** A function $g : X \to [0, +\infty]$ is a ***strong upper gradient*** for ϕ if every curve $v \in AC((a,b), X)$ the function $g \circ v$ is Borel measurable and

$$|\phi \circ v(t) - \phi \circ v(s)| \le \int_s^t g \circ v \cdot |v'| \, d\lambda \qquad \forall s,t \in (a,b) : s \le t. \tag{1.21}$$

The relation between these two definitions is as follows.

1.2.4 **Facts** Consider a function $g : X \to [0, +\infty]$.

(i) Let g be a strong upper gradient for ϕ. If moreover we have $g \circ v \cdot |v'| \in L^1((a,b), \mathbb{R})$ for every absolutely continuous curve $v : (a,b) \to X$, then $\phi \circ v$ is also absolutely continuous according to Definition 1.1.1. Hence, the derivative of $\phi \circ v$ is finite λ-a.e. in (a,b) and by Theorem A.2.6 one obtains

$$|(\phi \circ v)'(t)| \le g \circ v(t) \cdot |v'|(t) \qquad \lambda\text{-a.e. in } (a,b).$$

Thus, one can choose $\varphi = \phi \circ v$ in Definition 1.2.2 and as a result g is also a weak upper gradient for ϕ.

(ii) Now let g be a weak upper gradient for ϕ. If for every absolutely continuous curve $v : (a,b) \to X$ we have $\phi \circ v \in AC((a,b), \mathbb{R})$, then $\phi \circ v = \varphi$. In this case, clearly (1.21) holds. Hence, g is also a strong upper gradient for ϕ.

Among all possible choices of upper gradients we will consider particular ones, called *slopes*. Regarding the following definition, recall that $(f)^+$ and $(f)^-$ denote the *positive part* and the *negative part* of a real-valued function f, respectively.

1.2.5 **Definition** We call

$$|\partial\phi|(x) := \limsup_{y \to x} \frac{(\phi(x) - \phi(y))^+}{d(x,y)} \qquad x \in \operatorname{dom}\phi$$

the ***local slope*** of ϕ. Similarly,

$$\mathfrak{l}_\phi(x) := \sup_{y \neq x} \frac{(\phi(x) - \phi(y))^+}{d(x,y)} \qquad x \in \operatorname{dom}\phi$$

is called the ***global slope*** of ϕ.

1.2.6 Theorem (Slopes are upper gradients) *The local slope $|\partial\phi|$ is a weak upper gradient for ϕ.*
If ϕ is lower semi-continuous, *i.e.* $\liminf_{y\to x} \phi(y) \geq \phi(x)$ *for every $x \in X$, then $\mathfrak{l}_\phi$ is also a strong upper gradient for ϕ.*

For the proof of the theorem above we need the following lemma from the theory of *Sobolev spaces* (cf. Example A.2.10 for a brief overview of the occurring definitions):

1.2.7 Lemma *Let $\varphi \in L^1((a,b),\mathbb{R})$ and $g \in L^1\big((a,b),\mathbb{R}_0^+\big)$ be such that*

$$|\varphi(s) - \varphi(t)| \leq (g(s) + g(t))\,|s-t| \qquad \text{for } \lambda\text{-a.e. } s,t \in (a,b).$$

Then φ belongs to the Sobolev space $W^{1,1}((a,b))$.

Proof Define the linear functional $T : C_c^\infty((a,b)) \to \mathbb{R}$ with

$$T(\zeta) := \int_a^b \varphi \cdot \zeta' \,\mathrm{d}\lambda \qquad \zeta \in C_c^\infty((a,b))\,.$$

Due to *Lebesgue's dominated convergence theorem* as well as *change of variables* and, if necessary, arbitrarily extending the functions φ, ζ, g to the real line, we obtain the estimate

$$|T(\zeta)| = \lim_{h\to 0}\left|\int_a^b \varphi(t)\frac{\zeta(t+h)-\zeta(t)}{h}\,\mathrm{d}\lambda(t)\right| = \tag{1.22.a}$$

$$= \lim_{h\to 0}\left|\int_a^b \varphi(t)\frac{\zeta(t+h)}{h}\,\mathrm{d}\lambda(t) - \int_a^b \varphi(t)\frac{\zeta(t)}{h}\,\mathrm{d}\lambda(t)\right| = \tag{1.22.b}$$

$$= \lim_{h\to 0}\left|\int_a^b \frac{\varphi(s-h)-\varphi(s)}{h}\zeta(s)\,\mathrm{d}\lambda(s)\right| \leq \tag{1.22.c}$$

$$\leq \limsup_{h\to 0}\int_a^b (g(s-h)+g(s))\,|\zeta(s)|\,\mathrm{d}\lambda(s) = \tag{1.22.d}$$

$$= 2\int_a^b g(s)\,|\zeta(s)|\,\mathrm{d}\lambda(s) \leq 2\,\|g\|_1 \cdot \|\zeta\|_\infty\,. \tag{1.22.e}$$

Hence T is continuous on $C_c^\infty((a,b))$. Since $C_c(a,b)$ is a dense subspace of $C_0((a,b))$ with respect to $\|\cdot\|_\infty$, the space of continuous functions on (a,b) which vanish at a and b, we may invoke *continuous linear extension* (cf. a.e. Theorem 1.9.1 in [41]), i.e. there exists a unique extension of T to a continuous linear functional $\widetilde{T}$ on $C_0((a,b))$ such that $\|T\| = \|\widetilde{T}\|$.

Now we can apply the *Riesz-Markov theorem* (see Theorem A.3.3 in Appendix A). Thus, there exists a unique regular countably additive signed Borel measure μ on (a,b) such that

$$\widetilde{T}(\zeta) = \int_a^b \zeta\,\mathrm{d}\mu \qquad \forall \zeta \in C_0((a,b))\,.$$

Moreover, we have $\|\widetilde{T}\| = \|T\| = |\mu|((a,b)) \leq 2\,\|g\|_1$. Clearly, (1.22) also gives the estimate

$$\left|\int_a^b \zeta\,\mathrm{d}\mu\right| \leq 2\int_a^b |\zeta|\cdot|g|\,\mathrm{d}\lambda \qquad \forall \zeta \in C_c^\infty((a,b))\,.$$

Next, let $A \subseteq (a,b)$ be a Borel set. Then there exists a sequence $(\zeta_n)_{n\in\mathbb{N}}$ in $C_0((a,b))$ such that $\lim_{n\to\infty}\zeta_n(t) = \mathbb{1}_A(t)$ pointwise in (a,b). As a result,

$$|\mu(A)| = \lim_{n\to\infty}\left|\int_a^b \zeta_n\,\mathrm{d}\mu\right| \leq \lim_{n\to\infty} 2\int_a^b |\zeta_n|\cdot|g|\,\mathrm{d}\lambda = 2\int_A |g|\,\mathrm{d}\lambda.$$

Furthermore, for any finite partition $(A_i)_{i\in I}$ of A into disjoint measurable subsets A_i we have

$$\sum_{i\in I} |\mu(A_i)| \le 2 \sum_{i\in I} \int_{A_i} |g| \, \mathrm{d}\lambda = \int_A |g| \, \mathrm{d}\lambda.$$

Consequently, $|\mu| \ll \lambda$ and with regard to the *Radon-Nikodym theorem* one yields an integrable density $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \in L^1((a,b),\mathbb{R})$ with

$$\int_a^b \varphi \cdot \zeta' \, \mathrm{d}\lambda = \int_a^b \zeta \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \, \mathrm{d}\lambda \qquad \forall \zeta \in C_c^\infty((a,b))\,.$$

Hence, $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ is a weak derivative of φ and φ belongs to $W^{1,1}((a,b))$. ∎

Now we are ready to proof the main result.

Proof of Theorem 1.2.6 First, we show that $|\partial\phi|$ is a weak upper gradient for ϕ: Consider a curve $v \in AC((a,b),X)$ which satisfies (WUG1) and (WUG2) of Definition 1.2.2 and define the set

$$A := \{t \in (a,b) : \phi \circ v(t) = \varphi(t) \text{ and } \exists \varphi'(t) \text{ and } \exists\, |v'|\,(t)\}\,.$$

We immediately obtain that $(a,b) \setminus A$ is a λ-null set since the derivative φ' and the metric differential $|v'|$ exist λ-a.e. in (a,b).
Now consider a point $t \in A$. In case that $\varphi'(t) = 0$, the inequality $|\varphi'(t)| \le |\partial\phi|\,(t) \cdot |v'|\,(t)$ from Definition 1.2.2 is clearly satisfied. Therefore, assume $\varphi'(t) \neq 0$.
Fix a point $t \in A$ and note that

$$\varphi'(t) = \lim_{\substack{s\to t\\ s\in A}} \frac{\phi \circ v(s) - \phi \circ v(t)}{s-t} \neq 0.$$

Hence, there exists a neighborhood $U_\varepsilon(t)$ such that $v(s) \neq v(t)$, or equivalently $d(v(s),v(t)) \neq 0$, for all $s \in U_\varepsilon(t) \cap A \setminus \{t\}$. Let $\alpha_t := \operatorname{sgn} \varphi'(t)$. Then

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

This establishes the first result.

Next, we show that $\mathfrak{l}_\phi$ is lower semi-continuous in X: To this aim, consider two different points $x, y \in X$, $x \neq y$. Then for every sequence $(x_n)_{n\in\mathbb{N}}$ in X which converges to x, one has $x_n \neq y$ for almost all x_n. As we required ϕ to be lower semi-continuous, we obtain

$$\liminf_{n\to\infty} \mathfrak{l}_\phi(x_n) \ge \liminf_{n\to\infty} \frac{(\phi(x_n) - \phi(y))^+}{d(x_n,y)} \ge \frac{(\phi(x) - \phi(y))^+}{d(x,y)} \qquad \forall y \in X : y \neq x.$$

Henceforth, we infer

$$\liminf_{n\to\infty} \mathfrak{l}_\phi(x_n) \ge \sup_{y\neq x} \frac{(\phi(x) - \phi(y))^+}{d(x,y)} = \mathfrak{l}_\phi(x),$$

which establishes the lower semi-continuity of $\mathfrak{l}_\phi$ in X.

Now we are ready to check that $\mathfrak{l}_\phi$ is a strong upper gradient for ϕ: Assume a curve $v \in AC((a,b),X)$ is given such that $\mathfrak{l}_\phi \circ v \cdot |v'|$ belongs to $L^1((a,b),\mathbb{R})$.
Recall that a function $f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous iff for every $\alpha \in \mathbb{R} \cup \{+\infty\}$ the sublevel set $\{x \in X : f(x) \leq \alpha\}$ is closed in X (cf. a.e. Lemma 1.24.iii in [6]). Since v is continuous, this means that the composition $\mathfrak{l}_\phi \circ v$ is lower semi-continuous; in particular this shows that $\mathfrak{l}_\phi \circ v$ is Borel.
Due to Lemma 1.1.8.ii, there exists an arc-length reparametrization

$$\hat{v} = v \circ \tau : (0,L) \to X,$$

where τ is the quantile function of a absolutely continuous increasing map $\varsigma : [a,b] \to [0,L]$ as defined in (1.15), and L the length of the curve v. Define

$$\varphi := \phi \circ \hat{v} : (0,L) \to \mathbb{R} \cup \{+\infty\} \qquad \text{and} \qquad g := \mathfrak{l}_\phi \circ \hat{v} : (0,L) \to \mathbb{R} \cup \{+\infty\}.$$

As $|\hat{v}'| = 1$ λ-a.e. in $(0,L)$, $\hat{v}$ is 1-Lipschitz, due to Theorem 1.1.7 (MD3) and Fact 1.1.4.iii. Then by definition of the global slope $\mathfrak{l}_\phi$, one has

$$(\varphi(s) - \varphi(t))^+ = (\phi \circ \hat{v}(s) - \phi \circ \hat{v}(t))^+ \leq \frac{|s-t|}{d(\hat{v}(s),\hat{v}(t))} (\phi \circ \hat{v}(s) - \phi \circ \hat{v}(t))^+ \leq$$

$$\leq |s-t| \sup_{\substack{t \in (0,L) \\ \hat{v}(t) \neq \hat{v}(s)}} \frac{(\phi \circ \hat{v}(s) - \phi \circ \hat{v}(t))^+}{d(\hat{v}(s),\hat{v}(t))} \leq |s-t| \sup_{\substack{x \in X \\ x \neq \hat{v}(s)}} \frac{(\phi \circ \hat{v}(s) - \phi(x))^+}{d(\hat{v}(s),x)} = |s-t| g(s)$$

for every choice of points $s,t \in (0,L)$ such that $\hat{v}(s) \neq \hat{v}(t)$. This immediately gives us

$$(\varphi(s) - \varphi(t))^+ \leq |s-t| g(s) \qquad \forall s,t \in (0,L), \tag{1.23}$$

and

$$|\varphi(s) - \varphi(t)| = (\varphi(s) - \varphi(t))^+ + (\varphi(s) - \varphi(t))^+ \leq |s-t| (g(s) + g(t)) \qquad \forall s,t \in (0,L). \tag{1.24}$$

Moreover, via *change of variables* with transformation ς we have

$$\int_0^L g \,\mathrm{d}\lambda = \int_{\varsigma(a)}^{\varsigma(b)} \mathfrak{l}_\phi \circ \hat{v} \,\mathrm{d}\lambda = \int_a^b \mathfrak{l}_\phi \circ v \circ (\tau \circ \varsigma) \cdot \varsigma' \,\mathrm{d}\lambda = \int_a^b \mathfrak{l}_\phi \circ v \cdot |v'| \,\mathrm{d}\lambda < +\infty,$$

where we used that (1.14) establishes $|v'| = \varsigma'$ λ-a.e. in (a,b). Hence, g belongs to $L^1((0,L),\mathbb{R})$. Moreover, (1.24) allows us to apply Lemma 1.2.7 to obtain that φ belongs to the Sobolev space $W^{1,1}((0,L))$.

Now we check that φ is a continuous representative of the corresponding equivalence class in $W^{1,1}((0,L))$: To this end, fix a point $t_0 \in (0,L)$ and $\varepsilon > 0$ such that $B_{2\varepsilon}(t_0) \subseteq (0,L)$. By virtue of Lemma 1.1.2, there exists $\delta > 0$ such that $\delta \leq \varepsilon$ and

$$\int_{B_\delta(t)} g \,\mathrm{d}\lambda < \varepsilon \qquad \forall t \in B_\varepsilon(t_0).$$

As a consequence of this estimate and (1.23) we obtain

$$\frac{1}{2\delta} \int_\delta^\delta |\varphi(t+s) - \varphi(t)| \,\mathrm{d}\lambda(s) = \frac{1}{2\delta} \int_\delta^\delta (\varphi(t+s) - \varphi(t))^+ + (\varphi(t) - \varphi(t+s))^+ \mathrm{d}\lambda(s) \leq \tag{1.25.a}$$

$$\leq \frac{1}{2\delta} \int_\delta^\delta |s| (g(t+s) + g(t)) \,\mathrm{d}\lambda(s) \leq \frac{1}{2} \int_\delta^\delta g(t+s) \,\mathrm{d}\lambda(s) + \frac{1}{2} \int_\delta^\delta g(t) \,\mathrm{d}\lambda(s) < \varepsilon \tag{1.25.b}$$

for all points $t \in B_\varepsilon(t_0)$. This shows that the open interval $B_\varepsilon(t_0)$ consists only of Lebesgue points, i.e. φ satisfies

$$\lim_{\varepsilon \searrow 0} \int_{t-\varepsilon}^{t+\varepsilon} \varphi \, \mathrm{d}\lambda = 0 \qquad \forall t \in B_\varepsilon(t_0).$$

Moreover, the estimate in (1.25) is uniform with respect to $t \in B_\varepsilon(t_0)$. Due to Fact A.2.7.v, we infer that for every point t_0 in $(0, L)$ there exists a neighborhood $U \subseteq (0, L)$ of t_0 such that φ is uniformly continuous on U. In particular φ is continuous on $(0, L)$.

Now we may infer from Example A.2.10 that φ is indeed an absolutely continuous representative of the corresponding equivalence class in $W^{1,1}((0, L))$. Since Fact 1.1.4.iii shows that the composition $\varphi \circ \varsigma$ is absolutely continuous as well, we conclude from $\phi \circ v = \phi \circ \hat{v} = \varphi \circ \varsigma$ the absolute continuity of $\phi \circ v$. Finally, we apply Fact 1.2.4.i to $|\partial \phi|$ and use that $|\partial \phi| \leq \mathfrak{l}_\phi$ in X to ascertain that

$$|\phi \circ v(t) - \phi \circ v(s)| \leq \int_s^t |\partial \phi| \circ v \cdot |v'| \, \mathrm{d}\lambda \leq \int_s^t \mathfrak{l}_\phi \circ v \cdot |v'| \, \mathrm{d}\lambda \qquad \forall s, t \in (a, b) : s \leq t.$$

In the general case that the integrand $\mathfrak{l}_\phi \circ v \cdot |v'|$ is only Borel, it suffices to consider all integrable restrictions of the form $\left(\mathfrak{l}_\phi \circ v \cdot |v'|\right)\big|_{(\acute{a}, \acute{b})}$ to intervals $(\acute{a}, \acute{b}) \subseteq (a, b)$. Otherwise the inequality in (1.21) is trivially satisfied.

This establishes that $\mathfrak{l}_\phi$ is a strong upper gradient for ϕ. ■

1.3 Gradient Flows

In this section, our starting point is a gradient flow in Euclidean space. To this end, we identify the flow with its flow curves, i.e. we consider differentiable curves $v : (0, +\infty) \to \mathbb{R}^n$, starting from $\lim_{t \searrow 0} v(t) = x_0 \in \mathbb{R}^n$, which solve

$$v'(t) = -\nabla F\big(v(t)\big) \qquad \forall t \in (0, +\infty) \tag{1.26}$$

for a given differentiable potential $F : \mathbb{R}^n \to \mathbb{R}$. A brief introduction to the theory of gradient flows in the Euclidean setting is given in Section B.2 of Appendix B.

The aim of this section is to generalize the notion of gradient flows to a metric setting. This will be accomplished in two steps:

In *Step 1* we will focus on gradient flows in an Euclidean framework and introduce various equivalent characterizations of a gradient flow. Since (1.26) makes no sense in a general metric setting, we will make use of these characterizations in *Step 2* to establish a metric notion of gradient flows for ***λ**-convex* functionals.

Notation

In this section we will regard $\mathbb{R}^n$ as a real Hilbert space, endowed with a (not necessarily standard) inner product $\langle \cdot, \cdot \rangle$ and the norm $\|\cdot\|$ induced by $\langle \cdot, \cdot \rangle$.

Step 1 First we note that, for any given curve $v \in C^1\big((0, +\infty), \mathbb{R}^n\big)$, starting from $\lim_{t \searrow 0} v(t) = x_0 \in \mathbb{R}^n$, and potential $F \in C^1(\mathbb{R}^n, \mathbb{R})$, equation (1.26) can be utilized to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} F\big(v(t)\big) = \Big\langle \nabla F\big(v(t)\big), v'(t) \Big\rangle = -\left\|\nabla F\big(v(t)\big)\right\|^2 = -\|v'(t)\|^2 \qquad \forall t \in (0, +\infty). \tag{1.27}$$

This equation is called *energy identity* of the gradient flow. Note that (1.27) may also be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}F\big(v(t)\big) = -\frac{1}{2}\left\|\nabla F\big(v(t)\big)\right\|^2 - \frac{1}{2}\left\|v'(t)\right\|^2 \qquad \forall t \in (0,+\infty), \tag{1.28}$$

or its weaker integral form

$$F\big(v(s)\big) - F\big(v(t)\big) = \frac{1}{2}\int_s^t \|v'\|^2 \,\mathrm{d}\lambda + \frac{1}{2}\int_s^t \left\|\nabla F\big(v\big)\right\|^2 \mathrm{d}\lambda \qquad s,t \in (0,+\infty) : s \le t, \tag{1.29}$$

which are called *energy dissipation equality* and its weak formulation.
There exist also variants of (1.28) and (1.29); these are the *energy dissipation inequality*

$$\frac{\mathrm{d}}{\mathrm{d}t}F\big(v(t)\big) \le -\frac{1}{2}\left\|\nabla F\big(v(t)\big)\right\|^2 - \frac{1}{2}\left\|v'(t)\right\|^2 \qquad \forall t \in (0,+\infty), \tag{1.30}$$

and its weak formulation

$$F\big(v(s)\big) - F\big(v(t)\big) \ge \frac{1}{2}\int_s^t \|v'\|^2 \,\mathrm{d}\lambda + \frac{1}{2}\int_s^t \left\|\nabla F\big(v\big)\right\|^2 \mathrm{d}\lambda \qquad s,t \in (0,+\infty) : s \le t, \tag{1.31}$$

respectively.

A different characterization of (1.26) can be established by the notion of $\boldsymbol{\lambda}$*-convexity* which is a slight generalization of the usual concept of convexity.

1.3.1 Definition Let H be a real Hilbert space. We say that a function $F : H \to \mathbb{R} \cup \{+\infty\}$ is $\boldsymbol{\lambda}$***-convex*** with respect to $\boldsymbol{\lambda} \in \mathbb{R}$ if

$$F\left((1-t)x + ty\right) \le (1-t)F(x) + tF(y) - \frac{\boldsymbol{\lambda}}{2}t(1-t)\left\|x-y\right\|^2 \qquad \forall x,y \in H,\ \forall t \in [0,1].$$

In the case $\boldsymbol{\lambda} = 0$, we simply call the function F ***convex***.

1.3.2 Facts

(i) Let H be a real Hilbert space. For any two points $x, y \in H$ and $t \in \mathbb{R}$ we have the expressions

$$t(1-t)\left\|x-y\right\|^2 = \langle (1-t)(x-y), t(x-y)\rangle$$

and

$$\left\|(1-t)x + ty\right\|^2 = \langle x + t(y-x), x\rangle + \langle x + t(y-x), t(y-x)\rangle,$$

which can be added to obtain the identity

$$\left\|(1-t)x + ty\right\|^2 = (1-t)\left\|x\right\|^2 + t\left\|y\right\|^2 - t(1-t)\left\|x-y\right\|^2. \tag{1.32}$$

Note that, by setting $t = \frac{1}{2}$ above, one obtains the well-known *parallelogram law*:

$$\left\|x+y\right\|^2 + \left\|x-y\right\|^2 = 2\left\|x\right\|^2 + 2\left\|y\right\|^2.$$

Hence, (1.32) holds exactly when the norm $\|\cdot\|$ is induced by an scalar product. In this case, one can utilize (1.32) to infer that a functional $F : H \to \mathbb{R} \cup \{+\infty\}$ is $\boldsymbol{\lambda}$-convex iff the function

$$x \longmapsto F(x) - \frac{\boldsymbol{\lambda}}{2}\left\|x\right\|^2$$

is convex.

(ii) Consider a functional F which belongs to $C^2(\mathbb{R}^n, \mathbb{R})$. Then we have the following characterization of λ-convexity, due to Brunn and Hadamard (see Theorem 2.10 in [30] for the convex case):

$$F \text{ is } \lambda\text{-convex} \quad \text{iff} \quad y^\top \operatorname{Hess} F(x) y \geq \lambda y^\top y \quad \forall x, y \in \mathbb{R}^n. \tag{1.33}$$

Proof of (1.33) Assume that F is λ-convex and fix two points $\tilde{x}, \tilde{y} \in \mathbb{R}^n$. Then the function

$$G(h) := F\big((1-h)\tilde{x} + h\tilde{y}\big) = F\big(\tilde{x} + h(\tilde{y} - \tilde{x})\big)$$

belongs to the class $C^2(\mathbb{R}, \mathbb{R})$. λ-convexity of F gives us for

$$t = \frac{1}{2}, \quad x = \tilde{x} + h(\tilde{y} - \tilde{x}), \quad y = \tilde{x} - h(\tilde{y} - \tilde{x}), \quad h \in \mathbb{R},$$

the estimate

$$2F(\tilde{x}) \leq F\big(\tilde{x} + h(\tilde{y} - \tilde{x})\big) + F\big(\tilde{x} - h(\tilde{y} - \tilde{x})\big) - \lambda h^2 \|\tilde{y} - \tilde{x}\|^2 \qquad \forall h \in \mathbb{R},$$

which is equivalent to

$$\lambda \|\tilde{y} - \tilde{x}\|^2 \leq \frac{F\big(\tilde{x} + h(\tilde{y} - \tilde{x})\big) + F\big(\tilde{x} - h(\tilde{y} - \tilde{x})\big) - 2F(\tilde{x})}{h^2} = \frac{G(h) + G(-h) - 2G(0)}{h^2}.$$

As a result, we obtain for the second order derivative of G at 0 that

$$\lambda \|\tilde{y} - \tilde{x}\|^2 \leq G''(0). \tag{1.34}$$

On the other hand, the chain rule yields

$$G''(h) = \sum_{i,j} \frac{\partial^2 f\big(\tilde{x} + h(\tilde{y} - \tilde{x})\big)}{\partial x_i \partial x_j} \big(\tilde{y}_i - \tilde{x}_i\big)\big(\tilde{y}_j - \tilde{x}_j\big).$$

Together with (1.34) this establishes the result

$$(\tilde{y} - \tilde{x})^\top \operatorname{Hess} F(\tilde{x})(\tilde{y} - \tilde{x}) = \sum_{i,j} \frac{\partial^2 f(\tilde{x})}{\partial x_i \partial x_j} \big(\tilde{y}_i - \tilde{x}_i\big)\big(\tilde{y}_j - \tilde{x}_j\big) = G''(0) \geq \lambda (\tilde{y} - \tilde{x})^\top (\tilde{y} - \tilde{x}).$$

To prove the converse statement, assume that $\operatorname{Hess} F - \lambda \operatorname{Id}$ is positive semi-definite. Consider two points $x, y \in \mathbb{R}^n$. Then a second order Taylor expansion of F at $l_t := (1-t)x + ty$, $t \in [0,1]$ implies

$$F(x) = F(l_t) + (x - l_t)^\top \nabla F(l_t) + \frac{1}{2}(x - l_t)^\top \operatorname{Hess} F\big(l_t + \vartheta(x - l_t)\big)(x - l_t) \geq \tag{1.35.a}$$

$$\geq F(l_t) + t(x - y)^\top \nabla F(l_t) + \frac{\lambda}{2} t^2 \|y - x\|^2. \tag{1.35.b}$$

for a suitable $\vartheta \in [0,1]$. Similarly, we obtain

$$F(y) \geq F(l_t) - (1-t)(x - y)^\top \nabla F(l_t) + \frac{\lambda}{2}(1-t)^2 \|y - x\|^2. \tag{1.36}$$

Multiplying the inequalities (1.35) and (1.36) with $(1-t)$ and t, respectively, and adding them up results in

$$(1-t)F(x)+tF(y)\geq F(l_t)+\frac{\lambda}{2}t(1-t)\|y-x\|^2,$$

which corresponds to the λ-convexity of F.

(iii) Setting $t=1$ in (1.35) results in the so-called *subgradient inequality*

$$\langle\nabla F(y),x-y\rangle\geq F(x)-F(y)-\frac{\lambda}{2}\|y-x\|^2\qquad\forall x,y\in\mathbb{R}^n.\tag{1.37}$$

Note that for functionals F, which belong to $C^2(\mathbb{R}^n,\mathbb{R})$, inequality (1.37) holds precisely when F is λ-convex.

Yet another characterization of λ-convexity can be established by switching the positions of x and y in (1.37) and adding both versions, which results in

$$\langle\nabla F(x)-\nabla F(y),x-y\rangle\geq\lambda\|x-y\|^2\qquad\forall x,y\in\mathbb{R}^n.\tag{1.38}$$

This inequality is known as λ-*monotonicity* of ∇F. Observe that setting $x=y+tx_0$ in (1.38) leads to

$$\langle\nabla F(y+tx_0)-\nabla F(y),tx_0\rangle\geq\lambda\|tx_0\|^2\qquad\forall x_0,y\in\mathbb{R}^n,t>0.$$

Hence, dividing by $t^2>0$ results in

$$\frac{1}{t}\langle\nabla F(y+tx_0)-\nabla F(y),x_0\rangle\geq\lambda\|x_0\|^2,$$

where we can pass to the limit $(t\searrow 0)$ to obtain

$$x_0^\top\operatorname{Hess}F(y)x_0=\langle\operatorname{Hess}F(y)x_0,x_0\rangle\geq\lambda\|x_0\|^2\qquad\forall x_0,y\in\mathbb{R}^n.$$

We conclude that for C^2-functionals (1.38) is again equivalent to the λ-convexity of F.

(iv) Note that that for every $x\in\mathbb{R}^n$ the matrix $\operatorname{Hess}F(x)$ is symmetric. Hence, there exists a *spectral decomposition*

$$\operatorname{Hess}F(x)=\sum_{\lambda\in\sigma_x}\lambda P_\lambda$$

where σ_x denotes the *spectrum* of $\operatorname{Hess}F(x)$, consisting only of real eigenvalues, and P_λ the orthogonal projection onto the eigenspace corresponding to the eigenvalue λ. Therefore, we have

$$y^\top\operatorname{Hess}F(x)y=\sum_{\lambda\in\sigma_x}\lambda y^\top P_\lambda y\geq\min\sigma_x\sum_{\lambda\in\sigma_x}y^\top P_\lambda y=\min\sigma_x\cdot y^\top y\qquad\forall y\in\mathbb{R}^n.$$

Thus, by virtue of (1.33), the largest possible $\lambda\in\mathbb{R}$ such that F is λ-convex, can be characterized by

$$\lambda_{\max}=\inf_{x\in\mathbb{R}^n}\min\sigma_x.$$

Consider again a gradient flow as in (1.26) such that $v\in C^1\big((0,+\infty),\mathbb{R}^n\big)$ and $F\in C^2(\mathbb{R}^n,\mathbb{R})$. For every $y\in\mathbb{R}^n$ we can use the subgradient inequality from Fact 1.3.2.iii to obtain the estimate

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|v(t)-y\|^2 = \langle v'(t), v(t)-y\rangle = \big\langle \nabla F\big(v(t)\big), y-v(t)\big\rangle$$
$$\leq F(y) - F\big(v(t)\big) - \frac{\lambda}{2}\|v(t)-y\|^2 .$$

Thus

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|v(t)-y\|^2 \leq F(y) - F\big(v(t)\big) - \frac{\lambda}{2}\|v(t)-y\|^2 \qquad \forall y \in \mathbb{R}^n, \tag{1.39}$$

which is called *evolution variational inequality* of the gradient flow.

The following result summarizes the relations between the various characterizations of a gradient flow, presented so far.

1.3.3 Proposition *Let v be a curve which belongs to $C^1((0,+\infty),\mathbb{R}^n)$ and let $F : \mathbb{R}^n \to \mathbb{R}$ be a C^2-functional.*

(i) *The curve v is a solution of* (1.26) *iff it satisfies the energy dissipation inequality* (1.30) *or its weak formulation* (1.31).

(ii) *The curve v is a solution of* (1.26) *iff it satisfies the energy dissipation equality* (1.28) *or its weak formulation* (1.29).

(iii) *Under the additional assumption that the functional F is λ-convex, v is a solution of* (1.26) *iff it satisfies the energy variational inequality* (1.39).

Proof To prove (i), assume that v satisfies the weak form (1.31) of the energy dissipation inequality. The chain rule implies

$$F\big(v(t)\big) - F\big(v(s)\big) = \int_s^t \big\langle \nabla F(v), v'\big\rangle \,\mathrm{d}\lambda \qquad \forall s,t \in (0,+\infty) : s \leq t.$$

Thus (1.31) yields

$$\frac{1}{2}\int_s^t \big\|v' + \nabla F(v)\big\|^2 \,\mathrm{d}\lambda = \frac{1}{2}\int_s^t \|v'\|^2 + \big\|\nabla F(v)\big\|^2 \,\mathrm{d}\lambda + \int_s^t \big\langle \nabla F(v), v'\big\rangle \,\mathrm{d}\lambda \leq 0.$$

This means that $v'(t) = -\nabla F\big(v(t)\big)$ for λ-a.e. t in $(0,+\infty)$. Due to the C^1-smoothness of v, we infer that v solves (1.26) for all $t \in (0,+\infty)$.

For (ii) we show that, whenever the energy dissipation inequality (1.30) is satisfied, actually equality holds: Note that for any C^1-curve v the Cauchy-Schwartz inequality and the *inequality of arithmetic and geometric means (AM-GM inequality)* imply

$$\frac{\mathrm{d}}{\mathrm{d}t}F\big(v(t)\big) = \big\langle \nabla F\big(v(t)\big), v'(t)\big\rangle \geq -\big\|\nabla F\big(v(t)\big)\big\| \cdot \|v'(t)\| \geq -\frac{1}{2}\big\|\nabla F\big(v(t)\big)\big\|^2 - \frac{1}{2}\|v'(t)\|^2 .$$

Together with (1.30), this estimate yields (1.28). In a similar way one obtains the equivalence of the weak formulations of the energy dissipations inequality and equality.

Concerning (iii), let v be a curve which solves (1.39). Evaluating this inequality at the point $l_\varepsilon := v(t) + \varepsilon y$ for $\varepsilon > 0$ and $y \in \mathbb{R}^n$ establishes

$$-\langle v'(t), y\rangle = \frac{1}{\varepsilon}\langle v'(t), v(t) - l_\varepsilon\rangle \leq \frac{1}{\varepsilon}\Big(F(l_\varepsilon) - F\big(v(t)\big)\Big) - \varepsilon\frac{\lambda}{2}\|y\|^2 .$$

Passing to the limit ($\varepsilon \searrow 0$) implies

$$-\langle v'(t), y\rangle \leq \big\langle \nabla F\big(v(t)\big), y\big\rangle \qquad \forall y \in \mathbb{R}^n,$$

which means that $-v(t) = \nabla F\big(v(t)\big)$ for all $t \in (0,+\infty)$. ∎

It is worth mentioning that the evolution variational inequality along *every* flow curve enforces the λ-convexity of the gradient flow's potential.

1.3.4 Proposition *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a C^2-potential of a gradient flow as in* (1.26) *and fix $\lambda \in \mathbb{R}$. Suppose, for every $x_0 \in \mathbb{R}^n$ there exists a unique flow curve $v \in C^1\big((0,+\infty), \mathbb{R}^n\big)$ such that $\lim_{t\searrow 0} v(t) = x_0$ and v satisfies the evolution variational inequality* (1.39)*. Then F is λ-convex.*

Proof Let $l_\varepsilon := (1-\varepsilon)x + \varepsilon y$, $\varepsilon \in \mathbb{R}$ for arbitrary $x, y \in \mathbb{R}^n$ and let $v_\varepsilon(t)$ be the unique flow curve, starting from $\lim_{t\searrow 0} v_\varepsilon(t) = l_\varepsilon$. Then, by virtue of (1.26), we have

$$\langle \nabla F(x), y - x\rangle = -\langle v'_0(0), y - x\rangle = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\|v_0(t) - y\|^2\right)\bigg|_{t=0} \leq F(y) - F(x) - \frac{\lambda}{2}\|x - y\|^2 = \tag{1.40.a}$$

$$= -\big(F(x) - F(y)\big) - \frac{\lambda}{2}\|x - y\|^2 \leq -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\|x - v_1(t)\|^2\right)\bigg|_{t=0} - \lambda\|x - y\|^2 = \tag{1.40.b}$$

$$= -\langle v'_1(0), x - y\rangle - \lambda\|x - y\|^2 = \langle \nabla F(y), y - x\rangle - \lambda\|x - y\|^2, \tag{1.40.c}$$

where we used the evolution variational inequality in line (1.40.a) as well as in line (1.40.b). Note that the estimate in (1.40) implies

$$\langle \nabla F(y) - \nabla F(x), y - x\rangle \geq \lambda\|x - y\|^2 \qquad \forall x, y \in \mathbb{R}^n,$$

which is exactly the λ-monotonicity of ∇F, as seen in Fact 1.3.2.iv. This establishes the λ-convexity of the functional F. ∎

Before moving to the general metric framework, we investigate the connection between gradient flows and *subdifferentials*.

1.3.5 Remark The subgradient inequality (1.37) motivates the notion of a *subdifferential* with respect to a possibly non smooth λ-convex function. To this end, consider a real Hilbert space H and let $F : H \to \mathbb{R} \cup \{+\infty\}$ be a λ-convex functional. Then a vector $v \in H$ is called *subgradient* of F at the point $x \in \operatorname{dom} F$ if

$$\langle v, x - y\rangle \leq F(x) - F(y) - \frac{\lambda}{2}\|x - y\|^2 \qquad \forall y \in H. \tag{1.41}$$

The set of all subgradients of F at $x \in \operatorname{dom} F$ is called *subdifferential* of F at x, commonly denoted by $\partial F(x)$.

In the case that the functional F belongs to $C^2(\mathbb{R}^n, \mathbb{R})$, (1.37) shows that $\nabla F(x)$ is a subgradient of F at all points $x \in \mathbb{R}^n$.

The notion of subdifferentials of convex functions (i.e. $\lambda = 0$) has been introduced in the 1960s by Moreau and Rockafeller in [42] and [18] respectively, and is well established in the theory of convex optimization. With this concept at hand, one is able of establishing a generalized gradient flow with respect to Hilbert space valued λ-convex functionals:

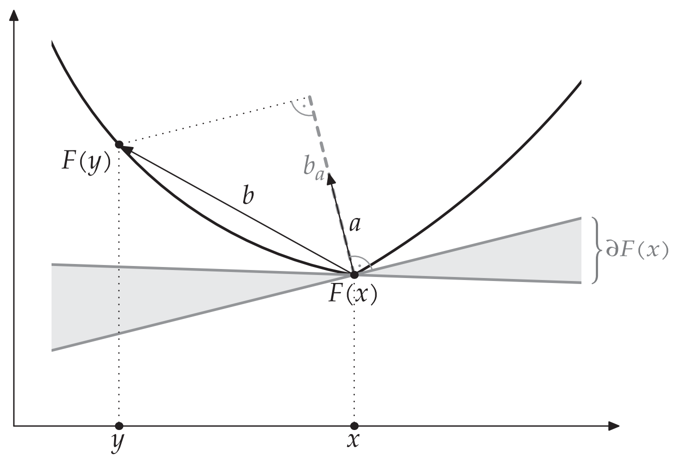


Figure 1.1 The subdifferential with respect to a λ-convex functional $F : \mathbb{R} \to \mathbb{R}$: In this case the underlying inequality (1.37) can be expressed as $\frac{\lambda}{2}|x - y|^2 \leq (-v, 1)^\top\big(x - y, F(x) - F(y)\big) = ab_a$. Here the vector $(-v, 1)$ with length a can be interpreted as normal of the line, going through the point $\big(x, F(x)\big)$ with slope v. Note that in the case $\lambda \geq 0$, as depicted above, the line has to be everywhere either touching or below the graph of F.

We say that $v : (0,+\infty) \to \mathbb{R}^n$ is the flow curve, starting from $x_0 \in H$, of a gradient flow for a λ-convex functional $F : H \to \mathbb{R}$ if v is absolutely continuous and

$$\lim_{t\searrow 0} v(t) = x_0 \in H, \tag{1.42.a}$$

$$v'(t) \in \partial F\big(v(t)\big) \qquad \lambda\text{-a.e. in } (0,+\infty). \tag{1.42.b}$$

A fruitful theory of such gradient flows in Hilbert spaces has been developed, under the additional requirement of F being *lower semi-continuous* by Kōmura [35], Crandall and Pazy [25], Crandall and Liggett [24], Bénilan [11], Lions [38] amongst others. We cite here some main uniqueness and existence results of Barbu [5] and Brézis [16]:

Assume that the functional $F : H \to \mathbb{R} \cup \{+\infty\}$ is λ-convex and lower semi-continuous. Then the following statements hold:

- *For all starting points $x_0 \in \overline{\operatorname{dom} F}$ the gradient flow* (1.42) *has a unique solution v which belongs to $AC((0,+\infty),H)$.*
- *For every $x \in H$ the subdifferential $\partial F(x)$ is closed and convex. Henceforth, as long as $\partial F(x)$ is non-empty, there exists a unique element $v_{\min}$ in $\partial F(x)$ such that the set $\{\|v\| : v \in \partial F(x)\}$ attains its minimum at $\|v_{\min}\|$. The element $v_{\min}$ is usually denoted by $\nabla F(x)$.*
- *The real-valued functions $t \mapsto \|v(t)\|$, $t \mapsto \big\|\nabla F\big(v(t)\big)\big\|$ belong to $L^2_{\mathrm{loc}}((0,+\infty),\mathbb{R})$ and $t \mapsto F\big(v(t)\big)$ belongs to $AC_{\mathrm{loc}}((0,+\infty),H)$, such that the energy dissipation equality holds:*

$$F\big(v(s)\big) - F\big(v(t)\big) = \frac{1}{2}\int_s^t \|v'\|^2 \,\mathrm{d}\lambda + \frac{1}{2}\int_s^t \big\|\nabla F\big(v\big)\big\|^2 \,\mathrm{d}\lambda \qquad s,t \in (0,+\infty) : s \le t.$$

- *The curve v is the unique solution of the evolution variational inequality*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\hat{v}(t) - y\|^2 \le F(y) - F\big(\hat{v}(t)\big) - \frac{\lambda}{2}\|v(t) - y\|^2 \qquad \forall y \in H,\ \lambda\text{-a.e. } t \in (0,+\infty),$$

amongst all curves $\hat{v}$ of the class $AC_{\mathrm{loc}}((0,+\infty),H)$, which start from $\lim_{t\searrow 0}\hat{v}(t) = x_0$.

Step 2 We move on to a metric framework: We already noted that (1.26) admits no direct formulation in such a setting. However, one may utilize the tools developed in Section 1.1 and Section 1.2 to carry over the gradient flow characterizations (1.31), (1.29) and (1.39) to a metric space (X,d). To this aim, it is sufficient to require that the curve v belongs to $AC_{\mathrm{loc}}((0,+\infty),X)$. Then Theorem 1.1.7 assures that the metric derivative $|v'|$ exists λ-a.e. in $(0,+\infty)$ and is Borel. Moreover, since the reverse triangle inequality implies, for fixed $y \in X$, that

$$\big|d\big(v(s),y\big) - d\big(v(t),y\big)\big| \le d\big(v(s),v(t)\big) \qquad \forall s,t \in (0,+\infty),$$

we obtain that the real-valued function $t \to d\big(v(t),y\big)$ is locally absolutely continuous and therefore differentiable λ-a.e. in $(0,+\infty)$. By virtue of Fact A.2.5.ii, this holds also true for the map $t \to d^2\big(v(t),y\big)$. Thus we can establish the following definitions:

1.3.6 **Definition** Assume that a functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ with proper effective domain $\operatorname{dom}\phi$ is given.

(EDI) A curve $v \in AC_{\mathrm{loc}}((0,+\infty),X)$, starting from $\lim_{t\searrow 0} v(t) = x_0 \in \operatorname{dom}\phi$, satisfies the ***energy dissipation inequality (EDI)*** if

$$\frac{1}{2}\int_0^t |v'|^2 \,\mathrm{d}\lambda + \frac{1}{2}\int_0^t |\partial\phi|^2 \circ v \,\mathrm{d}\lambda \le \phi(x_0) - \phi\big(v(t)\big) \qquad t \in (0,+\infty), \tag{1.43.a}$$

and

$$\frac{1}{2}\int_s^t |v'|^2 \, \mathrm{d}\lambda + \frac{1}{2}\int_s^t |\partial\phi|^2 \circ v \, \mathrm{d}\lambda \leq \phi\big(v(s)\big) - \phi\big(v(t)\big) \qquad \lambda\text{-a.e. } s,t \in (0,+\infty) : s \leq t. \tag{1.43.b}$$

By abuse of terminology, we call such a curve ***gradient flow in the EDI sense***.

(EDE) A curve $v \in AC_{\mathrm{loc}}((0,+\infty),X)$, starting from $\lim_{t\searrow 0} v(t) = x_0 \in \operatorname{dom}\phi$, satisfies the ***energy dissipation equality (EDE)*** if

$$\frac{1}{2}\int_s^t |v'|^2 \, \mathrm{d}\lambda + \frac{1}{2}\int_s^t |\partial\phi|^2 \circ v \, \mathrm{d}\lambda = \phi\big(v(s)\big) - \phi\big(v(t)\big) \qquad \forall s,t \in (0,+\infty) : s \leq t. \tag{1.44}$$

By abuse of terminology, such a curve is called ***gradient flow in the EDE sense***.

(EVI) A curve $v \in AC_{\mathrm{loc}}((0,+\infty),X)$, starting from $\lim_{t\searrow 0} v(t) = x_0 \in \overline{\operatorname{dom}\phi}$, satisfies the ***evolution variational inequality (EVI)*** with respect to a given $\boldsymbol{\lambda} \in \mathbb{R}$ if

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} d^2\big(v(t),y\big) \leq \phi(y) - \phi\big(v(t)\big) - \frac{\boldsymbol{\lambda}}{2} d^2\big(v(t),y\big) \qquad \forall y \in X, \lambda\text{-a.e. } t \in (0,+\infty). \tag{1.45}$$

By abuse of terminology, we say such a curve is a ***gradient flow in the EVI sense*** with respect to $\boldsymbol{\lambda}$.

Note that a gradient flow in the sense of the definition above, may not depict a flow as in Definition B.1.1. However, if one can show the existence and uniqueness of such a curve v_{x_0} with respect to every given initial datum x_0, then a global flow on X is induced via the identity

$$\varkappa^t(x_0) := \begin{cases} v_{x_0}(-t) & t \in (-\infty,0), \\ x_0 & t = 0, \\ v_{x_0}(t) & t \in (0,+\infty). \end{cases}$$

Therefore, we will explore minimal existence and uniqueness assumptions on the metric space X and the functional ϕ in the next section.

1.4 First Results in a General Metric Setting

In this section we provide some results on existence and uniqueness of a gradient flow according to Definition 1.3.6. However, the particular choice of the underlying definition (EDI),(EDE) or (EVI) plays an important role in the general metric framework. In particular, there exists no equivalent to Proposition 1.3.3 for gradient flows on metric spaces.

Our minimal assumption in this section will be the lower semi-continuity of the functional ϕ. In this context we will show that a gradient flow in the (EVI) sense is the strongest of the presented formulations. To this end, we regard the following result towards uniqueness of the (EVI) flow.

1.4.1 Proposition *Consider two curves $v,w :\in AC_{\mathrm{loc}}((0,+\infty),X)$, both of which solve the evolution variational inequality* (1.45) *with respect to some $\boldsymbol{\lambda} \in \mathbb{R}$ and assume that ϕ is a lower semi-continuous functional with proper effective domain. Then v and w satisfy the following $\boldsymbol{\lambda}$-contraction property:*

$$d\big(v(t),w(t)\big) \leq \mathrm{e}^{-\boldsymbol{\lambda}(t-s)} d\big(v(s),w(s)\big) \qquad s,t \in (0,+\infty) : s \leq t.$$

In particular, for every initial datum $x_0 \in \overline{\mathrm{dom}\,\phi}$ there exists at most one gradient flow in the EVI sense.

Proof Let $u : (0,+\infty) \to X$ be a locally absolutely continuous curve satisfying (EVI) and note that the map $\phi \circ u$ is lower semi-continuous on $(0,+\infty)$. Therefore u is Borel and bounded from below on every compact subset of $(0,+\infty)$. On the other hand, $\phi \circ u$ is bounded from above by a locally integrable function since the corresponding terms in (1.45) are locally absolutely continuous with respect to t. We conclude that $\phi \circ u$ belongs to $L^1_{\mathrm{loc}}((0,+\infty),\mathbb{R})$.

Now integration of (1.45) over the interval $(t-h,t)$ establishes the following weak formulation

$$\frac{1}{h}d^2\big(u(t),y\big) + \frac{1}{h}d^2\big(u(t-h),y\big) + \frac{1}{h}\int_{t-h}^{t} \lambda d^2\big(u(s),y\big) + 2\phi\big(u(s)\big)\,\mathrm{d}\lambda(s) \le 2\phi(y) \qquad \forall y \in X \tag{1.46}$$

for any $h > 0$. The lower semi-continuity of $\phi \circ u$ implies

$$\limsup_{h\searrow 0} \frac{1}{h}\int_{t-h}^{t} \phi \circ u \,\mathrm{d}\lambda \ge \limsup_{h\searrow 0} \int_{t-h}^{t} \Big(\inf_{\substack{s\neq t\\ s\in B_h(t)}} \phi \circ u(s)\Big)\,\mathrm{d}\lambda \ge \liminf_{h\to t} \phi \circ u(t) \ge \phi \circ u(t)$$

at every point $t \in (0,+\infty)$. Thus, we may invoke Fact A.2.7.ii and the the estimate given above, to pass to the limit superior ($h \searrow 0$) in (1.46) and yield

$$\frac{\mathrm{d}^-}{\mathrm{d}t}d^2\big(u(t),y\big) := \limsup_{h\searrow 0} \frac{d^2\big(u(t),y\big) + d^2\big(u(t-h),y\big)}{h} \le 2\phi(y) - \lambda d^2\big(u(t),y\big) - 2\phi\big(u(t)\big) \tag{1.47}$$

for *all* $t \in (0,+\infty)$ and $y \in X$. Here, the expression $\frac{\mathrm{d}^-}{\mathrm{d}t}d^2\big(u(t),y\big)$ denotes the *upper left Dini derivative* of the function $t \mapsto d^2\big(u(t),y\big)$.

In a similar fashion, one can integrate (1.45) over the interval $(t,t+h)$ and thus obtain an estimate for the right difference quotient

$$\frac{\mathrm{d}^+}{\mathrm{d}t}d^2\big(u(t),y\big) := \limsup_{h\searrow 0} \frac{d^2\big(u(t+h),y\big) + d^2\big(u(t),y\big)}{h} \le 2\phi(y) - \lambda d^2\big(u(t),y\big) - 2\phi\big(u(t)\big) \tag{1.48}$$

for *all* $t \in (0,+\infty)$ and $y \in X$, where $\frac{\mathrm{d}^+}{\mathrm{d}t}$ denotes the *upper right Dini derivative* of $t \mapsto d^2\big(u(t),y\big)$.

Now setting $u(t) := v(t)$, $y := w(t)$ in (1.47) and $u(t) := w(t)$, $y := v(t)$ in (1.48), and adding both inequalities up, implies

$$\frac{\mathrm{d}^-}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(s),w(t)\big) + \frac{\mathrm{d}^+}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(t),w(s)\big) \le -2\lambda d^2\big(v(t),w(t)\big) \qquad \forall t \in (0,+\infty). \tag{1.49}$$

We want to show that $\frac{\mathrm{d}}{\mathrm{d}t}đ^2(t)$ is a lower bound for the left-hand side of (1.49) on $(0,+\infty)$, where $đ(t) := d\big(v(t),w(t)\big)$. To this aim, note that $đ$ is locally absolutely continuous, due to the estimate

$$\begin{aligned}|đ(s) - đ(t)| &= \big|d\big(w(s),v(s)\big) - d\big(w(s),v(t)\big) + d\big(w(s),v(t)\big) - d\big(w(t),v(t)\big)\big| \le \\ &\le d\big(v(s),v(t)\big) + d\big(w(s),w(t)\big)\end{aligned}$$

for all $s,t \in (0,+\infty)$. Therefore, the map $đ^2$ also belongs to $AC((0,+\infty),\mathbb{R})$ by virtue of Fact 1.1.4.ii.

Next, fix a non-negative bump function $\zeta \in C_c^\infty((0,+\infty))$, a compact interval $[\acute{a},\acute{b}] \subset (0,+\infty)$ and $h > 0$ such that

$$\operatorname{dist}\big(\operatorname{supp}\zeta, \mathbb{R}\setminus(\acute{a},\acute{b})\big) > 2h.$$

Then *change of variables* (see Corollary A.2.8 in Appendix A) implies

$$-\int_0^\infty \bar{d}^2(t)\frac{\zeta(t+h)-\zeta(t)}{h}\,\mathrm{d}\lambda(t) = \int_0^\infty \bar{d}^2(t)\frac{\zeta(t)}{h}\,\mathrm{d}\lambda(t) - \int_0^\infty \bar{d}^2(t)\frac{\zeta(t+h)}{h}\,\mathrm{d}\lambda(t) = \tag{1.50.a}$$

$$= \int_0^\infty \zeta(s)\frac{\bar{d}^2(s)-\bar{d}^2(t-h)}{h}\,\mathrm{d}\lambda(s) = \int_0^\infty \zeta(s)\frac{d^2\big(w(s),v(s)\big)-d^2\big(w(s-h),v(s)\big)}{h}\,\mathrm{d}\lambda(s) + \tag{1.50.b}$$

$$+\int_0^\infty \zeta(s)\frac{d^2\big(w(s-h),v(s)\big)-d^2\big(w(s-h),v(s-h)\big)}{h}\,\mathrm{d}\lambda(s) = \tag{1.50.c}$$

$$= \int_0^\infty \zeta(s)\frac{d^2\big(w(s),v(s)\big)-d^2\big(w(s-h),v(s)\big)}{h}\,\mathrm{d}\lambda(s) + \tag{1.50.d}$$

$$+\int_0^\infty \zeta(r+h)\frac{d^2\big(w(r),v(r+h)\big)-d^2\big(w(r),v(r)\big)}{h}\,\mathrm{d}\lambda(r) \tag{1.50.e}$$

Clearly, one can apply the *dominated convergence theorem* to pass to the limit ($h \to 0$) in the first integral of (1.50.a). On the other hand, we can invoke Definition 1.1.1 for the curve w to obtain a function $m \in L^1_{\mathrm{loc}}((0,+\infty),\mathbb{R})$ and the estimate

$$\big|d\big(w(t),y\big) - d\big(w(t-h),y\big)\big| \le d\big(w(t),w(t-h)\big) \le \int_{t-h}^{t} m\,\mathrm{d}\lambda \qquad \forall t,h \in (0,+\infty), \tag{1.51}$$

uniformly with respect to all maps $t \mapsto d\big(w(t),y\big)$, $y \in X$. Thus we can apply Lemma A.2.3 to arrive at an uniform estimate for the map $t \mapsto d^2\big(w(t),y\big)$, i.e.

$$\big|d^2\big(w(t),y\big) - d^2\big(w(t-h),y\big)\big| \le \int_{t-h}^{t} n\,\mathrm{d}\lambda \qquad \forall y \in X,\ \forall t,h \in (0,+\infty),$$

where n belongs to $L^1_{\mathrm{loc}}((0,+\infty),\mathbb{R})$. Set $\tilde{n}(t) := \mathbb{1}_{(\acute{a}+h,\acute{b}-h)}(t)n(t)$ and assume that Z is a primitive of ζ on $(0,+\infty)$. By virtue of Theorem A.2.4, the real-valued map $t \mapsto \int_{t-h}^{t}\tilde{n}\,\mathrm{d}\lambda$ is absolutely continuous on $(\acute{a},\acute{b})$ with derivative $t \mapsto \tilde{n}(t)-\tilde{n}(t-h)$ λ-a.e. on $(\acute{a},\acute{b})$. Hence *integration by parts* (see Corollary A.2.9 in Appendix A) implies

$$\int_{\acute{a}}^{\acute{b}} \zeta(t)\Big(\int_{t-h}^{t}\tilde{n}\,\mathrm{d}\lambda\Big)\,\mathrm{d}\lambda(t) = -\int_{\acute{a}}^{\acute{b}} \mathrm{Z}(t)\big(\tilde{n}(t)-\tilde{n}(t-h)\big)\,\mathrm{d}\lambda(t) = \tag{1.52.a}$$

$$= \int_{\acute{a}}^{\acute{b}} \mathrm{Z}(t+h)\tilde{n}(t)\,\mathrm{d}\lambda(t) - \int_{\acute{a}}^{\acute{b}} \mathrm{Z}(t)\tilde{n}(t)\,\mathrm{d}\lambda(t) = \int_{\acute{a}}^{\acute{b}} \big(\mathrm{Z}(t+h)-\mathrm{Z}(t)\big)\tilde{n}(t)\,\mathrm{d}\lambda(t), \tag{1.52.b}$$

where we used the fact that $\tilde{n}$ vanishes outside of the interval $(\acute{a}+h,\acute{b}-h)$. Note that we can utilize the *mean value theorem* to obtain an estimate for the integrand on the right hand-side of (1.52.b):

$$\big|\big(\mathrm{Z}(t+h)-\mathrm{Z}(t)\big)\tilde{n}(t)\big| \le h\sup_{s\in\mathbb{R}}\zeta(s)\,|\tilde{n}(t)|\,.$$

As this bound clearly belongs to $L^1((0,+\infty),\mathbb{R})$, we can multiply (1.52) by h^{-1} and invoke the *dominated convergence theorem* to obtain

$$\lim_{h\searrow 0}\frac{1}{h}\int_0^\infty \zeta(t)\Big(\int_{t-h}^{t}\tilde{n}\,\mathrm{d}\lambda\Big)\,\mathrm{d}\lambda(t) = \int_0^\infty \zeta(t)\tilde{n}(t)\,\mathrm{d}\lambda(t). \tag{1.53}$$

Next, choosing $y = v(t)$ in (1.51) establishes

$$A_h(t) := \frac{d^2\big(w(t), v(t)\big) - d^2\big(w(t-h), v(t)\big)}{h} \le \frac{1}{h}\int_{t-h}^{t} \tilde{n}\,\mathrm{d}\lambda =: B_h(t) \qquad \forall t \in \operatorname{supp}\zeta;$$

in particular, we have $B_h - A_h \ge 0$ on $\operatorname{supp}\zeta$. Thus, we can invoke *Fatou's lemma* to pass to the limit superior $(h \searrow 0)$ in (1.50.d):

$$\int_0^\infty \zeta(t)\tilde{n}(t)\,\mathrm{d}\lambda(t) - \limsup_{h\searrow 0}\int_0^\infty \zeta(t)A_h(s)\,\mathrm{d}\lambda(t) = \liminf_{h\searrow 0}\int_0^\infty \zeta(t)\big(B_h(t) - A_h(t)\big)\,\mathrm{d}\lambda(t) \ge$$

$$\ge \int_0^\infty \zeta(t)\liminf_{h\searrow 0}\big(B_h(t) - A_h(t)\big)\,\mathrm{d}\lambda(t) = \int_0^\infty \zeta(t)\tilde{n}(t)\,\mathrm{d}\lambda(t) - \int_0^\infty \zeta(t)\frac{\mathrm{d}^-}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(s), w(t)\big)\,\mathrm{d}\lambda(t),$$

i.e.

$$\limsup_{h\searrow 0}\int_0^\infty \zeta(t)A_h(s)\,\mathrm{d}\lambda(t) \le \int_0^\infty \zeta(t)\frac{\mathrm{d}^-}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(s), w(t)\big)\,\mathrm{d}\lambda(t).$$

Since a similar argument allows one to pass to the limit superior $(h \searrow 0)$ in (1.50.e), we ultimately arrive at

$$\int_0^\infty \zeta(t)\Big(\frac{\mathrm{d}}{\mathrm{d}t}\bar{d}^{\,2}(t)\Big)\,\mathrm{d}\lambda(t) = -\int_0^\infty \Big(\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t)\Big)\bar{d}^{\,2}(t)\,\mathrm{d}\lambda(t) \le \tag{1.54.a}$$

$$\le \int_0^\infty \zeta(t)\frac{\mathrm{d}^-}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(s), w(t)\big)\,\mathrm{d}\lambda(t) + \int_0^\infty \zeta(t)\frac{\mathrm{d}^+}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(t), w(s)\big)\,\mathrm{d}\lambda(t), \tag{1.54.b}$$

where we used *integration by parts* to obtain the equality in (1.54.a). Since our choice of the non-negative test function $\zeta \in C_c^\infty(\mathbb{R})$ was arbitrary in (1.54), the inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{d}^{\,2}(t)\,\mathrm{d}\lambda(t) \le \frac{\mathrm{d}^-}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(s), w(t)\big) + \frac{\mathrm{d}^+}{\mathrm{d}s}\bigg|_{s=t} d^2\big(v(t), w(s)\big) \qquad \lambda\text{-a.e. } t \in (0, +\infty) \tag{1.55}$$

follows. Henceforth, plugging (1.49) and (1.55) together, results in

$$\frac{\mathrm{d}}{\mathrm{d}t}d^2\big(w(t), v(t)\big) + 2\boldsymbol{\lambda}d^2\big(w(t), v(t)\big) \le 0 \qquad \lambda\text{-a.e. } t \in (0, +\infty).$$

Moreover, multiplying this inequality with $\mathrm{e}^{2\boldsymbol{\lambda}t}$ and applying the *chain rule* yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{e}^{2\boldsymbol{\lambda}t}\,d^2\big(w(t), v(t)\big) \le 0 \qquad \lambda\text{-a.e. } t \in (0, +\infty).$$

As this means that the continuous map $t \mapsto \mathrm{e}^{2\boldsymbol{\lambda}t}\,d^2\big(w(t), v(t)\big)$ is non-increasing *everywhere* in $(0, +\infty)$, we conclude that

$$\mathrm{e}^{2\boldsymbol{\lambda}s}\,d^2\big(w(s), v(s)\big) \le \mathrm{e}^{2\boldsymbol{\lambda}t}\,d^2\big(w(t), v(t)\big) \qquad \forall s, t \in (0, +\infty) : s \le t. \tag{1.56}$$

Finally, in the case that the curves v and w have the same initial datum $\lim_{t\searrow 0} v(t) = \lim_{t\searrow 0} v(t) = x_0$, (1.56) shows that v and w necessarily coincide on $(0, +\infty)$. ■

1.4.2 Corollary *Assume that ϕ is lower semi-continuous functional with proper effective domain. If a curve $v \in AC_{\mathrm{loc}}((0, +\infty), X)$ solves* (1.45) *with respect to some $\boldsymbol{\lambda} \in \mathbb{R}$, then v is locally Lipschitz.*

Proof Note that for any fixed $h > 0$ the curve $\hat{v}_h(t) := v(t+h)$ belongs to $AC_{\text{loc}}((0,+\infty),X)$ and solves (1.45) as well. Hence, Proposition 1.4.1 implies that

$$\frac{d\big(v(t),\hat{v}_h(t)\big)}{h} \le \mathrm{e}^{-\lambda(t-s)}\,\frac{d\big(v(s),\hat{v}_h(s)\big)}{h} \qquad \forall s,t \in (0,+\infty) : s \le t. \tag{1.57}$$

Letting $(h \searrow 0)$ in (1.57), we can invoke Theorem 1.1.7 to obtain

$$|v'(t)| \le \mathrm{e}^{-\lambda(t-s)}\,|v'(s)| \qquad \lambda\text{-a.e. } s,t \in (0,+\infty) : s \le t.$$

In particular, we have for any compact interval $[\acute{a},\acute{b}] \subset (0,+\infty)$ and any point $c \in (0,\acute{a})$ where the metric differential $|v'|(s)$ exists, the estimate

$$d\big(v(s),v(t)\big) \le \int_{s\wedge t}^{s\vee t} |v'|\,\mathrm{d}\lambda \le |v'|(c)\,|s-t| \qquad \forall s,t \in [\acute{a},\acute{b}].$$

■

When comparing the notions of (EDI) and (EDE), the following implication is obvious:

1.4.3 **Fact** Let $\phi : X \to \mathbb{R}\cup\{+\infty\}$ be a functional with proper effective domain and $v \in AC_{\text{loc}}((0,+\infty),X)$ be a curve, starting from $x_0 \in X$. If v is a gradient flow in the EDE sense, then v satisfies also (EDI).

On the other hand, the following result, namely, that (EVI) is the strongest of the three notions of gradient flows in a metric setting, is a non-trivial consequence of Proposition 1.4.2.

1.4.4 **Proposition** *Let $\phi : X \to \mathbb{R}\cup\{+\infty\}$ be a lower semi-continuous functional with proper effective domain and $v \in AC_{\text{loc}}((0,+\infty),X)$ be a curve, starting from $x_0 \in X$. If v is a gradient flow in the EVI sense with respect to some $\lambda \in \mathbb{R}$, then v satisfies (EDI) and (EDE).*

Proof By means of Corollary 1.4.2, we may assume that v is a locally Lipschitz curve. In the first part of the proof we will show that in this case the map $\phi\circ v$ is also locally Lipschitz on $(0,+\infty)$:

We start with the *reverse triangle inequality*, implying for any $h > 0$ that

$$\frac{d\big(v(t+h),y\big) - d\big(v(t),y\big)}{h} \ge -\frac{d\big(v(t+h),v(t)\big)}{h} \qquad \forall y \in X,\ \forall t \in (0,+\infty). \tag{1.58}$$

Hence, by taking the limit $(h \searrow 0)$ in the inequality above, we infer for every $y \in X$ that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}d^2\big(v(t),y\big) = d\big(v(t),y\big)\frac{\mathrm{d}}{\mathrm{d}t}d\big(v(t),y\big) \ge -d\big(v(t),y\big)\,|v'|(t) \qquad \lambda\text{-a.e. } t \in (0,+\infty), \tag{1.59}$$

where we used the *chain rule* in the first equality above.

Fix a compact interval $[\acute{a},\acute{b}] \subseteq (0,+\infty)$ and let C be a Lipschitz constant of v on $[\acute{a},\acute{b}]$. Then we have $h^{-1}d\big(v(t+h),v(t)\big) \le L$ for all $t \in [\acute{a},\acute{b}]$. Hence, taking the limit $(h \searrow 0)$ establishes L as an upper bound for the metric differential $|v'|$ λ-a.e. in $[\acute{a},\acute{b}]$. In particular, (1.59) becomes

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}d^2\big(v(t),y\big) \ge -Cd\big(v(t),y\big) \qquad \forall y \in X,\ \lambda\text{-a.e. } t \in (0,+\infty).$$

Now plugging this inequality into (1.45) yields

$$\phi\big(v(t)\big) - \phi(y) \le Cd\big(v(t),y\big) - \frac{\lambda}{2}d^2\big(v(t),y\big) \qquad \forall y \in X, \tag{1.60}$$

for every $t \in [\acute{a},\acute{b}] \setminus N$, where N is a suitable λ-null set in $[\acute{a},\acute{b}]$. Since the lower semi-continuity of ϕ implies

$$\liminf_{\substack{s\to t\\ s\notin N}} \geq \phi(t) \qquad \forall t\in N,$$

we infer that **(1.60)** holds for *all* $t \in [\acute{a}, \acute{b}]$.
Choosing $t = s_1$ and $y = v(s_2)$ in **(1.60)** results in

$$\phi\big(v(s_1)\big) - \phi\big(v(s_2)\big) \leq Cd\big(v(s_1), v(s_2)\big) - \frac{\lambda}{2} d^2\big(v(s_1), v(s_2)\big) \qquad \forall s_1, s_2 \in (0, +\infty),$$

whereas choosing $t = s_2$ and $y = v(s_1)$ result in the same inequality with the terms $\phi\big(v(s_1)\big)$ and $\phi\big(v(s_2)\big)$ on the left-hand side reversed. Thus

$$\begin{aligned}\left|\phi\big(v(s_1)\big) - \phi\big(v(s_2)\big)\right| &\leq Cd\big(v(s_1), v(s_2)\big) - \frac{\lambda}{2} d^2\big(v(s_1), v(s_2)\big) \leq \\ &\leq C^2\left(1 + \frac{|\lambda|}{2}|s_1 - s_2|\right)|s_1 - s_2| \leq C^2\left(1 + \frac{|\lambda|}{2}(\acute{b} - \acute{a})\right)|s_1 - s_2|\end{aligned}$$

for all $s_1, s_2 \in (0, +\infty)$. Hence, we have established the local Lipschitz continuity of $\phi \circ v$.

In particular, the map $\phi \circ v$ has a derivative λ-a.e. in $(0, +\infty)$. Henceforth, the *AM-GM inequality* implies

$$-\frac{\mathrm{d}}{\mathrm{d}t}\phi\big(v(t)\big) = \lim_{h\to 0} \frac{\phi\big(v(t+h)\big) - \phi\big(v(t)\big)}{h} = \tag{1.61.a}$$

$$= \lim_{h\to 0}\left(\frac{\phi\big(v(t+h)\big) - \phi\big(v(t)\big)}{d\big(v(t+h), v(t)\big)} \cdot \frac{d\big(v(t+h), v(t)\big)}{h}\right) \leq \tag{1.61.b}$$

$$\leq \limsup_{h\to 0} \frac{\phi\big(v(t+h)\big) - \phi\big(v(t)\big)}{d\big(v(t+h), v(t)\big)} \limsup_{h\to 0} \frac{d\big(v(t+h), v(t)\big)}{h} \leq \tag{1.61.c}$$

$$\leq |\partial\phi|\left(v(t)\right)|v'|(t) \leq \frac{1}{2}|\partial\phi|^2\left(v(t)\right) + \frac{1}{2}|v'|^2(t) \tag{1.61.d}$$

for λ-a.e. $t \in (0, +\infty)$. Thus we have established that v satisfies a λ-a.e. pointwise formulation of (EDI).

To show that v is also a gradient flow in the EDE sense, we need an estimate for the local slope of ϕ along the curve v:
Plugging **(1.45)** and **(1.59)** together, leads to

$$-d\big(v(t), y\big)|v'|(t) \leq \phi(y) - \phi\big(v(t)\big) - \frac{\lambda}{2} d^2\big(v(t), y\big) \qquad \forall y \in X,\ \lambda\text{-a.e. } t \in (0, +\infty).$$

Therefore, dividing this inequality by $d\big(v(t), y\big)$ and taking the limit superior $(y \to v(t))$ pointwise for every $t \in (0, +\infty)$, gives the desired estimate

$$|\partial\phi|\left(v(t)\right) = \limsup_{y\to v(t)} \frac{\big(\phi\big(v(t)\big) - \phi(y)\big)^+}{d\big(v(t), y\big)} \leq |v'|(t) \qquad \lambda\text{-a.e. } t \in (0, +\infty). \tag{1.62}$$

Now we are well-equipped to prove that v satisfies (EDE):
Fix a compact interval $[\acute{a}, \acute{b}] \subset (0, +\infty)$ and a λ-null set $N \subset [\acute{a}, \acute{b}]$ such that the derivative of $\phi \circ v$ and the metric differential $|v'|$ exist at every point in $[\acute{a}, \acute{b}] \setminus N$. Via integration of **(1.45)** over $(t, t+h)$ for some $h > 0$ and afterwards setting $y = v(t)$, one obtains for every $t \in [\acute{a}, \acute{b}]$ the estimate

$$\frac{1}{2}d^2\big(v(t+h),v(t)\big) \leq \int_t^{t+h} \phi\big(v(t)\big) - \phi\big(v(s)\big)\,\mathrm{d}\lambda(s) - \frac{\lambda}{2}\int_t^{t+h} d^2\big(v(s),v(t)\big)\,\mathrm{d}\lambda(s) \leq \tag{1.63.a}$$

$$\leq \int_t^{t+h} \phi\big(v(t)\big) - \phi\big(v(s)\big)\,\mathrm{d}\lambda(s) - \frac{\lambda}{2}\int_t^{t+h} C^2\,|s-t|^2\,\mathrm{d}\lambda(s) \leq \int_t^{t+h} \phi\big(v(t)\big) - \phi\big(v(s)\big)\,\mathrm{d}\lambda(s) + \frac{|\lambda|\,C^2}{2h^3}, \tag{1.63.b}$$

where we used that v is Lipschitz continuous with Lipschitz constant C on $[\acute{a}-h,\acute{b}+h]$. Moreover, a *change of variables* in the last integral of **(1.63.b)** implies

$$\frac{d^2\big(v(t+h),v(t)\big)}{2h^2} \leq \int_0^1 r\,\frac{\phi\big(v(t)\big) - \phi\big(v(t+hr)\big)}{hr}\,\mathrm{d}\lambda(r) + \frac{|\lambda|\,C^2}{2h} \qquad \forall t \in [\acute{a},\acute{b}].$$

Since the expression $h^{-1}\left|\phi\big(v(t)\big) - \phi\big(v(t+hr)\big)\right|$ is bounded on $[\acute{a},\acute{b}]$ by C , we can invoke the *dominated convergence theorem* to take the limit $(h \to 0)$ in the inequality above and arrive at

$$\frac{1}{2}\left|v'\right|^2(t) \leq -\frac{\mathrm{d}}{\mathrm{d}t}\phi\big(v(t)\big)\int_0^1 r\,\mathrm{d}\lambda(r) = -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\phi\big(v(t)\big) \qquad \forall t \in [\acute{a},\acute{b}] \setminus N.$$

Now we can apply **(1.62)** and yield

$$-\frac{\mathrm{d}}{\mathrm{d}t}\phi\big(v(t)\big) \geq \frac{1}{2}\left|v'\right|^2(t) + \frac{1}{2}\left|v'\right|^2(t) \geq \frac{1}{2}\left|v'\right|^2(t) + \frac{1}{2}\left|\partial\phi\right|\big(v(t)\big) \qquad \lambda\text{-a.e. } t \in (0,+\infty).$$

Together with **(1.61)**, this estimate shows that v satisfies (EDE). ■

The following elementary example shows that the implication in Proposition 1.4.4 cannot be reversed.

1.4.5 Example Consider the space $\left(\mathbb{R}^2, \|\cdot\|_\infty\right)$ and define a smooth functional on $\mathbb{R}^2$ by $\phi(x_1,x_2) := x_1$. Clearly, ϕ is convex and $|\partial\phi| = \|\nabla\phi\|_\infty \equiv 1$.
Next define a family $(v_i)_{i\in\mathbb{R}^+}$ of smooth curves with joint initial datum $(0,0)$ by

$$v_i : [0,+\infty) \to \mathbb{R}^2, \qquad v_i(t) := \left(-t, \frac{t}{1+i}\right) \qquad \forall i \in [0,+\infty),$$

and note that $\phi\big(v_i(t)\big) = -t$ and $|v'|(t) = \|v'(t)\|_\infty = 1$ for all $i \in \mathbb{R}^+$. Now it is immediate to check that every v_i satisfies **(1.44)** as well as **(1.43)**. On the other hand, the lack of uniqueness of the flow curve v_i and Proposition 1.4.1 imply that the family $(v_i)_{i\in\mathbb{R}^+}$ does not belong to (EVI) as $v(t) := (t,0)$ depicts the unique gradient flow for ϕ in the EVI sense. In particular, Proposition 1.3.3 cannot be applied since the norm $\|\cdot\|_\infty$ does not induce any inner product on $\mathbb{R}^n$. ◀

The difference between (EDI) and (EDE) in general metric setting – apart from Fact 1.4.3 – is more subtle. We just refer to Example 3.15 in [2]. By making use of the *minimizing movement scheme* which will be introduced in the next section, this example shows that there exists a gradient flow in the EDI sense, that does not satisfy (EDE).

1.5 A Glimpse on the Minimizing Movements Scheme

In this short section we investigate an discrete approximating scheme which plays a major role in the existence theory of gradient flows in metric spaces.
At first we introduce a uniform partition of $(0,+\infty)$:

Notation

Denote with $P_\tau := (n\tau)_{n\in\mathbb{N}_0}$ the uniform partition of $(0,+\infty)$ into left-open, right-closed intervals $I_\tau^n := \big((n-1)\tau, n\tau\big]$, $n \in \mathbb{N}$ of size $\tau > 0$.

1.5.1 **Definition** Let a lower semi-continuous functional $\phi : X \to \mathbb{R}\cup\{+\infty\}$ with proper effective domain $\operatorname{dom}\phi$ be given, where (X,d) is a Polish metric space. Define the functional

$$\Phi : \mathbb{R}^+ \times \operatorname{dom}\phi \times X \longrightarrow \mathbb{R}\cup\{+\infty\} \tag{1.64.a}$$

$$(\tau, M, x) \longmapsto \frac{1}{2\tau} d^2(x,M) + \phi(x). \tag{1.64.b}$$

For any given *time step* $\tau > 0$ and *discrete initial datum* $M_\tau^0 \in \operatorname{dom}\phi$, a ***τ-discrete minimizing movement*** starting from M_τ^0 is a sequence $(M_\tau^n)_{n\in\mathbb{N}}$ in $\operatorname{dom}\phi$ which satisfies

$$\Phi(\tau, M_\tau^{n-1}, M_\tau^n) \leq \Phi(\tau, M_\tau^{n-1}, x) \qquad \forall x \in X,\ \forall n \in \mathbb{N}.$$

A ***discrete solution*** is the piecewise constant interpolant

$$\overline{M}_\tau(t) := \sum_{n=1}^{\infty} M_\tau^n \mathbb{1}_{I_\tau^n}(t) \qquad \forall t \in (0,+\infty).$$

In general, the existence of a discrete minimizing movement $(M_\tau^n)_{n\in\mathbb{N}}$ cannot be assured without further assumption on the functional ϕ. However, in case of existence of such a sequence for every $\tau > 0$, one hopes to find a limit curve as $(\tau \searrow 0)$ which satisfies the definition of a gradient flow in some sense.

For instance, a rather elementary convergence result could be obtained if one requires all sublevel sets of ϕ to be locally compact. Then there exists a sequence of times steps $(\tau_n)_{n\in\mathbb{N}}$ with corresponding discrete minimizing movements such that their discrete solutions converge to a gradient flow in the EDI sense as $(\tau_n \searrow 0)$ (cf. Theorem 3.14 in [2] or Corollary 2.4.11 in [3]).

Nevertheless, we are interested in the stronger notion of (EVI) gradient flows for the remainder of this section. We have already noted in Section 1.3 (see Proposition 1.3.4) that gradient flows in the sense of (EVI) are closely related to the λ-convexity of the underlying functional ϕ. However, in a general metric setting the existence of such a gradient flow does not only depend on ϕ but also on the geometrical structure of the metric space (X,d). This leads to the following the following definitions.

1.5.2 **Definition** Let (X,d) be a metric space. A curve $\gamma : [0,1] \to X$ is called ***(constant-speed) geodesic*** if

$$d\big(\gamma(s),\gamma(t)\big) = |t-s|\, d\big(\gamma(0),\gamma(1)\big) \qquad \forall s,t \in [0,1].$$

We call a functional $\phi : X \to \mathbb{R}\cup\{+\infty\}$ ***λ-convex along a curve*** $\gamma : [0,1] \to X$ if

$$\phi\big(\gamma(t)\big) \leq (1-t)\phi\big(\gamma(0)\big) + t\phi\big(\gamma(1)\big) - \frac{\lambda}{2} t(1-t) d^2\big(\gamma(0),\gamma(1)\big) \qquad t \in [0,1]. \tag{1.65}$$

In particular, ϕ is called ***geodesically λ-convex*** if for every pair of points $x,y \in \operatorname{dom}\phi$ there exists a geodesic γ with end-points $\gamma(0) = x$, $\gamma(1) = y$ such that ϕ is λ-convex along γ.

The definitions presented so far allow us to cite the following theorem due to Ambrosio, Gigli, Savaré regarding the existence of gradient flows.

1.5.3 **Theorem** (Existence of EVI gradient flows) *Let (X,d) be a Polish metric space and $\phi : X \to \mathbb{R}\cup\{+\infty\}$ be a lower semi-continuous functional with proper effective domain $\operatorname{dom}\phi$. Assume that Φ defined in* (1.64) *satisfies the following property:*

(CON) *For every triple of points $M, x_0, x_1 \in \operatorname{dom}\phi$ there exists a curve $\gamma : [0,1] \to X$ with end-points $\gamma(0) = x_0$, $\gamma(1) = x_1$ such that $\Phi(\tau, M, \cdot)$ is $(\tau^{-1} + \lambda)$-convex along γ for every $\tau > 0$ with $\frac{1}{\tau} > -\min\{0, \lambda\}$.*

Then the following statements hold:

(i) *For every discrete initial datum $M_\tau^0 = u_0 \in \overline{\operatorname{dom}\phi}$ and every time step $\tau > 0$ with $1 + \tau\lambda > 0$ there exists a discrete minimizing movement $(M_\tau^n)_{n\in\mathbb{N}}$ in $\operatorname{dom}\phi$.*

(ii) *The corresponding discrete solutions $\overline{M}_\tau$ converge compactly to a limit curve $u \in AC_{\mathrm{loc}}((0,+\infty), X)$ as $(\tau \searrow 0)$.*

(iii) *The limit curve u is the unique gradient flow in the EVI sense, starting from u_0.*

(iv) *For every $T \geq 0$ there exists a constant $C_{\lambda,T} > 0$ such that the following error estimate holds:*

$$d\big(u(t), \overline{M}_\tau(t)\big) \leq C_{\lambda,T}\, |\partial\phi|\,(u_0)\tau \qquad \forall t \in [0, T].$$

A proof of this theorem may be found in Chapter 5 of [3]. For a more accessible sketch of proof see also Theorem 3.25 in [2].

Finally, we mention a geometrical class of metric spaces in which the property (CON) seems to be very natural.

1.5.4 Remark (Non-positively curved geodesic spaces) We call a metric space (X, d) *geodesic space* if for every pair of points $x_0, x_1 \in X$ there exists a constant-speed geodesic γ with end-points $\gamma(0) = x_0$, $\gamma(1) = x_1$.

Then a geodesic space is said to be *non-positively curved (NPC) in the sense of Alexandrov* if for every constant speed geodesic γ and every point $x \in X$ the following inequality holds:

$$d^2\big(\gamma(t), x\big) \leq (1-t)d^2\big(\gamma(0), x\big) + td^2\big(\gamma(1), x\big) - t(1-t)d^2\big(\gamma(0), \gamma(1)\big) \qquad \forall t \in [0,1]. \tag{1.66}$$

Clearly, above inequality holds precisely when the functional $\frac{1}{2}d^2(\cdot, x)$ is 1-convex along γ.

Now assume that a geodesically λ-convex functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ on an NPC space (X, d) is given and fix $\tau > 0$. Then for every pair of points $x_0, x_1 \in X$ there exists a geodesic γ with end-points $\gamma(0) = x_0$, $\gamma(1) = x_1$ such that (1.65) holds. Therefore, we may add the inequalities (1.66) multiplied by $\frac{1}{2\tau}$ and (1.65) up to obtain that the functional $\Phi(\tau, M, \cdot)$ is $(\tau^{-1} + \lambda)$-convex along γ. Thus, in NPC spaces geodesics seem to be the natural choice of curves required in (CON).

However, it turns out that Wasserstein spaces over $\mathbb{R}^n$ do not satisfy (1.66). Hence, a different class of curves has to be considered in such spaces.

More information regarding NPC spaces and the geometry of metric spaces may be found in Papadopoulos [44] or in Burago, Burago, Ivanov [19].

2 Gradient Flows in Wasserstein Spaces

2.1 The Kantorovich Transportation Problem

Our starting point in this chapter is the *Kantorovich transportation problem*.
Recall that for any two given measurable spaces $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$, $(X_1 \times X_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ denotes the *product measurable space* equipped with the *tensor-product σ-algebra* $\mathcal{A}_1 \otimes \mathcal{A}_2$, i.e. the σ-algebra on $X_1 \times X_2$ generated by $\mathcal{A}_1 \times \mathcal{A}_2$. Since the projections $\pi^i : X_1 \times X_2 \to X_i$, $i \in \{1, 2\}$ are measurable maps, we can consider the *pushforward* $\pi^i_* \sigma := \sigma \circ (\pi^i)^{-1}$ of an arbitrary measure σ on $(X_1 \times X_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, which then induces a measure on $(X_i, \mathcal{A}_i)$.

Notation

For any two probability measures μ_1 on $(X_1, \mathcal{A}_1)$ and μ_2 on $(X_2, \mathcal{A}_2)$ we denote by $\Pi(\mu_1, \mu_2)$ the set of all probability measures σ on $(X_1 \times X_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ such that $\pi^i_* \sigma = \mu_i$ for $i \in \{1, 2\}$.

2.1.1 Definition (Kantorovich problem) Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be two measure spaces and assume that there is given a measurable map $h : X_1 \times X_2 \to \mathbb{R}_0^+ \cup \{+\infty\}$. The elements of $\Pi(\mu_1, \mu_2)$ are called ***admissible plans (of transportation)***. We say that $\sigma_{\min} \in \Pi(\mu_1, \mu_2)$ is an ***optimal plan (of transportation)*** if $\sigma_{\min}$ minimizes the functional

$$K(\mu_1, \mu_2, \sigma) := \int_{X_1 \times X_2} h \,\mathrm{d}\sigma \in \mathbb{R}_0 \cup \{+\infty\}\,, \tag{2.1}$$

i.e. $K(\mu_1, \mu_2, \sigma_{\min}) = \inf_{\sigma \in \Pi(\mu_1, \mu_2)} K(\mu_1, \mu_2, \sigma)$. The map h is called ***cost function*** of the Kantorovich problem (2.1).

It is clear that there always exists an admissible plan for the Kantorovich problem since the product measure $\mu_1 \times \mu_2$ belongs to $\Pi(\mu_1, \mu_2)$. However, note that $\inf_{\sigma \in \Pi(\mu_1, \mu_2)} K(\mu_1, \mu_2, \sigma)$ need not be finite. Nevertheless, the Kantorovich problem has a solution under rather general assumptions:

2.1.2 Theorem *For any two Radon probability measures μ_1 and μ_2 on Tychonoff spaces $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$, and any lower semi-continuous cost function $h : X_1 \times X_2 \to \mathbb{R}_0 \cup \{+\infty\}$ the Kantorovich problem admits an optimal plan.*

The proof of this result relies heavily on the fact that weak convergence of measures is closely related to the concept of *uniformly tight* measures, introduced in Section A.4 of Appendix A.

2.1.3 Lemma *Let $(\mu_i)_{i \in I}$ and $(\nu_i)_{i \in I}$ be uniformly tight families of Radon probability measures on topological spaces $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$. Then the set $\bigcup_{i \in I} \Pi(\mu_i, \nu_i)$ is uniformly tight.*

Proof For every $\varepsilon > 0$ there exist compact sets $K_i \subseteq X_i$ such that $\mu_i(X_1 \setminus K_1) < \varepsilon/2$ and $\nu_i(X_2 \setminus K_2) < \varepsilon/2$ for every $i \in I$. Therefore, we have

$$\begin{aligned}
\sigma_i\big((X_1 \times X_2) \setminus (K_1 \times K_2)\big) &= \sigma_i\left(\left(\pi^1\right)^{-1}(X_1 \setminus K_1) \cap \left(\pi^2\right)^{-1}(X_2 \setminus K_2)\right) \leq \\
&\leq \sigma_i\left(\left(\pi^1\right)^{-1}(X_1 \setminus K_1)\right) + \sigma_i\left(\left(\pi^2\right)^{-1}(X_2 \setminus K_2)\right) = \\
&= \pi^1_* \sigma_i(X_1 \setminus K_1) + \pi^2_* \sigma_i(X_2 \setminus K_2) = \mu_i(X_1 \setminus K_1) + \nu_i(X_2 \setminus K_2) < \varepsilon
\end{aligned}$$

uniformly for all $\sigma_i \in \Pi(\mu_i, \nu_i)$, $i \in I$. Clearly, $K_1 \times K_2$ is compact in $X_1 \times X_2$, due to *Tychonoff's theorem*. ■

The following result (cf. Theorem 8.6.7 in [12]) shows that one implication of *Prokhorov's theorem* (cf. Theorem A.4.7 in Appendix A) also holds under the more general assumption that the underlying space is merely Tychonoff.

2.1.4 Lemma *Let $\mathcal{N}$ be a uniformly tight family of non-negative Radon measures on a Tychonoff space $(X, \mathcal{T})$. If $\mathcal{N}$ is uniformly bounded in total variation, then the closure of $\mathcal{N}$ in the weak-* topology $\sigma\big(C_b'(X), C_b(X)\big)$ is compact and can be represented as a family of non-negative Radon measures.*

Proof We consider $\mathcal{N}$ as a linear subspace of the topological dual space $C_b'(X)$. As $\mathcal{N}$ is uniformly bounded in total variation, it is also bounded in $C_b'(X)$. Thus, we can invoke the *Banach-Alaoglu theorem* to infer that the closure of $\mathcal{N}$ is compact in $C_b'(X)$ with respect to the weak-* topology $\sigma\big(C_b'(X), C_b(X)\big)$.
Now let $(\sigma_i)_{i\in I}$ be a net in $\mathcal{N}$ such that $(\sigma_i)_{i\in I}$ converges weakly to some functional F in $C_b'(X)$. We have to show that F can be represented by a non-negative Radon measure. At this point the uniform tightness of $\mathcal{N}$ comes into play: For every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq X$ such that $|\sigma|(K_\varepsilon^c) < \varepsilon$ for all $\sigma \in \mathcal{N}$. Thus

$$|F(f)| = \left|\lim_{i\in I}\Big(\int_{K_\varepsilon^c} f \,\mathrm{d}\sigma_i\Big)\right| \leq \limsup_{i\in I}\left|\int_{K_\varepsilon^c} f \,\mathrm{d}\sigma_i\right| \leq \varepsilon \,\|f\|_\infty \qquad \forall f \in C_b(X) : f\big|_{K_\varepsilon} = 0$$

shows that condition (RAD) in Theorem A.3.5 is satisfied. Moreover, the functional F is positive, i.e. $F(f) \geq 0$ for all $f \in C_b(X_1 \times X_2)$, $f \geq 0$ since every σ_i, $i \in I$ depicts a positive functional in $C_b'(X)$. Therefore, F is represented by some non-negative Radon measure σ on X and $\sigma_i \xrightarrow{w^*} \sigma$, i.e.

$$\lim_{i\in I}\int_X f \,\mathrm{d}\sigma_i = \int_X f \,\mathrm{d}\sigma \qquad \forall f \in C_b(X)\,.$$

■

Proof of Theorem 2.1.2 First observe that the projections π^i, $i \in \{1, 2\}$ separate points on the product space $X_1 \times X_2$. Hence, $X_1 \times X_2$ inherits the Tychonoff property from the spaces X_1 and X_2. Clearly, for every $\sigma \in \Pi(\mu_1, \mu_2)$ we have total variation

$$|\mu|(X_1 \times X_2) = \mu(X_1 \times X_2) = 1.$$

Hence, $\Pi(\mu_1, \mu_2)$ is bounded in $C_b'(X_1 \times X_2)$. Moreover, $\Pi(\mu_1, \mu_2)$ is uniformly tight since the Radon property assures that each μ_1 and μ_2 is (uniformly) tight and we may invoke Lemma 2.1.3. As a result, Lemma 2.1.4 implies that $\Pi(\mu_1, \mu_2)$ has compact closure in the weak-* topology $\sigma\big(C_b'(X_1 \times X_2), C_b(X_1 \times X_2)\big)$.
Now we have to show that $\Pi(\mu_1, \mu_2)$ is closed and therefore compact in $C_b'(X_1 \times X_2)$. Let $(\sigma_i)_{i\in I}$ be a net in $\Pi(\mu_1, \mu_2)$ such that $(\sigma_i)_{i\in I}$ converges weakly to some non-negative Radon measure σ, i.e.

$$\lim_{i\in I}\int_{X_1\times X_2} f \,\mathrm{d}\sigma_i = \int_{X_1\times X_2} f \,\mathrm{d}\sigma \qquad \forall f \in C_b(X_1 \times X_2)\,. \tag{2.2}$$

In particular, we may choose $f \equiv 1$ in (2.2) to observe that $\sigma(X_1 \times X_2) = 1$. This shows that σ is a probability measure.
Since the projections π^1 and π^2 are continuous maps, we have

$$\int_{X_j} f \,\mathrm{d}\mu_j = \lim_{i\in I}\Big(\int_{X_j} f \,\mathrm{d}\pi^j_*\sigma_i\Big) = \lim_{i\in I}\Big(\int_{X_1\times X_2} f\circ\pi^j \,\mathrm{d}\sigma_i\Big) = \int_{X_1\times X_2} f\circ\pi^j \,\mathrm{d}\sigma = \int_{X_j} f \,\mathrm{d}\pi^j_*\sigma \qquad \forall f \in C_b(X_j)$$

for $j \in \{1, 2\}$. Hence, μ_j and $\pi^j_*\sigma$ represent the same functional in $C_b'(X_j)$. Thus, Theorem A.3.5 implies that $\mu_j = \pi^j_*\sigma$ on X_j for $j \in \{1, 2\}$. We conclude that σ belongs to $\Pi(\mu_1, \mu_2)$, which shows that $\Pi(\mu_1, \mu_2)$ is compact with respect to the weak-* topology.

Finally, it remains to prove that the functional

$$\begin{aligned}\Pi(\mu_1,\mu_2) &\longrightarrow \mathbb{R}\\ \sigma &\longmapsto K(\mu_1,\mu_2,\sigma)\end{aligned}$$

is lower semi-continuous on $\Pi(\mu_1,\mu_2)$ with respect to the weak-* topology. Once this is shown, it is easy to see that $K(\mu_1,\mu_2,\cdot)$ admits a minimum on the compact set $\Pi(\mu_1,\mu_2)$ (see a.e. Theorem 7.3.1 in [36]).

In case, the lower semi-continuous cost function h is bounded, Proposition A.4.5.ii immediately implies

$$\liminf_{\varsigma\to\sigma}\int_{X_1\times X_2} h\,\mathrm{d}\varsigma \geq \int_{X_1\times X_2} h\,\mathrm{d}\sigma \qquad \forall\sigma\in\Pi(\mu_1,\mu_2),$$

which is exactly the lower semi-continuity of $K(\mu_1,\mu_2,\cdot)$.

In the general situation, let $(h_n)_{n\in\mathbb{N}}$ be a sequence of bounded functions $h_n := \min(h,n)$. Apparently, we have

$$\liminf_{y\to x} h_n(y) = \min\big(\liminf_{y\to x} h(y),n\big) \geq \min\big(h(x),n\big) = h_n(x) \qquad \forall x\in X_1\times X_2,\ \forall n\in\mathbb{N},$$

which implies the lower semi-continuity of each h_n. Since we have already observed that the functionals

$$K_n(\sigma) := \int_{X_1\times X_2} h_n\mathrm{d}\sigma \qquad \forall\sigma\in\Pi(\mu_1,\mu_2),\ n\in\mathbb{N}$$

are lower semi-continuous for bounded cost functions h_n, we can invoke *monotone convergence* to write $K(\mu_1,\mu_2,\cdot)$ as

$$K(\mu_1,\mu_2,\sigma) = \int_{X_1\times X_2} \sup_{n\in\mathbb{N}} h_n\,\mathrm{d}\sigma = \sup_{n\in\mathbb{N}}\int_{X_1\times X_2} h_n\,\mathrm{d}\sigma \qquad \forall\sigma\in\Pi(\mu_1,\mu_2).$$

Hence, lower semi-continuity of $K(\mu_1,\mu_2,\cdot)$ follows. ■

Now we bring up an important relationship of the Kantorovich problem and its dual problem.

2.1.5 **Remark** (Kantorovich duality) In the setting of Definition 2.1.1, we formulate the related *dual problem* as follows: Consider the functional

$$J(\mu,\nu,\varphi,\psi) := \int_{X_1}\varphi\,\mathrm{d}\mu + \int_{X_2}\psi\,\mathrm{d}\nu \qquad \forall(\varphi,\psi)\in L^1(X_1,\mu,\mathbb{R})\times L^1(X_2,\nu,\mathbb{R})\,.$$

Then a pair $(\tilde\varphi,\tilde\psi)$ in the set

$$\Phi_h := \big\{(\varphi,\psi)\in L^1(X_1,\mu,\mathbb{R})\times L^1(X_2,\nu,\mathbb{R}) : \varphi(x_1)+\psi(x_2)\leq h(x_1,x_2)\big\}$$

is called *optimal* if $J(\mu,\nu,\tilde\varphi,\tilde\psi) = \sup_{(\varphi,\psi)\in\Phi_h} J(\mu,\nu,\varphi,\psi)$.

Now the *Kantorovich duality* asserts that under certain conditions the optimal value of the functional K equals the optimal value of J. The earliest result of such an equality goes back to Kantorovich's seminal paper [32]. Indeed, the dual problem is vividly inspired by linear programming where duality is well understood. However, the class of spaces, for which equality between the Kantorovich and its dual problem is archived, is considerably large (see a.e. a recent series of papers [7], [8], [9], [10] by Beiglböck, Schachermayer and Léonard).

Here we cite a version due to D. Edwards which admits a rather simple proof (cf. Theorem 4.1 in [28]):

Let $h : X_1 \times X_2 \to \mathbb{R}_0^+ \cup \{+\infty\}$ *be a lower semi-continuous cost function. Then*

$$\min_{\sigma \in \Pi(\mu_1,\mu_2)} K(\mu_1,\mu_2,\sigma) = \sup_{(\varphi,\psi)\in\Phi_h} J(\mu,\nu,\varphi,\psi). \tag{2.3}$$

Note that the value on the right-hand side of (2.3) need not be attained in Φ_h and the value $+\infty$ is not excluded.

The statement also holds for other function classes than Φ_h. For instance, one may consider

$$\Psi_h := \{(\varphi,\psi) \in \mathcal{B}_b(X_1) \times \mathcal{B}_b(X_2) : \varphi(x_1) + \psi(x_2) \le h(x_1,x_2)\},$$

where $\mathcal{B}_b(X_i)$ denotes all bounded Borel measurable real-valued functions on X_i, for $i \in \{1,2\}$. Then one may invoke *monotone convergence* to obtain an analogous result (cf. Corollary 3.2 in [29]):

Let $h : X_1 \times X_2 \to \mathbb{R}_0^+ \cup \{+\infty\}$ *be a lower semi-continuous cost function. Then*

$$\min_{\sigma \in \Pi(\mu_1,\mu_2)} K(\mu_1,\mu_2,\sigma) = \sup_{(\varphi,\psi)\in\Psi_h} J(\mu,\nu,\varphi,\psi). \tag{2.4}$$

At the end of this section, we mention the *Monge problem*, a transportation problem closely related to the one of Kantorovich.

2.1.6 **Remark** (Monge Problem) In the setting of the Kontorovich problem, consider two measure spaces $(X_1,\mathcal{A}_1,\mu_1)$ and $(X_2,\mathcal{A}_2,\mu_2)$ and let $h : X_1 \times X_2 \to \mathbb{R}_0^+ \cup \{+\infty\}$ be a measurable cost function. We denote by $\mathrm{T}(\mu_1,\mu_2)$ the class of all measurable maps $t : X_1 \to X_2$ with pushforward $t_*\mu_1 = \mu_2$. The elements in $\mathrm{T}(\mu_1,\mu_2)$ are called *admissible transport maps*.

A transport map $t_{\min} \in \mathrm{T}(\mu_1,\mu_2)$ is called *optimal* if it minimizes the *Monge problem*

$$M(\mu_1,\mu_2,t) := \int_{X_1} h\big(x,t(x)\big)\,\mathrm{d}\mu_1(x) \in \mathbb{R}_0 \cup \{+\infty\}, \tag{2.5}$$

i.e. $M(\mu_1,\mu_2,t_{\min}) = \inf_{t\in\mathrm{T}(\mu_1,\mu_2)} M(\mu_1,\mu_2,t)$.

It is clear, that every given transport map $t_{\min} \in \mathrm{T}(\mu_1,\mu_2)$ induces an admissible plan in the set $\Pi(\mu_1,\mu_2)$ of the corresponding Kantorovich problem by means of the pushforward $(\mathrm{Id},t)_*\mu_1$. However, unlike the Kantorovich problem, the Monge problem may not admit an optimal solution in even very simple settings.

For example, one may consider $X_1 = X_2 = [-1,1]$ with the *quadratic cost function* $h(x_1,x_2) := |x_1 - x_2|^2$ and measures $\mu_1 := \delta_0$, $\mu_2 := 2^{-1}(\delta_{-1} + \delta_1)$. Then $\Pi(\mu_1,\mu_2)$ consists only of one admissible plan $\sigma = 2^{-1}(\delta_{(0,-1)} + \delta_{(0,1)})$. Hence, the Kantorovich problem admits a unique solution.

On the other hand, there exists no admissible transport map $t \in \mathrm{T}(\mu_1,\mu_2)$ in the corresponding Monge problem since t would be required to take values at ± 1 at the same time (see Figure 2.1).

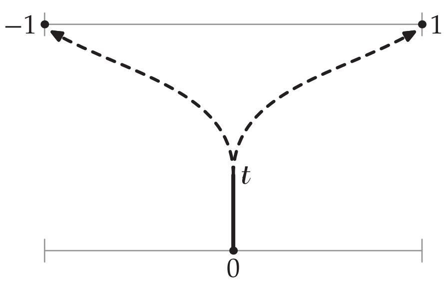


Figure 2.1 When μ_1 charges single points, a single-valued transport map t may not exist. This means that the Monge problem does not admit any split of mass, whereas mass splitting transport plans are generally admissible in the corresponding Kantorovich problem.

The key argument in the example given above is the fact that the measure μ_1 gives mass to a single point. If one avoids such situations, one can expect results on the existence of an optimal transport map.

We cite the following theorem due to Brennier [15] and McCann [40]. To this aim, recall that every convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous. In particular, the coordinate functions φ^i, $i \in \{1,\dots n\}$ are absolutely continuous on every compact interval

$[a,b] \subset \mathbb{R}$. Therefore, Fact A.2.5.i implies that the gradient $\nabla\varphi$ exists λ-a.e. in $\mathbb{R}^n$. Moreover, we denote by $\varphi^\star(x) := \sup_{y\in\mathbb{R}^n}\{\langle x,y\rangle - \varphi(y)\}$ the *convex conjugate* of φ.

2.1.7 **Theorem** (Brennier) *Let μ_1 and μ_2 be two Borel probability measures on $\mathbb{R}^n$ such that $|\cdot|^2 \in L^1(\mathbb{R}^n,\mu_i,\mathbb{R})$ for $i \in \{1,2\}$ and $\mu_1(B) = 0$ for every λ_n-null set $B \in \mathcal{B}(\mathbb{R}^n)$. Consider the quadratic cost function $h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_0^+$, $h(x_1,x_2) := |x_1 - x_2|^2$. Then the following statements hold:*

(i) *There exists a Borel measurable transport map $t \in \mathrm{T}(\mu_1,\mu_2)$, such that $t = \nabla\varphi$ for some convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$. The transport map t is uniquely determined up a μ_1-null set in $\mathbb{R}^n$.*

(ii) *$t = \nabla\varphi$ is the unique optimal transport map of the corresponding Monge problem.*

(iii) *Under the additional assumption that $\mu_2(B) = 0$ for every λ_n-null set $B \in \mathcal{B}(\mathbb{R}^n)$, there exists a map $\varsigma \in \mathrm{T}(\mu_2,\mu_1)$ with $\varsigma = \nabla\varphi^\star$ such that $\nabla\varphi^\star \circ \nabla\varphi = \mathrm{Id}$ μ_1-a.e. in $\mathbb{R}^n$ and $\nabla\varphi \circ \nabla\varphi^\star = \mathrm{Id}$ μ_2-a.e. in $\mathbb{R}^n$.*

For a proof see a.e. Theorem 2.12 in Villani [46].

2.2 The Structure of Wasserstein Spaces

Consider a Radon probability measure μ on a metric space (X,d) and set $đ_y(x) := d(x,y)$ for $x,y \in X$. It is clear that in the case that $đ_{y_0}$ belongs to $L^p(X,\mu,\mathbb{R})$ for some $y_0 \in X$, the triangle inequality

$$đ_y(x) = d(x,y) \le d(x,y_0) + d(y,y_0) \qquad \forall x,y \in X$$

implies $đ_y \in L^p(\mu,X,\mathbb{R})$ for *all* $y \in X$. This justifies the following notation.

Notation

Given a metric space (X,d), $\mathcal{P}_r^p(X)$ denotes the set of all Radon probability measures μ such that $x \mapsto d(x,y_0)$ belongs to $L^p(X,\mu,\mathbb{R})$ for some $y_0 \in X$. The set $\mathcal{P}_r^p(X)$ does not depend on the choice of $y_0 \in X$.

The set $\mathcal{P}_r^p(X)$ can be equipped with a certain family of metrics. The idea is to consider the Kantorovich problem (2.1) with the metric d or, more generally, d^p, $p \ge 1$ as cost functions. As every metric space is Tychonoff, there always exists an optimal plan.

2.2.1 **Definition** Let (X,d) be a metric space. Then for every $p \ge 1$ the quantity

$$W_p(\mu,\nu) := \inf_{\sigma\in\Pi(\mu,\nu)}\Big(\int_{X\times X} d^p(x,y)\,\mathrm{d}\sigma(x,y)\Big)^{1/p} \qquad \forall \mu,\nu \in \mathcal{P}_r^p(X) \tag{2.6}$$

is called ***Wasserstein distance*** or ***Kantorovich distance*** of order p on $\mathcal{P}_r^p(X)$. The space $\big(\mathcal{P}_r^p(X), W_p\big)$ is called ***Wasserstein space*** of order p over X.

The fact that W_p defines a metric on $\mathcal{P}_r^p(X)$ for all $p \ge 1$ is not completely trivial.

2.2.2 **Proposition** *Let (X,d) be a metric space. Then for every $p \ge 1$ the Wasserstein distance W_p defines a metric on $\mathcal{P}_r^p(X)$.*

For the verification that W_p satisfies the triangle inequality, we need the following lemma which assures that certain compatible measures can be "glued together".

2.2.3 **Lemma** (Coupling) *Let X_1, X_2, X_3 be Tychonoff spaces and assume, there are two Radon probability measures $\sigma_{1,2}$ and $\sigma_{2,3}$ on the product spaces $X_1 \times X_2$ and $X_2 \times X_3$ with projections ${}^{1,2}\pi^2 : X_1 \times X_2 \to X_2$ and ${}^{2,3}\pi^2 : X_2 \times X_3 \to X_2$ such that*

$${}^{1,2}\pi^2_*\sigma_{1,2} = {}^{2,3}\pi^2_*\sigma_{2,3}.$$

Then on the product space $X_1 \times X_2 \times X_3$ there exists a Radon probability measure η with projections $\pi^{1,2} : X_1 \times X_2 \times X_3 \to X_1 \times X_2$ and $\pi^{2,3} : X_1 \times X_2 \times X_3 \to X_2 \times X_3$ such that

$$\pi^{1,2}_* \eta = \sigma_{1,2} \qquad \textit{and} \qquad \pi^{2,3}_* \eta = \sigma_{2,3}.$$

Proof In the first step we assume that the spaces X_1, X_2, X_3 are compact. Consider the linear subspace

$$\mathcal{C} := \{f(x_1,x_2) + g(x_2,x_3) : f \in C(X_1 \times X_2), g \in C(X_2 \times X_3)\} \subset C(X_1 \times X_2 \times X_3),$$

where we can define the linear functional

$$L(f+g) := \int_{X_1 \times X_2} f \, \mathrm{d}\sigma_{1,2} + \int_{X_2 \times X_3} f \, \mathrm{d}\sigma_{2,3} \qquad \forall (f+g) \in \mathcal{C}.$$

We have to assure that L is well-defined. To this aim, let $f+g$ and $\tilde{f}+\tilde{g}$ be two representatives of the same element in $\mathcal{C}$, i.e.

$$f(x_1,x_2) + g(x_2,x_3) = \tilde{f}(x_1,x_2) + \tilde{g}(x_2,x_3).$$

Then $f - \tilde{f}$ does not depend on x_1 and we can write $\varphi(x_2) = f(x_1,x_2) - \tilde{f}(x_1,x_2)$. This means

$$\int_{X_1 \times X_2} f - \tilde{f} \, \mathrm{d}\sigma_{1,2} = \int_{X_1 \times X_2} \varphi \, \mathrm{d}\sigma_{1,2} = \int_{X_2 \times X_3} \varphi \, \mathrm{d}\sigma_{2,3} = \int_{X_2 \times X_3} f - \tilde{f} \, \mathrm{d}\sigma_{2,3},$$

which implies

$$L(f+g) - L(\tilde{f}+\tilde{g}) = \int_{X_1 \times X_2} f - \tilde{f} \, \mathrm{d}\sigma_{1,2} + \int_{X_2 \times X_3} g - \tilde{g} \, \mathrm{d}\sigma_{2,3} = \int_{X_2 \times X_3} f - \tilde{f} + g - \tilde{g} \, \mathrm{d}\sigma_{2,3} = 0.$$

Hence, L is well-defined.

Next, we show that the linear functional L is bounded: For any $(f+g) \in \mathcal{C}$ such that $f+g \leq 1$ we have

$$f(x_1,x_2) + \sup_{x_3 \in X_3} g(x_2,x_3) \leq 1 \qquad \text{and} \qquad g(x_2,x_3) - \sup_{x_3 \in X_3} g(x_2,x_3) \leq 0.$$

Thus, we obtain

$$f(x_1,x_2) + g(x_2,x_3) = f(x_1,x_2) + \sup_{x_3 \in X_3} g(x_2,x_3) + g(x_2,x_3) - \sup_{x_3 \in X_3} g(x_2,x_3) \leq 1.$$

Together with $L(1) = 1$, this estimate implies that $\|L\| = \sup_{\|f+g\|_\infty \leq 1} L(f+g) = 1$, which means that L is a contraction.

Now we can invoke (a corollary of) the *Hahn-Banach theorem* to extend L to a linear functional $Ł$ on $C(X_1 \times X_2 \times X_3)$ with norm $\|Ł\| = 1$. By virtue of Theorem A.3.3, the functional $Ł$ can be represented by some unique Radon measure ν with total variation $|\nu|(X_1 \times X_2 \times X_3) = 1$. Since $Ł(1) = L(1) = 1$, it is clear that ν depicts a Radon probability measure.

Finally, we obtain

$$\int_{X_1 \times X_2} f \, \mathrm{d}\sigma_{1,2} = L(f) = Ł\big(f \circ \pi^{1,2}\big) = \int_{X_1 \times X_2 \times X_3} f \circ \pi^{1,2} \, \mathrm{d}\nu = \int_{X_1 \times X_2} f \circ \, \mathrm{d}\pi^{1,2}_* \nu \qquad \forall f \in C(X_1 \times X_2) \tag{2.7.a}$$

and

$$\int_{X_2 \times X_3} g \, \mathrm{d}\sigma_{2,3} = L(g) = Ł\big(g \circ \pi^{2,3}\big) = \int_{X_1 \times X_2 \times X_3} g \circ \pi^{2,3} \, \mathrm{d}\nu = \int_{X_2 \times X_3} g \circ \, \mathrm{d}\pi^{2,3}_* \nu \qquad \forall g \in C(X_2 \times X_3). \tag{2.7.b}$$

Clearly, (2.7) implies that both $\sigma_{1,2}$ and $\pi^{1,2}_{*}\nu$, as well as $\sigma_{2,3}$ and $\pi^{2,3}_{*}\nu$ represent the same functionals in $C'(X_1 \times X_2)$ and $C'(X_2 \times X_3)$, respectively. Due to Theorem A.3.3, we infer

$$\sigma_{1,2} = \pi^{1,2}_{*}\nu \qquad \text{and} \qquad \sigma_{2,3} = \pi^{2,3}_{*}\nu,$$

which proofs the existence of the required coupling in the compact case.

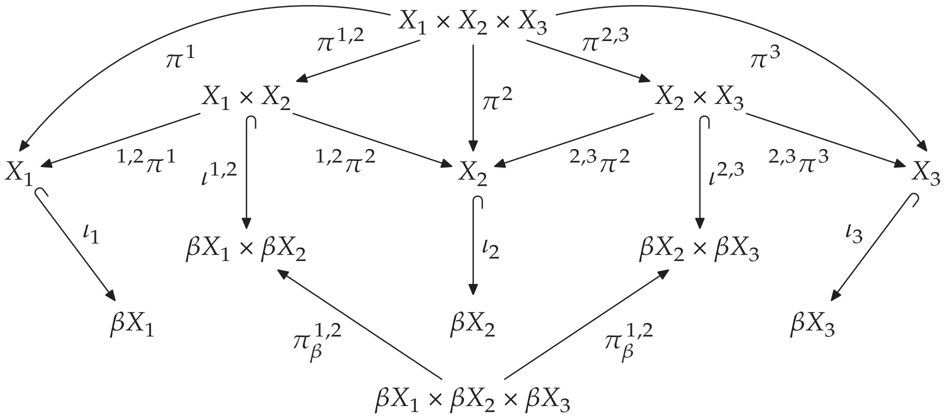


Figure 2.2 The underlying projections and embeddings of Lemma 2.2.3.

We turn now to the general case: Since the measures $\sigma_{1,2}$ and $\sigma_{2,3}$ are Radon, we may assume that the spaces X_1, X_2, X_3 may be represented as countable unions of compact sets: Indeed, there exists a sequence $\left(K^{1,2}_n\right)_{n\in\mathbb{N}}$ of compact sets $K^{1,2}_n \subseteq X_1 \times X_2$ such that

$$\sigma_{1,2}\big((X_1 \times X_2) \setminus K_n\big) < \frac{1}{n} \qquad \forall n \in \mathbb{N}.$$

Setting $K^{1,2} := \bigcup_{n\in\mathbb{N}} K^{1,2}_n$, this immediately implies that $\sigma_{1,2}(K^{1,2}) = \sigma_{1,2}(X_1 \times X_2)$. In particular, the measure $\sigma_{1,2}$ does not give any mass to sets in $(K^{1,2})^{\mathrm{c}}$. As countable union of measurable sets, $K^{1,2}$ belongs to $\mathcal{B}(X_1 \times X_2)$. Repeating this construction for the Radon measure $\sigma_{2,3}$ on the space $X_2\times X_3$, one also obtains a set $K^{2,3} \in \mathcal{B}(X_2\times X_3)$ with $\sigma_{2,3}(K^{2,3}) = \sigma_{2,3}(X_2 \times X_3)$.

Now we consider the Stone-Čech compactifications βX_i with continuous embeddings $\iota_i : X_i \to \beta X_i$ for $i \in \{1,2,3\}$ (cf. also with Example A.3.6 in Appendix A). The projections ${}^{1,2}\pi^1$, ${}^{1,2}\pi^2$, ${}^{2,3}\pi^3$, and therefore also the compositions $\iota_1 \circ {}^{1,2}\pi^1$, $\iota_2 \circ {}^{1,2}\pi^2$, $\iota_3 \circ {}^{2,3}\pi^3$, are continuous maps. Thus, the set

$$K^1 := \iota_1 \circ {}^{1,2}\pi^1(K^{1,2}) = \iota_1 \circ {}^{1,2}\pi^1\Big(\bigcup_{n\in\mathbb{N}} K^{1,2}_n\Big) = \bigcup_{n\in\mathbb{N}}\Big(\iota_1 \circ {}^{1,2}\pi^1\big(K^{1,2}_n\big)\Big)$$

turns out to be Borel in the space βX_1. Similarly, we obtain that $K^2 := \iota_2 \circ {}^{1,2}\pi^2(K^{1,2})$ and $K^3 := \iota_3 \circ {}^{2,3}\pi^3(K^{2,3})$ belong to $\mathcal{B}(\beta X_2)$ and $\mathcal{B}(\beta X_3)$, respectively. Since the measures $\sigma_{1,2}$ and $\sigma_{2,3}$ are concentrated on the Borel sets $\iota_1^{-1}(K^1) \times \iota_2^{-1}(K^2)$ and $\iota_2^{-1}(K^2) \times \iota_3^{-1}(K^3)$, we can extend these two Radon measures to $\beta X_1 \times \beta X_2$ and $\beta X_2 \times \beta X_3$ by the formulas

$$\sigma^{\beta}_{1,2}(B) := \iota^{1,2}_{*}\big(B \cap (K^1 \times K^2)\big) \qquad \forall B \in \mathcal{B}(\beta X_1 \times \beta X_2)$$

and

$$\sigma^{\beta}_{2,3}(B) := \iota^{2,3}_{*}\big(B \cap (K^2 \times K^3)\big) \qquad \forall B \in \mathcal{B}(\beta X_2 \times \beta X_3),$$

where $\iota^{1,2}(x_1, x_2) := \big(\iota_1(x_1), \iota_2(x_2)\big)$ and $\iota^{2,3}(x_2, x_3) := \big(\iota_2(x_2), \iota_3(x_3)\big)$. It is obvious that $\sigma^{\beta}_{1,2}$ and $\sigma^{\beta}_{2,3}$ are again Radon probability measures on the compact spaces $\beta X_1 \times \beta X_2$ and

$\beta X_2 \times \beta X_3$. Hence, we can appeal to the compact case in the first part of the proof to obtain a Radon probability measure ν^β on $\beta X_1 \times \beta X_2 \times \beta X_3$, where the projections

$$\pi_\beta^{1,2} : \beta X_1 \times \beta X_2 \times \beta X_3 \to \beta X_1 \times \beta X_2 \qquad \text{and} \qquad \pi_\beta^{2,3} : \beta X_1 \times \beta X_2 \times \beta X_3 \to \beta X_2 \times \beta X_3$$

induce the pushforwards

$$\sigma_{1,2}^\beta = \pi_{\beta *}^{1,2} \nu^\beta \qquad \text{and} \qquad \sigma_{2,3}^\beta = \pi_{\beta *}^{2,3} \nu^\beta. \tag{2.8}$$

As a result, we have

$$\nu^\beta(K^1 \times K^2 \times \beta X_3) = \pi_{\beta *}^{1,2} \nu^\beta(K^1 \times K^2) = \sigma_{1,2}^\beta(K^1 \times K^2) = 1$$

and

$$\nu^\beta(\beta X_1 \times K^2 \times K^3) = \pi_{\beta *}^{2,3} \nu^\beta(K^2 \times K^3) = \sigma_{2,3}^\beta(K^2 \times K^3) = 1,$$

which shows that ν^β is concentrated on $K^1 \times K^2 \times K^3$. Thus, the restriction $\nu^\beta|_{K^1 \times K^2 \times K^3}$ induces a Radon probability measure with the required pushforwards on $X_1 \times X_2 \times X_3$. ■

Proof of Proposition 2.2.2 We have to check that for every $p \geq 1$ the Wasserstein distance W_p satisfies the axioms of a metric on the set $\mathcal{P}_r^p(X)$:
First, to infer that W_p takes only values in $\mathbb{R}_0^+$, it suffices to show that $(x, y) \mapsto d(x, y)$ belongs to $L^p(X \times X, \sigma, \mathbb{R})$ for every choice of $\sigma \in \Pi(\mu, \nu)$. This follows from the estimate

$$\frac{1}{2^p} d^p(x, y) \leq \left(\frac{1}{2} d(x, y_0) + \frac{1}{2} d(y_0, y)\right)^p \leq \frac{1}{2} d^p(x, y_0) + \frac{1}{2} d^p(y_0, y), \tag{2.9}$$

where we used the convexity of $t \mapsto t^p$. Now (2.9) immediately implies

$$\begin{aligned} \int_{X \times X} d^p(x, y)\, \mathrm{d}\sigma(x, y) &\leq 2^{p-1} \int_{X \times X} d^p(x, y_0)\, \mathrm{d}\sigma(x, y) + 2^{p-1} \int_{X \times X} d^p(y_0, y)\, \mathrm{d}\sigma(x, y) = \\ &= 2^{p-1} \int_{X \times X} d^p\big(\pi^1(x, y), y_0\big)\, \mathrm{d}\sigma(x, y) + 2^{p-1} \int_{X \times X} d^p\big(y_0, \pi^2(x, y)\big)\, \mathrm{d}\sigma(x, y) = \\ &= 2^{p-1} \int_X d^p(x, y_0)\, \mathrm{d}\mu(x) + 2^{p-1} \int_X d^p(y_0, y)\, \mathrm{d}\nu(y) < +\infty. \end{aligned}$$

Clearly, we have $W_p(\mu, \nu) = W_p(\mu, \nu)$ by definition of (2.6).
Next, we show that $\mu = \nu$ precisely when $W_p(\mu, \nu) = 0$. In fact, the pushforward $g_* \mu$ under the diagonal embedding $g : x \mapsto (x, x)$ is an admissible plan since it satisfies $\pi_*^i(g_* \mu) = (g \circ \pi^i)_* \mu = \mu$ for $i \in \{1, 2\}$. As the metric d vanishes on the diagonal $Đ := \{(x, y) \in X \times X : x = y\}$ and the measure $g_* \mu$ is concentrated on $Đ$, we obtain $W_p(\mu, \mu) = 0$.
On the other hand, $W_p(\mu, \nu) = 0$ implies, that the optimal plan $\sigma_{\min} \in \Pi(\mu, \nu)$ (whose existence is guaranteed by Theorem 2.1.2) is concentrated on the diagonal $Đ$. Thus

$$\begin{aligned} \mu(B) = \pi_*^1 \sigma_{\min}(B) &= \sigma_{\min}\big(Đ \cap (\pi^1)^{-1}(B)\big) = \\ &= \sigma_{\min}\big(Đ \cap (\pi^2)^{-1}(B)\big) = \pi_*^2 \sigma_{\min}(B) = \nu(B) \end{aligned}$$

for all $B \in \mathcal{B}(X)$.
Finally, it remains to show the triangle inequality: For any given measures $\mu_1, \mu_2 \mu_3 \in \mathcal{P}_r^p(X)$, let $\sigma_{1,2}$ be an optimal plan in $\Pi(\mu_1, \mu_2)$; likewise, let $\sigma_{2,3}$ be optimal in $\Pi(\mu_2, \mu_3)$. By virtue of Lemma 2.2.3, there exists a Radon probability measure η on the product space

$X \times X \times X$ with pushforwards $\pi_*^{1,2}\eta = \sigma_{1,2}$ and $\pi_*^{2,3}\eta = \sigma_{2,3}$. Since the projections in Figure 2.2 commute accordingly, this implies $\pi_*^i\eta = \mu_i$ for all $i \in \{1,2,3\}$. In particular, $\pi_*^{1,3}\eta$ belongs to $\Pi(\mu_1,\mu_3)$. Henceforth,

$$\begin{aligned}W_p(\mu_1,\mu_3) &\leq \Big(\int_{X\times X} d^p(x_1,x_3)\,\mathrm{d}\pi_*^{1,3}\eta(x_1,x_3)\Big)^{1/p} = \Big(\int_{X\times X\times X} d^p(x_1,x_3)\,\mathrm{d}\eta(x_1,x_2,x_3)\Big)^{1/p} \leq\\ &\leq \Big(\int_{X\times X\times X}\big(d(x_1,x_2)+d(x_2,x_3)\big)^p\,\mathrm{d}\eta(x_1,x_2,x_3)\Big)^{1/p} \leq\\ &\leq \Big(\int_{X\times X\times X} d^p(x_1,x_2)\,\mathrm{d}\eta(x_1,x_2,x_3)\Big)^{1/p} + \Big(\int_{X\times X\times X} d^p(x_2,x_3)\,\mathrm{d}\eta(x_1,x_2,x_3)\Big)^{1/p} =\\ &= \Big(\int_{X\times X\times X} d^p(x_1,x_2)\,\mathrm{d}\sigma_{1,2}(x_1,x_2)\Big)^{1/p} + \Big(\int_{X\times X\times X} d^p(x_2,x_3)\,\mathrm{d}\sigma_{2,3}(x_2,x_3)\Big)^{1/p} =\\ &= W_p(\mu_1,\mu_2) + W_p(\mu_2,\mu_3).\end{aligned}$$

■

Next we show two important estimates for the Wasserstein distances.

2.2.4 **Proposition** *Let (X,d) be a metric space. Then for every $p \geq 1$ and all $\mu,\nu \in \mathscr{P}_r^p(X)$ we have the following inequalities:*

(i) *For every choice of $p \leq q$ the estimate $W_p(\mu,\nu) \leq W_q(\mu,\nu)$ holds.*

(ii) *For every $y_0 \in X$ we have*

$$W_p(\mu,\nu) \leq 2^{1/q}\Big(\int_X d^p(x,y_0)\,\mathrm{d}|\mu-\nu|(x)\Big)^{1/p}, \qquad \frac{1}{p}+\frac{1}{q} = 1.$$

Proof The inequality in (i) follows readily from *Hölder's inequality*: Due to Theorem 2.1.2, there exists an optimal plan $\sigma_{\min} \in \Pi(\mu,\nu)$ with respect to the cost function $d^q(x,y)$. Therefore,

$$\begin{aligned}W_p(\mu,\nu) &\leq \Big(\int_{X\times X} d^p(x,y)\,\mathrm{d}\sigma_{\min}(x,y)\Big)^{1/p} \leq\\ &\leq \Big(\int_{X\times X} 1\,\mathrm{d}\sigma_{\min}(x,y)\Big)^{1/(\tilde p q)}\Big(\int_{X\times X} d^{p\tilde q}(x,y)\,\mathrm{d}\sigma_{\min}(x,y)\Big)^{1/(p\tilde q)} = W_q(\mu,\nu),\end{aligned}$$

where $\tilde p, \tilde q \geq 1$ such that $q = p\tilde q$ and $\frac{1}{\tilde p}+\frac{1}{\tilde q} = 1$.

Concerning (ii), we define a measure σ on $X \times X$ by

$$\sigma(A\times B) := \mu(A\cap B) - \eta^+(A\cap B) + \frac{1}{c}\eta^+(A)\eta^-(B) \qquad \forall (A\times B) \in \mathcal{B}(X)\otimes\mathcal{B}(X),$$

where $\eta := \mu - \nu$ with a unique *Jordan decomposition* $\eta = \eta^+ - \eta^-$ and and a positive constant

$$c := \eta^+(X) = \eta^+(X) - (\mu-\nu)(X) = \eta^-(X).$$

Due to the *minimality property* of the Jordan decomposition, we have $\mu - \eta^+ \geq 0$. Together with $\sigma(X\times X) = 1$, this implies that σ is a probability measure. Moreover, we have

$$\sigma(B\times X) = \mu(B) - \eta^+(B) + \eta^+(B) = \mu(B) \qquad B \in \mathcal{B}(X)$$

and

$$\sigma(X\times B) = \mu(B) - \eta^+(B) + \eta^-(B) = \nu(B) \qquad B \in \mathcal{B}(X).$$

Hence, σ has the required pushforwards and belongs to $\Pi(\mu,\nu)$. Now recall that (2.9) implies the estimate

$$d^p(x,y_0) \le 2^{p-1}\big(d^p(x,y_0) + d^p(y_0,y)\big). \tag{2.10}$$

Thus,

$$\big(W_p(\mu,\nu)\big)^p \le \int_{X\times X} d^p(x,y)\,\mathrm{d}\sigma(x,y) = \tag{2.11.a}$$

$$= \int_X d^p(x,x)\,\mathrm{d}(\mu-\eta^+)(x) + \frac{1}{c}\iint_{X^2} d^p(x,y)\,\mathrm{d}\eta^+(x)\,\mathrm{d}\eta^-(y) \le \tag{2.11.b}$$

$$\le 2^{p-1}\frac{1}{c}\iint_{X^2} d^p(x,y_0) + d^p(y_0,y)\,\mathrm{d}\eta^+(x)\,\mathrm{d}\eta^-(y) \le \tag{2.11.c}$$

$$\le 2^{p-1}\Big(\int_X d^p(x,y_0)\,\mathrm{d}\eta^+(x) + \int_X d^p(y_0,y)\,\mathrm{d}\eta^-(y)\Big) = \tag{2.11.d}$$

$$= 2^{p-1}\int_X d^p(x,y_0)\,\mathrm{d}(\eta^+ + \eta^-)(x) = 2^{p-1}\int_X d^p(x,y_0)\,\mathrm{d}|\mu-\nu|(x), \tag{2.11.e}$$

where we used *Fubini's theorem*. ■

Now we investigate a useful characterization of the Wasserstein distance W_1.

2.2.5 Theorem (Kantorovich-Rubinstein) *Let (X,d) be a metric space. Then we have*

$$W_1(\mu,\nu) = \sup\Big\{\int_X f\,\mathrm{d}(\mu-\nu) : f \in Lip_1(X)\Big\} \qquad \forall \mu,\nu \in \mathscr{P}_r^1(X).$$

The proof of this theorem relies heavily on the Kantorovich duality, introduced in Remark 2.1.5.

Proof In the first part of the proof we show the identity

$$W_1(\mu,\nu) = \sup\Big\{\int_X l\,\mathrm{d}(\mu-\nu) : l \in C_b(X),\, l(x)-l(y) \le d(x,y)\Big\}. \tag{2.12}$$

To this aim, we may invoke (2.4) to obtain the inequality

$$\sup\Big\{\int_X l\,\mathrm{d}(\mu-\nu) : l \in C_b(X),\, l(x)-l(y) \le d(x,y)\Big\} \le \tag{2.13.a}$$

$$\le \sup\Big\{\int_X f\,\mathrm{d}\mu + \int_X g\,\mathrm{d}\nu : f,g \in \mathscr{B}_b(X), f(x)+g(y) \le d(x,y)\Big\} = W_1(\mu,\nu). \tag{2.13.b}$$

This means that for every $\varepsilon > 0$ there exist $f,g \in \mathscr{B}_b(X)$ with $f(x)+g(y) \le d(x,y)$ such that

$$W_1(\mu,\nu) - \varepsilon < \int_X f\,\mathrm{d}\mu + \int_X g\,\mathrm{d}\nu. \tag{2.14}$$

Now define a function $k : X \to \mathbb{R}$, $k(x) := \inf_{y\in X}\{d(x,y) - g(y)\}$. Then the estimate

$$\begin{aligned} k(x_0) - k(y_0) &= \inf_{x\in X}\{\sup_{y\in X}\{d(x_0,x) - d(y_0,y) - g(x) + g(y)\}\} \le \\ &\le \sup_{y\in X}\{d(x_0,y) - d(y_0,y)\} \le d(x_0,y_0) \end{aligned}$$

shows that $|k(x_0) - k(y_0)| \le d(x_0,y_0)$ for all $x_0,y_0 \in X$. Thus, k belongs to $Lip_1(X)$. Furthermore,

$$f(x) \leq k(x) \leq d(x,x) - g(x) = -g(x) \qquad \forall x \in X \tag{2.15}$$

and the fact that f, g belong to $\mathcal{B}_b(X)$ imply that k is bounded on X. In particular, (2.15) shows that $f(x) + g(y) \leq k(x) - k(y)$ for all $x, y \in X$. Together with (2.14) and (2.13), this inequality may be used to obtain

$$W_1(\mu,\nu) - \varepsilon < \int_X f \,\mathrm{d}\mu + \int_X g \,\mathrm{d}\nu = \int_{X\times X} f(x) + g(y) \,\mathrm{d}\sigma(x,y) \leq \int_{X\times X} k(x) - k(y) \,\mathrm{d}\sigma(x,y) =$$
$$= \int_X k \,\mathrm{d}\mu - \int_X k \,\mathrm{d}\nu \leq \sup\Big\{\int_X f \,\mathrm{d}(\mu - \nu) : f \in C_b(X), f(x) - f(y) \leq d(x,y)\Big\} \leq W_1(\mu,\nu),$$

where σ denotes an arbitrary admissible plan in $\Pi(\mu,\nu)$. This estimate shows (2.12) and completes the first part of the proof.

Next observe that for every $l \in Lip_1(X)$ and fixed $y_0 \in X$ we have

$$|l(x)| \leq l(y_0) + |l(x) - l(y_0)| \leq l(y_0) + d(x, y_0) \qquad \forall x \in X.$$

As μ, ν belong to $\mathcal{P}_r^1(X)$, this shows that $\pm l \in L^1(X,\mu,\mathbb{R}) \cap L^1(X,\nu,\mathbb{R})$. In particular, the pair $(l, -l)$ belongs to the class Φ_d as defined in Remark 2.1.5. Thus, (2.12) and (2.3) imply the final estimate

$$\begin{aligned} W_1(\mu,\nu) &= \sup\Big\{\int_X l \,\mathrm{d}(\mu-\nu) : l \in C_b(X),\, l(x) - l(y) \leq d(x,y)\Big\} = \\ &= \sup\Big\{\int_X l \,\mathrm{d}(\mu-\nu) : l \in C_b(X),\, |l(x) - l(y)| \leq d(x,y)\Big\} \leq \\ &\leq \sup\Big\{\int_X l \,\mathrm{d}(\mu-\nu) : l \in Lip_1(X)\Big\} \leq \\ &\leq \sup\Big\{\int_X f \,\mathrm{d}\mu - \int_X g \,\mathrm{d}\nu : (f,g) \in \Phi_d\Big\} = W_1(\mu,\nu). \end{aligned}$$

■

Wasserstein spaces inherit to a great extend the topological structure of the underlying metric space:

2.2.6 **Proposition** *Let (X,d) be a metric space. Then for every $p \geq 1$ the Wasserstein space $\big(\mathcal{P}_r^p(X), W_p\big)$ inherits the following properties from X:*

(i) *The space $\big(\mathcal{P}_r^p(X), W_p\big)$ is complete if (X,d) is complete.*

(ii) *The space $\big(\mathcal{P}_r^p(X), W_p\big)$ is separable if (X,d) is separable.*

In particular, the Wasserstein space over a Polish metric space is again Polish.

The key ingredient in the proof of (i) is the following lemma.

2.2.7 **Lemma** *Let (X,d) be a complete metric space and $p \geq 1$. Then every Cauchy sequence in $\big(\mathcal{P}_r^p(X), W_p\big)$ is tight.*

For the proof of this lemma recall that in a complete metric space (X,d) a set K is compact, precisely when K is closed and *totally bounded*, i.e. for every $\varepsilon > 0$, K may be written as finite union of of sets $(K_n)_{n\leq N}$, $K_n \subseteq K$ with diameter $\mathrm{diam}(K_n) < \varepsilon$.

Proof Denote by $(\mu_n)_{n\in\mathbb{N}}$ a Cauchy sequence in $\big(\mathcal{P}_r^p(X), W_p\big)$. By virtue of Proposition 2.2.4.i, $(\mu_n)_{n\in\mathbb{N}}$ is also a Cauchy sequence with respect to W_1.

First, we show that for every $\varepsilon > 0$ there exist finitely many points $x_1, \dots x_M$ in X such that

$$\mu_n\big(B_{2\varepsilon}(x_1) \cup \dots \cup B_{2\varepsilon}(x_M)\big) \geq 1 - 2\varepsilon \qquad \forall n \in \mathbb{N}. \tag{2.16}$$

To this end, let $\varepsilon > 0$ be given and fix $N \in \mathbb{N}$ such that

$$W_1(\mu_N, \mu_k) < \varepsilon^2 \qquad \forall k \in \mathbb{N} : k \geq N. \tag{2.17}$$

Since $\mathcal{N} := \{\mu_1, \dots \mu_N\}$ is uniformly tight as finite family of Radon probability measures, there exists a compact set $K_\varepsilon \subseteq X$ such that $\mu_i(K_\varepsilon^c) < \varepsilon$ for all $\mu_i \in \mathcal{N}$. The compactness of K_ε implies that there exist finitely many points $x_1, \dots x_M$ in X with

$$K \subset B_\varepsilon(x_1) \cup \dots \cup B_\varepsilon(x_M) =: U.$$

Now define

$$U_\varepsilon := \{x \in X : \operatorname{dist}(x, U) < \varepsilon\} \subseteq B_{2\varepsilon}(x_1) \cup \dots \cup B_{2\varepsilon}(x_M)$$

and set $\phi(x) := \big(1 - \varepsilon^{-1} \operatorname{dist}(x, U)\big)^+$. In case, that ϕ does not vanish at two points $x, y \in X$, the inverse triangle inequality implies

$$|\phi(x) - \phi(y)| = \frac{1}{\varepsilon} |\operatorname{dist}(x, U) - \operatorname{dist}(y, U)| \leq \frac{1}{\varepsilon} d(x, y). \tag{2.18}$$

On the other hand, if $\phi(x) = 0$, then $d(x, U) \geq \varepsilon$. Therefore

$$|\phi(x) - \phi(y)| = 1 - \frac{1}{\varepsilon} d(y, U) \leq \frac{1}{\varepsilon}\big(d(x, U) - d(y, U)\big) \leq \frac{1}{\varepsilon} d(x, y) \qquad \forall y \in X : \phi(y) \neq 0. \tag{2.19}$$

Together, (2.18) and (2.19) imply that ϕ is Lipschitz continuous with a Lipschitz constant ε^{-1}. Henceforth, we may assume that $\varepsilon\phi$ belongs to $Lip_1(X)$. Moreover, ϕ is bounded such that $\mathbb{1}_U \leq \phi \leq \mathbb{1}_{U_\varepsilon}$. Thus, we may invoke the *monotonicity of the integral* and Theorem 2.2.5 to obtain

$$\mu_n(U_\varepsilon) \geq \int_X \phi \, \mathrm{d}\mu_n = \int_X \phi \, \mathrm{d}\mu_i + \frac{1}{\varepsilon}\Big(\int_X \varepsilon\phi \, \mathrm{d}\mu_n - \int_X \varepsilon\phi \, \mathrm{d}\mu_i\Big) \geq \tag{2.20.a}$$

$$\geq \int_X \phi \, \mathrm{d}\mu_i - \frac{1}{\varepsilon} W_1(\mu_n, \mu_i) \geq \mu_i(U) - \frac{1}{\varepsilon} W_1(\mu_n, \mu_i) \tag{2.20.b}$$

for all $i, n \in \mathbb{N}$. We need to find an estimate for the right-hand side of (2.20): By choice of K_ε, we have

$$\mu_i(U) \geq \mu_i(K_\varepsilon) \geq 1 - \varepsilon \qquad \forall \mu_i \in \mathcal{N}. \tag{2.21}$$

Moreover, we may choose $i(n) := \min\{n, N\}$ for every $n \in \mathbb{N}$ in (2.20). Thus, (2.17) implies that $W_1(\mu_n, \mu_{i(k)}) < \varepsilon^2$ for every $n \in \mathbb{N}$. Together with (2.20), this estimate and (2.21) result in

$$\mu_n(U_\varepsilon) \geq 1 - \varepsilon - \frac{\varepsilon^2}{\varepsilon} = 1 - 2\varepsilon \qquad \forall n \in \mathbb{N},$$

which completes the first part of the proof.

In the second part of the proof we construct for every $\varepsilon > 0$ a compact set $S \subseteq X$ such that $\mu_n(X \setminus S) < \varepsilon$: By virtue of the first part, for all $k \in \mathbb{N}$ there exists a finite family of points $x_1^k, \dots x_{M_k}^k$ in X such that

$$\mu_n\Big(X \setminus \big(B_{2^{-k}}(x_1^k) \cup \dots \cup B_{2^{-k}}(x_{M_k}^k)\big)\Big) < 2^{-k}\varepsilon \qquad \forall n \in \mathbb{N}.$$

Hence, we may set

$$S := \bigcap_{k \in \mathbb{N}} \big(\overline{B_{2^{-k}}(x_1^k)} \cup \dots \cup \overline{B_{2^{-k}}(x_{M_k}^k)}\big),$$

to obtain that $\mu_n(X \setminus S) < \varepsilon$ for all $n \in \mathbb{N}$. It is clear that S is closed as intersection of closed sets. Furthermore, for every $\delta > 0$ the set S is covered by the finite union $B_\delta(x_1^k) \cup \ldots \cup B_\delta(x_{M_k}^k)$, where $k \in \mathbb{N}$ such that $2^{-k} < \delta$. This shows that S is totally bounded and therefore compact. ■

The following proof is closely related to the arguments used in the proof of Theorem 2.1.2.

Proof of Proposition 2.2.6 Concerning (i), let $(\mu_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\left(\mathcal{P}_r^p(X), W_p\right)$. By virtue of Lemma 2.2.7, we may assume that $(\mu_n)_{n\in\mathbb{N}}$ is uniformly tight. Invoking Lemma 2.1.4, we obtain a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ of non-negative Radon measures, weakly converging to a non-negative Radon measure μ on X. In particular this means that $\mu(X) = \mu_{n_k}(X) = 1$ for all $k \in \mathbb{N}$. Hence, μ is also a probability measure. However, it is still unclear, whether μ belongs to $\mathcal{P}_r^p(X)$.

Now denote by σ_{n_k,n_l} the optimal plan in $\Pi(\mu_{n_k}, \mu_{n_k})$ for $k, l \in \mathbb{N}$ such that

$$W_p(\mu_{n_k}, \mu_{n_k}) = \int_{X\times X} d^p(x,y)\,\mathrm{d}\sigma_{n_k,n_l}(x,y).$$

Due to uniform tightness of $(\mu_{n_k})_{k\in\mathbb{N}}$, Lemma 2.1.3 implies that for every $l \in \mathbb{N}$ the family $(\sigma_{n_k,n_l})_{k\in\mathbb{N}}$ is uniformly tight. Hence, for every choice of $l \in \mathbb{N}$, Lemma 2.1.4 may be applied once more to extract a weakly converging subsequence $(\sigma_{n_{\hat{k}},n_l})_{\hat{k}\in\mathbb{N}}$. One may argue as above to infer that the limit σ_{n_l} is again a Radon probability measure on $X \times X$.

In particular, $\sigma_{n_{\hat{k}},n_l} \xrightarrow{w^*} \sigma_{n_l}$ implies

$$\int_X f\,\mathrm{d}\mu = \lim_{\hat{k}\to\infty} \int_X f\,\mathrm{d}\mu_{n_{\hat{k}}} = \lim_{\hat{k}\to\infty} \int_{X\times X} f \circ \pi^1\,\mathrm{d}\sigma_{n_{\hat{k}},n_l} = \int_{X\times X} f \circ \pi^1\,\mathrm{d}\sigma_{n_l} \qquad \forall f \in C_\mathrm{b}(X)$$

and

$$\int_X f\,\mathrm{d}\mu_{n_l} = \lim_{\hat{k}\to\infty} \int_{X\times X} f \circ \pi^2\,\mathrm{d}\sigma_{n_{\hat{k}},n_l} = \int_{X\times X} f \circ \pi^2\,\mathrm{d}\sigma_{n_l} \qquad \forall f \in C_\mathrm{b}(X)\,.$$

Thus, one may apply Theorem A.3.5 to obtain $\mu = \pi^1_*\sigma_{n_l}$ and $\mu_{n_l} = \pi^2_*\sigma_{n_l}$. This means that σ_{n_l} belongs to $\Pi(\mu, \mu_{n_l})$ for every $l \in \mathbb{N}$.

Recall that we have already observed in the proof of Theorem 2.1.2 that Proposition A.4.5.ii holds also for non-negative integrands without an upper bound since the functional $\sigma \mapsto \int_{X\times X} d^p(x,y)\,\mathrm{d}\sigma(x,y)$ arises as supremum of lower semi-continuous functionals, i.e

$$\int_{X\times X} d^p(x,y)\,\mathrm{d}\sigma(x,y) = \sup_{N\in\mathbb{N}} \int_{X\times X} \min\{N, d^p(x,y)\}\,\mathrm{d}\sigma(x,y),$$

by *monotone convergence*. Thus, we may use that σ_{n_l} is an admissible plan to infer for every $l \in \mathbb{N}$ that

$$\left(W_p(\mu, \mu_{n_l})\right)^p \leq \int_{X_1\times X_2} d^p(x,y)\,\mathrm{d}\sigma_{n_l}(x,y) \leq \tag{2.22.a}$$

$$\leq \liminf_{\hat{k}\to\infty} \int_{X_1\times X_2} d^p(x,y)\,\mathrm{d}\sigma_{n_{\hat{k}},n_l}(x,y) = \liminf_{\hat{k}\to\infty} \left(W_p(\mu_{n_{\hat{k}}}, \mu_{n_l})\right)^p. \tag{2.22.b}$$

Since $(\mu_{n_l})_{l\in\mathbb{N}}$ is still a Cauchy sequence, for every $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that $W_p(\mu_{n_{\hat{k}}}, \mu_{n_l}) < \varepsilon$ for all $\hat{k}, l \geq N_\varepsilon$. Together with (2.22), this implies $\lim_{l\to\infty} W_p(\mu, \mu_{n_l}) = 0$. Moreover, this shows $\mu \in \mathcal{P}_r^p(X)$ since we can use (2.10) to obtain the estimate

$$\begin{aligned}\int_X d^p(x,y_0)\,\mathrm{d}\mu(x) &= \int_X d^p(x,y_0)\,\mathrm{d}\sigma_{n_l}(x,y) \le \\ &\le 2^{p-1}\int_{X\times X} d^p(x,y)\,\mathrm{d}\sigma_{n_l}(x,y) + 2^{p-1}\int_{X\times X} d^p(y_0,y)\,\mathrm{d}\sigma_{n_l}(x,y) \le \\ &\le 2^{p-1}\varepsilon + \int_X d^p(y_0,y)\,\mathrm{d}\mu_{n_l}(y) < +\infty\end{aligned}$$

for arbitrary $l \ge N_\varepsilon$.
Finally, $W_p(\mu,\mu_n) \le W_p(\mu,\mu_{n_l}) + W_p(\mu_{n_l},\mu_n)$ shows that the whole sequence $(\mu_n)_{n\in\mathbb{N}}$ converges to μ with respect to the Wasserstein metric W_p. Hence, $\big(\mathcal{P}_r^p(X), W_p\big)$ is complete.

Now we turn to (ii): Let $\varepsilon > 0$ and a measure $\mu \in \mathcal{P}_r^p(X)$ be given. Due to our premise, there exists a countable, dense subset $D \subseteq X$; fix $x_0 \in D$. By absolute continuity of the integral (see Lemma 1.1.2) and the Radon property of μ, there exists a compact set K_ε in X with

$$\int_{K_\varepsilon^c} d^p(x_0,x)\,\mathrm{d}\mu(x) < \varepsilon^p. \tag{2.23}$$

By compactness of K_ε, there exist finitely many points $x_1, \dots x_N$ in D such that $K_\varepsilon \subseteq B_\varepsilon(x_1) \cup \dots \cup B_\varepsilon(x_N)$. Via

$$\acute{B}_i := B_\varepsilon(x_i) \setminus \bigcup_{j<i} B_\varepsilon(x_j) \qquad \forall i \le N$$

we obtain a finite covering of K_ε with disjoint sets $\acute{B}_1, \dots \acute{B}_N$. Therefore, the function

$$\alpha(x) := \begin{cases} x_i & \text{if } x \in K_\varepsilon \cap \acute{B}_i, \\ x_0 & \text{if } x \notin K_\varepsilon, \end{cases}$$

is well-defined on X. In particular, we have $d(\alpha(x),x) < \varepsilon$ for all $x \in K_\varepsilon$ since $\acute{B}_i \subseteq B_\varepsilon(x_i)$. Together with (2.23), this implies

$$\int_X d^p(\alpha(x),x)\,\mathrm{d}\mu(x) = \int_{K_\varepsilon} d^p(\alpha(x),x)\,\mathrm{d}\mu(x) + \int_{K_\varepsilon^c} d^p(\alpha(x),x)\,\mathrm{d}\mu(x) \le \tag{2.24.a}$$

$$\le \int_{K_\varepsilon} \varepsilon^p\,\mathrm{d}\mu(x) + \int_{K_\varepsilon^c} d^p(x_0,x)\,\mathrm{d}\mu(x) < 2\varepsilon^p. \tag{2.24.b}$$

Clearly, the pushforward $\alpha_*\mu$ is a probability measure on X. Moreover, $\alpha_*\mu$ is Radon as the measure is concentrated on the compact set $\{x_i : 0 \le i \le N\}$. Since

$$\int_X d^p(x,x_0)\,\mathrm{d}\alpha_*\mu(x) = \int_{K_\varepsilon} d^p(\alpha(x),x_0)\,\mathrm{d}\mu(x) \le \varepsilon \max_{i\le N}\{d^p(x_i,x_0)\} < +\infty,$$

we infer that $\alpha_*\mu$ belongs to $\mathcal{P}_r^p(X)$. Define $\beta : X \to X \times X$, $\beta(x) := (\alpha(x), x)$. Then $\beta_*\mu$ depicts an admissible plan in $\Pi(\alpha_*\mu,\mu)$ and therefore (2.24) implies

$$W_p(\alpha_*\mu,\mu) \le \Big(\int_{X\times X} d^p(x,y)\,\mathrm{d}\beta_*\mu(x,y)\Big)^{1/p} = \Big(\int_X d^p(\alpha(x),x)\,\mathrm{d}\mu(x)\Big)^{1/p} < 2\varepsilon. \tag{2.25}$$

This shows that the measure μ may be approximated in $\big(\mathcal{P}_r^p(X), W_p\big)$ by measures $\alpha_*\mu$ which can be represented by Dirac measures in the form of

$$\alpha_*\mu = \sum_{i=0}^N a_i\delta_{x_i}, \qquad a_i := \alpha_*\mu\big(\{x_i\}\big). \tag{2.26}$$

However, the set of all measures of the form (2.25) is still not countable. To this end, let $b_0, \dots b_N$ be coefficients in $\mathbb{Q}$ such that $|a_i - b_i| < \varepsilon$ for all $0 \le i \le N$. Then both (2.25) and Proposition 2.2.4.ii imply

$$W_p\Big(\sum_{i=0}^N b_i \delta_{x_i}, \mu\Big) \le W_p\Big(\sum_{i=0}^N b_i \delta_{x_i}, \sum_{i=0}^N a_i \delta_{x_i}\Big) + W_p\Big(\sum_{i=0}^N a_i \delta_{x_i}, \mu\Big) \le$$

$$\le 2^{1/q}\Big(\sum_{i=1}^N d^p(x_i, x_0)\,|b_i - a_i|\Big)^{1/p} + 2\varepsilon \le 2^{1/q} \max_{i \le N} \{d(x_i, x_0)\}\,(N\varepsilon)^{1/p} + 2\varepsilon.$$

This shows that the countable set of measures of the form

$$\sum_{i=0}^N b_i \delta_{x_i}, \qquad b_i \in \mathbb{Q},\ x_i \in D,\ N \in \mathbb{N}_0$$

is dense in $\big(\mathcal{P}_r^p(X), W_p\big)$. ■

2.3 Towards the Existence of Gradient Flows in Wasserstein Spaces

In this section we invoke the minimizing movements scheme, introduced in Section 1.5, to obtain an existence result of gradient flows in the EVI sense in the quadratic Wasserstein space over $\mathbb{R}^n$. In order to simplify the existence theory to a great extend, we make the following assumptions on the functional ϕ:

Notation

From now on, ϕ denotes a non-negative, lower semi-continuous functional with proper effective domain $\operatorname{dom}\phi$ on the Wasserstein space $\big(\mathcal{P}_r^2(\mathbb{R}^n), W_2\big)$. Moreover, we will assume that ϕ meets the following requirements:

(REQ1) Every probability measure $\mu \in \operatorname{dom}|\partial\phi|$ is absolutely continuous with respect to the Lebesgue measure λ_n on $\mathbb{R}^n$.

(REQ2) There exists a constant $\tilde{\tau} > 0$ such that the functional $\Phi(\tau, \mu, \cdot)$ defined in (1.64) admits at least one minimum μ_τ in $\mathcal{P}_r^2(\mathbb{R}^n)$ for every choice of $\tau \in (0, \tilde{\tau})$ and $\mu \in \mathcal{P}_r^2(\mathbb{R}^n)$.

2.3.1 Facts

(i) Note that, since $\mathbb{R}^n$ depicts a Polish space, the class of Borel probability measures coincides with the class of Radon probability measures on $\mathbb{R}^n$. Nevertheless, we will stick with the notation $\mathcal{P}_r^2(\mathbb{R}^n)$ for consistency.

(ii) Let $(\mu_\tau)_{\tau<\tilde{\tau}}$ be a minimizer as in (REQ2). Then for all $\mu, \nu \in \mathcal{P}_r^2(\mathbb{R}^n)$ we obtain the estimate

$$\phi(\nu) - \phi(\mu_\tau) = \Phi(\tau, \mu, \nu) - \Phi(\tau, \mu, \mu_\tau) + \frac{1}{2\tau}\big(W_2^2(\mu_\tau, \mu) - W_2^2(\nu, \mu)\big) \ge \tag{2.27.a}$$

$$\ge \frac{1}{2\tau}\big(W_2^2(\mu_\tau, \mu) - W_2^2(\nu, \mu)\big) \ge \tag{2.27.b}$$

$$\ge \frac{1}{2\tau}\Big(W_2(\mu_\tau, \mu)\big|W_2(\mu_\tau, \nu) - W_2(\nu, \mu)\big| - W_2(\nu, \mu)\big(W_2(\mu_\tau, \mu) + W_2(\mu_\tau, \nu)\big)\Big) \ge \tag{2.27.c}$$

$$\ge -\frac{1}{2\tau} W_2(\mu_\tau, \nu)\big(W_2(\mu_\tau, \mu) + W_2(\nu, \mu)\big). \tag{2.27.d}$$

By a rearrangement of the terms in (2.27) we have

$$\frac{\phi(\mu_\tau) - \phi(\nu)}{W_2(\mu_\tau, \nu)} \le \frac{1}{2\tau}\big(W_2(\mu_\tau, \mu) + W_2(\nu, \mu)\big) \qquad \forall \nu \in \mathcal{P}_r^2(\mathbb{R}^n) : \nu \ne \mu_\tau.$$

Now taking the limit superior ($\nu \to \mu_\tau$) in the inequality above yields

$$|\partial\phi|(\mu_\tau) \leq \frac{W_2(\mu_\tau, \mu)}{2\tau} < +\infty. \qquad \forall \tau \in (0, \tilde{\tau}),\ \forall \nu \in \mathcal{P}_r^2(\mathbb{R}^n) : \nu \neq \mu_\tau.$$

Hence, (REQ1) implies that the minimizer $(\mu_\tau)_{\tau<\tilde{\tau}}$ is absolutely continuous with respect to the Lebesgue measure λ_n on $\mathbb{R}^n$.

As noted in Remark 1.5.4, the characterizing inequality (1.66) of NPC spaces, together with geodesic λ-convexity of ϕ, immediately implies that geodesics satisfy (CON) which is a key assumption for the existence of gradient flows in the EVI sense. However, the following example shows that geodesics may not be the right choice of interpolating curves in the Wasserstein space $\left(\mathcal{P}_r^2(\mathbb{R}^n), W_2\right)$.

2.3.2 **Example** The Wasserstein space $\left(\mathcal{P}_r^2(\mathbb{R}^2), W_2\right)$ is not an NPC space.

Proof Define the probability measures

$$\mu_0 := \frac{1}{2}\left(\delta_{(1,1)} + \delta_{(5,3)}\right),$$

$$\mu_1 := \frac{1}{2}\left(\delta_{(-1,1)} + \delta_{(-5,3)}\right),$$

$$\nu := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(0,-4)}\right).$$

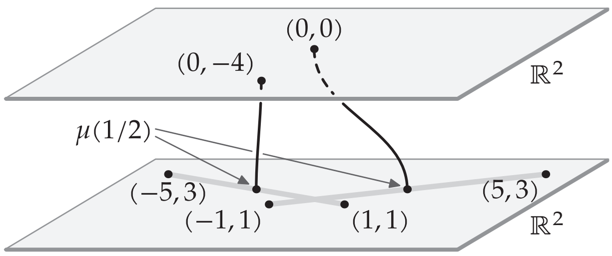


Figure 2.3 For $t = 1/2$, the links between the atoms of the measures $\mu(t)$ and ν represent an optimal plan of transportation with respect to quadratic costs $|\cdot|^2$. Note that the depicted optimal plan is not uniquely determined, due to particular symmetry of the transport problem at $t = 1/2$.

Clearly, μ_i belongs to $\mathcal{P}_r^2(\mathbb{R}^2)$ for $i \in \{1, 2, 3\}$. Since all admissible plans in each of the sets $\Pi(\mu_0, \mu_1)$, $\Pi(\mu_0, \nu)$, $\Pi(\mu_0, \nu)$ are concentrated on at most four points in $\mathbb{R}^4$, explicit computations of the distances are elementary and one obtains

$$W_2^2(\mu_0, \mu_1) = 40, \quad W_2^2(\mu_0, \nu) = 30, \quad W_2^2(\mu_1, \nu) = 30.$$

Moreover, one easily shows that

$$\mu(t) := \frac{1}{2}\left(\delta_{(1-6t, 1+2t)} + \delta_{(5-6t, 3-2t)}\right) \qquad \forall t \in [0, 1]$$

depicts a constant-speed geodesic with end-points $\mu(0) = \mu_0$ and $\mu(1) = \mu_1$ (see Figure 2.3). Now

$$W_2^2(\mu(1/2), \nu) = 40 > 20 = \frac{30}{2} + \frac{30}{2} - \frac{40}{4} = \frac{W_2^2(\mu_0, \nu)}{2} + \frac{W_2^2(\mu_1, \nu)}{2} - \frac{W_2^2(\mu_0, \mu_1)}{4}$$

shows that inequality (1.66) does not hold. ◀

In their seminal book [3] Ambrosio, Gigli, Savaré considered a different class of interpolating curves in $\left(\mathcal{P}_r^2(\mathbb{R}^2), W_2\right)$, which is more promising. To this aim, we will make use of the transport maps introduced in Remark 2.1.6.

2.3.3 **Definition** For $i \in \{0, 1\}$ let μ, σ_i be in $\mathcal{P}_r^2(\mathbb{R}^n)$ such that $\mu \ll \lambda_n$. We denote by $\boldsymbol{t}_\mu^{\sigma_i}$ the optimal μ-a.e. uniquely determined transport map in $\mathrm{T}(\mu, \sigma_i)$, such that

$$\int_{\mathbb{R}^n} \left|\boldsymbol{t}_\mu^{\sigma_i}(x) - x\right|^2 \mathrm{d}\mu(x) = W_2^2(\mu, \sigma_i). \tag{2.28}$$

The existence of $\boldsymbol{t}_\mu^\nu$ and validity of (2.28) is ensured by Theorem 2.1.7.

The interpolating curve

$$\sigma(s) := \big((1-s)t_\mu^{\sigma_0} + s t_\mu^{\sigma_1}\big)_* \mu \qquad \forall s \in [0,1] \tag{2.29}$$

in $\mathcal{P}_r^2(\mathbb{R}^n)$ is called ***generalized geodesic*** connecting σ_0 and σ_1 with ***base point*** μ.

Exacting a convexity assumption along generalized geodesics, we are eventually able give a relatively brief proof for existence of gradient flows in the EVI sense:

2.3.4 Theorem (Existence of gradient flows in the quadratic Wasserstein space) *Let $\phi : \mathcal{P}_r^2(\mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$ be a non-negative, lower semi-continuous functional with proper effective domain* $\operatorname{dom}\phi$. *Assume that ϕ satisfies* (REQ1), (REQ2) *and the following property:*

(SCON) *For every $\mu \in \operatorname{dom}|\partial\phi|$ and all $\sigma_0, \sigma_1 \in \operatorname{dom}\phi$ the functional ϕ is λ-convex along the generalized geodesic $\sigma(s)$ connecting σ_0 and σ_1 with base point μ.*

Then for every initial datum $\mu_0 \in \operatorname{dom}\phi$ there exists a unique gradient flow $\mu(s)$ in the sense of (EVI), *starting from $\lim_{s\searrow 0} \mu(s) = \mu_0$.*

Proof First, notice that the geodesic $\sigma(s)$ is well-defined since the base point of $\sigma(s)$ belongs to $\operatorname{dom}|\partial\phi|$ and (REQ1) assures that the optimal transport maps in (2.29) exist.
Next, (REQ1) implies that for every $\tau \in (0, \tilde\tau)$ and discrete initial datum $M_\tau^0 \in \operatorname{dom}\phi$ there exists a τ-discrete minimizing movement $(M_\tau^n)_{n\in\mathbb{N}}$ in $\mathcal{P}_r^2(\mathbb{R}^n)$, recursively defined by

$$M_\tau^n := \min_{M \in \mathcal{P}_r^2(\mathbb{R}^n)} \Phi\big(\tau, M_\tau^{n-1}, M\big) \qquad \forall n \in \mathbb{N},$$

where the functional Φ is defined as in (1.64). Moreover, Fact 2.3.1.ii yields that $M_\tau^n \ll \lambda_n$ for all $n \in \mathbb{N}_0$. Hence, the optimal transport maps $t_\tau^n := t_{M_\tau^n}^{M_\tau^{n-1}}$ are well-defined for every $n \in \mathbb{N}$, due to Theorem 2.1.7.

⇝ A basic inequality for the distance between μ and $\sigma(s)$: Inserting definition (2.29) of the generalized geodesic $\sigma(s)$ yields

$$W_2^2\big(\mu, \sigma(s)\big) = \int_{\mathbb{R}^n} \Big|(1-s)t_\mu^{\sigma_0}(x) + s t_\mu^{\sigma_1}(x) - x\Big|^2 \mathrm{d}\mu(x) = \tag{2.30.a}$$

$$= \int_{\mathbb{R}^n} \Big|(1-s)\big(t_\mu^{\sigma_0}(x) - x\big) + s\big(t_\mu^{\sigma_1}(x) - x\big)\Big|^2 \mathrm{d}\mu(x). \tag{2.30.b}$$

We may invoke the elementary identity (1.32) with respect to $(\mathbb{R}^n, |\cdot|)$ to write the integral in (2.30.b) in the form

$$\int_{\mathbb{R}^n} (1-s)\Big|t_\mu^{\sigma_0}(x) - x\Big|^2 + s\Big|t_\mu^{\sigma_1}(x) - x\Big|^2 - s(1-s)\Big|t_\mu^{\sigma_0}(x) - t_\mu^{\sigma_1}(x)\Big|^2 \mathrm{d}\mu(x) = \tag{2.31.a}$$

$$= (1-s)W_2^2(\mu, \sigma_0) + sW_2^2(\mu, \sigma_1) - s(1-s)\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 \,\mathrm{d}\big(t_\mu^{\sigma_0}, t_\mu^{\sigma_1}\big)_* \mu(x,y). \tag{2.31.b}$$

Since $\big(t_\mu^{\sigma_0}, t_\mu^{\sigma_1}\big)_* \mu$ is an admissible plan in $\Pi(\sigma_0, \sigma_1)$, we infer

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 \,\mathrm{d}\big(t_\mu^{\sigma_0}, t_\mu^{\sigma_1}\big)_* \mu(x,y) \le W_2^2(\sigma_0, \sigma_1).$$

Together with (2.30) and (2.31), this inequality shows

$$W_2^2\big(\mu, \sigma(s)\big) \le (1-s)W_2^2(\mu, \sigma_0) + sW_2^2(\mu, \sigma_1) - s(1-s)W_2^2(\sigma_0, \sigma_1) \qquad \forall s \in [0,1]. \tag{2.32}$$

⇝ A discrete version of (EVI) for the minimizing movements: Inserting Definition 1.5.2, our assumption (SCON) implies

$$\phi\big(\sigma(s)\big) \leq (1-s)\phi(\sigma_0) + s\phi(\sigma_1) - \frac{\lambda}{2}s(1-s)d^2(\sigma_0,\sigma_1) \leq (1-s)\phi(\sigma_0) + s\phi(\sigma_1).$$

Now adding this estimate and (2.32) multiplied by $\frac{1}{2\tau}$ up, yields

$$\frac{1}{2\tau}W_2^2\big(\mu,\sigma(s)\big) + \phi\big(\sigma(s)\big) \leq (1-s)\left(\frac{1}{2\tau}W_2^2(\mu,\sigma_0) + \phi(\sigma_0)\right) + \\ + s\left(\frac{1}{2\tau}W_2^2(\mu,\sigma_1) + \phi(\sigma_1)\right) - \frac{1}{2\tau}s(1-s)W_2^2(\sigma_0,\sigma_1),$$

which may be written more compactly as

$$\Phi\big(\tau,\mu,\sigma(s)\big) \leq (1-s)\Phi(\tau,\mu,\sigma_0) + s\Phi(\tau,\mu,\sigma_1) - \frac{1}{2\tau}s(1-s)W_2^2(\sigma_0,\sigma_1) \qquad \forall s \in [0,1]. \tag{2.33}$$

Choosing $\mu := M_\tau^{n-1}$, $\sigma_0 := M_\tau^n$ in (2.33) and invoking the definition of the minimizing movements $(M_\tau^n)_{n\in\mathbb{N}}$, we find that

$$\Phi\big(\tau,M_\tau^{n-1},M_\tau^n\big) \leq \Phi\big(\tau,M_\tau^{n-1},\sigma(s)\big) \leq \\ \leq (1-s)\Phi\big(\tau,M_\tau^{n-1},M_\tau^n\big) + s\Phi\big(\tau,M_\tau^{n-1},\sigma_1\big) - \frac{1}{2\tau}s(1-s)W_2^2\big(M_\tau^n,\sigma_1\big).$$

Actually, the terms in this estimate may be rearranged such that

$$0 \leq \Phi\big(\tau,M_\tau^{n-1},\sigma_1\big) - \Phi\big(\tau,M_\tau^{n-1},M_\tau^n\big) - \frac{1}{2\tau}(1-s)W_2^2\big(M_\tau^n,\sigma_1\big) \qquad \forall s \in (0,1]. \tag{2.34}$$

Now taking the limit $(s \searrow 0)$ in (2.34) implies

$$0 \leq \Phi\big(\tau,M_\tau^{n-1},\sigma_1\big) - \Phi\big(\tau,M_\tau^{n-1},M_\tau^n\big) - \frac{1}{2\tau}W_2^2\big(M_\tau^n,\sigma_1\big) \qquad \forall \sigma_1 \in \operatorname{dom}\phi. \tag{2.35}$$

Inserting the definition of the functional Φ, this inequality may also be written as a discrete analogue of (EVI):

$$\frac{1}{2\tau}\left(W_2^2\big(M_\tau^n,\sigma_1\big) - W_2^2\big(M_\tau^{n-1},\sigma_1\big)\right) \leq \phi(\sigma_1) - \phi\big(M_\tau^n\big) - \frac{1}{2\tau}W_2^2\big(M_\tau^n,\sigma_1\big) \quad \forall \sigma_1 \in \operatorname{dom}\phi. \tag{2.36}$$

⇝ A continuous version of (EVI) for the minimizing movements: In this step of the proof we move towards a continuous version of (2.36). To this aim, we introduce the *linear interpolant* on the interval $(a,b] \subset \mathbb{R}$ between two points $x,y \in \mathbb{R}$:

$$\boldsymbol{I}_{(a,b]}^{x,y}(t) := \frac{b-t}{b-a}x + \frac{t-a}{b-a}y \qquad \forall t \in (a,b].$$

This notion allows us two define the following piecewise linear functions with respect to the uniform partition P_τ of $(0,+\infty)$ as introduced at the beginning of Section 1.5:

$$\phi_\tau(t) := \sum_{n=1}^{\infty} \boldsymbol{I}_{I_\tau^n}^{\phi(M_\tau^{n-1}),\,\phi(M_\tau^n)}(t) \qquad \forall t \in (0,+\infty);$$

$$W_\tau^2(t,\sigma) := \sum_{n=1}^{\infty} \boldsymbol{I}_{I_\tau^n}^{W_2^2(M_\tau^{n-1},\sigma),\,W_2^2(M_\tau^n,\sigma)}(t) \qquad \forall t \in (0,+\infty),\ \forall \sigma \in \operatorname{dom}\phi.$$

Clearly, the function $W_\tau^2(\cdot,\sigma)$ is smooth on each interval I_τ^n. Therefore, we may write

$$\frac{\mathrm{d}}{\mathrm{d}t}W_\tau^2(t,\sigma) = \frac{1}{\tau}\left(W_2^2(M_\tau^n,\sigma) - W_2^2(M_\tau^{n-1},\sigma)\right) \qquad \forall t \in I_\tau^n,\ \forall n \in \mathbb{N}.$$

Together with **(2.36)** and choosing $\sigma_1 = \sigma$, this identity yields

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}W_\tau^2(t,\sigma) \le \phi(\sigma) - \phi(M_\tau^n) - \frac{1}{2\tau}W_2^2(M_\tau^n,\sigma) \le \phi(\sigma) - \phi(M_\tau^n) \qquad t \in \mathbb{R}^+ \setminus P_\tau. \tag{2.37}$$

Introducing the residual

$$\begin{aligned}\frac{1}{2}R_\tau(t) &:= \phi_\tau(t) - \phi(M_\tau^n) = \frac{n\tau - t}{\tau}\phi(M_\tau^{n-1}) + \frac{t-(n-1)\tau}{\tau}\phi(M_\tau^n) - \phi(M_\tau^n) = \\ &= \frac{n\tau - t}{\tau}\phi(M_\tau^{n-1}) + \frac{t-(n-1)\tau-\tau}{\tau}\phi(M_\tau^n) = \frac{n\tau - t}{\tau}\left(\phi(M_\tau^{n-1}) - \phi(M_\tau^n)\right),\end{aligned}$$

inequality **(2.37)** becomes

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}W_\tau^2(t,\sigma) \le \phi(\sigma_1) - \phi_\tau(t) + \frac{1}{2}R_\tau(t) \qquad t \in \mathbb{R}^+ \setminus P_\tau,\ \forall \sigma \in \operatorname{dom}\phi. \tag{2.38}$$

Finally, note that the residual R_τ is non-negative since setting $\sigma_1 = M_\tau^{n-1}$ in **(2.36)** immediately implies that $\phi(M_\tau^{n-1}) - \phi(M_\tau^n) \ge 0$, in case that $M_\tau^{n-1} \in \operatorname{dom}\phi$.

⇝ Comparison of two different time steps: Now we introduce a different time step $\eta > 0$ with corresponding minimizing movements $(M_\eta^k)_{k\in\mathbb{N}}$ on the partition P_η. For comparison with $(M_\tau^n)_{n\in\mathbb{N}}$, we consider yet another piecewise linear interpolating function, namely

$$W_{\tau,\eta}^2(t,s) := \sum_{k=1}^{\infty} \boldsymbol{I}_{I_\eta^k}^{W_\tau^2(t,M_\eta^k),\, W_\tau^2(t,M_\eta^{k-1})}(s) \qquad \forall t,s \in (0,+\infty). \tag{2.39}$$

Unwrapping the definitions of the terms in **(2.39)**, we observe for $t \in I_\tau^n$, $s \in I_\eta^k$ that

$$W_{\tau,\eta}^2(t,s) = \frac{k\eta - s}{\eta}W_\tau^2(t,M_\eta^k) + \frac{s-(k-1)\eta}{\eta}W_\tau^2(t,M_\eta^{k-1}) = \tag{2.40.a}$$

$$= \frac{s-(k-1)\eta}{\eta}\cdot\frac{(n-1)\tau - t}{\tau}W_2^2(M_\tau^{n-1},M_\eta^{k-1}) + \frac{k\eta - s}{\eta}\cdot\frac{t-(n-1)\tau}{\tau}W_2^2(M_\tau^n,M_\eta^k) + \tag{2.40.b}$$

$$+ \frac{k\eta - s}{\eta}\cdot\frac{n\tau - t}{\tau}W_2^2(M_\tau^{n-1},M_\eta^k) + \frac{s-(k-1)\eta}{\eta}\cdot\frac{t-(n-1)\tau}{\tau}W_2^2(M_\tau^n,M_\eta^{k-1}) = \tag{2.40.c}$$

$$= \frac{n\tau - t}{\tau}W_\eta^2(s,M_\tau^n) + \frac{t-(n-1)\tau}{\tau}W_\eta^2(s,M_\tau^{n-1}) = W_{\eta,\tau}^2(s,t). \tag{2.40.d}$$

Obviously, this computation shows that we have $W_{\tau,\eta}^2(t,s) = W_{\eta,\tau}^2(s,t)$ for all $t,s > 0$. Furthermore, we may choose $t = n\tau$, $s = k\eta$ in **(2.40)** to infer that $W_{\tau,\eta}^2(n\tau,k\eta) = W_2^2(M_\tau^n,M_\eta^k)$ for all $k,n \in \mathbb{N}$.

Now we take a convex combination of **(2.38)** on I_η^k: Fix $s \in I_\eta^k$; then adding the two inequalities

(2.38) with $\sigma = M_\eta^k$, multiplied by $\frac{k\eta - s}{\eta}$,

(2.38) with $\sigma = M_\eta^{k-1}$, multiplied by $\frac{s-(k-1)\eta}{\eta}$

up, yields

$$\begin{aligned}&\frac{1}{2}\frac{\partial}{\partial t}\left(\frac{k\eta - s}{\eta}W_\tau^2(t,M_\eta^k) + \frac{s-(k-1)\eta}{\eta}W_\tau^2(t,M_\eta^{k-1})\right) \le \\ &\le \frac{k\eta - s}{\eta}\phi(M_\eta^k) + \frac{s-(k-1)\eta}{\eta}\phi(M_\eta^{k-1}) - \phi_\tau(t) + \frac{1}{2}R_\tau(t),\end{aligned}$$

which may be also written more clearly as

$$\frac{1}{2}\frac{\partial}{\partial t}W^2_{\tau,\eta}(t,s) \le \phi_\eta(s) - \phi_\tau(t) + \frac{1}{2}R_\tau(t) \qquad \forall t,s \in (0,+\infty) : t \notin P_\tau. \tag{2.41}$$

Moreover, reversing the roles of τ and η in (2.41) and recalling the identity $W^2_{\tau,\eta}(t,s) = W^2_{\eta,\tau}(s,t)$, we also obtain

$$\frac{1}{2}\frac{\partial}{\partial s}W^2_{\tau,\eta}(t,s) \le \phi_\tau(t) - \phi_\eta(s) + \frac{1}{2}R_\eta(s) \qquad \forall t,s \in (0,+\infty) : s \notin P_\eta. \tag{2.42}$$

Thus, summing (2.41) and (2.42) results in

$$\frac{\partial}{\partial t}W^2_{\tau,\eta}(t,s) + \frac{\partial}{\partial s}W^2_{\tau,\eta}(t,s) \le R_\tau(t) + R_\eta(s) \qquad \forall t,s \in (0,+\infty) : t \notin P_\tau,\ s \notin P_\eta.$$

Setting $t = s$ in this inequality and applying the *chain rule*, we eventually arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t}W^2_{\tau,\eta}(t,t) \le R_\tau(t) + R_\eta(t) \qquad \forall t \in (0,+\infty) \setminus \big(P_\tau \cup P_\eta\big). \tag{2.43}$$

Since all terms in (2.43) are continuous outside the λ-null set $P_\tau \cup P_\eta$, we may write (2.43) in an integrated form, which results in

$$W^2_{\tau,\eta}(T,T) \le W^2_{\tau,\eta}(0,0) \int_0^T R_\tau + R_\eta \,\mathrm{d}\lambda \qquad \forall T \in (0,+\infty). \tag{2.44}$$

To establish an estimate for the right-hand side of this inequality, observe that, by definition of R_τ, we have for every $N \in \mathbb{N}$

$$\int_0^{N\tau} R_\tau \,\mathrm{d}\lambda = \sum_{n=1}^{N} \int_{I^n_\tau} R_\tau \,\mathrm{d}\lambda = \sum_{n=1}^{N} \Big(\phi\big(M^{n-1}_\tau\big) - \phi\big(M^n_\tau\big)\Big) \int_{I^n_\tau} \frac{n\tau - t}{\tau} \,\mathrm{d}\lambda(t) \le$$
$$\le \sum_{n=1}^{N} \tau \Big(\phi\big(M^{n-1}_\tau\big) - \phi\big(M^n_\tau\big)\Big) = \tau \Big(\phi\big(M^0_\tau\big) - \phi\big(M^N_\tau\big)\Big).$$

In particular, the non-negativity of the functional ϕ implies

$$\int_0^\infty R_\tau \,\mathrm{d}\lambda \le \tau\phi\big(M^0_\tau\big). \tag{2.45}$$

Since an analogous estimate also holds for η, (2.44) ultimately becomes

$$W^2_{\tau,\eta}(T,T) \le W^2_{\tau,\eta}(0,0) + \tau\phi\big(M^0_\tau\big) + \eta\phi\big(M^0_\eta\big) \qquad \forall T \in (0,+\infty). \tag{2.46}$$

⇝ Convergence of the minimizing movements: Choosing $\sigma_1 = M^{n-1}_\tau$ in (2.36) yields

$$0 < \frac{1}{\tau}W_2^2\big(M^n_\tau, M^{n-1}_\tau\big) \le \phi\big(M^{n-1}_\tau\big) - \phi\big(M^n_\tau\big) \le \phi\big(M^{n-1}_\tau\big) \qquad \forall n \in \mathbb{N}.$$

This estimate immediately implies $\phi\big(M^n_\tau\big) \le \phi\big(M^0_\tau\big)$ for all $n \in \mathbb{N}$, as well as

$$W_2^2\big(M^n_\tau, M^{n-1}_\tau\big) \le \phi\big(M^0_\tau\big) \quad \text{and} \quad W_2^2\big(M^n_\eta, M^{n-1}_\eta\big) \le \phi\big(M^0_\eta\big) \qquad \forall n \in \mathbb{N}. \tag{2.47}$$

We need an appropriate estimate for the interpolating function $W_{\tau,\eta}$. Therefore, we observe that, in general, we have

$$x \wedge y \leq \boldsymbol{I}^{x,y}_{(a,b]}(t) \leq x \vee y \qquad \forall t \in (a,b],\ \forall x,y \in \mathbb{R}.$$

Assuming for the moment that $x \leq y$, these bounds imply

$$y = \boldsymbol{I}^{x,y}_{(a,b]}(b) \leq \boldsymbol{I}^{x,y}_{(a,b]}(t) + (y-x) + \big(\boldsymbol{I}^{x,y}_{(a,b]}(t) - x\big) = 2\boldsymbol{I}^{x,y}_{(a,b]}(t) + (y-2x) \qquad \forall t \in (a,b].$$

In the general situation, this estimate becomes

$$x \vee y \leq 2\boldsymbol{I}^{x,y}_{(a,b]}(t) + \big(x \vee y - 2(x \wedge y)\big) \qquad \forall t \in (a,b]. \tag{2.48}$$

Recall that the elementary inequality (2.9) with $p = 2$ immediately yields the following variant of the reverse triangle inequality for our minimizing movements:

$$W_2^2(M_\tau^{n-1},\sigma) \vee W_2^2(M_\tau^n,\sigma) - 2\left(W_2^2(M_\tau^{n-1},\sigma) \wedge W_2^2(M_\tau^n,\sigma)\right) \leq 2W_2^2(M_\tau^{n-1},M_\tau^n) \tag{2.49}$$

for all $\sigma \in \operatorname{dom}\phi$ and $k,n \in \mathbb{N}$. Similarly,

$$W_2^2(M_\tau^n,M_\eta^{k-1}) \vee W_2^2(M_\tau^n,M_\eta^k) - 2\left(W_2^2(M_\tau^n,M_\eta^{k-1}) \wedge W_2^2(M_\tau^n,M_\eta^k)\right) \leq \tag{2.50.a}$$

$$\leq 2W_2^2(M_\eta^{k-1},M_\eta^k) \tag{2.50.b}$$

for all $k,n \in \mathbb{N}$. Now consider $P_\tau \cup P_\eta$ as refinement of both partitions P_τ and P_η and assume that $t \in I_\tau^n \cap I_\eta^k$. Then combining (2.48) with (2.50) implies

$$W_{\tau,\eta}^2(n\tau,k\eta) = \boldsymbol{I}_{I_\eta^k}^{W_\tau^2(n\tau,M_\eta^k),\,W_\tau^2(n\tau,M_\eta^{k-1})}(k\eta) \leq \tag{2.51.a}$$

$$\leq 2W_{\tau,\eta}^2(n\tau,t) + W_\tau^2(n\tau,M_\eta^{k-1}) \vee W_\tau^2(n\tau,M_\eta^k) - \tag{2.51.b}$$

$$-\,2\left(W_\tau^2(n\tau,M_\eta^{k-1}) \wedge W_\tau^2(n\tau,M_\eta^k)\right) = \tag{2.51.c}$$

$$= 2W_{\tau,\eta}^2(n\tau,t) + W_2^2(M_\tau^n,M_\eta^{k-1}) \vee W_2^2(M_\tau^n,M_\eta^k) - \tag{2.51.d}$$

$$-\,2\left(W_2^2(M_\tau^n,M_\eta^{k-1}) \wedge W_2^2(M_\tau^n,M_\eta^k)\right) \leq \tag{2.51.e}$$

$$\leq 2W_{\tau,\eta}^2(n\tau,t) + 2W_2^2(M_\eta^{k-1},M_\eta^k). \tag{2.51.f}$$

It remains to establish an estimate for the first term in (2.51.f): Repeating the same argument as before, we may use (2.48) together with (2.49) to obtain for $\sigma \in \{M_\eta^{k-1},M_\eta^k\}$ that

$$W_\tau^2(n\tau,\sigma) := \boldsymbol{I}_{I_\tau^n}^{W_2^2(M_\tau^{n-1},\sigma),\,W_2^2(M_\tau^n,\sigma)}(n\tau) \leq$$

$$\leq 2W_\tau^2(t,\sigma) + W_2^2(M_\tau^{n-1},\sigma) \vee W_2^2(M_\tau^n,\sigma) - 2\left(W_2^2(M_\tau^{n-1},\sigma) \wedge W_2^2(M_\tau^n,\sigma)\right) \leq$$

$$\leq 2W_\tau^2(t,\sigma) + 2W_2^2(M_\tau^{n-1},M_\tau^n)$$

This estimate readily implies

$$W_{\tau,\eta}^2(n\tau,t) = \frac{k\eta - t}{\eta} W_\tau^2(n\tau,M_\eta^k) + \frac{t-(k-1)\eta}{\eta} W_\tau^2(n\tau,M_\eta^k) \leq \tag{2.52.a}$$

$$\leq 2\frac{k\eta - t}{\eta} W_\tau^2(t,M_\eta^k) + 2\frac{t-(k-1)\eta}{\eta} W_\tau^2(t,M_\eta^k) + 2W_2^2(M_\tau^{n-1},M_\tau^n) = \tag{2.52.b}$$

$$= 2\boldsymbol{I}_{I_\eta^k}^{W_\tau^2(n\tau,M_\eta^k),\,W_\tau^2(n\tau,M_\eta^{k-1})}(t) + 2W_2^2(M_\tau^{n-1},M_\tau^n) = 2W_{\tau,\eta}^2(t,t) + 2W_2^2(M_\tau^{n-1},M_\tau^n). \tag{2.52.c}$$

Putting together (2.51) and (2.52), we finally arrive at

$$W_{\tau,\eta}^2(n\tau,k\eta) \leq 4W_{\tau,\eta}^2(t,t) + 4W_2^2(M_\tau^{n-1},M_\tau^n) + 2W_2^2(M_\eta^{k-1},M_\eta^k) \qquad \forall t \in I_\tau^n \cap I_\eta^k. \tag{2.53}$$

Using (2.47), the fact that the discrete solutions $\overline{M}_\tau$, $\overline{M}_\eta$ are constant on each $I_\tau^n \cap I_\eta^k$ and recalling from the beginning of the previous step the identity $W_2^2(M_\tau^n, M_\eta^k) = W_{\tau,\eta}^2(n\tau, k\eta)$, (2.53) results in

$$W_2^2(\overline{M}_\tau(t), \overline{M}_\eta(t)) \le 4W_{\tau,\eta}^2(t,t) + 4\tau\phi(M_\tau^0) + 2\tau\phi(M_\eta^0) \qquad \forall t \in (0, +\infty).$$

Combining this estimate with (2.46), we easily get

$$W_2^2(\overline{M}_\tau(t), \overline{M}_\eta(t)) \le 4W_2^2(M_\tau^0, M_\eta^0) + 8\tau\phi(M_\tau^0) + 6\tau\phi(M_\eta^0) \qquad \forall t \in (0, +\infty). \tag{2.54}$$

Now taking $M_\tau^0 = M_\eta^0$ in (2.54) yields

$$\sup_{t>0} W_2^2(\overline{M}_\tau(t), \overline{M}_\eta(t)) \le +8\tau\phi(M_\tau^0) + 6\tau\phi(M_\eta^0), \tag{2.55}$$

which shows that for every sequence $(\tau_n)_{n\in\mathbb{N}}$, converging to zero in $\mathbb{R}^+$, $(\overline{M}_{\tau_n}(t))_{\mathbb{N}}$ is a Cauchy sequence in the complete space $(\mathcal{P}_r^2(\mathbb{R}^n), W_2)$, uniformly with respect to $t \in \mathbb{R}_0^+$. Therefore, there exists a limit curve $\mu(t)$ in $(\mathcal{P}_r^2(\mathbb{R}^n))$ such that $\lim_{n\to\infty} \overline{M}_{\tau_n}(t) = \mu(t)$ for all $t \in \mathbb{R}_0^+$.
In particular, for every choice of $\tau > 0$, $t \in P_\tau$, we may pass to the limit $(\eta \searrow 0)$ in (2.46) to obtain a rough error estimate on every partition P_τ:

$$\sup_{t\in P_\tau} W_2^2(\overline{M}_\tau, \mu(t)) \le W_2^2(M_\tau^0, \mu(0)) + \tau\phi(M_\tau^0). \tag{2.56}$$

⇝ The limit curve $\mu(t)$ satisfies an integrated form of (EVI): Fix an arbitrary $\sigma \in \operatorname{dom}\phi$. Recalling that the expressions in (2.38) are continuous outside the λ-null set P_τ, we may integrate (2.38) over any bounded interval $(a,b) \subset \mathbb{R}^+$ to get

$$\frac{1}{2}W_\tau^2(b,\sigma) - \frac{1}{2}W_\tau^2(a,\sigma) + \int_a^b \phi_\tau \,\mathrm{d}\lambda \le \tag{2.57.a}$$

$$\le (b-a)\phi(\sigma) + \frac{1}{2}\int_a^b R_\tau \,\mathrm{d}\lambda \le (b-a)\phi(\sigma) + \frac{\tau}{2}\phi(M_\tau^0), \tag{2.57.b}$$

where we used (2.45) and the fact that R_τ depicts a non-negative function. Observe that we have

$$\lim_{\tau\searrow 0} W_\tau^2(t,\sigma) = W_2^2(\mu(t),\sigma), \quad \liminf_{\tau\searrow 0} \phi_\tau = \phi(\mu(t)), \quad \lim_{\tau\searrow 0} \phi(M_\tau^0) = \phi(\mu(0)) < +\infty. \tag{2.58}$$

Thus, Proposition A.4.5.ii yields

$$\int_a^b \max\{N, \phi(\mu(t))\}\,\mathrm{d}\lambda(t) \le \liminf_{\tau\searrow 0} \int_a^b \max\{N, \phi_\tau)\}\,\mathrm{d}\lambda(t) \qquad \forall N \in \mathbb{N}.$$

and therefore, by *monotone convergence*,

$$\int_a^b \phi(\mu(t))\,\mathrm{d}\lambda(t) = \liminf_{\tau\searrow 0} \int_a^b \phi_\tau \,\mathrm{d}\lambda(t).$$

Using this estimate together with (2.58), we may pass to the limit $(\tau \searrow 0)$ in (2.57) and immediately obtain the evolution variational inequality with respect to $\boldsymbol{\lambda}_0 = 0$ in the integrated form

$$\frac{1}{2}W_2^2(\mu(b),\sigma) - \frac{1}{2}W_2^2(\mu(a),\sigma) + \int_a^b \phi(\mu(t))\,\mathrm{d}\lambda(t) \le (b-a)\phi(\sigma) \qquad \forall \sigma \in \operatorname{dom}\phi. \tag{2.59}$$

$\mu(t)$ is a gradient flow in the EVI sense: It remains to be shown that $\mu(t)$ is an EVI gradient flow in the sense of Definition 1.3.6: First note that for every pair $s, t > 0$ we may choose $\tau > 0$ and $n \in \mathbb{N}$ such that $s, t \in I_\tau^n$. Thus, $\overline{M}_\tau(t) = \overline{M}_\tau(s)$ implies

$$W_2^2\big(\mu(t), \mu(s)\big) \leq 2W_2^2\big(\mu(t), \overline{M}_\tau(t)\big) + 2W_2^2\big(\mu(s), \overline{M}_\tau(s)\big).$$

Since the estimate in (2.55) is uniform with respect to t, we easily infer from the inequality given above that $\mu(t)$ is a continuous curve in $\mathcal{P}_r^n(\mathbb{R}^n)$. Consequently, $\phi\big(\mu(t)\big)$ is lower semi-continuous.

Our next claim is that $\phi\big(\mu(t)\big) \leq \phi\big(\mu(0)\big)$ for all $t > 0$. We prove this statement by contradiction: To this aim, assume that $\phi\big(\mu(t_0)\big) > \phi\big(\mu(0)\big)$ for some $t_0 > 0$. Consider

$$A := \big\{t \in (0, t_0) : \phi\big(\mu(t)\big) \leq \phi\big(\mu(0)\big)\big\} \subseteq \operatorname{dom} \phi$$

and set $T := \sup A$. Then the lower semi-continuity of $\phi\big(\mu(t)\big)$ implies

$$\phi\big(\mu(T)\big) \leq \liminf_{\substack{t \to T \\ t \in A}} \phi\big(\mu(t)\big) \leq \phi\big(\mu(0)\big).$$

Therefore, T belongs to the set A and $\mu(T) \in \operatorname{dom} \phi$. In particular, we have $\phi\big(\mu(s)\big) > \phi\big(\mu(0)\big)$ for all $s \in (T, t_0]$. Hence,

$$\int_T^{t_0} \phi\big(\mu(s)\big)\, \mathrm{d}\lambda(s) > (t_0 - T)\phi\big(\mu(T)\big). \tag{2.60}$$

On the other hand, choosing $(a, b) = (T, t_0)$ and $\sigma = \mu(T)$ in (2.59), immediately implies

$$\int_T^{t_0} \phi\big(\mu(s)\big)\, \mathrm{d}\lambda(s) \leq (t_0 - T)\phi\big(\mu(T)\big). \tag{2.61}$$

Since (2.60) and (2.61) contradict each other, our claim is proven.

It is clear from (2.59) that $\phi\big(\mu(t)\big)$ belongs to $L^1_{\mathrm{loc}}\big(\mathbb{R}_0^+, \mathbb{R}_0^+ \cup \{+\infty\}\big)$ and thus $\mu(t) \in \operatorname{dom} \phi$ for λ-a.e. $t \geq 0$. However, one readily infers that $\mu(t) \in \operatorname{dom} \phi$ for *all* $t \geq 0$: For every $t \geq 0$ there exists a sequence $(t_n)_{n \in \mathbb{N}}$ in $(0, +\infty)$ with $\lim_{n \to \infty} t_n = t$ and $\mu(t_n) \in \operatorname{dom} \phi$ for all $n \in \mathbb{N}$. Now we may use our claim stated above and the lower semi-continuity of $\phi\big(\mu(t)\big)$ to infer

$$\phi\big(\mu(t)\big) \leq \liminf_{n \to \infty} \phi\big(\mu(t_n)\big) \leq \phi\big(\mu(0)\big) < +\infty.$$

Finally, we are able to show that $\mu(t)$ belongs to $AC_{\mathrm{loc}}\big((0, +\infty), \mathcal{P}_r^n(\mathbb{R}^n)\big)$: Indeed, for any bounded interval $(a, b) \subset \mathbb{R}^+$, setting $\sigma = \mu(a)$ in (2.59) at once yields

$$W_2^2\big(\mu(a), \mu(b)\big) \leq 2(b - a)\phi\big(\mu(a)\big) \leq \int_a^b 2\phi\big(\mu(0)\big)\, \mathrm{d}\lambda(t) < +\infty.$$

Having established that $\mu(t)$ is locally absolutely continuous, the reverse triangle inequality and Fact 1.1.4.ii imply that $t \mapsto W_2^2\big(\mu(t), \sigma\big)$ belongs to $AC_{\mathrm{loc}}((0, +\infty), \mathbb{R})$ for all $\sigma \in \operatorname{dom} \phi$. Fix any Lebesgue point $t_0 \in (0, +\infty)$ of $\phi\big(\mu(t)\big)$ such that $t \mapsto W_2^2\big(\mu(t), \sigma\big)$ is differentiable at t. Then, recalling Fact A.2.7.ii, we may set $a = t_0 - \varepsilon$, $b = t_0 + \varepsilon$ in (2.59), divide this inequality by 2ε, and ultimately pass to the limit ($\varepsilon \searrow 0$) to obtain the variational evolution inequality in the desired form (1.45) with respect to $\boldsymbol{\lambda}_0 = 0$. ∎

Although the $\mathcal{P}_r^2(\mathbb{R}^n)$ does not belong to the class of NPC spaces, it is entirely possible to apply the general existence theory we have introduced in Section 1.5.

2.3.5 **Remark** (Convexity along generalized geodesics) Recall from (2.3.2) that (1.66) does not hold for every geodesic γ in $\mathscr{P}_r^2(\mathbb{R}^2)$. However, one can show that (1.66) does hold for generalized geodesics, which another fine property of these interpolating curves. Then one may require (SCON) to hold for a lower semi-continuous functional ϕ to deduce that $\Phi(\tau, M, \cdot)$ as defined in Definition 1.5.2 satisfies condition (CON). Thus, one is able to invoke the general existence theory of Theorem 1.5.3.

In particular, Theorem 1.5.3.iv yields the optimal error estimate $W_2\big(\mu(t), \overline{M}_\tau(t)\big) \leq \mathcal{O}(\tau)$, whereas the proof of Theorem 2.3.4 provides only the subpar bound $\mathcal{O}(\sqrt{\tau})$ for the rate of convergence by means of (2.56).

The approach, relying on the general existence theory, is due to Ambrosio, Gigli, Savaré and can be found in Section 9.2 of [3].

We conclude this chapter with an fundamental example that illustrates a promising application of gradient flows in $\mathscr{P}_r^2(\mathbb{R}^2)$ within the field of partial differential equations.

The subsequent discussion will only provide an informal prospect on an important application of the theory developed so far and will be kept at a sketchy level. However, everything can be done rigorously; we refer to Section 3.3 in [2] or Chapter 10 in [3] for the details.

2.3.6 **Example** (Heat equation) We introduce the *Entropy functional* $E : \mathscr{P}_r^2(\mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$, defined by

$$E(\nu) := \begin{cases} \int\limits_{\mathbb{R}^n} \frac{\mathrm{d}\nu}{\mathrm{d}\lambda_n} \log\Big(\frac{\mathrm{d}\nu}{\mathrm{d}\lambda_n}\Big)\, \mathrm{d}\lambda_n, & \text{if } \nu \ll \lambda_n \\ +\infty, & \text{otherwise,} \end{cases}$$

where $\frac{\mathrm{d}\nu}{\mathrm{d}\lambda_n}$ denotes the *Radon–Nikodym derivative* of ν with respect to λ_n. Now fix any initial datum $\rho_0 \in \mathscr{P}_r^2(\mathbb{R}^n)$ with $\rho_0 \ll \lambda_n$ and assume for every $\tau > 0$ there exists a minimizing movement ρ_τ, i.e.

$$\Phi(\tau, \rho_0, \rho_\tau) \leq \Phi(\tau, \rho_0, \nu) \qquad \forall \nu \in \mathscr{P}_r^2(\mathbb{R}^n).$$

Moreover, we assume that ρ_τ is absolutely continuous with respect to λ_n.

Now we claim that ρ_τ satisfies

$$\frac{1}{\tau}\int\limits_{\mathbb{R}^n} \varphi \,\mathrm{d}(\rho_\tau - \rho_0) = \int\limits_{\mathbb{R}^n} \Delta\phi \,\mathrm{d}\rho_\tau + \mathcal{O}(\tau) \qquad \forall \varphi \in C_c^\infty(\mathbb{R}^n)\,.$$

This identity shows that ρ_τ is a first order approximation of the distributional solution, starting at ρ_0, of the homogeneous heat equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \Delta\rho. \tag{2.62}$$

It can be shown that the Entropy functional E satisfies the assumptions of Theorem 2.3.4. Then one may infer that the discrete solutions of the corresponding minimizing movements with initial datum ρ_0 converge to a limit curve $\rho(t)$ which depicts a distributional solution of (2.62). Since Theorem 2.3.4 implies that $\rho(t)$ is a gradient flow in the EVI sense, $\rho(s)$ satisfies the contraction and regularizing properties obtained in Section 1.4. In particular, $\rho(t)$ is locally Lipschitz continuous and the unique solution of the heat equation. ◀

Appendix A

In this appendix we state some important results from basic measure theory without proof. Basically, we will follow the exposure of Bogachev, sections 5.2 to 5.6 of [12], although almost any other text about measure theory should cover the following results as well.

Section A.3 is devoted to measure theory on topological spaces with emphasis on the *Riesz representation theorems*. The material from this section is borrowed from Bogachev (chapter 7 in [12]), as well as Cohn (chapter 7 in [22]).

In Section A.4 a brief overview of the *weak convergence of measures* on topological spaces is given. An elaborate treatment of this topic can be found in chapter 8 of [12].

A.1 Functions of Bounded Variation

A.1.1 Definition A function $f : T \to \mathbb{R}$, $T \subseteq \mathbb{R}$ is of ***bounded variation*** if one has

$$V(f,T) := \sup \sum_{i=1}^{n} |f(t_{i+1}) - f(t_i)| < \infty,$$

where the supremum is taken over all finite collections $t_1 \leq t_2 \leq \cdots \leq t_{n+1}$ in T. The non-negative quantity $V(f,T)$ is called ***variation*** of f on T.

A function $f : T \to \mathbb{R}$ is of ***essential bounded variation*** if there exists a function $g : T \to \mathbb{R}$ of bounded variation such that $f(x) = g(x)$ λ-a.e. in T. In this case we set $V(f,T) := V(g,T)$.

A.1.2 Facts Let $f : [a,b] \to \mathbb{R}$ be of bounded variation.

(i) The functions $V : x \mapsto V(f,[a,x])$ and $U : x \mapsto V(x) - f(x)$ are non-decreasing on $[a,b]$.

(ii) The function V is continous at a point $x_0 \in [a,b]$ iff the function f is continuous at this point.

(iii) For every $c \in (a,b)$ one has $V(f,[a,b]) = V(f,[a,c]) + V(f,[c,b])$.

(iv) A continuous function of bounded variation is the difference of two continuous non-decreasing functions.

(v) Every function of bounded variation has at most countably many points of discontinuity.

Finally, we have the following important result, concerning differentiability of functions of bounded variation.

A.1.3 Proposition *Let $f : [a,b] \to \mathbb{R}$ be a function of bounded variation. Then f has a finite derivative almost everywhere on [a,b].*

A.2 Absolutely Continuous Functions on the Real Line

In this section we consider only functions with values in $\mathbb{R}$.

A.2.1 Definition A function $f : [a,b] \to \mathbb{R}$ is called ***absolutely continuous*** if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \varepsilon \tag{A.1}$$

for every finite collection of pairwise disjoint intervals $(a_i, b_i) \subset [a,b]$ with $\sum_{i=1}^{n} |b_i - a_i| < \delta$.

A.2.2 Facts Let absolutely continuous functions $f, g : [a,b] \to \mathbb{R}$ be given.

(i) We immediately deduce from the definition above that every absolutely continuous function is also uniformly continuous.

(ii) The functions fg and $f+g$ are absolutely continuous. Moreover, if $g \geq c > 0$, then f/g is absolutely continuous.

(iii) Consider two absolutely continuous functions $f : [a,b] \to \mathbb{R}$ and $g : [c,d] \to [a,b]$. Then, under the additional hypothesis that g is monotone on $[c,d]$, the composition $f \circ g$ is again absolutely continuous.

The statements in Fact 1.1.4.ii are simple implications of this rather technical lemma.

A.2.3 Lemma *Let $(f_i)_{i \in I}$ be a family of* uniformly absolutely continuous functions $f_i : [a,b] \to \mathbb{R}$, $i \in I$, *i.e. for every $\varepsilon > 0$ there exists $\delta > 0$ such that* (A.2) *holds for every for every f_i, $i \in I$ and every finite collection of pairwise disjoint intervals $(a_i, b_i) \subset [a,b]$ with $\sum_{i=1}^{n} |b_i - a_i| < \delta$. Then the following statement holds:*
Suppose that $\psi : U \to \mathbb{R}$ is a Lipschitz map and $(I_j)_{j \leq n}$ a finite collection of index sets $I_j \subseteq I$ such that $\big(f_{i_1}(x), \ldots, f_{i_n}(x)\big) \in U \subseteq \mathbb{R}^n$ for all $x \in [a,b]$ and all $i_j \in I_j$, $1 \leq j \leq n$. Then the function $\psi\big(f_{i_1}, \ldots, f_{i_n}\big)$ is absolutely continuous on the interval $[a,b]$ for every choice of $\big(f_{i_1}, \ldots, f_{i_n}\big)$.

Note that a finite family $(f_i)_{i \in I}$ absolutely continuous functions $f_i : [a,b] \to \mathbb{R}$, $i \in I$ is trivially uniformly absolutely continuous.

The following crucial characterization of absolutely continuous functions is a generalization of the classical *fundamental theorem of calculus*.

A.2.4 Theorem *A function $f : [a,b] \to \mathbb{R}$ is absolutely continuous iff there exists $g \in L^1([a,b], \mathbb{R})$ such that*

$$f(x) = f(a) + \int_a^x g \, \mathrm{d}\lambda, \qquad \forall x \in [a,b]. \tag{A.2}$$

We proceed in this section with some results which show that an absolutely continuous functions is closely related to its variation.

A.2.5 Facts Let $f : [a,b] \to \mathbb{R}$ be absolutely continuous.

(i) Then f is of bounded variation. In particular f has a finite derivative λ-almost everywhere on $[a,b]$ and necessarily $f'(x) = g(x)$ λ-almost everywhere on $[a,b]$, where g is given as in (A.2).

(ii) The function $V : x \mapsto V(f, [a,x])$ is absolutely continuous as well.

(iii) If (A.2) holds, then

$$V(f, [a,b]) = \int_a^b |g| \, \mathrm{d}\lambda.$$

Under the assumption that the integrand g in (A.2) is continuous, it is well known that g can be recovered by means of differentiation of f. Non surprisingly, a slightly weaker result holds under the broader assumption of mere integrability of g. Then one can show that the function $x \mapsto \int_a^x g \, \mathrm{d}\lambda$ is absolutely continuous on $[a,b]$ and one arrives at the so-called *Newton-Leibniz formula*.

A.2.6 Theorem (Newton-Leibniz) *Let a function $f \in L^1([a,b], \mathbb{R})$ be given. Then*

$$\frac{\mathrm{d}}{\mathrm{d}x} \int_a^x f \, \mathrm{d}\lambda = f(x) \quad \lambda\text{-almost everywhere on } [a,b]. \tag{A.3}$$

A.2.7 Facts Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ such that f is integrable on every ball $B_r(x)$ in $\mathbb{R}^n$.

(i) Since all balls $B_r(x)$ are relative compact sets in $\mathbb{R}^n$, f does not need to be integrable on the whole space. Nevertheless (A.3) holds in the case $n = 1$ because the differentiability of a function is a local property.

(ii) In the the one-dimensional case the derivative in (A.3) may be written in terms of a symmetric difference quotient which results in

$$\lim_{\varepsilon \searrow 0} \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} f \, d\lambda = f(x) \quad \text{for } \lambda\text{-almost every } x \in \mathbb{R}.$$

(iii) Moreover, in the general case $n \in \mathbb{N}$, Theorem A.2.6 admits a multidimensional generalization to

$$\lim_{r \searrow 0} \frac{1}{\lambda_n(B_r(x))} \int_{B_r(x)} f \, d\lambda_n = f(x) \quad \text{for } \lambda_n\text{-almost every } x \in \mathbb{R}^n. \tag{A.4}$$

(iv) The equality in (A.4) holds iff x is a *Lebesgue point* of f, i.e. iff x satisfies

$$\lim_{r \searrow 0} \frac{1}{\lambda_n(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| \, d\lambda_n(y) = 0. \tag{A.5}$$

Hence, λ_n-almost every point in $\mathbb{R}^n$ is a Lebesgue point of f.

(v) The function f is uniformly continuous on $\mathbb{R}^n$ iff the limit in (A.5) is uniform with respect to the variable x.
Note that the premise of uniformity of the limit cannot be weakened and one cannot even expect f to be continuous in this case (cf. chapter 2 in [33] for an counterexample and the generalization of this result to Hölder continuous functions).

A useful corollary of Theorem A.2.6 is the *change of variables formula* for absolutely continuous transformations.

A.2.8 Corollary (Change of variables) *Let $\varphi : [c,d] \to \mathbb{R}$ be an increasing absolutely continuous function. Then for every $f \in L^1([a,b],\mathbb{R})$ such that $F([c,d]) \subseteq [a,b]$ for $F(x) := \int_a^b f \, d\lambda$, the function $f \circ \varphi \cdot \varphi'$ also belongs to $L^1([c,d],\mathbb{R})$ and one has*

$$\int_{\varphi(c)}^{\varphi(d)} f \, d\lambda = \int_c^d f \circ \varphi \cdot \varphi' \, d\lambda.$$

This assertion remains true for unbounded intervals of the form $(-\infty, d]$, $[c, +\infty)$, and $(-\infty, +\infty)$.

There exists also a version of the *integration by parts formula* for absolutely continuous functions.

A.2.9 Corollary (Integration by parts) *For every choice of absolutely continuous functions $f, g : [a,b] \to \mathbb{R}$ we have*

$$\int_a^b f' g \, d\lambda = f(b)g(b) - f(a)g(a) - \int_a^b f g' \, d\lambda.$$

Another application of Theorem A.2.6 is the following characterization of the Sobolev space $W^{1,1}((a,b))$. For a more elaborate treatment confer section 7.2 in [37].

A.2.10 Example (Absolute continuity vs. Sobolev spaces) Recall that the *Sobolev space* $W^{1,1}((a,b))$ consists of the subset of functions $f \in L^1((a,b),\mathbb{R})$ such that their weak derivative also belong to $L^1((a,b),\mathbb{R})$. We say that $u \in L^1((a,b),\mathbb{R})$ is a *weak derivative* of the function $f \in L^1((a,b),\mathbb{R})$ if

$$\int_a^b u\zeta \, \mathrm{d}\lambda = -\int_a^b f\zeta' \, \mathrm{d}\lambda \qquad \forall \zeta \in C_c^\infty((a,b)) .$$

We denote the weak derivative of f, in accordance with the classical derivative, with f'.
For a given function $f \in W^{1,1}((a,b))$ we set

$$g(t) := \int_a^t f' \, \mathrm{d}\lambda \qquad \forall t \in (a,b).$$

We immediately notice that g absolutely continuous, due to Theorem A.2.4. Moreover, g belongs to $W^{1,1}((a,b))$ since integration by parts and the Newton-Leibniz formula give us

$$\int_a^b g\zeta' \, \mathrm{d}\lambda = -\int_a^b g'\zeta \, \mathrm{d}\lambda = -\int_a^b f'\zeta \, \mathrm{d}\lambda \qquad \forall \zeta \in C_c^\infty((a,b)) .$$

As $g' = f'$ λ-a.e. in (a,b), we obtain

$$\int_a^b (g-f)\zeta' \, \mathrm{d}\lambda = \int_a^b (g'-f')\zeta d\lambda = 0 \qquad \forall \zeta \in C_c^\infty((a,b)) , \tag{A.6}$$

Since $C_c^\infty((a,b))$ is dense in $C_0((a,b))$ with respect to the uniform norm $\|\cdot\|_\infty$, the equality in (A.6) extends to all functions $\zeta \in C_0((a,b))$. As a result, setting

$$\psi(x) := \eta(x) - \frac{1}{b-a}\int_a^b \eta \, \mathrm{d}\lambda \qquad \text{and} \qquad \phi(x) := \int_a^x \psi \, \mathrm{d}\lambda \qquad \forall x \in (a,b)$$

for $\eta \in C((a,b))$, we obtain

$$\int_a^b (g-f)\psi \, \mathrm{d}\lambda = \int_a^b (g-f)\phi' \, \mathrm{d}\lambda = 0.$$

Utilizing *Fubini's theorem*, this can be written in the form

$$0 = \int_a^b (g-f)\psi \, \mathrm{d}\lambda = \int_a^b \eta \Big(g - f + \frac{1}{b-a}\int_a^b g - f \, \mathrm{d}\lambda\Big) \, \mathrm{d}\lambda \qquad \forall \eta \in C((a,b)) .$$

Hence, we conclude that f is λ-a.e. equal to an absolutely continuous function $g + c$ with the constant $c = \frac{1}{b-a}\int_a^b g - f \, \mathrm{d}\lambda$. ◀

We conclude this section with the following example which shows the connection between absolute continuous and Lipschitz continuous functions.

A.2.11 Example (Absolute continuity vs. Lipschitz continuity on the real line)
↝ Consider a function $f : [a,b] \to \mathbb{R}$ which is Lipschitz continuous, i.e. there exists a constant $C_{Lip} > 0$ such that

$$|f(s) - f(t)| \le C_{Lip} |s - t| \qquad \forall s,t \in [a,b]. \tag{A.7}$$

Clearly, from (A.7) follows that f also satisfies Definition 1.1.1 of absolute continuity on the real line.

Moreover, recall from Fact A.2.5.i that f is differentiable λ-almost everywhere on $[a,b]$, and let $t \in [a,b]$ be a point where the derivative of f exists. Then we have the estimate

$$\left|\frac{f(s)-f(t)}{s-t}\right| \leq C_{Lip} \qquad \forall s,t \in [a,b] : s \neq t.$$

We infer that $|f'(t)| \leq C_{Lip}$.

⇝ On the other hand, let an absolute continuous function $f : [a,b] \to \mathbb{R}$ and a constant $C > 0$ be given such that $f'(t) \leq C$ λ-almost everywhere on $[a,b]$. Due to Theorem A.2.4, we obtain

$$|f(t) - f(s)| \leq \int_s^t |f'| \,\mathrm{d}\lambda \leq \int_s^t C \,\mathrm{d}\lambda = C\,|t-s| \qquad \forall s,t \in [a,b] : s \leq t.$$

Thus, f is Lipschitz continuous on $[a,b]$. ◀

A.3 Measures on Topological Spaces

A topological spaces carries a natural σ-algebra which is closely related to the topology of the space:

A.3.1 **Definition** The ***Borel σ-algebra*** $\mathcal{B}(X)$ of a topological space $(X,\mathcal{T})$ is the smallest σ-algebra which contains all open sets of X. The elements of $\mathcal{B}(X)$ are called the ***Borel sets*** in X.

A mapping $f : X \to Y$ between topological spaces $(X,\mathcal{T})$ and $(Y,\mathcal{O})$ is called ***Borel*** (or *Borel measurable*) if $f^{-1}(\mathcal{B}(Y)) \subseteq \mathcal{B}(X)$.

Clearly, every continuous function $f : X \to Y$ is also Borel.

Next, we will consider two important classes of *signed measures* on topological spaces:

A.3.2 **Definition** Let $(X,\mathcal{T})$ be a Hausdorff space.

(i) A countably additive signed measure on the Borel σ-algebra $\mathcal{B}(X)$ is called a ***Borel measure*** on X.

(ii) A Borel measure μ on X is called a ***(finite) Radon measure*** if its variation $|\mu|$ is finite and for every Borel set $B \in \mathcal{B}(X)$ and $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq B$ such that

$$|\mu|(B \setminus K_\varepsilon) < \varepsilon. \tag{A.8}$$

Obviously, every Radon measure is also a Borel measure. Although the converse statement is not true on arbitrary topological spaces, the class of Borel measures with finite variation and the class of Radon measures coincide on Polish spaces.

Every Radon measure μ on a compact Hausdorff space $(K,\mathcal{T})$ defines a continuous linear functional on the Banach space $C(K)$ equipped with the uniform norm $\|\cdot\|_\infty$ via the formula $f \mapsto \int_K f \,\mathrm{d}\mu$. The following version of the *Riesz representation theorem* shows that the converse statement is also valid.

A.3.3 **Theorem** *Let $(K,\mathcal{T})$ be a compact Hausdorff space. Then for every continuous linear functional L on the Banach space $C(K)$, there exists a unique finite Radon measure μ such that*

$$L(f) = \int_K f \,\mathrm{d}\mu \qquad \forall f \in C(K).$$

Moreover, L is a positive functional precisely if μ is a non-negative measure.

More generally, one can state a similar result for locally compact Hausdorff spaces. To this end, recall that $C_0(X)$ denotes the space of continuous functions which vanish at infinity, i.e. for every function $f \in C_0(X)$ and $\varepsilon > 0$, there exits a compact set $K_\varepsilon \subseteq X$ such that $|f(x)| < \varepsilon$ on $X \setminus K_\varepsilon$.

A.3.4 Theorem (Riesz-Markov) *Let $(X, \mathcal{T})$ be a locally compact Hausdorff space. Then for any continuous linear functional L on the space $C_0(X)$, there exists a unique Borel measure μ on X such that μ is Radon on all sets of finite measure and*

$$L(f) = \int_X f \, \mathrm{d}\mu \qquad \forall f \in C_0(X).$$

The functional L is positive precisely if the Borel measure μ is non-negative.

In the even more general case that the topological space $(X, \mathcal{T})$ is not locally compact, one can still establish a one-to-one correspondence between Radon measures and certain continuous linear functionals on $C_\mathrm{b}(X)$, provided that $(X, \mathcal{T})$ is Tychonoff.

Recall that a *Tychonoff space* is a Hausdorff space such that any closed set C and any point $x \notin C$ can be separated by a continuous function. It turns out that every locally compact Hausdorff space as well as every metric space is Tychonoff.

A.3.5 Theorem *Let $(X, \mathcal{T})$ be a Tychonoff space. Then*

$$L(f) = \int_X f \, \mathrm{d}\mu \qquad \forall f \in C_\mathrm{b}(X) \tag{A.9}$$

establishes a one-to-one correspondence between Radon measures μ on X and the class of continuous linear functionals L on $C_\mathrm{b}(X)$ which satisfy the following condition:

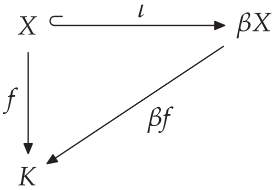
(RAD) *For every $\varepsilon > 0$ there exists a compact set K_ε such that*

$$|L(f)| \leq \varepsilon \, \|f\|_\infty \qquad \forall f \in C_\mathrm{b}(X) : f\big|_{K_\varepsilon} = 0.$$

In particular, formula (A.9) *establishes a one-to-one correspondence between non-negative Radon measures μ on X and all positive linear functionals L on $C_\mathrm{b}(X)$ which satisfy* (RAD).

Another option is to consider the *Stone-Čech compactification* of a non-compact Tychonoff space, where Theorem A.3.3 is applicable.

A.3.6 Example (Stone-Čech compactification) Consider a Tychonoff space $(X, \mathcal{T})$. Then there exists a compact Hausdorff space βX, together with an embedding $\iota : X \to \beta X$, i.e. the map ι is a homeomorphism onto its image, such that the following universal property is satisfied: For every compact Hausdorff space K and every continuous map $f : X \to K$ there exists a unique continuous lifting $\beta f : \beta X \to K$ such that the following diagram commutes:



The space βX is called *Stone-Čech compactification* of X. Note that in general $\iota(X)$ need not be open in βX. However, the embedding ι depicts an open map precisely when the underlying space X is locally compact Hausdorff.

In this situation, every set O which belongs to the subspace topology $\mathcal{T}_{\iota(X)}$, is also open in βX. Therefore, we obtain

$$\mathcal{B}(\mathcal{T}_{\iota(X)}) = \{B \cap \iota(X) : B \in \mathcal{B}(\beta X)\}.$$

Hence, one can extend any given Borel measure μ on $(X, \mathcal{T})$ to βX by the formula

$$\mu_\beta(B) := \iota_* \mu(B \cap X) = \mu\big(\iota^{-1}(B \cap X)\big) \qquad \forall B \in \mathcal{B}(\beta X).$$

In particular, μ_β is Radon iff μ is Radon. ◀

A.3.7 Remark Consider a finite Radon measure μ on a locally compact Hausdorff space $(X, \mathcal{T})$. Property (A.8) of μ means that the measure can be approximated on a Borel set $B \in \mathcal{B}(X)$ from within by compact subsets of B. A similar property holds for the approximation of μ from without by open supersets of B:

To this end, regard the complement $B^\complement := X \setminus B$ of a Borel set B. By definition, for every $\varepsilon > 0$ there exists a compact subset $K_\varepsilon \subseteq B^\complement$ such that

$$\mu\big(B^\complement \setminus K_\varepsilon\big) = \mu\big((B \cup K_\varepsilon)^\complement\big) = \mu\big(K_\varepsilon^\complement \setminus B\big) < \varepsilon,$$

where $K_\varepsilon^\complement$ is an open superset of B.

This property of μ extends to (possibly non-finite) Borel measures via

$$\mu(B) = \sup\{\mu(K) : K \subseteq B,\ K \text{ compact}\} \qquad \forall B \in \mathcal{B}(X). \tag{A.10.a}$$

Likewise, one can reformulate (A.8) for (possibly non-finite) Borel measures:

$$\mu(B) = \inf\{\mu(O) : O \in \mathcal{T},\ O \subseteq B\} \qquad \forall B \in \mathcal{B}(X). \tag{A.10.b}$$

A Borel measure μ which satisfies (A.10.a) is called *inner regular*. If μ satisfies (A.10.b), it is called *outer regular*. A measure which is both inner and outer regular is simply called *regular*.

By definition, finite Radon measures are regular. Hence, one can reformulate the theorems presented above for regular Borel measures with finite variation.

A.4 Weak Convergence of Measures

In this section we consider another important σ-algebra on topological spaces, different from the aforementioned Borel σ-algebra.

A.4.1 Definition The ***Baire σ-algebra*** $\mathcal{Ba}(X)$ on a topological space $(X, \mathcal{T})$ is the smallest σ-algebra on X with respect to which all functions $f \in C_b(X)$ are measurable. The elements of $\mathcal{Ba}(X)$ are called the ***Baire sets*** in X.

A countably additive signed measure on the Baire σ-algebra $\mathcal{Ba}(X)$ is called a ***Baire measure*** on X.

Since $\mathcal{Ba}(X)$ is generated by all sets in $\big\{f^{-1}(O) : f \in C_b(X),\ O \subseteq \mathbb{R} \text{ is open}\big\} \subseteq \mathcal{T}$, we clearly have $\mathcal{Ba}(X) \subseteq \mathcal{B}(X)$. In particular, $\mathcal{Ba}(X)$ coincides with $\mathcal{B}(X)$ if X is a metric space. Hence, a measure μ on a metric space is Baire, precisely when μ is Borel.

Recall that the *weak-* topology* on the continuous dual space $C_b'(X)$ is the *weak topology* $\sigma\big(C_b'(X), C_b(X)\big)$, i.e. the initial topology on $C_b'(X)$ with respect to $\iota(C_b(X))$, where $\iota : C_b(X) \to C_b^{**}(X)$ is the canonical embedding of $C_b(X)$ into the bidual space $C_b^{**}(X)$.

Denote by $\mathcal{M}_\sigma(X)$ the linear space of all Baire measures on $(X, \mathcal{T})$ with finite variation. Then we can identify $\mathcal{M}_\sigma(X)$ with a subspace of $C_b'(X)$ by means of formula (A.9) for all $\mu \in \mathcal{Ba}(X)$. Hence, we may consider convergence of Borel measures on X with respect to $\sigma\big(C_b'(X), C_b(X)\big)$

A.4.2 Definition A net $(\mu_i)_{i\in I}$ in $\mathcal{M}_\sigma(X)$ is called ***weakly convergent*** to a measure $\mu \in \mathcal{M}_\sigma(X)$ if $(\mu_i)_{i\in I}$ converges to μ with respect to the weak-* topology $\sigma\big(C_b'(X), C_b(X)\big)$. In this case we write $\mu_i \xrightarrow{w^*} \mu$.

A net $(\mu_i)_{i\in I}$ of Borel measures on X is said to be weakly convergent if the restriction $\mu_i|_{\mathcal{Ba}(X)}$ to the Baire σ-algebra $\mathcal{Ba}(X)$ converge weakly in the sense above.

A.4.3 Facts Let $(\mu_i)_{i\in I}$ be a net in $\mathcal{M}_\sigma(X)$.

(i) A more concise characterization of weak convergence can be given as follows: For every $\mu \in \mathcal{M}_\sigma(X)$ we have

$$\mu_i \xrightarrow{w^*} \mu \qquad \text{iff} \qquad \lim_{i\in I} \int_X f \,\mathrm{d}\mu_i = \int_X f \,\mathrm{d}\mu \quad \forall f \in C_b(X)\,.$$

(ii) A stronger notion of convergence of measures is given by the total variation $|\mu|(X)$ of a measure μ on X. Indeed, the space of all signed measures on X with finite variation forms a Banach space with respect to $|\cdot|(X)$.

If $(\mu_i)_{i\in I}$ converges in the total variation to a measure $\mu \in \mathcal{M}_\sigma(X)$, then $(\mu_i)_{i\in I}$ converges also weakly to μ. The following elementary example shows that the converse statement is not true in general.

A.4.4 Example Let $p \in L^1(\mathbb{R}, \mathbb{R})$ be a probability density and define probability measures ν_n with densities $p_n := np(nt)$, $n \in \mathbb{N}$. Then we infer that $(\nu_n)_{\mathbb{N}}$ is weakly convergent to the dirac measure δ_0 by means of *dominated convergence* applied to

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(t)p_n(t)\,\mathrm{d}\lambda(t) = \lim_{n\to\infty} \int_{\mathbb{R}} f(s/t)p(s)\,\mathrm{d}\lambda(s) = f(0) = \int_{\mathbb{R}} f\,\mathrm{d}\delta_0 \qquad \forall f \in C_b(\mathbb{R})\,.$$

On the other hand, we have $|\nu_n - \delta_0|(\mathbb{R}) = 2$ for all $n \in \mathbb{N}$. Hence, $(\nu_n)_{n\in\mathbb{N}}$ does not converge in total variation. ◀

In case, a given function on X is only semi-continuous, one implication of the characterization in Fact A.4.3.i may be somewhat relaxed.

A.4.5 Proposition *Suppose that a net $(\mu_i)_{i\in I}$ of Borel probability measures on a Tychonoff space $(X, \mathcal{T})$ converges weakly to a Radon probability measure μ. Then for every bounded function $f : X \to \mathbb{R}$ the following statements hold:*

(i) *If f is upper semi-continuous, then*

$$\limsup_{i\in I} \int_X f\,\mathrm{d}\mu_i \le \int_X f\,\mathrm{d}\mu.$$

(ii) *If f is lower semi-continuous, then*

$$\liminf_{i\in I} \int_X f\,\mathrm{d}\mu_i \ge \int_X f\,\mathrm{d}\mu.$$

Weak convergence of measures is closely related to uniform tightness of measures.

A.4.6 Definition A family $\mathcal{N}$ of finite Radon measures on a topological space $(X, \mathcal{T})$ is called ***uniformly tight*** if for every $\varepsilon > 0$ there exists a compact set K_ε such that

$$|\mu|(X \setminus K_\varepsilon) < \varepsilon \qquad \forall \mu \in \mathcal{N}.$$

In particular, every finite family $\mathcal{N}$ of Radon measures is uniformly tight since the finite union of compact sets is again compact in X.

The following theorem relates uniform tightness to relative compactness in the space of Radon measures on a complete metric space. Recall that the notions of Radon and Borel measures coincide on Polish spaces.

A.4.7 Theorem (Prokhorov) *Let X be a complete metric space and let $\mathcal{N}$ be a family of Radon measures on X. Then the following conditions are equivalent:*

(i) *Every sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{N}$ contains a weakly convergent subsequence.*

(ii) *The family $\mathcal{N}$ is uniformly tight and uniformly bounded in total variation norm $|\cdot|(X)$.*

Appendix B

This appendix is devoted to a brief introduction of *vector flows*. For simplicity we will eschew the notion of flows on smooth manifolds and confine ourselves to the basic theory for Euclidean space $\mathbb{R}^n$, right after we have established the fundamental definitions for general topological spaces.

To this end, we will follow Alongi and Nelson [1] where the theory of flows in the context of manifolds is treated as well.

B.1 Flows

B.1.1 Definition A ***local flow*** in a topological space X is a continuous function $\varkappa : (a,b)\times X \to X$ such that

(i) $0 \in (a,b)$ and $\varkappa(0,x) = x$ for all $x \in X$,
(ii) $\varkappa(t+s,x) = \varkappa(s,\varkappa(t,x))$ for all $x \in X$ and $s,t \in (a,b)$ such that $s+t \in (a,b)$.

In the case that (a,b) is the real line, $\varkappa$ is called a ***complete flow*** on X.

The topological space x is called ***phase space*** of the flow $\varkappa$.

Notation

It is custom to denote the flow by $\varkappa^t(x)$ instead of $\varkappa(t,x)$. As a result, (i) can be expressed as $\varkappa^0$ being the identity on x and (ii) becomes $\varkappa^{s+t} = \varkappa^s \circ \varkappa^t$.

B.1.2 Facts Let $\varkappa^t$ be a complete flow on a topological space X.

(i) Recall that a *(left) group action* of a group G on a set M is a mapping $G\times M \to M$, $(g,m) \mapsto g\cdot m$ such that

$$g_1 \cdot (g_2 \cdot m) = (g_1 g_2)\cdot m \quad \text{and} \quad e\cdot m = m \qquad \forall g_1, g_2 \in G,\ \forall m \in M,$$

where e denotes the identity element of G.

Hence, the flow $(t,x) \mapsto \varkappa^t(x)$ is a continuous group action of the topological group $(\mathbb{R},+)$ on X.

(ii) From (ii) in Definition B.1.1 follows that for each $t \in \mathbb{R}$ the flow $\varkappa^t$ is invertible with continuous inverse $(\varkappa^t)^{-1} = \varkappa^{-t}$. Thus, $\varkappa^t$ is a homeomorphism on X for every $t \in \mathbb{R}$.

In particular, the family $\{\varkappa^t : t \in \mathbb{R}\}$ forms a continuous group of continuous maps.

The notion of flows arises naturally in the study of ordinary differential equations, as the following example shows.

B.1.3 Example Let $F : \mathbb{R}^n :\to \mathbb{R}^n$ be a Lipschitz continuous vector field. By the well-known *Picard-Lindelöf theorem*, there exists a unique curve $v \in C^1\big((-\varepsilon,\varepsilon),\mathbb{R}^n\big)$, $\varepsilon > 0$, which is the solution of the *initial value problem*

$$\begin{aligned} v(0) &= x_0 \in \mathbb{R}^n, \\ v'(t) &= F\big(v(t)\big) \qquad \forall t \in (-\varepsilon,\varepsilon). \end{aligned}$$

Then $\varkappa^t(x_0) := v(t)$ is a *local flow of the vector field* F. In this case, the integral curves $t \mapsto \varkappa^t(x_0)$ are called *flow curves* of F. ◀

The image of a flow curve $t \mapsto \varkappa^t(x_0)$ leads to a natural generalization for flows on topological spaces:

B.1.4 Definition If $\varkappa^t$ is a complete flow on a topological space X and x a point of X, then the ***orbit*** of x with respect to $\varkappa^t$ is the set

$$\mathbb{R} \cdot x := \left\{ \varkappa^t(x) : t \in \mathbb{R} \right\};$$

the ***forward orbit*** of x with respect to $\varkappa^t$ is the set

$$\mathbb{R}_0^+ \cdot x := \left\{ \varkappa^t(x) : t \in \mathbb{R}_0^+ \right\};$$

the ***backward orbit*** of x with respect to $\varkappa^t$ is the set

$$\mathbb{R}_0^- \cdot x := \left\{ \varkappa^t(x) : t \in \mathbb{R}_0^- \right\}.$$

Obviously, the orbit, as well as the forward orbit and backward orbit, of a point with respect to a flow are path-connected spaces.
Moreover, two orbits coincide when they have non-empty intersection.

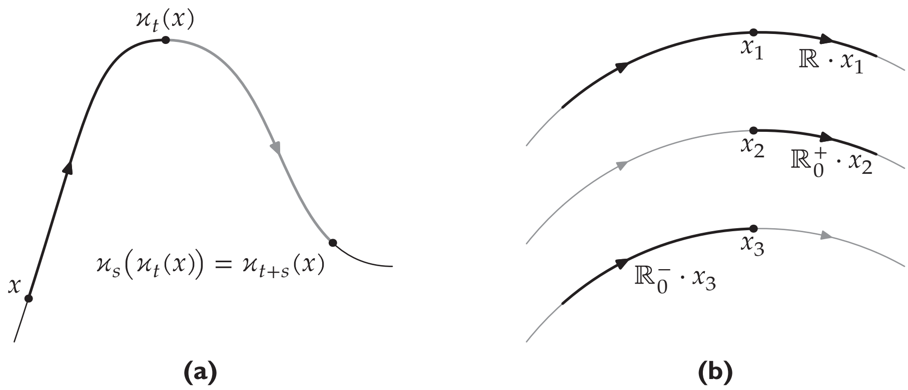


Figure B.1 (a) depicts the group property (ii) of Definition B.1.1; (b) illustrates the orbit, as well as the forward and backward orbit of a complete flow.

B.2 Gradient flows in Euclidean space

Consider a moving particle in Euclidean space $\mathbb{R}^n$ with smooth position function $v : \mathbb{R} \to \mathbb{R}^n$ with respect to time and consider the potential energy $\phi : \mathbb{R}^n \to \mathbb{R}$ of the particle as a smooth function of the particle's position.
Assuming that the particle moves to decrease its potential energy most rapidly, the ordinary differential equation

$$v'(t) = -\nabla\phi\big(v(t)\big)$$

models the particle's motion accordingly.
This gives rise to the following definition:

B.2.1 Definition A vector field $F : \mathbb{R}^n \to \mathbb{R}^n$ is called a ***gradient vector field*** if there exists a differentiable function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that

$$F = -\nabla\phi.$$

A complete flow $\varkappa^t$ is a ***gradient flow*** if $t \mapsto \varkappa^t(x)$ is differentiable for each $x \in \mathbb{R}^n$ and there exits a gradient vector field $F = -\nabla\phi$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\varkappa^t(x) = F\left(\varkappa^t(x)\right) = -\nabla\phi\left(\varkappa^t(x)\right).$$

In this case, ϕ is called ***potential*** for the vector field F or the flow $\varkappa^t$.

B.2.2 Example

⇝ Consider a complete flow $\varkappa^t$ on $\mathbb{R}^n$, defined by

$$\varkappa^t(x) := e^{-t}x,$$

and the function

$$\begin{aligned} \phi : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2}\|x\|_2^2 + c \end{aligned}$$

for an arbitrary $c \in \mathbb{R}$. Then $\varkappa^t$ is a gradient flow and ϕ is a potential for $\varkappa^t$.

⇝ On the other hand, define a smooth vector field $F : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\frac{\partial F_i}{\partial x^j} \neq \frac{\partial F_j}{\partial x^i}$$

for some indices $1 \leq i < j \leq n$. Then there exists no potential for F and $\varkappa^t$ cannot be expressed as a gradient flow. ◀

References

[1] J. M. Alongi and G. S. Nelson, *Recurrence and Topology*, volume 85. (American Mathematical Soc., 2007).

[2] L. Ambrosio and N. Gigli, A user's guide to optimal transport, in *Modelling and Optimisation of Flows on Networks* (Springer), p. 1–155.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*. (Springer, 2006).

[4] L. Ambrosio and G. Savaré, *Gradient flows of probability measures*, Handbook of differential equations: evolutionary equations **3** (2007), 1–136

[5] V. Barbu, *Nonlinear semigroups and differential equations in Banach spaces*. (Editura Academiei Bucharest-Noordhoff, Leyden, 1976).

[6] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. (Springer, 2011).

[7] M. Beiglböck, C. Léonard, and W. Schachermayer, *A general duality theorem for the monge–kantorovich transport problem*, arXiv preprint arXiv:0911.4347 (2009)

[8] M. Beiglböck, C. Léonard, and W. Schachermayer, *On the duality theory for the monge-kantorovich transport problem*, arXiv preprint arXiv:1010.5403 (2010)

[9] M. Beiglböck, C. Léonard, and W. Schachermayer, *A generalized dual maximizer for the monge–kantorovich transport problem*, ESAIM: Probability and Statistics **16** (2012), 306–323

[10] M. Beiglböck and W. Schachermayer, *Duality for borel measurable cost functions*, Transactions of the American Mathematical Society **363** (2011), no. 8, 4203–4224

[11] P. Bénilan, *Équations d'évolution dans un espace de Banach quelconque et applications*. (Paris XI -Orsay., 1972).

[12] V. Bogachev, *Measure Theory*, volume Vol. 1–2 of *Measure Theory*. (Springer, 2006).

[13] V. I. Bogachev and A. V. Kolesnikov, *The monge-kantorovich problem: achievements, connections, and perspectives*, Russian Mathematical Surveys **67** (2012), no. 5, 785

[14] F. Bolley, Separability and completeness for the wasserstein distance, in *Séminaire de probabilités XLI* (Springer), p. 371–377.

[15] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on pure and applied mathematics **44** (1991), no. 4, 375–417

[16] H. Brezis, *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. (Elsevier, 1973).

[17] H. Brezis, M. Crandall, and A. Pazy, *Perturbations of nonlinear maximal monotone sets in banach space*, Communications on Pure and Applied Mathematics **23** (1970), no. 1, 123–144

[18] A. Brøndsted and R. Rockafellar, *On the subdifferentiability of convex functions*, Proceedings of the American Mathematical Society **16** (1965), no. 4, 605–611

[19] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, volume 33. (American Mathematical Society Providence, 2001).

[20] P. Clément and W. Desch, *Some remarks on the equivalence between metric formulations of gradient flows*, Bollettino dell unione matematica italiana **3** (2010a), no. 3, 583–588

[21] P. Clément and W. Desch, *A "crandall–liggett" approach to "gradient flows" in metric spaces*, J. Abstr. Differ. Equ. Appl **1** (2010b), no. 1, 46–60

[22] D. L. Cohn, *Measure theory*. (Springer, 2013).

[23] K. Craig, *The exponential formula for the wasserstein metric*, arXiv preprint arXiv:1310.2912 (2013)

[24] M. G. Crandall and T. M. Liggett, *Generation of semi-groups of nonlinear transformations on general banach spaces*, American Journal of Mathematics (1971), 265–298

[25] M. G. Crandall and A. Pazy, *Semi-groups of nonlinear contractions and dissipative sets*, Journal of functional analysis **3** (1969), no. 3, 376–418

[26] S. Daneri and G. Savaré, *Lecture notes on gradient flows and optimal transport*, arXiv preprint arXiv:1009.3737 (2010)

[27] E. De Giorgi, *New problems on minimizing movements*, Ennio de Giorgi: Selected Papers (1993), 699–713

[28] D. A. Edwards, *A simple proof in monge–kantorovich duality theory*, Studia Math **200** (2010), no. 1, 67–77

[29] D. A. Edwards, *On the kantorovich–rubinstein theorem*, Expositiones Mathematicae **29** (2011), no. 4, 387–398
[30] P. M. Gruber, *Convex and discrete geometry*, volume 336. (Springer, 2007).
[31] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the fokker–planck equation*, SIAM journal on mathematical analysis **29** (1998), no. 1, 1–17
[32] L. V. Kantorovich *On the translocation of masses*, in *Dokl. Akad. Nauk SSSR* (1942), pp. 199–201.
[33] J. Karlsson, *Lebesgue points, hölder continuity and sobolev functions* (2009)
[34] B. Kirchheim, *Rectifiable metric spaces: local structure and regularity of the hausdorff measure*, Proceedings of the American Mathematical Society **121** (1994), no. 1, 113–123
[35] Y. Komura and others, *Nonlinear semi-groups in hilbert space*, Journal of the Mathematical Society of Japan **19** (1967), no. 4, 493–507
[36] A. J. Kurdila and M. Zabarankin, *Convex functional analysis*. (Springer, 2006).
[37] G. Leoni, *A first course in Sobolev spaces*, volume 105. (American Mathematical Soc., 2009).
[38] J. L. Lions, *Quelques méthodes de résolution des problemes aux limites non linéaires* (1969)
[39] U. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Communications in Analysis and Geometry **6** (1998), 199–254
[40] R. J. McCann and others, *Existence and uniqueness of monotone measure-preserving maps*, Duke Mathematical Journal **80** (1995), no. 2, 309–324
[41] R. E. Megginson, *An introduction to Banach space theory*, volume 183. (Springer, 1998).
[42] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la Société mathématique de France **93** (1965), 273–299
[43] H. Pajot, Y. Ollivier, and C. Villani, *Optimal Transportation: Theory and Applications*, volume 413. (Cambridge University Press, 2014).
[44] A. Papadopoulos, *Metric spaces, convexity and nonpositive curvature*. (2014).
[45] L. N. Vaserstein, *Markov processes over denumerable products of spaces, describing large systems of automata*, Problemy Peredachi Informatsii **5** (1969), no. 3, 64–72
[46] C. Villani, *Topics in optimal transportation*, volume 58. (American Mathematical Soc., 2003).
[47] C. Villani, *Optimal transport: old and new*, volume 338. (Springer, 2008).

Symbol Glossary

The page numbers on the right indicate the first time the symbol is defined or used.

$\mathbb{N}$	Natural numbers $1, 2, 3, 4, \ldots$	
$\mathbb{Z}$	Integers	
$\mathbb{Q}$	Rational numbers	
$\mathbb{R}$	Real numbers	
$\mathbb{C}$	Complex numbers	
$s \wedge t := \inf\{s, t\}$		
$s \vee t := \sup\{s, t\}$		
$\lvert\mu\rvert(X)$	Total variation of the measure μ on X	
$\mu \ll \nu$	Absolute continuity of the measure μ with respect to ν	
$\mu_i \xrightarrow{w^*} \mu$	Weak convergence of $(\mu_i)_{i\in I}$ to μ	66
$f_*\mu := \mu \circ f^{-1}$	Pushforward of a measure μ with respect to f	35
$AC((a,b),X)$	Space of absolutely continuous curves	7
$AC^p((a,b),X)$		7
$AC_{\mathrm{loc}}((a,b),X)$	Space of locally absolutely continuous functions	7
$AC^p_{\mathrm{loc}}((a,b),X)$		7
$\mathcal{A}_1 \otimes \mathcal{A}_2$	Tensor-product σ-algebra	35
$B(X,Y)$	Space of bounded linear operators from X to Y	
$B_r(x) := \{y \in X : d(x,y) \leq r\}$	Closed ball with radius r centered at a point x in X	
λ	Lebesgue measure	
λ_n	n-dimensional Lebesgue measure	
$\mathcal{B}_{\mathrm{b}}(X)$	Space of bounded Borel measurable functions on X	38
βX	Stone-Čech compactification of $(X, \mathcal{T})$	64
$\mathcal{B}(X)$	Borel σ-algebra on $(X, \mathcal{T})$	63
$\mathcal{B}a(X)$	Baire σ-algebra on $(X, \mathcal{T})$	65
$C_0(X)$	Space of continuous functions on $(X, \mathcal{T})$ which vanish at infinity	
$C_{\mathrm{b}}(X)$	Space of bounded continuous functions on $(X, \mathcal{T})$	
$C_{\mathrm{c}}^\infty((a,b))$	Space of infinitely differentiable functions with compact support in (a,b)	
$C(X)$	Space of continuous real-valued functions on X	
$\varphi^\star(x) := \sup_{y\in X}\{\langle x,y\rangle - \varphi(y)\}$	Convex conjugate of a functional φ on X	39
$\frac{\mathrm{d}^-}{\mathrm{d}t}f(t) := \limsup\limits_{h\searrow 0} \frac{f(t)-f(t-h)}{h}$	upper left Dini derivative of f at t	26
$\frac{\mathrm{d}^+}{\mathrm{d}t}f(t) := \limsup\limits_{h\searrow 0} \frac{f(t+h)-f(t)}{h}$	upper right Dini derivative of f at t	26
δ_x	Dirac measure with mass at the point x	
$I^{x,y}_{(a,b]}$	Linear interpolant between two points $x, y \in \mathbb{R}$	52
$I^n_\tau := \big((n-1)\tau, n\tau\big]$, $n \in \mathbb{N}$		32
$Lip_L(X)$	Space of Lipschitz continuous functions with Lipschitz constant at most L on X	
$L^p_{\mathrm{loc}}((a,b),X)$	Space of locally p-integrable functions	
$\lvert v'\rvert$	Metric differential of a curve $v : (a,b) \to X$	9
$\overline{M}_\tau$	Discrete solution	32
M^0_τ	Discrete initial datum	32
M^n_τ	τ-discrete minimizing movement	32
$\mathcal{M}_\sigma(X)$	Space of all finite Baire measures on $(X, \mathcal{T})$	65
$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$	Radon–Nikodym derivative of ν with respect to μ	
$O(n)$	Big O notation	

$(f)^{-} := -\min\{f, 0\}$	Negative part of f **14**
$(f)^{+} := \max\{f, 0\}$	Positive part of f **14**
Φ_h	**37**
$\Pi(\mu_1, \mu_2)$	Class of admissible plans **35**
$\mathscr{P}_r^p(X)$	**39**
Ψ_h	**38**
P_τ	Uniform partition of $(0, +\infty)$ with size $\tau > 0$ **32**
$\lvert\partial\phi\rvert$	local slope of ϕ **14**
$\mathfrak{l}_\phi$	global slope of ϕ **14**
$\partial F(x)$	Subdifferential of F at x **23**
$\sigma(X, Y)$	Weak topology on X with respect to $Y \subseteq X'$
$\mathrm{T}(\mu_1, \mu_2)$	Class of admissible transport maps **38**
$\mathcal{T}_Y := \{O \cap Y : O \in \mathcal{T}\}$	Subspace topology on $Y \subseteq X$ with respect to $(X, \mathcal{T})$
$\boldsymbol{t}_\mu^\nu$	Optimal transport map from μ to ν **50**
$\mathfrak{U}(x)$	Neighborhood filter at point x of a topological space $(X, \mathcal{T})$
$W^{(1,1)}((a, b))$	Sobolev space **62**
$W_p(\mu, \nu)$	Wasserstein/Kantorovich distance of order p **39**
X^*	Algebraic dual space of X
X'	Topological dual space of X
X^{**}	Double dual space of X

Index