Chapter 7

The Wasserstein Distance and its Behaviour along Geodesics

In this chapter we will introduce the p-th Wasserstein distance $W_p(\mu,\nu)$ between two measures $\mu,\nu \in \mathscr{P}_p(X)$. The first section is devoted to its preliminary properties, in connection with the optimal transportation problems studied in the previous chapter and with narrow convergence: the main topological results are valid in general metric spaces.

In the last two sections we will focus our attention to the case when X is an Hilbert space: we will characterize the (minimal, constant speed) geodesics with respect to the Wasserstein distance and, for $p = 2$ and a given $\nu \in \mathscr{P}_2(X)$, we will study the behaviour of the map $\mu \mapsto W_2^2(\mu,\nu)$ along geodesics: in particular, we will give a precise formula for its derivative along geodesics and and we will prove its semi-concavity, an important geometric property which is related to a metric version of suitable curvature inequalities.

7.1 The Wasserstein distance

Let X be a separable metric space satisfying the Radon property (5.1.9) and $p \geq 1$. The (p-th) *Wasserstein distance* between two probability measures μ^1, $\mu^2 \in \mathscr{P}_p(X)$ is defined by

$$\begin{aligned} W_p^p(\mu^1,\mu^2) &:= \min\Big\{ \int_{X^2} d(x_1,x_2)^p \, d\boldsymbol{\mu}(x_1,x_2) : \ \boldsymbol{\mu} \in \Gamma(\mu^1,\mu^2)\Big\} \\ &= \min\Big\{ \boldsymbol{d}(x_1,x_2)^p_{L^p(\boldsymbol{\mu};X)} : \ \boldsymbol{\mu} \in \Gamma(\mu^1,\mu^2)\Big\}. \end{aligned} \tag{7.1.1}$$

Using Remark 5.3.3 we can show that the function defined above is indeed a distance. Indeed, if $\mu^i \in \mathscr{P}_p(X)$ for $i = 1,2,3$, $\boldsymbol{\gamma}^{1\,2}$ is optimal between μ^1 and μ^2 and $\boldsymbol{\gamma}^{2\,3}$ is optimal between μ^2 and μ^3 we can find $\boldsymbol{\gamma} \in \mathscr{P}(X^3)$ such that

$\pi^{1\,2}_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}^{1\,2}$ and $\pi^{2\,3}_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}^{2\,3}$. The plan $\boldsymbol{\gamma}^{1\,3} := \pi^{1\,3}_{\#}\boldsymbol{\gamma}$ belongs to $\Gamma(\mu^1, \mu^3)$ and since

$$W_p(\mu^1, \mu^2) = \boldsymbol{d}(x_1, x_2)_{L^p(\boldsymbol{\gamma};X)}, \quad W_p(\mu^2, \mu^3) = \boldsymbol{d}(x_1, x_2)_{L^p(\boldsymbol{\gamma};X)}$$

and

$$\boldsymbol{d}(x_1, x_3)_{L^p(\boldsymbol{\gamma}^{1\,3};X)} = \boldsymbol{d}(x_1, x_3)_{L^p(\boldsymbol{\gamma};X)},$$

we immediately get $W_p(\mu^1, \mu^3) \leq W_p(\mu^1, \mu^2) + W_p(\mu^2, \mu^3)$ from the standard triangle inequality of the L^p distance.

In the particular case when $p = 1$ and μ and ν have a bounded support we can use the duality formula (6.1.1) and the fact that c-concavity coincides with 1-Lipschitz continuity and $\varphi^c = -\varphi$ for the cost $c(x, y) = d(x, y)$ to obtain

$$W_1(\mu, \nu) = \sup\left\{\int \varphi \, d(\mu - \nu) : \; \varphi : X \to \mathbb{R} \text{ 1-Lipschitz}\right\}. \tag{7.1.2}$$

We denote by $\Gamma_o(\mu^1, \mu^2) \subset \Gamma(\mu^1, \mu^2)$ (which also depends on p, even if we omit to indicate explicitly this dependence) the convex and narrowly compact set of *optimal plans* where the minimum is attained, i.e.

$$\boldsymbol{\gamma} \in \Gamma_o(\mu^1, \mu^2) \quad \Longleftrightarrow \quad \int_{X^2} d(x_1, x_2)^p \, d\boldsymbol{\gamma}(x_1, x_2) = W_p^p(\mu^1, \mu^2). \tag{7.1.3}$$

When $\Gamma_o(\mu^1, \mu^2)$ contains a unique plan $\boldsymbol{\gamma} = (\boldsymbol{i} \times \boldsymbol{r})_{\#}\mu^1$ induced by a transport map $\boldsymbol{r}$ as in (5.2.13), we will also denote $\boldsymbol{r}$ by $\boldsymbol{t}_{\mu^1}^{\mu^2}$; therefore $\boldsymbol{t}_{\mu^1}^{\mu^2}$ is characterized by

$$\boldsymbol{t}_{\mu^1}^{\mu^2} : X \to X, \quad \big(\boldsymbol{t}_{\mu^1}^{\mu^2}\big)_{\#}\mu^1 = \mu^2, \quad \Gamma_o(\mu^1, \mu^2) = \big\{\big(\boldsymbol{i} \times \boldsymbol{t}_{\mu^1}^{\mu^2}\big)_{\#}\mu^1\big\}, \tag{7.1.4}$$

it is the unique (strict) minimizer of the optimal transportation problem in the original Monge's formulation (6.0.1), and satisfies

$$\int_X d\big(x, \boldsymbol{t}_{\mu^1}^{\mu^2}(x)\big)^p \, d\mu^1(x) = W_p^p(\mu^1, \mu^2). \tag{7.1.5}$$

Given $\boldsymbol{\mu}$-measurable maps $\boldsymbol{r}, \boldsymbol{s} : \boldsymbol{X} \to X$, a very useful inequality giving an estimate from above of the Wasserstein distance is

$$W_p(\boldsymbol{r}_{\#}\boldsymbol{\mu}, \boldsymbol{s}_{\#}\boldsymbol{\mu}) \leq \boldsymbol{d}(\boldsymbol{r}, \boldsymbol{s})_{L^p(\boldsymbol{\mu};X)}. \tag{7.1.6}$$

It holds because $\boldsymbol{\gamma} = (\boldsymbol{r}, \boldsymbol{s})_{\#}\boldsymbol{\mu} \in \Gamma(\boldsymbol{r}_{\#}\boldsymbol{\mu}, \boldsymbol{s}_{\#}\boldsymbol{\mu})$ and $\int d(x_1, x_2)^p \, d\boldsymbol{\gamma} = \boldsymbol{d}(\boldsymbol{r}, \boldsymbol{s})^p_{L^p(\boldsymbol{\mu};X)}$.

From Theorem 6.1.4 we derive that $\boldsymbol{\mu}$ is optimal iff its support is $d(\cdot,\cdot)^p$-*monotone* according to Definition 6.1.3, i.e.

$$\sum_{k=1}^N d\big(x_1^k, x_2^k\big)^p \leq \sum_{k=1}^N d\big(x_1^k, x_2^{\sigma(k)}\big)^p \tag{7.1.7}$$

for every choice of $(x_1^k, x_2^k) \in \operatorname{supp}\boldsymbol{\mu}$, $k = 1, \dots, N$, and for every permutation $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$ (actually Theorem 6.1.4 shows only that $\boldsymbol{\mu}$ has to be concentrated on a c-monotone set, but since in this case the cost is continuous the c-monotonicity holds, by a density argument, for the whole support of $\boldsymbol{\mu}$).

Remark 7.1.1. It is not difficult to check that supports of optimal plans satisfy the slightly stronger property

$$\overline{\bigcup_{\boldsymbol{\gamma}\in\Gamma_o(\mu^1,\mu^2)} \operatorname{supp}\boldsymbol{\gamma}} \quad \text{is } d(\cdot,\cdot)^p - \text{monotone.} \tag{7.1.8}$$

For, we take a sequence $(\boldsymbol{\gamma}_n)$ narrowly dense in $\Gamma_o(\mu^1, \mu^2)$ and we consider the new plan $\bar{\boldsymbol{\gamma}} := \sum_n 2^{-n}\boldsymbol{\gamma}_n$. The plan $\bar{\boldsymbol{\gamma}}$ is optimal, too, and its support coincides with (7.1.8).

Remark 7.1.2 (Cyclical monotonicity in the case when X is Hilbert). When $p = 2$ and X is a (pre-)Hilbert space, condition (7.1.7) is equivalent to the classical *cyclical monotonicity* of $\operatorname{supp}\boldsymbol{\mu}$, i.e. for every cyclical choice of points $(x_1^k, x_2^k) \in \operatorname{supp}\boldsymbol{\mu}$, $k = 0, \dots, N$, with $(x_1^0, x_2^0) = (x_1^N, x_2^N)$, we have

$$\sum_{k=1}^{N} \langle x_1^k - x_1^{k-1}, x_2^k\rangle \ge 0. \tag{7.1.9}$$

In particular, if $\boldsymbol{r} = \nabla\phi$ for some convex C^1 function ϕ then $\boldsymbol{r}$ is a 2-optimal transport map for every measure $\mu \in \mathscr{P}_2(X)$ such that $\int |\boldsymbol{r}|^2\, d\mu < +\infty$.

A useful application of the necessary and sufficient optimality conditions is given by the following stability of optimality with respect to narrow convergence.

Proposition 7.1.3 (Stability of optimality and narrow lower semicontinuity). *Let $(\mu_n^1), (\mu_n^2) \subset \mathscr{P}_p(X)$ be two sequences narrowly converging to μ^1, μ^2 respectively, and let $\boldsymbol{\mu}_n \in \Gamma_o(\mu_n^1, \mu_n^2)$ be a sequence of optimal plans with $\int_{X^2} d(x_1, x_2)^p\, d\boldsymbol{\mu}_n$ bounded.*
Then $(\boldsymbol{\mu}_n)$ is narrowly relatively compact in $\mathscr{P}(X^2)$ and any narrow limit point $\boldsymbol{\mu}$ belongs to $\Gamma_o(\mu^1, \mu^2)$, with

$$\begin{aligned} W_p(\mu^1, \mu^2) &= \int_{X^2} d(x_1, x_2)^p\, d\boldsymbol{\mu}(x_1, x_2) \\ &\le \liminf_{n\to\infty} \int_{X^2} d(x_1, x_2)^p\, d\boldsymbol{\mu}_n(x_1, x_2) = \liminf_{n\to\infty} W_p(\mu_n^1, \mu_n^2). \end{aligned} \tag{7.1.10}$$

Proof. The relative compactness of the sequence $(\boldsymbol{\mu}_n)$ is a consequence of Lemma 5.2.2 and the "lim inf" inequality in (7.1.10) is a direct consequence of (5.1.15), which in particular yields $\int_{X^2} d(x_1, x_2)^p\, d\boldsymbol{\mu} < +\infty$.
Using proposition 5.1.8 it is immediate to check by approximation that the support of $\boldsymbol{\mu}$ is $d(\cdot,\cdot)^p$-monotone. □

When X is a Hilbert space, the Wasserstein distance is lower semicontinuous w.r.t. the weaker narrow convergence in $\mathscr{P}(X_\varpi)$:

Lemma 7.1.4 (Weak narrow lower semicontinuity of W_p in Hilbert spaces). *Let X be a (separable) Hilbert space and let (μ_n^1), $(\mu_n^2) \subset \mathscr{P}_p(X)$ be two weakly tight sequences (according to (5.1.32)) narrowly converging to μ^1, μ^2 in $\mathscr{P}(X_\varpi)$. Then*

$$W_p(\mu^1, \mu^2) \le \liminf_{n\to\infty} W_p(\mu_n^1, \mu_n^2). \tag{7.1.11}$$

Proof. The map $(x_1, x_2) \mapsto |x_1 - x_2|^p$ is weakly l.s.c. in $X \times X$: we simply argue as in the previous proof and we apply Lemma 5.1.12(c). Notice that in this case the first line of (7.1.10) is an inequality "$\le$", since we do not know that the limit plan μ is optimal any more; nevertheless, the inequality is sufficient to obtain (7.1.11). $\square$

Proposition 7.1.5 (Convergence, compactness and completeness). $\mathscr{P}_p(X)$ *endowed with the p-Wasserstein distance is a separable metric space which is complete if X is complete. A set $\mathcal{K} \subset \mathscr{P}_p(X)$ is relatively compact iff it is p-uniformly integrable and tight. In particular, for a given sequence $(\mu_n) \subset \mathscr{P}_p(X)$ we have*

$$\lim_{n\to\infty} W_p(\mu_n, \mu) = 0 \iff \begin{cases} \mu_n \text{ narrowly converge to } \mu, \\ (\mu_n) \text{ has uniformly integrable } p\text{-moments.} \end{cases} \tag{7.1.12}$$

Proof. Let us first prove the completeness of $\mathscr{P}_p(X)$, by assuming that X is complete. It suffices to show that any sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathscr{P}_p(X)$ such that

$$\sum_{n=1}^{\infty} W_p(\mu^n, \mu^{n+1}) < +\infty$$

is converging. We choose $\boldsymbol{\alpha}^{n\,(n+1)} \in \Gamma_o(\mu^n, \mu^{n+1})$ and use Lemma 5.3.4 to find $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$, with $\boldsymbol{X} = X^{\mathbb{N}}$, satisfying (5.3.8). It follows that

$$\sum_{n=1}^{\infty} \boldsymbol{d}(\pi^n, \pi^{n+1})_{L^p(\boldsymbol{\mu}; X)} < +\infty.$$

Therefore, (π^n) is a Cauchy sequence in $L^p(\boldsymbol{\mu}; X)$, which is a complete metric space, and admits a limit map $\pi^\infty \in L^p(\boldsymbol{\mu}; X)$. Setting $\mu_\infty := \pi^\infty_\# \boldsymbol{\mu} \in \mathscr{P}_p(X)$, we easily find

$$\begin{aligned} \limsup_{n\to\infty} W_p(\mu^n, \mu^\infty) &\le \limsup_{n\to\infty} \boldsymbol{d}(\pi^n, \pi^\infty)_{L^p(\boldsymbol{\mu}; X)} \\ &\le \limsup_{n\to\infty} \sum_{j=n}^{\infty} \boldsymbol{d}(\pi^{j+1}, \pi^j)_{L^p(\boldsymbol{\mu}; X)} = 0. \end{aligned}$$

We will prove now the equivalence (7.1.12) (a different argument in locally compact spaces, based on the duality formula (7.1.2), is available for instance in [126]).

First we suppose that $W_p(\mu_n, \mu) \to 0$. Arguing as before, we can choose optimal plans $\boldsymbol{\beta}^{1\,n} \in \Gamma_o(\mu, \mu_n,)$ and use Lemma 5.3.4 (with $\mu_1 := \mu$) to find $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$ satisfying (5.3.8). It follows that

$$\lim_{n\to\infty} \boldsymbol{d}(\pi^n, \pi^1)_{L^p(\boldsymbol{X},\boldsymbol{\mu};X)} = 0,$$

and therefore, for every continuous real function f with p-growth the Vitali dominated convergence theorem gives

$$\begin{aligned}\lim_{n\to\infty} \int_X f(x)\, d\mu_n(x) = \lim_{n\to\infty} \int_{\boldsymbol{X}} f(\pi^n(\boldsymbol{x}))\, d\boldsymbol{\mu}(\boldsymbol{x}) &= \int_{\boldsymbol{X}} f(\pi^1(\boldsymbol{x}))\, d\boldsymbol{\mu}(\boldsymbol{x}) \\ &= \int_X f(x)\, d\mu(x).\end{aligned}$$

By lemma 5.1.7 we obtain the narrow convergence and the uniform p-integrability of the sequence (μ_n).

Conversely, let us suppose that the sequence (μ_n) has uniformly integrable p-moments and it is narrowly converging to μ; in particular, by (5.4.7), the set $\{\mu, \mu_n : n \in \mathbb{N}\}$ is tight. As before, let us choose $\boldsymbol{\alpha}^{1\,n} \in \Gamma_o(\mu, \mu_n)$: it easy to check that the sequence $(\boldsymbol{\alpha}^{1\,n})$ is p-uniformly integrable and tight in $\mathscr{P}(X \times X)$ (see Lemma 5.2.2): a subsequence $k \mapsto n_k$ exists such that $\boldsymbol{\alpha}^{1\,n_k} \to \boldsymbol{\alpha}$ narrowly, with $\boldsymbol{\alpha} \in \Gamma_o(\mu, \mu)$ by Proposition 7.1.3. Applying Lemma 5.1.7 we get

$$\begin{aligned}\lim_{k\to\infty} W_p^p(\mu, \mu_{n_k}) &= \lim_{k\to\infty} \int_{X\times X} |x_1 - x_2|^p\, d\boldsymbol{\alpha}^{1\,n_k}(x_1, x_2) \\ &= \int_{X\times X} |x_1 - x_2|^p\, d\boldsymbol{\alpha}(x_1, x_2) = 0.\end{aligned}$$

Since the limit is independent of the subsequence n_k we get the convergence of μ_n with respect to the Wasserstein distance. Using (7.1.12) it is now immediate to check that convex combinations of Dirac masses with centers in a countable dense subset of X and with rational coefficients are dense in $\mathscr{P}_p(X)$, therefore $\mathscr{P}_p(X)$ is separable. $\square$

It is interesting to note that in the previous proof of the equivalence between narrow and Wasserstein topology (on sets with uniformly integrable p-moments), one implication (the topology induced by the Wasserstein distance is stronger than the narrow one) could be directly deduced from (7.1.2) via the approximation arguments discussed in Section 5.1, thus avoiding Lemma 5.3.4; this implication is therefore considerably easier than the converse one, which relies on the stability property 7.1.3 and therefore on the main characterization results of Chapter 6 for optimal transportation problems. However the argument via Lemma 5.3.2 seems to be necessary to get completeness, at least in infinite dimensions.

Remark 7.1.6 (Limit of the optimal plan). As a byproduct of the previous proof, we obtain that if $\mu_n \to \mu$ in $\mathscr{P}_p(X)$ and $\boldsymbol{\mu}_n \in \Gamma_o(\mu, \mu_n)$, then

$$\boldsymbol{\mu}_n \to (\boldsymbol{i} \times \boldsymbol{i})_{\#}\mu \quad \text{in } \mathscr{P}_p(X \times X). \tag{7.1.13}$$

Remark 7.1.7 ($\mathscr{P}(X)$ is a Polish space if X is Polish). By taking an equivalent *bounded* metric on X, all the Wasserstein distances induce the topology of narrow convergence between probability measures: as we already noticed in Remark 5.1.1, the narrow topology $\mathscr{P}(X)$ is metrizable; moreover, if X is a Polish space, then $\mathscr{P}(X)$ is a Polish space, too.

Remark 7.1.8 (Relative compactness of $\mathscr{P}_p(X)$-bounded sets). When X is infinite dimensional Hilbert space, bounded subset in $\mathscr{P}_p(X)$ are not relatively compact in $\mathscr{P}(X)$ any more, but they are relatively compact in $\mathscr{P}(X_\varpi)$.

Remark 7.1.9 ($\mathscr{P}_p(X)$ is locally compact only if X is compact). If X is not compact, the space $\mathscr{P}_p(X)$ is not locally compact, not even in the case when $X = \mathbb{R}^d$ is finite dimensional. Indeed, assume that for some $\epsilon > 0$ and $x_0 \in X$ the closed ball in $\mathscr{P}_p(X)$

$$\mathcal{B}_\varepsilon := \Big\{\mu \in \mathscr{P}_p(X) : W_p(\mu, \delta_{x_0}) \le \varepsilon\Big\} = \Big\{\mu \in \mathscr{P}_p(X) : \int_X d(x, x_0)^p \, d\mu(x) \le \varepsilon^p\Big\}$$

is compact and let us prove that an arbitrary sequence $(x_n) \in X$ admits a convergent subsequence. It is not restrictive to assume $\liminf_{n\to\infty} d(x_n, x_0) > 0$ (otherwise (x_n) admits a subsequence converging to x_0), and therefore $\inf_{n\in\mathbb{N}} d(x_n, x_0) = \delta > 0$. We consider the real numbers

$$m_n = \frac{(\delta \wedge \varepsilon)^p}{d(x_n, x_0)^p} \le 1, \quad \text{so that} \quad m_n d(x_n, x_0)^p = (\delta \wedge \varepsilon)^p;$$

the sequence of measures $\mu_n := (1 - m_n)\delta_{x_0} + m_n \delta_{x_n}$ belongs to $\mathcal{B}_\varepsilon$ since $W_p(\mu_n, \delta_{x_0}) = \varepsilon \wedge \delta$ and therefore admits a subsequence $(\mu_{n'})$ converging to some $\mu \neq \delta_{x_0}$ in $\mathscr{P}_p(X)$.
Since (m_n) is bounded, too, it is not restrictive to assume that $m_{n'} \to m \in [0, 1]$ which should be strictly positive, being $\mu \neq \delta_{x_0}$. By Proposition 5.1.8 (see also Corollary 5.1.9) it follows that μ takes the form $(1-m)\delta_{x_0} + m\delta_x$ for some $x \in X$, and therefore $x_{n'} \to x$.

Lemma 7.1.10 (Approximation by convolution). *Let $\mu \in \mathscr{P}_p(\mathbb{R}^d)$ and let $(\rho_\varepsilon) \subset C^\infty(\mathbb{R}^d)$ be a family of nonnegative mollifiers such that*

$$\rho_\varepsilon(x) := \varepsilon^{-d}\rho(x/\varepsilon), \ \int_{\mathbb{R}^d} \rho(x)\, dx = 1, \ \mathsf{m}_p^p(\rho) := \int_{\mathbb{R}^d} |x|^p \rho(x)\, dx < +\infty. \quad (7.1.14)$$

Then if $\mu_\varepsilon := \mu * \rho_\varepsilon$

$$W_p(\mu, \mu_\varepsilon) \le \varepsilon \mathsf{m}_p(\rho), \quad (7.1.15)$$

and therefore μ_ε converges to μ in $\mathscr{P}_p(\mathbb{R}^d)$ as $\varepsilon \downarrow 0$.

Proof. We introduce the family of plans $\boldsymbol{\gamma}_\varepsilon := \int \rho_\varepsilon(\cdot - x)\mathscr{L}^d \, d\mu(x)$ defined by

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x, y)\, d\boldsymbol{\gamma}_\varepsilon(x, y) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \phi(x, y)\rho_\varepsilon(y - x)\, dy\, d\mu(x)$$

which obviously satisfy $\boldsymbol{\gamma}_\varepsilon \in \Gamma(\mu, \mu_\varepsilon)$. Therefore

$$W_p^p(\mu, \mu_\varepsilon) \le \iint_{(\mathbb{R}^d)^2} |x-y|^p \, d\boldsymbol{\gamma}_\varepsilon(x,y) = \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |x-y|^p \rho_\varepsilon(y-x)\, dy \Big) \, d\mu(x)$$
$$= \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |z|^p \rho_\varepsilon(z)\, dz \Big) \, d\mu(x) = \int_{\mathbb{R}^d} |\varepsilon z|^p \rho(z)\, dz = \varepsilon^p \int_{\mathbb{R}^d} |z|^p \rho(z)\, dz \qquad \square$$

Remark 7.1.11. Combining Proposition 5.1.13 with $j(r) := r^p$, $1 < p < +\infty$, and Lemma 5.1.7 we get the following useful characterization of the convergence in $\mathscr{P}_p(X)$, which is particularly interesting when X is infinite dimensional Hilbert space:

$$\lim_{n\to\infty} W_p(\mu_n, \mu) = 0 \iff \begin{cases} \mu_n \text{ narrowly converge to } \mu \text{ in } \mathscr{P}(X_\varpi), \\ \displaystyle\lim_{n\to\infty} \int_X |x|^p \, d\mu_n(x) = \int_X |x|^p \, d\mu(x). \end{cases} \tag{7.1.16}$$

Since we have at our disposal new powerful results (which are consequences of the theory presented in Chapter 6) we conclude this section by showing a simpler proof of (7.1.16), which could be extended to the case of uniformly convex Banach spaces.

Proof. Let us consider the (Radon, separable) metric space X_ϖ with the distance induced by the norm $\|\cdot\|_\varpi$; since $\|\cdot\|_\varpi^p \le |\cdot|^p$, (7.1.16) and Lemma 5.1.7 show that $\|\cdot\|_\varpi^p$ is uniformly integrable w.r.t. the sequence (μ_n). Applying (7.1.12) of Proposition 7.1.5 in X_ϖ (this characterization does not require the completeness of the metric space), we obtain that μ_n converges to μ in the p-Wasserstein distance of $\mathscr{P}_p(X_\varpi)$. It follows by Remark 7.1.6 that any sequence of plans $\boldsymbol{\mu}_n \in \Gamma(\mu_n, \mu)$, optimal in $\mathscr{P}_p(X_\varpi)$, satisfies

$$\boldsymbol{\mu}_n \to (\boldsymbol{i} \times \boldsymbol{i})_{\#}\mu \quad \text{in } \mathscr{P}_p(X_\varpi \times X_\varpi) \quad \text{as } n \to \infty. \tag{7.1.17}$$

We suppose $p \ge 2$ and we integrate with respect to $\boldsymbol{\mu}_n$ the inequality (c_p is a strictly positive constant, $j_p(x_1) = |x_1|^{p-2}x_1$)

$$c_p|x_1 - x_2|^p \le \frac{1}{p}|x_2|^p - \frac{1}{p}|x_1|^p - \langle j_p(x_1), x_2 - x_1\rangle \quad \forall\, x_1, x_2 \in X,$$

which we will prove in Lemma 10.2.1; we obtain

$$c_p W_p^p(\mu, \mu_n) \le \int_{X\times X} c_p |x_1 - x_2|^p \, d\boldsymbol{\mu}_n(x_1, x_2) \tag{7.1.18a}$$
$$\le \int_{X\times X} \Big(\frac{1}{p}|x_2|^p - \frac{1}{p}|x_1|^p - \langle j_p(x_1), x_2 - x_1\rangle \Big) \, d\boldsymbol{\mu}_n(x_1, x_2)$$
$$= \frac{1}{p}\int_X |x_2|^p \, d\mu_n(x_2) - \frac{1}{p}\int_X |x_1|^p \, d\mu(x_1) - \int_{X\times X} \langle y_1, y_2\rangle \, d\hat{\boldsymbol{\mu}}_n(y_1, y_2), \tag{7.1.18b}$$

where

$$\hat{\boldsymbol{\mu}}_n := \big(j_p \circ \pi^1, \pi^2 - \pi^1\big)_\# \boldsymbol{\mu}_n.$$

Since the first marginal of $\hat{\boldsymbol{\mu}}_n$ is fixed in $\mathscr{P}_q(X)$, it is easy to check by Lemma 5.2.1 that

$$\hat{\boldsymbol{\mu}}_n \to \big((j_p)_\#\mu\big) \times \delta_0 \quad \text{in } \mathscr{P}(X \times X_\varpi) \quad \text{as } n \to \infty,$$

and that $(\boldsymbol{\mu}_n)$ satisfies the assumptions of Lemma 5.2.4; therefore, passing to the limit as $n \to \infty$ in (7.1.18a,b), the convergence of the moments (7.1.16) and Lemma 5.2.4 yield $W_p(\mu, \mu_n) \to 0$.
The case $p < 2$ follows by the same argument and inequality (10.2.5). □

7.2 Interpolation and geodesics

In this section we are assuming that X is a separable Hilbert space and $p > 1$, and we show that constant speed geodesics in $\mathscr{P}_p(X)$ coincide with a suitable class of interpolations obtained from optimal transport plans. Recall that a curve $\mu_t \in \mathscr{P}_p(X)$, $t \in [0,1]$, is a constant speed geodesic (see also (2.4.3)) if

$$W_p(\mu_s, \mu_t) = (t-s) W_p(\mu_0, \mu_1) \quad \forall\, 0 \le s \le t \le 1. \tag{7.2.1}$$

If $\boldsymbol{\mu} \in \mathscr{P}(X^N)$, $N \ge 2$, $1 \le i, j, k \le N$, and $t \in [0,1]$ we set

$$\pi_t^{i\to j} := (1-t)\pi^i + t\pi^j : X^N \to X, \tag{7.2.2}$$

$$\pi_t^{i\to j,k} := (1-t)\pi^{i,k} + t\pi^{j,k} : X^N \to X^2, \tag{7.2.3}$$

$$\mu_t^{i\to j} := (\pi_t^{i\to j})_\# \boldsymbol{\mu} \in \mathscr{P}(X), \tag{7.2.4}$$

$$\boldsymbol{\mu}_t^{i\to j,k} := (\pi_t^{i\to j,k})_\# \boldsymbol{\mu} \in \mathscr{P}(X^2). \tag{7.2.5}$$

It is well known that $\Gamma_o(\mu^1, \mu^2)$ can contain in general more than one element. In the following lemma we show that along a geodesic the optimal plans to the extreme points μ_0, μ_1 are unique if we consider μ_t, $t \in (0,1)$, as the initial measure.

Lemma 7.2.1. *Let $(\mu_t)_{t\in[0,1]}$ be a constant speed geodesic in $\mathscr{P}_p(X)$ and let $t \in (0,1)$. Then $\Gamma_o(\mu_t, \mu_1)$ (resp. $\Gamma_o(\mu_0, \mu_t)$) contains a unique plan $\boldsymbol{\mu}^{t\,1}$ (resp. $\boldsymbol{\mu}^{0\,t}$) and this plan (resp. $(\boldsymbol{\mu}^{0\,t})^{-1}$) is induced by a transport. Moreover, $\boldsymbol{\mu} = \boldsymbol{\mu}^{t\,1} \circ \boldsymbol{\mu}^{0\,t} \in \Gamma_o(\mu_0, \mu_1)$ and*

$$\boldsymbol{\mu}^{0\,t} = (\pi_t^{1,1\to 2})_\# \boldsymbol{\mu}, \quad \boldsymbol{\mu}^{t\,1} = (\pi_t^{1\to 2,2})_\# \boldsymbol{\mu}. \tag{7.2.6}$$

Proof. For $t \in (0,1)$ let $\boldsymbol{\gamma}$ (resp. $\boldsymbol{\eta}$) be optimal transport plans between μ_0 and μ_t (resp. μ_t and μ_1). In order to clarify the structure of the proof it is convenient to view μ_0, μ_t, μ_1 as measures in $\mathscr{P}(X_1)$, $\mathscr{P}(X_2)$, $\mathscr{P}(X_3)$, where X_i are distinct copies of X. Then, we can define

$$\boldsymbol{\lambda} := \int_{X_2} \gamma_{x_2} \times \eta_{x_2}\, d\mu_t(x_2) \in \Gamma(\mu_0, \mu_t, \mu_1)$$

where $\boldsymbol{\gamma} = \int_{X_2} \gamma_{x_2}\, d\mu_t$ and $\boldsymbol{\eta} = \int_{X_2} \eta_{x_2}\, d\mu_t$ are the disintegrations of $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ with respect to the common variable x_2. Then, since (recall the composition of plans in Remark 5.3.3)

$$\boldsymbol{\mu} = \boldsymbol{\eta} \circ \boldsymbol{\gamma} = \pi^{1,3}_{\#}\boldsymbol{\lambda} \in \Gamma(\mu_0, \mu_1)$$

we get

$$\begin{aligned} W_p(\mu_0, \mu_1) &\leq \|x_1 - x_3\|_{L^p(\boldsymbol{\mu};X)} \leq \|x_1 - x_2\|_{L^p(\boldsymbol{\lambda};X)} + \|x_2 - x_3\|_{L^p(\boldsymbol{\lambda};X)} \\ &= \|x_1 - x_2\|_{L^p(\boldsymbol{\gamma};X)} + \|x_2 - x_3\|_{L^p(\boldsymbol{\eta};X)} = W_p(\mu_0, \mu_1). \end{aligned}$$

This proves that $\boldsymbol{\mu}$ is optimal; moreover, since all inequalities are equalities and the L^p-norm is strictly convex, we get that there exists $\alpha > 0$ such that $x_2 - x_1 = \alpha(x_3 - x_1)$ for $\boldsymbol{\lambda}$-a.e. triple (x_1, x_2, x_3). Using the fact that $W_p(\mu_t, \mu_0) = tW_p(\mu_0, \mu_1)$ we obtain $\alpha = t$ and therefore

$$x_2 - x_1 = t(x_3 - x_1) \quad \boldsymbol{\lambda}\text{-a.e. in } X_1 \times X_2 \times X_3.$$

Denoting by $z(x_2)$ the barycenter of γ_{x_2}, the linearity of this relation w.r.t. x_1 yields

$$x_2 - z(x_2) = t(x_3 - z(x_2)) \quad \boldsymbol{\eta}\text{-a.e. in } X_2 \times X_3.$$

Hence $\boldsymbol{\eta}$ is induced by the transport $\boldsymbol{r}_t(x_2) = x_2/t - z(x_2)(1-t)/t$. Since z depends on $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ have been chosen independently, this proves that $\boldsymbol{\eta}$ is unique, so that $\boldsymbol{\eta} = \boldsymbol{\mu}^{t\,1}$, the measure defined in (7.2.6). Inverting the order of μ_0 and μ_1, we obtain the other identity. □

Theorem 7.2.2 (Characterization of constant speed geodesics). *If $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$ then the curve $t \mapsto \mu_t := \mu_t^{1\to2}$ is a constant speed geodesic connecting μ^1 to μ^2. Conversely, any constant speed geodesic $\mu_t : [0,1] \to \mathscr{P}_p(X)$ connecting μ^1 to μ^2 has this representation for a suitable $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$, which can be constructed from any point μ_t, $0 < t < 1$, as in the previous Lemma.*

Proof. By (7.1.6) we get

$$W_p(\mu_t, \mu_s) \leq (t - s) W_p(\mu^1, \mu^2) \quad \forall s, t \in (0,1),\ s \leq t. \tag{7.2.7}$$

If there is a strict inequality for some $s < t$ we immediately derive a contradiction by applying the triangle inequality with the points μ_0, μ_s, μ_t and μ_1. Therefore equality holds and μ_t is a constant speed geodesic.

Let μ_t be a constant speed geodesic and for a fixed $t \in (0,1)$ let $\boldsymbol{\mu} := \boldsymbol{\mu}^{t\,1} \circ \boldsymbol{\mu}^{0\,t}$ be as in Lemma 7.2.1. Since $\boldsymbol{\mu}^{0\,t} = (\pi_t^{1,1\to2})_\# \boldsymbol{\mu}$ is the unique element of $\Gamma_o(\mu_0, \mu_t)$ and the curve $s \mapsto \mu_{ts}$, $s \in [0,1]$ is a constant speed geodesic, we get

$$\mu_{st} = (\pi_s^{1\to2})_\# \boldsymbol{\mu}^{0\,t} = (\pi_s^{1\to2} \circ \pi_t^{1,1\to2})_\# \boldsymbol{\mu} = (\pi_{st}^{1\to2})_\# \boldsymbol{\mu}.$$

Inverting μ_0 with μ_1 we conclude. □

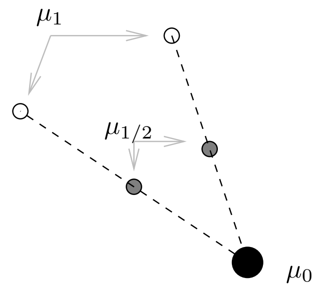

Figure 7.1: An example of geodesic: the mass of μ^0 splits into two parts

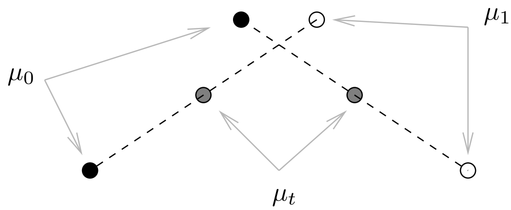

Figure 7.2: Another example of geodesic: the trajectories may intersect

In the case $X = \mathbb{R}$, using the explicit representation (6.0.3) for the Wasserstein distance in terms of the inverses of distribution functions, we get

$$F^{-1}_{\mu_t^{1\to 2}} = (1-t)F^{-1}_{\mu^1} + tF^{-1}_{\mu^2} \quad \mathscr{L}^1\text{-a.e. in } (0,1). \tag{7.2.8}$$

for any geodesic $\mu_t^{1\to 2}$ induced by $\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^2)$.

7.3 The curvature properties of $\mathscr{P}_2(X)$

In this section we consider the particular case $p = 2$ and we establish some finer geometric properties of $\mathscr{P}_2(X)$.

In particular we will prove in Theorem 7.3.2 the *semiconcavity inequality* of the Wasserstein distance from a fixed measure μ^3 along the constant speed geodesics $\mu_t^{1\to 2}$ connecting μ^1 to μ^2:

$$W_2^2(\mu_t^{1\to 2}, \mu^3) \geq (1-t)W_2^2(\mu^1, \mu^3) + tW_2^2(\mu^2, \mu^3) - t(1-t)W_2^2(\mu^1, \mu^2). \tag{7.3.1}$$

According to Aleksandrov's metric notion of curvature (see [5] and Section 12.3 in the Appendix), this inequality can be interpreted by saying that the Wasserstein space is a positively curved metric space (in short, a *PC*-space). This was already pointed out by a formal computation in [107], showing also that generically the inequality is strict (see Example 7.3.3). See also Section 12.3 in the Appendix, where we recall some basic facts of the theory of positively curved metric spaces.

For $\boldsymbol{\mu} \in \Gamma(\mu^1, \mu^2, \mu^3) \subset \mathscr{P}_2(X^3)$ and $i, j, k \in \{1, 2, 3\}$, $t \in [0, 1]$ we set

$$W^2_{\boldsymbol{\mu}}(\mu_t^{i\to j}, \mu^k) := \int_{X^3} |(1-t)x_i + tx_j - x_k|^2 \, d\boldsymbol{\mu}(x_1, x_2, x_3). \tag{7.3.2}$$

By (7.1.6) we get

$$W^2_2(\mu_t^{i\to j}, \mu^k) \le W^2_{\boldsymbol{\mu}}(\mu_t^{i\to j}, \mu^k). \tag{7.3.3}$$

Moreover, the Hilbertian identity

$$|(1-t)a + tb - c|^2 = (1-t)|a-c|^2 + t|b-c|^2 - t(1-t)|b-a|^2$$

gives

$$W^2_{\boldsymbol{\mu}}(\mu_t^{1\to 2}, \mu^3) = (1-t)W^2_{\boldsymbol{\mu}}(\mu^1, \mu^3) + tW^2_{\boldsymbol{\mu}}(\mu^2, \mu^3) - t(1-t)W^2_{\boldsymbol{\mu}}(\mu^1, \mu^2), \tag{7.3.4}$$

and the related differential identities

$$\frac{d}{dt} W^2_{\boldsymbol{\mu}}(\mu_t^{1\to 2}, \mu^3) = W^2_{\boldsymbol{\mu}}(\mu^2, \mu^3) - W^2_{\boldsymbol{\mu}}(\mu^1, \mu^3) + (2t-1)W^2_{\boldsymbol{\mu}}(\mu^1, \mu^2) \tag{7.3.5}$$

$$= \frac{1}{1-t}\Big(W^2_{\boldsymbol{\mu}}(\mu^2, \mu^3) - W^2_{\boldsymbol{\mu}}(\mu_t^{1\to 2}, \mu^2) - W^2_{\boldsymbol{\mu}}(\mu_t^{1\to 2}, \mu^3)\Big) \tag{7.3.6}$$

$$= \frac{1}{t}\Big(W^2_{\boldsymbol{\mu}}(\mu_t^{1\to 2}, \mu^1) + W^2_{\boldsymbol{\mu}}(\mu_t^{1\to 2}, \mu^3) - W^2_{\boldsymbol{\mu}}(\mu^1, \mu^3)\Big). \tag{7.3.7}$$

Proposition 7.3.1. *Let* $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2)$, $t \in (0,1)$ *and* $\boldsymbol{\mu}^{t\,3} \in \Gamma_o(\mu_t^{1\to 2}, \mu^3)$. *Then there exists a plan*

$$\boldsymbol{\mu}_t \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3) \quad \textit{such that} \quad (\pi_t^{1\to 2, 3})_\# \boldsymbol{\mu} = \boldsymbol{\mu}^{t\,3}, \tag{7.3.8}$$

and this plan is unique if $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$. *For each plan* $\boldsymbol{\mu}_t$ *satisfying* (7.3.8) *we have*

$$W^2_2(\mu_t^{1\to 2}, \mu^3) = (1-t)W^2_{\boldsymbol{\mu}_t}(\mu^1, \mu^3) + tW^2_{\boldsymbol{\mu}_t}(\mu^2, \mu^3) - t(1-t)W^2_{\boldsymbol{\mu}_t}(\mu^1, \mu^2). \tag{7.3.9}$$

Proof. Let $\Sigma_t : X^2 \to X^2$ and $\Lambda_t : X^3 \to X^3$ be the homeomorphisms defined by

$$\Sigma_t(x_1, x_2) := ((1-t)x_1 + tx_2, x_2), \quad \Lambda_t(x_1, x_2, x_3) = ((1-t)x_1 + tx_2, x_2, x_3)$$

and notice that $\boldsymbol{\mu}$ has the required properties if and only if $\boldsymbol{\nu} := \Lambda_{t\#}\boldsymbol{\mu}$ satisfies

$$\pi^{1,2}_\# \boldsymbol{\nu} = \Sigma_{t\#}\boldsymbol{\mu}^{1\,2}, \quad \pi^{1,3}_\# \boldsymbol{\nu} = \boldsymbol{\mu}^{t\,3}. \tag{7.3.10}$$

Then, Lemma 5.3.2 says that there exists a plan $\boldsymbol{\nu}$ fulfilling (7.3.10) and, since Λ_t is invertible, this proves the existence of $\boldsymbol{\mu}$. When $\boldsymbol{\mu}^{1\,2}$ is optimal, since $\Sigma_{t\#}\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu_t^{1\to 2}, \mu^2)$, we infer from Lemma 7.2.1 that $\Sigma_{t\#}\boldsymbol{\mu}^{1\,2}$ is unique and induced by a transport map and therefore $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ are uniquely determined. □

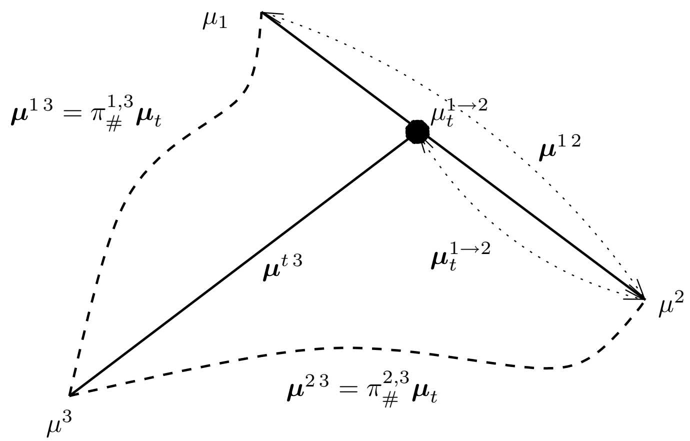

Figure 7.3: $\boldsymbol{\mu}^{1\,2}$ and $\boldsymbol{\mu}^{t\,3}$ are given optimal plans; $\boldsymbol{\mu}^{2\,3}$ and $\boldsymbol{\mu}^{1\,3}$ are not optimal, in general

Theorem 7.3.2 ($\mathscr{P}_2(X)$ is a PC-space). *For each choice of μ^1, μ^2, $\mu^3 \in \mathscr{P}_2(X)$ and $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2)$ we have*

$$W_2^2(\mu_t^{1\to 2}, \mu^3) \geq (1-t)W_2^2(\mu^1, \mu^3) + tW_2^2(\mu^2, \mu^3) - t(1-t)W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2) \quad (7.3.11)$$

and the map $t \mapsto W_2^2(\mu_t^{1\to 2}, \mu^3) - t^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2)$ is concave in $[0,1]$. In particular, choosing $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$ (see Figure 7.3) we have

$$W_2^2(\mu_t^{1\to 2}, \mu^3) \geq (1-t)W_2^2(\mu^1, \mu^3) + tW_2^2(\mu^2, \mu^3) - t(1-t)W_2^2(\mu^1, \mu^2) \quad (7.3.12)$$

and therefore $\mathscr{P}_2(X)$ is a PC-space.

Proof. (7.3.11) is a direct consequence of (7.3.9) and (7.3.3). In order to prove the concavity property we choose λ, t_1, $t_2 \in [0,1]$, $t := (1-\lambda)t_1 + \lambda t_2$, and we have only to develop the obvious calculations:

$$\begin{aligned}
&W_2^2(\mu_t^{1\to 2}, \mu^3) - t^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2) = W_2^2(\mu_\lambda^{t_1\to t_2}, \mu^3) - t^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2)\\
&\geq (1-\lambda)W_2^2(\mu^{t_1}, \mu^3) + \lambda W_2^2(\mu^{t_2}, \mu^3) - \Big(\lambda(1-\lambda)(t_2-t_1)^2 + t^2\Big)W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2)\\
&= (1-\lambda)\big[W_2^2(\mu_{t_1}^{1\to 2}, \mu^3) - t_1^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2)\big] + \lambda\big[W_2^2(\mu_{t_2}^{1\to 2}, \mu^3) - t_2^2 W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2)\big].
\end{aligned}$$

In the case $\boldsymbol{\mu}^{1\,2} \in \Gamma_o(\mu^1, \mu^2)$ is sufficient to note that $W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1, \mu^2) = W_2^2(\mu^1, \mu^2)$. □

Example 7.3.3 (Strict positivity of the sectional curvature). The following example shows that in general the inequality (7.3.1) is strict. Let

$$\mu^1 := \frac{1}{2}\left(\delta_{(1,1)} + \delta_{(5,3)}\right), \quad \mu^2 := \frac{1}{2}\left(\delta_{(-1,1)} + \delta_{(-5,3)}\right), \quad \mu^3 := \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(0,-4)}\right).$$

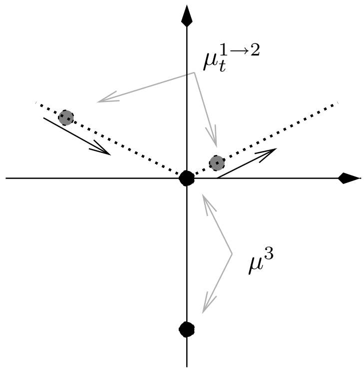

Figure 7.4: μ^3 is the sum of deltas on black dots, $\mu_t^{1\to 2}$ is moving along the dotted lines

Then, it is immediate to check that $W_2^2(\mu^1,\mu^2) = 40$, $W_2^2(\mu^1,\mu^3) = 30$, and $W_2^2(\mu^2,\mu^3) = 30$. On the other hand, the unique constant speed geodesic joining μ^1 to μ^2 is given by

$$\mu_t := \frac{1}{2}\left(\delta_{(1-6t,1+2t)} + \delta_{(5-6t,3-2t)}\right)$$

and a simple computation gives

$$24 = W_2^2(\mu_{1/2},\mu^3) > \frac{30}{2} + \frac{30}{2} - \frac{40}{4}.$$

Formula (7.3.11) is useful to evaluate the directional derivative of the Wasserstein distance. If $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1,\mu^2)$, general properties of concave maps ensures that for each point $t \in [0,1)$ there exists the right derivative

$$\frac{d}{dt+} W_2^2(\mu_t^{1\to 2},\mu^3) := \lim_{t'\downarrow t} \frac{W_2^2(\mu_{t'}^{1\to 2},\mu^3) - W_2^2(\mu_t^{1\to 2},\mu^3)}{t'-t}$$

and, for $t \in (0,1]$, the left derivative

$$\frac{d}{dt-} W_2^2(\mu_t^{1\to 2},\mu^3) := \lim_{t'\uparrow t} \frac{W_2^2(\mu_t^{1\to 2},\mu^3) - W_2^2(\mu_{t'}^{1\to 2},\mu^3)}{t-t'}$$

satisfying

$$\frac{d}{dt+} W_2^2(\mu_t^{1\to 2},\mu^3) \le \frac{d}{dt-} W_2^2(\mu_t^{1\to 2},\mu^3) \quad \forall\, t \in (0,1)$$

and, for a (at most) countable subset $\mathscr{N} \subset (0,1)$

$$\frac{d}{dt+} W_2^2(\mu_t^{1\to 2},\mu^3) = \frac{d}{dt-} W_2^2(\mu_t^{1\to 2},\mu^3) \quad \forall\, t \in (0,1) \setminus \mathscr{N}. \tag{7.3.13}$$

Corollary 7.3.4. *Let* μ^1, μ^2, $\mu^3 \in \mathscr{P}_2(X)$, $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2)$, $t \in [0,1]$, *and*

$\boldsymbol{\mu}_t \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3)$ *such that* $(\pi_t^{1\to 2,3})_{\#}\boldsymbol{\mu} \in \Gamma_o(\mu_t^{1\to 2}, \mu^3)$ *as in Proposition* 7.3.1.

Then

$$\begin{aligned}\frac{d}{dt_+} W_2^2(\mu_t^{1\to 2}, \mu^3) &\le W^2_{\boldsymbol{\mu}_t}(\mu^2, \mu^3) - W^2_{\boldsymbol{\mu}_t}(\mu^1, \mu^3) + (2t-1) W^2_{\boldsymbol{\mu}_t}(\mu^1, \mu^2)\\ &= \frac{1}{1-t}\Big(W^2_{\boldsymbol{\mu}_t}(\mu^2, \mu^3) - W^2_{\boldsymbol{\mu}_t}(\mu_t^{1\to 2}, \mu^2) - W_2^2(\mu_t^{1\to 2}, \mu^3)\Big)\\ &= \frac{1}{t}\Big(W^2_{\boldsymbol{\mu}_t}(\mu_t^{1\to 2}, \mu^1) + W_2^2(\mu_t^{1\to 2}, \mu^3) - W^2_{\boldsymbol{\mu}_t}(\mu^1, \mu^3)\Big)\\ &\le \frac{d}{dt_-} W_2^2(\mu_t^{1\to 2}, \mu^3).\end{aligned} \tag{7.3.14}$$

In particular, equality holds in the previous formula whenever t belongs to the set of differentiability of the distance, i.e. $t \in (0,1) \setminus \mathscr{N}$.

Proof. We simply observe that

$$W_2^2(\mu_{t'}^{1\to 2}, \mu^3) \le W^2_{\boldsymbol{\mu}_t}(\mu_{t'}^{1\to 2}, \mu^3) \quad \text{if } t' \neq t, \qquad W_2^2(\mu_t^{1\to 2}, \mu^3) = W^2_{\boldsymbol{\mu}_t}(\mu_t^{1\to 2}, \mu^3),$$

and we apply (7.3.9) and (7.3.5), (7.3.6), (7.3.7) to evaluate the right and left derivatives. □

We conclude this section by a precise characterization of the right derivative (7.3.14) at time $t = 0$; we need to introduce some more definitions.

Definition 7.3.5 (A new class of multiple plans). *Let* $\boldsymbol{\mu}^{1\,2} \in \mathscr{P}_2(X^2)$ *and* $\mu^3 \in \mathscr{P}_2(X)$. *We say that* $\boldsymbol{\mu} \in \Gamma(\boldsymbol{\mu}^{1\,2}, \mu^3)$ *belongs to* $\Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)$ *if* $\pi^{1,3}_{\#}\boldsymbol{\mu} \in \Gamma_o(\mu^1, \mu^3)$.

Proposition 7.3.6. *Let* $\boldsymbol{\mu}^{1\,2} \in \Gamma(\mu^1, \mu^2)$, $\mu^3 \in \mathscr{P}_2(X)$. *Then for every* $\boldsymbol{\mu} \in \Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)$ *such that*

$$\int_{X^3} |x_2 - x_3|^2 \, d\boldsymbol{\mu} = \min\Big\{\int_{X^3} |x_2 - x_3|^2 \, d\boldsymbol{\nu} : \boldsymbol{\nu} \in \Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu^3)\Big\} \tag{7.3.15}$$

we have

$$\begin{aligned}\frac{d}{dt_+} W_2^2(\mu_t^{1\to 2}, \mu^3)_{|t=0} &= \Big(W^2_{\boldsymbol{\mu}}(\mu^2, \mu^3) - W^2_{\boldsymbol{\mu}}(\mu^1, \mu^2) - W_2^2(\mu^1, \mu^3)\Big)\\ &= -2\int_{X^3} \langle x_2 - x_1, x_3 - x_1\rangle \, d\boldsymbol{\mu}.\end{aligned} \tag{7.3.16}$$

Proof. We already know by (7.3.14) that

$$\frac{d}{dt_+} W_2^2(\mu_t^{1\to 2}, \mu^3)_{|t=0} \le \Big(W^2_{\boldsymbol{\mu}}(\mu^2, \mu^3) - W^2_{\boldsymbol{\mu}}(\mu^1, \mu^2) - W_2^2(\mu^1, \mu^3)\Big)$$

so that we simply have to prove the opposite inequality. Let $\mathscr{N}$ be the negligible set defined by (7.3.13); thanks to (7.3.14) and to the semiconcavity of the squared distance map, we have

$$
\begin{aligned}
\frac{d}{dt+}W_2^2(\mu_t^{1\to 2},\mu^3) &= \lim_{t\downarrow 0, t\notin\mathscr{N}} \frac{d}{dt+}W_2^2(\mu_t^{1\to 2},\mu^3)\\
&= \lim_{t\downarrow 0, t\notin\mathscr{N}} \frac{1}{1-t}\Big(W^2_{\boldsymbol{\mu}_t}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}^{1\,2}}(\mu_t^{1\to 2},\mu^2) - W_2^2(\mu_t^{1\to 2},\mu^3)\Big)\\
&\geq \Big(W^2_{\boldsymbol{\mu}_0}(\mu^2,\mu^3) - W^2_{\boldsymbol{\mu}^{1\,2}}(\mu^1,\mu^2) - W_2^2(\mu^1,\mu^3)\Big),
\end{aligned}
$$

where $\boldsymbol{\mu}_0$ is any narrow accumulation point of $\boldsymbol{\mu}_t$ as $t\downarrow 0$. By Proposition 7.1.3 $\pi^{1\,2}_\#\boldsymbol{\mu}_0 = \boldsymbol{\mu}^{1\,2}$, $\pi^{1\,3}_\#\boldsymbol{\mu}_0 \in \Gamma_o(\mu^1,\mu^3)$. Invoking (7.3.14) again, we conclude. □

Since the integrals of $|x_1-x_2|^2$ and of $|x_1-x_3|^2$ do not depend on the choice of $\boldsymbol{\nu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)$, we can reformulate (7.3.16) as

$$
\frac{d}{dt_+}W_2^2(\mu_t^{1\to 2},\mu^3)|_{t=0} = \min_{\boldsymbol{\nu}\in\Gamma_o(\boldsymbol{\mu}^{1\,2},\mu^3)} -2\int_{X^3}\langle x_2-x_1, x_3-x_1\rangle\,d\boldsymbol{\nu}. \tag{7.3.17}
$$