# Chapter 12

# Appendix

## 12.1 Carathéodory and normal integrands

In this section we recall some standard facts about integrands depending on two variables, measurable w.r.t. the first one, and more regular w.r.t. the second one.

**Definition 12.1.1 (Carathéodory and normal integrands).** *Let $X_1$, $X_2$ be Polish spaces, let $\mu \in \mathscr{P}(X_1)$ and let $\mathscr{L}$ be the $\Sigma$-algebra of $\mu$-measurable subsets of $X_1$. We say that a $\mathscr{L} \times \mathscr{B}(X_2)$-measurable function $f : X_1 \times X_2 \to \mathbb{R}$ is a* Carathéodory *integrand if $x_2 \mapsto f(x_1, x_2)$ is continuous for $\mu$-a.e. $x_1 \in X_1$.*
*We say that a $\mathscr{L} \times \mathscr{B}(X_2)$-measurable function $f : X_1 \times X_2 \to [0, +\infty]$ is a normal integrand if $x_2 \mapsto f(x_1, x_2)$ is lower semicontinuous for $\mu$-a.e. $x_1 \in X_1$.*

In order to check that a given function $f$ is a Carathéodory integrand the following remark will often be useful.

**Remark 12.1.2.** Suppose that a function $f : X_1 \times X_2 \to \mathbb{R}$ satisfies

$$\begin{aligned} x_2 \mapsto f(x_1, x_2) &\quad \text{is continuous for } \mu\text{-a.e. } x_1 \in X_1, \\ x_1 \mapsto f(x_1, x_2) &\quad \text{is } \mathscr{L}\text{-measurable for each } x_2 \in X_2. \end{aligned} \tag{12.1.1}$$

Then $f$ is a Carathéodory integrand. Indeed we can approximate $f$ by the $\mathscr{L} \times \mathscr{B}(X_2)$-measurable functions

$$f_\varepsilon(x_1, x_2) := \sum_i f_\varepsilon(x_1, y_i) \chi_{V_i^\varepsilon}(x_2),$$

where $\{V_i^\varepsilon\}$ is a partition of $X_2$ into (at most) countably many Borel sets with diameter less than $\varepsilon$ and $y_i \in V_i^\varepsilon$. By the first condition in (12.1.1) the functions $f_\varepsilon$ pointwise converge to $u$ out of a set $N \times X_2$ with $\mu(N) = 0$. Therefore $f$ is $\mathscr{L} \times \mathscr{B}(X_2)$-measurable.

For the proof of the following theorem, we refer to [28, Thm. 1, Cor. 1, Thm. 2($(d) \Rightarrow (a)$)].

**Theorem 12.1.3 (Scorza–Dragoni).** *Let $X_1$, $X_2$ be Polish spaces and let $\mu \in \mathscr{P}(X_1)$; if $f$ is defined in $X_1 \times X_2$ with values in $\mathbb{R}$ (resp. in $[0,+\infty]$) is a Carathéodory (resp. normal) integrand, then for every $\varepsilon > 0$ there exists a continuous (resp. l.s.c. and bounded above by $f$) function $f_\varepsilon$ such that*

$$\mu\left(\{x_1 \in X_1 : \; f(x_1,x_2) \neq f_\varepsilon(x_1,x_2) \quad \text{for some } x_2 \in X_2\}\right) \leq \varepsilon. \tag{12.1.2}$$

## 12.2 Weak convergence of plans and disintegrations

In this section we examine more closely the relation between narrow convergence and disintegration for families of plans $\boldsymbol{\gamma}^n \in \mathscr{P}(X_1 \times X_2)$ whose first marginal is independent of $n$.

In the sequel we assume that $X_1$ and $X_2$ are Polish spaces, and $\mu_1 \in \mathscr{P}(X_1)$. We start by stating natural continuity and lower semicontinuity properties with respect to narrow convergence of Carathéodory and normal integrands.

**Theorem 12.2.1.** *Let $\boldsymbol{\gamma}^n \in \mathscr{P}(X_1 \times X_2)$ narrowly converging to $\boldsymbol{\gamma}$ and such that $\pi^1_\#\boldsymbol{\gamma}^n = \mu_1$. Then for every normal integrand $f$ we have*

$$\liminf_{n\to\infty} \int_{X_1\times X_2} f(x_1,x_2)\,d\boldsymbol{\gamma}^n(x_1,x_2) \geq \int_{X_1\times X_2} f(x_1,x_2)\,d\boldsymbol{\gamma}(x_1,x_2), \tag{12.2.1}$$

*and for every bounded Carathéodory integrand we have*

$$\lim_{n\to\infty} \int_{X_1\times X_2} f(x_1,x_2)\,d\boldsymbol{\gamma}^n(x_1,x_2) = \int_{X_1\times X_2} f(x_1,x_2)\,d\boldsymbol{\gamma}(x_1,x_2). \tag{12.2.2}$$

*Proof.* We simply apply Lemma 5.1.10 and the Scorza–Dragoni approximation theorem of the previous section. □

If $\boldsymbol{\gamma}^n$ narrowly converge to $\boldsymbol{\gamma}$ in $\mathscr{P}(X_1 \times X_2)$ and $\pi^1_\#\boldsymbol{\gamma}^n$ is independent of $n$, the following result provides a finer description of the limit $\boldsymbol{\gamma}$.

**Lemma 12.2.2.** *Let $X_1$, $X_2$ be Polish spaces and let $\boldsymbol{\gamma}^n \in \mathscr{P}(X_1 \times X_2)$ narrowly converging to $\boldsymbol{\gamma}$ and such that $\pi^1_\#\boldsymbol{\gamma}^n = \mu_1$ is independent of $n$. If $\{\gamma^n_{x_1}\}_{x_1\in X_1}$, $\{\gamma_{x_1}\}_{x_1\in X_1}$ are the disintegrations of $\boldsymbol{\gamma}^n$, $\boldsymbol{\gamma}$ w.r.t. $\mu_1$ and $G_{x_1} \subset \mathscr{P}(X_2)$ is the subset of all the narrow accumulation points of $(\gamma^n_{x_1})_{n\in\mathbb{N}}$, then we have*

$$\gamma_{x_1} \subset \overline{\operatorname{conv} G_{x_1}} \quad \text{for } \mu_1\text{-a.e. } x_1 \in X_1. \tag{12.2.3}$$

*In particular*

$$\operatorname{supp}\gamma_{x_1} \subset \overline{\bigcup_{\gamma\in G_{x_1}} \operatorname{supp}\gamma} \quad \text{for } \mu_1\text{-a.e. } x_1 \in X_1. \tag{12.2.4}$$

*Proof.* Taking into account Remark 5.1.5 we can find a function $\varphi : X_2 \to [0, +\infty]$ with compact sublevels, such that

$$\int_{X_1 \times X_2} \varphi(x_2)\, d\boldsymbol{\gamma}(x_1, x_2) \le \sup_{n \in \mathbb{N}} \int_{X_1 \times X_2} \varphi(x_2)\, d\boldsymbol{\gamma}^n(x_1, x_2) = S < +\infty. \quad (12.2.5)$$

In particular, for any open set $A \subset X_1$ and any continuous and bounded function $f : X_2 \to \mathbb{R}$ we have

$$\begin{aligned} \int_{A \times X_2} f(x_2)\, d\boldsymbol{\gamma}(x_1, x_2) + \varepsilon S &\ge \lim_{n \to +\infty} \int_{A \times X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\boldsymbol{\gamma}^n(x_1, x_2) \\ &\ge \int_A \Big( \inf_{\varepsilon > 0} \liminf_{n \to \infty} \int_{X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\gamma^n_{x_1}(x_2) \Big)\, d\mu^1(x_1) \end{aligned} \quad (12.2.6)$$

Passing to the limit as $\varepsilon \downarrow 0$ and observing that $A$ is arbitrary, we get

$$\int_{X_2} f(x_2)\, d\gamma_{x_1}(x_2) \ge \inf_{\varepsilon > 0} \liminf_{n \to \infty} \int_{X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\gamma^n_{x_1}(x_2) \quad \text{for } \mu\text{-a.e. } x_1$$

and it is not difficult to show using Prokhorov theorem that

$$\liminf_{n \to \infty} \int_{X_2} \big(f(x_2) + \varepsilon\varphi(x_2)\big)\, d\gamma^n_{x_1}(x_2) \ge \inf_{\gamma \in G_{x_1}} \int_{X_2} f(x_2)\, d\gamma(x_2) \quad (12.2.7)$$

and

$$\int_{X_2} f(x_2)\, d\gamma_{x_1}(x_2) \ge \inf_{\gamma \in G_{x_1}} \int_{X_2} f(x_2)\, d\gamma(x_2) \quad (12.2.8)$$

for $\mu^1$-a.e. $x_1 \in X_1$. Choosing $f$ in a countable set $\mathscr{C}_0$ satisfying (5.1.2a,b) we can find a $\mu^1$-negligible subset $N \subset X_1$ such that (12.2.8) holds for each $f \in \mathscr{C}$ and $x_1 \in X_1 \setminus N$. In fact the approximation property (5.1.2a,b) shows that (12.2.8) holds for each function $f \in C^0_b(X_2)$ and therefore Hahn–Banach theorem yields $\gamma_{x_1} \in \overline{\mathrm{conv}}\, G_{x_1}$ for $x_1 \in X_1 \setminus N$. □

We conclude this section with an useful convergence result:

**Lemma 12.2.3.** *Let $X_1$ be a Polish space, let $X_2$ be a separable Hilbert space, and let $f : X_2 \to [0, +\infty]$ be a l.s.c. strictly convex function. Suppose that $(\boldsymbol{\gamma}_n) \subset \mathscr{P}(X_1 \times X_2)$ narrowly converges to $\boldsymbol{\gamma} = \int_{X_1} \gamma_{x_1}\, d\mu_1(x_1)$, with $\mu_1 = \pi^1_\# \boldsymbol{\gamma}$; if the barycenter of $\boldsymbol{\gamma}$ $\bar{\boldsymbol{\gamma}}(x_1) = \int_{X_2} x_2\, d\boldsymbol{\gamma}_{x_1}(x_2)$ exists and satisfies*

$$\liminf_{n \to \infty} \int_{X_1 \times X_2} f(x_2)\, d\boldsymbol{\gamma}_n(x_1, x_2) = \int_{X_1} f(\bar{\boldsymbol{\gamma}}_{x_1})\, d\mu_1(x_1) \in \mathbb{R} \quad (12.2.9)$$

*then $\boldsymbol{\gamma} = (\boldsymbol{i} \times \bar{\boldsymbol{\gamma}})_\# \mu_1$. The same result holds if $\pi^1_\# \boldsymbol{\gamma}^n = \mu_1$ and $f : X_1 \times X_2 \to [0, +\infty]$ is a normal integrand such that $f(x_1, \cdot)$ is strictly convex for $\mu_1$-a.e. $x_1 \in X_1$; in this case the barycenters $\bar{\boldsymbol{\gamma}}_n$ converge to $\bar{\boldsymbol{\gamma}}$ in $\mu_1$-measure.*

*Proof.* Equality (12.2.9) yields

$$\begin{aligned}\int_{X_1}\Big(\int_{X_2} f(x_2)\,d\boldsymbol{\gamma}_{x_1}(x_2)\Big)\,d\mu_1(x_1) &= \int_{X_1\times X_2} f(x_2)\,d\boldsymbol{\gamma}(x_1,x_2)\\ &\le \liminf_{n\to+\infty}\int_{X_1\times X_2} f(x_2)\,d\boldsymbol{\gamma}_n(x_1,x_2)\\ &\le \int_{X_1} f(\bar{\boldsymbol{\gamma}}(x_1))\,d\gamma_1(x_1),\end{aligned}$$

so that Jensen inequality yields

$$\int_{X_2} f(x_2)\,d\boldsymbol{\gamma}_{x_1}(x_2) = f(\bar{\boldsymbol{\gamma}}(x_1)) \quad \text{for } \mu_1\text{-a.e. } x_1\in X_1$$

and the strict convexity of $f$ yields $\boldsymbol{\gamma}_{x_1} = \delta_{\bar{\boldsymbol{\gamma}}(x_1)}$. The second part of the statement can be proved in an analogous way. □

## 12.3 PC metric spaces and their geometric tangent cone

In this section we review some basic general facts about *positively curved* (in short PC) spaces in the sense of Aleksandrov [5, 40, 139], and we recall the related notion of tangent cone; in the last section we will discuss its relationships with the tangent space we introduced in Section 8.4 for the Wasserstein space $\mathscr{P}_2(X)$.

Let $(\mathscr{S}, d)$ be a metric space; a *constant speed geodesic* $\boldsymbol{x}^{1\to 2} : t\in[0,T]\mapsto x_t\in\mathscr{S}$ connecting $x^1$ to $x^2$ is a curve satisfying

$$x_0 = x^1,\quad x_T = x^2,\quad d(x_t,x_s) = \frac{t-s}{T}d(x^1,x^2)\quad \forall\, 0\le s\le t\le T. \tag{12.3.1}$$

In particular we are dealing with geodesics of minimal length whose metric derivative $|\boldsymbol{x}'|(t)$ is constant on $[0,T]$ and equal to $T^{-1}d(x^1,x^2)$.

We say that $\mathscr{S}$ is *geodesically complete* (or *length space*) if each couple of points can be connected by a constant speed geodesic.

**Definition 12.3.1 (*PC*-spaces).** *A geodesically complete metric space $(\mathscr{S},d)$ is* positively curved *(a PC-space) if for every $x^0\in\mathscr{S}$ and every constant speed geodesic $\boldsymbol{x}^{1\to 2} : t\in[0,1]\mapsto x_t^{1\to 2}$ connecting $x^1$ to $x^2$ it holds*

$$d^2(x_t^{1\to 2},x^0) \ge (1-t)d^2(x^1,x^0) + td^2(x^2,x^0) - t(1-t)d^2(x^1,x^2). \tag{12.3.2}$$

Observe that in an Hilbert space $X$ (12.3.2) is in fact an identity, since for $x_t^{1\to 2} = (1-t)x^1 + tx^2$ we have

$$|x_t^{1\to 2} - x^0|^2 = (1-t)|x^1-x^0|^2 + t|x^2-x^0|^2 - t(1-t)|x^1-x^2|^2. \tag{12.3.3}$$

Therefore condition (12.3.2) can be considered as a sort of comparison property for triangles: let us exploit this fact.

**Definition 12.3.2 (Triangles).** *A* triangle $\boldsymbol{x}$ *in* $\mathscr{S}$ *is a triple* $\boldsymbol{x} = (\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{2\to 3}, \boldsymbol{x}^{3\to 1})$ *of constant speed geodesics connecting (with obvious notation) three points* $x^1$, $x^2$, $x^3$ *in* $\mathscr{S}$. *We denote by* $\triangle = \triangle(\boldsymbol{x}) \subset \mathscr{S}$ *the image of the curves* $\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{2\to 3}, \boldsymbol{x}^{3\to 1}$.

To each triangle $\boldsymbol{x}$ in $\mathscr{S}$ we can consider a corresponding reference triangle (unique, up to isometric transformation) $\hat{\boldsymbol{x}} = (\hat{\boldsymbol{x}}^{1\to 2}, \hat{\boldsymbol{x}}^{2\to 3}, \hat{\boldsymbol{x}}^{3\to 1})$ in $\mathbb{R}^2$ connecting the points $\hat{x}^1, \hat{x}^2, \hat{x}^3 \in \mathbb{R}^2$ such that

$$|\hat{x}^i - \hat{x}^j| = d(x^i, x^j) \quad i, j = 1, 2, 3. \tag{12.3.4}$$

Two points $x \in \triangle$, $\hat{x} \in \hat{\triangle}$ are correspondent if

$$x = \boldsymbol{x}_t^{i\to j}, \quad \hat{x} = \hat{\boldsymbol{x}}_t^{i\to j} \quad \text{for some } t \in [0,1],\ i, j \in \{1, 2, 3\}.$$

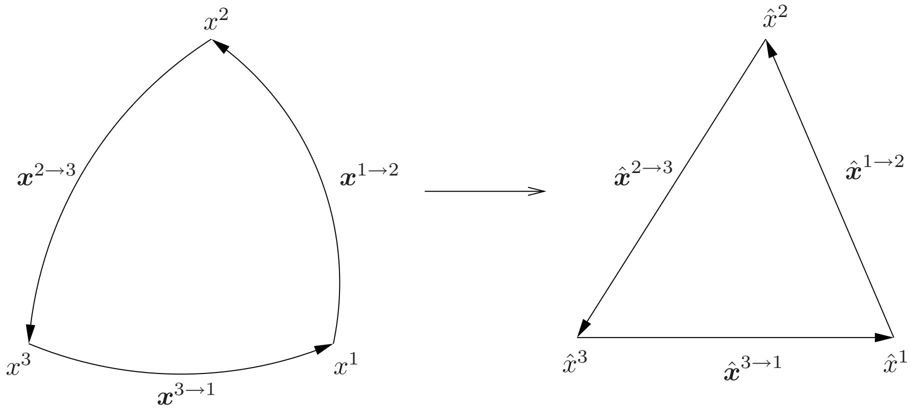


Figure 12.1: on the left the triangle on the *PC*-space and on the right its euclidean reference.

**Proposition 12.3.3 (Triangle comparison).** *If* $\mathscr{S}$ *is a PC-space and* $\triangle \subset \mathscr{S}, \hat{\triangle} \subset \mathbb{R}^2$ *are two corresponding triangles, then for each couples of correspondent points* $x, y \in \triangle, \hat{x}, \hat{y} \in \hat{\triangle}$ *we have*

$$d(x, y) \geq |\hat{x} - \hat{y}|. \tag{12.3.5}$$

*Proof.* When $x$ or $y$ is a vertex of the triangle, then (12.3.5) is just (12.3.2): thus we have to examine the case (up to permutation of the indexes) $x = x_t^{1\to 2}$, $y = x_s^{1\to 3}$, $t, s \in (0, 1)$. Denoting by $\boldsymbol{x}^{1\to t}$ the rescaled geodesic connecting $x^1$ to $x = x_t^{1\to 2}$ and by introducing a new geodesic $\boldsymbol{x}^{t\to 3}$ connecting $x$ to $x^3$, we can consider the new triangle $\boldsymbol{x}' = (\boldsymbol{x}^{1\to t}, \boldsymbol{x}^{t\to 3}, \boldsymbol{x}^{3\to 1})$ connecting $x^1, x, x^3$. The corresponding euclidean reference $\hat{\boldsymbol{x}}'$ can be constructed keeping fixed $\hat{x}^1$ and $\hat{x}^3$ (and therefore $\hat{y} = \hat{x}_s^{1\to 3}$) and introducing a new point $\hat{x}'$, which in general will be different from $\hat{x}$, such that $|\hat{x}' - \hat{x}^1| = d(x, x^1)$, $|\hat{x}' - \hat{x}^3| = d(x, x^3)$. Applying (12.3.2) we obtain

$$|\hat{x}' - \hat{x}^3| = d(x, x^3) \geq |\hat{x} - \hat{x}^3|$$

and applying the identity (12.3.3) we get

$$|\hat{x}' - \hat{y}|^2 = (1-s)|\hat{x}' - \hat{x}^1|^2 + s|\hat{x}' - \hat{x}^3|^2 - s(1-s)|\hat{x}^3 - \hat{x}^1|^2$$
$$\geq (1-s)|\hat{x} - \hat{x}^1|^2 + s|\hat{x} - \hat{x}^3|^2 - s(1-s)|\hat{x}^3 - \hat{x}^1|^2 = |\hat{x} - \hat{y}|^2$$

therefore, applying (12.3.2) again to the triangles $\boldsymbol{x}', \hat{\boldsymbol{x}}'$ we obtain

$$d(x,y) \geq |\hat{x}' - \hat{y}'| = |\hat{x}' - \hat{y}| \geq |\hat{x} - \hat{y}|. \qquad \square$$

In a Hilbert space $X$ the angle $\angle(\hat{\boldsymbol{x}}^{1\to 2}, \hat{\boldsymbol{x}}^{1\to 3}) \in [0,\pi]$ between the two segments joining $\hat{x}^1$ to $\hat{x}^2$ and $\hat{x}^1$ to $\hat{x}^3$ can be easily computed by the formula

$$\cos(\angle(\hat{\boldsymbol{x}}^{1\to 2}, \hat{\boldsymbol{x}}^{1\to 3})) = \frac{\langle \hat{x}^2 - \hat{x}^1, \hat{x}^3 - \hat{x}^1 \rangle}{|\hat{x}^2 - \hat{x}^1|\,|\hat{x}^3 - \hat{x}^1|} = \alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3), \tag{12.3.6}$$

where

$$\alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3) = \frac{|\hat{x}^2 - \hat{x}^1|^2 + |\hat{x}^3 - \hat{x}^1|^2 - |\hat{x}^3 - \hat{x}^2|^2}{2|\hat{x}^2 - \hat{x}^1|\,|\hat{x}^3 - \hat{x}^1|}. \tag{12.3.7}$$

In particular, if $\hat{x}^{1\to 2}_t := (1-t)\hat{x}^1 + t\hat{x}^2$ and $\hat{x}^{1\to 3}_s := (1-s)\hat{x}^1 + s\hat{x}^3$, we have

$$\alpha(\hat{x}^1; \hat{x}^{1\to 2}_t, \hat{x}^{1\to 3}_s) = \alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3) \quad \forall\, t,\, s \in (0,1]. \tag{12.3.8}$$

Taking into account of (12.3.7), in the case of a general $PC$-space, it is natural to introduce the function

$$\alpha(x^1; x^2, x^3) := \frac{d(x^2,x^1)^2 + d(x^3,x^1)^2 - d(x^3,x^2)^2}{2d(x^2,x^1)\,d(x^3,x^1)}, \quad x^1 \neq x^2, x^3 \tag{12.3.9}$$

and we have the following monotonicity result.

**Lemma 12.3.4 (Angle between geodesics).** *Let $(\mathscr{S}, d)$ be a $PC$-space and let $\boldsymbol{x}^{1\to 2}$, $\boldsymbol{x}^{1\to 3}$ be constant speed geodesics starting from $x^1$; then the function*

$$t, s \in (0,1] \mapsto \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s) \quad \textit{is nondecreasing in } s, t. \tag{12.3.10}$$

*The* angle $\angle(\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{1\to 3}) \in [0,\pi]$ *between $\boldsymbol{x}^{1\to 2}$ and $\boldsymbol{x}^{1\to 3}$ is thus defined by the formula*

$$\cos(\angle(\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{1\to 3})) := \inf_{s,t} \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s) = \lim_{s,t\downarrow 0} \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s). \tag{12.3.11}$$

*Proof.* It is sufficient to prove that $\alpha(x^1; x^2, x^3) \geq \alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s)$ for $s, t \in (0,1]$; if $\hat{\boldsymbol{x}}$ is a corresponding reference triangle with vertexes $\hat{x}^1, \hat{x}^2, \hat{x}^3$, we easily have by Proposition 12.3.3 and (12.3.8)

$$\alpha(x^1; x^{1\to 2}_t, x^{1\to 3}_s) \leq \alpha(\hat{x}^1; \hat{x}^{1\to 2}_t, \hat{x}^{1\to 3}_s) = \alpha(\hat{x}^1; \hat{x}^2, \hat{x}^3) = \alpha(x^1; x^2, x^3) \qquad \square$$

**Remark 12.3.5.** Notice that the separate limit as $t \downarrow 0$ is given by

$$\lim_{t\downarrow 0} \alpha(x^1; x_t^{1\to 2}, x_s^{1\to 3}) = \lim_{t\downarrow 0} \frac{t^2 d^2(x^1,x^2) + d^2(x^1, x_s^{1\to 3}) - d^2(x_t^{1\to 2}, x_s^{1\to 3})}{2ts\, d(x^1,x^2)\, d(x^1,x^3)}$$
$$= -\big(2sd(x^1,x^2)\, d(x^1,x^3)\big)^{-1} \frac{d}{dt}\Big(d^2(x_t^{1\to 2}, x_s^{1\to 3})\Big)_{|t=0+}$$

and therefore

$$\cos\big(\angle(\boldsymbol{x}^{1\to 2}, \boldsymbol{x}^{1\to 3})\big) = -\big(2d(x^1,x^2)\, d(x^1,x^3)\big)^{-1} \frac{\partial^2}{\partial s \partial t}\Big(d^2(x_t^{1\to 2}, x_s^{1\to 3})\Big)_{|t,s=0+}$$

For a fixed $x \in \mathscr{S}$ let us denote by $G(x)$ the set of all constant speed geodesics $\boldsymbol{x}$ starting from $x$ and parametrized in some interval $[0, T_{\boldsymbol{x}}]$; recall that the metric velocity of $\boldsymbol{x}$ is $|\boldsymbol{x}'| = d(\boldsymbol{x}(t), x)/t$, $t \in (0, T]$. We set

$$\begin{gathered} \|\boldsymbol{x}\|_x := |\boldsymbol{x}'|, \quad \langle \boldsymbol{x}, \boldsymbol{y}\rangle_x := \|\boldsymbol{x}\|_x \, \|\boldsymbol{y}\|_x \cos(\angle(\boldsymbol{x}, \boldsymbol{y})), \\ d_x^2(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{x}\|_x^2 + \|\boldsymbol{y}\|_x^2 - 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle_x. \end{gathered} \tag{12.3.12}$$

If $\boldsymbol{x} \in G(x)$ and $\lambda > 0$ we denote by $\lambda\boldsymbol{x}$ the geodesic

$$(\lambda \boldsymbol{x})_t := \boldsymbol{x}_{\lambda t}, \quad T_{\lambda \boldsymbol{x}} = \lambda^{-1} T_{\boldsymbol{x}}, \tag{12.3.13}$$

and we observe that for each $\boldsymbol{x}, \boldsymbol{y} \in G(x)$, $\lambda > 0$, it holds

$$\|\lambda \boldsymbol{x}\|_x = \lambda \|\boldsymbol{x}\|_x, \quad \langle \lambda \boldsymbol{x}, \boldsymbol{y}\rangle_x = \langle \boldsymbol{x}, \lambda \boldsymbol{y}\rangle_x = \lambda \langle \boldsymbol{x}, \boldsymbol{y}\rangle_x \tag{12.3.14}$$

Observe that the restriction of a geodesic is still a geodesic; we say that $\boldsymbol{x} \sim \boldsymbol{y}$ if there exist $\varepsilon > 0$ such that $\boldsymbol{x}|_{[0,\varepsilon]} = \boldsymbol{y}|_{[0,\varepsilon]}$.

**Theorem 12.3.6 (An abstract notion of Tangent cone).** *If $\boldsymbol{x}$, $\boldsymbol{y} : [0, T] \to \mathscr{S}$ are two geodesics starting from $x$ we have*

$$d_x(\boldsymbol{x}, \boldsymbol{y}) = \lim_{t\downarrow 0} \frac{d(\boldsymbol{x}_t, \boldsymbol{y}_t)}{t} = \sup_{t\in(0,T]} \frac{d(\boldsymbol{x}_t, \boldsymbol{y}_t)}{t}. \tag{12.3.15}$$

*In particular, the function $d_x$ defined by* (12.3.12) *is a distance on the quotient space $G(x)/\sim$. The completion of $G(x)/\sim$ is called the* tangent cone $\mathbf{Tan}_x\mathscr{S}$ *at the point $x$.*

*Proof.* (12.3.15) follows by a simple computation since for each $s > 0$ (12.3.11) yields

$$\cos\big(\angle(\boldsymbol{x}, \boldsymbol{y})\big) = \lim_{t\downarrow 0} \frac{d^2(\boldsymbol{x}_{ts}, x) + d^2(\boldsymbol{y}_{ts}, x) - d^2(\boldsymbol{x}_{ts}, \boldsymbol{y}_{ts})}{2d(\boldsymbol{x}_{ts}, x) d(\boldsymbol{y}_{ts}, x)}$$
$$= \frac{d^2(\boldsymbol{x}_s, x) + d^2(\boldsymbol{y}_s, x)}{2d(\boldsymbol{x}_s, x) d(\boldsymbol{y}_s, x)} - \lim_{t\downarrow 0} \frac{d^2(\boldsymbol{x}_{ts}, \boldsymbol{y}_{ts})}{2t^2 d(\boldsymbol{x}_s, x) d(\boldsymbol{y}_s, x)}$$

and therefore from (12.3.12) we have

$$d_x^2(\boldsymbol{x},\boldsymbol{y}) = \frac{d^2(\boldsymbol{x}_s,x)+d^2(\boldsymbol{y}_s,x)}{s^2} - 2\frac{d(\boldsymbol{x}_s,x)d(\boldsymbol{y}_s,x)}{s^2}\cos(\angle(\boldsymbol{x},\boldsymbol{y}))$$
$$= \lim_{t\downarrow 0}\frac{d^2(\boldsymbol{x}_{ts},\boldsymbol{y}_{ts})}{2t^2s^2}.$$

□

**Remark 12.3.7 (The tangent cone as Gromov-Hausdorff blow up of pointed spaces).** In the finite dimensional case $\mathbf{Tan}_x\mathscr{S}$ can also be characterized as the Gromov-Hasudorff limit of the sequence of pointed metric spaces $(\mathscr{S},x,n\cdot d)$ as $n\to\infty$. [40, 7.8.1]

## 12.4 The geometric tangent spaces in $\mathscr{P}_2(X)$

Taking into account of the abstract definition of Tangent cone 12.3.6 for $PC$-spaces and the fact proved in Section 7.3 that $\mathscr{P}_2(X)$ is a $PC$-space, we want an explicit representation of the abstract tangent space $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ induced by the 2-Wasserstein distance.

First of all we want to determine a precise expression for the angle between two geodesics. Observe that an optimal plan $\boldsymbol{\mu}\in\Gamma_o(\mu^1,\mu^2)$ is associated to the geodesic $\boldsymbol{\mu}^{1\to2}$ with $\mu_t^{1\to2}=(\pi_t^{1\to2})_\#\boldsymbol{\mu}$ whose velocity is equal to the distance between the end points $|\mu'|^2=\int|x_2-x_1|^2\,d\boldsymbol{\mu}$. If we want to represent each constant speed geodesics, it is convenient to introduce the new "velocity" plans

$$\boldsymbol{\gamma}_\lambda := \big(\pi^1,\lambda(\pi^2-\pi^1)\big)_\#\boldsymbol{\mu}, \tag{12.4.1}$$

that can be used to provide a natural parametrizations for the rescaled geodesic $(\lambda\cdot\boldsymbol{\mu}^{1\to2})_t := \boldsymbol{\mu}_{\lambda t}^{1\to2}$ as follows:

$$\mu_{\lambda t}^{1\to2} = \big((1-\lambda t)\pi^1+\lambda t\pi^2\big)_\#\boldsymbol{\mu} = (\pi^1+t\pi^2)_\#\boldsymbol{\gamma}_\lambda \qquad t\in\big[0,\lambda^{-1}\big]. \tag{12.4.2}$$

Therefore we can identify constant speed geodesics parametrized in some interval $[0,\lambda^{-1}]$ with transport plans $\boldsymbol{\gamma}$ of the type

$$\boldsymbol{\gamma} = \big(\pi^1,\lambda(\pi^2-\pi^1)\big)_\#\boldsymbol{\mu} \quad \text{for some } \textit{optimal plan } \boldsymbol{\mu}\in\mathscr{P}_2(X),$$

and therefore we set

$$\boldsymbol{G}(\mu) := \Big\{\boldsymbol{\gamma}\in\mathscr{P}_2(X^2):\ \pi^1_\#\boldsymbol{\gamma}=\mu,\ \big(\pi^1,\pi^1+\varepsilon\pi^2\big)_\#\boldsymbol{\gamma}\text{ is optimal, for some }\varepsilon>0\Big\}. \tag{12.4.3}$$

It easy to check that there is a one-to-one correspondence between $\boldsymbol{G}(\mu)$ and the quotient $G(\mu)/\sim$ introduced in the previous section: for, to each plan $\boldsymbol{\gamma}\in\boldsymbol{G}(\mu)$ we associate the (equivalence class of the) geodesic

$$\mu_t := (\pi^1+t\pi^2)_\#\boldsymbol{\gamma}, \quad 0\le t\le\varepsilon, \tag{12.4.4}$$

where $\varepsilon > 0$ is chosen as in (12.4.3). Conversely, if $\mu_t$, $t \in [0,T]$, is a curve such that $\mu_{|_{[0,\varepsilon]}}$ is a (minimal, constant speed) geodesic, then for every $\lambda^{-1} \in (0,\varepsilon]$ there exists a unique optimal plan $\boldsymbol{\mu}_\lambda \in \Gamma_o(\mu_0, \mu_{\lambda^{-1}})$ such that

$$\mu_t = \big(\pi^1 + \lambda t(\pi^2 - \pi^1)\big)_\# \boldsymbol{\mu}_\lambda \quad t \in [0, \lambda^{-1}];$$

by Theorem 7.2.2

$$0 < \lambda_1^{-1} < \lambda_2^{-1} \le \varepsilon \quad \Longrightarrow \quad \boldsymbol{\mu}_{\lambda_1} = \big(\pi^1, \pi^1 + \lambda_2/\lambda_1(\pi^2 - \pi^1)\big)_\# \boldsymbol{\mu}_{\lambda_2},$$

so that

$$\boldsymbol{\gamma} = \big(\pi^1, \lambda(\pi^2 - \pi^1)\big)_\# \boldsymbol{\mu}_\lambda \quad \text{is independent of } \lambda \text{, belongs to } \boldsymbol{G}(\mu), \tag{12.4.5}$$

and represents $\mu_t$ through (12.4.4).
Motivated by the above discussion, we introduce the following definition:

**Definition 12.4.1 (Exponential map in $\mathscr{P}_2(X)$).** *For $\mu \in \mathscr{P}(X)$ and $\boldsymbol{\gamma} \in \boldsymbol{G}(\mu)$ we define*

$$\lambda \cdot \boldsymbol{\gamma} := \big(\pi^1, \lambda\pi^2\big)_\# \boldsymbol{\gamma}, \quad \exp_\mu(\boldsymbol{\gamma}) := \big(\pi^1 + \pi^2\big)_\# \boldsymbol{\gamma}. \tag{12.4.6}$$

The notation is justified by the fact that the curve

$$t \mapsto \exp_\mu(t \cdot \boldsymbol{\gamma}) \quad \text{is a constant speed geodesic in some interval } [0,\varepsilon] \tag{12.4.7}$$

whenever $\boldsymbol{\gamma} \in \boldsymbol{G}(\mu)$.

For $\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3} \in \mathscr{P}_2(X^2)$ with $\pi^1_\# \boldsymbol{\gamma}^{1\,i} = \mu$, $i = 2, 3$, we set

$$\|\boldsymbol{\gamma}^{1\,2}\|_\mu^2 := \int_{X^2} |x_2|^2 \, d\boldsymbol{\gamma}^{1\,2}(x_1, x_2), \tag{12.4.8}$$

$$\langle \boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3} \rangle_\mu = \max\Big\{ \int_{X^3} \langle x_2, x_3 \rangle \, d\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) \Big\}, \tag{12.4.9}$$

$$W_\mu^2(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) = \min\Big\{ \int_{X^3} |x_2 - x_3|^2 \, d\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) \Big\}, \tag{12.4.10}$$

where $\Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ is the family of all 3-plans in $\boldsymbol{\gamma} \in \mathscr{P}(X^3)$ such that $\pi^{1,2}_\# \boldsymbol{\gamma} = \boldsymbol{\gamma}^{1\,2}$ and $\pi^{1,3}_\# \boldsymbol{\gamma} = \boldsymbol{\gamma}^{1\,3}$.

**Proposition 12.4.2.** *Suppose that $\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}$ belongs to $\boldsymbol{G}(\mu)$ so that they can be identified with the constant speed geodesics $\boldsymbol{\mu}^{1\to2}, \boldsymbol{\mu}^{1\to3}$ through* (12.4.4)*. Then the previous definitions coincide with the corresponding quantities introduced in* (12.3.12) *for general PC-metric spaces.*

*Proof.* The first identity of (12.4.8) is immediate. In order to prove the second one we apply Proposition 7.3.6, by taking into account Remark 12.3.5: thus we have

$$\langle \boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3} \rangle_\mu = \lim_{s \downarrow 0} 2s^{-1} \int_{X^3} \langle x_2 - x_1, x_3 - x_1 \rangle \, d\boldsymbol{\mu}_s,$$

where $\mu_s^{1\to 3} = \exp_\mu(s\boldsymbol{\gamma}^{1\,3})$ and $\boldsymbol{\mu}_s \in \Gamma_o(\boldsymbol{\mu}^{1\,2}, \mu_s^{1\to 3})$ is chosen among the minimizers of (7.3.15). It is easy to check that we can choose

$$\boldsymbol{\mu}_s = \big(\pi^1, \pi^1 + \pi^2, \pi^1 + s\pi^2\big)_{\#}\boldsymbol{\gamma},$$

where $\boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ realizes the maximum in (12.4.9) (or equivalently the minumum of (12.4.10)) and therefore

$$\begin{aligned}\lim_{s\downarrow 0} s^{-1}\int_{X^3}\langle x_2 - x_1, x_3 - x_1\rangle\, d\boldsymbol{\mu}_s &= \lim_{s\downarrow 0} s^{-1}\int_{X^3}\langle x_2, x_1 + s x_3 - x_1\rangle\, d\boldsymbol{\gamma}\\ &= \int_{X^3}\langle x_2, x_3\rangle\, d\boldsymbol{\gamma}.\end{aligned}$$

The last formula of (12.4.8) follows now directly by the definition (12.3.12). □

If either $\boldsymbol{\gamma}^{1\,2}$ or $\boldsymbol{\gamma}^{1\,3}$ are induced by a transport map $\boldsymbol{t}$, e.g. $\boldsymbol{\gamma}^{1\,2} = \big(\boldsymbol{i}\times\boldsymbol{t}\big)_{\#}\mu$, then the previous formulae are considerably simpler, since

$$\|\boldsymbol{\gamma}^{1\,2}\|_\mu^2 := \int_{X^2} |\boldsymbol{t}(x_1)|^2\, d\mu(x_1) = \|\boldsymbol{t}\|^2_{L^2(\mu;X)}, \tag{12.4.11}$$

$$\langle\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}\rangle_\mu = \int_{X^2}\langle \boldsymbol{t}(x_1), x_3\rangle\, d\boldsymbol{\gamma}^{1\,3}(x_1,x_3), \tag{12.4.12}$$

$$W_\mu^2(\boldsymbol{\gamma}^{1,2}, \boldsymbol{\gamma}^{1\,3}) = \int_{X^2} |\boldsymbol{t}(x_1) - x_3|^2\, d\boldsymbol{\gamma}^{1\,3}(x_1,x_3). \tag{12.4.13}$$

Finally, if also $\boldsymbol{\gamma}^{1\,3} = \big(\boldsymbol{i}\times\boldsymbol{s}\big)_{\#}\mu$, then (12.4.12) and (12.4.13) become

$$\langle\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}\rangle_\mu = \int_X \langle \boldsymbol{t}(x_1), \boldsymbol{s}(x_1)\rangle\, d\mu(x_1) = (\boldsymbol{t}, \boldsymbol{s})_{L^2(\mu;X)}, \tag{12.4.14}$$

$$W_\mu^2(\boldsymbol{\gamma}^{1,2}, \boldsymbol{\gamma}^{1\,3}) = \int_X |\boldsymbol{t}(x_1) - \boldsymbol{s}(x_1)|^2\, d\mu(x_1) = \|\boldsymbol{t} - \boldsymbol{s}\|^2_{L^2(\mu;X)}. \tag{12.4.15}$$

These results lead to the following definition.

**Definition 12.4.3 (Geometric tangent cone).** *The* geometric tangent cone $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ *to* $\mathscr{P}_2(X)$ *at* $\mu$ *is the* closure *of* $\boldsymbol{G}(\mu)$ *in* $\mathscr{P}_2(X^2)$ *with respect to the distance* $W_\mu(\cdot,\cdot)$.

In Section 8.4 we already introduced a notion of tangent space $\mathrm{Tan}_\mu\mathscr{P}_2(X)$ and we showed in Theorem 8.5.1 its equivalent characterization in terms of optimal transport maps

$$\mathrm{Tan}_\mu\mathscr{P}_2(X) = \overline{\big\{\lambda(\boldsymbol{r} - \boldsymbol{i}) : \quad (\boldsymbol{i}\times\boldsymbol{r})_\#\mu \in \Gamma_o(\mu, \boldsymbol{r}_\#\mu),\ \lambda > 0\big\}}^{L^2(\mu;X)}. \tag{12.4.16}$$

In order to compare these two notions, let us recall the Definition 5.4.2 of *barycentric projection* $\bar{\boldsymbol{\gamma}}$ of a plan $\boldsymbol{\gamma} \in \mathscr{P}_2(X^2)$ with $\pi^1_\#\boldsymbol{\gamma} = \mu$:

$$\boldsymbol{t} := \bar{\boldsymbol{\gamma}} \quad\Leftrightarrow\quad \boldsymbol{t}(x_1) = \int_X x_2\, d\boldsymbol{\gamma}_{x_1}(x_2), \quad \boldsymbol{t} \in L^2(\mu;X), \tag{12.4.17}$$

which is a nonexpansive map from $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ to $L^2(\mu;X)$. Indeed choosing $\boldsymbol{\gamma}\in\Gamma^1(\boldsymbol{\gamma}^1,\boldsymbol{\gamma}^2)$ and denoting by $\boldsymbol{\gamma}^1_{x_1}$ and $\boldsymbol{\gamma}^2_{x_1}$ the disintegrations of $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$ w.r.t. $\mu$ we have

$$\int_X |\bar{\boldsymbol{\gamma}}^1-\bar{\boldsymbol{\gamma}}^2|^2\,d\mu=\int_X\left|\int_{X^2}(x_2-x_3)\,d\boldsymbol{\gamma}_{x_1}\right|^2 d\mu\le\int_{X^3}|x_2-x_3|^2\,d\boldsymbol{\gamma},$$

so that

$$\|\bar{\boldsymbol{\gamma}}^1-\bar{\boldsymbol{\gamma}}^2\|_{L^2(\mu;X)}\le W_\mu(\boldsymbol{\gamma}^1,\boldsymbol{\gamma}^2). \tag{12.4.18}$$

We have the following result:

**Theorem 12.4.4.** *For every $\mu\in\mathscr{P}_2(X)$ the tangent space is the image of $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ through the barycentric projection. Moreover, if $\mu\in\mathscr{P}_2^r(X)$, then the barycentric projection is an isometric one-to-one correpondence between $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ and $\mathrm{Tan}_\mu\mathscr{P}_2(X)$.*

*Proof.* Let us first prove that $\bar{\boldsymbol{\gamma}}\in\mathrm{Tan}_\mu\mathscr{P}_2(X)$ for any $\boldsymbol{\gamma}\in\mathbf{Tan}_\mu\mathscr{P}_2(X)$. By the continuity of the barycentric projection and the identity $\overline{(\pi^1,\pi^1+\varepsilon\pi^2)_\#\boldsymbol{\gamma}}=\boldsymbol{i}+\varepsilon\bar{\boldsymbol{\gamma}}$, it suffices to show that $(\bar{\boldsymbol{\mu}}-\boldsymbol{i})\in\mathrm{Tan}_\mu\mathscr{P}_2(X)$ for any optimal plan $\boldsymbol{\mu}$ whose first marginal is $\mu$. We know that $\operatorname{supp}\boldsymbol{\mu}$ is contained in the graph of the subdifferential of a convex and l.s.c. function $\psi:X\to(-\infty,+\infty]$, i.e.

$$y\in\partial\psi(x)\qquad\text{for any }(x,y)\in\operatorname{supp}\boldsymbol{\mu}.$$

Since $\partial\psi(x)$ is a closed convex subset of $X$ for every $x\in D(\partial\psi)$, we obtain that $\bar{\boldsymbol{\mu}}(x)=\int_X y\,d\boldsymbol{\mu}_x(y)\in\partial\psi(x)$ for $\mu$-a.e. $x$; therefore $\bar{\boldsymbol{\mu}}$ is an optimal transport map and $(\bar{\boldsymbol{\mu}}-\boldsymbol{i})\in\mathrm{Tan}_\mu\mathscr{P}_2(X)$.

In order to show that the barycentric projection is onto it suffices to prove that the map $I:\mathrm{Tan}_\mu\mathscr{P}_2(X)\mapsto\mathscr{P}(X\times X)$ defined by $I(\boldsymbol{v}):=(\boldsymbol{i}\times\boldsymbol{v})_\#\mu$ takes its values in $\mathbf{Tan}_\mu\mathscr{P}_2(X)$ and to notice that it satisfies $\overline{I(\boldsymbol{v})}=\boldsymbol{v}$. Since the unique plan in $\Gamma^1\left(I(\boldsymbol{v}),I(\boldsymbol{v}')\right)$ is $(\boldsymbol{i}\times\boldsymbol{v}\times\boldsymbol{v}')_\#\mu$, we have

$$W_\mu^2\left(I(\boldsymbol{v}),I(\boldsymbol{v}')\right)=\int_X|\boldsymbol{v}-\boldsymbol{v}'|^2\,d\mu,$$

so that our thesis follows if $I(\boldsymbol{v})\in\boldsymbol{G}(\mu)$ for every $\boldsymbol{v}$ in the dense subset of $\mathrm{Tan}_\mu\mathscr{P}_2(X)$ introduced in (12.4.16): this last property follows trivially by the definition of $\boldsymbol{G}(\mu)$ (12.4.3). Finally in the case when $\mu$ is regular all optimal transport plans in $\boldsymbol{G}(\mu)$ are induced by transports: therefore $I$ is onto and it is the inverse of the barycentric projection. □

**Remark 12.4.5 (The exponential map and its inverse).** Observe that the exponential map is a contraction since

$$W_2(\exp_\mu(\boldsymbol{\mu}),\exp_\mu(\boldsymbol{\sigma}))\le W_\mu(\boldsymbol{\mu},\boldsymbol{\sigma}), \tag{12.4.19}$$

but in general, it is not injective, even if it is restricted to the tangent space. Nevertheless it admits a natural (multivalued) right inverse defined by

$$\exp_\mu^{-1}(\nu) := \Big\{\boldsymbol{\mu} \in \boldsymbol{G}(\mu) : \big(\pi^1, \pi^1 + \pi^2\big)_{\#}\boldsymbol{\mu} \in \Gamma_o(\mu, \nu)\Big\}. \tag{12.4.20}$$

We conclude this section with an explicit representation of the distance $W_\mu$ defined by (12.4.10).

**Proposition 12.4.6.** *Let $\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}$ be two plans in $\mathscr{P}_2(X^2)$ with the same first marginal $\mu$. Then $\boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ realizes the minimum in* (12.4.8) *if and only if its disintegration w.r.t. $\mu$ satisfies*

$$\boldsymbol{\gamma}_{x_1} \in \Gamma_o(\boldsymbol{\gamma}^{1\,2}_{x_1}, \boldsymbol{\gamma}^{1\,3}_{x_1}) \quad \textit{for } \mu\textit{-a.e. } x_1 \in X. \tag{12.4.21}$$

*Moreover*

$$W_\mu^2(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3}) = \int_X W_2^2(\boldsymbol{\gamma}^{1\,2}_{x_1}, \boldsymbol{\gamma}^{1\,3}_{x_1})\, d\mu(x_1). \tag{12.4.22}$$

*Proof.* For any $\boldsymbol{\gamma} \in \Gamma^1(\boldsymbol{\gamma}^{1\,2}, \boldsymbol{\gamma}^{1\,3})$ we clearly have

$$\int_{X^3} |x_2 - x_3|^2\, d\boldsymbol{\gamma} = \int_X \int_{X^2} |x_2 - x_3|^2\, d\boldsymbol{\gamma}_{x_1}\, d\mu(x_1) \geq \int_X W_2^2(\mu^{1,2}_{x_1}, \mu^{1,3}_{x_1})\, d\mu(x_1).$$

Equality and the necessary and sufficient condition for optimality follows immediately by Lemma 5.3.2 and by the next measurable selection result. □

**Lemma 12.4.7.** *Suppose that $(\mu^2_{x_1})_{x_1 \in X_1}, (\mu^3_{x_1})_{x_1 \in X_1}$ are Borel families of measures in $\mathscr{P}_p(X)$ defined in a Polish space $X_1$.*

*The map*

$$x_1 \mapsto W_p^p(\mu^2_{x_1}, \mu^3_{x_1}) \quad \textit{is Borel} \tag{12.4.23}$$

*and there exists a Borel family $\boldsymbol{\gamma}_{x_1} \in \mathscr{P}_p(X \times X)$ such that $\boldsymbol{\gamma}_{x_1} \in \Gamma_o(\mu^2_{x_1}, \mu^3_{x_1})$.*

*Proof.* We show first that $x \mapsto \sigma_x$ is a Borel map between $X_1$ and $\mathscr{P}_p(X)$ whenever $x \mapsto \sigma_x$ is Borel in the sense used in Section 5.3. Indeed by assumption $x \mapsto \sigma_x(A)$ is a Borel map for any open set $A \subset X$ and since

$$\int_X f\, d\sigma_x = \int_0^\infty \sigma_x(\{f > t\})\, dt - \int_{-\infty}^0 \sigma_x(\{f < t\})\, dt$$

and the integral can be approximated by Riemann sums, we have also that $x \mapsto \int_X f\, d\sigma_x$ is Borel for any $f \in C^0_b(X)$.

Let $\delta$ be the distance inducing the narrow convergence on $\mathscr{P}(X)$ introduced in (5.1.6). It follows that $x \mapsto \delta(\sigma_x, \sigma)$ is Borel for any $\sigma \in \mathscr{P}(X)$. By (7.1.12) it follows that the distance $\tilde{W}$ defined by

$$\tilde{W}^p(\mu, \sigma) := \delta^p(\mu, \sigma) + \left| \int |x|^p\, d\mu - \int |x|^p\, d\sigma \right|$$

induces the $p$-Wasserstein topology on $\mathscr{P}_p(X)$; we deduce that $x \mapsto \tilde{W}(\sigma_x, \sigma)$ is Borel for any $\sigma \in \mathscr{P}_p(X)$, therefore $x \mapsto \sigma_x$ is Borel, seen as a function with values in $\mathscr{P}_p(X)$.

In order to prove the second part of the statement, let us observe that the multivalued map $\mu^2, \mu^3 \in \mathscr{P}_p(X) \mapsto \Gamma_o(\mu^2, \mu^3) \subset \mathscr{P}_p(X \times X)$ is upper semicontinuous thanks to Proposition 7.1.3. In particular for each open set $G \subset \mathscr{P}_p(X \times X)$ the set

$$\Big\{(\mu^2, \mu^3) : \Gamma_o(\mu^2, \mu^3) \cap G \neq \varnothing\Big\}$$

is open in $\mathscr{P}_p(X) \times \mathscr{P}_p(X)$. Therefore classical measurable selection theorems (see for instance Theorem III.23 in [49]) give the thesis. □