Scuola Normale Superiore di Pisa
Classe di Scienze Matematiche, Fisiche e Naturali

PhD Thesis

On the geometry of the space of probability measures in $\mathbb{R}^n$ endowed with the quadratic optimal transport distance

Nicola Gigli

Advisor

prof. Luigi Ambrosio

2002 – 2004

per mia sorella
per mio fratello
e per me

Contents

Introduction

In the last years, starting from the seminal papers [**45, 16, 46, 15**] the theory of optimal transportation and the geometry induced by the optimal transportation problem on the space of probability measures have received a lot of attention (see [**36, 51, 52, 11**]). Motivations for this interest come from application to Optimization, Functional and Geometric Inequalities, Evolution PDE's, and, in even more recent times, Riemannian Geometry.

The aim of this thesis is to provide an comprehensive study of the geometry of the space of probability measures in $\mathbb{R}^d$ endowed with the quadratic optimal transportation distance. Some of the results described here are contained in the book [**11**], written with L. Ambrosio and G. Savaré (those in Chapter 3 and part of those in Chapter 4), while other results are presented here for the first time (in Chapter 4, 5 and 6). In particular we complete the analysis performed in [**11**] of a "calculus" with transport plans (considering a larger tangent bundle which includes plans as well), showing that a nice Hilbertian structure can be extended up to this level. In addition, still having in mind an Hilbertian analogy, we study the closure with respect to weak topologies of geodesically convex sets in the space of probability measures, showing that one of the assumptions made in [**11**], in the theory of gradient flows for convex functionals, is redundant. Finally, we try to make a complete analysis (with positive results, and counterexamples as well) of the problem of parallel transport in these spaces, showing that the parallel transport is well-defined and unique along a class of regular curves.

Now we pass to a more detailed description of the thesis's content.

In Chapter 1 we recall the measure-theoretic results we will need in the work. In particular, in the first section we study the definition and the first properties of the push-forward of measures: given $\mu \in \mathscr{P}(\mathbb{R}^d)$ and a μ-measurable function $f : \mathbb{R}^d \to \mathbb{R}^{d'}$, the push-forward $f_\#\mu$ of μ through f is the probability measure on $\mathbb{R}^{d'}$ defined by

$$f_\#\mu(A) := \mu(f^{-1}(A)), \qquad \forall \text{ Borel set } A \subset \mathbb{R}^{d'}.$$

In the last section we shortly discuss the Kantorovich formulation of the quadratic optimal transportation problem, i.e. the problem of finding, given two measures μ, ν in $\mathscr{P}(\mathbb{R}^d)$, the minimum of

$$\int |x-y|^2 d\boldsymbol{\gamma}(x,y), \tag{0.1}$$

among the set $\mathcal{A}dm(\mu,\nu)$ of all *admissible plans* $\boldsymbol{\gamma}$, i.e. the set of those probability measure $\boldsymbol{\gamma}$ on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\pi^1_\#\boldsymbol{\gamma} = \mu$, $\pi^2_\#\boldsymbol{\gamma} = \nu$, π^1, π^2 being the projections onto the first and the second coordinate, respectively. The aim is just to recall the cornerstones of the theory, without entering into the details of the proofs (for this, we refer for instance to [**11**] and to [**51**]). The key results are the existence of the minimum of the optimal transportation problem and Brenier's theorem: the former ensures that for any couple of measures μ, ν in $\mathscr{P}(\mathbb{R}^d)$ a minimizing plan always exists, the latter says that if μ and ν satisfy $\int |x|^2 d\mu$, $\int |x|^2 d\nu < \infty$ and μ is *regular* (i.e. $\mu(\Sigma) = 0$ for any $\mathcal{H}^{d-1}$ rectifiable set $\Sigma \subset \mathbb{R}^d$), then the minimizing plan $\boldsymbol{\gamma}$ is unique and there exists a (unique μ-a.e.) convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that $\boldsymbol{\gamma} = (Id, \nabla\varphi)_\# \mu$, Id being the identity operator.

In Chapter 2 we begin the study of the distance W, recalling some well-known facts. In the first section we introduce the space of probability measures with bounded second moment

$$\mathscr{P}_2(\mathbb{R}^d) := \Big\{ \mu \in \mathscr{P}(\mathbb{R}^d) \ : \ \int |x|^2 d\mu(x) < +\infty \Big\}, \tag{0.2}$$

and the distance W on $\mathscr{P}_2(\mathbb{R}^d)$:

$$W^2(\mu, \nu) := \min \Big\{ \int |x - y|^2 d\boldsymbol{\gamma} \ : \ \boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu) \Big\}. \tag{0.3}$$

We call *optimal* a plan $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$ which attains the minimum in the above expression.

In the second section we show that the space $(\mathscr{P}_2(\mathbb{R}^d), W)$ is a length space and we characterize its geodesics: we show that a curve $t \to \mu_t$ is a constant speed geodesic from μ_0 to μ_1 with constant velocity $W(\mu_0, \mu_1)$ if and only if there exists an optimal plan $\boldsymbol{\gamma}$ such that

$$\mu_t = ((1-t)\pi^1 + t\pi^2)_\# \boldsymbol{\gamma}. \tag{0.4}$$

The interpolation between two measures μ_0, μ_1 given by the above formula was first introduced by McCann in [**45**] to prove uniqueness of the minimum of a certain class of functionals in $\mathscr{P}(\mathbb{R}^d)$: for this purpose he studied the *displacement convexity* of those functionals (now also known as geodesic convexity on $(\mathscr{P}_2(\mathbb{R}^d), W)$), i.e. convexity along the curves $t \to \mu_t$, with μ_t defined as above.

In the third section we begin the study of the curvature of the space $(\mathscr{P}_2(\mathbb{R}^d), W)$ by showing that it is *positively curved* in the sense of Aleksandrov, that is: for every constant speed geodesic μ_t and every measure σ it holds

$$W^2(\mu_t, \sigma) \geq (1-t)W^2(\mu_0, \sigma) + tW^2(\mu_1, \sigma) - t(1-t)W^2(\mu_0, \mu_1), \tag{0.5}$$

(observe that in the flat Euclidean setting the squared distance satisfies the equality in this equation). A way to interpret equation (0.5) is: the funtion $W^2(\cdot, \sigma)$ is (-1)–concave along constant speed geodesics. It is then interesting to know that the same function is *not* λ–convex for any λ: we show this fact with an explicit counterexample. In the fourth section we introduce the weak topology τ on $\mathscr{P}_2(\mathbb{R}^d)$, which will play a central role in Chapter 5, as the topology such that a sequence (μ_n) on $\mathscr{P}_2(\mathbb{R}^d)$ converges to

μ iff

$$\begin{aligned} \int \varphi d\mu_n \to \int \varphi d\mu \quad & \forall \varphi \in C_c(\mathbb{R}^d, \mathbb{R}), \\ \int |x|^2 d\mu_n < C, \quad & \forall n \in \mathbb{N}. \end{aligned} \tag{0.6}$$

The reason of such a definition is that, at least from a formal viewpoint, many properties of $(\mathscr{P}_2(\mathbb{R}^d), W)$ are the natural generalization of properties valid in an Hilbert setting: in these cases the topology induced by W plays the role of the strong topology and τ plays the one of the weak topology (for example, it is possible to show that W is lower semicontinuous w.r.t. τ and that bounded sets are relatively compact w.r.t. τ). Moreover, recalling the first part of [**11**], it is worth noticing that the topology τ is the natural weak topology in the space $\mathscr{P}_2(\mathbb{R}^d)$ for the study, by compactness methods, of the minimizing movements scheme (the implict Euler scheme canonically used for the approximation of gradient flows). We do not investigate the application to PDE's in this work and we just point out that the introduction of τ does not add new results to [**11**]: the definition of τ simply clarifies the ideas, and helps putting everything in the right perspective. The last section is devoted to the development of some useful tools concerning the convergence of maps $T_n \in L^2_{\mu_n}$ when the base measures μ_n may depend on n.

Chapter 3 is the key one to start the study of the *Riemannian* structure of $(\mathscr{P}_2(\mathbb{R}^d), W)$, first studied in the pioneering work [**46**] and then formalized in [**11**]. In the first section we characterize absolutely continuous curves in $\mathscr{P}_2(\mathbb{R}^d)$ through the study of the continuity equation

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0. \tag{0.7}$$

Indeed we show that, for any absolutely continuous curve, for a.e. t there exists a vector $v_t \in L^2(\mu_t, \mathbb{R}^d)$ ($L^2_{\mu_t}$ in the sequel) such that (0.7) holds and the bound

$$\|v_t\|_{\mu_t} \le |\dot{\mu}_t| \tag{0.8}$$

is valid for a.e. t (where $\|v\|_\mu$ is the norm of the vector v in L^2_μ). Conversely, every solution (μ_t, v_t) of (0.7) with

$$\int_0^1 \|v_t\|_{\mu_t} dt < +\infty \tag{0.9}$$

gives a curve $t \to \mu_t$ which is a.e. equal to an absolutely continuous curve for which $|\dot{\mu}_t| \le \|v_t\|_{\mu_t}$ holds.

Having this characterization in mind, it is natural to think to the v_t's as the "velocity" vectors of the curve μ_t. The only problem is that, for a given μ_t, there is more than one family of v_t's which satisfies (0.7): indeed, it is sufficient to find vectors w_t such that $\nabla \cdot (w_t \mu_t) = 0$ to find a new family $v_t + w_t$ still satisfying (0.7). So we need to understand which v_t's we should think as *the right* ones. There are essentially two ways to answer to this question, both of them leading to the same result: on one hand, one can observe that by (0.7) the action of v_t is defined only on the gradient of smooth functions. On the other hand, one can observe that by the linearity of the expression $\nabla\cdot(v_t\mu_t)$ and the strict convexity of the L^2 norm, there exists only one family of vectors v_t of minimal norm satisfying (0.7): such vectors are characterized by $\int\langle v_t,w_t\rangle d\mu_t=0$ for every w_t such that $\nabla\cdot(w_t\mu_t)=0$ and satisfy, by what we said on the solution of (0.7), $\|v_t\|_{\mu_t}=|\dot\mu_t|$ for a.e. t. In both cases we are lead to define the velocity vectors of μ_t as those satisfying (0.7) and belonging to the *tangent space* at μ_t, defined, for a general measure μ, as

$$\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) := \overline{\left\{\nabla\varphi\ :\ \varphi\in C_c^\infty(\mathbb{R}^d)\right\}}^{L^2_\mu}, \tag{0.10}$$

which we endow with the L^2_μ norm.

It should be noted carefully that, with this definition, the space $(\mathscr{P}_2(\mathbb{R}^d),W)$ is *n*ot an infinite dimensional Riemannian manifold in the canonical sense: indeed, it can be shown that the natural exponential map $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))\ni v\to\exp_\mu(v):=(Id+v)_\#\mu$ has *injectivity radius* 0 for every μ. However, there are quite a lot analogies with that structure. The first one is that the following formula, known as Benamou-Brenier formula [**15**], holds:

$$W(\mu_0,\mu_1)=\min\left\{\int_0^1\|v_t\|_{\mu_t}dt\ :\ \frac{d}{dt}\mu_t+\nabla\cdot(v_t\mu_t)=0\right\}, \tag{0.11}$$

which shows that we can recover the distance between two measures by minimizing the length of the connecting curves.

We conclude the Chapter with a first study of the differential properties of $W^2(\cdot,\sigma)$, showing that this functional is differentiable at regular measures μ and that its differential is $v:=2(Id-T_\mu^\sigma)$ (where T_μ^ν is the unique optimal transport map from μ to ν). This means that the following equation holds:

$$\lim_{\nu\to\mu}\frac{W^2(\nu,\sigma)-W^2(\mu,\sigma)-\displaystyle\int\langle v,T_\mu^\nu-Id\rangle d\mu}{W(\mu,\nu)}=0. \tag{0.12}$$

In Chapter 4 we deeply analyze a topic introduced in the appendix of [**11**]: a more general tangent space which includes not only maps, as $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, but even plans. With the same terminology of [**11**], we call this space the *geometric* tangent space. The reason for the introduction of such a space is that the space $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is not completely satisfactory when dealing with non-regular measures: for example if $\mu=\delta_{x_0}$ is a Dirac mass, then $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is isomorphic to $\mathbb{R}^d$ and the image of the exponential map consists only of Dirac masses. If we want to define a tangent space with a surjective exponential map, we have to allow the "splitting" of masses, in the same spirit of Kantorovich relaxation, involving plans, of Monge's problem, involving maps. In the first section of the Chapter we introduce the geometric tangent space $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ via an analogy with the case of a regular manifold M embedded in $\mathbb{R}^N$: for an embedded manifold M, one may say that a vector v belongs to the tangent space of M in x_0 if and only if there exists a constant speed geodesic $t\to x_t$ such that $v=\lim_t(x_t-x_0)/t$, therefore the tangent space can be defined by studying only the behavior of geodesics. With the formal identification of points x_t of the manifold with measures μ_t in $\mathscr{P}_2(\mathbb{R}^d)$, of the vector $x_t - x_0$ with the plan $(\pi^1, \pi^2 - \pi^1)_\# \boldsymbol{\gamma}_t$, where $\boldsymbol{\gamma}_t$ is an optimal plan between μ_0 and μ_t, and of the rescaled vector $(x_t - x_0)/t$ with the "rescaled" plan $(\pi^1, (\pi^2 - \pi^1)/t)_\# \boldsymbol{\gamma}_t$, we know that our tangent space has to contain at least all the plans of the kind

$$\lim_{t \to 0^+} \left(\pi^1, \frac{\pi^2 - \pi^1}{t} \right)_\# \boldsymbol{\gamma}_t,$$

where $\boldsymbol{\gamma}_t$ is any choice of optimal plans from a fixed measure μ_0 and the point μ_t of a constant speed geodesic $t \to \mu_t$. Of course we have to define the topology of the limit. The characterization of geodesics given by equation (0.4) allows us to say that the previous limit is equal to (because eventually equal to) $(\pi^1, (\pi^2 - \pi^1)/T)_\# \boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ is an optimal plan and $T > 0$ is a positive number. Reversing the last formula we know that $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ has to contain at least the set $\mathcal{G}eod\,\mu$ defined as

$$\mathcal{G}eod\,\mu := \left\{ \boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu \ : \ (\pi^1, \pi^1 + T\pi^2)_\# \boldsymbol{\gamma} \text{ is optimal for some } T > 0 \right\},$$

where $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ is the set of all plans in $\mathscr{P}_2(\mathbb{R}^{2d})$ whose first marginal is μ. Given that we want the regular tangent space to be isometrically included in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ via the natural embedding $v \to \mathfrak{I}(v) := (Id, v)_\# \mu$, we are lead to endow $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ with the distance

$$W_\mu^2(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) := \int W^2((\boldsymbol{\gamma}_1)_x, (\boldsymbol{\gamma}_2)_x) d\mu(x),$$

where $\boldsymbol{\gamma}_x$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. its first marginal. The definition of the geometric tangent space will then be

$$\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) := \overline{\mathcal{G}eod\,\mu}^{W_\mu},$$

meaning that the closure is taken w.r.t. the distance W_μ. Given that

$$\|f - g\|_\mu = W_\mu(\mathfrak{I}(f), \mathfrak{I}(g)) \tag{0.13}$$

and that for every regular function φ the map $Id + t\nabla\varphi$ is optimal for sufficiently small t, we have that $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is naturally embedded in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. Equation (0.13) suggests a way to define the *norm* of plan and the *scalar product* between plans as following:

$$\|\boldsymbol{\gamma}\|^2 := W_\mu^2(\boldsymbol{\gamma}, \mathfrak{I}(0)) = \int |x_2|^2 d\boldsymbol{\gamma}(x_1, x_2), \tag{0.14a}$$

$$2\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle := \|\boldsymbol{\gamma}_1\|^2 + \|\boldsymbol{\gamma}_2\|^2 - W_\mu^2(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2). \tag{0.14b}$$

Furthermore, the exponential map $\boldsymbol{\gamma} \to \exp(\boldsymbol{\gamma}) := (\pi^1 + \pi^2)_\# \boldsymbol{\gamma}$ is clearly surjective, and the inverse function $\exp_\mu^{-1}(\nu)$ is well defined by

$$\exp_\mu^{-1}(\nu) := \left\{ \boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) : (\pi^1, \pi^1 + \pi^2)_\# \boldsymbol{\gamma} \text{ is optimal between } \mu \text{ and } \nu \right\},$$

(which is not the functional inverse of exp).

In the second section we derive a useful and interesting formula to compute the directional derivative of $W^2(\cdot, \sigma)$ along an interpolation curve $t \to \mu_t := (\pi^1, \pi^1 +$

$t\pi^2)_\#\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ is any plan in $\mathscr{P}_2(\mathbb{R}^{2d})$ not necessarily optimal. The formula is:

$$\frac{d}{dt}W^2(\mu_t,\sigma)|_{t=0} = -2 \max_{\boldsymbol{\sigma}\in\exp_\mu^{-1}(\sigma)} \langle \boldsymbol{\gamma}, \boldsymbol{\sigma}\rangle,$$

which is formally equivalent to the one valid in Riemannian manifolds.

In the third section, following the same approach given in [**11**] (see also [**4, 20, 49**]), we introduce an abstract notion of tangent cone for positively curved length spaces (i.e. length spaces which satisfy an inequality like (0.5)), showing that a notion of *angle* between geodesics and of norm of a geodesic are well defined, and that in the case of $(\mathscr{P}_2(\mathbb{R}^d), W)$ these notions are consistent with those defined in (0.14).

In the fourth section we conclude the study of the embedding

$$\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) \hookrightarrow \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)).$$

by showing the non trivial fact that the subset of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ of plans induced by maps is isomorphic through $\mathfrak{I}^{-1}$ to $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$: this means that every optimal map is tangent, regardless of any assumption on the regularity of the base measure. Conversely, we study the *barycentric projection* of a plan $\boldsymbol{\gamma}$, defined as

$$\mathscr{B}(\boldsymbol{\gamma})(x_1) := \int x_2 d(\boldsymbol{\gamma})_{x_1},$$

showing that it is a well defined map from $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ onto L^2_μ, and that the range of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, thus providing a natural left inverse to $\mathfrak{I}$.

The fifth section is devoted to the study of the algebraic properties of operations on $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$: we introduce a (multivalued) sum between plans as

$$\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2 := \Big\{(\pi^0, \pi^1+\pi^2)_\#\boldsymbol{\alpha} \ : \ \boldsymbol{\alpha}\in\mathscr{P}_2(\mathbb{R}^{3d}),\ \pi^{0,1}_\#\boldsymbol{\alpha} = \boldsymbol{\gamma}_1, \pi^{0,2}_\#\boldsymbol{\alpha} = \boldsymbol{\gamma}_2\Big\},$$

the product of a plan $\boldsymbol{\gamma}$ with a scalar λ as

$$\lambda\cdot\boldsymbol{\gamma} := (\pi^1, \lambda\pi^2)_\#\boldsymbol{\gamma},$$

and we show that these operations and the scalar product defined in (0.14), endow $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ with a structure very similar to the one of an Hilbert space. A remarkable result of this section is that the *opposite* $-1\cdot\boldsymbol{\gamma}$ of a tangent plan $\boldsymbol{\gamma}$, is tangent, too: for this reason we called $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ tangent *space* rather than tangent cone.

The last section of the Chapter studies the projection operator

$$\mathscr{P}_2(\mathbb{R}^{2d})_\mu \ni \boldsymbol{\gamma} \to \mathcal{P}(\boldsymbol{\gamma}) \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), \tag{0.15}$$

where $\mathcal{P}(\boldsymbol{\gamma})$ is the element of minimal distance from $\boldsymbol{\gamma}$ among those in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. We show that this operator is well-defined, that is characterized by the property

$$\langle\boldsymbol{\gamma},\boldsymbol{\eta}\rangle = \langle\mathcal{P}(\boldsymbol{\gamma}),\boldsymbol{\eta}\rangle, \quad \forall\boldsymbol{\eta}\in\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)),$$

and that, at least for plans $\boldsymbol{\gamma}$ induced by maps, it holds the identity

$$\lim_{t\to0^+}\frac{W(\exp(t\cdot\boldsymbol{\gamma}),\mu)}{t} = \|\mathcal{P}(\boldsymbol{\gamma})\|.$$

In Chapter 5 we study the τ-closure properties of *geodesically convex* sets and of *strongly geodesically convex* sets in $\mathscr{P}_2(\mathbb{R}^d)$, i.e. those sets C such that

- $((1-t)\pi^1 + t\pi^2)_\#\boldsymbol{\gamma} \in C$ for every $\mu, \nu \in C$ and some $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu,\nu)$,
- $((1-t)\pi^1 + t\pi^2)_\#\boldsymbol{\gamma} \in C$ for every $\mu, \nu \in C$ and every $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu,\nu)$,

respectively. The reason of such an interest comes from the fact that, as pointed out in [**11**], in order to have convergence of the minimizing movements scheme it is necessary to have some sort of weak compactness of the sublevels of the functional. On Hilbert spaces this is guaranteed when the functional is convex and lower semicontinuous, because closed and convex sets are weakly closed, thus it is natural to ask whether the same is true for *geodesically convex* functionals, i.e. functionals convex along constant speed geodesics, in $\mathscr{P}_2(\mathbb{R}^d)$. In [**11**] mainly two kind of geodesically convex functionals on $\mathscr{P}_2(\mathbb{R}^d)$ are studied: those *geodesically convex* and those *strongly geodesically convex*, defined respectively as those functionals F satisfying

- $F(\mu_t) \le (1-t)F(\mu_0) + tF(\mu_1)$ for every $\mu_t = ((1-t)\pi^1 + t\pi^2)_\#\boldsymbol{\gamma}$, for some optimal plan $\boldsymbol{\gamma}$,
- $F(\mu_t) \le (1-t)F(\mu_0) + tF(\mu_1)$ for every $\mu_t = ((1-t)\pi^1 + t\pi^2)_\#\boldsymbol{\gamma}$, for some plan $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu,\nu)$.

Clearly the sublevels of a (strongly) geodesically convex functional are (strongly) geodesically convex sets, thus in order to have the desired convergence of the minimizing movements technique one should prove some kind of $\tau-$compactness of bounded (strongly) geodesically convex sets. In [**11**] the problem is solved in two different ways for strongly and not strongly geodesically convex functionals:

- for strongly g.c. functionals an a priori estimate of the distance between two discrete solutions is given, proving the strong convergence of scheme without any compactness assumption,
- for g.c. functionals the $\tau-$compactness of sublevel sets is assumed as a further hypothesis.

The approach we present here is rather different: we focus on (strongly) g.c. sets and we imitate the proof of weak compactness of convex sets in Hilbert spaces to our space, to gain $\tau-$closure. This requires first the introduction of the "halfspaces" $\mathscr{H}^+_{\boldsymbol{\gamma};C}$ and $\mathscr{H}^-_{\boldsymbol{\gamma};C}$:

$$\mathscr{H}^+_{\boldsymbol{\gamma};C} := \left\{ \nu \in \mathscr{P}_2(\mathbb{R}^d) \ : \ \langle \boldsymbol{\eta}, \boldsymbol{\gamma} \rangle \ge C, \text{ for some } \boldsymbol{\eta} \in \exp_\mu^{-1}(\nu) \right\},$$
$$\mathscr{H}^-_{\boldsymbol{\gamma};C} := \left\{ \nu \in \mathscr{P}_2(\mathbb{R}^d) \ : \ \langle \boldsymbol{\eta}, \boldsymbol{\gamma} \rangle \le C, \text{ for some } \boldsymbol{\eta} \in \exp_\mu^{-1}(\nu) \right\}.$$

We prove the $\tau-$compactness of the sets $\mathscr{H}^+_{\boldsymbol{\gamma};C}$ and, when the base measure μ is regular, of $\mathscr{H}^-_{\boldsymbol{\gamma};C}$. And finally we prove that

- a strongly g.c. set may be written as the intersection of a family of hyperplanes of the kind $\mathscr{H}^+_{\boldsymbol{\gamma};C}$,
- a g.c. set *included in the set of regular measures* may be written as the intersection of a family of hyperplanes of the kind $\mathscr{H}^-_{\boldsymbol{\gamma};C}$.

This shows that strongly g.c. sets are always $\tau-$closed (consistently with what proved in [**11**]), and that g.c. sets made of regular measures are $\tau-$closed, too: this implies the redundancy of the hypothesis on $\tau-$compactness of the sublevels, at least for functionals which are equal to $+\infty$ on non-regular measures. Technical difficulties arise when trying to generalize the second result to general g.c. sets: it is not clear to the author whether the regularity hypothesis on the measures may be dropped or not.

In the final section of the Chapter we introduce the concept of subdifferential of functionals in $\mathscr{P}_2(\mathbb{R}^d)$ (already presented in [**11**]), in light of the results and of the notation developed in Chapter 4: again, here we do not introduce new concepts, but just review them in a (possibly) clearer way. The purpose is to show that of the two potential definitions of subdifferential of geodetically convex functional

$$\boldsymbol{\gamma} \in \partial^- F(\mu) \Leftrightarrow \forall \nu \in \mathscr{P}_2(\mathbb{R}^d)\ \forall \boldsymbol{\eta} \in \exp_\mu^{-1}(\nu) \quad F(\nu) \geq F(\mu) + \langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle,$$
$$\boldsymbol{\gamma} \in \partial^- F(\mu) \Leftrightarrow \forall \nu \in \mathscr{P}_2(\mathbb{R}^d)\ \forall \boldsymbol{\eta} \in \exp_\mu^{-1}(\nu) \quad F(\nu) \geq F(\mu) - \langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle,$$

only the second one (equivalent to the one adopted in [**11**]) is correct, if one aims to have a strong-weak closure property of the subdifferential: this is proved by an explicit counterexample. At the end of the section we apply the definitions given to the study of the differential properties of the functional $W^2(\cdot, \sigma)$.

In Chapter 6 we make a second-order analysis of $(\mathscr{P}_2(\mathbb{R}^d), W)$. Given that we know that $(\mathscr{P}_2(\mathbb{R}^d), W)$ has a "Riemannian like" structure, we try to answer to questions as: is there a Levi-Civita connection in this case? If yes, what is it? Does parallel transport of vectors exist? What about the curvature operator? These questions, mostly at a formal (smooth) level and on compact Riemannian manifolds, have been already addressed in [**43**]: here we focus also on some analytic aspects of these questions.

The Chapter is structured as follows: in the first section we recall the case of a regular manifold M embedded in $\mathbb{R}^N$, showing how the Levi-Civita connection, the parallel transport and the curvature operator look like in this well-known case. We also show an existence proof of the parallel transport based on an approximation argument, which we will be able to imitate in the Wasserstein setting. The idea is to show that, given a curve $[0,1] \ni t \to \gamma(t) \in M$, a tangent vector u_0 in $\gamma(0)$, and letting P_t be the projection of $\mathbb{R}^N$ onto the tangent space at M in $\gamma(t)$, there exists the limit of

$$\mathcal{P}(u_0) := P_1(P_{t_{n-i}}(\cdots P_{t_1}(u_0)))$$

over the directed set of partitions $\mathcal{P} = \{0 \leq t_0 < \cdots < t_n = 1\}$ of $[0,1]$ (where with some abuse of notation we wrote $\mathcal{P}$ both for the partition and for the induced operator) and that this limit is the parallel transport of u_0 to the tangent space in $\gamma(1)$ along the curve $\gamma(t)$. The key ingredient to prove this result is the inequality

$$|P_{t_3}(u) - P_{t_3}(P_{t_2}(u))| \leq C|u||t_3 - t_2||t_2 - t_1|, \tag{0.18}$$

valid for all vectors u in the tangent space of $\gamma(t_1)$, where C is a constant which depends on the regularity of the manifold and of the curve. The idea to prove this inequality is that in a regular setting the norm of the operators

$$V_t \ni u \to u - P_s(u),$$
$$V_t^{\perp} \ni u \to P_s(u),$$

are bounded by $C|s-t|$, where V_t, $V_t^{\perp}$ are the tangent and the orthogonal space to M at $\gamma(t)$, respectively. From inequality (0.18), it follows that for given a partition $\mathcal{P} = \{0 \le t_0 \le \cdots \le t_n = 1\}$ and any refinement $\mathcal{P}'$ it holds

$$|\mathcal{P}(u) - \mathcal{P}'(u)| \le C|u| \sum_{i=0}^{n-1} |t_{i+1} - t_i|^2,$$

from which the desired convergence of the approximation scheme follows.

In the second section we develop the analytical tools needed to replicate in the Wasserstein setting the construction made for embedded Riemannian manifolds. The key concept we introduce is the *angle between tangent spaces* through a map: for a given couple $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ and a map $T \in L^2_\mu$ such that $T_\# \mu = \nu$, we define the angle $\theta_T(\mu,\nu) \in [0, \pi/2]$ as

$$\sin\theta_T(\mu,\nu) := \sup_{\substack{v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|v\|_\nu = 1}} \|v \circ T - P_\mu(v \circ T)\|_\mu,$$

and we prove that it holds:

$$\sin\theta_T(\mu,\nu) \le \mathrm{Lip}(T - Id).$$

In the third section we introduce the class of *regular curves* which is the class of curves along which we are able to define the parallel transport and to show its existence. We will say that an absolutely continuous curve $[0,1] \ni t \to \mu_t$ is regular if its velocity vector field v_t satisfies:

$$\int_0^1 \mathrm{Lip}(v_t)dt < \infty,$$

where by $\mathrm{Lip}(v)$, $v \in L^2_\mu$, we mean the infimum of the Lipschitz constant among all the vector field $\mu-$a.e. equal to v.

The well-known Cauchy-Lipschitz theory ensures that for a regular curve there exists a unique family of maps $\mathbf{T}(t,s,\cdot) : \operatorname{supp}\mu_t \to \operatorname{supp}\mu_s$, which we call the *flow* of the curve, absolutely continuous in s and Lipschitz in x satisfying

$$\text{(0.20)} \qquad \begin{cases} \mathbf{T}(t,t,x) = x, & \forall x, t, \\ \dfrac{d}{ds}\mathbf{T}(t,s,x)_{|s=\overline{s}} = v_{\overline{s}}(\mathbf{T}(t,\overline{s},x)) & \forall x,t, \quad a.e.\ \overline{s}, \\ \mathbf{T}(s,r,\mathbf{T}(t,s,x)) = \mathbf{T}(t,r,x), & \forall x,t,s,r \\ (\mathbf{T}(t,s,x))_\# \mu_t = \mu_s & \forall t,s \in [0,1]. \end{cases}$$

The existence of the flow maps allows the introduction of the concept of *absolutely continuous* vector field along μ_t: we say that $t \to u_t \in L^2_{\mu_t}$ is absolutely continuous if for any $t_0 \in [0,1]$ the map $t \to u_t \circ \mathbf{T}(t_0,t,\cdot) \in L^2_{\mu_{t_0}}$ is absolutely continuous. Given that the composition with $\mathbf{T}(t_0,t_0',\cdot)$ is an isometry between $L^2_{\mu_{t_0'}}$ and $L^2_{\mu_{t_0}}$, to check the absolute continuity of a vector field, it is sufficient to check whether $t \to u_t \circ \mathbf{T}(t_0, t, \cdot) \in L^2_{\mu_{t_0}}$ is absolutely continuous for some $t_0 \in [0, 1]$.

The derivative $\frac{d}{dt} u_t \in L^2_{\mu_t}$ of an absolutely continuous vector field u_t is then defined as

$$\frac{d}{dt} u_t := \lim_{h \to 0} \frac{u_{t+h} \circ \mathbf{T}(t, t+h, \cdot) - u_t}{h} = \frac{d}{dr}(u_r \circ \mathbf{T}(s, r, \cdot))_{|r=t} \circ \mathbf{T}(t, s, \cdot), \quad \forall s \in [0, 1].$$

It is a simple consequence of the definition that the derivative exists for a.e. $t \in [0, 1]$ and that the function $t \to \|\frac{d}{dt} u_t\|_{\mu_t}$ is integrable. Another easy consequence of the definition is that the Leibniz rule holds in the following sense:

$$\frac{d}{dt}\langle u^1_t, u^2_t\rangle_{\mu_t} = \langle \frac{d}{dt} u^1_t, u^2_t\rangle_{\mu_t} + \langle u^1_t, \frac{d}{dt} u^2_t\rangle_{\mu_t}, \tag{0.21}$$

It is important to underline that this definition of derivative allows us to take derivative of a function u_t whose range belongs to different L^2 spaces as t varies: actually these spaces can be quite different from each other, if the support of μ_t depends on time.

We conclude the section with a proof of density of regular curves in the set of absolutely continuous curves: the delicate point in our approximation result is due to the fact that regularity imposes a Lipschitz condition on the *tangent* velocity field. The typical approximation schemes for solutions to the continuity equation, on the other hand, produce a regularized vector field that is compatible with the regularized density, but it is not tangent in general. Therefore a further projection of the regularized velocity on the tangent space is needed.

In section 4 we study the problem of the parallel transport: we prove that along regular curves a precise notion of parallel transport can be given, and we are able to prove its existence and uniqueness. Conversely, we show that under some circumstances the parallel transport "no matter the definition" doesn't exist if the curve fails to be regular.

Imitating the definition of parallel transport in the case of an embedded manifold, we say that $t \to u_t$ is a parallel transport along the regular curve μ_t if it is tangent, absolutely continuous and satisfies:

$$P_{\mu_t}\left(\frac{d}{dt} u_t\right) = 0, \quad a.e.\ t \in [0, 1], \tag{0.22}$$

where $P_{\mu_t} : L^2_{\mu_t} \to \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ is the orthogonal projection. In the case when $\mu_t = \rho_t \mathscr{L}^d$ ($\mathscr{L}^d$ being the Lebesgue measure), the PDE corresponding to the parallel transport of a gradient vector field $\nabla \varphi_t$ is, in accordance with the formal calculations made in [**43**],

$$\nabla \cdot \Big((\partial_t \nabla \varphi_t + \nabla^2 \varphi_t \cdot v_t) \rho_t \Big) = 0.$$

It is a simple consequence of the definition and of equation (0.21) that the parallel transport is linear and preserves the scalar product, and therefore is unique.

Existence is a more delicate issue. The key bound we use, which is a direct consequence of equations (0.20) is

$$\text{Lip}(\mathbf{T}(t,s,\cdot)-Id) \le \exp\left(\int_t^s \text{Lip}(v_r)dr\right) - 1. \tag{0.23}$$

The idea of the proof is to imitate the construction made for embedded manifolds: we introduce the operators $\mathscr{P}_t^s : \text{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \to \text{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$ as $\mathscr{P}_t^s(u) := P_{\mu_s}(u \circ \mathbf{T}(s,t,\cdot))$, then we observe that for any $u \in \text{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ the vector field $t \to \mathscr{P}_0^t(u)$ is a first order approximation of the parallel transport, in the sense that it satisfies equation (0.22) at $t=0$. Indeed from the bound (0.23) it follows

$$\|\mathscr{P}_t^0(\mathscr{P}_0^t(u))-u\|_{\mu_0} = \|\mathscr{P}_t^0(\mathscr{P}_0^t(u)-u\circ\mathbf{T}(t,0,\cdot))\|_{\mu_0} \le \|u\|_{\mu_0}\left(\int_0^t \text{Lip}(v_r)dr\right)^2 = o(t).$$

Then we define for any partition $\mathcal{P} := \{0=t_1,\dots,t_n=1\}$ of $[0,1]$ the map from $\text{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ to $\text{Tan}_{\mu_1}(\mathscr{P}_2(\mathbb{R}^d))$

$$\mathcal{P}(u) := \mathscr{P}_{t_{n-1}}^{t_n}(\cdots \mathscr{P}_{t_1}^{t_2}(u)),$$

and, thanks to the bound (0.23), we are able to prove that for any $u \in \text{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ there exists the limit $\mathcal{T}_0^1(u)$ of $\mathcal{P}(u)$ as P varies over the direct set of partitions. We then repeat this argument to the restriction of the curve to the intervals $[t,s]$, to produce a family of maps $\mathcal{T}_t^s : \text{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \to \text{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$. These maps are exactly the parallel transport maps along the regular curve μ_t, in the sense that for any $u \in \text{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ the vector field $t \to \mathcal{T}_0^t(u)$ is the parallel transport of u along μ_t.

We conclude the section with some explicit example of parallel transport and with a counterexample to its existence along a non regular curve.

In the fifth section we define the covariant derivative and the curvature tensor on the space $(\mathscr{P}_2(\mathbb{R}^d), W)$. The spirit is similar to the one of Lott in [**43**], although with a less formal approach.

Recalling that in a classical embedded Riemannian manifold the parallel transport $T_t^s : T_{\gamma(t)}M \to T_{\gamma(s)}M$ and the covariant derivative $\nabla_{\dot\gamma(t)}$ along a curve $\gamma(t)$ are related by

$$\nabla_{\dot\gamma(t)}u_t = \lim_{s\to t} P_t\left(\frac{T_s^t(u_s)-u_t}{s-t}\right),$$

we define analogously the covariant derivative in $\mathscr{P}_2(\mathbb{R}^d)$ along the regular curve μ_t as

$$\nabla_{v_t}u_t = \lim_{s\to t} P_{\mu_t}\left(\frac{\mathcal{T}_s^t(u_s)-u_t}{s-t}\right) = P_{\mu_t}(\frac{d}{dt}u_t),$$

for any absolutely continuous vector field $t \to u_t$ (here v_t is the velocity vector field of the curve μ_t). We are then able to prove that this derivative is the Levi-Civita on $\mathscr{P}_2(\mathbb{R}^d)$, proving that it is uniquely characterized by the compatibility with the metric and the torsion-free identity.

Having defined the Levi-Civita connection, we can study the curvature operator: we will give a useful representation formula for it (already appeared in [**43**]) and prove that the sectional curvatures of $\mathscr{P}_2(\mathbb{R}^d)$ are always non-negative.

In section 6 we continue analyzing the formal analogy between $\mathscr{P}_2(\mathbb{R}^d)$ and an embedded Riemannian manifold, by defining an analog of the Sasaki metric on the tangent bundle.

Recall that, for a Riemannian manifold M, it is possible to endow the tangent bundle TM with a natural Riemannian metric, the so-called Sasaki metric, which induces the following distance on the tangent bundle:

$$d^2\Big((p^1,u^1),(p^2,u^2)\Big) = \inf_{\gamma}\left\{\big(\mathcal{L}(\gamma)\big)^2 + |\mathcal{T}(u^1)-u^2|^2\right\},$$

where the infimum is taken among all the smooth curves $\gamma(t)$ in M connecting p^1 to p^2, $\mathcal{L}(\gamma)$ is the length of γ and $\mathcal{T}(u^1)$ is the parallel transport of u^1 along γ to the point p^2.

This definition has an analogous in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$, because it doesn't rely on a differential structure on the tangent bundle to be written. Therefore, for any couple of measures $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and any couple of tangent vectors $u \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, $v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d))$ we define

$$d^2\big((\mu,u),(\nu,v)\big) := \inf_{\mu_t}\left\{(\mathcal{L}(\mu_t))^2 + \|v-\mathcal{T}_0^1(u)\|_\nu^2\right\},$$

where the infimum is taken on the set of regular curves $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ such that $\mu_0 = \mu$ and $\mu_1 = \nu$, $\mathcal{L}(\mu_t)$ is the length of μ_t and $\mathcal{T}_t^s$ are the parallel transport maps along μ_t. In particular we define $d((\mu,u),(\nu,v)) := +\infty$ if there is no regular curve connecting μ to ν (which may happen, think for instance to the case in which the the support of μ and the one of ν have a different number of connected components: the fact that the flow maps are isomorphisms implies that there is no regular curve joining μ to ν).

In order to obtain a real valued distance we first define the lower semicontinuous envelope d^* of d by:

$$d^*\big((\mu,u),(\nu,v)\big) := \inf\Big\{\liminf_{n\to\infty} d\big((\mu_n,u_n),(\nu_n,v_n)\big) \ : \ (\mu_n,u_n)\to(\mu,u),\ (\nu_n,v_n)\to(\nu,v)\Big\},$$

where by convergence of (μ_n,u_n) to (μ,u) we mean: $W(\mu,\mu_n)\to 0$, $\int u_n\varphi d\mu_n \to \int u\varphi d\mu$, $\forall\varphi\in C^b(\mathbb{R}^d,\mathbb{R}^d)$ and $\|u_n\|_{\mu_n}\to\|u\|_\mu$. Then, since it is not clear whether the function d^* is sufficienly well-behaved (for instance whether the triangle inequality holds), we define the function $\mathcal{D}$ as:

$$\mathcal{D}\big((\mu,u),(\nu,v)\big) := \inf\left\{d^*\big((\mu,\nabla\varphi),(\nu,\nabla\psi)\big) + \|u-\nabla\varphi\|_\mu + \|v-\nabla\psi\|_\nu\right\},$$

where the infimum is taken among all $\varphi,\psi\in C_c^\infty(\mathbb{R}^d)$. We then prove that $\mathcal{D}$ is a distance on the set of couples (μ,u), $u\in\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, which metrizes the convergence described above. Furthermore, for any absolutely continuous curve (μ_t,u_t) w.r.t. $\mathcal{D}$

the curve μ_t is absolutely continuous in $\mathscr{P}_2(\mathbb{R}^d)$ and:

$$\lim_{s\to t} \frac{\mathcal{D}\big((\mu_s, u_s),(\mu_t, u_t)\big)}{|s-t|} \geq \|v_t\|_{\mu_t} \qquad \text{for a.e. } t, \tag{0.24}$$

where v_t is the tangent field of μ_t.
Conversely, if μ_t is a regular curve and u_t is a parallel transport along it, the map (μ_t, u_t) is absolutely continuous w.r.t. $\mathcal{D}$ and equality holds a.e. in (0.24).

Having this result in mind, we give the following definition: $t \to u_t$ is a weak parallel transport along the absolutely continuous curve $t \to \mu_t$ if the curve $t \to (\mu_t, u_t)$ is absolutely continuous w.r.t. $\mathcal{D}$ and equality holds in (0.24) for a.e. t. The final part of the section is devoted to the study of the properties of the weak parallel transport. However, it is important to underline that even this weaker notion of transport is not sufficient to gain existence: the counterexample to the existence of parallel transport we present in section 4, works as well for this weaker notion.

CHAPTER 1

Preliminary results

1. Transport of measures

For any separable metric space X we will denote by $\mathscr{B}(X)$ the family of all Borel subset of X, and by $\mathscr{P}(X)$ the set of Borel probability measures on X. Given two separable metric spaces X_1, X_2, a measure $\mu \in \mathscr{P}(X_1)$ and a $\mu-$measurable map $T : X_1 \to X_2$ the *push forward of* μ *through* T is the measure $T_\# \mu \in \mathscr{P}(X_2)$ defined as

$$T_\# \mu(E) := \mu(T^{-1}(E)) \quad \forall E \in \mathscr{B}(X_2). \tag{1.1}$$

Note that if f, g are $T_\#\mu$-measurable functions which differ on a $T_\#\mu$-negligible set E then $f \circ T, g \circ T$ differ on $T^{-1}(E)$ which is μ-negligible by definition of $T_\#\mu$. In particular, for every $T_\#\mu$-measurable function f, the function $f \circ T$ is $\mu-$measurable, and it follows from the definition that

$$\int_{X_2} f dT_\# \mu = \int_{X_1} f \circ T d\mu. \tag{1.2}$$

More precisely, one integral exists if and only if exists the other, and in this case they are equal. Observe that for any $p \geq 1$ the composition with T is a well defined map from $\mathscr{L}^p(X_2, T_\#\mu)$ to $\mathscr{L}^p(X_1, \mu)$: equation (1.2) implies that this map is an isometry. The same argument shows that if $S : X_2 \to X_3$ is a $T_\#\mu$ measurable map with values in a separable metric space X_3, then $T \circ S$ is a well defined $\mu-$measurable map. In this case it holds

$$(S \circ T)_\# \mu = S_\#(T_\# \mu).$$

From now on we will denote with $\operatorname{supp} \mu$ the *support* of the measure $\mu \in \mathscr{P}(X_1)$ defined as the complement of the largest open set U such that $\mu(U) = 0$, or equivalently, as

$$\operatorname{supp} \mu := \Big\{ x \in X_1 \ : \ \mu(U) > 0 \text{ for each neighborhood } U \text{ of } x \Big\}.$$

It is always true that the measure $T_\#\mu$ is concentrated on the image of T, however it may happen that $\operatorname{supp}(T_\#\mu) \subsetneq \overline{T(\operatorname{supp} \mu)}$ if the image of T is not closed. A simpler situation occurs if we let T be continuous: indeed in this case it holds

$$T(\operatorname{supp} \mu) \subset \operatorname{supp}(T_\# \mu) = \overline{T(\operatorname{supp} \mu)}.$$

Moreover it is easy to check that under this hypothesis the map $T_\# : \mathscr{P}(X_1) \to \mathscr{P}(X_2)$ is continuous w.r.t. the weak convergence in duality with continuous and bounded functions, that is:

$$\mu_n \rightharpoonup \mu \quad \Leftrightarrow \quad \int_X \varphi d\mu_n \to \int_X \varphi d\mu, \ \forall \varphi \in C_b(X).$$

In the following we will often consider product spaces $\boldsymbol{X} = \prod_1^N X_i$ where the X_i are separable metric spaces (mostly copies of $\mathbb{R}^d$), in this case we will usually write x_i and $\boldsymbol{x} = (x_1, \dots, x_N)$ for a variable in X_i and $\boldsymbol{X}$ respectively. The projections operators π^i and $\pi^{i_1,\dots,i_k}$ are defined as

$$\begin{array}{rclcrcl} \pi^i : \boldsymbol{X} & \to & X_i & \qquad & \pi^{i_1,\dots,i_k} : \boldsymbol{X} & \to & \prod_{j=1}^k X_{i_j} \\ \boldsymbol{x} & \to & x_i & & \boldsymbol{x} & \to & (x_{i_1}, \dots, x_{i_k}). \end{array}$$

Given a measure $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$ (usually called $N-$*plan*), its marginals are defined as $\pi^i_{\#}\boldsymbol{\mu} \in \mathscr{P}(X_i)$ and $\pi^{i_1,\dots,i_k}_{\#}\boldsymbol{\mu} \in \mathscr{P}(\prod_k X_{i_k})$. For $\mathscr{A}_i \subset \mathscr{P}(X_i)$, $i = 1, \dots, N$, the set of admissible $N-$plans is

$$\mathcal{ADM}(\mathscr{A}_1, \dots, \mathscr{A}_N) := \Big\{ \boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X}) \ : \ \pi^i_{\#}\boldsymbol{\mu} \in \mathscr{A}_i, \ \forall i = 1, \dots N \Big\}.$$

If each $\mathscr{A}_i$ is a singleton, $\mathscr{A}_i = \{\mu_i\}$, then we will write $\mathcal{ADM}(\mu_1, \dots, \mu_N)$ instead of $\mathcal{ADM}(\{\mu_1\}, \dots, \{\mu_N\})$.

REMARK 1.1. *If $(\boldsymbol{\mu}_n) \in \mathscr{P}(\boldsymbol{X})$ is a sequence of plans weakly converging in duality with C_b to some $\boldsymbol{\mu} \in \mathscr{P}(\boldsymbol{X})$ then the continuity of the projection implies that $\pi^i_{\#}\boldsymbol{\mu}_n$ weakly converge in duality with C_b to $\pi^i_{\#}\boldsymbol{\mu}$ for any $i = 1, \dots, N$.*

If $N = 2$ then a plan $\boldsymbol{\mu} \in \mathcal{ADM}(\mu_1, \mu_2)$ is also called a *transport plan* for the couple (μ_1, μ_2). If a transport plan satisfies

$$\boldsymbol{\mu} = (Id, T)_{\#}\mu_1 \tag{1.3}$$

for some map T, then we will say that the plan is *induced by the map T* from μ_1. It can be easily seen that a plan $\boldsymbol{\mu}$ is induced by a map if and only if it is concentrated on the graph of some μ_1-measurable map T, moreover in this case the map T is unique up to μ_1-negligible sets and (1.3) holds.

2. Tightness, 2-boundedness and 2-uniform integrability

From now on we will work in Euclidean spaces, so we assume $X = \mathbb{R}^d$. A subset $\mathscr{A} \subset \mathscr{P}(X)$ is said to be *tight* if for any $\varepsilon > 0$ there exists a compact subset K_ε of X such that $\mu(X \setminus K_\varepsilon) < \varepsilon$ for every $\mu \in \mathscr{A}$, or, equivalently, if

$$\overline{\lim_{R \to +\infty}} \sup_{\mu \in \mathscr{A}} \mu(X \setminus B_R) = 0.$$

A different, but equivalent, formulation of the definition of tightness is the following: $\mathscr{A}$ is tight if and only if there exists a positive function φ such that $\varphi(x) \to +\infty$ as $|x| \to +\infty$ for which it holds

$$\sup_{\mu \in \mathscr{A}} \int \varphi \, d\mu =: C < +\infty. \tag{1.4}$$

Indeed suppose (1.4) holds, then Chebichev inequality gives

$$\mu(\{\varphi > t\}) \leq \frac{C}{t},$$

and since the set $\{\varphi \le t\}$ is compact we get one implication. Conversely suppose that $\mathscr{A}$ is tight: let K_n be an increasing sequence of compact sets such that $\mu(X \setminus K_n) \le 2^{-n}$ for any $\mu \in \mathscr{A}$ and define

$$\varphi(x) := \inf \Big\{ n \ : \ x \in K_n \Big\}.$$

Clearly $\varphi \to +\infty$ tends to infinity as $|x| \to \infty$, moreover for any $\mu \in \mathscr{A}$ it holds

$$\int \varphi d\mu = \sum_n n(\mu(K_n \setminus K_{n-1})) = \sum_n n(\mu(X \setminus K_{n-1}) - \mu(X \setminus K_n)) \le \sum_n \frac{n}{2^{n-1}} = 1,$$

where $K_0 = \varnothing$.

THEOREM 1.2 (Prokhorov). *A set $\mathscr{A} \subset \mathscr{P}(X)$ is relatively compact w.r.t. the weak convergence in duality with C_b if and only if it is tight. Moreover a sequence $(\mu_n) \in \mathscr{P}(X)$ weakly converging in duality with C_c to a measure μ (possibly not in $\mathscr{P}(X)$) is tight if and only if*

$$\lim_n \int \varphi d\mu_n = \int \varphi d\mu, \quad \forall \varphi \in C_b(X). \tag{1.5}$$

Proof. Suppose that $\mathscr{A}$ is tight, choose a sequence $\mu_n \in \mathscr{A}$ and extract a subsequence (not relabeled) weakly converging in duality with C_c to some μ. Clearly $\mu(X) \le 1$, so we need to prove that $\mu(X) \ge 1$. Fix $\varepsilon > 0$, find a compact K_ε such that $\mu_n(K_\varepsilon) \ge 1 - \varepsilon$ for any n, and choose a function $\chi_\varepsilon \in C_c(X)$ such that $0 \le \chi_\varepsilon \le 1$ and identically 1 on K_ε. Since it holds

$$\mu(X) \ge \int \chi_\varepsilon d\mu = \lim_n \int \chi_\varepsilon d\mu_n \ge \varliminf \mu_n(K_\varepsilon) \ge 1 - \varepsilon, \tag{1.6}$$

we proved that tightness implies compactness. Let us turn to the converse implication: suppose $\mathscr{A}$ is not tight so that there exists $\varepsilon > 0$ such that for any compact set K it holds $\sup_n \mu_n(X \setminus K) > \varepsilon$. Up to pass to a subsequence, not relabeled, we may assume that (μ_n) weakly converge to some μ in duality with C_c. Choose $R_n \nearrow +\infty$ and, for any n, find a measure $\mu_n \in \mathscr{A}$ such that $\mu_n(B_{R_n}) \le 1 - \varepsilon$. Letting $\chi_n \in C_c(X)$ be a function such that $0 \le \chi_n \le 1$, $\chi \equiv 1$ on B_{R_n} and $\chi_n \equiv 0$ outside $B_{R_{n+1}}$ we have

$$\mu(X) = \sup_n \mu(B_{R_n}) \le \sup_n \int \chi_n d\mu = \sup_n \lim_m \int \chi_n d\mu_m \le 1 - \varepsilon.$$

For the second part of the proof note that choosing $\varphi \equiv 1$ in (1.5) we get $\mu(X) = 1$ and so, by the previous argument, that μ_n is tight. Suppose conversely that μ_n is tight, fix $\varphi \in C_b(X)$ and let $\chi_R \in C_c(x)$ be a cut-off function $0 \le \chi_R \le 1$ identically 1 in B_R and equal to 0 outside B_{2R}. We have

$$\begin{aligned} \left| \int \varphi d\mu_n - \int \varphi d\mu \right| &\le \left| \int \varphi \chi_R d\mu_n - \int \varphi \chi_R d\mu \right| + \left| \int \varphi (1 - \chi_R) d\mu_n \right| + \left| \int \varphi (1 - \chi_R) d\mu \right| \\ &\le \left| \int \varphi \chi_R d\mu_n - \int \varphi \chi_R d\mu \right| + \sup |\varphi| \Big(\sup_n \mu_n(X \setminus B_R) + \mu(X \setminus B_R) \Big), \end{aligned}$$

from which it follows

$$\varlimsup_n \left| \int \varphi d\mu_n - \int \varphi d\mu \right| \le \sup |\varphi| \Big(\sup_n \mu_n(X \setminus B_R) + \mu(X \setminus B_R) \Big).$$

Letting R tend to $+\infty$ we get the thesis. □

PROPOSITION 1.3. *Let* $\boldsymbol{X} = \prod_{i=1}^N X_i$ *be a product space. A set* $\mathscr{A} \subset \mathscr{P}(\boldsymbol{X})$ *is tight if and only if so are the sets* $\pi^i_\#(\mathscr{A}) \subset \mathscr{P}(X_i)$ *for any* $i = 1, \dots, n$.

Proof. One implication follows from the relative compactness of $\mathscr{A}$ and the continuity of the maps $\pi^i_\#$ w.r.t. the duality with C_b. For the other one fix $\varepsilon > 0$ and choose compact sets $K_i \subset X_i$ such that $\mu_i(X_i \setminus K_i) \leq \varepsilon/N$ for any $\mu_i \in \pi^i_\#(\mathscr{A})$, $i = 1, \dots, n$. The set $\prod_i K_i$ is compact in $\boldsymbol{X}$ and it holds

$$\mu(\boldsymbol{X} \setminus \prod_{i=1}^n K_i) \leq \sum_{i=1}^n \mu((\pi^i)^{-1}(X \setminus K_i)) \leq \varepsilon, \quad \forall \mu \in \mathscr{A}.$$

□

In the next chapters we will mostly work with the subset of $\mathscr{P}(X)$ of measures with finite second moment:

$$\mathscr{P}_2(X) := \Big\{ \mu \in \mathscr{P}(X) \ : \ \int |x|^2 d\mu < +\infty \Big\}.$$

Note that if $\mu \in \mathscr{P}_2(X)$ and φ is a Borel function with quadratic growth, i.e. a function such that

$$|\varphi(x)| \leq A|x|^2 + B \qquad \forall x \in X,$$

for some $A \geq 0$, $B \in \mathbb{R}$, then $\varphi \in \mathscr{L}^1(X, \mu)$ and so $\int \varphi d\mu$ is a well defined real number.

A set $\mathscr{A} \subset \mathscr{P}_2(X)$ is *2-uniformly integrable* if for every ε there exists a compact set K_ε such that

$$\int_{X \setminus K_\varepsilon} |x|^2 d\mu < \varepsilon, \qquad \forall \mu \in \mathscr{A}.$$

This definition is formally very close to the one of tightness (it is actually possible to introduce the 2-uniform integrability in terms of tightness of the set $\{(1 + |x|^2)\mu : \mu \in \mathscr{A}\}$), and it will be of primary importance in the study of compactness in the space $\mathscr{P}_2(X)$ endowed with the Wasserstein distance.

The proof of the next proposition is very similar to the one we gave before, so we will omit it.

PROPOSITION 1.4. *Given* $\mathscr{A} \subset \mathscr{P}_2(X)$ *the following three conditions are equivalent.*

i) $\mathscr{A}$ *is 2-uniformly integrable.*

ii)
$$\varlimsup_{R \to +\infty} \sup_{\mu \in \mathscr{A}} \int_{X \setminus B_R} |x|^2 d\mu(x) = 0.$$

iii)
$$\sup_{\mu \in \mathscr{A}} \int \varphi d\mu < +\infty,$$

for some positive function φ *with more than quadratic growth at infinity, i.e. a function such that* $\varphi(x)/|x|^2 \to +\infty$ *as* $|x| \to +\infty$.

Moreover for a given sequence $(\mu_n) \subset \mathscr{P}_2(X)$ *weakly converging in duality with* C_c *to some* $\mu \in \mathscr{P}_2(X)$, *the following conditions are equivalent:*

i) (μ_n) *is 2-uniformly integrable,*

ii) $\displaystyle\int |x|^2 d\mu = \lim \int |x|^2 d\mu_n$,

iii) $\displaystyle\int \varphi d\mu = \lim \int \varphi d\mu_n$ *for any continuous function* φ *with quadratic growth at infinity.*

A set $\mathscr{A} \subset \mathscr{P}_2(X)$ is *2-bounded* if

$$\sup_{\mu \in \mathscr{A}} \int |x|^2 d\mu(x) < +\infty.$$

Obviously a 2-uniformly integrable set is 2-bounded. It is easy to prove that a set is 2-bounded if and only if

$$\sup_{\mu \in \mathscr{A}} \int \varphi d\mu < +\infty,$$

for any φ with quadratic growth at infinity. Moreover for a 2-bounded sequence $(\mu_n) \subset \mathscr{P}_2(X)$ weakly converging to $\mu \in \mathscr{P}(X)$ it holds

$$\int \varphi d\mu = \lim_n \int \varphi d\mu_n, \tag{1.7}$$

for any continuous function φ with less than quadratic growth, i.e. a function such that $\varphi(x)/(1 + |x|^2) \to 0$ as $|x| \to +\infty$.

Proposition 1.5. *Let* $\mathscr{A} \subset \mathscr{P}_2(\boldsymbol{X})$ *be a set of probability measures on a product space* $\boldsymbol{X} = \prod X_i$, $i = 1, \dots, N$. *Then* $\mathscr{A}$ *is 2-uniformly integrable (2-bounded) if and only if so are the sets* $\pi^i_\#(\mathscr{A}) \subset X_i$, $i = 1, \dots, N$.

Proof. The assertions about 2-boundedness follow from

$$\int_{\boldsymbol{X}} \sum_{i=1}^N |x_i|^2 d\boldsymbol{\mu} = \sum_{i=1}^N \int_{X_i} |x_i|^2 d\pi^i_\# \boldsymbol{\mu},$$

so we turn to the 2-uniform integrability. Since

$$\int_{\boldsymbol{X} \setminus \prod K_i} \sum_{i=1}^N |x_i|^2 d\boldsymbol{\mu} \geq \sum_{i=1}^N \int_{X \setminus (\pi^i)^{-1}(K_i)} |x_i|^2 d\boldsymbol{\mu} = \sum_{i=1}^N \int_{X \setminus (\pi^i)^{-1}(K_i)} |x_i|^2 d\pi^i_\# \boldsymbol{\mu}$$

for any $i = 1, \dots, N$, the 2-uniform integrability of $\mathscr{A}$ implies the one of $\pi^i_\#(\mathscr{A})$ for any $i = 1, \dots, N$. For the other implication, by induction we can assume $N = 2$. Fix R, let $B^i_R \subset X_i$, $i = 1, 2$, be the balls with center in the origin and radius R and use the set equality

$$\boldsymbol{X} \setminus (B^1_R \times B^2_R) = \Big((X_1 \setminus B^1_R) \times B^2_R\Big) \bigcup \Big(B^1_R \times (X_2 \setminus B^2_R)\Big) \bigcup \Big((X_1 \setminus B^1_R) \times (X_2 \setminus B^2_R)\Big),$$

to easily obtain

$$\int_{\boldsymbol{X} \setminus (B^1_R \times B^2_R)} |x_1|^2 + |x_2|^2 d\boldsymbol{\mu} \leq 3 \int_{X_1 \setminus B^1_R} |x_1|^2 d\pi^1_\# \boldsymbol{\mu} + 3 \int_{X_2 \setminus B^2_R} |x_2|^2 d\pi^2_\# \boldsymbol{\mu},$$

from which the thesis follows. □

3. Dudley's lemma

A fundamental tool we will need in the sequel is Dudley's lemma. In this section we will prove it as a consequence of the disintegration theorem (recall that the spaces we are dealing with are Euclidean). Recall that a map $x_1 \in X_1 \to \mu_x \in \mathscr{P}(X_2)$ is called a Borel map if so is the map $x_1 \in X_1 \to \mu_{x_1}(A)$ for any open set $A \subset X_2$. Given such a Borel map and a measure $\mu \in \mathscr{P}(X_1)$, it is easy to see that the formula

$$I(\varphi(x_1, x_2)) := \int \int \varphi(x_1, x_2) d\mu_{x_1}(x_2) d\mu(x_1),$$

defines a continuous positive linear functional on $C_c(X_1 \times X_2)$ with norm 1, which is therefore a probability measure on $X_1 \times X_2$. It is natural to ask whether this statement has an inverse: the answer is positive and it is given by the following proposition, which is a direct consequence of the disintegration theorem.

PROPOSITION 1.6. *Let X_1, X_2 be two Euclidean spaces and $\boldsymbol{\mu} \in \mathscr{P}(X_1 \times X_2)$. Then there exists a Borel map $x_1 \to \mu_{x_1}$ on X_1 with values in $\mathscr{P}(X_2)$ such that*

$$\int \varphi(x_1, x_2) d\boldsymbol{\mu}(x_1, x_2) = \int \int \varphi(x_1, x_2) d\mu_{x_1}(x_2) d\pi^1_\# \boldsymbol{\mu},$$

for every $\varphi \in C_c(X_1 \times X_2)$. Moreover the family μ_{x_1} is uniquely determined up to equality $\pi^1_\# \boldsymbol{\mu}$ a.e..

We will refer to the family μ_{x_1} as the disintegration of $\boldsymbol{\mu}$ with respect to its first marginal (or w.r.t. the first variable) and write $d\boldsymbol{\mu}(x_1, x_2) = d\pi^1_\# \boldsymbol{\mu}(x_1) \otimes d\boldsymbol{\mu}_{x_1}(x_2)$ or $d\boldsymbol{\mu} = d\pi^1_\# \boldsymbol{\mu} \otimes d\boldsymbol{\mu}_{x_1}$ if no ambiguity occurs.

Let us now turn to the main result of this section.

LEMMA 1.7 (Dudley). *Let X_1, X_2, X_3 be three Euclidean spaces and let $\boldsymbol{\mu}^{1,2} \in \mathscr{P}(X_1 \times X_2)$, $\boldsymbol{\mu}^{1,3} \in \mathscr{P}(X_1 \times X_3)$ be two measures such that*

$$\pi^1_\# \boldsymbol{\mu}^{1,2} = \pi^1_\# \boldsymbol{\mu}^{1,3}. \tag{1.8}$$

Then there exists a plan $\boldsymbol{\mu}^{1,2,3} \in \mathscr{P}(X_1 \times X_2 \times X_3)$ such that

$$\pi^{1,2}_\# \boldsymbol{\mu}^{1,2,3} = \boldsymbol{\mu}^{1,2}, \tag{1.9a}$$

$$\pi^{1,3}_\# \boldsymbol{\mu}^{1,2,3} = \boldsymbol{\mu}^{1,3}. \tag{1.9b}$$

Proof. Observe that the condition (1.8) is necessary to the existence of $\boldsymbol{\mu}^{1,2,3}$. To show the sufficency let $\mu := \pi^1_\# \boldsymbol{\mu}^{1,2} = \pi^1_\# \boldsymbol{\mu}^{1,3}$ be the common marginal and $\boldsymbol{\mu}^{1,2}_{x_1}, \boldsymbol{\mu}^{1,3}_{x_1}$ be the disintegrations of $\boldsymbol{\mu}^{1,2}, \boldsymbol{\mu}^{1,3}$ w.r.t. μ. It is easy to check that the measure $\boldsymbol{\mu}^{1,2,3}$ defined by

$$d\boldsymbol{\mu}^{1,2,3}(x_1, x_2, x_3) := d\mu(x_1) \otimes d\Big(\boldsymbol{\mu}^{1,2}_{x_1}(x_2) \otimes \boldsymbol{\mu}^{1,3}_{x_1}(x_3)\Big) \tag{1.10}$$

satisfies the condition. □

Observe that if a measure $\boldsymbol{\mu}^{1,2,3}$ satisfies (1.9a), (1.9b) then it has to hold

$$\pi^2_\#\boldsymbol{\mu}^{1,2,3}_{x_1} = \boldsymbol{\mu}^{1,2}_{x_1},$$
$$\pi^3_\#\boldsymbol{\mu}^{1,2,3}_{x_1} = \boldsymbol{\mu}^{1,3}_{x_1}.$$

Together with the uniqueness result stated in 1.6 this fact implies that if either $\boldsymbol{\mu}^{1,2}$ or $\boldsymbol{\mu}^{1,3}$ is induced by a map on X_1, then there is a *unique* 3-plan satisfying the conditions of 1.7. More explicitely if $\boldsymbol{\mu}^{1,2} = (Id, T)_\#\mu$ then the unique $\boldsymbol{\mu}^{1,2,3}$ is given by

$$d\boldsymbol{\mu}^{1,2,3}(x_1, x_2, x_3) = d\mu(x_1) \otimes d\Big(\delta_{T(x_1)} \otimes \boldsymbol{\mu}^{1,3}_{x_1}(x_3)\Big).$$

It is usually said that Dudley's lemma allows to make the *composition* of plans, the reason being the following. It is just a matter of notation to observe that an analogous version of Dudley lemma holds if the given measures $\boldsymbol{\mu}^{1,2}, \boldsymbol{\mu}^{2,3}$ belong to $\mathscr{P}(X_1 \times X_2), \mathscr{P}(X_2 \times X_3)$ respectively and equation (1.8) is replaced by

$$\pi^2_\#\boldsymbol{\mu}^{1,2} = \pi^2_\#\boldsymbol{\mu}^{2,3}.$$

In this case if $\boldsymbol{\mu}^{1,2} = (Id, T)_\#\mu$ and $\boldsymbol{\mu}^{2,3} = (Id, S)_\#(T_\#\mu)$ for some measure $\mu \in \mathscr{P}(X_1)$ then the unique 3-plan with marginals $\boldsymbol{\mu}^{1,2}, \boldsymbol{\mu}^{2,3}$ is given by $\boldsymbol{\mu}^{1,2,3} = (Id, T, (S \circ T))_\#\mu$ and it satisfies $\pi^{1,3}_\#\boldsymbol{\mu}^{1,2,3} = (Id, (S \circ T))_\#\mu$, which is exactly the plan induced by the composition of the two given maps.

4. On the solution of the optimal transport problem

In this section we collect the basic results we need to introduce the Wasserstein distance. It is not our purpose to describe accurately how to solve the optimal transport problem neither the techniques used and we do not intend to state the results in their maximum generality. We simply list the main definitions and theorems in the form and setting we are interested in, and we refer to [**7**],[**11**], [**45**], [**15**] for a more detailed analysis.

We will mostly work with 2-bounded sets, therefore in the following we will not indicate if weak convergence is done w.r.t. C_b or C_c, exceptions will be indicated.

The optimal transport problem is the following: given two measures $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ find the infimum of

$$\mathcal{A}dm(\mu,\nu) \ni \boldsymbol{\gamma} \to \int |x_1 - x_2|^2 d\boldsymbol{\gamma}(x_1, x_2), \tag{1.11}$$

and answer to questions like: does there exist a minimizing plan? Is it unique? Is it induced by a map?

A standard tightness-semicontinuity argument ensures that there always exists a plan $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu,\nu)$ which attains the minimum, and simple counterexamples show that in general this plan need not to be unique.

We will denote with $\mathcal{O}pt(\mu,\nu)$ the set of *optimal plans* for the couple μ, ν:

$$\text{(1.12)}\quad \boldsymbol{\gamma} \in \mathcal{O}pt(\mu,\nu) \Leftrightarrow$$
$$\boldsymbol{\gamma} \in \mathcal{A}dm(\mu,\nu) \text{ and } \int |x_1 - x_2|^2 d\boldsymbol{\gamma} = \inf_{\boldsymbol{\gamma}' \in \mathcal{A}dm(\mu,\nu)} \int |x_1 - x_2|^2 d\boldsymbol{\gamma}'.$$

Note that by weak lower semicontinuity of $\boldsymbol{\gamma} \to \int |x_1 - x_2|^2 d\boldsymbol{\gamma}$ it follows that $\mathcal{O}pt(\mu, \nu)$ is weakly closed.

In order to understand the main properties of optimal plans we need the following definition.

Definition 1.8. *A set $A \subset (\mathbb{R}^d_1 \times \mathbb{R}^d_2)$ is* cyclically monotone *if for every $N \geq 1$ and any choice of $(x^i_1, x^i_2) \in A$, $i = 1, \ldots, N$, it holds*

$$\sum_{i=1}^{N} \langle x^i_1, x^i_2 \rangle \geq \sum_{i=1}^{N} \langle x^i_1, x^{i+1}_2 \rangle, \tag{1.13}$$

where $x^{N+1}_2 = x^1_2$.

It can be proved that the following theorem holds .

Theorem 1.9. *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^d_2)$ be a plan and $\mu := \pi^1_\# \boldsymbol{\gamma}, \nu := \pi^2_\# \boldsymbol{\gamma}$ be its marginals. Then $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu, \nu)$ if and only if its support is cyclically monotone.*

Remark 1.10. *The following property holds (a little bit stronger then the one stated in the previous theorem): for any $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ the set*

$$\overline{\bigcup_{\boldsymbol{\gamma} \in \mathcal{O}pt(\mu,\nu)} \operatorname{supp} \boldsymbol{\gamma}}, \tag{1.14}$$

is cyclically monotone. Indeed it is sufficient to consider a weakly dense countable subset $\{\boldsymbol{\gamma}_n\}_{n \in \mathbb{N}}$ of $\mathcal{O}pt(\mu, \nu)$, to consider the (optimal) plans $\boldsymbol{\gamma}'_N := 1/N \sum_1^N \boldsymbol{\gamma}_n$, to notice that $\operatorname{supp} \boldsymbol{\gamma}'_N = \cup_1^N \operatorname{supp} \boldsymbol{\gamma}_n$ and to pass to the limit to get (1.14).

A simple but of great importance consequence of this theorem is the following stability result.

Proposition 1.11. *Let $(\mu_n), (\nu_n) \subset \mathscr{P}_2(\mathbb{R}^d)$ be two sequences of measures weakly converging to μ, ν respectively and let $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu_n, \nu_n)$ be any choice of optimal plans. Then the sequence $(\boldsymbol{\gamma}_n)$ is tight and any weak accumulation point $\boldsymbol{\gamma}$ belongs to $\mathcal{O}pt(\mu, \nu)$.*

Proof. The tightness of the sequence follows from proposition 1.3 and remark 1.1 ensures that any accumulation point $\boldsymbol{\gamma}$ belongs to $\mathcal{A}dm(\mu, \nu)$. Now suppose that $\boldsymbol{\gamma}_n \rightharpoonup \boldsymbol{\gamma}$, observe that under this hypothesis the support of $\boldsymbol{\gamma}$ is contained in the Kuratowski lower limit of the supports of $\boldsymbol{\gamma}_n$ (i.e. any point of $\operatorname{supp} \boldsymbol{\gamma}$ can be approximated by points in $\operatorname{supp} \boldsymbol{\gamma}_n$): the continuity of the scalar product ensures that $\operatorname{supp} \boldsymbol{\gamma}$ is cyclically monotone, too. □

The following classical result of Rockafellar gives a very interesting characterization of cyclically monotone sets.

Theorem 1.12 (Rockafellar). *A subset $A \subset \mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone if and only if it is contained in the graph of the subdifferential of a convex function.*

From this theorem it follows the well known result of Brenier on the existence of optimal transport maps (i.e. maps which induce optimal transport plans). Recall that a set $C \subset \mathbb{R}^d$ is said to be $\mathcal{H}^k$ rectifiable if there exists a sequence f_n of Lipschitz maps from $\mathbb{R}^k$ to $\mathbb{R}^d$ such that

$$H^k\left(C \setminus \bigcup_n f_n(\mathbb{R}^k)\right) = 0,$$

where $\mathcal{H}^k$ is the $k-$Hausdorff measure, and the following property of convex functions (see [**3**]).

THEOREM 1.13. *Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then the set of points of non-differentiabilty is $\mathcal{H}^{d-1}$ rectifiable.*

DEFINITION 1.14 (Regular measures). *The set of regular measures on $\mathbb{R}^d$ is the subset of $\mathscr{P}_2(\mathbb{R}^d)$ defined as follows*

$$\mathscr{P}_2^r(\mathbb{R}^d) := \left\{\mu \in \mathscr{P}_2(\mathbb{R}^d) \ : \ \mu(C) = 0, \ \textit{for any } \mathcal{H}^{d-1} \textit{ rectifiable set } C\right\} \tag{1.15}$$

THEOREM 1.15 (Brenier). *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be given measures and suppose that μ is regular. Then there exists a unique optimal plan and this plan is induced by the gradient of a convex function.*

Proof. Use remark 1.10 and Rockafellar theorem to obtain the existence of a convex function φ such that every optimal plan is concentrated on the graph of $\partial^-\varphi$ (the subdifferential of φ). Now observe that since μ is regular $\partial^-\varphi(x)$ is a singleton for μ-a.e. x, therefore the disintegration of any optimal plan $\boldsymbol{\gamma}$ is given by $d\boldsymbol{\gamma}(x,y) = d\mu(x) \times \delta_{\nabla\varphi(x)}dy$. This implies that the optimal plan is unique and it is given by $(Id, \nabla\varphi)_\#\mu$. □

REMARK 1.16 (Discrete measures). *Another interesting case in which there is existence of an optimal transport map, although not necessarily unique, is the one in which the two measures μ and ν are both discrete, with finite support of the same cardinality, say N, and each point has mass $1/N$. Indeed, it is not hard to see that under these assumptions, the extremal points of the closed convex set of admissible plans are given exactly by those plans induced by maps. From the linearity of the cost, the claim follows.*

From now on, whenever $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$ is a regular measure, we will indicate by T_μ^ν the optimal transport map from μ to $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ given by the previous theorem. Observe that if ν is regular too, the uniqueness of the optimal plan ensures

$$\begin{aligned} T_\nu^\mu \circ T_\mu^\nu &= Id, \qquad \mu - a.e., \\ T_\mu^\nu \circ T_\nu^\mu &= Id, \qquad \nu - a.e.. \end{aligned}$$

In general, however, the group property does not hold:

$$T_\nu^\sigma \circ T_\mu^\nu \neq T_\mu^\sigma, \tag{1.16}$$

due to the fact that a composition of monotone maps on $\mathbb{R}^d$, $d > 1$, need not be monotone. As we will see in the next chapter, equation (1.16) is the one which in some sense is responsible of the non-flatness of the *manifold* $\mathscr{P}_2(\mathbb{R}^d)$, see proposition 2.13 and equations (2.19), (2.20).

CHAPTER 2

The distance W and its geometry

1. The distance W

In this chapter we define the distance W and we investigate the first properties of the space $(\mathscr{P}_2(\mathbb{R}^d), W)$ from both topological and geometrical viewpoints.

DEFINITION 2.1. *Let $\mu_1, \mu_2 \in \mathscr{P}_2(\mathbb{R}^d)$ be two measures. The Wasserstein distance $W(\mu_1, \mu_2)$ between them is defined by*

$$W^2(\mu_1, \mu_2) := \inf_{\boldsymbol{\gamma} \in \mathcal{A}dm(\mu_1, \mu_2)} \int |x_1 - x_2|^2 d\boldsymbol{\gamma}(x_1, x_2).$$

We will call the number $\sqrt{\int |x_1 - x_2|^2 d\boldsymbol{\gamma}}$ the *cost* of the plan $\boldsymbol{\gamma}$.

The discussion of the previous chapter ensures that infimum is always attained: recall that $\mathcal{O}pt(\mu_1, \mu_2)$ is the set of those plans that reach the minimum.

In order to let the notation be not too heavy, from now on we will indicate by $\|f(\boldsymbol{x})\|_{\boldsymbol{\gamma}}$ the L^2 norm w.r.t. the measure $\boldsymbol{\gamma}$ of the function f whose domain contains the support of $\boldsymbol{\gamma}$: for instance

$$\|x_1 - x_2\|_{\boldsymbol{\gamma}} = \sqrt{\int |\pi^1 - \pi^2|^2 d\boldsymbol{\gamma}} = \sqrt{\int |x_1 - x_2|^2 d\boldsymbol{\gamma}(x_1, x_2)}.$$

The set $\mathcal{O}pt(\mu_0, \mu_1)$ is therefore defined as

$$\boldsymbol{\gamma} \in \mathcal{O}pt(\mu_0, \mu_1) \quad \Leftrightarrow \quad \boldsymbol{\gamma} \in \mathcal{A}dm(\mu_1, \mu_2) \quad \text{and} \quad \|x_1 - x_2\|_{\boldsymbol{\gamma}} = W(\mu_0, \mu_1). \tag{2.1}$$

In the following we will extensively use the simple and crucial inequality

$$W(T_\# \mu, S_\# \mu) \le \|T - S\|_\mu, \tag{2.2}$$

which is a consequence of $(T, S)_\# \mu \in \mathcal{A}dm(T_\# \mu, S_\# \mu)$.

THEOREM 2.2 (W is a distance). *The map $(\mu, \nu) \to W(\mu, \nu)$ is a distance on $\mathscr{P}_2(\mathbb{R}^d)$.*

Proof. It is clear that W is symmetric. Considering the plan $(Id, Id)_\# \mu$ we get that $W(\mu, \mu) = 0$. Conversely if $W(\mu, \nu) = 0$, then there exists a plan $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$ such that $\int |x_1 - x_2|^2 d\boldsymbol{\gamma} = 0$: this means that the maps π^1, π^2 are equal in $L^2_{\boldsymbol{\gamma}}$, and therefore $\mu = \pi^1_\# \boldsymbol{\gamma} = \pi^2_\# \boldsymbol{\gamma} = \nu$. To prove the triangular inequality we need Dudley's lemma. Consider three measures μ_1, μ_2, μ_3 and two optimal plans $\boldsymbol{\gamma}^{1,2} \in \mathcal{O}pt(\mu_1, \mu_2), \boldsymbol{\gamma}^{2,3} \in$

$\mathcal{O}pt(\mu_2,\mu_3)$: then we know that there exists a 3-plan $\boldsymbol{\gamma}$ such that

$$\pi^{1,2}_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}^{1,2},$$
$$\pi^{2,3}_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}^{2,3}.$$

Since $\pi^{1,3}_{\#}\boldsymbol{\gamma} \in \mathcal{A}dm(\mu_1,\mu_3)$, the conclusion follows by

$$\begin{aligned}
W(\mu_1,\mu_3) \leq & \sqrt{\int |x_1-x_3|^2 d\pi^{1,3}_{\#}\boldsymbol{\gamma}} = \sqrt{\int |x_1-x_3|^2 d\boldsymbol{\gamma}} \\
\leq & \sqrt{\int |x_1-x_2|^2 d\boldsymbol{\gamma}} + \sqrt{\int |x_2-x_3|^2 d\boldsymbol{\gamma}} \\
= & \sqrt{\int |x_1-x_2|^2 d\pi^{1,2}_{\#}\boldsymbol{\gamma}} + \sqrt{\int |x_2-x_3|^2 d\pi^{2,3}_{\#}\boldsymbol{\gamma}} \\
= & W(\mu_1,\mu_2) + W(\mu_2,\mu_3).
\end{aligned}$$

To get that W is real valued, i.e. it never attains the value $+\infty$ use the triangular inequality and the remark below. □

REMARK 2.3 (Boundedness). *For any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ the plan $(Id,0)_{\#}\mu$ is the unique element of $\mathcal{A}dm(\mu,\delta_0)$ and therefore it is optimal. In particular we get that*

$$W^2(\mu,\delta_0) = \|x\|^2_{\mu},$$

and that a subset of $\mathscr{P}_2(\mathbb{R}^d)$ is bounded if and only if it is 2-bounded.

We collect in the next propositions the first properties about convergence and compactness.

PROPOSITION 2.4. *Let $(\mu^1_n),(\mu^2_n)$ be two sequences in $\mathscr{P}_2(\mathbb{R}^d)$ weakly converging in duality with C_b to μ^1,μ^2 respectively. Then it holds*

$$W(\mu^1,\mu^2) \leq \varliminf_{n\to+\infty} W(\mu^1_n,\mu^2_n).$$

Proof. Without loss of generality we can assume that the $\varliminf$ is a limit. Choosing $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu^1_n,\mu^2_n)$, proposition 1.3 ensures that this sequence is tight and so from every subsequence we can extract a subsequence $(\boldsymbol{\gamma}_{n_k})$ weakly converging to some $\boldsymbol{\gamma}$. By remark 1.1 we have $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu^1,\mu^2)$; since $|x_1-x_2|^2$ is continuous and bounded from below we get

$$W(\mu^1,\mu^2) \leq \|x_1-x_2\|_{\boldsymbol{\gamma}} \leq \varliminf_{k} \|x_1-x_2\|_{\boldsymbol{\gamma}_{n_k}}.$$

□

PROPOSITION 2.5. *The space $(\mathscr{P}_2(\mathbb{R}^d),W)$ is a complete and separable metric space. A sequence (μ_n) $W-$converges to μ if and only if it is 2-uniformly integrable and weakly convergent to μ in duality with C_b. A subset of $\mathscr{P}_2(\mathbb{R}^d)$ is relatively compact if and only if it is 2-uniformly integrable.*

Proof. Let us first check the completeness. Choose a Cauchy sequence (μ_n) and observe that such a sequence is bounded and therefore tight. Find a subsequence (μ_{n_k}) weakly converging to some μ and observe that

$$\lim_n W(\mu_n, \mu) \leq \lim_n \varliminf_k W(\mu_n, \mu_{n_k}) = 0.$$

Now suppose that $W(\mu_n, \mu) \to 0$, we want to prove that μ_n weakly converge to μ and that it is 2-uniformly integrable. Being the sequence bounded, we can find a subsequence μ_{n_k} weakly converging to some ν: the proof we just did on Cauchy sequences weakly converging gives $\nu = \mu$, and since this result is independent on the subsequence chosen we get the weak convergence of the whole sequence μ_n. To prove the 2-uniform integrability use proposition 1.4 and remark 2.3 to get

$$\|x\|_\mu = W(\mu, \delta_0) = \lim_n W(\mu_n, \delta_0) = \lim_{n\to\infty} \|x\|_{\mu_n}.$$

To show that weak convergence and 2-uniform integrability imply $W-$ convergence choose plans $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu_n, \mu)$ and observe that proposition 1.5 ensures the 2-uniform integrability of the sequence $\boldsymbol{\gamma}_n$. Choose a subsequence $\boldsymbol{\gamma}_{n_k}$ weakly converging to some $\boldsymbol{\gamma}$ and note that proposition 1.11 ensures that $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu, \mu)$, and therefore $\boldsymbol{\gamma} = (Id, Id)_\# \mu$, from which

$$\lim_{k\to\infty} W(\mu, \mu_{n_k}) = \lim_{k\to\infty} \|x_1 - x_2\|_{\boldsymbol{\gamma}_{n_k}} = \|x_1 - x_2\|_{\boldsymbol{\gamma}} = 0.$$

As usual, this is independent on the subsequence chosen and so we get $W(\mu, \mu_n) \to 0$. The characterization of relatively compact sets is now straightforward. In order to get a countable dense subset consider the set of convex combinations of Dirac masses at rational points with rational coefficients. □

It is important to note that the space $(\mathscr{P}_2(\mathbb{R}^d), W)$ is *not* locally compact: there exists no measure (not even one with compact support) with a compact neighborhood. This is essentially due to the fact that 2-boundedness and 2-uniform integrability are different concepts. In order to give an explicit example fix $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $\varepsilon > 0$ and consider the measures

$$\mu_n := \left(1 - \frac{1}{n^2}\right)\mu + \frac{1}{n^2}\delta_{x_n},$$

where x_n satisfies $|x_n| = n\varepsilon - \|x\|_\mu$. We claim that all of these measures belong to the ball with centre in μ and radius ε and that there is no subsequence convergent w.r.t. W. For the first part of the assertion observe that the plan

$$\boldsymbol{\gamma}_n := (Id, Id)_\# \left((1 - \frac{1}{n^2})\mu\right) + \frac{1}{n^2}(Id, C_{x_n})_\# \mu,$$

belongs to $\mathcal{A}dm(\mu, \mu_n)$ (where C_x is the constant function on $\mathbb{R}^d$ whose value is x) and calculate

$$W(\mu, \mu_n) \leq \|x_1 - x_2\|_{\boldsymbol{\gamma}_n} = \frac{1}{n}\|x - x_n\|_\mu \leq \frac{1}{n}\left(\|x\|_\mu + |x_n|\right) = \varepsilon.$$

The second part follows from

$$\varliminf_{n\to\infty} \int_{\mathbb{R}^d \setminus B_R} |x|^2 d\mu_n \geq \varliminf_{n\to\infty} \frac{|x_n|^2}{n^2} = \varepsilon^2.$$

2. Interpolation and geodesics

The interpolation between measures we define in this section is a fundamental tool for the next chapters: as we will see it is the natural interpolation that arises in the study of Wasserstein distance.

DEFINITION 2.6. *Let $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ be two measures and let $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu_0, \mu_1)$ be a plan. The interpolation from μ_0 to μ_1 through $\boldsymbol{\gamma}$ is the curve $[\boldsymbol{\gamma}](t) : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ defined as*

$$[\boldsymbol{\gamma}](t) := \Big((1-t)\pi^1 + t\pi^2\Big)_{\#}\boldsymbol{\gamma}. \tag{2.3}$$

Observe that $[\boldsymbol{\gamma}](0) = \mu_0$ and $[\boldsymbol{\gamma}](1) = \mu_1$.

It may be noticed that the measures μ_0, μ_1 are not needed in the definition, since the curve depends only on the plan $\boldsymbol{\gamma}$, however we kept them because in the applications the extreme points of the curve are usually given and one has to look for a suitable admissible plan to get the right interpolation curve.

If $\boldsymbol{\gamma} = (Id, T)_{\#}\mu_0$ is induced by a map equation (2.3) reduces to

$$[\boldsymbol{\gamma}](t) := \big((1-t)Id + tT\big)_{\#}\mu_0. \tag{2.4}$$

PROPOSITION 2.7. *For any $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ and any $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu_0, \mu_1)$ the curve $t \to [\boldsymbol{\gamma}](t)$ is Lipschitz with constant less or equal than the cost of $\boldsymbol{\gamma}$.*

Proof. It is sufficient to observe that for every $0 \le s < t \le 1$ the plan

$$\boldsymbol{\gamma}_{s,t} := ((1-s)\pi^1 + s\pi^2, (1-t)\pi^1 + t\pi^2)_{\#}\boldsymbol{\gamma}$$

belongs to $\mathcal{Adm}([\boldsymbol{\gamma}](s), [\boldsymbol{\gamma}](t))$ and that

$$W^2([\boldsymbol{\gamma}](s), [\boldsymbol{\gamma}](t)) \le \|x_1 - x_2\|^2_{\boldsymbol{\gamma}_{s,t}} = (t-s)^2\|x_1 - x_2\|^2_{\boldsymbol{\gamma}}. \tag{2.5}$$

□

It is interesting to observe that the usual linear interpolation $\mu_t := (1-t)\mu_0 + t\mu_1$ still defines a continuous curve (see proposition 2.5), but in general this curve is not absolutely continuous. For istance choose $\mu_0 := \delta_0$ and $\mu_1 := \delta_1$: it is not difficult to check that

$$W\big(\mu_t, \mu_s\big) = \sqrt{|t-s|}.$$

Observe that with the same choice of μ_0, μ_1 the unique admissible plan is $\delta_{(0,1)}$ and it gives the curve $t \to \delta_t$, which is an "horizontal" interpolation between the two deltas which is radically different from the usual "vertical" one. Proposition 2.7 suggests that for $\boldsymbol{\gamma} \in \mathcal{Opt}(\mu_0, \mu_1)$ the curve $\boldsymbol{\gamma}[t]$ is a geodesic from μ_0 to μ_1. The next theorem shows that this is actually the case, and that every constant speed geodesic comes from this construction.

THEOREM 2.8. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be two measures. Then a curve $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ is a constant speed geodesic from μ to ν if and only if $\mu_t = [\boldsymbol{\gamma}](t)$ for some optimal plan $\boldsymbol{\gamma} \in \mathcal{Opt}(\mu, \nu)$.*

Proof. Suppose that $\boldsymbol{\gamma} \in \mathit{Opt}(\mu, \nu)$ and note that equation (2.5) together with equality $\|x_1 - x_2\|_{\boldsymbol{\gamma}} = W(\mu, \nu)$ gives that $[\boldsymbol{\gamma}](t)$ is a constant speed geodesic (it cannot happen that the inequality in (2.5) is strict for some $s < t$, otherwise there would be a contradiction applying the triangular inequality to the measures $\mu = \mu_0, \mu_s, \mu_t, \mu_1 = \nu$).

The converse implication is more difficult. Suppose $\mu_0 = \mu$, $\mu_1 = \nu$, fix $0 < t < 1$ and choose plans $\boldsymbol{\gamma}_{0,t}, \boldsymbol{\gamma}_{t,1}$ in $\mathit{Opt}(\mu_0, \mu_t)$, and $\mathit{Opt}(\mu_t, \mu_1)$ respectively. In order to let the proof be more clear we will use different names for the variables of different measures: x_1, x_2, x_3 will be the variables of μ_0, μ_t, μ_1 respectively.

Use Dudley's lemma to find a 3-plan $\boldsymbol{\gamma}$ such that

$$\begin{aligned} \pi^{1,2}_{\#}\boldsymbol{\gamma} &= \boldsymbol{\gamma}_{0,t}, \\ \pi^{2,3}_{\#}\boldsymbol{\gamma} &= \boldsymbol{\gamma}_{t,1}, \end{aligned}$$

and observe that it holds

$$\begin{aligned} W(\mu_0, \mu_1) \leq \|x_1 - x_3\|_{\boldsymbol{\gamma}} &\leq \|x_1 - x_2\|_{\boldsymbol{\gamma}} + \|x_2 - x_3\|_{\boldsymbol{\gamma}} \\ &= W(\mu_0, \mu_t) + W(\mu_t, \mu_1) = W(\mu_0, \mu_1), \end{aligned} \tag{2.6}$$

which implies in particular that $\pi^{1,3}_{\#}\boldsymbol{\gamma} \in \mathit{Opt}(\mu_0, \mu_1)$. Our goal is to prove that $\mu_t = [\pi^{1,3}_{\#}\boldsymbol{\gamma}](t)$. Since all the inequalities in (2.6) are equalities and the L^2 norm is strictly convex, we get that there exists some $\lambda > 0$ such that $x_1 - x_2 = \lambda(x_2 - x_3)$ for $\boldsymbol{\gamma}$−a.e. (x_1, x_2, x_3). Moreover, since $W(\mu_0, \mu_t) = \frac{t}{1-t}W(\mu_t, \mu_1)$ we have $\lambda = \frac{t}{1-t}$ and so

$$x_3 = x_2 + \frac{1-t}{t}(x_2 - x_1), \qquad \boldsymbol{\gamma} - a.e.\ (x_1, x_2, x_3). \tag{2.7}$$

This equation implies that the 3-plan $\boldsymbol{\gamma}$ is actually a function of the 2-plan $\boldsymbol{\gamma}_{0,t}$, more explicitly it holds

$$\boldsymbol{\gamma} = \left(\pi^1, \pi^2, \frac{1}{t}\pi^2 - \frac{1-t}{t}\pi^1\right)_{\#}\boldsymbol{\gamma}_{0,t}. \tag{2.8}$$

Since it holds $\pi^{2,3}_{\#}\boldsymbol{\gamma} = \boldsymbol{\gamma}_{t,1}$ we get that $\boldsymbol{\gamma}_{t,1}$ is a function of $\boldsymbol{\gamma}_{0,t}$: since it was chosen independently we obtain that $\boldsymbol{\gamma}_{t,1}$ is unique. Arguing in similar way we get that $\boldsymbol{\gamma}_{0,t}$ is unique, too.

Up to now we know that $[\pi^{1,3}_{\#}\boldsymbol{\gamma}](t) = \mu_t$ (it follows from (2.8)) and we want to prove that the same result holds for $t' \neq t$. Suppose $t' < t$ (the other case is similar) and consider the curve $s \to \mu_{st}$: it is clearly a constant speed geodesic from μ_0 to μ_t. Choosing $s = t'/t$, a plan $\boldsymbol{\gamma}_{0,t'} \in \mathit{Opt}(\mu_0, \mu_{t'})$ (actually *the* plan) and arguing as before, the uniqueness of the optimal plan for (μ_0, μ_t) gives the formula

$$\boldsymbol{\gamma}_{0,t} = \left(\pi^1, \frac{1}{s}\pi^2 - \frac{1-s}{s}\pi^1\right)_{\#}\boldsymbol{\gamma}_{0,t'}. \tag{2.9}$$

We conclude with the following calculation

$$\begin{aligned}[\boldsymbol{\gamma}](t') =&[\boldsymbol{\gamma}](st) = ((1-st)\pi^1 + st\pi^2)_\# \boldsymbol{\gamma} \\ =&((1-st)\pi^1 + s(\pi^2 - (1-t)\pi^1))_\# \boldsymbol{\gamma}_{0,t} \\ =&((1-s)\pi^1 + s\pi^2)_\# \boldsymbol{\gamma}_{0,t} = \pi^2_\# \boldsymbol{\gamma}_{0,t'} = \mu_{t'}.\end{aligned}$$

□

The previous proposition characterizes geodesics parametrized in $[0,1]$, that is geodesics whose (constant) velocity is equal to the distance between the extreme measures. If we want to drop the restriction to $[0,1]$ it is sufficient to observe that $\mu_t : [0,\bar{t}] \to \mathscr{P}_2(\mathbb{R}^d)$ is a constant speed geodesic from μ to ν if and only if $t \to \mu_{t\bar{t}}$ is a constant speed geodesic parametrized on $[0,1]$, that is, if and only if $\mu_{t\bar{t}} = [\boldsymbol{\gamma}](t)$ for some optimal plan $\boldsymbol{\gamma} \in \mathit{Opt}(\mu,\nu)$, or, equivalently, iff

$$\mu_t = [\boldsymbol{\gamma}](t/\bar{t}) = \left(\left(1 - \frac{t}{\bar{t}}\right)\pi^1 + \frac{t}{\bar{t}}\pi^2\right)_\# \boldsymbol{\gamma}.$$

Introducing the plan

$$\boldsymbol{\gamma}_{\bar{t}} := \left(\pi^1, \pi^1 + \frac{\pi^2 - \pi^1}{\bar{t}}\right)_\# \boldsymbol{\gamma}, \tag{2.10}$$

the previous equation may be written as

$$\mu_t = [\boldsymbol{\gamma}_{\bar{t}}](t).$$

Since formula (2.10) is invertible, we get

$$\boldsymbol{\gamma} = \left(\pi^1, \pi^1 + \bar{t}(\pi^2 - \pi^1)\right)_\# \boldsymbol{\gamma}_{\bar{t}},$$

and we obtain the following result:

Corollary 2.9. *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a measure. Then $t \to \mu_t$ is a constant speed geodesic starting from μ parametrized in some right neighborhood of 0 if and only if $\mu_t = [\boldsymbol{\gamma}](t)$ for some plan $\boldsymbol{\gamma}$ such that*

$$\bar{\boldsymbol{\gamma}} := \left(\pi^1, \pi^1 + \bar{t}(\pi^2 - \pi^1)\right)_\# \boldsymbol{\gamma} \text{ is optimal for some } \bar{t} > 0.$$

Moreover for $t < \bar{t}$ the unique optimal plan from μ to μ_t is given by

$$(\pi^1, (1-t)\pi^1 + t\pi^2)_\# \boldsymbol{\gamma} = \left(\pi^1, \pi^1 + \frac{t}{\bar{t}}(\pi^2 - \pi^1)\right)_\# \bar{\boldsymbol{\gamma}}.$$

Remark 2.10 (The regular case). *Suppose that the measure μ is regular. Then theorems 2.8 and 1.15 ensure that for any $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a unique constant speed geodesic $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ from μ to ν, and this geodesic is given by (see also formula* (2.4)*):*

$$\mu_t = (Id + t(T_\mu^\nu - Id))_\# \mu.$$

Moreover the previous corollary gives that $t \to \mu_t$ is a constant speed geodesic starting from μ parametrized in some right neighborhood of 0 if and only if

$$\mu_t = \left(Id + \frac{t}{\bar{t}}(T - Id)\right)_\# \mu, \tag{2.11}$$

for some optimal map T and some $\bar{t} > 0$. In this case it holds

$$T_\mu^{\mu_t} = Id + \frac{t}{\bar{t}}(T - Id), \quad \text{for } t < \bar{t}. \tag{2.12}$$

In the next proposition we collect the first properties of geodesics.

PROPOSITION 2.11. *Let $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ be a geodesic.*

i) *For any $t \in (0,1)$ the sets $\mathit{Opt}(\mu_0, \mu_t)$, $\mathit{Opt}(\mu_t, \mu_1)$ contain only one element. Both of the optimal plan are induced by maps from μ_t. Those maps are Lipschitz with constant at most $\frac{1}{1-t}$ and $\frac{1}{t}$ respectively.*
ii) *The range of two different geodesics with the same extreme points are disjoint (except for the extreme points themselves, of course).*
iii) *If either μ_0 or μ_1 is regular then the geodesic $t \to \mu_t$ is unique. Moreover in this case μ_t is regular for every $t \in (0,1)$.*

Proof. To prove *i)* recall that we already observed that for a given geodesic $\boldsymbol{\gamma}[t]$, $0 \le t \le 1$ (where $\boldsymbol{\gamma}$ is an optimal plan) and for a fixed $0 < t < 1$ there exists only one optimal plan between $\boldsymbol{\gamma}[0]$ and $\boldsymbol{\gamma}[t]$ given by $\boldsymbol{\gamma}_{0,t} := (\pi^1, (1-t)\pi^1 + t\pi^2)_\# \boldsymbol{\gamma}$. To prove that $\boldsymbol{\gamma}_{0,t}$ is induced by a Lipschitz map from $\boldsymbol{\gamma}[t]$ observe that any point (z, w) in its support may be written as $(x, (1-t)x + ty)$ for some point (x, y) in the support of $\boldsymbol{\gamma}$. Choose now any couple of points $(z_1, w_1), (z_2, w_2)$ in the support of $\boldsymbol{\gamma}[t]$ and find points $(x_1, y_1), (x_2, y_2)$ in the support of $\boldsymbol{\gamma}$ such that $(z_i, w_i) = (x_i, (1-t)x_i + ty_i)$, $i = 1, 2$. Use the monotonicity property of the support of $\boldsymbol{\gamma}$ to get

$$\begin{aligned}(1-t)^2|z_1 - z_2|^2 \le& \langle (1-t)(x_1 - x_2), (1-t)(x_1 - x_2) + t(y_1 - y_2)\rangle \\ \le& (1-t)|x_1 - x_2||(1-t)(x_1 - x_2) + t(y_1 - y_2)| \\ =& (1-t)|z_1 - z_2||w_1 - w_2|.\end{aligned}$$

Statement *ii)* is another simple consequence of the proof of the previous proposition: indeed we showed that it is possible to recover an entire geodesic by knowing only one of its internal points. It is therefore impossible for two different geodesics to have a non empty intersection in their interior.

The first part of *iii)* was already noticed in remark 2.10. In order to prove the other assertion suppose that μ_0 is regular, then we know that defining $T_t := (1-t)Id + tT_{\mu_0}^{\mu_1}$ the geodesic is given by $\mu_t = (T_t)_\# \mu_0$. From point (i) we know that T_t is invertible and that T_t^{-1} is Lipschitz (with constant less than $(1-t)^{-1}$). Now let C be a $\mathcal{H}^{n-1}$ rectifiable set, since $(T_t)^{-1}$ is Lipschitz we know that $(T_t)^{-1}(C)$ is still a $\mathcal{H}^{n-1}$ rectifiable set, and so from $\mu_t(C) = \mu_0((T_t)^{-1}(C)) = 0$ we conclude. □

3. The curvature of $(\mathscr{P}_2(\mathbb{R}^d), W)$

Here we begin a deeper analysis of the geometry of the space $(\mathscr{P}_2(\mathbb{R}^d), W)$. The main result of this section is the semiconcavity inequality of W^2 along geodesics (see inequality (2.16)).

In a Euclidean setting the unique geodesic from x_0 to x_1 is given by $t \to x_t := (1-t)x_0 + tx_1$, a simple calculation shows that for any y it holds

$$|y - x_t|^2 = (1-t)|y - x_0|^2 + t|y - x_1|^2 - t(1-t)|x_0 - x_1|^2. \tag{2.13}$$

If we replace $|x-y|$ with $d(x,y)$ the objects involved make sense even in an arbitrary metric space with geodesics (in short a *length space*). Such a space is called *positively curved* (or PC space) in the sense of Aleksandrov if the inequality $\geq$ holds for any constant speed geodesic $t \to x_t$ and every $y \in E$, i.e.

$$d^2(y, x_t) \geq (1-t)d^2(y, x_0) + td^2(y, x_1) - t(1-t)d^2(x_0, x_1). \tag{2.14}$$

and it is called *non positively curved* (or NPC space) if $\leq$ always holds.

The space $(\mathscr{P}_2(\mathbb{R}^d), W)$ is a PC space, the proof is a simple consequence of the following variant of Dudley's lemma.

LEMMA 2.12. *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})$ be a plan, $t \in [0,1]$ a fixed point, and $\boldsymbol{\gamma}_t \in \mathscr{P}_2(\mathbb{R}^{2d})$ such that $\pi^1_\#\boldsymbol{\gamma}_t = [\boldsymbol{\gamma}](t)$. Then there exists a 3-plan $\boldsymbol{\mu}_t$ satisfying*

$$\pi^{1,2}_\# \boldsymbol{\mu}_t = \boldsymbol{\gamma}, \tag{2.15a}$$

$$((1-t)\pi^1 + t\pi^2, \pi^3)_\# \boldsymbol{\mu}_t = \boldsymbol{\gamma}_t. \tag{2.15b}$$

Proof. Use Dudley's lemma to find a 3-plan $\overline{\boldsymbol{\mu}}$ such that

$$\pi^{1,2}_\# \overline{\boldsymbol{\mu}} = (\pi^1, (1-t)\pi^1 + t\pi^2)_\# \boldsymbol{\gamma},$$
$$\pi^{2,3}_\# \overline{\boldsymbol{\mu}} = \boldsymbol{\gamma}_t,$$

and define

$$\boldsymbol{\mu}_t := \left(\pi^1, \pi^1 + \frac{\pi^2 - \pi^1}{t}, \pi^3\right)_\# \overline{\boldsymbol{\mu}}.$$

□

PROPOSITION 2.13 ($(\mathscr{P}_2(\mathbb{R}^d), W)$ is positively curved). *Let $\mu_0, \mu_1, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be three measures and $\boldsymbol{\gamma} \in \mathit{Opt}(\mu_0, \mu_1)$ be an optimal plan. Then*

$$W^2(\nu, [\boldsymbol{\gamma}](t)) \geq (1-t)W^2(\nu, \mu_0) + tW^2(\nu, \mu_1) - t(1-t)W^2(\mu_0, \mu_1), \ \forall t \in [0,1]. \tag{2.16}$$

Proof. Fix $t \in [0,1]$ and find, using lemma 2.12, a 3-plan $\boldsymbol{\mu}$ satisfying

$$\pi^{1,2}_\# \boldsymbol{\mu}_t = \boldsymbol{\gamma},$$
$$((1-t)\pi^1 + t\pi^2, \pi^3)_\# \boldsymbol{\mu}_t \in \mathit{Opt}([\boldsymbol{\gamma}](t), \nu).$$

Then it holds

$$\begin{aligned} W^2(\nu, [\boldsymbol{\gamma}](t)) &= \|(1-t)x_1 + tx_2 - x_3\|^2_{\boldsymbol{\mu}_t} \\ &= (1-t)\|x_1 - x_3\|^2_{\boldsymbol{\mu}_t} + t\|x_2 - x_3\|^2_{\boldsymbol{\mu}_t} - t(1-t)\|x_1 - x_2\|^2_{\boldsymbol{\mu}_t} \\ &\geq (1-t)W^2(\nu, \mu_0) + tW^2(\nu, \mu_1) - t(1-t)W^2(\mu_0, \mu_1). \end{aligned} \tag{2.17}$$

□

The proof of the previous proposition allows to understand the case of equality in (2.16). Indeed suppose that for some $t_0 \in (0,1)$ it holds

$$W^2([\boldsymbol{\gamma}](t_0), \nu) = (1-t_0)W^2(\nu, \mu_0) + t_0 W^2(\nu, \mu_1) - t_0(1-t_0)W^2(\mu_0, \mu_1),$$

then equation (2.17) shows that there exists a 3-plan $\boldsymbol{\mu}$ such that

$$\pi^{1,2}_{\#}\boldsymbol{\mu} = \boldsymbol{\gamma}, \tag{2.18a}$$

$$((1-t_0)\pi^1 + t_0\pi^2, \pi^3)_{\#}\boldsymbol{\mu} \in \mathcal{O}pt([\boldsymbol{\gamma}](t_0), \nu), \tag{2.18b}$$

$$\pi^{1,3}_{\#}\boldsymbol{\mu} \in \mathcal{O}pt(\mu_0, \nu), \tag{2.18c}$$

$$\pi^{2,3}_{\#}\boldsymbol{\mu} \in \mathcal{O}pt(\mu_1, \nu). \tag{2.18d}$$

Checking the cyclical monotonicity of the supports it is not difficult to verify that conditions (2.18c), (2.18d) imply that for every $t \in [0,1]$ the plan $((1-t)\pi^1 + t\pi^2, \pi^3)_{\#}\boldsymbol{\mu}$ has to be optimal (see also the proof of 4.25), moreover since $((1-t)\pi^1 + t\pi^2)_{\#}\boldsymbol{\mu} = [\boldsymbol{\gamma}](t)$ we get $((1-t)\pi^1 + t\pi^2, \pi^3)_{\#}\boldsymbol{\mu} \in \mathcal{O}pt([\boldsymbol{\gamma}](t), \nu)$ for any $t \in [0,1]$. In conclusion equality holds for some internal point if and only if it holds for any $t \in [0,1]$. Thus, last condition is satisfied if and only if there exists a 3-plan $\boldsymbol{\mu}$ such that

$$\begin{aligned} &\pi^{1,2}_{\#}\boldsymbol{\mu} = \boldsymbol{\gamma}, \\ &\pi^{1,3}_{\#}\boldsymbol{\mu} \in \mathcal{O}pt(\mu_0, \nu), \\ &\pi^{2,3}_{\#}\boldsymbol{\mu} \in \mathcal{O}pt(\mu_1, \nu). \end{aligned} \tag{2.19}$$

For regular measures the equations above reduce to

$$T_{\mu_0}^{\mu_1} \circ T_{\nu}^{\mu_0} = T_{\nu}^{\mu_1}. \tag{2.20}$$

It is for these reasons that at the end of the first chapter we interpreted the fact that in general $T_{\mu_0}^{\mu_1} \circ T_{\nu}^{\mu_0} \neq T_{\nu}^{\mu_1}$ as a manifestation of the curvature of $\mathscr{P}_2(\mathbb{R}^d)$.

Observe that equation (2.16) is equivalent to say that the function $W^2(\cdot, \sigma)$ is -1–concave along geodesics. We can then ask whether the same function is λ–convex for some λ. The answer is no, as shown by the following example.

EXAMPLE 2.14. *Let* $\nu := \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,-1)})$, $\mu_0 := \frac{1}{2}(\delta_{(-2,1)} + \delta_{(0,0)})$ *and* $\mu_1 := \frac{1}{2}(\delta_{(2,1)} + \delta_{(0,0)})$. *Then it is easy to see that there exists only one geodesic* μ_t *connecting* μ_0 *to* μ_1, *namely* $\mu_t := \frac{1}{2}(\delta_{(-2+2t,1-t)} + \delta_{(2t,t)})$. *Moreover it can be checked that the infimum of the optimal transport problem from* ν *to* μ_t *is always attained by one of the two maps* T_t^1, T_t^2 *defined as*

$$\begin{aligned} T_t^1(0,0) &= (-2+2t, 1-t), \qquad & T_t^1(0,-1) &= (2t,t), \\ T_t^1(0,0) &= (2t,t), & T_t^1(0,-1) &= (-2+2t, 1-t). \end{aligned}$$

Therefore the function $t \to W^2(\mu_t, \nu)$, *equal to*

$$W^2(\mu_t, \nu) = \frac{1}{2}\min\Big\{10t^2 - 8t + 6, 10t^2 - 12t + 8\Big\}, \tag{2.21}$$

has a concave cusp at $t = \frac{1}{2}$.

4. The weak topology of $\mathscr{P}_2(X)$

In this section we introduce a useful tool: the weak topology $(\mathscr{P}_2(X), \tau)$. In the following whenever we write $\mathscr{P}_2(X)$ without an explicit reference to a topology we will always assume we are working with the strong topology induced by W.

Let us start recalling the definition of *inductive limit of topologies*. Let $(\mathscr{X}_n, \tau_n)$, $n \in \mathbb{N}$, be a sequence of topological spaces such that $\mathscr{X}_n \subset \mathscr{X}_{n+1}$ with continuous inclusion for every n. The inductive limit of the topologies τ_n is the strongest topology τ on $\mathscr{X} := \cup_n \mathscr{X}_n$ which let the inclusions $\iota_n : \mathscr{X}_n \hookrightarrow \mathscr{X}$ be continuous for every $n \in \mathbb{N}$.

PROPOSITION 2.15. *It holds:*

i) *A subset $A \subset \mathscr{X}$ belongs to τ if and only if*

$$A \cap \mathscr{X}_n \in \tau_n \qquad \forall n \in \mathbb{N}. \tag{2.22}$$

The analogous property holds for closed sets, too.

ii) *Let $x \in \mathscr{X}$ be a point and N be the lowest integer such that $x \in \mathscr{X}_N$. A subset $U \subset \mathscr{X}$ is a neighborhood of x if and only if $U \cap \mathscr{X}_n$ is a neighborhood of x in $(\mathscr{X}_n, \tau_n)$ for every $n \geq N$.*

iii) *τ is Hausdorff if and only if each τ_n is Hausdorff.*

iv) *Suppose that τ is Hausdorff. A sequence $(x_n) \subset \mathscr{X}$ converges to some $x \in \mathscr{X}$ if and only if there exists N such that $x \in \mathscr{X}_N$, $x_n \in \mathscr{X}_N$ $\forall n$, and $x_n \to x$ in $(\mathscr{X}_N, \tau_N)$.*

Proof. Let $A \in \tau$ be an open set. Since $\iota_n^{-1}(A) = A \cap \mathscr{X}_n$ and ι_n is continuous we have that $A \cap \mathscr{X}_n$ is open for every n. Conversely the set identities

$$\begin{aligned} \left(\bigcup_{i\in I} A_i\right) \cap \mathscr{X}_n &= \bigcup_{i\in I} (A_i \cap \mathscr{X}_n) \qquad && \forall n \in \mathbb{N}, \\ (A_1 \cap A_2) \cap \mathscr{X}_n &= (A_1 \cap \mathscr{X}_n) \cap (A_2 \cap \mathscr{X}_n) \qquad && \forall n \in \mathbb{N}, \end{aligned}$$

imply that the set of A's which satisfy (2.22) is a topology. The statement on closed sets follows from

$$(\mathscr{X} \setminus A) \cap \mathscr{X}_n = \mathscr{X}_n \setminus A, \qquad \forall n \in \mathbb{N}.$$

ii) and *iii*) are straightforward consequences of *i*). The *if* implication of *iv*) is obvious, so we turn to the *only if*. Let $x_n \to x$ be a converging sequence in $\mathscr{X}$, clearly it is enough to show that there exists M such that $x_n \in \mathscr{X}_M$ for every n. Suppose on the contrary that this is not the case, i.e. suppose that for every k there exists $x_{n_k} \notin \mathscr{X}_k$, without loss of generality we can assume that $k \to n_k$ is an increasing sequence and $x_{n_k} \neq x$ $\forall k$. Now simply observe that the complement of $\{x_{n_k}\}_{k\in\mathbb{N}}$ is an open set (since its intersection with every $\mathscr{X}_n$ is the complement of a finite set) which contains x but contains none of the x_{n_k}. □

Observe that if one takes a sequence of sets $A_n \subset \mathscr{X}_n$ such that $A_n \in \tau_n$ for every $n \in \mathbb{N}$, it may happen that $A := \cup_n A_n$ is not an open set in $\mathscr{X}$: indeed since $A \cap \mathscr{X}_n = \cup_1^n A_i \cup ((\cup_{n+1}^\infty A_i) \cap \boldsymbol{x}_n)$, and the first term may not belong to τ_n, this set may fail to be open.

Let us now come back to our case. Define

$$\mathscr{X}_n := \left\{\mu \in \mathscr{P}(X)\ :\ \int |x|^2 d\mu \leq n\right\},$$

and let τ_n be the usual weak topology on $\mathscr{X}_n$. Clearly it holds $\mathscr{P}_2(X) = \cup \mathscr{X}_n$.

DEFINITION 2.16. *The weak topology τ of $\mathscr{P}_2(X)$ is the inductive limit of $(\mathscr{X}_n, \tau_n)$.*

Note that the definition makes sense because $\mathscr{X}_n \subset \mathscr{X}_{n+1}$ and the inclusion is continuous.

The first properties of $(\mathscr{P}_2(X), \tau)$ are collected below.

- The topology τ is *not* the weak topology on $\mathscr{P}_2(X)$ induced by the duality with continuous and bounded functions. Indeed by proposition 2.15 we have that $\mu_n \xrightarrow{\tau} \mu$ if and only if $\mu_n \rightharpoonup \mu$ and μ_n is 2-bounded.
- Closed balls in $(\mathscr{P}_2(X), W)$ are compact and sequentially compact in $(\mathscr{P}_2(X), \tau)$. Indeed the previous statement implies that balls are sequentially compact, and to get compactness observe that for any closed ball B there exists N such that $B \subset \mathscr{X}_N$: since $\mathscr{X}_N$ is compact the conclusion follows.
- $(\mathscr{P}_2(X), \tau)$ is *not* induced by a distance. Indeed *ii*) of proposition 2.15 suggests that $(\mathscr{P}_2(X), \tau)$ does not satisfy the first axiom of numerability. The following is an explicit counterexample.

  EXAMPLE 2.17. *Let $X := \mathbb{R}$, $\mu := \delta_0$, $\mu_n := (1 - \frac{1}{n^2})\delta_0 + \frac{1}{n^2}\delta_n$ and $\mu_{n,m} := (1 - \frac{1}{n^2} - \frac{1}{m^2})\delta_0 + \frac{1}{n^2}\delta_n + \frac{1}{m^2}\delta_{nm}$. Then $\mu_n \xrightarrow{\tau} \mu$ and for a fixed n it holds $\mu_{n,m} \xrightarrow{\tau} \mu_n$ (since $\int |x|^2 d\mu_{n,m} = 1 + n^2$ for every m), but we cannot find a sequence of elements of the form $\mu_{n,m}$ which $\tau-$converge to μ.*

- $(\mathscr{P}_2(X), \tau)$ is *not* locally sequentially compact. Indeed choose an open neighborhood U of δ_0 and observe that each sequence $(\mu_n) \subset (U \cap \mathscr{X}_n) \setminus \mathscr{X}_{n-1}$ does not have any converging subsequence (note that it eventually holds $U \cap \mathscr{X}_{n-1} \subsetneq U \cap \mathscr{X}_n$).
- W is sequentially lower semicontinuous w.r.t. τ. Indeed for any couple of sequences μ_n, ν_n $\tau-$converging to μ, ν respectively and any choice of optimal plans $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu_n, \nu_n)$ we have by proposition 1.5 that $(\boldsymbol{\gamma}_n)$ is bounded in $(\mathscr{P}_2(\mathbb{R}^{2d}), W)$. Therefore it is sequentially compact w.r.t. τ and for every subsequence $(\boldsymbol{\gamma}_{n_k})$ converging to a certain $\boldsymbol{\gamma}$ we have $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$ and therefore

$$W(\mu, \nu) \leq \|x_1 - x_2\|_{\boldsymbol{\gamma}} \leq \varliminf_{k\to\infty} \|x_1 - x_2\|_{\boldsymbol{\gamma}_{n_k}} = \varliminf_{k\to\infty} W(\mu_{n_k}, \nu_{n_k}). \tag{2.23}$$

  Being the result independent on the subsequence chosen we get the claim.

In the next propositions we analyze the relationship between convergence of measures, weak or strong, and convergence of optimal plans, in the same spirit of proposition 1.11. The first proposition below shows that the plan $\boldsymbol{\gamma}$ of equation (2.23) is actually optimal for the couple μ, ν.

PROPOSITION 2.18. *Let $\mu_n, \nu_n, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be given measures such that $\mu_n \xrightarrow{\tau} \mu$, $\nu_n \xrightarrow{\tau} \nu$, and let $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu_n, \nu_n)$ be any choice of optimal plans. Then the sequence*

(γ_n) is relatively compact in $(\mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^d_2), \tau)$ and every τ accumulation point belongs to $\mathit{Opt}(\mu, \nu)$.

Proof. The relative compactness follows from proposition 1.5 and remark 2.3, while the optimality of the limit plan is a consequence of proposition 1.11. □

The analogous statement for strong convergence holds with a weaker assumption on the plans γ_n:

PROPOSITION 2.19. *Let $\mu_n, \nu_n, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be given measures such that $\mu_n \xrightarrow{W} \mu$, $\nu_n \xrightarrow{W} \nu$, and let $\gamma_n \in \mathit{Adm}(\mu, \nu)$ be any choice of admissible plans such that*

$$\varlimsup_{n\to\infty} \|x_1 - x_2\|_{\gamma_n} \leq W(\mu, \nu).$$

Then the sequence γ_n is relatively compact in $(\mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^d_2), W)$ and any W accumulation point belongs to $\mathit{Opt}(\mu, \nu)$.

Proof. The relative compactness is a consequence of propositions 1.5, 2.5, while the optimality of the limit plan follows by

$$W(\mu, \nu) \leq \|x_1 - x_2\|_{\gamma} \leq \varliminf_{k\to\infty} \|x_1 - x_2\|_{\gamma_{n_k}} \leq W(\mu, \nu).$$

□

5. Convergence of maps

In the following we will usually need to find the limit map, in a certain sense, of a sequence of maps $T_n \in L^2(\mathbb{R}^d, \mathbb{R}^{d'}; \mu_n)$ when the measures μ_n are different. The result will be a couple (T, μ) with $T \in L^2(\mathbb{R}^d, \mathbb{R}^{d'}; \mu)$. We are going to describe two type of convergence which are in some sense similar to the strong and to the weak convergence in L^2.

DEFINITION 2.20 (Weak and strong convergence of maps). *Let (T, μ), (T_n, μ_n), $n \in \mathbb{N}$, be given couples such that $T_n \in L^2(\mathbb{R}^d, \mathbb{R}^{d'}; \mu_n)$ and $T \in L^2(\mathbb{R}^d, \mathbb{R}^{d'}; \mu)$. We say that (T_n, μ_n) weakly converge to (T, μ), and write $(T_n, \mu_n) \xrightarrow{\tau} (T, \mu)$, if*

$$\mu_n \xrightarrow{\tau} \mu, \tag{2.24}$$

$$T_n \mu_n \rightharpoonup T\mu, \tag{2.25}$$

$$\sup_{n\in\mathbb{N}} \|T_n\|_{\mu_n} < +\infty, \tag{2.26}$$

where the convergence in (2.25) is in duality with C_c, and, in (2.26), $\|T_n\|_{\mu_n}$ stands for the L^2 norm of T_n as element of $L^2(\mathbb{R}^d, \mathbb{R}^{d'}, \mu_n)$. We say that (T_n, μ_n) strongly converge (or simply converge) to (T, μ), and write $(T_n, \mu_n) \twoheadrightarrow (T, \mu)$, if

$$\mu_n \xrightarrow{\tau} \mu, \tag{2.27}$$

$$T_n \mu_n \rightharpoonup T\mu, \tag{2.28}$$

$$\|T_n\|_{\mu_n} \to \|T\|_{\mu}. \tag{2.29}$$

Observe that if $\mu_n = \mu$ for every n these definitions reduce to the usual weak and strong convergence in $L^2(\mathbb{R}^d, \mathbb{R}^{d'}, \mu)$. It is possible to define appropriate topologies which induce these distances with the same ideas used to construct τ, but we do not investigate them.

In the following when we write a couple (T, μ) it will be always intended that T is an L^2 map w.r.t. μ with values in some Euclidean space $\mathbb{R}^{d'}$, moreover we will sometimes write $T_n \twoheadrightarrow T$ (or $T_n \xrightarrow{\tau} T$) omitting the explicit reference to the measures μ_n, μ if there is no risk of confusion. Before studying the first properties of weak and strong convergence of maps we need the following definition.

Definition 2.21 (Barycentric projection). *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^{d'}_2)$ be a plan and let $d\boldsymbol{\gamma} = d\pi^1_\#\boldsymbol{\gamma} \otimes d\boldsymbol{\gamma}_{x_1}$ be its disintegration w.r.t. its first marginal. The* barycentric projection *of $\boldsymbol{\gamma}$ w.r.t. $\pi^1_\#\boldsymbol{\gamma}$ is the function $\mathscr{B}(\boldsymbol{\gamma})(x_1) : \mathbb{R}^d \to \mathbb{R}^{d'}$ defined as*

$$\mathscr{B}(\boldsymbol{\gamma})(x_1) := \int x_2 d\boldsymbol{\gamma}_{x_1}.$$

Observe that since

$$\|\mathscr{B}(\boldsymbol{\gamma})\|^2_{\pi^1_\#\boldsymbol{\gamma}} = \int |\mathscr{B}(\boldsymbol{\gamma})(x_1)|^2 d\pi^1_\#\boldsymbol{\gamma} = \int \left| \int x_2 d\boldsymbol{\gamma}_{x_1} \right|^2 d\pi^1_\#\boldsymbol{\gamma} \\ \leq \int |x_2|^2 d\boldsymbol{\gamma}_{x_1} d\pi^1_\#\boldsymbol{\gamma} = \int |x_2|^2 d\boldsymbol{\gamma} < +\infty, \tag{2.30}$$

$\mathscr{B}(\boldsymbol{\gamma})$ is a well defined function belonging to $L^2(\mathbb{R}^d, \mathbb{R}^{d'}; \pi^1_\#\boldsymbol{\gamma})$. The same equation implies that equality holds if and only if $\boldsymbol{\gamma}_{x_1} = \delta_{\mathscr{B}(\boldsymbol{\gamma})(x_1)}$ for $\pi^1_\#\boldsymbol{\gamma}$-a.e. x_1, i.e.

$$\|\mathscr{B}(\boldsymbol{\gamma})\|_{\pi^1_\#\boldsymbol{\gamma}} = \|x_2\|_{\boldsymbol{\gamma}} \quad \Leftrightarrow \quad \boldsymbol{\gamma} = (Id, \mathscr{B}(\boldsymbol{\gamma}))_\#(\pi^1_\#\boldsymbol{\gamma}). \tag{2.31}$$

The barycentric projection is also characterized by the following formula

$$\int \langle \varphi(x_1), x_2 \rangle d\boldsymbol{\gamma} = \int \langle \varphi(x_1), \mathscr{B}(\boldsymbol{\gamma})(x_1) \rangle d\pi^1_\#\boldsymbol{\gamma}, \quad \forall \varphi \in C^\infty_c(\mathbb{R}^d, \mathbb{R}^{d'}). \tag{2.32}$$

Proposition 2.22. *Let $(T_n, \mu_n), (T, \mu)$, $n \in \mathbb{N}$, be given couples. Then*

$$(Id, T_n)_\#\mu_n \xrightarrow{\tau} (Id, T)_\#\mu \Rightarrow (T_n, \mu_n) \xrightarrow{\tau} (T, \mu). \tag{2.33}$$

Conversely if $(T_n, \mu_n) \xrightarrow{\tau} (T, \mu)$ then the family of plans $(Id, T_n)_\#\mu_n$ is 2-bounded and for any weak accumulation point $\boldsymbol{\gamma}$ it holds

$$\mathscr{B}(\boldsymbol{\gamma}) = T. \tag{2.34}$$

Proof. Let us start with (2.33). Since $\pi^1_\#(Id, T_n)_\#\mu_n = \mu_n$ and $\|T_n\|^2_{\mu_n} = \int |x_2|^2 d(Id, T_n)_\#\mu_n$ conditions (2.24) and (2.26) are fulfilled. Moreover choosing test functions of the form $\psi(x_1, x_2) = \langle \varphi(x_1), x_2 \rangle$ we easily obtain (2.25). On the other hand the same choice of ψ and equation (2.32) yield to (2.34). □

Observe that this proposition yields the inequality

$$(T_n, \mu_n) \xrightarrow{\tau} (T, \mu) \quad \Rightarrow \quad \|T\|_\mu \leq \varliminf_{n\to\infty} \|T_n\|_{\mu_n}. \tag{2.35}$$

Indeed, from equation (2.30) we get

$$\|T\|_\mu = \|\mathscr{B}(\boldsymbol{\gamma})\|_\mu \le \|x_2\|_{\boldsymbol{\gamma}} \le \varliminf_{n\to\infty} \|x_2\|_{\boldsymbol{\gamma}_n} = \varliminf_{n\to\infty} \|T_n\|_{\mu_n}.$$

Therefore as in the usual case of Hilbert spaces the difference between strong and weak convergence is that in the first case there is no loss of norm.

Proposition 2.23. *The following statements are equivalent.*

i)

$$(T_n, \mu_n) \twoheadrightarrow (T, \mu). \tag{2.36}$$

ii) $\mu_n \xrightarrow{\mathcal{T}} \mu$ *and for any choice of* $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu_n, \mu)$ *it holds*

$$\|T_n(x_1) - T(x_2)\|_{\boldsymbol{\gamma}_n} \to 0. \tag{2.37}$$

iii) *For any continuous function* φ *with linear growth and values in some Euclidean space, it holds*

$$(\varphi \circ T_n, \mu_n) \twoheadrightarrow (\varphi \circ T, \mu). \tag{2.38}$$

iv) $\|T_n\|_{\mu_n} \to \|T\|_\mu$ *and*

$$(Id, T_n)_\# \mu_n \xrightarrow{\mathcal{T}} (Id, T)_\# \mu. \tag{2.39}$$

Proof. *iv)* $\Rightarrow$ *i)* follows easily from (2.33). Conversely suppose that $(Id, T_{n_k})_\# \mu_{n_k} \xrightarrow{\mathcal{T}} \boldsymbol{\gamma}$ and use the other implication of proposition 2.22 to get

$$\|T\|_\mu = \|\mathscr{B}(\boldsymbol{\gamma})\|_\mu \le \|x_2\|_{\boldsymbol{\gamma}} \le \varliminf_{k\to\infty} \|x_2\|_{\boldsymbol{\gamma}_{n_k}} = \varliminf_{k\to\infty} \|T_{n_k}\|_{\mu_{n_k}} = \|T\|_\mu.$$

Therefore equation (2.31) gives

$$\boldsymbol{\gamma} = (Id, \mathscr{B}(\boldsymbol{\gamma}))_\# \mu = (Id, T)_\# \mu.$$

Now we turn to *i)* $\Rightarrow$ *ii)*: note that proposition 1.11 implies $\boldsymbol{\gamma}_n \xrightarrow{\mathcal{T}} (Id, Id)_\# \mu$. Now fix $\psi \in C_b(\mathbb{R}^d, \mathbb{R}^{d'})$ and observe that since $\psi(x_1) - \psi(x_2) \in C_b(\mathbb{R}^d_1 \times \mathbb{R}^d_2, \mathbb{R}^{d'})$ it holds $\|\psi(x_1) - \psi(x_2)\|_{\boldsymbol{\gamma}_n} \to 0$, moreover

$$\varlimsup_{n\to\infty} \|T_n - \psi\|^2_{\mu_n} = \|T\|^2_\mu + \|\psi\|^2_\mu - 2 \varlimsup_{n\to\infty} \int \langle T_n, \psi \rangle d\mu_n = \|T - \psi\|^2_\mu.$$

Hence, we can conclude with

$$\begin{aligned} &\varlimsup_{n\to\infty} \|T_n(x_1) - T(x_2)\|_{\boldsymbol{\gamma}_n} \\ &\qquad \le \varlimsup_{n\to\infty} \|T_n - \psi\|_{\mu_n} + \varlimsup_{n\to\infty} \|\psi(x_1) - \psi(x_2)\|_{\boldsymbol{\gamma}_n} + \|\psi - T\|_\mu = 2\|\psi - T\|_\mu, \end{aligned}$$

by letting ψ tend to T in $L^2(\mathbb{R}^d, \mathbb{R}^{d'}, \mu)$.

To show that *ii)* $\Rightarrow$ *i)* begin observing that

$$\Big| \|T_n\|_{\mu_n} - \|T\|_\mu \Big| \le \|T_n(x_1) - T(x_2)\|_{\boldsymbol{\gamma}_n} \to 0,$$

to get (2.29). Then choose $\psi \in C_c(\mathbb{R}^d, \mathbb{R}^{d'})$ and note that since $\|\psi(x_1) - \psi(x_2)\|_{\boldsymbol{\gamma}_n} \to 0$ and $\|T_n(x_1) - T(x_2)\|_{\boldsymbol{\gamma}_n} \to 0$ equation (2.28) follows from

$$\int \langle \psi, T_n \rangle d\mu_n - \int \langle \psi, T \rangle d\mu = \int \langle \psi(x_1), T_n(x_1) \rangle - \langle \psi(x_2), T(x_2) \rangle d\boldsymbol{\gamma}_n(x_1, x_2)$$

$$= \int \langle \psi(x_1), T_n(x_1) - T(x_2) \rangle d\boldsymbol{\gamma}_n(x_1, x_2) + \int \langle \psi(x_1) - \psi(x_2), T(x_2) \rangle d\boldsymbol{\gamma}_n(x_1, x_2) \to 0.$$

Since it is clear that $iii) \Rightarrow i)$ it remains to prove that $i), ii), iv) \Rightarrow iii)$. We will show that if $(T_n, \mu_n), (T, \mu)$ satisfy (2.37) then $(\varphi \circ T_n, \mu), (\varphi \circ T, \mu)$ still satisfy (2.37). Since

$$\left| \int \varphi_1 \circ T - \varphi_2 \circ T d\mu \right| \leq \sup_x |\varphi_1(x) - \varphi_2(x)|,$$

it suffices to check (2.38) for Lipschitz functions φ. Under this hypothesis we have

$$\|\varphi(T_n(x_1)) - \varphi(T(x_2))\|_{\boldsymbol{\gamma}_n} \leq \mathrm{Lip}_\varphi \|T_n(x_1) - T(x_2)\|_{\boldsymbol{\gamma}_n} \to 0.$$

□

REMARK 2.24. *It is not hard to see that, with the same notation of the above proposition, if we assume $W(\mu_n, \mu) \to 0$ as $n \to \infty$, then the strong convergence of maps is equivalent to $\|T(x) - T_n(y)\|_{\boldsymbol{\gamma}_n} \to 0$ as $n \to \infty$, where $\boldsymbol{\gamma}_n \in \mathcal{Adm}(\mu, \mu_n)$ is any choice of plans satisfying $\|x - y\|_{\boldsymbol{\gamma}_n} \to 0$ as $n \to \infty$.*

Observe that if μ is regular equation (2.37) reads as follows

$$T_n \circ T_\mu^{\mu_n} \to T \qquad \text{in } L^2(\mathbb{R}^d, \mathbb{R}^{d'}, \mu).$$

COROLLARY 2.25. *Let (T_n, μ_n) be couples strongly converging to (T, μ). Then*

$$(T_n)_\# \mu_n \to T_\# \mu, \quad \text{in } \mathscr{P}_2(\mathbb{R}^{d'}, W).$$

Moreover, if in addition $\mu_n \to \mu$ then

$$(Id, T_n)_\# \mu_n \to (Id, T)_\# \mu, \quad \text{in } \mathscr{P}_2(\mathbb{R}^{d+d'}, W).$$

Proof. It is a straightforward consequence of the previous proposition. □

A consequence of propositions 2.18, 2.19 is the close link between convergence of measures and convergence of optimal maps given below.

PROPOSITION 2.26. *Let $\mu_n, \nu_n, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be given measures such that there exists optimal maps T_n from μ_n to ν_n and T from μ to ν. Suppose moreover that $\mu_n \xrightarrow{\tau} \mu, \nu_n \xrightarrow{\tau} \nu$ and that $(Id, T)_\# \mu$ is the unique element of $\mathcal{Opt}(\mu, \nu)$. Then*

$$(T_n, \mu_n) \xrightarrow{\tau} (T, \mu).$$

PROPOSITION 2.27. *Let $\mu_n, \nu_n, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be given measures such that $\mu_n \xrightarrow{W} \mu$, $\nu_n \xrightarrow{W} \nu$ and $(Id, T)_\# \mu$ is the unique element of $\mathcal{Opt}(\mu, \nu)$ for some map T. Then if $T_n \in L^2(\mu_n, \mathbb{R}^d)$ is any sequence satisfying $(T_n)_\# \mu_n = \nu_n$ and*

$$\overline{\lim_{n \to \infty}} \|T_n - Id\|_{\mu_n} \leq W(\mu, \nu), \tag{2.40}$$

it holds

$$(T_n, \mu_n) \twoheadrightarrow (T, \mu).$$

Proof. Proposition 2.26 is a simple consequence of proposition 2.18, 2.22, while proposition 2.27 follows from 2.19 and 2.23. □

Observe that the conditions on the given measures are fulfilled if they are regular; we will mostly use the previous propositions under this assumption.

The following lemma is a variant of the classical result in Hilbert spaces:

$$v_n \to v, \ w_n \rightharpoonup w \quad \Rightarrow \quad \langle v_n, w_n \rangle \to \langle v, w \rangle.$$

LEMMA 2.28. *Let $(T_n, \mu_n), (S_n, \mu_n)$, $n \in \mathbb{N}$, be given couples such that all the maps have values in the same $\mathbb{R}^{d'}$ and suppose that*

$$(T_n, \mu_n) \twoheadrightarrow (T, \mu), \tag{2.41}$$

$$(S_n, \mu_n) \xrightarrow{\tau} (S, \mu), \tag{2.42}$$

for some measure μ and maps T, S. Then

$$\lim_{n\to\infty} \int \langle T_n, S_n \rangle d\mu_n = \int \langle T, S \rangle d\mu. \tag{2.43}$$

Proof. Let $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu_n, \mu)$, $n \in \mathbb{N}$, be any choice of optimal plans and $\psi \in C_c(\mathbb{R}^d, \mathbb{R}^{d'})$ be a fixed function. Reasoning as in the proof of proposition 2.23, from (2.41) we know that

$$\lim_{n\to\infty} \|T_n - \psi\|_{\mu_n} = \|T - \psi\|_{\mu}, \tag{2.44}$$

which yield to

$$\begin{aligned}
\left| \int \langle T_n, S_n \rangle d\mu_n - \int \langle T, S \rangle d\mu \right| \le & \left| \int \langle T_n(x_1) - \psi(x_1), S_n(x_1) \rangle d\boldsymbol{\gamma}_n \right| \\
& + \left| \int \langle \psi(x_1), S_n(x_1) \rangle - \langle \psi(x_2), S(x_2) \rangle d\boldsymbol{\gamma}_n \right| \\
& + \left| \int \langle \psi(x_2) - T(x_2), S(x_2) \rangle d\boldsymbol{\gamma}_n \right|.
\end{aligned}$$

Taking the limit, using (2.44), (2.42) and letting $C := \sup_n \|S_n\|_{\mu_n}$ we get

$$\overline{\lim_{n\to\infty}} \left| \int \langle T_n, S_n \rangle d\mu_n - \int \langle T, S \rangle d\mu \right| \le 2C \|T - \psi\|_{\mu},$$

and by the arbitrarieness of ψ the thesis follows. □

We will mostly apply this lemma in the following form:

COROLLARY 2.29. *Let $(\boldsymbol{\gamma}_n)$ be a sequence of plans $\tau-$converging to some $\boldsymbol{\gamma}$ such that*

$$\pi^1_\# \boldsymbol{\gamma}_n \xrightarrow{W} \pi^1_\# \boldsymbol{\gamma}.$$

Then

$$\int \langle x_1, x_2 \rangle d\boldsymbol{\gamma}_n \to \int \langle x_1, x_2 \rangle d\boldsymbol{\gamma}.$$

Proof. It is sufficient to apply the previous lemma with $\mu_n := \gamma_n$, $T_n := \pi^1$, $S_n := \pi^2$. □

We conclude with the following proposition, which shows that bounded sequences of maps are weakly compact.

PROPOSITION 2.30 (Weak compactness of maps). *Let (μ_n) be a 2-bounded sequence of measures and let $T_n \in L^2(\mathbb{R}^d, \mathbb{R}^{d'}, \mu_n)$ be a sequence of maps such that $\|T_n\|_{\mu_n} \leq C < +\infty$. Then there exists a subsequence (T_{n_k}, μ_{n_k}) weakly converging to some couple (T, μ).*

Proof. Since the sequence of measures is 2-bounded, we can extract a subsequence, not relabeled, which weakly converges to a measure μ. Now consider a countable set $D \subset C_c^\infty(\mathbb{R}^d, \mathbb{R}^{d'})$ which is dense in L^2_μ and observe that it holds

$$\overline{\lim_{n\to\infty}} \left| \langle T_n, \varphi \rangle_{\mu_n} \right| \leq \overline{\lim_{n\to\infty}} \|T_n\|_{\mu_n} \|\varphi\|_{\mu_n} \leq C \|\varphi\|_\mu,$$

for any $\varphi \in D$. Therefore we can extract a subsequence (T_{n_k}) such that there exists the limit $L(\varphi)$ of $k \to \langle T_{n_k}, \varphi \rangle_{\mu_{n_k}}$ for any $\varphi \in D$. The functional L can be clearly extended to a linear functional on the whole L^2_μ whose norm is controlled by C (by the density of D), therefore from the Rietz representation theorem we get the existence of a map $T \in L^2_\mu$ such that $L(S) = \langle T, S \rangle_\mu$ for any $S \in L^2_\mu$. From the density of D in C_c we get that

$$\lim_{n\to\infty} \langle \varphi, T_n \rangle_{\mu_n} = \langle \varphi, T \rangle_\mu, \quad \forall \varphi \in C_c(\mathbb{R}^d, \mathbb{R}^{d'}),$$

therefore the proof is complete. □

REMARK 2.31. *In the following we will often use, without explicit mention, the fact that for a $\tau-$converging sequence of measures (μ_n) and a sequence of maps $f_n \in L^2_{\mu_n}$ such that $\sup_n \|f_n\|_{\mu_n} < \infty$, in order to check the weak convergence of f_n it is sufficient to check the convergence in duality with functions in $C_c^\infty(\mathbb{R}^d)$. This is an easy consequence of the tightness of (μ_n) and the uniform bound of the norms of f_n.*

CHAPTER 3

The regular tangent space

In this chapter we investigate the so-called *Riemannian structure* of $(\mathscr{P}_2(\mathbb{R}^d), W)$. The ideas presented here were first studied by Otto in [**46**], and subsequentely made rigorous in [**11**]: the approach we present follows the one of the latter work. The first step will be the study of the continuity equation

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0, \tag{3.1}$$

which enables us to characterize absolutely continuous curves. As we will see, this will naturally lead to the introduction of the *regular tangent* space as

$$\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) := \overline{\left\{ \nabla\varphi \ : \ \varphi \in C_c^\infty(\mathbb{R}^d) \right\}}^{L^2(\mu,\mathbb{R}^d)}.$$

This space is the object of analysis of the second section of the chapter.

1. Continuity equations and AC curves

Recall that a curve $t \to u_t$ on an interval I, with values in a metric space (E, d), is said to be absolutely continuous (AC) if there exists a function $g \in L^1(I)$ such that

$$d(u_t, u_s) \le \int_t^s g(r)dr, \quad \forall t \le s, \ t, s \in I. \tag{3.2}$$

It is a well known fact (see for instance [**13**]) that for a given AC curve $u(t)$ the limit

$$\lim_{h\to 0} \frac{d(u_{t+h}, u_t)}{h}$$

exists for a.e. $t \in I$ and defines a fuction which belongs to $L^1(I)$ and is minimal among the functions g satisfying (3.2). This function is usually denoted by $t \to |\dot{u}_t|$ and called *metric derivative*, or metric speed.

The *length* $\mathcal{L}(u_t)$ of a curve $t \to u_t$ is defined as

$$\mathcal{L}(u_t) := \sup \sum_{i=0}^{n-i} d(u_{t_i}, u_{t_{i+1}}),$$

where the sup is taken among all $n \in \mathbb{N}$ and partitions $\{t_0 \le \cdots \le t_n\}$ of the inteval I. For an AC curve it is possible to show that

$$\mathcal{L}(u_t) = \int_I |\dot{u}_t| dt.$$

In this section we will show that absolutely continuous curves in $\mathscr{P}_2(\mathbb{R}^d)$ are completely characterized by equation (3.1), meaning that:

- for every absolutely continuous curve $[0,1] \ni t \to \mu_t \in \mathscr{P}_2(\mathbb{R}^d)$ it is possible to find "velocity vectors" v_t such that $\|v_t\|_{\mu_t} \leq |\dot{\mu}_t|$ for a.e. t and (3.1) holds,
- conversely, every distributional solution of (3.1) with $\|v_t\|_{\mu_t} \in L^1(0,1)$ gives a curve μ_t which is a.e. equal to an absolutely continuous curve whose metric speed is bounded by $\|v_t\|_{\mu_t}$.

We will skip some technical details of the proofs, and refer to [**11**] for a more detailed analysis.

Before proving our main result we need a couple of lemmata.

LEMMA 3.1 (Time rescaling). *Let (μ_t, v_t) be a solution of (3.1), $t \to \gamma(t)$ a strictly increasing and absolutely continuous map and define $\overline{\mu}_t := \mu_{\gamma(t)}, \overline{v}_t := \gamma'(t)v_t$. Then $(\overline{\mu}_t, \overline{v}_t)$ is a solution of (3.1), too.*

Proof. Straightforward. □

PROPOSITION 3.2 (Uniqueness of solutions of continuity equation). *Let σ_t, $0 \leq t \leq 1$, be a family of signed measures solving*

$$\frac{d}{dt}\sigma_t + \nabla \cdot (v_t \sigma_t) = 0,$$

with $\sigma_0 \leq 0$,

$$\int_0^1 \int |v_t| d|\sigma_t| < +\infty$$

and

$$\int_0^1 \left(|\sigma_t|(B) + \sup_B |v_t| + \mathrm{Lip}(v_t, B) \right) dt < +\infty$$

for each compact set $B \subset \mathbb{R}^d$. Then $\sigma_t \leq 0$ for any $t \in [0,1]$.

Proof. Fix $\psi \in C_c^\infty(\mathbb{R}^d \times (0,1))$ with $0 \leq \psi \leq 1$, $R > 0$, and a smooth cut-off function $\chi_R(\cdot) = \chi(\cdot/R) \in C_c^\infty(\mathbb{R}^d)$ such that $0 \leq \chi_R \leq 1$, $|\nabla \chi_R| \leq 2/R$, $\chi_R \equiv 1$ on $B_R(0)$ and $\chi_R \equiv 0$ outside $B_{2R}(0)$. We define w_t so that $w_t = v_t$ on $B_{2R}(0) \times (0,1)$, $w_t = 0$ if $t \neq [0,1]$ and

$$\sup_{\mathbb{R}^d} |w_t| + \mathrm{Lip}(w_t, \mathbb{R}^d) \leq \sup_{B_{2R}(0)} |v_t| + \mathrm{Lip}(v_t, B_{2R}(0)) \quad \forall t \in [0,1].$$

Let w_t^ε be obtained from w_t by a double mollification with respect to the space and time variables: notice that w_t^ε satisfy

$$\sup_{\varepsilon \in (0,1)} \int_0^1 \left(\sup_{\mathbb{R}^d} |w_t^\varepsilon| + \mathrm{Lip}(w_t^\varepsilon, \mathbb{R}^d) \right) dt < +\infty.$$

We now build with the method of characteristics a smooth solution $\varphi^\varepsilon : \mathbb{R}^d \times [0,1] \to \mathbb{R}$ of the PDE

$$\frac{d}{dt}\varphi^\varepsilon + \langle w_t^\varepsilon, \nabla \varphi^\varepsilon \rangle = \psi \quad \text{in } \mathbb{R}^d \times (0,1), \quad \varphi^\varepsilon(x,1) = 0 \ \ \forall x \in \mathbb{R}^d.$$

It is not difficult to check (we skip the details) that $-1 \le \varphi^\varepsilon \le 0$ and that $|\nabla\varphi^\varepsilon|$ is uniformly bounded w.r.t. ε, t and x.

We insert now the test function $\varphi^\varepsilon\chi_R$ in the continuity equation and take into account that $\sigma_0 \le 0$ and $\varphi^\varepsilon \le 0$ to obtain

$$\begin{aligned}0 \ge -\int_{\mathbb{R}^d}\varphi^\varepsilon\chi_R d\sigma_0 &= \int_0^1\int_{\mathbb{R}^d}\chi_R\frac{d\varphi^\varepsilon}{dt} + \langle v_t, \chi_R\nabla\varphi^\varepsilon + \varphi^\varepsilon\nabla\chi_R\rangle d\sigma_t dt\\ &= \int_0^1\int_{\mathbb{R}^d}\chi_R(\psi + \langle v_t - w_t^\varepsilon, \nabla\varphi^\varepsilon\rangle)d\sigma_t dt + \int_0^1\int_{\mathbb{R}^d}\varphi^\varepsilon\langle\nabla\chi_R, v_t\rangle d\sigma_t dt\\ &\ge \int_0^1\int_{\mathbb{R}^d}\chi_R(\psi + \langle v_t - w_t^\varepsilon, \nabla\varphi^\varepsilon\rangle)d\sigma_t dt - \int_0^1\int_{\mathbb{R}^d}|\nabla\chi_R||v_t|d|\sigma_t|dt.\end{aligned}$$

Letting $\varepsilon \to 0^+$ and using the uniform bound on $|\nabla\varphi^\varepsilon|$ and the fact that $w_t = v_t$ on $\operatorname{supp}\chi_R \times [0,1]$, we get

$$\int_0^1\int_{\mathbb{R}^d}\chi_R\psi d\sigma_t dt \le \int_0^1\int_{\mathbb{R}^d}|\nabla\chi_R||v_t|d|\sigma_t|dt \le \frac{2}{R}\int_0^1\int_{R\le|x|\le 2R}|v_t|d|\sigma_t|dt.$$

Eventually letting $R \to \infty$ we obtain that $\int_0^1\int_{\mathbb{R}^d}\psi d\sigma_t dt \le 0$. Since ψ is arbitrary we get the thesis. □

LEMMA 3.3. *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a measure and let E be a $\mathbb{R}^m$ valued measure on $\mathbb{R}^d$ with finite total variation and absolutely continuous w.r.t. μ. Then for any convolution kernel ρ it holds*

$$\int\left|\frac{E*\rho}{\mu*\rho}\right|^2\mu*\rho dx \le \int\left|\frac{E}{\mu}\right|^2 d\mu$$

Proof. We use Jensen inequality in the following form: if $\Psi : \mathbb{R}^{m+1} \to [0,+\infty]$ is convex, l.s.c. and positively 1-homogeneous, then

$$\Psi\left(\int\psi(x)d\theta(x)\right) \le \int\Psi(\psi(x))d\theta(x),$$

for any Borel map $\psi : \mathbb{R}^d \to \mathbb{R}^{m+1}$ and any positive and finite measure θ in $\mathbb{R}^d$ (by rescaling θ to be a probability measure and looking at the image measure $\psi_\#\theta$ the formula reduces to the standard Jensen inequality). Fix $x \in \mathbb{R}^d$ and apply the inequality above with $\psi := (E/\mu, 1)$, $\theta = \rho(x-\cdot)\mu$ and

$$\Psi(z,t) := \begin{cases}\dfrac{|z|^2}{t} & \text{if } t > 0,\\ 0 & \text{if } (z,t) = (0,0),\\ +\infty & \text{elsewhere,}\end{cases}$$

to obtain

$$\begin{aligned}\left|\frac{E*\rho(x)}{\mu*\rho(x)}\right|^2\mu*\rho(x) =&\Psi\left(\int\frac{E}{\mu}(y)\rho(x-y)d\mu(y),\int\rho(x-y)d\mu(y)\right)\\ \leq&\int\Psi(\frac{E}{\mu}(y),1)\rho(x-y)d\mu(y)\\ =&\int\left|\frac{E}{\mu}\right|^2(y)\rho(x-y)d\mu(y).\end{aligned}$$

Integrating w.r.t. x we get the thesis. □

THEOREM 3.4 (Characterization of AC curves). *Let $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve. Then for a.e. $t \in [0,1]$ there exist vector fields $v_t \in L^2(\mu_t, \mathbb{R}^d)$ whose norm is controlled by the metric derivative of μ_t*

$$\|v_t\|_{\mu_t} \leq |\dot{\mu}_t|, \quad a.e.\ t \in [0,1],$$

such that the equation

$$\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0, \tag{3.3}$$

holds in the sense of distributions, i.e.

$$\int_0^1\int\left(\frac{d}{dt}\varphi(t,x) + \langle\nabla_x\varphi(x,t), v_t(x)\rangle\right)d\mu_t(x)dt = 0, \tag{3.4}$$

for all $\varphi \in C_c^\infty((0,1)\times\mathbb{R}^d)$.

Conversely if $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ satisfies equation (3.3) *for some family of vector fields $v_t \in L^2(\mu, \mathbb{R}^d)$ such that*

$$\int_0^1 \|v_t\|_{\mu_t} < +\infty,$$

then $t \to \mu_t$ admits an absolutely continuous representative $t \to \overline{\mu}_t$ satisfying

$$|\dot{\overline{\mu}}_t| \leq \|v_t\|_{\overline{\mu}_t}, \quad a.e.\ t \in [0,1].$$

Proof. Let us start assuming that μ_t is an absolutely continuous curve. Using lemma 3.1 we may assume that the metric derivative is constant. For any map $\varphi \in C_c^1(\mathbb{R}^d)$ consider the upper semicontinuous map $H_\varphi(x,y)$ defined as

$$H_\varphi(x,y) := \begin{cases} |\nabla\varphi(x)| & \text{if } x = y,\\ \dfrac{|\varphi(x)-\varphi(y)|}{|x-y|} & \text{if } x \neq y,\end{cases}$$

moreover let $\boldsymbol{\gamma}_t^s$ be any optimal plan for the couple μ_t, μ_s. With this notation we evaluate the following limit for every t in which μ_t is metrically differentiable

$$
\begin{aligned}
\overline{\lim_{h\to 0^+}} \frac{|\mu_{t+h}(\varphi) - \mu_t(\varphi)|}{h} &= \overline{\lim_{h\to 0^+}} \frac{1}{h}\int_{\mathbb{R}^{2d}} |\varphi(y) - \varphi(x)| d\boldsymbol{\gamma}_t^{t+h}(x,y) \\
&\le \overline{\lim_{h\to 0^+}} \frac{1}{h}\int_{\mathbb{R}^{2d}} |x-y| H_\varphi(x,y) d\boldsymbol{\gamma}_t^{t+h}(x,y) \\
&\le \overline{\lim_{h\to 0^+}} \frac{W(\mu_{t+h},\mu_t)}{h}\left(\int_{\mathbb{R}^{2d}} H_\varphi^2(x,y) d\boldsymbol{\gamma}_t^{t+h}(x,y)\right)^{1/2} \\
&\le |\dot\mu_t| \|\nabla\varphi\|_{\mu_t},
\end{aligned}
\tag{3.5}
$$

where the last inequality follows from the upper semicontinuity of H_φ and the fact that $\boldsymbol{\gamma}_t^{t+h}$ converge to $(Id, Id)_\# \mu_t$ as $h \to 0$. Consider now a function $\varphi \in C_c^1((0,1)\times\mathbb{R}^d)$ and let $\boldsymbol{\mu} \in \mathscr{P}_2((0,1)\times\mathbb{R}^d)$ be the measure whose disintegration w.r.t. t is μ_t, then we have

$$
\begin{aligned}
\int_{(0,1)\times\mathbb{R}^d} \frac{d}{dt}\varphi(t,x) d\boldsymbol{\mu}(t,x) &= \lim_{h\to 0^+} \int_{(0,1)\times\mathbb{R}^d} \frac{\varphi(t+h,x) - \varphi(t,x)}{h} d\boldsymbol{\mu} \\
&= \lim_{h\to 0^+} \int_{(0,1)} -\frac{\mu_t(\varphi(t,x)) - \mu_{t-h}\varphi(t,x)}{h} dt,
\end{aligned}
$$

therefore using equation (3.5) and Fatou's lemma we have

$$
\left| \int_{(0,1)\times\mathbb{R}^d} \frac{d}{dt}\varphi(t,x) d\boldsymbol{\mu}(t,x) \right| \le \int_{(0,1)} |\dot\mu_t| \|\nabla\varphi(t,x)\|_{\mu_t} dt. \tag{3.6}
$$

Now consider the space V defined as the closure w.r.t. $L^2_{\boldsymbol{\mu}}$ of the set $\left\{\nabla\varphi : \varphi \in C_c^1((0,1)\times\mathbb{R}^d)\right\}$ and consider the linear functional L defined on gradients of regular functions as

$$
L(\nabla\varphi) := -\int_{(0,1)\times\mathbb{R}^d} \frac{d}{dt}\varphi d\boldsymbol{\mu}.
$$

Equation (3.6) ensures that L can be uniquely extended to a bounded functional on the whole V whose norm is controlled by $(\int_0^1 |\dot\mu_t|^2)^{1/2}$ which is equal to $\int_0^1 |\dot\mu_t|$ by our assumption on the metric derivative of $t \to \mu_t$ (here we endow V with the L^2 norm given by $\boldsymbol{\mu}$). Since V is an Hilbert space, the functional L may be represented by the scalar product with a certain vector $v(t,x)$: defining $v_t(x) := v(t,x)$ we have

$$
\int_{(0,1)} \int \frac{d}{dt}\varphi(t,x) d\mu_t(x) dt = -L(\nabla\varphi) = \int_{(0,1)} \int \langle \nabla\varphi(t,x), v_t(x)\rangle d\mu_t(x) dt.
$$

To conclude we need just to show that $\|v_t\|_{\mu_t} \le |\dot\mu_t|$: we already know that $\int_0^1 \|v_t\|_{\mu_t} \le \int_0^1 |\dot\mu_t|$, to prove that $\int_t^s \|v_t\|_{\mu_t} \le \int_t^s |\dot\mu_t|$ holds for any $0 \le t \le s \le 1$ it is sufficient to repeat the argument restricting the curve to the interval (t,s). We will come out with a vector $\overline{v}(r,x) \in V_t^s$, where V_t^s is the closure of $\left\{\nabla\varphi : \varphi \in C_c^1((t,s)\times\mathbb{R}^d)\right\}$ w.r.t. the

L^2 norm given by $\boldsymbol{\mu}|_{(t,s)\times\mathbb{R}^d}$, such that $\int_t^s \|\overline{v}(r,\cdot)\|_{\mu_r} dr \leq \int_t^s |\dot{\mu}_r| dr$ and

$$\int_{(t,s)} \int \frac{d}{dr}\varphi(r,x) d\mu_r(x) dr = \int_{(t,s)} \int \langle \nabla\varphi(r,x), \overline{v}(r,x)\rangle d\mu_r(x) dr.$$

Therefore it must hold $\overline{v}(r,x) = v(r,x)$ $\boldsymbol{\mu}-$a.e. and the claim is proved.

Now we turn to the converse implication: we will assume that μ_t is a family of functions which satisfies equation (3.4) for a certain family of vectors $v_t(x) \in L^2_{\mu_t}$ satisfying $\int_0^1 \|v_t\|_{\mu_t} dt < \infty$ and we wil prove that μ_t is (equivalent to) an absolutely continuous curve in $\mathscr{P}_2(\mathbb{R}^d)$ whose metric derivative is bounded by $\|v_t\|_{\mu_t}$.

We will proceed by approximation: consider a family of gaussian mollifiers ρ^ε and define

$$\mu_t^\varepsilon := \mu * \rho^\varepsilon,$$
$$v_t^\varepsilon := \frac{(v\mu) * \rho^\varepsilon}{\mu_t^\varepsilon}.$$

For the measures μ_ε it clearly holds the equation

$$\frac{d}{dt}\mu_t^\varepsilon + \nabla\cdot(v_t^\varepsilon \mu_t^\varepsilon) = 0. \tag{3.7}$$

It is possible to prove (we skip the details, see Lemma 8.1.4 of [**11**] for a detailed analysis) that there exists a unique family of maps $X_t^\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ such that $\dot{X}_t^\varepsilon(x) = v_t^\varepsilon(X_t^\varepsilon(x))$ and $X_0^\varepsilon(x) = x$ and its just a matter of calculation to show that the measures $(X_t^\varepsilon)_\# \mu_0^\varepsilon$ satisfy equation (3.7), therefore by the uniqueness proposition 3.2 we get that $\mu_t^\varepsilon = (X_t^\varepsilon)_\# \mu_0^\varepsilon$ for a.e. t. This representation allows us to estimate the distance between μ_t^ε and μ_s^ε: indeed $(X_t^\varepsilon, X_s^\varepsilon)_\# \mu_0^\varepsilon \in \mathcal{Adm}(\mu_t^\varepsilon, \mu_s^\varepsilon)$ and therefore

$$W^2(\mu_t^\varepsilon, \mu_s^\varepsilon) \leq \int |X_t^\varepsilon - X_s^\varepsilon|^2 d\mu_0^\varepsilon \leq (s-t)\int\int_t^s |\dot{X}_r^\varepsilon| dr d\mu_0^\varepsilon = (s-t)\int_t^s \int |v_r^\varepsilon| d\mu_r^\varepsilon dr$$

for any $0 \leq t \leq s \leq 1$. Since clearly μ_t^ε $W-$converge to μ_t for every t as ε goes to 0, and since by lemma 3.3 (with $E := v_t\mu_t$) it holds $\|v_t^\varepsilon\|_{\mu_t^\varepsilon} \leq \|v_t\|_{\mu_t}$, the previous inequality implies

$$W^2(\mu_t, \mu_s) \leq (s-t)\int\int_t^s |v_r| d\mu_r dr,$$

which is the thesis. □

This theorem allows to prove the well-know Benamou-Brenier formula (see also [**15**]) for the Wasserstein distance:

$$W^2(\mu^0, \mu^1) = \inf \int_0^1 \|v_t\|^2_{\mu_t} dt,$$

where the infimum is taken among all the absolutely continuous curves such that $\mu_0 = \mu^0$, $\mu_1 = \mu^1$ and among all vector fields v_t such that (3.1) holds. Indeed on one hand it holds

$$\left(\int_0^1 \|v_t\|^2_{\mu_t} dt\right)^{1/2} \geq \int_0^1 \|v_t\|_{\mu_t} dt \geq \mathcal{L}(\mu_t) \geq W(\mu_0, \mu_1),$$

where in the second inequality we used the second part od the theorem.

On the other hand, taking as μ_t any constant speed geodesic connecting μ^0 to μ^1 and using the first part of the theorem we get

$$\left(\int_0^1 \|v_t\|^2_{\mu_t} dt\right)^{1/2} = \int_0^1 \|v_t\|_{\mu_t} dt \le \mathcal{L}(\mu_t) = W(\mu_0, \mu_1),$$

2. The regular tangent space

Let us analyze some of the consequences of theorem 3.4. Given an absolutely continuous curve $t \to \mu_t$ the set of those families of vectors $v_t \in L^2(\mu_t, \mathbb{R}^d)$ satisfying (3.3) is not uniquely determined: indeed it is sufficient to choose $w_t \in L^2(\mu_t, \mathbb{R}^d)$ such that $\nabla\cdot(w_t\mu_t) = 0$ and to add these vectors to the v_t's to find another solution. However the following considerations suggest that there is a natural choice for the v_t's.

- The second implication of theorem 3.4 shows that for any solution v_t it holds $\|v_t\|_{\mu_t} \ge |\dot\mu_t|$.
- The first implication ensures that there exists at least one choice such that

$$\|v_t\|_{\mu_t} = |\dot\mu_t|. \tag{3.8}$$

- The linearity w.r.t. v_t of equation (3.3) and the strict convexity of the L^2 norms give the uniqueness of those v_t's which satisfy (3.8).
- Equation (3.4) ensures that the vectors v_t act only (through the Riesz isomorphism of $L^2(\mu_t, \mathbb{R}^d)$) on the gradient of regular functions.

These facts lead us to the following definition.

Definition 3.5 (Regular tangent space). *The* regular tangent space *(or simply tangent space) to the space $\mathscr{P}_2(\mathbb{R}^d)$ at a measure μ is*

$$\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) := \overline{\left\{\nabla\varphi \ : \ \varphi \in C_c^\infty(\mathbb{R}^d)\right\}}^{L^2(\mu,\mathbb{R}^d)}.$$

Tangent vectors are characterized by the following property:

$$\|v\|_\mu \le \|v + w\|_\mu, \qquad \forall w \in L^2(\mu, \mathbb{R}^d) \text{ such that } \nabla\cdot(w\mu) = 0. \tag{3.9}$$

Indeed (3.9) holds if and only if

$$\int \langle v, w\rangle d\mu = 0, \qquad \forall w \in L^2(\mu, \mathbb{R}^d) \text{ such that } \nabla\cdot(w\mu) = 0, \tag{3.10}$$

and this is true if and only if $v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. Equation (3.9) ensures that the two possible choices for the vectors v_t suggested by the previous considerations (solution of minimal norm or gradient) actually give the same vectors. We will denote by P_μ the orthogonal projection of $L^2(\mu, \mathbb{R}^d)$ into $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, which is obviously characterized by

$$\int \langle v - P_\mu(v), v'\rangle d\mu = 0 \quad \forall v' \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)). \tag{3.11}$$

Let us also introduce the *normal space*, i.e. the orthogonal complement of $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$:

$$\mathrm{Tan}_\mu^\perp(\mathscr{P}_2(\mathbb{R}^d)) := \Big\{ w \in L^2(\mu, \mathbb{R}^d) \ : \ \nabla \cdot (w\mu) = 0 \Big\}. \tag{3.12}$$

The analogous of equation (3.9) is: $w \in \mathrm{Tan}_\mu^\perp(\mathscr{P}_2(\mathbb{R}^d))$ if and only if

$$\|v\|_\mu \leq \|v + w\|_\mu, \qquad \forall v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)). \tag{3.13}$$

From now on given an absolutely continuous curve μ_t we will refer to the vectors $v_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ which solve (3.3) as the *velocity vectors* of the curve.

In order to enlight notation we will write $\langle \cdot, \cdot \rangle_\mu$ for the scalar product in $L^2(\mu, \mathbb{R}^d)$:

$$\langle v, v' \rangle_\mu := \int \langle v, v' \rangle d\mu \qquad \forall v, v' \in L^2(\mu, \mathbb{R}^d).$$

It is important to underline immediately that the structure given by the tangent spaces does not yield really an infinite dimensional Riemannian manifold. This is due to the fact that the *injectivity radius* of this “manifold” is 0. Before studying this problem we define the exponential map. Let us start with the following observation: if a constant speed geodesic $[0,1] \ni t \to \mu_t$ is induced by an optimal map T, that is $\mu_t = (Id + t(T - Id))_\# \mu_0$, then the continuity equation is fulfilled with

$$v_t = (T - Id) \circ (Id + t(T - Id))^{-1},$$

and since $\|v_t\|_{\mu_t} = \|T - Id\|_{\mu_0} = W(\mu_0, \mu_1)$ for any $t \in [0,1]$, those v_t are tangent (at least for a.e. t). Since $v_0 = T - Id$ we are led to the following definition:

Definition 3.6 (The exponential map). *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a measure. The exponential map is defined as*

$$L^2(\mu, \mathbb{R}^d) \ni v \to \exp_\mu(v) := (Id + v)_\# \mu. \tag{3.14}$$

Since the defining formula makes sense for any vector field in $L^2(\mu, \mathbb{R}^d)$ we didn't restrict the definition to the tangent space.

Let us now come back to the question about the injectivity radius: it is clear that the map $t \to \exp_\mu(tv)$ defines a geodesic in the interval $[0, \bar{t}]$ if and only if $Id + \bar{t}v$ is an optimal map, and this is true if and only if $\mathcal{I} + \bar{t}Dv \geq 0$ in distributional sense (here $\mathcal{I}$ is the identity matrix). So $t \to \exp_\mu(tv)$ defines a geodesic in some interval if and only if Dv is bounded from below, but since $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is defined as the $L^2(\mu, \mathbb{R}^d)$ closure of the set of gradients without taking in account any regularity, simple counterexamples show that Dv is generally unbounded.

The same argument shows that for any $\varphi \in C_c^\infty(\mathbb{R}^d)$ the map $Id + t\nabla\varphi$ is optimal for sufficiently small t and therefore it holds

$$\lim_{t \to 0^+} \frac{W(\exp_\mu(t\nabla\varphi), \mu)}{t} = \|\nabla\varphi\|_\mu.$$

Now, using the inequality

$$W(\exp_\mu(tv), \exp_\mu(t\nabla\varphi)) \leq t\|v - \nabla\varphi\|_\mu,$$

it is easy to check that for every $v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ it holds

$$\lim_{t\to 0^+} \frac{W(\exp_\mu(tv),\mu)}{t} = \|v\|_\mu,$$

i.e. the infinitesimal behavior of the exponential map on the tangent space is the right one. We will show in the next chapter (see theorem 4.19) that the previous formula characterizes the tangent vectors.

The next proposition proves an important property of velocity vectors: they can be recovered as limit of optimal transport plans (or maps, if any). It is important to underline that even if there is no hypothesis on the regularity of the measures μ_t, the limit plan is induced by the map v_t. It is for this reason that in the studies of differential properties along absolutely continuous curves it suffices to consider the regular tangent space.

Proposition 3.7. *Let μ_t be an absolutely continuous curve and let v_t be its velocity vectors. Then for almost every $t \in [0,1]$ it holds: for any choice of $\boldsymbol{\gamma}_t^{t+h} \in \mathit{Opt}(\mu_t,\mu_{t+h})$ we have*

$$\lim_{h\to 0} \left(\pi^1, \frac{\pi^2-\pi^1}{h}\right)_{\#} \boldsymbol{\gamma}_t^{t+h} \to (Id, v_t)_{\#}\mu_t, \tag{3.15}$$

and

$$\lim_{h\to 0} \frac{W(\mu_{t+h},(Id+hv_t)_{\#}\mu_t)}{h} = 0. \tag{3.16}$$

In particular for almost every $t \in [0,1]$ such that $\mu_t \in \mathscr{P}_2^r(\mathbb{R}^d)$ it holds

$$\frac{T_{\mu_t}^{\mu_{t+h}} - Id}{h} \to v_t, \qquad \text{in } L^2(\mu_t,\mathbb{R}^d). \tag{3.17}$$

Proof. Theorem 3.4 ensures that it is sufficient to prove (3.15) and (3.16) for every $t \in [0,1]$ such that

$$\frac{W(\mu_{t+h},\mu_t)}{h} \to |\dot\mu_t| = \|v_t\|_{\mu_t}. \tag{3.18}$$

So, fix t satisfying (3.18), define

$$\boldsymbol{\eta}_h := \left(\pi^1, \frac{\pi^2-\pi^1}{h}\right)_{\#} \boldsymbol{\gamma}_t^{t+h},$$

and observe that the family $\boldsymbol{\eta}_h$ is bounded. Letting $\boldsymbol{\eta}$ be any τ accumulation point of $\boldsymbol{\eta}_h$ as $h \to 0$, the inequality

$$\|\mathscr{B}(\boldsymbol{\eta})\|_{\mu_t} \le \|x_2\|_{\boldsymbol{\eta}} \le \varliminf_{h\to 0} \|x_2\|_{\boldsymbol{\eta}_h} = \lim \frac{W(\mu_{t+h},\mu_t)}{h} = \|v_t\|_{\mu_t},$$

and equations (3.9), (2.31) ensure that in order to prove (3.15) it suffices to show that

$$\nabla\cdot((\mathscr{B}(\boldsymbol{\eta})-v_t)\mu_t) = 0. \tag{3.19}$$

Now fix $\varphi \in C_c^\infty(\mathbb{R}^d)$, observe that

$$\begin{aligned}\frac{\mu_{t+h}(\varphi)-\mu_t(\varphi)}{h} &= \frac{1}{h}\int \varphi(x_2)-\varphi(x_1)d\boldsymbol{\gamma}_t^{t+h}\\ &= \frac{1}{h}\int \varphi(x_1+hx_2)-\varphi(x_1)d\boldsymbol{\eta}_h = \int \langle \nabla\varphi(x_1), x_2\rangle d\boldsymbol{\eta}_h + R_h,\end{aligned}$$

where $|R_h| \le h \operatorname{Lip}_{\nabla\varphi} \|x_2\|_{\boldsymbol{\eta}_h}$, and let h tend to 0 to obtain

$$\int \langle \nabla\varphi, v_t\rangle d\mu_t = \int \langle \nabla\varphi(x_1), x_2\rangle d\boldsymbol{\eta} = \int \langle \nabla\varphi, \mathscr{B}(\boldsymbol{\eta})\rangle d\mu_t.$$

The arbitrariness of φ gives (3.19).

Now we turn to (3.16). Observe that since

$$(\pi^1 + hv_t\circ\pi^1, \pi^2)_{\#}\boldsymbol{\gamma}_t^{t+h} \in \mathcal{Adm}((Id+hv)\mu_t, \mu_{t+h})$$

it holds

$$\frac{W^2(\mu_{t+h}, (Id+hv_t)_{\#}\mu_t)}{h^2} \le \frac{1}{h^2}\|x_1 + hv_t(x_1) - x_2\|^2_{\boldsymbol{\gamma}_t^{t+h}} = \|v_t(x_1) - x_2\|^2_{\boldsymbol{\eta}_h}.$$

Now observe that for any $\varphi \in C_c^\infty$, equation (3.15) and the fact that the function $(x_1, x_2) \to (\nabla\varphi(x_1) - x_2)$ has linear growth, give $\|\nabla\varphi(x_1) - x_2\|_{\boldsymbol{\eta}_h} \to \|\nabla\varphi(x_1) - v_t(x_1)\|_{\mu_t}$. Therefore we have

$$\begin{aligned}\lim_{h\to 0}\frac{W(\mu_{t+h}, (Id+hv_t)_{\#}\mu_t)}{h} &\le \lim_{h\to 0}\|v_t(x_1) - x_2\|_{\boldsymbol{\eta}_h}\\ &\le \|v_t - \nabla\varphi\|_{\mu_t} + \lim_{h\to 0}\|\nabla\varphi(x_1) - x_2\|_{\boldsymbol{\eta}_h}\\ &= 2\|v_t - \nabla\varphi\|_{\mu_t}.\end{aligned}$$

Letting $\nabla\varphi$ tend to v_t in $L^2_{\mu_t}$ we get the thesis.

The last statement is a consequence of (3.15) and proposition 2.23. □

As a first application of the definition of tangent space we discuss the differentiability properties of the functional $W^2(\cdot, \sigma)$.

Proposition 3.8 (Differentiability of W^2 at regular measures). *Let $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$ be a fixed measure. Then the functional*

$$\mathscr{P}_2(\mathbb{R}^d) \ni \mu \to W^2(\mu, \sigma),$$

is differentiable at any regular measure μ and its differential is $2(Id - T_\mu^\sigma)$, i.e. the following equality holds for any $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$:

$$\lim_{\nu \xrightarrow{W} \mu} \frac{W^2(\nu,\sigma) - W^2(\mu,\sigma) - \langle 2(Id - T_\mu^\sigma), T_\mu^\nu - Id\rangle_\mu}{W(\nu,\mu)} = 0. \tag{3.20}$$

Proof. The $\overline{\lim}$ inequality follows by

$$W^2(\nu,\sigma) \le \|T_\mu^\nu - T_\mu^\sigma\|^2_\mu = W^2(\mu,\sigma) + W^2(\mu,\nu) + 2\langle T_\mu^\nu - Id, Id - T_\mu^\sigma\rangle_\mu.$$

To get the other one, fix any sequence $\nu_n \to \mu$ and choose $\boldsymbol{\alpha}_n$ such that

$$\pi^{1,2}_{\#}\boldsymbol{\alpha}_n = (Id, T_\mu^{\nu_n})_{\#}\mu,$$
$$\pi^{2,3}_{\#}\boldsymbol{\alpha}_n \in \mathcal{O}pt(\nu_n, \sigma),$$

to obtain

$$\begin{aligned} W^2(\mu,\sigma) &\leq \|x_1 - x_3\|^2_{\boldsymbol{\alpha}_n} \\ &= W^2(\mu,\nu_n) + W^2(\nu_n,\sigma) + 2\int \langle x_1 - T_\mu^{\nu_n}(x_1), T_\mu^{\nu_n}(x_1) - x_3\rangle d\boldsymbol{\gamma}_n(x_1,x_3), \end{aligned}$$

where $\boldsymbol{\gamma}_n := \pi^{1,3}_{\#}\boldsymbol{\alpha}_n$.

Conclude observing that since $\|x_2 - x_3\|_{\boldsymbol{\gamma}_n} \leq W(\mu,\nu_n) + W(\nu_n,\sigma)$, proposition 2.19 implies $\boldsymbol{\gamma}_n \stackrel{W}{\to} (Id, T_\mu^\sigma)_{\#}\mu$ and so it holds

$$\begin{aligned} &\left| \int \langle x_1 - T_\mu^{\nu_n}(x_1), T_\mu^{\nu_n}(x_1) - x_3\rangle d\boldsymbol{\gamma}_n(x_1,x_3) - \int \langle x_1 - T_\mu^{\nu_n}(x_1), x_1 - T_\mu^\sigma(x_1)\rangle d\mu(x_1)\right| \\ &\qquad = \left| \int \langle x_1 - T_\mu^{\nu_n}(x_1), T_\mu^{\nu_n}(x_1) - x_3 - x_1 + T_\mu^\sigma(x_1)\rangle d\boldsymbol{\gamma}_n \right| \\ &\qquad \leq W(\mu,\nu_n)\|T_\mu^{\nu_n}(x_1) - x_3 - x_1 + T_\mu^\sigma(x_1)\|_{\boldsymbol{\gamma}_n} = o(W(\mu,\nu_n)). \end{aligned}$$

□

It is interesting to notice the following formal analogy with the usual Riemannian case: if M is a Riemannian manifold, the differential of the function $x \to d^2(x,z)$, provided it exists, is $-2\exp_x^{-1}(z)$. This means that it holds

$$\lim_{y\to x} \frac{d^2(y,z) - d^2(x,z) - \langle -2\exp_x^{-1}(z), \exp_x^{-1}(y)\rangle_x}{d(x,y)} = 0,$$

where $\langle\cdot,\cdot\rangle_x$ is the scalar product on T_xM.

Now observe that in the space $(\mathscr{P}_2(\mathbb{R}^d), W)$, for a given $\mu \in \mathscr{P}_2^r(\mathbb{R}^d)$, the natural right inverse of $v \to \exp_\mu(v) = (Id + v)_{\#}\mu$ is $\nu \to \exp_\mu^{-1}(\nu) := T_\mu^\nu - Id$. This means that $T_\mu^\nu - Id$ is the only map $f \in L^2_\mu$ such that $[0,1] \ni t \to \exp_\mu(tf)$ is a geodesic connecting μ to ν. Therefore equation (3.20) is the perfect analogous of the formula valid in the usual Riemannian setting. We will investigate in the next Chapter how to generalize the concept of $\exp_\mu^{-1}$ to non-regular measures, and in Chapter 5 the general differentiability properties of $\mu \to W^2(\mu,\sigma)$ (see theorem 5.20 and corollary 5.22).

A simple consequence of the previous proposition and of (3.15) is the following formula for the derivative of $W^2(\cdot,\sigma)$ along absolutely continuous curves with range on regular measures:

$$\frac{d}{dt}W^2(\mu_t,\sigma) = 2\langle v_t, Id - T_{\mu_t}^\sigma\rangle_{\mu_t}, \quad \text{a.e. } t \in [0,1]$$

Since, as we will see in corollary 5.22, W^2 is generally *not* differentiable on non regular measures, it is quite surprising that a similar formula holds even without any assumption on μ_t:

PROPOSITION 3.9. *Let $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve, v_t its velocity vector field and $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$ a fixed measure. Then for a.e. t it holds:*

$$\frac{d}{dt} W^2(\mu_t, \sigma) = 2 \int \langle v_t(x_1), x_1 - x_2 \rangle d\boldsymbol{\gamma}(x_1, x_2), \quad \forall \boldsymbol{\gamma} \in \mathcal{O}pt(\mu_t, \sigma). \tag{3.21}$$

Proof. Observe that the map $t \to W(\mu_t, \sigma)$ is absolutely continuous, so the same is true for $t \to W^2(\mu_t, \sigma)$ and the left hand side of (3.21) exists for a.e. t. We will show that claimed property is true for every t such that (3.16) holds and the derivative of $W(\mu_t, \sigma)$ exists; fix such a t and observe that equation (3.16) ensures that

$$\frac{d}{ds} W^2(\mu_s, \sigma)_{|s=t} = \lim_{h \to 0} \frac{W^2((Id + hv_t)_\# \mu_t, \sigma) - W^2(\mu_t, \sigma)}{h}.$$

Now choose any $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu_t, \sigma)$ and note that since $(\pi^1 + hv_t \circ \pi^1, \pi^2)_\# \boldsymbol{\gamma} \in \mathcal{A}dm((Id + hv_t)_\# \mu_t, \sigma)$ it holds

$$\begin{aligned} W^2((Id + hv_t)_\# \mu_t, \sigma) \leq & \|x_1 + hv_t(x_1) - x_2\|^2_{\boldsymbol{\gamma}} \\ = & W^2(\mu_t, \sigma) + h^2 \|v_t\|^2_{\mu_t} + 2h \int \langle v_t(x_1), x_1 - x_2 \rangle d\boldsymbol{\gamma} \\ = & W^2(\mu_t, \sigma) + h^2 \|v_t\|^2_{\mu_t} - 2h \langle v_t, \mathscr{B}(\boldsymbol{\gamma}) - Id \rangle_{\mu_t}, \end{aligned}$$

and the thesis follows dividing by h and letting $h \uparrow 0$ and $h \downarrow 0$. □

Observe that the equation (3.21) may be written as

$$\frac{d}{dt} W^2(\mu_t, \sigma) = 2 \langle v_t, Id - \mathscr{B}(\boldsymbol{\gamma}) \rangle_{\mu_t}.$$

We will show in theorem 4.15 that the barycentric projection of an optimal plan belongs to the regular tangent space, and usually different optimal plans have different barycentric projections: so it is a non trivial fact that almost everywhere along an absolutely continuous curve the barycentric projections of optimal plans act in the same way on the velocity vector field.

CHAPTER 4

The geometric tangent space

1. An introduction to $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$

The regular tangent space is a very useful tool in the study of differentiability properties along absolutely continuous curves or on regular measures, however the study of the properties of geodesically convex functionals (see the next chapter) shows that in order to reproduce in the framework of the Wasserstein distance the classical results of convex analysis, an enlargement of $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is necessary. Even from a theoretical point of view the definition of $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is not completely satisfactory: for instance if $\mu = \delta_x$ is a delta then $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is isomorphic to $\mathbb{R}^d$ (and so it is very small) and the range of the exponential map is just the set of all δ's. This is due, of course, to the fact that working with maps does not allow to *split* masses. In this section we will describe another tangent space which we call $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, made with plans, which will be a subset of

$$\mathscr{P}_2(\mathbb{R}^{2d})_\mu := \Big\{\gamma \in \mathscr{P}_2(\mathbb{R}^{2d}) \ : \ \pi^1_\# \gamma = \mu\Big\}, \tag{4.1}$$

and we will show that for any measure μ the map

$$\begin{array}{rcl} \mathfrak{I}_\mu : \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) & \to & \mathscr{P}_2(\mathbb{R}^{2d})_\mu \\ v & \to & (Id, v)_\# \mu, \end{array} \tag{4.2}$$

will be an embedding of $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ into $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, and that this embedding is actually an isomorphism if the base measure μ is regular.

Before going on we want to underline that this section is the most technical part of the thesis. Working with plans is not very intuitive, and many proofs will need the introduction of multiple plans of dimension 4 or even 5. However the structure of $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ is formally very similar to the one of Hilbert spaces: hoping for a better understanding we will sometimes recall the proof of some basic theorems in Hilbert setting before proving, with the same ideas, their analogous in $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$.

To understand the heuristic idea behind the definition of the geometric tangent space let us think for a moment to the case of a Riemannian manifold M embedded in some $\mathbb{R}^d$. The tangent space at point $x \in M$ may be defined in the following way: for any geodesic $t \to x(t)$ starting from x, i.e. $x(0) = x$, we try to find the limit in $\mathbb{R}^d$ of

$$\frac{x(t) - x(0)}{t}, \tag{4.3}$$

and, whenever this limit exists, we put it into the tangent space, which therefore will be the set of all possible limits of the previous kind.

Trying to do the same in the Wasserstein framework arise the following problems: what is the natural analogous for equation (4.3) in our setting? And: what is the topology we use to calculate the limit?

For regular base measure μ the answer to the first question should be

$$\frac{T_\mu^{\mu_t} - Id}{t}, \tag{4.4}$$

where $t \to \mu_t$ is a given geodesic starting from μ parametrised in some right neighborhood of 0. Remark 2.10, and in particular (2.12), ensures that for some optimal map T and $\bar{t} > 0$ it holds

$$\frac{T_\mu^{\mu_t} - Id}{t} = \frac{T - Id}{\bar{t}}, \qquad \text{for } t < \bar{t}.$$

In particular there is no problem in finding the limit and we can affirm that the geometric tangent space at a regular measure μ should contain at least all the maps of the form

$$v := \frac{T - Id}{\bar{t}},$$

where T varies among all possible optimal maps and $\bar{t}$ among positive reals. Now observe that since all the maps belong to $L^2(\mu, \mathbb{R}^d)$ it is natural to equip the tangent space with the L^2 distance w.r.t. μ. Since we want to deal with a complete tangent space we define

$$\mathcal{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) := \overline{\left\{ v \in L^2(\mu, \mathbb{R}^d) : Id + tv \text{ is optimal for some } t > 0 \right\}}^{L^2(\mu, \mathbb{R}^d)} \tag{4.5}$$

Now let us come back to the general case, i.e. $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ is not necessarily regular. Let $t \to \mu_t$ be a constant speed geodesic starting from μ and parametrised in some right neighborhood of 0. Keeping in mind the correspondence

$$T \to (Id, T)_\# \mu,$$

from maps to plans, and equation (4.4), we are led to try to find the limit of

$$\left(\pi^1, \frac{\pi^2 - \pi^1}{t} \right)_\# \gamma_t, \qquad \gamma_t \in \mathcal{Opt}(\mu, \mu_t). \tag{4.6}$$

Using corollary 2.9 we know that there exist an optimal transport plan $\overline{\gamma}$ and a positive $\bar{t}$ such that for $t < \bar{t}$ it holds

$$\gamma_t = \left(\pi^1, \pi^1 + \frac{t}{\bar{t}} (\pi^2 - \pi^1) \right)_\# \overline{\gamma},$$

from which we get that the expression in (4.6) is once again independent on small t and equal to

$$\gamma := \left(\pi^1, \frac{1}{\bar{t}} (\pi^2 - \pi^1) \right)_\# \overline{\gamma}.$$

Inverting this equation we define the following set:

$$\mathcal{GEOD}_\mu := \left\{ \gamma \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu \ : \ (\pi^1, \pi^1 + t\pi^2)_\# \gamma \text{ is optimal for some } t > 0 \right\}.$$

Analogously to the previous case we have

$$\mathcal{GEOD}_\mu \subset \mathscr{P}_2(\mathbb{R}^{2d})_\mu,$$

and we want to endow $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ with some distance and then define the geometric tangent space as the closure of $\mathcal{GEOD}_\mu$ w.r.t. this distance. Since we want the map $\mathfrak{I}_\mu$ defined in (4.2) to be an isometry, we define, for any couple $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$, the distance $W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ as

$$W_\mu^2(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) := \int W^2\big((\boldsymbol{\gamma}_1)_x, (\boldsymbol{\gamma}_2)_x\big) d\mu(x), \tag{4.7}$$

where $(\boldsymbol{\gamma}_1)_x, (\boldsymbol{\gamma}_2)_x$ are the disintegrations of $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ w.r.t. μ.

Finally we give the following definition:

DEFINITION 4.1 (The geometric tangent space). *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a measure. The* geometric tangent space $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *of $\mathscr{P}_2(\mathbb{R}^d)$ at μ is the closure of $\mathcal{GEOD}_\mu$ in $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$with respect to the distance W_μ.*

Actually we have still to show that W_μ is a distance and that the geometric tangent space is complete (this is not automatic from the definition since we have to prove at first that $(\mathscr{P}_2(\mathbb{R}^{2d})_\mu, W_\mu)$ is complete). These are the goals of the foregoing discussion.

From now on $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ is a fixed measure, and its variable will always be x_0. A generic plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ will therefore belong to $\mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_i)$ for some $i = 1, 2, \ldots$. For a given couple $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ the set of *admissible 3-plan* is defined as

$$\mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) := \Big\{\boldsymbol{\alpha} \in \mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_1 \times \mathbb{R}^d_2) \;:\; \pi^{0,1}_\#\boldsymbol{\alpha} = \boldsymbol{\gamma}_1,\ \pi^{0,2}_\#\boldsymbol{\alpha} = \boldsymbol{\gamma}_2\Big\}.$$

The *cost* associated to a 3-plan $\boldsymbol{\alpha}$ is

$$\|x_1 - x_2\|_{\boldsymbol{\alpha}} = \left(\int |x_1 - x_2|^2 d\boldsymbol{\alpha}\right)^{1/2},$$

and the proposition below justifies this definition.

PROPOSITION 4.2. *Let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be any couple of plans. Then*

$$W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \min_{\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)} \|x_1 - x_2\|_{\boldsymbol{\alpha}}. \tag{4.8}$$

Proof. Let $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ be any admissible 3-plan and $d\boldsymbol{\alpha} = d\mu \times d\boldsymbol{\alpha}_{x_0}$ be its disintegration w.r.t. μ. Since for μ-a.e. x_0 it holds $\boldsymbol{\alpha}_{x_0} \in \mathcal{Adm}((\boldsymbol{\gamma}_1)_{x_0}, (\boldsymbol{\gamma}_2)_{x_0})$ we have

$$\begin{aligned}\|x_1 - x_2\|^2_{\boldsymbol{\alpha}} &= \int \|x_1 - x_2\|^2 d\boldsymbol{\alpha} = \int \Big(\|x_1 - x_2\|^2 d\boldsymbol{\alpha}_{x_0}\Big) d\mu(x_0) \\ &\geq \int W^2\big((\boldsymbol{\gamma}_1)_{x_0}, (\boldsymbol{\gamma}_2)_{x_0}\big) d\mu(x_0) = W_\mu^2(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2).\end{aligned}$$

To obtain the other inequality use the approximation lemma below on the plans $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ and observe that for every $n \in \mathbb{N}$ the set $\mathcal{Opt}((\boldsymbol{\gamma}^n_1)_{x_0}, (\boldsymbol{\gamma}^n_2)_{x_0},)$ contains only one element $\boldsymbol{\alpha}^n_{x_0}$ and that, by proposition 2.19 the function $x_0 \to \boldsymbol{\alpha}^n_{x_0}$ is continuous from

$\mathbb{R}^d$ to $\mathscr{P}_2(\mathbb{R}^d)$ endowed with the weak topology given by the duality with C_c functions. Therefore there exists a 3-plan $\boldsymbol{\alpha}^n$ whose disintegration w.r.t. x_0 is $\boldsymbol{\alpha}^n_{x_0}$, and it holds

$$\|x_1 - x_2\|^2_{\boldsymbol{\alpha}^n} = \int \|x_1 - x_2\|^2_{\boldsymbol{\alpha}^n_{x_0}} d\mu = \int W^2((\boldsymbol{\gamma}^n_1)_{x_0}, (\boldsymbol{\gamma}^n_2)_{x_0}) d\mu = W^2(\boldsymbol{\gamma}^n_1, \boldsymbol{\gamma}^n_2).$$

Letting n tend to ∞ the proof is achieved. □

LEMMA 4.3. *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$. Then there exists a sequence of plans $(\boldsymbol{\gamma}^n) \subset \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ such that*

i) *$(\boldsymbol{\gamma}^n)_{x_0}$ is regular for $\mu -$ a.e. x_0,*

ii) *$x \to (\boldsymbol{\gamma}^n)_{x_0}$ is continuous as a function from $\mathbb{R}^d$ to $(\mathscr{P}_2(\mathbb{R}^d), W)$,*

iii) $\int W^2((\boldsymbol{\gamma}^n)_{x_0}, \boldsymbol{\gamma}_{x_0}) d\mu(x_0) \to 0.$

Proof. Let $X_N \subset \mathscr{P}_2(\mathbb{R}^d)$ be the set of measures whose second moment is bounded by N endowed with the distance W (which induces the weak topology given by the duality with C_c functions). Define $Y_N \subset \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ to be the set of those plans $\overline{\boldsymbol{\gamma}}$ such that $(\overline{\boldsymbol{\gamma}})_{x_0} \in X_N$ for $\mu-$a.e. x_0, and assume for a moment that $\boldsymbol{\gamma} \in Y_N$ for a certain N. Then we know that the Borel function from $\mathbb{R}^d$ to X_N given by $x_0 \to \boldsymbol{\gamma}_{x_0}$ may be approximated $\mu-$a.e. with a sequence of continuous functions $x_0 \to \boldsymbol{\sigma}^n_{x_0} \in X_N$. Now let (ρ^n) be any sequence of mollifiers on $\mathbb{R}^d$ and define $\boldsymbol{\gamma}^n_{x_0} := \boldsymbol{\sigma}^n_{x_0} * \rho^n$. It is clear that $\boldsymbol{\gamma}^n_{x_0} \in X_{N'}$ for some $N' \geq N$ independent on n, and that $x_0 \to \boldsymbol{\gamma}^n_{x_0}$ is continuous. We claim that the plans $\boldsymbol{\gamma}^n$ satisfy the thesis. We have only to show that $\int W^2((\boldsymbol{\gamma}^n)_{x_0}, \boldsymbol{\gamma}_{x_0}) d\mu(x_0)$ tends to 0. Observe that

$$\left(\int W^2((\boldsymbol{\gamma}^n)_{x_0}, \boldsymbol{\gamma}_{x_0}) d\mu(x_0)\right)^{1/2}$$
$$\leq \left(\int W^2((\boldsymbol{\gamma}^n)_{x_0}, \boldsymbol{\sigma}^n_{x_0}) d\mu(x_0)\right)^{1/2} + \left(\int W^2((\boldsymbol{\sigma}^n)_{x_0}, \boldsymbol{\gamma}_{x_0}) d\mu(x_0)\right)^{1/2},$$

and that the first integrand is bounded by $\int |x|^2 \rho^n(x) dx$ for every x_0, while the second one converges to 0 for $\mu-$a.e. x_0 (by construction of $\boldsymbol{\sigma}^n_{x_0}$) and is bounded by $4N^2$. The dominated convergence theorem gives the thesis.

For the general case it is sufficent to approximate at first $\boldsymbol{\gamma}$ with plans $\overline{\boldsymbol{\gamma}}^n \in Y_n$ and to proceed with a standard diagonalization argument. □

The previous proposition shows that the infimum is always attained. We will denote by $\mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ the set of those 3-plans which realize the minimum, i.e.:

$$\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \iff W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \|x_1 - x_2\|_{\boldsymbol{\alpha}}. \tag{4.9}$$

The proof of the previous proposition characterizes optimal 3-plans as follows

$$\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \iff \boldsymbol{\alpha}_{x_0} \in \mathcal{O}pt((\boldsymbol{\gamma}_1)_{x_0}, (\boldsymbol{\gamma}_2)_{x_0}), \quad \mu\text{-a.e. } x_0.$$

THEOREM 4.4. *The function W_μ is a distance on $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$.*

Proof. Choosing $\boldsymbol{\alpha} := (\pi^0, \pi^1, \pi^1)_\# \boldsymbol{\gamma}$ it follows that $W_\mu(\boldsymbol{\gamma}, \boldsymbol{\gamma}) = 0$. Conversely suppose that $W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = 0$ and choose an optimal 3-plan $\boldsymbol{\alpha}$. The equality $\|x_1 - x_2\|_{\boldsymbol{\alpha}} = 0$ gives that for $\boldsymbol{\alpha}$-a.e. triple (x_0, x_1, x_2) it holds $x_1 = x_2$. In particular the two functions $\pi^{0,1}, \pi^{0,2}$ are equal in $L^2(\boldsymbol{\alpha}, \mathbb{R}^d)$ and therefore $\boldsymbol{\gamma}_1 = \pi^{0,1}_\# \boldsymbol{\alpha} = \pi^{0,2}_\# \boldsymbol{\alpha} = \boldsymbol{\gamma}_2$.

The equality $W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = W_\mu(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_1)$ is clear.

To get the triangle inequality argue as in the proof of triangle inequality for W: given $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ choose $\boldsymbol{\alpha}_{1,2} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, $\boldsymbol{\alpha}_{2,3} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$ and find a 4-plan $\boldsymbol{\beta}$ satisfying

$$\pi^{0,1,2}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_{1,2},$$
$$\pi^{0,2,3}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_{2,3}.$$

Since $\pi^{0,1,3}_\# \boldsymbol{\beta} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_3)$ we have

$$W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_3) \leq \|x_1 - x_3\|_{\boldsymbol{\beta}} \leq \|x_1 - x_2\|_{\boldsymbol{\beta}} + \|x_2 - x_3\|_{\boldsymbol{\beta}} = W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) + W_\mu(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3).$$

Finally since $W_\mu(\boldsymbol{\gamma}, (Id, 0)_\# \mu) = \|x_1\|_{\boldsymbol{\gamma}} < \infty$, the triangular inequality ensures that $W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) < +\infty$ for every couple $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$. □

THEOREM 4.5 (l.s.c. of W_μ and completeness of $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$). *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a measure. Then the distance $W_\mu(\cdot, \cdot)$ is lower semicontinuous with respect to the weak convergence in $\mathscr{P}_2(\mathbb{R}^{2d})$ and the space $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ is complete with respect to the distance $W_\mu(\cdot, \cdot)$.*

Proof. Let $(\boldsymbol{\gamma}^1_n), (\boldsymbol{\gamma}^2_n) \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be two sequences weakly converging to $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2$ in $\mathscr{P}_2(\mathbb{R}^{2d})$ respectively and let $\boldsymbol{\alpha}_n$ be optimal 3-plans between $\boldsymbol{\gamma}^1_n$ and $\boldsymbol{\gamma}^2_n$. It is easy to see that the sequence $\boldsymbol{\alpha}_n$ is tight in $\mathbb{R}^{3d}$ and so there exists a subsequence $(\boldsymbol{\alpha}_{n_k})$ weakly converging in $\mathscr{P}_2(\mathbb{R}^{3d})$ to a certain 3-plan $\boldsymbol{\alpha}$ and such that $\lim_k \|x_1 - x_2\|_{\boldsymbol{\alpha}_{n_k}} = \underline{\lim}_n \|x_1 - x_2\|_{\boldsymbol{\alpha}_n}$. Given that the limit 3-plan $\boldsymbol{\alpha}$ satisfies $\pi^{0,1}_\# \boldsymbol{\alpha} = \boldsymbol{\gamma}^1, \pi^{0,2}_\# \boldsymbol{\alpha} = \boldsymbol{\gamma}^2$, we have:

$$W_\mu\left(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2\right) \leq \|x_1 - x_2\|_{\boldsymbol{\alpha}} \leq \lim_{k\to\infty} \|x_1 - x_2\|_{\boldsymbol{\alpha}_{n_k}} = \varliminf_{n\to\infty} \|x_1 - x_2\|_{\boldsymbol{\alpha}_n} = \varliminf_{n\to\infty} W_\mu\left(\boldsymbol{\gamma}^1_n, \boldsymbol{\gamma}^2_n\right).$$

Now we turn to the completeness: assume that $(\boldsymbol{\gamma}_n)$ is a Cauchy sequence. Being this sequence bounded in $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ it is also tight and so there exists a subsequence $(\boldsymbol{\gamma}_{n_k})$ weakly converging to a plan $\boldsymbol{\gamma}$. The lower semicontinuity yields

$$\varlimsup_{m\to\infty} W_\mu(\boldsymbol{\gamma}_m, \boldsymbol{\gamma}) \leq \varlimsup_{m\to\infty} \varliminf_{k\to\infty} W_\mu\left(\boldsymbol{\gamma}_m, \boldsymbol{\gamma}_{n_k}\right) = 0.$$

□

COROLLARY 4.6. $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *is a complete metric space.*

It should be noted carefully that the topology induced by the distance W_μ is *not* the one induced by W on $\mathscr{P}_2(\mathbb{R}^{2d})$. For instance let $d = 1$, $\mu := \mathcal{L}^1|_{[0,1]}$ and $v_n := \sin(nx) \in L^2(\mu, \mathbb{R})$. Then it is not difficult to show that $n \to \mathfrak{I}_\mu(v_n)$ does not have any converging subsequence in $(\mathscr{P}_2(\mathbb{R}^{2d})_\mu, W_\mu)$. On the other hand some calculations show that $\mathfrak{I}_\mu(v_n)$ tends to $(1 - y^2)^{-1/2} \cdot \mathcal{L}^2|_{[0,1]^2}$ in $(\mathscr{P}_2(\mathbb{R}^{2d})_\mu, W)$.

There is a special case in which the two topologies give the same converging sequences: when the limit plan is induced by a map.

PROPOSITION 4.7. *Let $\boldsymbol{\gamma}_n, \boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be plans, and assume that $\boldsymbol{\gamma}$ s induced by a map. Then $\boldsymbol{\gamma}_n$ $W_\mu-$converge to $\boldsymbol{\gamma}$ if and only if $\boldsymbol{\gamma}_n$ $W-$converge to $\boldsymbol{\gamma}$.*

Proof. Since $W_\mu \geq W$ one implication is obvious. For the converse one, take a sequence of continuous functions f_n with bounded support such that $f_n \to f$ in L^2_μ, where f is the function which induces $\boldsymbol{\gamma}$. Observe the functions $|y - f_k(x)|^2$ are continuous on $\mathbb{R}^{2d}$, bounded from below and with quadratic growth, therefore propositions 2.5, 1.4 ensure that

$$\lim_{n\to\infty} W_\mu(\boldsymbol{\gamma}_n, \mathfrak{I}(f_k)) = \lim_{n\to\infty} \|y - f_k(x)\|_{\boldsymbol{\gamma}_n} = \|y - f_k(x)\|_{\boldsymbol{\gamma}} = \|f - f_k\|_\mu.$$

Therefore it holds

$$\overline{\lim_{n\to\infty}} W_\mu(\boldsymbol{\gamma}_n, \boldsymbol{\gamma}) \leq \overline{\lim_{k\to\infty}}\ \overline{\lim_{n\to\infty}} \Big(W_\mu(\boldsymbol{\gamma}_n, \mathfrak{I}(f_k)) + W_\mu(\mathfrak{I}(f_k), \boldsymbol{\gamma}) \Big) \leq 2 \overline{\lim_{k\to\infty}} \|f - f_k\|_\mu = 0.$$

□

2. Directional derivative of $W^2(\cdot, \nu)$

In this section we describe the first differentiability properties of $W^2(\cdot,\nu)$ along curves of the kind $[\boldsymbol{\gamma}](t)$, where $\boldsymbol{\gamma}$ is any plan in $\mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^d_2)$, not necessarily optimal.

Let us first observe that using lemma 2.12 and reasoning as in the proof of 2.13 it is possible to prove an inequality like (2.16) without any assumption on $\boldsymbol{\gamma}$:

$$W^2([\boldsymbol{\gamma}(t)], \nu) \geq (1-t)W^2([\boldsymbol{\gamma}](0), \nu) + tW^2(\boldsymbol{\gamma}[1], \nu) - t(1-t)\|x_1 - x_2\|^2_{\boldsymbol{\gamma}}. \tag{4.10}$$

It is just a matter of calculations to show that this equation implies that the function

$$t \to W^2([\boldsymbol{\gamma}](t), \nu) - t^2\|x_1 - x_2\|_{\boldsymbol{\gamma}}$$

is concave. Therefore from the general theory of concave function we know that for every $t \in [0, 1)$ there exists the right derivative

$$\frac{d}{dt_+} W^2([\boldsymbol{\gamma}](t), \nu) := \lim_{h\to 0^+} \frac{W^2([\boldsymbol{\gamma}](t+h), \nu) - W^2([\boldsymbol{\gamma}](t), \nu)}{h},$$

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

$$\frac{d}{dt_-} W^2([\boldsymbol{\gamma}](t), \nu) := \lim_{h\to 0^-} \frac{W^2([\boldsymbol{\gamma}](t+h), \nu) - W^2([\boldsymbol{\gamma}](t), \nu)}{h},$$

and that

$$\frac{d}{dt_+} W^2([\boldsymbol{\gamma}](t), \nu) \leq \frac{d}{dt_-} W^2([\boldsymbol{\gamma}](t), \nu), \quad \forall t \in (0, 1).$$

Moreover for every t not belonging to the set $\mathcal{N}$ of points of non differentiability (which is at most countable), it holds

$$\frac{d}{dt_+} W^2([\boldsymbol{\gamma}](t), \nu) = \frac{d}{dt_-} W^2([\boldsymbol{\gamma}](t), \nu).$$

LEMMA 4.8. *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^d_2)$ be a plan and let ν be a fixed measure. Then for every 3-plan $\boldsymbol{\mu}_t$ satisfying*

$$\pi^{1,2}_{\#}\boldsymbol{\mu}_t = \boldsymbol{\gamma}, \tag{4.11a}$$

$$((1-t)\pi^1 + t\pi^2, \pi^3)_{\#}\boldsymbol{\mu}_t \in \mathit{Opt}([\boldsymbol{\gamma}](t), \nu), \tag{4.11b}$$

it holds

$$\begin{aligned}\frac{d}{dt_+}W^2([\boldsymbol{\gamma}](t),\nu) \leq& \|x_2 - x_3\|^2_{\boldsymbol{\mu}_t} - \|x_1 - x_3\|^2_{\boldsymbol{\mu}_t} + (2t-1)\|x_1 - x_2\|^2_{\boldsymbol{\mu}_t} \\ \leq& \frac{d}{dt_-}W^2([\boldsymbol{\gamma}](t),\nu).\end{aligned} \tag{4.12}$$

In particular both inequalities are equalities if t belongs to the set of differentiability of the function.

Proof. Use lemma 2.12 to find, for every $t \in [0,1]$, a 3-plan $\boldsymbol{\mu}_t$ satisfying

$$\pi^{1,2}_{\#}\boldsymbol{\mu}_t = \boldsymbol{\gamma},$$
$$((1-t)\pi^1 + t\pi^2, \pi^3)_{\#}\boldsymbol{\mu}_t \in \mathit{Opt}([\boldsymbol{\gamma}](t), \nu).$$

Now fix t_0 and take derivatives on

$$\|(1-t)x_1 + tx_2 - x_3\|^2_{\boldsymbol{\mu}_{t_0}} = (1-t)\|x_1 - x_3\|^2_{\boldsymbol{\mu}_{t_0}} + t\|x_2 - x_3\|^2_{\boldsymbol{\mu}_{t_0}} - t(1-t)\|x_1 - x_2\|^2_{\boldsymbol{\mu}_{t_0}},$$

to get

$$\frac{d}{dt}\|(1-t)x_1 + tx_2 - x_3\|^2_{\boldsymbol{\mu}_{t_0}} = \|x_2 - x_3\|^2_{\boldsymbol{\mu}_{t_0}} - \|x_1 - x_3\|^2_{\boldsymbol{\mu}_{t_0}} + (2t-1)\|x_1 - x_2\|^2_{\boldsymbol{\mu}_{t_0}}.$$

To conclude it is sufficient to observe that

$$W^2([\boldsymbol{\gamma}](t),\nu) \leq \|(1-t)x_1 + tx_2 - x_3\|^2_{\boldsymbol{\mu}_{t_0}},$$
$$W^2([\boldsymbol{\gamma}](t_0),\nu) = \|(1-t_0)x_1 + t_0x_2 - x_3\|^2_{\boldsymbol{\mu}_{t_0}}.$$

□

From this lemma it is possible to derive a precise formula for the directional derivative. Let us first introduce the following set:

DEFINITION 4.9. *Given a plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^d_2)$ and measure $\nu \in \mathscr{P}_2(\mathbb{R}^d_3)$, the set $\mathit{Opt}(\boldsymbol{\gamma},\nu)$ is the set of 3-plans $\boldsymbol{\mu}$ satisfying*

$$\pi^{1,2}_{\#}\boldsymbol{\mu} = \boldsymbol{\gamma},$$
$$\pi^{1,3}_{\#}\boldsymbol{\mu} \in \mathit{Opt}(\pi^1_{\#}\boldsymbol{\gamma}, \nu).$$

PROPOSITION 4.10. *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^d_1 \times \mathbb{R}^d_2)$ be a plan and let $\nu \in \mathscr{P}_2(\mathbb{R}^d_3)$ be a fixed measure. Then the following formula holds:*

$$\frac{d}{dt}|_{t=0}W^2([\boldsymbol{\gamma}](t),\nu) = \inf_{\boldsymbol{\mu}\in \mathit{Opt}(\boldsymbol{\gamma},\nu)} \|x_2 - x_3\|^2_{\boldsymbol{\mu}} - \|x_1 - x_2\|^2_{\boldsymbol{\gamma}} - W^2([\boldsymbol{\gamma}](0),\nu) \tag{4.13a}$$

$$= -2\sup_{\boldsymbol{\mu}\in \mathit{Opt}(\boldsymbol{\gamma},\nu)} \langle x_2 - x_1, x_3 - x_1\rangle_{\boldsymbol{\mu}}. \tag{4.13b}$$

Proof. Observe that a plan $\boldsymbol{\mu}$ realizes the inf in (4.13a) if and only if realizes the sup in (4.13b).

The previous lemma ensures that for any $\boldsymbol{\mu} \in \mathcal{O}pt(\boldsymbol{\gamma}, \nu)$ it holds

$$\begin{aligned}\frac{d^+}{dt} W^2([\boldsymbol{\gamma}](t), \nu)_{|t=0} \leq & \|x_2 - x_3\|^2_{\boldsymbol{\mu}} - \|x_1 - x_2\|^2_{\boldsymbol{\gamma}} - W^2([\boldsymbol{\gamma}](0), \nu) \\ = & -2\langle x_2 - x_1, x_3 - x_1\rangle_{\boldsymbol{\mu}}.\end{aligned}$$

To prove the converse inequality choose, for any $t \in [0, 1]$, a 3-plan $\boldsymbol{\mu}_t$ satisfying (4.11), observe that this family is 2-bounded in $\mathscr{P}_2(\mathbb{R}^{3d})$ and let $\boldsymbol{\mu}_0$ be any weak accumulation point. Remark 1.1 and proposition 1.11 give $\boldsymbol{\mu}_0 \in \mathcal{O}pt(\boldsymbol{\gamma}, \nu)$, moreover from

$$W^2([\boldsymbol{\gamma}](t), \nu) = \|(1-t)x_1 + tx_2 - x_3\|^2_{\boldsymbol{\mu}_t} = \|x_1 - x_3\|^2_{\boldsymbol{\mu}_t} + o(1),$$

we get $\|x_1 - x_3\|^2_{\boldsymbol{\mu}_t} \to W^2([\boldsymbol{\gamma}](0), \nu)$ as $t \to 0^+$. We conclude with:

$$\begin{aligned}\frac{d}{dt_+}_{|t=0} W^2([\boldsymbol{\gamma}](t), \nu) = & \lim_{t \to 0^+} \frac{d}{dt_-} W^2([\boldsymbol{\gamma}](t), \nu) \\ \geq & \lim_{t \to 0^+} \|x_2 - x_3\|^2_{\boldsymbol{\mu}_t} - \|x_1 - x_3\|^2_{\boldsymbol{\mu}_t} + (2t-1)\|x_1 - x_2\|^2_{\boldsymbol{\mu}_t} \\ \geq & \|x_2 - x_3\|^2_{\boldsymbol{\mu}_0} - W^2([\boldsymbol{\gamma}](0), \nu) - \|x_1 - x_2\|^2_{\boldsymbol{\gamma}}.\end{aligned}$$

□

Before going deeper on the analysis of the properties of $(\mathscr{P}_2(\mathbb{R}^{2d})_\mu, W_\mu)$, we open a brief parenthesis and we describe in the following subsection another possible construction of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. As we will see, the definitions we gave are consistent with the abstract construction of tangent space for PC spaces.

3. Tangent cone for PC spaces

Recall that a length space (E, d) is positively curved in the sense of Aleksandrov if for any constant speed geodesic $x_t : [0, 1] \to E$ and every $y \in E$ it holds

$$d^2(x_t, y) \geq (1-t)d^2(x_0, y) + td^2(x_1, y) - t(1-t)d^2(x_0, x_1), \tag{4.14}$$

and that in a *flat* Hilbertian setting the equality always holds in the above expression.

In a Euclidean setting the angle $0 \leq \theta_x(y, z) \leq \pi$ between the two segments joining x to y and z respectively, is given by the formula

$$\cos\big(\theta_x(y, z)\big) := \frac{\langle y - x, z - x\rangle}{|y - x||z - x|} = \frac{|y - x|^2 + |z - x|^2 - |y - z|^2}{2|y - x||z - x|},$$

and it is easy to see that if $y_t := (1-t)x + ty$, $z_t := (1-t)x + tz$ are the segment joining x to y and z respectively, then it holds

$$\theta_x(y, z) = \theta_x(y_t, z_s), \quad \forall t, s \in (0, 1].$$

If we try to do the same in a PC space we are naturally lead to define the angle $0 \leq \theta_x(y, z) \leq \pi$ from which the couple (y, z) is seen from x as:

$$\cos\big(\theta_x(y, z)\big) := \frac{d^2(x, y) + d^2(x, z) - d^2(y, z)}{2d(x, y)d(x, z)}.$$

Now let $\boldsymbol{y}, \boldsymbol{z} : [0,1] \to E$ be constant speed geodesics joining x to y and z respectively, it is just a matter of calculations to see that both the functions

$$t \to \cos\Big(\theta_x(\boldsymbol{y}(t), \boldsymbol{z}(s))\Big),$$
$$s \to \cos\Big(\theta_x(\boldsymbol{y}(t), \boldsymbol{z}(s))\Big),$$

are non increasing. It is then possible and natural to define the angle $0 \le \theta_x(\boldsymbol{y}, \boldsymbol{z}) \le \pi$ between the constant speed geodesics $\boldsymbol{y}$ and $\boldsymbol{z}$ as

$$\theta_x(\boldsymbol{y}, \boldsymbol{z}) := \sup_{t,s \in (0,1]} \theta_x(\boldsymbol{y}(t), \boldsymbol{z}(s)) = \lim_{t,s \to 0^+} \theta_x(\boldsymbol{y}(t), \boldsymbol{z}(s)). \tag{4.15}$$

In this formula the geodesics do not need to be defined in the whole interval $[0,1]$: if we let $[0, T_{\boldsymbol{y}}]$ be the domain of $\boldsymbol{y}$ we can still define $\theta_x(\boldsymbol{y}, \boldsymbol{z})$ as the limit of $\theta_x(\boldsymbol{y}(t), \boldsymbol{z}(s))$ as $t, s \to 0^+$, or as the supremum of $\theta_x(\boldsymbol{y}(t), \boldsymbol{z}(s))$ among all $t \in (0, T_{\boldsymbol{y}}]$ and $s \in (0, T_{\boldsymbol{z}}]$. Observe moreover that $d(\boldsymbol{y}(t), x)/t$ is independent on $t \in (0, T_{\boldsymbol{y}}]$ and equal to the (constant) metric derivative $|\dot{\boldsymbol{y}}|$ of $\boldsymbol{y}$.

Following the analogy with the Hilbert case we define:

$$\|\boldsymbol{y}\|_x := |\dot{\boldsymbol{y}}|, \tag{4.16a}$$
$$\langle \boldsymbol{y}, \boldsymbol{z} \rangle_x := \|\boldsymbol{y}\|_x \|\boldsymbol{z}\|_x \cos\big(\theta_x(\boldsymbol{y}, \boldsymbol{z})\big), \tag{4.16b}$$
$$d_x^2(\boldsymbol{y}, \boldsymbol{z}) := \|\boldsymbol{y}\|_x^2 + \|\boldsymbol{z}\|_x^2 - 2\langle \boldsymbol{y}, \boldsymbol{z} \rangle_x. \tag{4.16c}$$

Let us now introduce the set $\mathscr{Geod}_x$ of all constant speed geodesics starting from x; on this set we define the equivalence relation

$$\boldsymbol{y} \sim \boldsymbol{z} \quad \Leftrightarrow \quad \boldsymbol{y}|_{[0,t]} = \boldsymbol{z}|_{[0,t]} \quad \text{for some } t > 0.$$

Since the objects defined in the formulas (4.16a), (4.16b) and (4.16c) depend only on the behavior of the geodesics near 0, it is clear that the function $d_x(\cdot,\cdot)$ induces a function, still denoted by $d_x(\cdot,\cdot)$, on $\mathscr{Geod}_x/\sim$: we are going to prove that this function is a distance.

Since

$$\begin{aligned} \cos\big(\theta_x(\boldsymbol{y}, \boldsymbol{z})\big) &= \lim_{t,s \to 0^+} \frac{d^2(\boldsymbol{y}(t), x) + d^2(\boldsymbol{z}(s), x) - d^2(\boldsymbol{y}(t), \boldsymbol{z}(s))}{2d(\boldsymbol{y}(t), x) d(\boldsymbol{z}(t), s)} \\ &= \frac{1}{2\|\boldsymbol{y}\|_x \|\boldsymbol{z}\|_x} \left(\lim_{t,s \to 0^+} \frac{t^2\|\boldsymbol{y}\|_x^2 + s^2\|\boldsymbol{z}\|_x^2 - d^2(\boldsymbol{y}(t), \boldsymbol{z}(s))}{ts} \right), \end{aligned} \tag{4.17}$$

applying definitions (4.16c) we obtain

$$d_x(\boldsymbol{y}, \boldsymbol{z}) = \lim_{t \to 0^+} \frac{d(\boldsymbol{y}(t), \boldsymbol{z}(t))}{t} = \sup_{t \in [0,T]} \frac{d(\boldsymbol{y}(t), \boldsymbol{z}(t))}{t}, \tag{4.18}$$

where $T := \min\{T_{\boldsymbol{y}}, T_{\boldsymbol{z}}\}$, and therefore d_x is a distance as claimed.

Definition 4.11 (Tangent cone for *PC* spaces). *The tangent cone to the space (E,d) at a point $x \in E$ is the completion of $\mathscr{Geod}_x/\sim$ with respect to the distance d_x.*

It is not difficult to check that the functions defined in (4.16) induce functions on $\mathcal{Geod}_x$ and that they extend continuously to $\mathbf{Tan}_x(E)$.

We now show that if we apply this abstract definition to the space $\mathscr{P}_2(\mathbb{R}^d)$, the result is consistent with the definition we gave in the first section.

THEOREM 4.12. *Let $(E, d) := (\mathscr{P}_2(\mathbb{R}^d), W)$. Then the tangent cone just defined is canonically isometric to $(\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), W_\mu)$. The isometry is the unique continuous extension of the following bijective map from $\mathcal{GEOD}_\mu$ to $\mathcal{Geod}\,\mu/\sim$: a plan $\boldsymbol{\gamma}$ correspond to the equivalence class of the geodesic $t \to (\pi^1 + t\pi^2)_\#\boldsymbol{\gamma}$.*

Proof. To get the claim it is sufficient to prove that for any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{GEOD}_\mu$ it holds

$$d_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2). \tag{4.19}$$

Fix such $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ and think $\boldsymbol{\gamma}_1 \in \mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_1)$ and $\boldsymbol{\gamma}_2 \in \mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_2)$. It is clear that the value $\|x_1\|_{\boldsymbol{\gamma}}$ is equal to $\|\boldsymbol{\gamma}\|_\mu$ defined in (4.16a), moreover we know that $W^2_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \|x_1 - x_2\|^2_{\boldsymbol{\alpha}} = \|x_1\|^2_{\boldsymbol{\gamma}_1} + \|x_2\|^2_{\boldsymbol{\gamma}_2} - 2\langle x_1, x_2\rangle_{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, and therefore our thesis may be written in the following way: for any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{GEOD}\mu$ and any $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ it holds

$$\langle x_1, x_2\rangle_{\boldsymbol{\alpha}} = \langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\rangle_\mu = \|\boldsymbol{\gamma}_1\|_\mu \|\boldsymbol{\gamma}_2\|_\mu \cos\big(\theta_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\big). \tag{4.20}$$

Let us examine the right hand side of this equation. Using the definitions (4.16) and equation (4.17) and taking separate limits, first w.r.t. s and then w.r.t. t, we obtain the following formula:

$$\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\rangle_\mu = -\lim_{t\to 0^+} \frac{1}{2t} \frac{d}{ds}|_{s=0}\big(W^2((\pi^0 + t\pi^1)_\#\boldsymbol{\gamma}_1, (\pi^0 + s\pi^2)_\#\boldsymbol{\gamma}_2)\big).$$

Using proposition 4.10 to compute the derivative in the right hand side and recalling (see theorem 2.8) that for small t the plan $(\pi^0, \pi^0 + t\pi^1)_\#\boldsymbol{\gamma}_1$ is the unique element of $\mathcal{Opt}(\mu, (\pi^0 + t\pi^1)_\#\boldsymbol{\gamma}_1)$, we get

$$\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\rangle_\mu = \lim_{t\to 0^+} \frac{1}{t}\langle x_1 - x_0, x_2 - x_0\rangle_{\boldsymbol{\alpha}_t}, \tag{4.21}$$

where $\boldsymbol{\alpha}_t$ minimize $\|x_1 - x_2\|^2_{\boldsymbol{\alpha}_t}$ among all 3-plans in $\mathcal{ADM}_\mu((\pi^0, \pi^0 + t\pi^1)_\#\boldsymbol{\gamma}_1, (\pi^0, \pi^0 + \pi^2)_\#\boldsymbol{\gamma}_2)$. Some manipulations show that $\boldsymbol{\alpha}_t$ has this property if and only if the 3-plan

$$\overline{\boldsymbol{\alpha}}_t := \left(\pi^0, \pi^1 - \pi^0, \frac{\pi^2 - \pi^0}{t}\right)_\# \boldsymbol{\alpha}_t,$$

realizes the minimum of $\|x_1 - x_2\|_{\overline{\boldsymbol{\alpha}}_t}$ among the set $\mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. This means that we can keep $\overline{\boldsymbol{\alpha}}_t$ constant in time and equal to some $\overline{\boldsymbol{\alpha}} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and obtain

$$\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\rangle_\mu = \langle x_1, x_2\rangle_{\overline{\boldsymbol{\alpha}}},$$

and therefore our thesis is proved. □

This characterization of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ leads us to the following definitions.

DEFINITION 4.13. *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu \subset \mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_1)$ be a plan. The* norm *of $\boldsymbol{\gamma}$ is*

$$\|\boldsymbol{\gamma}\|_\mu := \|x_1\|_{\boldsymbol{\gamma}}.$$

Let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be two plans. The scalar product *of them is given by*

$$2\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle := W_\mu^2(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) - \|\boldsymbol{\gamma}_1\|_\mu^2 - \|\boldsymbol{\gamma}_2\|_\mu^2 = 2\langle x_1, x_2 \rangle_{\boldsymbol{\alpha}}, \quad \forall \boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2).$$

These definitions reduce to the ones given in (4.16a), (4.16b) if the plans belong to $\mathcal{GEOD}_\mu$, here we extended them to the whole of $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$.

4. On the inclusion $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) \hookrightarrow \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$

In this section we analyse the relationship between the two tangent spaces we defined and study the properties of the exponential map.

A particular subset of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is the set of all plans induced by maps: since for such plans W_μ essentially reduces to the more common L^2 distance w.r.t. μ, this set is isomorphic through the embedding $\mathfrak{I}_\mu$ to the set

$$\mathcal{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) := \overline{\left\{ v \in L^2(\mu, \mathbb{R}^d) : Id + tv \text{ is optimal for some } t > 0 \right\}}^{L^2(\mu,\mathbb{R}^d)}.$$

In the first section we already introduced this set for regular measures μ, here we just dropped this assumption.

Observe that the definitions of $\mathcal{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ are quite different: the first one is obtained from the study of geodesics while the latter one was the result of the analysis of the continuity equation. It is therefore a non trivial fact that the following theorem holds, which shows that $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is an enlargement of $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. Observe that there is no regularity assumption of μ.

THEOREM 4.14. *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be any measure. Then*

$$\mathcal{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) = \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)).$$

Proof. The inclusion $\supset$ is quite obvious: for any $\varphi \in C_c^\infty(\mathbb{R}^d)$ there exists some $t > 0$ such that $D(Id + t\nabla\varphi) \geq 0$ in the sense of symmetric operators, this means that $Id + t\nabla\varphi$ is the gradient of a convex function and therefore optimal.

To prove the converse implication we will show that any monotone map $T \in L^2(\mu, \mathbb{R}^d)$ belongs to $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. Let us first suppose that $\mu \ll \mathcal{L}^d$, $\mathcal{L}^d \ll \mu$, that μ has bounded density and that T is bounded, say $|T(x)| \leq C < \infty$ for a.e. $x \in \mathbb{R}^d$. In this case for any family of mollifiers ρ_ε the maps $T_\varepsilon := T * \rho_\varepsilon$ are well defined, satisfy $|T_\varepsilon(x)| \leq C$ for any $x \in \mathbb{R}^d$ and $\varepsilon > 0$, and converge for $\mu-$a.e. x to $T(x)$ as $\varepsilon \to 0$. The fact that μ has bounded density implies that the costant functions are in L^2_μ, therefore by the dominated convergence theorem, the convergence a.e. of the equibounded functions T_ε implies convergence in L^2_μ. From the absolute continuity of μ and Brenier's theorem we have the existence of a convex map φ such that $T = \nabla\varphi$ μ-a.e., therefore it holds $T_\varepsilon = (\nabla\varphi) * \rho_\varepsilon = \nabla(\varphi * \rho_\varepsilon) \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. The $L^2_\mu-$convergence of T_ε to T gives $T \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.

Now drop the hypothesis on the boundedness of T. Define $\nu := T_{\#}\mu$ and find a sequence (ν_n) of measures with bounded support $W-$converging to ν. Let T_n be the optimal transport map from μ to ν_n: by what we just proved we have $T_n \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ for every $n \in \mathbb{N}$. Recall proposition 2.27 to obtain $\|T_n - T\|_\mu \to 0$ and therefore $T \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.

For the general case we can assume that $\mathfrak{I}_\mu(T)$ is the unique optimal plan from μ to $T_{\#}\mu$: otherwise substitute T with $(1-t)Id + tT$, $0<t<1$, use proposition 2.11 to get the uniqueness of $\mathfrak{I}_\mu((1-t)Id+tT)$, and let t go to 1. Under this hypothesis we will use equation (3.10) and show that $\langle T, w\rangle_\mu = 0$ for any $w \in L^2(\mu, \mathbb{R}^d)$ such that $\nabla\cdot(w\mu) = 0$. Let ρ_ε be as before a family of mollifiers and define

$$\mu_\varepsilon := \mu * \rho_\varepsilon,$$
$$w_\varepsilon := \frac{(w\mu)*\rho_\varepsilon}{\mu_\varepsilon}.$$

It is clear that $\mu_\varepsilon \xrightarrow{W} \mu$, $\mu_\varepsilon \ll \mathcal{L}^d$, $\mathcal{L}^d \ll \mu_\varepsilon$, $\sup \mu_\varepsilon(x) \le \sup \rho_\varepsilon(x)$, $(w_\varepsilon, \mu_\varepsilon) \twoheadrightarrow (w,\mu)$ (use lemma 3.3 to get $\|w_\varepsilon\|_{\mu_\varepsilon} \le \|w\|_\mu$) and that $\nabla\cdot(w_\varepsilon\mu_\varepsilon) = 0$. Moreover letting T_ε be the unique optimal map from μ_ε to $T_{\#}\mu$ applying proposition 2.27 we get $(T_\varepsilon, \mu_\varepsilon) \twoheadrightarrow (T,\mu)$. Since the previous result ensures that $\langle T_\varepsilon, w_\varepsilon\rangle_{\mu_\varepsilon} = 0$ the conclusion follows by lemma 2.28 (here both sequences converge strongly). □

With this theorem we proved that the natural embedding $\mathfrak{I}$ of L^2_μ into $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ maps the set $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ into the set $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. On the other hand we have the natural map $\mathscr{B} : \mathscr{P}_2(\mathbb{R}^{2d})_\mu \to L^2_\mu$ and we may ask if the image of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is contained or not in $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. The answer is affirmative and given by the following theorem.

Theorem 4.15. *Let* $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. *Then* $\mathscr{B}(\boldsymbol{\gamma}) \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.

Proof. Observe that since $\|\mathscr{B}(\boldsymbol{\gamma}_1) - \mathscr{B}(\boldsymbol{\gamma}_2)\|_\mu \le \|x_1 - x_2\|_{\boldsymbol{\alpha}}$ for any $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, by taking the infimum w.r.t. $\boldsymbol{\alpha}$ we get $\|\mathscr{B}(\boldsymbol{\gamma}_1) - \mathscr{B}(\boldsymbol{\gamma}_2)\|_\mu \le W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. Therefore in order to prove the thesis it is sufficient to consider the case of plans $\boldsymbol{\gamma}$ such that $(\pi^0, \pi^0 + t\pi^1)_{\#}\boldsymbol{\gamma}$ is optimal for some $t > 0$, or, which is the same, of plans $\boldsymbol{\gamma}$ which may be written as $(\pi^0, 1/t(\pi^1 - \pi^0))_{\#}\overline{\boldsymbol{\gamma}}$ for some optimal plan $\overline{\boldsymbol{\gamma}}$ and some t. For such plans it holds $\mathscr{B}(\boldsymbol{\gamma}) = (\mathscr{B}(\overline{\boldsymbol{\gamma}}) - Id)/t$, therefore our thesis will be proved if we show that $\mathscr{B}(\overline{\boldsymbol{\gamma}}) \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ for any optimal plan $\overline{\boldsymbol{\gamma}}$. Since $\overline{\boldsymbol{\gamma}}$ is optimal, we know that there exists a convex function φ such that $\overline{\boldsymbol{\gamma}}$ is concentrated on the graph of the subdifferential of φ. Disintegrating $\overline{\boldsymbol{\gamma}}$ w.r.t. the first variable we obtain that $\mathrm{supp}(\overline{\boldsymbol{\gamma}})_x \subset \partial^-\varphi(x)$. Given that $\partial^-\varphi(x)$ is a closed convex set, we have that $\int y d(\overline{\boldsymbol{\gamma}})_x(y) \in \partial^-\varphi(x)$, that is: the graph of $\mathscr{B}(\overline{\boldsymbol{\gamma}})$ is concentrated on the graph of the subdifferential of φ and therefore $\mathscr{B}(\overline{\boldsymbol{\gamma}})$ is an optimal map. By the previous theorem we have that $\mathscr{B}(\overline{\boldsymbol{\gamma}}) \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and therefore the thesis. □

Definition 4.16 (Multiplication of a plan with a scalar). *Let* $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ *be a plan and* $t \in \mathbb{R}$. *The plan* $t\cdot\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ *is*

$$t\cdot\boldsymbol{\gamma} := (\pi^1, t\pi^2)_{\#}\boldsymbol{\gamma}.$$

PROPOSITION 4.17. *For any $t \in \mathbb{R}$ the map $\boldsymbol{\gamma} \to t \cdot \boldsymbol{\gamma}$ defines an omothety on $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ with constant $|t|$, i.e.*

$$W_\mu(t \cdot \boldsymbol{\gamma}_1, t \cdot \boldsymbol{\gamma}_2) = |t| W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2). \tag{4.22}$$

Proof. It is sufficient to observe that for a given $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ the plan $(\pi^0, t\pi^1, t\pi^2)_\# \boldsymbol{\alpha}$ belongs to $\mathcal{ADM}_\mu(t \cdot \boldsymbol{\gamma}_1, t \cdot \boldsymbol{\gamma}_2)$, and that its cost is $|t|\ \|x_1 - x_2\|_{\boldsymbol{\alpha}}$. □

Now we characterize tangent plans through the infinitesimal behavior of the exponential map. Observe that the definition below is consistent with its analogous in $L^2(\mu, \mathbb{R}^d)$ (see 3.6).

DEFINITION 4.18 (The exponential map and its inverse). *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be a plan. The exponential map $\exp_\mu : \mathscr{P}_2(\mathbb{R}^{2d})_\mu \to \mathscr{P}_2(\mathbb{R}^d)$ is defined as*

$$\exp_\mu(\boldsymbol{\gamma}) := (\pi^1 + \pi^2)_\# \boldsymbol{\gamma}.$$

The inverse of $\exp_\mu$ is the multivalued function from $\mathscr{P}_2(\mathbb{R}^d)$ to $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ defined as

$$\boldsymbol{\gamma} \in \exp_\mu^{-1}(\nu) \Leftrightarrow (\pi^1, \pi^1 + \pi^2)_\# \boldsymbol{\gamma} \in \mathcal{O}pt(\mu, \nu).$$

Note that the functional inverse of $\exp_\mu$ is different from what we defined since it may happen for a plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ to satisfy $(\pi^1 + \pi^2)_\# \boldsymbol{\gamma} = \nu$ and $(\pi^1, \pi^1 + \pi^2)_\# \boldsymbol{\gamma} \notin \mathcal{O}pt(\mu, \nu)$

Note that as in the previous chapter we did not restrict the domain of $\exp_\mu$ to tangent plans.

With this notation we have that for any $\boldsymbol{\gamma} \in \mathcal{GEOD}_\mu$ the map $t \to \exp_\mu(t \cdot \boldsymbol{\gamma})$ is a constant speed geodesic in some right neighborhood of 0 whose metric derivative is $\|\boldsymbol{\gamma}\|_\mu$.

Observe moreover that since $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ implies $(\pi^0 + \pi^1, \pi^0 + \pi^2)_\# \boldsymbol{\alpha} \in \mathcal{A}dm(\exp_\mu(\boldsymbol{\gamma}_1), \exp_\mu(\boldsymbol{\gamma}_2))$ the exponential map is nonexpansive, i.e.

$$W(\exp_\mu(\boldsymbol{\gamma}_1), \exp_\mu(\boldsymbol{\gamma}_2)) \leq W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2). \tag{4.23}$$

This is consistent with the positive curvature of $(\mathscr{P}_2(\mathbb{R}^d), W)$.

THEOREM 4.19 (Infinitesimal behavior of $\exp_\mu$ on tangent plans). *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be a plan. Then $\boldsymbol{\gamma}$ is tangent if and only if*

$$\lim_{t \to 0^+} \frac{W(\exp_\mu(t \cdot \boldsymbol{\gamma}), \mu)}{t} = \|\boldsymbol{\gamma}\|_\mu. \tag{4.24}$$

More explicitely, if $\boldsymbol{\gamma}$ satisfies the previous equation, then for every choice of $\boldsymbol{\eta}_t \in \mathcal{O}pt(\mu, \exp_\mu(t \cdot \boldsymbol{\gamma}))$ the plans $\boldsymbol{\gamma}_t := (\pi^0, (\pi^1 - \pi^0)/t)_\# \boldsymbol{\eta}_t \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ converge to $\boldsymbol{\gamma}$ in $(\mathscr{P}_2(\mathbb{R}^{2d})_\mu, W_\mu)$.

Proof. We have already noticed that the $\boldsymbol{\gamma} \in \mathcal{GEOD}_\mu$ satisfy equation (4.24). The conclusion for general $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ follows by a density argument using equation (4.22) and inequality (4.23).

Let us turn to the converse implication: we will show that the plans $\boldsymbol{\gamma}_t$ defined in the statement $W_\mu-$converge to $\boldsymbol{\gamma}$. From the equation

$$W_\mu^2(\boldsymbol{\gamma}_t,\boldsymbol{\gamma}) = \|\boldsymbol{\gamma}_t\|^2 + \|\boldsymbol{\gamma}\|^2 - 2\langle\boldsymbol{\gamma}_t,\boldsymbol{\gamma}\rangle,$$

and equation (4.24) it follows that the thesis will be proved as soon as we prove that

$$\varliminf_{t\to 0^+} \langle\boldsymbol{\gamma}_t,\boldsymbol{\gamma}\rangle \geq \|\boldsymbol{\gamma}\|^2.$$

Let us evaluate the double derivative $\frac{d}{ds}\frac{d}{dt}W^2(\mu_s,\mu_t)$ at $(0,0)$. By proposition 4.10 and its proof we have that

$$\frac{d}{ds}|_{s=0^+}\frac{d}{dt}|_{t=0^+}W^2(\mu_t,\mu_s) = -2\lim_{s\to 0^+}\frac{1}{s}\langle\boldsymbol{\gamma},(\pi^0,\pi^1-\pi^0)_\#\boldsymbol{\eta}_s\rangle = -2\lim_{s\to 0^+}\langle\boldsymbol{\gamma},\boldsymbol{\gamma}_s\rangle. \tag{4.25}$$

On the other hand, by direct computation we have that

$$\begin{aligned}\frac{d}{ds}|_{s=0^+}\frac{d}{dt}|_{t=0^+}W^2(\mu_t,\mu_s) &= \lim_{s\to 0^+}\frac{1}{s}\lim_{t\to 0^+}\frac{W^2(\mu_t,\mu_s)-W^2(\mu_0,\mu_s)}{t}\\ &= \lim_{s\to 0^+}\lim_{t\to 0^+}\frac{W(\mu_t,\mu_s)-W(\mu_0,\mu_s)}{t}\frac{W(\mu_t,\mu_s)+W(\mu_0,\mu_s)}{s}\\ &=2\lim_{s\to 0^+}\lim_{t\to 0^+}\frac{W(\mu_t,\mu_s)-W(\mu_0,\mu_s)}{t}\lim_{s\to 0^+}\frac{W(\mu_0,\mu_s)}{s}.\end{aligned}$$

Applying the trivial inequality $W(\mu_t,\mu_s)\leq(s-t)\|\boldsymbol{\gamma}\|$, valid for all $0\leq t\leq s$, and recalling equation (4.24) it follows that

$$\frac{d}{ds}|_{s=0^+}\frac{d}{dt}|_{t=0^+}W^2(\mu_t,\mu_s)\leq -2\|\boldsymbol{\gamma}\|^2. \tag{4.26}$$

Putting together equations (4.25) and (4.26) we get $\lim_{s\to 0^+} -2\langle\boldsymbol{\gamma},\boldsymbol{\gamma}_s\rangle \leq -2\|\boldsymbol{\gamma}\|^2$ as desired. □

5. Operations on $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$

For general measures the geometric tangent space is different from the regular one and it is not a vector space. However, it is still possible to define the product of a plan with a real number (we already defined it in 4.16), a sum between plans and a scalar product (defined in 4.13) giving to this space a structure similar to the one of an Hilbert space. However the sum is a multivalued operator.

Definition 4.20. *Let $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\in\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be two plans and $\boldsymbol{\alpha}\in\mathcal{ADM}_\mu(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)$. The sum of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ through $\boldsymbol{\alpha}$ is*

$$\boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}} \boldsymbol{\gamma}_2 := (\pi^0,\pi^1+\pi^2)_\#\boldsymbol{\alpha}.$$

The set of all the possible values of $\boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}} \boldsymbol{\gamma}_2$ as $\boldsymbol{\alpha}$ varies in $\mathcal{ADM}_\mu(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)$ is denoted by $\boldsymbol{\gamma}_1\oplus\boldsymbol{\gamma}_2$.

We recall that for a given couple $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\in\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ of plans, their scalar product is defined as

$$2\langle\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\rangle := \|\boldsymbol{\gamma}_1\|^2+\|\boldsymbol{\gamma}_2\|^2-W_\mu^2(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2) = 2\langle x_1,x_2\rangle_{\boldsymbol{\alpha}}, \qquad \boldsymbol{\alpha}\in\mathcal{OPT}_\mu(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2). \tag{4.27}$$

An equivalent definition is given by the following formula:

$$\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle := \sup_{\boldsymbol{\alpha}' \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)} \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}'}.$$

In this definition we took the supremum (maximum) of $\langle x_1, x_2 \rangle_{\boldsymbol{\alpha}'}$, letting the inner product be a single valued operator. A different choice would be to deal with the multivalued operator given by the whole set of $\langle x_1, x_2 \rangle_{\boldsymbol{\alpha}'}$ as $\boldsymbol{\alpha}$ varies in $\mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$: in this case the operator has better *linearity* properties (for instance the thesis of proposition 4.27 would be true even for $\lambda < 0$). Choosing one definition or the other is mainly a matter of taste. We preferred to keep our definition because it is consistent with the scalar product on the tangent cone for PC spaces (see equation (4.16b), (4.16c)), but notice that here we defined the product for any couple of plans in $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ and not only for those in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.

Observe that a simple consequence of equation (4.27) is the continuity of the inner product. A better result is given by the following proposition.

Proposition 4.21. *Let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\eta} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be given plans. Then*

$$\Big| \langle \boldsymbol{\gamma}_1, \boldsymbol{\eta} \rangle - \langle \boldsymbol{\gamma}_2, \boldsymbol{\eta} \rangle \Big| \leq W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \|\boldsymbol{\eta}\|$$

Proof. Let $\boldsymbol{\alpha}_1 \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\eta})$, and $\boldsymbol{\alpha}_2 \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, be two optimal 3-plan, and consider a 4-plan $\boldsymbol{\beta}$ satisfying

$$\pi^{0,1,2}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_1,$$
$$\pi^{0,1,3}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_2.$$

Then we have $\pi^{0,2,3}_\# \boldsymbol{\beta} \in \mathcal{ADM}_\mu(\boldsymbol{\eta}, \boldsymbol{\gamma}_2)$ and therefore

$$\langle \boldsymbol{\gamma}_1, \boldsymbol{\eta} \rangle = \langle x_1, x_2 \rangle_{\boldsymbol{\beta}} = \langle x_1 - x_3, x_2 \rangle_{\boldsymbol{\beta}} + \langle x_3, x_2 \rangle_{\boldsymbol{\beta}} \leq W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)\|\boldsymbol{\eta}\| + \langle \boldsymbol{\gamma}_2, \boldsymbol{\eta} \rangle.$$

By exchanging the roles of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ the thesis is proven. □

In the next propositions we show that sum and product with scalar give a structure similar to that of a vectorial space.

Proposition 4.22. *Let $\lambda_1, \lambda_2 \in \mathbb{R}$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a measure and $\boldsymbol{\gamma}_i \in \mathscr{P}_2(\mathbb{R}^{2d})$, $i = 1, 2, 3$, be plans such that $\pi^1_\# \boldsymbol{\gamma}_i = \mu$, $i = 1, 2, 3$. Then we have:*

$$\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_2 \oplus \boldsymbol{\gamma}_1, \tag{4.28a}$$

$$\boldsymbol{\gamma}_1 \oplus (\boldsymbol{\gamma}_2 \oplus \boldsymbol{\gamma}_3) = (\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2) \oplus \boldsymbol{\gamma}_3, \tag{4.28b}$$

$$\mathbf{0}_\mu \oplus \boldsymbol{\gamma}_1 = \{\boldsymbol{\gamma}_1\}, \tag{4.28c}$$

$$0 \cdot \boldsymbol{\gamma}_1 = \mathbf{0}_\mu, \tag{4.28d}$$

$$1 \cdot \boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_1, \tag{4.28e}$$

$$\lambda_1 \cdot (\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2) = (\lambda_1 \cdot \boldsymbol{\gamma}_1) \oplus (\lambda_1 \cdot \boldsymbol{\gamma}_2), \tag{4.28f}$$

$$\lambda_1 \cdot (\lambda_2 \cdot \boldsymbol{\gamma}_1) = (\lambda_1 \lambda_2) \cdot \boldsymbol{\gamma}_1, \tag{4.28g}$$

$$(\lambda_1 + \lambda_2) \cdot \boldsymbol{\gamma}_1 \subset (\lambda_1 \cdot \boldsymbol{\gamma}_1) \oplus (\lambda_2 \cdot \boldsymbol{\gamma}_1), \tag{4.28h}$$

where $\mathbf{0}_\mu := (Id, 0)_\# \mu \in \mathscr{P}_2(\mathbb{R}^{2d})$.

Proof. (4.28a), (4.28d), (4.28e), (4.28g) are obvious; since $\mathscr{ADM}_\mu(\boldsymbol{\gamma}_1, \mathbf{0}_\mu)$ contains only the plan $(\pi^1, \pi^2, 0)_\# \boldsymbol{\gamma}_1$ we obtain (4.28c). If $\lambda_1 = 0$ then (4.28f) simply says $\mathbf{0} = \mathbf{0} \oplus \mathbf{0}_\mu$, for the other case note that $\boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ if and only if $(\pi^1, \lambda_1\pi^2, \lambda_1\pi^3)_\# \boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\lambda_1 \cdot \boldsymbol{\gamma}_1, \lambda_1 \cdot \boldsymbol{\gamma}_2)$. (4.28h) follows observing that the plan $(\pi^1, \lambda_1\pi^2, \lambda_2\pi^2)_\# \boldsymbol{\gamma}_1$ belongs to $\mathscr{ADM}_\mu(\lambda_1 \cdot \boldsymbol{\gamma}_1, \lambda_2 \cdot \boldsymbol{\gamma}_1)$. It remains to prove (4.28b), in order to make the proof more clear we think $\boldsymbol{\gamma}_i \in \mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_i)$. Choose $\boldsymbol{\alpha}_1 \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$, $\boldsymbol{\alpha}_2 \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 +_{\boldsymbol{\alpha}_1} \boldsymbol{\gamma}_3)$ and define $\boldsymbol{\alpha}'_1 := (\pi^0, \pi^1 + \pi^2, \pi^2)_\# \boldsymbol{\alpha}_1$. Dudley's lemma ensures the existence of a 4-plan $\boldsymbol{\alpha}$ such that $\pi^{0,1,2}_\# \boldsymbol{\alpha} = \boldsymbol{\alpha}_2$ and $\pi^{0,2,3}_\# \boldsymbol{\alpha} = \boldsymbol{\alpha}'_1$; it follows that the plan $\overline{\boldsymbol{\alpha}} := (\pi^0, \pi^1, \pi^2 - \pi^3, \pi^3)_\# \boldsymbol{\alpha}$ satisfies $\pi^{0,i}_\# \overline{\boldsymbol{\alpha}} = \boldsymbol{\gamma}_i$, $i = 1, 2, 3$, and defining $\boldsymbol{\beta}_1 := (\pi^0, \pi^1, \pi^2)_\# \overline{\boldsymbol{\alpha}}$, $\boldsymbol{\beta}_2 := (\pi^0, \pi^1 + \pi^2, \pi^3)_\# \overline{\boldsymbol{\alpha}}$ we obtain

$$\boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}_2} (\boldsymbol{\gamma}_2 +_{\boldsymbol{\alpha}_1} \boldsymbol{\gamma}_3) = (\boldsymbol{\gamma}_1 +_{\boldsymbol{\beta}_1} \boldsymbol{\gamma}_2) +_{\boldsymbol{\beta}_2} \boldsymbol{\gamma}_3.$$

This proves one inclusion, the other is analogous. □

Remark 4.23. *In general the inclusion in* (4.28h) *is strict: indeed if the plan* $\boldsymbol{\gamma}$ *is not induced by a map the set* $\mathscr{ADM}_\mu(\lambda_1 \cdot \boldsymbol{\gamma}, \lambda_2 \cdot \boldsymbol{\gamma})$ *does not contain only the plan* $(\pi^1, \lambda_1\pi^2, \lambda_2\pi^2)_\# \boldsymbol{\gamma}$. *The following is an explicit example: let* $\mu := \delta_0 \in \mathscr{P}_2(\mathbb{R})$ *be the base measure and* $\boldsymbol{\gamma} := 1/2(\delta_{(0,0)} + \delta_{(0,1)}) \in \mathscr{P}_2(\mathbb{R}^2)_\mu$. *Then* $2 \cdot \boldsymbol{\gamma} = 1/2(\delta_{(0,0)} + \delta_{(0,2)})$ *and*

$$\boldsymbol{\gamma} \oplus \boldsymbol{\gamma} = \{1/2(\lambda\delta(0,0) + (2 - 2\lambda)\delta_{(0,1)} + \lambda\delta_{(0,2)}) : \lambda \in [0,1]\}.$$

Proposition 4.24 (continuity of the operations). *Let* $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ *be a measure and let* $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ *be a plan. Then the map* $\lambda \to \lambda \cdot \boldsymbol{\gamma}$ *is Lipschitz continuous from* $\mathbb{R}$ *to* $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ *and the map* $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \to \boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2$ *is continuous from* $(\mathscr{P}_2(\mathbb{R}^{2d})_\mu)^2$ *to the set of compact subsets of* $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ *endowed with the Hausdorff distance.*

Proof. The first assertion follows by noticing that $(\pi^0, \lambda_1\pi^1, \lambda_2\pi^2)_\# \boldsymbol{\gamma}$ is an admissible 3-plan for the couple $(\lambda_1 \cdot \boldsymbol{\gamma}, \lambda_2 \cdot \boldsymbol{\gamma})$ and its cost is $|\lambda_1 - \lambda_2| \|\boldsymbol{\gamma}\|$. In order to prove the second one let us first observe that the compactness of $\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2$ follows from the compactness of $\mathscr{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and the continuity of the map $\boldsymbol{\alpha} \to (\pi^0, \pi^1 + \pi^2)_\# \boldsymbol{\alpha}$. Now fix $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_2 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$, we will show that for every $\boldsymbol{\gamma} \in \boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2$ there exists $\boldsymbol{\gamma}' \in \boldsymbol{\gamma}'_1 \oplus \boldsymbol{\gamma}'_2$ such that $W_\mu(\boldsymbol{\gamma}, \boldsymbol{\gamma}') \leq W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}'_1) + W_\mu(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}'_2)$: choose $\boldsymbol{\alpha}_i \in \mathscr{OPT}_\mu(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}'_i)$, $i = 1, 2$, $\boldsymbol{\gamma} \in \boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2$ and $\boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ such that $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}} \boldsymbol{\gamma}_2$ and find, thanks to Dudley's lemma, a 5-plan $\boldsymbol{\beta}$ such that

$$\pi^{0,1,2}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_1$$
$$\pi^{0,3,4}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_2$$
$$\pi^{0,1,3}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}.$$

Defining $\boldsymbol{\alpha}' := \pi^{0,2,4}_\# \boldsymbol{\beta}$, $\boldsymbol{\gamma}' := \boldsymbol{\gamma}'_1 +_{\boldsymbol{\alpha}'} \boldsymbol{\gamma}'_2$ and $\overline{\boldsymbol{\beta}} := (\pi^0, \pi^1 + \pi^3, \pi^2 + \pi^4)_\# \boldsymbol{\beta}$ we obtain:

$$W_\mu(\boldsymbol{\gamma}, \boldsymbol{\gamma}') \leq \|x_1 - x_2\|_{\overline{\boldsymbol{\beta}}} = \|x_1 + x_3 - x_2 - x_4\|_{\boldsymbol{\beta}} \leq$$
$$\|x_1 - x_2\|_{\boldsymbol{\beta}} + \|x_3 - x_4\|_{\boldsymbol{\beta}} = W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}'_1) + W_\mu(\boldsymbol{\gamma}_2, \boldsymbol{\gamma}'_2).$$

□

These operations and their properties are very similar to those which define the vector space structure, now we want to prove that for any measure $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ the geometric tangent space $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is a "subspace" of $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$, i.e. we will prove that $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is stable under these operations.

PROPOSITION 4.25. *For every $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ we have $\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2 \subset \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.*

Proof. By proposition 4.24 it is enough to prove the theorem under the additional hypothesis $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{GEOD}_\mu$, so we assume that there exists $t > 0$ such that $(\pi^1, \pi^1 + t\pi^2)_\# \boldsymbol{\gamma}_i$, $i = 1, 2$, is optimal. This means that for $i = 1, 2$, $n \in \mathbb{N}$, σ permutation of $\{1, \ldots, n\}$ and $(x_1, y_1), \ldots, (x_n, y_n) \in \operatorname{supp} \boldsymbol{\gamma}_i$ it holds

(4.29)
$$\begin{aligned}\sum_{j=1}^n |x_j|^2 + t\sum_{j=1}^n \langle x_j, y_j \rangle =& \sum_{j=1}^n \langle x_j, x_j + t y_j \rangle \\ \geq& \sum_{j=1}^n \langle x_j, x_{\sigma(j)} + t y_{\sigma(j)} \rangle = \sum_{j=1}^n \langle x_j, x_{\sigma(j)} \rangle + t \sum_{j=1}^n \langle x_j, y_{\sigma(j)} \rangle.\end{aligned}$$

Choose $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and define $\boldsymbol{\gamma} := \boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}} \boldsymbol{\gamma}_2$: we want to prove that $(\pi^1, \pi^1 + t/2\pi^2)_\# \boldsymbol{\gamma}$ is optimal, so we fix $n \in \mathbb{N}$, $\sigma \in S_n$ and $(x_j, y_j) \in \operatorname{supp} \boldsymbol{\gamma}$, $j = 1, \ldots, n$ and we will show that it holds

$$\sum_{j=1}^n |x_j|^2 + t/2 \sum_{j=1}^n \langle x_j, y_j \rangle \geq \sum_{j=1}^n \langle x_j, x_{\sigma(j)} \rangle + t/2 \sum_{j=1}^n \langle x_j, y_{\sigma(j)} \rangle.$$

We know that for every $\varepsilon > 0$ and every $(x, y) \in \operatorname{supp} \boldsymbol{\gamma}$ there exists $(x', y^1, y^2) \in \operatorname{supp} \boldsymbol{\alpha}$ such that $|(x, y) - (x', y^1 + y^2)| < \varepsilon$, in particular $(x', y^i) \in \operatorname{supp} \boldsymbol{\gamma}_i$, $i = 1, 2$. Choosing in this way (x'_j, y^1_j, y^2_j), $j = 1, \ldots, n$, from (x_j, y_j), $j = 1, \ldots, n$, and adding up the two inequalities (4.29) relatives to $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ obtained with these (x'_j, y^1_j) and (x'_j, y^2_j), $j = 1, \ldots, n$ we get

$$\begin{aligned}2\sum_{j=1}^n |x_j|^2 + t\sum_{j=1}^n \langle x_j, y_j \rangle \geq& 2\sum_{j=1}^n |x'_j|^2 + t\sum_{j=1}^n \langle x'_j, y^1_j + y^2_j \rangle - 8Mn\varepsilon \\ \geq& 2\sum_{j=1}^n \langle x'_j, x'_{\sigma(j)} \rangle + t\sum_{j=1}^n \langle x_j, y^1_{\sigma(j)} + y^2_{\sigma(j)} \rangle - 8Mn\varepsilon \\ \geq& 2\sum_{j=1}^n \langle x_j, x_{\sigma(j)} \rangle + t\sum_{j=1}^n \langle x_j, y_{\sigma(j)} \rangle - 16Mn\varepsilon,\end{aligned}$$

where $M := \max_j |(x_j, y_j)|$. Letting ε go to 0 we get the thesis. □

REMARK 4.26. *A similar computation shows that if $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ are optimal plans with the same first marginal, then each element of $\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2$ is an optimal plan, too. Since in addition if $\boldsymbol{\gamma}$ is optimal then even $\lambda \cdot \boldsymbol{\gamma}$ is optimal we get that any linear combination (w.r.t. the operations just defined) with positive coefficients of optimal plans is optimal.*

PROPOSITION 4.27. *Let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ and $\lambda \geq 0$. Then it holds*

$$\langle \lambda \cdot \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle = \langle \boldsymbol{\gamma}_1, \lambda \cdot \boldsymbol{\gamma}_2 \rangle = \lambda \langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle, \tag{4.30}$$

$$\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_3 \rangle + \langle \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 \rangle = \max \langle \boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 \rangle. \tag{4.31}$$

Proof. Equation (4.30) is clear if $\lambda = 0$, otherwise it is sufficient to observe that $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ iff $(\pi^0, \lambda\pi^1, \pi^2)_\# \boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\lambda \cdot \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and iff $(\pi^0, \pi^1, \lambda\pi^2)_\# \boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \lambda \cdot \boldsymbol{\gamma}_2)$.

To prove (4.31) let us first prove that the sum on the left is an element of the set on the right: observe that for every choice of $\boldsymbol{\alpha}_{i,3} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_3)$, $i = 1, 2$, there exists a 4-plan $\boldsymbol{\beta}$ satisfying

$$\pi^{0,1,3}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_{1,3},$$
$$\pi^{0,2,3}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_{2,3},$$

Define $\boldsymbol{\alpha}_{1,2} := \pi^{0,1,2}_\# \boldsymbol{\beta}$ and check that a consequence of the previous remark is

$$(\pi^0, \pi^1 + \pi^2, \pi^3)_\# \boldsymbol{\beta} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}_{1,2}} \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3),$$

from which it follows the claim. To prove that it holds (4.31), choose any element $\overline{\boldsymbol{\gamma}} \in \boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2$, say $\overline{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}} \boldsymbol{\gamma}_2$, and find a 4-plan $\boldsymbol{\beta}$ satisfying

$$\pi^{0,1,2}_\# \boldsymbol{\beta} = \boldsymbol{\alpha},$$
$$(\pi^0, \pi^1 + \pi^2, \pi^3)_\# \boldsymbol{\beta} \in \mathcal{OPT}_\mu(\overline{\boldsymbol{\gamma}}, \boldsymbol{\gamma}_3),$$

then conclude with

$$\langle \overline{\boldsymbol{\gamma}}, \boldsymbol{\eta} \rangle = \langle x_1 + x_2, x_3 \rangle_{\boldsymbol{\beta}} = \langle x_1, x_3 \rangle_{\boldsymbol{\beta}} + \langle x_2, x_3 \rangle_{\boldsymbol{\beta}} \leq \langle \boldsymbol{\gamma}_1, \boldsymbol{\eta} \rangle + \langle \boldsymbol{\gamma}_2, \boldsymbol{\eta} \rangle.$$

□

REMARK 4.28. *In general it is not true that $\langle \lambda \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle = \lambda \langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle$, for a negative λ. Think for instance to the case $\lambda = -1$: we have that*

$$\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}) \quad \Leftrightarrow \quad \overline{\boldsymbol{\alpha}} := (\pi^0, -\pi^1, \pi^2)_\# \boldsymbol{\alpha} \in \mathcal{ADM}_\mu(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta}),$$

and therefore it it holds

$$-1\langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle = -\left(\max_{\overline{\boldsymbol{\alpha}} \in \mathcal{ADM}_\mu(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta})} \langle x_1, x_2 \rangle_{\overline{\boldsymbol{\alpha}}} \right)$$
$$= \min_{\overline{\boldsymbol{\alpha}} \in \mathcal{ADM}_\mu(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta})} \langle -x_1, x_2 \rangle_{\overline{\boldsymbol{\alpha}}} = \min_{\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})} \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}}.$$

Since in general $\min_{\boldsymbol{\alpha}} \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}} < \max_{\boldsymbol{\alpha}} \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}}$, when $\boldsymbol{\alpha}$ varies in $\mathcal{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ we get the claim.

It is obvious from the definition that if $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ then $\lambda \cdot \boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ for any $\lambda > 0$, so the question is whether the same property holds for $\lambda < 0$. The answer is given by the following proposition.

PROPOSITION 4.29. *Let $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ be a plan. Then $-1 \cdot \boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.*

Proof. By density and positive homogeneity we can assume that $\overline{\boldsymbol{\gamma}} := (\pi^1, \pi^1 + \pi^2)_\# \boldsymbol{\gamma}$ is optimal. Since

$$-1 \cdot \boldsymbol{\gamma} = -1 \cdot \overline{\boldsymbol{\gamma}} \oplus (Id, Id)_\# \mu,$$

and since $(Id, Id)_\# \mu$ clearly belongs to $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, by proposition 4.25 our thesis become: the opposite of an optimal plan belongs to the geometric tangent space.

So let $\boldsymbol{\gamma}$ be an optimal plan, we know that there exists a convex function ψ such that $\operatorname{supp} \boldsymbol{\gamma}$ is contained in the graph of the subdifferential of ψ. Clearly $-1 \cdot \boldsymbol{\gamma}$ has the support contained in the graph of the superdifferential of the concave function $-\psi$.

Since we want to prove that $-1 \cdot \boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ we need to find a sequence of plans in $\mathcal{GEOD}_\mu$ which tends to $-1 \cdot \boldsymbol{\gamma}$.

Observe that a plan $\boldsymbol{\eta} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ belongs to $\mathcal{GEOD}_\mu$ if and only if for some $t > 0$ the support of $(\pi^1 + t\pi^2)_\# \boldsymbol{\eta}$ is cyclically monotone. By Rockafellar theorem this is true if and only if there exists a convex function φ such that $(\pi^1 + t\pi^2)_\# \boldsymbol{\eta}$ is concentrated on the graph of $\partial^- \varphi$, and therefore if and only if $\boldsymbol{\eta}$ is concentrated on the graph of $\partial^-((\varphi - Id)/t)$. Recalling that a function ϕ is said semi-convex if it exists some $t > 0$ such that $Id + t\phi$ is convex, we just proved that $\boldsymbol{\eta} \in \mathcal{GEOD}_\mu$ if and only if it is concentrated on the graph of the subdifferential of a semi-convex function.

Therefore part of the problem is to obtain the superdifferential of the concave function $-\psi$ as limit (in some sense) of subdifferential of semi-convex functions. The key idea here is the following statement: *for any convex set $C \subset \mathbb{R}^d$, any interior point (in the sense of convex sets) $p \in C$ it holds $-C \subset \cup_{n=0}^\infty (nC - (n+1)p)$.* This is true because letting $\tilde{C} := C - p$ the assertion becomes $-\tilde{C} \subset \cup_{n=0}^\infty n\tilde{C}$ which is obviously because the fact that 0 is an interior point of $\tilde{C}$ implies that $\cup_{n=0}^\infty n\tilde{C}$ is the whole linear span of $\tilde{C}$.

From this inclusion we can hope for an inclusion like $\partial^+(-\psi)(x_0) \subset \partial^-(n\psi - (n+1)\varphi)(x_0)$, where φ is a regular function such that $\nabla\varphi(x_0)$ is an interior point of $\partial^-\psi(x_0)$ for μ−a.e. x_0. However this cannot be done because such a function φ may not exist, so we have to proceed in two steps. In the first one we approximate $-1 \cdot \boldsymbol{\gamma}$ with measures $\boldsymbol{\eta}_n$ whose disintegration $(\boldsymbol{\eta}_n)_{x_0}$ are such that

$$\operatorname{supp}(\boldsymbol{\eta}_n)_{x_0} \subset n\partial^-\psi - (n+1)\mathscr{B}(\boldsymbol{\gamma})(x_0),$$

for μ−a.e. x_0. In the second step we approximate $\mathscr{B}(\boldsymbol{\gamma})$ with the gradient of regular functions φ_m and show that for any n there exists a sequence $(\boldsymbol{\eta}_{n,m})$ of measures which converges to $\boldsymbol{\eta}_n$ such that $\operatorname{supp} \boldsymbol{\eta}_{n,m} \subset \partial^-(n\psi - (n+1)\varphi_m)$.

Now we turn to the details.

step 1 Let $x_0 \to \boldsymbol{\gamma}_{x_0}$ be the disintegration of $\boldsymbol{\gamma}$ with respect to μ and, for every x_0, let $A_{x_0} \subset \mathbb{R}^d$ be the closed convex hull of $\operatorname{supp} \boldsymbol{\gamma}_{x_0}$. It is clear that $A_{x_0} \subset \partial^-\psi(x_0)$ for μ−a.e. x_0 and that $\mathscr{B}(\boldsymbol{\gamma})(x_0)$ is either the only point of A_{x_0} (if $\boldsymbol{\gamma}_{x_0}$ is not a Dirac mass) or an inteorior point of A_{x_0} in the sense of convex sets. Our previous statement about convex sets (with $C = A_{x_0}$ and $p = \mathscr{B}(\boldsymbol{\gamma})(x_0)$) gives that for almost every x_0 it holds

$$-A_{x_0} \subset \bigcup_{n=0}^{\infty} \Big(nA_{x_0} - (n+1)\mathscr{B}(\boldsymbol{\gamma})(x_0) \Big),$$

the inclusion being obvious if $A_{x_0} = \{\mathscr{B}(\boldsymbol{\gamma})(x_0)\}$. Let $r_n(x_0)$ be the function defined as

$$r_n(x_0) := \sup\{r \;:\; B_r(-\mathscr{B}(\boldsymbol{\gamma})(x_0)) \subset nA_{x_0} - (n+1)\mathscr{B}(\boldsymbol{\gamma})(x_0)\},$$

if $\boldsymbol{\gamma}_{x_0}$ is not a Dirac mass, and $r_n(x_0) := +\infty$ elsewhere. It is clear that

$$r_n(x_0) \xrightarrow{n} +\infty \qquad \text{for } \mu - \text{a.e. } \mathrm{x}_0. \tag{4.32}$$

Let $P^{r,x_0} : \mathbb{R}^d \to \mathbb{R}^d$ be the projection onto the ball $\overline{B}_r(-\mathscr{B}(\boldsymbol{\gamma})(x_0))$ and define

$$\boldsymbol{\eta}_{n,x_0} := (P^{r_n(x_0),x_0})_\# \boldsymbol{\gamma}_{x_0}.$$

and $\boldsymbol{\eta}_n := \mu \otimes \boldsymbol{\eta}_{n,x_0}$. Obviously it holds

$$\operatorname{supp} \boldsymbol{\eta}_{n,x_0} \subset nA_{x_0} - (n+1)\mathscr{B}(\boldsymbol{\gamma})(x_0),$$

for any n and μ−a.e. x_0. Equation (4.32) implies that for μ−a.e. x_0 it holds $W(\boldsymbol{\eta}_{n,x_0}, \boldsymbol{\gamma}_{x_0}) \to 0$ as $n \to \infty$, moreover we have

$$\begin{aligned} W^2(\boldsymbol{\eta}_{n,x_0}, \boldsymbol{\gamma}_{x_0}) &\leq \int |P^{r_n(x_0),x_0}(x_1) - x_1|^2 d\boldsymbol{\gamma}_{x_0}(x_1) \\ &\leq \int |x_1 - \mathscr{B}(\boldsymbol{\gamma})(x_0)|^2 d\boldsymbol{\gamma}_{x_0}(x_1) \leq \int |x_1|^2 d\boldsymbol{\gamma}_{x_0}(x_1). \end{aligned}$$

Since $\int |x_1|^2 d\boldsymbol{\gamma}_{x_0}(x_1) d\mu(x_0) = \int |x_1|^2 d\boldsymbol{\gamma}(x_0, x_1) < +\infty$ the dominated convergence theorem gives

$$W_\mu^2(\boldsymbol{\eta}_n, \boldsymbol{\gamma}) = \int W^2(\boldsymbol{\eta}_{n,x_0}, \boldsymbol{\gamma}_{x_0}) d\mu(x_0) \to 0.$$

step 2 We know (see theorem 4.15) that $\mathscr{B}(\boldsymbol{\gamma}) \in \operatorname{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, so there exists a sequence (φ_m) of functions in C_c^∞ such that $\nabla\varphi_m \to \mathscr{B}(\boldsymbol{\gamma})$ in L^2_μ. Consider for any n, m, x_0 the translation defined as

$$\tau^{n,m,x_0}(x_1) := x_1 + (n+1)(\mathscr{B}(\boldsymbol{\gamma})(x_0) - \nabla\varphi_m(x_0)),$$

and define $\boldsymbol{\eta}_{n,m,x_0} := (\tau^{n,m,x_0})_\# \boldsymbol{\eta}_{n,x_0}$. Since

$$\tau^{n,m,x_0}(nA_{x_0} - (n+1)\mathscr{B}(\boldsymbol{\gamma})(x_0)) = nA_{x_0} - (n+1)\nabla\varphi_m(x_0),$$

we have

$$\operatorname{supp} \boldsymbol{\eta}_{n,m,x_0} \subset nA_{x_0} - (n+1)\nabla\varphi_m(x_0) \subset \partial^-(n\psi - (n+1)\varphi_m)(x_0),$$

and so the measures $\boldsymbol{\eta}_{n,m} := \mu \otimes \boldsymbol{\eta}_{n,m,x_0}$ belong to $\mathcal{GEOD}_\mu$. Fix n and observe that

$$\begin{aligned} W_\mu^2\left(\boldsymbol{\eta}_{n,m}, \boldsymbol{\eta}_n\right) &= \int W^2(\boldsymbol{\eta}_{n,m,x_0}, \boldsymbol{\eta}_{n,x_0}) d\mu(x_0) \\ &\leq (n+1)\|\mathscr{B}(\boldsymbol{\gamma}) - \nabla\varphi_m\|_\mu^2 \to 0, \qquad \text{as } \; m \to \infty. \end{aligned}$$

From this fact and the convergence of $\boldsymbol{\eta}_n$ to $\boldsymbol{\gamma}$ in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))(\mathscr{P}_2(\mathbb{R}^d))$ it follows the existence of a sequence (m_n) such that $W_\mu\left(\boldsymbol{\eta}_{n,m_n}, \boldsymbol{\gamma}\right) \to 0$ as $n \to \infty$, and the proof is complete.

□

6. The projection onto $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$

In this section we show that there is a well defined projection operator from $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ to $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, and we describe its basic properties.

The first proposition below asserts that there is a well defined projection operator from $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ to $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. Since its proof is very similar to the one which says that there is a unique projection onto a closed and convex subset of an Hilbert space, in order to better follow the argument we will report first the proof of this well known fact.

HILBERT CASE 1. *Let K be a closed convex subset of an Hilbert space H. Then for each vector $v \in H$ there exists a unique $P(v) \in K$ which minimizes the distance from v among all $w \in K$.*

Proof. Let (w_n) be a minimizing sequence and let $d := \inf_{w\in K} |v - w|$. Since $(w_n + w_m)/2 \in K$ by the parallelogram rule we have

$$
\begin{aligned}
\left|\frac{w_n - w_m}{2}\right|^2 &= \frac{1}{2}|v - w_n|^2 + \frac{1}{2}|v - w_m|^2 - \left|v - \frac{w_n + w_m}{2}\right|^2 \\
&\le \frac{1}{2}|v - w_n|^2 + \frac{1}{2}|v - w_m|^2 - d^2,
\end{aligned}
\tag{4.33}
$$

from which it follows that (w_n) is a Cauchy sequence. From the minimizing property of (w_n), we get that its limit produces the (only) minimizer vector $P(v)$. □

PROPOSITION 4.30 (projection onto the tangent space). *Let $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be a plan. Then there exists a unique plan $\mathcal{P}(\boldsymbol{\gamma}) \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ (called projection of $\boldsymbol{\gamma}$ onto the tangent space) which minimizes the distance $W_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ among all $\boldsymbol{\eta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.*

Proof. Let $(\boldsymbol{\eta}_n)$ be a minimizing sequence and let $d := \inf W_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ among all $\boldsymbol{\eta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. For each $n, m \in \mathbb{N}$ let $\boldsymbol{\alpha}_{n,m} \in \mathscr{P}_2(\mathbb{R}^{4d})$ be a 4-plan such that

$$
\begin{aligned}
\pi^{0,1,2}_{\#}\boldsymbol{\alpha}_{n,m} &\in \mathit{OPT}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}_n), \\
\pi^{0,1,3}_{\#}\boldsymbol{\alpha}_{n,m} &\in \mathit{OPT}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}_m),
\end{aligned}
$$

and define $\boldsymbol{\beta}_{n,m} := \pi^{0,2,3}_{\#}\boldsymbol{\alpha}_{n,m}$ and $\boldsymbol{\eta}_{n,m} := 1/2 \cdot (\boldsymbol{\eta}_n +_{\boldsymbol{\beta}_{n,m}} \boldsymbol{\eta}_m)$. We know that $\boldsymbol{\eta}_{n,m} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ so $W_\mu\left(\boldsymbol{\gamma}, \boldsymbol{\eta}_{n,m}\right) \ge d$ and therefore

$$
\begin{aligned}
\frac{1}{4}W_\mu^2(\boldsymbol{\eta}_n, \boldsymbol{\eta}_m) &\le \frac{1}{4}\|x_2 - x_3\|^2_{\boldsymbol{\alpha}_{n,m}} \\
&\le \frac{1}{2}\|x_1 - x_2\|^2_{\boldsymbol{\alpha}_{n,m}} + \frac{1}{2}\|x_1 - x_3\|^2_{\boldsymbol{\alpha}_{n,m}} - \left\|x_1 - \frac{x_2 + x_3}{2}\right\|^2_{\boldsymbol{\alpha}_{n,m}} \\
&\le \frac{1}{2}W_\mu^2(\boldsymbol{\gamma}, \boldsymbol{\eta}_n) + \frac{1}{2}W_\mu^2(\boldsymbol{\gamma}, \boldsymbol{\eta}_m) - W_\mu^2\left(\boldsymbol{\gamma}, \boldsymbol{\eta}_{n,m}\right) \\
&\le \frac{1}{2}W_\mu^2(\boldsymbol{\gamma}, \boldsymbol{\eta}_n) + \frac{1}{2}W_\mu^2(\boldsymbol{\gamma}, \boldsymbol{\eta}_m) - d^2,
\end{aligned}
$$

and letting $n, m \to \infty$ we obtain that $(\boldsymbol{\eta}_n)$ is a Cauchy sequence. □

Now we are going to prove that there exists a unique optimal 3-plan between $\boldsymbol{\gamma}$ and $\mathcal{P}(\boldsymbol{\gamma})$ and that this plan is induced by map (see equation (4.36) for the definition). Before doing this we need the following lemma. In the sequel we denote by $\overline{\mathrm{Conv}}\, A$ the closed convex hull of $A \subset \mathbb{R}^d$.

LEMMA 4.31. *Let $\boldsymbol{\gamma}, \boldsymbol{\eta} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be two plans and suppose that $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and that for μ-a.e. x_0 it holds*

$$\operatorname{supp}(\boldsymbol{\eta}_{x_0}) \subset \overline{\mathrm{Conv}}\Big(\operatorname{supp}(\boldsymbol{\gamma}_{x_0})\Big). \tag{4.34}$$

Then $\boldsymbol{\eta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.

Proof. Since $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ we know that there exists a sequence $(\boldsymbol{\gamma}^n) \in \mathcal{GEOD}_\mu$ converging to $\boldsymbol{\gamma}$ w.r.t. W_μ. Up to extracting a subsequence, not relabeled, we may assume that for $\mu-$a.e. x_0 it holds $W(\boldsymbol{\gamma}^n_{x_0}\boldsymbol{\gamma}_{x_0}) \to 0$ as $n \to \infty$ and that the sequence of functions $x_0 \to \|\boldsymbol{\gamma}^n_{x_0}\|$ is uniformly dominated by an L^2 function g.

Now observe that since $\boldsymbol{\gamma}^n \in \mathcal{GEOD}_\mu$, there exists a sequence of semi-convex functions (φ^n) such that for any n and μ-a.e. x_0 it holds $\operatorname{supp}(\boldsymbol{\gamma}^n_{x_0}) \subset \partial^-\varphi^n(x_0)$ (see also the proof of proposition 4.29). Given that $\partial^-\varphi^n(x_0)$ is closed and convex for any $n \in \mathbb{N}$, $x_0 \in \mathbb{R}^d$, it holds

$$\overline{\mathrm{Conv}}\Big(\operatorname{supp}\big(\boldsymbol{\gamma}^n_{x_0}\big)\Big) \subset \partial^-\varphi^n(x_0), \quad \text{for } \mu\text{-a.e. } x_0.$$

Consider the projections $P^n_{x_0} : \mathbb{R}^d \to \partial^-\varphi^n(x_0)$ and observe that the map $(x_0, x_1) \to P^n_{x_0}(x_1)$ is measurable for every $n \in \mathbb{N}$. Define, for every $n \in \mathbb{N}$, the plan $\overline{\boldsymbol{\gamma}}^n \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ by defining its disintegration w.r.t. the first marginal by $(\overline{\boldsymbol{\gamma}}^n)_{x_0} := (P^n_{x_0})_\# \boldsymbol{\gamma}_{x_0}$. Estimating $W((\overline{\boldsymbol{\gamma}}^n)_{x_0}, \boldsymbol{\gamma}_{x_0})$ with $\|P^n_{x_0} - Id\|_{\boldsymbol{\gamma}_{x_0}}$ and using the minimizing properties of the projection and the fact that $\operatorname{supp}(\boldsymbol{\gamma}^n_{x_0}) \subset \partial^-\varphi^n(x_0)$, we obtain $W((\overline{\boldsymbol{\gamma}}^n)_{x_0}, \boldsymbol{\gamma}_{x_0}) \leq W(\boldsymbol{\gamma}^n_{x_0}, \boldsymbol{\gamma}_{x_0})$. Therefore $(\overline{\boldsymbol{\gamma}}^n)$ $W_\mu-$converges to $\boldsymbol{\gamma}$. As a consequence, we get

$$\big|(x_0 P^n_{x_0}(x_1)) - (x_0, x_1)\big| \to 0, \quad \text{as } n \to \infty \quad \text{for } \boldsymbol{\gamma}\text{-a.e. } (x_0, x_1). \tag{4.35}$$

Analogously, define the plans $\boldsymbol{\eta}^n \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ by defining their disintegrations w.r.t. their first marginal by $(\boldsymbol{\eta}^n)_{x_0} := (P^n_{x_0})_\# \boldsymbol{\eta}_{x_0}$. Given that clearly $\operatorname{supp}((\boldsymbol{\eta}^n)_{x_0}) \subset \partial^-\varphi^n(x_0)$, it holds $\boldsymbol{\eta}^n \in \mathcal{GEOD}_\mu$. Therefore our thesis will be proved if we show that $W_\mu(\boldsymbol{\eta}^n, \boldsymbol{\eta}) \to 0$ as $n \to \infty$.

Inclusion (4.34), the convexity of $\partial^-\varphi^n(x_0)$, (4.35) and the convexity of $(x, y) \to |x - y|$ give

$$\big|P^n_{x_0}(x_1) - (x_0, x_1)\big| \to 0, \quad \text{as } n \to \infty \quad \text{for } \boldsymbol{\eta}\text{-a.e. } (x_0, x_1).$$

To conclude it is sufficient to show that the sequence of functions $x_0 \to W(\boldsymbol{\eta}_{x_0}, \boldsymbol{\eta}^n_{x_0})$ is dominated by a function in L^2_μ. To this aim define the functions $f^n : \operatorname{supp}(\mu) \to \mathbb{R}^d$ letting $f^n(x_0)$ be the element of minimal norm of $\partial^-\varphi^n(x_0)$. Since $\boldsymbol{\gamma}^n \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ and $\operatorname{supp}(\boldsymbol{\gamma}^n_{x_0}) \subset \partial^-\varphi^n(x_0)$, we have that (f^n) is a sequence in L^2_μ dominated by the function g. Conclude estimating from above $W(\boldsymbol{\eta}_{x_0}, \boldsymbol{\eta}^n_{x_0})$ with the cost of the transport map $P^n_{x_0}$:

$$\Big(\int |x_1 - P^n_{x_0}(x_1)|^2 d\boldsymbol{\eta}_{x_0}(x_1)\Big)^{1/2} \le \Big(\int |x_1 - f_n(x_0)|^2 d\boldsymbol{\eta}_{x_0}(x_1)\Big)^{1/2}$$
$$\le \Big(\int |x_1|^2 d\boldsymbol{\eta}_{x_0}(x_1)\Big)^{1/2} + f^n(x_0)$$
$$\le 2g(x_0).$$

□

PROPOSITION 4.32. *For any $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu \subset \mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_1)$ there is a unique optimal 3-plan $\boldsymbol{\alpha}^{\mathcal{P}(\boldsymbol{\gamma})}_{\boldsymbol{\gamma}}$ in $\mathcal{OPT}_\mu(\boldsymbol{\gamma}, \mathcal{P}(\boldsymbol{\gamma}))$. Moreover this plan is induced by a map in the following sense: there exists a Borel map $P_{\boldsymbol{\gamma}} : \mathbb{R}^d_0 \times \mathbb{R}^d_1 \to \mathbb{R}^d_2$ such that*

$$\boldsymbol{\alpha}^{\mathcal{P}(\boldsymbol{\gamma})}_{\boldsymbol{\gamma}} = (\pi^0, \pi^1, P_{\boldsymbol{\gamma}} \circ (\pi^0, \pi^1))_\# \boldsymbol{\gamma}. \tag{4.36}$$

Observe that condition (4.36) may be given in the following equivalent way: for μ-a.e. x_0 the plan $(\boldsymbol{\alpha}^{\mathcal{P}(\boldsymbol{\gamma})}_{\boldsymbol{\gamma}})_{x_0} \in \mathcal{O}pt(\boldsymbol{\gamma}_{x_0}, (\mathcal{P}(\boldsymbol{\gamma}))_{x_0})$ is induced by the map $x_1 \to P_{\boldsymbol{\gamma}}(x_0, x_1)$.

Proof. To prove the thesis is sufficient to show that any optimal 3-plan $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}, \mathcal{P}(\boldsymbol{\gamma}))$ is induced by a map: indeed if $T_1, T_2 : \mathbb{R}^d_0 \times \mathbb{R}^d_1 \to \mathbb{R}^d_2$ were two different optimal maps, then the plan

$$\frac{1}{2}\Big((\pi^0, \pi^1, T_1 \circ (\pi^0, \pi^1))_\# \boldsymbol{\gamma} + (\pi^0, \pi^1, T_2 \circ (\pi^0, \pi^1))_\# \boldsymbol{\gamma}\Big),$$

would be optimal and not induced by a map.

So fix $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}, \mathcal{P}(\boldsymbol{\gamma}))$, let $d\boldsymbol{\alpha} = d\boldsymbol{\gamma} \otimes d\boldsymbol{\alpha}_{x_0,x_1}$ be its disintegration it w.r.t. $\boldsymbol{\gamma}$ and define

$$T(x_0, x_1) := \int x_2 d\boldsymbol{\alpha}_{x_0,x_1}(x_2),$$

and $\overline{\boldsymbol{\gamma}} := (\pi^0, T \circ (\pi^0, \pi^1))_\# \boldsymbol{\gamma}$. Since clearly $T(x_0, x_1) \in \overline{\mathrm{Conv}}\,(\mathrm{supp}(\boldsymbol{\alpha}_{x_0,x_1}))$ for $\boldsymbol{\gamma}$-a.e. (x_0, x_1), and

$$\big(\mathcal{P}(\boldsymbol{\gamma})\big)_{x_0} = \int \boldsymbol{\alpha}_{x_0,x_1} d(\pi^1_\# \boldsymbol{\alpha})(x_1),$$

we get

$$\mathrm{supp}\Big((\overline{\boldsymbol{\gamma}})_{x_0}\Big) \subset \overline{\mathrm{Conv}}\Bigg(\mathrm{supp}\Big(\big(\mathcal{P}(\boldsymbol{\gamma})\big)_{x_0}\Big)\Bigg).$$

Applying the previous lemma we get $\overline{\boldsymbol{\gamma}} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, moreover the plan $(\pi^0, \pi^1, T \circ (\pi^0, \pi^1))_\# \boldsymbol{\gamma}$ is admissible for the couple $\boldsymbol{\gamma}$, $\overline{\boldsymbol{\gamma}}$, and therefore

$$W^2_\mu(\boldsymbol{\gamma}, \overline{\boldsymbol{\gamma}}) \le \int |x_1 - T(x_0, x_1)|^2 d\boldsymbol{\gamma}(x_0, x_1)$$
$$= \int \Big| \int (x_1 - x_2) d\boldsymbol{\alpha}_{x_0,x_1}(x_2)\Big|^2 d\boldsymbol{\gamma}(x_0, x_1)$$
$$\le \int |x_1 - x_2|^2 d\boldsymbol{\alpha}_{x_0,x_1}(x_2) d\boldsymbol{\gamma}(x_0, x_1) = W^2_\mu(\boldsymbol{\gamma}, \mathcal{P}(\boldsymbol{\gamma})).$$

Since the second inequality is strict if $\boldsymbol{\alpha}_{x_0,x_1}$ is not a Dirac mass for $\boldsymbol{\gamma}$-a.e. (x_0,x_1), by the minimizing properties of $\mathcal{P}(\boldsymbol{\gamma})$ we get the validity of equation (4.36). $\square$

The first remarkable properties of the plan $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}$ are collected below.

THEOREM 4.33. *Let $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ be any plan. Then, for every 4-plan $\boldsymbol{\beta}$ satisfying*

$$\pi^{0,1,2}_{\#}\boldsymbol{\beta} = \boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}, \tag{4.37a}$$

$$\pi^{0,3}_{\#}\boldsymbol{\beta} = \boldsymbol{\eta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), \tag{4.37b}$$

it holds:

i) $\langle x_1, x_3\rangle_{\boldsymbol{\beta}} = \langle x_2, x_3\rangle_{\boldsymbol{\beta}}$.
ii) $\langle x_1, x_3\rangle_{\boldsymbol{\beta}} = \langle \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle$ *if and only if* $\langle x_2, x_3\rangle_{\boldsymbol{\beta}} = \langle \mathcal{P}(\boldsymbol{\gamma}), \boldsymbol{\eta}\rangle$.
iii) $\pi^{0,1,3}_{\#}\boldsymbol{\beta} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ *if and only if* $\pi^{0,2,3}_{\#}\boldsymbol{\beta} \in \mathcal{OPT}_\mu(\mathcal{P}(\boldsymbol{\gamma}), \boldsymbol{\eta})$.

Conversely let $\overline{\boldsymbol{\gamma}}$ be a plan in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, and assume that there exists a 3-plan $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}, \overline{\boldsymbol{\gamma}})$ such that for every 4-plan $\boldsymbol{\beta}$ satisfying

$$\pi^{0,1,2}_{\#}\boldsymbol{\beta} = \boldsymbol{\alpha}, \tag{4.38a}$$

$$\pi^{0,3}_{\#}\boldsymbol{\beta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), \tag{4.38b}$$

it holds

$$\langle x_1, x_3\rangle_{\boldsymbol{\beta}} = \langle x_2, x_3\rangle_{\boldsymbol{\beta}}.$$

Then $\overline{\boldsymbol{\gamma}} = \mathcal{P}(\boldsymbol{\gamma})$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}$.

Proof. It is clear that *ii*) and *iii*) follow directly from *i*). To prove *i*) observe that $\boldsymbol{\gamma}_\varepsilon := (\pi^0, \pi^2 + \varepsilon\pi^3)_{\#}\boldsymbol{\beta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ for any $\varepsilon \in \mathbb{R}$, and evaluate

$$W^2_\mu(\boldsymbol{\gamma}, \boldsymbol{\gamma}_\varepsilon) \le \|x_1 - x_2 - \varepsilon x_3\|^2_{\boldsymbol{\beta}} = \|x_1 - x_2\|^2_{\boldsymbol{\beta}} - 2\varepsilon\langle x_1 - x_2, x_3\rangle_{\boldsymbol{\beta}} + \varepsilon^2\|x_3\|_{\boldsymbol{\beta}}.$$

To get the thesis let ε tend to 0^+ and 0^- and use the minimality of $\boldsymbol{\gamma}_0 = \mathcal{P}(\boldsymbol{\gamma})$.

Let us turn to the converse implication: choose any 3-plan $\boldsymbol{\alpha}$ such that $\pi^{0,1}_{\#}\boldsymbol{\alpha} = \boldsymbol{\gamma}$, $\overline{\boldsymbol{\gamma}} := \pi^{0,2}_{\#}\boldsymbol{\alpha} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and assume that $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}$. We will show that there exists a 4-plan $\boldsymbol{\beta}$ satisfying (4.38), for which $\langle x_1 - x_2, x_3\rangle_{\boldsymbol{\beta}}$ is different from 0. Indeed for any $\overline{\boldsymbol{\beta}}$ such that

$$\pi^{0,1,2}_{\#}\overline{\boldsymbol{\beta}} = \boldsymbol{\alpha},$$
$$\pi^{0,1,3}_{\#}\overline{\boldsymbol{\beta}} = \boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})},$$

we have that (thanks to what we just proved, and the fact that $(\pi^0, \pi^2 - \pi^3)_{\#}\boldsymbol{\beta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)))$ $\langle x_1 - x_3, x_2 - x_3\rangle_{\overline{\boldsymbol{\beta}}} = 0$. Therefore we can conclude with

$$\begin{aligned} 0 \neq \|x_2 - x_3\|^2_{\overline{\boldsymbol{\beta}}} =& \langle x_2 - x_3, x_2 - x_1\rangle_{\overline{\boldsymbol{\beta}}} + \langle x_2 - x_3, x_1 - x_3\rangle_{\overline{\boldsymbol{\beta}}} \\ =& \langle x_2 - x_3, x_2 - x_1\rangle_{\overline{\boldsymbol{\beta}}} = \langle x_3, x_2 - x_1\rangle_{\boldsymbol{\beta}}, \end{aligned}$$

where $\boldsymbol{\beta} := (\pi^0, \pi^1, \pi^2, \pi^2 - \pi^3)_{\#}\overline{\boldsymbol{\beta}}$ satisfies (4.38). $\square$

COROLLARY 4.34. *Let $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ be any plan. The plan $\mathcal{P}(\boldsymbol{\gamma})$ is the unique element of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ with the following property:*

$$\langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle = \langle \mathcal{P}(\boldsymbol{\gamma}), \boldsymbol{\eta} \rangle, \quad \forall \boldsymbol{\eta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)). \tag{4.39}$$

Proof. From $i)$ and $ii)$ of the previous theorem it follows easily that $\mathcal{P}(\boldsymbol{\gamma})$ has the stated property. Now suppose that $\boldsymbol{\sigma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ satisfies (4.39), too, and observe that from

$$\|\mathcal{P}(\boldsymbol{\gamma})\|^2 = \langle \mathcal{P}(\boldsymbol{\gamma}), \boldsymbol{\gamma} \rangle = \langle \mathcal{P}(\boldsymbol{\gamma}), \boldsymbol{\sigma} \rangle = \langle \boldsymbol{\gamma}, \boldsymbol{\sigma} \rangle = \|\boldsymbol{\sigma}\|^2,$$

we obtain $\|\mathcal{P}(\boldsymbol{\gamma})\| = \|\boldsymbol{\sigma}\|$ and $\langle \boldsymbol{\sigma}, \mathcal{P}(\boldsymbol{\gamma}) \rangle = \|\mathcal{P}(\boldsymbol{\gamma})\|^2 = \|\boldsymbol{\sigma}\|^2$, from which the thesis follows. □

The theorem we just proved allows us to prove the "linearity" of $\mathcal{P}$ stated in the following theorem.

THEOREM 4.35 (Linearity of $\mathcal{P}$). *Let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ be plan in $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$, and let λ be a real number. Then it holds:*

$$\mathcal{P}(\lambda \cdot \boldsymbol{\gamma}_1) = \lambda \cdot \mathcal{P}(\boldsymbol{\gamma}_1), \tag{4.40a}$$

$$\mathcal{P}(\boldsymbol{\gamma}_1 \oplus \boldsymbol{\gamma}_2) = \mathcal{P}(\boldsymbol{\gamma}_1) \oplus \mathcal{P}(\boldsymbol{\gamma}_2). \tag{4.40b}$$

Proof. The first equation follows easily by the observation that for every couple of plan $\boldsymbol{\gamma}, \boldsymbol{\eta}$ it holds: $\boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ if and only if $\lambda \cdot \boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\lambda \cdot \boldsymbol{\gamma}, \lambda \cdot \boldsymbol{\eta})$, where $\lambda \cdot \boldsymbol{\alpha}$ is intended to be the plan $(\pi^0, \lambda\pi^1, \lambda\pi^2)_\# \boldsymbol{\alpha}$.

Now we turn to the second one. Fix a 3-plan $\boldsymbol{\alpha}$ admissible for the couple $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ and define $\boldsymbol{\gamma}_3 := \boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}} \boldsymbol{\gamma}_2 = (\pi^0, \pi^1 + \pi^2)_\# \boldsymbol{\alpha}$. Our goal is to prove that there exists a 3-plan which we call $\mathcal{P}(\boldsymbol{\alpha})$ admissible for the couple $\mathcal{P}(\boldsymbol{\gamma}_1), \mathcal{P}(\boldsymbol{\gamma}_2)$, such that $\mathcal{P}(\boldsymbol{\gamma}_1) +_{\mathcal{P}(\boldsymbol{\alpha})} \mathcal{P}(\boldsymbol{\gamma}_2) = \mathcal{P}(\boldsymbol{\gamma}_3)$. Let us use the variables x_1, x_2, x_3, x_4 for the plans $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \mathcal{P}(\boldsymbol{\gamma}_1), \mathcal{P}(\boldsymbol{\gamma}_2)$, respectively. Choose a 5-plan $\boldsymbol{\beta}$ satisfying

$$\pi^{0,1,2}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}, \tag{4.41a}$$

$$\pi^{0,1,3}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_{\boldsymbol{\gamma}_1}^{\mathcal{P}(\boldsymbol{\gamma}_1)}, \tag{4.41b}$$

$$\pi^{0,2,4}_\# \boldsymbol{\beta} = \boldsymbol{\alpha}_{\boldsymbol{\gamma}_2}^{\mathcal{P}(\boldsymbol{\gamma}_2)}, \tag{4.41c}$$

and define $\mathcal{P}(\boldsymbol{\alpha}) := \pi^{0,2,4}_\# \boldsymbol{\beta}$ (observe that since $\boldsymbol{\alpha}_{\boldsymbol{\gamma}_i}^{\mathcal{P}(\boldsymbol{\gamma}_i)}$ is induced by a map on each fibre - see proposition 4.32 - there is a unique $\boldsymbol{\beta}$ satifying the above conditions, so $\mathcal{P}(\boldsymbol{\alpha})$ is actually a well defined function of $\boldsymbol{\alpha}$). With this notation our claim becomes: prove that the projetion of $(\pi^0, \pi^1 + \pi^2)_\# \boldsymbol{\beta}$ is $(\pi^0, \pi^3 + \pi^4)_\# \boldsymbol{\beta}$. In order to prove this we will use the second part of the theorem 4.33: choose any 6-plan $\mathbf{B}$ satisfying:

$$\pi^{0,1,2,3,4}_\# \mathbf{B} = \boldsymbol{\beta},$$

$$\pi^{0,5}_\# \mathbf{B} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)).$$

All we have to prove is the equality

$$\langle x_1 + x_2, x_5 \rangle_{\mathbf{B}} = \langle x_3 + x_4, x_5 \rangle_{\mathbf{B}},$$

whose proof is obvious, since $\langle x_1, x_5 \rangle_{\mathbf{B}} = \langle x_3, x_5 \rangle_{\mathbf{B}}$ and $\langle x_2, x_4 \rangle_{\mathbf{B}} = \langle x_3, x_5 \rangle_{\mathbf{B}}$ because of theorem 4.33 applied to the couples $\boldsymbol{\gamma}_1, \mathcal{P}(\boldsymbol{\gamma}_1)$ and $\boldsymbol{\gamma}_2, \mathcal{P}(\boldsymbol{\gamma}_2)$. □

Observe that in particular we proved that

$$\mathcal{P}\big(\boldsymbol{\gamma}_1 +_{\boldsymbol{\alpha}} \boldsymbol{\gamma}_2\big) = \mathcal{P}(\boldsymbol{\gamma}_1) +_{\mathcal{P}(\boldsymbol{\alpha})} \mathcal{P}(\boldsymbol{\gamma}_2).$$

REMARK 4.36. *It is not clear to the author whether the plan $\mathcal{P}(\boldsymbol{\alpha})$ is optimal whenever $\boldsymbol{\alpha}$ is.*

In the following proposition we show that the map $\mathcal{P}$ is 1-Lipschitz.

HILBERT CASE 2. *Let $V \subset H$ be a closed subspace of an Hilbert space H, let $v_1, v_2 \in H$ be two fixed vectors and let $P : H \to V$ be the orthogonal projection. Then $|P(v_1) - P(v_2)| \leq |v_1 - v_2|$.*

Proof. We have that

$$\begin{aligned}\langle v_1 - P(v_1), P(v_1) - P(v_2)\rangle = 0,\\ \langle v_2 - P(v_2), P(v_1) - P(v_2)\rangle = 0.\end{aligned}$$

Therefore the thesis follows from

$$\begin{aligned}|v_1 - v_2|^2 =&|P(v_1) - P(v_2) + (v_1 - P(v_1)) + (v_2 - P(v_2))|^2\\ =&|P(v_1) - P(v_2)|^2 + 2\langle P(v_1) - P(v_2), v_1 - P(v_1)\rangle\\ &+ 2\langle P(v_1) - P(v_2), v_2 - P(v_2)\rangle + |v_1 - P(v_1) + v_2 - P(v_2)|^2\\ =&|P(v_1) - P(v_2)|^2 + |v_1 - P(v_1) + v_2 - P(v_2)|^2 \geq |P(v_1) - P(v_2)|^2.\end{aligned}$$

□

COROLLARY 4.37 (Non expansivity of $\mathcal{P}$). *Let $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be any couple of plans. Then it holds*

$$W_\mu(\mathcal{P}(\boldsymbol{\gamma}_1), \mathcal{P}(\boldsymbol{\gamma}_2)) \leq W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2).$$

Proof. Let x_1, x_2, x_3, x_4 be the variables of $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \mathcal{P}(\boldsymbol{\gamma}_1), \mathcal{P}(\boldsymbol{\gamma}_2)$. Choose $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and find a 5-plan $\boldsymbol{\beta}$ satisfying

$$\begin{aligned}\pi^{0,1,2}_{\#}\boldsymbol{\beta} &= \boldsymbol{\alpha},\\ \pi^{0,1,3}_{\#}\boldsymbol{\beta} &= \boldsymbol{\alpha}^{\mathcal{P}(\boldsymbol{\gamma}_1)}_{\boldsymbol{\gamma}_1},\\ \pi^{0,2,4}_{\#}\boldsymbol{\beta} &= \boldsymbol{\alpha}^{\mathcal{P}(\boldsymbol{\gamma}_2)}_{\boldsymbol{\gamma}_2}.\end{aligned}$$

Since $(\pi^0, \pi^3 - \pi^4)_{\#}\boldsymbol{\beta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ we know from theorem 4.33 that it holds

$$\begin{aligned}\langle x_3 - x_4, x_1 - x_3\rangle_{\boldsymbol{\beta}} = 0,\\ \langle x_3 - x_4, x_4 - x_2\rangle_{\boldsymbol{\beta}} = 0,\end{aligned}$$

therefore we conclude with:

$$\begin{aligned}W^2_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) =&\|x_1 - x_2\|^2_{\boldsymbol{\beta}} = \|(x_3 - x_4) + (x_1 - x_3) + (x_4 - x_2)\|^2_{\boldsymbol{\beta}}\\ =&\|x_3 - x_4\|^2_{\boldsymbol{\beta}} + 2\langle x_3 - x_4, x_1 - x_3\rangle_{\boldsymbol{\beta}}\\ &+ 2\langle x_3 - x_4, x_4 - x_2\rangle_{\boldsymbol{\beta}} + \|x_1 - x_3 + x_4 - x_2\|^2_{\boldsymbol{\beta}}\\ =&\|x_3 - x_4\|^2_{\boldsymbol{\beta}} + \|x_1 - x_3 + x_4 - x_2\|^2_{\boldsymbol{\beta}} \geq W_\mu(\mathcal{P}(\boldsymbol{\gamma}_1), \mathcal{P}(\boldsymbol{\gamma}_2)).\end{aligned}$$

□

Observe that from this corollary it follows in particular (taking $\boldsymbol{\gamma}_2 = \mathbf{0}_\mu$) that

$$\|\mathcal{P}(\boldsymbol{\gamma})\| \leq \|\boldsymbol{\gamma}\|.$$

We now analyze the properties of $\mathcal{P}$ w.r.t. some of the other functions we defined. Recall that P_μ is the orthogonal projection of L^2_μ onto $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, that $\mathfrak{I}(f) = (Id, f)_\# \mu$ and that $\mathscr{B}$ is the barycentric projection.

PROPOSITION 4.38 (Further properties of $\mathcal{P}$). *It holds*

$$\begin{aligned}
\mathcal{P}(\mathcal{P}(\boldsymbol{\gamma})) &= \mathcal{P}(\boldsymbol{\gamma}),\\
\langle \mathcal{P}(\boldsymbol{\gamma}_1), \boldsymbol{\gamma}_2\rangle &= \langle \boldsymbol{\gamma}_1, \mathcal{P}(\boldsymbol{\gamma}_2)\rangle,\\
\mathcal{P}(\mathfrak{I}(f)) &= \mathfrak{I}(P_\mu(f)),\\
P_\mu(\mathscr{B}(\boldsymbol{\gamma})) &= \mathscr{B}(\mathcal{P}(\boldsymbol{\gamma})).
\end{aligned}$$

Proof. The first identity is obvious. For the second one observe $\langle \mathcal{P}(\boldsymbol{\gamma}_1), \boldsymbol{\gamma}_2\rangle = \langle \mathcal{P}(\boldsymbol{\gamma}_1), \mathcal{P}(\boldsymbol{\gamma}_2)\rangle = \langle \boldsymbol{\gamma}_1, \mathcal{P}(\boldsymbol{\gamma}_2)\rangle$. To prove the third one observe that it is sufficient to show that the projection of a plan induced by a map is induced by a map, too: to prove this it is sufficient to recall that the plan $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}$ is induced by a map in the sense of equation (4.36). To prove the last equation recall the properties of the objects involved:

$$\begin{array}{lll}
\overline{\boldsymbol{\gamma}} = \mathcal{P}(\boldsymbol{\gamma}) \Leftrightarrow \overline{\boldsymbol{\gamma}} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), \text{ and } \langle \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle = \langle \overline{\boldsymbol{\gamma}}, \boldsymbol{\eta}\rangle, & \forall \boldsymbol{\eta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)),\\
f = \mathscr{B}(\boldsymbol{\gamma}) \Leftrightarrow \langle \boldsymbol{\gamma}, \mathfrak{I}(g)\rangle = \langle f, g\rangle_\mu, & \forall g \in L^2_\mu,\\
v = P_\mu(f) \Leftrightarrow v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) \text{ and } \langle f, g\rangle_\mu = \langle v, g\rangle_\mu, & \forall g \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)).
\end{array}$$

Therefore

$$\begin{aligned}
&v = P_\mu(\mathscr{B}(\boldsymbol{\gamma}))\\
&\qquad\Leftrightarrow\quad v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) \text{ and } \langle \mathscr{B}(\boldsymbol{\gamma}), g\rangle_\mu = \langle v, g\rangle_\mu \quad \forall g \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))\\
&\qquad\Leftrightarrow\quad v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) \text{ and } \langle \boldsymbol{\gamma}, \mathfrak{I}(g)\rangle = \langle v, g\rangle_\mu \quad \forall g \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)),
\end{aligned}$$

and on the other hand

$$\begin{aligned}
&\overline{v} = \mathscr{B}(\mathcal{P}(\boldsymbol{\gamma}))\\
&\qquad\Leftrightarrow\quad \langle \mathcal{P}(\boldsymbol{\gamma}), \mathfrak{I}(f)\rangle = \langle v, f\rangle_\mu \quad \forall f \in L^2_\mu,\\
&\qquad\Leftrightarrow\quad \langle \boldsymbol{\gamma}, \mathfrak{I}(P_\mu(f))\rangle = \langle v, f\rangle_\mu \quad \forall f \in L^2_\mu,
\end{aligned}$$

therefore clearly $v = \overline{v}$. □

Before proving our final result, which regards the infinitesimal behavior of the exponential map, we state a lemma regarding convergence of $\mathcal{P}$, which, although not used in the sequel, is of its own interest. We already showed that $\mathcal{P}$ is Lipschitz on $(\mathscr{P}_2(\mathbb{R}^{2d})_\mu, W_\mu)$, but what about the weak continuity? The following example shows that in general we cannot hope for weak continuity.

EXAMPLE 4.39. *Let* $\mu := (2\pi)^{-1}\mathcal{L}^1|_{[0,2\pi]}$, $v_n(x) := \sin nx$. *It is not difficult to see that, given that* $d = 1$, *it holds* $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) = L^2_\mu$. *In particular we have that* $v_n \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *which implies* $\mathcal{P}(\mathfrak{I}(v_n)) = \mathfrak{I}(v_n)$. *Let* $\boldsymbol{\gamma}$ *be the* $W-$*limit of* $(\mathfrak{I}(v_n))$:

$$d\boldsymbol{\gamma}(x_0, x_1) = \frac{1}{(2\pi^2)\sqrt{1 - x_1^2}} d(x_0, x_1)|_Q,$$

where $Q = [0, 2\pi] \times [-1, 1]$. In particular $\boldsymbol{\gamma}$ is not induced by a map. Given that μ is absolutely continuous, the space $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is isometric to $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ through the map $\mathfrak{I}$. It follows that $\boldsymbol{\gamma} \notin \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and therefore $\mathcal{P}(\mathfrak{I}(v_n)) = \mathfrak{I}(v_n)$ can't converge (not even weakly) to $\mathcal{P}(\boldsymbol{\gamma})$ since the $W-$limit of the sequence is $\boldsymbol{\gamma}$ and does not belong to $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.

LEMMA 4.40. *Let $(\boldsymbol{\gamma}_n) \subset \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be a sequence of plans $\tau-$converging to some plan $\boldsymbol{\gamma}$. Then for every $\tau-$accumulation point $\boldsymbol{\sigma}$ of $\mathcal{P}(\boldsymbol{\gamma}_n)$ it holds $\mathcal{P}(\boldsymbol{\gamma}) = \mathcal{P}(\boldsymbol{\sigma})$. In particular*

$$\|\mathcal{P}(\boldsymbol{\gamma})\| \leq \varliminf_{n\to\infty} \|\mathcal{P}(\boldsymbol{\gamma}_n)\|.$$

Proof. We can assume without loss of generality that $\boldsymbol{\alpha}_{\boldsymbol{\gamma}_n}^{\mathcal{P}(\boldsymbol{\gamma}_n)}$ $\tau-$converge to some $\boldsymbol{\alpha}$ with $\pi^{0,2}_\#\boldsymbol{\alpha} = \boldsymbol{\sigma}$.

We argue by contradiction: suppose that $\mathcal{P}(\boldsymbol{\gamma}) \neq \mathcal{P}(\boldsymbol{\sigma})$. Choose a 5-plan $\mathbf{B}$ satisfying

$$\begin{aligned}
\pi^{0,1,2}_\#\mathbf{B} &= \boldsymbol{\alpha}, \\
\pi^{0,1,3}_\#\mathbf{B} &= \boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}, \\
\pi^{0,2,4}_\#\mathbf{B} &= \boldsymbol{\alpha}_{\boldsymbol{\sigma}}^{\mathcal{P}(\boldsymbol{\sigma})},
\end{aligned}$$

and use part *i*) of the theorem 4.33 (observe that $(\pi^0, \pi^3 - \pi^4)_\#\mathbf{B} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$) to get

$$\begin{aligned}
0 < W_\mu^2(\mathcal{P}(\boldsymbol{\gamma}), \mathcal{P}(\boldsymbol{\sigma})) \leq& \|x_3 - x_4\|^2_{\mathbf{B}} \\
=& \langle x_3, x_3 - x_4\rangle_{\mathbf{B}} - \langle x_4, x_3 - x_4\rangle_{\mathbf{B}} \\
=& \langle x_1, x_3 - x_4\rangle_{\mathbf{B}} - \langle x_2, x_3 - x_4\rangle_{\mathbf{B}} \\
=& \langle x_1 - x_2, x_3 - x_4\rangle_{\mathbf{B}}.
\end{aligned}$$

Define the plan $\boldsymbol{\beta} := (\pi^0, \pi^1, \pi^2, \pi^3 - \pi^4)_\#\mathbf{B}$ and observe that we just proved

$$\langle x_1 - x_2, x_3\rangle_{\boldsymbol{\beta}} > 0. \tag{4.43}$$

We want to find a sequence $\boldsymbol{\beta}_n$ $\tau-$converging to $\boldsymbol{\beta}$ satisfying

$$\begin{aligned}
\pi^{0,1,2}_\#\boldsymbol{\beta}_n &= \boldsymbol{\alpha}_{\boldsymbol{\gamma}_n}^{\mathcal{P}(\boldsymbol{\gamma}_n)}, \\
\pi^{0,3}_\#\boldsymbol{\beta}_n &= \boldsymbol{\eta}.
\end{aligned} \tag{4.44}$$

In order to do so, find, for every $n \in \mathbb{N}$, a 7-plan $\mathbf{B}_n$ satisfying

$$\begin{aligned}
\pi^{0,1,2,3}_\#\mathbf{B}_n &= \boldsymbol{\beta}, \\
\pi^{0,1,2,4,5,6}_\#\mathbf{B}_n &\in \mathcal{O}pt(\boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}_n}^{\mathcal{P}(\boldsymbol{\gamma}_n)}),
\end{aligned}$$

where in the second equation we thought $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})} \in \mathscr{P}_2(\mathbb{R}^d_0 \times \mathbb{R}^d_1 \times \mathbb{R}^d_2)$ and $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})} \in \mathscr{P}_2(\mathbb{R}^d_4 \times \mathbb{R}^d_5 \times \mathbb{R}^d_6)$. The $\tau-$convergence of $\boldsymbol{\alpha}_{\boldsymbol{\gamma}_n}^{\mathcal{P}(\boldsymbol{\gamma}_n)}$ to $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}^{\mathcal{P}(\boldsymbol{\gamma})}$ implies that $(\mathbf{B}_n)$

$\tau-$converges to $(\pi^0,\pi^1,\pi^2,\pi^3,\pi^0,\pi^1,\pi^2)_\#\boldsymbol{\beta}$. The 4-plans $\boldsymbol{\beta}_n := (\pi^4,\pi^5,\pi^6,\pi^3)_\#\mathbf{B}_n$ satisfy (4.44), therefore from part *i*) of theorem 4.33 we have

$$\langle x_5 - x_6, x_3\rangle_{\mathbf{B}_n} = \langle x_5 - x_6, x_3\rangle_{\boldsymbol{\beta}_n} = 0.$$

Passing to the limit as $n\to\infty$ we get

$$0 = \lim_{n\to\infty}\langle x_6 - x_5, x_3\rangle_{\mathbf{B}_n} = \langle x_1 - x_2, x_3\rangle_{\boldsymbol{\beta}},$$

which contradicts (4.43). Therefore it must hold $\mathcal{P}(\boldsymbol{\gamma}) = \mathcal{P}(\boldsymbol{\sigma})$.

For the second statement observe that the $\tau-$convergence of $\boldsymbol{\alpha}_{\boldsymbol{\gamma}_n}^{\mathcal{P}(\boldsymbol{\gamma}_n)}$ to $\boldsymbol{\alpha}$ implies the $\tau-$convergence of $\mathcal{P}(\boldsymbol{\gamma}_n)$ to $\boldsymbol{\sigma}$, therefore:

$$\|\mathcal{P}(\boldsymbol{\gamma})\| = \|\mathcal{P}(\boldsymbol{\sigma})\| \le \|\boldsymbol{\sigma}\| \le \varliminf_{n\to\infty}\|\mathcal{P}(\boldsymbol{\gamma}_n)\|.$$

□

The next (and final) theorem regards the infinitesimal behavior of the exponential function: there we prove that the two curves $t\to\exp(t\cdot\boldsymbol{\gamma})$ and $t\to\exp(t\cdot\mathcal{P}(\boldsymbol{\gamma}))$ behave in the same way near 0 *under the assumption that the plan $\boldsymbol{\gamma}$ is induced by a map.* For the general case we are able only to prove only a partial result.

THEOREM 4.41 (Infinitesimal behavior of $\exp_\mu$). *Let $\boldsymbol{\gamma}\in\mathscr{P}_2(\mathbb{R}^{2d})_\mu$ be a plan, $\mu_t := \exp(t\cdot\boldsymbol{\gamma})$ (so that $\mu_0=\mu$). Let $\boldsymbol{\eta}_t$ be any choice of optimal plans between μ_0 and μ_t (i.e. $\boldsymbol{\eta}_t\in\mathcal{O}pt(\mu_0,\mu_t)$), $t\in[0,1]$, and define*

$$\boldsymbol{\sigma}_t := \left(\pi^0, \frac{\pi^1-\pi^0}{t}\right)_\#\boldsymbol{\eta}_t.$$

Then it holds

$$\varlimsup_{t\to 0^+}\langle\boldsymbol{\sigma}_t,\boldsymbol{\sigma}\rangle \le \langle\mathcal{P}(\boldsymbol{\gamma}),\boldsymbol{\sigma}\rangle, \quad \forall\boldsymbol{\sigma}\in\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)). \tag{4.45}$$

Proof. By density and positive 1-homogeneity it is sufficient to prove equation (4.45) for all the plans $\boldsymbol{\sigma}\in\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ such that $(\pi^1,\pi^1+\pi^2)_\#\boldsymbol{\sigma}$ is the unique optimal plan between its marginals. With this assumption let us evaluate $\frac{d}{dt}W^2(\mu_t,\sigma)$ at $t=0$, where $\sigma := (\pi^1+\pi^2)_\#\boldsymbol{\sigma}$. From proposition 4.10 we know that it is equal to $-2\langle\boldsymbol{\gamma},\boldsymbol{\sigma}\rangle = -2\langle\mathcal{P}(\boldsymbol{\gamma}),\boldsymbol{\sigma}\rangle$, and on the other hand taking any $\boldsymbol{\alpha}_t\in\mathcal{OPT}_\mu(\boldsymbol{\eta}_t,(\pi^1,\pi^1+\pi^2)_\#\boldsymbol{\sigma})$ we have

$$\begin{aligned}\frac{d}{dt}|_{t=0}W^2(\mu_t,\mu_0) &= \lim_{t\to 0^+}\frac{W^2(\mu_t,\nu)-W^2(\mu_0,\nu)}{t}\\ &\le \lim_{t\to 0^+}\frac{\|x_1-x_2\|^2_{\boldsymbol{\alpha}_t}-\|x_0-x_2\|^2_{\boldsymbol{\alpha}_t}}{t}\\ &= -2\langle\frac{x_1-x_0}{t},x_2-x_0\rangle_{\boldsymbol{\alpha}_t} = \langle\boldsymbol{\sigma}_t,\boldsymbol{\sigma}\rangle.\end{aligned}$$

□

It may seem curious that for general plan we obtain only an inequality in the limit of scalar products, instead of an equality, however (although the author doesn't know whether the equality holds or not) we want to point out why it is not possible to derive the equality from the inequality. More explicitely: for every μ and every $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ there exists $\boldsymbol{\gamma}' \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ such that $\langle \boldsymbol{\gamma}', \boldsymbol{\sigma} \rangle \geq \langle \boldsymbol{\gamma}, \boldsymbol{\sigma} \rangle$ for any $\boldsymbol{\sigma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$, with strict inequality for some $\boldsymbol{\sigma}$. To see this, choose any vector $v \in \mathbb{R}^d$ different from 0, define the functions $f^{\pm}(x_0, x_1) := (x_0, x_1 \pm v)$ and the plan $\boldsymbol{\gamma}' := 1/2(f^+_\# \boldsymbol{\gamma} + f^-_\# \boldsymbol{\gamma})$. Then it is easy to see that for every $\boldsymbol{\sigma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ and every 3-plan $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}, \boldsymbol{\sigma})$, there exists $\boldsymbol{\alpha}' \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}', \boldsymbol{\sigma})$ such that $\langle x_1, x_2 \rangle_{\boldsymbol{\alpha}} = \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}'}$: define $\boldsymbol{\alpha}' := 1/2(\overline{f}^+_\# \boldsymbol{\alpha} + \overline{f}^-_\# \boldsymbol{\alpha})$, where $\overline{f}^{\pm}(x_0, x_1, x_2) := (x_0, x_1 \pm v, x_2)$. This shows that the inequality $\langle \boldsymbol{\gamma}', \boldsymbol{\sigma} \rangle \geq \langle \boldsymbol{\gamma}, \boldsymbol{\sigma} \rangle$ always holds. To show that sometimes it has to be strict, observe that $\|\boldsymbol{\gamma}'\| > \|\boldsymbol{\gamma}\|$, therefore at least one of the two inequalities in $\|\boldsymbol{\gamma}'\|^2 \geq \langle \boldsymbol{\gamma}', \boldsymbol{\gamma} \rangle \geq \|\boldsymbol{\gamma}\|^2$ has to be strict.

We conclude the chapter by underlying an interesting fact regarding curves of the type $t \to \exp(t \cdot \boldsymbol{\gamma})$, which actually creates difficulties when trying to generalize the previous theorem to general plans and shows that the parallel transport does not always exist (see chapter 6).

Fix any plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ and consider the curve $\mu_t := \exp(t \cdot \boldsymbol{\gamma})$. This curve is clearly absolutely continuous, therefore considering its tangent vector field v_t (which exist for a.e. t), we know from proposition 3.9 that for every $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and a.e. t it holds

$$\frac{d}{dt^+} W^2(\mu_t, \nu) = -2 \langle v_t(x), \mathscr{B}(\boldsymbol{\eta}_t) - Id \rangle_{\mu_t},$$

where $\boldsymbol{\eta}_t$ is any plan in $\mathcal{O}pt(\mu_t, \nu)$. On the other hand proposition 4.10 ensures that the same derivative is equal to $-2\langle \boldsymbol{\gamma}_t, (\pi^0, \pi^1 - \pi^0)_\# \boldsymbol{\eta}_t \rangle$, where $\boldsymbol{\gamma}_t := (\pi^0 + t\pi^1, \pi^1)_\# \boldsymbol{\gamma}$. The characterization we did of the projection of a plan implies that for a.e. t it holds

$$\mathcal{P}(\boldsymbol{\gamma}_t) = v_t. \tag{4.46}$$

In order to prove the previous theorem for general plans we would like to write the previous equation for $t = 0$, however this cannot be done, because we don't know whether v_t is well defined at $t = 0$.

The interesting fact is that the metric derivative $|\dot{\mu}_t|$ can exists for *every* t and be discontinuous at 0. Given that for a.e. t it holds $|\dot{\mu}_t| = \|v_t\|_{\mu_t} = \|\mathcal{P}(\boldsymbol{\gamma}_t)\|$, this partially explains why it is difficult to find properties of $\mathcal{P}(\boldsymbol{\gamma})$ by taking the limit as $t \to 0$ of similar properties of $\mathcal{P}(\boldsymbol{\gamma}_t)$.

The following is an example of a plan $\boldsymbol{\gamma}$ which is induced by a map for which $|\dot{\mu}_0| \leq 1$, while $|\dot{\mu}_t|$ is costant and bigger than 1 for positive t; moreover for positive t the curve μ_t is locally a geodesic (but is not a geodesic in any interval $[0, T]$) and the plans (actually functions) $\boldsymbol{\gamma}_t$ are tangent in μ_t.

EXAMPLE 4.42. *Let* $Q = [0,1] \times [0,1]$ *be the unit square in* $\mathbb{R}^2$ *and* T_i, $i = 1, 2, 3, 4$ *be the four triangles in which* Q *is divided by its diagonal. Define the function* $f : Q \to \mathbb{R}^2$

as

$$f_{|T_1} := (1,1),$$
$$f_{|T_2} := (-1,1),$$
$$f_{|T_3} := (-1,-1),$$
$$f_{|T_4} := (1,-1).$$

Now define $\mu := \mathcal{L}^2{}_{|Q}$ *and observe that* $f \in L^2_\mu$ *and that for* $t > 0$, $\mu_t := (Id + tf)_\# \mu$ *is made of 4 connected components, each one the traslation of one of the* T_i.

The function $f_t := f \circ (Id + tf)^{-1}$ *is constant on each of those components and therefore clearly tangent. It follows that for positive* t *we have* $|\dot{\mu}_t| = \|f_t\|_{\mu_t} = \|f\|_\mu = \sqrt{2}$. *Our example will be concluded if we show that* $|\dot{\mu}_0| \le 1$. *To see this consider the function* g *on* Q *defined as*

$$g_{|T_1} := (0,1),$$
$$g_{|T_2} := (-1,0),$$
$$g_{|T_3} := (0,-1),$$
$$g_{|T_4} := (1,0),$$

and let $\nu_t := (Id + tg)_\# \mu$.

Then clearly $W(\mu_t, \nu_t) \le \|f - g\|_\mu = 1$, *therefore the conclusion will follow if we show that* $W(\nu_t, \mu) = o(t)$. *This can be proved either by applying theorem 4.41 and checking that* $P_\mu(g) = 0$, *or, equivalently but in a more geometrical way, by finding a family of transport maps whose cost goes to* 0 *faster than* t: *for the latter approach see the picture below (a mass of order* t *is moved of a distance of order* t, *giving a cost of order* $t^{3/2}$).

CHAPTER 5

Geodesically convex sets

The aim of this chapter is to study the $\tau-$closure properties of geodesically convex sets. The reasons of such an interest come from the fact that the distance W was recently studied because of the strict relations with some evolution PDE's which may be interpreted as curves of maximal slope of certain *geodesically* convex functionals, i.e. functionals that are convex along geodesics. Such an approach, introduced by Otto in [**46**] and then further analyzed by Carrillo-McCann-Villani in [**25**], by Agueh in [**2**] and by the author together with Ambrosio and Savaré in [**11**] (see [**11**] for more detailed references), leads to the study of the problem of existence and uniqueness of those curves: the theory of minimizing movements introduced by De Giorgi ([**29**]) provides a satisfactory answer to these questions under only weak compactness assumptions. In [**11**] there are mainly two theorems on existence of curves of maximal slope for geodesically convex functionals which rely on two different kind of assumptions on the functional F:

i) F is lower semicontinuous w.r.t. the *weak* topology and satisfies $F(\boldsymbol{\gamma}[t]) \leq (1-t)F(\boldsymbol{\gamma}[0]) + tF(\boldsymbol{\gamma}[1])$ for any *optimal* plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})$ (see Corollaries 2.4.11 and 2.4.12 of [**11**]),

ii) F is lower semicontinuous w.r.t. the *strong* topology and satisfies $F(\boldsymbol{\gamma}[t]) \leq (1-t)F(\boldsymbol{\gamma}[0]) + tF(\boldsymbol{\gamma}[1])$ for *any* plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})$ (see Theorem 4.0.4 of [**11**]).

The two notions of convexity along geodesics just introduced are strictly related to the following notions of geodesic convexity for sets:

Definition 5.1 (Geodesically convex sets). *We will say that a set $C \subset \mathscr{P}_2(\mathbb{R}^d)$ is* geodesically convex *if for any $\mu_1, \mu_0 \in C$ there exists a $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu_0, \mu_1)$ such that the whole segment joining μ_0 to μ_1 through $\boldsymbol{\gamma}$ belongs to C, that is:*

$$[\boldsymbol{\gamma}](t) \in C, \quad \forall t \in [0,1].$$

Definition 5.2 (Strongly geodesically convex sets). *We will say that a set $C \subset \mathscr{P}_2(\mathbb{R}^d)$ is* strongly geodesically convex *if for any $\mu_1, \mu_0 \in C$ and every $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu_0, \mu_1)$ the whole segment joining μ_0 to μ_1 through $\boldsymbol{\gamma}$ belongs to C, that is:*

$$[\boldsymbol{\gamma}](t) \in C, \quad \forall t \in [0,1].$$

It is easy to check that a functional F l.s.c. w.r.t. the strong topology is convex along geodesics in the sense of (i) (respectively, (ii)), if and only if its sublevels are geodesically convex (respectively, strongly geodesically convex).

The main result of this chapter is to show that in case (i) the assumption of lower semicontinuity w.r.t. τ is redundant and may be substituted with semicontinuity w.r.t. W, *provided we know that* F *attains the value* $+\infty$ *on non regular measures*. In order to prove this we will show that any $W-$closed geodesically convex subset of $\mathscr{P}_2(\mathbb{R}^d)$ is $\tau-$closed.

The chapter has the following structure: we first introduce the concept of *halfspace* and study its basic properties, after that we prove that any $W-$closed strongly geodesically convex set is $\tau-$closed (thus in consistency with the results proven under the first kind of hypothesis on the functional F), and finally we turn to $W-$closed geodesically convex subsets of $\mathscr{P}_2(\mathbb{R}^d)$ and we prove that those sets are $\tau-$closed, too. In the last section we will shortly discuss the definition of the subdifferential of a geodesically convex functional and we analyze the differentiability properties of $W^2(\cdot,\sigma)$.

As in the last sections of the previous chapter it will happen that a theorem stated in the Wasserstein space is the analogous of a theorem valid in Hilbert spaces, in this case we will anticipate the proof of our theorem with the simpler one valid in a Euclidean setting.

1. Halfspaces

In this section we define the halfspaces and discuss their basic properties.

Definition 5.3 (Halfspace). *Let* $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *be any tangent plan, and* $C \in \mathbb{R}$ *a constant. The two halfspaces identified by* $\boldsymbol{\gamma}$ *and* C *are*

$$\mathscr{H}^+_{\boldsymbol{\gamma};C} := \left\{ \nu \in \mathscr{P}_2(\mathbb{R}^d) \ : \ \langle \boldsymbol{\eta}, \boldsymbol{\gamma} \rangle \geq C, \ \text{for some } \boldsymbol{\eta} \in \exp_\mu^{-1}(\nu) \right\},$$
$$\mathscr{H}^-_{\boldsymbol{\gamma};C} := \left\{ \nu \in \mathscr{P}_2(\mathbb{R}^d) \ : \ \langle \boldsymbol{\eta}, \boldsymbol{\gamma} \rangle \leq C, \ \text{for some } \boldsymbol{\eta} \in \exp_\mu^{-1}(\nu) \right\}.$$

For regular measures μ *and* $v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *we will write* $\mathscr{H}^\pm_{v;C}$*, in this case the definition simplifies in*

$$\mathscr{H}^+_{v;C} := \left\{ \nu \ : \ \langle T_\mu^\nu - Id, v \rangle_\mu \geq C \right\},$$
$$\mathscr{H}^-_{v;C} := \left\{ \nu \ : \ \langle T_\mu^\nu - Id, v \rangle_\mu \leq C \right\}.$$

Proposition 5.4. *Let* $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *be a plan and* $C \in \mathbb{R}$ *a real number. Then it holds:*

i) *The halfspace* $\mathscr{H}^+_{\boldsymbol{\gamma};C}$ *is weakly closed (i.e. closed w.r.t.* τ*).*
ii) *If the plan* $\boldsymbol{\gamma}$ *is induced by a map (and in particular if the base measure* μ *is regular), then the halfspace* $\mathscr{H}^-_{\boldsymbol{\gamma};C}$ *is weakly closed, too.*

Proof. Let us start with i): consider a sequence ν_n $\tau-$converging to ν and belonging to $\mathscr{H}^+_{\boldsymbol{\gamma};C}$, and a sequence $(\boldsymbol{\eta}_n)$ such that $\boldsymbol{\eta}_n \in \exp_\mu^{-1}(\nu_n)$ such that $\langle \boldsymbol{\gamma}, \boldsymbol{\eta}_n \rangle \geq C$, for all n. Then choose $\boldsymbol{\alpha}_n \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}_n)$ and observe that there must exist a subsequence $\boldsymbol{\alpha}_{n_k}$ which $\tau-$converges to some $\boldsymbol{\alpha}$ such that $\boldsymbol{\eta} := \pi^{0,2}_\#\boldsymbol{\alpha} \in \exp_\mu^{-1}(\nu)$. Corollary 2.29 ensures that

$$\langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle \geq \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}} = \lim_k \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}_{n_k}} \geq C.$$

The claim *ii*) follows with the same idea. Use the same notation as before and assume that $\langle \boldsymbol{\gamma}, \boldsymbol{\eta}_n \rangle \leq C$. Observe that if $\boldsymbol{\gamma} = (Id, T)_{\#}\mu$, for any plan $\overline{\boldsymbol{\gamma}} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$ the set $\mathcal{OPT}_\mu(\boldsymbol{\gamma}, \overline{\boldsymbol{\gamma}})$ contains only one element, namely $(\pi^0, T \circ \pi^0, \pi^1)_{\#}\overline{\boldsymbol{\gamma}}$. Therefore the limit plan $\boldsymbol{\alpha}$ is optimal and it holds

$$\langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle = \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}} = \lim_k \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}_{n_k}} \leq C.$$

□

REMARK 5.5. *It is important to observe that, for non regular measure μ, an half-space $\mathscr{H}^-_{\boldsymbol{\gamma};C}$ may fail to be $\tau-$closed. Moreorer, in general this set may fail even to be $W-$closed. Indeed consider the following example: let $\mu \in \mathscr{P}(\mathbb{R}^2)$ be defined as the 1-dimensional Hausdorff measure restricted to the interval $[0,1] \times \{0\}$, consider the subsets A_n of $[0,1]$ defined as*

$$A_n := \bigcup_{i=0}^{2^n-1} \left[\frac{2i}{2^{n+1}}, \frac{2i+1}{2^{n+1}} \right],$$

and define the functions $f_n : \mathbb{R} \to \mathbb{R}$ as

$$f_n := \begin{cases} 1 & \text{if } x \in A_n, \\ -1 & \text{if } x \notin A_n. \end{cases}$$

Then it is not difficult to see that if we define $\nu_n := (Id, f_n)_{\#}\mu$ (where we think μ as a measure on $\mathbb{R}$), then the plans $\boldsymbol{\gamma}_n$ induced by the maps (Id, f_n) are the only optimal plans between μ an ν_n. Moreover it is clear that the v_n W-converge to the measure

$$\nu := \frac{1}{2} \left((Id, 1)_{\#}\mu + (Id, -1)_{\#}\mu \right),$$

having identified once again μ with its restriction to $\mathbb{R}$, and that there is only one optimal plan between μ and ν, namely the plan $\boldsymbol{\gamma}$ whose disintegration w.r.t. to $d\mu(x,y)$ is given by the family $1/2(\delta_{x,1} + \delta_{x,-1})$. Then, we have that $W_\mu(\boldsymbol{\gamma}_n, \boldsymbol{\gamma}) = 1$ for every n, and that $\|\boldsymbol{\gamma}_n\| = \|\boldsymbol{\gamma}\| = 1$, therefore from the equality

$$W^2_\mu(\boldsymbol{\gamma}_n, \boldsymbol{\gamma}) = \|\boldsymbol{\gamma}\|^2 + \|\boldsymbol{\gamma}_n\|^2 - 2\langle \boldsymbol{\gamma}, \boldsymbol{\gamma}_n \rangle$$

we gain $\lim \langle \boldsymbol{\gamma}, \boldsymbol{\gamma}_n \rangle = 1/2$. As a consequence, for any C satisfying $1/2 < C < 1$ the halfspace $\mathscr{H}^-_{\boldsymbol{\gamma};C}$ is not even $W-$closed, since $\nu_n \in \mathscr{H}^-_{\boldsymbol{\gamma};C}$ for every n, but $\nu \notin \mathscr{H}^-_{\boldsymbol{\gamma};C}$.

REMARK 5.6. *The definition of halfspace make sense for any $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$, and not only for those in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. We restricted the interest to tangent plans because of corollary 4.34, which ensures that $\mathscr{H}^\pm_{\boldsymbol{\gamma};C} = \mathscr{H}^\pm_{\mathcal{P}(\boldsymbol{\gamma});C}$, where $\mathcal{P}(\boldsymbol{\gamma})$ is the projection of $\boldsymbol{\gamma}$ onto $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.*

2. Strongly geodesically convex sets

In this section we assume that the $W-$closed set C satisfies the strong geodesic convexity hypothesis given in definition 5.2, and we will prove that it is weakly closed.

The first step of the proof is to prove that for any point $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a point in C of minimal distance from μ (the proof is very similar to that of proposition 4.30).

Proposition 5.7. *Let $C \subset \mathscr{P}_2(\mathbb{R}^d)$ be a $W-$closed strongly geodesically convex set and let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a measure. Then there exists a measure $\nu \in C$ which minimizes the distance from μ among all the measures in C.*

Proof. Let $D := W(\mu, C)$, let (ν_n) be a minimizing sequence, choose $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu, \nu_n)$ for any n
$in\mathbb{N}$ and fix $\boldsymbol{\alpha}_n^m \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_n, \boldsymbol{\gamma}_m)$. Since $\pi^{1,2}_\# \boldsymbol{\alpha}_n^m \in \mathcal{A}dm(\nu_n, \nu_m)$ we have $(\pi^1/2 + \pi^2/2)_\# \boldsymbol{\alpha}_n^m \in C$, therefore the Hilbertian identity

$$\left|\frac{b-c}{2}\right|^2 = \frac{1}{2}|a-b|^2 + \frac{1}{2}|a-c|^2 - \left|a - \frac{b+c}{2}\right|^2,$$

yields to

$$\begin{aligned}\frac{1}{4}W^2(\nu_n, \nu_m) \leq& \frac{1}{4}\|x_1 - x_2\|^2_{\boldsymbol{\alpha}_n^m} \\ =& \frac{1}{2}\|x_0 - x_1\|^2_{\boldsymbol{\alpha}_n^m} + \frac{1}{2}\|x_0 - x_2\|^2_{\boldsymbol{\alpha}_n^m} - \left\|x_0 - \frac{x_1 + x_2}{2}\right\|^2_{\boldsymbol{\alpha}_n^m} \\ \leq& \frac{1}{2}W^2(\mu, \nu_n) + \frac{1}{2}W^2(\mu, \nu_m) - D^2.\end{aligned}$$

It follows that (v_n) is a Cauchy sequence, and given that C is $W-$closed, that it converges to a point $v \in C$ which is clearly a (and actually *the*) point of minimal distance. □

Now we turn to the main result of this section.

Theorem 5.8. *Let $C \subset \mathscr{P}_2(\mathbb{R}^d)$ be a $W-$closed strongly geodesically convex set. Then it is $\tau-$closed.*

Proof. Fix a measure μ not in C, let ν be the minimizing point found by the previous proposition and choose a plan $\boldsymbol{\gamma} \in \exp_\mu^{-1}(\nu)$. We will show that $C \subset \mathscr{H}^+_{\boldsymbol{\gamma}; D^2}$, where $D := W(\mu, C)$: since clearly μ is outside that halfspace, by proposition 5.4 and the arbitrariness of μ we conclude.

Choose any $\sigma \in C$, a plan $\boldsymbol{\sigma} \in \exp_\mu^{-1}(\sigma)$ and a 3-plan $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\sigma})$. Since $\mu_t := (\pi^0 + (1-t)\pi^1 + t\pi^2)_\# \boldsymbol{\alpha} \in C$, the minimality property of ν yields to

$$\begin{aligned}0 \leq \frac{d^+}{dt}\frac{1}{2}W^2(\mu_t, \mu)_{|t=0} \leq& \frac{d^+}{dt}\frac{1}{2}\|(1-t)x_1 + tx_2\|^2_{\boldsymbol{\alpha}}{}_{|t=0} \\ =& \langle x_2 - x_1, x_1\rangle_{\boldsymbol{\alpha}} = \langle x_2, x_1\rangle_{\boldsymbol{\alpha}} - \|\boldsymbol{\gamma}\|^2 \leq \langle \boldsymbol{\gamma}, \boldsymbol{\sigma}\rangle - D^2.\end{aligned}$$

By the arbitrariness of σ the thesis is achieved. □

3. Geodesically convex subsets of $\mathscr{P}_2^r(\mathbb{R}^d)$

In this section we drop strong assumption 5.2 on the $W-$closed set C, substituting it with the weaker one 5.1. Furthermore, we assume that every measure in C is regular, and we will prove that C is weakly closed. It is not clear to the author if the thesis remains true dropping the regularity assumption.

This case is much more complex than the one analyzed in the previous section: in order to clarify the ides behind the proof, we recall here the main steps.

step 1 We prove that without loss of generality we can assume to deal with geodesically convex sets which are bounded: indeed recalling the definition of the topology τ, we have that a subset C of $\mathscr{P}_2(\mathbb{R}^d)$ is $\tau-$closed if and only if so are its intersections with the closed ball and center in δ_0. Proving that those intersections are closed, geodesically convex and bounded we will get the claim.

step 2 This is the most important step of the proof: we will show that for each closed subset C of a complete geodesic metric space (E, d), every $a < 1$ and every point $x \notin C$ there exists a geodesic from C to x "which moves away from C with velocity bigger than a", that is, there exists a geodesic parameterized with arc length $t \to x_t$ such that $x_0 \in C$ and $x_{d(x_0,x)} = x$ satisfying

$$\frac{d(x_t, C)}{t} \geq a, \quad \forall t \in (0, d(x_0, x)].$$

The interest of this claim comes from the lack of any kind of compactness hypothesis. Observe that if C is compact, the claim follows easily by picking any point $y \in C$ of minimal distance from x and any geodesic from y to x. Such a geodesic would have the stated property even for $a = 1$.

step 3 We assume that every measure in our bounded, geodesically convex, closed set C is regular. Under this assumption we will show that the set C may be written as the intersection of a family of weakly closed halfspaces. More in detail: we will use the previous step to show that for each measure $\mu \notin C$ and every $a < 1$ we can find a measure μ^a such that $t \to \mu_t^a := (Id + t(T_{\mu_a}^{\mu} - Id))_{\#}\mu_a$ satisfies $W(\mu_t^a, C) \geq at\, W(\mu_a, \mu)$. Then the geodesic convexity (this is the point in which we use this hypothesis) implies, with some calculation, that it holds

$$\langle T_{\mu_a}^{\mu} - Id, T_{\mu_a}^{\nu} - Id\rangle \leq C_a, \quad \forall \nu \in C,$$

where $C_a := \sqrt{1-a^2} R\, W(\mu_a, \mu)$ and R is the diameter of C. Since $C_a \to 0$ as $a \to 1$, we can find some a such that the set C is contained in the "halfspace" given by

$$\mathscr{H}^-_{T_{\mu_a}^{\mu} - Id; C_a} := \left\{ \nu \ : \ \langle T_{\mu_a}^{\mu} - Id, T_{\mu_a}^{\nu} - Id\rangle_{\mu_a} \leq C_a \right\},$$

while the measure μ is not.

Showing that such a set is weakly closed we will get the conclusion.

Let us now enter into details.

PROPOSITION 5.9 (Step 1). *Let C be a geodesically convex subset of $\mathscr{P}_2(\mathbb{R}^d)$. Then letting $B_R := \{\mu : \|x\|_\mu \leq R\}$, we have that $C \cap B_R$ is geodesically convex for every $R \in \mathbb{R}^+$. Moreover C is $\tau-$closed if and only if $C \cap B_R$ is $\tau-$closed for every $R \in \mathbb{R}^+$, where $B_R := \{\mu : \|x\|_\mu \leq R\}$.*

Proof. We prove at first that $C \cap B_R$ is geodesically convex. Let $\mu_1, \mu_2 \in C \cap B_R$ and find an optimal plan $\boldsymbol{\gamma}$ such that $\mu_t := \boldsymbol{\gamma}[t] \in C$ for every $t \in [0,1]$. It holds

$$\|x\|^2_{\mu_t} = \|(1-t)x_1 + tx_2\|^2_{\boldsymbol{\gamma}} \le (1-t)\|x_1\|^2_{\boldsymbol{\gamma}} + t\|x_2\|^2_{\boldsymbol{\gamma}} = (1-t)\|x_1\|^2_{\mu_1} + t\|x\|^2_{\mu_2} \le R^2,$$

which is the first claim.

The other claim follows directly from the definition of the topology τ and from proposition 2.15. □

PROPOSITION 5.10 (Step 2). *Let (E,d) be a complete geodesic metric space, $C \subset E$ a closed set, $P \in E \setminus C$, and $0 \le a < 1$. Then there exists a point $Q \in C$ such that*

$$\frac{d(Q_t, C)}{t} \ge a, \quad \forall t \in (0, d(P,Q)], \tag{5.1}$$

where $Q_t : [0, d(P,Q)] \to E$ is any choice of a geodesic connecting Q to P ($Q_0 = Q$, $Q_1 = P$) parameterized by arc length.

Proof. Let we fix a notation: for any point $R \in C$ let $\mathbf{R}$ be the set of geodesics connecting R to P parameterized by arc length, and let $R_t \in \mathbf{R}$ be a generic element of this set.

We will say that a point $R \in C$ has the property $\mathcal{G}$ iff for every $R_t \in \mathbf{R}$ it holds

$$\frac{d(R_t, C)}{t} \ge a, \quad \forall t \in (0, d(P,R)].$$

Our aim is to prove that a point with the property $\mathcal{G}$ exists.

Start choosing any point $R \in C$ and suppose that it doesn't have the property $\mathcal{G}$. Then there exists a point $R' \in C$ such that

$$d(R_t, R') < at, \quad \text{for some } t > 0 \text{ and some } R_t \in \mathbf{R}. \tag{5.2}$$

From this we get

$$d(R, R') \le d(R, R_t) + d(R_t, R') < t(a+1),$$

and $d(R', P) \le d(R', R_t) + d(R_t, P) < at + d(R,P) - t$ from which it follows

$$d(R,P) - d(R',P) > t(1-a). \tag{5.3}$$

Putting together the last two inequalities we get the key estimate

$$d(R,R') < \frac{1+a}{1-a}\big(d(R,P) - d(R',P)\big). \tag{5.4}$$

This inequality is all we need to prove the thesis: we will proceed by transfinite induction by using its telescopic property.

Let Ω be the first uncountable ordinal. Define a function

$$\Omega \to C$$
$$\alpha \to R_\alpha,$$

beginning by choosing $R_0 \in C$ in any way. Then if α is the successor of some ordinal, we have two cases:

i) $R_{\alpha-1}$ has the property $\mathcal{G}$,
ii) $R_{\alpha-1}$ does not have the property $\mathcal{G}$.

In the first case we put $R_\alpha := R_{\alpha-1}$, in the second one we choose R_α among those points R' satisfying (5.2) with $R = R_{\alpha-1}$. Finally, if α is a limit ordinal we let R_α be the limit of $R_{\alpha'}$ with $\alpha' < \alpha$.

We have to prove this to be a good definition, we will do this by proving at the same tima that the following "extended" version of (5.4) holds:

$$d(R_\alpha, R_\beta) < \frac{1+a}{1-a}\left(d(R_\alpha, P) - d(R_\beta, P)\right), \quad \forall \alpha \leq \beta. \tag{5.5}$$

We prove this inequality by transfinite induction on β: it is true for 0, and it is easy to see that if it holds for β then it holds for $\beta+1$. Indeed by construction and from the first part of the proof, $R_{\beta+1}$ satisfies (5.4) with $R = R_\beta$, $R' = R_{\beta+1}$, therefore combining (5.4) and (5.5) we get

$$\begin{aligned} d(R_\alpha, R_{\beta+1}) &\leq d(R_\alpha, R_\beta) + d(R_\beta, R_{\beta+1}) \\ &< \frac{1+a}{1-a}\left(d(R_\alpha, P) - d(R_\beta, P) + d(R_\beta, P) - d(R_{\beta+1}, P)\right) \\ &= \frac{1+a}{1-a}\left(d(R_\alpha, P) - d(R_{\beta+1}, P)\right), \quad \forall \alpha \leq \beta. \end{aligned}$$

Given that the case $\alpha = \beta + 1$ is obvious, we get the claim.

Now let β be a limit ordinal, observe that we can't write inequality (5.5) for such a β, yet, since we have still to prove that R_β exists: we are going to prove at the same time that R_β is well defined and that for this point (5.5) holds. Since $\beta < \Omega$ there exists an increasing sequence (α_n) converging to β; for every α_n the inequality (5.5) holds, therefore we have

$$d(R_{\alpha_m}, R_{\alpha_n}) < \frac{1+a}{1-a}\left(d(R_{\alpha_m}, P) - d(R_{\alpha_n}, P)\right), \quad \forall m \leq n.$$

Being the sequence $d(R_{\alpha_n}, P)$ non increasing (by equation (5.3)) and bounded from below, it is a Cauchy sequence and the previous inequality shows that the same is true for the sequence R_{α_n}, which therefore converges to some point we call R_β. Since the previous argument applies to every increasing sequence α_n, showing that the corrispond points R_{α_n} form a Cauchy sequence, we get that R_β is well defined (i.e. it does not depend on the particular sequence (α_n) chosen), that the function $\alpha \to R_\alpha$ is continuous (with respect to the order topology) and that (5.5) holds for any $\beta < \Omega$.

Observe that from inequality (5.3) it follows that if $R_{\alpha+1} \neq R_\alpha$, then $d(R_{\alpha+1}, P)$ is strictly less than $d(R_\alpha, P)$. We are almost done: since there is no strictly decreasing function from Ω to $\mathbb{R}$, we have that the map $\alpha \to R_\alpha$ has to be eventually constant, therefore for some α we have $R_\alpha = R_{\alpha+1}$, which means by construction that the point $Q = R_\alpha$ satisfies the thesis. □

Observe that we proved a statement stronger than the one climed at the beginning of the section: indeed we proved not only that there exists a geodesic from some point C to P with the stated property, but also that there exists a point Q such that every geodesic connecting it to P has the property.

Note that this proposition is a generalization of the Drop Theorem of Daneš valid in Banach spaces, see [**21**] for further reference.

After the work on thesis was completed, the author discovered that both the statement of the proposition and its proof are closely related to the Ekeland-Bishop-Phelps principle. Actually a shorter proof may be given with a direct application of the EBP principle: we present here one found by B.Kirchheim. Use EBP principle to find $Q \in C$ which is a minimizer of

$$x \to f(x) := d(x,P) + \frac{1-a}{1+a} d(x,Q).$$

Then such a Q has the claimed property. Indeed, if this is not the case, there exists $R \in C$ and $0 \leq t \leq d(P,Q)$ such that $d(R,Q_t) < at$. For such R we have the following bounds

$$\begin{aligned} d(R,P) &\leq d(R,Q_t) + d(Q_t,P) < at + d(P,Q) - t, \\ d(R,Q) &\leq d(R,Q_t) + d(Q_t,Q) < at + t. \end{aligned}$$

Therefore it holds

$$f(Q) = d(P,Q) = d(P,Q) - t(1-a) + \frac{1-a}{1+a} t(a+1) > d(R,P) + \frac{1-a}{1+a} d(R,Q) = f(R),$$

which contradicts the minimality of Q.

Hilbert case 3. *Let C be a strongly closed, convex subset of an Hilbert space H. Then C is weakly closed.*

Proof. Since C is convex and weakly closed if and only if so are the sets $C \cap B_R$, we may assume without loss of generality that C is bounded. Let R be its diameter. Choose any point $y \notin C$: the claim will be achieved if we show that there exist a point x and a constant $c \in \mathbb{R}$ such that

$$\langle z - x, y - x \rangle \leq c < |y - x|^2, \quad \forall z \in C.$$

Indeed in this case the set C would be included in the halfspace $\{z : \langle z - x, y - x\rangle \leq c\}$ which is weakly closed and does not contain y. By the arbitrarieness of y this would give the thesis.

Let us prove our claim. Fix $a < 1$ and apply proposition 5.10 with $P = y$ to find a point $x^a \in C$ satisfying

$$\inf_{z \in C} |(1-t)x^a + ty - z| \geq at|y - x^a|. \tag{5.6}$$

Now fix $z \in C$ and let $v^a := y - x^a$, $w := z - x^a$. Observe that for small $t > 0$ it holds

$$x^a + t\sqrt{1-a^2}\frac{|v^a|}{|w|} w \in C,$$

therefore we know from (5.6) that

$$\left| tv^a - t\sqrt{1-a^2}\frac{|v^a|}{|w|} w \right|^2 \geq a^2 t^2 |v^a|^2.$$

Some manipulations show that the last equation implies

$$\langle v^a, w \rangle \leq |v^a||w|\sqrt{1-a^2} \leq R|v^a|\sqrt{1-a^2}.$$

By choosing a near to 1 and observing that $|v^a| \leq d(y,C)+R$ we get that the last term of the previous inequality is close to 0. Therefore it is smaller than $d^2(y,C)$ which in turn is smaller than $|y-x^a|^2$ and the claim is achieved. □

PROPOSITION 5.11 (Step 3). *Let C be a strongly closed, geodesically closed subset of $\mathscr{P}_2^r(\mathbb{R}^d)$. Then C is $\tau-$closed.*

Proof. Applying proposition 5.9 we may assume without loss of generality that C is bounded. Let R be its diameter. Choose any measure $\nu \notin C$: the claim will be achieved if we show that there exists a measure $\mu \in C$ and a constant $c \in \mathbb{R}$ such that

$$\langle T_\mu^\nu - Id, T_\mu^\sigma - Id\rangle \leq c < \|T_\mu^\nu - Id\|_\mu^2 = W^2(\mu,\nu), \quad \forall \sigma \in C.$$

Indeed in this case the set C would be included in the halfspace $\mathscr{H}^-_{T_\mu^\nu - Id;c}$ which is weakly closed by proposition 5.4 and does not contain ν. By the arbitrariness of ν we can conclude.

Let us prove our claim. Fix $a<1$ and apply proposition 5.10 with $P=\nu$ to find a measure μ^a satisfying

$$W(\mu_t^a, C) \geq atW(\nu, \mu_t^a),$$

where $\mu_t^a := (Id+t(T_{\mu^a}^\nu - Id))_\# \mu^a$. Now fix $\sigma \in C$ and define $v^a = T_\mu^\nu - Id$, $w = T_\mu^\sigma - Id$. Observe that for small $t>0$ it holds

$$\sigma_t := \left(Id + t\sqrt{1-a^2}\frac{\|v^a\|_{\mu^a}}{\|w\|_{\mu^a}}w\right)_\# \mu^a \in C,$$

therefore we know that

$$W^2(\mu_t^a, \sigma_t) \geq a^2 t^2 \|v^a\|_{\mu^a}^2.$$

Recalling that for every couple of tangent plans $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ (see (4.18)) it holds

$$\lim_{t\to 0^+} \frac{W(\exp(t\cdot\boldsymbol{\gamma}_1), \exp(t\cdot\boldsymbol{\gamma}_2))}{t} = W_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2),$$

and that $v^a, w \in \mathrm{Tan}_{\mu^a}(\mathscr{P}_2(\mathbb{R}^d))$ we obtain

$$\left\|v^a - \sqrt{1-a^2}\frac{\|v^a\|_{\mu^a}}{\|w\|_{\mu^a}}w\right\|_{\mu^a}^2 \geq a^2\|v^a\|_{\mu^a}^2.$$

Some algebraic manipulations show that the previous inequality implies

$$\langle v^a, \sqrt{1-a^2}\frac{\|v^a\|_{\mu^a}}{\|w\|_{\mu^a}}w\rangle_{\mu^a} \leq \|v^a\|_{\mu^a}\|w\|_{\mu^a}\sqrt{1-a^2} \leq R\|v^a\|_{\mu^a}\sqrt{1-a^2}.$$

By choosing a near to 1 and observing that $\|v^a\|_{\mu^a} \leq W(\nu, C)+R$ we get that the last term of the previous inequality is close to 0. Therefore it is smaller than $d^2(\nu, C)$, which in turn is smaller then $W^2(\nu,\mu^a)$ and the claim is achieved. □

4. Subdifferentials and differentiability properties of W^2

We conclude this chapter with a short discussion on the right definition of subdifferential of geodesically convex functional (with the hope to clarify the one introduced in [**11**] in light of the notation introduced here) and to complete the analysis of the differentiability properties of the function $W^2(\cdot,\sigma)$. Although the results proven here are not needed in the sequel, they are interesting from a theoretical point of view, and provide an application of the theory developed up to here.

Definition 5.12 (geodesically convex functionals). *A functional $F:\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}\cup\{+\infty\}$ is geodesically convex if for every constant speed geodesic $t\to\mu_t$ parameterised in some interval, the function*

$$t\to F(\mu_t),$$

is convex.

Obviously the definition would be the same if we let the geodesics be parameterised in $[0,1]$. Moreover, it is a simple consequence of theorem 2.8 that F is geodesically convex if and only if for any couple of measures μ,ν there exists a constant speed geodesic $t\to\mu_t$ joining μ to ν along which F is convex.

In the following we will always deal with *p*roper functionals, i.e. functionals which are not constantly equal to $+\infty$.

A simple consequence of the theorem proved in the previous section is the following one.

Theorem 5.13. *Let F be a geodesically convex functional, lower semicontinuous w.r.t. W. Assume also that F satisfies one of the two conditions below:*

- *$F(\mu_t)$ is convex along curves $\mu_t=[\boldsymbol{\gamma}](t)$ even for non optimal $\boldsymbol{\gamma}$;*
- *$F(\mu)=+\infty$ for any measure $\mu\notin\mathscr{P}_2^r(\mathbb{R}^d)$.*

Then F is lower semicontinuous w.r.t. the weak topology τ.

Proof. In either case the sublevels of F are weakly closed because of what we proved in the last two sections. □

We will denote by $D(F)$ the domain of F, that is the set of those measure $\mu\in\mathscr{P}_2(\mathbb{R}^d)$ such that $F(\mu)<+\infty$.

Definition 5.14 (Subdifferential, superdifferential and differential). *Let $F:\mathscr{P}_2(\mathbb{R}^d)\to(-\infty,+\infty]$ be a functional and let $\mu\in D(F)$. We say that a plan $\boldsymbol{\gamma}\in\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ belongs to the subdifferential $\partial^-F(\mu)$ of F at μ if for every measure ν and every $\boldsymbol{\eta}\in\exp_\mu^{-1}(\nu)$ it holds*

$$F(\nu)\geq F(\mu)-\langle -1\cdot\boldsymbol{\gamma},\boldsymbol{\eta}\rangle+o(W(\nu,\mu)). \tag{5.7}$$

Analogously, we say that $\boldsymbol{\gamma}\in\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ belongs to the superdifferential $\partial^+F(\mu)$ of F at μ if for every measure ν and every $\boldsymbol{\eta}\in\exp_\mu^{-1}(\nu)$ it holds

$$F(\nu)\leq F(\mu)+\langle\boldsymbol{\gamma},\boldsymbol{\eta}\rangle+o(W(\nu,\mu)). \tag{5.8}$$

A plan $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *is said to be the differential* $\partial F(\mu)$ *of* F *at* μ *if it is both a subdifferential and a superdifferential (it is easy to see that a differential, whenever it exists, is unique).*

We will denote by $D(\partial^\pm F)$ *the domain superdifferential and subdifferential respectively, i.e. the set of those measures* μ *such that the set* $\partial^\pm F(\mu)$ *is not empty.*

Observe that for regular measures μ the previous definitions reduce to:

$$v \in \partial^- F(\mu) \Leftrightarrow F(\nu) \geq F(\mu) + \langle v, T_\mu^\nu - Id\rangle_\mu + o(W(\nu,\mu)),\ \forall \nu \in \mathscr{P}_2(\mathbb{R}^d)$$
$$v \in \partial^+ F(\mu) \Leftrightarrow F(\nu) \leq F(\mu) + \langle v, T_\mu^\nu - Id\rangle_\mu + o(W(\nu,\mu)),\ \forall \nu \in \mathscr{P}_2(\mathbb{R}^d).$$

A natural question arises: why the term $-\langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle$ instead of $\langle \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle$ in the definition of sub-differential? The answer is that the definition we have chosen is the right one to ensure the strong-weak closure of subdifferentials of geodesically convex functions: this will become clear with the proof of theorem 5.18 and example 5.19. Here we just observe that by the analogy with the Euclidean case a priori one assumes that for regular measures the linear term has to be similar to $\langle v, T_\mu^\nu - Id\rangle$ (which is actually the case) while there is no reason a priori to guess that when dealing with non-regular measures μ, the correct generalization is $\langle \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle$ instead of some other member of the set $\langle x_1, x_2\rangle_{\boldsymbol{\alpha}}$, when $\boldsymbol{\alpha}$ varies in the union of the sets $\mathscr{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ as $\boldsymbol{\eta}$ varies in $\exp_\mu^{-1}(\nu)$. If $\exp_\mu^{-1}(\nu)$ contains only one element $\boldsymbol{\eta}$, one may just hope for the right term to be either

$$\max_{\boldsymbol{\alpha}} \langle x_1, x_2\rangle_{\boldsymbol{\alpha}}, \quad \boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}) \text{ or,}$$
$$\min_{\boldsymbol{\alpha}} \langle x_1, x_2\rangle_{\boldsymbol{\alpha}}, \quad \boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}),$$

which lead respectively to (see also remark 4.28)

$$\langle \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle,$$
$$-\langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle.$$

Choosing a priori one definition instead of the other one is just a matter of taste, but, as we said before, the second one leads to the properties of the subdifferential we would like to have.

Observe that by virtue of our definition in order to prove that equation (5.7) is satisfied one has just to prove that equation

$$F(\nu) \geq F(\mu) + \langle x_1, x_2\rangle_{\boldsymbol{\alpha}} + o(W(\nu,\mu)),$$

holds for some $\boldsymbol{\alpha} \in \mathscr{ADM}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ with $\boldsymbol{\eta} \in \exp_\mu^{-1}(\nu)$.

In case we are dealing with geodesically convex lower semicontinuous functionals, the definition of subdifferential simplies a bit, as showed by the following proposition.

Proposition 5.15. *Let* F *be a g.c. and l.s.c. functional and let* $\mu \in D(F)$ *be a measure in its domain. Then* $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ *belongs to* $\partial^- F(\mu)$ *if and only if for every* $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ *there exists* $\boldsymbol{\eta} \in \exp_\mu^{-1}(\nu)$ *such that*

$$F(\nu) \geq F(\mu) - \langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta}\rangle. \tag{5.9}$$

Proof. We have to prove that we can drop the term $o(W(\mu,\nu))$ and that it is sufficient the existence of one element in $\exp_\mu^{-1}(\nu)$ satisfying (5.9), instead than asking the inequality to hold for any plan in $\exp_\mu^{-1}(\nu)$.

Let us first check the second property: take any $\boldsymbol{\eta} \in \exp_\mu^{-1}(\nu)$ and consider the curve $t \to \nu_t := \exp(t\cdot\boldsymbol{\eta})$ for $0 \le t \le 1$. We know that for $t < 1$ there exists only one element $\boldsymbol{\eta}_t$ in $\exp_\mu^{-1}(\nu_t)$, namely $(\pi^0, \pi^0 + t\pi^1)_{\#}\boldsymbol{\eta}$, therefore it has to hold

$$(1-t)F(\mu) + tF(\nu) \ge F(\nu_t) \ge F(\mu) - \langle -1\cdot\boldsymbol{\gamma}, \boldsymbol{\eta}_t\rangle + o(W(\nu_t,\mu)),$$

which leads to the claim by letting t go to 1.

For the other part of the proof, argue by contradiction, assuming that

$$F(\nu) \le F(\mu) - \langle -1\cdot\boldsymbol{\gamma}, \boldsymbol{\eta}\rangle - \varepsilon,$$

for some $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, $\boldsymbol{\eta} \in \exp_\mu^{-1}(\nu)$ and $\varepsilon > 0$. Then, defining $\nu_t := \exp_\mu(t\cdot\boldsymbol{\eta})$ we get that

$$F(\nu_t) \le (1-t)F(\mu) + tF(\nu) \le F(\mu) - \langle -1\cdot\boldsymbol{\gamma}, t\cdot\boldsymbol{\eta}\rangle - t\varepsilon,$$

which contradicts equation (5.7) by letting t go to 0. □

The definition of geodesically convex sets given before for subsets of $\mathscr{P}_2(\mathbb{R}^d)$ has an analogous for subsets of $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$: a set $K \subset \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is *geodesically convex* if for every couple of plans $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in K$ there exists a geodesic connecting them entirely contained in K. In other words, for any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in K$ there has to exist an $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ such that $\boldsymbol{\alpha}[t] \in K$ for any $0 \le t \le 1$.

PROPOSITION 5.16. *Let F be a functional on $\mathscr{P}_2(\mathbb{R}^d)$. Then for every $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ the set $\partial^- F(\mu)$ is $\tau-$sequentially closed and geodesically convex.*

Proof. We can assume that $\partial^- F(\mu)$ is not empty. To get the convexity apply proposition 4.27 to get that

$$\langle((1-t)\cdot\boldsymbol{\gamma}_1)\oplus(t\cdot\boldsymbol{\gamma}_2), \boldsymbol{\eta}\rangle \le \langle(1-t)\cdot\boldsymbol{\gamma}_1, \boldsymbol{\eta}\rangle + \langle t\cdot\boldsymbol{\gamma}_2, \boldsymbol{\eta}\rangle. \tag{5.10}$$

To prove the closure take any sequence $\boldsymbol{\gamma}_n \in \partial^- F(\mu)$ $\tau-$converging to some $\boldsymbol{\gamma}$, fix a plan $\boldsymbol{\eta} \in \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and consider any sequence $(\boldsymbol{\alpha}_n) \subset \mathcal{OPT}_\mu(-1\cdot\boldsymbol{\gamma}_n, \boldsymbol{\eta})$. Then, possibly extracting a subsequence, we can assume that $(\boldsymbol{\alpha}_n)$ $\tau-$converges to some $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(-1\cdot\boldsymbol{\gamma}, \boldsymbol{\eta})$. Then it holds

$$\lim_{n\to\infty}\langle -1\cdot\boldsymbol{\gamma}_n, \boldsymbol{\eta}\rangle = \lim_{n\to\infty}\langle x_1, x_2\rangle_{\boldsymbol{\alpha}_n} = \langle x_1, x_2\rangle_{\boldsymbol{\alpha}} \le \langle -1\cdot\boldsymbol{\gamma}, \boldsymbol{\eta}\rangle.$$

□

Observe that we proved a stronger kind of convexity for the set $\partial^- F(\mu)$: indeed equation (5.10) shows that for any $\boldsymbol{\alpha} \in \mathcal{ADM}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ (and not just for those in $\mathcal{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$) the whole segment $\boldsymbol{\alpha}[t]$ belongs to $\partial^- F(\mu)$. The weak closure of subdifferential will be more deeply analyzed in theorem 5.18 below.

COROLLARY 5.17. *Let F be a functional and $\mu \in D(\partial^- F)$. Then there exists a unique element in $\partial^- F(\mu)$ of minimal norm.*

Proof. The existence is a simple consequence of the $\tau-$compacteness of bounded sequences in $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and of the previous proposition. The uniqueness follows from the *strict* convexity of the norm on $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$: suppose that there exist two plans $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ of minimal norm and take any $\boldsymbol{\alpha} \in \mathscr{OPT}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. Then the previous proposition ensures that for any $0 < t < 1$ the plan $\boldsymbol{\alpha}[t]$ belongs to $\partial^- F(\mu)$. For such plans it holds

$$\|\boldsymbol{\alpha}[t]\|^2 = \|(1-t)x_1 + tx_2\|^2_{\boldsymbol{\alpha}} < (1-t)\|x_1\|^2_{\boldsymbol{\alpha}} + t\|x_2\|^2_{\boldsymbol{\alpha}} = (1-t)\|\boldsymbol{\gamma}_1\|^2 + t\|\boldsymbol{\gamma}_2\|^2,$$

contradicting the minimality. □

The following theorem shows that even our definition of subdifferential, reproduces the classical strong-weak closure property. A counterexample follows, showing that the other alternative doesn't work as well.

THEOREM 5.18 (Strong-weak closure of subdifferential). *Let F be a g.c. and l.s.c. functional, $\mu_n, \mu \in \mathscr{P}_2(\mathbb{R}^d)$, $n \in \mathbb{N}$, be given measures such that $W(\mu_n, \mu) \to 0$ as $n \to \infty$ and let $\boldsymbol{\gamma}_n \in \partial^- F(\mu_n)$ be τ converging to some $\boldsymbol{\gamma}$. Then $\boldsymbol{\gamma} \in \partial^- F(\mu)$.*

Proof. Obviously $\pi^1_\# \boldsymbol{\gamma} = \mu$. Fix $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and observe that the following inequality holds for every n:

$$F(\nu) \geq F(\mu_n) - \langle -1 \cdot \boldsymbol{\gamma}_n, \boldsymbol{\eta}_n \rangle,$$

where $\boldsymbol{\eta}_n$ is a certain plan in $\exp^{-1}_{\mu_n}(\nu)$. Since F is lower semicontinuous, to get the thesis we have just to prove that

$$\overline{\lim_n} \langle -1 \cdot \boldsymbol{\gamma}_n, \boldsymbol{\eta}_n \rangle \leq \langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle,$$

where $\boldsymbol{\eta}$ is some element of $\exp^{-1}_\mu(\nu)$. Possibly extracting a subsequence we can assume that $(\boldsymbol{\eta}_n)$ $W-$converge w.r.t. W to some $\boldsymbol{\eta}$ such that $\pi^1_\# \boldsymbol{\eta} = \mu$, $(\pi^1 + \pi^2)_\# \boldsymbol{\eta} = \nu$, moreover since the plans $(\pi^1, \pi^1 + \pi^2)_\# \boldsymbol{\eta}_n$ are optimal, the same is true for $\boldsymbol{\eta}$, which therefore belongs to $\exp^{-1}_\mu(\nu)$. Now choose $\boldsymbol{\alpha}_n \in \mathscr{OPT}_\mu(-1 \cdot \boldsymbol{\gamma}_n, \boldsymbol{\eta}_n)$ and take a subsequence (not relabeled) which τ converges to some $\boldsymbol{\alpha} \in \mathscr{ADM}_\mu(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta})$. Since $\pi^{1,3}_\# \boldsymbol{\alpha}_n = \boldsymbol{\eta}_n$ converge w.r.t. W to $\pi^{1,3}_\# \boldsymbol{\alpha} = \boldsymbol{\eta}$ we can apply corollary 2.29 to get

$$\lim_{n \to \infty} \langle -1 \cdot \boldsymbol{\gamma}_n, \boldsymbol{\eta}_n \rangle = \lim_{n \to \infty} \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}_n} = \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}} \leq \langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle.$$

□

The proof of the previous theorem suggests why the same technique doesn't work with the other possible definition of subdifferential: with the same notation used in the proof, if one tries to prove that $\underline{\lim} \langle \boldsymbol{\gamma}_n, \boldsymbol{\eta}_n \rangle \geq \langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle$ by taking $\boldsymbol{\alpha}_n \in \mathscr{OPT}_\mu(\boldsymbol{\gamma}_n, \boldsymbol{\eta}_n)$ and passing to the limit, one still has that $\langle x_1, x_2 \rangle_{\boldsymbol{\alpha}_n} = \langle x_1, x_2 \rangle_{\boldsymbol{\alpha}}$, but nothing ensures that the limit 3-plan $\boldsymbol{\alpha}$ is optimal for the couple $\boldsymbol{\gamma}, \boldsymbol{\eta}$. The following example shows that there is actually no hope to find a proof of the previous theorem with the other definition of subdifferential.

EXAMPLE 5.19. *For $t \geq 0$ let $\mu_t := (\delta_t + \delta_{-t})/2 \in \mathscr{P}_2(\mathbb{R})$ and define the functional F as*

$$F(\mu) := \begin{cases} -t, & \text{if } \mu = \mu_t \text{ for some } 0 \leq t \leq 1, \\ +\infty, & \text{elsewhere.} \end{cases}$$

It is not difficult to see that F is a g.c. and l.s.c. functional and that for $0 < t \leq 1$ $\boldsymbol{\gamma}_t := (\delta_{(t,-1)} + \delta_{(-t,1)})/2$ belongs to $\partial^- F(\mu_t)$ (actually it is the unique element): indeed for fixed $t, T > 0$ let $\boldsymbol{\eta}_t^T := (\delta_{(t,T-t)} + \delta_{(-t,-T+t)})/2$ be the unique element in $\exp_{\mu_t}^{-1}(\mu_T)$. Then for every $0 < t, T \leq 1$ it holds $\langle \boldsymbol{\gamma}_t, \boldsymbol{\eta}_t^T \rangle = t - T$ and therefore the inequality

$$F(\mu_T) \geq F(\mu_t) + \langle \boldsymbol{\gamma}_t, \boldsymbol{\eta}_t^T \rangle$$

holds (and it is an equality). But letting t go to 0, clearly $\boldsymbol{\gamma}_t$ converges to $\boldsymbol{\gamma} := (\delta_{(0,1)} + \delta_{(0,-1)})/2$ and we have that $\langle \boldsymbol{\gamma}, \boldsymbol{\eta}_0^T \rangle = T$, which is strictly bigger than the limit of $\langle \boldsymbol{\gamma}_t, \boldsymbol{\eta}_t^T \rangle$, and the previous equation does not hold:

$$F(\mu_T) = -T \ngeq T = F(\mu_0) + \langle \boldsymbol{\gamma}_0, \boldsymbol{\eta}_0^T \rangle.$$

Now we turn to the study of differential properties of W^2. Observe that by inequality (2.16) we know that $W^2(\cdot, \sigma)$ satisfies a *concavity* inequality (it is -1 geodesically concave, in a sense which can be specified), therefore we should expect to find superdifferentiability properties rather than subdifferentibility ones. The following theorem shows that actually the superdifferential is always not empty.

THEOREM 5.20. *Let $\mu, \sigma \in \mathscr{P}_2(\mathbb{R}^d)$ be two measures and fix $\boldsymbol{\gamma} \in \exp_\mu^{-1}(\sigma)$. Then $-2 \cdot \boldsymbol{\gamma}$ belongs to the superdifferential of the functional $\nu \to W^2(\nu, \sigma)$ at μ.*

Proof. Choose $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\eta} \in \exp_\mu^{-1}(\nu)$ and $\boldsymbol{\alpha} \in \mathcal{OPT}_\mu(\boldsymbol{\eta}, \boldsymbol{\gamma})$. Then $(\pi^0 + \pi^1, \pi^0 + \pi^2)_\# \boldsymbol{\alpha} \in \mathcal{Adm}(\nu, \sigma)$ and therefore it holds

$$\begin{aligned} W^2(\nu, \sigma) \leq \|x_1 - x_2\|_{\boldsymbol{\alpha}}^2 =& \|x_1\|_{\boldsymbol{\alpha}}^2 + \|x_2\|_{\boldsymbol{\alpha}}^2 - 2\langle x_1, x_2 \rangle_{\boldsymbol{\alpha}} \\ =& W^2(\mu, \nu) + W^2(\mu, \sigma) - \langle 2 \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle, \end{aligned}$$

from which the thesis follows. □

Once it is proved that the superdifferential is always not empty the natural question that arises is: what is the element of minimal norm? The non trivial answer is given in the following theorem.

THEOREM 5.21. *Let $\mu, \sigma \in \mathscr{P}_2(\mathbb{R}^d)$ be two measures. Then the element of minimal norm of the superdifferential of $W^2(\cdot, \sigma)$ at μ is $-2\mathscr{B}(\boldsymbol{\gamma}_0)$, where $\boldsymbol{\gamma}_0$ is the unique minimiser of $\|\mathscr{B}(\boldsymbol{\gamma})\|_\mu$ among all $\boldsymbol{\gamma}$ in $\exp_\mu^{-1}(\sigma)$.*

Proof. Let us first prove that there is only one minimiser of $\|\mathscr{B}(\boldsymbol{\gamma})\|_\mu$ in $\exp_\mu^{-1}(\sigma)$. The existence follows by $\tau-$compacteness of the set and lower semicontinuity of the norm of the barycentric projection. For uniqueness, observe that $\exp_\mu^{-1}(\sigma)$ is stable under (usual) convex combination of plans, and that $t \to \|\mathscr{B}((1-t)\boldsymbol{\gamma}_1 + t\boldsymbol{\gamma}_2)\|_\mu = \|(1-t)\mathscr{B}(\boldsymbol{\gamma}_1) + t\mathscr{B}(\boldsymbol{\gamma}_2)\|_\mu$ is strictly convex.

To prove the thesis we will first show that $-2\mathscr{B}(\boldsymbol{\gamma})$ is a superdifferential for every $\boldsymbol{\gamma} \in \exp_\mu^{-1}(\sigma)$, and then we prove that for every superdifferential $\boldsymbol{\gamma}$ it holds $P_\mu(\mathscr{B}(\boldsymbol{\gamma})) = -2\mathscr{B}(\boldsymbol{\eta})$ for some $\boldsymbol{\eta} \in \exp_\mu^{-1}(\sigma)$. Given that it holds $\|P_\mu(\mathscr{B}(\boldsymbol{\gamma}))\|_\mu \leq \|\mathscr{B}(\boldsymbol{\gamma})\|_\mu \leq \|\boldsymbol{\gamma}\|$ this will prove the thesis.

To prove the first claim choose a plan $\boldsymbol{\gamma} \in \exp_\mu^{-1}(\sigma)$, a measure ν, a plan $\boldsymbol{\eta} \in \exp_\mu^{-1}(\nu)$ and define the 3-plan $\boldsymbol{\alpha}$ to be the plan whose disintegration w.r.t. μ satisfies

$\boldsymbol{\alpha}_{x_0} = \boldsymbol{\gamma}_{x_0} \times \boldsymbol{\eta}_{x_0}$. Observe that $\boldsymbol{\alpha}$ satisfies $(\pi^0 + \pi^1, \pi^0 + \pi^2)_{\#}\boldsymbol{\alpha} \in \mathcal{A}dm(\sigma, \nu)$ and $\langle x_1, x_2 \rangle_{\boldsymbol{\alpha}} = \langle \mathfrak{I}(\mathscr{B}(\boldsymbol{\gamma})), \boldsymbol{\eta} \rangle$, therefore it holds

$$W^2(\nu, \sigma) \leq \|x_1 - x_2\|^2_{\boldsymbol{\alpha}} = W^2(\mu, \sigma) + W^2(\mu, \nu) - 2\langle \mathfrak{I}(\mathscr{B}(\boldsymbol{\gamma})), \boldsymbol{\eta} \rangle,$$

and the claim is proved.

To prove the second claim let us introduce the set

$$K := \left\{ -2\mathscr{B}(\boldsymbol{\eta}) \ : \ \boldsymbol{\eta} \in \exp_\mu^{-1}(\sigma) \right\},$$

and observe that K is closed, convex and included in $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ (because of theorem 4.15). Therefore, in order to prove that $P_\mu(\mathscr{B}(\boldsymbol{\gamma})) \in K$ for every superdifferential $\boldsymbol{\gamma}$ in μ, by the Hanh-Banach theorem applied to in $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ it is sufficient to show that

$$\langle \mathscr{B}(\boldsymbol{\gamma}), v \rangle_\mu \geq \min_{f \in K} \langle f, v \rangle_\mu, \quad \forall v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)). \tag{5.11}$$

By density it is sufficient to show that the previous inequality holds for every v such that $(Id, Id + Tv)_{\#}\mu$ is the unique optimal plan between its marginals for a certain $T > 0$. Let us assume this is the case, define $\mu_t := (Id + tv)_{\#}\mu$ and use proposition 4.10 to compute

$$\begin{aligned} \frac{d}{dt} W^2(\mu_t, \sigma)_{|t=0} &= -2 \sup_{\boldsymbol{\eta} \in \exp^-1_\mu(\sigma)} \langle \mathfrak{I}(v), \boldsymbol{\eta} \rangle \\ &= -2 \sup_{\boldsymbol{\eta} \in \exp^-1_\mu(\sigma)} \langle v, \mathscr{B}(\boldsymbol{\eta}) \rangle_\mu = \inf_{f \in K} \langle v, f \rangle_\mu. \end{aligned} \tag{5.12}$$

The same derivative can be evaluated using the fact that $\boldsymbol{\gamma}$ is a superdifferential, which gives the inequality (recall that $(Id, Id + tv)_{\#}\mu$ is the unique optimal plan between μ and μ_t for $t \leq T$):

$$W^2(\mu_t, \sigma) \leq W^2(\mu, \sigma) - \langle -1 \cdot \boldsymbol{\gamma}, \mathfrak{I}(tv) \rangle + o(t) = W^2(\mu, \sigma) + t\langle \mathscr{B}(\boldsymbol{\gamma}), v \rangle_\mu + o(t).$$

This leads to

$$\frac{d}{dt} W^2(\mu_t, \sigma)_{|t=0} \leq \langle \mathscr{B}(\boldsymbol{\gamma}), v \rangle_\mu. \tag{5.13}$$

Equations (5.12) and (5.13) give (5.11), and therefore the thesis. □

The last two theorems allow the characterization of differentiability points of $W^2(\cdot, \sigma)$.

Corollary 5.22. *The functional $W^2(\cdot, \sigma)$ is differentiable in μ if and only if the set $\exp_\mu^{-1}(\sigma)$ contains only one element, and this element is induced by a map.*

Proof. It is clear that if the superdifferential contains more than one element, then the point can't be a point of differentiability. Therefore, given that by the previous theorems we know that both $-2 \cdot \boldsymbol{\gamma}$ and $-2\mathscr{B}(\boldsymbol{\gamma})$ belong to the superdifferential at μ for every $\boldsymbol{\gamma} \in \exp_\mu^{-1}(\sigma)$, the points which do not satisfy the assumption of the statement are not point of differentiability.

To conclude, fix a point μ such that $\exp_\mu^{-1}(\sigma)$ contains only an element $\boldsymbol{\gamma}$ induced by a map, choose a sequence of points $\nu_n \in \mathscr{P}_2(\mathbb{R}^d)$, plans $\boldsymbol{\eta}_n \in \exp_\mu^{-1}(\nu_n)$ and 3-plans $\boldsymbol{\alpha}_n$ such that $(\pi^0, \pi^1 - \pi^0)_\# \boldsymbol{\alpha}_n = \boldsymbol{\eta}_n$ and $(\pi^1, \pi^2 - \pi^1)_\# \boldsymbol{\alpha}_n \in \exp_{\nu_n}^{-1}(\sigma)$. We have

$$\begin{aligned} W^2(\nu_n, \sigma) - W^2(\mu, \sigma) \geq & \|x_2 - x_1\|^2_{\boldsymbol{\alpha}_n} - \|x_2 - x_0\|^2_{\boldsymbol{\alpha}_n} \\ = & \|x_1 - x_0\|^2_{\boldsymbol{\alpha}_n} - 2\langle x_1 - x_0, x_2 - x_0 \rangle_{\boldsymbol{\alpha}_n} \\ \geq & W^2(\mu, \nu_n) - 2\langle \boldsymbol{\eta}_n, \boldsymbol{\gamma}_n \rangle, \end{aligned}$$

where $\boldsymbol{\gamma}_n := (\pi^0, \pi^2 - \pi^0)_\# \boldsymbol{\alpha}_n$. Given that $(\pi^0, \pi^0 + \pi^1)_\# \boldsymbol{\gamma}_n \in \mathcal{Adm}(\mu, \sigma)$ and that $\|\boldsymbol{\gamma}_n\| \leq \|x_1 - x_0\|_{\boldsymbol{\alpha}_n} + \|x_2 - x_1\|_{\boldsymbol{\alpha}_n} = W(\mu, \nu_n) + W(\nu_n, \sigma) \to W(\mu, \sigma)$ we have that $\boldsymbol{\gamma}_n$ converges to $\boldsymbol{\gamma}$ w.r.t. W (because of the uniqueness of $\boldsymbol{\gamma}$). By proposition 4.7 and the fact that $\boldsymbol{\gamma}$ is induced by a map, this implies that $\boldsymbol{\gamma}_n$ converges to $\boldsymbol{\gamma}$ w.r.t. W_μ, therefore the previous estimate gives

$$W^2(\nu_n, \sigma) - W^2(\mu, \sigma) \geq W^2(\mu, \nu_n) - 2\langle \boldsymbol{\eta}_n, \boldsymbol{\gamma} \rangle + W(\mu, \nu_n) o(1)$$

and the proof is achieved. $\square$

CHAPTER 6

An approach to second order analysis

In the previous chapters we developed a first order theory for the space $(\mathscr{P}_2(\mathbb{R}^d), W)$, showing that it admits a tangent space endowed with an inner product which makes $\mathscr{P}_2(\mathbb{R}^d)$ formally very similar to a Riemannian manifold. In order to go deeper in the study of this space some natural questions arise: is there a *Levi-Civita* connection for this Riemannian structure? Is there a parallel transport for a tangent vector along a curve? The aim of this chapter is to provide a first step in this direction and to show that under certain regularity assumptions both questions admit an affirmative answer. Part of the calculation made in this Chapter (more explicitely, part of those of section 5) already appeared in [**43**], where the equation of the covariant derivative and the expression for the curvature operator were computed at a formal level. The results proven here can be found in [**10**].

1. The case of a manifold embedded in $\mathbb{R}^d$

In this section we recall the Riemannian concepts we are going to study in the Wasserstein setting. An important ingredient is a proof of existence of the parallel transport that, as we will see later on, can be imitated to prove existence of the parallel transport along a dense class of curves in $\mathscr{P}_2(\mathbb{R}^d)$. For a deeper analysis of all the topics discussed in this section, apart the proof of existence of parallel transport we present, see [**33**].

Throughout this section M will be a C^∞ manifold embedded in $\mathbb{R}^d$ endowed with the induced Riemannian structure, $\gamma(t) : [0,1] \to M$ a fixed C^∞ curve and $v(t) = \dot\gamma(t) \in T_{\gamma(t)}M$, $t \in [0,1]$, the velocity vector of $\gamma(t)$. We will think to the tangent space $V_t := T_{\gamma(t)}M$ at the point $\gamma(t)$ as a linear subspace of $\mathbb{R}^d$ (i.e. we *translate* it to let the origin be included) and we let $P_t : \mathbb{R}^d \to V_t$ be the orthogonal projection of $\mathbb{R}^d$ onto V_t.

Recall that a connection on a M is a map $(v,u) \to D_v u$ on the set of smooth vector fields into itself, additive on each component and satisfying:

$$D_{fv}u = fD_v u, \tag{6.1a}$$

$$D_v(fu) = fD_v u + \langle \nabla f, v\rangle u, \tag{6.1b}$$

for any $f \in C^\infty(M)$.

Equation (6.1a) implies that for a given u, the value of $D_v u$ at a point p depends only on the value of v at p, for this reason we will sometime write $D_{v(p)}u$ for $D_v u(p)$. Equation (6.1b) gives that for a given v, $D_v u(p)$ depends only on the value of u along an integral curve $\gamma(t)$ of v starting at p, i.e. a curve γ on a certain right-neighborhood of 0 such that $\gamma(0) = p$, $\gamma'(t) = v(\gamma(t))$.

In general there is more than one connection, however if the manifold is Riemannian, as in our case, only one is *compatible with the metric* and *torsion free*, i.e. only one satisfies the following two conditions:

$$\frac{d}{dt}\Big(\langle u^1(\gamma(t)), u^2(\gamma(t))\rangle\Big) = \langle \nabla_{v(t)} u^1(\gamma(t)), u^2(\gamma(t))\rangle + \langle u^1(\gamma(t)), \nabla_{v(t)} u^2(\gamma(t))\rangle, \tag{6.2a}$$

$$\nabla_{u^1} u^2 - \nabla_{u^2} u^1 = [u^1, u^2]. \tag{6.2b}$$

where u^1, u^2 are tangent vector fields, and in (6.2a) $u^1(t), u^2(t)$ are their value in $\gamma(t)$, respectively. The fact that there is at most one connection $\nabla_v(u)$ for which the previous equations are satisfied is a consequence of the Koszul formula:

$$2\langle \nabla_v(u), w\rangle = v(\langle u, w\rangle) + u(\langle v, w\rangle) - w(\langle u, v\rangle) + \langle [u, v], w\rangle - \langle [u, w], v\rangle - \langle [v, w], u\rangle,$$

valid for any vector fields u, v, w defined on the whole M, and any connection $\nabla_v(u)$ satisfying equations (6.2) (in this formula we identified a tangent vector field with the derivation it induces). Given that the formula expresses the covariant derivative in terms of the Riemannian metric only, the uniqueness follows.

The unique connection satisfying (6.2) is called Levi-Civita connection. In our setting the Levi-Civita connection along a vector $v \in T_pM$ tangent to M at the point p is given by the following formula:

$$\nabla_v u(p) := P_p\left(\frac{d}{dt} u(\gamma(t))\right), \tag{6.3}$$

where u is a tangent vector field, γ is a smooth curve such that $\gamma(0) = p$, $\dot\gamma(0) = v$ and P_p is the orthogonal projection onto T_pM.

To prove that the connection we defined is the Levi-Civita one, observe that it holds

$$\frac{d}{dt}\langle u^1_t, u^2_t\rangle = \langle \frac{d}{dt} u^1_t, u^2_t\rangle + \langle u^1_t, \frac{d}{dt} u^2_t\rangle = \langle \nabla_{v(t)} u^1(t), u^2(t)\rangle + \langle u^1(t), \nabla_{v(t)} u^2(t)\rangle,$$

for any u^1_t, u^2_t tangent in $\gamma(t)$, and

$$\begin{aligned}
[u^1, u^2]F =& u^1(u^2(F)) - u^2(u^1(F)) = u^1(\langle \nabla F, u^2\rangle) - u^2(\langle \nabla F, u^1\rangle) \\
=& \langle \nabla^2 F \cdot u^1, u^2\rangle + \langle \nabla F, \nabla_{u^1} u^2\rangle - \langle \nabla^2 F \cdot u^2, u^1\rangle - \langle \nabla F, \nabla_{u^2} u^1\rangle \\
=& (\nabla_{u^1} u^2 - \nabla_{u^2} u^1)F,
\end{aligned}$$

for any two vector fields u^1, u^2 and any smooth functional F on M (in the calculations we did, we implicitly assumed F to be extended to a neighborhood of M in such a way to leave $\nabla F(p) \in T_pM$).

The vector field $t \to u_t$ is the parallel transport of the vector u_0 along γ if and only if

$$P_t\left(\frac{d}{dt} u_t\right) = 0. \tag{6.4}$$

Observe that it is easy to prove the uniqueness of the solution of this equation: indeed by linearity it is sufficient to show that the norm is preserved in time, and this follows by:

$$\frac{d}{dt}\|u_t\|^2 = 2\langle \frac{d}{dt}u_t, u_t\rangle = 2\langle P_t\Big(\frac{d}{dt}u_t\Big), u_t\rangle = 0.$$

Therefore the problem is to show the existence of a solution of (6.4) for a given initial datum u_0. This is usually done by using coordinates and solving an appropriate system of differential equations, however this technique cannot be applied to the space $\mathscr{P}_2(\mathbb{R}^d)$ (we have neither Cristoffel symbols, nor coordinates). Here we are going to show how the parallel transport can be constructed using tools which have a Wasserstein analogous.

One concept useful to understand the process we will use is the following:

Definition 6.1 (Angle between subspaces). *Let $V_0, V_1 \subset \mathbb{R}^d$ be two given subspaces, and let P_i, $i = 0, 1$, be the orthogonal projection of $\mathbb{R}^n$ onto V_i and $P_i^\perp := Id - P_i$. Then the angle $\theta(V_0, V_1) \in [0, \pi/2]$ is defined by the following formula:*

$$\sin\theta(V_0, V_1) = \|P_1^\perp|_{V_0}\|_{op} = \sup_{\substack{v_0\in V_0\\ |v_0|=1}} |P_1^\perp(v_0)|.$$

It is not difficult to see that, letting $V_i^\perp$, $i = 0, 1$, the orthogonal complement of V_i, it holds

$$\begin{aligned}\sin\theta(V_0, V_1) &= \sup_{\substack{v_0\in V_0\\ |v_0|=1}} |v_0 - P_1(v_0)| \\ &= \sup_{\substack{v_0\in V_0,\ |v_0|=1\\ v_1^\perp\in V_1^\perp,\ |v_1^\perp|=1}} \langle v_0, v_1^\perp\rangle = \sin\theta(V_1^\perp, V_0^\perp).\end{aligned}$$

It is important to observe that in general $\theta(V_0, V_1) = \theta(V_1, V_0)$ does *not* hold: for instance if $V_0 \subset V_1$ we have $\theta(V_0, V_1) = 0$ and $\theta(V_1, V_0) = \pi/2$.

By applying this concept to a smooth curve $\gamma(t)$ on M, with $V_t = T_{\gamma(t)}M$, we clearly have that both functions

$$(t, s) \to \theta(V_t, V_s),$$
$$(t, s) \to \theta(V_s, V_t),$$

are Lipschitz. Therefore for some constant C both the following inequalities hold:

(6.5a) $$|u - P_s(u)| \le C|u||s - t|, \qquad \forall t, s \in [0, 1] \text{ and } u \in V_t.$$

(6.5b) $$|P_s(u^\perp)| \le C|u^\perp||s - t|, \qquad \forall t, s \in [0, 1] \text{ and } u^\perp \in V_t^\perp$$

The idea under the construction we are going to introduce is given by the following formula:

(6.6) $$\nabla_{v(0)} P_t(u) = 0, \quad \forall u \in V_0.$$

That is: the vectors $P_t(u)$ are a first order approximation of the parallel transport. Equation (6.6) follows by applying inequalities (6.5) (note that $u - P_t(u) \in V_t^\perp$):

$$|P_0(u - P_t(u))| \le Ct|u - P_t(u)| \le C^2t^2|u|.$$

Now let $\mathfrak{P}$ be the direct set of all the partitions of $[0,1]$, where $\mathcal{P} \geq \mathcal{Q}$, $\mathcal{P},\mathcal{Q} \in \mathfrak{P}$, if $\mathcal{P}$ is a refinement of $\mathcal{Q}$. For $\mathcal{P} = \{0 = t_0 < t_1 < \cdots < t_N = 1\} \in \mathfrak{P}$ and $u \in V_0$ define $\mathcal{P}(u) \in V_1$ as:

$$\mathcal{P}(u) := P_1(P_{t_{N-1}}(\cdots P_0(u))).$$

Our first goal is to prove that the limit $\mathcal{P}(u)$ for $\mathcal{P} \in \mathfrak{P}$ exists. After this is done, it will be naturally defined a curve $t \to u_t \in V_t$ by taking partitions of $[0,t]$ instead of $[0,1]$: the final target is to show that this curve is actually the parallel transport of u along the curve γ.

The proof is based on the following lemma.

Lemma 6.2. *Let $0 \leq s_1 \leq s_2 \leq s_3 \leq 1$ be given numbers. Then it holds:*

$$\big|P_{s_3}(u) - P_{s_3}(P_{s_2}(u))\big| \leq C^2|u||s_1 - s_2||s_2 - s_3|, \quad \forall u \in V_{s_1}.$$

Proof. Since $P_{s_3}(u) - P_{s_3}(P_{s_2}(u)) = (P_{s_3}(Id - P_{s_2})(u))$, the proof is a straightforward application of inequalities (6.5). ☐

From this lemma it follows easily by induction that the following inequality holds for any $0 \leq s_1 < \cdots < s_N \leq 1$ and $u \in V_{s_1}$:

(6.7)

$$\begin{aligned}
&\big|P_{s_N}(u) - P_{s_N}(P_{s_{N-1}}\cdots(P_{s_2}(u)))\big| \\
&\quad\leq \big|P_{s_N}(u) - P_{s_N}(P_{s_{N-1}}(u))\big| + \big|P_{s_N}(P_{s_{N-1}}(u)) - P_{s_N}(P_{s_{N-1}}\cdots(P_{s_2}(u)))\big| \\
&\quad\leq C^2|u||s_1 - s_{N-1}||s_{N-1} - s_N| + \big|P_{s_{N-1}}(u) - P_{s_{N-1}}\cdots(P_{s_2}(u))\big| \\
&\quad\leq \cdots \\
&\quad\leq C^2|u|\sum_{i=2}^{N-1}|s_1 - s_i||s_i - s_{i+1}| \\
&\quad\leq C^2|u||s_1 - s_N|^2.
\end{aligned}$$

With this result, we can prove the existence of the limit of $P(u)$ as P varies in $\mathfrak{P}$.

Theorem 6.3. *For any $u \in V_0$ there exists the limit of $\mathcal{P}(u)$ as $\mathcal{P}$ varies in $\mathfrak{P}$.*

Proof. We have to prove that given $\varepsilon > 0$ there exists a partition $\mathcal{P}$ such that it holds

$$|\mathcal{P}(u) - \mathcal{Q}(u)| \leq |u|\varepsilon, \quad \forall \mathcal{Q} \geq \mathcal{P}. \tag{6.8}$$

In order to do so it is sufficient to find $0 = t_0 < t_1 < \cdots < t_N = 1$ such that $\sum_i |t_{i+1} - t_i|^2 \leq \varepsilon/C^2$, and repetedly apply equation (6.7) (we will come back to this proof with more details in theorem 6.31). ☐

Now we can introduce the maps $T_s^t : V_s \to V_t$ which associate to the vector $u \in V_s$ the limit of the process just described (taking into account partitions of $[s,t]$).

Theorem 6.4. *For any $t_1 \leq t_2 \leq t_3 \in [0,1]$ it holds*

$$T_{t_2}^{t_3} \circ T_{t_1}^{t_2} = T_{t_1}^{t_3}. \tag{6.9}$$

Moreover for any $u \in V_0$ the curve $t \to u_t := T_0^t(u) \in V_t$ is the parallel transport of u along γ.

Proof. To prove equation (6.9) it is sufficient to consider those partitions of $[t_1, t_3]$ which contain t_2 and pass to the limit first on $[t_1, t_2]$ and then on $[t_2, t_3]$.

To prove the second part of the statement observe that due to (6.9) it is sufficient to check that the covariant derivative is 0 at 0. Note that from (6.7) it follows

$$|P_t(u) - u_t| \leq C^2 t^2,$$

therefore the thesis follows from (6.6). □

2. On the angle between tangent spaces in $\mathscr{P}_2(\mathbb{R}^d)$

The construction we did on regular manifolds embedded in $\mathbb{R}^d$ shows that the key step which allows to prove the existence of the parallel transport is the Lipschitz property of the angle between tangent spaces. In this section we introduce the analogous notion of angle for the space $\mathscr{P}_2(\mathbb{R}^d)$ and analyze its properties.

An important difference with the case of a manifold embedded in $\mathbb{R}^d$, is that in our context the two spaces $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and $\mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d))$ are not (affine) subspaces of a bigger Hilbert space, therefore we cannot directly imitate the definition of angle given in the first section. However, a natural way to embed L^2_ν into L^2_μ is given by the composition with a map T pushing μ into ν: to a map $f \in L^2_\nu$ we can associate the "translated" map $f \circ T \in L^2_\mu$.

Clearly the translation through a map is an isometry from L^2_μ to L^2_ν. The definition of angle comes out naturally.

Definition 6.5 (Angle between tangent spaces through a map). *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ be two measures and let T be a transport map from μ to ν. Then the angle $\theta_T(\mu, \nu) \in [0, \pi/2]$ between the tangent spaces at μ and ν through the map T is given by*

$$\sin\theta_T(\mu, \nu) := \sup \|P^\perp_\mu(v \circ T)\|_\mu,$$

where $P^\perp_\mu = Id - P_\mu$ where is the orthogonal projection onto $\mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and the supremum is taken among all $v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d))$ such that $\|v\|_\nu = 1$.

It is important to note that the angle between the tangent spaces at two measures, strongly depends on the transport maps used.

Observe that, assuming that the transport map T is invertible, the angle $\theta_T(\mu, \nu)$ is in general *n*ot equal to the angle $\theta_{T^{-1}}(\nu, \mu)$: this reflects the fact that the are two angles between two subspaces V_1 and V_2 of $\mathbb{R}^d$, depending on whether we are considering projections from V_1 onto V_2 or from V_2 to V_1.

The fundamental bound on the angle we are going to use in the rest of the chapter is given by the following proposition: the key requirement is the Lipschitz property of the transport map, while there is *n*o regularity assumption for the measures involved.

From now on, for a given vector valued map $T \in L^2_\mu$ we will indicate with $\mathrm{Lip}(T)$ the infimum of the least Lipschitz constant among all the functions equal to T μ–a.e..

Proposition 6.6. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $T \in L^2_\mu$ a transport map from μ to ν. Suppose that T is (equivalent to) a Lipschitz function. Then it holds*

$$\sin\theta_T(\mu, \nu) \leq \mathrm{Lip}(T - Id). \tag{6.10}$$

Proof. Equation (6.10) is equivalent to

$$\|\nabla\varphi\circ T - P_\mu(\nabla\varphi\circ T)\|_\mu \leq \|\nabla\varphi\|_\nu \operatorname{Lip}(T-Id), \quad \forall\varphi\in C_c^\infty(\mathbb{R}^d,\mathbb{R}). \tag{6.11}$$

Let us suppose at first that $T - Id \in C_c^\infty(\mathbb{R}^d,\mathbb{R}^d)$. In this case $\varphi\circ T\in C_c^\infty(\mathbb{R}^d)$, and therefore its gradient $\nabla(\varphi\circ T) = \nabla T\cdot(\nabla\varphi)\circ T$ belongs to $\operatorname{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. From the minimality properties of the projection we get:

$$\begin{aligned}\|\nabla\varphi\circ T - P_\mu(\nabla\varphi\circ T)\|_\mu &\leq \|\nabla\varphi\circ T - \nabla T\cdot(\nabla\varphi)\circ T\|_\mu\\ &= \left(\int |(\mathcal{I}-\nabla T(x))\cdot\nabla\varphi(T(x))|^2 d\mu(x)\right)^{1/2}\\ &\leq \left(\int |\nabla\varphi(T(x))|^2\|\nabla(Id-T)(x)\|_{op}^2 d\mu(x)\right)^{1/2}\\ &\leq \|\nabla\varphi\|_\nu \operatorname{Lip}(T-Id),\end{aligned}$$

where $\mathcal{I}$ is the identity matrix and $\|\nabla(Id-T)(x)\|_{op}$ is the operator norm of the linear functional from $\mathbb{R}^d$ to $\mathbb{R}^d$ given by $v\to\nabla(Id-T)(x)\cdot v$.

Now turn to the general case. Find a sequence $(T_n - Id)\subset C_c^\infty(\mathbb{R}^d,\mathbb{R}^d)$ such that

$$\begin{aligned}&T_n\to T &&\text{uniformly on compact sets,}\\ &\overline{\lim_n}\operatorname{Lip}(T_n-Id)\leq\operatorname{Lip}(T-Id).\end{aligned}$$

It is clear that for such a sequence it holds $\|T-T_n\|_\mu\to 0$, therefore we have

$$\begin{aligned}\|\nabla\varphi\circ T - P_\mu(\nabla\varphi\circ T)\|_\mu &\leq \|\nabla\varphi\circ T - \nabla(\varphi\circ T_n)\|_\mu\\ &\leq \|\nabla\varphi\circ T-\nabla\varphi\circ T_n\| + \|\nabla\varphi\circ T_n - \nabla(\varphi\circ T_n)\|\\ &\leq \operatorname{Lip}(\nabla\varphi)\|T-T_n\|_\mu + \|\nabla\varphi\circ T_n\|_\mu\operatorname{Lip}(T_n-Id).\end{aligned}$$

Letting $n\to+\infty$ we get the thesis. □

In the rest of the section we study the infinitesimal behavior of the angle between two tangent spaces. The results proven, although not needed in the following, help understanding the behavior of tangent and normal vectors w.r.t. (weak) convergence of maps.

Start observing that the notion of translation through a map may be generalized to the case of translation through a plan: the only difference being that the translation through a plan is no more an isometry, but just a non-expansive map.

Definition 6.7 (Composition of a vector with a plan). *For any couple of measures $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$ and any admissible plan $\boldsymbol{\gamma}\in\mathcal{A}dm(\mu,\nu)$ we define the* composition *of a vector $f\in L^2_\nu$ with $\boldsymbol{\gamma}$ as the vector $f\circ\boldsymbol{\gamma}\in L^2_\mu$ defined by:*

$$f\circ\boldsymbol{\gamma}(x) := \int f(y)d\boldsymbol{\gamma}_x(y).$$

The inequality

$$\|f\circ\boldsymbol{\gamma}\|_\mu^2 = \int\left(\int f(y)d\boldsymbol{\gamma}_x(y)\right)^2 d\mu(x)\leq\int|f(y)|^2 d\boldsymbol{\gamma}(x,y) = \|f\|_\nu^2,$$

shows that the composition with a plan is a non-expansive map. It is worth noticing that, letting $\boldsymbol{\alpha}$ be the unique 3-plan satisfying

$$\pi^{1,2}_{\#}\boldsymbol{\alpha} = \boldsymbol{\gamma},$$
$$\pi^{2,3}_{\#}\boldsymbol{\alpha} = (Id, f)_{\#}\nu,$$

the composition $f \circ \boldsymbol{\gamma}$ is nothing but the barycentric projection of the marginal $\pi^{1,3}_{\#}\boldsymbol{\alpha}$.

REMARK 6.8. *Actually there is no particular need to take the barycentric projection instead than simply the marginal, if one does so, he defines a norm preserving map from L^2_ν to $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$. We preferred to take the projection since we are only interested in translation of maps, and therefore it is simpler to work only with maps, instead than with plans.*

REMARK 6.9. *Obviously it is possible to define the composition of a plan in $\mathscr{P}_2(\mathbb{R}^{2d})_\nu$ with a transport plan in $\mathcal{Adm}(\mu,\nu)$: it is what we already did a lot of times by using Dudley's lemma. What comes out is a multi-valued operator continuous w.r.t. the Hausdorff distance between compact subsets of $\mathscr{P}_2(\mathbb{R}^{2d})_\mu$.*

DEFINITION 6.10 (Angle between tangent spaces through a plan). *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ be two measures and $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu,\nu)$ an admissible plan. Then the angle $\theta_{\boldsymbol{\gamma}}(\mu,\nu) \in [0,\pi/2]$ between the tangent spaces at μ and ν through the plan $\boldsymbol{\gamma}$ is defined as:*

$$\sin\theta_{\boldsymbol{\gamma}}(\mu,\nu) := \sup \|P^\perp_\mu(v\circ\boldsymbol{\gamma})\|_\mu, \qquad v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d)),\ \|v\|_\nu = 1.$$

As for the angles through maps, in general it holds $\theta_{\boldsymbol{\gamma}}(\mu,\nu) \neq \theta_{\boldsymbol{\gamma}^{-1}}(\nu,\mu)$.

We want to understand whether the angles $\theta_{\boldsymbol{\gamma}}(\mu,\nu)$, $\theta_{\boldsymbol{\gamma}^{-1}}(\nu,\mu)$ converge to 0 when μ is a fixed measure and $\boldsymbol{\gamma}$ converges to $(Id,Id)_{\#}\mu$.

Let (μ_n) be a sequence of measures $W-$converging to a measure μ, and $\boldsymbol{\gamma}_n \in \mathcal{Adm}(\mu,\mu_n)$ such that $\|x_1 - x_2\|_{\boldsymbol{\gamma}_n} \to 0$ as $n \to \infty$. In these hypotheses in general the angle $\theta_{\boldsymbol{\gamma}_n}(\mu,\mu_n)$ does not converge to 0. Actually it can be proved something more, as the following proposition shows:

PROPOSITION 6.11. *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $u \in L^2_\mu$. Then there exists a sequence of measures (μ_n) and a sequence of maps $u_n \in \mathrm{Tan}_{\mu_n}(\mathscr{P}_2(\mathbb{R}^d))$ such that (u_n) strongly converges to u.*

Proof. Assume for a moment that $\mu \ll \mathcal{L}^d$ with positive density and that u is Lipschitz. For every $n \in \mathbb{N}$, define the "grid" G_n as the set of those points x in $\mathbb{R}^d$ such that at least one coordinate of nx is integer, and the set H_n as

$$H_n := \left\{ x \in \mathbb{R}^d \ :\ d(x, G_n) > \frac{1}{n^2} \right\}.$$

It is clear that $\mu(\mathbb{R}^d \setminus H_n) \to 0$ as $n \to \infty$. Define

$$\mu_n := \frac{1}{\mu(H_n)}\mu|_{H_n}.$$

Now define, for every $n \in \mathbb{N}$, the function $u_n \in L^2_{\mu_n}$ as the function which is constant on each connected component of $\mathbb{R}^d \setminus H_n$, with value equal to the average of u on the connected component itself. Clearly $u_n \in L^2_{\mu_n}$, moreover, since it is constant on each connected component, it holds $u_n \in \mathrm{Tan}_{\mu_n}(\mathscr{P}_2(\mathbb{R}^d))$. Let L be the Lipschitz constant of u and conclude with the estimate

$$\|u_n - u\|_{\mu_n} \le \frac{L\sqrt{d}}{n}.$$

For the general case, define $\mu_\varepsilon := \mu * \rho_\varepsilon$ and $u_\varepsilon := (u\mu) * \rho_\varepsilon / \mu_\varepsilon$, approximate u_ε in $L^2_{\mu_\varepsilon}$ with Lipschitz functions and conclude with a diagonalization argument. □

Observe that the approximation of the measure μ we used, has the following properties: if $\mu \ll \mathcal{L}^d$ then $\mu_n \ll \mathcal{L}^d$. Furthermore, in this case it holds $\overline{\lim} \sup |\mu_n/\mathcal{L}^d| \le \sup |\mu/\mathcal{L}^d|$.

As said, this proposition implies that the angle $\overline{\theta_{\gamma_n}(\mu, \mu_n)}$ doesn't converge to 0 as $n \to \infty$: indeed, observe that

$$\begin{aligned}
\sin \theta_{\gamma}(\mu,\nu) &= \sup_{\substack{v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|v\|_\nu = 1}} \|P^\perp_\mu(v \circ \gamma)\|_\mu \\
&= \sup_{\substack{v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|v\|_\nu = 1}} \sup_{\substack{u \in \mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|u\|_\mu = 1}} \langle u, v \circ \gamma \rangle_\mu \\
&= \sup_{\substack{v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|v\|_\nu = 1}} \sup_{\substack{u \in \mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|u\|_\mu = 1}} \int \langle u(x), v(y) \rangle d\gamma(x,y) \\
&= \sup_{\substack{u \in \mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|u\|_\mu = 1}} \sup_{\substack{v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|v\|_\nu = 1}} \langle u \circ \gamma^{-1}, v \rangle_\nu \\
&= \sup_{\substack{u \in \mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d)) \\ \|u\|_\mu = 1}} \|P_\nu(u \circ \gamma^{-1})\|_\nu.
\end{aligned}$$

Now choose $u \in \mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and find $\mu_n \in \mathscr{P}_2(\mathbb{R}^d)$, $u_n \in \mathrm{Tan}_{\mu_n}(\mathscr{P}_2(\mathbb{R}^d))$ such that (u_n) strongly converges to u. Use remark 2.24 to get that for any choice of plans $\gamma_n \in \mathcal{A}dm(\mu, \mu_n)$ satisfying $\lim_n \|x - y\|_{\gamma_n} = 0$ the sequence $(u \circ \gamma_n^{-1})$ strongly converges to u and it holds $\|u \circ \gamma_n^{-1} - u_n\|_{\gamma_n} \to 0$, and conclude by the minimizing properties of the projection:

$$\begin{aligned}
\underline{\lim}_{n\to\infty} \|P_\nu(u \circ \gamma_n^{-1})\|^2_\nu &= \underline{\lim}_{n\to\infty} \left(\|u \circ \gamma_n^{-1}\|^2_\nu - \|u \circ \gamma_n^{-1} - P_{\mu_n}(u \circ \gamma_n^{-1})\|^2_{\mu_n} \right) \\
&\ge \underline{\lim}_{n\to\infty} \|u \circ \gamma_n^{-1}\|^2_\nu - \overline{\lim}_{n\to\infty} \|u \circ \gamma_n^{-1} - u_n\|^2_{\mu_n} = \|u\|^2_\mu.
\end{aligned}$$

With the same notation, regarding the angle $\theta_{\gamma_n^{-1}}(\mu_n, \mu)$ we don't have a final word, however we can show that for any $u \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ the vectors $u \circ \gamma_n^{-1} - P_{\mu_n}(u \circ \gamma_n^{-1})$ strongly converge to 0.

PROPOSITION 6.12. *Let $\mu, \mu_n \in \mathscr{P}_2(\mathbb{R}^d)$, $n \in \mathbb{N}$, and $\gamma_n \in \mathcal{A}dm(\mu, \mu_n)$ weakly converging to $(Id, Id)_\# \mu$ as $n \to \infty$. Then*

$$\|u \circ \gamma_n^{-1} - P_{\mu_n}(u \circ \gamma_n^{-1})\|_{\mu_n} \to 0, \quad \forall u \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)). \tag{6.12}$$

Proof. Observe that by the uniform continuity of the operators $u \to u \circ \boldsymbol{\gamma}_n^{-1} - P_\nu(u \circ \boldsymbol{\gamma}_n^{-1})$ (their norm is less or equal to 1), it is sufficient to prove equation (6.12) for $u = \nabla\varphi$ as φ varies in $C_c^\infty(\mathbb{R}^d)$. Now notice that $\nabla\varphi \in \mathrm{Tan}_{\mu_n}(\mathscr{P}_2(\mathbb{R}^d))$ for every n, therefore by the minimizing property of the projection we have:

$$\begin{aligned}\|\nabla\varphi \circ \boldsymbol{\gamma}_n^{-1} - P_\nu(\nabla\varphi \circ \boldsymbol{\gamma}_n^{-1})\|_{\mu_n}^2 \leq& \|\nabla\varphi \circ \boldsymbol{\gamma}_n^{-1} - \nabla\varphi\|_{\mu_n}^2 \\ \leq& \int_{\mathrm{supp}(\varphi)^2} |\nabla\varphi(x_1) - \nabla\varphi(x_2)|^2 d\boldsymbol{\gamma}_n(x_1, x_2) \\ \leq& \mathrm{Lip}(\nabla\varphi) \int_{\mathrm{supp}(\varphi)^2} |x_1 - x_2|^2 d\boldsymbol{\gamma}_n(x_1, x_2),\end{aligned}$$

and the last term goes to 0 as $n \to \infty$ by the weak convergence of $\boldsymbol{\gamma}_n$ to $(Id, Id)_\# \mu$. □

We can summarize the previous propositions by saying:

- A sequence of tangent vectors strongly converging, does not necessarily converge to a tangent vector, while a sequence of normal vectors weakly converging, converges to a normal vector
- For a given sequence (μ_n) converging to μ, every tangent vector in μ can be approximated by tangent vectors in μ_n w.r.t. the strong convergence of maps, while a normal vector may not be approximated by normal vector, not even w.r.t. the weak convergence of maps.

It is therefore interesting to know under which conditions on the measures, we have that any limit (possibly weak) of tangent vectors is a tangent vector. We don't have a general answer on this, however proposition 6.6 and the proof of 4.14 give the sufficient conditions collected by the following proposition.

PROPOSITION 6.13. *Let (μ_n) be a sequence of measures weakly converging to a measure μ and $u_n \in \mathrm{Tan}_{\mu_n}(\mathscr{P}_2(\mathbb{R}^d))$. Assume that (u_n) weakly converge to some $u \in L^2_\mu$ and that one of the two following conditions holds:*

- *there exists a sequence of transport maps (not necessarily optimal) (T_n) from μ to μ_n such that $(\mathrm{Lip}(T_n - Id))$ converges to 0 as n goes to infinity,*
- *there exists a sequence (ρ_n) of regular measures weakly converging to δ_0 such that $\mu_n = \mu * \rho_n$.*

Then $u \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.

Proof. Suppose the first condition is true. Then from proposition 6.6 we have $\|u_n \circ T_n - P_\mu(u_n \circ T_n)\|_\mu \leq \mathrm{Lip}(T_n - Id)\|u_n\|_{\mu_n}$ therefore the sequence $(P_\mu(u_n \circ T_n))$ weakly converges to u, too. Since $P_\mu(u_n \circ T_n) \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ we get the conclusion.

Now assume the second condition is true and choose a vector $w \in \mathrm{Tan}_\mu^\perp(\mathscr{P}_2(\mathbb{R}^d))$: we will show that $\langle w, u \rangle_\mu = 0$. By the Radon-Nikodym derivation theorem and the absolute continuity of $(w\mu) * \rho_n$ w.r.t. $\mu_n = \mu * \rho_n$ we know that the function

$$w_n := \frac{(w\mu) * \rho_n}{\mu_n},$$

is well defined for any $n \in \mathbb{N}$. Moreover from lemma 3.3 we know that $\|w_n\|_{\mu_n} \leq \|w\|_\mu$. It is clear that (w_n) strongly converges to w and that $\nabla \cdot (w_n \mu_n) = 0$, therefore using lemma 2.28 we can pass to the limit in $\langle u_n, w_n \rangle_{\mu_n} = 0$ to get $\langle u, w \rangle_\mu = 0$ as claimed.
□

3. Regular curves

In this section we will introduce the *regular* curves, which is a dense class of curves, along which a parallel transport can be defined and exists.

By analogy with the classical Riemannian case, (see equation (6.4)), and due to the discussion of the preceding section, we wish to say that u_t is a parallel transport along a curve μ_t if

$$\lim_{h \to 0} P_{\rho_t} \left(\frac{u_{t+h} \circ \mathbf{T}(t, t+h, \cdot) - u_t}{h} \right) = 0, \quad \forall t \in [0, T],$$

for a suitable class of maps $\mathbf{T}(t, s, \cdot)$ satisfying $\mathbf{T}(t, s, \cdot)_{\#}\mu_t = \mu_s$. Given that we want to apply proposition 6.6 (to replicate the approximation scheme for the construction of parallel transport), it is natural to ask for the maps $\mathbf{T}(t, s, \cdot)$ to be Lipschitz. From a technical point of view, it seems necessary that $\mathbf{T}(t, s, \cdot)$ have the semi-group property, i.e. $\mathbf{T}(t, s, \mathbf{T}(r, t, \cdot)) = \mathbf{T}(r, s, \cdot)$ (see the proof of lemma 6.29 on the next section). Finally, in order to have the parallel transport to induce the Levi-Civita connection on $(\mathscr{P}_2(\mathbb{R}^d), W)$, it must hold

$$\frac{d}{ds} \mathbf{T}(t, s, x)_{|s=\bar{s}} = v_{\bar{s}}(\mathbf{T}(t, \bar{s}, x)), \quad \forall t, \bar{s} \in [0, 1], x \in \operatorname{supp}(\mu_t),$$

where v_t are the velocity vectors of μ_t (see the discussion in section 5 for a further explanation).

Since we know from the classical Cauchy-Lipschitz theory that whenever the velocity vectors v_t of ρ_t satisfy

$$\int_0^1 \operatorname{Lip}(v_t) dt < +\infty,$$

then the flow maps exist and are Lipschitz functions of the space variable, we give the following definition:

Definition 6.14 (Regular curves). *Let $[0, T] \ni t \to \mu_t \in \mathscr{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve and let v_t be its velocity vector field defined, for a.e. $t \in [0, T]$, by*

$$v_t \in \operatorname{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)),$$
$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0.$$

We say that μ_t is regular if it holds

$$\int_0^T \operatorname{Lip}(v_t) dt < +\infty.$$

Observe that we are making *no* regularity assumption on the measures μ_t.

It is worth noticing that the definition of regularity for a curve is independent from its (Lipschitz) parameterization. In the following we will always assume that a regular curve is parameterized in $[0,1]$.

As already pointed out, for a given regular curve it exists a unique family of maps $\mathbf{T}(t,s,x) : [0,1]\times[0,1]\times\mathbb{R}^d \to \mathbb{R}^d$, which we call the *flow* of the curve, absolutely continuous in s and Lipschitz in x satisfying

$$\begin{cases} \mathbf{T}(t,t,x) = x, & \forall x,t, \\ \dfrac{d}{ds}\mathbf{T}(t,s,x)_{|s=\overline{s}} = v_{\overline{s}}(\mathbf{T}(t,\overline{s},x)) & \forall x,t, \quad a.e.\ \overline{s}, \\ \mathbf{T}(s,r,\mathbf{T}(t,s,x)) = \mathbf{T}(t,r,x), & \forall x,t,s,r \\ (\mathbf{T}(t,s,x))_{\#}\mu_t = \mu_s & \forall t,s\in[0,1]. \end{cases} \tag{6.13}$$

Observe that we have the following bound on the Lipschitz constant of $\mathbf{T}(t,s,\cdot)-Id$:

$$\mathrm{Lip}(\mathbf{T}(t,s,\cdot)-Id) \le \exp\left(\left|\int_t^s \mathrm{Lip}(v_r)dr\right|\right) - 1, \quad t,s\in[0,1]. \tag{6.14}$$

This inequality is a well known consequence of the validity of equations (6.13), we recall the proof for the sake of completeness.

Proposition 6.15. *Let $\mathbf{T}(t,s,\cdot)$ be the flow maps of a regular curve. Then the following estimates hold:*

$$\begin{aligned} \mathrm{Lip}(\mathbf{T}(t,s,\cdot)) &\le \exp\left(\left|\int_t^s \mathrm{Lip}(v_r)dr\right|\right), \quad t,s\in[0,1], \\ \mathrm{Lip}(\mathbf{T}(t,s,\cdot)-Id) &\le \exp\left(\left|\int_t^s \mathrm{Lip}(v_r)dr\right|\right) - 1, \quad t,s\in[0,1]. \end{aligned}$$

Proof. The first equation follows by a direct application of Gronwall lemma to the inequality

$$\begin{aligned} \frac{d}{ds}|\mathbf{T}(t,s,x)-\mathbf{T}(t,s,y)|^2 =& 2\langle \mathbf{T}(t,s,x)-\mathbf{T}(t,s,y), v_s(\mathbf{T}(t,s,x)) - v_s(\mathbf{T}(t,s,y))\rangle \\ \le& 2|\mathbf{T}(t,s,x)-\mathbf{T}(t,s,y)|^2\,\mathrm{Lip}(v_s). \end{aligned}$$

For the second one, observe that it holds

$$\begin{aligned} &\frac{d}{ds}|\mathbf{T}(t,s,x)-x-\mathbf{T}(t,s,y)+y|^2 \\ &\qquad =2\langle \mathbf{T}(t,s,x)-x-\mathbf{T}(t,s,y)+y, v_s(\mathbf{T}(t,s,x)) - v_s(\mathbf{T}(t,s,y))\rangle \\ &\qquad \le 2|\mathbf{T}(t,s,x)-x-\mathbf{T}(t,s,y)+y||x-y|\,\mathrm{Lip}(v_s)\,\mathrm{Lip}(\mathbf{T}(t,s,\cdot)). \end{aligned}$$

therefore the conclusion follows by integrating from t to s the inequality

$$\frac{d}{ds}|\mathbf{T}(t,s,x)-x-\mathbf{T}(t,s,y)+y| \le |x-y|\,\mathrm{Lip}(v_s)\exp\left(\left|\int_t^s \mathrm{Lip}(v_r)dr\right|\right).$$

□

For a vector field $t \to u_t \in L^2_{\mu_t}$ along a regular curve, we give the following definition of absolute continuity.

DEFINITION 6.16 (Absolutely continuous vector fields). *Let $t \to \mu_t$ be a regular curve and $t \to u_t \in L^2_{\mu_t}$ a vector field along it. We say that $t \to u_t$ is absolutely continuous if the maps $t \to u_t \circ \mathbf{T}(s,t,\cdot) \in L^2_{\mu_s}$ are absolutely continuous for any $s \in [0,1]$, where $\mathbf{T}(t,s,\cdot)$ are the flow maps of μ_t. For an absolutely continuous vector field we will write $\frac{d}{dt}u_t \in L^2_{\mu_t}$ for its derivative defined by:*

$$\frac{d}{dt}u_t := \lim_{h\to 0}\frac{u_{t+h}\circ \mathbf{T}(t,t+h,\cdot) - u_t}{h} = \frac{d}{dt}(u_t \circ \mathbf{T}(s,t,\cdot))\circ \mathbf{T}(t,s,\cdot), \quad \forall s \in [0,1].$$

Given that the composition with $\mathbf{T}(t,s,\cdot)$ is an isometry from $L^2_{\mu_s}$ to $L^2_{\mu_t}$, it is clear that a vector field $t \to u_t$ is absolutely continuous if and only if for some $s \in [0,1]$ the curve $t \to u_t \circ \mathbf{T}(s,t,\cdot) \in L^2_{\mu_s}$ is absolutely continuous. A simple consequence of the definition is that for an absolutely continuous vector field u_t, the map $t \to \|u_t\|_{\mu_t}$ is absolutely continuous.

It is important to underline that the definition of derivative of an absolutely continuous vector field allows us to take derivative of a function $t \to u_t$ whose range belongs to different L^2 spaces as t varies and that this derivative is *not* the partial derivative of some function of the kind $(t,x) \to u_t(x)$. To see this, consider a case in which the measures μ_t are mutually singular (this happens, for instance, if the measure μ_t is concentrated on some $k-$dimensional submanifold of $\mathbb{R}^d$, $k < d$, which moves in time): in this case the function u_t is not defined on the support of μ_s for $s \neq t$. Nevertheless, for regular curves, the above provides a good definition of derivative of a vector field along the curve.

In the following we will extensively use, without explicit mention, the fact that for any $\varphi \in C^\infty_c(\mathbb{R}^d,\mathbb{R})$, the vector field $t \to \nabla\varphi \in L^2_{\mu_t}$ is tangent and absolutely continuous. Its derivative in the sense of the above definition is $\nabla^2\varphi \cdot v_t$.

Observe that the definition of derivative of an absolutely continuous vector field is consistent with the Leibnitz rule, as shown by the following calculation:

(6.15)

$$\begin{aligned}
\frac{d}{dt}\langle u^1_t, u^2_t\rangle_{\mu_t} &= \frac{d}{dt}\langle u^1_t \circ \mathbf{T}(0,t,\cdot), u^2_t \circ \mathbf{T}(0,t,\cdot)\rangle_{\mu_0} \\
&= \langle \frac{d}{dt}(u^1_t \circ \mathbf{T}(0,t,\cdot)), u^2_t \circ \mathbf{T}(0,t,\cdot)\rangle_{\mu_0} + \langle u^1_t \circ \mathbf{T}(0,t,\cdot), \frac{d}{dt}(u^2_t \circ \mathbf{T}(0,t,\cdot))\rangle_{\mu_0} \\
&= \langle \frac{d}{dt}u^1_t, u^2_t\rangle_{\mu_t} + \langle u^1_t, \frac{d}{dt}u^2_t\rangle_{\mu_t}
\end{aligned}$$

We conclude the section by analyzing the first properties of regular curves: topological constraints, stability and density in the class of absolutely continuous curves. This part of the work may be skipped in a first reading: if the reader is more interested in the construction of the parallel transport along regular curves, we suggest to go directly to section 4.

REMARK 6.17. *An important consequence of the proposition 6.15, is a strong topological restriction on the supports of the measures μ_t along a regular curve. Indeed, from the fact that the flow maps $\mathbf{T}(t,s,\cdot):\mathbb{R}^d\to\mathbb{R}^d$ are (actually, can be extended to) bi-Lipschitz homeomorphisms of $\mathbb{R}^d$ into itself and from $\operatorname{supp}(\mu_s)=\mathbf{T}(t,s,\operatorname{supp}\mu_t)$, we obtain that the supports of the measures μ_t are all homeomorphic.*

In particular, if $\operatorname{supp}\mu_0$ has n connected components, say $A_0^1,\dots,A_0^n$, then $A_t^i:=\mathbf{T}(t,0,\cdot)(A_0^i)$, $i=1,\dots,n$, are the connected components of $\operatorname{supp}(\mu_t)$ and it holds

$$\mu_t|_{A_t^i}=\mathbf{T}(t,0,\cdot)_\#(\mu_0|_{A_0^i}),\quad i=1,\dots,n.$$

It follows that the mass $\mu_t(A_t^i)$ of each connected component is constant in time. Furthermore, it is easy to check that the velocity vectors of μ_t^i are (the restriction of) the vectors v_t, thus the curves $t\to\mu_t^i:=\mu_0(A_i)^{-1}\mu_t|_{A_t^i}$, $i=1,\dots,n$ are regular, too, and their flow maps are (the restriction of) the flow maps of μ_t. Notice that from the definition of absolutely continuous vector field, it is easy to check that a vector field $t\to u_t\in L^2_{\mu_t}$ is absolutely continuous if and only if so are the vector fields $t\to u_t\in L^2_{\mu_t^i}$, $i=1,\dots,n$.

We want now to point out a stability result for regular curves. This result is a direct consequence of theorem 3.5 of [**12**], which gives sufficient conditions for the flow maps of a sequence of curves (μ_t^n) to converge to the flow maps of the limit curve μ_t. Here we skip the proof. Observe that the hypothesis we make are slightly weaker than those of 3.5 of [**12**], given that we only assume convergence w.r.t. the weak topology of $\mathscr{P}_2(\mathbb{R}^d)$, however it is easy to check, following the same proof, that the result is still valid.

THEOREM 6.18 (Stability of regularity). *Let (μ_t^n) be a sequence of regular curves such that for every t the sequence (μ_t^n) converges to some μ_t w.r.t. the weak topology of $\mathscr{P}_2(\mathbb{R}^d)$. Assume also that the limit curve $t\to\mu_t$ is absolutely continuous, let v_t^n, v_t be the velocity vector fields of μ_t^n, μ_t respectively, and suppose that*

$$\begin{aligned}\overline{\lim_{n\to\infty}}\,\|v_t^n\|_{\mu_t^n}\le\|v_t\|_{\mu_t},&\quad a.e.\ t\in[0,1],\\ \operatorname{Lip}(v_t^n)\le g(t),&\quad\forall n\in\mathbb{N},\ a.e.\ t\in[0,1],\end{aligned}$$

for some function $g\in L^1(0,1)$.

Then it holds $\operatorname{Lip}(v_t)\le g(t)$ for a.e. t, thus the limit curve μ_t is regular. Furthermore for a.e. t the sequence (v_t^n) strongly converges to v_t and for every $t,s\in[0,1]$ the flow maps $\mathbf{T}^n(t,s,\cdot)$ of μ_t^n strongly converge to $\mathbf{T}(t,s,\cdot)$.

We conclude this section with a proof of a density result for regular curves. It is well-known that the set

$$\mathscr{P}_2^a(\mathbb{R}^d):=\left\{\mu\in\mathscr{P}_2(\mathbb{R}^d)\,:\,\mu\ll\mathscr{L}^d\right\}$$

is a *geodesic* subspace of $\mathscr{P}_2(\mathbb{R}^d)$ (i.e. any geodesic between two points in $\mathscr{P}_2^a(\mathbb{R}^d)$ is entirely contained in $\mathscr{P}_2^a(\mathbb{R}^d)$) and the same is true for the subsets $\{\mu=\rho\mathscr{L}^d\,:$

$\|\rho\|_\infty \le C\}$. Our approximation will be obtained with measures in this class, and preserves these upper bounds on the densities, if any.

The delicate point in the approximation result is due to the fact that regularity imposes a Lipschitz condition on the *tangent* velocity field. The typical approximation schemes for solutions to the continuity equation, on the other hand, produce a regularized vector field that is compatible with the regularized density, but it is not tangent in general. Therefore a further projection of the regularized velocity on the tangent space is needed.

The following lemma will be used in the reduction to compactly supported measures.

LEMMA 6.19 (Monotone approximation). *Let $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ be absolutely continuous and let v_t be its tangent velocity field. Then there exist absolutely continuous curves $\mu_t^n : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ and $z_n \uparrow 1$ satisfying:*

(i) *$z_n\mu_t^n \uparrow \mu_t$ for all $t \in [0,1]$ and $\sup_t W_2(\mu_t^n, \mu_t) \to 0$ as $n \to \infty$;*
(ii) *the tangent velocity field of μ_t^n is v_t, and there exists a closed ball B_n such that $\operatorname{supp}\mu_t^n \subset B_n$ for all $t \in [0,1]$.*

PROOF. Let Ω be the Banach space of continuous maps from $[0,1]$ to $\mathbb{R}^d$ and let $e_t : \Omega \to \mathbb{R}^d$ be the evaluation maps at time t, i.e. $e_t(\omega) = \omega(t)$. According to [**11**, 8.2.1], we can represent μ_t as the law under e_t of a suitable probability measure η in Ω, concentrated in the set of absolutely continuous solutions of the equation $\dot\omega = v_t(\omega)$.

Let

$$\Omega_n := \left\{\omega \in \Omega \,:\, \omega \text{ is absolutely continuous, } |\omega(0)| \le n,\ \int_0^1 |\dot\omega| dt \le n\right\}$$

and set $\eta_n = \chi_{\Omega_n}\eta$, $z_n = \eta(\Omega_n) = \eta_n(\Omega)$ and $\mu_t^n = z_n^{-1}(e_t)_\#\eta_n$. It is easy to check condition (i), and that the support of μ_t^n is contained in the ball $\overline{B}_{2n}(0)$. Since also η_n is concentrated on curves solving the ODE $\dot\omega = v_t(\omega)$, it turns out that v_t is an admissible velocity field for μ_t^n (i.e. the continuity equation holds, see again [**11**, 8.2.1] for instance). We conclude that v_t is the tangent velocity fields noticing that, because of condition (i), $v_t \in \operatorname{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ implies $v_t \in \operatorname{Tan}_{\mu_t^n}(\mathscr{P}_2(\mathbb{R}^d))$. □

We can now state our approximation result.

THEOREM 6.20 (Approximation by regular curves). *Let $\mu_t : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve. Then there exist regular curves $\mu_t^n : [0,1] \to \mathscr{P}_2(\mathbb{R}^d)$ satisfying:*

(i) *$\sup_t W_2(\mu_t^n, \mu_t) \to 0$ as $n \to +\infty$;*
(ii) *$\mu_t^n = \rho_t^n \mathscr{L}^n$, $\sup_t \|\rho_t^n\|_\infty < +\infty$, ρ_t^n are smooth, the smooth tangent velocity fields v_t^n are gradients of smooth maps $\varphi_t^n : \mathbb{R}^d \to \mathbb{R}$ satisfying $\sup_t \operatorname{Lip}(v_t^n) < \infty$ and $\{\rho_t^n > 0\}$ is a bounded open set with a smooth boundary;*
(iii) *if $v_t \in L^2_{\mu_t}$ is the tangent field of μ_t, we have that $v_t^n\mu_t^n$ weakly converge to $v_t\mu_t$ and*

$$\lim_{n\to\infty} \int |v_t^n|^2 d\mu_t^n = \int |v_t|^2 d\mu_t.$$

PROOF. **Step 1.** Regularization of 1-periodic solutions.

By Lemma 6.19 and a diagonal argument we can assume that the supports of μ_t are contained in a fixed compact set. By a scaling argument, we can also assume with no loss of generality that the union of these supports is a compact subset K of $(0,1)^d$. We consider the 1-periodic extension $\mu_t^{\rm per}$ of μ_t, still solving the continuity equation with the 1-periodic extension $v_t^{\rm per}$ of v_t, and the regularized densities

$$\varrho_t^n := \mu_t^{\rm per} * \chi_n,$$

still 1-periodic. Here χ_n is a family of smooth and symmetric w.r.t. 0, convolution kernels converging to δ_0 whose support has a diameter *equal* to $2\sqrt{d}$. With this choice of χ_n, we have $\inf_t \inf \varrho_t^n > 0$, and standard properties of convolution yield

$$\sup \varrho_t^n \le \sup_t \mu_t^{\rm per}([-1,2]^d) \sup \chi_n = 3^d \sup \chi_n.$$

Analogous bounds hold, of course, for all higher order derivatives of ϱ_t^n. Passing to the velocity fields, we consider as in [**36, 11**] this regularization:

$$w_t^n := \frac{(v_t^{\rm per}\mu_t^{\rm per}) * \chi_n}{\varrho_t^n}$$

which satisfies, thanks to the lower bound on ϱ_t^n, $\sup_t {\rm Lip}(w_t^n) < \infty$ (and the same holds for higher order derivatives) and preserves the validity of the continuity equation. Eventually we consider the projection $v_t^n = \nabla\varphi_t^n$ of w_t^n on periodic gradients by solving the PDE

$$\nabla\cdot(\nabla\varphi_t^n\varrho_t^n) = \nabla\cdot(w_t^n\varrho_t^n).$$

From the variational formulation of the PDE we obtain

$$\int_{(0,1)^d} |v_n^t|^2\varrho_t^n\,dx \le \int_{(0,1)^d} |w_n|^2\varrho_t^n dx. \tag{6.16}$$

We can use standard elliptic regularity theory to obtain that $\sup_t {\rm Lip}(v_t^n) < \infty$. Moreover, using Jensen's inequality as in [**11**, 8.1.10], we have the local estimate

$$\int_{(0,1)^d} |w_t^n|^2\varrho_t^n dx \le \int\int_{(0,1)^d} |v_t|^2(y)\chi_n(x-y)dxd\mu_t^{\rm per}(y). \tag{6.17}$$

Step 2. Construction of the approximating sequence.

We build $\mu_t^n = \rho_t^n\mathscr{L}^d \in \mathscr{P}_2^a(\mathbb{R}^d)$, with the same velocity field v_t^n, from the periodic measures $\varrho_t^n\mathscr{L}^d$. To this aim, we shall first consider ϱ_t^n as measures in the flat d-dimensional torus $\mathbb{T}^d \sim (0,1)^d$, with velocity field v_t^n. We denote by $\mathbb{P}$ the Lebesgue measure on $\mathbb{T}^d$, by $\mathbf{X}^n(t,x) : [0,1]\times\mathbb{T}^d \to \mathbb{T}^d$ the smooth flow of v_t^n (starting from $s=0$), and by η_n the probability measure in $C([0,1];\mathbb{T}^d)$ defined by

$$\eta_n := \mathbf{X}^n(t,\cdot)_\#(\varrho_0^n\mathbb{P}).$$

Equivalently, η_n is the law of the random variable $x \mapsto \mathbf{X}^n(\cdot,x) \in C([0,1];\mathbb{T}^d)$ under $\varrho_0^n\mathbb{P}$. Classical representation results for solutions to the continuity equation with a Lipschitz vector field ensure that $\varrho_t^n\mathbb{P} = \mathbf{X}^n(t,\cdot)_\#(\varrho_0^n\mathbb{P})$, and since $e_t\circ\mathbf{X}^n(\cdot,x) = \mathbf{X}^n(t,x)$ we obtain

$$(e_t)_\#\eta_n = \varrho_t^n\mathbb{P} \qquad \forall t\in[0,1],\ n\in\mathbb{N}. \tag{6.18}$$

From (6.16) and (6.17) we get

$$\sup_n \int_0^1 \int_{\mathbb{T}^d} |v_t^n|^2 \varrho_t^n d\mathbb{P}(x) dt < \infty$$

and this, using Prokhorov theorem as in [**12, 44**], gives that (η_n) is a relatively compact sequence in $\mathscr{P}(C([0,1];\mathbb{T}^d))$. It is not restrictive, extracting if necessary a subsequence, to assume that (η_n) weakly converges, in the duality with continuous and bounded functions in $C([0,1];\mathbb{T}^d)$, to some probability measure η. Passing to the limit as $n \to \infty$ in (6.18) we obtain that $(e_t)_\# \eta = \mu_t^{\rm per}$ for all $t \in [0,1]$, and this means that η-almost all the paths ω are contained in $\tilde K$ (here we denote by $\tilde K$ the image of K in $\mathbb{T}^d$ and we consider $\mu_t^{\rm per}$ as probability measures in $\mathbb{T}^d$).

Now, let $\delta < 1$ be such that K is contained in the interior of $[\delta, 1-\delta]^d$ and define $\tilde\eta_n := z_n^{-1} \chi_{\Omega(\delta)} \eta_n$, where

$$\Omega(\delta) := \left\{\omega \in C([0,1];\mathbb{T}^d) : \omega(t) \bmod(1) \in (\delta, 1-\delta)^d \ \forall t \in [0,1]\right\}, \qquad z_n := \eta_n(\Omega(\delta))$$

(in other words, we remove the trajectories that cross $\partial(\delta, 1-\delta)^d$). Since η is supported on paths contained in $\tilde K$, we have that $z_n \to 1$ and still $\tilde\eta_n$ weakly converge to $\chi_{\Omega(\delta)}\eta = \eta$. We define

$$\mu_t^n := (\tilde e_t)_\# \tilde\eta_n$$

where $\tilde e_t(\omega(t)) = \omega_t \bmod(1) \in [0,1)^d$. The measures μ_t^n can also be represented by

$$\mu_t^n = z_n^{-1} \mathbf{Y}^n(t,\cdot)_\# (\chi_{E^n(\delta)} \varrho_0^n \mathscr{L}^d), \tag{6.19}$$

where $\mathbf{Y}^n(t,x) = \mathbf{X}^n(t,x) \bmod(1)$ and $E^n(\delta) = \{x \in (0,1)^d \,:\, \mathbf{X}^n(\cdot,x) \in \Omega(\delta)\}$.

By construction μ_t^n are probability measures in $\mathbb{R}^d$ concentrated on $[\delta, 1-\delta]^d$. It is immediate to check that the tangent field to μ_t^n is v_t^n (because η_n is concentrated on solutions to the ODE $\dot\omega = v_t^n(\omega)$ in $\mathbb{T}^d$, and v_t^n are gradients). In particular μ_t^n are regular curves and the convergence of μ_t^n to μ_t follows at once from the convergence of $\tilde\eta_n$ to η, using the evaluation map $\tilde e_t$. Notice also that the inequality $z_n \tilde\eta_n \le \eta_n$ and the fact that the mass of their difference is infinitesimal imply

$$z_n \mu_t^n \le \varrho_t^n \mathscr{L}^d \quad \text{and} \quad \lim_{n\to\infty} (\varrho_t^n \mathscr{L}^d - z_n \mu_t^n)((0,1)^d) = 0. \tag{6.20}$$

Step 3. Convergence of velocity fields.

Notice first that (6.16) and (6.17) give

$$\begin{aligned}
\limsup_{n\to\infty} \int |v_t^n|^2 d\mu_t^n &\le \limsup_{n\to\infty} \int |v_t^n(e_t(\omega))|^2 d\eta_n(\omega) = \limsup_{n\to\infty} \int |v_t^n|^2 \varrho_t^n d\mathbb{P} \\
&\le \limsup_{n\to\infty} \int |w_t^n|^2 \varrho_t^n d\mathbb{P} \\
&\le \limsup_{n\to\infty} \int \int_{(0,1)^d} |v_t|^2(y) \chi_n(x-y) dx d\mu_t^{\rm per}(y) = \int |v_t|^2 d\mu_t.
\end{aligned}$$

Now, recall (see for instance [**11**, 9.4.3]) that the functional

$$G(\nu,\mu) := \begin{cases} \displaystyle\int |f|^2 d\mu & \text{if } \nu = f\mu \text{ with } f \in L^2(\mu;\mathbb{R}^d), \\ +\infty & \text{otherwise} \end{cases}$$

is jointly lower semicontinuous in $\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)$ with respect to weak convergence in the duality with $C_b(\mathbb{R}^d)$, to obtain that any weak limit point σ of $v_t^n\mu_t^n$ as $n\to\infty$ has the form $\tilde v\mu_t$ for some $\tilde v \in L^2_{\mu_t}$ with $\|\tilde v\|_{\mu_t}\le \|v_t\|_{\mu_t}$. On the other hand, passing to the limit in $\nabla\cdot((w_t^n - v_t^n)\varrho_t^n) = 0$ and taking into account the weak convergence in the duality with $C_c(\mathbb{R}^d)$ of $w_t^n\varrho_t^n = (v_t^{\rm per}\mu_t^{\rm per}) * \chi_n$ to $v_t^{\rm per}\mu_t^{\rm per}$ and the convergence to 0 in $(0,1)^d$ of $\varrho_t^n\mathscr{L}^d - \mu_t^n$ (ensured, even in the strong sense, by (6.20)) we get $\nabla\cdot((\tilde v - v_t)\mu_t) = 0$. Since v_t is tangent and $\|\tilde v\|_{\mu_t}\le\|v_t\|_{\mu_t}$, it must be $\tilde v = v_t$. This proves the weak convergence of velocity fields that provides also, thanks to the lower semicontinuity of G, the $\liminf$ inequality

$$\liminf_{n\to\infty}\int |v_t^n|^2 d\mu_t^n \ge \int |v_t|^2 d\mu_t.$$

Step 4. Eventually we can regularize the characteristic function of the set $E^n(\delta)$ in (6.19), by smooth functions χ^n such that $\{\chi_n\varrho_0^n > 0\}$ is smooth and bounded, to approximate the curve μ_t^n by curves with the same velocity field and smooth densities with respect to $\mathscr{L}^d$, with smooth supports. □

The following is a useful technical result, which allows to approximate a regular curve with regular curves with range in the set of measures with finite support. The advantage of working with these measures μ, is that in this case an absolutely continuous curve is automatically regular: indeed, by continuity and compactness it is clear that $\inf_t \inf_{x_t\neq y_t} |x_t - y_t| > 0$, where $x_t, y_t \in \operatorname{supp}\mu_t$, therefore, since the velocity vectors v_t are defined only in a finite set, it is clear that $t\to \mathrm{Lip}(v_t)$ is integrable as soon as $t\to\|v_t\|_{\mu_t}$ is.

PROPOSITION 6.21. *Let $t\to\mu_t$ be a regular curve. Then there exists a sequence of regular curves (μ_t^n) such that for every n,t the support of μ_t^n is a finite set of points and it holds*

$$\lim_{n\to\infty}\sup_{t\in[0,1]} W(\mu_t^n,\mu_t) = 0,$$
$$\lim_{n\to\infty}\|v_t^n\|_{\mu_t^n} = \|v_t\|_{\mu_t}, \quad a.e.\ t\in[0,1],$$
$$\sup_{n\in\mathbb{N}}\mathrm{Lip}(v_t^n)\le \mathrm{Lip}(v_t), \quad a.e.\ t\in[0,1],$$

where v_t, v_t^n are the velocity vectors of $t\to\mu_t$ and $t\to\mu_t^n$ respectively.

It is clear that the cardinality of the support of the measures of a regular curve, cannot change in time. Therefore the bound on the cardinality of $\operatorname{supp}(\mu_t^n)$ is, for every n, uniform in t.

Proof. Let $\mathbf{T}(t,s,x)$ be the flow maps of μ_t. Find a sequence (μ_0^n) of measures with finite support such that $\mathrm{supp}(\mu_0^n) \subset \mathrm{supp}\,\mu_0$ and $W(\mu_0, \mu_0^n) \to 0$ as $n \to \infty$. Now define

$$\mu_t^n := \mathbf{T}(0,t,\cdot)_\# \mu_0^n.$$

It is clear that $t \to \mu_t^n$ is an absolutely continuous curve whose velocity vector field is v_t. Therefore the thesis will follow if we show the uniform convergence of $t \to \mu_t^n$ to $t \to \mu_t$. Observe that the maps $\mathbf{T}(0,t,\cdot)$ are equilipschitz in t, let L be a common Lipschitz constant, choose a plan $\boldsymbol{\gamma}_0^n \in \mathcal{Opt}(\mu_0, \mu_0^n)$ and define

$$\boldsymbol{\gamma}_t^n := (\mathbf{T}(0,t,\cdot), \mathbf{T}(0,t,\cdot))_\# \boldsymbol{\gamma}_0^n.$$

It holds $\boldsymbol{\gamma}_t^n \in \mathcal{Adm}(\mu_t, \mu_t^n)$, therefore

$$W(\mu_t^n, \mu_t) \le \|x_1 - x_2\|_{\boldsymbol{\gamma}_t^n} = \|\mathbf{T}(0,t,x_1) - \mathbf{T}(0,t,x_2)\|_{\boldsymbol{\gamma}_0^n} \le L\|x_1 - x_2\|_{\boldsymbol{\gamma}_0^n} = LW(\mu_0^n, \mu_0).$$

□

4. Parallel transport

In this section we define the parallel transport along regular curves in $\mathscr{P}_2(\mathbb{R}^d)$, show its existence, uniqueness and stability. We will also show by giving an explicit example that the lack of regularity may lead to the non-existence of the parallel transport, even along geodesics. At the end of the section we collect some examples of regular curves and parallel transports along them.

Definition 6.22 (Parallel transport along regular curves). *Let μ_t be a regular curve, $\mathbf{T}(t,s,x)$ its flow and $t \to u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ a tangent vector field defined along the curve. We say that $t \to u_t$ is a parallel transport if it is absolutely continuous and it holds*

$$P_{\mu_t}\left(\frac{d}{dt}u_t\right) = 0, \quad a.e.\ t \in [0,1]. \tag{6.21}$$

Equation (6.21) may be equivalently written as:

$$\lim_{h\to 0} P_{\mu_t}\left(\frac{u_{t+h}\circ \mathbf{T}(t,t+h,\cdot) - u_t}{h}\right) = 0, \quad in\ L^2_{\mu_t} a.e.\ t \in [0,1].$$

Another equivalent characterization, thanks to (6.15), is

$$\frac{d}{dt}\langle \nabla\eta, u_t\rangle_{\mu_t} = \langle \nabla^2\eta \cdot v_t, u_t\rangle_{\mu_t} \qquad \text{for a.e. } t \in (0,1), \text{ for all } \eta \in C_c^\infty(\mathbb{R}^d), \tag{6.22}$$

or, in integral form:

$$\langle \nabla\eta, u_s\rangle_{\mu_s} - \langle \nabla\eta, u_t\rangle_{\mu_t} = \int_t^s \langle \nabla^2\eta \cdot v_r, u_r\rangle_{\mu_r} dr, \quad \forall \eta \in C_c^\infty. \tag{6.23}$$

Observe that this equation makes sense even if the underlying curve $t \mapsto \mu_t$ is not regular, but only absolutely continuous. We will come back to this point in the last section.

It is easy to check that also the concept of parallel transport is invariant under reparameterization: if μ_t is a regular curve, u_t is a parallel transport along it and

$r : [0, R] \to [0, 1]$ is a Lipschitz reparameterization of $[0, 1]$, then $\tilde{\mu}_s := \mu_{r(s)}$ is regular in $[0, R]$ and $\tilde{u}_s := u_{r(s)}$ is a parallel transport along it.

PROPOSITION 6.23 (Linearity and conservation of scalar product). *Let u^1_t, u^2_t be two parallel transports along a regular curve $t \to \mu_t$ on $[0, 1]$ and $\lambda^1, \lambda^2 \in \mathbb{R}$. Then $t \to \langle u^1_t, u^2_r \rangle_{\mu_t}$ is constant and $t \to \lambda^1 u^1_t + \lambda^2 u^2_t$ is a parallel transport.*

Proof. The claim on $\lambda^1 u^1_t + \lambda^2 u^2_t$ follows directly by the linearity of the definition of absolutely continuous vector field and of equation (6.21).

To prove that the scalar product is constant, just observe that $t \to \langle u^1_t, u^2_t \rangle_{\mu_t} = \frac{1}{2}(\|u^1_t\|^2_{\mu_t} + \|u^2_t\|^2_{\mu_t} - \|u^1_t - u^2_t\|^2_{\mu_t})$ is absolutely continuous and that by the Leibnitz rule (6.15) its derivative is given by

$$\frac{d}{dt}\langle u^1_t, u^2_t \rangle_{\mu_t} = \langle \frac{d}{dt} u^1_t, u^2_t \rangle_{\mu_t} + \langle u^1_t, \frac{d}{dt} u^2_t \rangle_{\mu_t} = \langle P_{\mu_t}(\frac{d}{dt} u^1_t), u^2_t \rangle_{\mu_t} + \langle u^1_t, P_{\mu_t}(\frac{d}{dt} u^2_t) \rangle_{\mu_t} = 0.$$

□

As a direct consequence we get the uniqueness of the parallel transport.

COROLLARY 6.24 (Uniqueness of parallel transport). *Let μ_t be a regular curve and $u \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$. Then there exists at most one parallel transport $t \to u_t$ along μ_t such that $u_0 = u$.*

Proof. Suppose that u^i_t, $i = 1, 2$, are two parallel transports such that $u^i_0 = u$, $i = 1, 2$. Then we know from the previous proposition that $u^1_t - u^2_t$ is a parallel transport, too. Since we know also that the norm is preserved in time, we get $\|u^1_t - u^2_t\|_{\mu_t} = \|u - u\|_{\mu_0} = 0$. □

Observe that for parallel transports we have an explicit bound on the norm of $\frac{d}{dt} u_t$, which depends only on the Lipschitz constant of the vectors v_t.

PROPOSITION 6.25. *Let $t \to \mu_t$ be a regular curve and $t \to u_t$ a parallel transport along it. Then it holds*

$$\left\| \frac{d}{dt} u_t \right\|_{\mu_t} \le \|u_0\|_{\mu_0} \mathrm{Lip}(v_t), \quad a.e.\ t \in [0, 1]. \tag{6.24}$$

Proof. We will prove that equation (6.24) is fulfilled at any Lebesgue point t of the function $s \to \mathrm{Lip}(v_s)$. Fix such t and observe that it holds

$$\|u_s \circ \mathbf{T}(t, s, \cdot) - u_t\|_{\mu_t} \le \|P_{\mu_t}(u_s \circ \mathbf{T}(t, s, \cdot)) - u_t\|_{\mu_t} + \|P^\perp_{\mu_t}(u_s \circ \mathbf{T}(t, s, \cdot))\|_{\mu_t}.$$

Dividing by $|s - t|$ and letting $s \to t$ we have that the first term goes to 0 by definition of parallel transport, while for the second one we have the following estimate, based on proposition 6.6:

$$\begin{aligned} \lim_{s \to t^+} \frac{\|P^\perp_{\mu_t}(u_s \circ \mathbf{T}(t, s, \cdot))\|_{\mu_t}}{s - t} &\le \lim_{s \to t^+} \frac{1}{s - t} \|u_0\|_{\mu_0} \mathrm{Lip}(\mathbf{T}(t, s, \cdot) - Id) \\ &\le \lim_{s \to t^+} \|u_0\|_{\mu_0} \frac{\exp\left(\int_t^s \mathrm{Lip}(v_r) dr\right) - 1}{s - t} = \|u_0\|_{\mu_0} \mathrm{Lip}(v_t). \end{aligned}$$

The case $s \to t^-$ is analogous. □

A simple consequence of theorem 6.18 is the following stability result of the parallel transport.

THEOREM 6.26 (Stability of parallel transport). *Let (μ_t^n) be a sequence of regular curves such that for every t the sequence (μ_t^n) converges to some μ_t w.r.t. the weak topology of $\mathscr{P}_2(\mathbb{R}^d)$. Assume also that the limit curve μ_t is absolutely continuous and that it holds:*

$$\begin{aligned}\overline{\lim_{n\to\infty}}\|v_t^n\|_{\mu_t^n} &\leq \|v_t\|_{\mu_t}, \quad a.e.\ t\in[0,1],\\ \mathrm{Lip}(v_t^n) &\leq g(t), \quad \forall n\in\mathbb{N},\ a.e.\ t\in[0,1],\end{aligned}$$

where v_t^n, v_t are the velocity vectors of μ_t^n, μ_t respectively and $g\in L^1(0,1)$. Let $u_t^n\in\mathrm{Tan}_{\mu_t^n}(\mathscr{P}_2(\mathbb{R}^d))$ be a sequence of parallel transports along μ_t^n and assume that

i) *for every $t\in[0,1]$, every weak limit point of (u_t^n) is tangent,*
ii) *for some $t_0\in[0,1]$ $(u_{t_0}^n)$ weakly (resp. strongly) converges to some $u_{t_0}\in\mathrm{Tan}_{\mu_{t_0}}(\mathscr{P}_2(\mathbb{R}^d))$.*

Then for every $t\in[0,1]$ (u_t^n) weakly (resp. strongly) converges to some u_t and the limit vector field is tangent.

Proof. We know from theorem 6.18 that the limit curve is regular and that the maps $\mathbf{T}^n(t,s,\cdot)$ converge strongly to the map $\mathbf{T}(t,s,\cdot)$ for any $t,s\in[0,1]$.

Possibly extracting a subsequence, not relabeled, we may assume that (u_t^n) weakly converges to some vector $u_t\in L^2_{\mu_t}$ for every $t\in[0,1]\cap\mathbb{Q}$. In this hypothesis, the maps $u_s^n\circ\mathbf{T}^n(t,s,\cdot)$ converge weakly to the map $u_s\circ\mathbf{T}(t,s,\cdot)$ for any $s,t\in[0,1]\cap\mathbb{Q}$, indeed they are clearly equibounded and it holds

$$\langle u_s^n\circ\mathbf{T}^n(t,s,\cdot),\varphi\rangle_{\mu_t^n}=\langle u_s^n,\varphi\circ\mathbf{T}^n(s,t,\cdot)\rangle_{\mu_s^n}\to\langle u_s,\varphi\circ\mathbf{T}(s,t,\cdot)\rangle_{\mu_s}=\langle u_s\circ\mathbf{T}(t,s,\cdot),\varphi\rangle_{\mu_t},$$

having used proposition 2.23 to get the strong convergence of $\varphi\circ\mathbf{T}^n(s,t,\cdot)$ to $\varphi\circ\mathbf{T}(s,t,\cdot)$.

Therefore from proposition 6.25 we get

$$\|u_s\circ\mathbf{T}(t,s,\cdot)-u_t\|_{\mu_t}\leq\varliminf_{n\to\infty}\|u_s^n\circ\mathbf{T}^n(t,s,\cdot)-u_t^n\|_{\mu_t^n}\leq\varliminf_{n\to\infty}\int_t^s\mathrm{Lip}(v_r^n)dr\leq\int_t^s g(r)dr,$$

so the vector field $t\to u_t$ is absolutely continuous.

Now observe that since $t\to u_t^n$ is a parallel transport along μ_t^n, from equation (6.23) we know that it holds:

$$\langle u_s^n,\nabla\eta\rangle_{\mu_s^n}-\langle u_t^n,\nabla\eta\rangle_{\mu_t^n}=\int_t^s\langle u_r^n,\nabla^2\eta\cdot v_r^n\rangle_{\mu_r^n}dr,\quad\forall\eta\in C_c^\infty(\mathbb{R}^d),0\leq t\leq s\leq 1.$$

The strong convergence of (v_r^n) to v_r implies the strong convergence of $\nabla^2\eta\cdot v_r^n$ to $\nabla^2\eta\cdot v_r$, furthermore the $\overline{\lim}$ of the functions $r\to\langle u_r^n,\nabla^2\eta\cdot v_r^n\rangle_{\mu_r^n}$ is dominated by the $L^1(0,1)$ function $\overline{\lim}_n\|u_r^n\|_{\mu_r^n}\mathrm{Lip}(\nabla\eta)\|v_r^n\|_{\mu_r^n}\leq\sup_n\left(\|u_{t_0}^n\|_{\mu_{t_0}^n}\right)\mathrm{Lip}(\nabla\eta)\|v_r\|_{\mu_r}$, so by the dominated convergence theorem we can pass to the limit in the above equation and get that the limit vector field is a parallel transport.

Given that we know that $(u_{t_0}^n)$ weakly converges to u_{t_0} and that the parallel transport is unique, we get that this result is independent on the subsequence chosen, so for every $t\in[0,1]$ the whole sequence (u_t^n) weakly converges to u_t. If the convergence of

$(u^n_{t_0})$ to u_{t_0} is strong, from the fact that the parallel transport preserves the norm we get $\lim_n \|u^n_t\|_{\mu^n_t} = \lim_n \|u^n_{t_0}\|_{\mu^n_t} = \|u_{t_0}\|_{\mu_{t_0}} = \|u_t\|_{\mu_t}$ which gives the strong convergence of (u^n_t) to u_t. □

It is important to underline that hypothesis (i) is strictly needed, because it is not possible to deduce that the limit vector field is tangent assuming only that u_{t_0} is tangent for some $t_0 \in [0,1]$, not even if we assume strong convergence of (u^n_t) for every $t \in [0,1]$. The following is an explicit counterexample.

Example 6.27 (The limit vector field is not always tangent). *Observe that for a regular curve $t \to \mu_t$ and a given $u_0 \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$, the vector $u_0 \circ \mathbf{T}(1,0,\cdot)$ belongs to $L^2_{\mu_1}$ but is, in general, not tangent. So choose such a μ_t and u_0 and apply proposition 6.21 to approximate the curve with a sequence (μ^n_t) of regular curves with range on measures with finite support such that:*

$$\lim_{n\to\infty} \sup_{t\in[0,1]} W(\mu^n_t, \mu_t) = 0,$$
$$\mathrm{supp}(\mu^n_t) \subset \mathrm{supp}(\mu_t), \quad \forall t \in [0,1],$$
$$\frac{d}{dt}\mu^n_t + \nabla \cdot (v_t \mu^n_t) = 0,$$

where the v_t are the velocity vectors of $t \to \mu_t$.

It is clear that the (restriction of) the flow maps $\mathbf{T}(t,s,\cdot)$ of μ_t to $\mathrm{supp}(\mu^n_t)$ defines the flow maps of μ^n_t.

Now let $\boldsymbol{\gamma}^n \in \mathcal{O}pt(\mu_0, \mu^n_0)$ and define $u^n_0 := u_0 \circ (\boldsymbol{\gamma}^n)^{-1}$: it is easy to check that (u^n_0) strongly converges to u_0. Now let $u^n_t := u^n_0 \circ \mathbf{T}(t,0,\cdot)$ and observe that since $\mathrm{supp}\,\mu^n_t$ is finite, it holds $u^n_t \in \mathrm{Tan}_{\mu^n_t}(\mathscr{P}_2(\mathbb{R}^d))$, furthermore it is clear from the definitions that $t \to u^n_t$ is a parallel transport for any $n \in \mathbb{N}$.

It follows from the strong convergence of $(\mathbf{T}(t,0,\cdot), \mu^n_t)$ to $(\mathbf{T}(t,0,\cdot), \mu_t)$ that $(u^n_0 \circ \mathbf{T}(t,0,\cdot), \mu^n_t)$ strongly converge to $(u_0 \circ \mathbf{T}(t,0,\cdot), \mu_t)$ as $n \to \infty$.

Given that we assumed $u_0 \circ \mathbf{T}(1,0,\cdot) \notin \mathrm{Tan}_{\mu_1}(\mathscr{P}_2(\mathbb{R}^d))$ we have that the sequence of parallel transports $t \to u^n_t$ converges strongly for any t, but the limit vector field is tangent at $t = 0$ but not at $t = 1$.

In view if this example, it is interesting to know under which additional conditions the limit vector field is tangent: proposition 6.13 tells us that the two following hypothesis are both sufficient:

- For a.e. t and every n there exists a transport map T^n_t from μ_t to μ^n_t such that the sequence $(\mathrm{Lip}(T^n_t - Id))$ converges to 0 as $n \to \infty$,
- For a.e. t and every n there exists a regular mollifiers (ρ^n_t) (regular in the sense that the induced measure is in $\mathscr{P}^r_2(\mathbb{R}^d)$) such that $\mu^n_t = \mu_t * \rho^n_t$ and $W(\rho^n_t, \delta_0) \to 0$ as $n \to \infty$ for any t.

Now we turn to the proof of the existence of the parallel transport: $t \to \mu_t$ will be a fixed regular curve, v_t its velocity vector field and $\mathbf{T}(t,s,x)$ its flow.

In order to enlighten the notation we define

$$D(t,s) := \exp\left(\int_t^s \mathrm{Lip}(v_r)dr\right) - 1, \quad 0 \le t \le s \le 1,$$
$$D(t,s) := D(s,t) \qquad 0 \le s \le t \le 1.$$

Then we let τ_t^s be the "translation" from $L^2_{\mu_t}$ to $L^2_{\mu_s}$ given by $\mathbf{T}(t,s,\cdot)$:

$$\tau_t^s(f) := \tau_{\mathbf{T}(t,s,\cdot)}(f) = f \circ \mathbf{T}(t,s,\cdot).$$

Note that from the group property of $\mathbf{T}(t,s,\cdot)$ it follows

$$\tau_t^r = \tau_s^r \circ \tau_t^s, \quad \forall t,s,r \in [0,1]. \tag{6.25}$$

Moreover we define

$$\mathscr{P}_t^s(u) := P_{\mu_s}\Big(\tau_t^s(u)\Big).$$

Observe that the maps $\mathscr{P}_t^s$ are non-expansive and that by inequality (6.14) and proposition 6.6 we get:

$$\|\mathscr{P}_s^t(w)\|_{\mu_t} \le \|w\|_{\mu_s} D(t,s), \quad t,s \in [0,1], w \in \mathrm{Tan}^\perp_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d)), \tag{6.26a}$$

$$\|\tau_t^s(u) - \mathscr{P}_t^s(u)\|_{\mu_s} \le \|u\|_{\mu_t} D(t,s), \quad t,s \in [0,1], u \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)). \tag{6.26b}$$

To prove the existence of the transport we proceed as in the first section: let $\mathfrak{P}$ be the direct sets of all the partitions of $[0,1]$, where $\mathcal{Q} \ge \mathcal{P}$, $\mathcal{P}, \mathcal{Q} \in \mathfrak{P}$, if $\mathcal{Q}$ is a refinement of $\mathcal{P}$, and, for $\mathcal{P} = \{0 = t_0 \le t_1 \le \cdots \le t_N = 1\} \in \mathfrak{P}$ and $u \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ define $\mathcal{P}(u) \in \mathrm{Tan}_{\mu_1}(\mathscr{P}_2(\mathbb{R}^d))$ as:

$$\mathcal{P}(u) := \mathscr{P}^1_{t_{N-1}}(\mathscr{P}^{t_{N-1}}_{t_{N-2}}(\cdots \mathscr{P}^{t_1}_0(u))).$$

Finally, let $D^2(\mathcal{P}) := \sum_i D^2(t_i, t_{i+1})$.

We will prove first that there exists a unique limit $\mathcal{T}_0^1(u) \in \mathrm{Tan}_{\mu_1}(\mathscr{P}_2(\mathbb{R}^d))$ of $\mathcal{P}(u)$ as $\mathcal{P}$ varies in $\mathfrak{P}$, then we will define a curve $t \to u_t = \mathcal{T}_0^t(u) \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ by considering the restriction of $t \to \mu_t$ to $[0,t]$, and finally prove that this curve is the parallel transport of u along the curve $t \to \mu_t$.

Lemma 6.28. *It holds*

$$D(t_1,s_1) \le D(t_2,s_2), \quad \forall [t_1,s_1] \subset [t_2,s_2] \subset [0,1], \tag{6.27a}$$

$$\sum_{i=1}^{n-1} D(t_i, t_{t+1}) \le D(t,s), \quad t = t_1 \le \ldots \le t_n = s, \tag{6.27b}$$

$$\lim_{\mathcal{P} \in \mathfrak{P}} D^2(\mathcal{P}) = 0, \tag{6.27c}$$

$$\lim_{s \to t} \frac{D^2(t,s)}{|s-t|} = 0, \quad a.e.\ t \in [0,1]. \tag{6.27d}$$

Proof. Equation (6.27a) is clear. For (6.27b) we need to prove that $e^a - 1 + e^b - 1 \le e^{a+b} - 1$ for positive a, b, which is obvious.

The convexity of $r \to e^r - 1$ in $[0, \int_0^1 \mathrm{Lip}(v_t)dt]$ gives

$$D(t,s) \le \left(\frac{\int_0^1 \mathrm{Lip}(v_r)dr - 1}{\int_0^1 \mathrm{Lip}(v_r)dr}\right) \int_t^s \mathrm{Lip}(v_r)dr, \tag{6.28}$$

from which, taking the integrability of $\mathrm{Lip}(v_t)$ into account, (6.27d) follows at every Lebesgue point of $t \mapsto \mathrm{Lip}(v_t)$. Finally, from (6.28) we get

$$\sum_{i=0}^{N-1} D^2(t_{i+1}, t_i) \le C \sum_{i=0}^{N-1} \left(\int_{t_i}^{t_{i+1}} \mathrm{Lip}(v_r)dr\right)^2 \le C \max_i \left\{\int_{t_i}^{t_{i+1}} \mathrm{Lip}(v_r)\,dr\right\} \int_0^1 \mathrm{Lip}(v_r)\,dr,$$

from which (6.27c) follows, taking the absolute continuity property of the integral into account. □

LEMMA 6.29. *Let* $0 \le s_1 \le s_2 \le s_3 \le 1$ *and let* $u \in \mathrm{Tan}_{\mu_{s_1}}(\mathscr{P}_2(\mathbb{R}^d))$. *Then it holds:*

$$\left|\left|\mathscr{P}_{s_1}^{s_3}(u) - \mathscr{P}_{s_2}^{s_3}(\mathscr{P}_{s_1}^{s_2}(u))\right|\right|_{\mu_{s_3}} \le \|u\|_{\mu_{s_1}} D(s_1, s_2) D(s_2, s_3). \tag{6.29}$$

Proof. Observe that

$$\mathscr{P}_{s_1}^{s_3}(u) - \mathscr{P}_{s_2}^{s_3}(\mathscr{P}_{s_1}^{s_2}(u)) = \mathscr{P}_{s_2}^{s_3}(\tau_{s_1}^{s_2}(u) - \mathscr{P}_{s_1}^{s_2}(u)),$$

and that $\tau_{s_1}^{s_2}(u) - \mathscr{P}_{s_1}^{s_2}(u) \in \mathrm{Tan}^{\perp}_{\mu_{s_2}}(\mathscr{P}_2(\mathbb{R}^d))$. Therefore the thesis follows by a direct application of inequalities (6.26). □

COROLLARY 6.30. *Let* $\mathcal{P} = \{t = t_0 \le \cdots \le t_n = s\}$ *be a partition of* $[t,s] \subset [0,1]$ *and let* $\mathcal{Q}$ *be a refinement of* $\mathcal{P}$. *Then it holds*

$$\|\mathcal{P}(u) - \mathcal{Q}(u)\|_{\mu_1} \le \|u\|_{\mu_0} D^2(\mathcal{P}), \tag{6.30}$$

for every $u \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$.

Proof. Without loss of generality we may assume $[t,s] = [0,1]$. Fix $i < n$ such that (t_i, t_{i+1}) contains some element of $\mathcal{Q}$ and write $\mathcal{Q} \cap [t_i, t_{i+1}] = \{t_i = s_{i,0} < s_{i,1} < \cdots < s_{i,k(i)} = t_{i+1}\}$ for some $k(i) \ge 1$. Now, we claim that

$$\|\mathscr{P}_{s_{i,0}}^{s_{i,k(i)}}(u_{t_i}) - \mathscr{P}_{s_{i,k_i-1}}^{s_{i,k_i}}(\mathscr{P}_{s_{i,k_i-2}}^{s_{i,k_i-1}}(\cdots(\mathscr{P}_{s_{i,0}}^{s_{i,1}}(u_{t_i}))))\|_{\mu_{t_{i+1}}} \le \|u_{t_i}\|_{\mu_{t_i}} D^2(t_i, t_{i+1}) \tag{6.31}$$

for all $u_{t_i} \in \mathrm{Tan}_{\mu_{t_i}}(\mathscr{P}_2(\mathbb{R}^d))$. Indeed, the left hand side of (6.31) can be estimated by

$$\begin{aligned}
& \|\mathscr{P}_{s_{i,0}}^{s_{i,k(i)}}(u_{t_i}) - \mathscr{P}_{s_{i,k(i)-1}}^{s_{i_{k(i)}}}(\mathscr{P}_{s_{i,0}}^{s_{i,k(i)-1}}(u_{t_i}))\|_{\mu_{t_{i+1}}} \\
+ \quad & \|\mathscr{P}_{s_{i,k(i)-1}}^{s_{i,k(i)}}(\mathscr{P}_{s_{i,0}}^{s_{i,k(i)-1}}(u_{t_i})) - \mathscr{P}_{s_{i,k(i)-1}}^{s_{i,k(i)}}(\mathscr{P}_{s_{i,k(i)-2}}^{s_{i,k(i)-1}}(\cdots(\mathscr{P}_{s_{i,0}}^{s_{i,1}}(u_{t_i}))))\|_{\mu_{t_{i+1}}} \\
\le \quad & \|u_{t_i}\|_{\mu_{t_i}} D(s_{i,0}, s_{i,k(i)-1}) D(s_{i,k(i)-1}, s_{i,k(i)}) \\
+ \quad & \|\mathscr{P}_{s_{i,0}}^{s_{i,k(i)-1}}(u_{t_i}) - \mathscr{P}_{s_{i,k(i)-2}}^{s_{i,k(i)-1}}(\mathscr{P}_{s_{i,k(i)-3}}^{s_{i,k(i)-2}}(\cdots(\mathscr{P}_{t_i}^{s_{i,0}}(u_{t_i}))))\|_{\mu_{t_{i+1}}} \\
\le \quad & \cdots \\
\le \quad & \|u_{t_i}\|_{\mu_{t_i}} \sum_{j=0}^{k(i)-1} D(s_{i,0}, s_{i,j}) D(s_{i,j}, s_{i,j+1}) \le \|u_{t_i}\|_{\mu_{t_i}} D(t_i, t_{i+1}) \sum_{j=0}^{k(i)-1} D(s_{i,j}, s_{i,j+1}) \\
\le \quad & \|u_{t_i}\|_{\mu_{t_i}} D^2(t_i, t_{i+1}).
\end{aligned}$$

Now, let us assume that (t_0, t_1) contains some element of $\mathcal{Q}$ and let $\mathcal{P}' = [t_1, 1] \cap \mathcal{P}$, $\mathcal{Q}' = [t_1, 1] \cap \mathcal{Q}$, $u \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ and $v, w \in \mathrm{Tan}_{\mu_{t_1}}(\mathscr{P}_2(\mathbb{R}^d))$ be such that $\mathcal{P}(u) = \mathcal{P}'(v)$ and $\mathcal{Q}(u) = \mathcal{Q}'(w)$. Then, the inequality (6.31) with $i = 0$ reads

$$\|v - w\|_{\mu_{t_1}} \leq \|u\|_{t_0} D^2(t_0, t_1),$$

(the estimate is trivial if $\mathcal{Q}' \cap (t_0, t_1) = \emptyset$, because $v = w$) so that

$$\begin{aligned} \|\mathcal{P}(u) - \mathcal{Q}(u)|_{\mu_{t_n}} &\leq \|\mathcal{P}'(v) - \mathcal{Q}'(v)\|_{\mu_{t_n}} + \|\mathcal{Q}'(v) - \mathcal{Q}'(w)\|_{\mu_{t_n}} \\ &\leq \|\mathcal{P}'(v) - \mathcal{Q}'(v)\|_{\mu_{t_n}} + \|u\|_{t_0} D^2(t_0, t_1). \end{aligned}$$

Since $\|v\|_{t_1} \leq \|u\|_{t_0}$ we can apply repeatedly (6.31) in the intervals (t_i, t_{i+1}) to obtain $\|\mathcal{P}(u) - \mathcal{Q}(u)\|_{\mu_1} \leq \|u\|_{\mu_0} D^2(\mathcal{P})$. □

Theorem 6.31 (Existence of the limit of $\mathcal{P}(u_0)$). *Let $t \to \mu_t$ be a regular curve and $u_0 \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$. Then there exists the limit of $\mathcal{P}(u_0)$ as $\mathcal{P}$ varies in the direct set $\mathfrak{P}$.*

Proof. It follows directly from the previous corollary and from lemma 6.28. □

Define $\mathcal{T}_0^1(u_0)$ as the vector obtained by the limit process described above, and observe that by repeating the arguments to the restriction of $t \to \mu_t$ to the interval $[t, s]$, we can define a map $\mathcal{T}_t^s : \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \to \mathrm{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$. Furthermore, by considering the curve $t \to \mu_{1-t}$, we can define the maps $\mathcal{T}_t^s$ even for $t > s$.

Proposition 6.32 (Group property). *Let $t \to \mu_t$ be a regular curve and $\mathcal{T}_t^s : \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \to \mathrm{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$ be defined as above. Then*

$$\mathcal{T}_t^s \circ \mathcal{T}_r^t = \mathcal{T}_r^s, \quad \forall r, s, t \in [0, 1]. \tag{6.32}$$

Proof. Let us first assume $r \leq t \leq s$. In this case it is sufficient to observe that, by definition of limit over a direct set, to take the limit over all the partitions of $[r, s]$ is equal to take the limit over the set of partitions which contain the point t. The thesis then follows easily.

For the general case it is sufficient to prove that $\mathcal{T}_t^s = (\mathcal{T}_s^t)^{-1}$, or, without loss of generality, that $\mathcal{T}_0^1 = (\mathcal{T}_1^0)^{-1}$. The latter equation will follow if we show that

$$\lim_{\mathcal{P} \in \mathfrak{P}} \|u - \mathcal{P}^{-1}(\mathcal{P}(u))\|_{\mu_0} = 0, \quad \forall u \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d)), \tag{6.33}$$

where $\mathcal{P}^{-1} : \mathrm{Tan}_{\mu_1}(\mathscr{P}_2(\mathbb{R}^d)) \to \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ is defined as

$$\mathcal{P}^{-1}(u) := \mathscr{P}_{t_1}^0(\mathscr{P}_{t_2}^{t_1}(\cdots \mathscr{P}_1^{t_n-1}(u)))$$

for the partition $\mathcal{P} = \{0 = t_0 \leq \cdots t_n = 1\}$ (and in particular is *not* the functional inverse of $u \to \mathcal{P}(u)$).

Observe that for any $u \in \mathrm{Tan}_{\mu_{t_i}}(\mathscr{P}_2(\mathbb{R}^d))$ the identities $u = \mathscr{P}_{t_{i+1}}^{t_i}(\tau_{t_i}^{t_{i+1}}(u))$ and $\mathscr{P}_{t_i}^{t_{i+1}}(u) - \tau_{t_i}^{t_{i+1}}(u) \in \mathrm{Tan}_{\mu_{t_{i+1}}}^{\perp}(\mathscr{P}_2(\mathbb{R}^d))$, in conjunction with inequalities (6.26), yield

$$\begin{aligned} \|\mathscr{P}_{t_{i+1}}^{t_i}(\mathscr{P}_{t_i}^{t_{i+1}}(u)) - u\|_{\mu_{t_i}} &= \|\mathscr{P}_{t_{i+1}}^{t_i}(\mathscr{P}_{t_i}^{t_{i+1}}(u) - \tau_{t_i}^{t_{i+1}}(u))\|_{\mu_{t_i}} \\ &\leq \|\mathscr{P}_{t_i}^{t_{i+1}}(u) - \tau_{t_i}^{t_{i+1}}(u)\|_{\mu_{t_i}} D(t_i, t_{i+1}) \\ &\leq \|u\|_{\mu_{t_i}} D^2(t_i, t_{i+1}). \end{aligned}$$

For any $u \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ we obtain

$$\begin{aligned}\|u - \mathscr{P}^0_{t_1}(\cdots(\mathscr{P}^{t_{n-1}}_1(\mathcal{P}(u)))\|_{\mu_0} \\ &\le \|u - \mathscr{P}^0_{t_1}(\mathscr{P}^{t_1}_0(u))\|_{\mu_0} + \|\mathscr{P}^0_{t_1}(\mathscr{P}^{t_1}_0(u)) - \mathscr{P}^0_{t_1}(\cdots(\mathscr{P}^{t_{n-1}}_1(\mathcal{P}(u))))\|_{\mu_0} \\ &\le \|u\|_{\mu_0} D^2(0,t_1) + \|v - \mathscr{P}^{t_1}_{t_2}(\cdots(\mathscr{P}^1_{t_{n-1}}(\mathcal{P}'(v))))\|_{\mu_{t_1}},\end{aligned}$$

where $v = \mathscr{P}^{t_1}_0(u)$ and $\mathcal{P}' = \{t_1 < \cdots < t_n\}$ (so that $\mathcal{P}'(v) = \mathcal{P}(u)$). Since $\|v\|_{\mu_{t_1}} \le \|u\|_{\mu_0}$ we can continue in this way, to arrive at

$$\|u - \mathscr{P}^0_{t_1}(\cdots(\mathscr{P}^{t_{n-1}}_1(\mathcal{P}(u))))\|_{\mu_0} \le \|u\|_{\mu_0} D^2(\mathcal{P})$$

and this, taking (6.27c) into account, leads to (6.33). □

Proposition 6.33 (The limit process produces the parallel transport). *Let $t \to \mu_t$ be a regular curve, $u_0 \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ and let the maps $\mathcal{T}^s_t$ be defined as above. Then the vector field $t \to u_t := \mathcal{T}^t_0(u_0)$ is the parallel transport of u_0 along the curve.*

Proof. Consider any interval $[t,s] \subset [0,1]$, its trivial partition $\mathcal{P} = \{t,s\}$ and any (finer) partition Q. Applying inequality (6.30) and passing to the limit on $\mathcal{Q}$ we get

$$\|\mathscr{P}^s_t(u) - \mathcal{T}^s_t(u)\|_{\mu_s} \le \|u\|_{\mu_t} D^2(t,s), \quad \forall u \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)). \tag{6.34}$$

Coupling this equation with inequality (6.26b) we get

$$\begin{aligned}\|\tau^s_t(u) - \mathcal{T}^s_t(u)\|_{\mu_s} \le&\|\tau^s_t(u) - \mathscr{P}^s_t(u)\|_{\mu_s} + \|\mathscr{P}^s_t(u) - \mathcal{T}^s_t(u)\|_{\mu_s} \\ \le&\|u\|_{\mu_t} D(t,s)\,(1 + D(0,1))\,, \qquad \forall u \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)),\end{aligned}$$

from which it easily follows the absolute continuity of $t \to \mathcal{T}^t_0(u_0)$.

Now pick a Lebesgue point t of the function $t \to \mathrm{Lip}(v_t)$ and observe that inequality (6.34) gives

$$\lim_{s\to t} \frac{\|\mathscr{P}^s_t(u) - \mathcal{T}^s_t(u)\|_{\mu_s}}{|s-t|} = 0,$$

for every $u \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$, and so in particular for $u = \mathcal{T}^t_0(u)$, therefore to conclude it is sufficient to prove that

$$\lim_{s\to t} P_{\mu_t}\left(\frac{\tau^t_s(\mathscr{P}^s_t(u)) - u}{s-t}\right) = 0, \quad \forall u \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)).$$

Observe that $\mathscr{P}^s_t(u) - \tau^s_t(u) \in \mathrm{Tan}^\perp_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$, therefore from inequalities (6.26) we get

$$\begin{aligned}\|P_{\mu_t}(\tau^t_s(\mathscr{P}^s_t(u)) - u)\|_{\mu_t} &= \|\mathscr{P}^t_s(\mathscr{P}^s_t(u) - \tau^s_t(u))\|_{\mu_t} \\ &\le \|\mathscr{P}^s_t(u) - \tau^s_t(u)\|_{\mu_t} D(t,s) \le \|u\|_{\mu_t} D^2(t,s).\end{aligned}$$

□

Remark 6.34 (Parallel transport along a flow). *The above arguments works as well if, instead than assuming that the curve μ_t is regular, we assume the existence of a family of maps $X(t,s,x)$ satisfying $X(t,s,\cdot)_{\#}\mu_t = \mu_s$, having the group property $X(t,s,X(r,t,x)) = X(r,s,x)$ and such that the Lipschitz constant of $X(t,s,\cdot) - Id$ is bounded by a certain function $D(t,s)$ having the properties* (6.27) *(i.e. we just drop the condition that the velocity field of X is the* tangent *velocity vector of μ_t). The result would be the existence of a vector field $t \to u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ such that $t \to u_t \circ X(s,t,\cdot) \in L^2_{\mu_s}$ is absolutely continuous for any s and satisfying the equation*

$$P_{\mu_t}\left(\lim_{h\to 0}\frac{u_{t+h}\circ X(t,t+h,\cdot) - u_t}{h}\right) = 0,$$

for a.e. t. For such vector fields, it is possible to prove, with exactly the same arguments, uniqueness, linearity and conservation of scalar product.

We will show in the next section that our choice of the flows $\mathbf{T}$ related to the tangent vector field is more natural, because it induces the Levi-Civita connection on $\mathscr{P}_2(\mathbb{R}^d)$.

Before analyzing some examples of parallel transport, we want to point out that the construction we did of parallel tranport along a regular curve, allows a little generalization to the case of *forward* parallel transport along a *locally regular* curve. The question is the following. Consider an absolutely continuous curve μ_t on $[0,T]$ such that the function $t \to \mathrm{Lip}(v_t)$ belongs to $L^1_{\mathrm{loc}}((0,T])$ (we will see in example 6.39 that this is the case for a geodesic on $[0,1]$ for any $T \in (0,1)$). We say that $[0,T] \ni t \to u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ is a parallel transport if it is a parallel transport on the interval $(0,T]$ (which makes sense, due to the locality of the definition of parallel transport) and u_t strongly converge to u_0 as $t \to 0$. Having this definition in mind, two questions come out naturally: the first one is whether there exists the parallel transport along μ_t of a vector in $\mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$, which we call forward parallel transport, the second one is whether there exists the parallel transport of a vector in $\mathrm{Tan}_{\mu_T}(\mathscr{P}_2(\mathbb{R}^d))$, which we call backward parallel transport. We are going to prove here that the forward parallel transport always exists, while the same is *not* true for the backward one. This reflects what we already said at the end of section 2, where we proved that a sequence of tangent vectors doesn't necessarily converge to a tangent vector (and thus the backward parallel transport may not exist), and that every tangent vector may be approximated by tangent vectors (and thus the forward parallel transport exists).

For the proof of existence of the forward parallel transport, we will need the following technical result which is of its own interest.

Lemma 6.35. *Let $t \to \mu_t$ be a regular curve and $\mathcal{T}_t^s$ the optimal transport maps along it. Then it holds*

$$\|\mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi\|_{\mu_s} \le \mathrm{Lip}(\nabla\varphi)\mathcal{L}_t^s(\mu_r), \quad \forall \varphi \in C_c^\infty(\mathbb{R}^d), \tag{6.35}$$

where $\mathcal{L}_t^s(\mu_r)$ is the length of μ_r restricted to the interval $[t,s]$.

Proof. Observe that $s \to \mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi \in \mathrm{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$ is an absolutely continuous vector field along μ_t. The conclusion follows from the differential inequality:

$$\begin{aligned}
\frac{d}{ds}\|\mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi\|^2_{\mu_s} &= 2\langle \mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi, \frac{d}{ds}\big(\mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi\big)\rangle_{\mu_s} \\
&= 2\langle \mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi, \frac{d}{ds}\big(\mathcal{T}_t^s(\nabla\varphi)\big) - \nabla^2\varphi\cdot v_s\rangle_{\mu_s} \\
&= 2\langle \mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi, P_{\mu_s}\big(\frac{d}{ds}(\mathcal{T}_t^s(\nabla\varphi)) - \nabla^2\varphi\cdot v_s\big)\rangle_{\mu_s} \\
&= -2\langle \mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi, P_{\mu_s}(\nabla^2\varphi\cdot v_s)\rangle_{\mu_t} \\
&= -2\langle \mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi, \nabla^2\varphi\cdot v_s\rangle_{\mu_t} \\
&\le 2\|\mathcal{T}_t^s(\nabla\varphi) - \nabla\varphi\|_{\mu_s}\,\mathrm{Lip}(\nabla\varphi)\|v_s\|_{\mu_s}.
\end{aligned}$$

□

Proposition 6.36 (Forward parallel transport). *Let $[0,T] \ni t \to \mu_t$ be an absolutely continuous curve such that the function $\mathrm{Lip}(v_t)$ belongs to $L^1_{\mathrm{loc}}((0,T])$ and let $u_0 \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$. Then there exists a parallel transport $(0,T) \ni t \to u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ such that u_t converges strongly to u_0 as $t \to 0$.*

Proof. Start assuming that u_0 is the gradient of $\varphi \in C_c^\infty(\mathbb{R}^d)$. Fix $\varepsilon > 0$, think $\nabla\varphi$ as a vector in $\mathrm{Tan}_{\mu_\varepsilon}(\mathscr{P}_2(\mathbb{R}^d))$ and define the vectors $u_t^\varepsilon := \mathcal{T}_\varepsilon^t(\nabla\varphi)$ for any $t \in [\varepsilon, T]$, so that we have $u_\varepsilon^\varepsilon = \nabla\varphi$. From

$$\|u_t^{\varepsilon'} - u_t^\varepsilon\|_{\mu_t} = \|u_\varepsilon^{\varepsilon'} - u_\varepsilon^\varepsilon\|_{\mu_\varepsilon} = \|\mathcal{T}_{\varepsilon'}^\varepsilon(\nabla\varphi) - \nabla\varphi\|_{\mu_\varepsilon} \le \mathrm{Lip}(\nabla\varphi)\omega(\varepsilon) \qquad 0 < \varepsilon' \le \varepsilon \le t \le T,$$

with $\omega(\varepsilon) := \int_0^\varepsilon \|v_t\|_{\mu_t}dt$, we get that for any t, the family $\{u_t^\varepsilon\}$ converges in $\mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$, as $\varepsilon \to 0$, to a vector u_t satisfying $\|u_t^\varepsilon - u_t\|_{\mu_t} \le \mathrm{Lip}(\nabla\varphi)\omega(\varepsilon)$. The limit vector field u_t is easily seen to be a parallel transport in the interval $(0,T]$: indeed from the uniform bound (6.24) we get its local absolute continuity, and we conclude by the stability of the solutions of (6.23).

From

$$\|u_t\|_{\mu_t} = \lim_\varepsilon \|u_t^\varepsilon\|_{\mu_t} = \lim_\varepsilon \|u_\varepsilon^\varepsilon\|_{\mu_\varepsilon} = \lim_\varepsilon \|\nabla\varphi\|_{\mu_\varepsilon}$$

we get that the norm of u_t is constant, and equal to $\|\nabla\varphi\|_{\mu_0}$. Finally it holds

$$\langle u_\varepsilon, \eta\rangle_{\mu_\varepsilon} = \langle u_\varepsilon - u_\varepsilon^\varepsilon, \eta\rangle_{\mu_\varepsilon} + \langle u_\varepsilon^\varepsilon, \eta\rangle_{\mu_\varepsilon} = R_\varepsilon + \langle \nabla\varphi, \eta\rangle_{\mu_\varepsilon} \qquad \forall \eta \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d),$$

where the term R_ε is bounded by $\|u_t - u_t^\varepsilon\|_{\mu_t}\sup|\eta| \le \omega(\varepsilon)\,\mathrm{Lip}(\nabla\varphi)\sup|\eta|$.

For the general case, just approximate u_0 with smooth gradients u_0^n, apply the construction above to obtain the existence of forward parallel transports $t \to u_t^n$ of u_0^n and use the fact that (clearly) $\|u_t^n - u_t^m\|_{\mu_t} = \|u_0^n - u_0^m\|_{\mu_0}$ to get that for any t the sequence (u_t^n) strongly converges to some u_t such that $\|u_t\|_{\mu_t} = \|u_0\|_{\mu_0}$. By the stability argument used above we get that $t \to u_t$ is a parallel transport on $(0,T]$, so we need just to prove that u_t weakly converges to u_0 as $t \to 0$. To prove this, observe that since $[0,T] \ni t \to u_t^n$ is a forward parallel transport, passing to the limit as $t \to 0$

in (6.23) we get that for every $\eta \in C_c^\infty(\mathbb{R}^d)$ it holds:

$$\langle u_t^n, \nabla\eta\rangle_{\mu_t} - \langle u_0^n, \nabla\eta\rangle_{\mu_0} = \int_0^t \langle u_r^n, \nabla^2\eta \cdot v_r\rangle_{\mu_r} dr \leq \|u_t^n\|_{\mu_t} \operatorname{Lip}(\nabla\eta)\omega(t).$$

Letting $n \to \infty$ in the above inequality the weak convergence follows. □

Now we turn to the counterexample to the existence of the backward parallel transport. As we will see, it is possible that a parallel transport u_t exists for positive t, that the vectors u_t converge strongly to some vector u_0 as $t \to 0$, but such that the vector u_0 is *not* a tangent vector: this shows that the general problem of existence of the parallel transport is, in general, intrisically prohibited by the geometry of $\mathscr{P}_2(\mathbb{R}^d)$. Observe that the curve considered is a geodesic.

Example 6.37. *Let $Q = [0,1]\times[0,1]$ be the unit square in $\mathbb{R}^2$ and let T_i, $i = 1,2,3,4$, be the four open triangles in which Q is divided by its diagonals. Let $\mu_0 := \chi_Q\mathscr{L}^2$ and define the function $v : Q \to \mathbb{R}^2$ as the gradient of the convex map* $\max\{|x|,|y|\}$*, as in the figure. Set also $u = v^\perp$, the rotation by $\pi/2$ of v, in Q and $u = 0$ out of Q. Notice that u is orthogonal to $\operatorname{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, since it holds $\nabla\cdot(u\mu) = 0$.*

Set $\mu_t := (Id+tv)_\#\mu_0$ and observe that, for positive t, the support Q_t of μ_t is made of 4 connected components, each one the translation of one of the sets T_i, and that $\mu_t = \chi_{Q_t}\mathscr{L}^2$.

It is immediate to check that the velocity vectors of μ_t are given by $v_t := v\circ(Id + tv)^{-1}$, so that $\operatorname{Lip}(v_t) = t^{-1}$ and μ_t is locally regular in $(0,1)$, and that the flow maps of μ_t in $(0,1]$ are given by

$$\mathbf{T}(t,s,\cdot) = (Id + sv)\circ(Id + tv)^{-1}, \quad \forall t,s \in (0,1].$$

Now, set $u_t := u\circ\mathbf{T}(t,0,\cdot)$ and notice that u_t is tangent at μ_t, because u_t is constant in the connected components of the support of μ_t. Since $u_{t+h}\circ\mathbf{T}(t,t+h,\cdot) = u_t$, we obtain that u_t is a parallel transport in $(0,1]$. Furthermore, since u_t converges to u as $t \to 0$ and $u \notin \operatorname{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, there is no way to extend u_t to a continuous tangent *vector field on the whole* $[0,1]$.

In the rest of the section we analyze some examples of parallel transport.

Example 6.38 (Equation in the smooth case). *Assume that $u_t(x) = \nabla\varphi(t,x)$ for a certain function $\varphi \in C^2([0,1]\times\mathbb{R}^d,\mathbb{R})$ with uniformly bounded second derivatives. Then it is easy to see that equation (6.21) becomes:*

$$\nabla\cdot\Big(\mu_t(\partial_t\nabla\varphi + \nabla_x^2\varphi\cdot v_t)\Big) = 0. \tag{6.36}$$

or, which is the same

$$\int_0^1\int \langle\partial_t\nabla\varphi + \nabla_x^2\varphi\cdot v_t, \nabla\eta\rangle d\mu_t dt = 0, \quad \forall\eta \in C_c^\infty((0,1)\times\mathbb{R}^d,\mathbb{R}) \tag{6.37}$$

Observe that these equations are valid without any regularity assumption on the measures involved.

EXAMPLE 6.39 (Geodesics). *Consider a geodesic* $t \to \mu_t$ *defined on the interval* $[0,1]$*: we want to prove that in any interval of the form* $[\varepsilon, 1-\varepsilon]$ *it is regular and its velocity vectors are a parallel transport on each interval of this kind.*

Fix $t \in (0,1)$*; we know from proposition 2.11 that there exists only one optimal plan between* μ_t *and* μ_1 *and that this plan is induced by a Lipschitz map* T_t *with Lipschitz constant bounded by* t^{-1}*. We know also that for* $s \in [t,1]$ *it holds* $\mu_s = (Id + \frac{s-t}{1-t}(T_t - Id))_{\#}\mu_t$*, the transport map being optimal. Calculating the velocity vector* v_t *as limit of the optimal transport maps, we get* $v_t = (1-t)^{-1}(T_t - Id)$*, therefore its Lipschitz constant is bounded by* $(1+t)(t(1-t))^{-1}$ *(actually, since* T_t *is monotone, it can be proved that* $\mathrm{Lip}(v_t) \leq (t(1-t))^{-1}$*). Our claim on the regularity of* μ_t *in* $[\varepsilon, 1-\varepsilon]$ *follows.*

Now assume that the geodesic $[0,1] \ni t \to \mu_t$ *is induced by a Lipschitz optimal map* T*. In this case its flow is given by*

$$\mathbf{T}(t,s,\cdot) = (Id + s(T - Id)) \circ (Id + t(T - Id))^{-1},$$

and the velocity vectors satisfy

$$v_s = v_t \circ T(t,s,\cdot),$$

therefore a direct calculation shows that v_t *is a parallel transport.*

EXAMPLE 6.40 (Constant vector fields). *Let* $t \to \mu_t$ *be a regular curve,* $v \in \mathbb{R}^d$ *and* C_v *the function on* $\mathbb{R}^d$ *constantly equal to* v*. Define* $u_0 := C_v \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$*. We claim that the parallel transport* u_t *of* u_0 *along* μ_t *is given by* $u_t = C_v$*, for all* $t \in [0,1]$*.*

The proof is immediate: it is sufficient to observe that $u_t(x) = \nabla\varphi(t,x)$*, where* $\varphi(t,x) = \langle x, v\rangle$*, and to verify that* φ *satisfies equation* (6.36).

EXAMPLE 6.41 (Translations). *Choose* $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$ *and an absolutely continuous curve* $[0,1] \ni t \to \gamma(t) \in \mathbb{R}^d$ *such that* $\gamma(0) = 0$*. As before, for any vector* $v \in \mathbb{R}^d$*, let* $C_v : \mathbb{R}^d \to \mathbb{R}^d$ *be the function constantly equal to* v*. Define* $\mu_t := (Id + C_{\gamma(t)})_{\#}\mu_0$*, so that* μ_t *is the translation of* μ_0 *by the vector* $\gamma(t)$*.*

It is clear that in this case the curve μ_t *is regular and that its flow maps are given by*

$$\mathbf{T}(t,s,x) = x + \gamma(s) - \gamma(t). \tag{6.38}$$

Furthermore, it is easy to check that the composition with $\mathbf{T}(t,s,x)$ *defines an isometry from* $\mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ *to* $\mathrm{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$*, therefore (by construction) the parallel transport maps coincide with the composition with the flow maps: i.e. for a given* $u_0 \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$*, the parallel tranport* u_t *of* u_0 *along* μ_t *is given by*

$$u_t(x) = u_0(x - \gamma(t)). \tag{6.39}$$

EXAMPLE 6.42 (Separate sets). *The parallel transport behaves independently on different connected components of* $\operatorname{supp}\mu_t$*. Indeed, recall remark 6.17 and observe that, with the same notation, a simple cut-off argument shows that* $u \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ *if and only if* $u \in \mathrm{Tan}_{\mu_t^i}(\mathscr{P}_2(\mathbb{R}^d))$ *for any* $i = 1, \ldots, n$*. Therefore a straightforward application* *of the definition of parallel transport gives that* $t \to u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ *is a parallel transport along* μ_t *if and only if* $t \to u_t \in \mathrm{Tan}_{\mu_t^i}(\mathscr{P}_2(\mathbb{R}^d))$ *is a parallel transport along* μ_t^i *for any* $i = 1, \dots, n$.

5. Covariant derivative and curvature operator

In this section we analyze the covariant derivative and the curvature tensor induced by the parallel transport we defined.

It is well known that, in the classical Riemannian setting, the definition of parallel transport leads to the one of covariant derivative via the formula

$$\nabla_{\dot\gamma(t)} u(t) := \lim_{s \to t} \frac{T_s^t(u(s)) - u(t)}{s - t}, \tag{6.40}$$

where $\gamma(t)$ is a smooth curve, $u(t) \in T_{\gamma(t)}M$ is a smooth vector field and T_s^t, for any s, t, is the parallel transport map from $T_{\gamma(s)}M$ to $T_{\gamma(t)}M$ along γ.

The same construction may be used in the Wasserstein setting:

Definition 6.43 (Covariant derivative). *Let* μ_t *be a regular curve, let* $v_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ *be its velocity vector and let* $u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ *be an absolutely continuous vector field along* μ_t. *The covariant derivative of* u_t *along* μ_t *is:*

$$\nabla_{v_t} u_t := \lim_{s \to t} \frac{\mathcal{T}_s^t(u_s) - u_t}{s - t},$$

where $\mathcal{T}_s^t$ *are the parallel transport maps along* μ_t *and the derivative takes place in* $L^2_{\mu_t}$.

Using the definition of absolutely continuous vector field, it is not difficult to check that the covariant derivative exists for a.e. t and that the function $t \mapsto \|\nabla_{v_t} u_t\|_{\mu_t}$ is integrable. Indeed, inequality (6.34) implies that the covariant derivative satisfies:

$$\nabla_{v_t} u_t = \lim_{s \to t} \frac{\mathscr{P}_s^t(u_s) - u_t}{s - t} = P_{\mu_t}\left(\lim_{s \to t} \frac{u_s \circ \mathbf{T}(t, s, \cdot) - u_t}{s - t} \right). \tag{6.41}$$

If the vector field u_t is given by the gradient of smooth functions, i.e. if $u_t = \nabla_x \varphi_t(x)$ for some $\varphi_t \in C_c^\infty(\mathbb{R}^d)$ smoothly varying in time, the previous equation reads as

$$\nabla_{v_t} u_t = P_{\mu_t}\big(\partial_t \nabla \varphi_t + \nabla_x^2 \varphi_t \cdot v_t\big). \tag{6.42}$$

Equation (6.42) and the analogous one (6.36) were first given in [**43**], although from a formal viewpoint and under stronger assumptions on the measures μ_t.

Having defined the covariant derivative, our first goal is to prove that it is the Levi-Civita connection on $(\mathscr{P}_2(\mathbb{R}^d), W_2)$. Recalling the discussion made for the classical case of Riemanniann manifolds, we need to prove that it is *compatible with the metric* and *torsion-free.* The compatibility with the metric is a simple consequence of the definition: indeed, for a given couple of absolutely continuous vector fields u_t^1, $u_t^2 \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$

along the regular curve μ_t, we have:

$$
\begin{aligned}
\frac{d}{dt}\langle u_t^1, u_t^2\rangle_{\mu_t} &= \langle \frac{d}{dt}u_t^1, u_t^2\rangle_{\mu_t} + \langle u_t^1, \frac{d}{dt}u_t^2\rangle_{\mu_t} \\
&= \langle P_{\mu_t}\left(\frac{d}{dt}u_t^1\right), u_t^2\rangle_{\mu_t} + \langle u_t^1, P_{\mu_t}\left(\frac{d}{dt}u_t^2\right)\rangle_{\mu_t} \\
&= \langle \nabla_{v_t} u_t^1, u_t^2\rangle_{\mu_t} + \langle u_t^1, \nabla_{v_t} u_t^2\rangle_{\mu_t},
\end{aligned} \tag{6.43}
$$

having used the Leibniz rule (6.15) and the fact that both vector fields are tangent.

To prove the torsion-free identity, we need first to understand how to calculate the Lie bracket of two vector fields. To this aim, let μ_t^i, $i = 1, 2$, be two regular curves such that $\mu_0^1 = \mu_0^2 =: \mu$ and let $u_t^i \in \mathrm{Tan}_{\mu_t^i}(\mathscr{P}_2(\mathbb{R}^d))$ be two absolutely continuous vector fields along μ_t^i satisfying $u_0^i = v_0^{1-i}$, where v_t^i are the tangent velocity fields of μ_t^i. We assume that both the velocity fields v_t^i of μ_t^i and the derivative of u_t^i exists for *all* t and are continuous in time w.r.t. the strong convergence of maps, so that the Leibniz rule (6.15) holds for *all* the times (and not just as functional equality) and the initial condition makes sense.

Let us consider vector fields as derivations, and the functional $\mu \mapsto F_\eta(\mu) := \int \eta d\mu$, for $\eta \in C_c^\infty(\mathbb{R}^d)$ fixed. By the continuity equation, the derivative of F_η along u_t^2 is equal to $\langle \nabla\eta, u_t^2\rangle_{\mu_t^2}$, therefore from the fact that $u_0^1 = v_0^2$ we get:

$$
\begin{aligned}
u^1(u^2(F_\eta))(\mu) &= \frac{d}{dt}\langle \nabla\eta, u_t^2\rangle_{\mu_t^2}|_{t=0} = \langle \nabla^2\eta \cdot v_0^2, u_0^2\rangle_\mu + \langle \nabla\eta, \frac{d}{dt}u_t^2|_{t=0}\rangle_\mu \\
&= \langle \nabla^2\eta \cdot u_0^1, u_0^2\rangle_\mu + \langle \nabla\eta, \nabla_{v_0^2} u_t^2\rangle_\mu.
\end{aligned}
$$

Subtracting the analogous term $u^2(u^1(F_\eta))(\mu)$ and using the symmetry of $\nabla^2\eta$ and the identities $u_0^i = v_0^{1-i}$, $i = 0, 1$, we get

$$
[u^1, u^2](F_\eta)(\mu) = \langle \nabla\eta, \nabla_{u_0^1} u_t^2 - \nabla_{u_0^2} u_t^1\rangle_\mu.
$$

Given that the set $\{\nabla\eta\}_{\eta\in C_c^\infty}$ is dense in $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, the above equation characterizes $[u^1, u^2]$ as:

$$
[u^1, u^2] = \nabla_{u_0^1} u_t^2 - \nabla_{u_0^2} u_t^1, \tag{6.44}
$$

which proves the torsion-free identity for the covariant derivative.

In the case of the parallel transport along a flow, considered in Remark 6.34, given that the right composition with $X(t, s, \cdot)$ is an isometry from $L^2_{\mu_t}$ to $L^2_{\mu_s}$, it holds

$$
\langle u_s^1, u_s^2\rangle_{\mu_s} = \langle u_s^1 \circ X(t, s, \cdot), u_s^2 \circ X(t, s, \cdot)\rangle_{\mu_t},
$$

subtracting $\langle u_t^1, u_t^2\rangle_{\mu_t}$, dividing both terms by $s - t$ and letting $s \to t$ we get that the Leibniz rule holds even using the maps $X(t, s, \cdot)$:

$$
\frac{d}{dt}\langle u_t^1, u_t^2\rangle_{\mu_t} = \langle \frac{d}{ds}u_s^1 \circ X(t, s, \cdot)|_{s=t}, u_t^2\rangle_{\mu_t} + \langle u_t^1, \frac{d}{ds}u_s^2 \circ X(t, s, \cdot)|_{s=t}\rangle_{\mu_t},
$$

for any couple of vector fields u_t^i such that $t \mapsto u_t^i \circ X(s,t,\cdot)$ is absolutely continuous for $i = 1, 2$. From this formula it follows that the parallel transport along *any* flow X compatible with μ_t preserves the scalar product.

Of course, different parallel transports define different covariant derivatives $\tilde{\nabla}_{v_t} u_t$ via (6.41): they are expressed by

$$\tilde{\nabla}_{v_t} u_t := P_{\mu_t}\left(\frac{d}{ds} u_s \circ X(t,s,\cdot)_{|s=t}\right).$$

Denoting by $\tilde{v}_t$ the velocity field of X, we get that the covariant derivative of the vector field $u_t := \nabla\varphi$, $\varphi \in C_c^\infty(\mathbb{R}^d)$, is given by $\tilde{\nabla}_{v_t}\nabla\varphi = P_{\mu_t}(\nabla^2\varphi \cdot \tilde{v}_t)$. It is easy to check that a generic covariant derivative is not torsion-free. Indeed, assume that it is and observe that in this case the two following equations hold:

$$\langle \nabla_{\nabla\varphi^1}\nabla\varphi^2, \nabla\varphi^3\rangle_\mu + \langle \nabla\varphi^2, \nabla_{\nabla\varphi^1}\nabla\varphi^3\rangle_\mu = \langle \tilde{\nabla}_{\nabla\varphi^1}\nabla\varphi^2, \nabla\varphi^3\rangle_\mu + \langle \nabla\varphi^2, \tilde{\nabla}_{\nabla\varphi^1}\nabla\varphi^3\rangle_\mu,$$
$$\nabla_{\nabla\varphi^1}\nabla\varphi^2 - \nabla_{\nabla\varphi^2}\nabla\varphi^1 = \tilde{\nabla}_{\nabla\varphi^1}\nabla\varphi^2 - \tilde{\nabla}_{\nabla\varphi^2}\nabla\varphi^1,$$

for any $\varphi^i \in C_c^\infty(\mathbb{R}^d)$, $i = 1, 2, 3$. From these equalities, with some algebraic manipulations (more explicitely, by following the calculations indicated in the Koszul formula), it follows that $\langle \nabla_{\nabla\varphi^1}\nabla\varphi^2, \nabla\varphi^3\rangle_\mu = \langle \tilde{\nabla}_{\nabla\varphi^1}\nabla\varphi^2, \nabla\varphi^3\rangle_\mu$, so that the two covariant derivatives coincide.

It remains to prove that if a different flow X induces the same covariant derivative of the flow $\mathbf{T}$, then $X = T$. To prove this observe that in the identity

$$\langle \nabla^2\varphi \cdot \tilde{v}_t, \nabla\eta\rangle_{\mu_t} = \langle \nabla^2\varphi \cdot v_t, \nabla\eta\rangle_{\mu_t} \qquad \forall \varphi, \eta \in C_c^\infty(\mathbb{R}^d)$$

we can use test functions $\varphi \in C^2(\mathbb{R}^d)$ with uniformly bounded second derivatives (by a simple approximation argument based on the finiteness of the second moments of μ_t). Choosing $\varphi(x) = |\langle x, \xi\rangle|^2$ gives

$$\int \frac{\partial\eta}{\partial\xi}\langle \tilde{v}_t - v_t, \xi\rangle d\mu_t = 0 \qquad \forall \eta \in C_c^\infty(\mathbb{R}^d),\ \xi \in \mathbb{R}^d.$$

This means the symmetric part of the distributional derivative of the vector-valued distribution $(\tilde{v}_t - v_t)\mu_t$ vanishes; Korn's inequality gives that the distribution is equivalent to a constant. By integrability, this constant must be 0, i.e. $\tilde{v}_t = v_t$ μ_t-a.e. in $\mathbb{R}^d$.

The definition of covariant derivative allows us to define the curvature tensor and to check, at least formally, that $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ is positively curved by proving that its sectional curvatures are always non-negative. The spirit of the foregoing discussion and the calculations we do, are basically borrowed from Lott's work [**43**].

Given four vector fields $\mu \mapsto \nabla\varphi_\mu^i \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, $i = 1, \ldots, 4$, the curvature tensor $\mathcal{R}$ calculated on them is defined as:

$$\begin{aligned}\langle \mathcal{R}(\nabla\varphi_\mu^1, \nabla\varphi_\mu^2)(\nabla\varphi_\mu^3), \nabla\varphi_\mu^4\rangle_\mu :&= \langle \nabla_{\nabla\varphi_\mu^1}(\nabla_{\nabla\varphi_\mu^2}\nabla\varphi_\mu^3), \nabla\varphi_\mu^4\rangle_\mu \\ &- \langle \nabla_{\nabla\varphi_\mu^2}(\nabla_{\nabla\varphi_\mu^1}\nabla\varphi_\mu^3), \nabla\varphi_\mu^4\rangle_\mu \\ &- \langle \nabla_{[\nabla\varphi_\mu^1, \nabla\varphi_\mu^2]}\nabla\varphi_\mu^3, \nabla\varphi_\mu^4\rangle_\mu.\end{aligned}$$

With the same calculation used in the classical Riemannian case, it is easy to check that $\mathcal{R}$ is actually a tensor, i.e. that its value at the measure μ depends only on the value of the four vector fields at μ. Therefore in order to evaluate it, we can consider the simpler vector fields $\mu \mapsto \nabla\varphi^i \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, $i = 1, \ldots, 4$, where the functions φ^i do not depend on the base measure μ. This will simplify the calculations. Under this assumption we have

$$\nabla_v \nabla\varphi = P_\mu(\nabla^2\varphi \cdot v) \qquad \forall v \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)). \tag{6.45}$$

In order to give an explicit formula for $\mathcal{R}$, it is useful to introduce the function $\xi_\mu(\varphi^1, \varphi^2) \in L^2_\mu$ as

$$\xi_\mu(\varphi^1, \varphi^2) := P_\mu^\perp(\nabla^2\varphi^1 \cdot \nabla\varphi^2) = \nabla^2\varphi^1 \cdot \nabla\varphi^2 - \nabla_{\nabla\varphi^2}\nabla\varphi^1(\mu).$$

Observe that from $\nabla^2\varphi^1 \cdot \nabla\varphi^2 + \nabla^2\varphi^2 \cdot \nabla\varphi^1 = \nabla(\langle\nabla\varphi^1, \nabla\varphi^2\rangle) \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ we get $\xi_\mu(\varphi^1, \varphi^2) = -\xi_\mu(\varphi^2, \varphi^1)$.

Proposition 6.44. *The curvature tensor is given by*

$$\begin{aligned}\langle\mathcal{R}(\nabla\varphi^1, \nabla\varphi^2)(\nabla\varphi^3), \nabla\varphi^4\rangle_\mu =& \langle\xi_\mu(\varphi^1, \varphi^4), \xi_\mu(\varphi^2, \varphi^3)\rangle_\mu - \langle\xi_\mu(\varphi^1, \varphi^3), \xi_\mu(\varphi^2, \varphi^4)\rangle_\mu \\ &- 2\langle\xi_\mu(\varphi^1, \varphi^2), \xi_\mu(\varphi^3, \varphi^4)\rangle_\mu.\end{aligned}$$

Proof. Define $\mu_t := (Id + t\nabla\varphi^1)_\#\mu$ and $F(\nu) := \int \eta\, d\nu$ with $\eta := \langle\nabla^2\varphi^3 \cdot \nabla\varphi^2, \nabla\varphi^4\rangle$. Evaluate the derivative at $t = 0$ of $F(\mu_t)$ to get

$$\frac{d}{dt}F(\mu_t)_{|t=0} = \frac{d}{dt}\int \eta \circ (Id + t\nabla\varphi^1)\, d\mu_{|t=0} = \langle\nabla\eta, \nabla\varphi^1\rangle_\mu.$$

On the other hand, using equations (6.45) and (6.43) we have

$$\begin{aligned}\frac{d}{dt}F(\mu_t)_{|t=0} &= \frac{d}{dt}\langle\nabla^2\varphi^3 \cdot \nabla\varphi^2, \nabla\varphi^4\rangle_{\mu_t|t=0} \\ &= \frac{d}{dt}\langle\nabla_{\nabla\varphi^2}\nabla\varphi^3(\mu_t), \nabla\varphi^4\rangle_{\mu_t|t=0} \\ &= \langle\nabla_{\nabla\varphi^1}(\nabla_{\nabla\varphi^2}\nabla\varphi^3), \nabla\varphi^4\rangle_\mu + \langle\nabla_{\nabla\varphi^2}\nabla\varphi^3, \nabla_{\nabla\varphi^1}\nabla\varphi^4\rangle_\mu.\end{aligned}$$

Coupling the last two equations and then using the trivial identity $\langle P_\mu(v), P_\mu(w)\rangle_\mu = \langle v, w\rangle_\mu - \langle P_\mu^\perp(v), P_\mu^\perp(w)\rangle_\mu$, valid for any v, $w \in L^2_\mu$, we obtain the equality

$$\begin{aligned}\langle\nabla_{\nabla\varphi^1}(\nabla_{\nabla\varphi^2}\nabla\varphi^3), \nabla\varphi^4\rangle_\mu &= \langle\nabla\big(\langle\nabla^2\varphi^3 \cdot \nabla\varphi^2, \nabla\varphi^4\rangle\big), \nabla\varphi^1\rangle_\mu - \langle\nabla_{\nabla\varphi^2}\nabla\varphi^3, \nabla_{\nabla\varphi^1}\nabla\varphi^4\rangle_\mu \\ &= \langle\nabla\big(\langle\nabla^2\varphi^3 \cdot \nabla\varphi^2, \nabla\varphi^4\rangle\big), \nabla\varphi^1\rangle_\mu - \langle\nabla^2\varphi^3 \cdot \nabla\varphi^2, \nabla^2\varphi^4 \cdot \nabla\varphi^1\rangle_\mu \\ &+ \langle\xi_\mu(\varphi^3, \varphi^2), \xi_\mu(\varphi^4, \varphi^1)\rangle_\mu.\end{aligned}$$

The computation of the gradient of $\langle\nabla^2\varphi^3 \cdot \nabla\varphi^2, \nabla\varphi^4\rangle$ gives

$$\begin{aligned}\langle\nabla_{\nabla\varphi^1}(\nabla_{\nabla\varphi^2}\nabla\varphi^3), \nabla\varphi^4\rangle_\mu &= \int \nabla^3\varphi^3(\nabla\varphi^2, \nabla\varphi^4, \nabla\varphi^1)\, d\mu + \langle\nabla^2\varphi^3 \cdot \nabla\varphi^4, \nabla^2\varphi^2 \cdot \nabla\varphi^1\rangle_\mu \\ &+ \langle\xi_\mu(\varphi^3, \varphi^2), \xi_\mu(\varphi^4, \varphi^1)\rangle_\mu.\end{aligned} \tag{6.46}$$

Analogously, it holds:

$$\langle \nabla_{\nabla\varphi^2}(\nabla_{\nabla\varphi^1}\nabla\varphi^3), \nabla\varphi^4\rangle_\mu = \int \nabla^3\varphi^3(\nabla\varphi^1, \nabla\varphi^4, \nabla\varphi^2)d\mu + \langle \nabla^2\varphi^3 \cdot \nabla\varphi^4, \nabla^2\varphi^1 \cdot \nabla\varphi^2\rangle_\mu + \langle \xi_\mu(\varphi^3, \varphi^1), \xi_\mu(\varphi^4, \varphi^2)\rangle_\mu, \tag{6.47}$$

so that, subtracting (6.47) from (6.46), the symmetry of $\nabla^3\varphi^3$ gives

$$\begin{aligned} &\langle \nabla_{\nabla\varphi^1}(\nabla_{\nabla\varphi^2}\nabla\varphi^3), \nabla\varphi^4\rangle_\mu - \langle \nabla_{\nabla\varphi^2}(\nabla_{\nabla\varphi^1}\nabla\varphi^3), \nabla\varphi^4\rangle_\mu \\ &= \langle \nabla^2\varphi^3 \cdot \nabla\varphi^4, \nabla^2\varphi^2 \cdot \nabla\varphi^1\rangle_\mu - \langle \nabla^2\varphi^3 \cdot \nabla\varphi^4, \nabla^2\varphi^1 \cdot \nabla\varphi^2\rangle_\mu \\ &+ \langle \xi_\mu(\varphi^3, \varphi^2), \xi_\mu(\varphi^4, \varphi^1)\rangle_\mu - \langle \xi_\mu(\varphi^3, \varphi^1), \xi_\mu(\varphi^4, \varphi^2)\rangle_\mu. \end{aligned} \tag{6.48}$$

Recalling equation (6.44) we get

$$\begin{aligned} \langle \nabla_{[\nabla\varphi^1, \nabla\varphi^2]}\nabla\varphi^3, \nabla\varphi^4\rangle_\mu &= \langle \nabla^2\varphi^3 \cdot P_\mu(\nabla^2\varphi^2 \cdot \nabla\varphi^1 - \nabla^2\varphi^1 \cdot \nabla\varphi^2), \nabla\varphi^4\rangle_\mu \\ &= \langle P_\mu(\nabla^2\varphi^2 \cdot \nabla\varphi^1 - \nabla^2\varphi^1 \cdot \nabla\varphi^2), \nabla^2\varphi^3 \cdot \nabla\varphi^4\rangle_\mu \\ &= \langle \nabla^2\varphi^3 \cdot \nabla\varphi^4, \nabla^2\varphi^2 \cdot \nabla\varphi^1 - \nabla^2\varphi^1 \cdot \nabla\varphi^2\rangle_\mu \\ &- \langle \xi_\mu(\varphi^2, \varphi^1), \xi_\mu(\varphi^3, \varphi^4)\rangle_\mu + \langle \xi_\mu(\varphi^1, \varphi^2), \xi_\mu(\varphi^3, \varphi^4)\rangle_\mu. \end{aligned}$$

Subtracting the last equations from (6.48), all the terms except those involving the functions ξ_μ cancel, and the thesis follows. □

From the representation formula of the curvature tensor, it follows immediately that the sectional curvatures of $\mathscr{P}_2(\mathbb{R}^d)$ are non-negative (for the definition see, for instance, [**33**], Chapter 4, section 3). Indeed, it holds:

$$K(\nabla\varphi, \nabla\psi)(\mu) = \frac{\langle \mathcal{R}(\nabla\varphi, \nabla\psi)\nabla\psi, \nabla\varphi\rangle_\mu}{\|\nabla\varphi\|_\mu^2\|\nabla\psi\|_\mu^2 - \langle \nabla\varphi, \nabla\psi\rangle_\mu^2} = \frac{3\|\xi_\mu(\varphi^1, \varphi^2)\|_\mu^2}{\|\nabla\varphi\|_\mu^2\|\nabla\psi\|_\mu^2 - \langle \nabla\varphi, \nabla\psi\rangle_\mu^2} \geq 0.$$

6. A distance on the tangent bundle

Recall that, for a Riemannian manifold M, it is possible to endow the tangent bundle TM with a natural Riemannian metric, the so-called Sasaki metric, in the following way (see also [**33**], Chapter 3, exercise 2).

Fix a point $(p, u) \in TM$ and choose two regular curves $[0, 1] \ni t \to \alpha^i(t) \in TM$, $i = 1, 2$, such that $\alpha^1(0) = \alpha^2(0) = (p, u)$. Let $(p^i(t), u^i(t)) := \alpha^i(t)$ and $v^i(t) := (p^i(t))'$, $i = 1, 2$. Clearly $V^i := (\alpha^i)'(0) \in T_{(p,u)}(TM)$, $i = 1, 2$. The scalar product $\langle \cdot, \cdot\rangle^*$ between V^1 and V^2 is defined as

$$\langle V^1, V^2\rangle^* := \langle v^1(0), v^2(0)\rangle + \langle \nabla_{v^1}u^1(0), \nabla_{v^2}u^2(0)\rangle.$$

It is possible to show that this is a good definition, that is, it depends only on V^1, V^2 and not on the particular curves $\alpha^1(t)$, $\alpha^2(t)$, therefore it defines a metric tensor on TM. It is then easy to see that the distance d on TM induced by this metric tensor is given by

$$d^2\Big((p^1, u^1), (p^2, u^2)\Big) = \inf_\gamma \big(\mathcal{L}(\gamma)\big)^2 + |\mathcal{T}(u^1) - u^2|^2, \tag{6.49}$$

where the infimum is taken among all the smooth curves $\gamma(t)$ in M connecting p^1 to p^2, $\mathcal{L}(\gamma)$ is the length of γ and $\mathcal{T}(u^1)$ is the parallel transport of u^1 along γ to the point p^2. This definition has a Wasserstein analogous, as it doesn't rely on a differentiable structure of the tangent bundle to be written.

So turn back to the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ and recall that for a given absolutely continuous curve $t \to \mu_t$ on $[0,1]$ we define $\mathcal{L}_{t_1}^{t_2}(\mu_t)$, $0 \leq t_1 \leq t_2 \leq 1$, as the length of the curve on the interval $[t_1, t_2]$: $\mathcal{L}_{t_1}^{t_2}(\mu_t) := \int_{t_1}^{t_2} \|v_t\|_{\mu_t} dr$, where v_t is the velocity vector field of μ_t. We will also set $\mathcal{L}(\mu_t) := \mathcal{L}_0^1(\mu_t)$ for a curve parametrized in $[0,1]$. Recall also that for a regular curve $t \to \mu_t$, we define $\mathcal{T}_t^s : \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \to \mathrm{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$ as the parallel transport maps.

The tangent bundle is defined as

$$\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)) := \Big\{(\mu, u) \ : \ \mu \in \mathscr{P}_2(\mathbb{R}^d),\ u \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))\Big\}.$$

Throughout all this section, we say that a sequence (μ_n, u_n) converges to (μ, u) if the measures converge w.r.t. W and the maps converge strongly, i.e. if:

$$\begin{aligned} &\lim_{n\to\infty} W(\mu_n, \mu) = 0, \\ &\lim_{n\to\infty} \|u_n\|_{\mu_n} = \|u\|_\mu, \\ &\lim_{n\to\infty} \langle u_n, \nabla\varphi\rangle_{\mu_n} = \langle u, \nabla\varphi\rangle_\mu \quad \forall \varphi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}). \end{aligned} \tag{6.50}$$

Remark 6.45. *Actually in general, to have strong convergence of maps we should ask for the third of the above equations to hold for any smooth vector valued map $\eta \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$, and not just for gradients. However, since the vectors $(u_n), u$ are tangent, this more restrictive choice is still sufficient to get the desired convergence. Indeed, let $\overline{u}$ be any weak limit of (u_n) (which exists, because the second equation gives a uniform bound on the norms), and observe that the third equation implies $P_\mu(\overline{u}) = u$, therefore the claim follows from*

$$\|u\|_\mu \leq \|\overline{u}\|_\mu \leq \varliminf_{n\to\infty} \|u_n\|_{\mu_n} = \|u\|_\mu.$$

In the following we will, with some abuse of notation, say u_n converges to u instead than (μ_n, u_n) converge to (μ, u).

By analogy with equation (6.49), we define the following function on $[\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))]^2$:

$$d^2((\mu, u), (\nu, v)) := \inf \left\{ (L_0^1(\mu_t))^2 + \|v - \mathcal{T}_0^1(u)\|_\nu^2 \right\},$$

where the infimum is taken on the set of regular curves $t \to \mu_t$ on $[0,1]$ such that $\mu_0 = \mu$ and $\mu_1 = \nu$. In particular we define $d((\mu, u), (\nu, v)) := +\infty$ if there is no regular curve connecting μ and ν.

The first properties of d are collected below.

PROPOSITION 6.46. *The function d is non-negative, symmetric and satisfies the triangular inequality. Moreover for any couple $(\mu, u), (\nu, v) \in \operatorname{Tan}(\mathscr{P}_2(\mathbb{R}^d))$ and every $\varphi \in C_c^\infty(\mathbb{R}^d)$ it holds:*

$$
\begin{aligned}
W(\mu,\nu) &\le d((\mu,u),(\nu,v)),\\
\Big|\|u\|_\mu - \|v\|_\nu\Big| &\le d((\mu,u),(\nu,v)),\\
\Big|\langle u,\nabla\varphi\rangle_\mu - \langle v,\nabla\varphi\rangle_\nu\Big| &\le d((\mu,u),(\nu,v))\big(\|u\|_\mu \operatorname{Lip}(\nabla\varphi) + \|\nabla\varphi\|_\nu\big).
\end{aligned}
$$

Proof. The non-negativity and the triangle property are clear. For the symmetry it is enough to observe that for any regular curve connecting μ and ν it holds $\|\mathcal{T}_0^1(u)-v\|_\nu = \|u-\mathcal{T}_1^0(v)\|_\mu$, given that the parallel transport is norm preserving.

The first inequality is obvious. For the second, notice that $\|u\|_\mu = \|\mathcal{T}_0^1(u)\|_\nu$, therefore $\big|\|u\|_\mu - \|v\|_\nu\big| = \big|\|\mathcal{T}_0^1(u)\|_\nu - \|v\|_\nu\big| \le \|\mathcal{T}_0^1(u)-v\|_\nu$. To prove the last one start observing that for any regular curve $t \to \mu_t$ connecting μ to ν, the function $t \to \langle \mathcal{T}_0^t(u), \nabla\varphi\rangle_{\mu_t}$ is absolutely continuous and its derivative is given by

$$
\frac{d}{dt}\langle \mathcal{T}_0^t(u),\nabla\varphi\rangle_{\mu_t} = \langle \mathcal{T}_0^t(u), \nabla^2\varphi\cdot v_t\rangle_{\mu_t} \le \operatorname{Lip}(\nabla\varphi)\|u\|_\mu\|v_t\|_{\mu_t},
$$

where v_t are the velocity vectors of $t \to \mu_t$. Integrate to obtain

$$
\Big|\langle u,\nabla\varphi\rangle_\mu - \langle \mathcal{T}_0^1(u),\nabla\varphi\rangle_\nu\Big| \le \operatorname{Lip}(\nabla\varphi)\|u\|_\mu\mathcal{L}(\mu_t).
$$

Finally observe that

$$
\Big|\langle \mathcal{T}_0^1(u),\nabla\varphi\rangle_\nu - \langle v,\nabla\varphi\rangle_\nu\Big| \le \|\mathcal{T}_0^1(u)-v\|_\nu\|\nabla\varphi\|_\nu.
$$

□

The first properties stated above suggest that d behaves like a distance on $\operatorname{Tan}(\mathscr{P}_2(\mathbb{R}^d))$, the only problem being that it is not real valued. Given that the regular curves are dense in the set of absolutely continuous curves, a natural candidate for a relaxation of d is its lower semicontinuous envelope d^*, defined by:

$$
d^*\Big((\mu,u),(\nu,v)\Big) := \inf_{\substack{(\mu_n,u_n)\to(\mu,u)\\(\nu_n,v_n)\to(\nu,v)}} \varliminf_{n\to\infty} d\Big((\mu_n,u_n),(\nu_n,v_n)\Big).
$$

However, it seems that the function d^* does not have sufficient regularity properties. For instance it is not clear - to the author - whether d^* satisfies the triangle inequality. Therefore we modify a bit the definition, and we introduce the function $\mathcal{D}$ as:

$$
\mathcal{D}\Big((\mu,u),(\nu,v)\Big) := \inf_{\varphi,\psi\in C_c^\infty(\mathbb{R}^d)} \left\{ d^*\Big((\mu,\nabla\varphi),(\nu,\nabla\psi)\Big) + \|u-\nabla\varphi\|_\mu + \|v-\nabla\psi\|_\nu \right\}.
$$

With the introduction of $\mathcal{D}$ we are allowed to regularize the vectors u, v provided we pay the L^2 difference between the regularizations and the vectors themselves.

We want to prove that $\mathcal{D}$ is a distance on $\operatorname{Tan}(\mathscr{P}_2(\mathbb{R}^d))$ with metrizes the convergence of maps.

It is easy to check that the following inequalities are valid for any couple $(\mu,u),(\nu,v)\in \mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))$ and every $\varphi\in C_c^\infty$:

$$W(\mu,\nu)\le d^*((\mu,u),(\nu,v)),$$
$$\Big|\|u\|_\mu-\|v\|_\nu\Big|\le d^*((\mu,u),(\nu,v)),$$
$$\Big|\langle u,\nabla\varphi\rangle_\mu-\langle v,\nabla\varphi\rangle_\nu\Big|\le d^*((\mu,u),(\nu,v))\big(\|u\|_\mu\,\mathrm{Lip}(\nabla\varphi)+\|\nabla\varphi\|_\nu\big).$$

Indeed all of terms, apart d^*, are continuous w.r.t. to convergence in $\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))$: therefore the inequalities follow directly from the analogous inequalities valid for d.

From the latter inequalities it is just a matter of simple algebraic manipulations to show that $\mathcal{D}$ satisfies:

$$W(\mu,\nu)\le\mathcal{D}\Big((\mu,u),(\nu,v)\Big), \tag{6.50a}$$

$$\Big|\|u\|_\mu-\|v\|_\nu\Big|\le\mathcal{D}\Big((\mu,u),(\nu,v)\Big), \tag{6.50b}$$

(6.50c)
$$\Big|\langle u,\nabla\varphi\rangle_\mu-\langle v,\nabla\varphi\rangle_\nu\Big|\le\mathcal{D}\Big((\mu,u),(\nu,v)\Big)\big(\|u\|_\mu\,\mathrm{Lip}(\nabla\varphi)+2\|\nabla\varphi\|_\nu+\|\nabla\varphi\|_\mu\big).$$

PROPOSITION 6.47 ($\mathcal{D}$ is a distance). *The function $\mathcal{D}$ is a distance on* $\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))$.

Proof. Clearly $\mathcal{D}$ is non-negative, symmetric and $\mathcal{D}((\mu,u),(\mu,u))=0$. Equations (6.50) show that if $\mathcal{D}((\mu,u),(\nu,v))=0$, then $\mu=\nu$ and $u=v$.

Consider now a geodesic μ_t parametrized in $[0,1]$ connecting μ to ν and find a vector field (not necessarily continuous) u_t such that (μ_t,u_t) converges to (μ,u) as $t\to0$ and to (ν,v) as $t\to1$. Recalling that any geodesic is regular if restricted to the interval $[\varepsilon,1-\varepsilon]$ we can bound d^* from above with:

$$\begin{aligned}d^*\Big((\mu,u),(\nu,v)\Big)&\le\varliminf_{\varepsilon\to0}d\Big((\mu_\varepsilon,u_\varepsilon),(\mu_{1-\varepsilon},u_{1-\varepsilon})\Big)\\&\le\varliminf_{\varepsilon\to0}W(\mu,\nu)+\|\mathcal{T}_\varepsilon^{1-\varepsilon}(u_\varepsilon)-u_{1-\varepsilon}\|_{\mu_{1-\varepsilon}}\\&\le W(\mu,\nu)+\|u\|_\mu+\|v\|_\nu.\end{aligned}$$

Being d^* real valued, so is $\mathcal{D}$.

It remains to prove the triangle inequality. Fix three couples (μ^i,u^i), $i=1,2,3$, and choose a minimizing sequences for $\mathcal{D}((\mu^1,u^1),(\mu^2,u^2))$, that is: two sequences of smooth functions $u_n^1,u_n^{2,a}\in C_c^\infty(\mathbb{R}^d,\mathbb{R})$, $n\in\mathbb{N}$, and, for every $n\in\mathbb{N}$, two sequences of measures $(\mu^1_{n,m})$, $(\mu^{2,a}_{n,m})$, and a sequence of regular curves $t\to\nu^a_{n,m}(t)$ connecting $\mu^1_{n,m}$ to $\mu^{2,a}_{n,m}$, such that: $\lim_{m\to\infty}W(\mu^1_{n,m},\mu^1)=0$, $\lim_{m\to\infty}W(\mu^{2,a}_{n,m},\mu^2)=0$ and

$$d^*\Big((\mu^1,\nabla u_n^1),(\mu^2,\nabla u_n^{2,a})\Big)=\lim_{m\to\infty}\Big(\mathcal{L}^2(\nu^a_{n,m})+\|(\mathcal{T}^a_{n,m})_0^1(\nabla u_n^1)-\nabla u_n^{2,a}\|^2_{\mu^{2,a}_{n,m}}\Big)^{1/2},$$

$$\begin{aligned}\mathcal{D}\Big((\mu^1,u^1),(\mu^2,u^2)\Big)=\lim_{n\to\infty}&\Big(d^*\Big((\mu^1,\nabla u_n^1),(\mu^2,\nabla u_n^2)\Big)\\&+\|\nabla u_n^1-u^1\|_{\mu^1}+\|\nabla u_n^{2,a}-u^2\|_{\mu^2}\Big),\end{aligned}$$

where $(\mathcal{T}^a_{n,m})^s_t$, $t,s\in[0,1]$, are the parallel transport maps along $t\to\nu^a_{n,m}(t)$. It is easy to check, using proposition 6.21, that the curves $t\to\nu^a_{n,m}(t)$ may be chosen such that for every n,m the support of the measure $\nu^a_{n,m}(t)$ is made of exactly m points, each of mass $1/m$. Pick an analogous minimizing sequence for $D((\mu^2,u^2),(\mu^3,u^3))$, made of smooth functions $u^{2,b}_n, u^3_n$, measures $\mu^{2,b}_{n,m},\mu^3_{n,m}$ and curves $t\to\nu^b_{n,m}(t)$ from $\mu^{2,b}_{n,m}$ to $\mu^3_{n,m}$. Observe that $W(\mu^{2,a}_{n,m},\mu^{2,b}_{n,m})\to 0$ as $m\to\infty$, because both the sequences converge to μ^2, and that for every n,m, both the measures $\mu^{2,a}_{n,m}$ and $\mu^{2,b}_{n,m}$ are discrete and the mass of their points is $1/m$. In this case we know (see also remark 1.16) that there exists an optimal transport map from $\mu^{2,a}_{n,m}$ to $\mu^{2,b}_{n,m}$. Let $[0,1]\ni t\to\sigma_{n,m}(t)$ be the geodesic induced by this map and observe that it is regular on the whole $[0,1]$.

Define a regular curve $t\to\mu_{n,m}(t)$ on $[0,3]$ by joining $\nu^a_{n,m}(t),\sigma_{n,m}(t),\nu^b_{n,m}(t)$: its extreme points are $\mu^1_{n,m}$, $\mu^3_{n,m}$, which converge to μ^1 and μ^3 as $m\to\infty$. Therefore, for a fixed n we can use these curves to estimate $d^*((\mu^1,\nabla u^1_n),(\mu^3,\nabla u^3_n))$ as follows ($(\mathcal{T}_{n,m})^s_t$ are the parallel transport maps along $t\to\mu_{n,m}(t)$):

$$
\begin{aligned}
d^*\Big((\mu^1,\nabla u^1_n),(\mu^3,\nabla u^3_n)\Big) \le\ & \varlimsup_{m\to\infty}\Big(\mathcal{L}^2(\mu_{n,m})+\|(\mathcal{T}_{n,m})^3_0(\nabla u^1_n)-\nabla u^3_n\|_{\mu^3_{n,m}}\Big)^{1/2}\\
\le\ & \varlimsup_{m\to\infty}\Big(\mathcal{L}^2(\nu^1_{n,m})+\|(\mathcal{T}_{n,m})^1_0(\nabla u^1_n)-\nabla u^{2,a}_n\|_{\mu^{2,a}_{n,m}}\Big)^{1/2}\\
&+\varlimsup_{m\to\infty}\Big(\mathcal{L}^2(\sigma_{n,m})+\|(\mathcal{T}_{n,m})^2_1(\nabla u^{2,a}_n)-\nabla u^{2,b}_n\|_{\mu^{2,b}_{n,m}}\Big)^{1/2}\\
&+\varlimsup_{m\to\infty}\Big(\mathcal{L}^2(\nu^2_{n,m})+\|(\mathcal{T}_{n,m})^3_2(\nabla u^{2,b}_n)-\nabla u^3_n\|_{\mu^3_{n,m}}\Big)^{1/2}\\
\le\ & d^*\Big((\mu^1,\nabla u^1_n),(\mu^2,\nabla u^{2,a}_n)\Big)+d^*\Big((\mu^2,\nabla u^{2,b}_n),(\mu^3,\nabla u^3_n)\Big)\\
&+\|\nabla u^{2,a}_n-\nabla u^{2,b}_n\|_{\mu^2}\\
\le\ & d^*\Big((\mu^1,\nabla u^1_n),(\mu^2,\nabla u^{2,a}_n)\Big)+d^*\Big((\mu^2,\nabla u^{2,b}_n),(\mu^3,\nabla u^3_n)\Big)\\
&+\|\nabla u^{2,a}_n-u^2\|_{\mu^2}+\|\nabla u^2-u^{2,b}_n\|_{\mu^2},
\end{aligned}
$$

where in the third step we used lemma 6.35 and the fact that $W(\mu^{2,a}_{n,m},\mu^{2,b}_{n,m})\to 0$ as $m\to\infty$.

Now choose $\nabla u_n^1, \nabla u_n^3$ as regular functions to bound from above $\mathcal{D}((\mu^1, u^1), (\mu^3, u^3))$ with

$$\begin{aligned}
&\mathcal{D}\Big((\mu^1, u^1), (\mu^3, u^3)\Big) \\
&\quad \leq \varliminf_{n\to\infty} \left(d^*\Big((\mu^1, \nabla u_n^1), (\mu^3, \nabla u_n^3)\Big) + \|u^1 - \nabla u_n^1\|_{\mu^1} + \|u^3 - \nabla u_n^3\|_{\mu^3} \right) \\
&\quad \leq \varliminf_{n\to\infty} \left(d^*\Big((\mu^1, \nabla u_n^1), (\mu^2, \nabla u_n^{2,a})\Big) + d^*\Big((\mu^2, \nabla u_n^{2,b}), (\mu^3, \nabla u_n^3)\Big) \right. \\
&\qquad \left. + \|\nabla u_n^{2,a} - u^2\|_{\mu^2} + \|\nabla u^2 - u_n^{2,b}\|_{\mu^2} + \|u^1 - \nabla u_n^1\|_{\mu^1} + \|u^3 - \nabla u_n^3\|_{\mu^3} \right) \\
&\quad \leq \mathcal{D}\Big((\mu^1, u^1), (\mu^2, u^2)\Big) + \mathcal{D}\Big((\mu^2, u^2), (\mu^3, u^3)\Big).
\end{aligned}$$

□

Proposition 6.48 ($\mathcal{D}$ metrizes the convergence in $\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))$). *A sequence (μ_n, u_n) converges to (μ, u) in $\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))$ in the sense of equations (6.50) if and only if $\mathcal{D}((\mu_n, u_n), (\mu, u))$ converges to 0.*

Proof. Let $(\mu, u), (\mu_n, u_n)$, $n \in \mathbb{N}$, be given couples in $\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))$.

Suppose that $\mathcal{D}((\mu, u), (\mu_n, u_n)) \to 0$. Then equation (6.50a) implies that $W(\mu, \mu_n) \to 0$, equation (6.50c) that the maps u_n converge weakly to u and equation (6.50b) gives the convergence of norms. Therefore the maps converge in $\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d))$.

For the converse implication, fix a function $\varphi \in C_c^\infty(\mathbb{R}^d)$ and choose, for every n, a geodesic $\mu_n(t)$ from μ to μ_n and let, for any $0 < \varepsilon < 1$, $\mathcal{T}_n^\varepsilon$ be the parallel transport map from $\mu_n(\varepsilon)$ to $\mu_n(1-\varepsilon)$. Fix $n \in \mathbb{N}$ and observe that we have the following bound:

$$\begin{aligned}
\mathcal{D}\Big((\mu, u), (\mu_n, u_n)\Big) &\leq d^*\Big((\mu, \nabla\varphi), (\mu_n, \nabla\varphi)\Big) + \|u - \nabla\varphi\|_\mu + \|u_n - \nabla\varphi\|_{\mu_n} \\
&\leq \lim_{\varepsilon\to 0} \Big((1-2\varepsilon)^2 W^2(\mu, \mu_n) + \|\mathcal{T}_n^\varepsilon(\nabla\varphi) - \nabla\varphi\|_{\mu_n(1-\varepsilon)} \Big)^{1/2} \\
&\qquad + \|u - \nabla\varphi\|_\mu + \|u_n - \nabla\varphi\|_{\mu_n} \\
&\leq W(\mu, \mu_n) + W(\mu, \mu_n)\,\mathrm{Lip}(\nabla\varphi) + \|u - \nabla\varphi\|_\mu + \|u_n - \nabla\varphi\|_{\mu_n},
\end{aligned}$$

having used lemma 6.35. Letting $n \to \infty$ we get

$$\varlimsup_{n\to\infty} \mathcal{D}\Big((\mu, u), (\mu_n, u_n)\Big) \leq 2\|u - \nabla\varphi\|_\mu,$$

therefore the thesis follows by letting $\nabla\varphi \to u$ in L^2_μ. □

Remark 6.49. *With the same spirit it is possible to extend $\mathcal{D}$ to a function on* $\mathbf{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)) := \{(\mu, \boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu\}$, *provided we define what do we mean for convergence of a sequence of maps $u_n \in \mathrm{Tan}_{\mu_n}(\mathscr{P}_2(\mathbb{R}^d))$ to a plan $\boldsymbol{\gamma} \in \mathscr{P}_2(\mathbb{R}^{2d})_\mu$. However, independently from this definition, it is possibile to show that the resulting space* $(\mathbf{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)), \mathcal{D})$ *is* not *the completion of the space* $(\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)), \mathcal{D})$. *Indeed, the problem is that a sequence of tangent maps may converge strongly to a non tangent* *map: in proposition 6.11 we showed that actually* any *vector, tangent or not, may be approximated in the sense of strong convergence of maps with tangent vectors.*

The distance $\mathcal{D}$ is closely related to the behavior of parallel transport along regular curves, as the following proposition shows.

PROPOSITION 6.50. *Let $t \to (\mu_t, u_t)$ be an absolutely continuous curve in $(\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)), \mathcal{D})$. Then $t \to \mu_t$ is an absolutely continuos curve in $(\mathscr{P}_2(\mathbb{R}^d), W)$ and it holds*

$$\lim_{s \to t} \frac{\mathcal{D}\Big((\mu_s, u_s), (\mu_t, u_t)\Big)}{|s-t|} \geq \|v_t\|_{\mu_t}, \tag{6.51}$$

for a.e. t, where the v_t are the velocity vectors of μ_t.

Conversely, if $t \to \mu_t$ is a regular curve and $t \to u_t$ a parallel transport along it, then $t \to (\mu_t, u_t)$ is absolutely continuous in $(\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)), \mathcal{D})$ and the equality holds in the above limit for a.e. t.

Proof. The first claim follows directly from inequality (6.50a). For the second one just observe that $\mathcal{D} \leq d$ and estimate $d\Big((\mu_s, u_s), (\mu_t, u_t)\Big)$ from above by using the regular curve $r \to \mu_{(1-r)t+rs}$. □

This proposition suggests a way to generalize the definition of parallel transport in the following way:

DEFINITION 6.51 (Weak parallel transport). *Let $t \to \mu_t$ be an absolutely continuous curve and v_t its velocity vectors. We say that $t \to u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ is a weak parallel transport along μ_t if $t \to (\mu_t, u_t)$ is absolutely continuous in $(\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)), \mathcal{D})$ and equation (6.51) holds for a.e. t.*

For the weak parallel transport we are able to prove that it is norm preserving and satisfies equation (6.22), which characterizes the parallel transport along regular curves without the flow maps.

PROPOSITION 6.52 (Properties of weak parallel transport). *Let μ_t be an absolutely continuous curve and let $u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ be a weak parallel transport along it. Then $t \to \|u_t\|_{\mu_t}$ is constant, $t \to \langle u_t, \nabla\eta\rangle_{\mu_t}$ is absolutely continuous for every $\eta \in C_c^\infty(\mathbb{R}^d)$ and it holds*

$$\frac{d}{dt}\langle u_t, \nabla\eta\rangle_{\mu_t} = \langle u_t, \nabla^2\eta \cdot v_t\rangle_{\mu_t}, \quad \textit{a.e. } t.$$

Proof. Start observing that from inequality (6.50b) and the absolute continuity of $t \to (\mu_t, u_t) \in (\mathrm{Tan}\,(\mathscr{P}_2(\mathbb{R}^d)), \mathcal{D})$ it follows that $t \to \|u_t\|_{\mu_t}$ is absolutely continuous, too.

Now let $\Sigma \subset [0,1]$ be the set of t such that the metric derivative $|\dot{\mu}_t|$ and the velocity vector v_t of μ_t are defined and the equation of weak parallel transport is satisfied, i.e.

the set of those $t \in [0,1]$ such that:

$$\begin{aligned}
\lim_{h\to 0}\frac{\int \varphi d\mu_{t+h} - \int \varphi d\mu_t}{h} &= \langle v_t, \nabla\varphi\rangle_{\mu_t}, \quad \forall \varphi \in C_c^\infty(\mathbb{R}^d),\\
\lim_{h\to 0}\frac{W(\mu_{t+h},\mu_t)}{h} &= \|v_t\|_{\mu_t},\\
\lim_{h\to 0}\frac{\mathcal{D}\big((\mu_t,u_t),(\mu_{t+h},u_{t+h})\big)}{h} &= \|v_t\|_{\mu_t},
\end{aligned}$$

and recall that by theorem 3.4 we know that $\mathcal{L}^1([0,1]\setminus\Sigma)=0$. Therefore to prove that $t\to\|u_t\|^2_{\mu_t}$ is constant, it is sufficient to show that its derivative is 0 at any $t\in\Sigma$.

Fix such a t and a sequence $h^n\downarrow 0$. By definition of $\mathcal{D}$ and with a diagunalization argument, we can find first two sequences of functions $(\varphi^n),(\psi^n)\subset C_c^\infty(\mathbb{R}^d)$ then a sequence of regular curves parametrized by arc length $[0,k^n]\ni r\to\nu^n_r$ such that it holds:

$$\overline{\lim_{n\to\infty}}\,\frac{W(\mu_t,\nu^n_0)(1+\mathrm{Lip}(\nabla\varphi^n))}{h^n}=0, \tag{6.52}$$

$$\overline{\lim_{n\to\infty}}\,\frac{W(\mu_{t+h^n},\nu^n_{k^n})(1+\mathrm{Lip}(\nabla\psi^n))}{h^n}=0, \tag{6.53}$$

and

$$\begin{aligned}
\mathcal{D}\big((\mu_t,u_t),(\mu_{t+h^n},u_{t+h^n})\big) \le&\sqrt{(k^n)^2+\|(\mathcal{T}^n)_0^{k^n}(\nabla\varphi^n)-\nabla\psi^n\|^2_{\nu_1^n}}\\
&+\|\nabla\varphi^n-u_t\|_{\mu_t}+\|\nabla\psi^n-u_{t+h^n}\|_{\mu_{t+h^n}}\\
\le&\mathcal{D}\big((\mu_t,u_t),(\mu_{t+h^n},u_{t+h^n})\big)+h^n w(n),
\end{aligned}\tag{6.54}$$

where $(\mathcal{T}^n)_{r_1}^{r_2}$ are the parallel transport maps along ν^n_r and $w(n)\to 0$ as $n\to\infty$.

From equations (6.52), (6.53), inequality $k^n\ge W(\nu^n_0,\nu^n_{k^n})$ and by our choice of t, it follows $\underline{\lim}_n k^n/h^n\ge\|v_t\|_{\mu_t}$. Therefore inequalities (6.54) imply

$$\begin{aligned}
&\sup_n\Big\{\|\nabla\psi\|_{\mu_{t+h^n}},\|\nabla\psi\|_{\nu^n_{k^n}},\|\nabla\varphi\|_{\mu_t},\|\nabla\varphi\|_{\nu^n_0}\Big\}<\infty,\\
&\lim_{n\to\infty}\frac{k^n}{h^n}=\|v_t\|_{\mu_t},\\
&\lim_{n\to\infty}\frac{\|(\mathcal{T}^n)_0^{k^n}(\nabla\varphi^n)-\nabla\psi^n\|_{\nu_1^n}}{h^n}=0,\\
&\lim_{n\to\infty}\frac{\|\nabla\varphi^n-u_t\|_{\mu_t}}{h^n}=0,\\
&\lim_{n\to\infty}\frac{\|\nabla\psi^n-u_{t+h^n}\|_{\mu_{t+h^n}}}{h^n}=0.
\end{aligned}\tag{6.55}$$

Now define $A_n := \|\nabla\varphi^n - u_t\|_{\mu_t}$, $B_n := \|\nabla\psi^n - u_{t+h^n}\|_{\mu_{t+h^n}}$, $C_n := \|(\mathcal{T}^n)_0^{k^n}(\nabla\varphi^n) - \nabla\psi^n\|^2_{\nu_1^n}$ and observe that it holds

$$\begin{aligned}
\left|\|u_{t+h^n}\|^2_{\mu_{t+h^n}} - \|u_t\|^2_{\mu_t}\right| =& \Big|\langle u_{t+h^n} + \nabla\psi^n, u_{t+h^n} - \nabla\psi^n\rangle_{\mu_{t+h^n}} \\
&+ \|\nabla\psi^n\|^2_{\mu_{t+h^n}} - \|\nabla\psi^n\|^2_{\nu^n_{k^n}} \\
&+ \langle \nabla\psi^n - (\mathcal{T}^n)_0^{k^n}(\nabla\varphi^n), \nabla\psi^n + (\mathcal{T}^n)_0^{k^n}(\nabla\varphi^n)\rangle_{\nu^n_{k^n}} \\
&+ \|\nabla\varphi^n\|^2_{\nu_0^n} - \|\nabla\varphi^n\|^2_{\mu_t} \\
&- \langle u_t + \nabla\varphi^n, u_t - \nabla\varphi^n\rangle_{\mu_t}\Big| \\
\le& B_n\left|\|u_{t+h^n}\|_{\mu_{t+h^n}} + \|\nabla\psi^n\|_{\mu_{t+h^n}}\right| \\
&+ \mathrm{Lip}(\nabla\psi^n) W(\mu_{t+h^n}, \nu^n_{k^n})\Big(\|\nabla\psi^n\|_{\mu_{t+h^n}} + \|\nabla\psi^n\|_{\nu^n_{k^n}}\Big) \\
&+ C_n\left|\|\nabla\psi\|_{\nu^n_{k^n}} + \|\nabla\varphi\|_{\nu_0^n}\right| \\
&+ \mathrm{Lip}(\nabla\varphi^n) W(\mu_t, \nu_0^n)\Big(\|\nabla\varphi^n\|_{\mu_t} + \|\nabla\varphi^n\|_{\nu_0^n}\Big) \\
&+ A_n\left|\|u_t\|_{\mu_t} + \|\nabla\varphi\|_{\mu_t}\right|.
\end{aligned}$$

Dividing by h^n, letting $n \to \infty$ and using equations (6.55), we get that

$$\lim_{n\to\infty} \frac{\|u_{t+h^n}\|^2_{\mu_{t+h^n}} - \|u_t\|^2_{\mu_t}}{h^n} = 0.$$

By the arbitrariness of $t \in \Sigma$ and $h_n \downarrow 0$ we conclude.

Now we turn to the study of $t \to \langle u_t, \nabla\eta\rangle_{\mu_t}$ for a given $\eta \in C_c^\infty(\mathbb{R}^d)$. Start observing that from the fact that $t \to \|u_t\|_{\mu_t}$ is constant and by inequality (6.50c) we get that $t \to \langle u_t, \nabla\eta\rangle_{\mu_t}$ is absolutely continuous. With the same notation as above and with calculations similar to the ones just done, it is possible to check that it holds

$$\lim_{n\to\infty} \frac{\langle u_{t+h^n}, \nabla\eta\rangle_{\mu_{t+h^n}} - \langle u_t, \nabla\eta\rangle_{\mu_t}}{h^n} = \lim_{n\to\infty} \frac{\langle (\mathcal{T}^n)_0^{k^n}(\nabla\varphi^n), \nabla\eta\rangle_{\nu^n_{k^n}} - \langle \nabla\varphi^n, \nabla\eta\rangle_{\nu_0^n}}{h^n}.$$

Now observe that the right hand side can be evaluated using equation (6.23):

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\langle (\mathcal{T}^n)_0^{k^n}(\nabla\varphi^n), \nabla\eta\rangle_{\nu^n_{k^n}} - \langle \nabla\varphi^n, \nabla\eta\rangle_{\nu_0^n}}{h^n} \\
&= \lim_{n\to\infty} \frac{1}{h^n}\int_0^{k^n} \langle (T^n)_0^r(\nabla\varphi^n), \nabla^2\eta \cdot v_r^n\rangle_{\nu_r^n} dr,
\end{aligned}$$

where $v_r^n \in \mathrm{Tan}_{\nu_r^n}(\mathscr{P}_2(\mathbb{R}^d))$ are the velocity vectors of $r \to \nu_r^n$. By equation (6.55) it follows that for any $r \in [0, k^n]$ the sequence $\big((T^n)_0^r(\nabla\varphi^n)\big)$ strongly converges to u_t as $n \to \infty$. Therefore our thesis will be proved if we show that $V^n := (h^n)^{-1}\int_0^{k^n} v_r^n \circ (\mathcal{T}^n)_0^r dr \in L^2_{\nu_0^n}$ weakly converges to v_t as $n \to \infty$. To prove this, fix $\xi \in C_c^\infty(\mathbb{R}^d)$ and observe that from equations (6.55) and the fact that $\mathrm{Lip}(\xi), \mathrm{Lip}(\nabla\xi) < \infty$ it follows that

$$\langle \nabla\xi, v_t \rangle_{\mu_t} = \lim_{n\to\infty} \frac{\int \xi d\mu_{t+h^n} - \int \xi d\mu_t}{h^n} = \lim_{n\to\infty} \frac{\int \xi d\nu^n_{k^n} - \int \xi d\nu^n_0}{h^n} = \lim_{n\to\infty} \langle \nabla\xi, V^n \rangle_{\nu^n_0}.$$

Let V be any weak limit of (V^n) and observe that the above equations are equivalent to $P_{\mu_t}(V) = v_t$. To prove that V is tangent, and therefore equal to v_t, observe that:

$$\|v_t\|_{\mu_t} = \|P_{\mu_t}(V)\|_{\mu_t} \leq \|V\|_{\mu_t} \leq \varliminf_{n\to\infty} \|V^n\|_{\nu^n_0}$$
$$\leq \varliminf_{n\to\infty} \frac{\int_0^{k^n} \|v^n_r\|_{\nu^n_r} dr}{h^n} = \varliminf_{n\to\infty} \frac{k^n}{h^n} = \|v_t\|_{\mu_t},$$

hence $\|V\|_{\mu_t} = \|P_{\mu_t}(V)\|_{\mu_t}$ and the thesis is proved. □

It should be noticed that it is not clear - to the author - whether the weak parallel transport preserves the scalar product, nor if it is unique. From a tecnhical point of view, the big obstacle is that, if u^1_t, u^2_t are two absolutely continuous vector fields w.r.t. $\mathcal{D}$ along the same absolutely continuous curve μ_t, it is not clear whether the vector field $u^1_t + u^2_t$ is absolutely continuous or not.

Furthermore, the density of regular curves is not enough to gain existence of weak parallel transport through an approximation argument. The key problem is that the space $(\mathrm{Tan}(\mathscr{P}_2(\mathbb{R}^d)), \mathcal{D})$ is *not* complete. Example 6.37 shows that it might be impossible to extend a (weak) parallel transport "backward" to the initial point of a geodesic.

We conclude with a result regarding the lower semicontinuity of the function d. Given the definition of $\mathcal{D}$ and d^*, a natural question is the following: is $\mathcal{D}((\mu,u),(\nu,v)) = d^*((\mu,u),(\nu,v)) = d((\mu,u),(\nu,v))$ on those couples such that $d((\mu,u),(\nu,v)) < +\infty$? The answer is no, as we are going to show now.

The key fact is the possibility to "separate" the supports on the approximating curves. Observe that it always holds $d^* \leq d$, and, if the vectors are smooth, $\mathcal{D} \leq d^*$, therefore we prove our claim if we build an example in which

$$d^*((\mu,u),(\nu,v)) < d((\mu,u),(\nu,v)),$$

such that $u, v \in C^\infty_c(\mathbb{R}^d, \mathbb{R})$ and the right hand side is finite.

Examples exist in every dimension, however, for the sake of simplicity, we show an explicit construction only in $\mathbb{R}^2$.

Fix two positive numbers a, b. Define $A := \{-1\} \times [0,a] \subset \mathbb{R}^2$, $B := \{1\} \times [0,a] \subset \mathbb{R}^2$ and $\mu^0 = \mu^1 := (2a)^{-1}\mathcal{H}^1|_{A\cup B}$, where $\mathcal{H}^1$ is the 1-dimensional Hausdorff measure. Now define

$$u^0(x) := \begin{cases} (0,b), & x \in A, \\ (0,-b), & x \in B, \end{cases}$$

and $u^1(x) := -u^0(x)$.

Clearly $u^i \in \mathrm{Tan}_{\mu^i}(\mathscr{P}_2(\mathbb{R}^d)) \cap C^\infty_c(\mathbb{R}^d, \mathbb{R})$, $i = 0, 1$. It is clear that there exists at least one regular curve $t \to \mu_t$ connecting μ^0 to μ^1 (consider for instance the constant curve $t \to \mu^0$), therefore $d\Big((\mu^0, u^0), (\mu^1, u^1)\Big) < \infty$. We claim that for a, b large enough it holds

$$d^*\Big((\mu^0, u^0), (\mu^1, u^1)\Big) < d\Big((\mu^0, u^0), (\mu^1, u^1)\Big).$$

We start with a lower bound for $d\Big((\mu^0, u^0), (\mu^1, u^1)\Big)$. Choose a regular curve $t \to \mu_t$ connecting μ^0 to μ^1, let $\mathbf{T}(t, s, x)$ be its flow maps and set $\mathbf{T}(x) := \mathbf{T}(0, 1, x)$. Since $\mathrm{supp}(\mu^0) = \mathrm{supp}(\mu^1) = A \cup B$, it must hold $\mathbf{T}(A \cup B) = A \cup B$. Moreover, since $\mathbf{T}$ is an homeomorphism of $\mathbb{R}^2$ into itself and the sets A, B are separated it must be true one of the following:

case 1 $\mathbf{T}(A) = A$ and $\mathbf{T}(B) = B$,

case 2 $\mathbf{T}(A) = B$ and $\mathbf{T}(B) = A$.

In case 1, examples 6.40 and 6.42 ensure $\mathcal{T}_0^1(u^0) = u^0$, therefore we obtain the bound

$$((L_0^1(\mu_t))^2 + \|\mathcal{T}_0^1(u^0) - u^1\|_{\mu^1}^2)^{1/2} \geq \|u^0 - u^1\|_{\mu^1} = 2b.$$

In case 2, proposition 6.53 below, ensures that the length of $t \to \mu_t$ is bounded from below by $a/8$: we postpone the proof of this to avoid losing continuity in our discussion - here we just observe that this bound exists because if A and B exchange in a continuous way, then "one has to pass above the other". Therefore it holds:

$$((L_0^1(\mu_t))^2 + \|\mathcal{T}_0^1(u^0) - u^1\|_{\mu^1}^2)^{1/2} \geq \mathcal{L}_0^1(\mu_t) \geq a/8.$$

In summary, we have the following lower bound on d:

$$d\Big((\mu^0, u^0)(\mu^1, u^1)\Big) \geq \min\left\{\frac{a}{8}, 2b\right\}. \tag{6.56}$$

Now we want to build a sequence of measures (μ_n), such that $W(\mu^0, \mu_n) \to 0$, and two sequences of functions $(u_n^i) \in \mathrm{Tan}_{\mu_n}(\mathscr{P}_2(\mathbb{R}^d))$, $i = 0, 1$, which converge strongly to u^i, $i = 0, 1$, such that

$$\varliminf_{n\to\infty} d\Big((\mu_n, u_n^0)(\mu_n, u_n^1)\Big) < \min\left\{\frac{a}{8}, 2b\right\}.$$

In order to do so, find, for every $n \in \mathbb{N}$, a family of intervals $I_i^n \subset \mathbb{R}$, $i = 1, \ldots, n$ such that

$$I_i^n \cap I_j^n = \emptyset, \quad \forall n \in \mathbb{N}, 1 \leq i \leq n,$$
$$\lim_{n\to\infty} \sup_i \mathcal{H}^1(I_i^n) = 0,$$
$$\lim_{n\to\infty} I^n = [0, a],$$

where $I^n := \cup_{i=1}^n I_i^n$ and the limit is intended w.r.t. the Hausdorff distance on compact sets. Now define $A_n := \{-1\} \times I^n \subset \mathbb{R}^2$, $B^n := \{1\} \times I^n \subset \mathbb{R}^2$ and

$$\mu_n := \frac{1}{2\mathcal{H}^1(I^n)} \mathcal{H}^1|_{A^n \cup B^n}.$$

Finally, define u_n^0 as

$$u_n^0(x) := \begin{cases} (0,b), & x \in A^n, \\ (0,-b), & x \in B^n, \end{cases}$$

and $u_n^1(x) := -u_n^0(x)$. It is clear from the definition that (μ_n) converges to $\mu^0 = \mu^1$ w.r.t. W and (u_n^i) to u^i w.r.t. the strong convergence of maps.

Corollary 6.55 below ensures that, for every $n \in \mathbb{N}$, there exists a regular curve $t \to \mu_n(t)$ whose length is given by

$$\mathcal{L}_0^1(\mu_n(t)) = \frac{2}{\mathcal{H}^1(I^n)} \sum_{i=1}^{n} \mathcal{H}^1(I_i^n) \left(1 + \frac{(\mathcal{H}^1(I_i^n))}{3}\right), \tag{6.57}$$

such that

$$\mathbf{T}_n(A^n) = B^n, \tag{6.58a}$$

$$\mathbf{T}_n(B^n) = A^n, \tag{6.58b}$$

where $\mathbf{T}_n(x) := \mathbf{T}_n(0,1,x)$ and $\mathbf{T}_n(t,s,x)$ are the flow maps of $t \to \mu_n(t)$. Equations (6.58), examples 6.40 and 6.42 ensure that $\mathcal{T}_0^1(u_n^0) = u_n^1$. Therefore we obtain the inequality

$$\begin{aligned} d\Big((\mu_n, u_n^0), (\mu_n, u_n^1)\Big) &\leq \big(\mathcal{L}_0^1(\mu_n(t))^2 + \|\mathcal{T}_0^1(u_n^0) - u_n^1\|_{\mu_n(1)}\big)^{1/2} \\ &= \mathcal{L}_0^1(\mu_n(t))^2 \leq \frac{2}{\mathcal{H}^1(I^n)} \sum_{i=1}^{n} \mathcal{H}^1(I_i^n) \left(1 + \frac{(\mathcal{H}^1(I_i^n))}{3}\right). \end{aligned}$$

Passing to the limit we get

$$\varliminf_{n\to\infty} d\Big((\mu_n, u_n^0), (\mu_n, u_n^1)\Big) \leq 2,$$

therefore choosing $a > 16$ and $b > 1$ we get

$$d^*\Big((\mu^0, u^0), (\mu^1, u^1)\Big) \leq \varliminf_{n\to\infty} d\Big((\mu_n, u_n^0), (\mu_n, u_n^1)\Big) \leq 2 < d\Big((\mu^0, u^0), (\mu^1, u^1)\Big),$$

as desired.

Observe that, with minor but tedious modifications, it is possible to define the measures μ^i, μ_n^i, $i = 0, 1$, to be absolutely continuous with C^∞ density.

PROPOSITION 6.53. *With the same notation as above, suppose that for a certain regular curve $t \to \mu_t$ we are in case 2: that is, suppose that $\mathbf{T}(A) = B$ and $\mathbf{T}(B) = A$, where $\mathbf{T}(x) := \mathbf{T}(0,1,x)$ and the maps $\mathbf{T}(t,s,x)$ are the flow maps of $t \to \mu_t$. Then the length of the curve is at least $a/8$.*

Proof. For any two points $x_A \in A$ and $x_B \in B$ define the curve $[0,1] \ni t \to f(x_A, x_b, t) \in S^1$ (S^1 being the unit circle in $\mathbb{R}^2$) as

$$f(x_A, x_B, t) := \frac{\mathbf{T}(0,t,x_B) - \mathbf{T}(0,t,x_A)}{|\mathbf{T}(0,t,x_B) - \mathbf{T}(0,t,x_A)|}.$$

Being the maps $x \to \mathbf{T}(0,t,x)$ homeomorphism, they are injective: therefore the one above is a good definition and the function $t \to f(x_A, x_B, t)$ is continuous. Now let $P := (0, -1)$ and extend the function f to the interval $[-1, 2]$ by defining

$$f(x_A, x_B, t) := \frac{(t+1)(x_B - x_A) - tP}{|(t+1)(x_B - x_A) - tP|}, \quad \forall x \in [-1, 0),$$
$$f(x_A, x_B, t) := \frac{(2-t)(x_A - x_B) + (t-1)P}{|(2-t)(x_A - x_B) + (t-1)P|}, \quad \forall x \in (1, 2].$$

Given that the first component of $x_B - x_A$ is strictly positive, the above equations define, for each x_A, x_B, a continuous map from $[-1, 2]$ to S^1. Since we have $f(x_A, x_B, -1) = P = f(x_A, x_B, 2)$, f is a closed loop, and therefore defines an element of the omothopy group of S^1. Now observe that f varies continuously in x_A and x_B, so the map which associates to (x_A, x_B) the element of the omothopy group is continuous. Since the omothopy group of S^1 is discrete, this map has to be constant.

As a consequence we have that (at least) one of the following two must be true:

- for every $x_A \in A$, $x_B \in B$ the point P is in the image of $[0,1] \ni t \to f(x_A, x_B, t)$,
- for every $x_A \in A$, $x_B \in B$ the point $-P$ is in the image of $[0,1] \ni t \to f(x_A, x_B, t)$.

Suppose the first case is true (the other being similar). Define $B' := \{1\} \times [3a/4, a] \subset B$ and $A' := \{-1\} \times [0, a/4] \subset A$ and observe that for every $x_B \in B'$ and $x_A \in A'$ the second coordinate of $x_B - x_A$ is at least $a/2$, while our assumption implies that for some $t \in [0,1]$ the second coordinate of $\mathbf{T}(0,t,x_B) - \mathbf{T}(0,t,x_A)$ is negative. Therefore, defining the curves $[0,1] \ni t \to \gamma_x(t) := \mathbf{T}(0,t,x)$ for every $x \in A \cup B$, we have that $\mathcal{L}_0^1(\gamma_{x_A}(t)) + \mathcal{L}_0^1(\gamma_{x_B}(t)) \geq a/2$. The conclusion follows by the chain of inequalities:

$$\begin{aligned}
\int_0^1 \|v_t\|_{\mu_t} dt &= \int_0^1 \left(\int_{\mathbb{R}^2} |v_t(x)|^2 d\mu_t(x) \right)^{1/2} dt \geq \int_0^1 \int_{\mathbb{R}^2} |v_t(x)| d\mu_t(x) dt \\
&= \int_{\mathbb{R}^2} \int_0^1 |v_t(\mathbf{T}(0,t,x))| dt d\mu_0(x) = \int_{\mathbb{R}^2} \int_0^1 |\frac{d}{dt}\mathbf{T}(0,t,x)| dt d\mu_0(x) \\
&= \int_{\mathbb{R}^2} \mathcal{L}_0^1(\gamma_x) d\mu_0(x) \geq \int_{A' \cup B'} \mathcal{L}_0^1(\gamma_x) d\mu_0(x) \geq \frac{a}{8}.
\end{aligned}$$

□

PROPOSITION 6.54. *With the same notation of the example above, there exists a regular curve $t \to \mu_t$, such that $\mathbf{T}(A) = B$ and $\mathbf{T}(B) = A$ whose length is given by* $2(1 + a/3)$.

Proof. The idea is to move the mass as shown in the figure.

Step 1 Let $A' := \{-1\} \times [0, a/3]$, $B' := \{1\} \times [2/3a, a]$ and $T \in L^2_{\mu^0}$ be defined by

$$T(x_1, x_2) = \begin{cases} (x_1, x_2/3)\,, & (x_1, x_2) \in A, \\ (x_1, (2 + x_2)/3)\,, & (x_1, x_2) \in B. \end{cases}$$

Observe that the curve $t \to (Id + t(T - Id))_{\#}\mu^0$ is regular, that the support of the endpoint $\sigma := T_{\#}\mu^0$ is $A' \cup B'$ and that its length is equal to $\|T - Id\|_{\mu^0} = a/3$

Step 2 Define $S \in L^2_\sigma$ by

$$S(x_1, x_2) = \begin{cases} (1, x_2)\,, & (x_1, x_2) \in A', \\ (-1, x_2)\,, & (x_1, x_2) \in B'. \end{cases}$$

The curve $t \to (Id + t(S - Id))_{\#}\sigma$ is regular and its length is 2. Observe that the endpoint of the curve is the transport of σ through the reflection on the vertical axis.

Step 3 We are in a situation symmetric to the one of the first step. Reversing the construction given before we obtain a regular curve whose length is $a/3$.

Given that a piecewise regular curve is regular, the proof is achieved. □

COROLLARY 6.55. *With the same notation of the example above, there exists a regular curve $t \to \mu_n(t)$ whose flow maps satisfy equation* (6.58) *and whose length is given by equation* (6.57).

Proof. Fix n, then apply the construction given in the previous proposition to each one of the couples $(\{-1\} \times I^n_i, \{1\} \times I^n_i)$ for $i = 1, \ldots, n$. Given that during these movements the distance between the connected components of the support of μ_t is bounded from below by a positive constant uniformly on t, the resulting curve is regular. To conclude, observe that each couple $(\{-1\} \times I^n_i, \{1\} \times I^n_i)$ gives a contribution to the total mass of $\mathcal{H}^1(I^n_i)/\mathcal{H}^1(I^n)$. □

Bibliography

[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations, 10 (2005), No.3, pp. 309–360.

[2] M. Agueh, *Asymptotic behavior for doubly degenerate parabolic equations*, C.R. Math. Acad. Sci. Paris, 337 (2003), pp. 331–336.

[3] G. Alberti and L.Ambrosio, *A geometrical approach to monotone functions in* $\mathbb{R}^n$, Math. Z., 230 (1999), pp. 259–316.

[4] A. D. Aleksandrov, *A theorem on triangles in a metric space and some of its applications, inTrudy Mat. Inst. Steklov.*, v 28, Izdst. Akad. Nauk SSSR, Moscow, 1951, pp. 5-23.

[5] L. Ambrosio, *Metric space valued functions of bounded variation*, Ann. Sc. Norm. Sup. Pisa, 17 (1990), pp. 439–478.

[6] L. Ambrosio, *Minimizing movements*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5), 19 (1995), pp. 191–246.

[7] L. Ambrosio, *Lecture notes on optimal transport problem*, in Mathematical aspects of evolving interfaces, CIME summer school in Madeira (Pt), P.Colli and J.Rodrigues, eds., vol 1812, Spinger, 2003, pp. 1–52.

[8] L. Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math., 158 (2004), pp. 227–260.

[9] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variations and free discontinuity problems*, Oxford Mathematical Monographs, Clarendon Press, Oxford, 2000.

[10] L. Ambrosio and N. Gigli, *Construction of the parallel transport in the Wasserstein space*, Methods and Applications of Analysis, to appear, 2008.

[11] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Birkhäuser (2005).

[12] L. Ambrosio, S. Lisini and G. Savaré, *Stability of flows associated to gradient vector fields and convergence of iterated transport maps*, Manuscripta Math., 121 (2006), pp. 1–50.

[13] L. Ambrosio and P. Tilli, *Selected Topics on "analysis in metric spaces"*, Scuola Normale Superiore, Pisa, 2000.

[14] C. Baiocchi, *Discretization of evolution variational inequalities*, in Partial Differential Equations and the calculus of variations, Vol. I, F. Colombini, A. Marino, L. Modica, and S. Spagnolo, eds., Birkhäuser Boston, Boston, MA, 1989, pp. 59–92.

[15] J. -D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[16] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), pp. 375–417.

[17] H. Brezis, *Propriétés régularisantes de certains semi-groupes non linéaires*, Israel J. Math., 9 (1971), pp. 513–534.

[18] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No.5. Notas Matematica (50).

[19] H. Brezis, *Interpolation classes for monotone operatios*, Pacific Journal of Math, vol. 55, No.2, 1974, pp. 335–341.

[20] Y. Burago, M. Gromov, and G. Perel'man, *A. D. Aleksandrov spaces with curvatures bounded below*, Uspekhi Mat. Nauk, vol. 47, 1992, pp. 3–51,222.

[21] A. Brondsted, *On a lemma of Bishop and Phelps*,in Partial Differential Equations and Related Topics, J. A. Goldstein, ed., vol. 446 of Lecture Notes in Mathematics, Spinger, Berlin, 1975, pp. 65–74.

[22] Y. Burago, M. Gromov and G. Perel'man, *A. D. Aleksandrov spaces with curvatures bounded below*, Uspekhi Mat. Nauk, 47 (1992), pp. 3–51, 222.

[23] L. A. Caffarelli, M. Feldman and R. J. McCann, *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs*, J. Amer. Math. Soc., 15 (2002), pp. 1–26 (electronic).

[24] E. A. Carlen and W. Gangbo, *Constrained steepest descent in the 2-Wasserstein metric*, Arch. Ration. Mech Anal., 172 (2004), pp. 21–64.

[25] J. A. Carrillo, R. J. McCann and C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation nd mass transportation estimates*, Rev. Mat. Iberoamericana, 19 (2003), pp. 971–1018.

[26] J. A. Carrillo, R. J. McCann and C. Villani, *Contractions in the 2-Wassertein space and thermalization of granular media*, Arch. Ration. Mech. Anal., 179 (2006), No.2, pp. 217–263.

[27] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions*, Springer-Verlag, Berlin, 1977. Lecture Notes in Mathematics, Vol.580.

[28] D. Cordero-Erasquin, B. Nazaret and C. Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo-Niremberg inequalities*, Adv. Math., 182 (2004), pp. 307–332.

[29] E. De Giorgi, *New problems on minimizing movements*, in Boundary Value Problems for PDE and Applications, C. Baiocchi and J.L. Lions, eds., Masson, 1993, pp. 81-98.

[30] E. De Giorgi, A. Marino and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8), 69 (1980), pp. 180–187.

[31] H. Dietrich, *Zur c-Konvexität und c-Subdifferenzierbarkeit von Funktionalen*, Optimization, 19 (1998), pp. 355–371.

[32] R. J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math., 98 (1989), pp. 511–547.

[33] M. P. Do Carmo, *Riemannian Geometry*, Mathematics: Theory & Application, Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese Edition by Francis Flaherty.

[34] R. M. Dudley, *Probabilities and metrics*, Matematisk Institut, Aarhus Universitet, Aarhus, 1976. Convergence of laws on metric spaces, with a view to statistical testing, Lecture Notes Series, No. 45.

[35] R. M. Dudley, *Real analysis and probability*, Wadswoth & Brooks/Cole, Pacific Grove, California, 1989.

[36] L. C. Evans and W. Gangbo, *Differential equations methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., 137 (1999), pp. viii+66.

[37] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

[38] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969

[39] W. Gangbo, *The Monge mass transfer problem and its applications*, in Monge Ampere equation: applications to geometry an optimization (Deerfield Beach, FL, 1997), vol. 266 of Contemp. Math., Amer. Math. Soc., Providence, RI, 1999, pp. 79–104.

[40] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Math., 177 (1996), pp. 355–405.

[41] R. Jordan, D. Kinderlehrer and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17 (electronic).

[42] J. Jost, *Nonpositive curvature: geometric and analytic aspects*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1997.

[43] J. Lott, *Some geometric calculation on Wasserstein space*, Comm. Math. Phys., 277 (2008), pp. 423–427.

[44] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann.Math., (to appear).

[45] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.

[46] F. Otto, *Doubly degenerate diffusion equations as steepest descent*, Manuscript, (1996).

[47] F. Otto, *Evolution of microstructure in unstable media flow: a relaxational approach*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[48] R. Rossi and G. Savaré, *Gradient flow of non convex functionals in Hilbert spaces and applications*, tech. rep., IMATI-CNR, Pavia, 2004.

[49] K. T. Sturm, *Metric spaces of lower bounded curvature*, Exposition. Math., 17 (1999) pp. 35–47.

[50] C. Villani, *Optimal transportation, dissipative PDE's and functional inequalities*, in Optimal transportation and application (Martina Franca, 2001), vol. 1813 of Lecture Notes in Math., Springer, Berlin, 2003, pp. 53–89.

[51] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[52] C. Villani, *Optimal transport, old and new*, Springer Verlag, 2008.