

# THE TANGENT SPACE TO THE WASSERSTEIN SPACE: PARALLEL TRANSPORT AND OTHER APPLICATIONS

CHARLES BERTUCCI <sup>A</sup>

**ABSTRACT.** We propose a new notion of the *formal* tangent space to the Wasserstein space  $\mathcal{P}(X)$  at a given measure. Modulo an integrability condition, we say that this tangent space is made of functions over  $X$  which are valued in the probability measures over the tangent bundle to  $X$ . This generalization of previous concepts of tangent spaces allows us to define appropriate notions of parallel transport,  $\mathcal{C}^{1,\alpha}$  regularity over  $\mathcal{P}(X)$  and translation of a curve over  $\mathcal{P}(X)$ .

## CONTENTS

|   |    |
|---|----|
| Introduction  | 1  |
| 1. Notation and standard results  | 2  |
| 2. Tangent space to $\mathcal{P}_p(\mathbb{R}^d)$   | 3  |
| 3. Parallel transport on $\mathcal{P}_2(\mathbb{R}^d)$  | 5  |
| 4. Links with Lions' lift   | 10 |
| 5. Tangent space and parallel transport on $\mathcal{P}(\mathcal{O})$   | 11 |
| 6. Comparing elements in different tangent spaces and regularity of elements of $\mathcal{P}(\mathcal{O}) \rightarrow \mathbb{R}$ | 12 |
| 7. Translated curves and application in optimal control problems  | 18 |
| Conclusions   | 26 |
| Acknowledgments   | 26 |
| References  | 26 |

## INTRODUCTION

We introduce a new notion of tangent space to the space of probability measures over different finite dimensional metric sets, when it is equipped with a Wasserstein metric (i.e. a Wasserstein space). This notion is more general than the one previously considered in the literature in the sense that it contains strictly more elements. Our notion basically sees elements in the tangent space as measured valued functions, instead of simply functions, in the same spirit as the famous relaxations of either Young or Kantorovich.

We then proceed by using this point of view to define a notion of parallel transport on the space of probability measures (Section 3). Later, we build on this notion of parallel transport to define  $\mathcal{C}^{1,\alpha}$  spaces of real valued functions over the space of probability measures equipped with Wasserstein metrics (Section 6). We finally use a similar point

of view to explain how we can translate curves and how it is helpful in optimal control (Section 7).

We feel the need to insist upon the fact that, despite the maybe poor use of the differential geometry vocabulary in what follows (tangent space, parallel transport,...), we do not claim that the Wasserstein space is a smooth Riemannian manifold onto which such concepts are well defined. Nonetheless, we claim that adopting such a geometric point of view leads to the definition of tools that are useful in the analysis of several problems.

More detailed bibliographical comments are given in the core of the paper, but we now present briefly the context. Since the seminal work of Otto [20] and the influential book of Ambrosio, Gigli and Savaré [2], Wasserstein spaces are often thought of equipped with a formal Riemannian structure. The tangent space to the Wasserstein space  $\mathcal{P}(X)$  is widely taken as the closure of gradients of smooth functions over the metric space  $X$ , closure which is taken for some metric which depends on the probability measure at which we are taking the tangent space. This point of view is mostly induced by an Eulerian description of flows on  $\mathcal{P}(X)$ . It has been extremely fruitful. It also allowed to define parallel transports along flows which can be written in an Eulerian way with sufficient regularity [1, 18]. We provide here a more Lagrangian point of view which allows us to define parallel transport quite generally, along Lagrangian flows. We then present some natural consequences of this (Lagrangian) notion of parallel transport. Our point of view is clearly in line with the lift of Lions on the space of random variables introduced in [19], which we generalize to more involved settings, see Section 4.

## 1. NOTATION AND STANDARD RESULTS

Let  $d \geq 1$  be an integer. For a topological space  $O$ ,  $\mathcal{P}(O)$  stands for the set of Borel probability measures on  $O$ . For  $p \in [1, \infty)$ , we consider the space  $\mathcal{P}_p(\mathbb{R}^d) := \{m \in \mathcal{P}(\mathbb{R}^d) | \int_{\mathbb{R}^d} |x|^p m(dx) < \infty\}$  equipped with the  $p$ -Wasserstein distance  $W_p$  defined by

$$W_p(\mu, \nu) = \left( \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^{2d}} |x - y|^p \gamma(dx, dy) \right)^{\frac{1}{p}},$$

where  $\Pi(\mu, \nu)$  is the set of couplings between  $\mu$  and  $\nu$ , i.e. the set of measures  $\gamma \in \mathcal{P}(\mathbb{R}^{2d})$  such that  $(\pi_1)_\# \gamma = \mu$  and  $(\pi_2)_\# \gamma = \nu$ , where, for  $x, y \in \mathbb{R}^d$ ,  $\pi_1(x, y) = x$  and  $\pi_2(x, y) = y$ , and  $f_\# m$  denotes the image of the measure  $m$  by the map  $f$ . The maps  $\pi_1$  and  $\pi_2$  shall be used more generally on any couple as the natural projection maps. The set  $(\mathcal{P}_p(\mathbb{R}^d), W_p)$  is a complete separable metric space.

Below,  $\mathcal{O}$  shall be the closure of a smooth open bounded subset of  $\mathbb{R}^d$ . For  $p \in [1, \infty)$ ,  $(\mathcal{P}(\mathcal{O}), W_p)$  is a complete separable compact metric space.

For  $p \in (1, \infty)$ , we denote  $p' = p/(p-1)$ . We then define  $\mathcal{P}_{pp'}(\mathbb{R}^d \times \mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^{2d}) | (\pi_1)_\# \mu \in \mathcal{P}_p(\mathbb{R}^d), (\pi_2)_\# \mu \in \mathcal{P}_{p'}(\mathbb{R}^d)\}$ . This set can be equipped with a sort of  $(p, p')$ -Wasserstein distance (see Section 10.3 in [2]), that we do not detail here.

Below,  $M$  shall denote a smooth bounded manifold without boundary. Given  $x \in M$ ,  $T_x M$  is the tangent space to  $M$  at  $x$  and  $TM := \cup_{x \in M} \{x\} \times T_x M$  is the tangent bundle of  $M$ , which shall also be identified sometimes with  $\cup_{x \in M} T_x M$ . Given a smooth curve  $\theta : [0, 1] \rightarrow M$ , and  $p_0 \in T_{\theta(0)} M$ , the parallel transport of  $p_0$  along  $\theta$  is characterized

as the unique curve  $p : [0, 1] \rightarrow TM$  such that  $\nabla_{\theta'(t)} p(t) = 0$  together with  $p(0) = p_0$ . The set  $(\mathcal{P}(M), W_p)$  is a complete separable compact metric space for any  $p \in [1, \infty)$ , where  $W_p$  is defined by replacing  $|x - y|$  by the geodesic distance between  $x$  and  $y$  in the definition above.

## 2. TANGENT SPACE TO $\mathcal{P}_p(\mathbb{R}^d)$

**2.1. Main definitions.** The study of evolution equations on  $\mathcal{P}_p(\mathbb{R}^d)$  lead to the need of working with the tangent space to  $\mathcal{P}_p(\mathbb{R}^d)$  at some point  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ . An early and influential work in this direction is [20] which was in fact concerned with  $L^1$  spaces. Following [2] Section 8.4, several works including namely [16, 11] have then argue that the tangent space to  $(\mathcal{P}_p(\mathbb{R}^d), W_p)$  at  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$  is

$$(1) \quad \overline{\nabla_x C^\infty(\mathbb{R}^d, \mathbb{R})}^{L^{p'}((\mathbb{R}^d, \mu), \mathbb{R}^d)}.$$

Here, we present a different point of view and consider instead the tangent space to  $\mathcal{P}_p(\mathbb{R}^d)$  at  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$  to be given by

$$(2) \quad T_\mu(\mathcal{P}_p(\mathbb{R}^d)) := \left\{ \psi : \mathbb{R}^d \rightarrow \mathcal{P}(\mathbb{R}^d), \text{ measurable } \left| \int_{\mathbb{R}^{2d}} |z|^{p'} \psi_x(dz) \mu(dx) < \infty \right. \right\}.$$

By disintegration, it is equivalent to considering the set of probability measures  $\gamma \in \mathcal{P}_{pp'}(\mathbb{R}^{2d})$  such that  $(\pi_1)_\# \gamma = \mu$  and we shall sometimes say with a slight abuse of notation that some  $\gamma(dx, dz) = \mu(dx) \psi_x(dz)$  belongs to  $T_\mu(\mathcal{P}_p(\mathbb{R}^d))$  if  $\psi$  does. Pursuing in this direction, we define  $T\mathcal{P}_p(\mathbb{R}^d) := \cup_{\mu \in \mathcal{P}_p(\mathbb{R}^d)} \{ \mu \} \times T_\mu \mathcal{P}_p(\mathbb{R}^d)$  which will be identified with  $\mathcal{P}_{pp'}(\mathbb{R}^{2d})$ .

Of course, for all elements  $\phi$  in the tangent space defined in (1), we have an associated elements in  $T_\mu(\mathcal{P}_p(\mathbb{R}^d))$  by considering the map  $\psi : x \rightarrow \delta_{\phi(x)}$ , such elements shall be called deterministic.

We provide in the next sections applications which we believe justify precisely this notion of tangent space. We finish this section by giving some heuristics to support our definition of tangent space.

### 2.2. Heuristics.

**2.2.1. What can be found in the literature.** A similar notion of tangent space was in fact considered in [17] and Section 12.4 in [2] (the so-called velocity plans), with the restriction to maps  $\psi$  satisfying some sort of optimality, arising from the fact that only couplings arising from geodesics of optimal transport were considered. In regard to their definition, our main contribution is simply to avoid this restriction of  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ . Furthermore, we point out the work [17, 3] which, despite not identifying the same tangent space as us, studies it and provides interesting results concerning the decomposition of  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  into (1) and what could be interpreted as its orthogonal space.

Somehow, the comparison result we established in [5] is a natural application of the choice of tangent space. Indeed, by considering extended super-differentials (see Section 10.3 in [2]), we showed that comparison principles are immediate for Hamilton-Jacobi equation in  $\mathcal{P}(\mathbb{T}^d)$ . Hence, we allowed to seek for elements in super-differentials which

are in  $T\mathcal{P}_p(\mathbb{R}^d)$ , and not only in some set defined as in (1). This sort of ideas were later also used in [10, 8].

More generally, the notion of extended super-differentials of [2] basically seeks to define what does it mean for an element of  $T_\mu(\mathcal{P}_p(\mathbb{R}^d))$  to be in the super-differential of a function  $U : \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathbb{R}$  at the point  $\mu$ .

Furthermore, recent developments [12, 13, 14] in the precise characterization of not so smooth dynamics on  $\mathcal{P}_p(\mathbb{R}^d)$  have made a systematic use of so-called multi valued probability vector fields, which are somehow vector fields on the tangent space we defined above.

**2.2.2. Young measures.** The main issue when dealing with (1) is that the latter space behaves quite badly with  $\mu$ . The next example is a good illustration of this phenomenon.

**Example 2.1.** Let  $d = 1$ ,  $\mu = \delta_0$  and consider for  $\epsilon > 0$ ,  $\mu_\epsilon := \mathcal{N}(0, \epsilon)$ , the Gaussian distribution of mean 0 and variance  $\epsilon$ . Clearly  $W_2(\mu, \mu_\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . For any  $\epsilon > 0$ , the function  $f(x) = \text{sign}(x)$  if  $x \neq 0$ , is in  $L^2((\mathbb{R}, \mu_\epsilon), \mathbb{R})$ . However, the statement  $f \in L^2((\mathbb{R}, \mu), \mathbb{R})$  is far from being clear. More generally  $L^2((\mathbb{R}, \mu), \mathbb{R})$  has not much to do with  $L^2((\mathbb{R}, \mu_\epsilon), \mathbb{R})$ , even when  $\epsilon$  is small but non-zero.

Hence, it seems that when dealing with (1), we loose a lot of information at some point compared to what can happen in a close neighborhood. In our opinion, it is reminiscent of an old problem faced in calculus of variations which has been solved by Young [21]. He introduced what are now called Young measures, which allow to capture some oscillatory, random or complex behavior of sequences of functions by considering objects similar to the ones of (2). Usually in Young measures, the reference measure  $\mu$  is fixed and we are simply taking limits of functions. Here, we claim that if we allow  $\mu$  to vary also, then we naturally end up with  $T_\mu(\mathcal{P}_p(\mathbb{R}^d))$  if we want to keep all this information.

Typical problems in Young measures are to characterize the limit of functions  $f_n : [0, 1] \rightarrow \mathbb{R}$  defines as the almost everywhere derivative of  $g_n$ , a piecewise affine function which takes value 0 on  $\frac{k}{2^n}$  if  $k$  is even, and value 1 if  $k$  is odd, and which is smooth outside of these points. Typically, Young measures allow to state that, in some sense,  $f_n$  converges toward the measured valued map  $\psi : [0, 1] \rightarrow \mathcal{P}(\mathbb{R})$  which is constant and takes value  $\frac{1}{2}(\delta_{-1} + \delta_1)$ .

In Example 2.1, in some sense, it is natural to say that the limit version of  $f \in L^2((\mathbb{R}, \mu_\epsilon), \mathbb{R})$  is to say that the function defined only on  $\{0\}$  by  $(0 \rightarrow \frac{1}{2}(\delta_{-1} + \delta_1))$  is an element of  $T_\mu(\mathcal{P}(\mathbb{R}))$ .

The point of view of Young measures can be interpreted as the fact that a relaxation of a functional space usually leads to considering measured value functions. This is quite in the spirit of the powerful Kantorovich relaxation of optimal transport, in fact it is an omnipresent idea in optimization, analysis or calculus of variations that we feel deserved to be highlighted here.

**2.2.3. Sheaf structure of the tangent spaces.** When trying to equip  $(\mathcal{P}_2(\mathbb{R}^d), W_2)$  with the tangent space of (1), it is clear that we are quite far from obtaining the structure

of a tangent bundle. Indeed, the nature of the candidate tangent space varies too much with  $\mu$  and makes any sort of local trivialization impossible.

However, the identification of  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  with a subset of  $\mathcal{P}(\mathbb{R}^{2d})$  that we proposed hints a sheaf structure for this tangent space, that we establish in the next proposition.

**Proposition 2.2.** *The functor  $\mathcal{F}$  which associates to any open set  $U \subset (\mathcal{P}_2(\mathbb{R}^d), W_2)$   $\mathcal{F}(U) := \{\varphi : U \rightarrow \mathcal{P}_2(\mathbb{R}^{2d}) | \varphi \text{ is continuous}, (\pi_1)_\# \circ \varphi = Id\}$  defines a sheaf on  $\mathcal{P}_2(\mathbb{R}^d)$  when associated to the standard restriction morphisms. Furthermore, the stalk  $\mathcal{F}_\mu$  of  $\mathcal{F}$  over some  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  is given by  $\{\gamma \in \mathcal{P}_2(\mathbb{R}^{2d}) | (\pi_1)_\# \gamma = \mu\}$ , that is by  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ .*

*Proof.* The result is trivial as all the properties of a sheaf are automatically verified. Furthermore, the map  $(\pi_1)_\# \cdot$  is continuous on  $\mathcal{P}_2(\mathbb{R}^{2d})$ , which naturally yields the structure of the stalk.  $\square$

**2.2.4. Lagrangian vs Eulerian dynamics.** In finite dimension, the tangent space to a manifold  $M$  at a point  $x \in M$  can be thought of as being the set of speeds  $\dot{\theta}(0)$  of smooth curves  $\theta$  such that  $\theta(0) = x$ . In  $\mathcal{P}_p(\mathbb{R}^d)$ , restricting our attention to geodesics and having in mind the Benamou-Brenier dynamic reformulation of optimal transport [4], we know that, formally geodesics  $(m_t)_{t \in [-1,1]}$  are given as solution of continuity equations

$$\partial_t m + \operatorname{div}_x(\nabla_x \phi(t, x)m) = 0 \text{ in } (-1, 1) \times \mathbb{R}^d,$$

for some real valued function  $\phi : [-1, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$ . Hence, the space (1), which thus inherits the Eulerian formulation of geodesics on  $\mathcal{P}_p(\mathbb{R}^d)$ , is the closure of possible speeds  $\nabla_x \phi(0, \cdot)$ .

A more Lagrangian approach consists in looking at dynamics  $(m_t)_{t \in [-1,1]}$  expressed as the law of processes  $(X_t)_{t \in [-1,1]}$ , on a standard probabilistic space  $(\Omega, \mathcal{A}, \mathbb{P})$ . At a Lagrangian level, smoothness implies that  $X_t(\omega)$  is a smooth function of  $t$ , for almost every  $\omega$ . In this context, the speed of the curve is the collection  $(\dot{X}_t(\omega))_{\omega \in \Omega}$ . Keeping track of all the speeds then amounts to being interested in the set of pairs  $(X_0(\omega), \dot{X}_0(\omega))_{\omega \in \Omega}$ , and the fact that we are only interested in the law of such objects leads us to consider  $\gamma := \mathcal{L}(X_0, \dot{X}_0) \in \mathcal{P}(\mathbb{R}^{2d})$ . We expand on this in Sections 4 and 7.1 below.

**2.2.5. Particles approximation.** In order to tackle problems on  $\mathcal{P}_p(\mathbb{R}^d)$ , approximations of measures by combinations of  $N$  Dirac masses can be considered before passing to the limit  $N \rightarrow \infty$ . At the level  $N < \infty$ , we want to approximate  $\mathcal{P}_p(\mathbb{R}^d)$  with  $(\mathbb{R}^d)^N$  equipped with an appropriate  $\ell^p$  topology. For  $x \in (\mathbb{R}^d)^N$ , the tangent space  $T_x((\mathbb{R}^d)^N)$  is simply given by  $(\mathbb{R}^d)^N$ . Let  $(x_N)_{N \geq 1}$  be a sequence such that  $x_N \in (\mathbb{R}^d)^N$  and  $\lim_{N \rightarrow \infty} \frac{1}{N} \delta_{x_N^i} = \mu \in \mathcal{P}_p(\mathbb{R}^d)$  and  $(y_N)_{N \geq 1}$  be such that  $y_N \in T_{x_N}(\mathbb{R}^d)^N$  for all  $N \geq 1$ . We then say that  $(y_N)$  is going to approximate an element  $\psi \in T_\mu(\mathcal{P}_p(\mathbb{R}^d))$  if  $\lim_{N \rightarrow \infty} \frac{1}{N} \delta_{(x_N^i, y_N^i)} = \mu(dx)\psi_x(dz)$ . We then observe the difference with eventual deterministic tangent spaces such as (1) as the fact that there is no restriction of the type  $x_N^i = x_N^j \Rightarrow y_N^i = y_N^j$ .

### 3. PARALLEL TRANSPORT ON $\mathcal{P}_2(\mathbb{R}^d)$

The first application of our concept of tangent space is to define a notion of parallel transport on  $\mathcal{P}_p(\mathbb{R}^d)$ . The notion of parallel transport usually requires three main

concepts such as: tangent spaces (or more generally vector fields), a notion of smooth curves, and a notion of derivative along a smooth curve  $\theta$ . A parallel transport is then defined as a curve valued in the tangent space over the curve  $\theta$  whose derivative in the direction that follows  $\theta$  is 0. We make precise such an idea below.

We shall first present the notion of smooth curves along which are going to parallel transport our vectors. As the geometry of  $\mathbb{R}^d$  is euclidean, the notion of derivative along a curve shall be quite trivial. The case of  $\mathcal{P}(M)$  treated below will require some care in this aspect. Furthermore, we restrict the following to the case  $p = 2$  to lighten notation and comment at the end of this section on the case  $p \neq 2$ .

We finish this section by comparing our notion to the one introduced and studied in [1].

**3.1. Smooth curves.** A curve on  $\mathcal{P}_2(\mathbb{R}^d)$  is typically a map  $\theta : [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ . Saying that a curve is continuous or Lipschitz is clear in this setting. We want to consider curves which are more regular than simply Lipschitz. Since  $\mathcal{P}_2(\mathbb{R}^d)$  is not a proper normed vector space, there is no intrinsic way to do so and we explain below the underlying difficulties and the choice we make.

**3.1.1. What we do not do.** Given the nature of  $\mathcal{P}_2(\mathbb{R}^d)$ , it is not obvious how to define  $C^1$  curves. For instance, we could understand curves  $\theta$  satisfying  $\theta'(t) = cst$ , when embedding  $\mathcal{P}_2(\mathbb{R}^d)$  in the space of signed measures equipped with total variation. But by doing so we would miss curves of the form  $(\delta_{x(t)})_{t \in [0, 1]}$  for  $(x(t))_{t \in [0, 1]}$ , a smooth curve on  $\mathbb{R}^d$ . The latter type of curve is typically characterized by being the unique solution of equations of the form

$$(3) \quad \partial_t m + \operatorname{div}_x(\phi(t, x)m) = 0 \text{ on } (0, 1) \times \mathbb{R}^d,$$

where  $\operatorname{div}_x$  is the divergence operator on  $\mathbb{R}^d$  and  $\phi : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a smooth vector field, Lipschitz in  $x$ , say uniformly in  $t$ . However, restricting our attention to this type of curves would make us miss the geodesics of the distance  $W_2$ , which are typically characterized as being the solution of such equations for non-smooth functions  $\phi$ . The latter makes the discussion about what is a smooth curve quite complex, thus we do not adopt this point of view here.

**3.1.2. What we do.** We take a so-called Lagrangian point of view and consider more general curves  $\theta$  on  $\mathcal{P}_2(\mathbb{R}^d)$ , namely one which are of the form  $\theta(t) = (e_t)_\# \eta$  for some  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$ , where  $e_t(f) = f(t)$ , for  $f \in \mathcal{C}([0, 1], \mathbb{R}^d)$ , i.e.  $e_t$  is the evaluation mapping at time  $t$ . We say that  $\eta$  is  $\mathcal{C}^1$  if it is concentrated on  $\mathcal{C}^1$  curves. Such a notion of regularity does not naturally translate to a regularity for  $\theta$ , for instance because of the clear multiplicity of  $\eta$  which generate the same  $\theta$ .

In fact, through Ambrosio's superposition principle which is presented in details in Section 7, there is a correspondance between this Lagrangian point of view and the previous Eulerian one. It just happens that the Lagrangian point of view is much more convenient for what follows.

**3.2. Vector fields and their derivatives.** A vector field above  $\mathcal{P}_2(\mathbb{R}^d)$  is typically a map  $V$  defined on  $\mathcal{P}_2(\mathbb{R}^d)$  such that  $V(\mu) \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  for all  $\mu$ . To define parallel transport, we are interested in vector field along a curve, which we decided to be given

by a Lagrangian formulation, hence we shall naturally be interested in Lagrangian vector field.

A vector field along a Lagrangian curve  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  is a probability measure  $\Psi \in \mathcal{P}(\mathcal{C}[0, 1], \mathbb{R}^{2d})$  such that  $(p_1)_\# \Psi = \eta$ , where  $p_1 : \mathcal{C}([0, 1], \mathbb{R}^{2d}) \rightarrow \mathcal{C}([0, 1], \mathbb{R}^d)$  is defined by  $p_1(z)(t) = (\pi_1)(z(t))$ . If  $\Psi$  is concentrated along smooth curves in  $\mathbb{R}^{2d}$ , its derivative along  $(p_1)_\# \Psi$  is by definition the curve  $(\partial_2)_\# \Psi$  where  $\partial_2 : \mathcal{C}^1([0, 1], \mathbb{R}^{2d}) \rightarrow \mathcal{C}([0, 1], \mathbb{R}^d)$  is defined by  $\partial_2(z)(t) = \frac{d}{dt}(\pi_2(z(t)))$ . More generally,  $\partial_2$  is well defined on curves on  $\mathbb{R}^{2d}$  such that the second component is in  $\mathcal{C}^1$ . Such a space is denoted  $\mathcal{C}_y^1([0, 1], \mathbb{R}^{2d})$ .

**3.3. Main definition and properties.** Since the geometry of  $\mathbb{R}^d$  is rather trivial, the parallel transport in  $\mathbb{R}^d$  simply consists in translating the tangent vector. At the level of  $\mathcal{P}_2(\mathbb{R}^d)$ , this simplicity translates into the following definition.

**Definition 3.1.** Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  such that  $(e_0)_\# \eta = \mu$ . A parallel transport of  $\psi$  along  $\eta$  is a measure  $\Psi \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^{2d}))$  concentrated on  $\mathcal{C}_y^1([0, 1], \mathbb{R}^{2d})$  such that

- $(p_1)_\# \Psi = \eta$ ,
- $(\partial_2)_\# \Psi = \delta_0$ ,
- $(e_0)_\# \Psi(dx, dz) = \mu(dx)\psi_x(dz)$ .

**Remark 3.2.** Remark that no integrability condition is required here, as it follows from the definition as the next result shows.

The following result makes transparent that a parallel transport yields a curve in the tangent sheaf  $T\mathcal{P}_2(\mathbb{R}^d)$ .

**Proposition 3.3.** Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ ,  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  such that  $(e_0)_\# \eta = \mu$  and  $\Psi$  a parallel transport of  $\psi$  along  $\eta$ . Then, for all  $t \in [0, 1]$ ,  $(e_t)_\# \Psi = (e_t)_\# \eta(dx)\psi_{t,x}(dz)$  for some  $\psi_t \in T_{(e_t)_\# \eta}(\mathcal{P}_2(\mathbb{R}^d))$ .

*Proof.* Let  $(X, Z)$  be a process of law  $\Psi$ . For  $t \in [0, 1]$ ,  $(X_t, Z_t)$  is such that  $\mathcal{L}(X_t) = (e_t)_\# \eta$ . Moreover, since  $Z_t = Z_0$  and that  $\mathcal{L}(Z_0) \in \mathcal{P}_2(\mathbb{R}^d)$  since  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ , the result follows.  $\square$

The existence of parallel transport is proved next to be a simple consequence of the gluing Lemma.

**Proposition 3.4.** Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  such that  $(e_0)_\# \eta = \mu$ . There exists a parallel transport of  $\psi$  along  $\eta$ .

*Proof.* By gluing together  $(e_0, Id)_\# \eta(d\theta)$  and  $\mu(dx)\psi_x(dz)$  along their first coordinate, we obtain a probability measure  $\Gamma : \mathbb{R}^d \times \mathcal{C}([0, 1], \mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathcal{C}([0, 1], \mathbb{R}^{2d})$ ,  $(x, \theta, z) \rightarrow (t \rightarrow (\theta(t), z))$ . The probability  $\mathcal{T}_\# \Gamma$  is a required parallel transport.  $\square$

The next example shows what parallel transport is along curves which are simply constructed by interpolation of a coupling between two measures. In particular, such curves model all geodesics in  $\mathcal{P}_2(\mathbb{R}^d)$ .

**Example 3.5.** Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ . Let  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  being defined as the law of  $((1-t)X' + tY')$  for  $(X', Y')$  a couple of random variables of joint law  $\gamma \in \mathcal{P}_2(\mathbb{R}^{2d})$ . Then the construction of the parallel transport is somehow more understandable. Consider  $(X, Y, Z)$  random variables such that  $\mathcal{L}(X, Y) = \gamma$  and  $\mathcal{L}(X, Z) = \mu(dx)\psi_x(dz)$ . Then a parallel transport of  $\psi$  along  $\eta$  is given by the law of  $t \mapsto ((1-t)X + tY, Z)$ . Note that here also, the existence of  $(X, Y, Z)$  is guaranteed by the gluing Lemma.

The next result relies heavily on the fact that the parallel transport in  $\mathbb{R}^d$  is trivial, in the sense that it only depends on the starting and end points.

**Proposition 3.6.** Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ ,  $\eta_1, \eta_2 \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  such that  $(e_0)_\# \eta_i = \mu$  for  $i = 1, 2$  and  $\Psi$  a parallel transport of  $\psi$  along  $\eta_1$ . If  $(e_0, e_1)_\# \eta_1 = (e_0, e_1)_\# \eta_2$ , then  $\Psi$  is also a parallel transport of  $\psi$  along  $\eta_2$ .

We leave this immediate proof to the reader.

**Remark 3.7.** This result justifies that, when working on probability measures over  $\mathbb{R}^d$ , we can talk about a parallel transport  $\Psi$  of  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  along a coupling  $\gamma \in \mathcal{P}(\mathbb{R}^{2d})$  when  $(\pi_1)_\# \gamma = \mu$ .

In general, there is no uniqueness of a parallel transport with our definition as the next example shows.

**Example 3.8.** Let  $d = 1$  and  $\mu := \delta_0$ . Let  $\eta$  be given by  $\frac{1}{2}(\delta_{\theta^1} + \delta_{\theta^2})$  where  $\theta^i(t) = 2(\frac{3}{2} - i)t$ . Let  $\psi_0 = \frac{1}{2}(\delta_0 + \delta_2)$ . Then the laws of the three following processes are all parallel transports of  $\psi$  along  $\eta$ ,

$$\begin{cases} \{W(t) = (t, 0), t \in [0, 1]\} \text{ with probability } \frac{1}{2}, \\ \{W(t) = (-t, 2), t \in [0, 1]\} \text{ with probability } \frac{1}{2}, \end{cases} \quad \begin{cases} \{M(t) = (t, 2), t \in [0, 1]\} \text{ with probability } \frac{1}{2}, \\ \{M(t) = (-t, 0), t \in [0, 1]\} \text{ with probability } \frac{1}{2}, \end{cases}$$

$$\begin{cases} \{L(t) = (t, 2), t \in [0, 1]\} \text{ with probability } \frac{1}{4}, \\ \{L(t) = (t, 0), t \in [0, 1]\} \text{ with probability } \frac{1}{4}, \\ \{L(t) = (-t, 0), t \in [0, 1]\} \text{ with probability } \frac{1}{4}, \\ \{L(t) = (-t, 2), t \in [0, 1]\} \text{ with probability } \frac{1}{4}. \end{cases}$$

This example shows that as soon as there are multiple ways to obtain our main coupling through the gluing Lemma in the proof of Proposition 3.4, there will be multiple parallel transports. On the converse, the next result shows that when one of the two couplings involved in the proof of Proposition 3.4 is deterministic, then there is a unique parallel transport. Similar remarks about the uniqueness of the law given by the gluing Lemma are of course already well-known, see for instance Lemma 5.3.2 in [2].

**Proposition 3.9.** Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  such that  $(e_0)_\# \eta = \mu$ . Assume that at least one of the two following statements holds.

- The function  $\psi$  is of the form  $x \rightarrow \delta_{v(x)}$  for some measurable  $v \in L^2_\mu(\mathbb{R}^d, \mathbb{R}^d)$ .
- There exists  $f \in L_t^\infty(W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d))$  such that  $\eta$  is concentrated on curves  $\theta$  solutions of  $\theta'(t) = f(t, \theta(t))$ .

Then, there exists a unique parallel transport of  $\psi$  along  $\eta$ .

*Proof.* Consider two parallel transport  $\Psi_1$  and  $\Psi_2$  of  $\psi$  along  $\eta$ , and two processes  $(X^1, Z^1), (X^2, Z^2)$  of respective law  $\Psi_1$  and  $\Psi_2$ .

In the first case, we find that for  $i = 1, 2$ , the process  $(X^i, Z^i)$  is of the form  $(X_t^i, v(X_0^i))_{t \in [0,1]}$ . Since  $X^1$  and  $X^2$  both have law  $\eta$ , it follows that  $\Psi_1 = \Psi_2$ .

In the second case, since the flow  $\phi : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  associated to the ODE  $\theta'(t) = f(t, \theta(t))$  is well defined, it follows that  $X^i$  is of the form  $(\phi(t, X_0^i))_{t \in [0,1]}$ . Hence, for  $i = 1, 2$ , the process  $(X^i, Z^i)$  is of the form  $((\phi(t, X_0^i), Z^i))_{t \in [0,1]}$ . Since  $(X_0^1, Z^1)$  and  $(X_0^2, Z^2)$  have law  $\mu(dx)\psi_x(dz)$ , it also follows that  $\Psi_1 = \Psi_2$ .  $\square$

**Remark 3.10.** More general condition on  $\eta$  are sufficient. In fact, it suffices in this case, that  $\eta$  is concentrated on curves  $\theta$  such that there exists a measurable map  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $\theta(1) = T(\theta(0))$ .

Since parallel transport is uniquely defined when we start from a deterministic element of the tangent space, one might wonder why bother with the "extended" tangent space that we proposed. The next example shows that, even if we start with a deterministic element of the tangent space, and thus that the parallel transport is uniquely defined, the transported element might end up being a non-deterministic element of the tangent space at the final point.

**Example 3.11.** Let  $d = 1$ ,  $\mu$  be the normal distribution and  $\psi : x \rightarrow \delta_x \in T_\mu(\mathcal{P}_2(\mathbb{R}))$ . Let  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}))$  be given by the coupling  $\mu(dx)\delta_0(dy)$ . Then, the (unique) parallel transport  $\Psi$  of  $\psi$  along  $\eta$  is such that  $(e_1)_\# \Psi(dx, dz) = \delta_0(dx)\mu(dz)$ . In particular, the associated element  $\psi' \in T_{\delta_0}(\mathcal{P}_2(\mathbb{R}))$  is not deterministic, as it associates  $\mu$  to the point 0.

We also have the following obvious time reversibility of parallel transports.

**Proposition 3.12.** Let  $\tau(\theta)$  be the reverse parametrization by time of a curve  $\theta \in \mathcal{C}([0, 1], X)$  for some topological space  $X$ , i.e.  $\tau(\theta)(t) = \theta(1-t)$ .

If  $\Psi$  is a parallel transport along  $\eta$ , then  $\tau_\# \Psi$  is a parallel transport along  $\tau_\#\eta$ .

Finally, we present another example to highlight the fact that our notion of parallel transport indeed depends on the Lagrangian path.

**Example 3.13.** Let  $\mu$  be the uniform distribution on  $[0, 1]$ , and consider  $\psi : x \rightarrow \delta_x \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ . Let  $U$  be a random variable of law  $\mu$ ,  $\eta_1 := \mathcal{L}((U)_{t \in [0,1]})$  and  $\eta_2 = \mathcal{L}((tU + (1-t)(1-U))_{t \in [0,1]})$ . Then, for all  $t \in [0, 1]$ ,  $(e_t)_\# \eta_1 = (e_t)_\# \eta_2$ , however, the parallel transport of  $\psi$  along  $\eta_1$  and  $\eta_2$  are clearly different.

**3.4. The case of  $\mathcal{P}_p(\mathbb{R}^d)$ .** The extension of the above to  $\mathcal{P}_p(\mathbb{R}^d)$  does not raise any particular difficulty, mainly because the integrability conditions follows from the definition, as Proposition 3.3 shows. Hence, we do not give more details here.

**3.5. Link with a previous notion of parallel transport on  $\mathcal{P}_2(\mathbb{R}^d)$ .** In [1, 18], the authors introduced another notion of parallel transport on  $\mathcal{P}_2(\mathbb{R}^d)$  (and in fact on the space of probability measure over a manifold as well) that we do not present in full details, but rather give our interpretation of it, in order to compare it with our notion.

In [1, 18], the authors are mostly interested with what we can call Eulerian paths on  $\mathcal{P}_2(\mathbb{R}^d)$ , that is those that can be represented as solution to a continuity equation of the form

$$\partial_t m + \operatorname{div}_x(\alpha_t m) = 0 \text{ in } (0, 1) \times \mathbb{R}^d,$$

where  $(m_t)_{t \in [0, 1]} \in \mathcal{C}([0, 1], \mathcal{P}_2(\mathbb{R}^d))$  is the path in question and  $\alpha : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ , the velocity that moves the underlying particles. We refer to Section 7.1 below for more insights and references on such dynamics.

The authors then show that under certain regularity assumptions on  $\alpha$ , the ODE  $\dot{x} = \alpha(t, x)$  defines a sufficiently smooth flow, along which they can then parallel transport the deterministic elements of the tangent space, while projecting it onto the space of gradients, namely to ensure that the new vector field will have a 0 rotational.

Our point of view is that the two approaches coincide (up to the projection that we do want to use here) when the one of [1, 18] is well defined, but that we are able to avoid quite a lot of technical details because we do not use the Eulerian formulation of the path at interest. Of course, the downside of our approach is that we lose uniqueness of parallel transport in certain pathological cases. Furthermore, we believe our different examples justify this loss of uniqueness as natural.

#### 4. LINKS WITH LIONS' LIFT

The definitions and results of the previous section might seem rather obscure in terms of their notation. This section somehow enlightens them by making transparent the translation in terms of random variables, following the powerful lifting approach developed by P.-L. Lions in [19], that we now recall. Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a standard atomless probabilistic space and  $H = L^2(\Omega, \mathbb{R}^d)$  the separable Hilbert space of square integrable  $\mathbb{R}^d$  valued random variables on  $\Omega$ . The point of view developed by P.-L. Lions, mainly to study partial differential equations over  $\mathcal{P}_2(\mathbb{R}^d)$ , is simply to lift problems on  $\mathcal{P}_2(\mathbb{R}^d)$  into problems on  $H$ , an approach that we follow in this section.

Since  $H$  is flat, parallel transport in  $H$  is rather obvious. The tangent plan at  $X \in H$  is always  $H$  itself and the parallel transport of a vector  $Z \in H$  above some initial point  $X \in H$  along a curve  $\theta : [0, 1] \rightarrow H, t \mapsto \theta(t) := X_t$  is the constant function  $Z$ .

The main link between the two approaches is the following proposition, which is a tautology.

- Proposition 4.1.**
- Consider a  $\mathbb{R}^d$  valued process  $(X_t)_{t \in [0, 1]}$ , which is almost surely continuous in  $t$ , such that  $\sup_{t \in [0, 1]} \|X_t\|_H < \infty$  and  $Z \in H$ . Note  $\mathcal{L}(X_0, Z) := \mu(dx)\psi_x(dz)$ . Then,  $\mathcal{L}((X_t, Z)_{t \in [0, 1]})$  is a parallel transport of  $\psi$  along  $\mathcal{L}((X_t)_{t \in [0, 1]})$ .
  - Conversely, consider  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  such that  $(e_0)_\# \eta = \mu$  and a parallel transport  $\Psi$  of  $\psi$  along  $\eta$ . Then, any  $(X_t, Z)_{t \in [0, 1]}$  of law  $\Psi$  is the parallel transport in  $H$  of  $Z$  along  $(X_t)_{t \in [0, 1]}$ .

The above transposition is so simple that it can be tempting to restrict ourselves to working on  $H$  and then pull back all the information we have at the level of  $\mathcal{P}_2(\mathbb{R}^d)$ . This might be misleading if we work with probability measures on other domains than  $\mathbb{R}^d$ , as the underlying Hilbertian structure collapses. The next section extends this notion of parallel transport to cases of probability measures on smooth domain of  $\mathbb{R}^d$

or on manifolds, showing that in Lions' approach, the lift is more important than the Hilbertian structure.

## 5. TANGENT SPACE AND PARALLEL TRANSPORT ON $\mathcal{P}(\mathcal{O})$

We expand rapidly the notion of parallel transport on  $\mathcal{P}(\mathcal{O})$  when  $\mathcal{O}$  is either a smooth bounded domain of  $\mathbb{R}^d$  or a smooth compact Riemannian manifold  $M$ .

**5.1. The case of a smooth domain of  $\mathbb{R}^d$ .** Assume that  $\mathcal{O}$  is the closure of a smooth bounded domain of  $\mathbb{R}^d$ . Because  $\mathcal{O}$  is bounded, we simply work on  $\mathcal{P}(\mathcal{O})$  and there is no need to specify integrability conditions on the measures. In this somehow simpler case, the geometry of  $\mathcal{O}$  is still flat, hence the parallel transport on  $\mathcal{O}$  is trivial.

In this case, we define the tangent space of  $\mu \in \mathcal{P}(\mathcal{O})$  as

$$T_\mu(\mathcal{P}(\mathcal{O})) := \left\{ \psi : \mathcal{O} \rightarrow \mathcal{P}(\mathbb{R}^d) \mid \int_{\bar{\mathcal{O}} \times \mathbb{R}^d} |z| \psi_x(dz) \mu(dx) < \infty \right\}.$$

Note that the integrability condition is of the  $L^1$  type, this is natural from a duality standpoint, since  $\mathcal{O}$  is bounded. This notion of tangent space can be justified by the same heuristics as the ones we gave in Section 2.2. Moreover, in this case, it can be identified with  $\{\gamma \in \mathcal{P}_1(\mathcal{O} \times \mathbb{R}^d) \mid (\pi_1)_\# \gamma = \mu\}$ .

Parallel transport is here defined as follows.

**Definition 5.1.** Let  $\mu \in \mathcal{P}(\mathcal{O})$ ,  $\psi \in T_\mu(\mathcal{P}(\mathcal{O}))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], \mathcal{O}))$  such that  $(e_0)_\# \eta = \mu$ . A parallel transport of  $\psi$  along  $\eta$  is a measure  $\Psi \in \mathcal{P}(\mathcal{C}([0, 1], \mathcal{O} \times \mathbb{R}^d))$  concentrated on  $C_y^1([0, 1], \mathcal{O} \times \mathbb{R}^d)$  such that

- $(p_1)_\# \Psi = \eta$ .
- $(\partial_2)_\# \Psi = \delta_0$ .
- $(e_0)_\# \Psi(dx, dz) = \mu(dx) \psi_x(dz)$ .

All the results and example established in the case of  $\mathcal{P}_2(\mathbb{R}^d)$  follow in a similar manner and their details are omitted to avoid repetitions.

**5.2. The case of smooth Riemannian manifold.** Let  $(M, g)$  be smooth bounded Riemannian manifold. The tangent space to  $M$  at  $x \in M$  is denoted  $T_x M$ , and  $TM$  is the tangent bundle. The notion of parallel transport on  $(M, g)$  is the one associated to the Levi-Civita connection. Recall that given a  $C^1$  curve  $\theta : [0, 1] \rightarrow M$ , and  $\xi \in T_{\theta(0)} M$ , there exists a unique parallel transport of  $\xi$  along  $\theta$ , and it will be noted  $\tau_\theta(\xi)$ , which is, to fix a convention here, a curve in  $TM$ .

For  $\mu \in \mathcal{P}(M)$ , the tangent space to  $\mathcal{P}(M)$  at  $\mu$  is

$$(4) \quad T_\mu(\mathcal{P}(M)) := \left\{ \psi : M \rightarrow \cup_{x \in M} \mathcal{P}(T_x M), \text{ measurable}, \right. \\ \left. \forall x \in M, \psi_x \in \mathcal{P}(T_x M), \int_M \int_{T_x M} |z|_x \psi_x(dz) \mu(dx) < \infty \right\}.$$

The interpretation of the previous is quite clear: we keep the idea of allowing for a probabilistic description of the tangent vectors over every points. The integrability condition is in  $L^1$  because  $M$  is bounded. In this setting, one can identify, for  $\mu \in \mathcal{P}(M)$ ,

$T_\mu(\mathcal{P}(M))$  with  $\{\gamma \in \mathcal{P}_1(TM) | (\pi_1)_\# \gamma = \mu\}$ . Note that the norm  $|z|_x$  in (4) is the one of  $T_x M$  induced by  $g_x$ .

In this setting, parallel transport on  $\mathcal{P}(M)$  is defined as follows.

**Definition 5.2.** Let  $\mu \in \mathcal{P}(M)$ ,  $\psi \in T_\mu(\mathcal{P}(M))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], M))$ , concentrated on  $C^1$  curves and such that  $(e_0)_\# \eta = \mu$ . A parallel transport of  $\psi$  along  $\eta$  is a measure  $\Psi \in \mathcal{P}(\mathcal{C}([0, 1], TM))$  concentrated on  $C^1$  curves such that

- $(p_1)_\# \Psi = \eta$ .
- $\Psi$  is concentrated on parallel transports on  $M$ .
- $(e_0)_\# \Psi(dx, dz) = \mu(dx)\psi_x(dz)$ .

Once again, similar results as in the  $\mathcal{P}_2(\mathbb{R}^d)$  case hold and their proof are similar so we do not give them, except for the following integrability and existence results.

**Proposition 5.3.** Let  $\mu \in \mathcal{P}(M)$ ,  $\psi \in T_\mu(\mathcal{P}(M))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], M))$ , concentrated on  $C^1$  curves, such that  $(e_0)_\# \eta = \mu$  and  $\Psi$  a parallel transport of  $\psi$  along  $\eta$ . Then, for any  $t \in [0, 1]$ ,  $(e_t)_\# \Psi \in \mathcal{P}_1(TM)$ .

*Proof.* Since parallel transport on  $M$  preserves the norm, if we consider a process  $(X_t, Z_t)_{t \in [0, 1]}$  of law  $\Psi$ , we obtain that almost surely  $t \mapsto |Z_t|_{X_t}$  is constant in  $[0, 1]$ . Hence, the result follows by taking the expectation of this quantity.  $\square$

**Proposition 5.4.** Let  $\mu \in \mathcal{P}(M)$ ,  $\psi \in T_\mu(\mathcal{P}(M))$  and  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], M))$ , concentrated on  $C^1$  curves and such that  $(e_0)_\# \eta = \mu$ . There exists a parallel transport of  $\psi$  along  $\eta$ .

*Proof.* Once again, existence relies on the gluing Lemma, which we use to get the existence of a probability measure  $\Gamma(dx, dp, d\theta) \in \mathcal{P}(M \times TM \times \mathcal{C}([0, 1], M))$  such that  $(\pi_3)_\# \Gamma = \eta$ ,  $(\pi_1, \pi_2)_\# \Gamma = \mu(dx)\psi_x(dz)$  and  $(\pi_1, e_0 \circ \pi_3)_\# \Gamma = (Id, Id)_\# \mu$ . Then,  $(\tau_{\pi_3}(\pi_2))_\# \Gamma$  is a required parallel transport.  $\square$

**Remark 5.5.** In order to use the notion of parallel transport on  $M$ , we need to restrict ourselves to  $\eta \in \mathcal{P}(\mathcal{C}([0, 1], M))$  being supported on  $C^1$  curves. Recall that parallel transport on  $M$  cannot be defined along a curve which is simply continuous. Note finally that this regularity shall not guarantee the uniqueness of parallel transport on  $\mathcal{P}(M)$ , see Example 3.8 which can be adapted to this case for instance.

## 6. COMPARING ELEMENTS IN DIFFERENT TANGENT SPACES AND REGULARITY OF ELEMENTS OF $\mathcal{P}(\mathcal{O}) \rightarrow \mathbb{R}$

One of the main advantages of parallel transport is that it allows to compare elements in different tangent spaces, namely by parallel transporting one into the tangent space of the other, usually along a geodesic. We explain here how such an idea can be used to define regularity classes for functions  $U : \mathcal{P} \rightarrow \mathbb{R}$ . In particular, we are interested in  $\mathcal{C}^{1,\alpha}$  spaces, that is in the regularity of the differential, which is naturally an element of the tangent space. For the sake of clarity, we restrict ourselves to the case of  $\mathcal{P}_2(\mathbb{R}^d)$  here, and discuss extensions at the end of the section.

We organise our results as follows. We start by introducing a natural notion of distance on the tangent space  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  for some fixed  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  and show how, with the help of parallel transport, we can construct a way of comparing elements in different tangent spaces. We then proceed by giving another natural way of comparing elements

in different tangent spaces, by equipping  $T\mathcal{P}_2(\mathbb{R}^d) := \cup_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \{\mu\} \times T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  with a distance directly, commenting on the links with the first approach. We then proceed to define regularity spaces for functions  $\mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ .

The first steps of our program are clearly aligned with basic objects in differential geometry such as metric on both the base space and the tangent bundle (or sheaf here).

**6.1. A distance on the tangent space.** Before comparing two elements in different tangent spaces, we need to be able to do so for elements in the same one. Let  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ . Recall that there is an equivalence between choosing  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  and choosing  $\gamma \in \mathcal{P}(\mathbb{R}^{2d})$  such that  $(\pi_1)_\# \gamma = \mu$ , where  $\gamma(dx, dz) = \mu(dx)\psi_x(dz)$ . This could argue in favor of using a simple Wasserstein distance between  $\mu(dx)\psi_x(dz)$  and  $\mu(dx)\psi'_x(dz)$ , for  $\psi, \psi' \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ , however this does not work and we refer to the proof of Proposition 6.4 below for a convincing argument. The distance we shall use requires a bit more rigidity. It was already considered by Gigli and it is given by either a probabilistic reformulation or a coupling one.

For  $\psi, \psi' \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ , define  $d_\mu(\psi, \psi')$  by

$$d_\mu^2(\psi, \psi') := \inf_{\gamma} \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |z - z'|^2 \gamma(dx, dz, dz'),$$

where the infimum is taken over probability measures  $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$  such that  $(\pi_1, \pi_2)_\# \gamma = \mu(dx)\psi_x(dz)$  and  $(\pi_1, \pi_3)_\# \gamma = \mu(dx)\psi'_x(dz)$ . Equivalently, this quantity can be defined by

$$d_\mu^2(\psi, \psi') := \inf_{(X, Z, Z')} \mathbb{E}[|Z - Z'|^2],$$

where the infimum is taken over triplets of random variables  $(X, Z, Z')$  such that  $\mathcal{L}(X, Z) = \mu(dx)\psi_x(dz)$  and  $\mathcal{L}(X, Z') = \mu(dx)\psi'_x(dz)$ .

Note that if  $\psi$  and  $\psi'$  are deterministic, represented by respectively  $\alpha, \alpha' \in L_\mu^2(\mathbb{R}^d, \mathbb{R}^d)$ , then

$$d_\mu^2(\psi, \psi') = \|\alpha - \alpha'\|_{L_\mu^2}^2.$$

This distance has already been introduced in [17], and it can also be defined a sort of metric on  $T_\mu \mathcal{P}_2(\mathbb{R}^d)$  through the relation  $\langle \psi, \psi' \rangle_\mu := d_\mu^2(\psi, \delta_0) + d_\mu^2(\delta_0, \psi') - d_\mu^2(\psi, \psi')$ .

**6.2. Comparing elements in different tangent spaces.** Building on this distance, there is a natural way to compare two elements in different tangent spaces  $T_\mu \mathcal{P}_2(\mathbb{R}^d)$  and  $T_\nu \mathcal{P}_2(\mathbb{R}^d)$ : parallel transport one along a path connecting  $\mu$  and  $\nu$  and use the previous distance. Note that there is a choice to make, as there are several paths connecting  $\mu$  and  $\nu$  and there are no reason that they should give the same result. Furthermore, there is no uniqueness of the parallel transport either, even if the path is fixed, as we saw earlier. Quite naturally, we focus on geodesics here, but we shall see later that looking at more paths can also be natural. In any case, we are interested in the following quantity.

**Definition/Proposition 6.1.** *We note  $\mathcal{E} : T\mathcal{P}_2(\mathbb{R}^d) \times T\mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}_+$  the map defined by either*

$$(5) \quad \mathcal{E}((\mu, \phi), (\nu, \psi)) = \inf_{\Psi} d_\nu^2(\phi', \psi) = \inf_{\Psi'} d_\mu^2(\phi, \psi'),$$

where the first infimum is taken over parallel transports  $\Psi$  of  $\phi$  along geodesics joining  $\mu$  toward  $\nu$ , the second one over parallel transports  $\Psi'$  of  $\psi$  along geodesics joining  $\nu$  toward  $\mu$  and where we use the notations  $(e_1)_\# \Psi(dx, dz) = \nu(dx)\phi'_x(dz)$  and  $(e_1)_\# \Psi'(dx, dz) = \mu(dx)\psi'_x(dz)$ .

*Proof.* Define

$$I_1 := \inf_{\Psi} d_\nu^2(\phi', \psi), \quad I_2 := \inf_{\Psi'} d_\mu^2(\phi, \psi').$$

We need to prove  $I_1 = I_2$ . Let  $\Psi$  be a parallel transport of  $\phi$  along a geodesic  $\theta$  joining  $\mu$  toward  $\nu$ . Furthermore, consider an optimal coupling between  $(e_1)_\# \Psi$  and  $\nu(dx)\psi_x(dz)$ . By gluing the two together, we are allowed to consider random variables  $(X, Z, Y, Z')$  such that  $(X, Y)$  is an optimal coupling for  $W_2(\mu, \nu)$ ,  $\mathcal{L}(X, Z) = \mu(dx)\phi_x(dz)$ ,  $\mathcal{L}(Y, Z') = \nu(dx)\psi_x(dz)$  and  $(Y, Z, Z')$  is optimal for  $d_\nu^2(\phi', \psi)$ . The cost in  $I_1$  associated to  $\Psi$  is thus  $\mathbb{E}[|Z - Z'|^2]$ . Note that  $\mathcal{L}((1-t)Y + tX, Z')_{t \in [0,1]}$  provides a parallel transport of  $(\nu, \psi)$  along a geodesic joining  $\nu$  toward  $\mu$ . Hence, if we call  $\Psi'$  this parallel transport, its associated cost in  $I_2$  is  $d_\mu^2(\phi, \psi') \leq \mathbb{E}[|Z - Z'|^2]$ . Hence, we deduce that  $I_2 \leq I_1$ . By symmetry, we conclude that  $I_1 = I_2$ .  $\square$

**Remark 6.2.** We could also have defined  $\mathcal{E}$  with a supremum over geodesics, and we would have arrived at a somehow similar notion.

**6.3. A distance on  $T\mathcal{P}_2(\mathbb{R}^d)$ .** There is no reason that a distance on  $T\mathcal{P}_2(\mathbb{R}^d)$  can be derived immediately from  $\mathcal{E}$ . This type of phenomena also occurs in differential geometry. We propose here a distance on  $T\mathcal{P}_2(\mathbb{R}^d)$  and present links with the quantity  $\mathcal{E}$ . Let  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$  and  $\phi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ ,  $\psi \in T_\nu(\mathcal{P}_2(\mathbb{R}^d))$ . Recall that by disintegration, considering  $(\mu, \phi)$  is equivalent to considering  $\gamma(dx, dz) = \mu(dx)\phi_x(dz) \in \mathcal{P}_2(\mathbb{R}^{2d})$ . Thus, we identify the tangent sheaf with  $\mathcal{P}_2(\mathbb{R}^{2d})$ . Hence, measuring the distance between  $(\mu, \phi)$  and  $(\nu, \psi)$  can be done by means of the Wasserstein distance on  $\mathcal{P}_2(\mathbb{R}^{2d})$ , which yields the definition of our distance:

$$\mathcal{D}^2((\mu, \phi), (\nu, \psi)) := W_2^2(\mu(dx)\phi_x(dz), \nu(dx)\psi_x(dz)).$$

The standard properties of the Wasserstein distance imply the following.

**Proposition 6.3.** The function  $\mathcal{D}$  defines a metric on  $\mathcal{P}_2(\mathbb{R}^{2d})$ . Furthermore, for any  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$  and  $\psi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ ,  $\phi \in T_\nu(\mathcal{P}_2(\mathbb{R}^d))$ ,  $\mathcal{D}((\mu, \phi), (\nu, \psi)) \geq W_2^2(\mu, \nu)$ .

A bound from above on  $\mathcal{D}^2$  can be obtained as follows.

**Proposition 6.4.** Let  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ , and  $\phi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ ,  $\psi \in T_\nu(\mathcal{P}_2(\mathbb{R}^d))$ . Then

$$(6) \quad \mathcal{D}^2((\mu, \phi), (\nu, \psi)) \leq W_2^2(\mu, \nu) + \mathcal{E}((\mu, \phi), (\nu, \psi)).$$

Furthermore, the inequality can be strict and we do not have in general  $\mathcal{E} \leq \mathcal{D}^2$ .

*Proof.* Consider a parallel transport of  $\phi$  along a geodesic between  $\mu$  and  $\nu$ . This naturally defines a coupling between  $\mu(dx)\phi_x(dz)$  and  $\nu(dx)\psi_x(dz)$ . Hence, the left hand side is smaller than the right hand side in (6).

When  $(\mu, \phi) = \mathcal{L}((U, \lambda U))$  and  $(\mu, \psi) = \mathcal{L}((1-U, \lambda U))$  for  $U$  a uniform random variable on  $[0, 1]$  (hence  $\nu = \mu$ ), we can compute  $\mathcal{E}((\mu, \phi), (\mu, \psi)) = \lambda^2/3$  whereas  $\mathcal{D}^2((\mu, \phi), (\mu, \psi)) \leq 1/3$ .  $\square$

The previous result highlights the fact that there is a choice to be made when comparing different elements in  $T\mathcal{P}_2(\mathbb{R}^d)$ . A point of view purely based on parallel transport should rely on the quantity  $\mathcal{E}$ . On the other hand, taking inspiration from differential geometry,  $\mathcal{D}$  seems to behave like the natural geodesic distance on  $T\mathcal{P}_2(\mathbb{R}^d)$ . Outside of this section, we shall focus on the approach with parallel transport, namely because it is more natural when being interested in the regularity of functions  $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ .

We do not go into the full details of constructing an equivalent of the Sasaki metric on the tangent sheaf  $T\mathcal{P}_2(\mathbb{R}^d)$  (it is a natural metric when we try to look at  $T\mathcal{P}_2(\mathbb{R}^d)$  as a Riemannian manifold), as we do not need it for applications. Nonetheless, we briefly comment on it since  $\mathcal{D}$  looks like what the Sasaki distance could be. Building on the structure of  $T\mathcal{P}_2(\mathbb{R}^d)$  which can be seen as  $\mathcal{P}(T\mathbb{R}^d)$  with  $T\mathbb{R}^d = \mathbb{R}^{2d}$ , we know that  $T(T\mathcal{P}_2(\mathbb{R}^d))$  can be seen as  $\mathcal{P}(\mathbb{R}^{4d})$ . Furthermore, the normal crossing of the vertical part of the tangent space to  $T\mathcal{P}_2(\mathbb{R}^d)$  of the parallel transport can be stated formally quite simply, namely as the tangent part of the parallel transport is constant in this case in flat geometry. The precise decomposition of  $T_\gamma T\mathcal{P}_2(\mathbb{R}^d)$  into an horizontal and a vertical part seems to be the most interesting question here. Clearly, by the construction we proposed for parallel transport, it is immediate to check that  $(\frac{d}{dt})^\# \Psi$  is concentrated on curves  $(x(t), v(t))$  on  $T\mathbb{R}^d$  such that  $v'(t)$  crosses normally the tangent space  $T_{x(t)}\mathbb{R}^d$ .

**6.4. Reminders: Derivatives of function of probability measures.** Let  $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ . The function  $U$  is said to be differentiable at  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  if there exist a *deterministic* element  $\phi \in T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  and a modulus of continuity  $\omega$  such that, for any  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\gamma \in \Pi(\mu, \nu)$

$$\left| U(\nu) - U(\mu) - \int_{\mathbb{R}^{2d}} (y - x) \cdot \phi(x) \gamma(dx, dy) \right| \leq C_2(\gamma) \omega(C_2(\gamma)),$$

recall that  $C_2(\gamma) := (\int_{\mathbb{R}^{2d}} |x - y|^2 \gamma(dx, dy))^{\frac{1}{2}}$ . In such a case we note  $D_\mu U(\mu) = \phi \in L^2((\mathbb{R}^d, \mu), \mathbb{R}^d)$ .

**Remark 6.5.** *There are different notions of differentiability on  $\mathcal{P}_2(\mathbb{R}^d)$  that can be found in the literature, see chapter 5 in [11] for instance. Quite often, in the previous definition, the requirement is only made for coupling  $\gamma$  which are optimal for  $W_2(\mu, \nu)$ . In fact, there is an equivalence between the two notions, see Theorem 3.21 in [16], and we do not know any short proof of this fact. Here, we prefer to stick to the formulation with all couplings, as it is a nice non-Hilbertian formulation of the derivative introduced by P.-L. Lions on the lift of the function. In particular, one can check that we do not use in the following the Hilbertian structure of  $L^2(\Omega, \mathbb{R}^d)$ .*

**Remark 6.6.** *The previous definition of differentiability contains the fact that  $D_\mu U(\mu)$  is a deterministic element of the tangent space. This fact is not obvious. It has been proven by Lions in [19] by introducing its lift on the space of random variables.*

**6.5.  $\mathcal{C}^{1,\alpha}$  regularity on  $\mathcal{P}_2(\mathbb{R}^d)$ .** Let  $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  be a differentiable map at any  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ . The aim of this section is to define a notion of continuity, Hölder or Lipschitz regularity for the map  $\mu \rightarrow D_\mu U(\mu)$ . Of course, since for every  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $D_\mu U(\mu) \in$

$L^2((\mathbb{R}^d, \mu), \mathbb{R}^d)$ , this map is valued in  $\cup_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} L^2((\mathbb{R}^d, \nu), \mathbb{R}^d) \subset T\mathcal{P}_2(\mathbb{R}^d)$  and a suitable notion of parallel transport is needed to compare, for  $\mu \neq \nu$   $D_\mu U(\mu) \in L^2((\mathbb{R}^d, \mu), \mathbb{R}^d)$  and  $D_\mu U(\nu) \in L^2((\mathbb{R}^d, \nu), \mathbb{R}^d)$ . We adopt the following definition, in which we make constant use of the notation  $\gamma \in \Pi(\mu, \nu)$ , of Proposition 3.6 and of the notation  $\mathcal{T}^\gamma(\psi)$  to denote the parallel transport of  $\psi$  along the coupling  $\gamma$ .

**Definition 6.7.** Let  $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  be a differentiable map at any  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ . We say that it is  $\mathcal{C}^1$  if, for any  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$\lim_{C_2(\gamma) \rightarrow 0} d_\nu(\mathcal{T}^\gamma(D_\mu U(\mu)), D_\mu U(\nu)) = 0.$$

For  $\alpha \in (0, 1)$ , we say that this function is in  $\mathcal{C}^{1,\alpha}(\mathcal{P}_2(\mathbb{R}^d))$  if

$$(7) \quad I_\alpha(D_\mu U) := \sup_{\gamma, 0 < C_2(\gamma) \leq 1} \frac{d_\nu(\mathcal{T}^\gamma(D_\mu U(\mu)), D_\mu U(\nu))}{C_2(\gamma)^\alpha} < \infty.$$

The space  $\mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$  is defined in a similar way except for the fact that the supremum is taken over all  $\gamma$  such that  $C_2(\gamma) > 0$ .

**Remark 6.8.** Remark that, for  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\phi, \psi$  being deterministic elements of respectively  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ ,  $T_\nu(\mathcal{P}_2(\mathbb{R}^d))$ , and  $\gamma \in \Pi(\mu, \nu)$ ,

$$d_\nu^2(\mathcal{T}^\gamma(D_\mu U(\mu)), D_\mu U(\nu)) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |\phi(x) - \psi(y)|^2 \gamma(dx, dy).$$

This definition is perfectly equivalent to standard  $\mathcal{C}^{1,\alpha}$  regularity of the lift  $\tilde{U}$  of  $U$  in  $L^2(\Omega, \mathbb{R}^d)$  introduced in Section 4, except that this language allows to define such a solution even when the state space is not  $\mathbb{R}^d$ . We believe it is a strong justification of the regularity of the regularization of  $W_2^2$  introduced in [9].

**Remark 6.9.** Below we shall comment on the choice of the quantity  $C_2(\gamma)^\alpha$  in (7), namely that we could also have used  $C_{2\alpha}(\gamma)^\alpha$ .

The following holds.

**Proposition 6.10.** For  $\alpha \in (0, 1]$ ,  $\mathcal{C}^{1,\alpha}(\mathcal{P}_2(\mathbb{R}^d)) \cap L^\infty(\mathcal{P}_2(\mathbb{R}^d))$  is a normed vector space when equipped with the norm  $\mathcal{N}$  defined by

$$\mathcal{N}(U) := \|U\|_\infty + \sup_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \{d_\mu(D_\mu U(\mu), 0)\} + I_\alpha(D_\mu U).$$

**Remark 6.11.** • The same result holds for  $\mathcal{C}^{1,0}(\mathcal{P}_2(\mathbb{R}^d))$  with the convention that  $I_0(D_\mu U) = 0$  for any  $U$ .

• The boundedness of  $U$  and  $D_\mu U$  can be relaxed in the usual way: by considering weighted norms which allow for a fixed growth in  $W_2(\mu, \delta_0)$  (polynomial e.g.).

*Proof.* The probabilistic reformulation of  $d_\mu$  implies that  $I_\alpha$  indeed satisfies  $I_\alpha(\lambda D_\mu) = |\lambda| I_\alpha(D_\mu)$ . Hence, to verify that  $\mathcal{N}$  is a norm, we only need to check the triangular inequality. This does not raise any issue, for instance it follows easily from the probabilistic representation of  $d_\mu^2$ . The fact that  $\mathcal{N}$  is a norm then implies the vector space structure.

Hence, it only remains to verify the completeness. Let  $(U_n)_{n \geq 0}$  be a Cauchy sequence for  $\mathcal{N}$ . Since it is one for the  $\|\cdot\|_\infty$  norm, we know that it converges uniformly toward some function  $U$ , which is continuous. To prove the convergence of the derivative, consider the map  $\phi_n : \mu \mapsto \mu(dx)\delta_{D_\mu U_n(\mu)(x)}(dz)$ , valued in  $\mathcal{P}_2(\mathbb{R}^{2d})$ . Remark that for any  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $(\phi_n(\mu))_{n \geq 0}$  is a Cauchy sequence. Indeed, since we can write  $d_\mu(D_\mu U_n(\mu) - D_\mu U_m(\mu), 0) = W_2(\phi_n, \phi_m)$ . Hence,  $(\phi_n(\mu))_{n \geq 0}$  converges to some  $g(\mu) \in \mathcal{P}_2(\mathbb{R}^{2d})$  with first marginal  $\mu$ . Furthermore, this convergence is uniform in  $\mu$ . We now need to check that  $g = D_\mu U$ . Since, for all  $n \geq 0$ ,  $U_n$  is  $\mathcal{C}^1$ , we obtain that for any  $\gamma \in \Pi(\mu, \nu)$  for  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$U_n(\nu) - U_n(\mu) = \int_0^1 \int_{\mathbb{R}^{2d}} D_\mu U_n(((1-t)\pi_1 + t\pi_2)_\# \gamma)(x)(y-x)\gamma(dx, dy)dt.$$

Passing to the limit in this expression, and denoting  $\gamma(dx, dy) = \mu(dx)k_x(dy)$ , we deduce that

$$U(\nu) - U(\mu) = \int_0^1 \int_{\mathbb{R}^{3d}} z \cdot (y-x)g(((1-t)\pi_1 + t\pi_2)_\# \gamma)(dx, dz)k_x(dy)dt.$$

From this, we deduce that, if we denote  $g(\mu)(dx, dz) = \mu(dx)\psi_{\mu,x}(dz)$ , then  $x \mapsto \int_{\mathbb{R}^d} z\psi_{\mu,x}(dz) = D_\mu U(\mu)$ . In fact, we can show that  $g(\mu)$  is in fact of the form  $(Id, \phi)_\# \mu$  for some map  $\phi$ , which is then nothing else than  $D_\mu U(\mu)$ . In this setting, this is a consequence of Theorem 3.17 in [16] (it is even a consequence of a weaker results of Lions in [19] under our assumptions). Hence, we obtained that  $U$  is differentiable everywhere. Passing to the limit in the quotient which defined  $I_\alpha(U_n)$ , we easily obtained that  $I_\alpha(U) < \infty$ , hence, the result is proved.  $\square$

We also recall Proposition 5.36, vol I in [11], stated with other notation, but which here takes the form.

**Proposition 6.12.** *Let  $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  be such that  $I_1(D_\mu U) < \infty$ . Then, for any  $\mu$ , there exists a  $I_1(D_\mu U)$  Lipschitz function  $\phi_\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $D_\mu U = \phi_\mu$   $\mu$  almost everywhere.*

We do not reproduce its rather heavy proof and we leave the reader to check that its proof does not use the Hilbert structure of  $L^2(\Omega, \mathbb{R}^d)$ , and that it can be thus extended instantly to the case of a function  $U : \mathcal{P}(\mathcal{O}) \rightarrow \mathbb{R}$  for  $\mathcal{O}$  a smooth domain.

**6.6. Discussion on the  $\mathcal{C}^{1,\alpha}$  norm and its consequences.** There are mainly two points that we want to insist upon in the definition of  $I_\alpha$  in (7).

The first one is the fact that the supremum is taken over all couplings and not only over optimal couplings. As we mentioned above, the same choice/problem arises when defining the notion of derivative, and it has been shown in [16] that the two notions are equivalent. Of course, for  $I_\alpha$ , one implies the other, but it is not clear for us to show an equivalence between the two notions and we leave this question for future research. In particular, the proof of [16] uses in a crucial manner the fact that the term we want to estimate can be seen as a part of a Taylor expansion, which is not true here.

The second point we want to insist upon is that, when  $\alpha \in (0, 1)$  the choice to divide by  $C_2(\gamma)^\alpha$  might not be the only natural one. Indeed, one could have also defined

$$J_\alpha(D_\mu U) := \sup_{\gamma, 0 < C_2(\gamma) < 1} \frac{d_\nu(\mathcal{T}^\gamma(D_\mu U(\mu)), D_\mu U(\nu))}{C_{2\alpha}(\gamma)^\alpha}.$$

Since, by Jensen's inequality  $C_{2\alpha}(\gamma)^\alpha \leq C_2(\gamma)^\alpha$ ,  $J_\alpha(D_\mu U) \geq I_\alpha(D_\mu U)$ . Furthermore, the proof of Proposition 5.36, vol I in [11] can be carried out in a similar manner to obtain the following.

**Proposition 6.13.** *Let  $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  be such that  $J_\alpha(D_\mu U) < \infty$ . Then, for any  $\mu$ , there exists a  $\mathcal{C}^\alpha(\mathbb{R}^d, \mathbb{R}^d)$  continuous function  $\phi_\mu$ , such that  $\|\phi_\mu\|_{\mathcal{C}^\alpha} \leq J_\alpha(D_\mu U)$  and  $D_\mu U = \phi_\mu$   $\mu$  almost everywhere.*

*Proof.* We do not reproduce this quite long and tedious proof. We mention that following the exact same argument as in [11], we arrive at the required result by, instead of doing the Taylor expansion around the element of interest  $z_0 \in \mathbb{R}^d$  (with the notation of [11]), we instead bound the quotient which appears in the expression of the  $\mathcal{C}^\alpha$  norm of  $D_\mu U$ .  $\square$

Furthermore, we were not able to construct a function  $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  such that  $J_\alpha(D_\mu U) = \infty$  while  $I_\alpha(D_\mu U) < \infty$ , also not knowing if such a function exists. We also leave these questions for future research.

**6.7. Extensions to other cases than  $\mathcal{P}_2(\mathbb{R}^d)$ .** The results and definitions above translate immediately if, instead of working on  $\mathcal{P}_2(\mathbb{R}^d)$ , we are interested in either  $\mathcal{P}_p(\mathbb{R}^d)$  or  $\mathcal{P}(\mathcal{O})$  for a smooth bounded domain  $\mathcal{O} \subset \mathbb{R}^d$ . The case of probability measures over a smooth manifold requires a bit more care, as we constantly made use of Proposition 3.6 in the previous, which of course has no reason to be true in this setting. Thus, when defining the equivalent of  $I_\alpha$  in this setting, one has to be careful with the fact that the supremum has to be taken over all Lagrangian path  $\eta$  as follows.

$$\sup_{\eta \in LP(\mu, \nu)} \frac{d_\nu(\mathcal{T}^\eta(D_\mu U(\mu)), D_\mu U(\nu))}{C_2(\eta)^\alpha},$$

where

$LP(\mu, \nu) := \{\eta \in \mathcal{P}(\mathcal{C}([0, 1], M)), (e_0)_\# \eta = \mu, (e_1)_\# \eta = \nu, \eta \text{ is concentrated on } C^1([0, 1], M)\}$ , and, for  $\eta \in LP(\mu, \nu)$ ,

$$C_2(\eta) := \left( \int_{\mathcal{C}([0, 1], M)} \int_0^1 |y'(t)|_{y(t)}^2 dt \eta(dy) \right)^{\frac{1}{2}}.$$

Except for this change, the previous sections can be extended quite straightforwardly.

## 7. TRANSLATED CURVES AND APPLICATION IN OPTIMAL CONTROL PROBLEMS

In this section, we explain and prove how ideas introduced above yields quite simply regularity of the value function of some optimal control problems in the space of probability measures. Once again, we focus here on the case of  $\mathcal{P}_2(\mathbb{R}^d)$ . We comment on several extensions at the end of the section. The core idea of this section is to translate

curves in  $\mathcal{P}_2(\mathbb{R}^d)$ . We do not claim we shall be able to translate any curves so we start by explaining in details the curves with which we are going to be interested in. Then, we explain how to translate them, and finally how to use this translations in optimal control problems.

Given a curve  $(m_t)_{t \in [0,1]}$  in  $\mathcal{P}_2(\mathbb{R}^d)$  we would like to translate it to another curve  $(\tilde{m}(t))_{t \in [0,1]}$ . By translation, we mean that if the two curves are sufficiently smooth to have time derivative, then, in some sense,  $\tilde{m}'(t) \in T_{\tilde{m}(t)}$  should be a parallel transport of  $m'(t) \in T_{m(t)}(\mathcal{P}_2(\mathbb{R}^d))$  along a given path between  $m(t)$  and  $\tilde{m}(t)$ . More precisely, since we are motivated by control problems, our fundamental objective is in fact the following: given a curve  $(m(t))_{t \in [0,1]}$  generated by a control  $(\alpha_t)_{t \in [0,1]}$  (typically  $\alpha_t \in T_{m(t)}(\mathcal{P}_2(\mathbb{R}^d))$ ), can we generate another curve  $(\tilde{m}(t))_{t \in [0,1]}$  based on another control  $(\beta_t)_{t \in [0,1]}$  which is given so that for all  $t$ ,  $\beta_t$  is a parallel transport of  $\alpha_t$  along a certain path between  $m(t)$  and  $\tilde{m}(t)$ ?

**7.1. Curves that we want to translate.** We focus first on the construction of a curve given a control. When restricting our attention to controls which are deterministic elements of the tangent space, the question is well understood and we recall some facts about it here.

Consider a curve  $(m_t)_{t \in [0,1]}$  which is a weak solution of

$$(8) \quad \partial_t m + \operatorname{div}_x(\alpha m) = 0 \text{ in } (0, 1) \times \mathbb{R}^d,$$

for some measurable vector field  $\alpha : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ . By weak solution, we mean that we can test the previous relation against  $\mathcal{C}^{1,c}([0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d)$  test functions and obtain the usual relations. Assume that

$$(9) \quad \int_0^1 \|\alpha(s, \cdot)\|_{L^2(m_s)}^2 ds < \infty.$$

Condition (9) is far from sufficient to guarantee uniqueness of solutions  $m$  of (8), nonetheless, it is sufficient to obtain absolute continuity of the curve, and thus a precise sense in which we can understand  $m'(t) \in T_{m(t)}(\mathcal{P}_2(\mathbb{R}^d))$ . It is somehow summarized in Ambrosio's superposition principle that we recall here.

**Theorem 7.1** (Thm 8.2.1 in [2]). *If  $(m_t)_{t \in [0,1]}$  is a curve in  $\mathcal{P}_2(\mathbb{R}^d)$  which is a weak solution of (8) and if (9) is satisfied, then there exists a process  $(X_t)_{t \in [0,1]}$  such that  $m_t = \mathcal{L}(X_t)$  for all  $t$  and almost surely*

$$(10) \quad \forall t \in [0, 1], X_t = X_0 + \int_0^t \alpha(s, X_s) ds.$$

*Conversely, if  $(X_t)_{t \in [0,1]}$  is a solution of (10) and  $\int_0^1 \mathbb{E}[|\alpha(s, X_s)|^2] ds < \infty$ , then  $(\mathcal{L}(X_t))_{t \in [0,1]}$  is a weak solution of (8).*

Since there is no uniqueness of solutions  $m$  of (8) under the sole condition (9), this approach does not allow to talk about the evolution  $m$  of associated to a control  $\alpha$  but simply about an evolution  $m$  associated to the control  $\alpha$ .

Since we saw that the parallel transport of a deterministic element of the tangent space may fail to be deterministic, we want to consider an analogous of the previous approach for general elements of the tangent space.

This objective puts us into the territory of a quite usual problem in stochastic control: how to generate sufficient noise ? In other words, if we consider general controls  $\psi : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{P}(\mathbb{R}^d)$ , can we consider a natural associated evolution  $(m_t(dx)\psi_{t,x}(dz))_{t \in [0,1]}$  such that

$$\int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} |z|^2 \psi_{s,x}(dz) m_s(dx) ds < \infty \quad ?$$

The answer to this question is no in general. Consider for instance an hypothetical case in which, for all  $t \geq 0$ ,  $m_t(dx)\psi_{t,x}(dz) = m_t(dx)\rho(dz)$  for  $\rho$  the normal distribution. This would correspond to a situation in which, at any time, we are able to randomize the speed of the particles, independently of their current position, hence of what has been done before. The natural such candidate would be given by a process  $(X_t, Z_t)_{t \in [0,1]}$  where  $\mathcal{L}((X_t, Z_t)) = m_t(dx)\rho(dz)$  and  $X_t = X_0 + \int_0^t Z_s ds$ . We do not know how to construct such a process. On the other hand, we can characterize situations which we can represent as the next result shows.

**Proposition 7.2.** *Let  $(X_t, Z_t)_{t \in [0,1]}$  be a  $\mathbb{R}^d \times \mathbb{R}^d$  process such that,*

$$\int_0^1 \mathbb{E}[|Z_t|^2] dt < \infty$$

*and, almost surely, for all  $t \in [0, 1]$*

$$(11) \quad X_t = X_0 + \int_0^t Z_s ds.$$

*Then, there exists a probabilistic space  $K \subset \mathbb{R}^N$ , a measurable map  $b : [0, T] \times \mathbb{R}^d \times K \rightarrow \mathbb{R}^d$  and a continuous path  $(\mu_t)_{t \in [0,1]}$ , valued in  $\mathcal{P}(\mathbb{R}^d \times K)$  such that*

$$(12) \quad ((\pi_1, b(t, \cdot, \pi_1, \pi_2)_\# \mu_t)_{t \in [0,1]} = (\mathcal{L}(X_t, Z_t))_{t \in [0,1]}$$

*and  $(\mu, b)$  is a weak solution of the continuity equation*

$$(13) \quad \partial_t \mu + \operatorname{div}_x(b\mu) = 0 \text{ in } (0, 1) \times \mathbb{R}^d \times K,$$

*with initial condition  $\mu_0$  such that  $(\pi_1)_\# \mu_0 = \mathcal{L}(X_0)$ . Conversely, given a probabilistic set  $(K \subset \mathbb{R}^N, \mathcal{B}, \mathbb{P}')$  and a solution  $(\mu, b)$  of (13) satisfying*

$$\int_0^1 \int_{\mathbb{R}^d \times K} |b(t, x, k)|^2 \mu_t(dx, dk) dt < \infty,$$

*there exists a process  $(X_t, Z_t)_{t \in [0,1]}$  such that (12) and (11) hold.*

*Proof.* The proof is rather simple. On the first side, we note  $(\Omega, \mathcal{A}, \mathbb{P})$  the atomless probabilistic space onto which the process  $(X_t, Z_t)_{t \in [0,1]}$  is defined. Without loss of generality, we can assume that  $(\Omega, \mathcal{A}, \mathbb{P}) = ([0, 1]^N, \mathcal{B}, \operatorname{Leb})$ . Taking  $K = \Omega$ ,  $b(t, x, k) = Z_t(k)$  for  $(t, x, k) \in [0, 1] \times \mathcal{O} \times K$  and  $\mu_t = \mathcal{L}(X_t, Z_t)$  yields the solution.

On the other side, we use Ambrosio's superposition principle (in  $\mathbb{R}^d \times \mathbb{R}^N$ ) to obtain a process  $(X_t, k_t)_{t \in [0,1]}$  from  $\mu$  on a probabilistic space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Observe that almost surely  $k$  is constant. We then define  $Z_t(\omega) = b(t, X_t(\omega), k_t(\omega))$  and the result follows.  $\square$

**Remark 7.3.** Note that the set  $K$  is somehow free in the previous statement, as it can be changed into any non-atomic probabilistic space.

**Remark 7.4.** Clearly, the curves satisfying the assumptions of Theorem 7.1 also satisfy the assumptions of Proposition 7.2. Furthermore, consider  $(b, \mu, K)$  as in Proposition 7.2. Applying the more general version of Theorem 7.1 to the process given by Proposition 7.2, we obtain that there exists a measurable  $\alpha : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $m_t := (\pi_1)_\# \mu_t$  solves (8). In fact,  $\alpha(t, \cdot)$  is given  $m_t$  almost everywhere by  $\alpha(t, x) := \mathbb{E}[Z_t | X_t = x]$ .

The interpretation of this reformulation is that the set  $K$  is the set used to randomize the speed we associated to particles located at  $x \in \mathbb{R}^d$ . Ambrosio's superposition principle basically states that we can forget this space if we are only interested in the evolution of the marginal in  $x \in \mathbb{R}^d$  of (13). Since it is not the case here, this generalization shall have some importance.

The main role of the previous result is to avoid pathological situation such as the one mentioned before the Proposition.

**7.2. Translating the curves.** Given a  $\mathcal{C}^1$  curve  $(x_t)_{t \in [0,1]}$  in  $\mathbb{R}^d$ , and a point  $y_0 \in \mathbb{R}^d$ , we can easily understand the translation of  $(x_t)_{t \in [0,1]}$  to  $y_0$  as the curve  $(y_t)_{t \in [0,1]}$  defined by  $y_t := y_0 + \int_0^t \dot{x}_s ds$ . We now define a similar notion for curves on  $\mathcal{P}_2(\mathbb{R}^d)$  satisfying the assumptions of Proposition 7.2. Ideally, we would like to say that we simply need, at each time, to parallel transport the "speed" of the curve toward another point. However, this strategy runs into the following difficulty: a proper parallel transport is not simply defined between two measures, but along a path (or a minima a coupling in this case). Thus, we need to treat with care the way we are going to connect our two curves as time passes. We shall do so by not only seeking for the translated curve  $(\tilde{m}_t)_{t \in [0,1]}$  of  $(m_t)_{t \in [0,1]}$ , but rather by seeking for curves which are couplings between them.

We now propose the following notion of translated curve.

**Definition 7.5.** For any curve  $(m_t)_{t \in [0,1]}$  which can be represented through  $b, K, \mu$  as in Proposition 7.2, any  $\tilde{m}_0 \in \mathcal{P}_2(\mathbb{R}^d)$  and coupling  $\gamma_0 \in \Pi(m_0, \tilde{m}_0)$ , a coupling curve  $(\eta_t)_{t \in [0,1]}$  is a curve in  $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times K)$ , weak solution of

$$\partial_t \eta + \text{div}_x(b\eta) + \text{div}_y(b\eta) = 0 \text{ in } (0, 1) \times \mathbb{R}^d \times \mathbb{R}^d \times K,$$

with initial condition  $\eta_0(dx, dy, dk) = \gamma_0(dx, dy)\psi_x(dk)$  where  $m_0(dx)\psi_x(dk) := \mu_0(dx, dk)$ . The curve  $((\pi_2)_\# \eta_t)_{t \in [0,1]}$  is called a translation of  $(m_t)_{t \in [0,1]}$  along  $\gamma_0$ . Furthermore, the curve  $((\pi_1, \pi_2)_\# \eta_t)_{t \in [0,1]}$  is called the associated translated pair.

**Remark 7.6.** Of course, if  $(m_t)_{t \in [0,1]}$  is represented (as it is usually the case), without a probabilistic space  $K$ , then it can be omitted in the previous. Note however that in such a case, it is still likely that the base space  $\mathbb{R}^d$  shall play the role of  $K$  for the translation  $((\pi_2)_\# \eta_t)_{t \in [0,1]}$ .

Coupling curves exist in general as the next result shows.

**Proposition 7.7.** Let  $(m_t)_{t \in [0,1]}$  be a curve which can be represented through  $b, K, \mu$  as in Proposition 7.2. For any  $\gamma_0 \in \mathcal{P}(\mathbb{R}^{2d})$  with  $(\pi_1)_\# \gamma = m_0$ , there exists an associated coupling curve, translating  $(m_t)_{t \in [0,1]}$  along  $\gamma_0$ . If  $(\gamma_t)_{t \in [0,1]}$  is the translated pair, then  $C_2(\gamma_t)$  does not depend on time. More precisely, if  $(X_t, Y_t)_{t \in [0,1]}$  is a process of law  $(\gamma_t)_{t \in [0,1]}$ . Then, almost surely,  $X_t - Y_t$  is constant in time.

*Proof.* By the gluing Lemma, we can consider both a process  $(X_t(\omega), k(\omega))_{t \in [0,1]}$  of marginal law  $(\mu_t)_{t \in [0,1]}$  on a probabilistic space  $(\Omega, \mathcal{A}, \mathbb{P})$  such that, almost surely,  $X_t = \int_0^t b(s, X_s, k) ds + X_0$ , and a random variable  $Y_0$  such that  $(X_0, Y_0) \sim \gamma_0$ . Define the process  $Y_t = Y_0 + \int_0^t b(s, X_s, k) ds$ . Setting  $\eta_t := \mathcal{L}(X_t, Y_t, k)$  completes the proof. The rest follows easily.  $\square$

Note that there is no uniqueness of such a procedure as the next example shows. This is mainly due to similar reasons as the non-uniqueness of parallel transport on  $\mathcal{P}_2(\mathbb{R}^d)$ .

**Example 7.8.** For all  $t \in [0, 1]$ , define  $m_t = \frac{1}{2}(\delta_t + \delta_{-t})$ . Consider  $\gamma_0(dx, dy) = \frac{1}{2}\delta_0(dx)(\delta_{-2}(dy) + \delta_2(dy))$ . Remark that  $(m_t)_{t \in [0,1]}$  is a solution of (8) with  $\alpha(t, x) = 2\mathbb{1}_{x \geq 0} - 1$ . The three following trajectories are all translations of  $(m_t)_{t \in [0,1]}$  along  $\gamma_0$

$$\eta_t^1 = \frac{1}{2}(\delta_{-2-t} + \delta_{2+t}); \quad \eta_t^2 = \frac{1}{2}(\delta_{-2+t} + \delta_{2-t}); \quad \eta_t^3 = \frac{1}{4}(\delta_{-2-t} + \delta_{2+t} + \delta_{-2+t} + \delta_{2-t}).$$

As in the gluing Lemma, in certain "deterministic" situations, we are able to obtain uniqueness of the translation in the sense that all possible translations yields the same translated curve.

**Proposition 7.9.** Assume that  $(m_t)_{t \in [0,1]}$  is a weak solution of (8) for some  $\alpha \in L^1((0, 1), W^{1,\infty}(\mathbb{R}^d))$  and consider  $\tilde{m}_0 \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\gamma_0 \in \Pi(m_0, \tilde{m}_0)$ . Then, there is a unique translating coupling of  $(m_t)_{t \in [0,1]}$  along  $\gamma_0$ .

*Proof.* The result is trivial, since, by standard Cauchy Lipschitz theory, the flow of the ODE is uniquely defined, there is thus no room for any non-uniqueness.  $\square$

**Remark 7.10.** Note that we have here only an assumption on the regularity of the vector field, compared to the "or" type of assumption of Proposition 3.9. This is due to the time dependent nature of the present problem. We leave as a simple exercise to the interested reader to establish that deterministic initial coupling is not sufficient, if the velocity field is not smooth, for instance by considering the case of the curve  $(m_t)_{t \in [0,1]}$  given by  $m_t = \frac{1}{2}(\delta_{f(t)} + \delta_{-f(t)})$  where  $f(t) = \max(-\frac{1}{2} - t; 0; -\frac{1}{2} + t)$ .

This translation enjoys several other properties.

**Proposition 7.11.** The following hold.

- (i) Let  $(x_t)_{t \in [0,1]}$  be a  $\mathcal{C}^1$  curve on  $\mathbb{R}^d$ ,  $\gamma(dx, dy) = \delta_{x_0}(dx)\delta_{y_0}(dy)$  for some  $y_0 \in \mathbb{R}^d$ . Then the translation of  $(\delta_{x_t})_{t \in [0,1]}$  along  $\gamma$  is given by  $(\delta_{y_t})_{t \in [0,1]}$ .
- (ii) Let  $(m_t)_{t \in [0,1]}$  be a curve as in Proposition 7.7 and  $\gamma$  some optimal coupling for  $W_2$  whose first marginal is  $m_0$ . Let  $\tilde{m}_t$  be a translation of  $(m_t)_{t \in [0,1]}$  along  $\gamma$ . Then  $t \mapsto W_2(m_t, \tilde{m}_t)$  is bounded by  $C_2(\gamma)$ .

*Proof.* For the first point, it suffices to remark that there is a unique possible translated curve. The second point follows directly from point (v) in Proposition 7.7.  $\square$

Note that there are cases in which the  $W_2$  distance between a curve and its translation actually decreases before increasing again, as the previous example showed. A less pathological example is the following.

**Example 7.12.** Let  $m_t := \frac{1}{2}(\delta_{-4+3t} + \delta_{4-3t})$ . It is clearly a solution of (8) for some smooth  $\alpha$ , constant in time, such that  $\alpha(x) = -3$  if  $x \geq 1$  and  $\alpha(x) = 3$  if  $x \leq -1$ . Let  $\gamma(dx, dy) = \frac{1}{2}(\delta_{-4}(dx)\delta_{-1}(dy) + \delta_4(dx)\delta_1(dy))$ . It is clearly an optimal coupling for  $W_2$  and one can check that  $\tilde{m}_t := \frac{1}{2}(\delta_{-1+3t} + \delta_{1-3t})$  is the unique translation of  $(m_t)_{t \in [0,1]}$  along  $\gamma$ . Then

$$W_2^2(m_t, \tilde{m}_t) = \begin{cases} 9 & \text{if } t \leq \frac{1}{3}, \\ (5 - 6t)^2 & \text{else.} \end{cases}$$

This last example is yet another illustration (if it was still needed at this point) that the geometry of  $(\mathcal{P}_2(\mathbb{R}^d), W_2)$  is not flat at all...

**7.3. Applications in optimal control.** Consider  $T > 0$ ,  $L : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}_+$ ,  $G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}_+$  and the optimal control problem defined for all  $t \in [0, T]$ ,  $\bar{m} \in \mathcal{P}_2(\mathbb{R}^d)$

$$V(t, \bar{m}) := \inf_{(\alpha, m)} \left\{ \int_t^T \int_{\mathbb{R}^d} L(s, x, \alpha(s, x), m_s) m_s(dx) ds + G(m_T) \right\},$$

where the infimum is taken over all pairs  $(\alpha, m)$  such that  $m : [t, T] \mapsto \mathcal{P}_2(\mathbb{R}^d)$  is continuous,  $\alpha : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{P}(\mathbb{R}^d)$  is measurable and they solve (8) with  $m_t = \bar{m}$ . The function  $V$  is called the value function of the problem. Typically, some growth condition is assumed on  $L$  in its third argument, so that  $\alpha$  has to exhibit certain bounds. Quite often, those bounds are that for almost every  $s \in [t, T]$ ,  $\alpha \in L^2((\mathbb{R}^d, m_s), \mathbb{R}^d)$ .

The interpretation of such a problem is that one can control the measure  $m$  through the continuity equation by choosing the speed  $\alpha$ . This interpretation has to be manipulated with care since, because no particular regularity condition on  $\alpha$  has been imposed, the continuity is not necessarily uniquely solvable, thus the presence of  $m$  in the infimum as well (since, given an  $\alpha$ , multiple  $m$  can be solutions). We refer to [6] for more details on this kind of problem.

In view of [5] and the notion of parallel transport above, it may seem arbitrary to restrict ourselves to  $\alpha$  which are deterministic elements of the tangent space. Hence it is also natural to consider the following problem

$$(14) \quad U(t, m) := \inf_{(b, \mu, K)} \left\{ \int_t^T \int_{\mathbb{R}^d \times K} L(s, x, b(s, x, k), (\pi_1)_\# \mu_s) \mu_s(dx, dk) ds + G(m_T) \right\},$$

where the infimum is taken over all  $(b, \mu, K)$  such that (13) is satisfied and such that  $(\pi_1)_\# \mu_0 = m$ .

We always have  $U \leq V$ , and when  $L$  is convex in its third argument, we also have that  $V = U$  because of Jensen's inequality and Remark 7.4. We focus in the rest of this section on  $U$  as we believe it is a more natural object, for essentially all the arguments we gave when choosing our tangent space above. Our main result on  $U$  is the following.

**Theorem 7.13.** Assume that there exists  $C > 0$  such that  $L$  and  $G$  satisfy

$$\begin{aligned} \|G\|_\infty &\leq C, \quad \forall \mu, \nu, \quad |G(\mu) - G(\nu)| \leq CW_2(\mu, \nu), \\ \forall s, x, z, m, \quad C^{-1}|z|^2 - C &\leq L(s, x, z, m), \\ \forall s, x, y, z, m, m', \quad |L(s, x, z, m) - L(s, x, z, m')| &\leq C(1 + |z|)(|x - y| + W_2(m, m')) \end{aligned}$$

Then, there exists  $C_0$  depending only on  $C$  and  $T$  such that for all  $t \in [0, T], m, m' \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$|U(t, m) - U(t, m')| \leq C_0 W_2(m, m').$$

*Proof.* Let  $t \in [0, T], m, m' \in \mathcal{P}_2(\mathbb{R}^d)$  and  $\gamma$  optimal for  $W_2(m, m')$ . The proof consists in taking an almost optimal curve for  $U(t, m)$  and translating it along  $\gamma$ . Using this translation for  $U(t, m')$  shall then yield the result.

Note first that  $U(t, m)$  is indeed well defined since we can simply choose the pair  $(b, \mu, K)$  such that  $K$  is trivial,  $b \equiv 0$  and  $\mu_s(dx) = m(dx)$ . Let  $\epsilon > 0$  and an  $\epsilon$  optimal control  $(b, \mu, K)$  for  $U(t, \mu)$ . Consider a translating coupling  $\eta$  of  $(b, \mu, K)$  along  $\gamma$ . Note that  $m_s = (\pi_1)_\# \eta_s$  and define  $m'_s = (\pi_2)_\# \eta_s$ . By  $\epsilon$  optimality in  $U(t, m)$  and admissibility in  $U(t, m')$ , we obtain the following bound.

$$\begin{aligned} U(t, m') - U(t, m) &\leq G(\tilde{m}_T) - G(m_T) + \epsilon + \\ &\quad + \int_t^T \int_{\mathbb{R}^d \times \mathbb{R}^d \times K} L(s, y, b(s, x, k), m'_s) - L(s, x, b(s, x, k), m_s) \eta_s(dx, dy, dk) ds \\ &\leq \int_t^T \int_{\mathbb{R}^d \times \mathbb{R}^d \times K} C(1 + |b(s, x, k)|)(|x - y| + W_2(m_s, m'_s)) \eta_s(dx, dy, dk) ds \\ &\quad + CW_2(m'_T, m_T) + \epsilon. \end{aligned}$$

Remark now that because of the growth condition on  $L$ , both

$$I := \int_t^T \int_{\mathbb{R}^d \times \mathbb{R}^d \times K} |b(s, x, k)|^2 \eta_s(dx, dy, dk) ds < \infty$$

and  $I \leq C(C + U(t, \mu))$ . Thus, using Cauchy-Schwarz inequality, and recalling that for any  $s \in [t, T]$ ,  $W_2^2(m_s, m'_s) \leq \mathbb{E}[|X_s - Y_s|^2] = \mathbb{E}[|X_t - Y_t|^2] = W_2^2(m, m')$ , we obtain the required result.  $\square$

The previous result was already known, and it was namely proven by controllability argument [5]. This proof makes transparent the link with standard proofs in finite dimensional cases, see for instance [15].

**Remark 7.14.** Our definition of  $U$  is a sort of randomization of the controls allowed in  $V$ , it can be analyzed in regards of what is called relaxed controls, or mixed strategies in game theory. It could have been formulated in terms of random variables, similarly to problem of optimal control of Mc-Kean Vlasov dynamics, namely by defining  $U$  as

$$U(t, m) := \inf_{(X, Z)} \left\{ \int_t^T \mathbb{E}[L(s, X_s, Z_s, \mathcal{L}(X_s))] ds + G(\mathcal{L}(X_T)) \right\},$$

where the infimum is taken over  $(X_s, Z_s)_{s \in [t, T]}$  such that  $\mathcal{L}(X_t) = m$  and, almost surely,  $X_s = X_t + \int_t^s Z_u du$ . Thanks to Proposition 7.2, the two definitions are equivalent.

We preferred the definition (14) as it makes transparent the fact that  $K$  is chosen by the controller. In the case of two players extensions of such problems, this has an importance, as the players can both randomize their strategies, but not on the same probabilistic space [7].

**7.4. Extensions.** We comment on several extensions of the previous method, namely to highlight that even if our formulation with stochastic processes is helpful, it does not rely by any mean on the Hilbertian structure of  $L^2(\Omega, \mathbb{R}^d)$ .

**7.4.1. More general controls than directly the speed.** In several practical situations, the control  $\alpha$  does not affect directly the continuity equation as in (8), but rather intervene through another function  $\beta$  as in

$$\partial_t m + \operatorname{div}_x(\beta(x, \alpha)m) = 0 \text{ in } (0, 1) \times \mathbb{R}^d.$$

This does not play a strong role in the previous. Using our method our translation on the process  $(X_s, Z_s)$  where this time  $z$  does not represent the speed directly but rather the control, we are still able to transport the control to another initial condition. Note that such a transport is then not the one of parallel transport on  $T\mathcal{P}_2(\mathbb{R}^d)$ , but rather a similar transport on a sheaf (see Section 2.2.3) which is not the tangent one. Under standard smoothness assumption on  $\beta$  (globally Lipschitz for instance), we are then able to recover similar estimates.

**7.4.2. Other integrability conditions than  $\mathcal{P}_2(\mathbb{R}^d)$ .** The restriction to 2nd order moments plays no fundamental role in the previous, and similar results could be obtained for  $\mathcal{P}_p(\mathbb{R}^d)$  with  $p \neq 2$ .

**7.4.3. Other cases than  $\mathcal{P}_2(\mathbb{R}^d)$ .** Cases in which we are interested in probability measures over a subdomain of  $\mathbb{R}^d$  works exactly in the same way, formally, except for the technical treatment of the boundary of the domain.

When dealing with a smooth manifold, a similar approach can also yield some results, although several difficulties may appear. In this case, the construction of translated curves is more involved, and we need to use the appropriate notion of parallel transport on the manifold. We briefly explain the main challenge at the level of stochastic processes, restricting also our attention to optimal couplings for the sake of clarity. Consider thus a smooth bounded manifold  $M$  of dimension  $d$ , a  $M$ -valued stochastic process  $(X_t)_{t \in [0,1]}$  and a  $\mathbb{R}^d$  valued process  $(Z_t)_{t \in [0,1]}$  which is interpreted as being in  $T_{X_t} M$  for almost every  $t, \omega$  and such that, almost surely,

$$\frac{dX_t}{dt} = Z_t.$$

Let  $Y_0$  be a  $M$ -valued random variable, and consider  $\eta \in \mathcal{P}(C([0,1], M))$ , concentrated on geodesics of  $M$ , such that  $(e_0)_\# \eta = \mathcal{L}(X_0)$  and  $(e_1)_\# \eta = \mathcal{L}(Y_0)$ . A translation of  $(\mathcal{L}(X_t))_{(t \in [0,1])}$  along  $\eta$  should then naturally be given by  $(Y_t)_{t \in [0,1]}$ , defined as being almost surely a solution of

$$(15) \quad \frac{dY_t}{dt} = \mathcal{T}^{\vartheta_t}(Z_t),$$

with initial condition  $Y_0$ , where  $\mathcal{T}^\theta(z)$  denotes the parallel transport of  $z$  along the path  $\theta$  and  $(\vartheta_t)_{t \in [0,1]}$  is random process taking value in the set of geodesics of  $M$ , such that

almost surely, for almost every  $t \in [0, 1]$ ,  $\vartheta_t(\omega)$  connects  $X_t(\omega)$  to  $Y_t(\omega)$ . The main issue is that, there is no canonical way to choose  $\vartheta$ , even with the knowledge of  $\eta$ , contrary to the flat case. The possible multiplicity of geodesics then implies that the parallel transport of  $Z_t(\omega)$  along these geodesics is not a Lipschitz function of  $Y_t(\omega)$ , which thus makes more challenging the resolution of (15).

We believe there are two main ways to tackle this problem. The first one is to prove existence of weaker notion of solution to (15), just as Theorem 7.1 provides a probabilistic solution to a singular ODE. The second one is to restrict our attention to cases in which  $X_t(\omega)$  and  $Y_t(\omega)$  are sufficiently close. Note that to prove Lipschitz estimates as in Theorem 7.13, we are mainly interested in  $\mu, \nu$  such that  $W_2(\mu, \nu)$  is small, and, as is the distance gets smaller and smaller, the probability that a unique geodesic connects  $X_t(\omega)$  and  $Y_t(\omega)$  get closer to one as the points are getting closer to one another with larger and larger probability.

In any case, we leave this question for future research.

## CONCLUSIONS

We have presented another definition of the tangent space  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$  than the one which is usually considered. It allowed us sufficient flexibility to tackle the different problems we were interested in. It seems that the only price to pay for such a generalization is the loss of uniqueness of several concepts (parallel transports or translated curves namely). We provided examples of non-uniqueness which highlight the fact that this non-uniqueness arise more from the nature of the space of probability measure than by something that we would have missed when defining  $T_\mu(\mathcal{P}_2(\mathbb{R}^d))$ .

Several times, we made an crucial use of working at the level of  $\mathcal{P}(\mathcal{C}([0, 1], \mathbb{R}^d))$  rather than at the one of  $\mathcal{C}([0, 1], \mathcal{P}(\mathbb{R}^d))$ , therefore adopting a Lagrangian point of view rather than an Eulerian one. This strategy, which is aligned with the lift on the space of random variables, seems to be quite fruitful. Without trying to diminish the obvious advantages of the Eulerian formulation (conciseness, numerical methods,...), we believe that for theoretical purposes, the Lagrangian formulation is quite often more helpful.

## ACKNOWLEDGMENTS

The author is thankful to Luigi Ambrosio, Averil Aussedat, Jean-Michel Lasry and Pierre-Louis Lions for helpful discussions and/or comments on the manuscript. The author also acknowledges a partial support from the Chair FDD (ILB) and the Lagrange Mathematics and Computing Research Center. This work was partially funded by the ERC project PaDiESeM.

## REFERENCES

- [1] Luigi Ambrosio and Nicola Gigli. Construction of the parallel transport in the Wasserstein space. *Methods Appl. Anal.*, 15(1):1–30, 2008.
- [2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Springer, 2005.
- [3] Averil Aussedat. On the structure of the geometric tangent cone to the Wasserstein space. *Journal of Differential Equations*, 442:113520, 2025.

- [4] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3): 375–393, 2000.
- [5] Charles Bertucci. Stochastic optimal transport and Hamilton-Jacobi-Bellman equations on the set of probability measures. *Annales de l'Institut Henri Poincaré C*, 2024.
- [6] Charles Bertucci. *Analysis on spaces of measures: Hamilton-Jacobi-Bellman and mean field games master equations*. Forthcoming, 2026.
- [7] Charles Bertucci. Hamilton-Jacobi-Isaacs equation on the Wasserstein space. work in progress.
- [8] Charles Bertucci and Giacomo Ceccherini Silberstein. Optimal control of positive measures through Hamilton-Jacobi-Bellman equations. to appear.
- [9] Charles Bertucci and Pierre Louis Lions. An approximation of the squared Wasserstein distance and an application to Hamilton-Jacobi equations. *arXiv preprint arXiv:2409.11793*, 2024.
- [10] Charles Bertucci and Giacomo Ceccherini Silberstein. On the doubling of variables technique in first order hamilton-jacobi equations. *arXiv preprint arXiv:2512.03652*, 2025.
- [11] René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications I-II*, volume 3. Springer, 2018.
- [12] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. Dissipative probability vector fields and generation of evolution semigroups in Wasserstein spaces. *Probability Theory and Related Fields*, 185(3):1087–1182, 2023.
- [13] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. A lagrangian approach to totally dissipative evolutions in Wasserstein spaces. *arXiv preprint arXiv:2305.05211*, 2023.
- [14] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. Stochastic Euler schemes and dissipative evolutions in the space of probability measures. *arXiv preprint arXiv:2505.20801*, 2025.
- [15] Wendell H Fleming and Halil Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.
- [16] Wilfrid Gangbo and Adrian Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 125:119–174, 2019.
- [17] Nicola Gigli. On the geometry of the space of probability measures in  $\mathbb{R}^n$  endowed with the quadratic optimal transport distance. *Tesi, Scuola Normale Superiore*, 2008.
- [18] Nicola Gigli. *Second Order Analysis on  $\mathcal{P}_2(M), W_2$* . American Mathematical Soc., 2012.
- [19] Pierre-Louis Lions. Cours au College de France. [www.college-de-france.fr](http://www.college-de-france.fr), 2011.
- [20] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. 2001.
- [21] Laurence Chisholm Young. Generalized curves and the existence of an attained absolute minimum in the calculus of variations. *Comptes Rendus de la Societe des Sci. et des Lettres de Varsovie*, 30:212–234, 1937.

<sup>A</sup> CEREMADE, CNRS, UNIVERSITÉ PARIS DAUPHINE-PSL, UMR 7534, 75016 PARIS, FRANCE.