



HAL
open science

Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition

Frédéric Jean, Othmane Jerhaoui, Hasnaa Zidani

► **To cite this version:**

Frédéric Jean, Othmane Jerhaoui, Hasnaa Zidani. Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition. 2022. hal-03564787v1

HAL Id: hal-03564787

<https://ensta-paris.hal.science/hal-03564787v1>

Preprint submitted on 10 Feb 2022 (v1), last revised 14 Sep 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition

Frédéric Jean^a, Othmane Jerhaoui^{a,b}, Hasnaa Zidani^b

^aUMA, ENSTA Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France

^bNormandie Univ, INSA Rouen Normandie, Laboratoire de Mathématiques Appliquées (LMI), 76000 Rouen, France

Abstract

In this article, we study an optimal control problem on a compact Riemannian manifold M with imperfect information on the initial state of the system. The lack of information is modelled by a Borel probability measure along which the initial state is distributed. The state space of this problem is the space of Borel probability measures over M . We define a notion of viscosity in this space by taking as test functions a subset of the set of functions that can be written as a difference of two semi-convex functions. With this choice of test functions, we extend the notion of viscosity solution to Hamilton-Jacobi-Bellman equations in Wasserstein space, we also establish that the value function of the control problem with imperfect information is the unique viscosity solution of a Hamilton-Jacobi-Bellman equation in the space of Borel probability measures.

Keywords: Hamilton-Jacobi equations in the Wasserstein space, viscosity solutions, differentiability in the Wasserstein space, optimal control theory

2010 MSC: 35R15, 49L25, 35D40

1. Introduction

Let $T > 0$, M be a compact Riemannian manifold and consider a dynamical system defined by

$$\begin{cases} \dot{Y}(t) = f(Y(t), u(t)), & t \in [t_0, T], \\ Y(t_0) = x_0, & u(t) \in U, \end{cases} \quad (1)$$

where $f : M \times U \rightarrow TM$ is the dynamics, $x_0 \in M$ and $t_0 \in [0, T]$. The set U is the set of admissible control values which is assumed to be a compact subset of some metric space. We associate to a measurable function $u : [t_0, T] \rightarrow U$ an absolutely continuous trajectory $t \rightarrow Y_t^{t_0, x_0, u}$ solution to (1). A classical control setting consists in minimizing a final cost

$$l(Y_T^{t_0, x_0, u}),$$

where $l : M \rightarrow \mathbb{R}$ is a given function, over all trajectories that are solutions to (1). This optimal control problem with perfect information has been extensively investigated, see for instance [34, 14, 5] when the dynamical system is defined in a vector space, and [1, 21, 22, 28] when the dynamical system is defined in a differentiable manifold. Here, the dynamical system is defined in a compact

Email addresses: Frederic.Jean@ensta-paris.fr (Frédéric Jean), Othmane.Jerhaoui@ensta-paris.fr (Othmane Jerhaoui), Hasnaa.Zidani@insa-rouen.fr (Hasnaa Zidani)

Riemannian manifold. This case appears in the modeling of some optimal control problems related to robotics [9], mechanical systems [7] or quantum systems.

In this paper, we consider optimal control problems with imperfect information. More precisely, the initial condition in the state space is not perfectly known and is distributed along some Borel probability measure μ_0 . We assume that both the dynamics and the probability measure are known. The lack of information is very specific in this case, it comes only from the initial condition. The final cost is the expectation of the deterministic final cost with respect to μ_0

$$L(\mu_0) = \int_M l(Y_T^{t_0, x_0, u}) d\mu_0(x_0). \quad (2)$$

Optimal control problems with partial information are of fundamental importance both in terms of real world applications and mathematical theory. Indeed, in many applications, the initial state of the system (1) is not perfectly known, either due to the lack of measurements, errors of measurements, or even due to the nature of the system itself, meaning that the uncertainties are inevitable. As for the theoretical interest, since the optimal control problem with partial information is defined in the space of Borel probability measures, we need to develop proper tools in this space to describe the problem. In fact, we want to study the evolution of the lack of information on the initial condition in (1), modeled by a Borel probability measure μ_0 . The evolution can be seen as a control system in the Wasserstein space $\mathcal{P}_2(M)$, where the trajectories in $\mathcal{P}_2(M)$ are obtained by pushing forward μ_0 by the trajectories of (1). Therefore, the trajectories are of the form

$$\begin{cases} \mu_t = Y_t^{t_0, \cdot, u} \# \mu_0, & t \in [t_0, T], \quad \text{and} \quad x \mapsto Y_t^{t_0, x, u} \text{ is the flow of (1),} \\ \mu_{t_0} = \mu_0. \end{cases}$$

To the above control system defined on $\mathcal{P}_2(M)$, we assign the following value function:

$$\vartheta(t, \mu) := \inf_{u(\cdot) \in U} \int l(Y_T^{t, x, u}) d\mu(x), \quad \text{for } t \in [0, T], \mu \in \mathcal{P}_2(M).$$

In [20], the authors studied a more general optimal control problem in Wasserstein space over \mathbb{R}^N by using non-smooth analysis techniques such as differential inclusions and sub/super differentials. A new approach was also considered in [27] to define dynamical systems on the space of probability measures based on tools similar to the theory of Young measures (see [8] for a detailed study on the subject).

The goal of this paper is to characterize the value function as a unique solution to a Hamilton-Jacobi-Bellman (HJB) equation defined on the Wasserstein space $\mathcal{P}_2(M)$. Ideally, the HJB equation should have the following form:

$$\begin{cases} \partial_t v + H(\mu, D_\mu v) = 0, & (t, \mu) \in [0, T) \times \mathcal{P}_2(M), \\ v(T, \mu) = L(\mu). \end{cases} \quad (3)$$

Furthermore, we want to transpose the techniques commonly used in viscosity theory [15, 6] to the space $\mathcal{P}_2(M)$. To achieve this goal, it is necessary to go through several steps. First, we define a notion of classical solutions to the HJB equation. Then, we define the set of test functions for viscosity super- and sub-solutions. Finally, we prove a comparison principle that holds for

any supersolution and any subsolution, which will guarantee uniqueness of the viscosity solution. Besides, we want to prove that the value function verifies a dynamic programming principle and is the unique solution of a HJB equation. The difficulty here is that the space of Borel probability measures does not have a smooth structure. This space is infinite dimensional and when equipped with the Wasserstein distance d_W , it is a compact and separable geodesic space (see for instance [33, 32, 3, 4]).

In [17, 13, 18, 20] and references therein, viscosity solutions are defined by using a notion of viscosity sub/super gradient. Our approach is different. We aim at defining a viscosity notion (and test functions) in the Wasserstein space $(\mathcal{P}_2(M), d_W)$ in a way that generalizes the one already known in Euclidean spaces or Riemannian manifolds. In $\mathcal{P}_2(M)$, real valued Lipschitz functions that can be represented as a difference of two semi-convex functions (DC functions in short) admit directional derivatives at every point $\mu \in \mathcal{P}_2(M)$. The key idea of this paper is to consider DC functions as test functions, and define the notion of viscosity solutions using this set of functions. To do so, we use the characterization of geodesics in $\mathcal{P}_2(M)$ to define directional derivatives in $\mathcal{P}_2(M)$ and we define an appropriate notion of tangent space at any point $\mu \in \mathcal{P}_2(M)$ that encodes all the information about initial velocities of geodesics in the Wasserstein space.

This paper is organized as follows: in section 2, we introduce the optimal control problem and we prove that the value function is Lipschitz continuous and verifies a dynamic programming principle. In section 3, we recall some geometric properties of the Wasserstein space $(\mathcal{P}_2(M), d_W)$. In particular (see [19, 24]), we recall an abstract notion of tangent cone at a point μ in $\mathcal{P}_2(M)$. We then define a notion of differentiable functions in $\mathcal{P}_2(M)$ and we discuss the relation between the “analytic tangent space” usually used in optimal transport and the tangent cone at μ . In section 4, we present the main contribution of the present paper. We go back to our optimal control problem and we prove that the value function is the unique viscosity solution to an HJB equation of the form (3) by transposing viscosity theory techniques to the Wasserstein space $\mathcal{P}_2(M)$.

2. Setting of the problem

Throughout this manuscript, $(M, \langle \cdot, \cdot \rangle)$ is a finite dimensional, compact and connected Riemannian manifold without boundary. We denote by $|\cdot|$ the associated norm on the tangent bundle TM , and by $d(\cdot, \cdot)$ its Riemannian distance on M . The metric space (M, d) , is a complete and compact space and its topology is equivalent to the topology of the differentiable manifold M . The tangent bundle TM is itself a complete Riemannian manifold when endowed with the Sasaki metric [30]. We denote by $d_{TM}(\cdot, \cdot)$ the Riemannian distance on TM associated with the Sasaki metric (see Appendix B).

We denote by $\mathcal{P}(M)$ the set of Borel probability measures over M and $\mathcal{P}_2(M)$ the set of Borel probability measures with bounded second moment

$$\mathcal{P}_2(M) := \{ \mu \in \mathcal{P}(M) : \int d^2(x, x_0) d\mu(x) < \infty, \quad \forall x_0 \in M \}.$$

Actually, since M is compact, we have $\mathcal{P}_2(M) = \mathcal{P}(M)$ but we will keep using the notation $\mathcal{P}_2(M)$. Recall that for any two topological spaces X and Y , any Borel probability measure μ on X and any Borel function $g : X \rightarrow Y$, the pushforward measure $g\#\mu$ on Y is defined by

$$g\#\mu(A) = \mu(g^{-1}(A)) \quad \forall A \subset Y, \text{ a Borel set,}$$

or equivalently,

$$\int h dg\# \mu = \int h \circ g d\mu, \quad \forall h : Y \rightarrow \mathbb{R}, \text{ Borel measurable and bounded.}$$

We define the Wasserstein distance $d_W(\cdot, \cdot)$ over $\mathcal{P}_2(M)$ by

$$d_W(\mu, \nu) := \sqrt{\inf \left\{ \int d^2(x, y) d\gamma(x, y) \right\}},$$

where the infimum is taken over all Borel probability measures of $M \times M$ that have marginals μ and ν , i.e.

$$\gamma(A \times M) = \mu(A) \quad \text{and} \quad \gamma(M \times B) = \nu(B) \quad \forall A, B, \text{ Borel sets of } M.$$

Such Borel probability measures are called admissible plans of μ and ν and the set of such plans is denoted $Adm(\mu, \nu)$. It is well known that d_W verifies all the axioms of a distance and that the infimum is always reached [3, Theorem 1.5]. The admissible plans where the minimum is achieved are called optimal transport plans and the set of such plans is denoted $Opt(\mu, \nu) \subset Adm(\mu, \nu)$.

Let $T > 0$ and U be a compact subset of a metric space. Consider the dynamics, defined for $T > t_0 \geq 0$ and $x_0 \in M$, as

$$\begin{cases} \dot{Y}(t) = f(Y(t), u(t)), & t \in [t_0, T], \\ Y(t_0) = x_0, u(t) \in U, \end{cases} \quad (4)$$

where $f : M \times U \rightarrow TM$ satisfies the following assumptions:

$$(\mathbf{H}) \begin{cases} f : M \times U \rightarrow TM \text{ is continuous and Lipschitz continuous with respect to the state, i.e.} \\ \exists k > 0 : \quad d_{TM}(f(x, u), f(y, u)) \leq k d(x, y), \quad \forall u \in U, (x, y) \in M \times M. \end{cases}$$

$(\mathbf{H})_{co}$: for all $x \in M$, the set $f(x, U) := \{f(x, u) : u \in U\}$ is convex.

Remark 2.1. Hypotheses (\mathbf{H}) and $(\mathbf{H})_{co}$ are classical for optimal control problems (see [14, 5] for a detailed study). Also, since M and U are compact, then the vector field f is bounded. Furthermore, the Lipschitz assumption on $f(\cdot, u)$ in Hypothesis (\mathbf{H}) is equivalent to the following: there exists $k' > 0$ such that for all $u \in U$, $x, y \in M$ and every smooth curve $\alpha : [0, 1] \rightarrow M$ joining x and y , we have

$$|\tau_{x,y}^\alpha(f(x, u)) - f(y, u)| \leq k' \text{length}(\alpha),$$

with $\tau_{x,y}^\alpha$ is the parallel transport of $f(x, u)$ along the curve α and $\text{length}(\alpha)$ is the Riemannian length of the curve α (see [31, Lemma II.A.2.4]). We set

$$Lip(f) := \max(k, k').$$

We define the set of open-loop controls by

$$\mathcal{U} := \{u : [0, T] \rightarrow U : u(\cdot) \text{ is measurable}\}.$$

Under the assumption **(H)**, classical results of ordinary differential equations hold. In particular, for any control $u(\cdot) \in \mathcal{U}$ and $x_0 \in M$, there exists a unique Lipschitz trajectory $t \mapsto Y_t^{t_0, x_0, u}$ defined on all $[t_0, T]$. Moreover, we have the following estimates:

Proposition 2.2. *There exist $C_1, C_2 > 0$ positive constants such that for all $x_0, z_0 \in M$, for all $t_0 \in [0, T]$, and $t \mapsto Y_t^{t_0, x_0, u}$, $t \mapsto Y_t^{t_0, z_0, u}$ be solutions of (4), it holds:*

$$\begin{aligned} d(Y_T^{t_0, x_0, u}, Y_T^{t_0, z_0, u}) &\leq C_1 d(x_0, z_0), \\ d(Y_t^{t_0, x_0, u}, x_0) &\leq C_2 |t - t_0|, \quad t \in [t_0, T]. \end{aligned}$$

Proof. (Sketch). Since M is compact, then all the statements are local in nature. The global result is obtained by compactness of M and of $[t_0, T]$. First, by using Nash embedding theorem, M can be embedded isometrically into a Euclidean space $(\mathbb{R}^N, \|\cdot\|)$, with $N > 0$ big enough. Let $x_0 \in M$ and V be a small enough open neighborhood of x_0 . Then for $z_0 \in V$ we can apply the usual theory in \mathbb{R}^N and get

$$\|Y_T^{t_0, x_0, u} - Y_T^{t_0, z_0, u}\| \leq e^{Lip(f)T} \|x_0 - z_0\|.$$

Then by using the fact that the Euclidean distance is equivalent to the Riemannian distance in V , we get the result. The second assertion can be established with similar arguments, by taking t small enough so that $Y_t^{t_0, x_0, u} \in V$. \square

The control problem aims at minimizing the final cost

$$L(\mu_0) = \int l(Y_T^{t_0, x_0, u}) d\mu_0(x_0),$$

over all trajectories that are solutions of the dynamics (4) with the initial condition $x_0 \in M$, distributed along the measure $\mu_0 \in \mathcal{P}_2(M)$. We consider the following assumption:

(H_l) $l : M \rightarrow \mathbb{R}$ is Lipschitz continuous with constant $Lip(l)$.

When μ_0 is equal to the Dirac mass δ_{x_0} , the resulting system corresponds to the classical case without uncertainties on the initial condition. This problem is thoroughly studied in the literature. In particular, it is known that the set of trajectories of (4) is a compact set of $C([t_0, T], M)$ and therefore the optimal control problem admits a solution. When μ_0 is any probability measure of $\mathcal{P}_2(M)$, it is better to see this problem as an optimal control problem defined on the space of Borel probability measures $\mathcal{P}_2(M)$. First we rewrite the final cost the following way

$$L(\mu_0) = \int l(Y_T^{t_0, x_0, u}) d\mu_0(x_0) = \int l(y) dY_t^{t_0, \dots, u} \# \mu_0(y),$$

and we minimize this cost over the set of trajectories $t \mapsto \mu_t^{\mu_0, u}$ of the space $\mathcal{P}_2(M)$ that verify

$$\begin{cases} \mu_t^{\mu_0, u} = Y_t^{t_0, \dots, u} \# \mu_0, & t \in [t_0, T], \text{ and } x \mapsto Y_t^{t_0, x, u} \text{ is the flow of (4),} \\ \mu_{t_0}^{\mu_0, u} = \mu_0. \end{cases}$$

Furthermore, since for any $t \mapsto u(t)$, the vector field $f(\cdot, u(t))$ is uniformly Lipschitz continuous (see Remark 2.1), then the trajectory $t \mapsto \mu_t^{\mu_0, u}$ is the unique continuous solution of the continuity equation (see [4, 8])

$$\begin{cases} \partial_t \mu_t^{\mu_0, u} + \operatorname{div}(f(\cdot, u(t)) \mu_t^{\mu_0, u}) = 0, & t \in (t_0, T), \\ \mu_{t_0}^{\mu_0, u} = \mu_0, \end{cases}$$

and the equation is understood in the sense of distributions, i.e.

$$\int_0^T \int_M (\partial_t \phi(t, x) + \langle \nabla_x \phi(t, x), f(x, u(t)) \rangle) d\mu_t(x) dt = 0, \quad \forall \phi \in C_c^\infty((0, T) \times M).$$

Therefore, the above optimal control problem can be rewritten as

$$\begin{cases} \inf_{u \in \mathcal{U}} \int l(y) d\mu_T^{\mu_0, u}(y), \\ \text{such that } \begin{cases} \partial_t \mu_t^{\mu_0, u} + \operatorname{div}(f(\cdot, u(t)) \mu_t^{\mu_0, u}) = 0, & t \in (t_0, T), \\ \mu_{t_0}^{\mu_0, u} = \mu_0, \end{cases} \end{cases} \quad (5)$$

and the infimum is reached. It is worth pointing out that unlike the existing literature on this particular problem (see [13] and references therein) or more generally optimal control problems in the space of probability measures [29], the base space here is not the Euclidean space but rather a compact Riemannian manifold.

Under hypotheses **(H)** and **(H_l)**, we can already prove two properties of the value function.

Theorem 2.3 (Dynamic programming principle). *Let $\mu \in \mathcal{P}_2(M)$, $t \in [0, T]$ and $h \in [t, T - t]$. Then it holds*

$$\vartheta(t, \mu) = \inf_{u \in \mathcal{U}} \vartheta(t + h, Y_{t+h}^{t, \cdot, u} \# \mu).$$

Proof. Let $u_0(\cdot) \in U$ be such that

$$\vartheta(t, \mu) = \int l(Y_T^{t, x, u_0}) d\mu(x).$$

We have

$$\begin{aligned} \vartheta(t, \mu) &= \int l(Y_T^{t, x, u_0}) d\mu(x) = \int l(Y_T^{t+h, x, u_0}) dY_{t+h}^{t, \cdot, u_0} \# \mu(x) d\mu(x) \\ &\geq \inf_{u \in U} \int l(Y_T^{t+h, x, u}) dY_{t+h}^{t, \cdot, u} \# \mu(x) = \vartheta(t + h, Y_{t+h}^{t, \cdot, u} \# \mu). \end{aligned}$$

It remains to prove the other inequality. There exists, $u(\cdot) : [t, T] \rightarrow U$ such that the restriction $u_h : u|_{[t+h, T]}$ verifies

$$\vartheta(t + h, Y_{t+h}^{t, \cdot, u} \# \mu) = \int l(Y_T^{t+h, x, u_h}) dY_{t+h}^{t, \cdot, u_h} \# \mu(x).$$

On the other hand, we have

$$\vartheta(t, \mu) \leq \int l(Y_T^{t, x, u}) d\mu(x) = \int l(Y_T^{t+h, x, u_h}) dY_{t+h}^{t, \cdot, u_h} \# \mu(x) \leq \vartheta(t + h, Y_{t+h}^{t, \cdot, u} \# \mu),$$

hence the result. \square

Proposition 2.4. *Assume (H) and (H₁). Then, the value function ϑ is Lipschitz continuous on $[0, T] \times \mathcal{P}_2(M)$. In particular, ϑ is bounded.*

Proof. Let $t \in [0, T]$, $\mu, \sigma \in \mathcal{P}_2(M)$. There exists a trajectory $s \mapsto Y_s^{t,x,u}$ such that

$$\int l(Y_T^{t,x,u}) d\sigma(x) = \vartheta(t, \sigma).$$

Hence, we have

$$\vartheta(t, \mu) - \vartheta(t, \sigma) \leq \int l(Y_T^{t,x,u}) d\mu(x) - \int l(Y_T^{t,x,u}) d\sigma(x).$$

Let $\gamma \in Opt(\mu, \sigma)$. Then we get

$$\begin{aligned} \int l(Y_T^{t,x,u}) d\mu(x) - \int l(Y_T^{t,x,u}) d\sigma(x) &= \int (l(Y_T^{t,x,u}) - l(Y_T^{t,y,u})) d\gamma(x, y) \\ &\leq Lip(l) C_1 \int d(x, y) d\gamma(x, y) \\ &\leq Lip(l) C_1 \sqrt{\int d^2(x, y) d\gamma(x, y)} = Lip(l) C_1 d_W(\mu, \sigma), \end{aligned}$$

where $C_1 > 0$ is defined in Proposition 2.2. Thus we get

$$\vartheta(t, \mu) - \vartheta(t, \sigma) \leq Lip(l) C_1 d_W(\mu, \sigma).$$

We can exchange the roles of σ and μ to get the exact same inequality. Therefore, we get the Lipschitz continuity with respect to the state variable. To prove Lipschitz continuity with respect to time, let $t, s \in [0, T]$. We assume, without loss of generality, that $0 \leq t < s \leq T$. By Theorem 2.3, there exists a trajectory $r \mapsto Y_r^{t,x,u}$ such that

$$\vartheta(t, \sigma) = \vartheta(s, Y_s^{t,\cdot,u} \# \sigma).$$

We have

$$\begin{aligned} |\vartheta(s, \sigma) - \vartheta(t, \sigma)| &= |\vartheta(s, \sigma) - \vartheta(s, Y_s^{t,\cdot,u} \# \sigma)| \\ &\leq Lip(l) C_1 d_W(\sigma, Y_s^{t,\cdot,u} \# \sigma) \\ &\leq C_2 Lip(f) C_1 |t - s|, \end{aligned}$$

where $C_1, C_2 > 0$ are defined in Proposition 2.2. Thus ϑ is Lipschitz continuous with respect to the time variable, and the proof is completed. \square

In the classical theory of viscosity, the value function is the unique viscosity solution of the Hamilton Jacobi Bellman equation. The goal of the next two sections is to show that the value function, in this setting, is also a viscosity solution to a Hamilton Jacobi Bellman equation of the form

$$\begin{cases} \partial_t v + H(\mu, D_\mu v) = 0, & (t, \mu) \in [0, T] \times \mathcal{P}_2(M), \\ v(T, \mu) = L(\mu), \end{cases}$$

where the Hamiltonian will be defined precisely later. To do so, we will define a notion of differentiable functions in the Wasserstein space so that the notation $D_\mu v$ will become rigorous. Furthermore, we will define the test functions we are going to use to define viscosity solutions. This requires us to first study the geometric structure of the Wasserstein space.

3. Wasserstein space over compact Riemannian Manifolds

The results of this section are taken from [19, 2]. The first subsection aims to describe the geodesics in Wasserstein space. In the second subsection we give the definition of the tangent cone at a point $\mu \in \mathcal{P}_2(M)$ obtained using tools of metric geometry. Finally in the last subsection we give the definition of differentiable functions in Wasserstein spaces and we give an explicit expression of the differential of the squared Wasserstein distance. All these tools are going to be necessary to give a precise definition of the *Hamiltonian*, *classical solutions* and *viscosity notion* for HJB equations in $\mathcal{P}_2(M)$.

3.1. Geometric and topological properties of Wasserstein space

The Wasserstein space $(\mathcal{P}_2(M), d_W)$ shares many geometric and topological properties with the base space (M, d) . Indeed, since (M, d) is a Polish space (because it is a complete and separable metric space), then $(\mathcal{P}_2(M), d_W)$ is a Polish space. Also, since M is compact, then $(\mathcal{P}_2(M), d_W)$ is also compact (see [33, chapter 6]).

Next, we recall some general definitions on metric spaces. Let (X, d_X) be a metric space. A curve $\alpha : [0, 1] \rightarrow X$ is called a *minimizing constant speed geodesic* if

$$d_X(\alpha_t, \alpha_s) = |t - s|d_X(\alpha_0, \alpha_1).$$

The metric space (X, d_X) is said to be a *geodesic space* if any two points of X are connected by at least one minimizing constant speed geodesic. In what follows, we intend by ‘geodesic’, a minimizing constant speed geodesic. Note that the metric spaces (M, d) and (TM, d_{TM}) are geodesic spaces. Furthermore, the Wasserstein space $(\mathcal{P}_2(M), d_W)$ inherits this property from (M, d) and is also a geodesic space (see [4] or [33]).

We denote by $\mathcal{P}(TM)$ the set of Borel probability measures over TM . Since (TM, d_{TM}) is a complete geodesic space, we can define the Wasserstein space over TM

$$\mathcal{P}_2(TM) = \left\{ \eta \in \mathcal{P}(TM) : \int d_{TM}^2((x, v), (x_0, v_0)) d\eta(x, v) < \infty, \quad \forall (x_0, v_0) \in TM \right\} \quad (6)$$

endowed with the following Wasserstein distance for any $\eta, \gamma \in \mathcal{P}_2(TM)$:

$$d_W(\gamma, \eta) := \sqrt{\inf \left\{ \int d_{TM}^2(x, y) d\alpha(x, y) \right\}},$$

the infimum is taken over all admissible plans α with marginals γ and η . Notice that we kept the same notation for the Wasserstein distance. It would be clear from the context which base space is considered. Moreover, it is sufficient that the condition

$$\int d_{TM}^2((x, v), (x_0, v_0)) d\eta(x, v) < \infty, \quad \forall (x_0, v_0) \in TM$$

in (6) be verified for only one point $(x_0, v_0) \in TM$. Thus if we take $(x_0, 0_{x_0}) \in TM$, then this condition is equivalent to

$$\int |v|^2 d\eta(x, v) < \infty.$$

For $\mu \in \mathcal{P}_2(M)$, we denote by $\mathcal{P}_2(TM)_\mu \subset \mathcal{P}_2(TM)$, the set of plans γ such that $\pi^M \# \gamma = \mu$, where $\pi^M : TM \rightarrow M$ is the canonical projection onto M . This set is equivalent to the set of plans $\gamma \in \mathcal{P}(TM)$ such that

$$\pi^M \# \gamma = \mu, \quad \text{and} \quad \int |v|^2 d\gamma(x, v) < \infty.$$

Let $\exp : TM \rightarrow M$ be the exponential map of $(M, \langle \cdot, \cdot \rangle)$. The exponential $\mathbf{exp}_\mu(\gamma)$ of a plan $\gamma \in \mathcal{P}_2(TM)_\mu$ is defined by

$$\mathbf{exp}_\mu(\gamma) := \exp \# \gamma \in \mathcal{P}_2(M).$$

We define the map $\mathbf{exp}_\mu^{-1} : \mathcal{P}_2(M) \rightarrow \mathcal{P}_2(TM)_\mu$ by

$$\mathbf{exp}_\mu^{-1}(\nu) := \{ \gamma \in \mathcal{P}_2(TM)_\mu : \mathbf{exp}_\mu(\gamma) = \nu \text{ and } \int |v|^2 d\gamma(x, v) = (d_W(\mu, \nu))^2 \},$$

or in other words, the set of plans $\gamma \in \mathcal{P}_2(TM)$ such that $(\pi^M, \exp) \# \gamma$ is an optimal plan from μ to ν and $\int |v|^2 d\gamma(x, v) = (d_W(\mu, \nu))^2$. We introduce the following notation

$$\Delta_t(x, v) = (x, tv), \quad \forall t \in \mathbb{R}, (x, v) \in TM.$$

Remark 3.1. *The map \mathbf{exp}_μ^{-1} is not really an inverse map to \mathbf{exp}_μ since only optimal plans in the inverse image of ν are considered. While this might seem confusing, the map \mathbf{exp}_μ^{-1} is defined this way so that for all $\gamma \in \mathbf{exp}_\mu^{-1}(\nu)$, the curve $t \rightarrow \exp(\Delta_t) \# \gamma$ is a geodesic connecting μ and ν , see the theorem below.*

Theorem 3.2. *([19, Theorem 1.11]) Let $\mu, \nu \in \mathcal{P}_2(M)$. A curve $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}_2(M)$ is a geodesic connecting μ to ν if and only if there exists a plan $\gamma \in \mathbf{exp}_\mu^{-1}(\nu)$ such that*

$$\mu_t := \exp \circ \Delta_t \# \gamma, \quad \forall t \in [0, 1]. \tag{7}$$

The plan γ is uniquely identified by the geodesic. Moreover, for any $t \in (0, 1)$ there exists a unique optimal plan from μ to μ_t . Finally, if there exist two different geodesics connecting μ to ν , they do not intersect in intermediate times (i.e. on $(0, 1)$).

Introducing the following rescaling of a plan:

$$t \cdot \gamma = \Delta_t \# \gamma, \quad \forall t \in \mathbb{R}, \gamma \in \mathcal{P}_2(TM),$$

equation (7) can be rewritten in a more elegant way as

$$\mu_t = \exp \circ \Delta_t \# \gamma = \mathbf{exp}_\mu(t \cdot \gamma), \quad \forall t \in [0, 1].$$

From Theorem 3.2, we get the following result about geodesics emanating from any $\mu \in \mathcal{P}_2(M)$.

Proposition 3.3. *([19, Proposition 1.12]) Let $\mu \in \mathcal{P}_2(M)$ and let $(\mu_t)_t$ be a geodesic emanating from μ and defined in some interval $[0, \varepsilon]$, with $\varepsilon > 0$. Then there exists a unique plan $\gamma \in \mathcal{P}_2(TM)_\mu$ such that*

$$\mu_t = \mathbf{exp}_\mu(t \cdot \gamma), \quad \forall t \in [0, \varepsilon].$$

In what follows, we will need to use couplings of plans in $\mathcal{P}_2(TM)_\mu$. We denote by T^2M the set

$$T^2M := \left\{ (x, v_1, v_2) : v_1, v_2 \in T_xM \right\}.$$

We equip this set with the following distance:

$$d_{T^2M}^2\left((x, v_1, v_2), (y, w_1, w_2)\right) := d_{TM}^2\left((x, v_1), (y, w_1)\right) + d_{TM}^2\left((x, v_2), (y, w_2)\right).$$

We define the following three projections π^M, π^1, π^2

$$\pi^M(x, v_1, v_2) = x \in M, \quad \pi^1(x, v_1, v_2) = (x, v_1) \in TM, \quad \pi^2(x, v_1, v_2) = (x, v_2) \in TM.$$

Then, consider the Wasserstein space $\mathcal{P}_2(T^2M)$ over $(T^2M, d_{T^2M}^2)$. A plan $\alpha \in \mathcal{P}_2(T^2M)$ is said to be an admissible coupling of $\gamma_1, \gamma_2 \in \mathcal{P}_2(TM)_\mu$ if

$$\pi^1\#\alpha = \gamma_1, \quad \pi^2\#\alpha = \gamma_2,$$

and we write $\alpha \in \text{Adm}_\mu(\gamma_1, \gamma_2)$.

To summarize, we have seen in this section that the Wasserstein space $(\mathcal{P}_2(M), d_W)$ is a compact geodesic space, and its geodesics can be characterized by plans of $\mathcal{P}_2(TM)$ as shown in Proposition 3.3. Next, we are going to define a notion of ‘‘tangent space’’ of the Wasserstein space.

3.2. Tangent cone of Wasserstein space

The space $(\mathcal{P}_2(M), d_W)$ has a formal Riemannian structure. This has been first pointed out by Otto in [25]. loosely speaking, it was noticed that given an absolutely continuous curve $I \ni t \mapsto \mu_t$ in $\mathcal{P}_2(M)$, there exists a family of vector fields (v_t) in M , defined for almost every t , such that $v_t \in L_{\mu_t}^2$ (with $L_{\mu_t}^2$ is the space of squared integrable vector fields with respect to μ_t) and the following continuity equation

$$\frac{d}{dt}\mu_t + \text{div}(v_t\mu_t) = 0, \tag{8}$$

is satisfied in the distributional sense, i.e.

$$\int_I \int_M (\partial_t \phi(t, x) + \langle \nabla_x \phi(t, x), v_t(x) \rangle) d\mu_t(x) dt = 0, \quad \forall \phi \in C_c^\infty(I \times M).$$

First, notice that there is no unique choice of v_t . Indeed, all the vector fields of the form

$$v_t + \{\nabla\psi : \psi \in C^\infty(M)\}^\perp$$

also satisfy (8) in the distributional sense (the set $\{\nabla\psi : \psi \in C^\infty(M)\}^\perp$ is the orthogonal set to $\{\nabla\psi : \psi \in C^\infty(M)\}$ in $L_{\mu_t}^2$). This means that v_t acts only on gradients of smooth functions in the continuity equation. Furthermore, the previous observation means that one can select the v_t 's to belong to the set of gradients. Hence, one could consider the following *space of gradients* at some $\mu \in \mathcal{P}_2(M)$

$$\text{SpGr}_\mu(\mathcal{P}_2(M)) := \overline{\{\nabla\phi : \phi \in C^\infty(M)\}}^{L_\mu^2},$$

to be the tangent space of $\mathcal{P}_2(M)$ at μ . This construction is analytical and it has the advantage to retain the link between the tangent space and the continuity equation (8). On the other hand, it was shown in [19, 24] that the space of gradients is linked to the so-called *tangent cone*, a notion that is central in *metric geometry*, much like the notion of tangent space in Riemannian geometry, and could be defined for any geodesic space. The tangent cone is constructed in a purely geometric way using only germs of geodesics. Roughly speaking, geodesics in the Wasserstein space are of the following form:

$$\mu_t = \mathbf{exp}_\mu(t \cdot \gamma), \quad \gamma \in \mathcal{P}_2(TM)_\mu.$$

These plans can be seen as “initial velocities” of Wasserstein geodesics. The set of such plans that produce geodesics starting from μ_0 generates the tangent cone at μ_0 . Actually, we will see that we have the following analogy between M and $\mathcal{P}_2(M)$:

Space	M	$\mathcal{P}_2(M)$
Geodesics	$t \rightarrow \exp_x(tv)$	$t \rightarrow \mathbf{exp}_\mu(t \cdot \gamma)$
Tangent space	$T_x M$	$\gamma \in \mathcal{P}_2(TM)_\mu$, such that $t \rightarrow \mathbf{exp}_\mu(t \cdot \gamma)$ is a geodesic
Tangent bundle	TM	$\mathcal{P}_2(TM)$

We start first by introducing the tangent cone at any point $\mu \in \mathcal{P}_2(M)$, then we will define a notion of *differentiable functions* in $\mathcal{P}_2(M)$ and show that the squared Wasserstein distance is differentiable. Finally, we will highlight the link between the space of gradients and the tangent cone.

Let $\mu \in \mathcal{P}_2(M)$. We denote by $\mathbf{Geo}_\mu(\mathcal{P}_2(M))$ the set of all geodesics emanating from μ . Let $(\mu_t), (\nu_t) \in \mathbf{Geo}_\mu(\mathcal{P}_2(M))$ be two geodesics emanating from μ . We define the following equivalence relation on $\mathbf{Geo}_\mu(\mathcal{P}_2(M))$:

$$(\mu_t) \mathcal{R} (\nu_t) \quad \text{if and only if} \quad \mu_t = \nu_t \text{ in some right neighborhood } [0, \varepsilon], \text{ with } \varepsilon > 0.$$

The quotient space $\mathbf{Geo}_\mu(\mathcal{P}_2(M))/\mathcal{R}$, is called the *spaces of directions*. We denote the equivalence class of the geodesic (μ_t) starting from μ by $[\mu_t]$. This equivalence class represents the initial velocities of geodesics starting from μ .

Proposition 3.4. ([19, Proposition 3.6]). *Let $\mu \in \mathcal{P}_2(M)$ and $(\mu_t), (\nu_t) \in \mathbf{Geo}_\mu(\mathcal{P}_2(M))$ be two geodesics emanating from μ and defined in some right neighborhood of 0. Then the following limit exists:*

$$d_\mu^W([\mu_t], [\nu_t]) := \lim_{\varepsilon \downarrow 0} \frac{d_W(\mu_{\varepsilon t}, \nu_{\varepsilon t})}{\varepsilon}.$$

This limit defines a distance in the space of directions $\mathbf{Geo}_\mu(\mathcal{P}_2(M))/\mathcal{R}$.

The completion of $\mathbf{Geo}_\mu(\mathcal{P}_2(M))/\mathcal{R}$ with respect to d_μ^W is going to play the role of the tangent space. Actually, this set does not have the structure of a vector space but rather a structure of a cone. Hence it is called the *tangent cone*.

Definition 3.5. ([19, Definition 3.7] *Tangent cone*). *Let $\mu \in \mathcal{P}_2(M)$. The tangent cone at μ of $\mathcal{P}_2(M)$ is defined as:*

$$T_\mu \mathcal{P}_2(M) := \overline{\mathbf{Geo}_\mu(\mathcal{P}_2(M))/\mathcal{R}}^{d_\mu^W},$$

the closure is with respect to the distance d_μ^W .

This notion of tangent cone is abstract and it is constructed only from the geodesics and the Wasserstein distance. Actually, this construction is valid for any geodesic space (see [12, 10, 11, 2] for more details). From its definition, the tangent cone is defined only up to an isometry. Therefore, we are going to choose an isometric representative specific to the Wasserstein space.

Definition 3.6 (The distance W_μ). *Let $\mu \in \mathcal{P}_2(M)$ and $\gamma, \eta \in \mathcal{P}_2(TM)_\mu$. We define the following distance on $\mathcal{P}_2(TM)_\mu$:*

$$W_\mu^2(\gamma, \eta) := \int \left(d_W(\gamma_x, \eta_x) \right)^2 d\mu(x),$$

where $(\gamma_x)_{x \in M}$ and $(\eta_x)_{x \in M}$ are the disintegration with respect to the projection π^M (see Appendix A).

Proposition 3.7. ([19, Proposition 5.2]). *Let $\mu \in \mathcal{P}_2(M)$ and $\gamma, \eta \in \mathcal{P}_2(TM)_\mu$. Then*

$$W_\mu(\gamma, \eta) = \sqrt{\inf_{\alpha \in \text{Adm}_\mu(\gamma, \eta)} \int |v_1 - v_2|^2 d\alpha(x, v_1, v_2)},$$

where $\alpha \in \mathcal{P}_2(T^2M)$ is an admissible coupling of γ and η . The infimum is achieved, W_μ defines a distance on $\mathcal{P}_2(TM)_\mu$ and the metric space $(\mathcal{P}_2(TM)_\mu, W_\mu)$ is complete and separable.

Now, we get to the isometric representative of the tangent cone.

Theorem 3.8. ([19, theorem 5.5] Representation of the tangent cone). *Let $\mu \in \mathcal{P}_2(M)$. The space of directions $(\text{Geo}_\mu(\mathcal{P}_2(M))/\mathcal{R}, d_\mu^W)$ is (isometric to) the following subset of $(\mathcal{P}_2(TM)_\mu, W_\mu)$,*

$$\text{Dir}_\mu := \left\{ \gamma \in \mathcal{P}_2(TM)_\mu : t \mapsto \mathbf{exp}_\mu(t \cdot \gamma) \text{ is a geodesic in some neighborhood } [0, \varepsilon]. \right\}.$$

Given two geodesics $t \mapsto \mu_t = \mathbf{exp}_\mu(t \cdot \gamma)$, $t \mapsto \nu_t = \mathbf{exp}_\mu(t \cdot \eta)$, starting from μ and defined in a right neighborhood of 0, we have the following equality:

$$d_\mu^W([\mu_t], [\nu_t]) = \lim_{\varepsilon \downarrow 0} \frac{d_W(\mathbf{exp}_\mu(\varepsilon t \cdot \gamma), \mathbf{exp}_\mu(\varepsilon s \cdot \eta))}{\varepsilon} = \sqrt{\int \left(d_W(\gamma_x, \eta_x) \right)^2 d\mu(x)} = W_\mu(\gamma, \eta).$$

Hence, we write

$$[\mu_t] = \eta, \quad [\nu_t] = \gamma,$$

to be the directions of (μ_t) and (ν_t) respectively.

The tangent cone $(T_\mu \mathcal{P}_2(M), d_\mu^W)$ is (isometric to) the following subset

$$\begin{aligned} T_\mu \mathcal{P}_2(M) &= \overline{\text{Dir}_\mu}^{W_\mu} \\ &= \overline{\left\{ \gamma \in \mathcal{P}_2(TM)_\mu : t \mapsto \mathbf{exp}_\mu(t \cdot \gamma) \text{ is a geodesic in some neighborhood } [0, \varepsilon]. \right\}}^{W_\mu} \end{aligned} \quad (9)$$

with the closure taken with respect to the distance W_μ .

Furthermore, one can see clearly the structure of a cone on $T_\mu \mathcal{P}_2(M)$ since we have

$$\forall \gamma \in T_\mu \mathcal{P}_2(M), \forall \lambda \in \mathbb{R}^+, \quad \lambda \cdot \gamma \in T_\mu \mathcal{P}_2(M).$$

We end this section with a definition of velocity of curves.

Definition 3.9 (Velocity of curves). *Let $\alpha : [0, a) \rightarrow \mathcal{P}_2(M)$ be a curve in $\mathcal{P}_2(M)$ such that $\alpha_0 = \mu$. We say that $\gamma \in T_\mu \mathcal{P}_2(M)$ is the derivative of α at 0, if there exists a sequence of $\gamma^{(n)} \in \text{Dir}_\mu$ such that $\gamma^{(n)}$ converges to γ , and a sequence of geodesics $\alpha^{(n)}$ such that $\alpha_t^{(n)} = \mathbf{exp}_\mu(t \cdot \gamma^{(n)})$ and*

$$\lim_{n \rightarrow \infty} \left[\limsup_{\varepsilon \rightarrow 0} \frac{d(\alpha_\varepsilon, \alpha_\varepsilon^{(n)})}{\varepsilon} \right] = 0.$$

In this case, γ is also called velocity of α at 0.

Notice that if α is a geodesic, then the definition is trivially verified and $\gamma \in \text{Dir}_\mu$.

3.3. Differentiable functions and DC functions

In $\mathcal{P}_2(M)$, semiconvex and semiconcave functions admit directional derivatives at every point. Furthermore, their directional derivatives are used to construct a notion of a differential at every point. These functions are going to serve us as test functions in the definition of viscosity solutions. Moreover, the squared distance function $(d_W(\cdot, \sigma))^2$ (for some $\sigma \in \mathcal{P}_2(M)$ fixed) is a semiconcave function and an explicit formula will be given for its directional derivatives at every point.

Definition 3.10 (Differential). *Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a given function. We say that $\phi : T_\mu \mathcal{P}_2(M) \rightarrow \mathbb{R}$ is the differential of F at $\mu \in \mathcal{P}_2(M)$ if for any map $\alpha : [0, a) \rightarrow \mathcal{P}_2(M)$ such that $\alpha(0) = \mu$ and α admits a velocity at 0 (in the sense of Definition 3.9), denoted $\gamma \in T_\mu \mathcal{P}_2(M)$, the derivative at 0 of the curve $F \circ \alpha$ is equal to $\phi(\gamma)$. The function ϕ is denoted by $D_\mu F$.*

Real valued Lipschitz functions of $\mathcal{P}_2(M)$ that can be represented as a difference of semiconvex functions behave well in this setting.

Definition 3.11 (Semiconvex functions). *Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a given function. We say that F is semiconvex if there exists $\lambda \in \mathbb{R}$ such that for every geodesic $\alpha : [0, 1] \rightarrow \mathcal{P}_2(M)$ the following inequality holds*

$$F(\alpha_t) \leq (1-t)F(\alpha_0) + tF(\alpha_1) - \frac{\lambda}{2}t(1-t)d_W^2(\alpha_0, \alpha_1).$$

Definition 3.12 (Directional derivative). *Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a Lipschitz and semiconvex function. Let $\mu \in \mathcal{P}_2(M)$. The directional derivative of F at μ along a geodesic α emanating from μ is defined as:*

$$\sigma_\mu F(\alpha) := \lim_{h \downarrow 0} \frac{F(\alpha_h) - F(\alpha_0)}{h}.$$

The limit exists by the monotonicity of incremental ratios of convex functions. The fact that the limit is finite comes from the Lipschitz assumption on F .

Proposition 3.13. *([2, Proposition 5.7.2] Differential of semiconvex functions). Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a Lipschitz and semiconvex function. Let $\mu \in \mathcal{P}_2(M)$. Then F is differentiable at μ .*

Moreover, the differential $D_\mu F : T_\mu \mathcal{P}_2(M) \rightarrow \mathbb{R}$, is such that its restriction to the dense set Dir_μ is equal to

$$D_\mu F \cdot \gamma = \sigma_\mu F(\alpha) \quad \forall \gamma \in \text{Dir}_\mu, \text{ and } \alpha_t = \mathbf{exp}_\mu(t \cdot \gamma) \text{ in a right neighborhood of 0.}$$

Furthermore, the differential is Lipschitz and positively 1-homogeneous, i.e.

$$D_\mu F \cdot (\lambda \cdot \gamma) = \lambda D_\mu F \cdot \gamma, \quad \text{for any } \gamma \in T_\mu \mathcal{P}_2(M) \text{ and } \lambda \geq 0.$$

Definition 3.14 (Semiconcave functions). *Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a function. We say that F is semiconcave if and only if $(-F)$ is semiconvex.*

From Proposition 3.13, if F is Lipschitz and semiconcave, then it is differentiable and the differential is Lipschitz and positively 1-homogeneous.

Definition 3.15 (DC functions). *Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a given function. We say that F is DC if it can be represented as a difference of Lipschitz and semiconvex functions.*

From Proposition 3.13, if F is DC, then it is differentiable and the differential is Lipschitz and positively 1-homogeneous.

Definition 3.16 (Notation). *For $\mu \in \mathcal{P}_2(M)$, we denote by $Lip_1(T_\mu\mathcal{P}_2(M))$ the class of Lipschitz and positively 1-homogeneous functions of $T_\mu\mathcal{P}_2(M)$ and we set*

$$Lip_1(T\mathcal{P}_2(M)) := \bigcup_{\mu \in \mathcal{P}_2(M)} \{\mu\} \times Lip_1(T_\mu\mathcal{P}_2(M)).$$

Next, we define the differential of the squared Wasserstein distance. First, we recall some facts about the Riemannian distance $d^2(\cdot, y)$, for some $y \in M$ fixed. Since M is compact, then the squared distance function is semiconcave, i.e. there exists a constant $C \in \mathbb{R}^+$ such that for any geodesic $\alpha : [0, 1] \rightarrow M$ emanating from $x \in M$, we have

$$d^2(y, \alpha_t) \geq (1-t)d^2(y, \alpha_0) + td^2(y, \alpha_1) - Ct(1-t)d^2(\alpha_0, \alpha_1). \quad (10)$$

From equality (10), one can show the following.

Proposition 3.17. *([24, Proposition 3.1], [19, Proposition 4.1]). Let $\mu \in \mathcal{P}_2(M)$, $\gamma \in \mathcal{P}_2(TM)_\mu$, and a curve $t \mapsto \mathbf{exp}_\mu(t \cdot \gamma)$, not necessarily a geodesic, defined on $[0, 1]$. Then the function $t \mapsto (d_W(\sigma, \mu_t))^2$, for some $\sigma \in \mathcal{P}_2(M)$ fixed, is semiconcave with constant C , i.e.*

$$\left(d_W(\sigma, \mu_t)\right)^2 \geq (1-t)\left(d_W(\sigma, \mu_0)\right)^2 + t\left(d_W(\sigma, \mu_1)\right)^2 - Ct(1-t)\left(d_W(\mu_0, \mu_1)\right)^2.$$

In particular, we get that the function $(d_W(\cdot, \sigma))^2$ is semiconcave in the sense of definition 3.14, if we take $\gamma \in Dir_\mu$ (i.e. $t \mapsto \mathbf{exp}_\mu(t \cdot \gamma)$ is a geodesic). Hence it is differentiable in the sense of Definition 3.10. In fact, we have a more general result summarized in the following theorem:

Theorem 3.18. *([19, Theorem 4.2]). Let $\mu, \sigma \in \mathcal{P}_2(M)$, Let $\gamma \in \mathcal{P}_2(TM)_\mu$, and any curve $t \mapsto \mathbf{exp}_\mu(t \cdot \gamma)$ starting from μ , not necessarily a geodesic. Then it holds*

$$\left.\frac{d}{dt}\right|_{t=0} (d_W(\mathbf{exp}_\mu(t \cdot \gamma), \sigma))^2 = -2 \sup_{\beta} \int \langle v_1, v_2 \rangle d\beta(x, v_1, v_2), \quad (11)$$

where the supremum is taken over all $\beta \in \mathcal{P}_2(T^2M)$, such that

$$\pi^1 \# \beta = \gamma, \quad \text{and} \quad \pi^2 \# \beta \in \mathbf{exp}_\mu^{-1}(\sigma).$$

Notice that if γ is of the form $\gamma = g\sharp\mu$, with g a squared integrable vector field with respect to μ , then equation (11) becomes

$$\frac{d}{dt}\Big|_{t=0} (d_W(\mathbf{exp}_\mu(t \cdot \gamma), \sigma))^2 = -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle g(x), v \rangle d\zeta(x, v). \quad (12)$$

If $\gamma \in T_\mu \mathcal{P}_2(M)$, then by Proposition 3.13, we get

$$D_\mu(d^W(\cdot, \sigma))^2 \cdot \gamma = \frac{d}{dt}\Big|_{t=0} (d_W(\mathbf{exp}_\mu(t \cdot \gamma), \sigma))^2.$$

3.4. Link with the space of gradients

Usually, the ‘space of gradients’ is the one taken as the tangent space. We recall that the space of gradients at a point $\mu \in \mathcal{P}_2(M)$ is

$$SpGr_\mu(\mathcal{P}_2(M)) := \overline{\{\nabla\phi : \phi \in C^\infty(M)\}}^{L_\mu^2},$$

where $C^\infty(M)$ is the space of smooth functions and L_μ^2 is the space of squared integrable vector fields with respect to μ . There is a deep link between the tangent cone and the space of gradients, given by the barycentric projection: first observe that there is a natural embedding $\varsigma_\mu : L_\mu^2 \mapsto \mathcal{P}_2(TM)_\mu$ given by

$$\varsigma_\mu(g) = g\sharp\mu,$$

and it is an isometry. Consequently, a transport plan $\gamma \in \mathcal{P}_2(TM)_\mu$ is induced by a map if $\gamma = \varsigma_\mu(g)$, for some $g \in L_\mu^2$. The right inverse of $\varsigma_\mu(\cdot)$ is the barycentric projection

$$\mathcal{B}(\gamma)(x) = \int v d\gamma_x(v),$$

where $\{\gamma_x\}_{x \in M}$ is the disintegration of γ with respect to the projection π^M (see appendix A). The barycentric projection is characterized by the following equality,

$$\int \langle u(x), v \rangle d\gamma(x, v) = \int \langle u(x), \int v d\gamma_x(v) \rangle d\pi^M \sharp \gamma(x) = \int \langle u, \mathcal{B}(\gamma) \rangle d\mu, \quad \forall u \in L_\mu^2.$$

The barycentric projection links the tangent cone with the space of gradients in the following way.

Proposition 3.19. ([19, Corollary 6.4]). *Let $\mu \in \mathcal{P}_2(M)$, then*

$$\mathcal{B}(T_\mu \mathcal{P}_2(M)) = SpGr_\mu(\mathcal{P}_2(M)),$$

which is equivalent to

$$SpGr_\mu(\mathcal{P}_2(M)) = \{g \in L_\mu^2 : \varsigma_\mu(g) \in T_\mu \mathcal{P}_2(M)\}.$$

Remark 3.20. If $\mu \in \mathcal{P}_2(M)$ is an absolutely continuous measure, then Proposition 3.19 shows that there is a 1 to 1 correspondence between the space of gradients and the tangent cone via the isometry ς_μ [19, Corollary 6.6]. Indeed, in this case, a result due to McCann [23] shows that all geodesics emanating from μ are of the form

$$\mu_t = \exp(t \nabla \phi) \# \mu, \quad \text{with } \phi : M \rightarrow \mathbb{R} \text{ is a semiconvex function.}$$

In particular, the tangent cone is a Hilbert space when μ is absolutely continuous. Another consequence of Proposition 3.19 is that given $g \in L_\mu^2$ and $\pi^\mu : L_\mu^2 \rightarrow SpGr_\mu(\mathcal{P}_2(M))$ is the orthogonal projection onto $SpGr_\mu(\mathcal{P}_2(M))$, the plan $\pi^\mu \circ g \# \mu$ belongs to the tangent cone, i.e.

$$\pi^\mu \circ g \# \mu \in T_\mu \mathcal{P}_2(M),$$

since $\pi^\mu \circ g \in SpGr_\mu(\mathcal{P}_2(M))$.

From Proposition 3.19, we know that if $\zeta \in T_\mu \mathcal{P}_2(M)$, then $\mathcal{B}(\zeta) \in SpGr_\mu(\mathcal{P}_2(M))$. Hence, we can deduce that equation (12) is equal to ($\gamma = g \# \mu$)

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} (d_W(\mathbf{exp}_\mu(t \cdot \gamma), \sigma))^2 &= -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle g(x), \int v d\zeta_x(v) \rangle d\mu(x) \\ &= -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle g(x), \mathcal{B}(\zeta)(x) \rangle d\mu(x) \\ &\stackrel{(3.19)}{=} -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle \pi^\mu \circ g(x), \mathcal{B}(\zeta)(x) \rangle d\mu(x) \\ &= -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle \pi^\mu \circ g(x), v \rangle d\zeta(x, v). \end{aligned}$$

Remark 3.21. The above equality shows that for any $\mu \in \mathcal{P}_2(M)$, the derivative of the squared Wasserstein distance along plans of the form $\gamma = g \# \mu$, with $g \in L_\mu^2$, acts only on their projection onto the space of gradients. Consequently, one can always consider that the derivative along transport plans of the form $\gamma = g \# \mu$, is a derivative in the sense of Proposition 3.13, i.e.

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} (d_W(\mathbf{exp}_\mu(t \cdot \gamma), \sigma))^2 &= \frac{d}{dt} \Big|_{t=0} (d_W(\mathbf{exp}(t g) \# \mu, \sigma))^2 \\ &= -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle \pi^\mu \circ g(x), v \rangle d\zeta(x, v) \\ &= \frac{d}{dt} \Big|_{t=0} (d_W(\mathbf{exp}(t \pi^\mu \circ g) \# \mu, \sigma))^2 \\ &= D_\mu(d^W(\cdot, \sigma))^2 \cdot (\pi^\mu \circ g \# \mu), \end{aligned}$$

since $\pi^\mu \circ g \# \mu \in T_\mu \mathcal{P}_2(M)$ by Proposition 3.19.

4. Time-dependent Hamilton Jacobi Bellman equation in $\mathcal{P}_2(M)$

We have defined all the elements we need to give a precise definition of the Hamiltonian, viscosity solutions and test functions. In this section, we prove that the value function is the unique

viscosity solution to a Hamilton Jacobi Bellman equation. We recall the following notation from Section 3: $Lip_1(T(\mathcal{P}_2(M))) = \bigcup_{\mu \in \mathcal{P}_2(M)} \{\mu\} \times Lip_1(T_\mu(\mathcal{P}_2(M)))$. Consider the following Hamiltonian $H : Lip_1(T(\mathcal{P}_2(M))) \rightarrow \mathbb{R}$ defined by

$$H(\mu, D_\mu v) = \inf_{u \in U} D_\mu v \cdot \left(\pi^\mu \circ f(\cdot, u) \# \mu \right), \quad (13)$$

where $v : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ is a differentiable function in the sense of Definition 3.10. Notice that by Remark 3.21, $\pi^\mu \circ f(\cdot, u) \# \mu \in T_\mu \mathcal{P}_2(M)$, for any $\mu \in \mathcal{P}_2(M)$, and that the quantity

$$D_\mu v \cdot \left(\pi^\mu \circ f(\cdot, u) \# \mu \right)$$

is well defined for any differentiable function v .

Remark 4.1. *The Hamiltonian (13) has the same form as in the classical case. In fact, if $\mu = \delta_x$, is the Dirac measure for some $x \in M$, then we have*

$$L_{\delta_x}^2 = T_x M, \quad (\pi^\mu \circ f(\cdot, u)) \# \delta_x = \delta_{f(x,u)}, \quad H(\delta_x, D_{\delta_x} v) = \inf_{u \in U} D_{\delta_x} v \cdot \left(\delta_{f(x,u)} \right),$$

which is the expression of the Hamiltonian in the absence of uncertainty on the initial condition.

The Hamiltonian used in [13, 20] is similar to ours. In fact, let $v : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be the differentiable function

$$v(\mu) = -(d_W(\mu, \sigma))^2, \quad \forall \mu \in \mathcal{P}_2(M), \text{ and } \sigma \in \mathcal{P}_2(M) \text{ fixed.}$$

Then from Theorem 3.18 and Remark 3.21, we get that for all $\mu \in \mathcal{P}_2(M)$ an absolutely continuous measure, the Hamiltonian (13) is equal to

$$H(\mu, D_\mu v) = \inf_{u \in U} \int \langle \nabla \phi(x), f(x, u) \rangle d\mu(x),$$

with $\phi : M \rightarrow \mathbb{R}$ is a Lipschitz and semiconvex function such that the curve

$$t \mapsto \exp(t \nabla \phi) \# \mu, \quad \forall t \in [0, 1],$$

is the geodesic connecting μ and σ . The above expression of the Hamiltonian is similar to the one used in [13, 20]. This motivates us to use test functions that are in form of v . We will see below the choice of test functions based on this observation. We consider the following Hamilton Jacobi equation:

$$\begin{cases} \partial_t v + H(\mu, D_\mu v) = 0, & (t, \mu) \in [0, T) \times \mathcal{P}_2(M), \\ v(T, \mu) = L(\mu). \end{cases} \quad (14)$$

In order for the partial differentials $\partial_t v$ and $D_\mu v$ to be well defined in equation (14), we need the following lemma.

Lemma 4.2. *[10, Proposition I.5.3] (Product of geodesic spaces). Let (Y, d_Y) and (Z, d_Z) be two generic geodesic spaces. Then the product $(Y \times Z, d_{Y \times Z})$ equipped with the distance*

$$d_{Y \times Z}^2((y_1, z_1), (y_2, z_2)) := d_Y^2(y_1, y_2) + d_Z^2(z_1, z_2),$$

is a geodesic space. Furthermore, a curve $c : [0, 1] \rightarrow Y \times Z$ such that $c_t = (c_t^{(1)}, c_t^{(2)})$ is a geodesic if and only if $c^{(1)}$ and $c^{(2)}$ are geodesics in Y and Z respectively.

Lastly, if the tangent cones in (Y, d_Y) and (Z, d_Z) are defined at every point, then the tangent cone in $(Y \times Z, d_{Y \times Z})$ is defined in a natural way and is equal to the product of tangent cones of (Y, d_Y) and (Z, d_Z) .

We mention that the tangent cone of $[0, T)$ at 0 is isometric to \mathbb{R}^+ and is isometric to \mathbb{R} at the other points. Furthermore, notice that, by means of Proposition 4.2, the tangent cone of the product $[0, T) \times \mathcal{P}_2(M)$ at every point is well defined and is equal to the product of tangent cones of $[0, T)$ and $\mathcal{P}_2(M)$. Obviously, the product space $[0, T) \times \mathcal{P}_2(M)$ is equipped with the distance defined in Lemma 4.2. Hence, $\partial_t v$ and $D_\mu v$ denote the partial differentials of v with respect to the time variable and measure variable respectively. The partial derivatives at some $(s, \mu) \in [0, T) \times \mathcal{P}_2(M)$ are calculated using geodesics of the form $t \rightarrow (c_t^{(1)}, \mu)$ for the time variable and using geodesics of the form $t \rightarrow (s, c_t^{(2)})$ for the measure variable. Moreover, Notice that the partial derivative with respect to time is the usual right derivative in $[0, T)$. Hence, we will take test functions that are continuously differentiable with respect to the time variable and in the class of DC functions with respect to the measure variable, in order to define the notions of viscosity supersolution and viscosity subsolution.

Definition 4.3. (*Test functions*).

Let \mathcal{TEST}_1 be the set defined as:

$$\mathcal{TEST}_1 := \{(t, \mu) \rightarrow a + bt + c((d_W(\mu, \sigma))^2 + |t - s|^2) : a, b \in \mathbb{R}, c \in \mathbb{R}^+ \text{ and } (s, \sigma) \in \mathbb{R}^+ \times \mathcal{P}_2(M)\}.$$

We set $\mathcal{TEST}_2 = -\mathcal{TEST}_1 := \{-\phi : \phi \in \mathcal{TEST}_1\}$.

Definition 4.4. (*Viscosity solutions*).

- We say that a function $v : [0, T) \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ satisfies the inequality

$$\partial_t v + H(\mu, D_\mu v) \geq 0,$$

at $(t, \mu) \in [0, T) \times \mathcal{P}_2(M)$ in the viscosity sense if v is upper semicontinuous and for all \mathcal{TEST}_1 functions $\phi : [0, T) \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ such that $v - \phi$ attains a maximum at (t, μ) , we have

$$\partial_t \phi + H(\mu, D_\mu \phi) \geq 0.$$

A function v satisfying $\partial_t v + H(\mu, D_\mu v) \geq 0$ on $[0, T) \times \mathcal{P}_2(M)$ in the viscosity sense is called a viscosity subsolution of (14).

- Similarly, we say that a function $v : [0, T) \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ satisfies the inequality

$$\partial_t v + H(\mu, D_\mu v) \leq 0,$$

at $(t, \mu) \in [0, T) \times \mathcal{P}_2(M)$ in the viscosity sense if v is lower semicontinuous and for all \mathcal{TEST}_2 functions $\phi : [0, T) \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ such that $v - \phi$ attains a minimum at (t, μ) , then

$$\partial_t \phi + H(\mu, D_\mu \phi) \leq 0.$$

A function v satisfying $\partial_t v + H(\mu, D_\mu v) \leq 0$ on $[0, T) \times \mathcal{P}_2(M)$ in the viscosity sense is called a viscosity supersolution of (14).

- We say that a continuous function $v : [0, T] \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ is a viscosity solution of (14) if it is both a supersolution and a subsolution on $[0, T] \times \mathcal{P}_2(M)$ and verifies

$$v(T, \mu) = L(\mu).$$

In [13, 20], the notion of viscosity solution is defined by introducing sub/super gradients in the Wasserstein space. In here, we define test functions that can be considered as “classical solutions” in Wasserstein space. Furthermore, the comparison result obtained in [13, 20] holds only for uniformly continuous subsolutions and supersolutions. Here, using the approach of test functions, we prove the comparison principle for equation (14) for any bounded upper semicontinuous subsolution and bounded lower semicontinuous supersolution. First, we need two key results.

Proposition 4.5. *For all $\sigma, \mu \in \mathcal{P}_2(M)$ and $c > 0$, we have:*

$$H(\mu, cD_\mu(d_W(\sigma, \cdot))^2) - H(\sigma, -cD_\sigma(d_W(\mu, \cdot))^2) \leq 2c \text{Lip}(f) (d_W(\mu, \sigma))^2.$$

Proof. For any $(x, v) \in TM$, let $\tau_{x, \exp_x(v)}$ be the parallel transport from x to $\exp_x(v)$ along the curve $[0, 1] \ni t \rightarrow \exp_x(tv)$ (see Appendix B). First, since the parallel transport $\tau_{x, \exp_x(v)}$ preserves the Riemannian metric, we have

$$\forall (x, v) \in TM, \quad \langle f(x, u), v \rangle = \langle \tau_{x, \exp_x(v)}(f(x, u)), \tau_{x, \exp_x(v)}(v) \rangle \quad \text{and} \quad |\tau_{x, \exp_x(v)}(v)| = |v|.$$

Furthermore, since $f(\cdot, u)$ is Lipschitz, then by Remark 2.1 we have

$$\forall x \in M, \forall v \in T_x M, \quad |\tau_{x, \exp_x(v)}(f(x, u)) - f(\exp_x(v), u)| \leq \text{Lip}(f) |v|.$$

Thus we get for every $(x, v) \in TM$

$$\begin{aligned} \langle \tau_{x, \exp_x(v)}(f(x, u)), -\tau_{x, \exp_x(v)}(v) \rangle &\leq \langle f(\exp_x(v), u), -\tau_{x, \exp_x(v)}(v) \rangle + \text{Lip}(f) |v| |\tau_{x, \exp_x(v)}(v)| \\ &= \langle f(\exp_x(v), u), -\tau_{x, \exp_x(v)}(v) \rangle + \text{Lip}(f) |v|^2 \end{aligned}$$

Let $\sigma, \mu \in \mathcal{P}_2(M)$, $c > 0$, $\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)$. Then we have

$$\begin{aligned} - \int \langle f(x, u), v \rangle d\zeta(x, v) &= - \int \langle \tau_{x, \exp_x(v)}(f(x, u)), \tau_{x, \exp_x(v)}(v) \rangle d\zeta(x, v) \\ &\leq \int \langle f(\exp_x(v), u), -\tau_{x, \exp_x(v)}(v) \rangle d\zeta(x, v) + \text{Lip}(f) \int |v|^2 d\zeta(x, v) \\ &= \int \langle f(\exp_x(v), u), -\tau_{x, \exp_x(v)}(v) \rangle d\zeta(x, v) + \text{Lip}(f) (d_W(\sigma, \mu))^2, \end{aligned}$$

where the last equality holds since $\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)$. Let $\Gamma : TM \rightarrow TM$ defined for every $(x, v) \in TM$ by

$$\Gamma(x, v) = (\exp_x(v), -\tau_{x, \exp_x(v)}(v)).$$

Then it comes

$$\begin{aligned} - \int \langle f(x, u), v \rangle d\zeta(x, v) &\leq \int \langle f(\exp_x(v), u), -\tau_{x, \exp_x(v)}(v) \rangle d\zeta(x, v) + \text{Lip}(f) (d_W(\sigma, \mu))^2 \\ &= \int \langle f(x, u), v \rangle d\Gamma\# \zeta(x, v) + \text{Lip}(f) (d_W(\sigma, \mu))^2. \end{aligned}$$

Set $\tilde{\zeta} = \Gamma\sharp\zeta$. Notice that we have

$$\pi^M\sharp\tilde{\zeta} = \exp\sharp\zeta = \sigma, \quad \exp\sharp\tilde{\zeta} = \pi^M\sharp\zeta = \mu, \quad \int |v|^2 d\tilde{\zeta}(x, v) = (d_W(\mu, \sigma))^2,$$

since

$$\forall (x, v) \in TM, \quad \pi^M \circ \Gamma(x, v) = \exp_x(v), \quad \exp \circ \Gamma(x, v) = x,$$

and

$$\int |v|^2 d\tilde{\zeta}(x, v) = \int |-\tau_{x, \exp_x(v)}(v)|^2 d\zeta(x, v) = \int |v|^2 d\zeta(x, v) = (d_W(\mu, \sigma))^2.$$

Thus $\tilde{\zeta} \in \mathbf{exp}_\sigma^{-1}(\mu)$, and therefore it follows from Theorem 3.18 and Remark 3.21 that

$$\begin{aligned} D_\mu(d_W(\sigma, \cdot))^2 \cdot (\pi^\mu \circ f(\cdot, u)\sharp\mu) &\leq -2 \int \langle f(x, u), v \rangle d\zeta(x, v) \\ &\leq 2 \int \langle f(x, u), v \rangle d\tilde{\zeta}(x, v) + 2Lip(f)(d_W(\sigma, \mu))^2 \\ &\leq -D_\sigma(d_W(\mu, \cdot))^2 \cdot (\pi^\mu \circ f(\cdot, u)\sharp\sigma) + 2Lip(f)(d_W(\sigma, \mu))^2. \end{aligned}$$

By multiplying by c and taking the infimum over $u \in U$, we get the desired result. \square

Remark 4.6. *The above result is of fundamental importance to prove the comparison principle. Indeed, it allows us to use the variable doubling technique without requiring any additional regularity assumptions on the Hamiltonian. Furthermore, this result does not need any compactness assumptions on the Wasserstein space. The proof can also be adapted if for example the base space is the Euclidean space \mathbb{R}^N , rather than the compact manifold M . The reason is that the squared Wasserstein distance in $\mathcal{P}_2(\mathbb{R}^N)$ is a semiconvex function and its directional derivatives have an expression similar to (11) (see [4, Theorem 7.3.2 and Proposition 7.3.6]).*

Proposition 4.7. *Let \mathcal{O} be a subset of a metric space (X, d_X) , $\Phi : \mathcal{O} \rightarrow \mathbb{R}$ be an upper semicontinuous and $\Psi : \mathcal{O} \rightarrow \mathbb{R}$ be a lower semicontinuous, $\Psi \geq 0$, and*

$$M_a = \sup_{\mathcal{O}} \{ \Phi(x) - a\Psi(x) \},$$

with $a > 0$. Suppose $-\infty < \lim_{a \rightarrow \infty} M_a < +\infty$ and let $x_a \in \mathcal{O}$ be chosen such that

$$\lim_{a \rightarrow \infty} (M_a - (\Phi(x_a) - a\Psi(x_a))) = 0.$$

Then the following holds:

$$\begin{cases} (i) & \lim_{a \rightarrow +\infty} a\Psi(x_a) = 0, \\ (ii) & \Psi(\hat{x}) = 0 \text{ and } \lim_{a \rightarrow +\infty} M_a = \Phi(\hat{x}) = \sup_{\{\Psi(x)=0\}} \Phi(x), \text{ whenever } \hat{x} \in \mathcal{O} \text{ is a limit of the sequence } (x_a)_a. \end{cases}$$

Proof. The proof is exactly the same as in [15, Proposition 3.7], even though it was asserted only for the Euclidean case. We give here below the proof for the sake of completeness. Let

$$\delta_a = M_a - (\Phi(x_a) - a\Psi(x_a)),$$

so that $\lim_{a \rightarrow \infty} \delta_a = 0$. Since $\Psi \geq 0$, M_a decreases as a increases and $\lim_{a \rightarrow +\infty} M_a$ exists and is finite by assumption. Furthermore, we have:

$$M_{\frac{a}{2}} \geq \Phi(x_a) - \frac{a}{2} \Psi(x_a) \geq \Phi(x_a) - a \Psi(x_a) + \frac{a}{2} \Psi(x_a) = M_a - \delta_a + \frac{a}{2} \Psi(x_a),$$

which implies that $a \Psi(x_a) \leq 2(\delta_a + M_{\frac{a}{2}} - M_a)$, hence $\lim_{a \rightarrow +\infty} a \Psi(x_a) = 0$.

Suppose now $a_n \rightarrow +\infty$ and $x_{a_n} \rightarrow \hat{x} \in \mathcal{O}$. Then $\lim_{a_n \rightarrow +\infty} \Psi(x_{a_n}) = 0$ and by lower semicontinuity $\Psi(\hat{x}) = 0$. Moreover, since

$$\Phi(x_{a_n}) - a_n \Psi(x_{a_n}) = M_{a_n} - \delta_{a_n} \geq \sup_{\{\Psi(x)=0\}} \Phi(x) - \delta_{a_n},$$

and Φ is upper semicontinuous, the result holds. \square

Remark 4.8. *Proposition 4.7 is a very general statement. It only requires assumptions on the topology of the considered space. Furthermore, this result holds for non locally compact metric spaces.*

Theorem 4.9 (Comparison principle). *Assume (\mathbf{H}) and (\mathbf{H}_1) . Let $u, v : [0, T] \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be respectively a bounded upper semicontinuous subsolution and a bounded lower semicontinuous supersolution on $[0, T] \times \mathcal{P}_2(M)$. Then it holds:*

$$\sup_{[0, T] \times \mathcal{P}_2(M)} (v - w)_+ \leq \sup_{\{T\} \times \mathcal{P}_2(M)} (v - w)_+,$$

where $(a)_+ = \max(a, 0)$.

Proof. Let $M := \sup_{[0, T] \times \mathcal{P}_2(M)} (v - w)_+$. First, by replacing v by $v - \sup_{\{T\} \times \mathcal{P}_2(M)} (v - w)_+$, which is still a subsolution, it suffices to prove that $M \leq 0$. Suppose by contradiction that $M > 0$. Let $0 < \alpha \leq M$ and let $v^\alpha(t, \mu) = v(t, \mu) + \alpha t$. For α small enough, v^α is still a subsolution of (14). We construct test functions the following way:

$$\psi_a(t, s, \mu, \sigma) = v^\alpha(t, \mu) - w(s, \sigma) - \frac{a}{2}((d_W(\mu, \sigma))^2 + |t - s|^2).$$

Since v, w are bounded, $v - w$ is upper semicontinuous, and $[0, T] \times \mathcal{P}_2(M)$ is compact, then $M_a = \sup \psi_a$ is reached.

Let $(t_a, s_a, \mu_a, \sigma_a)$ be such that M_a is reached. We have

$$\lim_{a \rightarrow +\infty} (M_a - \psi_a(t_a, s_a, \mu_a, \sigma_a)) = 0, \quad \text{and} \quad -\infty < \lim_{a \rightarrow +\infty} M_a < +\infty.$$

Without loss of generality, we can suppose that $(t_a, s_a, \mu_a, \sigma_a)$ converges as $a \rightarrow +\infty$ (take a subsequence if necessary). Therefore, we can apply Proposition 4.7 and we get

$$\begin{cases} (i) & \lim_{a \rightarrow +\infty} \frac{a}{2}((d_W(\mu_a, \sigma_a))^2 + |t_a - s_a|^2) = 0, \\ (ii) & \lim_{a \rightarrow +\infty} M_a = M. \end{cases}$$

Hence, when a is big enough, we have $t_a, s_a \notin \{T\}$. Then we get

$$-\alpha + a(t_a - s_a) + H(\mu_a, \frac{a}{2} D_{\mu_a}(d_W(\cdot, \sigma_a))^2) \geq 0 \geq a(t_a - s_a) + H(\sigma_a, -\frac{a}{2} D_{\sigma_a}(d_W(\cdot, \mu_a))^2).$$

Thus we get from Proposition 4.5

$$0 \leq -\alpha + H(\mu_a, \frac{a}{2}D_{\mu_a}d_W(\cdot, \sigma_a)^2) - H(\sigma_a, -\frac{a}{2}D_{\sigma_a}d_W(\cdot, \mu_a)^2) \leq -\alpha + aLip(f)(d_W(\mu_a, \sigma_a))^2.$$

By letting a tend to infinity, we get $\alpha \leq 0$, a contradiction. \square

Before proving existence of the viscosity solution, we need the following proposition.

Proposition 4.10. *Let $t \rightarrow Y_t^{t_0, x_0, u}$ be a trajectory of (1). Let $\mu, \sigma \in \mathcal{P}_2(M)$. Then, there exists a subsequence, $(t_n)_n \downarrow t_0$ and a vector field $b \in L_\mu^2$, such that*

$$\forall x_0 \in M, \quad b(x_0) \in \overline{\text{co}}\{f(x_0, u) : u \in U\},$$

where $\overline{\text{co}}$ stands for the closed convex hull of the set, and verifies

$$\lim_{t_n \downarrow t_0} \frac{\left(d_W(Y_{t_n}^{t_0, x_0, u} \# \mu, \sigma)\right)^2 - \left(d_W(\mu, \sigma)\right)^2}{t_n - t_0} = \lim_{t_n \downarrow t_0} \frac{\left(d_W(\mathbf{exp}_\mu((t_n - t_0) \cdot b \# \mu), \sigma)\right)^2 - \left(d_W(\mu, \sigma)\right)^2}{t_n - t_0}$$

Proof. First, notice that if such a vector field $b(\cdot) \in L_\mu^2$ exists, then we have

$$\begin{aligned} & \left| \frac{\left(d_W(Y_{t_n}^{t_0, x_0, u} \# \mu, \sigma)\right)^2 - \left(d_W(\mathbf{exp}_\mu((t_n - t_0) \cdot b \# \mu), \sigma)\right)^2}{t_n - t_0} \right| \leq \\ & \frac{d_W(Y_{t_n}^{t_0, \dots, u_0} \# \mu, \mathbf{exp}_\mu((t_n - t_0) \cdot b) \# \mu)}{t_n - t_0} \left(d_W(\mathbf{exp}_\mu((t_n - t_0) \cdot b \# \mu), \sigma) + d_W(Y_{t_n}^{t_0, x_0, u} \# \mu, \sigma) \right). \end{aligned}$$

Hence it suffices to prove that

$$\lim_{t_n \downarrow t_0} \frac{d_W(Y_{t_n}^{t_0, \dots, u_0} \# \mu, \mathbf{exp}_\mu((t_n - t_0) \cdot b \# \mu))}{t_n - t_0} = 0.$$

By Nash embedding theorem, we can assume that M is isometrically embedded into a Euclidean space $(\mathbb{R}^N, \|\cdot\|)$ with $N > 0$ big enough. we have $Y_t^{t_0, x_0, u} = x_0 + \int_{t_0}^t f(Y_s^{t_0, x_0, u}, u(s))ds$, and the quantity

$$\frac{1}{t - t_0} \int_{t_0}^t f(Y_s^{t_0, x_0, u}, u(s))ds$$

is bounded independently of t . Hence there exists a subsequence $(t_n)_n \downarrow t_0$ and a sequence $\varepsilon_n \downarrow 0$ such that

$$b_n(x_0) := \frac{1}{t_n - t_0} \int_{t_0}^{t_n} f(Y_s^{t_0, x_0, u}, u(s))ds \rightarrow b(x_0), \quad \text{and} \quad b_n(x_0) \in \overline{\text{co}}\left(\bigcup_{d(z, x_0) \leq \varepsilon_n} \{f(z, u) : u \in U\}\right),$$

where $\overline{\text{co}}$ stands for the closed convex hull of the set. Hence $b(x_0) \in \overline{\text{co}}\{f(x_0, u) : u \in U\}$ and $b(\cdot)$ is measurable and bounded. Consider the curve $t \mapsto \mathbf{exp}_{x_0}((t - t_0)b(x_0))$. We have

$$\|\mathbf{exp}_{x_0}((t - t_0)b(x_0)) - (x_0 + (t - t_0)b(x_0))\| = o(|t - t_0|),$$

since the two curves are smooth and have the same position and velocity at t_0 . Then, we get

$$\lim_{t_n \downarrow t_0} \frac{1}{t_n - t_0} \left\| Y_{t_n}^{t_0, x_0, u} - \exp_{x_0}((t_n - t_0)b(x_0)) \right\| = \lim_{t_n \downarrow t_0} \left\| \frac{1}{t_n - t_0} \int_{t_0}^{t_n} f(Y_s^{t_0, x_0, u}, u(s)) ds - b(x_0) \right\| = 0.$$

On the other hand, since Nash embedding is biLipschitz, we get

$$\lim_{t_n \downarrow t_0} \frac{1}{t_n - t_0} d\left(Y_{t_n}^{t_0, x_0, u}, \exp_{x_0}((t_n - t_0)b(x_0))\right) = \lim_{t_n \downarrow t_0} \frac{1}{t_n - t_0} \left\| Y_{t_n}^{t_0, x_0, u} - \exp_{x_0}((t_n - t_0)b(x_0)) \right\| = 0.$$

Thus we get

$$\begin{aligned} & \lim_{t_n \downarrow t_0} \frac{1}{(t_n - t_0)^2} \left(d_W\left(Y_{t_n}^{t_0, \cdot, u_0} \# \mu, \exp_{\mu}((t_n - t_0) \cdot b \# \mu)\right) \right)^2 \leq \\ & \lim_{t_n \downarrow t_0} \frac{1}{(t_n - t_0)^2} \int d^2\left(Y_{t_n}^{t_0, x_0, u_0}, \exp_{x_0}((t_n - t_0)b(x_0))\right) d\mu(x_0) = 0, \end{aligned}$$

by dominated convergence, which implies the result. \square

Theorem 4.11. *Assume (\mathbf{H}) , (\mathbf{H}_l) and (\mathbf{H}_{co}) . Then the value function ϑ is the unique continuous viscosity solution to (14).*

Proof. Let $\phi \in \mathcal{TEST}_2$, such that $\vartheta - \phi$ attains a minimum at $(t_0, \mu_0) \in [0, T] \times \mathcal{P}_2(M)$.

So there exists, $a_1, a_2 \in \mathbb{R}$, $c \in \mathbb{R}^-$, $(s, \sigma) \in \mathbb{R}^+ \times \mathcal{P}_2(M)$ such that

$$\phi(t, \mu) = a_1 + a_2 t + c(|t - s|^2 + (d_W(\mu, \sigma))^2),$$

and

$$\forall (t, \mu) \in [0, T] \times \mathcal{P}_2(M), \phi(t, \mu) - \phi(t_0, \mu_0) \leq \vartheta(t, \mu) - \vartheta(t_0, \mu_0).$$

Let $t \mapsto Y_t^{t_0, x, u}$ be a trajectory of (1) such that $\vartheta(t_0, \mu_0) = \vartheta(t_0 + h, Y_{t_0+h}^{t_0, \cdot, u} \# \mu_0)$. So we get for all $h \in [t_0, T - t_0)$,

$$\phi(t_0 + h, Y_{t_0+h}^{t_0, \cdot, u} \# \mu_0) - \phi(t_0, \mu_0) \leq \vartheta(t_0 + h, Y_{t_0+h}^{t_0, \cdot, u} \# \mu_0) - \vartheta(t_0, \mu_0) \leq 0.$$

Thus along a subsequence $(h_n)_n \rightarrow 0$, by dividing by h_n and letting h_n tend to 0, we get by Proposition 4.10 and Theorem 3.18,

$$\partial_t \phi + \inf_{u \in U} D_\mu \phi \cdot (\pi^\mu \circ f(\cdot, u)) \# \mu_0 = \partial_t \phi + \inf_{v(\cdot) \in \overline{co}\{f(\cdot, u)\}} D_\mu \phi \cdot (\pi^\mu \circ v) \# \mu_0 \leq \partial_t \phi + D_\mu \phi \cdot (\pi^\mu \circ b) \# \mu_0 \leq 0,$$

where the first equality is obtained by Hypothesis $(\mathbf{H})_{co}$.

To prove that ϑ is a supersolution, let $\phi \in \mathcal{TEST}_1$, such that $\vartheta - \phi$ attains a maximum at $(t_0, \mu_0) \in [0, T] \times \mathcal{P}_2(M)$. So there exists, $a_1, a_2 \in \mathbb{R}$, $c \in \mathbb{R}^+$, $(s, \sigma) \in \mathbb{R}^+ \times \mathcal{P}_2(M)$ such that

$$\phi(t, \mu) = a_1 + a_2 t + c(|t - s|^2 + (d_W(\mu, \sigma))^2),$$

and

$$\forall (t, \mu) \in [0, T] \times \mathcal{P}_2(M), \phi(t, \mu) - \phi(t_0, \mu_0) \geq \vartheta(t, \mu) - \vartheta(t_0, \mu_0).$$

Let $t \mapsto Y_t^{t_0, x, u}$ be a trajectory that verifies the dynamics (4) with constant control $u \in U$. So we get for all $h \in [t_0, T - t_0)$,

$$\phi(t_0 + h, Y_{t_0+h}^{t_0, \cdot, u} \# \mu_0) - \phi(t_0, \mu_0) \geq \vartheta(t_0 + h, Y_{t_0+h}^{t_0, \cdot, u} \# \mu_0) - \vartheta(t_0, \mu_0) \geq 0.$$

On the other hand, by the same reasoning as in Proposition 4.10, we get

$$\lim_{h \downarrow 0} \frac{\left(d_W(Y_{t_0+h}^{t_0, x_0, u} \# \mu, \sigma)\right)^2 - \left(d_W(\mu, \sigma)\right)^2}{h} = \lim_{h \downarrow 0} \frac{\left(d_W(\mathbf{exp}(h \cdot (f(\cdot, u) \# \mu)), \sigma)\right)^2 - \left(d_W(\mu, \sigma)\right)^2}{h}.$$

Therefore, by dividing by h and letting h tend to 0, we get by Theorem 3.18

$$\partial_t \phi + D_\mu \phi \cdot (\pi^\mu \circ f(\cdot, u)) \# \mu_0 \geq 0.$$

By taking the infimum over $u \in U$, we get the result.

Finally, the final condition of (14) is trivially verified by ϑ . Hence, the value function ϑ is a solution to (14) and it is unique by Theorem 4.9. \square

Appendices

A. Disintegration theorem

We recall here the disintegration theorem. For more details, we refer to [4, theorem 5.3.1].

Theorem A.1. *Let X, Y be two Polish spaces (i.e. complete and separable metric spaces), $\mu \in \mathcal{P}(X)$, let $r : X \rightarrow Y$ be a Borel map and let $\nu = r \# \mu \in \mathcal{P}(Y)$. Then, there exists a ν -a.e. uniquely determined Borel family of probability measures $\{\mu_y\}_{y \in Y} \subset \mathcal{P}(X)$ such that:*

$$\mu_y(X \setminus r^{-1}(y)) = 0, \quad \text{for } \nu\text{-a.e. } y \in Y,$$

and

$$\int_X f(x) d\mu(x) = \int_Y \left(\int_{r^{-1}(y)} f(x) d\mu_y(x) \right) d\nu(y), \quad \text{for every Borel map } f : X \rightarrow [0, +\infty].$$

B. Riemannian manifolds

We recall some standard notions of Riemannian geometry. Some classical references are for example [26, 16]. We consider a connected differentiable manifold M with empty boundary endowed with a Riemannian metric $\langle \cdot, \cdot \rangle$ and we assume that $(M, \langle \cdot, \cdot \rangle)$ is a complete Riemannian manifold. Let $d(\cdot, \cdot)$ be the Riemannian distance on $(M, \langle \cdot, \cdot \rangle)$. The metric space (M, d) is a complete space and its topology is equivalent to the topology of the manifold M . For any $x \in M$, we denote by $T_x M$ the tangent space of M at x , by $TM := \cup_{x \in M} \{x\} \times T_x M$ the tangent bundle and by $\pi^M : TM \rightarrow M$ the canonical projection. Let ∇ be the Levi-Civita connection associated to $(M, \langle \cdot, \cdot \rangle)$. A vector field $V : M \rightarrow TM$ is a mapping such that

$$\pi^M \circ V(x) = x, \quad \forall x \in M.$$

Let $\alpha : [a, b] \rightarrow M$ be a smooth curve. The connection ∇ induces a linear isometry between $T_{\alpha(a)}M$ and $T_{\alpha(t)}M$, for all $t \in [a, b]$. More precisely, for all $v \in T_{\alpha(a)}$, there exists a unique vector field V along α , satisfying

$$\nabla_{\dot{\alpha}(t)}V(\alpha(t)) = 0, \quad \forall t \in [a, b], \text{ and } V(\alpha(a)) = v.$$

The resulting isometry, called the *parallel transport* along α from $\alpha(a)$ to $\alpha(b)$, and denoted by $\tau_{\alpha(a),\alpha(b)}^\alpha$ is defined by

$$\tau_{\alpha(a),\alpha(b)}^\alpha(v) = V(\alpha(b)), \quad \forall v \in T_{\alpha(a)}M.$$

There holds that $\tau_{\alpha(b_1),\alpha(b_2)}^\alpha \circ \tau_{\alpha(a),\alpha(b_1)}^\alpha = \tau_{\alpha(a),\alpha(b_2)}^\alpha$ and $(\tau_{\alpha(a),\alpha(b)}^\alpha)^{-1} = \tau_{\alpha(b),\alpha(a)}^\alpha$. For convenience, we will drop the superscript α , whenever it is clear from the context which curve α is used.

Let $\exp : TM \rightarrow M$ be the exponential map. For every $x \in M$, the function \exp maps straight lines of T_xM , $x \in M$, passing through $0_x \in T_xM$ to geodesics of M passing through x . Since $(M, \langle \cdot, \cdot \rangle)$ is supposed to be complete, it is a consequence of Hopf-Rinow theorem, that the exponential map is defined on all the tangent bundle. However it may not be a diffeomorphism.

The tangent bundle TM is itself a complete Riemannian manifold when endowed with the Sasaki metric [30]. The Riemannian distance d_{TM} on TM associated with the Sasaki metric is defined by

$$\forall (u, v) \in TM \times TM, \quad d_{TM}^2(u, v) := \inf \{ (\text{length}(\alpha))^2 + |\tau_{\pi^M(u),\pi^M(v)}^\alpha(u) - v|^2 \},$$

where the infimum is taken over all smooth curves $\alpha : [0, 1] \rightarrow M$ connecting $\pi^M(u)$ and $\pi^M(v)$ and its length is defined by

$$\text{length}(\alpha) := \int_0^1 \sqrt{\langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle} dt = \int_0^1 |\dot{\alpha}(t)| dt,$$

where $|\cdot|$ is the norm associated to the Riemannian metric $\langle \cdot, \cdot \rangle$ on the tangent bundle TM .

References

- [1] A. A. Agrachev and Y. Sachkov. *Control theory from the geometric viewpoint*, volume 87. Springer Science & Business Media, 2013.
- [2] S. Alexander, V. Kapovitch, and A. Petrunin. *Alexandrov geometry. Book in preparation*, 2017.
- [3] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.
- [4] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.
- [5] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer Science & Business Media, 2008.
- [6] G. Barles. Solutions de viscosité des équations de Hamilton-Jacobi. *Collection SMAI*, 1994.
- [7] J. Z. Ben-Asher. *Optimal control theory with aerospace applications*. American institute of aeronautics and astronautics, 2010.
- [8] P. Bernard. Young measures, superposition and transport. *Indiana University mathematics journal*, pages 247–275, 2008.
- [9] J. E. Bobrow, S. Dubowsky, and J. S. Gibson. Time-optimal control of robotic manipulators along specified paths. *The international journal of robotics research*, 4(3):3–17, 1985.

- [10] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2013.
- [11] D. Burago, I. D. Burago, Y. Burago, S. Ivanov, S. V. Ivanov, and S. A. Ivanov. *A course in metric geometry*, volume 33. American Mathematical Soc., 2001.
- [12] Y. Burago, M. Gromov, and G. Perel'man. A.D. Alexandrov spaces with curvature bounded below. *Russian mathematical surveys*, 47(2):1–58, 1992.
- [13] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review*, 10(01):1–16, 2008.
- [14] F. Clarke. *Functional analysis, calculus of variations and optimal control*, volume 264. Springer Science & Business Media, 2013.
- [15] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67, 1992.
- [16] M. P. Do Carmo. *Riemannian geometry*. Springer Science & Business Media, 2013.
- [17] W. Gangbo, T. Nguyen, A. Tudorascu, et al. Hamilton-Jacobi equations in the Wasserstein space. *Methods and Applications of Analysis*, 15(2):155–184, 2008.
- [18] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 125:119–174, 2019.
- [19] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of $(p, 2)$ -Wasserstein spaces. *Methods and Applications of Analysis*, 18(2):127–158, 2011.
- [20] C. Jimenez, A. Marigonda, and M. Quincampoix. Optimal control of multiagent systems in the Wasserstein space. *Calculus of Variations and Partial Differential Equations*, 59(2):1–45, 2020.
- [21] V. Jurdjevic, J. Velimir, and V. Durdević. *Geometric control theory*. Cambridge university press, 1997.
- [22] A. D. Lewis. Tautological control systems. In *53rd IEEE Conference on Decision and Control*, pages 4832–4837. IEEE, 2014.
- [23] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geometric & Functional Analysis GFA*, 11(3):589–608, 2001.
- [24] S.-I. Ohta. Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *American journal of mathematics*, 131(2):475–516, 2009.
- [25] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. 2001.
- [26] P. Petersen. *Riemannian geometry*, volume 171. Springer, 2006.
- [27] B. Piccoli. Measure differential inclusions. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1323–1328. IEEE, 2018.
- [28] M. Pouryayevali and H. Radmanesh. Trajectories of differential inclusions on Riemannian manifolds and invariance. *Set-Valued and Variational Analysis*, 27(1):241–264, 2019.
- [29] M. Quincampoix and V. M. Veliov. Optimal control of uncertain systems with incomplete information for the disturbances. *SIAM J. Control & Optim.*, 43(4):1373–1399, 2004.
- [30] S. Sasaki. On the differential geometry of tangent bundles of Riemannian manifolds ii. *Tohoku Mathematical Journal, Second Series*, 14(2):146–155, 1962.
- [31] W. P. Thurston, S. Patterson, et al. *Fundamentals of Hyperbolic Manifolds: Selected Expositions*, volume 328. Cambridge University Press, 2006.
- [32] C. Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.
- [33] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [34] R. B. Vinter. *Optimal control*. Springer, 2010.