



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. *SIAM Journal on Mathematical Analysis*, 2023, 56 (3), pp.3326-3356. 10.1137/23M1575251 . hal-03564787v2

HAL Id: hal-03564787

<https://ensta-paris.hal.science/hal-03564787v2>

Submitted on 14 Sep 2022

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 PROBLEM ON RIEMANNIAN MANIFOLDS UNDER PROBABILITY KNOWLEDGE OF THE INITIAL CONDITION*

FRÉDÉRIC JEAN[†], OTHMANE JERHAOUI[†], AND HASNAA ZIDANI[‡]

Abstract. In this article, we study a Mayer optimal control problem on the space of Borel probability measures over a compact Riemannian manifold M . This is motivated by certain situations where a central planner of a deterministic controlled system has only imperfect information on the initial state of the system. The lack of information here is very specific. It is described by a Borel probability measure along which the initial state is distributed. We define a new notion of viscosity in this space by taking test functions that are directionally differentiable and can be written as a difference of two semiconvex functions. With this choice of test functions, we extend the notion of viscosity to Hamilton-Jacobi-Bellman equations in Wasserstein spaces and we establish that the value function is the unique viscosity solution of a Hamilton-Jacobi-Bellman equation in the Wasserstein space over M .

Key words. Optimal Control, Viscosity solutions, Hamilton Jacobi Bellman equation, Wasserstein spaces, Multi-agent systems.

MSC codes. 49J15, 49L25, 49K27

1 Introduction. The study of optimal control problems and viscosity theory in Wasserstein spaces has gained more and more momentum in the last decade, due to its potential real-world applications in modeling multi-agent systems. The potential real-world applications include crowd dynamics modeling [21], opinion formation process modeling [10], herd analysis [41], autonomous multi-vehicle navigation [37] and modeling uncertainties on the initial state of a deterministic controlled system [30, 15]. These problems look into the evolution of a large number of agents, considered to be indistinguishable from one another, subject to local and nonlocal interactions that depend on the density of the distribution of all agents.

A suitable way to model these problems is through a macroscopic approach, where we consider the collection of all agents that belong to a state space denoted X (typically the Euclidean space or a Riemannian manifold), to be a density that evolves through time. If we assume further that the total number of all agents remains constant at all time, then we can normalize the density and assume that the total mass of the system is equal to 1 at all time. Therefore, the evolution of the system, seen as a probability density in the space of Borel probability measures over X and denoted $\mathcal{P}(X)$, is described by a curve $t \mapsto \mu_t \in \mathcal{P}(X)$, where μ_t is the probability density of the system at time $t \geq 0$. The conservation of the mass of the system at all time $t \geq 0$ is described by the *continuity equation*

$$(1.1) \quad \partial_t \mu_t + \operatorname{div}(w_t(\cdot)\mu_t) = 0,$$

where $w_t(\cdot)$ is a time dependent vector field and the equation is understood in the sense of distributions.

In this paper, we propose to study a deterministic controlled system on a compact Riemannian manifold M with imperfect information on the initial condition of the system, i.e. the controller

*Submitted to the editors DATE.

[†]UMA, ENSTA Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France (frederic.jean@ensta-paris.fr, othmane.jerhaoui@ensta-paris.fr).

[‡]Normandie Univ, INSA Rouen Normandie, Laboratoire LMI, 76000 Rouen, France (hasnaa.zidani@insa-rouen.fr).

only knows the initial condition through a Borel probability measure $\mu_0 \in \mathcal{P}(M)$, along which the initial state is distributed. This could be regarded as multi-agent system where the nonlocal interactions between the agents are not considered. More precisely, Let $T > 0$ and consider the following controlled equation

$$(1.2) \quad \begin{cases} \dot{Y}(t) = f(Y(t), u(t)), & t \in [t_0, T], \\ Y(t_0) = x_0, & u(t) \in U, \end{cases}$$

where $f : M \times U \rightarrow TM$ is the dynamics, assumed to be Lipschitz with respect to the first variable and continuous with respect to the second variable, $x_0 \in M$ 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. The control function $u(\cdot) \in U$ is a Borel measurable function $u : [t_0, T] \rightarrow U$. To emphasize the dependence of trajectories of the controlled system (1.2) on the control function $u(\cdot)$, the initial time t_0 and the initial position x_0 , we denote them by

$$t \mapsto Y_t^{t_0, x_0, u}.$$

The main feature of this problem is that the initial state x_0 is not perfectly known, but rather distributed along the probability measure μ_0 . The evolution curves of the initial uncertainty, denoted $t \mapsto \mu_t^{t_0, \mu_0, u}$, are obtained by the pushforward of μ_0 with the flow at time t of the controlled equation (1.2). Therefore, the curves are of the form

$$\begin{cases} \mu_t^{t_0, \mu_0, u} = Y_t^{t_0, \cdot, u} \# \mu_0, & t \in [t_0, T], \quad \text{and} \quad x \mapsto Y_t^{t_0, x, u} \text{ is the flow of (1.2),} \\ \mu_{t_0}^{t_0, \mu_0, u} = \mu_0. \end{cases}$$

Furthermore, notice that since $f(\cdot, u(t))$ is Lipschitz continuous and bounded, then it is a known fact that the evolution trajectory $t \mapsto \mu_t^{t_0, \mu_0, u}$, of the uncertainty μ_0 , is the unique solution to the continuity equation

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

in the distributional sense [4, 11]. The controller aims at minimizing the following final cost:

$$L(\mu) = \int \ell(y) d\mu(y),$$

where $\ell : M \rightarrow \mathbb{R}$ is a Lipschitz function. An immediate consequence of this assumption is that the function $L : \mathcal{P}(M) \rightarrow \mathbb{R}$ inherits the Lipschitz property from ℓ as well. The quantity $L(\mu_T^{t_0, \mu_0, u})$ represents the expectation of the deterministic final cost with respect to the measure $\mu_T^{t_0, \mu_0, u}$. To this optimal control problem, we associate the following value function

$$\vartheta(t_0, \mu_0) = \inf_{u(\cdot) \in U} L(\mu_T^{t_0, \mu_0, u}).$$

In the literature, A similar problem was studied by various authors in the space of Borel probability measures over the Euclidean space \mathbb{R}^N . In particular, it was addressed in [15] a differential

game problem with uncertainties on the initial condition and in [30] a Mayer optimal control problem with uncertainties on the initial condition. We stress on the difference between the set of trajectories $t \mapsto \mu_t$ considered here and the set of trajectories considered in [30]. Indeed, in the latter case, the set of the evolution curves $t \mapsto \mu_t$ of the initial uncertainty appears to be larger. The trajectories $t \mapsto \mu_t$ are solutions to the following continuity equation

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(w_t \mu_t) = 0, & t \in [t_0, T], \\ \mu_{t_0} = \mu_0, \end{cases}$$

where $w_t(\cdot)$ is a vector field such that

$$w_t(\cdot) \in \{f(\cdot, u) : u \in U\},$$

which gives rise to trajectories $t \mapsto \mu_t$ that may not be obtained by the pushforward of the initial uncertainty μ_0 by the flow at time t of the controlled equation (1.2). In this manuscript, we want to study the evolution of the lack of information on the initial condition in (1.2), modeled by a Borel probability measure μ_0 . Hence, we only consider trajectories $t \mapsto \mu_t$ that are obtained by the pushforward of the initial uncertainty μ_0 with the flow at time t of the controlled equation (1.2). Finally, we mention that more general controlled systems on the space of Borel probability measures over the Euclidean space were studied in [26, 12, 7, 36], where the nonlocal interactions between the agents are taken into account.

The first main goal of this paper is to study the properties and the regularity of the value function. In particular, we show that the value function is Lipschitz continuous with respect to both its variables and that it verifies the dynamic programming principle. The second main goal of this paper is to characterize the value function as a unique viscosity solution to a Hamilton-Jacobi-Bellman equation (HJB in short) defined on the Wasserstein space $\mathcal{P}(M)$. Ideally, the HJB equation should have the following form:

$$(1.3) \quad \begin{cases} \partial_t v + H(\mu, D_\mu v) = 0, & (t, \mu) \in [0, T] \times \mathcal{P}(M), \\ v(T, \mu) = L(\mu). \end{cases}$$

The Hamiltonian H and the derivative with respect to the measure variable $D_\mu v$ are to be defined in a suitable way. Furthermore, we want to define a viscosity notion for time-dependent Hamilton Jacobi equations in $\mathcal{P}(M)$ so that we can prove a comparison principle that holds for any bounded upper semicontinuous subsolution and any bounded lower semicontinuous supersolution. To give a precise definition of all these notions in $\mathcal{P}(M)$, we will rely on the pseudo-Riemannian structure of $\mathcal{P}(M)$, presented in Section 3.

Concerning viscosity theory in Wasserstein spaces, several notions have been introduced in the literature to study first order Hamilton Jacobi equations in the Wasserstein space over the Euclidean space $\mathcal{P}_2(\mathbb{R}^N)$. One approach relies on introducing a generalization of sub/super differentials to the space $\mathcal{P}_2(\mathbb{R}^N)$ [30, 26, 15]. Another approach is to define the notion of viscosity, in an extrinsic way, by “lifting” the Hamilton Jacobi equation to a Hilbert space, then use the viscosity theory in Banach spaces, developed in [19, 20]. All these approaches only give a comparison principle that holds only for any uniformly continuous subsolution and any uniformly continuous supersolution. In this paper, we use a different approach. We aim at transposing the viscosity theory techniques that

are used in the classical theory ([17]) to the space of Borel probability measures $\mathcal{P}(M)$. In particular, we define a suitable notion of viscosity using a class of real-valued functions that admit *directional derivatives* at all points $\mu \in \mathcal{P}(M)$. We then prove a comparison principle that holds for any bounded upper semicontinuous subsolution and any bounded lower semicontinuous supersolution. Finally, we prove that the value function is the unique viscosity solution to the above HJB equation by using the dynamic programming principle verified by the value function.

The paper is structured as follows. In Section 2, we formulate the Mayer problem in the space of probability measures and we give the main properties of the value functions. In Section 3, we recall some results of optimal transport theory and the geometry of $\mathcal{P}(M)$. In particular, we give a characterization of the geodesics in $\mathcal{P}(M)$, we describe the pseudo-Riemannian structure of $\mathcal{P}(M)$ and we give a definition of real-valued directionally differentiable functions in $\mathcal{P}(M)$. Section 4 is devoted to the study of a suitable HJB equation that characterizes the value function. In particular, we define the Hamiltonian we are going to work with, then we define a notion of viscosity using the class of functions that are directionally differentiable, we prove a comparison principle that holds for any bounded upper semicontinuous subsolution and any bounded lower semicontinuous supersolution and we prove that the value function is the viscosity of the HJB equation via the dynamic programming principle.

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 [38]. We denote by $d_{TM}(\cdot, \cdot)$ its Riemannian distance on TM associated to 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 Z , any Borel probability measure μ on X and any Borel function $g : X \rightarrow Z$, the pushforward measure $g\#\mu$ on Z is defined by

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

or equivalently,

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

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.

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

Such Borel probability measures γ 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 controlled system, defined for $T > t_0 \geq 0$ and $x_0 \in M$, as

$$(2.1) \quad \begin{cases} \dot{Y}(t) = f(Y(t), u(t)), & \text{for almost every } t \in [t_0, T], \\ Y(t_0) = x_0, u(t) \in U, \end{cases}$$

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

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

(\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 [16, 8] 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 [40, 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 (\mathbf{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 $[t_0, T]$ that is a solution to the controlled system (2.1). 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 (2.1), it holds:*

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

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

$$\int \ell(Y_T^{t_0, x_0, u}) d\mu_0(x_0),$$

over all trajectories that are solutions of the dynamics (2.1) with the initial condition at time t_0 is $x_0 \in M$, distributed along the measure $\mu_0 \in \mathcal{P}_2(M)$. We consider the following assumption:

(H $_\ell$) $\ell : M \rightarrow \mathbb{R}$ is Lipschitz continuous with constant $Lip(\ell)$.

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 (see for example [16, 5]). 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

$$\int \ell(Y_T^{t_0, x_0, u}) d\mu_0(x_0) = \int \ell(y) dY_T^{t_0, \cdot, u} \# \mu_0(y),$$

and we minimize this cost over the set of trajectories $t \mapsto \mu_t^{t_0, \mu_0, u}$ of the space $\mathcal{P}_2(M)$ that verify

$$\begin{cases} \mu_t^{t_0, \mu_0, u} = Y_t^{t_0, \cdot, u} \# \mu_0, & t \in [t_0, T], \text{ and } x \mapsto Y_t^{t_0, x, u} \text{ is the flow of (2.1),} \\ \mu_{t_0}^{t_0, \mu_0, u} = \mu_0. \end{cases}$$

Hence if we set

$$\forall \mu \in \mathcal{P}_2(M), \quad L(\mu) = \int \ell(y) d\mu(y),$$

then the final cost becomes

$$\int \ell(y) dY_T^{t_0, \cdot, u} \# \mu_0(y) = L(\mu_T^{t_0, \mu_0, u}).$$

For any $u(\cdot) \in \mathcal{U}$, the map $x \mapsto f(x, u(t))$ is Lipschitz continuous and bounded independently of t . Hence, the curve $t \mapsto \mu_t^{t_0, \mu_0, u}$ is the unique continuous solution of the continuity equation (see [4, 11])

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

in the sense of distributions, i.e.

$$\begin{cases} \int_{t_0}^T \int_M (\partial_t \phi(t, x) + \langle \nabla_x \phi(t, x), f(x, u(t)) \rangle) d\mu_t^{t_0, \mu_0, u}(x) dt = 0, & \forall \phi \in C_c^\infty([t_0, T] \times M), \\ \mu_{t_0}^{t_0, \mu_0, u} = \mu_0, \end{cases}$$

where $C_c^\infty([t_0, T] \times M)$ is the class of smooth functions of $[t_0, T] \times M$ with compact support. Therefore, the above optimal control problem can be rewritten as

$$(2.2) \quad \begin{cases} \inf_{u(\cdot) \in \mathcal{U}} L(\mu_T^{t_0, \mu_0, u}), \\ \text{such that } \begin{cases} \partial_t \mu_t^{t_0, \mu_0, u} + \operatorname{div}(f(\cdot, u(t)) \mu_t^{t_0, \mu_0, u}) = 0, & t \in [t_0, T], \\ \mu_{t_0}^{t_0, \mu_0, u} = \mu_0, \end{cases} \end{cases}$$

and the infimum is reached, since the set of trajectories of (2.1) is compact in the topology of uniform convergence under Hypotheses **(H)** and **(H_{co})** [5, Theorem 1, pp 60]. The associated *value function* to the above optimal control problem is defined as

$$\vartheta(t_0, \mu_0) := \inf_{u(\cdot) \in \mathcal{U}} L(\mu_T^{t_0, \mu_0, u}) = \inf_{u(\cdot) \in \mathcal{U}} \int l(y) d\mu_T^{t_0, \mu_0, u}(y).$$

Under Hypotheses **(H)**, **(H_ℓ)** and **(H_{co})**, 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 $h \in [t, T - t]$. Assume **(H)**, **(H_ℓ)** and **(H_{co})**. Then it holds*

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

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

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

We have

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

It remains to prove the other inequality. Let $u : [t, T] \rightarrow U$ and $u_{opt} : [t+h, T] \rightarrow U$ be such that

$$\int \ell d\mu_T^{t+h, \mu_{t+h}^{t, \mu, u}, u_{opt}} = \vartheta(t+h, \mu_{t+h}^{t, \mu, u})$$

Let $u^* : [0, T] \rightarrow U$ be the control function defined By

$$u^*(s) = \begin{cases} u(s), & \text{if } s \in [t, t+h], \\ u_{opt}(s), & \text{if } s \in [t+h, T]. \end{cases}$$

Thus we get

$$\begin{aligned} \vartheta(t, \mu) &\leq \int \ell d\mu_T^{t, \mu, u^*} = \int \ell d\mu_T^{t+h, \mu_{t+h}^{t, \mu, u}, u_{opt}} \\ &= \vartheta(t+h, \mu_{t+h}^{t, \mu, u}). \end{aligned}$$

By taking the infimum over $u(\cdot) \in \mathcal{U}$ we get the result. \square

PROPOSITION 2.4. *Assume (\mathbf{H}) , (\mathbf{H}_ℓ) and (\mathbf{H}_{co}) . 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

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

Hence, we have

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

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

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

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

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

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

$$\begin{aligned}
|\vartheta(s, \sigma) - \vartheta(t, \sigma)| &= |\vartheta(s, \sigma) - \vartheta(s, Y_s^{t, \dots, u} \# \sigma)| \\
&\leq Lip(\ell) C_1 d_W(\sigma, Y_s^{t, \dots, u} \# \sigma) \\
&\leq Lip(\ell) C_1 \sqrt{\int d^2(x, Y_s^{t, x, u}) d\sigma(x)} \\
&\leq Lip(\ell) C_1 C_2 |t - s|,
\end{aligned}$$

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 [9]. 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

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

In order to define the Hamiltonian and the notation $D_\mu v$ rigorously, we will analyse the geometric structure of Wasserstein spaces in the next section.

3 Wasserstein space over compact Riemannian Manifolds. The first subsection aims to give some geometric and topological properties of the Wasserstein space and to give a characterization of the geodesics in the Wasserstein space. In the second subsection we describe the pseudo-Riemannian structure that the Wasserstein space enjoys. In particular, we shed some light on where this structure behaves “nicely” and where it degenerates. Finally, in the last subsection we give the definition of directionally differentiable functions in the Wasserstein space. All these tools are going to be necessary to give a precise definition of the *Hamiltonian* and *viscosity notion* for Hamilton Jacobi equations in $\mathcal{P}_2(M)$.

3.1 Geometric and topological properties of Wasserstein space. The Wasserstein space $(\mathcal{P}_2(M), d_W)$ inherits many geometric and topological properties from 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 ([42, Chapter 6]).

Next, we recall the definition of geodesic 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 [42]).

We denote by $\mathcal{P}(TM)$ the set of Borel probability measures over TM . We define the Wasserstein space over (TM, d_{TM}) by

$$(3.1) \quad \mathcal{P}_2(TM) = \left\{ \eta \in \mathcal{P}(TM) : \int d_{TM}^2 \left((x, v), (x_0, v_0) \right) d\eta(x, v) < \infty, \quad \forall (x_0, v_0) \in TM \right\}$$

endowed with its corresponding Wasserstein distance. It is sufficient that the condition

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

in (3.1) to be verified for only one point $(x_0, v_0) \in TM$. Thus if we take $(x_0, 0_{x_0}) \in TM$, then from the definition of d_{TM} (see Appendix B), 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 measures 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 measures $\gamma \in \mathcal{P}(TM)$ 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 measure $\gamma \in \mathcal{P}_2(TM)_\mu$ 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

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

or in other words, the set of measures $\gamma \in \mathcal{P}_2(TM)$ 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

$$\Delta_t(x, v) = (x, tv), \quad \forall t \in \mathbb{R}, (x, v) \in TM \quad \text{and} \quad t \cdot \gamma = \Delta_t \# \gamma, \quad \forall t \in \mathbb{R}, \gamma \in \mathcal{P}_2(TM).$$

LEMMA 3.1. ([25, 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 measure $\gamma \in \mathbf{exp}_\mu^{-1}(\nu)$ such that*

$$(3.2) \quad \mu_t := (\exp \circ \Delta_t) \# \gamma = \mathbf{exp}_\mu(t \cdot \gamma), \quad \forall t \in [0, 1].$$

The measure γ uniquely defines the geodesic $(\mu_t)_{t \in [0,1]}$.

Remark 3.2. The map \mathbf{exp}_μ^{-1} is not really an inverse map to \mathbf{exp}_μ since only the measures $\gamma \in \mathcal{P}_2(TM)_\mu$ such that $(\pi^M, \mathbf{exp})\# \gamma$ are optimal plans of μ and ν are considered in the definition of \mathbf{exp}_μ^{-1} . 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 $t \mapsto \mathbf{exp}_\mu(t \cdot \gamma)$ is a geodesic connecting μ and ν .

From Lemma 3.1, we get the following result about geodesics emanating from any $\mu \in \mathcal{P}_2(M)$.

PROPOSITION 3.3. ([25, 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 measure $\gamma \in \mathcal{P}_2(TM)_\mu$ such that*

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

To summarize, we have seen in this section that the Wasserstein space $(\mathcal{P}_2(M), d_W)$ is a compact geodesic space and each geodesic starting from $\mu \in \mathcal{P}_2(M)$ can be characterized by a measure $\gamma \in \mathcal{P}_2(TM)_\mu$ as shown in Proposition 3.3. We stress on the fact that *not* all the curves of the form

$$t \mapsto \mathbf{exp}_\mu(t \cdot \gamma), \quad t \in [0, \varepsilon], \quad \gamma \in \mathcal{P}_2(TM)_\mu$$

are geodesics but all the geodesics are of this form.

3.2 The space of gradient vector fields and the tangent cone in $\mathcal{P}_2(M)$. The space $(\mathcal{P}_2(M), d_W)$ has a pseudo-Riemannian structure. This has been first pointed out by Otto in [33] and was justified rigorously by Ambrosio-Gigli-Savaré in the case of Wasserstein spaces over the Euclidean space $\mathcal{P}_2(\mathbb{R}^N)$ [4, Chapter 8]. We give hereafter the construction in $\mathcal{P}_2(M)$ following [23]. Let $L^2(\mu, TM)$ be the space of squared integrable vector fields with respect to $\mu \in \mathcal{P}_2(M)$, i.e. vector fields $w : M \rightarrow TM$ such that

$$\|w\|_{L^2(\mu, TM)}^2 := \int_M \langle w(x), w(x) \rangle d\mu(x) < +\infty.$$

The *space of gradient vector fields* at μ is the following Hilbert space:

$$SpGr_\mu(\mathcal{P}_2(M)) := \overline{\{\nabla\phi : \phi \in C_c^\infty(M)\}}^{L^2(\mu, TM)}.$$

Here $C_c^\infty(M)$ is the space of smooth functions of M with compact support. We denote by

$$\pi^\mu : L^2(\mu, TM) \rightarrow SpGr_\mu(\mathcal{P}_2(M))$$

the orthogonal projection map onto the space of gradient vector fields. The space of gradient vector fields is linked to the continuity equation in the following way. Let $I \subset \mathbb{R}$ be an open interval and $(\mu_t)_{t \in I}$ be a Lipschitz curve in $\mathcal{P}_2(M)$. Then, following [23, Proposition 2.5], there exists a family of vector fields $w : I \times M \rightarrow TM$, $(t, x) \mapsto w_t(x) \in T_x M$ for almost all $t \in I$, such that

$$\int_r^s \|w_t\|_{L^2(\mu_t, TM)}^2 dt < \infty, \quad \forall r < s \in I$$

and the continuity equation

$$(3.3) \quad \frac{d}{dt} \mu_t + \operatorname{div}(w_t \mu_t) = 0,$$

is satisfied in the distributional sense, i.e.

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

The family of vector fields w is not unique in general. Indeed, notice that if the family of vector fields w defined above verifies the continuity equation (3.3) in the sense of distributions, then the family of vector fields $(t, x) \mapsto \pi^{\mu_t} \circ w_t(x)$ also verifies the continuity equation (3.3) in the sense of distributions. However, it can be shown [23, Propositions 2.4 and 2.5] that $(t, x) \mapsto \pi^{\mu_t} \circ w_t(x)$ is the unique vector field that verifies (3.3) with minimal $L^2(\mu_t, TM)$ norm for almost all $t \in I$ and

$$\pi^{\mu_t} \circ w_t(\cdot) \in SpGr_{\mu_t}(\mathcal{P}_2(M)), \quad \text{for almost all } t \in I.$$

One can think of $\pi^{\mu_t} \circ w_t(\cdot)$ as the *velocity vector field* at time t for the curve $(\mu_t)_{t \in I}$. The construction of the space of gradient vector fields is analytical and it has the advantage to retain the link between Lipschitz curves of $\mathcal{P}_2(M)$ and the continuity equation (3.3). We will use this point of view to justify the expression of the Hamiltonian associated to the optimal control problem (2.2).

On the other hand, there is another point of view that also justifies rigorously the pseudo-Riemannian structure of $(\mathcal{P}_2(M), d_W)$ which consists in using tools of *metric geometry*. In short, for sufficiently well-behaved metric spaces, one can define a *tangent cone* at every point of the space. The tangent cone is the metric counterpart of the tangent space for Riemannian manifolds. For $(\mathcal{P}_2(M), d_W)$, it has been shown in [25, 32] that the notion of tangent cone is well-defined at every point. We give hereafter the definition of the tangent cone in $(\mathcal{P}_2(M), d_W)$ following [25]. First we define the space of directions at a point. Let $\mu \in \mathcal{P}_2(M)$. The space of directions at μ is the set of “initial velocities” of geodesics emanating from μ :

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

This definition is a direct consequence of Proposition 3.3. In this point of view, The measures γ are seen as the “initial velocities” of the corresponding geodesics starting from μ in analogy with Riemannian geometry. Next, we are going to define the *tangent cone* at μ . Following [25, Section 3], we define the following distance W_μ on Dir_μ . For all $\gamma, \eta \in Dir_\mu$, the quantity

$$W_\mu(\gamma, \eta) := \lim_{t \downarrow 0} \frac{d_W(\mathbf{exp}_\mu(t \cdot \gamma), \mathbf{exp}_\mu(t \cdot \eta))}{t}.$$

exists and defines a distance on Dir_μ [25, Corollary 5.6].

DEFINITION 3.4. (*Tangent cone*). Let $\mu \in \mathcal{P}_2(M)$. The tangent cone $T_\mu \mathcal{P}_2(M)$ is the following set

$$(3.4) \quad T_\mu \mathcal{P}_2(M) := \overline{Dir_\mu}^{W_\mu} = \overline{\left\{ \gamma \in \mathcal{P}_2(TM)_\mu : t \mapsto \mathbf{exp}_\mu(t \cdot \gamma) \text{ is a geodesic defined in some } [0, \varepsilon] \right\}}^{W_\mu}$$

with the closure taken with respect to the distance W_μ . By closure we intend the abstract completion of Dir_μ with respect to W_μ . One can see clearly the structure of a cone on $T_\mu \mathcal{P}_2(M)$ since we have

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

The above definition of W_μ is necessary for the tangent cone to be well-defined for reasons we won't develop here. This construction is not specific to Wasserstein spaces. It is valid for a large class of metric spaces. An important example of these spaces are geodesic spaces with one curvature bound in the sense of Alexandrov [1]. In the case of Wasserstein spaces, the interested reader can check [25, 32] for more details. The important idea to retain here is that the tangent cone is always defined as the completion of the space of directions with respect to this distance. Next, we highlight the connexion between the tangent cone and the space of gradient vector fields following [25, Section 6]. For any $\gamma \in \mathcal{P}_2(TM)_\mu$, We define its *barycentric projection* the following way

$$\mathcal{B} : \mathcal{P}_2(TM)_\mu \rightarrow L^2(\mu, TM) : \mathcal{B}(\gamma)(x) = \int v d\gamma_x(v),$$

where $\{\gamma_x\}_x$ is the disintegration of γ with respect to the projection π^M (see appendix A). The barycentric projection is the right inverse to the pushforward map

$$\forall g(\cdot) \in L^2(\mu, TM), \quad g\#_\mu \in \mathcal{P}_2(TM)_\mu.$$

Furthermore, the barycentric projection is characterized by the following equality

$$(3.5) \quad \begin{aligned} \int \langle w(x), v \rangle d\gamma(x, v) &= \int \langle w(x), \int v d\gamma_x(v) \rangle d\pi^M\#_\mu\gamma(x) \\ &= \int \langle w(x), \mathcal{B}(\gamma)(x) \rangle d\mu(x), \quad \forall w \in L^2(\mu, TM). \end{aligned}$$

Following [25, Corollary 6.4 and Proposition 6.3], the barycentric projection and the pushforward map link the tangent cone with the space of gradient vector fields in the following way:

$$(3.6) \quad \begin{aligned} SpGr_\mu(\mathcal{P}_2(M)) &= \{g(\cdot) \in L^2(\mu, TM) : g\#_\mu \in T_\mu\mathcal{P}_2(M)\}, \\ &= \{\mathcal{B}(\gamma)(x) : \gamma \in T_\mu\mathcal{P}_2(M)\}. \end{aligned}$$

Consequently, the space of gradient vector fields can be seen as a subset of the tangent cone since we trivially get from (3.6)

$$\forall \mu \in \mathcal{P}_2(M), \quad w(\cdot) \in SpGr_\mu(\mathcal{P}_2(M)) \iff w\#_\mu \in T_\mu\mathcal{P}_2(M) \text{ and } w(\cdot) \in L^2(\mu, TM).$$

More generally, given $g(\cdot) \in L^2(\mu, TM)$, the measure $\pi^\mu \circ g\#_\mu$ belongs to the tangent cone, i.e.

$$\pi^\mu \circ g\#_\mu \in T_\mu\mathcal{P}_2(M), \quad \text{since } \pi^\mu \circ g(\cdot) \in SpGr_\mu(\mathcal{P}_2(M)).$$

A natural question that raises itself is when the tangent cone and the space of gradient vector fields are equal. This question was answered by Gigli in [25]. It was shown that the two sets are equal at some $\mu \in \mathcal{P}_2(M)$ if and only if μ is a ‘‘regular measure’’, meaning that it gives zero measure to any hypersurface of M which, locally, is the graph of the difference of two convex functions [25, Corollary 6.6]. A regular measure μ is characterized by the following property: for any $\sigma \in \mathcal{P}_2(M)$, there exists a unique optimal transport plan between μ and σ and it is induced by a map, i.e. there exists a Borel measurable map $T : M \rightarrow M$ such that $(Id, T)\#_\mu$ is the optimal transport plan between μ and σ . This is a refinement of Brenier-McCann's result [31] in which the same property was proven to be true for the case where μ is absolutely continuous with respect to the Riemannian volume form. Intuitively, it means that when μ is a regular measure, the Riemannian structure

on $\mathcal{P}_2(M)$ behaves nicely, since the tangent cone is equal to the space of gradient vector fields, so it is a Hilbert space, in contrast with when μ is not a regular measure where the structure of the tangent cone degenerates. This distinction is important for us because we want to build a robust viscosity notion for first order Hamilton Jacobi equations that will allow us to treat them in all $\mathcal{P}_2(M)$. Therefore, we will use the tangent cone to define directionally differentiable functions since the tangent cone encodes all the information about initial velocities of geodesics starting from μ . We will then use directionally differentiable functions to define the viscosity notion.

3.3 Semiconvex/semiconcave/DC functions In $\mathcal{P}_2(M)$, Real-valued Lipschitz semiconvex or semiconcave functions admit directional derivatives at every point. These functions are going to serve us as test functions in the definition of viscosity notion. Moreover, the squared Wasserstein distance function $d_W^2(\cdot, \sigma)$ (for some $\sigma \in \mathcal{P}_2(M)$ fixed) is Lipschitz and semiconcave. An explicit formula will be given for its directional derivatives at every point.

Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a function and $\mu \in \mathcal{P}_2(M)$. We say that F has a *directional derivative* at μ along a geodesic $\alpha : [0, \varepsilon] \rightarrow \mathcal{P}_2(M)$ emanating from μ , with $\varepsilon > 0$, if the limit

$$\left. \frac{d}{dt} \right|_{t=0} F(\alpha_t) = \lim_{t \downarrow 0} \frac{F(\alpha_t) - F(\alpha_0)}{t}$$

exists and is finite. A particular class of functions that admit directional derivatives are Lipschitz functions that can be represented as a difference of semiconvex functions. We refer to them as Lipschitz and *DC functions*. We define them hereafter.

DEFINITION 3.5. *Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a function.*

- *We say that F is semiconcave if there exists $\lambda \in \mathbb{R}$ such that for every geodesic $\alpha : [0, 1] \rightarrow \mathcal{P}_2(M)$ the following inequality holds*

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

- *Similarly, we say that F is semiconvex if and only if $-F$ is semiconcave.*
- *Finally, we say that F is a DC function if it can be represented as a difference of two semiconvex functions.*

In particular, every semiconvex function is a DC function and every semiconcave function is also a DC function.

Let $\mu \in \mathcal{P}_2(M)$ and $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be a Lipschitz and semiconcave function. The directional derivative of F at μ along a geodesic α emanating from μ

$$\left. \frac{d}{dt} \right|_{t=0} F(\alpha_t) = \lim_{t \downarrow 0} \frac{F(\alpha_t) - F(\alpha_0)}{t}.$$

exists and is finite by [1, Proposition 6.14]. Furthermore, by Proposition 3.3, every geodesic α emanating from μ is of the following form

$$\alpha_t = \mathbf{exp}_\mu(t \cdot \gamma), \quad \text{for some } \gamma \in \text{Dir}_\mu \text{ and } t \in [0, \varepsilon].$$

So we define the *differential function* of F on Dir_μ , denoted $D_\mu F(\mu)$ by

$$\forall \gamma \in \text{Dir}_\mu, \quad D_\mu F(\mu)(\gamma) := \lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t}.$$

Moreover, following [1, Proposition 6.14] the differential function

$$\gamma \mapsto D_\mu F(\mu)(\gamma)$$

is Lipschitz and positively homogeneous in (Dir_μ, W_μ) , with a Lipschitz constant inferior or equal to the Lipschitz constant of F . Consequently, we can extend the *differential function* $D_\mu F(\mu)(\cdot)$ to be defined on the whole $(T_\mu \mathcal{P}_2(M), W_\mu)$ by density.

Similarly, if $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ is Lipschitz and semiconvex, then it is directionally differentiable and its differential function is Lipschitz and positively homogeneous and is defined by

$$D_\mu F(\mu)(\cdot) = -D_\mu(-F)(\mu)(\cdot).$$

Finally, if $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ is a Lipschitz and DC function then it is directionally differentiable and its differential is Lipschitz and positively homogeneous.

For $\mu \in \mathcal{P}_2(M)$, we denote by $\mathcal{C}_\mu(\mathcal{P}_2(M))$ the class of Lipschitz and positively homogeneous functions of $T_\mu \mathcal{P}_2(M)$ and we set

$$\mathcal{C}(\mathcal{P}_2(M)) := \bigcup_{\mu \in \mathcal{P}_2(M)} \{\mu\} \times \mathcal{C}_\mu(\mathcal{P}_2(M)),$$

to be the *metric cotangent bundle* of $\mathcal{P}_2(M)$. Next, we give an explicit expression of directional derivatives of the squared Wasserstein distance. The next result shows that the squared Wasserstein distance is a *semiconcave* function.

PROPOSITION 3.6. ([25, Proposition 4.1]). *Let $\sigma \in \mathcal{P}_2(M)$ be fixed. Then the squared Wasserstein distance*

$$\mathcal{P}_2(M) \ni \nu \mapsto d_W^2(\nu, \sigma)$$

is a Lipschitz and semiconcave function.

In particular, the squared Wasserstein distance function is directionally differentiable. In fact, a much more general result holds: if $F(\cdot) = d^2(\cdot, \sigma)$, then the limit

$$\lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t}$$

exists for all curves of the form

$$(3.7) \quad t \mapsto \mathbf{exp}_\mu(t \cdot \gamma), \quad \text{for some } \gamma \in \mathcal{P}_2(TM)_\mu \text{ and } t \in [0, \varepsilon].$$

even though they are not geodesics. We will only give a weaker version of the expression of the above limit, when

$$\gamma = g\#\mu, \quad g(\cdot) \in L^2(\mu, TM).$$

The general result can be found in [25, Theorem 4.2].

PROPOSITION 3.7. (*Derivative of the squared Wasserstein distance*) *Let $\mu, \sigma \in \mathcal{P}_2(M)$, and $g(\cdot) \in L^2(\mu, TM)$. Let $\gamma = g\#\mu \in \mathcal{P}_2(TM)_\mu$. Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be the function*

$$\forall \nu \in \mathcal{P}_2(M), \quad F(\nu) = d_W^2(\nu, \sigma).$$

Then it holds

$$(3.8) \quad \lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t} = -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle g(x), v \rangle d\zeta(x, v).$$

We stress on the fact that equality (3.8) holds for all curves of the form (3.7) even though they are not geodesics. Next, we show the following result concerning the differential of the squared Wasserstein distance, which is a consequence of Proposition 3.7 and the properties of the barycentric projection.

THEOREM 3.8. *Let $\mu, \sigma \in \mathcal{P}_2(M)$, and $g(\cdot) \in L^2(\mu, TM)$. Let $\gamma = \pi^\mu \circ g\# \mu \in T_\mu \mathcal{P}_2(M)$. Let $F : \mathcal{P}_2(M) \rightarrow \mathbb{R}$ be the function*

$$\forall \nu \in \mathcal{P}_2(M), \quad F(\nu) = d_W^2(\nu, \sigma).$$

Then it holds

$$\begin{aligned} D_\mu F(\mu)(\gamma) &= \lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t} = -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle \pi^\mu \circ g(x), v \rangle d\zeta(x, v) \\ &= -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle g(x), v \rangle d\zeta(x, v). \end{aligned}$$

Proof. First, we show that

$$D_\mu F(\mu)(\gamma) = \lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t}.$$

Since $\gamma \in T_\mu \mathcal{P}_2(M)$ and $D_\mu F(\mu)(\cdot)$ is Lipschitz, then there exists a sequence of measures $(\gamma_n)_n \subset \text{Dir}_\mu$ such that $W_\mu(\gamma_n, \gamma) \rightarrow 0$ and $D_\mu F(\mu)(\gamma_n) \rightarrow D_\mu F(\mu)(\gamma)$ as n tends to infinity. We have

$$\begin{aligned} \left| D_\mu F(\mu)(\gamma) - \lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t} \right| &= \left| \lim_{n \rightarrow \infty} D_\mu F(\mu)(\gamma_n) - \lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t} \right| \\ &= \left| \lim_{n \rightarrow \infty} \lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma_n)) - F(\mathbf{exp}_\mu(t \cdot \gamma))}{t} \right| \\ &\leq \text{Lip}(F) \lim_{n \rightarrow \infty} \lim_{t \downarrow 0} \frac{d_W(\mathbf{exp}_\mu(t \cdot \gamma_n), \mathbf{exp}_\mu(t \cdot \gamma))}{t} \\ &= \text{Lip}(F) \lim_{n \rightarrow \infty} W_\mu(\gamma_n, \gamma) = 0, \end{aligned}$$

where $\text{Lip}(F)$ is the Lipschitz constant of F . This implies the result. Furthermore, Proposition 3.7 gives us

$$\lim_{t \downarrow 0} \frac{F(\mathbf{exp}_\mu(t \cdot \gamma)) - F(\mu)}{t} = -2 \sup_{\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)} \int \langle \pi^\mu \circ g(x), v \rangle d\zeta(x, v).$$

It remains to prove the last equality. First, notice that from Lemma 3.1 we trivially get

$$\mathbf{exp}_\mu^{-1}(\sigma) \subset T_\mu \mathcal{P}_2(M).$$

Furthermore, we know from equality (3.6) that if $\zeta \in T_\mu \mathcal{P}_2(M)$, then $\mathcal{B}(\zeta) \in \text{SpGr}_\mu(\mathcal{P}_2(M))$.

Hence, from (3.5) and (3.6) we deduce that for any $\zeta \in \mathbf{exp}_\mu^{-1}(\sigma)$ we have

$$\begin{aligned} \int \langle g(x), v \rangle d\zeta(x, v) &= \int \langle g(x), \int v d\zeta_x(v) \rangle d\mu(x) \\ &= \int \langle g(x), \mathcal{B}(\zeta)(x) \rangle d\mu(x) \\ &= \int \langle \pi^\mu \circ g(x), \mathcal{B}(\zeta)(x) \rangle d\mu(x) \\ &= \int \langle \pi^\mu \circ g(x), v \rangle d\zeta(x, v), \end{aligned}$$

which implies the last equality. \square

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 and the viscosity notion. In this section, we prove that the value function is the unique viscosity solution to a Hamilton Jacobi Bellman equation. First, we give a justification for the Hamiltonian we are going to work with, based on Otto's point of view of the pseudo-Riemannian structure of $\mathcal{P}_2(M)$. We recall from Section 2 that the value function ϑ is equal to

$$\forall (t_0, \mu_0) \in [0, T] \times \mathcal{P}_2(M), \quad \vartheta(t_0, \mu_0) = \begin{cases} \inf_{u \in \mathcal{U}} L(\mu_T^{t_0, \mu_0, u}) \\ \text{such that } \begin{cases} \partial_t \mu_t^{t_0, \mu_0, u} + \text{div}(f(\cdot, u(t)) \mu_t^{t_0, \mu_0, u}) = 0, \\ \mu_{t_0}^{t_0, \mu_0, u} = \mu_0, \quad t \in [t_0, T], \end{cases} \end{cases}$$

and the continuity equation

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

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

$$\begin{cases} \int_{t_0}^T \int_M (\partial_t \phi(t, x) + \langle \nabla_x \phi(t, x), f(x, u(t)) \rangle) d\mu_t^{t_0, \mu_0, u}(x) dt = 0, & \forall \phi \in C_c^\infty([t_0, T] \times M), \\ \mu_{t_0}^{t_0, \mu_0, u} = \mu_0. \end{cases}$$

Moreover, by the discussion made at the beginning of Section 3.2, every trajectory $t \mapsto \mu_t^{t_0, \mu_0, u}$ is also a solution to the continuity equation

$$\begin{cases} \partial_t \mu_t^{\mu_0, u} + \text{div}(\pi^{\mu_t} \circ f(x, u(t)) \mu_t^{\mu_0, u}) = 0, & t \in [t_0, T], \\ \mu_{t_0}^{\mu_0, u} = \mu_0, \end{cases}$$

in the distributional sense. Hence, the quantity

$$\pi^{\mu_t} \circ f(\cdot, u(t)) \# \mu_t^{\mu_0, u} \in T_{\mu_t^{\mu_0, u}} \mathcal{P}_2(M)$$

can be seen as the *velocity* at time t of the trajectories $t \mapsto \mu_t^{\mu_0, u}$. This heuristic argument motivates us to consider the following Hamiltonian $H : \mathcal{C}(\mathcal{P}_2(M)) \rightarrow \mathbb{R}$ defined by

$$(4.1) \quad \forall (\mu, p_\mu) \in \mathcal{C}(\mathcal{P}_2(M)), \quad H(\mu, p_\mu) = \inf_{u \in \mathcal{U}} p_\mu \left(\pi^\mu \circ f(\cdot, u) \# \mu \right).$$

The definition of the Hamiltonian here resembles the one we usually encounter when Hamilton Jacobi equations are studied on a differentiable manifold. The only difference here is that since $(\mathcal{P}_2(M), d_W)$ is a metric space, the Hamiltonian is defined on the metric cotangent bundle. We consider the following Hamilton Jacobi equation

$$(4.2) \quad \begin{cases} \partial_t v + H(\mu, D_\mu v) = 0, & (t, \mu) \in [0, T] \times \mathcal{P}_2(M), \\ v(T, \mu) = L(\mu) = \int \ell d\mu. \end{cases}$$

We will take test functions that are twice 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 4.1. (*Test functions*).

Let $\mathcal{T}\mathcal{E}\mathcal{S}\mathcal{T}_1$ be the set defined as:

$$\mathcal{T}\mathcal{E}\mathcal{S}\mathcal{T}_1 := \{(t, \mu) \mapsto \psi(t) + a d_W^2(\mu, \sigma) : a \in \mathbb{R}^+, \sigma \in \mathcal{P}_2(M) \text{ and } \psi(\cdot) \in C^2([0, T], \mathbb{R})\}.$$

We set

$$\mathcal{T}\mathcal{E}\mathcal{S}\mathcal{T}_2 = -\mathcal{T}\mathcal{E}\mathcal{S}\mathcal{T}_1 = \{(t, \mu) \mapsto \psi(t) + a d_W^2(\mu, \sigma) : a \in \mathbb{R}^-, \sigma \in \mathcal{P}_2(M) \text{ and } \psi(\cdot) \in C^2([0, T], \mathbb{R})\}.$$

DEFINITION 4.2. (*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 $\mathcal{T}\mathcal{E}\mathcal{S}\mathcal{T}_1$ 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 (4.2).

- 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 $\mathcal{T}\mathcal{E}\mathcal{S}\mathcal{T}_2$ 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 (4.2).

- We say that a continuous function $v : [0, T] \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ is a viscosity solution of (4.2) if it is both a supersolution and a subsolution on $[0, T] \times \mathcal{P}_2(M)$ and verifies

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

Discussion on the notion of viscosity. The notion of viscosity was introduced in [18] to prove well-posedness of Hamilton Jacobi equations in the Euclidean space \mathbb{R}^N , where the test functions used were continuously differentiable functions for both the supersolution and subsolution. In this case, one can identify the differential of the test functions with its gradient and the Hamiltonian is assumed to be a continuous mapping from $\mathbb{R}^N \times \mathbb{R}^N$ to \mathbb{R} . Shortly after, the notion was extended to any Banach space, denoted V , which possesses the Radon-Nikodym property [19, 20]. The test functions used in this setting were Fréchet differentiable functions for both the supersolution and subsolution. The Fréchet differential belongs to the dual space of V , denoted V^* , and the Hamiltonian is assumed to be a continuous mapping from $V \times V^*$ to \mathbb{R} . The notion of viscosity can also be extended to Riemannian manifolds using continuously differentiable functions and the Hamiltonian is assumed to be a continuous mapping from the cotangent bundle to \mathbb{R} [6].

The common features between these definitions are that the state spaces considered in all these examples possess a structure rich enough so that one can assume continuity of the Hamiltonian with respect to both its variables in the topology of the product of the space and its dual space in the case of Banach spaces or the topology of the cotangent bundle in the case of differentiable manifolds. Furthermore, the Fréchet differentiability/continuous differentiability can be defined in these spaces and Fréchet differentiable/continuously differentiable functions exist “in abundance” in order to use them as test functions. In particular, the squared distance function of the state spaces considered is always differentiable (at least locally). This function is particularly important in viscosity theory because it is used to apply the *variable doubling* technique to obtain the comparison results that guarantee uniqueness of the viscosity solution. Furthermore, one can derive existence of the solution from the comparison results using the so-called Perron’s method (see for example [17, 9]).

In $(\mathcal{P}_2(M), d_W)$ this approach seems to be less straightforward. On the one hand, $(\mathcal{P}_2(M), d_W)$ is a metric space that does not have any bundle structure that can be exploited to assume continuity of the Hamiltonian on an interesting topology (indeed, the metric cotangent bundle defined in Section 3.3 can be endowed with the disjoint union distance which is not very useful). On the other hand, the notion of Fréchet differentiability/continuous differentiability is not well-defined in this space, due to the fact that the structure of the tangent cone at $\mu \in \mathcal{P}_2(M)$ degenerates when μ is not a regular measure.

The most known approach to circumvent these difficulties in Wasserstein spaces is through the so-called Lions differentiability [28]. The idea is the following: given a real-valued function of $\mathcal{P}_2(M)$, one considers its “lift” to the space of squared integrable random variables of a probability space equipped with an atomless probability measure (for example, a closed ball of M equipped with the normalized volume form). The lifted function depends on the random variables only through their law in $\mathcal{P}_2(M)$. One then defines Lions differentiable functions in $\mathcal{P}_2(M)$ as the set of functions such that their lift is Fréchet differentiable in the space of squared integrable random variables. This approach was studied in detail in [24] for the space $\mathcal{P}_2(\mathbb{R}^N)$. However, the functions that verify this notion of differentiability are not “abundant” in Wasserstein spaces. For example, the squared Wasserstein distance is not differentiable according to this definition. In fact, it was shown in [2] that the squared Wasserstein distance is differentiable according to this notion at some $\mu \in \mathcal{P}_2(\mathbb{R}^N)$ if and only if μ is a regular measure. This result is not surprising since the pseudo-Riemannian structure degenerates whenever μ is not regular. This presents a major issue for studying Hamilton Jacobi equations in Wasserstein spaces since we can no longer extend viscosity-type techniques (variable doubling, Perron’s method) in this setting.

Two possible approaches can be considered to solve this problem. The first approach would con-

sist in restricting the treatment of Hamilton Jacobi equations to the set of regular measures. The difficulty using this method is that the set of regular measures is not locally compact and not geodesically convex as it was shown in [27]. The second approach would be to relax the Lions differentiability condition and look for test functions that would still be differentiable in a suitable sense and exist “in abundance” in $\mathcal{P}_2(M)$. The latter approach is the one adopted in this manuscript.

The notion of differentiability that is most suitable in $\mathcal{P}_2(M)$ is directional differentiability, presented in Section 3.3. Indeed, all Lipschitz and DC functions are directionally differentiable at every point. The class of DC functions includes the squared distance function and most of the known functionals in $\mathcal{P}_2(M)$ (the internal energy functional, the potential energy functional, the interaction energy functional, the entropy functional...) [42, Chapter 15]. The test functions chosen in Definition 4.1 constitute a subset of the class of DC functions. More precisely, we choose a subset of semiconcave functions to test subsolutions and a subset of semiconvex functions to test supersolutions. However, this approach comes with a major difficulty which is that the Hamiltonian (4.1) is not continuous in this setting. This is the most delicate part to deal with. Luckily for us, for Hamiltonians of type (4.1), and the test functions chosen in Definition 4.1 We have enough information to guarantee well-posedness of the Hamilton Jacobi equation (4.2) without any further assumptions on the regularity of the Hamiltonian.

Finally, we mention that this approach could be adopted in more general metric spaces. More precisely, the notion of directional differentiability presented in Section 3.3 is the same for any geodesic space where the tangent cone is well-defined [1, Chapter 5]. This includes all geodesic spaces with one curvature bound in the sense of Alexandrov. Furthermore, the squared distance function is a DC function in these spaces. In fact, it is semiconvex for spaces with curvature bounded from above and semiconcave for spaces with curvature bounded from below in the sense of Alexandrov [1, 13, 14, 35]. In the particular case of $\mathcal{P}_2(M)$, it is an Alexandrov space with curvature bounded from below by 0 in the sense of Alexandrov if M is a compact connected Riemannian manifold with nonnegative sectional curvature [29, Theorem A.8], [39, Proposition 2.10-(iv)]. More generally, if M is any compact connected Riemannian manifold, then $\mathcal{P}_2(M)$ enjoys a “2-uniform structure” which can be regarded as a generalization of the nonnegatively curved property in the sense of Alexandrov [32].

Next, we prove a comparison principle that holds for any bounded upper semicontinuous subsolution and any bounded lower semicontinuous supersolution. First, we need two key results.

PROPOSITION 4.3. *For all $\sigma, \mu \in \mathcal{P}_2(M)$ and $a > 0$, we have:*

$$H(\mu, a D_\mu(d_W^2(\mu, \sigma))) - H(\sigma, -a D_\sigma(d_W^2(\mu, \sigma))) \leq 2a \text{Lip}(f) d_W^2(\mu, \sigma).$$

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 $(x, v) \in TM$

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

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

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

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

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

Then it comes

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

Set $\tilde{\zeta} = \beta\#\zeta$. Notice that we have

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

since

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

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 3.8 that

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

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

Remark 4.4. The above result is of fundamental importance to prove the comparison principle. Indeed, it will allow us to use the variable doubling technique without assuming any extra-regularity on the Hamiltonian. Furthermore, 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 semiconvave function and its directional derivatives have an expression similar to (3.8) (see [4, Theorem 7.3.2 and Proposition 7.3.6]).

PROPOSITION 4.5. Let \mathcal{O} be a subset of a metric space (X, d_X) , $\Phi : \mathcal{O} \rightarrow \mathbb{R}$ be upper semicontinuous, $\Psi : \mathcal{O} \rightarrow \mathbb{R}$ be lower semicontinuous $\Psi \geq 0$ and

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

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

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

Then the following holds:

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

Proof. The proof is exactly the same as in [17, 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

$$\varepsilon_a = \Gamma_a - (\Phi(x_a) - a \Psi(x_a)),$$

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

$$\Gamma_{\frac{a}{2}} \geq \Phi(x_a) - \frac{a}{2} \Psi(x_a) \geq \Phi(x_a) - a \Psi(x_a) + \frac{a}{2} \Psi(x_a) = \Gamma_a - \varepsilon_a + \frac{a}{2} \Psi(x_a),$$

which implies that $a \Psi(x_a) \leq 2(\varepsilon_a + \Gamma_{\frac{a}{2}} - \Gamma_a)$, 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

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

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

Remark 4.6. Proposition 4.5 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 4.7 (Comparison principle). Assume **(H)** and **(H_ℓ)**. Let $v, w : [0, T] \times \mathcal{P}_2(M) \rightarrow \mathbb{R}$ 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 $(r)_+ = \max(r, 0)$.

Proof. Let $\Gamma := \sup_{[0, T] \times \mathcal{P}_2(M)} (v - w)$. 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 $\Gamma \leq 0$.

By contradiction, suppose that $\Gamma > 0$. Let $0 < \alpha \leq \Gamma$ and let

$$v^\alpha(t, \mu) = v(t, \mu) + \alpha(t - T).$$

v^α is still a subsolution of (4.2). Furthermore, we take α small enough so that

$$\Gamma^\alpha := \sup_{[0,T] \times \mathcal{P}_2(M)} (v^\alpha - w) > 0.$$

We construct test functions the following way:

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

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

Let $(t_a, s_a, \mu_a, \sigma_a)$ be such that Γ_a is reached. Without loss of generality, we can suppose that $(t_a, s_a, \mu_a, \sigma_a)$ converges and $\lim_{a \rightarrow +\infty} \Gamma_a$ exists as $a \rightarrow +\infty$ (take a subsequence if necessary). We have

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

Therefore, we can apply Proposition 4.5 via the correspondences

$$X = \mathcal{O} = [0, T] \times \mathcal{P}_2(M), \quad \Phi(x) = v^\alpha(t, \mu) - w(s, \sigma), \quad \Psi(x) = \frac{1}{2}(d_W^2(\mu, \sigma) + |t - s|^2),$$

and we get

$$\begin{cases} (i) & \lim_{a \rightarrow +\infty} \frac{a}{2}(d_W^2(\mu_a, \sigma_a) + |t_a - s_a|^2) = 0, \\ & \mu_a, \sigma_a \rightarrow \hat{\mu} \in \mathcal{P}_2(M), \quad t_a, s_a \rightarrow \hat{t} \in [0, T], \text{ as } a \rightarrow \infty, \\ (ii) & \lim_{a \rightarrow +\infty} \Gamma_a = \psi_a(\hat{t}, \hat{t}, \hat{\mu}, \hat{\mu}) = \Gamma^\alpha. \end{cases}$$

Hence, when a is big enough, we have $t_a, s_a \notin \{T\}$ since $v^\alpha(\hat{t}, \hat{\mu}) - w(\hat{t}, \hat{\mu}) > 0$. Then we get

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

Thus we get from Proposition 4.3

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

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

A similar comparison result was obtained in [15, 26] for similar Hamilton Jacobi equations defined in the Wasserstein space over the Euclidean space. However, it holds only for uniformly continuous subsolutions and supersolutions. Here, with the new definition of viscosity, the comparison principle holds for equation (4.2) for any bounded upper semicontinuous subsolution and bounded lower semicontinuous supersolution. Before proving existence of the solution for equation (4.2), we need the following proposition.

PROPOSITION 4.8. *Let $t \mapsto Y_t^{t_0, x_0, u}$ be a trajectory of (1.2). Let $\mu, \sigma \in \mathcal{P}_2(M)$. Then, there exists a subsequence, $(t_n)_n \downarrow t_0$ and a vector field $b(\cdot) \in L^2(\mu, TM)$, such that*

$$\text{for } \mu \text{ almost every } x_0 \in M, \quad b(x_0) \in \overline{\text{co}}\{f(x_0, u) : u \in U\},$$

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 $b(\cdot) \in L^2_\mu$ exists, then we have

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

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

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

is uniformly bounded independently of t and x_0 . Let $(t_n)_n \downarrow t_0$, and let $b_n(\cdot)$ be the sequence of functions defined as

$$\forall x_0 \in M, \quad b_n(x_0) := \frac{1}{t_n - t_0} \int_{t_0}^{t_n} f(Y_s^{t_0, x_0, u}, u(s)) ds.$$

The sequence $(b_n(\cdot))_n$ is uniformly bounded. Furthermore, it is equiLipschitz. Indeed we have

$$\begin{aligned} \forall x_0, y_0 \in M, \quad \|b_n(x_0) - b_n(y_0)\| &\leq \frac{1}{t_n - t_0} \int_{t_0}^{t_n} \|f(Y_s^{t_0, x_0, u}, u(s)) - f(Y_s^{t_0, y_0, u}, u(s))\| ds \\ &\leq C_1 \text{Lip}(f) d(x_0, y_0), \end{aligned}$$

where C_1 is the constant from Proposition 2.2. Hence, by Arzelà–Ascoli Theorem, there exists a subsequence of $(t_n)_n$ (not relabeled here) and a function $b(\cdot)$ such that

$$\forall x_0 \in M, \quad b_n(x_0) \rightarrow b(x_0), \quad \text{as } n \text{ tends to infinity.}$$

Moreover, $b(\cdot) \in L^2(\mu, TM)$ since it is the pointwise limit of measurable and uniformly bounded functions. On the other hand, there exists $(\varepsilon_n) \downarrow 0$ such that

$$b_n(x_0) \in \overline{\text{co}} \left(\bigcup_{d(z, x_0) \leq \varepsilon_n} \{f(z, u) : u \in U\} \right),$$

where $\overline{\text{co}}$ stands for the closed convex hull of the set. Hence $b(x_0) \in \overline{\text{co}} \{f(x_0, u) : u \in U\}$. Consider the curve $t \mapsto \exp_{x_0}((t - t_0)b(x_0))$. For any $x_0 \in M$, We have

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

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 obtain

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

by dominated convergence, which implies the result. \square

THEOREM 4.9. *Assume (\mathbf{H}) , (\mathbf{H}_ℓ) and (\mathbf{H}_{co}) . Then the value function ϑ is the unique continuous viscosity solution to (4.2).*

Proof. First we prove that ϑ is a supersolution. Let $\phi \in \mathcal{TES}\mathcal{T}_2$, such that $\vartheta - \phi$ attains a minimum at $(t_0, \mu_0) \in [0, T] \times \mathcal{P}_2(M)$.

So there exists, $(a, \sigma) \in \mathbb{R}^- \times \mathcal{P}_2(M)$ and $\psi(\cdot) \in C^2([0, T], \mathbb{R})$ such that

$$\phi(t, \mu) = \psi(t) + a d_W^2(\mu, \sigma),$$

and

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

Let $t \mapsto Y_t^{t_0, x, u}$ be a trajectory of (1.2) such that $\vartheta(t_0, \mu_0) = \vartheta(t_0 + h, Y_{t_0+h}^{t_0, \dots, u} \# \mu)$. So we get for all $h \in [t_0, T - t_0]$,

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

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 4.8 and Theorem 3.8,

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

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

To prove that ϑ is a subsolution, let $\phi \in \mathcal{TES}\mathcal{T}_1$, such that $\vartheta - \phi$ attains a maximum at $(t_0, \mu_0) \in [0, T] \times \mathcal{P}_2(M)$. So there exists $(a, \sigma) \in \mathbb{R}^+ \times \mathcal{P}_2(M)$ and $\psi(\cdot) \in C^2([0, T], \mathbb{R})$ such that

$$\phi(t, \mu) = \psi(t) + a d_W^2(\mu, \sigma),$$

and

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

Let $t \mapsto Y_t^{t_0, x, u}$ be a trajectory that verifies the controlled system (2.1) with constant control $u \in U$. So we get for all $h \in [t_0, T - t_0]$,

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

On the other hand, by the same reasoning as in Proposition 4.8, we get

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

Therefore, by dividing by h and letting h tend to 0, we get by Theorem 3.8

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

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

Finally, the final condition of (4.2) is trivially verified by ϑ . Hence, the value function ϑ is a continuous bounded solution to (4.2) and it is unique by Theorem 4.7. \square

Conclusion To conclude, we have used Lipschitz and DC functions of $\mathcal{P}_2(M)$ to define a viscosity notion for the Hamilton-Jacobi equation (4.2). With this new framework, we showed that the value function is the unique viscosity solution of (4.2) and we proved a comparison principle that holds for any bounded upper semicontinuous subsolution and any bounded lower semicontinuous supersolution. This suggests that we should favor this class of functions to define the notion of viscosity for more general Hamilton Jacobi equations. In the future, We want to extend these results to the case of Wasserstein spaces defined over noncompact Riemannian manifolds. Furthermore, we want to study more general Hamilton Jacobi equations using the same framework. In particular, we want to study the well posedness of Hamilton Jacobi equation with nonconvex Hamiltonians and Hamiltonians coming from more general multi-agent controlled systems.

Appendix 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

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

Appendix B. Riemannian manifolds. We recall some standard notions of Riemannian geometry. Some classical references are for example [34, 22]. 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 [38]. 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] S. Alexander, V. Kapovitch, and A. Petrunin. Alexandrov geometry: foundations. *arXiv preprint arXiv:1903.08539*, 2022.
- [2] A. Alfonsi and B. Jourdain. Squared quadratic Wasserstein distance: optimal couplings and Lions differentiability. *ESAIM: Probability and Statistics*, 24:703–717, 2020.
- [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] J.-P. Aubin and A. Cellina. *Differential inclusions: set-valued maps and viability theory*, volume 264. Springer Science & Business Media, 2012.
- [6] D. Azagra, J. Ferrera, and B. Sanz. Viscosity solutions to second order partial differential equations on riemannian manifolds. *Journal of Differential Equations*, 245(2):307–336, 2008.
- [7] Z. Badreddine and H. Frankowska. Solutions to Hamilton–Jacobi equation on a Wasserstein space. *Calculus of Variations and Partial Differential Equations*, 61(1):1–41, 2022.
- [8] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer Science & Business Media, 2008.
- [9] G. Barles. Solutions de viscosité des équations de Hamilton-Jacobi. *Collection SMAI*, 1994.
- [10] N. Bellomo, M. Herrero, and A. Tosin. On the dynamics of social conflicts: Looking for the black swan. *Kinetic and Related Models*, 6:459–479, 02 2013.

- [11] P. Bernard. Young measures, superposition and transport. *Indiana University mathematics journal*, pages 247–275, 2008.
- [12] B. Bonnet and H. Frankowska. Differential inclusions in Wasserstein spaces: The Cauchy-Lipschitz framework. *Journal of Differential Equations*, 271:594–637, 2021.
- [13] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2013.
- [14] Y. Burago, M. Gromov, and G. Perel'man. A.D. Alexandrov spaces with curvature bounded below. *Russian mathematical surveys*, 47(2):1–58, 1992.
- [15] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review*, 10(01):1–16, 2008.
- [16] F. Clarke. *Functional analysis, calculus of variations and optimal control*, volume 264. Springer Science & Business Media, 2013.
- [17] 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.
- [18] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983.
- [19] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions I. Uniqueness of viscosity solutions. *Journal of functional analysis*, 62(3):379–396, 1985.
- [20] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions. II. Existence of viscosity solutions. *Journal of Functional Analysis*, 65(3):368–405, 1986.
- [21] E. Cristiani, B. Piccoli, and A. Tosin. *Multiscale modeling of pedestrian dynamics*, volume 12. Springer, 2014.
- [22] M. P. Do Carmo. *Riemannian geometry*. Springer Science & Business Media, 2013.
- [23] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. In *Annales de l'IHP Probabilités et statistiques*, volume 46, pages 1–23, 2010.
- [24] 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.
- [25] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of $(\mathcal{P}_2(M), W_2)$. *Methods and Applications of Analysis*, 18(2):127–158, 2011.
- [26] 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.
- [27] N. Juillet. On displacement interpolation of measures involved in Brenier's theorem. *Proceedings of the American Mathematical Society*, 139(10):3623–3632, 2011.
- [28] P.-L. Lions. Cours au college de france. Available at www.college-de-france.fr, 2007.
- [29] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.
- [30] A. Marigonda and M. Quincampoix. Mayer control problem with probabilistic uncertainty on initial positions. *Journal of Differential Equations*, 264(5):3212–3252, 2018.
- [31] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geometric & Functional Analysis GFA*, 11(3):589–608, 2001.
- [32] S.-I. Ohta. Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *American journal of mathematics*, 131(2):475–516, 2009.
- [33] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. 2001.
- [34] P. Petersen. *Riemannian geometry*, volume 171. Springer, 2006.

- [35] A. Petrunin. Semiconcave functions in Alexandrov's geometry. *arXiv preprint arXiv:1304.0292*, 2013.
- [36] B. Piccoli. Measure differential inclusions. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1323–1328. IEEE, 2018.
- [37] W. Ren and R. W. Beard. *Distributed consensus in multi-vehicle cooperative control*, volume 27. Springer, 2008.
- [38] S. Sasaki. On the differential geometry of tangent bundles of Riemannian manifolds ii. *Tohoku Mathematical Journal, Second Series*, 14(2):146–155, 1962.
- [39] K.-T. Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.
- [40] W. P. Thurston, S. Patterson, et al. *Fundamentals of Hyperbolic Manifolds: Selected Expositions*, volume 328. Cambridge University Press, 2006.
- [41] C. M. Topaz and A. L. Bertozzi. Swarming patterns in a two-dimensional kinematic model for biological groups. *SIAM Journal on Applied Mathematics*, 65(1):152–174, 2004.
- [42] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.