



**HAL**  
open science

# Viscosity solutions of centralized control problems in measure spaces

Othmane Jerhaoui, Averil Prost, Hasnaa Zidani

► **To cite this version:**

Othmane Jerhaoui, Averil Prost, Hasnaa Zidani. Viscosity solutions of centralized control problems in measure spaces. 2023. hal-04335852

**HAL Id: hal-04335852**

**<https://hal.science/hal-04335852v1>**

Preprint submitted on 11 Dec 2023

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# Viscosity solutions of centralized control problems in measure spaces

Othmane Jerhaoui\*, Averil Prost†, Hasnaa Zidani†

## Abstract

This work focuses on a control problem in the Wasserstein space of probability measures over  $\mathbb{R}^d$ . Our aim is to link this control problem to a suitable Hamilton-Jacobi-Bellman (HJB) equation. We explore a notion of viscosity solution using test functions that are locally Lipschitz and locally semiconvex or semiconcave functions. This regularity allows to define a notion of viscosity and a Hamiltonian function relying on directional derivatives. Using a generalization of Ekeland's principle, we show that the corresponding HJB equation admits a comparison principle, and deduce that the value function is the unique solution in this viscosity sense. The PDE tools are developed in the general framework of Measure Differential Equations.

**Keywords:** Hamilton-Jacobi, Wasserstein, viscosity solutions.

**MSC Classification:** 35F21, 35R06, 49Lxx.

## 1 Introduction

In this paper, we consider a Mayer control problem over the metric space of probability measures equipped with the Wasserstein distance. This class of problems is particularly suited for modeling physical situations where the state variable is known only up to a density of probability [Coy21]. It also provides a convenient formalism for problems involving the motion of populations, encompassing both discrete and continuous formulations ([PT09; Cor16]; see also the survey [CCH14]).

These problems have been extensively studied in cases where the state variable lies in some finite-dimensional vector space. An effective approach is to connect the control problem with a Hamilton-Jacobi-Bellman (HJB) partial differential equation. These equations are typically understood in the sense of viscosity solutions, which is a dedicated weak formulation adapted to the nonlinear nature of HJB equations. This approach originated in the work of Crandall, Ishii, and Lions [CIL92; Ish85]. The numerical methods developed for the HJB equation can then be employed to solve the original problem. Our objective with this work is to contribute to the extension of Hamilton-Jacobi techniques into the space of measures.

Considering probability measures as the state space poses several difficulties. We examine an infinite-dimensional subset of probability measures endowed with the Monge-Kantorovich distance, also known as the *Wasserstein distance*, derived from optimal transport theory. The Wasserstein space is not a Banach space, and defining a partial differential equation (PDE) within it is not straightforward. Additional technical challenges arise from the lack of local compactness and convexity of the distance function. Despite these challenges, measures are rich objects that can be viewed as points in a geodesic space, laws of random variables, or generalizations of densities. Each interpretation brings its own set of techniques from geometry, analysis, or probability theory.

Establishing a differential calculus in the Wasserstein space has been an actively researched problem over the past two decades. The foundational work by Otto introduced a pseudo-Riemannian calculus [Ott01; Vil09], offering striking reformulations of the porous medium equation as a gradient flow. Simultaneously, gradient flows in the Wasserstein space were investigated using techniques from general metric spaces [AGS05], leading to the construction of a natural tangent cone [Gig08].

The emergence of mean-field games [Lio06] emphasized the necessity for proper generalizations of gradients for functions dependent on measures, as well as the need for a second-order calculus. Two parallel approaches were pursued: the probabilistic school treated measures as laws of random variables and defined a surrogate for the gradient in the Wasserstein space by *lifting* mappings into Hilbert spaces  $L^2_{\mathbb{P}}(E; \mathbb{R}^d)$ , where  $(E, \mathcal{E}, \mathbb{P})$  represents a "reference" atomless probability

---

\*Université de Rennes, INSA Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France.

†INSA Rouen Normandie, Normandie Univ, LMI UR 3226, F-76000 Rouen, France. Corresponding author: [averil.prost@insa-rouen.fr](mailto:averil.prost@insa-rouen.fr)

space. This theory has achieved the existence and uniqueness of classical solutions for its primary problems, namely the mean-field games system and the master equation [Car13; CCD15; PW18; BY19; CP20].

On the other hand, a geometric approach relied on a "regular" tangent cone. This perspective primarily focused on mechanical systems arising from the minimization of energy over the space of measures, with the tangent cone emerging as a natural space for the dynamical systems of the minimizers. The elements of the tangent cone allow for the definition of sub and superdifferentials, whose intersection reduces to the so-called *Wasserstein gradient* [GNT08; CQ08; AG08; GŚ14; MQ18].

These two approaches were reconciled in [CD18] and [GT19], where it is shown that the gradient obtained through the lifting technique and the Wasserstein gradient coincide under quite general assumptions. Thus, both theories share the same limitations, as pointed out in [AF14] regarding the challenge of connecting the metric interpretation of an Eikonal equation with its interpretation using the regular tangent cone. A suitable notion of a tangent cone should describe the set of available directions in which a particle could move starting from a given point. However, for measures that possess atoms, the regular tangent cone does not enable the splitting of mass. This results in inconsistency when formulating an Eikonal equation such as  $\partial_t u + |\nabla u| = 0$ , where the gradient term encodes the variation of  $u$  along all available directions, which belong to the general tangent cone introduced in [Gig08]. As the theory of Eikonal-type equations enjoys a rich corpus of existence, uniqueness, and representation results [GHN15; GŚ15; HK15], such inconsistencies are undesirable. Let us mention the recent work of [DS23], in which an entropic penalization is used to restrain to the dense set of regular measures, in which regular and general tangent cone coincide.

In this work, we explore a reformulation of Hamilton-Jacobi equations that takes into account the entire set of directions. The formal derivation of Hamilton-Jacobi-Bellman equations shows that one can solely rely on the information provided by directional derivatives of the solution. This has the advantage of completely avoiding the discussion about the existence of a gradient or any other generalization of linearization. In the Wasserstein space, this line of investigation was initiated in [JJZ; Jer22], where the underlying space was taken as a compact manifold. The contributions of the present work are the following: we allow the dynamics to depend on the measure variable, which may have full support in the non-compact space  $\mathbb{R}^d$ . We give a formulation of the Hamilton-Jacobi equation that supports a strong comparison principle, and we show that in the case of Hamilton-Jacobi-Bellman equation, the unique solution is the value function of the control problem. In particular, the comparison principle is valid on Hamiltonians that are defined using directional derivatives along elements of the general tangent cone, and could be used in the context of Eikonal equations.

The paper is organized as follows. The setting of the problem is detailed in Section 2. Section 3 gathers the essential elements of the theory of the Wasserstein space needed in the subsequent sections. The control problem is studied in Section 4, where we introduce the value function and discuss its properties. Section 5 focuses on a general Hamilton-Jacobi equation, providing the definition of a viscosity solution and discussing the comparison principle. Finally, the case of Hamilton-Jacobi-Bellman equations is addressed in Section 6.

## 2 Setting of the problem

In this section, we establish the notations, outline the problem under investigation, and specify the assumptions of the paper.

**Notations. Space of measures** For any Polish space  $(\Omega, d)$ ,  $\mathcal{P}(\Omega)$  will denote the space of Borel probability measures on  $\Omega$ . In the sequel, we consider the subset  $\mathcal{P}_2(\Omega)$  of measures with finite second moment, i.e., measures such that for some (thus any)  $o \in \Omega$ , there holds

$$\mathcal{P}_2(\Omega) := \left\{ \mu \in \mathcal{P}(\Omega) \mid \int_{x \in \mathbb{R}^d} d^2(o, x) d\mu(x) < \infty \right\}.$$

For any couple  $(\mu, \nu) \in \mathcal{P}_2(\Omega)^2$ , we denote by  $\Gamma(\mu, \nu)$  the set of probability measures on  $\Omega \times \Omega$  with marginals  $\mu$  and  $\nu$ , further referred to as the set of *transport plans*. An example of such plans, or couplings, is given by the product measure  $\mu \otimes \nu$ , showing that the set  $\Gamma(\mu, \nu)$  is never empty. We endow  $\mathcal{P}_2(\Omega)$  with a metric, here chosen as the Monge-Kantorovitch distance with  $p = 2$  - also called *2-Wasserstein distance* in the literature - and defined as

$$d_{\mathcal{W}, \Omega}(\mu, \nu) := \sqrt{\inf_{\eta \in \Gamma(\mu, \nu)} \int_{(x, y) \in \Omega^2} d^2(x, y) d\eta(x, y)}.$$

The set of plans realizing the infimum is denoted  $\Gamma_o(\mu, \nu)$ . According to Theorem 4.1 in [Vil09], this set is always nonempty. In the sequel, we will simply denote  $d_{\mathcal{W}}$  the Wasserstein distance when the state space is clear from the context.

Throughout this paper, we consider  $\mathbb{R}^d$  as the underlying space. Its tangent space is defined as  $T\mathbb{R}^d = \bigcup_{x \in \mathbb{R}^d} \{x\} \times T_x \mathbb{R}^d$ , where  $T_x \mathbb{R}^d$  represents the tangent space to  $\mathbb{R}^d$  at the point  $x$ . To maintain a clear distinction between points and velocities, we refrain from identifying  $T_x \mathbb{R}^d$  with  $\mathbb{R}^d$ . Let  $T > 0$  be a fixed final time-horizon. We will use the notation

$$X := ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d), \quad d_X^2((t, \mu), (s, \nu)) := |t - s|^2 + d_{\mathcal{W}}^2(\mu, \nu).$$

**Trajectories in the Wasserstein space** Consider a controlled dynamical system of the form

$$\partial_s \mu_s + \operatorname{div}(f(\cdot, \mu_s, u(s)) \# \mu_s) = 0, \quad s \in [t, T], \quad \mu_t = \nu \quad (1)$$

where  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$  is an initial configuration of the system at a time  $t \in [0, T]$ . We denote  $(\mu_s^{t, \nu, u})_{s \in [t, T]}$  the solution of (1), which is understood in the sense of distributions. Here, the control input  $u(\cdot)$  is supposed to be a measurable function, i.e.

$$u(\cdot) \in L^0([t, T]; U) := \{v(\cdot) : [t, T] \mapsto U \mid v(\cdot) \text{ is Lebesgue-measurable}\},$$

where  $U \subset \mathbb{R}^k$  is a set of admissible controls, and  $f : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times U \mapsto T\mathbb{R}^d$  is a given controlled and measure-dependant dynamic. The study of the dynamical system (1) will be carried out in Section 4.1.

**The control problem and the value function** Now, consider a terminal cost  $\mathfrak{J} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ . The control problem we address in this paper is in Mayer form, and it consists of the following minimization problem:

$$\text{Find } u \in L^0([t, T]; U) \quad \text{such that } \mathfrak{J}(\mu_T^{t, \nu, u}) \leq \mathfrak{J}(\mu_T^{t, \nu, v}) \quad \forall v \in L^0([t, T]; U).$$

The *value function* associated to this control problem is defined as

$$V : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}, \quad V(t, \nu) := \inf \{ \mathfrak{J}(\mu_T^{t, \nu, u}) \mid u \in L^0([t, T]; U) \}. \quad (2)$$

The aim of the paper is to characterize the value function as the unique *viscosity solution* of a suitable Hamilton-Jacobi-Bellman (HJB) equation of the form

$$-\partial_t V(t, \mu) + H(\mu, D_\mu V(t, \mu)) = 0, \quad V(T, \mu) = \mathfrak{J}(\mu).$$

The definition of the Hamiltonian  $H$  and the meaning of the derivative  $D_\mu V(t, \mu)$  will be made precise in Sections 5 & 6. In these sections, we will also develop the notion of viscosity solution in the Wasserstein space.

**Running assumptions** Let us precise the main assumptions of the paper. We say that an application  $m : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is a modulus of continuity if it is continuous, nondecreasing and if  $m(0) = 0$ .

**Assumption [A1]** (Control set). *The set  $U \subset \mathbb{R}^k$  is compact.*

**Assumption [A2]** (Structure of the dynamic). *There exists constants  $[f], |f|_{0, \infty}$  such that*

- *$f$  is Lipschitz-continuous in the space and measure variables and locally Lipschitz in the control variable, in the sense that for all  $(x, y, \mu, \nu, u) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \times U$ ,*

$$|f(x, \mu, u) - f(y, \nu, u)| \leq [f] (|x - y| + d_{\mathcal{W}}(\mu, \nu)).$$

- *there exists a modulus of continuity  $m_{f, u} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that for all  $(x, \mu, u, v) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times U \times U$ ,*

$$|f(x, \mu, u) - f(x, \mu, v)| \leq (1 + |x|) m_{f, u}(|u - v|).$$

- *For all  $(\mu, u) \in \mathcal{P}_2(\mathbb{R}^d) \times U$ , there holds  $|f(0, \mu, u)| \leq |f|_{0, \infty}$ .*

**Assumption [A3]** (Regularity of the terminal cost). *The terminal cost  $\mathfrak{J} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is locally uniformly continuous, i.e. for each  $R > 0$ , there exists a modulus of continuity  $m_{\mathfrak{J}, R} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that*

$$|\mathfrak{J}(\mu) - \mathfrak{J}(\nu)| \leq m_{\mathfrak{J}, R}(d_{\mathcal{W}}(\mu, \nu)) \quad \forall (\mu, \nu) \in \left( \overline{\mathcal{B}}(\delta_0, R) \right)^2,$$

where  $\mathcal{B}(\delta_0, R)$  stands for the ball centered in the Dirac measure  $\delta_0$  and with radius  $R$ .

Under these assumptions, we will aim to demonstrate that the value function is the unique solution to an HJB equation. In the theory of HJ equations in a Banach space, the Hamiltonian is generally defined as a function on gradients, or more generally, on the space of linear functions. Here, we will define the Hamiltonian using a *metric cotangent bundle*  $\mathbb{T}$  as a set of local approximations of sufficiently smooth maps, replacing the set of  $p \mapsto \langle \nabla \phi, p \rangle$  parametrized by any  $\nabla \phi$ . This metric cotangent bundle is a subset of continuous and positively homogeneous applications, that may not enjoy the linear properties of their more traditional counterparts. Indeed, in the context of Hamilton-Jacobi-Bellman equations, this linearity is not essential for achieving results regarding the uniqueness and characterization of the value function. The core interpretation of HJB equations lies in imposing growth conditions along the characteristics of the control problem, and this, in turn, only necessitates one-sided derivatives.

The challenges arising in the Wasserstein space are twofold: firstly, the space lacks local compactness. This issue can be addressed by employing adapted Ekeland principles, as previously demonstrated in Hilbert spaces in [FGS17]. Secondly, the Wasserstein space exhibits positive curvature, which proves to be unfavorable for stability, contrasting with the reasoning applicable to negatively curved spaces. The critical aspect here lies in the fact that, broadly speaking, the directional derivative of a convex map enjoys lower semicontinuity, while concave maps have upper semicontinuous directional derivatives. This makes Hypothesis 3.4 in [Jer22], which assumes some upper/lower semicontinuity of the Hamiltonian, unattainable in  $\mathcal{P}_2(\mathbb{R}^d)$ .

### 3 Preliminaries on the Wasserstein space

Consider  $\mathcal{B}_{\mathbb{R}^d}$  the Borel  $\sigma$ -algebra of  $\mathbb{R}^d$ , and let  $g : \mathbb{R}^d \mapsto \mathbb{R}$  be a Borel-measurable function. The notation  $\#$  will be used to denote the push-forward operator on measures, defined for any Borel measure  $\mu$  as

$$g\#\mu(A) := \mu(g^{-1}(A)) \quad \forall A \in \mathcal{B}_{\mathbb{R}}.$$

The space  $\mathcal{P}_2(\mathbb{R}^d)$ , when endowed with the Wasserstein distance, is a geodesic space. A *constant speed geodesic parameterized over*  $[0, 1]$ , or in short a *geodesic*, is a curve  $(\mu_t)_{t \in [0, 1]} \subset \mathcal{P}_2(\mathbb{R}^d)$  satisfying  $d_{\mathcal{W}}(\mu_t, \mu_s) \leq |t - s| d(\mu_0, \mu_1)$  for all  $(s, t) \in [0, 1]^2$ . We first recall some results on geodesics, and then define directional derivatives along them.

#### 3.1 Representation of geodesics of $\mathcal{P}_2(\mathbb{R}^d)$

Let  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ , and  $\pi_x, \pi_y : (\mathbb{R}^d)^2 \mapsto \mathbb{R}^d$  be the canonical projections  $\pi_x((x, y)) = x$  and  $\pi_y((x, y)) = y$ . In  $\mathcal{P}_2(\mathbb{R}^d)$ , it is known (see [AGS05, Theorem 7.2.2]) that constant speed geodesics coincide with trajectories of the form

$$\mu_t = ((1 - t)\pi_x + t\pi_y)\#\eta \quad \forall t \in [0, 1], \quad \eta = \eta(x, y) \in \Gamma_o(\mu, \nu). \quad (3)$$

The uniqueness of geodesics in the space  $\mathbb{R}^d$  also allows for another equivalent representation, by means of probability measures over the tangent space  $\text{T}\mathbb{R}^d = \bigcup_{x \in \mathbb{R}^d} \{x\} \times \text{T}_x \mathbb{R}^d$ . We will denote  $(x, v) \in \text{T}\mathbb{R}^d$  a generic tangent element, and  $\pi_x, \pi_v$  the canonical projections. Let  $\mathcal{P}_2(\text{T}\mathbb{R}^d)_\mu$  be the set of *initial velocities*

$$\mathcal{P}_2(\text{T}\mathbb{R}^d)_\mu = \left\{ \gamma \in \mathcal{P}_2(\text{T}\mathbb{R}^d) \mid \pi_x\#\gamma = \mu \right\}.$$

Define the scalar multiplication  $\cdot$  of velocities as  $t \cdot \gamma = (\pi_x, t\pi_v)\#\gamma$ . To each element  $\gamma \in \mathcal{P}_2(\text{T}\mathbb{R}^d)_\mu$ , we associate a curve of measures by the *exponential map*

$$\exp_\mu : \mathcal{P}_2(\text{T}\mathbb{R}^d)_\mu \mapsto \mathcal{P}_2(\mathbb{R}^d), \quad \exp_\mu(t \cdot \gamma) := (\pi_x + t\pi_v)\#\gamma.$$

The sets  $\mathcal{P}_2(\text{T}\mathbb{R}^d)_\mu$  and  $\Gamma(\mu, \mathcal{P}_2(\mathbb{R}^d)) = \bigcup_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Gamma(\mu, \nu)$  are in bijection through the map

$$\Psi : \Gamma(\mu, \mathcal{P}_2(\mathbb{R}^d)) \mapsto \mathcal{P}_2(\text{T}\mathbb{R}^d)_\mu, \quad \Psi(\eta) = (\pi_x, \pi_y - \pi_x)\#\eta, \quad \Psi^{-1}(\gamma) = (\pi_x, \pi_x + \pi_v)\#\gamma. \quad (4)$$

The mapping  $\Psi$  is bicontinuous in the respective Wasserstein topologies.

**Remark 1** (Manifold case). *Let us stress that  $\Psi$  is a bijection owing to the uniqueness of geodesics in  $\mathbb{R}^d$ . Indeed, if  $E$  is a manifold over which an exponential map  $\exp : TE \mapsto E$  is defined, there can be several initial velocities of geodesics in the set  $\{(x, v) \in TE \mid \exp_x(v) = y\}$ . The corresponding theory is developed in [Gig09] (see in particular Definition 1.4), and in [JJZ] for the associated geodesic viscosity and HJB equations in the compact case.*

Any geodesic induced by an optimal transport plan  $\eta \in \Gamma_o(\mu, \nu)$  via (3) is equivalently represented using (4) as

$$\exp_\mu(t \cdot \Psi(\eta)) = (\pi_x + t\pi_\nu)\#\Psi(\eta) = (\pi_x + t(\pi_\nu - \pi_x))\#\eta = \mu_t.$$

We define the set of *initial velocities of geodesics* through the identification (4):

$$\mathcal{P}_2(\mathbb{T}\mathbb{R}^d)_{\mu,o} := \left\{ \Psi(\eta) \mid \eta \in \Gamma_o(\mu, \nu), \nu \in \mathcal{P}_2(\mathbb{R}^d) \right\}.$$

Following the notation of [Gig11], we will denote  $\exp_\mu^{-1}(\nu)$  the set of *initial velocities of geodesics* linking  $\mu$  to  $\nu$ , i.e.

$$\exp_\mu^{-1}(\nu) := \left\{ \gamma \in \mathcal{P}_2(\mathbb{T}\mathbb{R}^d)_{\mu,o} \text{ with } (\pi_x + \pi_\nu)\#\gamma = \nu \right\}. \quad (5)$$

We may equivalently define  $\exp_\mu^{-1}$  as

$$\exp_\mu^{-1}(\nu) = \left\{ \gamma \in \mathcal{P}_2(\mathbb{R}^d)_\mu \mid \exp_\mu(\gamma) = \nu \text{ and } \int_{(x,v) \in \mathbb{T}\mathbb{R}^d} |v|^2 d\gamma(x, v) = d_{\mathcal{V}}^2(\mu, \nu) \right\}.$$

This set is always nonempty, since  $\Gamma_o(\mu, \nu)$  is nonempty (see for instance Theorem 1.7 of [San15]). It is compact in  $(\mathcal{P}_2(\mathbb{T}\mathbb{R}^d), d_{\mathcal{V}})$  as the image of the compact  $\Gamma_o(\mu, \nu)$  through the continuous identification (4). We refer to it as the set of initial velocities of geodesics issued from  $\mu$  and reaching  $\nu$ .

### 3.2 The tangent cone

In the sequel, we denote

$$\mathbb{T}^2 \mathbb{R}^d := \left\{ (x, v_1, v_2) \mid x \in \mathbb{R}^d, v_i \in \mathbb{T}_x \mathbb{R}^d \right\}, \quad |(x, v_1, v_2) - (\bar{x}, \bar{v}_1, \bar{v}_2)|^2 := |x - \bar{x}|^2 + \sum_{i=1}^2 |v_i - \bar{v}_i|^2.$$

Given  $(\xi_1, \xi_2) \subset \mathcal{P}_2(\mathbb{R}^d)_\mu$ , we define

$$\Gamma_\mu(\xi_1, \xi_2) := \left\{ \alpha \in \mathcal{P}(\mathbb{T}^2 \mathbb{R}^d) \mid (\pi_x, \pi_{v_i})\#\alpha = \xi_i, i \in \{1, 2\} \right\}.$$

This particular set of transport plans is only allowing transfer of mass between pairs  $(x, v)$  and  $(y, w)$  such that  $x = y$ . For each  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ , define the application

$$W_\mu : \left( \mathcal{P}_2(\mathbb{R}^d)_\mu \right)^2 \rightarrow \mathbb{R}^+, \quad W_\mu^2(\xi, \zeta) := \inf_{\alpha \in \Gamma_\mu(\xi, \zeta)} \int_{(x,v,w) \in \mathbb{T}^2 \mathbb{R}^d} |v - w|^2 d\alpha(x, v, w). \quad (6)$$

As per [Gig08, Theorem 4.4],  $W_\mu$  is a metric over  $\mathcal{P}_2(\mathbb{R}^d)_\mu$  and the infimum is always attained. Moreover, disintegrating  $\xi = \xi_x \otimes \mu$  and  $\zeta = \zeta_x \otimes \mu$  allows to get a representation of  $W_\mu$  as ([Gig08, Proposition 4.2])

$$W_\mu^2(\xi, \zeta) := \int_{x \in \mathbb{R}^d} d_{\mathcal{V}}^2(\xi_x, \zeta_x) d\mu(x). \quad (7)$$

As an useful particular case, we record that

$$W_\mu^2(\xi, 0_\mu) = \int_{(x,v) \in \mathbb{T}^2 \mathbb{R}^d} |v|^2 d\xi(x, v) =: \|\xi\|_\mu^2. \quad (8)$$

We denote  $\Gamma_{\mu,o}(\xi, \zeta)$  the subset of  $\Gamma_\mu(\xi, \zeta)$  where the infimum of (6) is realized.

**Definition 1** (Tangent cone [Gig08, Definition 4.1]). *For each  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ , we define*

$$\mathbf{Tan}'_\mu \mathcal{P}_2(\mathbb{R}^d) := \left\{ \alpha \cdot \xi \mid \alpha \in \mathbb{R}^+, \xi \in \exp_\mu^{-1}(\sigma) \text{ for some } \sigma \in \mathcal{P}_2(\mathbb{R}^d) \right\}, \quad \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) := \overline{\mathbf{Tan}'_\mu \mathcal{P}_2(\mathbb{R}^d)}^{W_\mu}.$$

The pre-tangent cone  $\mathbf{Tan}'_\mu$  is the set of velocities  $\xi$  such that there exists  $\varepsilon > 0$  with  $s \mapsto \exp_\mu(s \cdot \xi)$  being a geodesic between its endpoints over the time interval  $[0, \varepsilon]$ . Owing to [Gig08, Proposition 4.30], there exists a well-defined projection

$$\pi^\mu : \mathcal{P}_2(\mathbb{T}\mathbb{R}^d)_\mu \rightarrow \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d), \quad \pi^\mu(\gamma) = \underset{\bar{\gamma} \in \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)}{\operatorname{argmin}} W_\mu(\gamma, \bar{\gamma}). \quad (9)$$

Note that by picking some  $\alpha \in \Gamma_{\mu,o}(\xi, \zeta)$ , we directly have

$$d_{\mathcal{W}, \mathbb{T}\mathbb{R}^d}^2(\xi, \zeta) \leq \int_{((x,v),(y,w)) \in (\mathbb{T}\mathbb{R}^d)^2} |x-y|^2 + |v-w|^2 d[(\pi_x, \pi_v), (\pi_x, \pi_w)] \# \alpha((x,v), (y,w)) = W_\mu^2(\xi, \zeta). \quad (10)$$

Moreover, each transport plan  $\alpha \in \Gamma(t \cdot \xi, t \cdot \zeta)$  induces a transport plan between  $\exp_\mu(t \cdot \xi)$  and  $\exp_\mu(t \cdot \zeta)$  by  $\beta := (\pi_x + t\pi_v, \pi_x + t\pi_w) \# \alpha$ . Consequently

$$d_{\mathcal{W}}^2(\exp_\mu(t \cdot \xi), \exp_\mu(t \cdot \zeta)) \leq W_\mu^2(t \cdot \xi, t \cdot \zeta) = t^2 W_\mu(\xi, \zeta). \quad (11)$$

Let  $(\mu, \nu) \in (\mathcal{P}_2(\mathbb{R}^d))^2$ . Following [Pic19, Definition 4.1], we define an application  $W_{(\mu,\nu)} : \mathcal{P}(\mathbb{T}\mathbb{R}^d)_\mu \times \mathcal{P}(\mathbb{T}\mathbb{R}^d)_\nu \rightarrow \mathbb{R}^+$  by

$$W_{(\mu,\nu)}^2(\xi, \zeta) := \inf \left\{ \int_{(x,v) \in \mathbb{T}\mathbb{R}^d, (y,w) \in \mathbb{T}\mathbb{R}^d} |v-w|^2 d\omega(x,v,y,w) \mid \omega \in \Gamma(\xi, \zeta), \pi_{(x,y)} \# \omega \in \Gamma_o(\mu, \nu) \right\}. \quad (12)$$

The map  $W_{(\mu,\nu)}$  computes the difference between  $\xi$  and  $\zeta$  by taking only paths whose projection on the base space is a geodesic. It is coherent with the tangent cone structure, since  $W_{(\mu,\mu)}(\xi, \zeta) = W_\mu(\xi, \zeta)$  for all  $\xi, \zeta \in \mathcal{P}(\mathbb{T}\mathbb{R}^d)_\mu$ . However, the application  $W_{(\mu,\nu)}$  does not satisfy the triangular inequality (see [Pic19, Remark 4]).

### 3.3 The metric cotangent bundle

**Definition 2** (Directionally differentiable map). *We say that an application  $\varphi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is directionally differentiable at  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  if for all  $\xi \in \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ , the limit*

$$\lim_{t \searrow 0} \frac{\varphi(\exp_\mu(t \cdot \xi)) - \varphi(\mu)}{t} =: D_\mu \varphi(\xi) \quad (13)$$

*exists. The application  $D_\mu \varphi : \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is called the differential at  $\mu$  of  $\varphi$ .*

Notice that in (13), we do not assume the limit to be uniform in  $\xi \in \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ . It is immediate that  $D_\mu \varphi$  is positively homogeneous, i.e.  $D_\mu \varphi(\alpha \cdot \xi) = \alpha D_\mu \varphi(\xi)$  for any  $\alpha \geq 0$ . Moreover, assume that  $\varphi$  is Lipschitz with constant  $[\varphi]$  in some ball centered in  $\mu$ . Then using (11),

$$\left| D_\mu \varphi(\xi) - D_\mu \varphi(\bar{\xi}) \right| \leq [\varphi] \lim_{t \searrow 0} \frac{d_{\mathcal{W}}(\exp_\mu(t \cdot \xi), \exp_\mu(t \cdot \bar{\xi}))}{t} \leq [\varphi] \lim_{t \searrow 0} \frac{t W_\mu(\xi, \bar{\xi})}{t} = [\varphi] W_\mu(\xi, \bar{\xi}). \quad (14)$$

Hence  $D_\mu \varphi$  is Lipschitz in  $(\mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d), W_\mu)$ . The above leads us to the following definition.

**Definition 3** (Metric cotangent bundle). *Let*

$$\mathbb{T}_\mu := \left\{ p : \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \mid p \text{ is Lipschitz in } W_\mu \text{ and positively homogeneous} \right\}, \quad \mathbb{T} := \bigcup_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \{\mu\} \times \mathbb{T}_\mu. \quad (15)$$

*The sets  $\mathbb{T}_\mu$  are stable by the pointwise operations  $(p+q)(\xi) := p(\xi) + q(\xi)$  and  $(\alpha p)(\xi) := \alpha p(\xi)$ . We endow  $\mathbb{T}$  with the application*

$$\|\cdot\| : \mathbb{T} \rightarrow \mathbb{R}^+, \quad (\mu, p) \mapsto \|p\|_\mu := \sup_{\xi \in \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d), \|\xi\|_\mu = 1} |p(\xi)|. \quad (16)$$

The application  $\|\cdot\|_\mu$  induces a norm on  $\mathbb{T}_\mu$ , and we recover  $|p(\xi)| \leq \|p\|_\mu \|\xi\|_\mu$ . The metric cotangent bundle contains all the infinitesimal approximations of "sufficiently smooth maps", generalizing the set of linear applications. A partial differential equation in the space of measures involves elements of  $\mathbb{T}$ , and we will naturally define the Hamiltonian as a function of  $\mathbb{T}$  into  $\mathbb{R}$ .

## 4 The control problem

### 4.1 Trajectories

The celebrated results of [AGS05] indicate that absolutely continuous curves in the Wasserstein space coincide with the solutions of the continuity equation in the sense of distributions. The recent work of [BF21; BF23] raised the theory of

continuity equations and continuity inclusions in  $\mathcal{P}_2(\mathbb{R}^d)$  to a level comparable to that of the Caratheodory differential inclusions in  $\mathbb{R}^d$ . Let us mention that the study of dynamical systems driven by measure-valued, discontinuous dynamics is drawing attention (see the Measure Differential Equations (MDE) of [CMP18; Pic19]), although it is known that in the Lipschitz setting, solutions of MDEs and continuity equations coincide ([CMP18, Theorem 1]). In this section, we first reformulate the controlled dynamical system (1) in order to apply the results of [BF23], gather some estimates and properties needed in the sequel, and study the properties of the value function.

#### 4.1.1 Existence, representation and regularity

Denote  $\mathcal{P}(U)$  the set of probability measures over the compact  $U$ , which is itself a compact set when endowed with the squared Wasserstein distance. Let  $L^0([t, T]; \mathcal{P}(U))$  be the space of Lebesgue-measurable curves  $\omega : [t, T] \rightarrow \mathcal{P}(U)$ . Define

$$F : \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}(U) \rightarrow \mathcal{C}(\mathbb{R}^d; \mathbb{T}\mathbb{R}^d), \quad (\mu, \omega) \mapsto F_\omega[\mu] := \int_{u \in U} f(\cdot, \mu, u) d\omega(u).$$

Under [A2], routine computations show that for all  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ , all  $\omega, \bar{\omega} \in \mathcal{P}(U)$  and  $x \in \mathbb{R}^d$ , there holds

$$|F_\omega[\mu](x) - F_{\bar{\omega}}[\mu](x)| \leq [f] (1 + |x|) \sqrt{\inf_{\alpha \in \Gamma(\omega, \bar{\omega})} \int_{(u, v) \in U^2} m_{f, u}^2 (|u - v|) d\alpha(u, v)}.$$

As  $m_{f, u}$  is continuous and  $U$  compact, the Wasserstein distance with cost  $m_{f, u}$  is continuous with respect to  $d_{\mathcal{W}, U}$ . Consequently, the application  $\omega \mapsto F_\omega[\mu]$  is continuous from  $(\mathcal{P}(U), d_{\mathcal{W}, U})$  to  $\mathcal{C}(\mathbb{R}^d; \mathbb{T}\mathbb{R}^d)$  endowed with the topology of convergence over compact sets. Therefore,  $[t, T] \ni s \mapsto F_{\omega(s)}[\mu]$  is Lebesgue-measurable for each  $\omega \in L^0([t, T]; \mathcal{P}(U))$ . For some fixed  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ , consider the associated continuity equation

$$\partial_s \mu_s + \operatorname{div} (F_{\omega(s)}[\mu_s] \# \mu_s) = 0 \quad s \in [t, T], \quad \mu_t = \nu. \quad (17)$$

Combining Theorems 2.18, Proposition 2.22 and Theorem 4.2 of [BF23], we get the following.

**THEOREM 1** (EXISTENCE, UNIQUENESS AND REPRESENTATION OF THE SOLUTION). Assume [A1] and [A2]. For each  $\omega \in L^0([t, T]; \mathcal{P}(U))$ , there exists a unique trajectory  $(\mu_s^{t, \nu, \omega})_{s \in [t, T]} \in \operatorname{AC}([t, T]; \mathcal{P}_2(\mathbb{R}^d))$  solution of (17) in the sense of distributions. Moreover  $s \mapsto F_{\omega(s)}[\mu_s^{t, \nu, \omega}]$  is Lebesgue-measurable, there exist constants  $m = m_{f, T}$  and  $M = M_{\nu, f, T}$  such that

$$d_{\mathcal{W}}(\mu_s^{t, \nu, \omega}, \delta_0) \leq m (1 + d_{\mathcal{W}}(\nu, \delta_0)), \quad d_{\mathcal{W}}(\mu_s^{t, \nu, \omega}, \mu_\tau^{t, \nu, \omega}) \leq M |\tau - s| \quad \forall t \leq s, \tau \leq T,$$

and the solution is given by the pushforward  $\mu_s^{t, \nu, \omega} = \Phi_s^t \# \nu$ , where  $\Phi_s^t : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is the well-defined flow of

$$\frac{d}{ds} \Phi_s^t(x) = F_{\omega(s)}[\mu_s^{t, \nu, \omega}](\Phi_s^t(x)), \quad \Phi_t^t(x) = x. \quad (18)$$

Choosing  $\omega(s) = \delta_{u(s)}$  for some  $u \in L^0([t, T]; U)$ , Theorem 1 brings well-posedness of the controlled system (1). The pushforward representation allows to obtain various estimates directly from the underlying dynamical system. In particular, a Grönwall estimate yields that

$$|\Phi_s^{0, x}(x) - x| \leq s \left( [f] |x| + |f|_{0, \infty} \right) e^{[f]s}, \quad \text{and} \quad d_{\mathcal{W}}(\mu_{t+s}^{t, \nu, \omega}, \nu) \leq s \left( [f] d_{\mathcal{W}}(\nu, \delta_0) + |f|_{0, \infty} \right) e^{[f]s} \quad (19)$$

for all  $\omega \in L^0([t, T]; \mathcal{P}(U))$ ,  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $x \in \mathbb{R}^d$  and  $0 \leq t, s, t + s \leq T$  (see Appendix D).

Define the reachable sets from  $(t, \nu)$  at time  $T$  by the flow of (17) as

$$\mathcal{R}_T^{t, \nu} := \{ \mu_T^{t, \nu, \omega} \mid \omega \in L^0([t, T]; \mathcal{P}(U)) \}. \quad (20)$$

**Lemma 1** (Lipschitz-continuity of the reachable sets). Assume [A1] and [A2]. There exists a constant  $[\mathcal{R}]$  depending only on  $f$  and  $T$  such that

$$\max \left( \sup_{\mu \in \mathcal{R}_T^{t, \nu}} \inf_{\bar{\mu} \in \mathcal{R}_T^{t, \bar{\nu}}} d_{\mathcal{W}}(\mu, \bar{\mu}), \sup_{\bar{\mu} \in \mathcal{R}_T^{t, \bar{\nu}}} \inf_{\mu \in \mathcal{R}_T^{t, \nu}} d_{\mathcal{W}}(\mu, \bar{\mu}) \right) \leq [\mathcal{R}] d_{\mathcal{W}}(\nu, \bar{\nu}) \quad \forall t \in [0, T], \nu, \bar{\nu} \in \mathcal{P}_2(\mathbb{R}^d).$$

*Demonstration.* Using chained Grönwall estimates (see Appendix D), we have that two solutions  $s \mapsto \mu_s^{t, \nu, \omega}$  and  $s \mapsto \mu_s^{t, \bar{\nu}, \omega}$  associated to the same control  $\omega \in L^0([t, T]; \mathcal{P}(U))$  satisfy

$$d_{\mathcal{W}} \left( \mu_T^{t, \nu, \omega}, \mu_T^{t, \bar{\nu}, \omega} \right) \leq \exp \left( [f] (T - t) \left( e^{[f](T-t)} + 1 \right) \right) d_{\mathcal{W}}(\nu, \bar{\nu}).$$

The claim follows by approximating each  $\mu = \mu_T^{t, \nu, \omega} \in \mathcal{R}_T^{t, \nu}$  by  $\mu_T^{t, \bar{\nu}, \omega}$ , and defining  $[\mathcal{R}] := \exp \left( [f] T \left( \exp([f] T) + 1 \right) \right)$ .  $\square$



### 4.1.2 Convex relaxation of the dynamic

Let  $\overline{\text{co}}F: \mathcal{P}_2(\mathbb{R}^d) \rightrightarrows \mathcal{C}(\mathbb{R}^d; \mathbb{T}\mathbb{R}^d)$  be given by

$$\overline{\text{co}}F[\mu] := \{F_\omega[\mu] \mid \omega \in \mathcal{P}(U)\} = \left\{ \int_{u \in U} f(\cdot, \mu, u) d\omega(u) \mid \omega \in \mathcal{P}(U) \right\}. \quad (21)$$

For each  $\mu$ , the set  $\overline{\text{co}}F[\mu] \subset \mathcal{C}(\mathbb{R}^d; \mathbb{T}\mathbb{R}^d)$  is closed in the topology of uniform convergence on compact sets; indeed, as this topology is metrizable, it suffices to show that  $(b_n)_n \subset \overline{\text{co}}F[\mu]$  and  $b_n \rightarrow_n b$  uniformly over the compacts implies  $b \in \overline{\text{co}}F[\mu]$ . As  $\mathcal{P}(U)$  is compact ([AGS05, Proposition 7.1.5]), some subsequence of  $\omega_n$  converges to a measure  $\omega \in \mathcal{P}(U)$ , and **[A2]** yields that for each  $R > 0$ ,

$$\sup_{x \in \mathbb{R}^d, |x| \leq R} \left| \int_{u \in U} f(x, \mu, u) d[\omega_n - \omega](u) \right| \leq (1+R) \inf_{\alpha \in \Gamma(\omega_n, \omega)} \int_{(u,v) \in U^2} m_{f,u}(|u-v|) d\alpha(u,v) \xrightarrow{n \rightarrow \infty} 0.$$

By uniqueness of the limit,  $b(x) = \int_{u \in U} f(x, \mu, u) d\omega(u)$ , and  $b \in \overline{\text{co}}F[\mu]$ .

**Remark 2** (Link with the closed convex envelope of [BF23]). *In our case,  $\overline{\text{co}}F[\mu]$  is equal to the closure of  $\mathcal{M}$  in the topology of uniform convergence over compact sets, where*

$$\mathcal{M} := \bigcup_{N \in \mathbb{N}} \left\{ \sum_{i=1}^N \alpha_i f(\cdot, \mu, u_i) \mid \sum_{i=1}^N \alpha_i = 1, \alpha_i \geq 0, u_i \in U \right\}.$$

Indeed, we trivially have  $\mathcal{M} \subset \overline{\text{co}}F$  through the representation  $\omega := \sum_{i=1}^N \alpha_i \delta_{u_i}$ . On the other hand, let  $\bar{b} = F_\omega[\mu] \in \overline{\text{co}}F$  for some  $\omega \in \mathcal{P}(U)$ . For each  $n \in \mathbb{N}$ , cover the compact  $U$  by a finite measurable partition  $(U_i^n)_{i \in [1, N_n]}$  of diameter inferior to  $1/n$ , and pick  $u_i^n \in U_i^n$ . Owing to **[A2]**, there holds for each  $R > 0$

$$\begin{aligned} \left\| \int_{u \in U} f(\cdot, \mu, u) d\omega(u) - \sum_{i=1}^{N_n} \omega(U_i^n) f(\cdot, \mu, u_i^n) \right\|_{\mathcal{C}} &\leq \sup_{x \in \mathbb{R}^d, |x| \leq R} \int_{u \in U} |f(x, \mu, u) - f(x, \mu, u_i^n)| d\omega(u) \\ &\leq (1+R) \int_{u \in U} m_{f,u}(|u - u_i^n|) d\omega(u) \leq (1+R) m_{f,u} \left( \frac{1}{n} \right), \end{aligned}$$

and  $\bar{b}$  is the uniform limit over the each compact of the sequence  $\left( \sum_{i=1}^{N_n} \omega(U_i^n) f(\cdot, \mu, u_i^n) \right)_n \subset \mathcal{M}$ . Thus  $\bar{b} \in \mathcal{M}$ , and equality holds.

Using a selection argument, the set of solutions  $\{(\mu_s^{t,v,\omega})_{s \in [t,T]} \mid \omega \in L^0([t,T]; \mathcal{P}(U))\}$  coincides with the set of solutions of the continuity inclusion

$$\partial_s \mu_s \in -\text{div}(\overline{\text{co}}F[\mu_s] \# \mu_s) \quad s \in [t, T], \quad \mu_t = \nu.$$

Consequently, we have the following.

**THEOREM 2** (RELAXATION (THEOREMS 4.5 AND 4.6 OF [BF23])). Assume **[A1]** and **[A2]**, and let  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ . The set

$$\{(\mu_s^{t,v,\omega})_{s \in [t,T]} \mid \omega \in L^0([t,T]; \mathcal{P}(U))\} \subset \text{AC}([t,T]; \mathcal{P}_2(\mathbb{R}^d))$$

is compact in the topology of the uniform convergence, and is the closure in this topology of the set of trajectories of (1), namely

$$\{(\mu_s^{t,v,u})_{s \in [t,T]} \mid u \in L^0([t,T]; U)\}.$$

### 4.1.3 Linearization of the trajectory

The following technical Lemma allows us to elude the lack of differentiability of a solution of the dynamical system (17), by approximating the said curve only along some sequence.

**Lemma 2** (Right linear approximation). *Assume **[A1]** and **[A2]**. Let  $\bar{s} > 0$ ,  $(\mu_s)_{s \in [0, \bar{s}]}$  be the solution of (17) for some control  $\omega \in L^0([0, \bar{s}]; \mathcal{P}(U))$ . Then there exists  $b \in \overline{\text{co}}F[\mu_0]$  and a sequence  $(s_n) \searrow 0$  such that*

$$\lim_{n \rightarrow \infty} \frac{d_{\mathcal{N}}(\mu_{s_n}, m_{s_n} \# \mu_0)}{s_n} = 0, \quad \text{where } m_s: \mathbb{R}^d \mapsto \mathbb{R}^d \text{ is given by } m_s(x) := x + sb(x).$$

*Demonstration.* Let  $(s_n)_n \searrow 0$ , and define  $b_n : \mathbb{R}^d \mapsto T\mathbb{R}^d$  by

$$b_n(x) := \frac{1}{s_n} \int_{s=0}^{s_n} F_{\omega(s)}[\mu_0](x) ds = F_{\bar{\omega}_n}[\mu_0](x), \quad \text{where} \quad \bar{\omega}_n := \frac{1}{s_n} \int_{s=0}^{s_n} \omega(s) ds \in \mathcal{P}(U).$$

Since  $U$  is compact, so is  $\mathcal{P}(U)$  (see [Vil09], Remark 6.19). Then, along a (non relabeled) subsequence,  $\bar{\omega}_n$  converges to some  $\bar{\omega} \in \mathcal{P}(U)$  for any Monge-Kantorovitch distance, and in particular for the Wasserstein distance over  $\mathcal{P}(U)$ . Let  $b \in \overline{c\bar{O}F}(\mu_0)$  be given by  $b = F_{\bar{\omega}}[\mu_0]$ . For each  $n \in \mathbb{N}$ , let  $\eta_n \in \Gamma(\omega_n, \bar{\omega})$  such that

$$d_{\mathcal{W}, m_{f,u}}(\omega_n, \bar{\omega}) := \inf_{\eta \in \Gamma(\omega_n, \bar{\omega})} \int_{(u,v) \in U^2} m_{f,u}(|u-v|) d\alpha(x, v) \geq \int_{(u,v) \in U^2} m_{f,u}(|u-v|) d\eta_n(x, v) - \frac{1}{n}.$$

Using [A2], we have that

$$|b_n(x) - b(x)| \leq \int_{(u,v) \in U^2} |f(x, \mu_0, u) - f(x, \mu_0, v)| d\eta_n(u, v) \leq (1 + |x|) \left[ d_{\mathcal{W}, m_{f,u}}(\omega_n, \bar{\omega}) + \frac{1}{n} \right],$$

and we conclude to the local uniform convergence of  $b_n$  towards  $b$ . We know from Theorem 1 that  $\mu_s = \Phi_s^{0,\omega} \# \mu_0$ , where the semigroup  $\Phi_s^{0,\omega}$  is defined in (18). Let  $m_s : \mathbb{R}^d \mapsto \mathbb{R}^d$  be given by  $m_s(x) = x + sb(x)$ . Along the sequence  $(s_n)_n$ , we have that

$$\begin{aligned} |\Phi_{s_n}^{0,\omega}(x) - m_{s_n}(x)| &= \left| \int_{s=0}^{s_n} F_{\omega(s)}[\mu_s](\Phi_s^{0,\omega}(x)) ds - s_n b(x) \right| \\ &\leq \int_{s=0}^{s_n} |F_{\omega(s)}[\mu_s](\Phi_s^{0,\omega}(x)) - F_{\omega(s)}[\mu_0](x)| ds + s_n |F_{\bar{\omega}_n}[\mu_0](x) - F_{\bar{\omega}}[\mu_0](x)| \\ &\leq [f] \int_{s=0}^{s_n} d_{\mathcal{W}}(\mu_s, \mu_0) + |\Phi_s^{0,\omega}(x) - x| ds + s_n [f] (1 + |x|) \left( d_{\mathcal{W}, m_{f,u}}(\bar{\omega}_n, \bar{\omega}) + \frac{1}{n} \right). \end{aligned}$$

By (19), we have  $d_{\mathcal{W}}(\mu_s, \mu_0) \leq se^{[f]s} \left( [f] d_{\mathcal{W}}(\mu_0, \delta_0) + |f|_{0,\infty} \right)$ . Plugging this into the above, we get after simplification

$$\frac{d_{\mathcal{W}}(\mu_{s_n}, m_{s_n} \# \mu_0)}{s_n} \leq 2 [f] \left[ s_n e^{[f]s_n} \left( [f] d_{\mathcal{W}}(\mu_0, \delta_0) + |f|_{0,\infty} \right) + d_{\mathcal{W}, m_{f,u}}(\bar{\omega}_n, \bar{\omega}) + \frac{1}{n} \right] \xrightarrow{n \rightarrow \infty} 0.$$

Hence the result.  $\square$

## 4.2 Properties of the value function

Recall that the value function is defined as

$$V : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}, \quad V(t, \nu) := \inf \{ \mathfrak{J}(\mu_T^{t,\nu,u}) \mid u \in L^0([t, T]; U) \}.$$

From Theorem 2 and the assumption [A3] of local uniform continuity of  $\mathfrak{J}$ , we obtain that the set of solutions of the relaxed system (17) may be substituted to the set of solutions of the original problem (1) without changing the value function, that is,

$$V(t, \nu) = \inf \{ \mathfrak{J}(\mu_T^{t,\nu,\omega}) \mid \omega \in L^0([t, T]; \mathcal{P}(U)) \} = \inf_{\mu \in \mathcal{R}_T^{t,\nu}} \mathfrak{J}(\mu),$$

where the reachable set  $\mathcal{R}_T^{t,\nu}$  is defined in (20). Notice that this equality would stand as well with a running cost, since the relaxation result concerns the whole trajectories and not only the reachable sets. In this deterministic setting, we retrieve the classical Dynamic Programming Principle (DPP): for each  $0 < h \leq T - t$ ,

$$V(t, \nu) = \inf_{\mu \in \mathcal{R}_{t+h}^{t,\nu}} V(t+h, \mu). \tag{22}$$

**Lemma 3** (Local uniform continuity of the value function). *Under the assumptions [A1], [A2] and [A3], the function  $V$  is locally uniformly continuous in time and space, i.e. for all  $R > 0$ , there exists a modulus  $m_{V,R} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that*

$$|V(s, \nu) - V(t, \mu)| \leq m_{V,R}(|t-s| + d_{\mathcal{W}}(\mu, \nu)) \quad \forall (t, s) \in [0, T]^2, (\mu, \nu) \in \left( \overline{\mathcal{B}}(\delta_0, R) \right)^2.$$

*Demonstration.* Let  $R > 0$ , and denote  $R_T := R + T \exp[f] T \left( [f] R + |f|_{0,\infty} \right)$  a radius large enough so that  $\mathcal{R}_T^{0,\nu} \subset \overline{\mathcal{B}}(\delta_0, R_T)$  for all  $\nu \in \overline{\mathcal{B}}(\delta_0, R)$ . Let  $m_{\mathfrak{J}, R_T}$  be a local modulus of continuity of  $\mathfrak{J}$  in the ball  $\overline{\mathcal{B}}(\delta_0, R_T)$ . According to the  $[\mathcal{R}]$ -Lipschitz continuity of the reachable sets given by Lemma 1, we have for all  $t \in [0, T]$  and  $\nu, \bar{\nu} \in \overline{\mathcal{B}}(\delta_0, R)$  that

$$V(t, \nu) - V(t, \bar{\nu}) \leq \sup_{\bar{\mu} \in \mathcal{R}_T^{t, \bar{\nu}}} \inf_{\mu \in \mathcal{R}_T^{t, \nu}} \mathfrak{J}(\mu) - \mathfrak{J}(\bar{\mu}) \leq \sup_{\bar{\mu} \in \mathcal{R}_T^{t, \bar{\nu}}} \inf_{\mu \in \mathcal{R}_T^{t, \nu}} m_{\mathfrak{J}, R_T} (d_{\mathcal{W}}(\mu, \bar{\mu})) \leq m_{\mathfrak{J}, R_T}([\mathcal{R}]d_{\mathcal{W}}(\nu, \bar{\nu})).$$

On the other hand, let  $0 \leq t \leq s \leq T$  and  $\nu \in \overline{\mathcal{B}}(\delta_0, R)$ . The DPP (22) and the Grönwall estimate (19) give us

$$\begin{aligned} V(t, \nu) - V(s, \nu) &= \inf_{\mu \in \mathcal{R}_s^{t, \nu}} V(s, \mu) - V(s, \nu) \leq \inf_{\mu \in \mathcal{R}_s^{t, \nu}} m_{\mathfrak{J}, R_T}([\mathcal{R}]d_{\mathcal{W}}(\mu, \nu)) \leq m_{\mathfrak{J}, R_T}([\mathcal{R}]|s-t| \exp([f]T) ([f]R + |f|_{0,\infty})), \\ V(s, \nu) - V(t, \nu) &= \sup_{\mu \in \mathcal{R}_s^{t, \nu}} V(s, \mu) - V(s, \nu) \leq \sup_{\mu \in \mathcal{R}_s^{t, \nu}} m_{\mathfrak{J}, R_T}([\mathcal{R}]d_{\mathcal{W}}(\mu, \nu)) \leq m_{\mathfrak{J}, R_T}([\mathcal{R}]|s-t| \exp([f]T) ([f]R + |f|_{0,\infty})). \end{aligned}$$

Hence  $V$  is locally uniformly continuous with a modulus depending only on  $\mathfrak{J}$ ,  $f$  and  $T$ .  $\square$

## 5 The Hamilton-Jacobi equation

In this section, we are interested into the following generic Hamilton-Jacobi equation

$$-\partial_t u(t, \mu) + H(\mu, D_\mu u(t, \mu)) = 0, \quad u(T, \mu) = \mathfrak{J}(\mu). \quad (23)$$

Regularity assumptions on the Hamiltonian  $H: \mathbb{T} \rightarrow \mathbb{R}$  will be precised further.

### 5.1 Notion of viscosity solutions

Recall that  $X := ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d)$  and  $d_X^2((t, \mu), (s, \nu)) := |s-t|^2 + d_{\mathcal{W}}^2(\mu, \nu)$ . In this section, we precise the definition of a viscosity solution of (23). To this aim, we will use a class of *test functions*, that will be more regular than the viscosity solution in order to bear the derivatives. The time variable and the measure variable of the test functions do not play symmetric roles, as weaker regularity on the measure dimension will be compensated by stronger assumptions on the time dimension.

#### 5.1.1 Regularity in the measure variable

**Definition 4** (Locally semiconcave/convex maps). *An application  $u: \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is locally semiconcave in  $\mathcal{P}_2(\mathbb{R}^d)$  if for all  $R > 0$ , there exists  $\lambda_R \in \mathbb{R}$  such that for all  $\mu, \nu \in \overline{\mathcal{B}}(\delta_0, R)$  and  $\eta = \eta(x, y) \in \Gamma_o(\mu, \nu)$ , there holds*

$$u((1-h)\pi_x + h\pi_y) \# \eta \geq (1-h)u(\mu) + hu(\nu) - \frac{\lambda_R}{2} h(1-h) d_{\mathcal{W}}^2(\mu, \nu) \quad \forall h \in [0, 1].$$

*An application  $u: \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is semiconvex if  $-u$  is semiconcave.*

Locally semiconcave/convex maps are directionally differentiable at all points. As an important example, the squared Wasserstein distance is directionally differentiable (see [Gig08, §4.2]) and for any  $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$  and  $\xi \in \mathcal{P}_2(\mathbb{T}\mathbb{R}^d)_\mu$ ,

$$D_\mu d_{\mathcal{W}}^2(\cdot, \sigma)(\xi) = \inf_{\eta \in \exp_\mu^{-1}(\sigma)} \inf_{\alpha \in \Gamma_\mu(\xi, \eta)} -2 \int_{(x, v, w) \in \mathbb{T}^2 \mathbb{R}^d} \langle v, w \rangle d\alpha(x, v, w). \quad (24)$$

Moreover, given  $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ , we may compute  $\|D_\mu d_{\mathcal{W}}^2(\cdot, \sigma)\|_\mu$  as defined in (16). Indeed,

$$\begin{aligned} |D_\mu d_{\mathcal{W}}^2(\cdot, \sigma)(\xi)| &= \left| \inf_{\alpha \in \Gamma_\mu(\xi, \exp_\mu^{-1}(\sigma))} \int_{(x, v, w)} -2 \langle v, w \rangle d\alpha(x, v, w) \right| \leq 2 \sup_{\alpha \in \Gamma_o(\xi, \exp_\mu^{-1}(\sigma))} \int_{(x, v, w)} |v| |w| d\alpha(x, v, w) \\ &\leq 2 \left( \int_{(x, v) \in \mathbb{T}\mathbb{R}^d} |v|^2 d\xi(x, v) \right)^{1/2} \sup_{\eta \in \exp_\mu^{-1}(\sigma)} \left( \int_{(x, w) \in \mathbb{T}\mathbb{R}^d} |w|^2 d\eta(x, w) \right)^{1/2} = 2 \|\xi\|_\mu d_{\mathcal{W}}(\mu, \sigma). \end{aligned}$$

Hence  $\|D_\mu d_{\mathcal{W}}^2(\cdot, \sigma)\|_\mu \leq 2d_{\mathcal{W}}(\mu, \sigma)$ . On the other hand, if  $\sigma \neq \mu$ , letting  $\xi \in \exp_\mu^{-1}(\sigma)$ ,

$$\|D_\mu d_{\mathcal{W}}^2(\cdot, \sigma)\|_\mu \geq \left| \frac{D_\mu d_{\mathcal{W}}^2(\cdot, \sigma)(\xi)}{\|\xi\|_\mu} \right| = \frac{1}{d_{\mathcal{W}}(\mu, \sigma)} \left| \lim_{h \searrow 0} \frac{(1-h)^2 d_{\mathcal{W}}^2(\mu, \sigma) - d_{\mathcal{W}}^2(\mu, \sigma)}{h} \right| = 2d_{\mathcal{W}}(\mu, \sigma), \quad (25)$$

showing equality.

**Remark 3** (Composition rule). *Let  $\varphi \in \mathcal{C}^2(\mathbb{R}^+; \mathbb{R}^+)$  be nondecreasing, and consider the composition  $\psi : \mu \mapsto \varphi(d_{\mathcal{W}}^2(\delta_0, \mu))$ . Denote  $\lambda_R$  a local constant of semiconcavity of  $\varphi$  over  $[0, 3R]$ , and  $[\varphi_R]$  a local constant of Lipschitz-continuity of  $\varphi$  over the same domain. Then  $\psi$  is semiconcave with modulus  $R\lambda_R + [\varphi_R]$  (see [CS04, Proposition 2.1.12]).*

In addition to the above, we ask for the following consistency with the tangent cone.

**Definition 5** (Geometrically consistent application). *We say that an application  $u : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is consistent with the geometric structure if it is directionally differentiable and if*

$$\forall \mu \in \mathcal{P}_2(\mathbb{R}^d), \quad \forall \xi \in \mathcal{P}_2(\mathbb{T}\mathbb{R}^d)_\mu, \quad D_\mu u(\xi) = D_\mu u(\pi^\mu \xi).$$

The fact that the squared Wasserstein distance is geometrically consistent is given by [JJZ, Theorem 3.8].

### 5.1.2 Regularity in the time variable

**Definition 6** (Locally Lipschitz time derivative). *We say that  $\varphi : X \rightarrow \mathbb{R}$  has locally Lipschitz time derivative if for all  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ ,  $\varphi(\cdot, \mu) \in \mathcal{C}^1([0, T]; \mathbb{R})$ , and if the application  $(t, \mu) \mapsto \partial_t \varphi(t, \mu)$  is locally Lipschitz in  $(X, d_X)$ .*

**Lemma 4** (Partial derivatives). *Let  $\varphi : X \rightarrow \mathbb{R}$  satisfy Definition 6. Let  $(t, \mu) \in X$  such that  $\varphi(t, \cdot)$  is directionally differentiable at  $\mu$ , and  $\xi \in \mathbf{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ . Then*

$$\lim_{h \searrow 0} \frac{\varphi(t+h, \exp_\mu(h \cdot \xi)) - \varphi(t, \mu)}{h} = \partial_t \varphi(t, \mu) + D_\mu \varphi(t, \mu)(\xi).$$

*Demonstration.* Let  $C$  be a Lipschitz constant for  $\partial_s \varphi$  in the ball  $\mathcal{B}\left((t, \mu), \sqrt{1 + \|\xi\|_\mu^2}\right)$ . Then for all  $0 < h \leq 1$ ,

$$\begin{aligned} \frac{\varphi(t+h, \exp_\mu(h \cdot \xi)) - \varphi(t, \mu)}{h} &= \frac{\varphi(t+h, \exp_\mu(h \cdot \xi)) - \varphi(t, \exp_\mu(h \cdot \xi))}{h} + \frac{\varphi(t, \exp_\mu(h \cdot \xi)) - \varphi(t, \mu)}{h} \\ &= \frac{1}{h} \int_{r=t}^{t+h} \partial_r \varphi(r, \exp_\mu(h \cdot \xi)) dr + \frac{\varphi(t, \exp_\mu(h \cdot \xi)) - \varphi(t, \mu)}{h} \\ &\in \partial_t \varphi(t, \mu) \pm \frac{C}{h} \int_{r=t}^{t+h} |r-t| + d_{\mathcal{W}}(\exp_\mu(h \cdot \xi), \mu) dr + \frac{\varphi(t, \exp_\mu(h \cdot \xi)) - \varphi(t, \mu)}{h} \\ &\subset \partial_t \varphi(t, \mu) \pm Ch(1 + \|\xi\|_\mu) + \frac{\varphi(t, \exp_\mu(h \cdot \xi)) - \varphi(t, \mu)}{h}. \end{aligned}$$

Letting  $h \searrow 0$  and using the directional differentiability of  $\varphi(t, \cdot)$  at  $\mu$ , we obtain the result.  $\square$

### 5.1.3 Locally uniform upper semicontinuity

Due to the lack of local compactness of  $\mathcal{P}_2(\mathbb{R}^d)$ , we consider a stronger definition than upper semicontinuity.

**Definition 7** (Locally uniformly upper semicontinuous). *Let  $(Y, d)$  be a complete metric space. A locally bounded application  $u : Y \rightarrow \mathbb{R}$  is said to be locally uniformly upper semicontinuous (luusc) if for any decreasing family of closed bounded sets  $(B_n)_{n \in \mathbb{N}}$  such that  $B := \bigcap_{n \in \mathbb{N}} B_n \neq \emptyset$  and  $\lim_{n \rightarrow \infty} \sup_{x \in B_n} \inf_{y \in B} d_Y(x, y) \rightarrow 0$ , there holds*

$$\lim_{n \rightarrow \infty} \sup_{y \in B_n} u(y) \leq \sup_{x \in B} u(x). \quad (26)$$

Similarly, we say that  $u$  is locally uniformly lower semicontinuous (lulsc) if  $-u$  is luusc.

**Remark 4** (Link with other notions of upper semicontinuity). *We gather here some properties, whose proofs are postponed to Appendix A.*

- Definition 7 is strictly weaker than continuity.
- In general, Definition 7 is strictly stronger than upper semicontinuity. However both definitions whenever  $Y$  is locally compact.
- Let  $\mathcal{S}$  be the set of nonempty closed and bounded subsets of  $Y$ . Then Definition 7 is equivalent to the upper semicontinuity of the set function  $U : B \mapsto \sup_{x \in B} u(x)$  in the Hausdorff topology. This definition makes sense in connection with the  $(\min, +)$  interpretation of Hamilton-Jacobi equations, as it exactly says that the Maslov measure of density  $u$  is upper semicontinuous (see [KM97; DMD99]).
- The applications that are simultaneously luusc and lulsc are exactly the locally uniformly continuous applications.
- In  $Y = \mathcal{P}_2(\mathbb{R}^d)$ , there is no comparison with upper semicontinuity in the narrow topology (see counterexamples in Appendix A).

#### 5.1.4 Definition of viscosity solutions

Gathering the above definitions, we arrive at the following.

**Definition 8** (Test functions). *We define*

$$\mathcal{T}_+ := \left\{ \varphi : ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \mid \begin{array}{l} \varphi \text{ and } \partial_t \varphi \text{ are locally Lipschitz, and } \forall s \in [0, T], \\ \varphi(s, \cdot) \text{ is locally semiconcave and geometrically consistent} \end{array} \right\}.$$

Similarly, we denote  $\mathcal{T}_- := \{-\varphi \mid \varphi \in \mathcal{T}_+\}$ , that is,

$$\mathcal{T}_- := \left\{ \varphi : ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \mid \begin{array}{l} \varphi \text{ and } \partial_t \varphi \text{ are locally Lipschitz, and } \forall s \in [0, T], \\ \varphi(s, \cdot) \text{ is locally semiconvex and geometrically consistent} \end{array} \right\}.$$

Distinguished members of  $\mathcal{T}_\pm$  are the applications of the form  $\varphi(t, \mu) = \psi(t) \pm \alpha d_{\mathcal{V}}^2(\mu, \sigma)$ , where  $\psi \in C^2([0, T]; \mathbb{R})$ ,  $\alpha \geq 0$  and  $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$  is fixed. Once given  $\mathcal{T}_\pm$ , the definition of viscosity solutions is a natural generalization of the finite-dimensional case.

**Definition 9** (Viscosity solutions). *We say that  $u : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is a viscosity*

- *subsolution if it is locally uniformly upper semicontinuous, if  $u(T, \cdot) \leq \mathfrak{J}$ , and if for any  $\varphi \in \mathcal{T}_+$  such that  $u - \varphi$  reaches a maximum at  $(t, \mu) \in X = ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d)$ , there holds*

$$-\partial_t \varphi(t, \mu) + H(\mu, D_\mu \varphi(t, \mu)) \leq 0. \quad (27)$$

- *supersolution if it is locally uniformly lower semicontinuous, if  $u(T, \cdot) \geq \mathfrak{J}$ , and if for any  $\psi \in \mathcal{T}_-$  such that  $u - \psi$  reaches a minimum at  $(t, \mu) \in X$ , there holds*

$$-\partial_t \psi(t, \mu) + H(\mu, D_\mu \psi(t, \mu)) \geq 0. \quad (28)$$

- *solution if it is both a sub and a supersolution.*

## 5.2 Comparison principle

The comparison principles, or maximum principles in the literature of elliptic equations, are used in the viscosity theory to provide uniqueness of the viscosity solutions. They draw their name from the corresponding results used over the viscous approximations of the PDE, and evolved jointly with the growing scope of HJB equations. When addressing equations in non-locally compact spaces, it is now common to rely on variations over Ekeland's variational principle [Eke74]: see [LY95; AF14; GS15]. This is not the only strategy in use in the literature: one could also modify the definition in order to stay over compact sets, as in [FK09] or [WZ19] in the pathwise setting.

The perturbed optimization principle will bring, as announced, perturbations. To cope with these additional terms, [MQ18; JM20] consider an "enlarged" set of semidifferentials, and a strengthened notion of viscosity solutions. We take another point of view by using the Borwein-Preiss principle, also called *smooth Ekeland principle*, that allows to choose the perturbation in a way that they can be embedded into the test functions. To ease the reading, we factorize the application of this theorem in the following Lemma.

**Lemma 5** (Perturbed optimization). Denote  $Y = \overline{X}^2 = [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ , endowed with the distance

$$d_Y^2((t, \mu, s, \nu), (\bar{t}, \bar{\mu}, \bar{s}, \bar{\nu})) := |t - \bar{t}|^2 + d_{\mathcal{V}}^2(\mu, \bar{\mu}) + |s - \bar{s}|^2 + d_{\mathcal{V}}^2(\nu, \bar{\nu}).$$

Let  $\Phi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$  be upper semicontinuous, proper and upper bounded,  $z^0 \in Y$  be fixed such that  $A := \sup \Phi - \Phi(z^0) < \infty$ , and assume that there exists  $R > 0$  such that

$$\{z \in Y \mid \Phi(z) \geq \Phi(z^0)\} \subset \overline{\mathcal{B}}(z^0, R). \quad (29)$$

Hence for each  $n \in \mathbb{N}_*$ , there exists  $z_n \in Y$  and a perturbation  $p_n : Y \rightarrow \mathbb{R}^+$  such that

1. the perturbed map  $\Phi - p_n$  reaches a global strict maximum in  $z_n$ ,
2. The map  $(t, \mu) \mapsto p_n(t, \mu, s_n, \nu_n)$  belongs to  $\mathcal{T}_+$ , and  $(s, \nu) \mapsto -p_n(t_n, \mu_n, s, \nu)$  belongs to  $\mathcal{T}_-$ ,
3. There exists an application  $\omega_{T,R,A} : \mathbb{N} \rightarrow \mathbb{R}^+$  such that

$$\sum_{r \in \{t, s\}} |\partial_r p_n(z_n)| + \sum_{\sigma \in \{\mu, \nu\}} (1 + d_{\mathcal{V}}(\sigma, \delta_0)) \|D_\sigma p_n(z_n)\|_\sigma \leq \omega_{T,R,A}(n) \xrightarrow{n \rightarrow \infty} 0,$$

4. There holds  $\sup \Phi \leq \Phi(z_n) + C_{R,A}(n)$ , where  $C_{R,A}$  is decreasing towards 0 when  $n \rightarrow \infty$ .

The proof of Lemma 5 is delayed to Appendix B.

**Assumption [A4]** (Structure of the Hamiltonian). There exists a constant  $[H]$  such that for any  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ , any  $a > 0$  and  $p, q \in \mathbb{T}_\mu$ ,

$$|H(\mu, p + q) - H(\mu, p)| \leq [H] (1 + d_{\mathcal{V}}(\mu, \delta_0)) \|q\|_\mu, \quad (30)$$

$$H(\mu, -aD_\mu d_{\mathcal{V}}^2(\cdot, \nu)) - H(\nu, aD_\nu d_{\mathcal{V}}^2(\mu, \cdot)) \leq 2a [H] d_{\mathcal{V}}^2(\mu, \nu). \quad (31)$$

The condition (30) is classical, and traduces a locally Lipschitz behavior of the Hamiltonian. The condition (31) could be interpreted as a one-sided Lipschitz control on the variation with respect to the first argument only.

**THEOREM 3** (COMPARISON PRINCIPLE). Assume [A4]. Let  $u : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  be a subsolution of (23) and  $v : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  be a supersolution of (23), which are such that  $u(T, \mu) \leq v(T, \mu)$  for all  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ . Then

$$u(t, \mu) \leq v(t, \mu) \quad \forall (t, \mu) \in ]0, T] \times \mathcal{P}_2(\mathbb{R}^d). \quad (32)$$

This proof builds on the ideas of [FGS17, Theorem 3.50, p. 206] developed in Hilbert spaces. The structure is the following: assume by contradiction that the inequality (32) is not satisfied. Thus we have an information on the sign of the maximum  $(t, \mu, s, \nu) \mapsto u(t, \mu) - v(s, \nu)$  on the diagonal  $t = s, \mu = \nu$ . As in the classical proof, this maximum over the diagonal is approximated by the maximum over the doubled space  $\overline{X}^2$  of a perturbation of  $u \ominus v$  that penalizes the distance to the diagonal. At the point of maximum, freezing two of the four variables of  $\overline{X}^2$  allows to apply the definition of viscosity sub and supersolution, giving two inequalities whose combination will eventually lead to a contradiction.

The proof below has several specificities with respect to the canonical arguments (see [CIL92, Section 3]). In the argument adapted to  $\mathbb{R}^d$ , the points of maximum over the doubled space exist, and admit cluster points on the diagonal. In  $\overline{X} = [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ , balls are not compact, and there is no reason for the maxima to be attained. This is circumvented by the use of variational principle in the vein of Ekeland's principle. Secondly, the so-constructed sequence of "almost maxima" do not necessarily contain converging subsequences. This is where the stronger form of upper semicontinuity of Definition 7 replaces the standard extraction, by working with limits over sets that decrease towards the diagonal. As opposite to the Hilbertian setting, we are not able to use a linear perturbation into the Ekeland principle, and we have to manipulate series of squared distances in order to get sufficient smoothness to embed the perturbation into test functions.

*Demonstration.* Denote again  $X = ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d)$ . By abuse of notation, let  $X^2 := ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d) \times ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d)$  and

$$d_X^2((t, \mu), (\bar{t}, \bar{\mu})) := |t - \bar{t}|^2 + d_{\mathcal{V}}^2(\mu, \bar{\mu}), \quad d_{X^2}^2((t, \mu, s, \nu), (\bar{t}, \bar{\mu}, \bar{s}, \bar{\nu})) := d_X^2((t, \mu), (\bar{t}, \bar{\mu})) + d_X^2((s, \nu), (\bar{s}, \bar{\nu})).$$

Assume by contradiction that

$$\Gamma := \min \left( 1, \sup \left\{ u(t, \mu) - v(t, \mu) \mid (t, \mu) \in \overline{X} \right\} \right) > 0. \quad (33)$$

**Penalizations** As  $u$  and  $-v$  are locally uniformly upper semicontinuous, we may build an application  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  controlling the growth of both  $u$  and  $-v$ , i.e. such that

$$\max(u(t, \mu), -v(t, \mu)) \leq g(d_{\mathcal{W}}(\mu, \delta_0)) \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (34)$$

Up to regularization, we may assume that  $g$  is increasing, of class  $\mathcal{C}^2$  and with  $g'(\cdot) \geq 1$ . Denote

$$h : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad h(t, r) := g^2((1+r)e^{-4[H]t}) + \frac{1}{t}.$$

As the composition  $\mu \mapsto h(t, d_{\mathcal{W}}^2(\delta_0, \mu))$  grows strictly faster than  $u$  and  $-v$  by (34), for any  $\iota > 0$ ,

$$\max[u(t, \mu), -v(t, \mu)] - \iota h(t, d_{\mathcal{W}}^2(\delta_0, \mu)) \leq \sup_{r \in \mathbb{R}^+} g(r) - \iota g^2((1+r^2)e^{-4[H]T}) =: \lVert\lVert g - \iota h \rVert\rVert < \infty. \quad (35)$$

Since  $g \geq 1$  and  $g' \geq 1$ , there holds

$$\partial_t h(t, r) = -8[H](1+r)e^{-4[H]t}(gg')((1+r)e^{-4[H]t}) - \frac{1}{t^2} \leq -8[H]e^{-4[H]T}. \quad (36)$$

Moreover, for any  $t > 0$  and  $r \in \mathbb{R}^+$ ,

$$2[H](1+r)\partial_r h(t, r) = 4[H](1+r)g((1+r)e^{-4[H]t})\partial_r g((1+r)e^{-4[H]t})e^{-4[H]t} = -\frac{1}{2}\partial_t h(t, r) - \frac{1}{2t^2} \leq -\frac{1}{2}\partial_t h(t, r). \quad (37)$$

As  $g$  is  $\mathcal{C}^2$ , the map  $h$  has locally Lipschitz time derivatives in  $X=0, T[\times \mathcal{P}_2(\mathbb{R}^d)]$ . Moreover, owing to Remark 3, the composition  $\mu \mapsto h(t, d_{\mathcal{W}}^2(\delta_0, \mu))$  is locally semiconcave for all  $t$ , and locally Lipschitz. Therefore  $(t, \mu) \mapsto \pm h(t, d_{\mathcal{W}}^2(\mu, \delta_0))$  belongs to  $\mathcal{F}_{\pm}$ .

We consider

$$\Gamma_\iota := \limsup_{r \rightarrow 0} \left\{ u(t, \mu) - v(s, \nu) - \iota(h(t, d_{\mathcal{W}}^2(\mu, \delta_0)) + h(s, d_{\mathcal{W}}^2(\nu, \delta_0))) \mid (t, \mu, s, \nu) \in \bar{X}^2, d_X((t, \mu), (s, \nu)) \leq r \right\}, \quad (38)$$

$$\Gamma_{\iota\varepsilon} := \sup \left\{ u(t, \mu) - v(s, \nu) - \iota(h(t, d_{\mathcal{W}}^2(\mu, \delta_0)) + h(s, d_{\mathcal{W}}^2(\nu, \delta_0))) - \frac{d_X^2((t, \mu), (s, \nu))}{\varepsilon} \mid (t, \mu, s, \nu) \in \bar{X}^2 \right\}. \quad (39)$$

Then one has

$$\Gamma_{\iota\varepsilon} \searrow_{\varepsilon \searrow 0} \Gamma_\iota \quad \text{for all } \iota > 0, \quad \text{and} \quad \Gamma_\iota \nearrow_{\iota \searrow 0} \Gamma_0 \geq \Gamma. \quad (40)$$

Here  $\Gamma_0$  may be equal to  $+\infty$ . The arguments of (40) are easy but tedious, and devolved in Appendix C to lighten the presentation.

**The perturbed maximization** Let  $\Phi_{\iota\varepsilon} : \bar{X}^2 \rightarrow \{-\infty\} \cup \mathbb{R}$  be given by

$$\Phi_{\iota\varepsilon}(t, \mu, s, \nu) := u(t, \mu) - v(s, \nu) - \iota(h(t, d_{\mathcal{W}}^2(\mu, \delta_0)) + h(s, d_{\mathcal{W}}^2(\nu, \delta_0))) - \frac{d_X^2((t, \mu), (s, \nu))}{\varepsilon}.$$

Let  $z^0 := (T, \delta_0, T, \delta_0)$ . For each fixed  $\iota, \varepsilon$ , the application  $\Phi_{\iota\varepsilon}$  is upper semicontinuous, proper and - using (35) - upper bounded in the complete metric space  $(\bar{X}^2, d_{X^2})$ . Moreover, if  $\Phi_{\iota\varepsilon}(z) \geq \Phi_{\iota\varepsilon}(z^0)$ , then

$$\frac{\iota}{2}(h(t, d_{\mathcal{W}}^2(\mu, \delta_0)) + h(s, d_{\mathcal{W}}^2(\nu, \delta_0))) \leq \lVert\lVert u - \frac{\iota}{2}h \rVert\rVert + \lVert\lVert -v - \frac{\iota}{2}h \rVert\rVert + 0 - (u(T, \delta_0) - v(T, \delta_0) - 2\iota h(T, 0) - 0), \quad (41)$$

and there exists  $R_\iota > 0$  such that  $d_{X^2}(z, z^0) \leq R_\iota$ . Notice that

$$\sup \Phi_{\iota\varepsilon} - \Phi_{\iota\varepsilon}(z^0) \leq \lVert\lVert u - \iota h \rVert\rVert + \lVert\lVert -v - \iota h \rVert\rVert - (u(T, \delta_0) - v(T, \delta_0) - 2\iota h(T, 0)) =: A_\iota$$

does not depend on  $\varepsilon$ .

Hence we may apply Lemma 5: for each  $n \in \mathbb{N}_*$ , there exist  $z_{\iota\varepsilon n} \in \bar{X}^2$  and a perturbation  $p_{\iota\varepsilon n} : \bar{X}^2 \rightarrow \mathbb{R}^+$  such that  $\Phi_{\iota\varepsilon} - p_{\iota\varepsilon n}$  reaches a maximum at  $z_{\iota\varepsilon n}$ , the partial functions  $(t, \mu) \mapsto p_{\iota\varepsilon n}(t, \mu, s_{\iota\varepsilon n}, \nu_{\iota\varepsilon n})$  and  $(s, \nu) \mapsto -p_{\iota\varepsilon n}(t_{\iota\varepsilon n}, \mu_{\iota\varepsilon n}, s, \nu)$  belong respectively to  $\mathcal{F}_+, \mathcal{F}_-$ , and there exist maps  $\omega_\iota, C_\iota : \mathbb{N}_* \rightarrow \mathbb{R}^+$  satisfying

$$\sum_{r \in \{s, t\}} |\partial_r p_{\iota\varepsilon n}(z_{\iota\varepsilon n})| + \sum_{\sigma \in \{\mu, \nu\}} (1 + d_{\mathcal{W}}(\sigma_{\iota\varepsilon n}, \delta_0)) \|D_\sigma p_{\iota\varepsilon n}(z_{\iota\varepsilon n})\|_{\sigma_{\iota\varepsilon n}} \leq \omega_\iota(n) \xrightarrow{n \rightarrow \infty} 0, \quad (42)$$

$$\Gamma_{\iota\varepsilon} - \Phi_{\iota\varepsilon}(z_{\iota\varepsilon n}) = \sup \Phi_{\iota\varepsilon} - \Phi_{\iota\varepsilon}(z_{\iota\varepsilon n}) \leq C_\iota(n) \xrightarrow{n \rightarrow \infty} 0.$$

Notice that

$$\Gamma_{i\varepsilon} + \frac{d_X^2((t_{i\varepsilon n}, \mu_{i\varepsilon n}), (s_{i\varepsilon n}, \nu_{i\varepsilon n}))}{2\varepsilon} \leq \Phi_{i\varepsilon}(z_{i\varepsilon n}) + C_i(n) + \frac{d_X^2((t_{i\varepsilon n}, \mu_{i\varepsilon n}), (s_{i\varepsilon n}, \nu_{i\varepsilon n}))}{2\varepsilon} \leq \Gamma_{i\frac{\varepsilon}{2}} + C_i(n)$$

so that for each fixed  $\iota$ , there holds by (40)

$$\frac{d_X^2((t_{i\varepsilon n}, \mu_{i\varepsilon n}), (s_{i\varepsilon n}, \nu_{i\varepsilon n}))}{\varepsilon} \leq 2\left(\Gamma_{i\frac{\varepsilon}{2}} - \Gamma_{i\varepsilon} + C_i(n)\right) \xrightarrow{\varepsilon \rightarrow 0, n \rightarrow \infty} 0. \quad (43)$$

**Staying away from  $T$**  Let us show that for sufficiently small  $\iota$  and large  $n$ , the points  $t_{i\varepsilon n}, s_{i\varepsilon n}$  belong to  $]0, T[$ . Since  $h(0, r) = +\infty$  for all  $r \geq 0$ , the construction of  $z_{i\varepsilon n}$  implies  $t_{i\varepsilon n}, s_{i\varepsilon n} > 0$ . On the other hand, recalling (40), let  $\iota_0 > 0$  be small enough so that  $\Gamma_\iota \geq \Gamma/2$  for all  $0 < \iota \leq \iota_0$ . Hence, as  $\varepsilon \mapsto \Gamma_{i\varepsilon}$  is nonincreasing,

$$u(t_{i\varepsilon n}, \mu_{i\varepsilon n}) - v(s_{i\varepsilon n}, \nu_{i\varepsilon n}) \geq \Phi_{i\varepsilon}(z_{i\varepsilon n}) \geq \Gamma_{i\varepsilon} - C_i(n) \geq \Gamma_\iota - C_i(n) \geq \frac{\Gamma}{2} - C_i(n)$$

and there exists  $n_\iota \in \mathbb{N}_*$  large enough such that  $u(t_{i\varepsilon n}, \mu_{i\varepsilon n}) - v(s_{i\varepsilon n}, \nu_{i\varepsilon n}) \geq \frac{\Gamma}{4}$  for all  $n \geq n_\iota$ .

Then for each  $0 < \iota \leq \iota_0$ ,  $\varepsilon > 0$  sufficiently small and  $n$  sufficiently large,  $t_{i\varepsilon n}, s_{i\varepsilon n} < T$  simultaneously. Indeed, recall from (41) that  $\mu_{i\varepsilon n}, \nu_{i\varepsilon n} \in \overline{\mathcal{B}}(\delta_0, R_\iota)$  for all  $\varepsilon, n$ . If there exists a sequence  $(\varepsilon_m, n_m) \rightarrow_m (0, \infty)$  such that  $n_m \geq n_\iota$  and  $T \in \{t_{i\varepsilon_m n_m}, s_{i\varepsilon_m n_m}\}$  for all  $m$ , then using (43),

$$\begin{aligned} \frac{\Gamma}{4} &\leq \liminf_{m \rightarrow \infty} u(t_{i\varepsilon_m n_m}, \mu_{i\varepsilon_m n_m}) - v(s_{i\varepsilon_m n_m}, \nu_{i\varepsilon_m n_m}) \\ &\leq \limsup_{r \searrow 0} \left\{ u(t, \mu) - v(s, \nu) \mid d_X^2((t, \mu), (s, \nu)) \leq r, \mu, \nu \in \overline{\mathcal{B}}(\delta_0, R_\iota), (t, s) \in [T-r, T]^2 \right\} \\ &\leq \sup \left\{ u(t, \mu) - v(s, \nu) \mid t = s = T, \mu = \nu \in \overline{\mathcal{B}}(\delta_0, R_\iota) \right\} = \sup_{\mu \in \overline{\mathcal{B}}(\delta_0, R_\iota)} u(T, \mu) - v(T, \mu). \end{aligned}$$

The last inequality holds since  $(t, \mu, s, \nu) \mapsto u(t, \mu) - v(s, \nu)$  is locally uniformly upper semicontinuous (see Definition 7). But this contradicts the assumption that  $u(T, \cdot) - v(T, \cdot) \leq 0$ .

**Application of the definition of semisolutions** Define  $\varphi_{i\varepsilon n}, \psi_{i\varepsilon n} : X = ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  as

$$\begin{aligned} \varphi_{i\varepsilon n}(t, \mu) &:= v(s_{i\varepsilon n}, \nu_{i\varepsilon n}) + \iota \left( h(t, d_{\mathcal{V}}^2(\mu, \delta_0)) + h(s_{i\varepsilon n}, d_{\mathcal{V}}^2(\nu_{i\varepsilon n}, \delta_0)) \right) + \frac{d_X^2((t, \mu), (s_{i\varepsilon n}, \nu_{i\varepsilon n}))}{\varepsilon} + p_{i\varepsilon n}(t, \mu, s_{i\varepsilon n}, \nu_{i\varepsilon n}), \\ \psi_{i\varepsilon n}(s, \nu) &:= u(t_{i\varepsilon n}, \mu_{i\varepsilon n}) - \iota \left( h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)) + h(s, d_{\mathcal{V}}^2(\nu, \delta_0)) \right) - \frac{d_X^2((t_{i\varepsilon n}, \mu_{i\varepsilon n}), (s, \nu))}{\varepsilon} - p_{i\varepsilon n}(t_{i\varepsilon n}, \mu_{i\varepsilon n}, s, \nu). \end{aligned}$$

By construction,  $\varphi_{i\varepsilon n} \in \mathcal{T}_+$  and  $\psi_{i\varepsilon n} \in \mathcal{T}_-$ . Recalling that  $u - \varphi_{i\varepsilon n}$  reaches a maximum at  $(t_{i\varepsilon n}, \mu_{i\varepsilon n}) \in X$ , we have

$$\begin{aligned} & -\iota \partial_t h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)) - \frac{2(t_{i\varepsilon n} - s_{i\varepsilon n})}{\varepsilon} - \partial_t p_{i\varepsilon n}(z_{i\varepsilon n}) \\ & + H \left( \mu_{i\varepsilon n}, \iota \partial_r h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)) D_\mu d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0) + \frac{D_\mu d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \nu_{i\varepsilon n})}{\varepsilon} + D_\mu p_{i\varepsilon n}(z_{i\varepsilon n}) \right) \leq 0. \end{aligned} \quad (44)$$

Let momentarily  $c_{i\varepsilon n} := [H](1 + d_{\mathcal{V}}(\mu_{i\varepsilon n}, \delta_0))$ . Using the Lipschitz assumption (30) and the estimate (25) on the differential of the squared distance, since  $\partial_r h \geq 0$ ,

$$H \left( \mu_{i\varepsilon n}, \iota \partial_r h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)) D_\mu d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0) + \frac{D_\mu d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \nu_{i\varepsilon n})}{\varepsilon} + D_\mu p_{i\varepsilon n}(z_{i\varepsilon n}) \right) \quad (45)$$

$$\geq H \left( \mu_{i\varepsilon n}, \frac{D_\mu d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \nu_{i\varepsilon n})}{\varepsilon} \right) - c_{i\varepsilon n} (\iota \partial_r h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)) \|D_\mu d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)\|_{\mu_{i\varepsilon n}} + \|D_\mu p_{i\varepsilon n}(z_{i\varepsilon n})\|_{\mu_{i\varepsilon n}}) \quad (46)$$

$$\geq H \left( \mu_{i\varepsilon n}, \frac{D_\mu d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \nu_{i\varepsilon n})}{\varepsilon} \right) - c_{i\varepsilon n} (2\iota d_{\mathcal{V}}(\mu_{i\varepsilon n}, \delta_0) \partial_r h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)) + \|D_\mu p_{i\varepsilon n}(z_{i\varepsilon n})\|_{\mu_{i\varepsilon n}}). \quad (47)$$

Recalling that the partial derivatives of  $h$  satisfy (37), we get that

$$2[H](1 + d_{\mathcal{V}}(\mu_{i\varepsilon n}, \delta_0)) d_{\mathcal{V}}(\mu_{i\varepsilon n}, \delta_0) \partial_r h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)) \leq -\frac{1}{2} \partial_t h(t_{i\varepsilon n}, d_{\mathcal{V}}^2(\mu_{i\varepsilon n}, \delta_0)). \quad (48)$$



On the other hand,  $v - \psi_{t_{\varepsilon n}}$  admits a minimum in  $(s_{t_{\varepsilon n}}, v_{t_{\varepsilon n}}) \in X$ . Applying the definition of supersolution and repeating the argument of (47) and (42), we get

$$\begin{aligned} & \iota \partial_s h(s_{t_{\varepsilon n}}, d_{\mathcal{V}}^2(v_{t_{\varepsilon n}}, \delta_0)) + \frac{2(s_{t_{\varepsilon n}} - t_{\varepsilon n})}{\varepsilon} + \partial_s p_{t_{\varepsilon n}}(z_{t_{\varepsilon n}}) - \frac{\iota}{2} \partial_s h(s_{t_{\varepsilon n}}, d_{\mathcal{V}}^2(v_{t_{\varepsilon n}}, \delta_0)) \\ & + [H](1 + d_{\mathcal{V}}(v_{t_{\varepsilon n}}, \delta_0)) \|D_v p_{t_{\varepsilon n}}(z_{t_{\varepsilon n}})\|_{v_{t_{\varepsilon n}}} + H\left(v_{t_{\varepsilon n}}, -\frac{D_v d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, v_{t_{\varepsilon n}})}{\varepsilon}\right) \geq 0. \end{aligned} \quad (49)$$

Taking the difference between (49) and the inequality resulting from plugging (47) and (42) into (44), we get after simplification

$$-\frac{\iota}{2} [\partial_s h(s_{t_{\varepsilon n}}, d_{\mathcal{V}}^2(v_{t_{\varepsilon n}}, \delta_0)) + \partial_t h(t_{\varepsilon n}, d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, \delta_0))] \leq H\left(v_{t_{\varepsilon n}}, -\frac{D_v d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, v_{t_{\varepsilon n}})}{\varepsilon}\right) - H\left(\mu_{t_{\varepsilon n}}, \frac{D_\mu d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, v_{t_{\varepsilon n}})}{\varepsilon}\right) \quad (50)$$

$$+ \partial_s p_{t_{\varepsilon n}}(z_{t_{\varepsilon n}}) + \partial_t p_{t_{\varepsilon n}}(z_{t_{\varepsilon n}}) + [H](1 + d_{\mathcal{V}}(v_{t_{\varepsilon n}}, \delta_0)) \|D_v p_{t_{\varepsilon n}}(z_{t_{\varepsilon n}})\|_{v_{t_{\varepsilon n}}} + [H](1 + d_{\mathcal{V}}(\mu_{t_{\varepsilon n}}, \delta_0)) \|D_\mu p_{t_{\varepsilon n}}(z_{t_{\varepsilon n}})\|_{\mu_{t_{\varepsilon n}}}. \quad (51)$$

**Estimates and conclusion** Recall from (36) that  $-\frac{1}{2} \partial_t h(t, r) \geq 4[H] e^{-4[H]T} > 0$  for all  $t > 0$  and  $r \in \mathbb{R}^+$ . Using the assumption (31) on  $H$ , we get

$$H\left(v_{t_{\varepsilon n}}, -\frac{D_v d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, v_{t_{\varepsilon n}})}{\varepsilon}\right) - H\left(\mu_{t_{\varepsilon n}}, \frac{D_\mu d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, v_{t_{\varepsilon n}})}{\varepsilon}\right) \leq 2[H] \frac{d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, v_{t_{\varepsilon n}})}{\varepsilon}.$$

Using (42) to estimate (51), we arrive at

$$4\iota [H] e^{-4[H]T} \leq 2[H] \frac{d_{\mathcal{V}}^2(\mu_{t_{\varepsilon n}}, v_{t_{\varepsilon n}})}{\varepsilon} + \omega_t(n). \quad (52)$$

Hence keeping  $\iota$  fixed, letting  $\varepsilon \rightarrow 0$  and  $n \rightarrow \infty$ , we get from (43) that  $4\iota [H] e^{-4[H]T} \leq 0$ , which is absurd. Thus  $\Gamma \leq 0$ .  $\square$

## 6 The case of Hamilton-Jacobi-Bellman equations

We now return to the case of control problems. Consider the dynamic  $\overline{\text{co}}F$  defined in (21), and let

$$H: \mathbb{T} \rightarrow \mathbb{R}, \quad H(\mu, p) := \sup_{f \in \overline{\text{co}}F[\mu]} -p(\pi^\mu(f\#\mu)). \quad (53)$$

The Hamilton-Jacobi-Bellman equation associated to (53) then writes

$$\begin{cases} -\partial_t u(t, \mu) + \sup_{f \in \overline{\text{co}}F[\mu]} -D_\mu u(t, \mu)(\pi^\mu(f\#\mu)) = 0 & (t, \mu) \in X = ]0, T[ \times \mathcal{P}_2(\mathbb{R}^d), \\ u(T, \mu) = \mathfrak{J}(\mu) & \mu \in \mathcal{P}_2(\mathbb{R}^d). \end{cases} \quad (54a)$$

$$u(T, \mu) = \mathfrak{J}(\mu) \quad \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (54b)$$

In the sequel, we verify that the control Hamiltonian (53) satisfies the assumptions of the comparison principle, and we then show that the value function is a solution of (54).

### 6.1 Properties of the control Hamiltonian

Let  $G: \mathcal{P}_2(\mathbb{R}^d) \rightrightarrows \mathcal{P}_2(\mathbb{T}\mathbb{R}^d)$  the probability vector field (in the spirit of [Pic19, Definition 2.1]) given by

$$G[\mu] := \{b\#\mu \mid b \in \overline{\text{co}}F[\mu]\}.$$

Under **[A2]**, this PVF is Lipschitz-continuous in the Hausdorff sense with respect to the application  $W_{(\mu, \nu)}$  defined in (12), with constant  $2[f]$ . Indeed, given  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$  and  $b_0 \in \overline{\text{co}}F[\mu_0]$  defined by  $b_0(x) = \int_{u \in U} f(x, \mu_0, u) d\omega(u)$ , define  $b_1 := \int_{u \in U} f(x, \mu_1, u) d\omega(u)$ . Denote  $\eta \in \Gamma_o(\mu, \nu)$ . Then

$$W_{(\mu_0, \mu_1)}^2(b_0\#\mu_0, b_1\#\mu_1) \leq \int_{(x, y) \in (\mathbb{R}^d)^2} \int_{u \in U} |f(x, \mu_0, u) - f(y, \mu_1, u)|^2 d\omega(u) d\eta(x, y) \leq 4[f]^2 d_{\mathcal{V}}^2(\mu_0, \mu_1).$$

**Remark 5** (Why using  $G$ ). *Our motivation in using a general probability vector field is to recover the connection between the metric slope and the directional derivatives. Indeed, in [AF14, Section 4.3], it is explicit that the regular tangent cone does not provide enough directions, in the sense that there exists measures  $\mu, \nu$  such that*

$$\sup_{f \in \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)} D_\mu d_{\mathcal{W}}^2(\cdot, \nu)(f\#\mu) < \sup_{\xi \in \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)} D_\mu d_{\mathcal{W}}^2(\cdot, \nu)(\xi) = |D_\mu d_{\mathcal{W}}^2(\cdot, \nu)|.$$

Here the last term is the metric slope. The strict inequality comes from the strong convexity of the squared distance of  $\mathbb{R}^d$ , which makes it sometimes more optimal to split mass than not to, as in the example  $\mu = \delta_0$  and  $\nu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$  in dimension one. Hence, although we restrain to  $\text{Tan}_\mu$  by using the theory of continuity equations to formulate the control problem, our HJB tools are formulated in the general case.

**Lemma 6** (Locally Lipschitz behavior of  $H$ ). *Assume [A2]. There exists a constant  $[H]$  such that*

$$|H(\mu, p+q) - H(\mu, p)| \leq [H] (1 + d_{\mathcal{W}}(\mu, \delta_0)) \|q\|_\mu \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d), p, q \in \mathbb{T}_\mu.$$

*Demonstration.* Using that  $|\sup_{a \in A} f(a) - \sup_{b \in A} g(b)| \leq \sup_{a \in A} |f(a) - g(a)|$ , one gets

$$|H(\mu, p+q) - H(\mu, p)| \leq \sup_{\xi \in G[\mu]} |-p(\pi^\mu \xi) - q(\pi^\mu \xi) + p(\pi^\mu \xi)| \leq \|q\|_\mu \sup_{\xi \in G[\mu]} \|\pi^\mu \xi\|_\mu.$$

Using the non-expansivity of the projection (see [Gig08, Corollary 4.37]),

$$\begin{aligned} \sup_{\xi \in G[\mu]} \|\pi^\mu \xi\|_\mu &= \sup_{f \in \text{co}F[\mu]} \|\pi^\mu(f\#\mu)\|_\mu \leq \sup_{f \in \text{co}F[\mu]} \|f\#\mu\|_\mu = \sup_{f \in \text{co}F[\mu]} \left( \int_{x \in \mathbb{R}^d} |f(x)|^2 d\mu(x) \right)^{1/2} \\ &\leq \left( |f|_{0,\infty} + [f] d_{\mathcal{W}}^2(\mu, \delta_0) \right)^{1/2} \leq \sqrt{|f|_{0,\infty} + [f]} (1 + d_{\mathcal{W}}(\mu, \delta_0)). \end{aligned}$$

We may then take  $[H] := \sqrt{|f|_{0,\infty} + [f]}$ . □

**Lemma 7** (Behavior on the squared distance). *Assume [A2]. There exists a constant  $C_H$  such that*

$$H(\mu, -aD_\mu d_{\mathcal{W}}^2(\cdot, \nu)) - H(\nu, aD_\nu d_{\mathcal{W}}^2(\mu, \cdot)) \leq 2aC_H d_{\mathcal{W}}^2(\mu, \nu) \quad \forall a \geq 0, (\mu, \nu) \in \mathcal{P}_2(\mathbb{R}^d)^2.$$

*Demonstration.* As  $H$  is positively homogeneous, we may take  $a = 1$ . Let  $\xi \in G[\mu]$  and  $\zeta \in G[\nu]$ . Using the bijection between  $\exp_\mu^{-1}(\nu)$  and  $\Gamma_o(\mu, \nu)$ , the directional derivative of the squared Wasserstein distance (24) writes

$$\begin{aligned} D_\mu d_{\mathcal{W}}^2(\cdot, \nu)(\pi^\mu \xi) &= D_\mu d_{\mathcal{W}}^2(\cdot, \nu)(\xi) = \inf_{\eta \in \Gamma_o(\mu, \nu)} \inf_{\alpha \in \Gamma(\xi, \nu), (\pi_x, \pi_y)\#\alpha = \eta} \int_{(x, v, y)} \langle v, -2(y-x) \rangle d\alpha, \\ D_\nu d_{\mathcal{W}}^2(\nu, \cdot)(\pi^\nu \zeta) &= D_\nu d_{\mathcal{W}}^2(\mu, \cdot)(\zeta) = \inf_{\eta \in \Gamma_o(\mu, \nu)} \inf_{\beta \in \Gamma(\zeta, \mu), (\pi_y, \pi_x)\#\beta = \eta} \int_{(x, v, y)} \langle v, -2(y-x) \rangle d\beta. \end{aligned}$$

By disintegration, for each  $\eta \in \Gamma_o(\mu, \nu)$  and  $(\alpha, \beta) \in \Gamma(\xi, \nu) \times \Gamma(\zeta, \mu)$  such that  $(\pi_x, \pi_y)\#\alpha = (\pi_y, \pi_x)\#\beta = \eta$ , there exists at least one plan  $\omega = \omega(x, v, y, w) \in \Gamma(\xi, \zeta) \subset \mathcal{P}((\mathbb{T}\mathbb{R}^d)^2)$  such that  $(\pi_x, \pi_v, \pi_y) = \alpha$  and  $(\pi_y, \pi_w, \pi_x)\#\omega = \beta$ . Then

$$\begin{aligned} D_\mu d_{\mathcal{W}}^2(\cdot, \nu)(\pi^\mu \xi) + D_\nu d_{\mathcal{W}}^2(\mu, \cdot)(\pi^\nu \zeta) &= \inf_{\eta \in \Gamma_o(\mu, \nu)} \inf_{\omega \in \Gamma(\xi, \zeta), (\pi_x, \pi_y)\#\omega = \eta} \int_{(x, v, y, w)} \langle v, -2(y-x) \rangle + \langle w, -2(x-y) \rangle d\omega \\ &= \inf_{\eta \in \Gamma_o(\mu, \nu)} \inf_{\omega \in \Gamma(\xi, \zeta), (\pi_x, \pi_y)\#\omega = \eta} 2 \int_{(x, v, y, w)} \langle v-w, x-y \rangle d\omega \\ &\leq \inf_{\eta \in \Gamma_o(\mu, \nu)} \inf_{\omega \in \Gamma(\xi, \zeta), (\pi_x, \pi_y)\#\omega = \eta} 2 \sqrt{\int_{(x, v, y, w)} |v-w|^2 d\omega} \int_{(x, y)} |x-y|^2 d\eta \\ &= 2W_{(\mu, \nu)}(\xi, \zeta) d_{\mathcal{W}}(\mu, \nu). \end{aligned}$$

Hence

$$\begin{aligned} H(\mu, -D_\mu d_{\mathcal{W}}^2(\cdot, \nu)) - H(\nu, D_\nu d_{\mathcal{W}}^2(\mu, \cdot)) &= \sup_{\xi \in G[\mu]} D_\mu d_{\mathcal{W}}^2(\cdot, \nu)(\pi^\mu \xi) - \sup_{\zeta \in G[\nu]} -D_\nu d_{\mathcal{W}}^2(\mu, \cdot)(\pi^\nu \zeta) \\ &= \sup_{\xi \in G[\mu]} \inf_{\zeta \in G[\nu]} D_\mu d_{\mathcal{W}}^2(\cdot, \nu)(\xi) + D_\nu d_{\mathcal{W}}^2(\mu, \cdot)(\zeta) \\ &\leq 2d_{\mathcal{W}}(\mu, \nu) \sup_{\xi \in G[\mu]} \inf_{\zeta \in G[\nu]} W_{(\mu, \nu)}(\xi, \zeta) \\ &\leq 4[f] d_{\mathcal{W}}^2(\mu, \nu) \end{aligned}$$

by the Lipschitz-continuity in  $W_{(\mu, \nu)}$  of  $G$ . Taking  $C_H = 2[f]$  proves the claim. □

## 6.2 Characterization of the solution

**THEOREM 4** (COMPLETE CHARACTERIZATION OF (54)). Assume [A1], [A2] and [A3]. The value function is the unique viscosity solution in the sense of Definition 9 of the Hamilton-Jacobi-Bellman equation (54).

*Demonstration.* Let us show that it is a viscosity solution of (54). By Lemma 3,  $V$  is locally uniformly continuous, hence simultaneously luusc and lulsc. By definition,  $V(T, \cdot) = \mathfrak{J}$ , so that we only have to verify inequalities (27) and (28).

**Subsolution inequality** Let  $\varphi \in \mathcal{T}_+$  such that  $V - \varphi$  reaches a maximum at  $(t, \mu) \in X$ , and  $[\varphi]$  a local Lipschitz constant in a ball of radius  $\max(MT, [f](|f|_{0,\infty} + d_{\mathcal{W}}(v, \delta_0)))$ , where  $M$  is the constant of Theorem 1. For each  $\varpi \in \mathcal{P}(U)$ , the constant measure  $\omega \in L^0([t, T]; \mathcal{P}(U))$  given by  $\omega(s) \equiv \varpi$  generates a smooth solution  $(\mu_s^{t,v,\omega})_{s \in [t, T]}$ , in the sense that

$$\lim_{s \searrow t} \frac{d_{\mathcal{W}}(\mu_s^{t,v,\omega}, (id + (s-t)F_{\varpi}[v])\#v)}{s} = 0.$$

This is easily seen by recalling that  $\mu_s^{t,v,\omega}$  is then the pushforward of  $v$  by the flow of an autonomous ODE associated to a globally Lipschitz dynamic. Consequently, along the curve  $h \mapsto \hat{\mu}_{t+h}^{t,v,\omega} := (id + hF_{\varpi})\#v$ , the DPP (22) gives

$$\frac{\varphi(t+h, \hat{\mu}_{t+h}^{t,v,\omega}) - \varphi(t, v)}{h} \geq \frac{\varphi(t+h, \hat{\mu}_{t+h}^{t,v,\omega}) - \varphi(t+h, \mu_{t+h}^{t,v,\omega})}{h} + \frac{V(t+h, \hat{\mu}_{t+h}^{t,v,\omega}) - V(t, \mu)}{h} \geq -[\varphi] \frac{d_{\mathcal{W}}(\mu_{t+h}^{t,v,\omega}, \hat{\mu}_{t+h}^{t,v,\omega})}{h} + 0.$$

Multiplying by  $-1$  and using the chain rule of Lemma 4 to take the limit in  $h \searrow 0$ , we obtain

$$-\partial_t \varphi(t, v) - D_v \varphi(t, v)(F_{\varpi}[v]\#v) \leq 0.$$

As  $\varphi(t, \cdot)$  is chosen geometrically consistent, we have  $D_v \varphi(t, v)(F_{\varpi}[v]\#v) = D_v \varphi(t, v)(\pi^\mu F_{\varpi}[v]\#v)$ . As this is valid for all  $b = F_{\omega_t}[v] \in \overline{\text{co}}F[v]$ , we may take the supremum to recover the inequality (27), so that  $V$  is a subsolution.

**Supersolution inequality** Let  $\psi \in \mathcal{T}_-$  such that  $V - \psi$  reaches a minimum in  $(t, v) \in X$ . Since under [A2], the set of solutions issued from  $(t, \mu)$  is compact in the topology of uniform convergence (see [BF23, Theorem 4.5]), we may find  $\omega \in L^0([t, T]; \mathcal{P}(U))$  such that  $V(t, v) = V(t+h, \mu_{t+h}^{t,v,\omega})$  for all  $h \in [0, T-t]$ . Let  $[\psi]$  be a local Lipschitz constant of  $\psi$  as above.

Applying Lemma 2, there exist  $(h_n)_n \searrow 0$  and  $b \in \overline{\text{co}}F[v]$  such that  $d_{\mathcal{W}}(\mu_{h_n}^{t,v,\omega}, \hat{\mu}_{h_n}) = o(h_n)$ , where  $\hat{\mu}_{h_n} := (id + h_n b)\#v$ . Then

$$-[\psi] \frac{d_{\mathcal{W}}(\mu_{h_n}^{t,v,\omega}, \hat{\mu}_{h_n})}{h_n} + \frac{\psi(t+h_n, \hat{\mu}_{h_n}) - \psi(t, v)}{h_n} \leq \frac{\psi(t+h_n, \mu_{h_n}^{t,v,\omega}) - \psi(t, v)}{h_n} \leq \frac{V(t+h_n, \mu_{h_n}^{t,v,\omega}) - V(t, v)}{h_n} = 0.$$

Taking the limit in  $n \rightarrow \infty$  and using again the chain rule of Lemma 4, we get  $\partial_t \psi(t, v) + D_v \psi(b) \leq 0$ . Multiplying by  $-1$  and taking the supremum over  $b \in \overline{\text{co}}F[v]$ , we obtain

$$-\partial_t \psi(t, v) + \sup_{b \in \overline{\text{co}}F[v]} -D_v \psi(b) = -\partial_t \psi(t, v) + \sup_{b \in \overline{\text{co}}F[v]} -D_v \psi(\pi^\mu b) = -\partial_t \psi(t, v) + H(v, D_v \psi(t, v)) \geq 0.$$

To conclude, assume that there exists another viscosity solution  $W : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  of (54). By Lemmata 6 and 7, the control Hamiltonian defined in (53) satisfies the assumption [A4]. Applying Theorem 3 to the couples  $(V, W)$  and  $(W, V)$ , we have  $V \geq W$  and  $W \geq V$  pointwise over  $]0, T[ \times \mathcal{P}_2(\mathbb{R}^d)$ , so that they coincide. As both are continuous, the equality extends to  $t = 0$ , and the solution is unique.  $\square$

# Appendices

## Appendix A Details on locally uniform upper semicontinuity

Recall that a locally bounded map  $u : Y \rightarrow \mathbb{R}$  of a complete metric space  $(Y, d_Y)$  is luusc if for any decreasing family of closed sets  $(B_n)_n$  such that  $B = \bigcap_n B_n \neq \emptyset$  and  $\sup_{x \in B_n} \inf_{y \in B} d_Y(x, y) \rightarrow_n 0$ , there holds

$$\lim_{n \rightarrow \infty} \sup_{y \in B_n} u(y) \leq \sup_{x \in B} u(x). \quad (55)$$

It turns out that we have the following.

**Lemma 8** (Link with other notions of upper semicontinuity). *The following holds.*

1. *The condition (55) is strictly weaker than continuity.*
2. *Let  $\mathcal{S}$  be the set of nonempty closed and bounded subsets of  $Y$ . Then (55) is equivalent to the upper semicontinuity of the set function  $U : \mathcal{S} \rightarrow \sup_{x \in B} u(x)$  in the Hausdorff topology.*
3. *The applications that are luusc and lulsc are exactly the locally uniformly continuous applications.*
4. *In general, the condition (55) is strictly stronger than upper semicontinuity. However both definitions coincide in locally compact spaces.*
5. *In  $Y = \mathcal{P}_2(\mathbb{R}^d)$ , there is no comparison with upper semicontinuity in the narrow topology.*

*Demonstration.* Point 1 is easily seen with the luusc map  $u := \mathbb{1}_{\{o\}}$ , where  $o \in Y$  is some fixed point. Consider the notations of Point 2. If  $U$  is locally upper semicontinuous, then (55) directly stands. On the other hand, assume that  $u$  is luusc. Consider a sequence of nonempty closed sets  $A_n \subset Y$  that converge in the Hausdorff distance, that is, there exists  $A \subset \mathcal{S}$  such that

$$d_{H,Y}(A_n, A) := \max \left( \sup_{x \in A_n} \inf_{y \in A} d_Y(x, y), \sup_{y \in A} \inf_{x \in A_n} d_Y(x, y) \right) \xrightarrow{n \rightarrow \infty} 0.$$

As  $A$  is bounded, the sequence  $(A_n)_n$  is contained in some bounded set. Consider  $B_n := \overline{\bigcup_{m \geq n} A_m}$ . Then  $(B_n)_n$  is a family of closed nonempty sets, uniformly bounded, with nonempty intersection equal to  $A$  and such that

$$d_{H,Y}(B_n, A) = \sup_{x \in B_n} \inf_{y \in A} d_Y(x, y) = \sup_{m \geq n} \sup_{x \in A_n} \inf_{y \in A} d_Y(x, y) = \sup_{m \geq n} d_{H,Y}(A_m, A) \xrightarrow{n \rightarrow \infty} 0.$$

Hence

$$\overline{\lim}_{n \rightarrow \infty} U(A_n) = \lim_{n \rightarrow \infty} \sup_{m \geq n} \sup_{y \in A_m} u(y) \leq \lim_{n \rightarrow \infty} \sup_{y \in B_n} u(y) \leq \sup_{x \in A} u(x) = U(A),$$

and  $U$  is locally upper semicontinuous, proving 2.

We turn to Point 3. If  $u$  is locally uniformly continuous, let  $m_u : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  be a local modulus of continuity in a ball containing all the  $B_n$ . If  $(y_n)_n$  is a maximizing sequence with  $y_n \in B_n$ , there exists  $x_n \in B$  such that  $d_Y(y_n, x_n) \leq \varepsilon_n \rightarrow_n 0$ , and

$$\lim_{n \rightarrow \infty} \sup_{y \in B_n} u(y) = \lim_{n \rightarrow \infty} u(y_n) \leq \overline{\lim}_{n \rightarrow \infty} u(x_n) + m(d_Y(y_n, x_n)) \leq \sup_{x \in B} u(x) + 0.$$

Thus  $u$  satisfies (55). On the other hand, if a locally bounded map  $u$  is both luusc and lulsc, then  $(x, y) \mapsto s(u(x) - u(y))$  is luusc for  $s \in \{-1, 1\}$ . The definition of luusc implies then that for some fixed  $o \in Y$  and all  $R > 0$ ,

$$\limsup_{r \rightarrow 0} \left\{ |u(x) - u(y)| \mid x, y \in \overline{\mathcal{B}}(o, R), d(x, y) \leq r \right\} \leq \max_{s \in \{-1, 1\}} \lim_{n \rightarrow \infty} \sup_{\substack{(x, y) \in \overline{\mathcal{B}}(o, R)^2 \\ d(x, y) \leq 1/n}} s(u(x) - u(y)) \leq 0.$$

Any continuous modulus superior to  $r \mapsto \sup \left\{ |u(x) - u(y)| \mid x, y \in \overline{\mathcal{B}}(o, R), d(x, y) \leq r \right\}$  furnishes a local modulus of continuity.

Points 4 and 5 will use similar counterexamples. Notice first that taking  $B_n = \overline{\mathcal{B}}(x, 1/n)$  in (55) for each  $x \in \mathbb{R}^d$ , we see that luusc always imply usc. On the other hand, if  $Y$  is locally compact, let  $u : Y \rightarrow \mathbb{R}$  be upper semicontinuous. Then any maximizing sequence  $(y_n)_n$  with  $y_n \in B_n$  contains a converging subsequence, whose limit belongs to  $\overline{B} = B$  owing to the uniform approximation of  $B$  by  $B_n$ , and (55) is satisfied.

Stays to exhibit an usc map that is not luusc. By now, we take  $Y = \mathcal{P}_2(\mathbb{R}^d)$ , and we define

$$G_r := \left\{ \mu_\alpha := \left( 1 - \frac{1}{\alpha^2} \right) \delta_0 + \frac{1}{\alpha^2} \delta_{(\alpha, 0, \dots, 0)} \mid \alpha \geq r \right\}, \quad \text{and} \quad H := \left\{ \nu_\beta := \left( 1 - \frac{1}{\beta} \right) \mu_\beta + \frac{1}{\beta} \delta_{(0, 1, 0, \dots, 0)} \mid \beta \geq 1 \right\}.$$

We have that for all  $r \geq 1$ , the sets  $G_r, H$  are nonempty, disjoint and included in  $\overline{\mathcal{B}}(\delta_0, 1)$ . Moreover,  $G_r$  and  $H$  are closed in  $\mathcal{P}_2(\mathbb{R}^d)$  for each  $r \geq 1$ : indeed, let  $(\mu_{\alpha_i})_i \subset G_r$  be a Cauchy sequence. If  $(\alpha_i)_i$  is unbounded, then  $(\mu_{\alpha_i})_i$  should converge to its narrow limit  $\delta_0$ : but this is absurd since  $d_{\mathcal{V}}(\mu_\alpha, \delta_0) = 1$  for all  $\alpha > 0$ . Thus  $(\alpha_i)_i$  is bounded by some constant  $C > 0$ , and some (non relabeled) subsequence converges towards some  $\bar{\alpha} \geq r$ . Computing

$$d_{\mathcal{V}}(\mu_{\alpha_n}, \mu_{\bar{\alpha}}) = \sqrt{2 \left( 1 - \frac{\alpha_n \wedge \bar{\alpha}}{\alpha_n \vee \bar{\alpha}} \right)},$$

we see that  $\mu_{\alpha_i} \rightarrow_i \mu_\alpha \in G_r$ . Let us note that here, we proved that the sets  $\{\mu_\alpha \mid \alpha \in [r, C]\}$  are compact in  $\mathcal{P}_2(\mathbb{R}^d)$ . By a similar argument, the set  $H$  is closed as well.

Consider  $u : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  given by  $u(\mu) = \mathbb{1}_{G_1}$ . As the indicator of a closed set,  $u$  is upper semicontinuous. We claim that it is not locally uniformly upper semicontinuous: indeed, define

$$B_n := \left\{ \mu \in \overline{\mathcal{B}}(\delta_0, 1) \mid |\text{Bary}_{\mathbb{R}^d}(\mu)| := \left| \int_{x \in \mathbb{R}^d} x d\mu(x) \right| \leq \frac{1}{n} \right\}.$$

It is easily verified that  $\mu \mapsto |\text{Bary}_{\mathbb{R}^d}(\mu)|$  is continuous, so that  $(B_n)_n$  is a globally bounded decreasing sequence of closed sets, whose intersection  $B = |\text{Bary}_{\mathbb{R}^d}(\cdot)|^{-1}(\{0\}) \cap \overline{\mathcal{B}}(\delta_0, 1)$  is nonempty, and such that

$$\begin{aligned} \sup_{\mu \in B_n} \inf_{\nu \in B} d_{\mathcal{W}}(\mu, \nu) &\leq \sup_{\mu \in B_n} d_{\mathcal{W}}\left(\mu, \left(\frac{n}{\sqrt{1+n^2}}(id - \text{Bary}_{\mathbb{R}^d}(\mu))\right) \# \mu\right) \\ &\leq \sup_{\mu \in B_n} \sqrt{\left(1 - \frac{n}{\sqrt{1+n^2}}\right)^2 d_{\mathcal{W}}^2(\mu, \delta_0) + \frac{n^2}{1+n^2} |\text{Bary}_{\mathbb{R}^d}(\mu)|^2} \\ &\leq \sqrt{\left(1 - \frac{n}{\sqrt{1+n^2}}\right)^2 + \frac{1}{1+n^2}} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

As  $|\text{Bary}_{\mathbb{R}^d}(\mu_\alpha)| = 1/\alpha$ , for any  $n$ , the intersection  $B_n \cap G_1$  is nonempty. However  $B \cap G_1 = \emptyset$ , so that

$$\lim_{n \rightarrow \infty} \sup_{\mu \in B_n} u(\mu) \geq \lim_{n \rightarrow \infty} u(\mu_n) = 1, \quad \sup_{x \in B} u(\mu) = 0.$$

Hence  $u$  is not luusc, proving Point 4.

We finally turn to Point 5. Let us build an application that is bounded and narrowly upper semicontinuous, but not locally uniformly upper semicontinuous. The strategy is similar to that of Point 4, with the additional requirement to work with narrowly closed sets. Let us show that the narrow closure of  $G_1$  is  $G_1 \cup \{\delta_0\}$ : indeed,  $(\mu_{\alpha_i})_i \subset G_1$  be a narrowly converging sequence. If  $(\alpha_i)_i$  is unbounded, the narrow limit is  $\delta_0$ . If  $(\alpha_i)_i$  is bounded, then we showed in Point 4 that  $\{\mu_\alpha \mid \alpha \in [1, C]\}$  is compact in  $\mathcal{P}_2(\mathbb{R}^d)$ , thus narrowly compact, and the narrow limit stays in  $G_1$ .

Consider the bounded and narrowly upper semicontinuous function  $u := \mathbb{1}_{\{\delta_0\} \cup G_1}$ . To show that  $u$  is not luusc, we consider the family of sets

$$B_n := H \cup G_n.$$

We immediately have that  $(B_n)_n$  is a decreasing family of nonempty closed sets, whose intersection is  $H$ , and all contained in the Wasserstein unit ball centered in  $\delta_0$ . Moreover,

$$\sup_{\mu \in B_n} \inf_{\nu \in B} d_{\mathcal{W}}(\mu, \nu) = \sup_{\alpha \geq n} \inf_{\nu \in H} d_{\mathcal{W}}(\mu_\alpha, \nu) \leq \sup_{\alpha \geq n} d_{\mathcal{W}}\left(\mu_\alpha, \left(1 - \frac{1}{\alpha}\right)\mu_\alpha + \frac{1}{\alpha}\delta_{(0,1,0,\dots,0)}\right) \leq \sup_{\alpha \geq n} \frac{1}{\alpha} = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0.$$

For each  $n$ , we have  $B_n \cap \{[\delta_0] \cup G_1\} = G_n \neq \emptyset$ . However,  $H \cap [\delta_0] \cup G_1 = \emptyset$ . Then

$$\lim_{n \rightarrow \infty} \sup_{\mu \in B_n} u(\mu) = 1 \quad \text{but} \quad \sup_{\mu \in H} u(\mu) = 0.$$

Hence  $u$  is not locally uniformly upper semicontinuous. The application  $\mu \mapsto d_{\mathcal{W}}^2(\mu, \delta_0)$  furnishes an example of map that is luusc, since locally uniformly continuous, but not narrowly upper semicontinuous. Consequently, there is no hierarchy between narrow upper semicontinuity and Definition 7.  $\square$

## Appendix B Perturbed optimization

**Lemma 9** (Perturbed optimization). *Denote  $Y = \overline{X}^2 = [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ , endowed with the distance*

$$d_Y^2((t, \mu, s, \nu), (\bar{t}, \bar{\mu}, \bar{s}, \bar{\nu})) := |t - \bar{t}|^2 + d_{\mathcal{W}}^2(\mu, \bar{\mu}) + |s - \bar{s}|^2 + d_{\mathcal{W}}^2(\nu, \bar{\nu}).$$

*Let  $\Phi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$  be upper semicontinuous, proper and upper bounded,  $z^0 \in Y$  be fixed such that  $A := \sup \Phi - \Phi(z_0) < \infty$ , and assume that there exists  $R > 0$  such that*

$$\{z \in Y \mid \Phi(z) \geq \Phi(z^0)\} \subset \overline{\mathcal{B}}(z^0, R). \quad (56)$$

*Hence for each  $n \in \mathbb{N}_*$ , there exists  $z_n \in Y$  and a perturbation  $p_n : Y \rightarrow \mathbb{R}^+$  such that*

1. the perturbed map  $\Phi - p_n$  reaches a global strict maximum in  $z_n$ ,

2. There exists an application  $\omega_{T,R,A} : \mathbb{N} \rightarrow \mathbb{R}^+$  such that

$$\sum_{r \in \{t,s\}} |\partial_r p_n(z_n)| + \sum_{\sigma \in \{\mu, \nu\}} (1 + d_{\mathcal{W}}(\sigma, \delta_0)) \|D_\sigma p_n(z_n)\|_\sigma \leq \omega_{T,R,A}(n) \xrightarrow{n \rightarrow \infty} 0,$$

3. The map  $(t, \mu) \mapsto p_n(t, \mu, s_n, \nu_n)$  belongs to  $\mathcal{F}_+$ , and  $(s, \nu) \mapsto -p_n(t_n, \mu_n, s, \nu)$  belongs to  $\mathcal{F}_-$ ,

4. There holds  $\sup \Phi \leq \Phi(z_n) + C_{R,A}(n)$ , where  $C_{R,A}$  is decreasing towards 0 when  $n \rightarrow \infty$ .

*Demonstration.* The metric space  $(Y, d_Y)$  is complete, and  $\Phi$  satisfies all the assumptions of the Ekeland-Borwein-Preiss-Zhu theorem [BZ05, Theorem 2.5.2]. We consider the gauge-type function  $d_Y^2$ , and the choice of ponderation

$$\alpha_{n,m} := \frac{1}{n2^{m+1}}, \quad \text{so that} \quad \alpha_{n,0} = \frac{1}{2n} \quad \text{and} \quad \sum_{m \in \mathbb{N}} \alpha_{n,m} = \frac{1}{n}.$$

Applying Ekeland-Borwein-Preiss-Zhu, we get the existence of some  $z_n \in Y$  and a sequence  $(z_{n,m})_{m \in \mathbb{N}} \subset Y$  such that

$$\left\{ \begin{array}{ll} d_Y^2(z_0, z_n) \leq \frac{A}{\alpha_0}, & d_Y^2(z_{n,m}, z_n) \leq \frac{A}{2^m \alpha_0} & \forall m \in \mathbb{N}, & (57a) \\ \Phi(z_n) \geq \Phi(z^0) + \sum_{m \in \mathbb{N}} \alpha_{n,m} d_Y^2(z^0, z_{n,m}), & & & (57b) \\ \Phi(z_n) - \sum_{m \in \mathbb{N}} \alpha_{n,m} d_Y^2(z_n, z_{n,m}) > \Phi(z) - \sum_{m \in \mathbb{N}} \alpha_{n,m} d_Y^2(z, z_{n,m}) & & \forall z \neq z_n. & (57c) \end{array} \right.$$

Define  $p_n : z \mapsto \sum_{m \in \mathbb{N}} \alpha_{n,m} d_Y^2(z, z_{n,m}) \geq 0$ . Then using (57a),

$$\begin{aligned} p_n(z) &\leq \sum_{m \in \mathbb{N}} \alpha_{n,m} d_Y^2(z, z_{n,m}) \leq 2 \sum_{m \in \mathbb{N}} \alpha_{n,m} (d_Y^2(z, z_n) + d_Y^2(z_n, z_{n,m})) \\ &\leq 2 \frac{d_Y^2(z, z_n)}{n} + \frac{A}{n\alpha_0} \sum_{m=0}^{\infty} 4^{-m} = 2 \frac{d_Y^2(z, z_n)}{n} + \frac{8A}{3} < \infty. \end{aligned}$$

Hence the application  $p_n$  is well-defined from  $Y$  to  $\mathbb{R}^+$ . By (57c),  $\Phi - p_n$  reaches a global strict maximum in  $z_n$ .

We turn to Points 2 and 3. The application  $p_n(\cdot, \mu, s, \nu)$  is of the form  $c + \sum_{m \in \mathbb{N}} 2^{-m-1} \frac{|t-t_{n,m}|^2}{n}$ , over a bounded interval, so uniformly convergent. By direct computation, its derivative is Lipschitz in  $[0, T]$  with constant  $2/n$ , and

$$|\partial_t p_n(t, \mu_n, s_n, \nu_n)| \leq \frac{1}{n} \sum_{m \in \mathbb{N}} 2^{-m} |t - t_{n,m}| \leq \frac{2T}{n} \quad \forall t \in [0, T]. \quad (58)$$

As  $(t, \mu) \mapsto p_n(t, \mu, s_n, \nu_n)$  writes as a sum of time and measure contributions, its derivative with respect to  $t$  is Lipschitz in the whole domain  $\bar{X}$ . Moreover, by (57b),  $z_n \in \{\Phi \geq \Phi(z^0)\}$ . As  $\mu_{n,m} \rightarrow_m \mu_n$ , the sequence  $(\mu_{n,m})_m$  stays in a bounded set of  $\mathcal{P}_2(\mathbb{R}^d)$ , and the partial function  $p_n(t, \cdot, s, \nu)$  is uniformly convergent for each  $n$ . Using the semiconcavity of  $d_{\mathcal{W}}^2(\cdot, \mu_{n,m})$  (see [AGS05, Theorem 7.3.2]), there holds for any  $(\sigma, \xi) \in \mathbf{Tan} \mathcal{P}_2(\mathbb{R}^d)$  and  $h \in [0, 1]$  that

$$\begin{aligned} p_n(t_n, \exp_\sigma(h \cdot \xi), s_n, \nu_n) &\geq c + \sum_{m \in \mathbb{N}} \alpha_{n,m} [(1-h)d_{\mathcal{W}}^2(\sigma, \mu_{n,m}) + h d_{\mathcal{W}}^2(\exp_\sigma(\xi), \mu_{n,m}) - h(1-h)d_{\mathcal{W}}^2(\sigma, \exp_\sigma(\xi))] \\ &= c + (1-h)p_n(t_n, \sigma, s_n, \nu_n) + h p_n(t_n, \exp_\sigma(\xi), s_n, \nu_n) - \frac{h(1-h)}{n} d_{\mathcal{W}}^2(\sigma, \exp_\sigma(\xi)). \end{aligned}$$

Thus  $p_n(t_n, \cdot, s_n, \nu_n)$  is locally semiconcave. As its directional derivative is a combination of that of squared distances, it is geometrically consistent in the sense of Definition 5. To prove that  $(t, \mu) \mapsto p_n(t, \mu, s_n, \nu_n) \in \mathcal{F}_+$ , there only stays to show the local Lipschitzianity in the measure variable. By direct computation, for any  $S > 0$  and  $\mu, \sigma \in \bar{\mathcal{B}}(\delta_0, S)$ , one has

$$\begin{aligned} |p_n(t_n, \mu, s_n, \nu_n) - p_n(t_n, \sigma, s_n, \nu_n)| &\leq d_{\mathcal{W}}(\mu, \sigma) \sum_{m \in \mathbb{N}} \alpha_{n,m} (d_{\mathcal{W}}(\mu, \mu_{n,m}) + d_{\mathcal{W}}(\sigma, \mu_{n,m})) \\ &\leq d_{\mathcal{W}}(\mu, \sigma) \sum_{m \in \mathbb{N}} 2\alpha_{n,m} (S + d_{\mathcal{W}}(\delta_0, \mu_n) + d_{\mathcal{W}}(\mu_n, \mu_{n,m})) \\ &\leq d_{\mathcal{W}}(\mu, \sigma) \frac{S + \sqrt{2nA}}{n} \sum_{m \in \mathbb{N}} \frac{1 + 2^{-m/2}}{2^m}. \end{aligned}$$

Here we used (57a). Let  $R > 0$  be given by the assumption (56) such that  $d_Y(z_n, z^0) \leq R$  independantly of  $n$ . By the above, there holds

$$(1 + d_{\mathcal{W}}(\mu_n, \delta_0)) \|D_\mu p_n(z_n)\|_{\mu_n} \leq (1 + R) \frac{R + \sqrt{2nA}}{n} \sum_{m \in \mathbb{N}} \frac{1 + 2^{-m/2}}{2^m}. \quad (59)$$

Gathering (58) and (59), we obtain the application  $\omega_{T,R,A}$  that decreases in  $n^{-1/2}$ . The reasoning over  $p_n(t, \mu, \cdot, \nu)$  and  $p_n(t, \mu, s, \cdot)$  is symmetric.

Finally, notice that the supremum of  $\Phi$  over  $Y$  is the same as the supremum of  $\Phi$  over  $\overline{\mathcal{B}}(z^0, R)$ . In consequence, (57c) gives

$$\begin{aligned} \sup \Phi &\leq \Phi(z_n) + \sup_{z \in \overline{\mathcal{B}}(z^0, R)} \sum_{m \in \mathbb{N}} \alpha_{n,m} [d_Y^2(z, z_{n,m}) - d_Y^2(z_n, z_{n,m})] \\ &\leq \Phi(z_n) + \sup_{z \in \overline{\mathcal{B}}(z^0, R)} \sum_{m \in \mathbb{N}} \alpha_{n,m} [d_Y(z, z_{n,m}) + d_Y(z_n, z_{n,m})] d_Y(z, z_n) \\ &\leq \Phi(z_n) + \sum_{m \in \mathbb{N}} \alpha_{n,m} [2R + 2d_Y(z_n, z_{n,m})] 2R \leq \Phi(z_n) + \sum_{m \in \mathbb{N}} \frac{1}{2^{m+1}n} \left[ 2R + 2\sqrt{\frac{nA}{2^{m+1}}} \right] 2R. \end{aligned}$$

Hence Point 4 by choosing  $C_{R,A} : n \mapsto \frac{4R^2}{n} + \frac{2R\sqrt{A}}{\sqrt{n}(2\sqrt{2}-1)}$ .  $\square$

## Appendix C Monotonicities

**Lemma 10** (Monotonicities). *Assume (33), and that  $u, -v$  are bounded from above. Let  $\Gamma, \Gamma_\iota$  and  $\Gamma_{\iota\varepsilon}$  be defined as in (33), (38) and (39). Then*

$$\Gamma_{\iota\varepsilon} \searrow_{\varepsilon \searrow 0} \Gamma_\iota \quad \text{for all } \iota > 0, \quad \text{and} \quad \Gamma_\iota \nearrow_{\iota \searrow 0} \Gamma_0 \geq \Gamma.$$

*Demonstration.* Define the additional variables

$$\begin{aligned} \Gamma_r &:= \min \left( 1, \sup \left\{ u(t, \mu) - v(s, \nu) \mid (t, \mu, s, \nu) \in \overline{X}^2, d_X((t, \mu), (s, \nu)) \leq r \right\} \right) \\ \Gamma_{lr} &:= \sup \left\{ u(t, \mu) - v(s, \nu) - \iota \left( h(t, d_{\mathcal{W}}^2(\mu, \delta_0)) + h(s, d_{\mathcal{W}}^2(\nu, \delta_0)) \right) \mid (t, \mu, s, \nu) \in \overline{X}^2, d_X((t, \mu), (s, \nu)) \leq r \right\}, \end{aligned}$$

so that  $\Gamma = \lim_{r \rightarrow 0} \Gamma_r$  and  $\Gamma_\iota = \lim_{r \rightarrow 0} \Gamma_{lr}$ . For each fixed  $\iota > 0$ , using the growth of  $h$ , the variables  $\Gamma_{lr}$  and  $\Gamma_{\iota\varepsilon}$  are upper bounded. Restricting  $\iota, \varepsilon$  and  $r$  to  $]0, 1]$ , (33) gives us that each term is lower bounded. Moreover, we have the monotonicities

$$\Gamma_r \searrow_r, \quad \Gamma_\iota \nearrow_\iota, \quad \Gamma_{lr} \searrow_r, \quad \Gamma_{\iota\varepsilon} \searrow_\varepsilon.$$

In consequence, the limits  $\lim_r \Gamma_r$ ,  $\lim_r \Gamma_{lr}$  and  $\lim_\varepsilon \Gamma_{\iota\varepsilon}$  are finite, and  $\lim_\iota \Gamma_\iota$  exists in  $\mathbb{R} \cup \{+\infty\}$ .

Assume that  $\Gamma_0 := \lim_{\iota \searrow 0} \Gamma_\iota < \Gamma$ . Then there exists  $r_0$  and  $\alpha > 0$  sufficiently small so that  $\Gamma_\iota \leq \Gamma_{lr} \leq \Gamma - \alpha \leq \Gamma_r - \alpha$  for all  $0 < r \leq r_0$  and  $\iota$ . Consequently, for some  $\frac{\alpha}{2}$ -optimal point  $z_\alpha = (t_\alpha, \mu_\alpha, s_\alpha, \nu_\alpha)$  for the definition of  $\Gamma_r$ , we have (independantly of  $\iota$ )

$$u(t_\alpha, \mu_\alpha) - v(s_\alpha, \nu_\alpha) - \iota \left( h(t_\alpha, d_{\mathcal{W}}^2(\mu_\alpha, \delta_0)) + h(s_\alpha, d_{\mathcal{W}}^2(\nu_\alpha, \delta_0)) \right) \leq u(t_\alpha, \mu_\alpha) - v(s_\alpha, \nu_\alpha) - \frac{\alpha}{2} \quad \forall \iota.$$

Letting  $\iota \searrow 0$ , we obtain a contradiction. Hence  $\Gamma_0 \geq \Gamma$ .

Assume now that for some fixed  $\iota$ ,  $\Gamma_{\iota 0} := \lim_{\varepsilon \searrow 0} \Gamma_{\iota\varepsilon} < \Gamma_\iota$ . Then for sufficiently small  $\varepsilon$ , there exists  $\alpha > 0$  such that  $\Gamma_{\iota\varepsilon} \leq \Gamma_\iota - \alpha \leq \Gamma_{lr} - \alpha$  for all  $r$ . For each  $r$ , denote  $z_{\alpha r}$  some  $\frac{\alpha}{2}$ -optimal point for the definition of  $\Gamma_{lr}$ , with  $d_X(t_{\alpha r}, \mu_{\alpha r}, s_{\alpha r}, \nu_{\alpha r}) \leq r$  by construction. There holds

$$\begin{aligned} u(t_{\alpha r}, \mu_{\alpha r}) - v(s_{\alpha r}, \nu_{\alpha r}) - \iota \left( h(t_{\alpha r}, d_{\mathcal{W}}^2(\mu_{\alpha r}, \delta_0)) + h(s_{\alpha r}, d_{\mathcal{W}}^2(\nu_{\alpha r}, \delta_0)) \right) &= \frac{d_X^2((t_{\alpha r}, \mu_{\alpha r}), (s_{\alpha r}, \nu_{\alpha r}))}{\varepsilon} \\ &\leq u(t_{\alpha r}, \mu_{\alpha r}) - v(s_{\alpha r}, \nu_{\alpha r}) - \iota \left( h(t_{\alpha r}, d_{\mathcal{W}}^2(\mu_{\alpha r}, \delta_0)) + h(s_{\alpha r}, d_{\mathcal{W}}^2(\nu_{\alpha r}, \delta_0)) \right) - \frac{\alpha}{2}. \end{aligned}$$

Hence

$$\frac{\alpha}{2} \leq \frac{d_X^2((t_{\alpha r}, \mu_{\alpha r}), (s_{\alpha r}, \nu_{\alpha r}))}{\varepsilon} \leq \frac{r^2}{\varepsilon},$$

and letting  $r \searrow 0$ , we obtain the desired contradiction.

Finally, assume that for some fixed  $\iota$ ,  $\Gamma_{\iota 0} := \lim_{\varepsilon \searrow 0} \Gamma_{\iota\varepsilon} > \Gamma_\iota$ . Then for sufficiently small  $r$ , there exists  $\alpha > 0$  such that  $\Gamma_{\iota\varepsilon} \geq \Gamma_{\iota 0} \geq \Gamma_{lr} + \alpha \geq \Gamma_\iota$  for all  $\varepsilon > 0$ . For each  $\varepsilon > 0$ , let  $z_{\varepsilon\alpha} \in X^2$  be  $\frac{\alpha}{2}$ -optimal for the definition of  $\Gamma_{\iota\varepsilon}$ . Hence

$$u(t_{\varepsilon\alpha}, \mu_{\varepsilon\alpha}) - v(s_{\varepsilon\alpha}, \nu_{\varepsilon\alpha}) - \iota \left( h(t_{\varepsilon\alpha}, d_{\mathcal{W}}^2(\mu_{\varepsilon\alpha}, \delta_0)) + h(s_{\varepsilon\alpha}, d_{\mathcal{W}}^2(\nu_{\varepsilon\alpha}, \delta_0)) \right) - \frac{d_X^2((t_{\varepsilon\alpha}, \mu_{\varepsilon\alpha}), (s_{\varepsilon\alpha}, \nu_{\varepsilon\alpha}))}{\varepsilon} \geq \Gamma_{\iota\varepsilon} - \frac{\alpha}{2} \geq \Gamma_\iota + \frac{\alpha}{2}$$

implies

$$\frac{d_X^2((t_{\varepsilon\alpha}, \mu_{\varepsilon\alpha}), (s_{\varepsilon\alpha}, \nu_{\varepsilon\alpha}))}{\varepsilon} \leq \|u\| + \|\nu\| + 0 - \Gamma_l + \frac{\alpha}{2},$$

and for sufficiently small  $\varepsilon$ , we have  $d_X((t_{\varepsilon\alpha}, \mu_{\varepsilon\alpha}), (s_{\varepsilon\alpha}, \nu_{\varepsilon\alpha})) < r$ . For this choice of parameters, we get

$$\begin{aligned} & u(t_{\varepsilon\alpha}, \mu_{\varepsilon\alpha}) - v(s_{\varepsilon\alpha}, \nu_{\varepsilon\alpha}) - l(h(t_{\varepsilon\alpha}, d_{\mathcal{W}}^2(\mu_{\varepsilon\alpha}, \delta_0)) + h(s_{\varepsilon\alpha}, d_{\mathcal{W}}^2(\nu_{\varepsilon\alpha}, \delta_0))) - \frac{d_{X_1}^2((t_{\varepsilon\alpha}, \mu_{\varepsilon\alpha}), (s_{\varepsilon\alpha}, \nu_{\varepsilon\alpha}))}{\varepsilon} \\ & \geq u(t_{\varepsilon\alpha}, \mu_{\varepsilon\alpha}) - v(s_{\varepsilon\alpha}, \nu_{\varepsilon\alpha}) - l(h(t_{\varepsilon\alpha}, d_{\mathcal{W}}^2(\mu_{\varepsilon\alpha}, \delta_0)) + h(s_{\varepsilon\alpha}, d_{\mathcal{W}}^2(\nu_{\varepsilon\alpha}, \delta_0))) + \frac{\alpha}{2}, \end{aligned}$$

a flagrant contradiction.  $\square$

## Appendix D Grönwall estimates

**Lemma 11** (Grönwall estimates). *Assume [A1] and [A2]. Let  $(\mu_s^{t,v,\omega})_{s \in [t, T]}$  denote the solution of (17) issued from  $(t, v) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$  and driven by the control  $\omega \in L^0([t, T]; \mathcal{P}(U))$ . Let  $0 \leq t \leq \bar{s} \leq s \leq T$ ,  $(v, \bar{v}) \in (\mathcal{P}_2(\mathbb{R}^d))^2$  and  $(\omega, \bar{\omega}) \in (L^0([t, T]; \mathcal{P}(U)))^2$ . Then*

$$\begin{aligned} d_{\mathcal{W}}(\mu_s^{t,v,\omega}, \mu_{\bar{s}}^{t,\bar{v},\bar{\omega}}) & \leq (s - \bar{s}) \left( [f](\bar{s} - t) \left( [f] d_{\mathcal{W}}(\delta_0, \nu) + |f|_{0,\infty} \right) e^{(\bar{s}-t)[f]} + [f] d_{\mathcal{W}}(\delta_0, \nu) + |f|_{0,\infty} \right) e^{(s-\bar{s})[f]} \\ & \quad + e^{[f](\bar{s}-t)} (1 + e^{[f](\bar{s}-t)}) \left( d_{\mathcal{W}}(\nu, \bar{\nu}) + E_{t,\bar{s},\bar{v},\omega,\bar{\omega}} \right), \end{aligned}$$

where

$$E_{t,\bar{s},\bar{v},\omega,\bar{\omega}} := \left( 1 + (\bar{s} - t) \left( |f|_{0,\infty} + [f] d_{\mathcal{W}}(\delta_0, \bar{\nu}) \right) e^{(\bar{s}-t)[f]} \right) \int_{r=t}^{\bar{s}} d_{\mathcal{W},m_{f,u}}(\omega(r), \bar{\omega}(r)) dr,$$

and

$$d_{\mathcal{W},m_{f,u}}^2(\bar{\omega}, \bar{\omega}) := \inf_{\alpha \in \Gamma(\bar{\omega}, \bar{\omega})} \int_{(u,v) \in U^2} m_{f,u}^2(|u - v|) d\alpha(u, v).$$

As particular cases, we record that

$$d_{\mathcal{W}}(\mu_{t+h}^{t,v,\omega}, \nu) \leq h \left( [f] d_{\mathcal{W}}(\delta_0, \nu) + |f|_{0,\infty} \right) e^{h[f]} \quad \text{and} \quad d_{\mathcal{W}}(\mu_T^{t,v,\omega}, \mu_T^{t,\bar{v},\bar{\omega}}) \leq e^{[f](T-t)} (1 + e^{[f](T-t)}) d_{\mathcal{W}}(\nu, \bar{\nu}).$$

*Demonstration.* Assume w.l.o.g that  $s \geq \bar{s}$ . Denote  $(\Phi_\tau^{t,v,\omega})_{\tau \in [t, s]}$  and  $(\Phi_\tau^{t,\bar{v},\bar{\omega}})_{\tau \in [t, \bar{s}]}$  the respective fluxes of the ODEs

$$\dot{y}_\tau = F_{\omega(\tau)}[\mu_\tau^{t,v,\omega}](y_\tau), \quad \dot{\bar{y}}_\tau = F_{\bar{\omega}(\tau)}[\mu_\tau^{t,\bar{v},\bar{\omega}}](\bar{y}_\tau).$$

On the one hand, for  $t \leq r \leq \tau \leq s$ ,

$$\begin{aligned} |\Phi_\tau^{t,v,\omega}(x) - \Phi_\tau^{t,\bar{v},\bar{\omega}}(x)| & \leq \int_{\theta=r}^\tau \int_{u \in U} \left| f \left( \Phi_\theta^{t,v,\omega}(x), \mu_\theta^{t,v,\omega}, u \right) \right| d\omega(\theta)(u) d\theta \\ & \leq \int_{\theta=r}^\tau \int_{u \in U} \left| f \left( \Phi_\theta^{t,v,\omega}(x), \mu_\theta^{t,v,\omega}, u \right) - f \left( \Phi_\theta^{t,\bar{v},\bar{\omega}}(x), \mu_\theta^{t,\bar{v},\bar{\omega}}, u \right) \right| + \left| f \left( \Phi_\theta^{t,v,\omega}(x), \mu_\theta^{t,v,\omega}, u \right) \right| d\omega(\theta)(u) d\theta \\ & \leq \int_{\theta=r}^\tau \left[ [f] \left| \Phi_\theta^{t,v,\omega}(x) - \Phi_\theta^{t,\bar{v},\bar{\omega}}(x) \right| + [f] \left| \Phi_\theta^{t,v,\omega}(x) \right| + |f|_{0,\infty} \right] d\theta, \end{aligned}$$

so that a Grönwall lemma yields

$$|\Phi_\tau^{t,v,\omega}(x) - \Phi_\tau^{t,\bar{v},\bar{\omega}}(x)| \leq (\tau - r) \left( [f] \left| \Phi_r^{t,v,\omega}(x) \right| + |f|_{0,\infty} \right) \exp((\tau - r)[f]). \quad (60)$$

In particular,

$$\begin{aligned} |\Phi_\tau^{t,v,\omega}(x) - \Phi_\tau^{t,\bar{v},\bar{\omega}}(x)| & \leq (\tau - r) \left( [f] \left| \Phi_r^{t,v,\omega}(x) - x \right| + [f] |x| + |f|_{0,\infty} \right) \exp((\tau - r)[f]) \\ & \leq (\tau - r) \left( [f] (r - t) \left( [f] |x| + |f|_{0,\infty} \right) \exp((r - t)[f]) + [f] |x| + |f|_{0,\infty} \right) \exp((\tau - r)[f]). \end{aligned}$$



Taking the square of each side and integrating with respect to  $v$ , we get

$$\begin{aligned} d_{\mathcal{W}}(\mu_r^{t,v,\omega}, \mu_r^{t,v,\omega}) &\leq \sqrt{\int_{x \in \mathbb{R}^d} |\Phi_r^{t,v,\omega}(x) - \Phi_r^{t,v,\omega}(x)|^2 dv(x)} \\ &\leq (\tau - r) \left( [f](r - t) \left( [f] d_{\mathcal{W}}(\delta_0, v) + |f|_{0,\infty} \right) \exp((r - t)[f]) + [f] d_{\mathcal{W}}(\delta_0, v) + |f|_{0,\infty} \right) \exp((\tau - r)[f]). \end{aligned} \quad (61)$$

On the other hand, for  $\tau \in [t, \bar{s}]$ ,

$$\begin{aligned} & \left| \Phi_r^{t,v,\omega}(x) - \Phi_r^{t,\bar{v},\bar{\omega}}(y) \right| \\ & \leq |x - y| + \int_{r=t}^{\tau} \left| \int_{u \in U} f(\Phi_r^{t,v,\omega}(x), \mu_r^{t,v,\omega}, u) d\omega(r)(u) - \int_{v \in U} f(\Phi_r^{t,\bar{v},\bar{\omega}}(y), \mu_r^{t,\bar{v},\bar{\omega}}, v) d\bar{\omega}(r)(v) \right| dr \\ & \leq |x - y| + \int_{r=t}^{\tau} \left| \int_{u \in U} \left[ f(\Phi_r^{t,v,\omega}(x), \mu_r^{t,v,\omega}, u) - f(\Phi_r^{t,\bar{v},\bar{\omega}}(y), \mu_r^{t,\bar{v},\bar{\omega}}, u) \right] d\omega(r)(u) \right| dr \\ & \quad + \int_{r=t}^{\tau} \left( 1 + |\Phi_r^{t,\bar{v},\bar{\omega}}(x)| \right) d_{\mathcal{W},m_{f,u}}(\omega(r), \bar{\omega}(r)) dr \\ & \leq |x - y| + [f] \int_{r=t}^{\tau} \left| \Phi_r^{t,v,\omega}(x) - \Phi_r^{t,\bar{v},\bar{\omega}}(y) \right| + d_{\mathcal{W}}(\mu_r^{t,v,\omega}, \mu_r^{t,\bar{v},\bar{\omega}}) dr \\ & \quad + \left( 1 + (\tau - t) \left( |f|_{0,\infty} + [f]|y| \right) e^{(\tau-t)[f]} \right) \int_{r=t}^{\tau} d_{\mathcal{W},m_{f,u}}(\omega(r), \bar{\omega}(r)) dr. \end{aligned}$$

Applying a second Grönwall lemma,

$$\left| \Phi_r^{t,v,\omega}(x) - \Phi_r^{t,\bar{v},\bar{\omega}}(y) \right| \leq \left( |x - y| + \int_{r=t}^{\tau} [f] d_{\mathcal{W}}(\mu_r^{t,v,\omega}, \mu_r^{t,\bar{v},\bar{\omega}}) dr + E_{t,\tau,y,\omega,\bar{\omega}} \right) \exp([f](\tau - t)),$$

where  $E_{t,\tau,y,\omega,\bar{\omega}} := \left( 1 + (\tau - t) \left( |f|_{0,\infty} + [f]|y| \right) e^{(\tau-t)[f]} \right) \int_{r=t}^{\tau} d_{\mathcal{W},m_{f,u}}(\omega(r), \bar{\omega}(r)) dr$ . Now, let  $\eta \in \Gamma_o(v, \bar{v})$ . The plan

$$\eta_{\bar{s}} := \left( \Phi_{\bar{s}}^{t,v,\omega}, \Phi_{\bar{s}}^{t,\bar{v},\bar{\omega}} \right) \# \eta$$

belongs to  $\Gamma(\mu_{\bar{s}}^{t,v,\omega}, \mu_{\bar{s}}^{t,\bar{v},\bar{\omega}})$ , so that

$$\begin{aligned} d_{\mathcal{W}}(\mu_{\bar{s}}^{t,v,\omega}, \mu_{\bar{s}}^{t,\bar{v},\bar{\omega}}) &\leq \sqrt{\int_{(x,y) \in (\mathbb{R}^d)^2} \left| \Phi_{\bar{s}}^{t,v,\omega}(x) - \Phi_{\bar{s}}^{t,\bar{v},\bar{\omega}}(y) \right|^2 d\eta(x,y)} \\ &\leq \exp([f](\bar{s} - t)) \sqrt{\int_{(x,y) \in (\mathbb{R}^d)^2} \left( |x - y| + \int_{r=t}^{\bar{s}} [f] d_{\mathcal{W}}(\mu_r^{t,v,\omega}, \mu_r^{t,\bar{v},\bar{\omega}}) dr + E_{t,\bar{s},y,\omega,\bar{\omega}} \right)^2 d\eta(x,y)} \\ &\leq \exp([f](\bar{s} - t)) \left( d_{\mathcal{W}}(v, \bar{v}) + \int_{r=t}^{\bar{s}} [f] d_{\mathcal{W}}(\mu_r^{t,v,\omega}, \mu_r^{t,\bar{v},\bar{\omega}}) dr + \sqrt{\int_{y \in \mathbb{R}^d} E_{t,\bar{s},y,\omega,\bar{\omega}}^2 d\bar{v}(y)} \right). \end{aligned}$$

As

$$\begin{aligned} \sqrt{\int_{y \in \mathbb{R}^d} E_{t,\bar{s},y,\omega,\bar{\omega}}^2 d\bar{v}(y)} &= \sqrt{\int_{y \in \mathbb{R}^d} \left( 1 + (\bar{s} - t) \left( |f|_{0,\infty} + [f]|y| \right) e^{(\bar{s}-t)[f]} \right)^2 \left( \int_{r=t}^{\bar{s}} d_{\mathcal{W},m_{f,u}}(\omega(r), \bar{\omega}(r)) dr \right)^2 d\bar{v}(y)} \\ &\leq \left( 1 + (\bar{s} - t) \left( |f|_{0,\infty} + [f] d_{\mathcal{W}}(\delta_0, \bar{v}) \right) e^{(\bar{s}-t)[f]} \right) \int_{r=t}^{\bar{s}} d_{\mathcal{W},m_{f,u}}(\omega(r), \bar{\omega}(r)) dr =: E_{t,\bar{s},\bar{v},\omega,\bar{\omega}} < \infty, \end{aligned}$$

we are ready to apply our third Grönwall lemma to get

$$d_{\mathcal{W}}(\mu_{\bar{s}}^{t,v,\omega}, \mu_{\bar{s}}^{t,\bar{v},\bar{\omega}}) \leq \exp([f](\bar{s} - t)) \left( d_{\mathcal{W}}(v, \bar{v}) + E_{t,\bar{s},\bar{v},\omega,\bar{\omega}} \right) \exp(\exp([f](\bar{s} - t)) [f](\bar{s} - t)). \quad (62)$$

Combining (61) and (62), we get

$$\begin{aligned} d_{\mathcal{W}}(\mu_{\bar{s}}^{t,v,\omega}, \mu_{\bar{s}}^{t,\bar{v},\bar{\omega}}) &\leq d_{\mathcal{W}}(\mu_{\bar{s}}^{t,v,\omega}, \mu_{\bar{s}}^{t,v,\omega}) + d_{\mathcal{W}}(\mu_{\bar{s}}^{t,v,\omega}, \mu_{\bar{s}}^{t,\bar{v},\bar{\omega}}) \\ &\leq (s - \bar{s}) \left( [f](\bar{s} - t) \left( [f] d_{\mathcal{W}}(\delta_0, v) + |f|_{0,\infty} \right) e^{(\bar{s}-t)[f]} + [f] d_{\mathcal{W}}(\delta_0, v) + |f|_{0,\infty} \right) e^{(s-\bar{s})[f]} \\ &\quad + e^{[f](\bar{s}-t)} \left( d_{\mathcal{W}}(v, \bar{v}) + E_{t,\bar{s},\bar{v},\omega,\bar{\omega}} \right) e^{[f](\bar{s}-t)}, \end{aligned}$$

the desired result.  $\square$

## References

- [AF14] L. Ambrosio and J. Feng. On a class of first order Hamilton–Jacobi equations in metric spaces. *Journal of Differential Equations*, April 2014.
- [AG08] L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Communications on Pure and Applied Mathematics*, 2008.
- [AGS05] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows*. Lectures in Mathematics ETH Zürich. Birkhäuser-Verlag, Basel, 2005.
- [BF21] B. Bonnet and H. Frankowska. Differential inclusions in Wasserstein spaces: The Cauchy-Lipschitz framework. *Journal of Differential Equations*, January 2021.
- [BF23] B. Bonnet and H. Frankowska. Caratheodory Theory and A Priori Estimates for Continuity Inclusions in the Space of Probability Measures, May 2023.
- [BY19] A. Bensoussan and S. C. P. Yam. Control problem on space of random variables and master equation. *ESAIM: Control, Optimisation and Calculus of Variations*, 2019.
- [BZ05] J. M. Borwein and Q. J. Zhu. *Techniques of Variational Analysis*. CMS Books in Mathematics. Springer-Verlag, New York, 2005.
- [Car13] P. Cardaliaguet. Notes on Mean Field Games, 2013.
- [CCD15] J.-F. Chassagneux, D. Crisan, and F. Delarue. A Probabilistic approach to classical solutions of the master equation for large population equilibria, April 2015.
- [CCH14] J. A. Carrillo, Y.-P. Choi, and M. Hauray. The derivation of swarming models: Mean-field limit and Wasserstein distances. In *Collective Dynamics from Bacteria to Crowds: An Excursion Through Modeling, Analysis and Simulation*, CISM International Centre for Mechanical Sciences. Springer, Vienna, 2014.
- [CD18] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I*. Probability Theory and Stochastic Modelling. Springer International Publishing, Cham, 2018.
- [CIL92] M. G. Crandall, H. Ishii, and P.-L. Lions. User’s guide to viscosity solutions of second order partial differential equations, June 1992.
- [CMP18] G. Cagnani, A. Marigonda, and B. Piccoli. Superposition Principle for Differential Inclusions. In *Large-Scale Scientific Computing*. Springer International Publishing, Cham, 2018.
- [Cor16] A. Corbetta. Multiscale crowd dynamics: physical analysis, modeling and applications, 2016.
- [Coy21] R. Coyaud. *Study of Approximations of Optimal Transport Problems and Application to Physics*. These de Doctorat, Paris Est, January 2021.
- [CP20] P. Cardaliaguet and A. Porretta. An Introduction to Mean Field Game Theory. In *Mean Field Games: Cetraro, Italy 2019*, Lecture Notes in Mathematics. Springer International Publishing, Cham, 2020.
- [CQ08] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review*, March 2008.
- [CS04] P. Cannarsa and C. Sinestrari. *Semiconcave Functions, Hamilton–Jacobi Equations, and Optimal Control*. Birkhäuser, Boston, MA, 2004.
- [DMD99] P. Del Moral and M. Doisy. Maslov Idempotent Probability Calculus, I. *Theory of Probability & Its Applications*, January 1999.
- [DS23] S. Daudin and B. Seeger. A comparison principle for semilinear Hamilton-Jacobi-Bellman equations in the Wasserstein space, August 2023.
- [Eke74] I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications*, August 1974.
- [FGŚ17] G. Fabbri, F. Gozzi, and A. Święch. *Stochastic Optimal Control in Infinite Dimension*. Probability Theory and Stochastic Modelling. Springer International Publishing, Cham, 2017.
- [FK09] J. Feng and M. Katsoulakis. A Comparison Principle for Hamilton–Jacobi Equations Related to Controlled Gradient Flows in Infinite Dimensions. *Archive for Rational Mechanics and Analysis*, May 2009.
- [GHN15] Y. Giga, N. Hamamuki, and A. Nakayasu. Eikonal equations in metric spaces. *Transactions of the American Mathematical Society*, January 2015.

- [Gig08] N. Gigli. *On the Geometry of the Space of Probability Measures Endowed with the Quadratic Optimal Transport Distance*. PhD thesis, Scuola Normale Superiore di Pisa, Pisa, 2008.
- [Gig09] N. Gigli. Second order analysis on  $(P_2(M), W_2)$ , 2009.
- [Gig11] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of  $(P_2(M), W_2)$ . *Methods and Applications of Analysis*, 2011.
- [GNT08] W. Gangbo, T. Nguyen, and A. Tudorascu. Hamilton-Jacobi Equations in the Wasserstein Space. *Methods and Applications of Analysis*, 2008.
- [GŚ14] W. Gangbo and A. Świech. Optimal transport and large number of particles. *Discrete and Continuous Dynamical Systems*, 2014.
- [GŚ15] W. Gangbo and A. Świech. Metric viscosity solutions of Hamilton–Jacobi equations depending on local slopes. *Calculus of Variations and Partial Differential Equations*, September 2015.
- [GT19] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, May 2019.
- [HK15] R. Hynd and H. K. Kim. Value functions in the Wasserstein spaces: finite time horizons. *Journal of Functional Analysis*, August 2015.
- [Ish85] H. Ishii. *Hamilton-Jacobi Equations with Discontinuous Hamiltonians on Arbitrary Open Sets*, 1985.
- [Jer22] O. Jerhaoui. *Viscosity Theory of First Order Hamilton Jacobi Equations in Some Metric Spaces*. PhD thesis, Institut Polytechnique de Paris, Paris, 2022.
- [JJZ] F. Jean, O. Jerhaoui, and H. Zidani. Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition.
- [JMQ20] C. Jimenez, A. Marigonda, and M. Quincampoix. Optimal control of multiagent systems in the Wasserstein space. *Calculus of Variations and Partial Differential Equations*, March 2020.
- [KM97] V. N. Kolokoltsov and V. P. Maslov. *Idempotent Analysis and Its Applications*. Springer Netherlands, Dordrecht, 1997.
- [Lio06] P.-L. Lions. *Jeux à champ moyen*, 2006.
- [LY95] X. Li and J. Yong. *Optimal Control Theory for Infinite Dimensional Systems*. Birkhäuser, Boston, MA, 1995.
- [MQ18] A. Marigonda and M. Quincampoix. Mayer control problem with probabilistic uncertainty on initial positions. *Journal of Differential Equations*, March 2018.
- [Ott01] F. Otto. The Geometry of Dissipative Evolution Equations: The Porous Medium Equation. *Communications in Partial Differential Equations*, January 2001.
- [Pic19] B. Piccoli. Measure Differential Equations. *Archive for Rational Mechanics and Analysis*, September 2019.
- [PT09] B. Piccoli and A. Tosin. Pedestrian flows in bounded domains with obstacles. *Continuum Mechanics and Thermodynamics*, July 2009.
- [PW18] H. Pham and X. Wei. Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM: Control, Optimisation and Calculus of Variations*, January 2018.
- [San15] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, Cham, 2015.
- [Vil09] C. Villani. *Optimal Transport*. Grundlehren Der Mathematischen Wissenschaften. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [WZ19] C. Wu and J. Zhang. Viscosity Solutions to Parabolic Master Equations and McKean-Vlasov SDEs with Closed-loop Controls, April 2019.