ESAIM: COCV 30 (2024) 91
https://doi.org/10.1051/cocv/2024081

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

VISCOSITY SOLUTIONS OF CENTRALIZED CONTROL PROBLEMS IN MEASURE SPACES

Averil Aussedat[1,*], Othmane Jerhaoui[2] and Hasnaa Zidani[1]

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.

Mathematics Subject Classification. 35F21, 35R06, 49Lxx.

Received March 4, 2024. Accepted November 7, 2024.

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 modelling physical situations where the state variable is known only up to a density of probability [1]. It also provides a convenient formalism for problems involving the motion of populations, encompassing both discrete and continuous formulations ([2, 3]; see also the survey [4]).

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 [5, 6]. 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,

Keywords and phrases: Hamilton–Jacobi, Wasserstein, viscosity solutions.

[1] INSA Rouen Normandie, Normandie Univ, LMI UR 3226, 76000 Rouen, France.
[2] Université de Rennes, INSA Rennes, CNRS, IRMAR – UMR 6625, 35000 Rennes, France.

[*] Corresponding author: averil.aussedat@insa-rouen.fr

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 [7, 8], 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 [9], leading to the construction of a natural tangent cone [10]. The emergence of mean-field games [11] emphasized the necessity for proper generalizations of solutions of partial differential equations with a measure variable, laying the ground for the extension of viscosity solutions to this setting.

The approach by test functions is very classical when working in $\mathbb{R}^d$, but entered only recently in the literature over the Wasserstein space. The reason is that the squared distance does not admit a Wasserstein gradient, or equivalently an L-derivative [12], at all measures. Recent works succeeded in finding smoothing procedures in order to provide sufficiently smooth test functions [13–15]. With this technology, [14] obtains a comparison principle for bounded and continuous semisolutions, and [13] a strong comparison principle in the particular case of a Hamiltonian involving a degenerate divergence term, thus calling for an *ad hoc* notion. A strong comparison principle is also proved with a test function approach in [16] in the presence of noise, which helps to control the regularity of the maxima.

In a close spirit, the works [17–20] consider a modification of the definition using test functions and suitable upper/lower envelopes or approximations of the Hamiltonian. [20] obtains a strong comparison principle by relying on a penalization by an energy with compact sublevels, an assumption that is removed in [17–19] with the help of the gradient flow structure of their problems. The latter series of work does apply in more general spaces than the Wasserstein space, but seems difficult to extend out of variational problems.

On the other hand, the approach by semidifferentials is more common, since it corresponds to allowing non-differentiable test functions. Here one can distinguish two tendencies. First, one can define semidifferentials as elements of $L^2_\mu(\mathbb{R}^d; \mathrm{T}\,\mathbb{R}^d)$ or Tan_μ, the common feature being that these elements are supported on graphs [12, 21–25]. In all these works, comparison principles are obtained under the assumption that at least one of the semisolutions to be compared is uniformly continuous. Secondly, one could enlarge the definition to allow for measure-valued elements in the semidifferentials, as in [26, 27]. The first of these works obtains a strong comparison principle for Eikonal-type equations by successfully drawing a link with metric viscosity, and the second one a strong comparison principle on the torus for a Hamiltonian arising in mean-field games. Let us point that the latter work seems the closest in spirit to ours, with the conceptual difference that the structure of differentiability is build with the Lions lifting procedure, and does not relate to the geodesic structure, hence preventing generalization to other classes of space.

The present paper provides a notion of viscosity solutions that relies mostly on the geodesic structure of the space, and a strong comparison principle that takes advantage of the allowance of the squared distance in the test function sets. The class of Hamiltonian under investigation allow to treat first-order control problems with sufficiently regular dynamic, although it does not seem to extend to the case where diffusion enters the continuity equations. In the Wasserstein space, this line of investigation was initiated in [28, 29], 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. Consider $\mathcal{B}_{\mathbb{R}^{\mathrm{d}}}$ the Borel $\sigma-$algebra of $\mathbb{R}^{\mathrm{d}}$, and let $g : \mathbb{R}^{\mathrm{d}} \mapsto \mathbb{R}$ be a Borel-measurable function. The notation $\#$ will be used to denote the push-forward operator on measures, defined for any Borel measure μ as

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

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

$$\mathscr{P}_2(\Omega) := \left\{\mu \in \mathscr{P}(\Omega) \;\middle|\; \int_{x\in\mathbb{R}^{\mathrm{d}}} d^2(o,x)\mathrm{d}\mu(x) < \infty\right\}.$$

For any couple $(\mu, \nu) \in \mathscr{P}_2(\Omega)^2$, we denote by $\Gamma(\mu,\nu)$ the set of probability measures on $\Omega\times\Omega$ with marginals μ and ν, 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 $\mathscr{P}_2(\Omega)$ with a metric, here chosen as the Monge–Kantorovich distance with $p = 2$ - also called $2-$*Wasserstein distance* in the literature - and defined as

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

The set of plans realizing the infimum is denoted $\Gamma_o(\mu,\nu)$. According to Theorem 4.1 in [8], 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 $\mathbb{R}^{\mathrm{d}}$ as the underlying space. Its tangent space is defined as $\mathrm{T}\,\mathbb{R}^{\mathrm{d}} = \bigcup_{x\in\mathbb{R}^{\mathrm{d}}}\{x\}\times\mathrm{T}_x\,\mathbb{R}^{\mathrm{d}}$, where $\mathrm{T}_x\,\mathbb{R}^{\mathrm{d}}$ represents the tangent space to $\mathbb{R}^{\mathrm{d}}$ at the point x. To maintain a clear distinction between points and velocities, we refrain from identifying $\mathrm{T}_x\,\mathbb{R}^{\mathrm{d}}$ with $\mathbb{R}^d$. Let $T > 0$ be a fixed final time-horizon. We will use the notation

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

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

$$\partial_s\mu_s + \mathrm{div}\,\left(f(\cdot,\mu_s,u(s))\#\mu_s\right) = 0, \quad s\in[t,T], \qquad \mu_t = \nu \tag{2.1}$$

where $\nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ is an initial configuration of the system at a time $t \in [0,T]$. Here the divergence operator is distribution-valued, and defined as

$$\mathrm{div}\,(\xi)(\varphi) := -\int_{(x,v)\in\mathrm{T}\,\mathbb{R}^{\mathrm{d}}} \langle\nabla_x\varphi, v\rangle\,\mathrm{d}\xi(x,v) \qquad \forall\varphi\in\mathcal{C}_c^\infty(\mathbb{R}^{\mathrm{d}};\mathbb{R}).$$

We denote $(\mu_s^{t,\nu,u})_{s\in[t,T]}$ the solution of (2.1), which is understood in the sense of distributions. Here, the control input $u(\cdot)$ is supposed to be a measurable function, *i.e.*

$$u(\cdot) \in L^1([t,T];U) \coloneqq \{v(\cdot) : [t,T] \mapsto U \mid v(\cdot) \text{ is Lebesgue-measurable}\},$$

where $U \subset \mathbb{R}^\kappa$ is a set of admissible controls, and $f : \mathbb{R}^{\mathrm{d}} \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U \mapsto \mathrm{T}\,\mathbb{R}^{\mathrm{d}}$ is a given controlled and measure-dependant dynamic. The study of the dynamical system (2.1) will be carried out in Section 4.1.

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

$$\text{Find } u \in L^1([t,T];U) \qquad \text{such that} \quad \mathfrak{J}\left(\mu_T^{t,\nu,u}\right) \leqslant \mathfrak{J}(\mu_T^{t,\nu,v}) \quad \forall v \in L^1([t,T];U). \tag{2.2}$$

The adjustments in the case of a Bolza problem with Lipschitz running cost are discussed in Section 6.3. The *value function* associated to this control problem is defined as

$$V : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}, \qquad V(t,\nu) \coloneqq \inf\left\{\mathfrak{J}\left(\mu_T^{t,\nu,u}\right) \;\middle|\; u \in L^1([t,T];U)\right\}. \tag{2.3}$$

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

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

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 $m : \mathbb{R}^+ \to \mathbb{R}^+$ is a modulus of continuity if it is continuous, nondecreasing and if $m(0) = 0$.

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

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

- f is Lipschitz-continuous in the space and measure variables with constant $[f]$, in the sense that for all $(x,y,\mu,\nu,u) \in \mathbb{R}^{\mathrm{d}} \times \mathbb{R}^{\mathrm{d}} \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U$,
$$|f(x,\mu,u) - f(y,\nu,u)| \leqslant [f]\left(|x-y| + d_{\mathcal{W}}(\mu,\nu)\right).$$
- there exists a modulus of continuity $m_{f,u} : \mathbb{R}^+ \to \mathbb{R}^+$ such that for all $(x,\mu,u,v) \in \mathbb{R}^{\mathrm{d}} \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U \times U$,
$$|f(x,\mu,u) - f(x,\mu,v)| \leqslant (1+|x|)m_{f,u}(|u-v|).$$
- For all $(\mu,u) \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U$, there holds $|f(0,\mu,u)| \leqslant |f|_{0,\infty}$.

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

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

where $\mathscr{B}(\delta_0,R)$ stands for the ball centred in the Dirac measure δ_0 and with radius R.

Under these assumptions, our objective is to show that the value function uniquely solves a HJB equation (in an appropriate sense). In the classical theory of HJ equations in Banach spaces, the Hamiltonian is typically defined as a function of gradients or, more generally, over the space of linear functionals. In this work, we introduce the Hamiltonian through a *metric cotangent bundle* $\mathbb{T}$, which serves as an alternative to the conventional mapping $p \mapsto \langle \nabla\phi, p\rangle$. The key idea behind the construction of $\mathbb{T}$ is to encapsulate local approximations of test functions in a manner that accommodates the geometric features of the underlying space. Unlike traditional linear structures, the metric cotangent bundle is composed of continuous and positively homogeneous mappings, which may lack linearity but remain suitable for our framework. In the context of HJB equations, linearity is not a strict requirement for establishing uniqueness and characterizing the value function. Instead, the essential interpretation of HJB equations revolves around enforcing growth conditions along the trajectories of the control problem, which relies fundamentally on one-sided derivatives rather than full gradient information.

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 [30]. Secondly, the Wasserstein space exhibits positive curvature, which proves to be unfavourable 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 upper semicontinuity, while concave maps have lower semicontinuous directional derivatives. This makes Hypothesis 3.4 in [29], which assumes some upper/lower semicontinuity of the Hamiltonian, unattainable in $\mathscr{P}_2(\mathbb{R}^d)$.

3. Preliminaries on the Wasserstein space

The space $\mathscr{P}_2(\mathbb{R}^d)$, 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 $(\mu_t)_{t\in[0,1]} \subset \mathscr{P}_2(\mathbb{R}^d)$ satisfying $d_{\mathcal{W}}(\mu_t,\mu_s) \leqslant |t-s|\, d(\mu_0,\mu_1)$ 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 $\mathscr{P}_2(\mathbb{R}^d)$

Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, 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 $\mathscr{P}_2(\mathbb{R}^d)$, it is known (see [9], Thm. 7.2.2) that constant speed geodesics coincide with trajectories of the form

$$\mu_t = ((1-t)\pi_x + t\pi_y)\#\eta \qquad \forall t \in [0,1], \quad \eta = \eta(x,y) \in \Gamma_o(\mu,\nu). \tag{3.1}$$

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 $\mathrm{T}\,\mathbb{R}^d = \bigcup_{x\in\mathbb{R}^d}\{x\}\times \mathrm{T}_x\,\mathbb{R}^d$. We will denote $(x,v) \in \mathrm{T}\,\mathbb{R}^d$ a generic tangent element, and π_x,π_v the canonical projections. Let $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^d)_\mu$ be the set of *initial velocities*

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

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

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

The sets $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^d)_\mu$ and $\Gamma(\mu,\mathscr{P}_2(\mathbb{R}^d)) = \bigcup_{\nu\in\mathscr{P}_2(\mathbb{R}^d)}\Gamma(\mu,\nu)$ are in bijection through the map $\Psi : \Gamma(\mu,\mathscr{P}_2(\mathbb{R}^d)) \mapsto \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^d)_\mu$ given by

$$\Psi(\eta) = (\pi_x, \pi_y - \pi_x)\#\eta, \qquad \Psi^{-1}(\gamma) = (\pi_x, \pi_x + \pi_v)\#\gamma. \tag{3.2}$$

The mapping Ψ is bicontinuous in the respective Wasserstein topologies.

Remark 3.1 (Manifold case)**.** Let us stress that Ψ is a bijection owing to the uniqueness of geodesics in $\mathbb{R}^{\mathrm{d}}$. 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 [31] (see in particular Def. 1.4), and in [28] 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 (3.1) is equivalently represented using (3.2) as

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

We define the set of *initial velocities of geodesics* through the identification (3.2):

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

Following the notation of [32], we will denote $\exp_\mu^{-1}(\nu)$ the set of *initial velocities of geodesics* linking μ to ν, *i.e.*

$$\exp_\mu^{-1}(\nu) \coloneqq \left\{\gamma \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_{\mu,o} \text{ with } (\pi_x + \pi_v)\#\gamma = \nu\right\}. \tag{3.3}$$

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

$$\exp_\mu^{-1}(\nu) = \left\{\gamma \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})_\mu \;\middle|\; \exp_\mu(\gamma) = \nu,\ \int_{(x,v)\in \mathrm{T}\,\mathbb{R}^{\mathrm{d}}} |v|^2\, d\gamma(x,v) = d^2_{\mathcal{W}}(\mu,\nu)\right\}.$$

This set is always nonempty, since $\Gamma_o(\mu,\nu)$ is nonempty (see for instance Thm. 1.7 of [33]). It is compact in $(\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}}), d_{\mathcal{W}})$ as the image of the compact $\Gamma_o(\mu,\nu)$ through the continuous identification (3.2). We refer to it as the set of initial velocities of geodesics issued from μ and reaching ν.

3.2. The tangent cone

In the sequel, we denote

$$\mathrm{T}^2\,\mathbb{R}^{\mathrm{d}} \coloneqq \left\{(x,v_1,v_2) \;\middle|\; x \in \mathbb{R}^{\mathrm{d}},\ v_i \in \mathrm{T}_x\,\mathbb{R}^{\mathrm{d}}\right\},$$

where $|(x,v_1,v_2) - (\overline{x},\overline{v}_1,\overline{v}_2)|^2 \coloneqq |x-\overline{x}|^2 + \sum_{i=1}^2 |v_i - \overline{v}_i|^2$. Given $\xi_1,\xi_2 \subset \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})_\mu$, we define

$$\Gamma_\mu(\xi_1,\xi_2) \coloneqq \left\{\alpha \in \mathscr{P}(\mathrm{T}^2\,\mathbb{R}^{\mathrm{d}}) \;\middle|\; (\pi_x,\pi_{v_i})\#\alpha = \xi_i,\ i \in \{1,2\}\right\}.$$

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 $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, let $W_\mu : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})_\mu \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})_\mu \to \mathbb{R}^+$ be defined by

$$W_\mu^2(\xi,\zeta) \coloneqq \inf_{\alpha\in\Gamma_\mu(\xi,\zeta)} \int_{(x,v,w)\in \mathrm{T}^2\,\mathbb{R}^{\mathrm{d}}} |v-w|^2\, \mathrm{d}\alpha(x,v,w). \tag{3.4}$$

As per [10], Theorem 4.4, W_μ is a metric over $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})_\mu$ 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_μ as (see [10], Prop. 4.2)

$$W_\mu^2(\xi,\zeta) \coloneqq \int_{x\in\mathbb{R}^{\mathrm{d}}} d^2_{\mathcal{W}}(\xi_x,\zeta_x)\, \mathrm{d}\mu(x). \tag{3.5}$$

As an useful particular case, we record that

$$W_\mu^2(\xi, 0_\mu) = \int_{(x,v)\in \mathrm{T}^2\,\mathbb{R}^{\mathrm{d}}} |v|^2 \,\mathrm{d}\xi(x,v) =: \|\xi\|_\mu^2. \tag{3.6}$$

We denote $\Gamma_{\mu,o}(\xi,\zeta)$ the subset of $\Gamma_\mu(\xi,\zeta)$ where the infimum of (3.4) is realized.

Definition 3.2 (Tangent cone [10], Def. 4.1)**.** For each $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, we define

$$\mathbf{Tan}'_\mu\,\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \coloneqq \left\{\alpha\cdot\xi \;\middle|\; \alpha\in\mathbb{R}^+, \xi\in\exp_\mu^{-1}(\sigma) \text{ for some } \sigma\in\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})\right\},$$

and $\mathbf{Tan}_\mu\,\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \coloneqq \overline{\mathbf{Tan}'_\mu\,\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})}^{W_\mu}$.

The pre-tangent cone $\mathbf{Tan}'_\mu$ is the set of velocities ξ 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 [10], Proposition 4.30, there exists a well-defined projection

$$\pi^\mu : \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu \to \mathbf{Tan}_\mu\,\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}), \qquad \pi^\mu(\gamma) = \underset{\overline{\gamma}\in\mathbf{Tan}_\mu\,\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})}{\operatorname{argmin}} W_\mu(\gamma,\overline{\gamma}). \tag{3.7}$$

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

$$d^2_{\mathcal{W},\mathrm{T}\,\mathbb{R}^{\mathrm{d}}}(\xi,\zeta) \leqslant \int_{(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})^2} |x-y|^2 + |v-w|^2 \, d[(\pi_x,\pi_v),(\pi_x,\pi_w)]\#\alpha = W_\mu^2(\xi,\zeta). \tag{3.8}$$

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

$$d^2_{\mathcal{W}}\left(\exp_\mu(t\cdot\xi), \exp_\mu(t\cdot\zeta)\right) \leqslant W_\mu^2(t\cdot\xi, t\cdot\zeta) = t^2 W_\mu^2(\xi,\zeta). \tag{3.9}$$

Let $(\mu,\nu)\in(\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}))^2$. Following [34], Definition 4.1, we define an application $W_{(\mu,\nu)} : \mathscr{P}(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu \times \mathscr{P}(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\nu \to \mathbb{R}^+$ by

$$W^2_{(\mu,\nu)}(\xi,\zeta) \coloneqq \inf\left\{ \int_{(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})^2} |v-w|^2 \,\mathrm{d}\omega(x,v,y,w) \;\middle|\; \begin{array}{c}\omega\in\Gamma(\xi,\zeta),\\ \pi_{(x,y)}\#\omega\in\Gamma_o(\mu,\nu)\end{array}\right\}. \tag{3.10}$$

The map $W_{(\mu,\nu)}$ computes the difference between ξ and ζ 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 $\xi,\zeta\in\mathscr{P}(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$. However, the application $W_{(\mu,\nu)}$ does not satisfy the triangular inequality (see [34], Rem. 4).

3.3. The metric cotangent bundle

Definition 3.3 (Directionally differentiable map)**.** We say that an application $\varphi : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ is directionally differentiable at $\mu\in\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ if for all $\xi\in\mathbf{Tan}_\mu\,\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, the limit

$$\lim_{t\searrow 0} \frac{\varphi(\exp_\mu(t\cdot\xi)) - \varphi(\mu)}{t} =: D_\mu\varphi(\xi) \tag{3.11}$$

exists. The application $D_\mu\varphi : \mathbf{Tan}_\mu\,\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ is called the differential at μ of φ.

Notice that in (3.11), we do not assume the limit to be uniform in $\xi \in \mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$. 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 $\alpha \geqslant 0$. Moreover, assume that φ is Lipschitz with constant $[\varphi]$ in some ball centred in μ. Then using (3.9),

$$\left|D_\mu\varphi(\xi) - D_\mu\varphi(\bar\xi)\right| \leqslant [\varphi] \lim_{t\searrow 0} \frac{d_{\mathcal{W}}(\exp_\mu(t\cdot\xi), \exp_\mu(t\cdot\bar\xi))}{t} = [\varphi]\, W_\mu(\xi,\bar\xi). \tag{3.12}$$

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

Definition 3.4 (Metric cotangent bundle)**.** Let

$$\mathbb{T}_\mu \coloneqq \left\{ p : \mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \;\middle|\; \begin{array}{c} p \text{ Lipschitz w.r.t. } W_\mu, \\ \text{and positively homogeneous} \end{array} \right\}, \tag{3.13}$$

and $\mathbb{T} \coloneqq \bigcup_{\mu\in\mathscr{P}_2(\mathbb{R}^d)} \{\mu\}\times\mathbb{T}_\mu$. The sets $\mathbb{T}_\mu$ are stable by the pointwise operations $(p+q)(\xi) \coloneqq p(\xi)+q(\xi)$ and $(\alpha p)(\xi) \coloneqq \alpha p(\xi)$. We endow $\mathbb{T}$ with the application

$$\|\cdot\| : \mathbb{T} \to \mathbb{R}^+, \qquad (\mu,p) \mapsto \|p\|_\mu \coloneqq \sup_{\xi\in\mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d), \|\xi\|_\mu = 1} |p(\xi)|\,. \tag{3.14}$$

The application $\|\cdot\|_\mu$ induces a norm on $\mathbb{T}_\mu$, and we recover $|p(\xi)| \leqslant \|p\|_\mu\|\xi\|_\mu$. 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 [9] 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 [35, 36] raised the theory of continuity equations and continuity inclusions in $\mathscr{P}_2(\mathbb{R}^d)$ 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 [34, 37]), although it is known that in the Lipschitz setting, solutions of MDEs and continuity equations coincide ([37], Them. 1). In this section, we first reformulate the controlled dynamical system (2.1) in order to apply the results of [36], 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 $\mathscr{P}(U)$ the set of probability measures over the compact U, which is itself a compact set when endowed with the squared Wasserstein distance. Let $L^1([t,T];\mathscr{P}(U))$ be the space of Lebesgue-measurable curves $\omega : [t,T] \to \mathscr{P}(U)$. Define

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

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

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

As $m_{f,u}$ is continuous and U compact, the right hand-side of the above inequality is continuous with respect to $d_{\mathcal{W},U}$. Consequently, the application $\omega \mapsto F_\omega[\mu]$ is continuous from $(\mathscr{P}(U), d_{\mathcal{W},U})$ to $\mathcal{C}\left(\mathbb{R}^d; \mathrm{T}\,\mathbb{R}^d\right)$ endowed with the topology of convergence over compact sets. Therefore, $[t,T] \ni s \mapsto F_{\omega(s)}[\mu]$ is Lebesgue-measurable for each $\omega \in L^1([t,T];\mathscr{P}(U))$. For some fixed $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, consider the associated continuity equation

$$\partial_s \mu_s + \mathrm{div}\,\left(F_{\omega(s)}[\mu_s]\#\mu_s\right) = 0 \quad s \in [t,T], \qquad \mu_t = \nu. \tag{4.1}$$

Combining Theorems 2.18, Proposition 2.22 and Theorem 4.2 of [36], we get the following.

Theorem 4.1 (Existence, uniqueness and representation of the solution)**.** *Assume [A1] and [A2]. For each $\omega \in L^1([t,T];\mathscr{P}(U))$, there exists an unique trajectory $(\mu_s^{t,\nu,\omega})_{s\in[t,T]} \in AC\big([t,T];\mathscr{P}_2(\mathbb{R}^d)\big)$ solution of (4.1) 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 $\forall t \leqslant s, \tau \leqslant T$,*

$$d_{\mathcal{W}}(\mu_s^{t,\nu,\omega}, \delta_0) \leqslant m\,(1 + d_{\mathcal{W}}(\nu,\delta_0))\,, \qquad d_{\mathcal{W}}(\mu_s^{t,\nu,\omega}, \mu_\tau^{t,\nu,\omega}) \leqslant M\,|\tau - s|\,.$$

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

$$\frac{\mathrm{d}}{\mathrm{d}s}\Phi_s^t(x) = F_{\omega(s)}[\mu_s^{t,\nu,\omega}](\Phi_s^t(x)), \qquad \Phi_t^t(x) = x. \tag{4.2}$$

Choosing $\omega(s) = \delta_{u(s)}$ for some $u \in L^1([t,T];U)$, Theorem 4.1 brings well-posedness of the controlled system (2.1). The pushforward representation allows to obtain various estimates directly from the underlying dynamical system. In particular, a Grönwall estimate yields that

$$\left|\Phi_s^{0,x}(x) - x\right| \leqslant s\left([f]\,|x| + |f|_{0,\infty}\right)\mathrm{e}^{[f]s}, \tag{4.3}$$

and

$$d_{\mathcal{W}}(\mu_{t+s}^{t,\nu,\omega}, \nu) \leqslant s\left([f]\,d_{\mathcal{W}}(\nu,\delta_0) + |f|_{0,\infty}\right)\mathrm{e}^{[f]s} \tag{4.4}$$

for all $\omega \in L^1([t,T];\mathscr{P}(U))$, $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, $x \in \mathbb{R}^d$ and $0 \leqslant t, s, t+s \leqslant T$ (see Appendix E).Define the reachable sets from (t,ν) at time T by the flow of (4.1) as

$$\mathcal{R}_T^{t,\nu} \coloneqq \left\{\mu_T^{t,\nu,\omega} \;\middle|\; \omega \in L^1([t,T];\mathscr{P}(U))\right\}. \tag{4.5}$$

Lemma 4.2 (Lipschitz-continuity of the reachable sets)**.** *Assume [A1] and [A2]. There exists a constant $[\mathcal{R}]$ depending only on f and T such that for all $t \in [0,T]$ and $\nu, \overline{\nu} \in \mathscr{P}_2(\mathbb{R}^d)$,*

$$\max\left(\sup_{\mu\in\mathcal{R}_T^{t,\nu}}\inf_{\overline{\mu}\in\mathcal{R}_T^{t,\overline{\nu}}} d_{\mathcal{W}}(\mu,\overline{\mu}),\ \sup_{\overline{\mu}\in\mathcal{R}_T^{t,\overline{\nu}}}\inf_{\mu\in\mathcal{R}_T^{t,\nu}} d_{\mathcal{W}}(\mu,\overline{\mu})\right) \leqslant [\mathcal{R}]d_{\mathcal{W}}(\nu,\overline{\nu}).$$

Proof. Using chained Grönwall estimates (see Appendix E), we have that two solutions $s \mapsto \mu_s^{t,\nu,\omega}$ and $s \mapsto \mu_s^{t,\overline{\nu},\omega}$ associated to the same control $\omega \in L^1([t,T]; \mathscr{P}(U))$ satisfy

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

The claim follows by approximating each $\mu = \mu_T^{t,\nu,\omega} \in \mathcal{R}_T^{t,\nu}$ by $\mu_T^{t,\overline{\nu},\omega}$, and defining $[\mathcal{R}] \coloneqq \mathrm{e}^{[f]T\left(\mathrm{e}^{[f]T}+1\right)}$. □

Lemma 4.3 (Smooth case)**.** *Let $0 \leqslant t \leqslant T$ and $\omega \in L^1([t,T]; \mathscr{P}(U))$ be a constant control $\omega(s) \equiv \varpi \in \mathscr{P}(U)$. Then the unique solution $(\mu_s^{t,\nu,\omega})_{s\in[t,T]}$ satisfy*

$$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,\omega}, \exp_\mu(h\cdot F_\varpi[\nu])\right)}{h} = 0.$$

The proof relies only on Grönwall estimates, and is deferred to Appendix A.

4.1.2. Convex relaxation of the dynamic

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

$$\overline{\mathrm{co}}F[\mu] \coloneqq \{F_\omega[\mu] \mid \omega \in \mathscr{P}(U)\} = \left\{\int_{u\in U} f(\cdot,\mu,u)\mathrm{d}\omega(u) \;\middle|\; \omega \in \mathscr{P}(U)\right\}. \tag{4.6}$$

For each μ, the set $\overline{\mathrm{co}}F[\mu] \subset \mathcal{C}(\mathbb{R}^{\mathrm{d}}; \mathrm{T}\,\mathbb{R}^{\mathrm{d}})$ is closed in the topology of uniform convergence on compact sets; indeed, as this topology is metrizable, it suffices to show that $(b_n)_n \subset \overline{\mathrm{co}}F[\mu]$ and $b_n \to_n b$ uniformly over the compacts implies $b \in \overline{\mathrm{co}}F[\mu]$. As $\mathscr{P}(U)$ is compact ([9], Prop. 7.1.5), some subsequence of ω_n converges to a measure $\omega \in \mathscr{P}(U)$, and [A2] yields that for each $R > 0$,

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

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

Remark 4.4 (Link with the closed convex envelope of [36])**.** In our case, $\overline{\mathrm{co}}F[\mu]$ is equal to the closure of $\mathcal{M}$ in the topology of uniform convergence over compact sets, where

$$\mathcal{M} \coloneqq \bigcup_{N\in\mathbb{N}}\left\{\sum_{i=1}^N \alpha_i f(\cdot,\mu,u_i) \;\middle|\; \sum_{i=1}^N \alpha_i = 1, \alpha_i \geqslant 0, u_i \in U\right\}.$$

Indeed, we trivially have $\mathcal{M} \subset \overline{\mathrm{co}}F$ through the representation $\omega \coloneqq \sum_{i=1}^N \alpha_i\delta_{u_i}$. On the other hand, let $\overline{b} = F_\omega[\mu] \in \overline{\mathrm{co}}F$ for some $\omega \subset \mathscr{P}(U)$. 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 $R > 0$

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

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 $\left\{(\mu_s^{t,\nu,\omega})_{s\in[t,T]} \;\middle|\; \omega \in L^1([t,T];\mathscr{P}(U))\right\}$ coincides with the set of solutions of the continuity inclusion

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

Consequently, we have the following.

Theorem 4.5 (Relaxation (Thms. 4.5 and 4.6 of [36]))**.** *Assume [A1] and [A2], and let $\nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. The set*

$$\left\{(\mu_s^{t,\nu,\omega})_{s\in[t,T]} \;\middle|\; \omega \in L^1([t,T];\mathscr{P}(U))\right\} \subset AC([t,T];\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}))$$

is compact in the topology of the uniform convergence, and is the closure in this topology of the set of trajectories of (2.1), namely

$$\left\{(\mu_s^{t,\nu,u})_{s\in[t,T]} \;\middle|\; u \in L^1([t,T];U)\right\}.$$

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 (4.1), by approximating the said curve only along some sequence.

Lemma 4.6 (Right linear approximation)**.** *Assume [A1] and [A2]. Let $\overline{s} > 0$, $(\mu_s)_{s\in[0,\overline{s}]}$ be the solution of (4.1) for some control $\omega \in L^1([0,\overline{s}];\mathscr{P}(U))$. Then there exists $b \in \overline{\text{co}}F[\mu_0]$ and a sequence $(s_n) \searrow 0$ such that*

$$\lim_{n\to\infty} \frac{d_{\mathcal{W}}\left(\mu_{s_n}, m_{s_n}\#\mu_0\right)}{s_n} = 0, \quad \textit{where} \quad m_s : \mathbb{R}^{\mathrm{d}} \ni x \mapsto x + sb(x) \in \mathbb{R}^{\mathrm{d}}.$$

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

$$b_n(x) := \frac{1}{s_n}\int_{s=0}^{s_n} F_{\omega(s)}[\mu_0](x)\mathrm{d}s = F_{\overline{\omega}_n}[\mu_0](x), \quad \text{where} \quad \overline{\omega}_n := \frac{1}{s_n}\int_{s=0}^{s_n} \omega(s)\mathrm{d}s.$$

Since U is compact, so is $\mathscr{P}(U)$ (see [8], Rem. 6.19). Then, along a (non relabeled) subsequence, $\overline{\omega}_n$ converges to some $\overline{\omega} \in \mathscr{P}(U)$ for the Wasserstein distance over $\mathscr{P}(U)$. Let $b \in \overline{\text{co}}F(\mu_0)$ be given by $b = F_{\overline{\omega}}[\mu_0]$. For each $n \in \mathbb{N}$, let $\eta_n \in \Gamma(\omega_n, \omega)$ such that

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

Using [A2], we have that

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

and we conclude to the local uniform convergence of b_n towards b. We know from Theorem 4.1 that $\mu_s = \Phi_s^{0,\omega}\#\mu_0$, where the semigroup $\Phi_s^{0,\omega}$ is defined in (4.2). Let $m_s : \mathbb{R}^{\mathrm{d}} \mapsto \mathbb{R}^{\mathrm{d}}$ be given by $m_s(x) = x + sb(x)$. Along the sequence $(s_n)_n$, we have that

$$\begin{aligned}
\left|\Phi^{0,\omega}_{s_n}(x) - m_{s_n}(x)\right| &= \left|\int_{s=0}^{s_n} F_{\omega(s)}[\mu_s]\left(\Phi^{0,\omega}_s(x)\right)\mathrm{d}s - s_n b(x)\right| \\
&\leqslant \int_{s=0}^{s_n} \left|F_{\omega(s)}[\mu_s]\left(\Phi^{0,\omega}_s(x)\right) - F_{\omega(s)}[\mu_0]\,(x)\right|\mathrm{d}s + s_n \left|F_{\overline{\omega}_n}[\mu_0](x) - F_{\overline{\omega}}[\mu_0](x)\right| \\
&\leqslant [f]\int_{s=0}^{s_n} d_{\mathcal{W}}(\mu_s,\mu_0) + \left|\Phi^{0,\omega}_s(x) - x\right|\mathrm{d}s + s_n\,[f]\,(1+|x|)\left(d_{\mathcal{W},m_{f,u}}(\overline{\omega}_n,\overline{\omega}) + \frac{1}{n}\right).
\end{aligned}$$

By (4.4), we have $d_{\mathcal{W}}(\mu_s,\mu_0) \leqslant s\mathrm{e}^{[f]s}\left([f]\,d_{\mathcal{W}}(\mu_0,\delta_0) + |f|_{0,\infty}\right)$. Plugging this into the above, we get after simplification

$$\frac{d_{\mathcal{W}}\left(\mu_{s_n}, m_{s_n}\#\mu_0\right)}{s_n} \leqslant 2\,[f]\left[s_n\mathrm{e}^{[f]s_n}\left([f]\,d_{\mathcal{W}}(\mu_0,\delta_0) + |f|_{0,\infty}\right) + \widetilde{d_{\mathcal{W}}}(\overline{\omega}_n,\overline{\omega}) + \frac{1}{n}\right],$$

which goes to 0 when n goes to ∞. Hence the result. □

4.2. Properties of the value function

Recall that the value function is defined as

$$V : [0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}, \qquad V(t,\nu) := \inf\left\{\,\mathfrak{J}(\mu^{t,\nu,u}_T)\;\middle|\; u \in L^1([t,T];U)\,\right\}.$$

From Theorem 4.5 and the assumption [A3] of local uniform continuity of $\mathfrak{J}$, we obtain that the set of solutions of the relaxed system (4.1) may be substituted to the set of solutions of the original problem (2.1) without changing the value function, that is,

$$V(t,\nu) = \inf\left\{\,\mathfrak{J}(\mu^{t,\nu,\omega}_T)\;\middle|\;\omega \in L^1([t,T];\mathscr{P}(U))\,\right\} = \inf_{\mu\in\mathcal{R}^{t,\nu}_T}\mathfrak{J}(\mu),$$

where the reachable set $\mathcal{R}^{t,\nu}_T$ is defined in (4.5). 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 $0 < h \leqslant T-t$,

$$V(t,\nu) = \inf_{\mu\in\mathcal{R}^{t,\nu}_{t+h}} V(t+h,\mu). \tag{4.7}$$

Lemma 4.7 (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 $m_{V,R}:\mathbb{R}^+\to\mathbb{R}^+$ such that for all $(t,s)\in[0,T]^2$ and $(\mu,\nu)\in\left(\overline{\mathscr{B}}(\delta_0,R)\right)^2$,*

$$|V(s,\nu) - V(t,\mu)| \leqslant m_{V,R}\left(|t-s| + d_{\mathcal{W}}(\mu,\nu)\right).$$

Proof. 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 $\mathcal{R}^{0,\nu}_T \subset \overline{\mathscr{B}}(\delta_0,R_T)$ for all $\nu\in\overline{\mathscr{B}}(\delta_0,R)$. Let $m_{\mathfrak{J},R_T}$ be a local modulus of continuity of $\mathfrak{J}$ in the ball $\overline{\mathscr{B}}(\delta_0,R_T)$. According to the $[\mathcal{R}]-$Lipschitz continuity of the reachable sets given by Lemma 4.2, we have for all $t\in[0,T]$ and $\nu,\overline{\nu}\in$

$\overline{\mathscr{B}}(\delta_0, R)$ that

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

On the other hand, let $0 \leqslant t \leqslant s \leqslant T$ and $\nu \in \overline{\mathscr{B}}(\delta_0, R)$. The DPP (4.7) and the Grönwall estimate (4.4) give us

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

and similarly

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

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

A close look at the proof reveals that if $\mathfrak{J}$ is uniformly continuous and f is globally bounded, the value function would also be globally uniformly continuous. However, since we do not assume this for $\mathfrak{J}$ and f, the value function is only locally uniformly continuous.

5. The Hamilton–Jacobi equation

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

$$-\partial_t u(t,\mu) + H\left(\mu, D_\mu u(t,\mu)\right) = 0, \qquad u(T,\mu) = \mathfrak{J}(\mu). \tag{5.1}$$

Regularity assumptions on the Hamiltonian $H : \mathbb{T} \to \mathbb{R}$ will be made precise further.

5.1. Notion of viscosity solutions

Recall that $X :=]0,T[\times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ 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 (5.1). 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 5.1 (Locally semiconcave/convex maps)**.** An application $u : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ is locally semiconcave in $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ if for all $R > 0$, there exists $\lambda_R \in \mathbb{R}$ such that for all $\mu,\nu \in \overline{\mathscr{B}}(\delta_0, R)$ and $\eta = \eta(x,y) \in \Gamma_o(\mu,\nu)$, there holds for all $h \in [0,1]$

$$u\left(((1-h)\pi_x + h\pi_y)\#\eta\right) \geqslant (1-h)u(\mu) + hu(\nu) - \frac{\lambda_R}{2}h(1-h)d_{\mathcal{W}}^2\left(\mu,\nu\right).$$

An application $u : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ 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 [10], Sect. 4.2) and for any $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$ and $\xi \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^d)_\mu$,

$$D_\mu d_{\mathcal{W}}^2(\cdot,\sigma)(\xi) = \inf_{\eta\in\exp_\mu^{-1}(\sigma)} \inf_{\alpha\in\Gamma_\mu(\xi,\eta)} -2\int_{(x,v,w)\in\mathrm{T}^2\,\mathbb{R}^d} \langle v,w\rangle\,\mathrm{d}\alpha(x,v,w). \tag{5.2}$$

Remark 5.2. Moreover, given $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$, we may compute $\|D_\mu d_{\mathcal{W}}^2(\cdot,\sigma)\|_\mu$ as defined in (3.14). Indeed,

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

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

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

showing equality.

Remark 5.3 (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 λ_R a local constant of semiconcavity of φ over $[0,3R]$, and $[\varphi_R]$ a local constant of Lipschitz-continuity of φ over the same domain. Then ψ is semiconcave with modulus $R\lambda_R + [\varphi_R]$ (see [38], Prop. 2.1.12).

The squared Wasserstein distance enjoys another interesting property: its directional derivative along ξ is always equal to that along $\pi^\mu\xi$, the projection onto the tangent cone. This follows from the explicit expression (5.2), which can be rewritten as the infimum of $-2\langle\alpha,\xi\rangle$, where $\langle\cdot,\cdot\rangle$ is the metric scalar product from [10], Definition 4.13. Furthermore, Corollary 4.34 of the same reference states that $\langle\alpha,\xi\rangle = \langle\alpha,\pi^\mu\xi\rangle$, since α is the optimal velocity of a geodesic. This additional structure prompts the following definition.

Definition 5.4 (Geometrically consistent applications)**.** An application $u : \mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is geometrically consistent if it is directionally differentiable and if

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

This property will be used in Section 6 in connection with the choice of the Hamiltonian for the control problem; see Remark 6.1 for further details.

5.1.2. Regularity in the time variable

Definition 5.5 (Locally Lipschitz time derivative)**.** We say that $\varphi : X\to\mathbb{R}$ has locally Lipschitz time derivative if for all $\mu\in\mathscr{P}_2(\mathbb{R}^d)$, $\varphi(\cdot,\mu)\in\mathcal{C}^1(]0,T[;\mathbb{R})$, and if the application $(t,\mu)\mapsto\partial_t\varphi(t,\mu)$ is locally Lipschitz in (X,d_X).

Lemma 5.6 (Partial derivatives)**.** *Let $\varphi : X\to\mathbb{R}$ satisfy Definition 5.5. Let $(t,\mu)\in X$ such that $\varphi(t,\cdot)$ is directionally differentiable at μ, and $\xi\in\mathbf{Tan}_\mu\,\mathscr{P}_2(\mathbb{R}^d)$. Then*

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

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

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

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

5.1.3. *Locally uniform upper semicontinuity*

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

Definition 5.7 (Locally uniformly upper semicontinuous)**.** Let (Y,d) be a complete metric space. A locally bounded application $u:Y\to\mathbb{R}$ 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 $B\coloneqq\bigcap_{n\in\mathbb{N}}B_n\neq\emptyset$ and $\lim_{n\to\infty}\sup_{x\in B_n}\inf_{y\in B}d_Y(x,y)\to 0$, there holds

$$\lim_{n\to\infty}\sup_{y\in B_n}u(y)\leqslant\sup_{x\in B}u(x). \tag{5.3}$$

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

Remark 5.8 (Link with other notions of upper semicontinuity)**.** We gather here some properties, whose proofs are postponed to Appendix B.

- Definition 5.7 is strictly weaker than continuity.
- In general, Definition 5.7 is strictly stronger than upper semicontinuity. However both definitions agree whenever the closed balls of Y are compact.
- Let $\mathcal{S}$ be the set of nonempty closed and bounded subsets of Y. Then Definition 5.7 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 [39, 40]).
- The applications that are simultaneously luusc and lulsc are exactly the locally uniformly continuous applications.
- In $Y=\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, there is no comparison with upper semicontinuity in the narrow topology (see counterexamples in Appendix B).

5.1.4. *Definition of viscosity solutions*

Gathering the above definitions, we arrive at the following.

Definition 5.9 (Test functions)**.** We define $\mathscr{T}_+$ as the set of $\varphi:]0,T[\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})\to\mathbb{R}$ such that

- φ and $\partial_t\varphi$ are locally Lipschitz from $]0,T[\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ to $\mathbb{R}$,
- for all $t\in]0,T[$, $\varphi(t,\cdot)$ is geometrically consistent in the sense of Definition 5.4,

- for all $t \in]0,T[$, $\varphi(t,\cdot)$ is locally semiconcave.

Similarly, we denote $\mathscr{T}_- := \{-\varphi \mid \varphi \in \mathscr{T}_+\}$.

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

Definition 5.10 (Viscosity solutions)**.** We say that $u : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ is a viscosity

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

$$-\partial_t \varphi(t,\mu) + H\left(\mu, D_\mu \varphi(t,\mu)\right) \leqslant 0. \tag{5.4}$$

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

$$-\partial_t \psi(t,\mu) + H\left(\mu, D_\mu \psi(t,\mu)\right) \geqslant 0. \tag{5.5}$$

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

Let us put this notion in context with respect to the existing literature. As customary with viscosity solutions, the specificities of each problem require slight adaptations, leading to a great variety of definitions. Here we only mention the main lines, with no claim to be exhaustive.

From our perspective, there are essentially two distinct and non-equivalent approaches to defining viscosity solutions in the Wasserstein space, which depend on the selection of the tangent cone. The option that is most commonly found in the literature is the *regular tangent cone*

$$\operatorname{Tan}_\mu \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) := \overline{\{\nabla\varphi \mid \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^{\mathrm{d}};\mathbb{R})\}}^{L^2_\mu(\mathbb{R}^{\mathrm{d}};\mathrm{T}\,\mathbb{R}^{\mathrm{d}})} = \left\{ f \in L^2_\mu(\mathbb{R}^{\mathrm{d}};\mathrm{T}\,\mathbb{R}^{\mathrm{d}}) \;\middle|\; f\#\mu \in \mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \right\}.$$

There are at least two compelling reasons to consider Tan_μ as an appropriate tangent cone. First, it is sufficient for describing the velocities of absolutely continuous curves, albeit only almost everywhere (see [9]). Second, it serves as the natural space in which two distinct theories have defined their gradients, namely the Lions differentiability [41–43] and the Wasserstein gradient [12, 44]. These constructions aim to establish a Hilbert-like structure on the tangent space and are most effectively applied to the dense subset of regular measures, as defined in [32]).

For this reason, the works [14, 24, 25, 44] consider either gradients, or sub- and superdifferentials *valued in this tangent cone.* From a slightly different perspective, [21] explores Hadamard semidifferentials, which coincide with the same definition in sufficiently smooth scenarios. However, it is also possible to define semidifferentials within the general tangent cone $\mathbf{Tan}_\mu$. As noted in [26], in general, these two approaches do not provide equivalent definitions of viscosity solutions. The problem lies in the computation of the metric slope, over which the definition of viscosity solutions for Eikonal-type equations of [26]. If one considers $\varphi : \nu \mapsto d^2_{\mathcal{W}}\left(\nu, \frac{\delta_{-a}+\delta_a}{2}\right)$ for some $a \in \mathbb{R}^{\mathrm{d}} \setminus \{0\}$, then the metric slope of φ at $\mu = \delta_0$ writes as

$$\varlimsup_{\nu\to\mu} \frac{|\varphi(\nu) - \varphi(\mu)|}{d_{\mathcal{W}}(\mu,\nu)} = \sup_{\xi \in \mathbf{Tan}_\mu \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}), \|\xi\|_\mu = 1} |D_\mu \varphi(\xi)|\,.$$

In this case, the supremum is reached at $\xi = \frac{\delta_{(0,-a/|a|)}+\delta_{(0,a/|a|)}}{2}$, with a metric slope equal to $2|a|$. However, the directional derivative of φ at μ along any direction of the form $\xi = (id,f)\#\mu$ vanishes. This simple example shows that relying on Tan_μ may imply additional smoothness assumptions if one wishes to ensure that both theories are equivalent.

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 [45]: see for instance [19, 26, 46, 47]. 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 [20] or [48] in the pathwise setting.

The perturbed optimization principle will bring, as announced, perturbations. To cope with these additional terms, [24, 25] 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 5.11 (Perturbed optimization)**.** *Denote $Y = \overline{X}^2 = [0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})\times[0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, endowed with the distance*

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

Let $\Phi : Y \to \mathbb{R}\cup\{-\infty\}$ 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

$$\left\{z\in Y \;\middle|\; \Phi(z)\geqslant\Phi(z^0)\right\}\subset\overline{\mathscr{B}}(z^0,R). \tag{5.6}$$

Hence for each $n\in\mathbb{N}_*$, there exists $z_n\in Y$ and a perturbation $p_n : Y\to\mathbb{R}^+$ 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 $\mathscr{T}_+$, and $(s,\nu)\mapsto -p_n(t_n,\mu_n,s,\nu)$ belongs to $\mathscr{T}_-$,*
3. *There exists an application $\omega_{T,R,A} : \mathbb{N}\to\mathbb{R}^+$ such that*

$$\sum_{r\in\{t,s\}}\left|\partial_r p_n(z_n)\right| + \sum_{\sigma\in\{\mu,\nu\}}\left(1+d_{\mathcal{W}}(\sigma,\delta_0)\right)\left\|D_\sigma p_n(z_n)\right\|_\sigma \leqslant \omega_{T,R,A}(n) \underset{n\to\infty}{\longrightarrow} 0,$$

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

The proof of Lemma 5.11 is delayed to Appendix C.

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

$$\left|H\left(\mu,p+q\right)-H\left(\mu,p\right)\right| \leqslant [H]\left(1+d_{\mathcal{W}}(\mu,\delta_0)\right)\|q\|_\mu, \tag{5.7}$$

$$H\left(\mu,-aD_\mu d_{\mathcal{W}}^2(\cdot,\nu)\right) - H\left(\nu,aD_\nu d_{\mathcal{W}}^2(\mu,\cdot)\right) \leqslant [H]\,d_{\mathcal{W}}(\mu,\nu)\left(a d_{\mathcal{W}}(\mu,\nu)+1\right). \tag{5.8}$$

The condition [A4] indicates a locally Lipschitz behavior of the Hamiltonian. In particular, the condition (5.8) can be seen as a one-sided Lipschitz bound on the variations with respect to the first argument only. In the context of the Hamilton–Jacobi–Bellman equation associated with the control problem (2.2), the Bellman

Hamiltonian satisfies the requirement [A4] as long as the standing hypothesis [A2] is satisfied. This assertion will be demonstrated in Lemmata 6.3 and 6.4.

Theorem 5.12 (Comparison principle)**.** *Assume [A4]. Let* $u : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ *be a locally bounded subsolution of (5.1) and* $v : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ *be a locally bounded supersolution of (5.1), which are such that* $u(T,\mu) \leqslant v(T,\mu)$ *for all* $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$*. Then*

$$u(t,\mu) \leqslant v(t,\mu) \qquad \forall (t,\mu) \in]0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}). \tag{5.9}$$

This proof builds on the ideas of [30], Theorem 3.50, p. 206 developed in Hilbert spaces. The structure is the following: assume by contradiction that the inequality (5.9) 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 compared to the classical arguments (see [5], Sect. 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 $\overline{X} = [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, 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 5.7 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.

Proof. Denote again $X =]0,T[\times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. By abuse of notation, let $X^2 :=]0,T[\times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times]0,T[\times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and define the distances

$$\begin{aligned} d_X^2((t,\mu),(\bar t,\bar\mu)) &:= \left|t-\bar t\right|^2 + d_{\mathcal{W}}^2(\mu,\bar\mu), \\ d_{X^2}^2((t,\mu,s,\nu),(\bar t,\bar\mu,\bar s,\bar\nu)) &:= d_X^2((t,\mu),(\bar t,\bar\mu)) + d_X^2((s,\nu),(\bar s,\bar\nu)). \end{aligned}$$

Assume by contradiction that

$$\Gamma := \min\left(1, \sup\left\{u(t,\mu) - v(t,\mu) \;\middle|\; (t,\mu) \in \overline{X}\right\}\right) > 0. \tag{5.10}$$

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

$$\max\left(u(t,\mu), -v(t,\mu)\right) \leqslant g\left(d_{\mathcal{W}}(\mu,\delta_0)\right) \qquad \forall \mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}). \tag{5.11}$$

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

$$h : \mathbb{R}^2 \to \mathbb{R}, \qquad h(t,r) := g^2\left((1+r)\,\mathrm{e}^{-4[H]t}\right) + \frac{1}{t}.$$

The application h will play the role of a penalization through the composition

$$h_{\circ d} : \mathbb{R} \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}), \qquad h_{\circ d}(t,\mu) := h(t, d_{\mathcal{W}}^2(\delta_0,\mu)).$$

As $h_{\circ d}(t,\cdot)$ grows strictly faster than u and $-v$ by (5.11), for any $\iota > 0$,

$$\begin{aligned}&\max\left[u(t,\mu),-v(t,\mu)\right]-\iota h_{\circ d}(t,\mu)\\&\leqslant \sup_{r\in\mathbb{R}^+} g\left(r\right)-\iota g^2\left((1+r^2)\mathrm{e}^{-4[H]T}\right)=:\llbracket g-\iota h\rrbracket<\infty.\end{aligned}\tag{5.12}$$

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

$$\partial_t h(t,r)=-8\left[H\right](1+r)\mathrm{e}^{-4[H]t}(gg')\left((1+r)\,\mathrm{e}^{-4[H]t}\right)-\frac{1}{t^2}\leqslant -8\left[H\right]\mathrm{e}^{-4[H]T}.\tag{5.13}$$

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

$$\begin{aligned}2\left[H\right](1+r)\,\partial_r h(t,r)&=4\left[H\right](1+r)\,(gg')\left((1+r)\,\mathrm{e}^{-4[H]t}\right)\mathrm{e}^{-4[H]t}\\&=-\frac{1}{2}\partial_t h(t,r)-\frac{1}{2t^2}\leqslant -\frac{1}{2}\partial_t h(t,r).\end{aligned}\tag{5.14}$$

As g is $\mathcal{C}^2$, the map h has a locally Lipschitz time derivative. Moreover, owing to Remark 5.3, the application $\mu \mapsto h_{\circ d}(t,\mu)$ is locally semiconcave for all t, and locally Lipschitz. Hence it may enter in the composition of test functions.

Let $\phi_\iota : \overline{X}^2 \to \{-\infty\} \cup \mathbb{R}$ be given by

$$\phi_\iota(t,\mu,s,\nu) \coloneqq u(t,\mu)-v(s,\nu)-\iota\left(h_{\circ d}(t,\mu)+h_{\circ d}(s,\nu)\right).$$

We consider

$$\Gamma_\iota \coloneqq \lim_{r\to 0}\sup\left\{\phi_\iota(t,\mu,s,\nu)\;\middle|\;(t,\mu,s,\nu)\in\overline{X}^2,\ d_X((t,\mu),(s,\nu))\leqslant r\right\},\tag{5.15}$$

$$\Gamma_{\iota\varepsilon} \coloneqq \sup\left\{\phi_\iota(t,\mu,s,\nu)-\frac{d_X^2((t,\mu),(s,\nu))}{\varepsilon}\;\middle|\;(t,\mu,s,\nu)\in\overline{X}^2\right\}.\tag{5.16}$$

Then one has

$$\Gamma_{\iota\varepsilon}\underset{\varepsilon\searrow 0}{\searrow}\Gamma_\iota \qquad \text{for all } \iota>0, \qquad \text{and} \qquad \Gamma_\iota \underset{\iota\searrow 0}{\nearrow}\Gamma_0 \geqslant \Gamma.\tag{5.17}$$

Here Γ_0 may be equal to $+\infty$. The arguments of (5.17) are easy but tedious, and devolved in Appendix D to lighten the presentation.

The perturbed maximization. Let $\Phi_{\iota\varepsilon} : \overline{X}^2 \to \{-\infty\} \cup \mathbb{R}$ be given by

$$\Phi_{\iota\varepsilon}(t,\mu,s,\nu) \coloneqq \varphi_\iota(t,\mu,s,\nu)-\frac{d_X^2((t,\mu),(s,\nu))}{\varepsilon}.$$

Let $z^0 \coloneqq (T,\delta_0,T,\delta_0)$. For each fixed ι,ε, the application $\Phi_{\iota\varepsilon}$ is upper semicontinuous, proper and - using (5.12) - upper bounded in the complete metric space $\left(\overline{X}^2, d_{X^2}\right)$. Moreover, if $\Phi_{\iota\varepsilon}(z) \geqslant \Phi_{\iota\varepsilon}(z^0)$, then

$$\frac{\iota}{2}\left(h_{\circ d}\left(t,\mu\right)+h_{\circ d}\left(s,\nu\right)\right)\leqslant 2\llbracket g-\frac{\iota}{2}h\rrbracket-\left(u(T,\delta_0)-v(T,\delta_0)-2\iota h(T,0)\right),\tag{5.18}$$

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

$$\sup \Phi_{\iota\varepsilon} - \Phi_{\iota\varepsilon}(z^0) \leqslant 2[\![g - \iota h]\!] - (u(T, \delta_0) - v(T, \delta_0) - 2\iota h(T, 0)) =: A_\iota$$

does not depend on ε.

Hence we may apply Lemma 5.11: for each $n \in \mathbb{N}_*$, there exist $z_{\iota\varepsilon n} \in \overline{X}^2$ and a perturbation $p_{\iota\varepsilon n} : \overline{X}^2 \to \mathbb{R}^+$ 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 $\mathscr{T}_+, \mathscr{T}_-$, and there exist maps $\omega_\iota, C_\iota : \mathbb{N}_* \to \mathbb{R}^+$ such that

$$\sum_{r \in \{s,t\}} |\partial_r p_{\iota\varepsilon n}(z_{\iota\varepsilon n})| + \sum_{\sigma \in \{\mu,\nu\}} (1 + d_{\mathcal{W}}(\sigma_{\iota\varepsilon n}, \delta_0)) \|D_\sigma p_{\iota\varepsilon n}(z_{\iota\varepsilon n})\|_{\sigma_{\iota\varepsilon n}} \leqslant \omega_\iota(n), \tag{5.19}$$
$$\Gamma_{\iota\varepsilon} - \Phi_{\iota\varepsilon}(z_{\iota\varepsilon n}) = \sup \Phi_{\iota\varepsilon} - \Phi_{\iota\varepsilon}(z_{\iota\varepsilon n}) \leqslant C_\iota(n)$$

and $\omega_\iota(n), C_\iota(n) \underset{n\to\infty}{\longrightarrow} 0$. Notice that

$$\begin{aligned}\Gamma_{\iota\varepsilon} + \frac{d_X^2((t_{\iota\varepsilon n}, \mu_{\iota\varepsilon n}), (s_{\iota\varepsilon n}, \nu_{\iota\varepsilon n}))}{2\varepsilon} &\leqslant \Phi_{\iota\varepsilon}(z_{\iota\varepsilon n}) + C_\iota(n) + \frac{d_X^2((t_{\iota\varepsilon n}, \mu_{\iota\varepsilon n}), (s_{\iota\varepsilon n}, \nu_{\iota\varepsilon n}))}{2\varepsilon} \\ &\leqslant \Gamma_{\iota\frac{\varepsilon}{2}} + C_\iota(n)\end{aligned}$$

so that for each fixed ι, there holds by (5.17)

$$\frac{d_X^2((t_{\iota\varepsilon n}, \mu_{\iota\varepsilon n}), (s_{\iota\varepsilon n}, \nu_{\iota\varepsilon n}))}{\varepsilon} \leqslant 2\left(\Gamma_{\iota\frac{\varepsilon}{2}} - \Gamma_{\iota\varepsilon} + C_\iota(n)\right) \underset{\varepsilon\to 0, n\to\infty}{\longrightarrow} 0. \tag{5.20}$$

To lighten the notation, we shorten $a_{\iota\varepsilon n}$ in a_* in the sequel, for any $a \in \{t, s, \mu, \nu, p, z, \cdots\}$.

Staying away from T. Let us show that for sufficiently small ι and large n, the points t_*, s_* belong to $]0, T[$. Since $h(0, r) = +\infty$ for all $r \geqslant 0$, the construction of z_* implies $t_*, s_* > 0$. On the other hand, recalling (5.17), let $\iota_0 > 0$ be small enough so that $\Gamma_\iota \geqslant \Gamma/2$ for all $0 < \iota \leqslant \iota_0$. Hence, as $\varepsilon \mapsto \Gamma_{\iota\varepsilon}$ decreases when ε decreases,

$$u(t_*, \mu_*) - v(s_*, \nu_*) \geqslant \Phi_{\iota\varepsilon}(z_*) \geqslant \Gamma_{\iota\varepsilon} - C_\iota(n) \geqslant \Gamma_\iota - C_\iota(n) \geqslant \frac{\Gamma}{2} - C_\iota(n)$$

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

Then for each $0 < \iota \leqslant \iota_0$, $\varepsilon > 0$ sufficiently small and n sufficiently large, $t_{\iota\varepsilon n}, s_{\iota\varepsilon n} < T$ simultaneously. Indeed, recall from (5.18) that $\mu_{\iota\varepsilon n}, \nu_{\iota\varepsilon n} \in \overline{\mathscr{B}}(\delta_0, R_\iota)$ for all ε, n. If there exists a sequence $(\varepsilon_m, n_m) \to_m (0, \infty)$ such that $n_m \geqslant n_\iota$ and $T \in \{t_{\iota\varepsilon_m n_m}, s_{\iota\varepsilon_m n_m}\}$ for all m, then using (5.20),

$$\begin{aligned}\frac{\Gamma}{4} &\leqslant \varliminf_{m\to\infty} u(t_{\iota\varepsilon_m n_m}, \mu_{\iota\varepsilon_m n_m}) - v(s_{\iota\varepsilon_m n_m}, \nu_{\iota\varepsilon_m n_m}) \\ &\leqslant \lim_{r \searrow 0} \sup\left\{u(t,\mu) - v(s,\nu) \;\middle|\; d_X^2((t,\mu),(s,\nu)) \leqslant r,\ \mu, \nu \in \overline{\mathscr{B}}(\delta_0, R_\iota),\ t, s \in [T - r, T]\right\} \\ &\leqslant \sup\left\{u(t,\mu) - v(s,\nu) \;\middle|\; t = s = T,\ \mu = \nu \in \overline{\mathscr{B}}(\delta_0, R_\iota)\right\}.\end{aligned}$$

The last inequality holds since $(t, \mu, s, \nu) \mapsto u(t, \nu) - v(s, \nu)$ is locally uniformly upper semicontinuous (see Def. 5.7). But this contradicts the assumption that $u(T, \cdot) - v(T, \cdot) \leqslant 0$.

Application of the definition of semisolutions. Define $\varphi_*, \psi_* : X =]0,T[\times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ as

$$\varphi_*(t,\mu) \coloneqq v(s_*,\nu_*) + \iota\left(h_{\circ d}(t,\mu) + h_{\circ d}(s_*,\nu_*)\right) + \frac{d_X^2((t,\mu),(s_*,\nu_*))}{\varepsilon} + p_*(t,\mu,s_*,\nu_*)$$

and

$$\psi_*(s,\nu) \coloneqq u(t_*,\mu_*) - \iota\left(h_{\circ d}(t_*,\mu_*) + h_{\circ d}(s,\nu)\right) - \frac{d_X^2((t_*,\mu_*),(s,\nu))}{\varepsilon} - p_*(t_*,\mu_*,s,\nu).$$

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

$$\begin{aligned} &-\iota\,\partial_t h_{\circ d}(t_*,\mu_*) - \frac{2(t_*-s_*)}{\varepsilon} - \partial_t p_*(z_*) \\ &+ H\left(\mu_*, \iota D_\mu h_{\circ d}(t_*,\mu_*) + \frac{D_\mu d_{\mathcal{W}}^2(\mu_*,\nu_*)}{\varepsilon} + D_\mu p_*(z_*)\right) \leqslant 0. \end{aligned} \tag{5.21}$$

Let us momentaneously denote $c_* \coloneqq [H]\left(1 + d_{\mathcal{W}}(\mu_*,\delta_0)\right)$. Using the Lipschitz assumption (5.7),

$$\begin{aligned} &H\left(\mu_*, \iota D_\mu h_{\circ d}(t_*,\mu_*) + \frac{D_\mu d_{\mathcal{W}}^2(\mu_*,\nu_*)}{\varepsilon} + D_\mu p_*(z_*)\right) \\ &\geqslant H\left(\mu_*, \frac{D_\mu d_{\mathcal{W}}^2(\mu_*,\nu_*)}{\varepsilon}\right) - c_*\left(\iota\|D_\mu h_{\circ d}(t_*,\mu_*)\|_{\mu_*} + \|D_\mu p_*(z_*)\|_{\mu_*}\right). \end{aligned}$$

Using the regularity of h and the estimate of Remark 5.2 on the differential of the squared distance, we have

$$\begin{aligned} \|D_\mu h_{\circ d}(t_*,\mu_*)\|_{\mu_*} &= \left|\partial_r h(t_*, d_{\mathcal{W}}^2(\mu_*,\delta_0))\right| \|D_\mu d_{\mathcal{W}}^2(\mu_*,\delta_0)\|_{\mu_*} \\ &= \partial_r h(t_*, d_{\mathcal{W}}^2(\mu_*,\delta_0)) 2 d_{\mathcal{W}}(\mu_*,\delta_0) \end{aligned}$$

since $\partial_r h \geqslant 0$. Recalling that the partial derivatives of h satisfy (5.14), we get that

$$2\,[H]\,(1 + d_{\mathcal{W}}(\mu_*,\delta_0)) d_{\mathcal{W}}(\mu_*,\delta_0) \partial_r h(t_*, d_{\mathcal{W}}^2(\mu_*,\delta_0)) \leqslant -\frac{1}{2}\partial_t h\left(t_*, d_{\mathcal{W}}^2(\mu_*,\delta_0)\right). \tag{5.22}$$

Combining the four last estimates, we obtain

$$\begin{aligned} &-\iota\,\partial_t h_{\circ d}(t_*,\mu_*) - \frac{2(t_*-s_*)}{\varepsilon} - \partial_t p_*(z_*) \\ &+ H\left(\mu_*, \frac{D_\mu d_{\mathcal{W}}^2(\mu_*,\nu_*)}{\varepsilon}\right) + \frac{\iota}{2}\,\partial_t h_{\circ d}(t_*,\mu_*) - c_*\|D_\mu p_*(z_*)\|_{\mu_*} \leqslant 0. \end{aligned} \tag{5.23}$$

On the other hand, $v - \psi_*$ admits a minimum in $(s_*,\nu_*) \in X$. Applying the definition of supersolution and repeating the argument from (5.21) to (5.23), we get

$$\begin{aligned} &\iota\partial_s h_{\circ d}\left(s_*,\nu_*\right) + \frac{2(s_*-t_*)}{\varepsilon} + \partial_s p_*(z_*) - \frac{\iota}{2}\partial_s h_{\circ d}\left(s_*,\nu_*\right) \\ &+ [H]\left(1 + d_{\mathcal{W}}(\nu_*,\delta_0)\right)\|D_\nu p_*(z_*)\|_{\nu_*} + H\left(\nu_*, -\frac{D_\nu d_{\mathcal{W}}^2(\mu_*,\nu_*)}{\varepsilon}\right) \geqslant 0. \end{aligned} \tag{5.24}$$

Taking the difference between (5.23) and (5.24) and the inequality resulting from plugging (5.23) and (5.19) into (5.21), we get after simplification

$$0 \leqslant H\left(\nu_*, -\frac{D_\nu d_{\mathcal{W}}^2(\mu_*, \nu_*)}{\varepsilon}\right) - H\left(\mu_*, \frac{D_\mu d_{\mathcal{W}}^2(\mu_*, \nu_*)}{\varepsilon}\right) + \frac{\iota}{2}\partial_s h_{\circ d}\,(s_*, \nu_*) \tag{5.25}$$

$$+\frac{\iota}{2}\partial_t h_{\circ d}(t_*, \mu_*) + \sum_{\tau\in\{s,t\}} \partial_\tau p_*(z_*) + [H] \sum_{\sigma\in\{\mu,\nu\}} (1 + d_{\mathcal{W}}(\sigma_*, \delta_0))\, \|D_\sigma p_*(z_*)\|_{\sigma_*}. \tag{5.26}$$

Estimates and conclusion. Recall from (5.13) that $-\frac{1}{2}\partial_t h(t,r) \geqslant 4\,[H]\,\mathrm{e}^{-4[H]T} > 0$ for all $t > 0$ and $r \in \mathbb{R}^+$. Using the assumption (5.8) on H, we get

$$H\left(\nu_*, -\frac{D_\nu d_{\mathcal{W}}^2(\mu_*, \nu_*)}{\varepsilon}\right) - H\left(\mu_*, \frac{D_\mu d_{\mathcal{W}}^2(\mu_*, \nu_*)}{\varepsilon}\right) \leqslant [H]\, d_{\mathcal{W}}(\mu_*, \nu_*) \left(\frac{d_{\mathcal{W}}(\mu_*, \nu_*)}{\varepsilon} + 1\right).$$

Using (5.19) to estimate (5.26), we arrive at

$$4\iota\,[H]\,\mathrm{e}^{-4[H]T} \leqslant [H]\, d_{\mathcal{W}}(\mu_*, \nu_*) \left(\frac{d_{\mathcal{W}}(\mu_*, \nu_*)}{\varepsilon} + 1\right) + \omega_\iota(n). \tag{5.27}$$

Keeping ι fixed, letting $\varepsilon \to 0$ and $n \to \infty$, we get from (5.20) that $4\iota\,[H]\,\mathrm{e}^{-4[H]T} \leqslant 0$, which is absurd. Thus $\Gamma \leqslant 0$. □

6. The case of Hamilton–Jacobi–Bellman equations

We now return to the case of control problems. Recall that $X =]0, T[\times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. Consider the dynamic $\overline{\mathrm{co}}F$ defined in (4.6), and let

$$H : \mathbb{T} \to \mathbb{R}, \qquad H\,(\mu, p) := \sup_{f\in\overline{\mathrm{co}}F[\mu]} -p\,(\pi^\mu(f\#\mu))\,. \tag{6.1}$$

The Hamilton–Jacobi–Bellman equation associated to (6.1) then writes

$$\begin{cases} -\partial_t u(t,\mu) + \sup\limits_{f\in\overline{\mathrm{co}}F[\mu]} -D_\mu u(t,\mu)\,(\pi^\mu(f\#\mu)) = 0 & (t,\mu) \in X, \qquad (6.2\mathrm{a}) \\ u(T,\mu) = \mathfrak{J}(\mu) & \mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}). \qquad (6.2\mathrm{b}) \end{cases}$$

Remark 6.1 (The Hamiltonian and geometric consistency)**.** The construction of the Hamiltonian in (6.1), with a projection onto the tangent cone, is the only reason to impose geometric consistency to the test functions. In particular, geometric consistency does not intervene in the comparison principle. One could remove the projection; in this case, the elements of the metric cotangent bundle should be defined on $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$. The choice of using π^μ has almost no impact, but follows an emerging consensus in the literature to restrict differential objects to (subsets of) $\mathbf{Tan}_\mu$ or Tan_μ.

In the sequel, we verify that the control Hamiltonian (6.1) satisfies the assumptions of the comparison principle, and we then show that the value function is a solution of (6.2).

6.1. Properties of the control Hamiltonian

Let $G : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \rightrightarrows \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})$ the probability vector field (in the spirit of [34], Def. 2.1) given by

$$G[\mu] \coloneqq \{b\#\mu \mid b \in \overline{\mathrm{co}}F[\mu]\}\,.$$

Under [A2], this PVF is Lipschitz-continuous in the Hausdorff sense with respect to the application $W_{(\mu,\nu)}$ defined in (3.10), with constant $2\,[f]$. Indeed, given $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and $b_0 \in \overline{\mathrm{co}}F[\mu_0]$ defined by $b_0(x) = \int_{u\in U} f(x,\mu_0,u)\mathrm{d}\omega(u)$, define $b_1 \coloneqq \int_{u\in U} f(x,\mu_1,u)\mathrm{d}\omega(u)$. Denote $\eta \in \Gamma_o(\mu,\nu)$. Then

$$\begin{aligned}W^2_{(\mu_0,\mu_1)}(b_0\#\mu_0, b_1\#\mu_1) &\leqslant \int_{(\mathbb{R}^{\mathrm{d}})^2}\int_{u\in U} |f(x,\mu_0,u) - f(y,\mu_1,u)|^2\,\mathrm{d}\omega(u)\mathrm{d}\eta(x,y)\\ &\leqslant 4\,[f]^2\, d^2_{\mathcal{W}}(\mu_0,\mu_1).\end{aligned}$$

Remark 6.2 (PVF formalism)**.** Our motivation in using a general probability vector field is to recover the connection between the metric slope and the directional derivatives. Indeed, in [26], Section 4.3, it is explicit that the regular tangent cone does not provide enough directions, in the sense that there exists measures μ,ν such that

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

Here the last term is the metric slope. The strict inequality comes from the strong convexity of the squared distance of $\mathbb{R}^{\mathrm{d}}$, 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 Tan_μ by using the theory of continuity equations to formulate the control problem, our HJB tools are formulated in the general case.

Lemma 6.3 (Locally Lipschitz behavior of H)**.** *Assume [A2]. There exists a constant* $[H]$ *such that*

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

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

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

Using the non-expansivity of the projection (see [10], Cor. 4.37),

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

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

Lemma 6.4 (Behavior on the squared distance)**.** *Assume [A2]. There exists a constant* C_H *such that for all* $a \geqslant 0$ *and* $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$*,*

$$H\left(\mu, -aD_\mu d^2_{\mathcal{W}}(\cdot,\nu)\right) - H\left(\nu, aD_\nu d^2_{\mathcal{W}}(\mu,\cdot)\right) \leqslant 2aC_H d^2_{\mathcal{W}}(\mu,\nu).$$

Proof. As H is positively homogeneous, we may take $a=1$. Let $\xi \in G[\mu]$ and $\zeta \in G[\nu]$. By [28], Theorem 3.8, there holds $D_\mu d^2_{\mathcal{W}}(\cdot,\nu)(\pi^\mu \xi) = D_\mu d^2_{\mathcal{W}}(\cdot,\nu)(\xi)$. Moreover, using the bijection between $\exp_\mu^{-1}(\nu)$ and $\Gamma_o(\mu,\nu)$, the directional derivative of the squared Wasserstein distance (5.2) writes

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

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 $\omega = \omega(x,v,y,w) \in \Gamma(\xi,\zeta) \subset \mathscr{P}\left((\mathrm{T}\,\mathbb{R}^{\mathrm{d}})^2\right)$ such that $(\pi_x,\pi_v,\pi_y) = \alpha$ and $(\pi_y,\pi_w,\pi_x)\#\omega = \beta$. Then

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

Hence

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

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

Remark 6.5 (Game Hamiltonians)**.** Using the elementary inequality $\inf_A f - \inf_A g \leqslant \sup_A (f-g)$ if all terms are finite, one can show that the Hamiltonian $H : \mathbb{T} \to \mathbb{R}$ given by

$$H(\mu,p) \coloneqq \inf_{a\in A} \sup_{b\in B} -p\left(\pi^\mu(f(\cdot,a,b)\#\mu)\right) - \ell(a,b,\mu)$$

satisfies the assumption [A4] if A, B are compact sets of controls, f satisfies [A2] and ℓ is Lipschitz with respect to the μ variable, uniformly in a, b. Similarly, one could relax f to consider a probability vector field. We do not insist more on the game setting.

6.2. Characterization of the solution

Theorem 6.6 (Complete characterization of (6.2))**.** *Assume [A1], [A2] and [A3]. The value function is the unique viscosity solution in the sense of Definition 5.10 of the Hamilton–Jacobi-Bellman equation (6.2).*

Proof. Let us show that it is a viscosity solution of (6.2). By Lemma 4.7, 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 (5.4) and (5.5).

Subsolution inequality. Let $\varphi \in \mathscr{T}_+$ such that $V - \varphi$ reaches a maximum at $(t,\mu) \in X$, and $[\varphi]$ a local Lipschitz constant in a ball of radius $\max(MT, [f]\,(|f|_{0,\infty} + d_{\mathcal{W}}(\nu,\delta_0)))$, where M is the constant of Theorem 4.1. By Lemma 4.3, each constant relaxed control $\omega \in L^1([t,T];\mathscr{P}(U))$ given by $\omega(s) \equiv \varpi \in \mathscr{P}(U)$ generates a smooth solution $(\mu_s^{t,\nu,\omega})_{s\in[t,T]}$, in the sense that

$$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,\omega}, (id + hF_\varpi[\nu])\,\#\nu\right)}{h} = 0.$$

Consequently, along the curve $h \mapsto \hat{\mu}_h \coloneqq (id + hF_\varpi)\,\#\nu$, the DPP (4.7) gives

$$\begin{aligned}\frac{\varphi(t+h,\hat{\mu}_h) - \varphi(t,\nu)}{h} &\geqslant \frac{\varphi(t+h,\hat{\mu}_h) - \varphi(t+h,\mu_{t+h}^{t,\nu,\omega})}{h} + \frac{V(t+h,\hat{\mu}_h) - V(t,\mu)}{h}\\ &\geqslant -[\varphi]\,\frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,\omega},\hat{\mu}_h\right)}{h} + 0.\end{aligned}$$

Multiplicating by -1 and using the chain rule of Lemma 5.6 to take the limit in $h \searrow 0$, we obtain

$$-\partial_t\varphi(t,\nu) - D_\nu\varphi(t,\nu)(F_\varpi[\nu]\#\nu) \leqslant 0.$$

As $\varphi(t,\cdot)$ is geometrically consistent, we have $D_\nu\varphi(t,\nu)(F_\varpi[\nu]\#\nu) = D_\nu\varphi(t,\nu)(\pi^\nu F_\varpi[\nu]\#\nu)$. Taking the supremum over $b = F_\varpi[\nu] \in \overline{\mathrm{co}}F[\nu]$, one recovers the inequality (5.4), so that V is a subsolution.

Supersolution inequality. Let $\psi \in \mathscr{T}_-$ such that $V - \psi$ reaches a minimum in $(t,\nu) \in X$. Since under [A2], the set of solutions issued from (t,μ) is compact in the topology of uniform convergence (see [36], Thm. 4.5), we may find $\omega \in L^1([t,T];\mathscr{P}(U))$ such that $V(t,\nu) = V(t+h,\mu_{t+h}^{t,\nu,\omega})$ for all $h \in [0,T-t]$. Let $[\psi]$ be a local Lipschitz constant of ψ as above. Applying Lemma 4.6, there exist $(h_n)_n \searrow 0$ and $b \in \overline{\mathrm{co}}F[\nu]$ such that $d_{\mathcal{W}}\left(\mu_{h_n}^{t,\nu,\omega},\hat{\mu}_{h_n}\right) = o(h_n)$, where $\hat{\mu}_{h_n} \coloneqq (id + h_n b)\,\#\nu$. Then

$$\begin{aligned}&-[\psi]\,\frac{d_{\mathcal{W}}\left(\mu_{h_n}^{t,\nu,\omega},\hat{\mu}_{h_n}\right)}{h_n} + \frac{\psi\left(t+h_n,\hat{\mu}_{h_n}\right) - \psi(t,\nu)}{h_n}\\ &\leqslant \frac{\psi\left(t+h_n,\mu_{h_n}^{t,\nu,\omega}\right) - \psi(t,\nu)}{h_n} \leqslant \frac{V\left(t+h_n,\mu_{h_n}^{t,\nu,\omega}\right) - V(t,\nu)}{h_n} = 0.\end{aligned}$$

Taking the limit in $n \to \infty$ and using again the chain rule of Lemma 5.6, we get $\partial_t\psi(t,\nu) + D_\nu\psi\,(b) \leqslant 0$. Multiplicating by -1, taking the supremum over $b \in \overline{\mathrm{co}}F[\nu]$, and using the geometric consistency of φ, we obtain

$$\begin{aligned}-\partial_t\psi(t,\nu) + \sup_{b\in\overline{\mathrm{co}}F[\nu]} -D_\nu\psi(b) &= -\partial_t\psi(t,\nu) + \sup_{b\in\overline{\mathrm{co}}F[\nu]} -D_\nu\psi(\pi^\nu b)\\ &= -\partial_t\psi(t,\nu) + H\left(\nu, D_\nu\psi(t,\nu)\right) \geqslant 0.\end{aligned}$$

To conclude, assume that there exists another viscosity solution $W : [0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ of (6.2). By Lemmata 6.3 and 6.4, the control Hamiltonian defined in (6.1) satisfies the assumption [A4]. Applying Theorem 5.12 to the couples (V,W) and (W,V), we have $V \geqslant W$ and $W \geqslant V$ pointwise over $]0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, so that they coincide. As both are continuous, the equality extends to $t = 0$, and the solution is unique. □

6.3. The case of a Bolza problem

To conclude, let us comment on the adjustments that should be made of one considers a running cost. Assume that the value function is defined as

$$V(t,\mu) = \inf_{u(\cdot)\in L^1(t,T;U)} \int_{s=t}^{T} \ell(\mu_s, u(s))\mathrm{d}s + \mathfrak{J}(\mu_T).$$

We suppose that $\ell : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U \to \mathbb{R}$ is Lipschitz with respect to μ and continuous with respect to u. Here a dependence of ℓ with respect to the space variable $x \in \mathbb{R}^{\mathrm{d}}$ may be encapsulated, for instance by choosing $\ell(\mu,u) = \int_{x\in\mathbb{R}^{\mathrm{d}}} \widetilde{\ell}(x,\mu,u)\mathrm{d}\mu(x)$.

Relaxed system. Following the course of the paper, let us consider the same relaxation of the control problem. Unlike in the Mayer problem, the relaxation procedure makes the relaxed control explicitly appear in the cost of the control problem: the value function is equal to

$$V(t,\mu) = \inf_{\omega(\cdot)\in L^1(t,T;\mathscr{P}(U))} \int_{s=t}^{t+h} \int_{u\in U} \ell(\mu_s^{t,\nu,\omega}, u)\mathrm{d}\omega(s)(u)\mathrm{d}s + \mathfrak{J}(\mu_T^{t,\nu,\omega}).$$

This may be seen by introducing an additional variable $z_\tau = \int_{s=t}^{\tau} \ell(y_s, u(s))\mathrm{d}s$ and applying the relaxation theorem to the augmented dynamical system in variables (y,z).

HJB equation. In view of the above, one expects the value function to satisfy the Hamilton–Jacobi equation with

$$H(\mu,p) \coloneqq \sup_{\omega\in\mathscr{P}(U)} \left[-p(\pi^\mu(F_\omega[\mu]\#\mu)) + \int_{u\in U} \ell(\mu,u)\mathrm{d}\omega(u) \right].$$

Since $\ell(\cdot,u)$ is Lipschitz, the first item of Assumption [A4] is satisfied. However, H is no longer positively homogeneous, and the second item of Assumption [A4] has to be modified in

$$H\left(\mu, -aD_\mu d_{\mathcal{W}}^2(\cdot,\nu)\right) - H\left(\nu, a d_{\mathcal{W}}^2(\mu,\cdot)\right) \leqslant 2[H]\left(a d_{\mathcal{W}}^2(\mu,\nu) + [\ell] d_{\mathcal{W}}(\mu,\nu)\right),$$

which is satisfied by H. The last step of the proof of the comparison principle Theorem 5.12 will see a term $2[H]d_{\mathcal{W}}(\mu_*,\nu_*)$ appear in (5.27), which goes to 0 with ε, n, and does not change the argument.

Characterization of the value function. Regarding the satisfaction of the HJB equation by the value function, one should pay attention to limits. In the first part of Theorem 6.6, the inequality

$$\frac{V(t+h,\hat{\mu}_h) - V(t,\mu)}{h} \geqslant \frac{1}{h}\int_{s=t}^{t+h}\int_{u\in U} \ell(\hat{\mu}_s,u)\mathrm{d}\varpi(u)\mathrm{d}s \geqslant \int_{u\in U} \ell(\nu,u)\mathrm{d}\varpi(u) - o(s)$$

follows from the uniform continuity of ℓ, and the fact that the control is taken constant. In the supersolution part, one considers instead

$$\frac{V(t+h,\hat{\mu}_h) - V(t,\mu)}{h} = \frac{1}{h}\int_{s=t}^{t+h}\int_{u\in U} \ell(\mu_s^{t,\nu,\omega},u)\mathrm{d}\frac{1}{h}\int_{s=t}^{t+h}(u)\mathrm{d}s = \left\langle \ell(\nu,\cdot), \frac{1}{h}\int_{s=t}^{t+h}\omega(s)\mathrm{d}s \right\rangle + o(h),$$

and by a close inspection of the proof of Lemma 4.6, one is allowed to consider that the time-averaged measure converges to some relaxed control along a vanishing subsequence $(h_n)_n$, so that the rest of the argument is unchanged.

Appendix A. Approximation of smooth trajectories

Lemma A.1 (Smooth case)**.** *Let $0 \leqslant t \leqslant T$ and $\omega \in L^1([t,T]; \mathscr{P}(U))$ be a constant control $\omega(s) \equiv \varpi \in \mathscr{P}(U)$. Then the unique solution $(\mu_s^{t,\nu,\omega})_{s\in[t,T]}$ satisfy*

$$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,\omega}, \exp_\mu(h \cdot F_\varpi[\nu])\right)}{h} = 0.$$

Proof. Let again $\Phi_\cdot^t$ be the flow of the ODE (4.2), such that $\mu_s^{t,\nu,\omega} = \Phi_s^t \# \nu$. There holds

$$\begin{aligned}
I &:= d_{\mathcal{W}}^2\left(\mu_{t+h}^{t,\nu,\omega}, \exp_\mu(h\cdot F_\varpi[\nu])\right) \\
&\leqslant \int_{x\in\mathbb{R}^d} \left|\Phi_{t+h}^t(x) - (x + hF_\varpi[\nu](x))\right|^2 \mathrm{d}\nu(x) \\
&\leqslant \int_{x\in\mathbb{R}^d} h \int_{s=t}^{t+h} \left|F_\varpi[\mu_s^{t,\nu,\omega}](\Phi_s^t(x)) - F_\varpi[\nu](x)\right|^2 \mathrm{d}s\mathrm{d}\nu(x) \\
&\leqslant \int_{x\in\mathbb{R}^d} h \int_{s=t}^{t+h} \int_{u\in U} \left|f(\Phi_s^t(x), \mu_s^{t,\nu,\omega}, u) - f(x,\nu,u)\right|^2 \mathrm{d}\varpi(u)\mathrm{d}s\mathrm{d}\nu(x) \\
&\leqslant \int_{x\in\mathbb{R}^d} h \int_{s=t}^{t+h} [f]^2 \left(d_{\mathcal{W}}(\mu_s^{t,\nu,\omega},\nu) + \left|\Phi_s^t(x) - x\right|\right)^2 \mathrm{d}s\mathrm{d}\nu(x).
\end{aligned}$$

Using (4.3) and (4.4),

$$\begin{aligned}
I &\leqslant \int_{x\in\mathbb{R}^d} h \int_{s=t}^{t+h} [f]^2 \left((s-t)\left([f]\left(|x| + d_{\mathcal{W}}(\nu,\delta_0)\right) + 2\,|f|_{0,\infty}\right)\mathrm{e}^{[f](s-t)}\right)^2 \mathrm{d}s\mathrm{d}\nu(x) \\
&\leqslant \int_{x\in\mathbb{R}^d} h^2[f]^2 \left(h\left([f]\left(|x| + d_{\mathcal{W}}(\nu,\delta_0)\right) + 2\,|f|_{0,\infty}\right)\mathrm{e}^{[f]h}\right)^2 \mathrm{d}\nu(x) \\
&\leqslant \left(2h^2\,[f]\left([f]\,d_{\mathcal{W}}(\nu,\delta_0) + |f|_{0,\infty}\right)\mathrm{e}^{[f]h}\right)^2.
\end{aligned}$$

Hence

$$\lim_{h\searrow 0} \frac{d_{\mathcal{W}}\left(\mu_{t+h}^{t,\nu,\omega}, \exp_\mu(h\cdot F_\varpi[\nu])\right)}{h} \leqslant \lim_{h\searrow 0} 2h\,[f]\left([f]\,d_{\mathcal{W}}(\nu,\delta_0) + |f|_{0,\infty}\right)\mathrm{e}^{[f]h} = 0.$$

□

Appendix B. Details on locally uniform upper semicontinuity

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

$$\lim_{n\to\infty} \sup_{y\in B_n} u(y) \leqslant \sup_{x\in B} u(x). \tag{B.1}$$

It turns out that we have the following.

Lemma B.1 (Link with other notions of upper semicontinuity)**.** *The following holds.*

1. *The condition (B.1) is strictly weaker than continuity.*

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

Proof. 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 (B.1) directly stands. On the other hand, assume that u is luusc. Consider a sequence of bounded nonempty closed sets $A_n \subset Y$ that converge in the Hausdorff distance, that is, there exists $A \subset \mathcal{S}$ such that

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

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

$$d_{H,Y}(B_n, A) = \sup_{x\in B_n}\inf_{y\in A} d_Y(x,y) = \sup_{m\geqslant n}\sup_{x\in A_n}\inf_{y\in A} d_Y(x,y) = \sup_{m\geqslant n} d_{H,Y}(A_n, A) \underset{n\to\infty}{\longrightarrow} 0.$$

Hence

$$\overline{\lim_{n\to\infty}}\, U(A_n) = \lim_{n\to\infty}\sup_{m\geqslant n}\sup_{y\in A_m} u(y) \leqslant \lim_{n\to\infty}\sup_{y\in B_n} u(y) \leqslant \sup_{x\in A} u(x) = U(A),$$

and U is locally upper semicontinuous, proving 2.

We turn to Point 3. If u is locally uniformly continuous, let $m_u : \mathbb{R}^+ \to \mathbb{R}^+$ 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 $d_Y(y_n, x_n) \leqslant \varepsilon_n \to_n 0$, and

$$\lim_{n\to\infty}\sup_{y\in B_n} u(y) = \lim_{n\to\infty} u(y_n) \leqslant \overline{\lim_{n\to\infty}}\, u(x_n) + m\left(d_Y(y_n, x_n)\right) \leqslant \sup_{x\in B} u(x) + 0.$$

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

$$\begin{aligned}&\lim_{r\to 0}\sup\left\{|u(x)-u(y)| \;\middle|\; x,y\in\overline{\mathscr{B}}(o,R),\ d(x,y)\leqslant r\right\}\\ &\leqslant \max_{s\in\{-1,1\}}\lim_{n\to\infty}\sup_{\substack{(x,y)\overline{\mathscr{B}}(o,R)^2\\ d(x,y)\leqslant 1/n}} s(u(x)-u(y)) \leqslant 0.\end{aligned}$$

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

Points 4 and 5 will use similar counterexamples. Notice first that taking $B_n = \overline{\mathscr{B}}(x, 1/n)$ in (B.1) for each $x \in \mathbb{R}^{\mathrm{d}}$, we see that luusc always imply usc. On the other hand, if the closed balls of Y are compact, let $u : Y \to \mathbb{R}$ be upper semicontinuous. Any maximizing sequence $(y_n)_n$ with $y_n \in B_n$ is in particular contained in a sufficiently large closed ball that is independent of n, hence contains a converging subsequence, whose limit belongs to $\overline{B} = B$ owing to the uniform approximation of B by B_n. By upper semicontinuity of y, (B.1) is satisfied.

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

$$G_r := \left\{ \mu_\alpha := \left(1 - \frac{1}{\alpha^2}\right)\delta_0 + \frac{1}{\alpha^2}\delta_{(\alpha,0,\cdots,0)} \;\middle|\; \alpha \geqslant r \right\},$$

and

$$H := \left\{ \nu_\beta := \left(1 - \frac{1}{\beta}\right)\mu_\beta + \frac{1}{\beta}\delta_{(0,1,0,\cdots,0)} \;\middle|\; \beta \geqslant 1 \right\}.$$

We have that for all $r \geqslant 1$, the sets G_r, H are nonempty, disjoint and included in $\overline{\mathscr{B}}(\delta_0, 1)$. Moreover, G_r and H are closed in $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ for each $r \geqslant 1$: 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 δ_0: but this is absurd since $d_{\mathcal{W}}(\mu_\alpha, \delta_0) = 1$ for all $\alpha > 0$. Thus $(\alpha_i)_i$ is bounded by some constant $C > 0$, and some (non relabeled) subsequence converges towards some $\overline{\alpha} \geqslant r$. Computing

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

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

Consider $u : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ 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

$$B_n := \left\{ \mu \in \overline{\mathscr{B}}(\delta_0, 1) \;\middle|\; \left|\mathrm{Bary}_{\mathbb{R}^{\mathrm{d}}}(\mu)\right| := \left|\int_{x \in \mathbb{R}^{\mathrm{d}}} x \mathrm{d}\mu(x)\right| \leqslant \frac{1}{n} \right\}.$$

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

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

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

$$\lim_{n\to\infty} \sup_{\mu \in B_n} u(\mu) \geqslant \lim_{n\to\infty} u(\mu_n) = 1, \qquad \sup_{x \in B} u(\mu) = 0.$$

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 δ_0. If $(\alpha_i)_i$ is bounded, then we showed in Point 4 that $\{\mu_\alpha \mid \alpha \in [1, C]\}$ is compact in $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, thus narrowly compact, and the narrow limit stays in G_1.

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

$$B_n \coloneqq H \cup G_n.$$

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 δ_0. Moreover,

$$\begin{aligned}\sup_{\mu\in B_n} \inf_{\nu\in B} d_{\mathcal{W}}(\mu_n, \nu) &= \sup_{\alpha\geqslant n} \inf_{\nu\in H} d_{\mathcal{W}}(\mu_\alpha, \nu) \\ &\leqslant \sup_{\alpha\geqslant n} d_{\mathcal{W}}\left(\mu_\alpha, \left(1-\frac{1}{\alpha}\right)\mu_\alpha + \frac{1}{\alpha}\delta_{(0,1,0,\cdots,0)}\right) \leqslant \sup_{\alpha\geqslant n}\frac{1}{\alpha} \underset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

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

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

Hence u is not locally uniformly upper semicontinuous. The application $\mu \mapsto d^2_{\mathcal{W}}(\mu, \delta_0)$ 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 5.7. □

Appendix C. Perturbed optimization

Lemma C.1 (Perturbed optimization)**.** *Denote $Y = \overline{X}^2 = [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, endowed with the distance*

$$d_Y^2((t,\mu,s,\nu),(\bar{t},\overline{\mu},\overline{s},\overline{\nu})) \coloneqq \left|t-\bar{t}\right|^2 + d^2_{\mathcal{W}}(\mu,\overline{\mu}) + |s-\overline{s}|^2 + d^2_{\mathcal{W}}(\nu,\overline{\nu}).$$

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

$$\left\{z \in Y \;\middle|\; \Phi(z) \geqslant \Phi(z^0)\right\} \subset \overline{\mathscr{B}}(z^0, R). \tag{C.1}$$

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

1. *the perturbed map $\Phi - p_n$ reaches a global strict maximum in z_n,*
2. *There exists an application $\omega_{T,R,A} : \mathbb{N} \to \mathbb{R}^+$ such that*

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

3. *The map $(t,\mu) \mapsto p_n(t,\mu,s_n,\nu_n)$ belongs to $\mathscr{T}_+$, and $(s,\nu) \mapsto -p_n(t_n,\mu_n,s,\nu)$ belongs to $\mathscr{T}_-$,*
4. *There holds $\sup \Phi \leqslant \Phi(z_n) + C_{R,A}(n)$, where $C_{R,A}$ is decreasing towards 0 when $n \to \infty$.*

Proof. The metric space (Y, d_Y) is complete, and Φ satisfies all the assumptions of the Ekeland–Borwein–Preiss–Zhu theorem [49], Theorem 2.5.2. We consider the gauge-type function d_Y^2, and the choice of ponderation

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

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

$$\begin{cases} d_Y^2(z_0, z_n) \leqslant \dfrac{A}{\alpha_{n,0}}, \qquad d_Y^2(z_{n,m}, z_n) \leqslant \dfrac{A}{2^m \alpha_{n,0}} \qquad \forall m \in \mathbb{N}, & \text{(C.2a)} \\ \Phi(z_n) \geqslant \Phi(z^0) + \displaystyle\sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(z^0, z_{n,m}), & \text{(C.2b)} \\ \Phi(z_n) - \displaystyle\sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(z_n, z_{n,m}) > \Phi(\cdot) - \displaystyle\sum_{m\in\mathbb{N}} \alpha_{n,m} d_Y^2(\cdot, z_{n,m}) \quad \forall z \neq z_n. & \text{(C.2c)} \end{cases}$$

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

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

Hence the application p_n is well-defined from Y to $\mathbb{R}^+$. By (C.2c), $\Phi - p_n$ 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 $c + \sum_{m\in\mathbb{N}} 2^{-m-1} \frac{|\cdot - t_{n,m}|^2}{n}$, over a bounded interval, so uniformly convergent. By direct computation, its derivative is Lipschitz in $[0, T]$ with constant $2/n$, and

$$|\partial_t p_n(t, \mu_n, s_n, \nu_n)| \leqslant \frac{1}{n} \sum_{m\in\mathbb{N}} 2^{-m} |t - t_{n,m}| \leqslant \frac{2T}{n} \qquad \forall t \in [0, T]. \tag{C.3}$$

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 $\overline{X}$. Moreover, by (C.2b), $z_n \in \{\Phi \geqslant \Phi(z^0)\}$. As $\mu_{n,m} \to_m \mu_n$, the sequence $(\mu_{n,m})_m$ stays in a bounded set of $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, 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 [9], Thm. 7.3.2), there holds for any $(\sigma, \xi) \in \mathbf{Tan}\, \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and $h \in [0, 1]$ that

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

Thus $p_n(t_n, \cdot, s_n, \nu_n)$ is locally semiconcave. As a combination of squared distance, it is geometrically consistent. To prove that $(t, \mu) \mapsto p_n(t, \mu, s_n, \nu_n)$ belongs to $\mathscr{T}_+$, there only stays to show the local Lipschitzianity in the measure variable. By direct computation, for any $S > 0$ and $\mu, \sigma \in \overline{\mathscr{B}}(\delta_0, S)$, one has

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

Here we used (C.2a). Let $R > 0$ be given by the assumption (C.1) such that $d_Y\left(z_n, z^0\right) \leqslant R$ independantly of n. By the above, there holds

$$(1 + d_{\mathcal{W}}(\mu_n,\delta_0))\,\|D_\mu p_n(z_n)\|_{\mu_n} \leqslant (1+R)\,\frac{R+\sqrt{2nA}}{n}\sum_{m\in\mathbb{N}}\frac{1+2^{-m/2}}{2^m}. \tag{C.4}$$

Gathering (C.3) and (C.4), 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 Φ over Y is the same as the supremum of Φ over $\overline{\mathscr{B}}(z^0,R)$. In consequence, (C.2c) gives

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

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

Appendix D. Monotonicities

Lemma D.1 (Monotonicities)**.** *Assume (5.10), and that $u, -v$ are bounded from above. Let Γ, Γ_ι and $\Gamma_{\iota\varepsilon}$ be defined as in (5.10), (5.15) and (5.16). Then*

$$\Gamma_{\iota\varepsilon} \underset{\varepsilon\searrow 0}{\searrow} \Gamma_\iota \qquad \textit{for all } \iota > 0, \qquad \textit{and} \qquad \Gamma_\iota \underset{\iota\searrow 0}{\nearrow} \Gamma_0 \geqslant \Gamma.$$

Proof. Define the additional variables

$$\begin{aligned}\Gamma_r &:= \min\left(1, \sup\left\{u(t,\mu) - v(s,\nu)\;\middle|\;(t,\mu,s,\nu)\in\overline{X}^2,\; d_X((t,\mu),(s,\nu)) \leqslant r\right\}\right),\\ \Gamma_{\iota r} &:= \sup\left\{\phi_\iota(t,\mu,s,\nu)\;\middle|\;(t,\mu,s,\nu)\in\overline{X}^2,\; d_X((t,\mu),(s,\nu)) \leqslant r\right\}\end{aligned}$$

so that $\Gamma = \lim_{r\to 0}\Gamma_r$ and $\Gamma_\iota = \lim_{r\to 0}\Gamma_{\iota r}$. For each fixed $\iota > 0$, using the growth of h, the variables $\Gamma_{\iota r}$ and $\Gamma_{\iota\varepsilon}$ are upper bounded. Restricting ι, ε and r to $]0,1]$, (5.10) gives us that each term is lower bounded. Moreover, we have the monotonicities

$$\Gamma_r \searrow_r, \qquad \Gamma_\iota \nearrow_\iota, \qquad \Gamma_{\iota r} \searrow_r, \qquad \Gamma_{\iota\varepsilon} \searrow_\varepsilon .$$

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

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

$$u(t_\alpha,\mu_\alpha) - v(s_\alpha,\nu_\alpha) - \iota\left(h_{\circ d}(t_\alpha,\mu_\alpha) + h_{\circ d}(s_\alpha,\nu_\alpha)\right) \leqslant u(t_\alpha,\mu_\alpha) - v(s_\alpha,\nu_\alpha) - \frac{\alpha}{2}.$$

Letting $\iota \searrow 0$, we obtain a contradition. Hence $\Gamma_0 \geqslant \Gamma$.

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

$$\phi_\iota(z_{\alpha r}) - \frac{d_X^2((t_{\alpha r},\mu_{\alpha r}),(s_{\alpha r},\nu_{\alpha r}))}{\varepsilon} \leqslant \phi_\iota(z_{\alpha r}) - \frac{\alpha}{2}.$$

Hence

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

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

Finally, assume that for some fixed ι, $\Gamma_{\iota 0} \coloneqq \lim_{\varepsilon\searrow 0}\Gamma_{\iota\varepsilon} > \Gamma_\iota$. Then for sufficiently small r, there exists $\alpha > 0$ such that $\Gamma_{\iota\varepsilon} \geqslant \Gamma_{\iota 0} \geqslant \Gamma_{\iota r} + \alpha \geqslant \Gamma_\iota$ 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

$$\phi_\iota(z_{\varepsilon\alpha}) - \frac{d_X^2((t_{\varepsilon\alpha},\mu_{\varepsilon\alpha}),(s_{\varepsilon\alpha},\nu_{\varepsilon\alpha}))}{\varepsilon} \geqslant \Gamma_{\iota\varepsilon} - \frac{\alpha}{2} \geqslant \Gamma_\iota + \frac{\alpha}{2}$$

implies

$$\frac{d_X^2((t_{\varepsilon\alpha},\mu_{\varepsilon\alpha}),(s_{\varepsilon\alpha},\nu_{\varepsilon\alpha}))}{\varepsilon} \leqslant [\![u]\!] + [\![-v]\!] + 0 - \Gamma_\iota + \frac{\alpha}{2},$$

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

$$\phi_\iota(z_{\varepsilon\alpha}) - \frac{d_{X_1}^2((t_{\varepsilon\alpha},\mu_{\varepsilon\alpha}),(s_{\varepsilon\alpha},\nu_{\varepsilon\alpha}))}{\varepsilon} \geqslant \phi_\iota(z_{\varepsilon\alpha}) + \frac{\alpha}{2},$$

a flagrant contradiction. □

Appendix E. Grönwall estimates

Lemma E.1 (Grönwall estimates). *Assume [A1] and [A2]. Let $(\mu_s^{t,\nu,\omega})_{s\in[t,T]}$ denote the solution of (4.1) issued from $(t,\nu) \in [0,T]\times\mathscr{P}_2(\mathbb{R}^d)$ and driven by the control $\omega \in L^1([t,T];\mathscr{P}(U))$. Let $0 \leqslant t \leqslant \bar{s} \leqslant s \leqslant T$,*

$(\nu,\overline{\nu}) \in \left(\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})\right)^2$ *and* $(\omega,\overline{\omega}) \in \left(L^1([t,T];\mathscr{P}(U))\right)^2$. *Then*

$$d_{\mathcal{W}}\left(\mu_s^{t,\nu,\omega},\mu_{\overline{s}}^{t,\overline{\nu},\overline{\omega}}\right) \leqslant (s-\overline{s})\left([f]\,(\overline{s}-t)\mathrm{e}^{(s-t)[f]}+\mathrm{e}^{(s-\overline{s})[f]}\right)\left([f]\,d_{\mathcal{W}}(\delta_0,\nu)+|f|_{0,\infty}\right) + \mathrm{e}^{[f](\overline{s}-t)\left(1+\mathrm{e}^{[f](\overline{s}-t)}\right)}\left(d_{\mathcal{W}}(\nu,\overline{\nu})+E_{t,\overline{s},\overline{\nu},\omega,\overline{\omega}}\right),$$

where

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

and

$$d_{\mathcal{W},m_{f,u}}(\varpi,\overline{\varpi}) \coloneqq \inf_{\alpha\in\Gamma(\varpi,\overline{\varpi})}\int_{(u,v)\in U^2} m_{f,u}(|u-v|)\mathrm{d}\alpha(u,v).$$

As particular cases, we record that

$$d_{\mathcal{W}}(\mu_{t+h}^{t,\nu,\omega},\nu) \leqslant h\left([f]\,d_{\mathcal{W}}(\delta_0,\nu)+|f|_{0,\infty}\right)\mathrm{e}^{h[f]},$$
$$d_{\mathcal{W}}\left(\mu_T^{t,\nu,\omega},\mu_T^{t,\overline{\nu},\omega}\right) \leqslant \mathrm{e}^{[f](T-t)\left(1+\mathrm{e}^{[f](T-t)}\right)}d_{\mathcal{W}}(\nu,\overline{\nu}).$$

Proof. Assume w.l.o.g that $s \geqslant \overline{s}$. Denote $(\Phi_\tau^{t,\nu,\omega})_{\tau\in[t,s]}$ and $\left(\Phi_\tau^{t,\overline{\nu},\overline{\omega}}\right)_{\tau\in[t,\overline{s}]}$ the respective fluxes of the ODEs

$$\dot{y}_\tau = F_{\omega(\tau)}[\mu_\tau^{t,\nu,\omega}](y_\tau), \qquad \dot{\overline{y}}_\tau = F_{\overline{\omega}(\tau)}[\mu_\tau^{t,\overline{\nu},\overline{\omega}}](\overline{y}_\tau).$$

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

$$\left|\Phi_\tau^{t,\nu,\omega}(x)-\Phi_r^{t,\nu,\omega}(x)\right| \leqslant \int_{\theta=r}^{\tau}\int_{u\in U}\left|f\left(\Phi_\theta^{t,\nu,\omega}(x),\mu_\theta^{t,\nu,\omega},u\right)\right|\mathrm{d}\omega(\theta)(u)\mathrm{d}\theta$$
$$\leqslant \int_{\theta=r}^{\tau}\left[[f]\left|\Phi_\theta^{t,\nu,\omega}(x)-\Phi_r^{t,\nu,\omega}(x)\right|+[f]\left|\Phi_r^{t,\nu,\omega}(x)\right|+|f|_{0,\infty}\right]\mathrm{d}\theta,$$

so that a Grönwall lemma yields

$$\left|\Phi_\tau^{t,\nu,\omega}(x)-\Phi_r^{t,\nu,\omega}(x)\right| \leqslant (\tau-r)\left([f]\left|\Phi_r^{t,\nu,\omega}(x)\right|+|f|_{0,\infty}\right)\mathrm{e}^{(\tau-r)[f]}. \tag{E.1}$$

In particular,

$$\left|\Phi_\tau^{t,\nu,\omega}(x)-\Phi_r^{t,\nu,\omega}(x)\right| \leqslant (\tau-r)\left([f]\left|\Phi_r^{t,\nu,\omega}(x)-x\right|+[f]\,|x|+|f|_{0,\infty}\right)\mathrm{e}^{(\tau-r)[f]}$$
$$\leqslant (\tau-r)\left([f]\,(r-t)\mathrm{e}^{(r-t)[f]}+1\right)\left([f]\,|x|+|f|_{0,\infty}\right)\mathrm{e}^{(\tau-r)[f]}.$$

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

$$\begin{aligned} d_{\mathcal{W}}(\mu_\tau^{t,\nu,\omega},\mu_r^{t,\nu,\omega}) &\leqslant \sqrt{\int_{x\in\mathbb{R}^{\mathrm{d}}}\left|\Phi_\tau^{t,\nu,\omega}(x)-\Phi_r^{t,\nu,\omega}(x)\right|^2\mathrm{d}\nu(x)} \\ &\leqslant (\tau-r)\left([f]\,(r-t)\mathrm{e}^{(r-t)[f]}+1\right)\left([f]\,d_{\mathcal{W}}(\delta_0,\nu)+|f|_{0,\infty}\right)\mathrm{e}^{(\tau-r)[f]}. \end{aligned} \tag{E.2}$$

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

$$\begin{aligned}
&\left|\Phi_\tau^{t,\nu,\omega}(x) - \Phi_\tau^{t,\overline{\nu},\overline{\omega}}(y)\right| \\
&\leqslant |x-y| + [f] \int_{r=t}^{\tau} \left|\Phi_r^{t,\nu,\omega}(x) - \Phi_r^{t,\overline{\nu},\overline{\omega}}(y)\right| + d_{\mathcal{W}}\left(\mu_r^{t,\nu,\omega}, \mu_r^{t,\overline{\nu},\overline{\omega}}\right) dr \\
&\qquad + \left(1 + (\tau - t)\left(|f|_{0,\infty} + [f]\,|y|\right) \mathrm{e}^{(\tau-t)[f]}\right) \int_{r=t}^{\tau} d_{\mathcal{W},m_{f,u}}\left(\omega(r), \overline{\omega}(r)\right) dr.
\end{aligned}$$

Applying a second Grönwall lemma,

$$\left|\Phi_\tau^{t,\nu,\omega}(x) - \Phi_\tau^{t,\overline{\nu},\overline{\omega}}(y)\right| \leqslant \left(|x-y| + [f] \int_t^{\tau} d_{\mathcal{W}}\left(\mu_r^{t,\nu,\omega}, \mu_r^{t,\overline{\nu},\overline{\omega}}\right) dr + E_{t,\tau,y,\omega,\overline{\omega}}\right) \mathrm{e}^{[f](\tau-t)},$$

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

$$\eta_{\overline{s}} \coloneqq \left(\Phi_{\overline{s}}^{t,\nu,\omega}, \Phi_{\overline{s}}^{t,\overline{\nu},\overline{\omega}}\right) \# \eta$$

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

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

As

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

we are ready to apply our third Grönwall lemma to get

$$d_{\mathcal{W}}(\mu_{\overline{s}}^{t,\nu,\omega}, \mu_{\overline{s}}^{t,\overline{\nu},\overline{\omega}}) \leqslant \mathrm{e}^{[f](\overline{s}-t)}\left(d_{\mathcal{W}}(\nu, \overline{\nu}) + E_{t,\overline{s},\overline{\nu},\omega,\overline{\omega}}\right) \mathrm{e}^{[f](\overline{s}-t)\mathrm{e}^{[f](\overline{s}-t)}}. \tag{E.3}$$

Combining (E.2) and (E.3), we get the desired result. □

References

[1] R. Coyaud, *Study of Approximations of Optimal Transport Problems and Application to Physics*. These de doctorat, Paris Est (2021).

[2] A. Corbetta, *Multiscale Crowd Dynamics: Physical Analysis, Modeling and Applications*. PhD thesis, Eindhoven University of Technology (2016).

[3] B. Piccoli and A. Tosin, Pedestrian flows in bounded domains with obstacles. *Continuum Mech. Thermodyn.* **21** (2009) 85–107.

[4] 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*, edited by A. Muntean and F. Toschi. CISM International Centre for Mechanical Sciences. Springer, Vienna (2014) 1–46.

[5] M.G. Crandall, H. Ishii and P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc.* **27** (1992) 1–67.

[6] H. Ishii, Hamilton–Jacobi equations with discontinuous Hamiltonians on arbitrary open sets. *Bull. Fac. Sci. Eng.* **28** (1985) 33–77.

[7] F. Otto, The geometry of dissipative evolution equations: the porous medium equation. *Commun. Part. Differ. Equ.* **26** (2001) 101–174.

[8] C. Villani, Optimal transport. Vol. 338 of *Grundlehren Der Mathematischen Wissenschaften.* Springer Berlin Heidelberg, Berlin, Heidelberg (2009).

[9] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows. Lectures in Mathematics ETH Zürich. Birkhäuser-Verlag, Basel (2005).

[10] 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).

[11] P.-L. Lions, Jeux à champ moyen, 2006/2007. Conférences au Collège de France.

[12] W. Gangbo and A. Tudorascu, On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *J. Math. Pures Appl.* **125** (2019) 119–174.

[13] C. Bertucci and P.L. Lions, An approximation of the squared Wasserstein distance and an application to Hamilton–Jacobi equations, September 2024. Preprint, available at http://arxiv.org/abs/2409.11793.

[14] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham and M. Rosestolato, Master Bellman equation in the Wasserstein space: uniqueness of viscosity solutions. *Trans. Am. Math. Soc.* **377** (2024) 31–83.

[15] A. Cosso and M. Martini, On smooth approximations in the Wasserstein space. *Electron. Commun. Probab.* **28** (2023) 1–11.

[16] S. Daudin, J. Jackson and B. Seeger, Well-posedness of Hamilton–Jacobi equations in the Wasserstein space: non-convex Hamiltonians and common noise, Dec. 2023. Preprint (arXiv:2312.02324).

[17] G. Conforti, R. Kraaij and D. Tonon, Hamilton–Jacobi equations for controlled gradient flows: Cylindrical test functions, 2023. Preprint, available at https://arxiv.org/abs/2302.06571.

[18] G. Conforti, R.C. Kraaij, L. Tamanini and D. Tonon, Hamilton-Jacobi equations for Wasserstein controlled gradient flows: existence of viscosity solutions (2024).

[19] G. Conforti, R.C. Kraaij and D. Tonon, Hamilton-Jacobi equations for controlled gradient flows: The comparison principle. *J. Funct. Anal.* **284** (2023). 1–53.

[20] J. Feng and M. Katsoulakis, A comparison principle for Hamilton–Jacobi equations related to controlled gradient flows in infinite dimensions. *Arch. Rational Mech. Anal.* **192** (2009) 275–310.

[21] Z. Badreddine and H. Frankowska, Solutions to Hamilton–Jacobi equation on a Wasserstein space. *Calc. Var. Part. Differ. Equ.* **61** (2021) 9.

[22] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *Int. Game Theory Rev.* **10** (2008) 1–16.

[23] S. Daudin and B. Seeger, A comparison principle for semilinear Hamilton–Jacobi–Bellman equations in the Wasserstein space, Aug. 2023. Preprint (arXiv:2308.15174).

[24] C. Jimenez, A. Marigonda and M. Quincampoix, Optimal control of multiagent systems in the Wasserstein space. *Calc. Var. Part. Differ. Equ.* **59** (2020) 1–45.

[25] A. Marigonda and M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions. *J. Differ. Equ.* **264** (2018) 3212–3252.

[26] L. Ambrosio and J. Feng, On a class of first order Hamilton–Jacobi equations in metric spaces. *J. Differ. Equ.* **256** (2014) 2194–2245.

[27] C. Bertucci, Stochastic optimal transport and Hamilton–Jacobi–Bellman equations on the set of probability measures, June 2023. Preprint, available at http://arxiv.org/abs/2306.04283.

[28] F. Jean, O. Jerhaoui and H. Zidani, Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition. *SIAM J. Math. Anal.* **56** (2024) 3326–3356.

[29] O. Jerhaoui, *Viscosity Theory of First Order Hamilton Jacobi Equations in Some Metric Spaces.* PhD thesis, Institut Polytechnique de Paris, Paris (2022).

[30] G. Fabbri, F. Gozzi and A. Święch, Stochastic optimal control in infinite dimension. Vol. 82 of *Probability Theory and Stochastic Modelling.* Springer International Publishing, Cham (2017).

[31] N. Gigli, Second order analysis on $(\mathrm{P}_2(\mathrm{M}),\mathrm{W}_2)$. *Mem. Am. Math. Soc.* **216** (2009) 1–147.

[32] N. Gigli, On the inverse implication of Brenier–McCann theorems and the structure of $(\mathrm{P}_2(\mathrm{M}),\mathrm{W}_2)$. *Methods Applic. Anal.* **18** (2011) 127–158.

[33] F. Santambrogio, Optimal transport for applied mathematicians. Vol. 87 of *Progress in Nonlinear Differential Equations and Their Applications.* Springer International Publishing, Cham (2015).

[34] B. Piccoli, Measure differential equations. *Arch. Rational Mech. Anal.* **233** (2019) 1289–1317.

[35] B. Bonnet and H. Frankowska, Differential inclusions in Wasserstein spaces: The Cauchy–Lipschitz framework. *J. Differ. Equ.* **271** (2021) 594–637.

[36] B. Bonnet and H. Frankowska, Caratheodory Theory and A Priori Estimates for Continuity Inclusions in the Space of Probability Measures, May 2023. Preprint (arXiv:2302.00963).

[37] G. Cavagnari, A. Marigonda and B. Piccoli, Superposition Principle for Differential Inclusions, in *Large-Scale Scientific Computing*, Vol. 10665, edited by I. Lirkov and S. Margenov. Springer International Publishing, Cham (2018) 201–209.

[38] P. Cannarsa and C. Sinestrari, *Semiconcave Functions, Hamilton—Jacobi Equations, and Optimal Control.* Birkhäuser, Boston, MA (2004).

[39] P. Del Moral and M. Doisy, Maslov Idempotent Probability Calculus, I. *Theory Probab. Applic.* **43** (1999) 562–576.

[40] V.N. Kolokoltsov and V.P. Maslov, Idempotent Analysis and Its Applications. Springer Netherlands, Dordrecht (1997).

[41] P. Cardaliaguet, Notes on Mean Field Games. Available at https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf, 2013.

[42] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions, The Master Equation and the Convergence Problem in Mean Field Games. Vol. 201 of *Annals of Mathematics Studies.* Princeton University Press, Princeton, NJ (2019).

[43] R. Carmona and F. Delarue, Probabilistic Theory of Mean Field Games with Applications II. Vol. 84 of *Probability Theory and Stochastic Modelling.* Springer International Publishing (2018).

[44] W. Gangbo, T. Nguyen and A. Tudorascu, Hamilton–Jacobi equations in the Wasserstein space. *Methods Applic. Anal.* **15** (2008) 155–184.

[45] I. Ekeland, On the variational principle. *J. Math. Anal. Applic.* **47** (1974) 324–353.

[46] W. Gangbo and A. Święch, Metric viscosity solutions of Hamilton–Jacobi equations depending on local slopes. *Calc. Var. Part. Differ. Equ.* **54** (2015) 1183–1218.

[47] X. Li and J. Yong, Optimal Control Theory for Infinite Dimensional Systems. Birkhäuser, Boston, MA (1995).

[48] C. Wu and J. Zhang, Viscosity solutions to parabolic master equations and McKean–Vlasov SDEs with closed-loop controls. *Ann. Appl. Probab.* **30** (2020) 936–986.

[49] J.M. Borwein and Q.J. Zhu, Techniques of Variational Analysis. CMS Books in Mathematics. Springer-Verlag, New York (2005).