# A MINIMALITY PROPERTY OF THE VALUE FUNCTION IN OPTIMAL CONTROL ON SPACES OF PROBABILITY MEASURES*

AVERIL AUSSEDAT† AND CRISTOPHER HERMOSILLA‡

**Abstract.** An optimal control problem with (possibly) unbounded terminal cost is considered in $\mathscr{P}_2(\mathbb{R}^d)$, the space of Borel probability measures with finite second moment. We consider a suitable weak topology rendering $\mathscr{P}_2(\mathbb{R}^d)$ locally compact. In this setting, we show that the value function of a control problem is the minimal viscosity supersolution of an appropriate Hamilton-Jacobi-Bellman (HJB) equation. Additionally, if the terminal cost is bounded and continuous, we show that the value function is the unique viscosity solution of the HJB equation.

**Key words.** Wasserstein space, Optimal control problems, viscosity solutions, weak topology

**MSC codes.** 35F21, 35R06, 49Lxx

**1. Introduction.** We consider a Hamilton-Jacobi equation arising from an optimal control problem whose state space is a set of measures. More precisely, the unknown of our equation belongs to $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, the space of Borel probability measures on $\mathbb{R}^{\mathrm{d}}$ with finite second moment. It is well-known that this setting is suitable for the modelling of optimal control of population dynamics in crowd motion [35, 16] or biology [14]. In these approaches, the configuration at time $s \in [0,T]$ of a population is represented by a measure $\boldsymbol{\mu}_s \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, and the evolution in time is assumed to satisfy a continuity equation of the form

$$\partial_s \boldsymbol{\mu}_s + \operatorname{div}(\mathsf{f}_s \boldsymbol{\mu}_s) = 0, \quad s \in [0,T], \qquad \boldsymbol{\mu}_0 = \nu. \tag{1.1}$$

The equation (1.1) is understood in the sense of distributions, and is the measure counterpart of an ODE with initial term $\nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and vector fields $\mathsf{f}_s \in \mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$. In this work, we study a controlled version of the continuity equation. Given some set $U \subset \mathbb{R}^{\kappa}$ and a dynamic $f : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U \to \mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$ we consider, for $u : [0,T] \to U$ measurable and $(t,\nu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, the associated *controlled continuity equation*

$$\partial_s \boldsymbol{\mu}_s + \operatorname{div}\left(f[s,\boldsymbol{\mu}_s,u(s)]\ \boldsymbol{\mu}_s\right) = 0, \quad s \in [t,T], \qquad \boldsymbol{\mu}_t = \nu. \tag{1.2}$$

Let $(\boldsymbol{\mu}_s^{t,\nu,u})_{s\in[t,T]}$ be the solution of (1.2), whose meaning and wellposedness will be discussed in Section 3; see also [8]. We are concerned with a Mayer type problem associated with a cost $\mathfrak{J} : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{+\infty\}$, which may enforce terminal state constraints when taking the value $+\infty$. To compute the optimal control, a general approach is to study the value function $V : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{\pm\infty\}$ given by

$$V(t,\nu) \coloneqq \inf\left\{ \mathfrak{J}(\boldsymbol{\mu}_T^{t,\nu,u}) \;\middle|\; u : [t,T] \to U \text{ is measurable} \right\}. \tag{1.3}$$

From the theory of Euclidean spaces, it is expected that the value function given in (1.3) satisfies the *Dynamic Programming Principle* (DPP)

$$V(t,\nu) = \inf\left\{ V\left(s,\boldsymbol{\mu}_s^{t,\nu,u}\right) \mid u : [s,T] \to U \text{ is measurable} \right\}, \text{ for } 0 \le t \le s \le T. \tag{1.4}$$

*Submitted to the editors DATE.

**Funding:** This work was supported by the Center for Mathematical Modeling (CMM) and ANID-Chile under BASAL funds for Center of Excellence FB210005 and Fondecyt Regular 1231049.

†INSA Rouen Normandie, Normandie Univ, LMI UR 3226, F-76000 Rouen, France. (averil.aussedat@insa-rouen.fr).

‡Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. (cristopher.hermosill@usm.cl).

The infinitesimal version of the DPP is the so-called *Hamilton-Jacobi-Bellman* (HJB) equation. A classical result of the HJB theory in Euclidian spaces is that whenever the end-point cost is real-valued and uniformly continuous, the corresponding value function is the unique solution of the HJB equation in the sense of viscosity solutions [18, 19]. In the case where the end-point cost may take infinite values, the value function may still be characterized as the smallest supersolution in the classical (Euclidean) sense. The purpose of this work is to transpose these results to problems where the state evolves in the space $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, that is, our main results is concerned with showing that the value function defined in (1.3) is the smallest supersolution of the HJB equation in an ad-hoc sense for the space $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$.

To be more precise, the HJB equation we consider in this paper is

$$-\partial_t v(t,\mu) + H\left(\mu, D_\mu v(t,\mu)\right) = 0. \tag{1.5}$$

Here $D_\mu v(t,\mu)$ is the directional derivative at $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ of $v(t,\cdot)$ along geodesics, which is a mapping defined on $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$, the subset of measures in $\mathscr{P}_2(\mathbb{R}^{2\mathrm{d}})$ whose first marginal is $\mu$; details are given in subsection 2.2. Accordingly, the (autonomous) *control Hamiltonian* in (1.5) is defined over the maps $p : \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu \to \mathbb{R}$ as follows

$$H(\mu,p) \coloneqq \sup_{u\in U} -p\left((\mathrm{Id}, f[\mu,u])\#\mu\right).$$

Here the symbol # is the pushforward operator. Similar definitions of $D_\mu v$ and $H$ are used in [30] to study viscosity solutions of general Hamilton-Jacobi equations in CAT(0) spaces, and in [29, 6] to treat the Wasserstein space by using its curvature.

The HJB equation we are interested in is a particular instance of a PDE on spaces of probability measures, which has attracted a lot of interest since the seminal work of Otto [34] on the porous media equation. The HJB equation, as infinitesimal version of the DPP, involves "derivatives" of functions defined over $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. This is one of the first issues we face when studying Hamilton-Jacobi equations on $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, since there is not a unique way to introduce differentiability in $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, and different notions of derivatives may be considered; in our case for instance we have chosen directional derivatives along geodesics. The corpus of results of the theory concerns gradient flows in the space of measures [4, 15], the master equation in the theory of mean-field games [13, 11], Hamiltonian systems and flow equations [3, 2] and optimal control problems [33, 31, 17, 7, 20]. As far as classical solutions are concerned, the prominent theory is the *Lions differentiability* [10], that gives a suitable definition of the gradient of a function defined on $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. The strength of this idea is to embed measures into an external Hilbert space, and use the Hilbertian structure therein. This strategy proved successful to study the master equation whenever the data is sufficiently smooth to expect a smooth solution [12, Thm. 1.5]. If such regularity is not achievable, the most effective strategy so far uses *semidifferentials* defined in an appropriate tangent space. Indeed, viscosity solutions may be defined by imposing some sign conditions on the sub and superdifferential of the solution at any point, thus accounting for the non-existence of a gradient. A standard choice in the literature is the *regular tangent space*; see [4]. This space comes from the study of continuity equations, and may be used to define viscosity solutions [25]. Let us mention also that the case of Eikonal-type equations can be treated more directly, as in [37] with Sobolev norms, or the metric viscosity notions of [2, 26, 28].

The notion of viscosity solution we consider is based on sub and supersolutions as usual; however, the space of test functions is made of functions that are only directionally differentiable along geodesics with respect to (w.r.t.) the measure variable;

see subsection 4.1 for details. This allows us to consider a large class of test functions, including the squared Wasserstein distance, and to avoid any smoothing argument.

Another of the main issues one needs to deal with when studying the HJB equation in $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ is the regularity of the value function and the associated topology. Recall that $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ may be endowed with different topologies, for example the narrow topology or the one induced by so-called *Wasserstein distance*; see [4, Chapter 7]. The main advantage of the narrow topology is that bounded sets w.r.t. the Wasserstein distance (i.e., sets with uniformly bounded second moment) are narrowly relatively compact. Therefore, for any $n \in \mathbb{N}$, each

$$\mathbf{K}_n := \left\{\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \mid \mathbf{m}_2(\mu) \leq n\right\}, \quad \text{where } \mathbf{m}_2(\mu) := \int_{\mathbb{R}^{\mathrm{d}}} |x|^2 d\mu, \tag{1.6}$$

is narrowly compact. However, it does not hold that any narrowly converging sequence is bounded w.r.t. the Wasserstein distance.

The results reported in [6] include a strong comparison principle for problems with terminal costs that are continuous w.r.t. the Wasserstein distance. To cope with the lack of local compactness of the Wasserstein space, the authors in [6] need to consider a restrictive notion of upper semicontinuity over the semisolutions. It is worthy to point out that many of the technicalities in [6] could be avoided using that $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ is narrowly locally compact. This in turn would enforce the continuity of the value function, not w.r.t. the Wasserstein topology, but w.r.t. a weaker topology. The main novelty of this work is precisely that, the use of an inductive topology on $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ that circumvents the issue described above. This topology is an intermediate one on $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, weaker than the Wasserstein but stronger than the narrow topology, built as the inductive limit of narrowly compact topological spaces.

Definition 1.1 (Inductive Topology [27, Definition 2.16]). *For each $n \in \mathbb{N}$, let $\mathbf{K}_n$ be given by* (1.6) *and denote $\mathbf{\Theta}_n$ the topology on $\mathbf{K}_n$ induced by the narrow topology. The topology $\mathbf{\Theta}$ is the strongest topology on $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ that lets the canonical injections $\iota_n : \mathbf{K}_n \to \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ be continuous from $(\mathbf{K}_n, \mathbf{\Theta}_n)$ to $\big(\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}), \mathbf{\Theta}\big)$.*

The interplay between this inductive topology and the notion of viscosity solution we have taken in this paper, allows us to show *mutatis mutandis*, and using ideas from [32], that the (discontinuous) value function given by (1.3) can be characterized as the smallest $\mathbf{\Theta}$-lower semicontinuous function that is a supersolution of the HJB equation (4.1), with terminal condition $V(T, \cdot) = \mathfrak{J}$ on $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$.

**1.1. Main contributions and standing assumptions.** The main results of the paper are in Section 5, and are proved for autonomous systems ($f$ does not depend on the time variable). However, the basic results concerning the value function in Section 3 are proved for non-autonomous systems. We describe now our main results for autonomous systems. First, assume that the $f[\cdot]$ is Lipschitz, that each set $f[\mu, U] \subset \mathcal{C}(\mathbb{R}^{\mathrm{d}}; \mathbb{R}^{\mathrm{d}})$ is convex, and that the cost $\mathfrak{J}$ satisfies the following.

*Hypothesis* 1.2. The function $\mathfrak{J} : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{+\infty\}$ is lower bounded, $\mathbf{\Theta}$-lower semicontinuous and it is not identically $+\infty$.

Then the value function $V$ given in (1.3) is the minimal supersolution of the HJB equation (1.5) in the sense of Definition 4.1; see Theorem 5.4. Second, in the case where $\mathfrak{J}$ is additionally bounded and $\mathbf{\Theta}$-continuous, we are able to prove that $V$ is actually the unique viscosity solution of (1.5). Additionally, we provide a strong comparison principle by a rather direct generalization of the arguments of [18]. The difficulty then lies in proving that the value function is itself $\mathbf{\Theta}$-lower semicontinuous.

In order to present the standing assumptions on the dynamics, we need to introduce some notation. The key point here is that the dynamics takes values in a proper subspace of $\mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$. This is due to the linear growth condition in Hypothesis 1.3. To be more precise, we define $\mathbf{X} := \left\{b \in \mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}}) \mid |b|_{\mathrm{ucc}} < \infty\right\}$ where

$$|b|_{\mathrm{ucc}} := \|b(0)\| + \sum_{n\in\mathbb{N}} \frac{1}{2^n} \sup_{x\in\mathbb{R}^{\mathrm{d}}} \{\|b(x)\| \mid \|x\| \leq n\}, \qquad \forall b \in \mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}}). \tag{1.7}$$

It is well-known that $\mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$ endowed with the topology of convergence on compact sets is a Fréchet space (see, e.g., [23, Fact 3.76]). The topology induced by $|\cdot|_{\mathrm{ucc}}$ in $\mathbf{X}$ is that of uniform convergence over compact sets. Since $(\mathbf{X}, |\cdot|_{\mathrm{ucc}})$ is a closed subspace of $\mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$, it is a a separable Banach space. It is worth mentioning that it is also possible to work directly with $\mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$ as a Fréchet space, provided that it is endowed with an appropriate metric as done in [8]. Apart from some technical issues, the analysis should be completely equivalent.

In the sequel, we make the following standing assumptions on the dynamics. When not explicitly stated otherwise, measurable is meant as Lebesgue-measurable.

*Hypothesis* 1.3. The dynamic $f : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U \to \mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$ is such that

- there is $m_f : [0,1] \to [0,+\infty)$ integrable such that for a.e. $t \in [0,T]$ and any $(x,\mu) \in \mathbb{R}^{\mathrm{d}} \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ we have $\sup_{u\in U} |f[t,\mu,u](x)| \leq m_f(t)(1+|x|+\mathbf{m}_2(\mu))$.
- $(t,u) \mapsto f[t,\mu,u]$ is a Carathéodory function for any $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ fixed.
- $f[\cdot,U]$ is a measurable/Lipschitz map[1] with convex images, which are also closed in the topology of local uniform convergence in $\mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$.
- $\mu \mapsto f[t,\mu,U]$ is $\boldsymbol{\Theta}$-upper semicontinuous for any $t \in [0,T]$ fixed, that is, for any $\varepsilon > 0$ and $\bar{\mu} \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, there is $\mathcal{O} \in \boldsymbol{\Theta}$ such that $\bar{\mu} \in \mathcal{O}$ and

$$\sup_{u\in U} \inf_{\bar{u}\in U} |f[t,\mu,u] - f[t,\bar{\mu},\bar{u}]|_{\mathrm{ucc}} \leq \varepsilon, \qquad \forall \mu \in \mathcal{O}.$$

- there is $L_f : [0,T] \to [0,+\infty)$ integrable such that for a.e. $t \in [0,T]$ and any $x,y \in \mathbb{R}^{\mathrm{d}}$ we have $\sup_{\mu\in\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})} \sup_{u\in U} |f[t,\mu,u](x) - f[t,\mu,u](y)| \leq L_f(t)|x-y|$.

Note that Hypothesis 1.3 implies that $f[\cdot,U]$ satisfies [8, Hypotheses (P)]. Thus, the wellposedness of (1.2) follows from [8, Thm. 3.2] combined with an *ad hoc* selection theorem; see Proposition 3.4. It also follows that the set-valued map $f[\cdot,U]$ has images in $\mathbf{X}$ because for a.e. $t \in [0,T]$ and any $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$

$$\sup_{u\in U} |f[t,\mu,u]|_{\mathrm{ucc}} \leq m_f(t)\left(2 + 2\mathbf{m}_2(\mu) + \sum_{n\in\mathbb{N}} \frac{n}{2^n}\right). \tag{1.8}$$

Hypothesis 1.3 implies that the images of $f[\cdot,U]$ are closed in $(\mathbf{X}, |\cdot|_{\mathrm{ucc}})$, and are sets of equiLipschitz and equibounded maps. The images of $f[\cdot,U]$ are thus compact in $(\mathbf{X}, |\cdot|_{\mathrm{ucc}})$. Indeed, by Arzelà-Ascoli, for any compact set $K$, $\{b|_K \mid b \in f[s,\mu,U]\}$ is compact in $\big(\mathcal{C}(K;\mathbb{R}^{\mathrm{d}}); |\cdot|_{\mathrm{ucc}}\big)$. Then a diagonal argument proves the claim.

The rest of the paper is organized as follows. Section 2 gathers the definition of the Wasserstein space and the metric differential structure over it, as well as the

[1] The mapping $t \mapsto f[t,\cdot]$ is measurable and there is $\kappa_f : [0,T] \to [0,+\infty)$ integrable such that for a.e. $t \in [0,T]$ and any $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ we have $\sup_{u\in U} |f(t,\mu,u) - f(t,\nu,u)|_{\mathrm{ucc}} \leq \kappa_f(t) W_2(\mu,\nu)$.

basic properties of the inductive topology. In Section 3, we study the trajectories of (1.2) and the continuity properties of the value function in the topology $\mathbf{\Theta}$. The HJB equation (1.5) is revisited in Section 4, where we define viscosity solutions and prove a strong comparison principle. Section 5 is devoted to the link between the value function and the HJB equation, and contains our main results.

**2. Preliminary material.** If $A$ and $B$ are two Polish spaces, the symbol $\#$ is used to denote the pushforward operator, which to any probability measure $\mu \in \mathscr{P}(A)$ and measurable map $g : A \to B$, associates another probability measure $g\#\mu \in \mathscr{P}(B)$ given by $(g\#\mu)(\Omega) = \mu\left(g^{-1}(\Omega)\right)$ for any measurable $\Omega \subset B$. We say that a sequence $(\mu_n)_n \subset \mathscr{P}(A)$ converges narrowly to $\mu \in \mathscr{P}(A)$ if $\int_A \varphi d\mu_n \to \int_A \varphi d\mu$ for any continuous bounded function $\varphi : A \to \mathbb{R}$; we denote it by $\mu_n \rightharpoonup_n \mu$.

**2.1. The Wasserstein space.** Let $\mathbb{R}^{2\mathrm{d}} \cong \mathbb{R}^{\mathrm{d}} \times \mathbb{R}^{\mathrm{d}}$ and $\pi_1, \pi_2 : \mathbb{R}^{2\mathrm{d}} \to \mathbb{R}^{\mathrm{d}}$ denote the canonical projections, i.e. $\pi_1(a,b) = a$ and $\pi_2(a,b) = b$. Given $\mu, \nu \in \mathscr{P}(\mathbb{R}^{\mathrm{d}})$, let $\Gamma(\mu,\nu) \subset \mathscr{P}(\mathbb{R}^{2\mathrm{d}})$ denote the set of *transport plans* $\eta$ with first marginal $\pi_1\#\eta$ equal to $\mu$, and second marginal $\pi_2\#\eta$ equal to $\nu$. Given $p \geq 1$, we say that $\mu$ has finite $p$-moment if $\mathbf{m}_p(\mu) := \int_{\mathbb{R}^{\mathrm{d}}} |x|^p d\mu$ is finite, and denote $\mathscr{P}_p(\mathbb{R}^{\mathrm{d}})$ the set of such measures. This set is endowed with the so-called *$p-$Wasserstein distance*

$$W_p^p(\mu,\nu) := \inf_{\eta \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^{2\mathrm{d}}} |\pi_1 - \pi_2|^p d\eta \qquad \forall \mu, \nu \in \mathscr{P}_p(\mathbb{R}^{\mathrm{d}}).$$

The set of optimal transport plans is denoted $\Gamma_o^p(\mu,\nu)$. For brevity, the $2-$Wasserstein distance will be simply called the *Wasserstein distance* and we set $\Gamma_o(\mu,\nu) = \Gamma_o^2(\mu,\nu)$. It is well-known that the curves $[0,1] \ni t \mapsto \eta_t := ((1-t)\pi_1 + t\pi_2)\#\eta$ parametrized by $\eta \in \Gamma_o(\mu,\nu)$ exactly describe the geodesics linking $\mu$ to $\nu$; see [4, Section 7]. The squared Wasserstein distance happens to be semiconcave along geodesics, i.e. for all $\sigma \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and $\eta \in \Gamma_o(\mu,\nu)$, it follows that

$$W_2^2(\eta_t,\sigma) \geqslant (1-t)W_2^2(\mu,\sigma) + tW_2^2(\nu,\sigma) - t(1-t)W_2^2(\mu,\nu), \quad \forall t \in [0,1]. \tag{2.1}$$

**2.2. Tangent and cotangent bundles.** Let $\mathrm{T}\,\mathbb{R}^{\mathrm{d}} \cong \mathbb{R}^{2\mathrm{d}}$ be the tangent bundle of $\mathbb{R}^{\mathrm{d}}$. For each $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, we write $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu := \left\{\xi \in \mathscr{P}_2(\mathbb{R}^{2\mathrm{d}}) \;\middle|\; \pi_1\#\xi = \mu\right\}$. This set can be understood as the largest set of velocities issued from $\mu$, that can be scaled with the operation $\lambda \cdot \xi := (\pi_1, \lambda\pi_2)\#\xi$. Then the curve $t \mapsto ((1-t)\pi_1 + t\pi_2)\#\eta$ coincides with $t \mapsto \exp_\mu(t \cdot \xi) := (\pi_1 + t\pi_2)\#\xi$. This curve generalizes the application $t \mapsto x + tv$, by sending the mass that $\xi$ puts over $(x,v)$ to the point $x + tv$. Following usual notations, we define the partial inverse of the exponential map $\exp_\mu$ by

$$\exp_\mu^{-1}(\nu) := \left\{\xi \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu \;\middle|\; \exp_\mu(\xi) = \nu, \quad \int_{\mathbb{R}^{2\mathrm{d}}} |\pi_2|^2 d\xi = W_2^2(\mu,\nu)\right\}.$$

To measure the distance between $\xi, \zeta \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$, one introduces a set of 3-plans

$$\Gamma_\mu(\xi,\zeta) := \left\{\alpha \in \mathscr{P}(\mathbb{R}^{3\mathrm{d}}) \;\middle|\; (\pi_1,\pi_2)\#\alpha = \xi,\ (\pi_1,\pi_3)\#\alpha = \zeta\right\},$$

and the application $W_\mu : \left(\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu\right)^2 \to \mathbb{R}^+$ given by

$$W_\mu^2(\xi,\zeta) := \inf_{\alpha \in \Gamma_\mu(\xi,\zeta)} \int_{\mathbb{R}^{3\mathrm{d}}} |\pi_2 - \pi_3|^2 d\alpha.$$

which is a distance in $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$ (see [27, Chap. 4]). Let $(\pi_1,0)\#\mu \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$ be the probability over the tangent bundle concentrated on the null velocity. We denote

$$|\cdot|_\mu : \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu \to \mathbb{R}^+, \qquad |\xi|_\mu \coloneqq W_\mu\left(\xi, (\pi_1,0)\#\mu\right).$$

In particular, if $\xi = g_1\#\mu$ and $\zeta = g_2\#\mu$ for some fields $g_1, g_2 \in L^2_\mu(\mathbb{R}^{\mathrm{d}}; \mathrm{T}\,\mathbb{R}^{\mathrm{d}})$, then $W_\mu(g_1\#\mu, g_2\#\mu) = \|g_1 - g_2\|_{L^2_\mu}$.

Definition 2.1. *For a function $p : \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu \to \mathbb{R}$, set*

$$\|p\|_\mu \coloneqq \sup\left\{|p(\xi)| \mid \xi \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu,\ |\xi|_\mu = 1\right\}.$$

*We define the metric cotangent bundle $\mathbb{T}$ as $\mathbb{T} \coloneqq \bigcup_\mu \{\mu\} \times \mathbb{T}_\mu$, where*

$$\mathbb{T}_\mu \coloneqq \left\{ p : \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu \to \mathbb{R} \;\middle|\; \begin{array}{c} \|p\|_\mu < \infty, \quad p(\lambda\xi) = \lambda p(\xi) \quad \forall \lambda \geqslant 0, \\ \text{and } p \text{ is Lipschitz w.r.t. } W_\mu. \end{array} \right\}.$$

We then recover $|p(\xi)| \leqslant \|p\|_\mu |\xi|_\mu$, for any $\xi \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$ and $p \in \mathbb{T}_\mu$. Elements of $\mathbb{T}_\mu$ may be built from directional derivatives of sufficiently smooth maps. Here, the directional derivative of $\varphi : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R}$ at $\mu$ along $\xi \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$ is given by

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

As an important example, let $\mu, \sigma \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. Due to the semiconcavity (2.1), the application $[0,1] \ni h \mapsto \frac{1}{h}\left(W_2^2(\exp_\mu(h \cdot \xi), \sigma) - W_2^2(\mu,\sigma)\right)$ is bounded and monotone, so that $W_2^2(\cdot, \sigma)$ is directionally differentiable along all $\xi$ such that $s \to \exp_\mu(s \cdot \xi)$ follows a geodesic. It turns out that the squared distance is directionally differentiable along any $\xi$ in $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$. According to [4, Prop. 7.3.6] (c.f. [27, Prop. 4.10]) we have $\|D_\mu W_2^2(\cdot,\sigma)\|_\mu = 2W_2(\mu,\sigma)$ and there holds

$$D_\mu W_2^2(\mu,\sigma)(\xi) = \inf_{\eta \in \exp_\mu^{-1}(\sigma)} \inf_{\alpha \in \Gamma_\mu(\xi,\eta)} \int_{\mathbb{R}^{3\mathrm{d}}} \langle \pi_2, \pi_3 \rangle \, d\alpha.$$

*Remark* 2.2. Here, we consider the space $\mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$ as a surrogate tangent cone on the Wasserstein space, instead of the tangent cone of [27, Chap. 4] relying on the curvature of $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. While more geometric, the general tangent cone does not bring nor withdraw anything to the argument, and we spare the reader its technicalities.

**2.3. The topology $\boldsymbol{\Theta}$ over $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$.** As pointed out in the introduction, one of the main novelties of this work is the use of an inductive topology on $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. We now review some of its main properties which are taken from [27, Chapter 2].

Lemma 2.3. *Let $\boldsymbol{\Theta}$ be given by Definition* 1.1.

1. *A set $A \subset \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ is closed in $\boldsymbol{\Theta}$ if and only if each $A \cap \mathbf{K}_n$ is closed in $\boldsymbol{\Theta}_n$.*
2. *A sequence $(\mu_n)_n \subset \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ converges in $\boldsymbol{\Theta}$ towards some $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ if and only if $\mu_n \rightharpoonup_n \mu$ and $\sup_{n\in\mathbb{N}} \mathbf{m}_2(\mu_n) < \infty$. We then denote $\mu_n \xrightarrow{\boldsymbol{\Theta}}_n \mu$.*
3. *Wasserstein closed balls are compact and sequentially compact in $\boldsymbol{\Theta}$.*
4. *The squared Wasserstein distance is sequentially $\boldsymbol{\Theta}$-lower semicontinuous.*

Recall that in an arbitrary topological space, any closed set is also sequentially closed. The converse is true for metrizable spaces, but not necessarily for any topological space. The spaces that satisfy this property are called *sequential*. Since the

narrow topology is metrizable (see [4, Remark 5.1.1]), each topological space $(\mathbf{K}_n, \mathbf{\Theta}_n)$ is sequential. It is important to mention that this property is inherited by the inductive topology $\mathbf{\Theta}$ as shown in Lemma 2.4 below; it is actually a general fact, sequential spaces are stable by inductive limits [24, Corollary 1.7].

LEMMA 2.4. $\left(\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}), \mathbf{\Theta}\right)$ *is a sequential space.*

*Proof.* Let $A \subset \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ be sequentially $\mathbf{\Theta}$-closed. From Item 1 in Lemma 2.3, it is enough to prove that for each $n \in \mathbb{N}$, the set $A_n \coloneqq A \cap \mathbf{K}_n$ is narrowly closed. Let $(\mu_m)_m \subset A_n$ be a sequence converging narrowly to $\mu \in \mathbf{K}_n$. By the continuity of $\iota_n$, the sequence $(\iota_n(\mu_m))_m \subset A$ converges to $\iota_n(\mu)$, and as $A$ is sequentially $\mathbf{\Theta}$-closed, $\iota_n(\mu) \in A$. The conclusion follows by noticing that $\iota_n(\mu) = \mu$ because $\mu \in \mathbf{K}_n$. ∎

Consequently, in the sequel, we make no distinction between lower (resp. upper) semicontinuity and sequential lower (resp. upper) semicontinuity for the topology $\mathbf{\Theta}$. In particular, the squared Wasserstein distance turns out to be $\mathbf{\Theta}$-lower semicontinuous thanks to Item 4 in Lemma 2.3.

**3. Trajectories in the Wasserstein space.** We now discuss the wellposedness of the controlled continuity equation (1.2) and properties of the set of solutions. Apart from Lemma 3.9, all the results in this section are written for non-autonomous systems, where the dependency of $f$ on the time variable is merely measurable. Considering an autonomous setting does not simplify the arguments, while it will demand more details for proving Corollary 3.7. Here, and in the sequel, solutions of (1.2) are understood in the distributional sense.

DEFINITION 3.1. *For each $(t, \nu) \in [0, T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and each measurable control $u : [t, T] \to U$, we say that a function $[t, T] \ni s \mapsto \boldsymbol{\mu}_s \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ is a solution of* (1.2) *if it is absolutely continuous (see e.g. [4, Definition 1.1.1]), $\boldsymbol{\mu}_t = \nu$ and*

$$\int_t^T \int_{\mathbb{R}^{\mathrm{d}}} \left[\partial_s \varphi + \langle \nabla_x \varphi, f[s, \boldsymbol{\mu}_s, u(s)]\rangle\right] d\boldsymbol{\mu}_s ds = 0, \quad \forall \varphi \in \mathcal{C}_c^\infty\left((t, T) \times \mathbb{R}^{\mathrm{d}}\right), \tag{3.1}$$

*Remark* 3.2. For any $\varphi \in \mathcal{C}_c^\infty\left((t, T) \times \mathbb{R}^{\mathrm{d}}\right)$, the integral term in (3.1) is well-defined. Indeed, let $u : [t, T] \to U$ be a measurable control, take $r > 0$ such that $\varphi(\cdot, x) \equiv 0$ for all $|x| \geq r$ and let $[t, T] \ni s \mapsto \boldsymbol{\mu}_s \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ be absolutely continuous; in particular $s \mapsto W_2(\boldsymbol{\mu}_s, \boldsymbol{\mu}_t)$ is bounded. Notice that $\sqrt{\mathbf{m}_2(\boldsymbol{\mu}_s)} \leq W_2(\boldsymbol{\mu}_s, \boldsymbol{\mu}_t) + \sqrt{\mathbf{m}_2(\boldsymbol{\mu}_t)}$, then $M \coloneqq \sup_{s \in [t,T]} \mathbf{m}_2(\boldsymbol{\mu}_s) < +\infty$. From Hypothesis 1.3 we deduce that the vector field $(s, x) \mapsto f[s, \boldsymbol{\mu}_s, u(s)](x)$ is $\mathcal{L} \otimes \mathcal{B}^{\mathrm{d}}$ measurable. Thanks to the Monotone Class Lemma [1, Thm. 4.13], the mapping

$$s \mapsto \int_{\mathbb{R}^{\mathrm{d}}} \langle \nabla_x \varphi(s, x), f[s, \boldsymbol{\mu}_s, u(s)](x)\rangle d\boldsymbol{\mu}_s \text{ is measurable.}$$

The integrability comes from Hypothesis 1.3 and the fact that

$$\int_{\mathbb{R}^{\mathrm{d}}} \langle \nabla_x \varphi(s, x), f[s, \boldsymbol{\mu}_s, u(s)](x)\rangle d\boldsymbol{\mu}_s \leq m_f(s)(1 + r + M), \qquad \text{for a.e. } s \in [t, T].$$

As in the case of Euclidean spaces, properties of the solutions of the control system (1.2) are obtained by appealing to a more general dynamical system, a *differential inclusion*. In our case, our vehicle will be the theory developed in [8].

*Remark* 3.3. For each $(t, \nu) \in [0, T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, consider the continuity inclusion

$$\partial_s \boldsymbol{\mu}_s \in -\operatorname{div}\left(f[s, \boldsymbol{\mu}_s, U]\ \boldsymbol{\mu}_s\right), \qquad s \in [t, T], \quad \boldsymbol{\mu}_t = \nu. \tag{3.2}$$

According to [8, Definition 2.25], a function $[t,T] \ni s \mapsto \boldsymbol{\mu}_s \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ is a solution of (3.2) if it is absolutely continuous, $\boldsymbol{\mu}_t = \nu$ and if there is a measurable selection $\mathbf{b}(s) \in f[s, \boldsymbol{\mu}_s, U]$ such that that $(\boldsymbol{\mu}_s)_{s\in[t,T]}$ is a solution of (1.1) with $\mathsf{f}_s = \mathbf{b}(s)$. It is not difficult to see that any solution of (1.2) in the sense of Definition 3.1 is a solution of (3.2), because Hypothesis 1.3 implies that the vector field $s \mapsto f[s, \boldsymbol{\mu}_s, u(s)]$ is measurable. On the other hand, from the Filippov Implicit Function Theorem [1, Thm. 18.17], for any measurable selection $\mathbf{b}(s) \in f[s, \boldsymbol{\mu}_s, U]$, there is a measurable function $u : [t,T] \to U$ such that $\mathbf{b}(s) = f[s, \boldsymbol{\mu}_s, u(s)]$ for a.e. $s \in [t,T]$. Thus, any solution of (3.2) provides a solution for (1.2). Also, thanks to Hypothesis 1.3, from [8, Prop. 2.21 & Remark 2.2], we have that for each $(t,\nu) \in [0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and each measurable control $u : [t,T] \to U$, $[t,T] \ni s \mapsto \boldsymbol{\mu}_s \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ is a solution of (1.2), and

$$\mathbf{m}_2(\boldsymbol{\mu}_s) \leq c_2 \left( \mathbf{m}_2(\nu) + \int_t^s m_f(r)(1 + \mathbf{m}_2(\boldsymbol{\mu}_r))dr \right) \qquad \forall s \in [t,T],$$

where $c_2 \coloneqq \sqrt{2} \exp\left( \|m_f\|_{L^1}^2 \right)$. So, from Grönwall's inequality, we get

$$\mathbf{m}_2(\boldsymbol{\mu}_s) \leq (1 + c_2 \mathbf{m}_2(\nu)) \exp\left( c_2 \int_t^s m_f(r)dr \right), \qquad \forall s \in [t,T]. \tag{3.3}$$

**3.1. Existence and properties of the trajectories.** We now move to the wellposedness of (1.2) and the representation of its solutions. Let us point out that the representation formula for the continuity equation can be deduced directly from [4, Prop. 8.1.8] in the case $(s,x) \mapsto f[s, \boldsymbol{\mu}_s, u(s)](x)$ is a Borel vector field. This is not necessarily the case in our setting, however the arguments for proving [4, Prop. 8.1.8] can be adapted; the key point here is that the flows of (3.4) are globally defined. For the sake of completeness, we provide the details of the proof.

Proposition 3.4 (Wellposedness and representation). *Assume Hypothesis* 1.3. *For each $(t,\nu) \in [0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and each measurable control $u : [t,T] \to U$, there is a unique solution of* (1.2)*, which we denote by $(\boldsymbol{\mu}_s^{t,\nu,u})_{s\in[t,T]}$. Moreover, for any $s \in [t,T]$, we have that $\boldsymbol{\mu}_s^{t,\nu,u} = Y_s^{t,u} \# \nu$ where, for any given $x \in \mathbb{R}^{\mathrm{d}}$, the mapping $[t,T] \ni s \mapsto Y_s^{t,u}(x) \in \mathbb{R}^{\mathrm{d}}$ is the unique solution of*

$$\frac{d}{ds} y(s) = f[s, \boldsymbol{\mu}_s, u(s)](y(s)), \quad y(t) = x. \tag{3.4}$$

*Proof.* Take a measurable control $u : [t,T] \to U$ and consider the set-valued map $G(s,\mu) = \{f[s,\mu,u(s)]\}$ defined for $(s,\mu) \in [t,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ with values in $\mathcal{C}(\mathbb{R}^{\mathrm{d}}; \mathbb{R}^{\mathrm{d}})$. Observe that Hypothesis 1.3 implies that $G$ satisfies [8, Hypotheses (P)]. Thus from [8, Thm.s 3.2] we get the existence of an absolutely continuous function $[t,T] \ni s \mapsto \boldsymbol{\mu}_s \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $\boldsymbol{\mu}_t = \nu$ and that satisfies (3.1); the latter is due to the fact that $s \mapsto f[s, \boldsymbol{\mu}_s, u(s)]$ is the only possible measurable selection of $s \mapsto G(s, \boldsymbol{\mu}_s)$. Notice that, thanks to [38, Corollary 2.4.5], for any $x \in \mathbb{R}^{\mathrm{d}}$ the Cauchy problem (3.4) has a unique solution, denoted $(Y_s^{t,u}(x))_{s\in[t,T]}$, because the vector field $(s,y) \mapsto g_{\boldsymbol{\mu}}(s,y) \coloneqq f[s, \boldsymbol{\mu}_s, u(s)](y)$ is measurable/Lipschitz in the sense that

$$\|g_{\boldsymbol{\mu}}(s,y_1) - g_{\boldsymbol{\mu}}(s,y_2)\| \leq L_f(s)\|y_1 - y_2\|, \text{ for a.e. } s \in [t,T], \ \forall y_1, y_2 \in \mathbb{R}^{\mathrm{d}}.$$

Therefore, for any $\varphi \in \mathcal{C}_c^\infty\left((t,T) \times \mathbb{R}^{\mathrm{d}}\right)$ and $x \in \mathbb{R}^{\mathrm{d}}$, the maps $s \mapsto \varphi(s, Y_s^{t,u}(x))$ are absolutely continuous in $(t,T)$. Consequently, proceeding as in [4, Lemma 8.1.6], we can show that the mapping $[t,T] \ni s \mapsto Y_s^{t,u} \# \nu$ is a solution of the continuity equation (1.1) with $\mathsf{f}_s = g_{\boldsymbol{\mu}}(s,\cdot)$. The fact that $\boldsymbol{\mu}_s = Y_s^{t,u} \# \nu$ for any $s \in [t,T]$

is then a consequence of [4, Prop. 8.1.7]. Finally, we turn into the uniqueness of solutions. Suppose that $\boldsymbol{\mu}_s^1 = Y_s\#\nu$ and $\boldsymbol{\mu}_s^2 = Z_s\#\nu$ are solutions of (3.1), where $Y$ and $Z$ are the flows of (3.4) associated with $g_1(s,y) = f[s,\boldsymbol{\mu}_s^1,u(s)](y)$ and $g_2(s,y) = f[s,\boldsymbol{\mu}_s^2,u(s)](y)$. From Hypothesis 1.3, for a.e. $s\in[t,T]$ and any $x\in\mathbb{R}^{\mathrm{d}}$ we have

$$\|g_1(s,Y_s(x)) - g_2(s,Z_s(x))\| \leq L_f(s)\|Y_s(x)-Z_s(x)\| + \kappa_f(s)W_2(\boldsymbol{\mu}_s^1,\boldsymbol{\mu}_s^2).$$

Therefore, from Grönwall's inequality [38, Lemma 2.4.4] and [4, (7.1.6)] we obtain

$$W_2(\boldsymbol{\mu}_s^1,\boldsymbol{\mu}_s^2) \leq \|Y_s - Z_s\|_{L^2(\nu)} \leq \exp\left(\int_t^T L_f(r)dr\right)\int_t^s \kappa_f(r)W_2(\boldsymbol{\mu}_r^1,\boldsymbol{\mu}_r^2)dr.$$

Using Grönwall's inequality again, we get $W_2(\boldsymbol{\mu}_s^1,\boldsymbol{\mu}_s^2)\leq 0$, and the conclusion follows. □

*Remark* 3.5. Since the dynamic $\tilde{f}:[0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})\times U\to\mathcal{C}(\mathbb{R}^{\mathrm{d}};\mathbb{R}^{\mathrm{d}})$ given by $\tilde{f}(s,\mu,u) \coloneqq -f(T-s,\mu,u)$ satisfies the hypotheses of Proposition 3.4, it is not difficult to see that if $(\boldsymbol{\mu}_s^{t,\nu,u})_{s\in[t,T]}$ is the solution given by Proposition 3.4 associated with $(t,\nu)\in[0,T)\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and the measurable control $u:[t,T]\to U$, then for any $\tilde{t}\in[0,t)$ and any measurable control $\tilde{u}:[\tilde{t},T]\to U$ such that $\tilde{u}|_{[t,T]} = u$ a.e., there is $\tilde{\nu}\in\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ for which $\boldsymbol{\mu}_t^{\tilde{t},\tilde{\nu},\tilde{u}} = \nu$. Here $(\boldsymbol{\mu}_s^{\tilde{t},\tilde{\nu},\tilde{u}})_{s\in[\tilde{t},T]}$ is the solution given by Proposition 3.4 associated with $(\tilde{t},\tilde{\nu})$ and the measurable control $\tilde{u}$.

**3.2. The value function and its properties.** We turn our attention to the properties of the value function. Before going further, we need to prove a convergence theorem for trajectories of (1.2). The following result is akin to [8, Thm. 4.5]. The main difference is that here convergence of the trajectories is not considered w.r.t. $W_2-$distance, but rather w.r.t. a weaker one ($W_p$ with $p\in[1,2)$), while the dependence on the initial condition is taken w.r.t. the inductive topology $\boldsymbol{\Theta}$.

Lemma 3.6 (Convergence of trajectories). *Assume Hypothesis* 1.3 *and take* $(t,\nu)\in[0,T)\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. *Let* $(\nu^n)_{n\in\mathbb{N}}\subset\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ *be such that* $\nu^n \xrightarrow{\boldsymbol{\Theta}}_n \nu$. *Let* $(\boldsymbol{\mu}_s^n)_{s\in[t,T]}$ *be the solution of* (1.2) *provided by Proposition* 3.4, *starting from* $\nu^n$ *and associated with the measurable control* $u^n:[t,T]\to U$. *Then, for any* $p\in[1,2)$ *there is a measurable control* $u:[t,T]\to U$ *such that*

$$\liminf_{n\to+\infty}\sup_{s\in[t,T]} W_p\left(\boldsymbol{\mu}_s^n,\boldsymbol{\mu}_s^{t,\nu,u}\right) = 0.$$

*Here* $(\boldsymbol{\mu}_s^{t,\nu,u})_{s\in[t,T]}$ *is the unique solution of* (1.2) *provided by Proposition* 3.4.

*Proof.* The proof is rather technical, and we divide it into several steps.

**Step 1:** We construct a candidate to solution. Let us fix $p\in[1,2)$. Since $\nu^n \xrightarrow{\boldsymbol{\Theta}}_n \nu$, from Lemma 2.3 there is $R\geqslant 0$ such that $\mathbf{m}_2(\nu_n)\leqslant R$ for all $n\in\mathbb{N}$. From (3.3) we deduce that there is $R_T\geqslant 0$ such that $\mathbf{m}_2(\boldsymbol{\mu}_s^n)\leqslant R_T$ for all $(s,n)\in[t,T]\times\mathbb{N}$. Therefore, thanks to [4, §5.1.1. and Lemma 5.1.7], for any $s\in[t,T]$, each section $(\boldsymbol{\mu}_s^n)_{n\in\mathbb{N}}$ has uniformly integrable $p-$moments and it is tight; in particular $(\boldsymbol{\mu}_s^n)_{s\in[t,T]}\subset\mathscr{P}_p(\mathbb{R}^{\mathrm{d}})$. From Prokhorov's Theorem, combined with [4, Prop.s 7.1.4 and 7.1.5], we get that the sections $(\boldsymbol{\mu}_s^n)_{n\in\mathbb{N}}$ are relatively compact in $(\mathscr{P}_p(\mathbb{R}^{\mathrm{d}}),W_p)$. We will therefore obtain the candidate as a limit point by using Arzelà-Ascoli Theorem in $(\mathscr{P}_p(\mathbb{R}^{\mathrm{d}}),W_p)$. Notice that from [4, (7.1.6)] we have

$$W_2(\boldsymbol{\mu}_s^n,\boldsymbol{\mu}_r^n) \leq \|Y_s^{t,u^n} - Y_r^{t,u^n}\|_{L^2(\nu^n)}, \qquad \forall r,s\in[t,T],$$

where $Y^{t,u^n}$ is given by Proposition 3.4. Assume for simplicity that $r < s$. By Grönwall's inequality [38, Lemma 2.4.4] we get

$$|Y_s^{t,u^n}(x) - Y_r^{t,u^n}(x)| \leq (1 + R_T + |x|) \exp\left(\int_r^s m_f(z)dz\right) \int_r^s m_f(z)dz, \ \forall x \in \mathbb{R}^{\mathrm{d}}.$$

It follows that $(\boldsymbol{\mu}^n)_{n\in\mathbb{N}}$ is equi-continuous w.r.t. the $W_2-$distance. Hölder's inequality implies that $W_p \leq W_2$. Consequently, $(\boldsymbol{\mu}_s^n)_{s\in[t,T]} \subset \mathscr{P}_p(\mathbb{R}^{\mathrm{d}})$ is equi-continuous w.r.t. the $W_p-$distance. By Arzelà-Ascoli Theorem, there is an absolutely continuous map $[t,T] \ni s \mapsto \boldsymbol{\mu}_s \in \mathscr{P}_p(\mathbb{R}^{\mathrm{d}})$ and a (non relabeled) subsequence $(\boldsymbol{\mu}^n)_{n\in\mathbb{N}}$ so that

$$\lim_{n\to+\infty} \sup_{s\in[t,T]} W_p(\boldsymbol{\mu}_s^n, \boldsymbol{\mu}_s) = 0 \quad \text{and} \quad \sup_{s\in[t,T]} \mathbf{m}_2(\boldsymbol{\mu}_s) \leq R_T. \tag{3.5}$$

It remains to prove that $\boldsymbol{\mu} = \boldsymbol{\mu}^{t,\nu,u}$ for some measurable control $u : [t,T] \to U$. **Step 2:** We study now the convergence of the vector fields $(s,y) \mapsto \boldsymbol{b}^n(s,y) := f[s, \boldsymbol{\mu}_s^n, u(s)](y)$. It follows from (1.8) that $\boldsymbol{b}^n \in L^1([t,T];\mathbf{X})$; here $L^q([t,T];\mathbf{X})$ stands for the usual Lebesgue-Bochner space for $q \in [1,+\infty]$ (see e.g., [22]). Moreover, since $f[s,\boldsymbol{\mu}_s,U]$ is compact in $(\mathbf{X}, |\cdot|_{\mathrm{ucc}})$, by [21, Corollary 2.6] (see also [8, Thm. 2.1]), there is a (non relabeled) subsequence that converges weakly to some $\boldsymbol{b} \in L^1([t,T];\mathbf{X})$.

Let us show that $\boldsymbol{b}_s \in A_s := f[s,\boldsymbol{\mu}_s,U]$ for a.e. $s \in [t,T]$. First, let us prove that

$$\int_t^T \boldsymbol{p}_s(\boldsymbol{b}_s)ds \leqslant \int_t^T \sigma(s,\boldsymbol{p}_s)ds, \qquad \forall \boldsymbol{p} \in L^\infty([t,T];\mathbf{X}^*), \tag{3.6}$$

where $\sigma(s,p) := \sup\{p(\beta) \mid \beta \in A_s\}$ for any $(s,p) \in [t,T] \times \mathbf{X}^*$. Notice that $\sigma$ is a Carathéodory function. Indeed, for $s \in [t,T]$ fixed, $\sigma(s,\cdot)$ is a finite convex function, which is also bounded above on any neighborhood of $p = 0$ thanks to (1.8), thus it is continuous. Also, since $s \mapsto A_s$ is measurable and has closed images, for any $p \in \mathbf{X}^*$ fixed $\sigma(\cdot,p)$ is measurable (see e.g., [5, Thm. 8.2.14]). Observe that the same arguments allow to deduce that for any $n \in \mathbb{N}$, the mapping $\sigma^n(s,p) := \sup\{p(\beta) \mid \beta \in f[s,\boldsymbol{\mu}_s^n,U]\}$ is Carathéodory as well. In particular, the mappings $s \mapsto \sigma(s,\boldsymbol{p}_s)$ and $s \mapsto \sigma^n(s,\boldsymbol{p}_s)$ are measurable for any $\boldsymbol{p} \in L^\infty([t,T];\mathbf{X}^*)$ fixed (see [5, Lemma 8.2.3]), and the integrals in (3.6) and (3.7) below have meaning. As a matter of fact, both integrals are finite thanks to (1.8).

By weak convergence, for any $\boldsymbol{p} \in L^\infty([t,T];\mathbf{X}^*) \subset [L^1([t,T];\mathbf{X})]^*$, we have that

$$\int_t^T \boldsymbol{p}_s(\boldsymbol{b}_s)ds = \limsup_{n\to+\infty} \int_t^T \boldsymbol{p}_s(\boldsymbol{b}_s^n)ds \leqslant \limsup_{n\to+\infty} \int_t^T \sigma^n(s,\boldsymbol{p}_s)ds. \tag{3.7}$$

On the other hand, from (3.5), we also have that $\boldsymbol{\mu}_s^n \xrightarrow{\boldsymbol{\Theta}}_n \boldsymbol{\mu}_s$ for each $s \in [t,T]$ fixed, so that by upper semi-continuity of $\mu \mapsto f[s,\mu,U]$, we get

$$\limsup_{n\to\infty} \sigma^n(s,\boldsymbol{p}_s) \leqslant \sigma(s,\boldsymbol{p}_s).$$

Using (1.8) and Fatou's Lemma we get (3.6). To complete this part of the proof, we use a result inspired by the theory of convex integral functionals developed by Rockafellar, in particular by [36, Thm. 2]. Let $\boldsymbol{\chi}_{A_s}$ be the characteristic function of $\boldsymbol{\chi}_{A_s}$ , i.e. $\boldsymbol{\chi}_{A_s}(b) = 0$ if $b \in \boldsymbol{\chi}_{A_s}$, and $\boldsymbol{\chi}_{A_s}(\nu) = +\infty$ otherwise. We claim that

$$\sup_{\boldsymbol{p}\in L^\infty([t,T];\mathbf{X}^*)} \int_t^T [\boldsymbol{p}_s(\boldsymbol{b}_s) - \sigma(s,\boldsymbol{p}_s)]\, ds = \int_t^T \boldsymbol{\chi}_{A_s}(\boldsymbol{b}_s)ds. \tag{3.8}$$

Since $A_s$ is a nonempty convex closed subset of $\mathbf{X}$, by the standard theory of convex analysis (see [39] for example), we have that $\sup_{p\in\mathbf{X}^*} p(\beta)-\sigma(s,p)=\boldsymbol{\chi}_{A_s}(\beta)$ for any $\beta\in\mathbf{X}$. Thus, the inequality "$\leqslant$" is straightforward. To prove the other inequality, for any $k\in\mathbb{N}$, we define $\boldsymbol{\chi}^k(s)=\sup_{p\in\mathbf{X}^*}\{p(\boldsymbol{b}_s)-\sigma(s,p)\mid \|p\|_{\mathbf{X}^*}\leq k\}$ for a.e. $s\in[t,T]$. Clearly $\boldsymbol{\chi}^k(s)\nearrow\boldsymbol{\chi}_{A_s}(b_s)$ for a.e. $s\in[t,T]$. Notice that, since $\mathbf{X}$ is separable, the set $\{p\in\mathbf{X}^*\mid\|p\|_{\mathbf{X}^*}\leq k\}$ is compact and metrizable in the weak* topology on $\mathbf{X}^*$ ([9, Thm. 3.28]), therefore it is separable. From the Marginal Map Theorem [5, Thm. 8.2.11], each $\boldsymbol{\chi}^k$ is a measurable function and the marginal mapping $R^k(s)=\{p\in\mathbf{X}^*\mid\|p\|_{\mathbf{X}^*}\leq k,\ \boldsymbol{\chi}^k(s)=p(\boldsymbol{b}_s)-\sigma(s,p)\}$ is measurable, and since for $s\in[t,T]$ fixed, $\sigma(s,\cdot)$ is weak* lower semicontinuous in $\mathbf{X}^*$, $R^k(s)$ is nonempty and closed in the weak* topology on $\mathbf{X}^*$. In particular, from the Measurable Selection Theorem [5, Thm. 8.1.3], for any $k\in\mathbb{N}$, there is $\boldsymbol{p}^k\in L^\infty([t,T];\mathbf{X}^*)$ such that $\|\boldsymbol{p}^k_s\|_{\mathbf{X}^*}\leq k$ and $\boldsymbol{\chi}^k(s)=\boldsymbol{p}^k_s(\boldsymbol{b}_s)-\sigma(s,\boldsymbol{p}^k_s)$ for a.e. $s\in[t,T]$. It follows then that

$$\sup_{\boldsymbol{p}\in L^\infty([t,T];\mathbf{X}^*)}\int_t^T[\boldsymbol{p}_s(\boldsymbol{b}_s)-\sigma(s,\boldsymbol{p}_s)]\,ds\geq\int_t^T\left[\boldsymbol{p}^k_s(\boldsymbol{b}_s)-\sigma(s,\boldsymbol{p}^k_s)\right]ds=\int_t^T\boldsymbol{\chi}^k(s)ds.$$

Since, $\boldsymbol{\chi}^k(s)\nearrow\boldsymbol{\chi}_{A_s}(b_s)$ for a.e. $s\in[t,T]$, we have that $s\mapsto\boldsymbol{\chi}_{A_s}(b_s)$ is measurable function, and therefore, we get (3.8) from the Monotone Convergence Theorem.

Combining then (3.6) and (3.8), we deduce that $\boldsymbol{b}_s\in f[s,\boldsymbol{\mu}_s,U]$ for a.e. $s\in[t,T]$.

**Step 3:** We prove finally that $\boldsymbol{\mu}$ is a solution to the continuity equation. Take $\varphi\in\mathcal{C}^\infty_c((t,T)\times\mathbb{R}^{\mathrm{d}};\mathbb{R})$ arbitrary, but fixed, and let $\boldsymbol{n}_\varphi\in\mathbb{N}$ be large enough so that $\varphi(\cdot,x)=0$ whenever $|x|>\boldsymbol{n}_\varphi$. Since $(\boldsymbol{\mu}^n_s)_{s\in[t,T]}$ solves (1.2), for any $n\in\mathbb{N}$ we have

$$\int_t^T\int_{\mathbb{R}^{\mathrm{d}}}[\partial_s\varphi+\langle\nabla_x\varphi,\boldsymbol{b}_s\rangle]\,d\boldsymbol{\mu}_s ds=\boldsymbol{I}^n_1+\boldsymbol{I}^n_2+\boldsymbol{I}^n_3,$$

where

$$\boldsymbol{I}^n_1:=\int_t^T\left[\int_{\mathbb{R}^{\mathrm{d}}}\partial_s\varphi d\boldsymbol{\mu}_s-\int_{\mathbb{R}^{\mathrm{d}}}\partial_s\varphi d\boldsymbol{\mu}^n_s\right]ds,\quad \boldsymbol{I}^n_2:=\int_t^T\int_{\mathbb{R}^{\mathrm{d}}}\langle\nabla_x\varphi,\boldsymbol{b}_s-\boldsymbol{b}^n_s\rangle\,d\boldsymbol{\mu}_s ds$$

and

$$\boldsymbol{I}^n_3:=\int_t^T\left[\int_{\mathbb{R}^{\mathrm{d}}}\langle\nabla_x\varphi,\boldsymbol{b}^n_s\rangle\,d\boldsymbol{\mu}_s-\int_{\mathbb{R}^{\mathrm{d}}}\langle\nabla_x\varphi,\boldsymbol{b}^n_s\rangle\,d\boldsymbol{\mu}^n_s\right]ds.$$

Since $\boldsymbol{\mu}^n_s\xrightarrow{\boldsymbol{\Theta}}_n\boldsymbol{\mu}_s$ for each $s\in[t,T]$ fixed, $\int_{\mathbb{R}^{\mathrm{d}}}\partial_s\varphi d\boldsymbol{\mu}^n_s\to\int_{\mathbb{R}^{\mathrm{d}}}\partial_s\varphi d\boldsymbol{\mu}_s$ if $n\to+\infty$. Thus from the Dominated Convergence Theorem, we get that $\boldsymbol{I}^n_1\to0$ when $n\to+\infty$. Also, given that $\boldsymbol{b}^n$ converges weakly to $\boldsymbol{b}$ in $L^1([t,T];\mathbf{X})$ we get that $\boldsymbol{I}^n_2\to0$ if $n\to+\infty$. To see this it suffices to note that the real-valued linear operator

$$\beta\mapsto\int_t^T\int_{\mathbb{R}^{\mathrm{d}}}\langle\nabla_x\varphi,\beta_s\rangle\,d\boldsymbol{\mu}_s ds$$

defined on $L^1([t,T];\mathbf{X})$ is bounded. Indeed, for any $\beta\in L^1([t,T];\mathbf{X})$ we have

$$\int_{\mathbb{R}^{\mathrm{d}}}\langle\nabla_x\varphi,\beta_s\rangle\,d\boldsymbol{\mu}_s\leqslant\int_{\{\|x\|\leq\boldsymbol{n}_\varphi\}}|\nabla_x\varphi|\,\|\beta_s\|d\boldsymbol{\mu}_s\leqslant 2^{\boldsymbol{n}_\varphi}\|\nabla_x\varphi\|_\infty|\beta_s|_{\mathrm{ucc}}.$$

Integrating over $s\in[t,T]$, we get that the operator is bounded, thus an element of $[L^1([t,T];\mathbf{X})]^*$. Therefore, $\boldsymbol{I}^n_2\to0$ when $n\to+\infty$.

The final step is to prove that $\boldsymbol{I}_3^n \to 0$ if $n \to +\infty$. This will be a consequence of the fact that $\boldsymbol{\mu}^n$ converges uniformly to $\boldsymbol{\mu}$ for the $W_p-$distance. For any $n \in \mathbb{N}$, let $[t,T] \ni s \mapsto \boldsymbol{\eta}_s^n \in \mathscr{P}_2(\mathbb{R}^{2\mathrm{d}})$ be such that $\boldsymbol{\eta}_s^n \in \Gamma_o^p(\boldsymbol{\mu}_s, \boldsymbol{\mu}_s^n)$ for a.e. $s \in [t,T]$. Then

$$\boldsymbol{I}_3^n = \int_t^T \int_{\mathbb{R}^{2\mathrm{d}}} \left[\langle \nabla_x \varphi(s,x), \boldsymbol{b}_s^n(x)\rangle - \langle \nabla_x \varphi(s,y), \boldsymbol{b}_s^n(y)\rangle\right] d\boldsymbol{\eta}_s^n(x,y) ds.$$

From Hypothesis 1.3 and (3.3), for any $n \in \mathbb{N}$, we have $\|\boldsymbol{b}_s^n(x)\| \leq m_f(s)(1+\boldsymbol{n}_\varphi + R_T)$ for a.e. $s \in [t,T]$ and whenever $\|x\| \leq \boldsymbol{n}_\varphi$. Let $A_\varphi = \{(x,y) \in \mathbb{R}^{2\mathrm{d}} \mid \|x\|, \|y\| \leq \boldsymbol{n}_\varphi\}$ and let $L_\varphi > 0$ be a Lipschitz constant for $\nabla_x \varphi$. From Hypothesis 1.3, we have

$$\boldsymbol{I}_3^n \leq \int_t^T [m_f(s) L_\varphi (1 + \boldsymbol{n}_\varphi + R_T) + L_f(s) \|\nabla_x \varphi\|_\infty] \int_{A_\varphi} \|x-y\| d\boldsymbol{\eta}_s^n(x,y) ds.$$

From Hölder's inequality, we get

$$\boldsymbol{I}_3^n \leq \sup_{s\in[t,T]} W_p\left(\boldsymbol{\mu}_s, \boldsymbol{\mu}_s^n\right) \int_t^T [m_f(s) L_\varphi (1 + \boldsymbol{n}_\varphi + R_T) + L_f(s) \|\nabla_x \varphi\|_\infty] ds.$$

This completes the proof. ∎

Observe that for any given measurable control $u : [0,T] \to U$, the dynamics $f_u : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times U \to \mathcal{C}(\mathbb{R}^{\mathrm{d}}; \mathbb{R}^{\mathrm{d}})$, defined via the formula $f_u(s,\mu) := f(s,\mu,u(s))$ for a.e. $s \in [0,T]$ and $\mu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, satisfies the assumption of Lemma 3.6. Therefore, by the uniqueness of solutions to (1.2), the following statements can be readily deduced.

Corollary 3.7. *Assume Hypothesis* 1.3 *and take* $(t,\nu) \in [0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ *and a measurable control* $u : [t,T] \to U$. *Let* $(\boldsymbol{\mu}_s^{t,\nu,u})_{s\in[t,T]}$ *be the unique solution of* (1.2) *provided by Proposition* 3.4. *Then, for any sequence* $(\nu^n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ *such that* $\nu^n \overset{\boldsymbol{\Theta}}{\to}_n \nu$ *and any* $p \in [1,2)$, *we have* $W_p\left(\boldsymbol{\mu}_T^{t,\nu_n,u}, \boldsymbol{\mu}_T^{t,\nu,u}\right) \to 0$ *as* $n \to +\infty$.

*Proof.* It is enough to observe that, from Lemma 3.6, the only accumulation point of $W_p\left(\boldsymbol{\mu}_T^{t,\nu_n,u}, \boldsymbol{\mu}_T^{t,\nu,u}\right)$ is zero. ∎

We are now in a position to present a regularity result for the value function.

Lemma 3.8 (Regularity of the value function). *Assume Hypotheses* 1.2 *and* 1.3. *Let* $V : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{\pm\infty\}$ *be the value function given by* (1.3) *Then each* $V(t,\cdot)$ *is not identically* $+\infty$ *and* $V$ *is lower bounded and* $\boldsymbol{\Theta}$*-lower semicontinuous. Moreover, if* $\mathfrak{J}$ *is bounded and* $\boldsymbol{\Theta}$*-continuous, then so is* $V$.

*Proof.* Lower boundedness of $V$ follows from that of $\mathfrak{J}$ and by its definition. Let $\nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $\mathfrak{J}(\nu) < +\infty$. From Remark 3.5, for any control $u : [0,T] \to U$, there is $\tilde{\nu} \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ for which $\boldsymbol{\mu}_T^{0,\tilde{\nu},u} = \nu$. Therefore, $V(t, \boldsymbol{\mu}_t^{0,\tilde{\nu},u}) \leqslant \mathfrak{J}(\nu) < +\infty$.
**Step I:** From Lemma 2.4, to prove that $V$ is $\boldsymbol{\Theta}$-lower semicontinuous, we only have to show that it is sequentially lower semicontinuous for the inductive topology $\boldsymbol{\Theta}$. Let $(t_n,\nu_n)_n \subset [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ be such that $t_n \to t \in [0,T]$ and $\nu_n \overset{\boldsymbol{\Theta}}{\to}_n \nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. For each $n \in \mathbb{N}$, there is a measurable control $u_n : [t_n,T] \to U$ such that $V(t_n,\nu_n) \geqslant \mathfrak{J}(\boldsymbol{\mu}_T^{t_n,\nu_n,u_n}) - 1/n$.

We claim that there is $\tilde{\nu}_n \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $\tilde{\nu}_n \overset{\boldsymbol{\Theta}}{\to}_n \nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and a measurable control $\tilde{u}_n : [t,T] \to U$ such that $\boldsymbol{\mu}_T^{t,\tilde{\nu}_n,\tilde{u}_n} = \boldsymbol{\mu}_T^{t_n,\nu_n,u_n}$. Indeed, if $t_n < t$, set $\tilde{u}_n = u_n|_{[t,T]}$ a.e. and $\tilde{\nu}_n = \boldsymbol{\mu}_t^{t_n,\nu_n,u_n}$. Similarly as done in the proof of Lemma 3.6 (Step 1), we can find $c > 0$ such that $W_2(\nu_n, \tilde{\nu}_n) \leq c \int_{t_n}^t m_f(r) dr$.

Therefore $W_p(\widetilde{\nu}_n, \nu) \to 0$ as $n \to +\infty$, for any $p \in [1,2)$, and so $\tilde{\nu}_n \overset{\boldsymbol{\Theta}}{\to}_n \nu$ as well. On the other hand, if $t_n > t$, it is enough take any $\tilde{u}_n : [t,T] \to U$ such that $\tilde{u}_n = u_n|_{[t,T]}$ a.e. and take $\tilde{\nu}_n \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ given by Remark 3.5. Similarly as mentioned above, we can find $c > 0$ such that $W_2(\nu_n, \widetilde{\nu}_n) \leq c \int_t^{t_n} m_f(r)dr$, and so $\tilde{\nu}_n \overset{\boldsymbol{\Theta}}{\to}_n \nu$ as $n \to +\infty$.

Therefore, from Lemma 3.6, possibly along a subsequence, there is a measurable control $u : [t,T] \to U$ such that $W_1\left(\boldsymbol{\mu}_T^{t_n,\nu_n,u_n}, \boldsymbol{\mu}_T^{t_n,\nu,u}\right) \to 0$ as $n \to +\infty$. Thus, in particular, $\boldsymbol{\mu}_T^{t_n,\nu_n,u_n} \overset{\boldsymbol{\Theta}}{\to}_n \boldsymbol{\mu}_T^{t,\nu,u}$. Then, by lower semicontinuity of $\mathfrak{J}$ in $\boldsymbol{\Theta}$, we get

$$\liminf_{n\to\infty} V(t_n,\nu_n) \geqslant \liminf_{n\to\infty} \mathfrak{J}(\boldsymbol{\mu}_T^{t_n,\nu_n,u_n}) - \frac{1}{n} \geqslant \mathfrak{J}(\boldsymbol{\mu}_T^{t,\nu,u}) \geqslant V(t,\nu).$$

**Step II:** Assume now that $\mathfrak{J}$ is bounded and $\boldsymbol{\Theta}$-continuous. Then $V$ shares the same bound by definition. To prove that $V$ is $\boldsymbol{\Theta}$-upper semicontinuous, fix $(t,\nu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and take a sequence $(t_n,\nu_n)_n \subset [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $t_n \to t$ and $\nu_n \overset{\boldsymbol{\Theta}}{\to}_n \nu$. Up to extraction of a subsequence, we may assume that $\limsup_{n\to\infty} V(t_n,\nu_n) = \lim_{n\to\infty} V(t_n,\nu_n)$. For each $\varepsilon > 0$, there is a measurable control $u_\varepsilon : [t,T] \to U$ such that $V(t,\nu) \geqslant \mathfrak{J}(\boldsymbol{\mu}_T^{t,\nu,u_\varepsilon}) - \varepsilon$. Notice that if $t_n < t$, then from (1.4), we have that $V(t_n,\nu_n) \leq V(t, \boldsymbol{\mu}_t^{t_n,\nu_n,\tilde{u}_n})$ for any measurable control $\tilde{u}_n : [t_n,T] \to U$. As described above in **Step I**, it follows that $\tilde{\nu}_n := \boldsymbol{\mu}_t^{t_n,\nu_n,\tilde{u}} \overset{\boldsymbol{\Theta}}{\to}_n \nu$ as $n \to +\infty$. Therefore, if we take $\tilde{u}_n$ such that $\tilde{u}_n|_{[t,T]} = u_\varepsilon$, from Corollary 3.7 we have that $\boldsymbol{\mu}_T^{t,\tilde{\nu}_n,u_\varepsilon} \overset{\boldsymbol{\Theta}}{\to}_n \boldsymbol{\mu}_T^{t,\nu,u_\varepsilon}$. Then, since $\mathfrak{J}$ is $\boldsymbol{\Theta}$-continuous,

$$\lim_{n\to\infty} V(t_n,\nu_n) \leqslant \limsup_{n\to\infty} V(t,\tilde{\nu}_n) \leqslant \limsup_{t\to\infty} \mathfrak{J}(\boldsymbol{\mu}_T^{t,\tilde{\nu}_n,u_\varepsilon}) = \mathfrak{J}(\boldsymbol{\mu}_T^{t,\nu,u_\varepsilon}) \leqslant V(t,\nu) + \varepsilon.$$

On the other hand, if $t < t_n$, using Remark 3.5 and the arguments described above in **Step I**, there is $\tilde{\nu}_n \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $\tilde{\nu}_n \overset{\boldsymbol{\Theta}}{\to}_n \nu$ as $n \to +\infty$ with $\boldsymbol{\mu}_{t_n}^{t,\tilde{\nu}_n,u_\varepsilon} = \nu_n$. Notice that $\boldsymbol{\mu}_T^{t_n,\nu_n,u_\varepsilon} = \boldsymbol{\mu}_T^{t,\tilde{\nu}_n,u_\varepsilon}$, and moreover, $\boldsymbol{\mu}_T^{t_n,\tilde{\nu}_n,u_\varepsilon} \overset{\boldsymbol{\Theta}}{\to}_n \boldsymbol{\mu}_T^{t,\nu,u_\varepsilon}$ thanks to Corollary 3.7. Thus, by continuity of $\mathfrak{J}$ we have

$$\lim_{n\to\infty} V(t_n,\nu_n) \leqslant \lim_{t\to\infty} \mathfrak{J}(\boldsymbol{\mu}_T^{t_n,\nu_n,u_\varepsilon}) = \lim_{t\to\infty} \mathfrak{J}(\boldsymbol{\mu}_T^{t_n,\tilde{\nu}_n,u_\varepsilon}) = \mathfrak{J}(\boldsymbol{\mu}_T^{t,\nu,u_\varepsilon}) \leqslant V(t,\nu) + \varepsilon.$$

Letting $\varepsilon \searrow 0$ in both cases, we conclude that $V$ is $\boldsymbol{\Theta}$-continuous. ∎

**3.3. Approximation along a subsequence.** Recall that, as mentioned in the introduction, the HJB equation is the infinitesimal version of the DPP stated in (1.4), and as such, it requires to work with the time linearization of the mapping $s \mapsto \boldsymbol{\mu}_s^{t,\nu,u}$ at $s = t$. From the representation formula given in Proposition 3.4, we have that $\boldsymbol{\mu}_s^{t,\nu,u} = Y_s^{t,u}\#\nu$ where, for any given $x \in \mathbb{R}^{\mathrm{d}}$, the mapping $s \mapsto Y_s^{t,u}(x) \in \mathbb{R}^{\mathrm{d}}$ is the unique solution of (3.4). This mapping is merely absolutely continuous, since the associated vector field $s \mapsto f[s, \boldsymbol{\mu}_s, u(s)](x)$ is just measurable. However, in the autonomous case, we can still approximate a trajectory issued from $\nu$ by a "linear" curve along some given sequence of times.

As a similar result has been proved in [6], we just provide a sketch of the proof.

LEMMA 3.9. *Let $(t,\nu) \in [0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and $u : [t,T] \to U$ be a measurable control. Assume that $f$ satisfies Hypothesis* 1.3*, and is autonomous. Let $(\boldsymbol{\mu}_s^{t,\nu,u})_{s\in[t,T]}$ be the unique solution of* (1.2) *provided by Proposition* 3.4*. There exist $\overline{b} \in f[\nu, U]$ and a vanishing sequence $(h_n)_{n\in\mathbb{N}} \subset (0, T-t]$ such that*

$$\lim_{n\to\infty} \frac{W_2\left(\boldsymbol{\mu}_{t+h_n}^{t,\nu,u}, \exp_\nu(h_n \cdot (Id, \overline{b})\#\nu)\right)}{h_n} = 0.$$

*Proof.* For $h > 0$ such that $t + h \leqslant T$, consider $\overline{b}^h(x) \coloneqq \dfrac{1}{h}\displaystyle\int_t^{t+h} f[\boldsymbol{\mu}_s, u(s)](x)ds$ for any $x \in \mathbb{R}^{\mathrm{d}}$. By convexity, $\overline{b}^h \in f[\nu, U]$ and by compactness of $f[\nu, U]$ in the topology of uniform convergence over compact sets, there exists a sequence $(h_n)_n \searrow 0$ and some $\overline{b} \in f[\nu, U]$ such that $\overline{b}^{h_n}$ converges uniformly over compact sets towards $\overline{b}$. From [4, (7.1.6)] and the representation formula given in Proposition 3.4, we have

$$W_2(\boldsymbol{\mu}_s^{t,\nu,u}, Z_n\#\nu) \leq \|Y_s^{t,u} - Z_n\|_{L^2(\nu)}, \qquad \forall s \in [t, T],$$

where $Z_n = \mathrm{Id} + h_n\overline{b}$. The conclusion then follows from appropriate estimates of the right-hand side; see [6, Lemma 4.6] for details. ∎

**4. The Hamilton-Jacobi equation.** Let $H : \mathbb{T} \to \mathbb{R}$, and consider the HJ equation

$$-\partial_t v(t,\mu) + H\left(\mu, D_\mu v(t,\mu)\right) = 0 \quad t \in (0,T), \qquad v(T,\cdot) = \mathfrak{J}. \tag{4.1}$$

This section is devoted to the notion of solution adapted to (4.1). We first introduce a definition of viscosity solutions using test functions, and then prove a comparison principle that implies the uniqueness of the viscosity solution of (4.1).

**4.1. Definition of viscosity solutions.** We employ two distinct sets of test functions for the sub and supersolutions. Denote

$$\mathscr{T}_\pm \coloneqq \left\{ (t,\mu) \mapsto \psi(t) \pm \sum_{i=1}^N \alpha_i W_2^2(\mu,\nu_i) \;\middle|\; \begin{array}{c} \psi \in \mathcal{C}^1((0,T);\mathbb{R}),\ N \in \mathbb{N},\ \text{and} \\ (\alpha_i,\nu_i)_{i\in\{1,\cdots,N\}} \subset \mathbb{R}^+ \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}). \end{array} \right\}$$

In particular, test functions in $\mathscr{T}_+$ are $\boldsymbol{\Theta}$-lower semicontinuous, locally Lipschitz and directionally differentiable everywhere, and $\mathscr{T}_- = -\mathscr{T}_+$. As each term of the finite sum of the measure component is directionally differentiable, so is each $\varphi(t,\cdot)$ for $\varphi \in \mathscr{T}_\pm$, and there holds $D_\mu\varphi(t,\mu)(\xi) = \pm\sum_{i=1}^N \alpha_i D_\mu W_2^2(\mu,\nu_i)(\xi)$.

We consider the following definition.

Definition 4.1 (Viscosity solution). *$v : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{\pm\infty\}$ is called*

- *a viscosity subsolution of (4.1) if it is $\boldsymbol{\Theta}$-upper semicontinuous, does not take the value $+\infty$, and for each $\varphi \in \mathscr{T}_+$ such that $v - \varphi$ reaches a finite maximum in $(t,\mu) \in (0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, there holds*

$$-\partial_t\varphi(t,\mu) + H\left(\mu, D_\mu\varphi(t,\mu)\right) \leqslant 0. \tag{4.2}$$

- *a viscosity supersolution of (4.1) if it is $\boldsymbol{\Theta}$-lower semicontinuous, does not take the value $-\infty$, and for each $\varphi \in \mathscr{T}_-$ such that $v - \varphi$ reaches a finite minimum in $(t,\mu) \in (0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, there holds*

$$-\partial_t\varphi(t,\mu) + H\left(\mu, D_\mu\varphi(t,\mu)\right) \geqslant 0. \tag{4.3}$$

- *a viscosity solution of (4.1) if it is both a sub and supersolution, and if $v(T,\mu) = \mathfrak{J}(\mu)$.*

This notion of viscosity solution takes ideas from [29, 6] on the use of *locally Lipschitz and semiconcave test functions* in place of smooth test functions. As $\mathcal{C}^2$ functions are locally semiconcave and semiconvex, this corresponds to choosing the correct condition in order to match the properties of the squared distance. Symmetrically, in non-negatively curved spaces, one may choose semiconvex test functions.

**4.2. Comparison principle.** The comparison principle is the key result in the viscosity theory. It essentially says that subsolutions are always smaller than supersolutions in the pointwise sense. This gives the uniqueness of the viscosity solution, and in the classical theory, also allows to obtain existence for general nonconvex Hamiltonians. Owing to the local compactness of $\left(\mathscr{P}_2(\mathbb{R}^{\mathrm{d}}), \mathbf{\Theta}\right)$, our strategy to obtain a comparison principle is quite close to that of [18]. In particular, we base our proof in the so-called *doubling variables method*. This technique in turn relies on a suitable *penalization lemma*. We state an adaptation of [18, Prop. 3.7] to our case; its proof is essentially the same as in [18], so we omit it.

LEMMA 4.2 ([18, Prop. 3.7]). *Let $(Z,\tau)$ be a topological space, $\mathcal{O} \subset Z$ be a nonempty set, $\Phi : \mathcal{O} \to \mathbb{R} \cup \{-\infty\}$ be $\tau$-upper semicontinuous and proper in $\mathcal{O}$, $\Psi : \mathcal{O} \to \mathbb{R}$ be $\tau$-lower semicontinuous and nonnegative. For any $a > 0$, define $\Gamma_a \coloneqq \sup_{x\in\mathcal{O}} [\Phi(x) - a\Psi(x)]$. Assume that $-\infty < \lim_{a\to\infty} \Gamma_a < \infty$, and let $x_a \in \mathcal{O}$ be chosen such that $\lim_{a\to\infty} (\Gamma_a - (\Phi(x_a) - a\Psi(x_a))) = 0$. Then the following holds:*

1. $\lim_{a\to\infty} a\Psi(x_a) = 0$,
2. *whenever $\hat{x} \in \mathcal{O}$ is a limit point of $(x_a)_a$ in $(Z,\tau)$, then $\Psi(\hat{x}) = 0$ and $\lim_{a\to\infty} \Gamma_a = \Phi(\hat{x}) = \sup_{\Psi(x)=0} \Phi(x)$.*

The comparison principle will rely on the next assumptions on the Hamiltonian.

*Hypothesis* 4.3. There is $C_H \geqslant 0$ such that for all $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and $p, q \in \mathbb{T}_\mu$,

$$|H(\mu, p+q) - H(\mu, p)| \leqslant C_H \left(1 + \sqrt{\mathbf{m}_2(\mu)}\right) \|p\|_\mu, \tag{4.4}$$

and for all $a \geqslant 0$,

$$H\left(\mu, -aD_\mu W_2^2(\mu,\nu)\right) - H\left(\nu, aD_\nu W_2^2(\mu,\nu)\right) \leqslant aC_H W_2^2(\mu,\nu). \tag{4.5}$$

The fact that the Hamiltonian issued from the control problem (1.3) and given in section 5, equation (5.1), satisfies Hypothesis 4.3 is proved in [6, Section 6, Lemmata 6 and 7]. The argument relies on the explicit formula of the differential of the squared distance, and on the construction of a suitable plan by disintegration. The Hamiltonian in the quoted reference features an additional projection on the geometric tangent cone; however the directional derivative of the squared Wasserstein distance along the projection is the same as along the original vector, as is proved in [27, Corollary 4.34]. We are now in a position to state our comparison principle.

PROPOSITION 4.4 (Comparison principle). *Assume Hypotheses* 1.2, 1.3, *and* 4.3. *Let $v : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{-\infty\}$ be a subsolution of* (4.1) *bounded from above, and $w : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{+\infty\}$ be a supersolution of* (4.1) *bounded from below. Assume that there exists $\sigma \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $v(T,\sigma) - w(T,\sigma) \in \mathbb{R}$. Then*

$$\Gamma \coloneqq \sup_{(t,\mu)\in[0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})} (v(t,\mu) - w(t,\mu)) \leqslant \sup_{\mu\in\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})} (v(T,\mu) - w(T,\mu)) =: \Gamma_T.$$

*Proof.* The proof is based on the doubling variables method [18]. The underlying idea is the following. Assume by contradiction that $\Gamma_T > \Gamma$. From this condition, deduce the existence of $(t,\mu) \in (0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ and of some test functions $\varphi$ and $\psi$ such that: (i) $v - \varphi$ attains its maximum at $(t,\mu)$, (ii) $w - \psi$ attains its minimum at $(t,\mu)$, and (iii) $-\partial_t\varphi + H(\mu, D_\mu\varphi) > -\partial_t\psi + H(\mu, D_\mu\psi)$. As long as $t > 0$, the definition of viscosity sub and super solution yields, respectively, to $-\partial_t\varphi + H(\mu, D_\mu\varphi) \leqslant 0$

and $-\partial_t\psi + H(\mu, D_\mu\psi) \geqslant 0$. This leads to a contradiction, and completes the proof. However, such a $(t,\mu)$ may not exist, and one has to proceed by approximation.

By assumption, $\Gamma_T$ and $\Gamma$ are finite. Since $v$ is a subsolution and $w$ is a supersolution, $\Gamma_T \leq 0$. Up to replacing $v$ by $v - \Gamma_T$, we may assume that $\Gamma_T = 0$. Assume by contradiction that $\Gamma > 0$. Define $v^\alpha(t,\mu) \coloneqq v(t,\mu) + \alpha(t-T)$ for any $\alpha > 0$. Note that $v^\alpha$ is also a subsolution and, since $\Gamma > 0$, there is $\alpha > 0$ so that $\Gamma^\alpha \coloneqq \sup\{v^\alpha(t,\mu) - w(t,\mu) \mid (t,\mu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})\} > 0$. Let now $\sigma \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ be as in the statement, and take a vanishing sequence $(\varepsilon_i)_i \subset (0,1)$ such that

$$\Gamma_i^\alpha \coloneqq \sup_{(t,\mu)\in[0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})} \left(v^\alpha(t,\mu) - w(t,\mu) - 2\varepsilon_i\left(W_2^2(\sigma,\mu) + \frac{1}{t}\right)\right) > 0, \qquad \forall i \in \mathbb{N}.$$

The existence of the sequence is ensured by the fact that $\Gamma^\alpha > 0$. The sequence $(\Gamma_i^\alpha)_i$ is bounded above by $\Gamma^\alpha$, nondecreasing and converges to $\Gamma^\alpha$. For each $i, n \in \mathbb{N}$, let

$$\Phi_{i,n}(t,s,\mu,\nu) \coloneqq v^\alpha(t,\mu) - w(s,\nu) - n\frac{W_2^2(\mu,\nu) + |t-s|^2}{2} - \varepsilon_i\left(W_2^2(\sigma,\mu) + W_2^2(\sigma,\nu) + \frac{1}{t} + \frac{1}{s}\right).$$

We divide the rest of the proof into several parts.

**Point of maximum** Notice that if $\Phi_{i,n}(t,s,\mu,\nu) \geqslant \Phi_{i,n}(T,T,\sigma,\sigma)$, then

$$\varepsilon_i\left(W_2^2(\sigma,\mu) + W_2^2(\sigma,\nu)\right) \leqslant \sup(v^\alpha) - \inf(w) + \frac{2\varepsilon_i}{T} - (v^\alpha(T,\sigma) - w(T,\sigma)) < \infty.$$

Then there is $N_i \in \mathbb{N}$ such that $\{\Phi_{i,n} \geqslant \Phi_{i,n}(T,T,\sigma,\sigma)\} \subset [0,T]^2 \times \mathbf{K}_{N_i} \times \mathbf{K}_{N_i}$. From Lemmas 2.3 and 2.4 , the set $[0,T] \times [0,T] \times \mathbf{K}_{N_i} \times \mathbf{K}_{N_i}$ is (sequentially) compact in the product topology $\tau \coloneqq \tau_{[0,T]} \otimes \tau_{[0,T]} \otimes \mathbf{\Theta} \otimes \mathbf{\Theta}$, where $\tau_{[0,T]}$ is the usual metric topology on $[0,T]$. Moreover, since $\Phi_{i,n}$ is not identically $+\infty$, it is upper bounded and upper semicontinuous in this topology, there exists a maximum point $z_{i,n} \coloneqq (t_{i,n}, s_{i,n}, \mu_{i,n}, \nu_{i,n})$ of $\Phi_{i,n}$ over its domain. As $N_i$ is independent of $n \in \mathbb{N}$, there is a subsequence (which we do not relabel) of $(z_{i,n})_n$ that converges to some $z_i \coloneqq (t_i, s_i, \mu_i, \nu_i) \in [0,T]\times[0,T]\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})\times\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. Consequently, we may assume that $z_{i,n} \to z_i$ in the topology $\tau$ and that $(\Phi_{i,n}(z_{i,n}))_n$ converges.

Applying Lemma 4.2, we get that possibly along a subsequence,

$$\lim_{n\to+\infty} n\left(W_2^2(\mu_{i,n},\nu_{i,n}) + |t_{i,n} - s_{i,n}|^2\right) = 0 \quad \text{and} \quad \lim_{n\to+\infty} \Phi_{i,n}(z_{i,n}) = \Gamma_i^\alpha. \tag{4.6}$$

In particular, it follows that $t_i = s_i$ and moreover, $\mu_i = \nu_i$ because $W_2^2$ is $\mathbf{\Theta}$-lower semicontinuous (see Lemma 2.3).

**Staying away from the boundary** By construction, $t_{i,n} > 0$ and $s_{i,n} > 0$ for each $i, n \in \mathbb{N}$. On the other hand, for each $i \in \mathbb{N}$, there exists $n_i \in \mathbb{N}$ large enough so that $t_{i,n}, s_{i,n} < T$ for all $n \geq n_i$. Indeed, if it was not the case, for some $i \in \mathbb{N}$ there would exist an increasing sequence of natural numbers $(n_j)_j$ such that either $t_{i,n_j} = T$ or $s_{i,n_j} = T$. From (4.6), we get that $t_{i,n_j} \to T$ and $s_{i,n_j} \to T$ when $j \to \infty$, and using the uppercontinuity of $v^\alpha$ and the lower semicontinuity of $w$, we get

$$\Gamma_i^\alpha = \lim_{j\to\infty} \Phi_{i,n_j}(z_{i,n_j}) \leqslant \limsup_{j\to\infty} v^\alpha(t_{i,n_j},\mu_{i,n_j}) - w(s_{i,n_j},\nu_{i,n_j}) \leqslant v^\alpha(T,\mu_i) - w(T,\nu_i).$$

This is absurd because $\Gamma_i^\alpha > 0$ and $v^\alpha(T,\mu_i) - w(T,\nu_i) \leq \Gamma_T = 0$, and we get that $(t_{i,n}, s_{i,n}) \in (0,T)^2$ for $n \in \mathbb{N}$ large enough. Passing into a subsequence if necessary, we may assume this is true for any $i, n \in \mathbb{N}$.

**Applying the definition of semisolutions** For each $i, n \in \mathbb{N}$, define

$$\begin{aligned}\varphi(t,\mu) &\coloneqq \frac{n}{2}\left(W_2^2(\mu,\nu_{i,n}) + |t - s_{i,n}|^2\right) + \varepsilon_i\left(W_2^2(\sigma,\mu) + \frac{1}{t}\right) - \alpha t,\\ \psi(s,\nu) &\coloneqq -\frac{n}{2}\left(W_2^2(\mu_{i,n},\nu) + |t_{i,n} - s|^2\right) - \varepsilon_i\left(W_2^2(\sigma,\nu) + \frac{1}{s}\right).\end{aligned}$$

By construction, $\varphi \in \mathscr{T}_+$ and $\psi \in \mathscr{T}_-$. As $v - \varphi$ reaches a maximum in $(t_{i,n}, \mu_{i,n}) \in (0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, applying the definition of subsolution, we get

$$\alpha + \frac{\varepsilon_i}{t_{i,n}^2} - n(t_{i,n} - s_{i,n}) + H\left(\mu_{i,n}, \frac{n}{2} D_\mu W_2^2(\mu_{i,n},\nu_{i,n}) + \varepsilon_i D_\mu W_2^2(\sigma,\mu_{i,n})\right) \leqslant 0.$$

Using that $\frac{\varepsilon_i}{t_{i,n}^2} \geqslant 0$, the assumption (4.4) on the Hamiltonian and the estimate $\|D_\mu W_2^2(\mu,\nu)\|_\mu \leqslant 2W_2(\mu,\nu)$, we get

$$\begin{aligned}\alpha - n(t_{i,n} - s_{i,n}) + H\left(\mu_{i,n}, \frac{n}{2} D_\mu W_2^2(\mu_{i,n},\nu_{i,n})\right)\\ -2\varepsilon_i W_2(\sigma,\mu_{i,n}) C_H(1 + \sqrt{\mathbf{m}_2(\mu_{i,n})}) \leqslant 0.\end{aligned} \tag{4.7}$$

Similarly, $w - \psi$ reaches a minimum in $(s_{i,n}, \nu_{i,n})$. Using the same reasoning as above,

$$\begin{aligned}n(s_{i,n} - t_{i,n}) + H\left(\nu_{i,n}, -\frac{n}{2} D_\nu W_2^2(\mu_{i,n},\nu_{i,n})\right)\\ +2\varepsilon_i W_2(\sigma,\nu_{i,n}) C_H(1 + \sqrt{\mathbf{m}_2(\nu_{i,n})}) \geqslant 0.\end{aligned} \tag{4.8}$$

Combining (4.7) and (4.8) and using the assumption (4.5), there holds for any $i, n \in \mathbb{N}$

$$\alpha \leqslant nC_H W_2^2(\mu_{i,n},\nu_{i,n}) + 2\varepsilon_i C_H \sum_{\varpi \in \{\mu_{i,n},\nu_{i,n}\}} W_2(\sigma,\varpi)(1 + \sqrt{\mathbf{m}_2(\varpi)}). \tag{4.9}$$

**Vanishing perturbation** Recall that $z_{i,n} \to z_i$ as $n \to +\infty$, where the convergence is understood in $\boldsymbol{\Theta}$ for the measure coordinates. Passing to the limit in (4.9) will not give useful information, since the squared Wasserstein distance is only $\boldsymbol{\Theta}$-lower semicontinuous, and we will not obtain an inequality on $z_i$. Therefore, we extract a diagonal sequence. Proceeding by induction and using (4.6), we may build an increasing sequence of natural numbers $(n_i)_i$ for which

$$n_i W_2^2(\mu_{i,n_i},\nu_{i,n_i}) \leqslant \frac{1}{i}, \quad \sup(\Phi_{i,n_i}) \geqslant \Gamma_i^\alpha - \frac{1}{i}, \quad \sup\left(\Phi_{i+1,n_{i+1}}\right) \geqslant \sup(\Phi_{i,n_i}).$$

The sequence $\left(\sup(\Phi_{i,n_i})\right)_i$ is nondecreasing and upper bounded by $\Gamma^\alpha$, thus converges. On the other hand,

$$\begin{aligned}0 &\leqslant \frac{\varepsilon_i}{2}\left(W_2^2(\sigma,\mu_{i,n_i}) + W_2^2(\sigma,\nu_{i,n_i}) + 0\right) \leqslant \Phi_{i+1,n_{i+1}}(z_i) - \Phi_{i,n_i}(z_i)\\ &\leqslant \sup\left(\Phi_{i+1,n_{i+1}}\right) - \sup\left(\Phi_{i+1,n_{i+1}}\right) \underset{i\to\infty}{\longrightarrow} 0.\end{aligned}$$

Evaluating (4.9) along the subsequence $(n_i)_i$ and passing to the limit, we obtain $\alpha \leqslant 0$, which is absurd. Consequently, $\Gamma \leqslant 0$. ∎

**5. Characterisation of the solution in the case of control problems.** Let us assume from now on that $f$ is autonomous. From the available theory on Euclidean spaces, we may expect the value function given by (1.3) to be a viscosity supersolution of the HJB equation (4.1) for the Hamiltonian

$$H : \mathbb{T} \to \mathbb{R}, \qquad H(\mu, p) \coloneqq \sup_{u \in U} -p\left((Id, f[\mu, u])\#\mu\right), \tag{5.1}$$

and a solution whenever $\mathfrak{J}$ is real-valued and $\boldsymbol{\Theta}$-continuous. From this point onward, unless otherwise stated, we assume that the Hamiltonian in (4.1) is given by (5.1).

Let us present an example. Let $\mathfrak{J} : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{+\infty\}$ be given by

$$\mathfrak{J}(\mu) \coloneqq \mathbf{m}_4(\mu) \coloneqq \int_{\mathbb{R}^{\mathrm{d}}} |x|^4 \, d\mu(x).$$

The function $\mathfrak{J}$ is finite over $\mathscr{P}_4(\mathbb{R}^{\mathrm{d}})$, the space of Borel probability measures on $\mathbb{R}^{\mathrm{d}}$ with finite 4-moment. Notice that $\mathscr{P}_4(\mathbb{R}^{\mathrm{d}})$ is strictly contained in $\mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, and closed with respect to $\boldsymbol{\Theta}$, since $\mu \mapsto \mathbf{m}_4(\mu)$ is narrowly lower semicontinuous; see [4, Lemma 7.1.4]. Let the dynamic $f : \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \times [0,1] \to \mathcal{C}(\mathbb{R}^{\mathrm{d}}; \mathbb{R}^{\mathrm{d}})$ be given by

$$f[\mu, u](x) \coloneqq -ux, \qquad \forall x \in \mathbb{R}^{\mathrm{d}},\ u \in [0,1].$$

Then $f[\mu, U]$ is convex, compact in the topology of uniform convergence on compact sets, and $|f[\mu, u](0) + \mathrm{Lip}\,(f[\mu, u])| \leqslant 1$. For each $u : [0,T] \to [0,1]$ measurable, the flow of the underlying ODE is given by $Y_s^{t,u}(x) = \exp\left(-\int_t^s u(r)dr\right) x$. Consequently,

$$\mathfrak{J}\left(Y_s^{t,u}\#\mu\right) = \int_{\mathbb{R}^{\mathrm{d}}} \exp\left(-4\int_t^s u(r)dr\right)|x|^4 \, d\mu(x) = \exp\left(-4\int_t^s u(r)dr\right)\mathfrak{J}(\mu),$$

and minimizing over $u : [0,T] \to [0,1]$ measurable, the value function is given by

$$V(t,\mu) = e^{-4(T-t)}\mathfrak{J}(\mu) \in \mathbb{R} \cup \{+\infty\}. \tag{5.2}$$

Let us compute the Hamiltonian on $V$. Let $(t,\mu) \in \mathrm{dom}\ V$. From the expression of $V$ in (5.2), the derivative with respect to time is seen to be equal to $\partial_t V(t,\mu) = 4V(t,\mu)$. On the other hand, the directional derivative along curves of measures of the form $h \mapsto \exp_\mu(h \cdot (Id, f[\mu,u])\#\mu)$ can be explicitly computed by

$$\mathfrak{J}(\exp_\mu(h \cdot (Id, f[\mu,u])\#\mu)) = \int_{\mathbb{R}^{\mathrm{d}}} |x + h(-ux)|^4 \, d\mu = (1 - 4hu + o(h))\,\mathfrak{J}(\mu),$$

which yields $D_\mu V(t,\mu)((Id, f[\mu,u])\#\mu) = -4uV(t,\mu)$. Gathering the above computations, there holds

$$-\partial_t V(t,\mu) + \sup_{u\in[0,1]} -D_\mu V(t,\mu)((Id, f[\mu,u])\#\mu) = -4V(t,\mu) + \sup_{u\in[0,1]} 4uV(t,\mu) = 0.$$

This directly implies that $V$ is a viscosity supersolution of (4.1) in the sense of Definition 4.1. Indeed, if $\varphi \in \mathscr{T}_-$ is such that $V - \varphi$ reaches a minimum in $(t,\mu) \in (0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$, then $\partial_t\varphi(t,\mu) = \partial_t V(t,\mu)$ and $D_\mu\varphi(t,\mu)(\xi) \leqslant D_\mu V(t,\mu)(\xi)$ along each $\xi \in \mathscr{P}_2(\mathrm{T}\,\mathbb{R}^{\mathrm{d}})_\mu$ such that $D_\mu V(t,\mu)(\xi)$ exists. Hence the inequality

$$-\partial_t\varphi(t,\mu) + \sup_{u\in U} -D_\mu\varphi(t,\mu)((Id, f[\mu,u])\#\mu) \geqslant 0.$$

The aim of this section is to show that this remains valid in our setting. We begin by the case where $\mathfrak{J}$ may be unbounded, and then restrict to a more regular case.

**5.1. General case.** We now prove that $V : [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}) \to \mathbb{R} \cup \{\pm\infty\}$, the value function given by (1.3), is a supersolution of (4.1) for the Hamiltonian (5.1).

THEOREM 5.1. *Assume Hypotheses* 1.2 *and* 1.3*, and that $f$ is autonomous. Then the value function is a supersolution of* (4.1) *in the sense of Definition* 4.1.

*Proof.* By Lemma 3.8, $V$ is lower bounded and $\boldsymbol{\Theta}$-lower semicontinuous. Let $\varphi \in \mathscr{T}_-$ such that $V - \varphi$ reaches a minimum in $(t,\nu) \in (0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. Let $\left((\boldsymbol{\mu}_s^{t,\nu,u^n})_{s\in[t,T]}\right)_n$ be a minimizing sequence for the optimization (1.3). By Lemma 3.6, there is a measurable control $u : [t,T] \to U$ such that $\boldsymbol{\mu}_T^{t,\nu,u^n} \xrightarrow{\boldsymbol{\Theta}} \boldsymbol{\mu}_T^{t,\nu,u}$ (possibly along a subsequence which we do not relabel). Since $\mathfrak{J}$ is $\boldsymbol{\Theta}$-lower semicontinuous we have

$$V(t,\nu) = V\left(t+h, \boldsymbol{\mu}_{t+h}^{t,\nu,u}\right) \qquad \forall h \in [0, T-t].$$

Recall that $\varphi(t,\mu) = \psi(t) + g(\mu)$, where $\psi \in \mathcal{C}^1((0,T);\mathbb{R})$ and $g$ is locally Lipschitz, directionally differentiable and $\boldsymbol{\Theta}$-lower semicontinuous. Thus, it follows that

$$\psi(t+h) - \psi(t) + g\left(\boldsymbol{\mu}_{t+h}^{t,\nu,u}\right) - g(\nu) \leqslant V\left(t+h, \boldsymbol{\mu}_{t+h}^{t,\nu,u}\right) - V(t,\nu) = 0, \quad \forall h \in [0, T-t].$$

Using Lemma 3.9, there exists a sequence $(h_n)_n \subset (0, T-t]$ with $h_n \searrow 0$, and some $b \in f[\nu, U]$ such that $W_2\left(\boldsymbol{\mu}_{t+h_n}^{t,\nu,u}, \exp_\nu(h_n \cdot (Id,b)\#\nu)\right) = o(h_n)$. Dividing the above by $h_n > 0$, and denoting $\mathrm{Lip}\,(g)$ a local Lipschitz constant of $g$ in a ball centered in $\nu$ and containing all $\boldsymbol{\mu}_{t+h_n}^{t,\nu,u}$ and $\exp_\nu(h_n \cdot (Id,b)\#\nu)$, we have

$$\frac{\psi(t+h_n) - \psi(t)}{h_n} + \frac{g(\exp_\nu(h_n \cdot (Id,b)\#\nu)) - g(\nu)}{h_n} \leqslant \mathrm{Lip}\,(g)\, O(h_n),$$

where $O(h_n) = \frac{W_2\left(\exp_\nu(h_n\cdot(Id,b)\#\nu), \boldsymbol{\mu}_{t+h_n}^{t,\nu,u}\right)}{h_n}$. Taking the limit in $n \to \infty$ and using the respective differentiability of $\psi$ and $g$,

$$\partial_t \psi(t) + D_\nu g(\nu)\left((Id,b)\#\nu\right) \leqslant 0.$$

Multiplying by $-1$ and taking the maximum over all $b \in f[\nu,U]$, we obtain that $-\partial_t\varphi(t,\nu) + H\left(\nu, D_\nu\varphi(t,\nu)\right) \geqslant 0$, which is the desired property. ∎

**5.2. Case of continuous and bounded terminal cost.** We show that in the case where $\mathfrak{J}$ is bounded and $\boldsymbol{\Theta}$-continuous, the value function is also a subsolution of (4.1). As it satisfies the terminal condition by definition, the comparison principle then guarantees that it will then be the unique viscosity solution.

THEOREM 5.2. *Assume Hypotheses* 1.2 *and* 1.3*. Suppose that $f$ is autonomous and that $\mathfrak{J}$ is bounded and $\boldsymbol{\Theta}$-continuous. Then $V$ is a subsolution of* (4.1).

*Proof.* By Lemma 3.8, the value function is bounded and $\boldsymbol{\Theta}$-upper semicontinuous. Let $\varphi \in \mathscr{T}_+$, with $\varphi(s,\nu) = \psi(s) + g(\nu)$, and $(t,\mu) \in (0,T) \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $V - \varphi$ reaches a maximum in $(t,\mu)$. Let $b \in f[\mu,U]$ be arbitrary but fixed. Then the flow $s \to Y_s^t$ of the autonomous ODE $\frac{d}{ds} y_s = b(y_s)$ is of class $\mathcal{C}^1$, and there holds

$$\lim_{h\searrow 0} \frac{W_2\left(Y_{t+h}^t \#\mu, \exp_\mu(h\cdot(Id,b)\#\mu)\right)}{h} = 0.$$

Denote $\boldsymbol{\mu}_s \coloneqq Y_s^t \#\mu$. Using the dynamic programming principle,

$$0 \leqslant V\left(s, \boldsymbol{\mu}_s\right) - V(t,\mu) \leqslant \varphi(s,\boldsymbol{\mu}_s) - \varphi(t,\mu) = \psi(s) - \psi(t) + g(\boldsymbol{\mu}_s) - g(\mu).$$

As $g$ is locally Lipschitz, dividing by $s-t$ and sending $s \to t$, we get

$$\partial_t \psi(t) + D_\mu g(\mu)((Id, b)\#\mu) \geqslant 0.$$

Multiplying by $-1$ and taking the sup on $b \in f[\mu, U]$ yields that $V$ is a subsolution. □

In the general case, $V$ may take the value $+\infty$ and has no chance to be a viscosity subsolution of the HJB equation (4.1). However, we may still prove that it is the smallest supersolution in the pointwise sense. The argument proceeds by truncation and regularization, relying on the following result.

LEMMA 5.3. *Let $\mathfrak{J} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ be lower bounded and $\boldsymbol{\Theta}$-lower semicontinuous. Then for each $N \in \mathbb{N}$, there is a nondecreasing sequence of bounded $\boldsymbol{\Theta}$-continuous maps $\mathfrak{J}_n : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ that converge pointwise towards $\mathfrak{J}$ over $\mathbf{K}_N$.*

*Proof.* Denote $\boldsymbol{\chi}_N : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ the characteristic function of $\mathbf{K}_N$, i.e. $\boldsymbol{\chi}_N(\nu) = 0$ if $\mathbf{m}_2(\nu) \leqslant N$, and $\boldsymbol{\chi}_N(\nu) = +\infty$ otherwise. Since $\mathbf{K}_N$ is $\boldsymbol{\Theta}-$compact, $\boldsymbol{\chi}_N$ is $\boldsymbol{\Theta}-$lower semicontinuous. Moreover, the function $\nu \mapsto \mathfrak{J}(\nu) + \boldsymbol{\chi}_N(\nu)$ is narrowly lower semicontinuous. Indeed, this is due to the fact that the topology $\boldsymbol{\Theta}$ coincides with the narrow topology on $\mathbf{K}_N$. Let $\mathbb{d} : \mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^+$ be a metric inducing the topology of narrow convergence over $\mathscr{P}_2(\mathbb{R}^d)$ (e.g. [4, Section 5.1]), and

$$\mathfrak{J}_n(\mu) \coloneqq \min\left(n, \inf_{\nu \in \mathscr{P}_2(\mathbb{R}^d)} (\mathfrak{J} + \boldsymbol{\chi}_N)(\nu) + n\mathbb{d}(\mu, \nu)\right).$$

We directly have $\mathfrak{J}_n(\mu) \leqslant \min(n, \mathfrak{J}(\mu)) \leqslant \mathfrak{J}(\mu)$ for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. As $\mathfrak{J}_n$ is lower bounded, using that $\inf(g_1) - \inf(g_2) \leqslant \sup(g_1 - g_2)$, for each $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ we get

$$\mathfrak{J}_n(\mu_0) - \mathfrak{J}_n(\mu_1) \leqslant \max\left(0, \sup_{\nu \in \mathscr{P}_2(\mathbb{R}^d)} n\left(\mathbb{d}(\mu_0, \nu) - \mathbb{d}(\mu_1, \nu)\right)\right) \leqslant n\mathbb{d}(\mu_0, \mu_1).$$

By symmetry, $\mathfrak{J}_n$ is $n$-Lipschitz w.r.t. $\mathbb{d}$, thus $\boldsymbol{\Theta}$-continuous. It is moreover bounded with values in $[\min(0, \inf(\mathfrak{J})), n]$. To prove pointwise convergence, let $\mu \in \mathbf{K}_N$ be fixed. Assume by contradiction that there exists $M < \mathfrak{J}(\mu)$ such that $\mathfrak{J}_n(\mu) \leqslant M$ for all $n$. Let $\varepsilon \coloneqq \min(1, \mathfrak{J}(\mu) - M) > 0$. Since $\mathfrak{J} + \boldsymbol{\chi}_N$ is narrowly lower semicontinuous, there exists $r > 0$ such that $\mathbb{d}(\mu, \nu) < r$ implies $(\mathfrak{J} + \boldsymbol{\chi}_N)(\nu) \geqslant M + \varepsilon/2$. In particular, $\inf_{\mathbb{d}(\mu,\nu)<r} (\mathfrak{J} + \boldsymbol{\chi}_N)(\nu) + n\mathbb{d}(\mu, \nu) \geq M + \varepsilon/2$. From the definition of $\mathfrak{J}_n$ we have

$$\mathfrak{J}_n(\mu) = \min\left(n, \inf_{\mathbb{d}(\mu,\nu)<r} (\mathfrak{J} + \boldsymbol{\chi}_N)(\nu) + n\mathbb{d}(\mu, \nu), \inf_{\mathbb{d}(\mu,\nu)\geqslant r} (\mathfrak{J} + \boldsymbol{\chi}_N)(\nu) + n\mathbb{d}(\mu, \nu)\right),$$

and so

$$\mathfrak{J}_n(\mu) \geq \min\left(n, M + \frac{\varepsilon}{2}, \inf_{\mathbb{d}(\mu,\nu)\geqslant r} (\mathfrak{J} + \boldsymbol{\chi}_N)(\nu) + nr)\right) \geq \min\left(n, M + \frac{\varepsilon}{2}, \inf \mathfrak{J} + nr\right).$$

Taking $n$ large enough so that $nr \geqslant M + \frac{\varepsilon}{2} - \inf(\mathfrak{J})$ and $n \geqslant M + \varepsilon/2$, we get $\mathfrak{J}_n(\mu) \geqslant M + \frac{\varepsilon}{2}$, which is absurd. Thus the claim. □

This regularization allows us to implement the strategy of [32, Prop. 1.1].

THEOREM 5.4. *Assume Hypotheses* 1.2 *and* 1.3. *Suppose that $f$ is autonomous. Then for any supersolution $v : [0,T] \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ of* (4.1) *such that $v(T, \cdot)$ is not identically $+\infty$ there holds*

$$v(t, \nu) \geqslant V(t, \nu), \qquad \forall (t, \nu) \in [0, T] \times \mathscr{P}_2(\mathbb{R}^d). \tag{5.3}$$

*Consequently, the value function $V$ is the smallest viscosity supersolution of* (4.1).

*Proof.* Let $(\bar{t},\bar{\nu}) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$. From (3.3), we can find $N \in \mathbb{N}$, such that $\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u} \in \mathbf{K}_N$ for any measurable control $u : [\bar{t},T] \to U$. If $v(\bar{t},\bar{\nu}) = \infty$, the inequality (5.3) is trivially satisfied. Assume now that $v(\bar{t},\bar{\nu}) < \infty$. Let $(\mathfrak{J}_n)_n$ be given by Lemma 5.3. By Theorems 5.1 and 5.2, the HJB equation

$$-\partial_t v_n(t,\mu) + H\left(\mu, D_\mu v_n(t,\mu)\right) = 0, \qquad v_n(T,\mu) = \mathfrak{J}_n(\mu) \tag{5.4}$$

admits a unique solution given by

$$V_n(t,\nu) \coloneqq \inf\left\{ \mathfrak{J}_n(\boldsymbol{\mu}_T^{t,\nu,u}) \;\middle|\; u : [t,T] \to U \text{ is measurable} \right\},\ \forall (t,\nu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^{\mathrm{d}}).$$

Note that the map $v$ is a supersolution of each regularized problem (5.4). Let $\sigma \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$ such that $v(T,\sigma) \in \mathbb{R}$: since $v(T,\sigma) \geqslant \mathfrak{J}(\sigma) \geqslant \mathfrak{J}_n(\sigma) = V_n(T,\sigma)$, we have $-\infty < V_n(T,\sigma) - v(T,\sigma) \leqslant 0$. In consequence, we can apply Proposition 4.4, and deduce that $v(t,\nu) \geqslant V_n(t,\nu)$ for any $(t,\nu) \in \mathscr{P}_2(\mathbb{R}^{\mathrm{d}})$.

By Proposition 4.4, the solutions $V_n$ are ordered in the sense that $V_{n+1}(t,\nu) \geqslant V_n(t,\nu)$ for all $n$. Moreover, $\mathfrak{J}_n \leqslant \mathfrak{J}$ implies that the subsolutions $V_n$ are smaller than the supersolution $V$. Hence the sequence $(V_n(\bar{t},\bar{\nu}))_n$ is nondecreasing and upper bounded by $v(\bar{t},\bar{\nu}) < \infty$, and converges. For each $n \in \mathbb{N}$, let $u^n : [t,T] \to U$ be a measurable control such that $V_n(\bar{t},\bar{\nu}) \geqslant \mathfrak{J}_n\left(\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u^n}\right) - \frac{1}{n}$. Using Lemma 3.6, some (non relabeled) subsequence converges in $\boldsymbol{\Theta}$ to $\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u}$ for some measurable control $u : [t,T] \to U$. Using the monotonicity of the family $(\mathfrak{J}_n)_n$ and the continuity in $\boldsymbol{\Theta}$ of each $\mathfrak{J}_m$ for a fixed $m$,

$$\lim_{n\to\infty} V_n(\bar{t},\bar{\nu}) \geqslant \liminf_{n\to\infty} \mathfrak{J}_n\left(\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u^n}\right) - \frac{1}{n} \geqslant \liminf_{n\to\infty, n\geqslant m} \mathfrak{J}_m\left(\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u^n}\right) - \frac{1}{n} = \mathfrak{J}_m\left(\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u}\right).$$

Since $\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u^n} \in \mathbf{K}_N$, the conclusion follows from taking the limit in $m \to \infty$ to obtain

$$v(\bar{t},\bar{\nu}) \geqslant \lim_{n\to\infty} V_n(t,\nu) \geqslant \mathfrak{J}\left(\boldsymbol{\mu}_T^{\bar{t},\bar{\nu},u}\right) \geqslant V(\bar{t},\bar{\nu}).$$

∎

**Acknowledgments.** The second author would like to thank the Center for Mathematical Modeling, in particular to the CMM-PhD Visiting Program 2022 for the opportunity to conduct this research at Universidad Técnica Federico Santa María.

## REFERENCES

[1] C. Aliprantis and K. Border, *Infinite Dimensional Analysis*, Springer-Verlag, 2006.
[2] L. Ambrosio and J. Feng, *On a class of first order Hamilton–Jacobi equations in metric spaces*, Journal of Differential Equations, 256 (2014), pp. 2194–2245.
[3] L. Ambrosio and W. Gangbo, *Hamiltonian ODEs in the Wasserstein space of probability measures*, Communications on Pure and Applied Mathematics, 61 (2008), pp. 18–53.
[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows*, Birkhäuser-Verlag, 2005.
[5] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, 1990.
[6] A. Aussedat, O. Jerhaoui, and H. Zidani, *Viscosity solutions of centralized control problems in measure spaces*, ESAIM: Control, Optimisation and Calculus of Variations, 30 (2024).
[7] E. Bayraktar, I. Ekren, and X. Zhang, *Comparison of viscosity solutions for a class of second order PDEs on the Wasserstein space*, Communications in Partial Differential Equations, 50 (2025), pp. 570–613.
[8] B. Bonnet-Weill and H. Frankowska, *Carathéodory theory and a priori estimates for continuity inclusions in the space of probability measures*, Nonlinear Analysis, 247 (2024).
[9] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer New York, 2011.

[10] P. Cardaliaguet, *Notes on Mean Field Games*, (2013).
[11] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games*, Princeton University Press, 2019.
[12] P. Cardaliaguet and A. Porretta, *An Introduction to Mean Field Game Theory*, in Mean Field Games: Cetraro, Italy 2019, Y. Achdou et al., eds., Springer International Publishing, 2020, pp. 1–158.
[13] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I*, Springer International Publishing, 2018.
[14] 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, A. Muntean and F. Toschi, eds., Springer, 2014, pp. 1–46.
[15] G. Cavagnari, G. Savaré, and G. Sodini, *A Lagrangian approach to totally dissipative evolutions in Wasserstein spaces*, (2023), https://doi.org/10.48550/arXiv.2305.05211.
[16] A. Corbetta, *Multiscale Crowd Dynamics: Physical Analysis, Modeling and Applications*, PhD thesis, Eindhoven University of Technology, 2016.
[17] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato, *Master Bellman equation in the Wasserstein space: Uniqueness of viscosity solutions*, Transactions of the American Mathematical Society, 31–83 (2024), pp. 31–83.
[18] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, 1992.
[19] M. G. Crandall and P.-L. Lions, *Hamilton-Jacobi equations in infinite dimensions I. uniqueness of viscosity solutions*, Journal of Functional Analysis, 62 (1985), pp. 379–396.
[20] S. Daudin, J. Jackson, and B. Seeger, *Well-posedness of Hamilton-Jacobi equations in the Wasserstein space: non-convex Hamiltonians and common noise*, Communications in Partial Differential Equations, 50 (2023), pp. 1–52.
[21] J. Diestel, W. M. Ruess, and W. Schachermayer, *On weak compactness in $L^1(\mu, x)$*, Proceedings of the American Mathematical Society, 118 (1993), pp. 447–453.
[22] J. Diestel and J. Uhl, *Vector Measures*, American Mathematical Society, 1977.
[23] M. Fabian, P. Habala, P. Hájek, V. Montesinos, and V. Zizler, *Banach space theory: The basis for linear and nonlinear analysis*, Springer Science & Business Media, 2011.
[24] S. P. Franklin, *Spaces in which sequences suffice*, Fund. Math., 57 (1965), pp. 107–115.
[25] W. Gangbo and A. Tudorascu, *On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations*, Journal de Mathématiques Pures et Appliquées, 125 (2019), pp. 119–174.
[26] Y. Giga, N. Hamamuki, and A. Nakayasu, *Eikonal equations in metric spaces*, Transactions of the American Mathematical Society, 367 (2015), pp. 49–66.
[27] 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.
[28] R. Hynd and H. K. Kim, *Value functions in the Wasserstein spaces: Finite time horizons*, Journal of Functional Analysis, 269 (2015), pp. 968–997.
[29] F. Jean, O. Jerhaoui, and H. Zidani, *Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition*, SIAM Journal on Mathematical Analysis, 56 (2024), pp. 3326–3356.
[30] O. Jerhaoui and H. Zidani, *Viscosity solutions of Hamilton-Jacobi equations in proper $CAT(0)$ spaces*, Journal of Geometric Analysis, 34 (2024).
[31] C. Jimenez, A. Marigonda, and M. Quincampoix, *Optimal control of multiagent systems in the Wasserstein space*, Calculus of Variations and Partial Differential Equations, 59 (2020).
[32] P.-L. Lions and P. Souganidis, *Differential Games, Optimal Control and Directional Derivatives of Viscosity Solutions of Bellman's and Isaacs' Equations.*, SIAM Journal on Control and Optimization, 23 (1984), pp. 566–583.
[33] A. Marigonda and M. Quincampoix, *Mayer control problem with probabilistic uncertainty on initial positions*, Journal of Differential Equations, 264 (2018), pp. 3212–3252.
[34] F. Otto, *The Geometry of Dissipative Evolution Equations: The Porous Medium Equation*, Communications in Partial Differential Equations, 26 (2001), pp. 101–174.
[35] B. Piccoli and A. Tosin, *Pedestrian flows in bounded domains with obstacles*, Continuum Mechanics and Thermodynamics, 21 (2009), pp. 85–107.
[36] R. Rockafellar, *Integrals which are convex functionals*, Pacific journal of mathematics, 24 (1968), pp. 525–539.
[37] H. M. Soner and Q. Yan, *Viscosity Solutions of the Eikonal Equation on the Wasserstein Space*, Applied Mathematics & Optimization, 90 (2024).
[38] R. Vinter, *Optimal Control*, Springer, 2010.
[39] C. Zalinescu, *Convex analysis in general vector spaces*, World Scientific Publishing Co., 2002.