# GENERALIZED DYNAMIC PROGRAMMING PRINCIPLE AND SPARSE MEAN-FIELD CONTROL PROBLEMS

GIULIA CAVAGNARI, ANTONIO MARIGONDA, AND BENEDETTO PICCOLI

Abstract. In this paper we study optimal control problems in Wasserstein spaces, which are suitable to describe macroscopic dynamics of multi-particle systems. The dynamics is described by a parametrized continuity equation, in which the Eulerian velocity field is affine w.r.t. some variables. Our aim is to minimize a cost functional which includes a control norm, thus enforcing a *control sparsity* constraint. More precisely, we consider a nonlocal restriction on the total amount of control that can be used depending on the overall state of the evolving mass. We treat in details two main cases: an instantaneous constraint on the control applied to the evolving mass and a cumulative constraint, which depends also on the amount of control used in previous times. For both constraints, we prove the existence of optimal trajectories for general cost functions and that the value function is viscosity solution of a suitable Hamilton-Jacobi-Bellmann equation. Finally, we discuss an abstract Dynamic Programming Principle, providing further applications in the Appendix.

## 1. Introduction

The investigation of optimal control problems in the space of measures is attracting an increasing interest by the mathematical community in the last years, due to the potential applications in the study of complex systems, or multi-agent systems. See for example [21] for crowd dynamics models, and [5] for social network analysis.

Many physical and biological phenomena can be modelled by *multi-particle system* with a large number of particles. At the *microscopic* level, the behaviour of each particle is determined by local and nonlocal interactions. As the number of particles increases, the complexity of the system grows very fast. It turns out that only a *macroscopic (statistical)* description of the state of such a system can be actually provided, in the sense that an observer can deduce the state of the system only by measuring averages of suitable quantities (as, e.g., in statistical mechanics).

Assuming that there are neither creation nor loss of agents during the evolution, it is natural to describe the state of the system by a time-dependent probability measure $\mu_t$ on $\mathbb{R}^d$, in the sense that for every Borel set $A \subseteq \mathbb{R}^d$ and $t \geq 0$ the quantity $\mu_t(A)$ corresponds to the fraction of the total agents that are cointained in $A$ at time $t$.

An alternative point of view is the following: suppose that the state of the system is expressed by a time-dependent real valued map $\Phi_t$ defined on the set of continuous and bounded functions on $\mathbb{R}^d$. In this sense, $\Phi_t(\varphi) \in \mathbb{R}$ expresses the result of the observer's measurement of the quantity $\varphi(\cdot)$ on the system at time $t$. If we assume that $\Phi_t$ is positive, linear and continuous, by Riesz representation theorem, we have that $\Phi_t$ - and hence the state of the system - can be uniquely represented by a Borel measure $\mu_t$ on $\mathbb{R}^d$, in the sense

*Date*: April 22, 2019.

2010 *Mathematics Subject Classification.* 34A60, 49J15.

*Key words and phrases.* multi-agent mean field sparse control, Hamilton-Jacobi equation in Wasserstein space, control with uncertainty, dynamic programming principle.

Declaration of interest: none.

that

$$\Phi_t(\varphi) = \int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x).$$

In [14, 16, 17] time-optimal control problems in the space of probability measures $\mathscr{P}(\mathbb{R}^d)$ are addressed, by considering systems without interactions among the agents. The authors were able to extend to this framework some classical results, among which an Hamilton-Jacobi-Bellman (briefly HJB) equation solved by the minimum-time function in a suitable viscosity sense. However, a full characterization of the minimum time function in terms of uniqueness of the solution of the corresponding HJB equation is still missing.

In [13], the author focuses on results concerning conditions for attainability and regularity of the minimum time function in $\mathscr{P}(\mathbb{R}^d)$. A set of tangent distributions in the 2-Wasserstein space is introduced in [4] for the study of viability results, i.e. weak invariance properties of subsets of probability measures for mean-field type controlled evolutions, with a dynamics taking into account possible interactions among the agents.

The study of viscosity solutions for general Hamilton-Jacobi equations in the space of measures (see e.g. [12, 25, 26]) deeply involves the definition of suitable notions of sub/super-differentials (see [2, 11]). Recently, in [27], a Mayer problem without interactions was studied. For such a problem, under strong regularity assumptions, the authors provided a full characterization of the value function as the unique viscosity solution of a suitable HJB equation in the class of Lipschitz continuous functions. The comparison principle used to establish the uniqueness of the viscosity solution has been extended in the forthcoming [28], where much milder assumptions are required.

The characterization of the value function as the unique solution of a PDE in Wasserstein spaces is currently one of the most challenging issues in control problems for time-evolving measures. From the point of view of the necessary conditions, in [6] the authors provide a Pontryagin Maximum principle for a problem with interactions encoded in the dynamics by means of a nonlocal vector field, under strong regularity on the control function. Finally, in [23] the authors investigate $\Gamma$-convergence results for a controlled nonlocal dynamics, proving thus the consistency of the mean field model in a measure-theoretic setting with the corresponding one in finite dimension.

In this paper we study an optimal control problem for a multi-particle system subject to the influence of an external controller, who aims to minimize a general cost function satisfying some basic properties. The evolution of the system starting from an initial configuration $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$ is described by an absolutely continuous curve $t \mapsto \mu_t$ in the space of probability measures, and, recalling that the total mass is preserved along the evolution, the macroscopic dynamics will be expressed by the continuity equation

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0, \\ \mu_{|t=0} = \mu_0. \end{cases}$$

The link between the macroscopic dynamics and the controller's influence on each individual at the microscopic level is expressed by the constraint on the Borel vector field $v$, requiring that

$$v_t(x) \in F(x) \text{ for a.e. } t \in [0, T] \text{ and } \mu_t\text{-a.e. } x \in \mathbb{R}^d,$$

where the set-valued map

$$F(x) := \{ f(x, u) := f_0(x) + A(x)\, u \,:\, u \in U \}$$

represents the admissible velocities for the particles that transit through the point $x \in \mathbb{R}^d$, $U \subset \mathbb{R}^m$ is a given convex and compact *control set*, with $0 \in U$, and $A(x)$ is a $d \times m$ matrix whose columns are the evaluations at $x \in \mathbb{R}^d$ of continuous vector fields $f_i : \mathbb{R}^d \to \mathbb{R}^d$, $i = 1, \dots m$, with uniform linear growth.

The starting point of the problem we discuss is similar to the *sparse control strategy* for multi-agent systems addressed in [10, 24], where the authors study the evolution of the entire population in the case of a control action concentrated only on a small portion of the total number of agents.

In this paper, we consider a similar interaction constraint but, instead of dealing with a nonlocal dynamics, we consider it as part of the cost functional to be minimized. More precisely, we define the *control magnitude density* $\Psi : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty]$ by

$$\Psi(x, v) := \begin{cases} \inf\{|u| : u \in U,\ f(x, u) = v\}, & \text{if } v \in F(x), \\ \\ +\infty, & \text{otherwise}, \end{cases}$$

which, under suitable assumptions, turns out to be continuous on $\operatorname{Graph} F$. This map gives the minimum norm for a control generating an admissible velocity $v$ at $x$. In particular, the integral

$$\int_{\mathbb{R}^d} \Psi(x, v_t(x))\, d\mu_t,$$

can be considered as a measure of the *effort* of the controller, combining the magnitude of the control used to implement the vector field $v_t(\cdot)$ and the quantity $\mu_t$ of particles to be controlled at time $t$. The above integral can thus be naturally used to impose *control sparsity constraints*, in the form of upper bounds on the controller's effort to drive the mass of particles.

To this aim, the constraint we will consider in this paper will be upper bounds on the $L^\infty$ norm or on the $L^1$ norm of the map

$$t \mapsto \int_{\mathbb{R}^d} \Psi(x, v_t(x))\, d\mu_t.$$

In the case of $L^\infty$ bound, at every instant of time the controller must choose how to use the maximum amount of effort available to control the population: we notice that a weak action distributed on a large number of particles, or a strong action on a few number of individuals may require the same effort. In the case of $L^1$ bound, the effects are cumulated in time, thus all the past history of the evolution must be taken into account.

We stress out that the two kinds of interaction considered are at the *macroscopic* level, not at the microscopic one. Indeed, they do not involve directly the behavior of each individual w.r.t. the others (like in the case of attraction/repulsion potentials). Instead, we are interested in the particles' *collective evolution* and on the *effort* required to generate it.

Considering both our cases of interest, the main goals of this paper are the following:

- to study compactness results for the set of feasible trajectories;
- to prove the existence of optimal trajectories for general cost functions;
- to provide necessary conditions in the form of an HJB equation solved by the value function in a suitable viscosity sense.

In order to treat in a unified way the two cases, in Section 4, we provide an abstract framework for optimal control problems proving a general Dynamic Programming Principle. Further possible applications of this framework are discussed in Appendix A.

The paper is organized as follows: in Section 2 we recall some basic notion about Wasserstein spaces and fix the notation; in Section 3 we outline the problem providing some preliminary results; in Section 4 we prove a general Dynamic Programming Principle; in Section 5 we analyze the two control-sparsity problems ($L^\infty$-time averaged sparsity and $L^1$-time averaged sparsity cases) proving existence of optimal trajectories for general cost

functions. In Section 6, we study for each of the considered cases an Hamilton-Jacobi-Bellman equation solved by the value function in some suitable viscosity sense. Finally, in Appendix A we discuss other applications of the framework outlined in Section 4, while in Appendix B we recall some estimates borrowed from [13, 16].

## 2. Preliminaries and notation

As main references concerning optimal transport and measure theory, the reader may refer to [2, 30, 31].

We will use the following notation.

| | |
|---|---|
| $B(x,r)$ | the open ball of radius $r$ of a normed space $X$, i.e., $B(x,r) := \{y \in X : \|y-x\|_X < r\}$; |
| $\overline{K}$ | the closure of a subset $K$ of a topological space $X$; |
| $\mathrm{Id}_X(\cdot)$ | the identity function of $X$, i.e. $\mathrm{Id}_X(x) = x$ for all $x \in X$; |
| $I_K(\cdot)$ | the indicator function of $K$, i.e. $I_K(x) = 0$ if $x \in K$, $I_K(x) = +\infty$ if $x \notin K$; |
| $\chi_K(\cdot)$ | the characteristic function of $K$, i.e. $\chi_K(x) = 1$ if $x \in K$, $\chi_K(x) = 0$ if $x \notin K$; |
| $C^0_b(X;Y)$ | the set of continuous bounded function from a Banach space $X$ to $Y$, endowed with $\|f\|_\infty = \sup_{x\in X} \|f(x)\|$ (if $Y = \mathbb{R}$, $Y$ will be omitted); |
| $C^0_c(X;Y)$ | the set of compactly supported functions of $C^0_b(X;Y)$, with the topology induced by $C^0_b(X;Y)$; |
| $\Gamma_I$ | the set of continuous curves from a real interval $I$ to $\mathbb{R}^d$; |
| $\Gamma_T$ | the set of continuous curves from $[0,T]$ to $\mathbb{R}^d$; |
| $e_t$ | the evaluation operator $e_t : \mathbb{R}^d \times \Gamma_I$ defined by $e_t(x,\gamma) = \gamma(t)$ for all $t \in I$; |
| $\mathscr{P}(X)$ | the set of Borel probability measures on a separable metric space $X$, endowed with the weak* topology induced by $C^0_b(X)$; |
| $\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)$ | the set of vector-valued Borel measures on $\mathbb{R}^d$ with values in $\mathbb{R}^d$, endowed with the weak* topology induced by $C^0_c(\mathbb{R}^d;\mathbb{R}^d)$; |
| $\operatorname{supp}\mu$ | the support of the measure $\mu$; |
| $\|\nu\|$ | the total variation of a measure $\nu \in \mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)$; |
| $\ll$ | the absolutely continuity relation between measures defined on the same $\sigma$-algebra; |
| $\mathrm{m}_p(\mu)$ | the $p$-th moment of a probability measure $\mu \in \mathscr{P}(X)$, i.e., $\mathrm{m}_p(\mu) = \|\mathrm{Id}_X\|^p_{L^p_\mu}$; |
| $r\sharp\mu$ | the push-forward of the measure $\mu \in \mathscr{P}(X)$ by the Borel map $r$ (see Definition 2.1); |
| $\mu \otimes \eta_x$ | the product measure of $\mu \in \mathscr{P}(X)$ with the Borel family of measures $\{\eta_x\}_{x\in X}$ (see Section 5.3 in [2]); |
| $\pi_i$ | the $i$-th projection map $\pi_i(x_1,\dots,x_N) = x_i$; |
| $\pi_{ij}$ | the $i,j$-th projection map $\pi_{ij}(x_1,\dots,x_N) = (x_i,x_j)$; |
| $\mathscr{P}_p(X)$ | the subset of the elements $\mathscr{P}(X)$ with finite $p$-moment, |
| $W_p(\mu,\nu)$ | the $p$-Wasserstein distance between $\mu$ and $\nu$ (see Definition 2.2); |
| $\Pi(\mu,\nu)$ | the set of admissible transport plans from $\mu$ to $\nu$ (see Definition 2.2); endowed with the $p$-Wasserstein distance; |
| $\mathscr{L}^d$ | the Lebesgue measure on $\mathbb{R}^d$; |
| $\frac{\nu}{\mu}$ | the Radon-Nikodym derivative of the measure $\nu$ w.r.t. the measure $\mu$; |
| $\mathrm{Lip}(f)$ | the Lipschitz constant of a function $f$. |

**Definition 2.1** (Pushforward measure)**.** Given two separable metric spaces $X, Y$, $\mu \in \mathscr{P}(X)$, and a Borel map $r : X \to Y$, we define the *push forward measure* $r\sharp\mu \in \mathscr{P}(Y)$ by

$r\sharp\mu(B) := \mu(r^{-1}(B))$ for all Borel sets $B \subseteq Y$, or equivalently,

$$\int_X f(r(x))\, d\mu(x) = \int_Y f(y)\, dr\sharp\mu(y),$$

for every bounded (or $r\sharp\mu$-integrable) Borel function $f : Y \to \mathbb{R}$.

We refer to Chapter 5, Section 2 of [2] for the main properties of the pushforward measures.

The Wasserstein distance and its basic properties are recalled below.

**Definition 2.2** (Wasserstein distance)**.** Given $\mu_1, \mu_2 \in \mathscr{P}(\mathbb{R}^d)$, $p \geq 1$, we define the *p-Wasserstein distance* between $\mu_1$ and $\mu_2$ by setting

$$W_p(\mu_1, \mu_2) := \left( \inf \left\{ \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^p \, d\pi(x_1, x_2) \, : \, \pi \in \Pi(\mu_1, \mu_2) \right\} \right)^{1/p}, \tag{2.1}$$

where the set of *admissible transport plans* $\Pi(\mu_1, \mu_2)$ is defined by

$$\Pi(\mu_1, \mu_2) := \left\{ \pi \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d) : \begin{array}{l} \pi(A_1 \times \mathbb{R}^d) = \mu_1(A_1), \\ \pi(\mathbb{R}^d \times A_2) = \mu_2(A_2), \end{array} \right.$$
$$\left. \text{for all } \mu_i\text{-measurable sets } A_i,\ i = 1, 2 \right\}.$$

**Proposition 2.3.** *$\mathscr{P}_p(\mathbb{R}^d)$ endowed with the p-Wasserstein metric $W_p(\cdot, \cdot)$ is a complete separable metric space. Moreover, given a sequence $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathscr{P}_p(\mathbb{R}^d)$ and $\mu \in \mathscr{P}_p(\mathbb{R}^d)$, we have that the following are equivalent*

(1) $\lim_{n\to\infty} W_p(\mu_n, \mu) = 0$,
(2) $\mu_n \rightharpoonup^* \mu$ *and* $\{\mu_n\}_{n\in\mathbb{N}}$ *has uniformly integrable p-moments.*

*Proof.* See Proposition 7.1.5 in [2]. □

Concerning disintegration results for measures, widely used in this paper, we refer the reader to Section 5.3 in [2]. The following result is Theorem 5.3.1 in [2].

**Theorem 2.4** (Disintegration)**.** *Given a measure $\mu \in \mathscr{P}(\mathbb{X})$ and a Borel map $r : \mathbb{X} \to X$, there exists a family of probability measures $\{\mu_x\}_{x\in X} \subseteq \mathscr{P}(\mathbb{X})$, uniquely defined for $r\sharp\mu$-a.e. $x \in X$, such that $\mu_x(\mathbb{X} \setminus r^{-1}(x)) = 0$ for $r\sharp\mu$-a.e. $x \in X$, and for any Borel map $\varphi : X \times Y \to [0, +\infty]$ we have*

$$\int_{\mathbb{X}} \varphi(z)\, d\mu(z) = \int_X \left[ \int_{r^{-1}(x)} \varphi(z)\, d\mu_x(z) \right] d(r\sharp\mu)(x).$$

*We will write $\mu = (r\sharp\mu) \otimes \mu_x$. If $\mathbb{X} = X \times Y$ and $r^{-1}(x) \subseteq \{x\} \times Y$ for all $x \in X$, we can identify each measure $\mu_x \in \mathscr{P}(X \times Y)$ with a measure on $Y$.*

## 3. Setting of the problem and preliminary results

We study an optimal control problem in the space of probability measures with a *control sparsity* constraint. We develop separately two specific constraints: an $L^\infty$-time averaged sparsity condition (Section 5.1) and an $L^1$-time averaged sparsity constraint (Section 5.2). In both cases, the sparsity-interaction term is encoded in the cost functional to be minimized. In the first case, at a.e. time instant we impose an upper bound on the magnitude of control to be used on the evolving mass. In the second case, this constraint is $L^1$ in time. Both these problems are strongly motivated by applications to multi-particle systems.

In this section we introduce and discuss some preliminary properties regarding the dynamics and the objects that will be used in Section 5 and also in Appendix A to describe some control sparsity constraints.

**Definition 3.1** (Standing assumptions)**.** Let $I \subseteq \mathbb{R}$ be a nonempty compact interval, $U \subseteq \mathbb{R}^m$ be a convex compact subset, with $0 \in U$, called the *control set*. Let $f_i \in C^0(\mathbb{R}^d; \mathbb{R}^d)$, $i = 0, \dots, m$ satisfying the following conditions

(1) *growth condition:*

$$C := \sup_{x \in \mathbb{R}^d} \left\{ \frac{1}{|x|+1} \sum_{i=0}^{m} |f_i(x)| \right\} < +\infty;$$

(2) *rank condition:* the $d \times m$ matrix $A(x) := (f_1(x), f_2(x), \dots f_m(x))$ has rank independent of $x \in \mathbb{R}^d$.

We define the set-valued map $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ by setting

$$F(x) := \{ f(x,u) := f_0(x) + A(x)\, u \ : \ u \in U \}.$$

This multifunction $F$ will govern the controlled dynamics in terms of a differential inclusion, as described in Definition 3.3. The *graph* of $F(\cdot)$ is the set

$$\operatorname{Graph} F := \{ (x,v) \in \mathbb{R}^d \times \mathbb{R}^d : \ v \in F(x) \}.$$

We define the *control magnitude density* $\Psi : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty]$ by

$$\Psi(x,v) = \begin{cases} \inf \{ |u| : \ u \in U, \ f(x,u) = v \}, & \text{if } v \in F(x), \\ +\infty, & \text{otherwise.} \end{cases}$$

**Lemma 3.2.** *In the framework of Definition 3.1, the following properties hold.*

(1) *The set-valued map $F(\cdot)$ is continuous w.r.t. the Hausdorff metric, and it has nonempty, compact and convex values at every $x \in \mathbb{R}^d$. In particular,* $\operatorname{Graph} F$ *is closed. Moreover, $F$ satisfies the linear growth condition, i.e., there exists a constant $D > 0$ such that $F(x) \subseteq \overline{B(0, D(|x|+1))}$ for every $x \in \mathbb{R}^d$.*

(2) *The map $\Psi : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty]$ satisfies the following properties*

   (*i*) *for all $v \in F(x)$, there exists $w \in \mathbb{R}^m$ such that*

$$\Psi(x,v) = |A(x)^+ (v - f_0(x)) + (I - A(x)^+ A(x))\, w|,$$

   *where $A(x)^+$ is the Moore-Penrose pseudo-inverse of $A(x)$, and $I$ is the $m \times m$ identity matrix. Moreover $\Psi(\cdot,\cdot)$ is continuous on its domain, i.e. on* $\operatorname{Graph} F$*;*

   (*ii*) *$\Psi(\cdot,\cdot)$ is l.s.c. in $\mathbb{R}^d \times \mathbb{R}^d$;*

   (*iii*) *$v \mapsto \Psi(x,v)$ is convex for any $x \in \mathbb{R}^d$.*

*Proof.* Item (1) follows directly from the definition, recalling the continuity of $f_i$, $i = 1, \dots, m$, the compactness of $U$ and the standing assumptions. We can take $D = C \cdot R_U$, where we define $R_U := \max\{|u| : u \in U\}$.

We prove (*i*). Let $x \in \mathbb{R}^d$ be fixed, $v \in F(x)$ and let us denote with $A(x)^+$ the Moore-Penrose pseudo-inverse of $A(x)$, which exists and is unique (see [29]). Then, by hypothesis, there exists a solution $u$ of $A(x)\, u = v - f_0(x)$ and by pseudo-inverse properties, we can characterize any such a solution by $u = A(x)^+ (v - f_0(x)) + (I - A(x)^+ A(x))\, w$, $w \in \mathbb{R}^m$.

Let us define the map $g : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^m$, $g(x,v,w) := A(x)^+ (v - f_0(x)) + (I - A(x)^+ A(x))\, w$. We have that $g$ is continuous, indeed by the rank condition in Definition 3.1 and pseudo-inverses properties, since $x \mapsto A(x)$ is continuous, so is $x \mapsto A(x)^+$.

Thus, by Proposition 1.4.14 in [3], we have that the multifunction $M : \mathbb{R}^d \times \mathbb{R}^d \rightrightarrows \mathbb{R}^m$, defined by $M(x,v) := \{ g(x,v,w) \ : \ w \in \overline{B_{R_U}(0)} \}$ is continuous. Hence, so is the set-valued function $Q : \mathbb{R}^d \times \mathbb{R}^d \rightrightarrows \mathbb{R}^m$, $Q(x,v) := M(x,v) \cap U$. By Corollary 9.3.3 in [3], we get the continuity of the minimal norm selection of $Q$, thus the continuity of $\Psi$ on its domain, i.e. on $\operatorname{Graph} F$, since

$$\Psi(x,v) = \min_{z \in Q(x,v)} |z|,$$

when $(x,v) \in \operatorname{Graph} F$.

Finally, $(ii)$ follows by the continuity of $\Psi$ on $\operatorname{dom}\Psi \equiv \operatorname{Graph} F$ and closedness of $\operatorname{Graph} F$.

We pass now to the proof of $(iii)$. Let $(x, v_j) \in \mathbb{R}^d \times \mathbb{R}^d$, $j = 1, 2$. If $(x, v_j) \notin \operatorname{Graph} F$, for some $j = 1, 2$, the convexity inequality is trivially satisfied, since $\Psi$ takes the value $+\infty$, so we assume that $(x, v_j) \in \operatorname{Graph} F$, $j = 1, 2$. We notice that, since for all $u_j \in Q(x, v_j)$, $j = 1, 2$, we have $v_j - f_0(x) = A(x)\, u_j$, then $\lambda u_1 + (1-\lambda)u_2 \in Q(x, \lambda v_1 + (1-\lambda)v_2)$ for all $\lambda \in [0, 1]$. Recalling the triangular inequality, we have for all $u_j \in Q(x, v_j)$, $j = 1, 2$,

$$\begin{aligned}\Psi(x, \lambda v_1 + (1-\lambda)v_2) =& \min\{|u| : u \in Q(x, \lambda v_1 + (1-\lambda)v_2)\}\\ \leq& |\lambda u_1 + (1-\lambda)u_2| \leq \lambda|u_1| + (1-\lambda)|u_2|,\end{aligned}$$

By taking the minimum on $u_j \in Q(x, v_j)$, $j = 1, 2$, we obtain that $\Psi(x, \cdot)$ is convex. □

Following the same line as in [13,14,16,17,27], we define the set of *admissible trajectories* as follows.

**Definition 3.3** (Admissible trajectories)**.** Let $I \subseteq \mathbb{R}$ be a compact and nonempty interval. In the setting of Definition 3.1, we say that $(\boldsymbol{\mu}, \boldsymbol{\nu})$ is an *admissible trajectory* defined on $I$, and we write $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_I$, if the time-depending Borel probability measure $\boldsymbol{\mu} = \{\mu_t\}_{t\in I} \subseteq \mathscr{P}(\mathbb{R}^d)$ and the time-depending Borel vector-valued measure $\boldsymbol{\nu} = \{\nu_t\}_{t\in I} \subseteq \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)$ satisfy the following properties

(A1) *continuity equation*: we have $\partial_t \mu_t + \operatorname{div} \nu_t = 0$ in the sense of distributions in $I \times \mathbb{R}^d$;

(A2) *velocity constraint*: $|\nu_t| \ll \mu_t$ for a.e. $t \in I$ and the Radon-Nikodym derivative satisfies $\dfrac{\nu_t}{\mu_t}(x) \in F(x)$, for $\mu_t$-a.e. $x \in \mathbb{R}^d$ and a.e. $t \in I$. Equivalently, we ask

$$\int_I \mathcal{J}_F(\mu_t, \nu_t)\, dt < +\infty,$$

where $\mathcal{J}_F : \mathscr{P}(\mathbb{R}^d) \times \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d) \to [0, +\infty]$ is defined by

$$\mathcal{J}_F(\mu, E) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} I_{F(x)}\left(\frac{E}{\mu}(x)\right) d\mu(x), & \text{if } |E| \ll \mu,\\ +\infty, & \text{otherwise.}\end{cases}$$

Given $\bar{\mu} \in \mathscr{P}(\mathbb{R}^d)$, we say that $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_I$ is an admissible trajectory *starting from* $\bar{\mu}$ if $\mu_{|t=\min I} = \bar{\mu}$, and we write $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_I(\bar{\mu})$. It can be proved (see [2]) that every $\boldsymbol{\mu}$ such that $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_I$ for some $\boldsymbol{\nu}$ admits a narrowly continuous representative. Thus from now on we will always refer to it.

*Remark* 3.4. Let $\bar{\mu} \in \mathscr{P}_p(\mathbb{R}^d)$ with $p \geq 1$, and $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_I(\bar{\mu})$. Then, due to the growth assumption on $F(\cdot)$, it is possible to bound the moments of $\mu_t$ in terms of the moments of $\mu_0$ (see Proposition B.1), and hence we have $\boldsymbol{\mu} \subseteq \mathscr{P}_p(\mathbb{R}^d)$.

Recalling the Superposition Principle (Theorem 8.2.1 in [2]) and its extension to differential inclusions provided in Theorem 1 in [20], we refer to the following.

**Definition 3.5** (Probabilistic representations)**.** Let $\boldsymbol{\mu} = \{\mu_t\}_{t\in I} \subseteq \mathscr{P}_p(\mathbb{R}^d)$ be an absolutely continuous trajectory and $\boldsymbol{\nu} = \{\nu_t\}_{t\in I}$ a family of Borel vector-valued measures such that $\partial_t \mu_t + \operatorname{div} \nu_t = 0$, $t \in I$. We say that a probability measure $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_I)$

(1) *represents the pair* $(\boldsymbol{\mu}, \boldsymbol{\nu})$, if $\boldsymbol{\eta}$ is concentrated on the couples $(x, \gamma) \in \mathbb{R}^d \times \Gamma_I$ where $\gamma$ satisfies $\dot{\gamma}(t) = \dfrac{\nu_t}{\mu_t}(\gamma(t))$, $\gamma(\min I) = x$, and $\mu_t = e_t \sharp \boldsymbol{\eta}$ for all $t \in I$;

(2) *represents* $\boldsymbol{\mu}$, if $\boldsymbol{\eta}$ is concentrated on the couples $(x, \gamma) \in \mathbb{R}^d \times \Gamma_I$ where $\gamma$ satisfies $\dot{\gamma}(t) \in F(\gamma(t))$, $\gamma(\min I) = x$, and $\mu_t = e_t \sharp \boldsymbol{\eta}$ for all $t \in I$.

We notice that in general if $(\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{A}_I$, then $\boldsymbol{\mu}$ can have more than a representation (see [20] for an example of this situation).

Recalling Theorem 8.2.1 in [2], and Theorem 1 in [20], we have that

**Lemma 3.6** (Equivalence)**.** *In the setting of Definition 3.1.*

(1) *Every $(\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{A}_I$ admits a representation $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_I)$ according to Definition 3.5*(1).

(2) *Every $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_I)$ concentrated on the couples $(x,\gamma) \in \mathbb{R}^d \times \Gamma_I$, where $\dot\gamma(t) \in F(\gamma(t))$ and $\gamma(0) = x$, represents an admissible trajectory $(\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{A}_I$, with $\boldsymbol{\nu} = \{\nu_t\}_{t\in I}$ with $\nu_t = v_t\mu_t$ and*

$$v_t(y) = \int_{e_t^{-1}(y)} \dot\gamma(t)\, d\eta_{t,y}(x,\gamma),$$

*for a.e. $t \in I$ and $\mu_t$-a.e. $x \in \mathbb{R}^d$, where the Borel family of measures $\{\eta_{t,y}\}_{y\in\mathbb{R}^d} \subseteq \mathscr{P}(\mathbb{R}^d \times \Gamma_I)$ is the disintegration of $\boldsymbol{\eta}$ w.r.t. the evaluation operator $e_t$.*

*Remark* 3.7. Lemma 3.6 allows us to consider equivalently an admissible trajectory defined as in Definition 3.3, or a probability measure $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ satisfying the property of Lemma 3.6 (2).

**Definition 3.8.** Given $\bar\mu \in \mathscr{P}(\mathbb{R}^d)$, we define

$$\mathcal{R}_I^{\mathcal{A}}(\bar\mu) := \left\{\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_I) \,:\, \exists (\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{A}_I(\bar\mu) \text{ s.t. } \boldsymbol{\eta} \text{ represents } (\boldsymbol{\mu},\boldsymbol{\nu})\right\}.$$

*Remark* 3.9. We stress that in general

$$\mathcal{R}_I^{\mathcal{A}}(\bar\mu) \subsetneq \left\{\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_I) \,:\, \exists (\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{A}_I(\bar\mu) \text{ s.t. } \boldsymbol{\eta} \text{ represents } \boldsymbol{\mu}\right\}.$$

Indeed, in the left-hand set we are requiring $\boldsymbol{\eta}$ to be a representation for the pair $(\boldsymbol{\mu},\boldsymbol{\nu})$ as in Definition 3.5(1), while in the right-hand side we are exploiting the various possibilities for the construction of a representation recalled in Definition 3.5(2). We borrow the following clarifying example from [20].

*Example* 3.10. In $\mathbb{R}^2$, let

- $\mathscr{A} = \{\gamma_{x,y}(\cdot)\}_{(x,y)\in\mathbb{R}^2} \subseteq AC([0,2])$ where $\gamma_{x,y}(t) = (x+t, y - t\,\mathrm{sgn}\, y)$ for any $(x,y) \in \mathbb{R}^2$, $t \in [0,2]$, with $\mathrm{sgn}(0) = 0$;
- $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$, $F(x,y) \equiv [-1,1] \times [-1,1]$ for all $(x,y) \in \mathbb{R}^2$;
- $\mu_0 = \dfrac{1}{2}\delta_0 \otimes \mathscr{L}^1_{|[-1,1]} \in \mathscr{P}(\mathbb{R}^2)$, $\boldsymbol{\eta} = \mu_0 \otimes \delta_{\gamma_{x,y}} \in \mathscr{P}(\mathbb{R}^2 \times \Gamma_2)$, $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,2]}$ with $\mu_t = e_t\sharp\boldsymbol{\eta}$;
- $Q$ be the open square of vertice $\{(0,0),(1,0),(1/2,\pm 1/2)\}$.

By construction we have that

- $F$ is in the form of Definition 3.1 (by taking for instance as $f_0$ the null function, $A(x)$ be the $2\times 2$ identity matrix and $\mathbb{R}^d \ni U = [-1,1]\times[-1,1]$) and $\dot\gamma(t) \in F(\gamma(t))$ for all $\gamma \in \mathscr{A}$ and $t \in ]0,2[$.
- $\boldsymbol{\mu}$ is an admissible trajectory and we denote with $\boldsymbol{\nu} = \{\nu_t\}_{t\in[0,2]}$ its driving family of Borel vector-valued measures.

Denoted by $v_t = \dfrac{\nu_t}{\mu_t}$ the mean vector field, this implies $v_t(x,y) = (1,0)$ for all $(x,y) \in Q \setminus (\mathbb{R} \times \{0\})$ and $t = x$. Now, consider the associated characteristics $\dot{\tilde\gamma}_y(t) = v_t(\tilde\gamma_y(t))$, $\tilde\gamma_y(0) = (0,y)$, $y \in [-1,1]$, and let us build $\tilde{\boldsymbol{\eta}} = \mu_0 \otimes \delta_{\tilde\gamma_y} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_{[0,2]})$.

We notice that, by construction, $\tilde{\boldsymbol{\eta}}$ represents $(\boldsymbol{\mu},\boldsymbol{\nu})$ (in particular it represents $\boldsymbol{\mu}$), while $\boldsymbol{\eta}$ represents $\boldsymbol{\mu}$ but not the pair $(\boldsymbol{\mu},\boldsymbol{\nu})$, since it is not constructed on the mean vector field $\dfrac{\nu_t}{\mu_t}$.

In Theorem 3 in [16] the authors give sufficient conditions providing compactness of the set $\mathcal{A}_I(\bar{\mu})$ w.r.t the uniform convergence of curves in $W_p$, with $\bar{\mu} \in \mathscr{P}_p(\mathbb{R}^d)$, $p \geq 1$. While Proposition 1 in [20] states the compactness of the set $\mathcal{R}_I^{\mathcal{A}}(\bar{\mu})$.

**Lemma 3.11** (Norm-minimal control density)**.** *Given $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_I$, there exists a Borel map $u : I \times \mathbb{R}^d \to U$, defined $\mu_t$-a.e. $x \in \mathbb{R}^d$ and a.e. $t \in I$, such that $\boldsymbol{\nu} = \{\nu_t = v_t \mu_t\}_{t \in I}$, with*

$$v_t(x) = f_0(x) + A(x)\, u(t,x), \quad \text{for } \mu_t\text{-a.e. } x \text{ and a.e. } t \in I,$$
$$\int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_t}{\mu_t}(x)\right) d\mu_t(x) = \int_{\mathbb{R}^d} |u(t,x)|\, d\mu_t(x), \quad \text{for a.e. } t \in I.$$

*We will call $u(t,x)$ the* norm-minimal control density *associated with the admissible trajectory $(\boldsymbol{\mu}, \boldsymbol{\nu})$.*

*Proof.* By assumption, there exists $v_t \in L^1_{\mu_t}$ such that $\boldsymbol{\nu} = \{\nu_t = v_t \mu_t\}_{t \in I}$, $v_t(x) \in F(x)$ for $\mu_t$-a.e. $x \in \mathbb{R}^d$ and a.e. $t \in I$. Then, by Lemma 3.2 for $\mu_t$-a.e. $x$ and a.e. $t$ there exists a unique minimum-norm solution $u(t,x) \in U$ for $v_t(x) = f_0(x) + A(x)\, u$. It is defined by $u(t,x) := A(x)^+ (v_t(x) - f_0(x)) + (I - A(x)^+ A(x))\, w$, for some $w \in \mathbb{R}^m$ and it satisfies $\Psi(x, v_t(x)) = |u(t,x)|$ for $\mu_t$-a.e. $x$ and a.e. $t$. By construction, $u : I \times \mathbb{R}^d \to U$ is a well-defined Borel map for $\mu_t$-a.e. $x$ and a.e. $t$. □

The following result allows us to prove the existence of an admissible trajectory with given (admissible) initial velocity and satisfying further properties which will be used later on in Section 6 to provide an HJB result in our framework.

**Lemma 3.12.** *Let $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, $T > 0$, $v_0 : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel map which is also a $L^2_{\mu_0}$-selection of $F(\cdot)$. Then for all $\beta > 0$, there exists an admissible pair $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_{[0,T]}(\mu_0)$ and a representation $\boldsymbol{\eta}$ for the pair $(\boldsymbol{\mu}, \boldsymbol{\nu})$ in the sense of Definition 3.5(1) satisfying*

(1) *for all $p \in L^2_{\mu_0}(\mathbb{R}^d)$*
$$\lim_{t \to 0^+} \int_{\mathbb{R}^d \times \Gamma_{[0,T]}} \langle p \circ e_0(x,\gamma), \frac{e_t(x,\gamma) - e_0(x,\gamma)}{t} \rangle\, d\boldsymbol{\eta}(x,\gamma) = \int_{\mathbb{R}^d} \langle p(x), v_0(x) \rangle\, d\mu_0(x);$$
(2) *$\nu_t \rightharpoonup^* v_0 \mu_0$ as $t \to 0^+$;*
(3) *for all $t \in [0,T]$ we have*
$$\int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_t}{\mu_t}(x)\right) d\mu_t(x) \leq \int_{\mathbb{R}^d} \Psi\left(x, v_0(x)\right) d\mu_0(x);$$
(4) *the following bound holds*
$$\int_0^T \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_t}{\mu_t}(x)\right) d\mu_t(x) dt \leq \beta.$$

*Proof.* Let $u_0 : \mathbb{R}^d \to U$ be a Borel map such that $v_0(x) = f(x, u_0(x))$ for all $x \in \mathbb{R}^d$ and $\Psi(x, v_0(x)) = |u_0(x)|$. Notice that such a map $u_0$ exists by the same argument used in the proof of Lemma 3.11. Define the map $G : \mathbb{R}^d \rightrightarrows \Gamma_T$

$$G(\bar{x}) = \left\{ \gamma \in AC([0,T]) : \gamma(t) = \bar{x} + \int_0^t f\left(\gamma(s), u_0(\bar{x}) e^{-s}\right) ds, \ \forall t \in [0,T] \right\}.$$

According to Theorem 8.2.9 p.315 in [3], to prove the measurability of this map it is sufficient to notice that the map $g : \mathbb{R}^d \times \Gamma_T \to \Gamma_T$ defined by

$$g(\bar{x}, \gamma)(t) := \gamma(t) - \bar{x} + \int_0^t f\left(\gamma(s), u_0(\bar{x}) e^{-s}\right) ds,$$

is a Carathéodory map, i.e., $x \mapsto g(x,\gamma)$ is Borel for every $\gamma \in \Gamma_T$ and $\gamma \mapsto g(x,\gamma)$ is continuous for every $x \in \mathbb{R}^d$.

By Theorem 8.1.3 p. 308 in [3], since $G(\cdot)$ is Borel, it admits a Borel selection $x \mapsto \gamma_x \in G(x)$. Define $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]}$ by setting $\mu_t = e_t\sharp\boldsymbol{\eta}$ where

$$\boldsymbol{\eta} = \mu_0 \otimes \delta_{\gamma_x} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T).$$

According to Theorem 1 in [20], we have that $(\boldsymbol{\mu}, \boldsymbol{\nu} = \{\nu_t\}_{t\in[0,T]}) \in \mathcal{A}_{[0,T]}$, where $\nu_t$ is defined by

$$\frac{\nu_t}{\mu_t}(y) =: v_t(y) = \int_{e_t^{-1}(y)} \dot{\gamma}(t)\, d\eta_{t,y}(x,\gamma),$$

for a.e. $t \in [0,T]$ and $\mu_t$-a.e. $y \in \mathbb{R}^d$, where we used the disintegration $\boldsymbol{\eta} = \mu_t \otimes \eta_{t,y}$.

We prove (1) following a similar procedure as for the proof of Proposition 2.5 in [27]. By Proposition B.1, we have that $e_0, \frac{e_t - e_0}{t} \in L^2_{\boldsymbol{\eta}}$ for all $t \in [0,T]$. Thus, for all $p \in L^2_{\mu_0}(\mathbb{R}^d)$, by the definition of $\boldsymbol{\eta}$, we have

$$\lim_{t\to 0^+} \int_{\mathbb{R}^d\times\Gamma_{[0,T]}} \langle p\circ e_0(x,\gamma), \frac{e_t(x,\gamma) - e_0(x,\gamma)}{t}\rangle\, d\boldsymbol{\eta}(x,\gamma) = \lim_{t\to 0^+} \int_{\mathbb{R}^d} \langle p(x), \frac{\gamma_x(t) - \gamma_x(0)}{t}\rangle\, d\mu_0(x).$$

By Dominated Convergence Theorem, we obtain

$$\begin{aligned}\lim_{t\to 0^+} \int_{\mathbb{R}^d} \langle p(x), \frac{\gamma_x(t) - \gamma_x(0)}{t}\rangle\, d\mu(x) &= \int_{\mathbb{R}^d} \langle p(x), \lim_{t\to 0^+} \frac{\gamma_x(t) - \gamma_x(0)}{t}\rangle\, d\mu_0(x)\\ &= \int_{\mathbb{R}^d} \langle p(x), v_0(x)\rangle\, d\mu_0(x),\end{aligned}$$

thanks to the uniform bound on $\dfrac{e_t - e_0}{t}$ in terms of the 2-moment of $\mu_0$ (see Proposition B.1).

We prove (2). By the definition of $\boldsymbol{\eta}$, we have

$$v_t(y) = \int_{e_t^{-1}(y)} f(y, u_0(\gamma(0))e^{-t})\, d\eta_{t,y}(x,\gamma).$$

For any $\varphi \in C^0_b(\mathbb{R}^d;\mathbb{R}^d)$, we then have

$$\int_{\mathbb{R}^d} \langle \varphi(y), v_t(y)\rangle\, d\mu_t(y) = \int_{\mathbb{R}^d\times\Gamma_T} \varphi(\gamma(t)) \cdot f(\gamma(t), u_0(\gamma(0))e^{-t})\, d\boldsymbol{\eta}(x,\gamma). \tag{3.1}$$

We observe that we can use the Dominated Convergence Theorem, indeed

$$\begin{aligned} f(\gamma(t), u_0(\gamma(0))e^{-t}) &= (1 + |\gamma(t)|)\frac{f(\gamma(t), u_0(\gamma(0))e^{-t})}{1 + |\gamma(t)|}\\ &\le C(1 + |\gamma(t)|) \cdot \max\{1, \operatorname{diam} U\},\end{aligned}$$

which is $\boldsymbol{\eta}$-integrable since we can estimate the 2-moment of $\mu_t$ in terms of the 2-moment of $\mu_0$ by Proposition B.1. Thus, by passing to the limit under the integral sign in (3.1), we obtain

$$\lim_{t\to 0^+} \int_{\mathbb{R}^d} \langle \varphi(y), v_t(y)\rangle\, d\mu_t(y) = \int_{\mathbb{R}^d} \langle \varphi(x), f(x, u_0(x))\rangle\, d\mu_0 = \int_{\mathbb{R}^d} \langle \varphi(x), v_0(x)\rangle\, d\mu_0(x).$$

We prove (3). Recalling the affine structure of $f$, we have

$$v_t(y) = f\left(y, \int_{e_t^{-1}(y)} u_0(\gamma(0))e^{-t}\, d\eta_{t,y}(x,\gamma)\right),$$

thus

$$\begin{aligned}\int_{\mathbb{R}^d} \Psi(y, v_t(y))\, d\mu_t(y) &\le \int_{\mathbb{R}^d} \int_{e_t^{-1}(y)} |u_0(\gamma(0))|e^{-t}\, d\eta_{t,y}(x,\gamma)\, d\mu_t(x)\\ &\le \int_{\mathbb{R}^d\times\Gamma_T} |u_0(\gamma(0))|\, d\boldsymbol{\eta}(x,\gamma) = \int_{\mathbb{R}^d} |u_0(x)|\, d\mu_0(x)\end{aligned}$$

$$= \int_{\mathbb{R}^d} \Psi(x, v_0(x))\, d\mu_0(x).$$

The last formula shows that the map

$$t \mapsto \int_{\mathbb{R}^d} \Psi(y, v_t(y))\, d\mu_t(y)$$

belongs to $L^1([0,T])$. In particular, for every $\beta > 0$ there exists $\tau > 0$ such that

$$\int_0^\tau \int_{\mathbb{R}^d} \Psi(y, v_t(y))\, d\mu_t(y)\, dt \leq \beta.$$

We then consider any solution $\tilde{\boldsymbol{\mu}} = \{\tilde{\mu}_t\}_{t\in[\tau,T]}$ of the equation

$$\begin{cases} \partial_t \tilde{\mu}_t + \operatorname{div}(f(x,0)\tilde{\mu}_t) = 0, \\ \tilde{\mu}_\tau = \mu_\tau. \end{cases}$$

By Lemma 4.4 in [22], the juxtaposition of $\boldsymbol{\mu}$ restricted to $[0,\tau]$ with $\tilde{\boldsymbol{\mu}}$, and the juxtaposition of the corresponding families of Borel vector-valued measures $\boldsymbol{\nu}$ restricted to $[0,\tau]$ with $\tilde{\boldsymbol{\nu}} = \{f(\cdot,0)\tilde{\mu}_t\}_{t\in[\tau,T]}$, yields an admissible trajectory satisfying (4). □

## 4. General dynamic programming principle

In this section we present an abstract Dynamic Programming Principle which holds in quite general frameworks: this will allow us to treat the optimal control problems proposed in Section 5 and in Appendix A in a unified way. The proposed structure establish a common framework to check the validity of a Dynamic Programming Principle for problems of different nature.

**Definition 4.1.** A *generalized control system* is a quadruplet $(X, \Sigma, c, c_f)$ where $X$, $\Sigma$ are nonempty sets, $c_f : X \to [0, +\infty]$, and $c : X \times X \times \Sigma \to [0, +\infty]$, is a map satisfying the following properties

**($C_1$)** for every $x, y, z \in X$, $\sigma_1, \sigma_2 \in \Sigma$, there exists $\sigma' \in \Sigma$ such that

$$c(x, z, \sigma') \leq c(x, y, \sigma_1) + c(y, z, \sigma_2).$$

**($C_2$)** for every $x, z \in X$, $\sigma \in \Sigma$, there exist $y' \in X$, $\sigma_1', \sigma_2' \in \Sigma$ such that

$$c(x, z, \sigma) \geq c(x, y', \sigma_1') + c(y', z, \sigma_2').$$

Given $x \in X$ we define the *reachable set from* $x$ by

$$\mathscr{R}_x := \left\{ y \in X : \inf_{\sigma\in\Sigma} c(x, y, \sigma) < +\infty \right\},$$

and if $y \in \mathscr{R}_x$ we say that $y$ can be reached from $x$. Notice that if $y \in \mathscr{R}_x$ and $z \in \mathscr{R}_y$, property **($C_1$)** implies that $z \in \mathscr{R}_x$, hence the position

$$R_\Sigma := \{(x, y) \in X \times X : y \in \mathscr{R}_x\},$$

defines a transitive relation $R_\Sigma$ on $X$.

If we define

$$X_\Sigma := \{x \in X : (x, x) \in R_\Sigma\},$$

we have that the restriction of $R_\Sigma$ on $X_\Sigma$ is a partial order on $\Sigma$. Equivalently, we have that $x \in X_\Sigma$ if and only if there exists $\sigma \in \Sigma$ such that $c(x, x, \sigma) < +\infty$, i.e., $x \in \mathscr{R}_x$.

Given $x \in X$ and $y \in \mathscr{R}_x$, we define the set of *admissible transitions* from $x$ to $y$ by

$$\mathscr{A}(x, y) := \{\sigma \in \Sigma : c(x, y, \sigma) < +\infty\}.$$

and if $\sigma \in \mathscr{A}(x, y)$, we call $c(x, y, \sigma)$ the *cost of the admissible transition* $\sigma$. We call $c_f(y)$ the *exit cost* at the state $y$.

We define the *value function* $V : X \to [0, +\infty]$ by setting

$$V(x) = \inf_{\substack{y \in X \\ \sigma \in \Sigma}} \{c(x, y, \sigma) + c_f(y)\},$$

and if $V(x) < +\infty$, we have $V(x) = \inf_{y \in \mathscr{R}_x} \inf_{\sigma \in \mathscr{A}(x,y)} \{c(x, y, \sigma) + c_f(y)\}$.

We prove a Dynamic Programming Principle for this general framework.

**Theorem 4.2** (Dynamic Programming Principle)**.** *For every $x \in X$ we have*

$$V(x) = \inf_{\substack{y \in X \\ \sigma \in \Sigma}} \{c(x, y, \sigma) + V(y)\}.$$

*Proof.* Set $W(x) = \inf_{\substack{y \in X \\ \sigma \in \Sigma}} \{c(x, y, \sigma) + V(y)\}$.

(1) We prove that $V(x) \geq W(x)$. If $V(x) = +\infty$ there is nothing to prove. So assume $V(x) < +\infty$. For all $\varepsilon > 0$ there exist $y_\varepsilon \in X$, $\sigma_\varepsilon \in \Sigma$ such that

$$V(x) + \varepsilon \geq c(x, y_\varepsilon, \sigma_\varepsilon) + c_f(y_\varepsilon).$$

According to $(\boldsymbol{C_2})$, there are $y'_\varepsilon \in X$, $\sigma'_{\varepsilon,1}, \sigma'_{\varepsilon,2} \in \Sigma$ such that

$$\begin{aligned} V(x) + \varepsilon \geq & c(x, y_\varepsilon, \sigma_\varepsilon) + c_f(y_\varepsilon) \geq c(x, y'_\varepsilon, \sigma'_{\varepsilon,1}) + c(y'_\varepsilon, y_\varepsilon, \sigma'_{\varepsilon,2}) + c_f(y_\varepsilon) \\ \geq & c(x, y'_\varepsilon, \sigma'_{\varepsilon,1}) + V(y'_\varepsilon) \geq \inf_{\substack{y \in X \\ \sigma \in \Sigma}} \{c(x, y, \sigma) + V(y)\} = W(x), \end{aligned}$$

and we conclude by letting $\varepsilon \to 0^+$.

(2) We prove that $V(x) \leq W(x)$. If $W(x) = +\infty$ there is nothing to prove. So assume $W(x) < +\infty$. For all $\varepsilon > 0$ there exist $y_\varepsilon, y'_\varepsilon \in X$, $\sigma_\varepsilon, \sigma'_\varepsilon \in \Sigma$ such that

$$W(x) + \varepsilon \geq c(x, y_\varepsilon, \sigma_\varepsilon) + V(y_\varepsilon) \geq c(x, y_\varepsilon, \sigma_\varepsilon) + c(y_\varepsilon, y'_\varepsilon, \sigma'_\varepsilon) + c_f(y'_\varepsilon) - \varepsilon.$$

According to $(\boldsymbol{C_1})$, there exists $\sigma''_\varepsilon \in \Sigma$ such that

$$\begin{aligned} W(x) + \varepsilon \geq & c(x, y'_\varepsilon, \sigma''_\varepsilon) + c_f(y'_\varepsilon) - \varepsilon \\ \geq & \inf_{\substack{y \in X \\ \sigma \in \Sigma}} \{c(x, y, \sigma) + c_f(y)\} - \varepsilon = V(x) - \varepsilon, \end{aligned}$$

and we conclude by letting $\varepsilon \to 0^+$.

□

**Definition 4.3** (Generalized admissible trajectory)**.** Let $(I, \leq_I)$ be a totally orderered set admitting a maximal element $b \in I$ and a minimal element $a \in I$ w.r.t. the order $\leq_I$. We endow $I$ with the order topology, and use the notation $I = [a, b]$. Given $x \in X$, a *generalized admissible trajectory* starting from $x$ defined on $I$ is a pair $(\gamma, \sigma)$ of maps $\gamma : I \to X$, $\sigma : I \to \Sigma$, satisfying

(1) $\gamma(a) = x$;
(2) $c(x, \gamma(t), \sigma(t)) < +\infty$ for all $t \in I$;
(3) for all $t_1, t_2 \in I$ with $t_1 \leq_I t_2$ there exists $\sigma_{t_1 \to t_2} \in \Sigma$ such that

$$c(x, \gamma(t_2), \sigma(t_2)) \geq c(x, \gamma(t_1), \sigma(t_1)) + c(\gamma(t_1), \gamma(t_2), \sigma_{t_1 \to t_2})$$

In particular, by taking $t = b$ in (2), we must have $\gamma(b) \in \mathscr{R}_x$. Moreover, from (3) we deduce that $c(\gamma(t_1), \gamma(t_2), \sigma_{t_1 \to t_2}) < +\infty$, so $\gamma(t_2) \in \mathscr{R}_{\gamma(t_1)}$ for all $t_1, t_2 \in I$ with $a \leq_I t_1 \leq_I t_2 \leq_I b$.

*Remark* 4.4. We notice that if $(\gamma, \sigma)$ is a generalized admissible trajectory defined on $I$, and $\sigma_{a\to a} \in \Sigma$ satisfies Definition 4.3 (3) with $t_1 = t_2 = a$, we can define $\sigma'(t) = \sigma(t)$ for $t \neq a$ and $\sigma'(a) = \sigma_{a\to a}$. In this case, we have that $(\gamma, \sigma')$ is still a generalized admissible trajectory, and, from Definition 4.3 (3), recalling that $\sigma'(a) = \sigma_{a\to a}$, we have $c(x, x, \sigma'(a)) = 0$. Thus, without loss of generality, given a generalized admissible trajectory $(\gamma, \sigma)$ defined on $I = [a, b]$ we always assume that $c(x, x, \sigma(a)) = 0$.

**Definition 4.5** (Optimal transitions and optimal trajectories)**.** Given $x, y \in X$, $\sigma \in \Sigma$, we say that $\sigma$ is an *optimal transition* from $x$ to $y$ if

$$V(x) = c(x, y, \sigma) + V(y).$$

A generalized admissible trajectory $(\gamma, \sigma)$ defined on $I = [a, b]$ is called an *optimal trajectory* if for all $t \in I$ we have that $\sigma(t)$ is an optimal transition from $\gamma(a)$ to $\gamma(t)$, i.e.,

$$V(\gamma(a)) = c(\gamma(a), \gamma(t), \sigma(t)) + V(\gamma(t)), \text{ for all } t \in I.$$

**Corollary 4.6** (DPP for generalized admissible trajectories)**.** *Let $x \in X$ and $(\gamma, \sigma)$ be a generalized admissible trajectory starting from $x$ defined on the totally ordered set $I$. Then the map $h : I \to [0, +\infty]$ defined as*

$$h(t) := c(x, \gamma(t), \sigma(t)) + V(\gamma(t)),$$

*is monotone increasing, and it is constant if and only if the trajectory is optimal. Moreover, if the trajectory is optimal, for all $t, s \in I$ with $t \leq s$ we have that any $\sigma_{t\to s} \in \Sigma$ satisfying* (3) *in Definition 4.3 is an optimal transition from $\gamma(t)$ to $\gamma(s)$.*

*Proof.* Let $(\gamma, \sigma)$ be a generalized admissible trajectory, we prove that $h(\cdot)$ is monotone increasing: indeed, recalling Theorem 4.2,

$$V(z_1) \leq c(z_1, z_2, \sigma) + V(z_2), \text{ for all } \sigma \in \Sigma,\ z_1, z_2 \in X,$$

thus, choosing $z_1 = \gamma(t)$, $z_2 = \gamma(s)$, $\sigma = \sigma_{t\to s}$, from Definition 4.3 (3), we have

$$V(\gamma(t)) - V(\gamma(s)) \leq c(\gamma(t), \gamma(s), \sigma_{t\to s}) \leq c(x, \gamma(s), \sigma(s)) - c(x, \gamma(t), \sigma(t)), \tag{4.1}$$

hence

$$h(t) = V(\gamma(t)) + c(x, \gamma(t), \sigma(t)) \leq V(\gamma(s)) + c(x, \gamma(s), \sigma(s)) = h(s),$$

as desired.

(1) We prove that, if $h(\cdot)$ is constant, then the trajectory is optimal. If $h$ is constant, we have

$$h(t) = V(\gamma(t)) + c(x, \gamma(t), \sigma(t)) = V(\gamma(s)) + c(x, \gamma(s), \sigma(s)) = h(s),$$

and so

$$V(\gamma(t)) - V(\gamma(s)) = c(x, \gamma(s), \sigma(s)) - c(x, \gamma(t), \sigma(t)).$$

In particular, all the inequalities in (4.1) are fulfilled as equality, thus $\sigma_{t\to s}$ is an optimal transition between $\gamma(t)$ and $\gamma(s)$. By taking $t = a$, and recalling that we can always assume that $c(x, x, \sigma(a)) = 0$ (see Remark 4.4), we have that $(\gamma, \sigma)$ is optimal.

(2) We prove that, if the trajectory is optimal, then $h(\cdot)$ is constant. Since the trajectory is optimal, we have

$$V(x) = c(x, \gamma(t), \sigma(t)) + V(\gamma(t)), \qquad V(x) = c(x, \gamma(s), \sigma(s)) + V(\gamma(s)),$$

hence

$$V(\gamma(t)) - V(\gamma(s)) = c(x, \gamma(s), \sigma(s)) - c(x, \gamma(t), \sigma(t)),$$

thus $h(\cdot)$ is constant, and again all the inequalities in (4.1) are fulfilled as equality, so $\sigma_{t\to s}$ is an optimal transition between $\gamma(t)$ and $\gamma(s)$.

□

**Corollary 4.7.** *Let $(\gamma, \sigma)$ be a generalized admissible trajectory defined on the totally ordered set $I = [a, b]$.*

(1) *if $\inf_{\sigma \in \Sigma} c(\gamma(b), \gamma(b), \sigma) = 0$, we have $V(\gamma(b)) \leq c_f(\gamma(b))$.*

(2) *if $(\gamma, \sigma)$ is optimal and $V(\gamma(b)) = c_f(\gamma(b))$ then $\gamma(b) \in X$ and $\sigma(b) \in \Sigma$ realize the infimum in the definition of $V(\gamma(a))$, i.e.,*

$$V(\gamma(a)) = c(\gamma(a), \gamma(b), \sigma(b)) + c_f(\gamma(b)).$$

(3) *if $\inf_{\sigma \in \Sigma} c(\gamma(b), \gamma(b), \sigma) = 0$ and $\gamma(b) \in X$ and $\sigma(b) \in \Sigma$ realize the infimum in the definition of $V(\gamma(a))$, i.e.,*

$$V(\gamma(a)) = c(\gamma(a), \gamma(b), \sigma(b)) + c_f(\gamma(b)),$$

*then $(\gamma, \sigma)$ is optimal.*

*Proof.*

(1) By assumption, for all $\varepsilon \geq 0$ there exists $\sigma_\varepsilon \in \Sigma$ with

$$V(\gamma(b)) \leq c(\gamma(b), \gamma(b), \sigma_\varepsilon) + c_f(\gamma(b)) \leq \varepsilon + c_f(\gamma(b)),$$

and we conclude by letting $\varepsilon \to 0^+$ to obtain $V(\gamma(b)) \leq c_f(\gamma(b))$.

(2) Recalling Theorem 4.2, we have

$$V(\gamma(a)) = c(\gamma(a), \gamma(b), \sigma(b)) + V(\gamma(b)) = c(\gamma(a), \gamma(b), \sigma(b)) + c_f(\gamma(b)).$$

(3) Conversely, since $V(\gamma(b)) \leq c_f(\gamma(b))$ by item 1, we have

$$V(\gamma(a)) = c(\gamma(a), \gamma(b), \sigma(b)) + c_f(\gamma(b)) \geq c(\gamma(a), \gamma(b), \sigma(b)) + V(\gamma(b)).$$

but, according to the Theorem 4.2, the opposite inequality holds, and so

$$V(\gamma(a)) = c(\gamma(a), \gamma(b), \sigma(b)) + V(\gamma(b)).$$

Recalling that $c(\gamma(a), \gamma(a), \sigma(a)) = 0$, by Corollary 4.6 we obtain for all $t \in [a, b]$

$$\begin{aligned} c(\gamma(a), \gamma(a), \sigma(a)) + V(\gamma(a)) &\leq c(\gamma(a), \gamma(t), \sigma(t)) + V(\gamma(t)) \\ &\leq c(\gamma(a), \gamma(b), \sigma(b)) + V(\gamma(b)), \end{aligned}$$

and since the first and the last terms are equal, we conclude that for all $t \in [a, b]$ we have $V(\gamma(a)) = c(\gamma(a), \gamma(t), \sigma(t)) + V(\gamma(t))$, and so the trajectory is optimal by Corollary 4.6.

□

This completes the proof of the Dynamic Programming Principle.

## 5. Control sparsity problems

In this section we use the notation and the setting introduced in Section 3 to formulate and analyze two problems involving a control-sparsity constraint. For both of them we will implement the following strategy:

- we describe the *control sparsity* constraint that will be included in the cost functional to be minimized;
- we prove a compactness property of the set of *feasible* trajectories, i.e. admissible trajectories satisfying the control sparsity constraint;
- we use the results of Section 4 to prove a Dynamic Programming for the value function of the problem;
- we prove the existence of an optimal trajectory;
- we characterize the set of initial velocities for a feasible trajectory.

The last step is essential in order to provide necessary conditions in form of an Hamilton-Jacobi-Bellman equation satisfied by the value function (see Section 6).

5.1. **The $L^\infty$-time averaged feasibility case.**

Let $\alpha \geq 0$ be fixed, $p \geq 1$. Referring to the notation introduced in Section 4, we set

$$X = \mathscr{P}_p(\mathbb{R}^d), \quad \Sigma = \bigcup_{\substack{I \subseteq \mathbb{R} \\ I \text{ compact interval}}} \left[ AC(I; \mathscr{P}_p(\mathbb{R}^d)) \times \mathrm{Bor}(I; \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)) \right].$$

Observe that the set $\mathcal{A}_I$ of admissible trajectories starting by a measure in $\mathscr{P}_p(\mathbb{R}^d)$ (see Definition 3.3 and Proposition B.1) is a subset of $\Sigma$.

On the set $\Sigma_I := AC(I; \mathscr{P}_p(\mathbb{R}^d)) \times \mathrm{Bor}(I; \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d))$ we will consider the topology of sequentially a.e. pointwise $w^*$-convergence, i.e., given $\{\rho^n\}_{n\in\mathbb{N}} := \{(\boldsymbol{\mu}^n, \boldsymbol{\nu}^n)\}_n \subseteq \Sigma_I$, and $\rho := (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \Sigma_I$, we say that $\rho^n \rightharpoonup^* \rho$ if and only if $(\mu_t^n, \nu_t^n) \rightharpoonup^* (\mu_t, \nu_t)$ for a.e. $t \in I$.

**Definition 5.1** ($L^\infty$-time feasible trajectories)**.** Given $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_I$, we define the map $\theta_\rho : I \to [0, +\infty]$ by setting

$$\theta_\rho(s) := \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x),$$

where $\Psi$ is the *control magnitude density*. Given $\bar{\mu} \in \mathscr{P}_p(\mathbb{R}^d)$, we set

$$\mathcal{F}_I^\infty(\bar{\mu}) := \{\rho \in \mathcal{A}_I(\bar{\mu}) : \theta_\rho(s) \leq \alpha \text{ for a.e. } s \in I\} = \{\rho \in \mathcal{A}_I(\bar{\mu}) : \|\theta_\rho\|_{L^\infty(I)} \leq \alpha\}$$

$$\mathcal{R}_I^\infty(\bar{\mu}) := \left\{ \boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_I) : \exists (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{F}_I^\infty(\bar{\mu}) \text{ s.t. } \boldsymbol{\eta} \text{ represents } (\boldsymbol{\mu}, \boldsymbol{\nu}) \right\},$$

and we define the set of *$\alpha$-feasible trajectories* defined on $I$ by

$$\mathcal{F}_I^\infty := \bigcup_{\bar{\mu} \in \mathscr{P}_p(\mathbb{R}^d)} \mathcal{F}_I^\infty(\bar{\mu}) \subseteq \Sigma_I.$$

Finally, notice that

$$\|\theta_\rho\|_{L^\infty(I)} \leq \alpha \text{ if and only if } \int_I \mathcal{E}(\mu_t, \nu_t)\, dt < +\infty,$$

where $\mathcal{E} : \mathscr{P}(\mathbb{R}^d) \times \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d) \to [0, +\infty]$ is defined by

$$\mathcal{E}(\mu, E) := \begin{cases} I_{[0,\alpha]}\left( \displaystyle\int_{\mathbb{R}^d} \Psi\left(x, \frac{E}{\mu}(x)\right) d\mu(x) \right), & \text{if } |E| \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases}$$

*Remark* 5.2. The quantity $\theta_\rho(s)$ represents the total magnitude of control acting on the mass at time $s$. Thus, the feasibility constraint imposes a restriction on the amount of control to be used w.r.t. the portion of controlled mass: in particular, at every instant of time the controller must decide if it is more convenient to control all the mass with a reduced amount of control, or to act on few individuals with a greater amount of control (*control sparsity*). In some cases (depending on the cost functional) the two strategies are actually equivalent. We refer to the surveys [10, 24] for some applications of a sparse control strategy in the framework of multi-agent systems.

The main topological properties of the set of feasible trajectories are summarized as follows, and are natural extensions of the same properties proved for the admissibility set $\mathcal{A}_I$, respectively in Proposition 3 and Theorem 3 in [16] and in Proposition 1 in [20].

**Proposition 5.3.** *Let $I \subseteq \mathbb{R}$ be a compact nonempty interval, $p \geq 1$, $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, $C_1 \geq 0$. Then*

1. *$\mathcal{F}_I^\infty$ is closed w.r.t. the topology of $\Sigma_I$;*
2. *for any $\mathscr{B} \subseteq \mathcal{F}_I^\infty$, $C_1 > 0$ such that for all $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathscr{B}$ with $\boldsymbol{\mu} = \{\mu_t\}_{t\in I}$ it holds $\mathrm{m}_p(\mu_0) \leq C_1$, we have that the closure of $\mathscr{B}$ in $\Sigma_I$ is contained in $\mathcal{F}_I^\infty$;*
3. *$\mathcal{F}_I^\infty(\mu_0)$ is compact in the topology of $\Sigma_I$*

(4) $\mathcal{R}_I^\infty(\mu_0)$ *is compact in the narrow topology.*

*Proof.* The proof is essentially based on the variational characterization of the feasibility constraint.

*Step 1:* The functional $\mathscr{F} : \mathscr{P}(\mathbb{R}^d) \times \mathscr{M}(\mathbb{R}^d;\mathbb{R}^d) \to [0,+\infty]$ defined by

$$\mathscr{F}(\mu, E) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} \Psi\left(x, \frac{E}{\mu}(x)\right) d\mu(x), & \text{if } |E| \ll \mu, \\ +\infty, & \text{otherwise} \end{cases} \tag{5.1}$$

is l.s.c. w.r.t. w*-convergence.

*Proof of Step 1:* By Lemma 3.2, the function $\Psi : \mathbb{R}^d \times \mathbb{R}^d \to [0,+\infty]$ is l.s.c. and $\Psi(x,\cdot)$ is convex for any $x \in \mathbb{R}^d$ with bounded domain. So adopting the notation in [7], we have that $\Psi_\infty(x,v) = 0$ if $v = 0$ and $\Psi_\infty(x,v) = +\infty$ if $v \neq 0$, where $\Psi_\infty(x,\cdot)$ denotes the *recession function* for $\Psi(x,\cdot)$. By l.s.c. of $F$, there exists a continuous selection $z : \mathbb{R}^d \to \mathbb{R}^d$ of $F$ (Michael's Theorem). Thus, by continuity of $\Psi(\cdot,\cdot)$ in $\operatorname{Graph} F$ (see Lemma 3.2), we have that $x \mapsto \Psi(x, z(x))$ is continuous and finite. We conclude by Lemma 2.2.3, p. 39, Theorem 3.4.1, p.115, and Corollary 3.4.2 in [7] or Theorem 2.34 in [1]. ◇

*Step 2:* Let $\rho^n := (\boldsymbol{\mu}^n, \boldsymbol{\nu}^n) \in \mathcal{F}_I^\infty$ for all $n \in \mathbb{N}$, $\rho := (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \Sigma_I$ be such that $\rho^n$ converges to $\rho$ in $\Sigma_I$. Then $\rho \in \mathcal{F}_I^\infty$.

*Proof of Step 2:* By convexity and l.s.c. of the indicator function $I_{[0,\alpha]}(\cdot)$ and l.s.c. of $\mathscr{F}(\cdot,\cdot)$, we have that the functional $\mathcal{E}(\cdot,\cdot)$ is l.s.c w.r.t. w*-convergence. The l.s.c. of the functional $\mathcal{J}_F(\cdot,\cdot)$, defined in Definition 3.3($A2$), was already proved in Lemma 3 in [16]. By Proposition 3 in [16] we have $\rho \in \mathcal{A}_I$. Now, fix $t \in I$ such that $(\mu_t^n, \nu_t^n) \rightharpoonup^* (\mu_t, \nu_t)$ and $\mathcal{E}(\mu_t^n, \nu_t^n) + \mathcal{J}_F(\mu_t^n, \nu_t^n) = 0$ for all $n \in \mathbb{N}$. By l.s.c. of $\mathcal{E}$, $\mathcal{J}_F$ and the fact that $\mathcal{E}, \mathcal{J}_F \geq 0$, we have

$$0 \leq \mathcal{E}(\mu_t, \nu_t) + \mathcal{J}_F(\mu_t, \nu_t) \leq \liminf_{n\to+\infty} \mathcal{E}(\mu_t^n, \nu_t^n) + \mathcal{J}_F(\mu_t^n, \nu_t^n) = 0.$$

By applying Fatou's lemma we deduce that

$$\int_I (\mathcal{E}(\mu_t, \nu_t) + \mathcal{J}_F(\mu_t, \nu_t))\, dt = 0,$$

hence $\rho \in \mathcal{F}_I^\infty$. ◇

All the assertion now follows recalling that uniformly boundedness of the moments along a sequence implies existence of a narrowly convergent subsequence (see e.g. Chapter 5 in [2]). □

We pass now to the analysis of the value function and of the Dynamic Programming Principle in this setting. To this aim, we will refer to the abstract results proved in Section 4.

**Definition 5.4** (Concatenation and restriction)**.**

(1) Let $I_i = [a_i, b_i] \subset \mathbb{R}$, $i = 1,2$, with $b_1 = a_2$, and $I := I_1 \cup I_2$. Let $\rho^i = (\boldsymbol{\mu}^i, \boldsymbol{\nu}^i) \in \mathcal{F}_{I_i}^\infty$ with $\mu_{b_1}^1 = \mu_{a_2}^2$. The *concatenation* $\rho^1 \star \rho^2 = (\boldsymbol{\mu}, \boldsymbol{\nu})$ of $\rho^1$ and $\rho^2$ is defined by setting $\mu_t = \mu_t^i$ and $\nu_t = \nu_t^i$ when $t \in I_i$ for $i = 1,2$, and noticing that we can assume $\nu_{b_1}^1 = \nu_{a_2}^2$ by changing the driving vector field in a $\mathscr{L}^1$-negligible set. We recall that this implies that $\rho^1 \star \rho^2 \in \mathcal{F}_I^\infty$. Indeed, by Lemma 4.4 in [22] we have that the set of solutions of the continuity equation is closed w.r.t. juxtaposition operations. The admissibility property of the resulting trajectory follows straightforwardly, as observed also in Theorem 6 in [16] and so does the feasibility one.

(2) Let $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{F}_I^\infty$. The *restriction* $\rho_{|I_1}$ of $\rho$ to a compact and nonempty interval $I_1 \subset I$, where $\rho_{|I_1} = (\boldsymbol{\mu}^1 = \{\mu_t^1\}_{t\in I_1}, \boldsymbol{\nu}^1 = \{\nu_t^1\}_{t\in I_1})$ is defined by setting $\mu_t^1 := \mu_t$ and $\nu_t^1 := \nu_t$ for all $t \in I_1$. Clearly we have $\rho_{|I_1} \in \mathcal{F}_{I_1}^\infty$.

*Remark* 5.5. We can always extend an $\alpha$-feasible trajectory $(\boldsymbol{\mu}^{(1)}, \boldsymbol{\nu}^{(1)}) \in \mathcal{F}^{\infty}_{[a,b]}$ to an $\alpha$-feasible trajectory defined in the extended time-interval $[a, c]$, for any $0 \le a \le b \le c$. It is sufficient to take $\bar{v}_t(\cdot) = f(\cdot, 0)$ for all $t \in [b, c]$ and to consider the solution $\boldsymbol{\mu}^{(2)} = \{\mu^{(2)}_t\}_{t\in[b,c]}$ of the continuity equation $\partial_t \mu_t + \operatorname{div}(\nu^{(2)}_t) = 0$ for $t \in [b, c]$, with $\mu_{|t=b} = \mu^{(1)}_{|t=b}$, $\nu^{(2)}_t = \bar{v}_t \mu_t$. We have that $(\boldsymbol{\mu}, \boldsymbol{\nu}) := (\boldsymbol{\mu}^{(1)}, \boldsymbol{\nu}^{(1)}) \star (\boldsymbol{\mu}^{(2)}, \boldsymbol{\nu}^{(2)})$ is an $\alpha$-feasible trajectory on $[a, c]$.

**Definition 5.6.** Let $c$ and $c_f$ be as in Definition 4.1, satisfying the following additional properties

- **($C_3$)** $c(\mu^{(1)}, \mu^{(2)}, \hat{\rho}) < +\infty$ if and only if $\hat{\rho} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}) \in \mathcal{F}^{\infty}_{I}(\mu^{(1)})$, with $\hat{\mu}_{|t=\max I} = \mu^{(2)}$ for some compact and nonempty interval $I \subset \mathbb{R}$;
- **($C_4$)** let $0 \le a \le b \le c$, $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{F}^{\infty}_{[a,c]}$. Then $c : X \times X \times \Sigma \to [0, +\infty]$ is superadditive by restrictions, i.e.

$$c(\mu_a, \mu_c, \rho) \ge c(\mu_a, \mu_b, \rho_{|[a,b]}) + c(\mu_b, \mu_c, \rho_{|[b,c]}).$$

Let $\bar{\mu} \in \mathscr{P}_p(\mathbb{R}^d)$, $I \subset \mathbb{R}$ nonempty and compact interval, and $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{F}^{\infty}_{I}(\bar{\mu})$. We define the set $\mathscr{G}^{\infty}_{\rho}$ made of the pairs $(\gamma, \sigma)$ defined as follows

1. $\gamma : I \to X$, $\gamma(t) := \mu_t$ for all $t \in I$;
2. $\sigma : I \to \Sigma$, $\sigma(t) := \rho_{|[\min I, t]}$ for all $t \in I$.

Finally, we define the set

$$\mathscr{G}^{\infty}_{I}(\bar{\mu}) := \left\{(\gamma, \sigma) \in \mathscr{G}^{\infty}_{\rho} \,:\, \rho \in \mathcal{F}^{\infty}_{I}(\bar{\mu})\right\}.$$

**Theorem 5.7** (DPP for $L^{\infty}$-time feasibility case)**.** *Let $V : \mathscr{P}_p(\mathbb{R}^d) \to [0, +\infty]$ be as in Definition 4.1. For any $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ we have*

$$V(\mu_0) = \inf_{\substack{\rho=(\boldsymbol{\mu},\boldsymbol{\nu})\in\mathcal{F}^{\infty}_{I}(\mu_0)\\ I\subseteq\mathbb{R} \text{ compact interval}}} \left\{c(\mu_0, \mu_{|t=\max I}, \rho) + V(\mu_{|t=\max I})\right\}.$$

*Proof.* The proof follows by Theorem 4.2 and **($C_3$)**. □

*Remark* 5.8. Any $(\gamma, \sigma) \in \mathscr{G}^{\infty}_{\rho}$, s.t. $\rho \in \mathcal{F}^{\infty}_{I}(\bar{\mu})$, is generalized admissible from $\bar{\mu}$, according to Definition 4.3. It is sufficient to take $\sigma_{t_1 \to t_2} := \rho_{|[t_1,t_2]}$ for any $0 \le t_1 \le t_2 \le T$, and observe that **($C_4$)** implies item (3) in Definition 4.3. Thus, Corollaries 4.6, 4.7 hold in this framework. Furthermore, since we can indentify any $\hat{\rho} \in \mathcal{F}^{\infty}_{[t,t]}$ with its restriction $\hat{\rho}_{|[t,t]}$, then by **($C_4$)** we have that $c(\gamma(t), \gamma(t), \sigma_{t\to t}) = 0$ for all $t \in I$.

**Proposition 5.9** (Existence of minimizers)**.** *Assume properties **($C_1$)** – **($C_4$)**. Let $p \ge 1$ and $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$. If $c(\mu_0, \cdot, \cdot)$ and $c_f$ are l.s.c. w.r.t. $w^*$-convergence and $V(\mu_0) < +\infty$, then there exist $I \subset \mathbb{R}$ nonempty and compact interval and an optimal trajectory $(\gamma, \sigma) \in \mathscr{G}^{\infty}_{I}(\mu_0)$, according to Definition 4.5.*

*Proof.* By finiteness of $V(\mu_0)$ and **($C_3$)**, we have that for all $n \in \mathbb{N}$, $n > 0$, there exist $(\mu^n, \rho^n) \in X \times \Sigma$ and $I^n \subset \mathbb{R}$ non empty compact interval such that $\rho^n = (\boldsymbol{\mu}^n, \boldsymbol{\nu}^n) \in \mathcal{F}^{\infty}_{I^n}(\mu_0)$ with $\mu^n_{|t=\max I^n} = \mu^n$ and

$$V(\mu_0) + \frac{1}{n} \ge c(\mu_0, \mu^n, \rho^n) + c_f(\mu^n).$$

Let $(\gamma^n, \sigma^n) \in \mathscr{G}^{\infty}_{\rho^n}$. Without loss of generality, we can assume $I^n = [0, T_n]$ for all $n \in \mathbb{N}$. Let $T = \liminf_{n\to+\infty} T_n$, then there exists a subsequence such that $T = \lim_{k\to+\infty} T_{n_k}$ and $T_{n_k} \ge T - \dfrac{1}{k}$ for all $k \ge 1$.

Let us consider the restrictions $\rho^{n_k}_{|[0,T-\frac{1}{k}]} \in \mathcal{F}^{\infty}_{[0,T-\frac{1}{k}]}(\mu_0)$ and any of their extensions in $[T - \frac{1}{k}, T]$ preserving the feasibility constraint (see Remark 5.5). Denote with $\hat{\rho}^{nk} := (\hat{\boldsymbol{\mu}}^{nk}, \hat{\boldsymbol{\nu}}^{nk}) \in \mathcal{F}^{\infty}_{[0,T]}(\mu_0)$ such an object. By compactness of $\mathcal{F}^{\infty}_{[0,T]}(\mu_0)$ proved in Proposition

5.3, $\hat{\rho}^{nk} \rightharpoonup^* \hat{\rho} := (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}) \in \mathcal{F}^{\infty}_{[0,T]}(\mu_0)$, i.e. $(\hat{\mu}^{nk}_t, \hat{\nu}^{nk}_t) \rightharpoonup^* (\hat{\mu}_t, \hat{\nu}_t)$ for a.e. $t \in [0,T]$. Furthermore, similarly to the proof of Theorem 3 in [16], by the standing assumption in Definition 3.1 and Remark 3.4, for any $k$ there exists a sequence $t_i \to T$ such that $\hat{\nu}^{nk}_{t_i} \rightharpoonup^* \hat{\nu}^{nk}_T$ for $i \to +\infty$, and $\hat{\mu}^{nk}_{t_i} \rightharpoonup^* \hat{\mu}^{nk}_T$ by absolute continuity of $\hat{\boldsymbol{\mu}}^{nk}$.

By a diagonal argument, we have that $\hat{\rho}^{nk}_{|[0,T-\frac{1}{k}]} \rightharpoonup^* \hat{\rho}$, up to subsequences. Property (3) in Definition 4.3 leads to

$$c(\mu_0, \mu^{n_k}, \rho^{n_k}) \geq c(\mu_0, \hat{\mu}^{nk}_{|t=T-\frac{1}{k}}, \hat{\rho}^{nk}_{|[0,T-\frac{1}{k}]}).$$

By passing to the limit up to subsequences we get

$$V(\mu_0) = c(\mu_0, \hat{\mu}, \hat{\rho}) + c_f(\hat{\mu}).$$

Thus, $(\gamma, \sigma) \in \mathscr{G}^{\infty}_{\hat{\rho}}$ is optimal by Corollary 4.7(3). □

In Section 7 we see a concrete example of cost and value functions satisfying $\boldsymbol{(C_1)}-\boldsymbol{(C_4)}$.

The last part of this subsection is devoted to the characterization of the set of initial velocities for feasible trajectories in $\mathscr{P}_2(\mathbb{R}^d)$. This is a fundamental ingredient to deal with the formulation of an Hamilton-Jacobi-Bellman equation for this problem, which is discussed in Section 6.1.

**Definition 5.10.** Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $0 \leq s \leq T$.

(1) Given $\boldsymbol{\eta} \in \mathcal{R}^{\infty}_{[s,T]}(\mu)$, we define

$$\mathscr{V}_{[s,T]}(\boldsymbol{\eta}) := \left\{ w_{\boldsymbol{\eta}} \in L^2_{\boldsymbol{\eta}}(\mathbb{R}^d \times \Gamma_{[s,T]}; \mathbb{R}^d) \ : \ \begin{array}{l} \exists \{t_i\}_{i\in\mathbb{N}} \subseteq ]s,T[, \text{ with } t_i \to s^+ \text{ and} \\ \dfrac{e_{t_i} - e_s}{t_i - s} \rightharpoonup w_{\boldsymbol{\eta}} \text{ weakly in } L^2_{\boldsymbol{\eta}}(\mathbb{R}^d \times \Gamma_{[s,T]}; \mathbb{R}^d) \end{array} \right\},$$

and $\mathscr{V}_{[s,T]}(\boldsymbol{\eta}) \neq \emptyset$ as already observed in Definition 11 in [16].

(2) We set

$$\mathscr{V}^{\infty}_{[s,T]}(\mu) := \left\{ x \mapsto \int_{\Gamma^x_{[s,T]}} w_{\boldsymbol{\eta}}(x,\gamma)\, d\eta_x \ : \ \boldsymbol{\eta} \in \mathcal{R}^{\infty}_{[s,T]}(\mu),\ w_{\boldsymbol{\eta}} \in \mathscr{V}_{[s,T]}(\boldsymbol{\eta}) \right\},$$

where we denoted with $\{\eta_x\}_{x\in\mathbb{R}^d}$ the disintegration of $\boldsymbol{\eta}$ w.r.t. the map $e_s$.

(3) We define the set

$$\mathscr{Z}^{\infty}(\mu) := \left\{ v \in L^2_{\mu}(\mathbb{R}^d; \mathbb{R}^d) \ : \ \begin{array}{l} v(x) \in F(x) \text{ for } \mu\text{-a.e. } x \in \mathbb{R}^d \\ \text{and } \displaystyle\int_{\mathbb{R}^d} \Psi(x, v(x))\, d\mu(x) \leq \alpha \end{array} \right\}.$$

**Lemma 5.11** (Initial velocity)**.** *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $0 \leq s \leq T$. Then $\mathscr{V}^{\infty}_{[s,T]}(\mu) \equiv \mathscr{Z}^{\infty}(\mu)$.*

*Proof.* From items (1) and (3) in Lemma 3.12, we deduce that $\mathscr{Z}^{\infty}(\mu) \subseteq \mathscr{V}^{\infty}_{[s,T]}(\mu)$.

Let us now prove the other inclusion, i.e. $\mathscr{V}^{\infty}_{[s,T]}(\mu) \subseteq \mathscr{Z}^{\infty}(\mu)$. Without loss of generality, let us consider $s = 0$. Let $\boldsymbol{\eta} \in \mathcal{R}^{\infty}_{[0,T]}(\mu)$. For any $(\boldsymbol{\mu}, \boldsymbol{\nu})$ represented by $\boldsymbol{\eta}$, we have in particular that $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_{[0,T]}$, then $x \mapsto \int_{\Gamma^x_{[0,T]}} w_{\boldsymbol{\eta}}(x,\gamma)\, d\eta_x$, with $w_{\boldsymbol{\eta}} \in \mathscr{V}_{[0,T]}(\boldsymbol{\eta})$, is an $L^2_{\mu}$-selection of $F$ by Lemma 5 in [16] and convexity of $F(x)$. Hence, it remains to prove that

$$\int_{\mathbb{R}^d} \Psi\left(x, \int_{\Gamma^x_{[0,T]}} w_{\boldsymbol{\eta}}(x,\gamma)\, d\eta_x\right) d\mu(x) \leq \alpha.$$

By feasibility of $(\boldsymbol{\mu}, \boldsymbol{\nu})$, for a.e. $t \in [0,T]$ we have

$$\int_{\mathbb{R}^d \times \Gamma_{[0,T]}} \Psi(\gamma(t), \dot{\gamma}(t))\, d\boldsymbol{\eta}(x,\gamma) = \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_t}{\mu_t}(x)\right) d\mu_t(x) \leq \alpha.$$

By hypothesis and Lemma 3.2, $\Psi(\cdot,\cdot)$ is uniformly continuous on Graph $F$ when the first variable ranges in a compact set $K \subset \mathbb{R}^d$.

Then, let us fix $n \in \mathbb{N}$ and consider the closed ball $\overline{B_n(0)} = \overline{B(0,n)} \subset \mathbb{R}^d$. We have

$$\begin{aligned}
\int_{\overline{B_n(0)}} \Psi\left(x, \int_{\Gamma^x_{[0,T]}} w_{\boldsymbol{\eta}}(x,\gamma)\, d\eta_x\right) d\mu(x) \leq \\
&\leq \liminf_{t\to 0^+} \int_{\overline{B_n(0)}} \Psi\left(x, \int_{\Gamma^x_{[0,T]}} \frac{e_t - e_0}{t}\, d\eta_x\right) d\mu(x) \\
&\leq \liminf_{t\to 0^+} \int_{\overline{B_n(0)}\times\Gamma_{[0,T]}} \Psi\left(e_0, \frac{e_t - e_0}{t}\right) d\boldsymbol{\eta}(x,\gamma) \\
&= \liminf_{t\to 0^+} \int_{\overline{B_n(0)}\times\Gamma_{[0,T]}} \Psi\left(\gamma(0), \frac{1}{t}\int_0^t \dot\gamma(r)\, dr\right) d\boldsymbol{\eta},
\end{aligned}$$

where we used l.s.c. of $\Psi$ in the second variable and Fatou's Lemma for the first inequality, and Jensen's inequality together with convexity of $\Psi$ in the second variable for the second inequality. Moreover,

$$\begin{aligned}
\liminf_{t\to 0^+} \int_{\overline{B_n(0)}\times\Gamma_{[0,T]}} \Psi\left(\gamma(0), \frac{1}{t}\int_0^t \dot\gamma(r)\, dr\right) d\boldsymbol{\eta} \leq \\
&\leq \liminf_{t\to 0^+} \frac{1}{t}\int_0^t \int_{\overline{B_n(0)}\times\Gamma_{[0,T]}} \Psi(\gamma(0), \dot\gamma(r))\, d\boldsymbol{\eta}\, dr \\
&\leq \liminf_{t\to 0^+} \frac{1}{t}\int_0^t \int_{\overline{B_n(0)}\times\Gamma_{[0,T]}} \Psi(\gamma(r), \dot\gamma(r))\, d\boldsymbol{\eta}\, dr + \varepsilon \\
&\leq \alpha + \varepsilon,
\end{aligned}$$

for any $\varepsilon$ small enough. Where we used again Jensen's inequality and convexity of $\Psi$ in the second variable for the first inequality. Finally we used uniform continuity of $\Psi(\cdot,\cdot)$ in the compact set $\overline{B_n(0)}$, together with uniform continuity of the set of all $\gamma \in \Gamma_{[0,T]}$ s.t. $(\gamma(0),\gamma) \in \operatorname{supp}\boldsymbol{\eta}$. Indeed, by the standing assumption in Definition 3.1 and compactness of $U$, for all $(x,\gamma) \in \operatorname{supp}\boldsymbol{\eta}$, we have

$$\begin{aligned}
|\gamma(t) - \gamma(0)| &\leq \int_0^t |\dot\gamma(s)|\, ds \leq C\int_0^t |\gamma(s)|\, ds + Ct \\
&\leq C\int_0^t |\gamma(s) - \gamma(0)|\, ds + Ct\,(1 + |\gamma(0)|),
\end{aligned}$$

and so, by Gronwall's inequality, and recalling that $\gamma(0) \in \overline{B(0,n)}$,

$$|\gamma(t) - \gamma(0)| \leq Ct\,(1 + |\gamma(0)|)e^{Ct} \leq CT(1+n)e^{CT}.$$

We conclude by letting $\varepsilon \to 0^+$ in the former estimate, and noticing that this holds for all $n \in \mathbb{N}$, thus by passing to the limit we have

$$\int_{\mathbb{R}^d} \Psi\left(x, \int_{\Gamma^x_{[0,T]}} w_{\boldsymbol{\eta}}(x,\gamma)\, d\eta_x\right) d\mu(x) = \sup_{n\in\mathbb{N}} \int_{\overline{B_n(0)}} \Psi\left(x, \int_{\Gamma^x_{[0,T]}} w_{\boldsymbol{\eta}}(x,\gamma)\, d\eta_x\right) d\mu(x) \leq \alpha.$$

□

5.2. **The $L^1$-time averaged feasibility case.**

Let $\alpha \geq 0$ be fixed, $p \geq 1$. Referring to the notation of Section 4, we take

$$X = \mathscr{P}_p(\mathbb{R}^d)\times[0,+\infty[, \quad \Sigma = \bigcup_{\substack{I\subseteq\mathbb{R} \\ I \text{ compact interval}}} \Big[AC(I;\mathscr{P}_p(\mathbb{R}^d)) \times \operatorname{Bor}(I;\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d))\Big]\times[0,+\infty[.$$

On the set $\Sigma_I := AC(I;\mathscr{P}_p(\mathbb{R}^d)) \times \mathrm{Bor}(I;\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)) \times [0,+\infty[$ we will consider the topology given by the sequentially a.e. $w^*$-convergence and the convergence in $\mathbb{R}$. More precisely, given $\{\rho^n\}_{n\in\mathbb{N}} := \{(\boldsymbol{\mu}^n,\boldsymbol{\nu}^n,\omega^n)\}_{n\in\mathbb{N}} \subseteq \Sigma_I$, and $\rho := (\boldsymbol{\mu},\boldsymbol{\nu},\omega) \in \Sigma_I$, we say that $\rho^n$ *converges in* $w^*/\mathbb{R}$ *to* $\rho$, and we write $\rho^n \rightharpoonup^{*/\mathbb{R}} \rho$, if and only if $(\mu_t^n,\nu_t^n) \rightharpoonup^* (\mu_t,\nu_t)$ for a.e. $t \in I$, and $\omega^n \to \omega$ in $\mathbb{R}$.

**Definition 5.12** ($L^1$-time feasible trajectories)**.** Given $\rho = (\boldsymbol{\mu},\boldsymbol{\nu},\omega) \in \mathcal{A}_I \times [0,+\infty[$, we define the map $\omega_\rho : I \to [0,+\infty]$ by setting

$$\omega_\rho(t) := \begin{cases} \omega + \displaystyle\int_{\min I}^t \theta_\rho(s)\,ds, & \text{if } \theta_\rho \in L^1([\min I,t]); \\ \\ +\infty, & \text{otherwise,} \end{cases}$$

where $\theta_\rho : I \to [0,+\infty]$ is defined as in Definition 5.1.

Given $(\bar\mu,\bar\omega) \in \mathscr{P}_p(\mathbb{R}^d) \times [0,+\infty[$, we set

$$\mathcal{F}_I^1(\bar\mu,\bar\omega) := \{\rho \in \mathcal{A}_I(\bar\mu) \times [0,+\infty[ \colon \omega_\rho(\max I) \le \alpha,\ \omega_\rho(\min I) = \bar\omega\}$$

$$\mathcal{R}_I^1(\bar\mu,\bar\omega) := \left\{\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_I) \,:\, \exists(\boldsymbol{\mu},\boldsymbol{\nu},\omega) \in \mathcal{F}_I^1(\bar\mu,\bar\omega) \text{ s.t. } \boldsymbol{\eta} \text{ represents } (\boldsymbol{\mu},\boldsymbol{\nu})\right\}.$$

and we define the set of $\alpha$-*feasible trajectories* defined on $I$ by

$$\mathcal{F}_I^1 := \bigcup_{\substack{\bar\mu \in \mathscr{P}_p(\mathbb{R}^d) \\ \bar\omega \ge 0}} \mathcal{F}_I^1(\bar\mu,\bar\omega) \subseteq \Sigma_I.$$

The counterpart of Proposition 5.3 is the following.

**Proposition 5.13.** *Let $I \subseteq \mathbb{R}$ be a compact nonempty interval, $p \ge 1$, $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, $C_1 \ge 0$, $\omega \in [0,\alpha]$. Then*

(1) *$\mathcal{F}_I^1$ is closed w.r.t. the topology of $\Sigma_I$;*
(2) *for any $\mathscr{B} \subseteq \mathcal{F}_I^\infty$, $C_1 > 0$ such that for all $(\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathscr{B}$ with $\boldsymbol{\mu} = \{\mu_t\}_{t\in I}$ it holds $\mathrm{m}_p(\mu_0) \le C_1$, we have that the closure of $\mathscr{B}$ in $\Sigma_I$ is contained in $\mathcal{F}_I^\infty$;*
(3) *$\mathcal{F}_I^1(\mu_0,\omega)$ is compact in the topology of $\Sigma_I$*
(4) *$\mathcal{R}_I^1(\mu_0,\omega)$ is compact in the narrow topology.*

*Proof.* It is enough to notice that, given $\rho := (\boldsymbol{\mu},\boldsymbol{\nu},\omega) \in \Sigma_I$, we have $\omega_\rho(\max I) \le \alpha$ if and only if $\mathcal{E}'(\rho) < +\infty$, where $\mathcal{E}' : AC(I,\mathscr{P}(\mathbb{R}^d)) \times \mathrm{Bor}(I,\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)) \times [0,+\infty[ \to [0,+\infty]$ is defined by

$$\mathcal{E}'(\boldsymbol{\mu},\boldsymbol{\nu},\omega) := \begin{cases} I_{[0,\alpha]}\left(\omega + \displaystyle\int_I\int_{\mathbb{R}^d} \Psi\left(x,\frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\,ds\right), & \text{if } |\nu_s| \ll \mu_s \text{ for a.e. } s \in I, \\ & \text{and } \theta_\rho \in L^1(I) \\ \\ +\infty, & \text{otherwise.} \end{cases}$$

Moreover, by applying Fatou's lemma and recalling the l.s.c. of the functional $\mathscr{F}$ in the proof of Proposition 5.3, we have that the functional $\mathscr{G} : AC(I,\mathscr{P}(\mathbb{R}^d)) \times \mathrm{Bor}(I,\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)) \times [0,+\infty[ \to [0,+\infty]$ defined by

$$\mathscr{G}(\boldsymbol{\mu},\boldsymbol{\nu},\omega) := \begin{cases} \omega + \displaystyle\int_I\int_{\mathbb{R}^d} \Psi\left(x,\frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\,ds, & \text{if } |\nu_s| \ll \mu_s \text{ for a.e. } s \in I, \\ & \text{and } \theta_\rho \in L^1(I) \\ \\ +\infty, & \text{otherwise.} \end{cases}$$

is l.s.c. w.r.t. $w^*/\mathbb{R}$-convergence. Thus the functional $\mathcal{E}'$ is l.s.c. The other assertions follows by an easy adaption of the proof of Proposition 5.3. □

With the following definition, we notice that Remark 5.5 can be easily applied also in this setting.

**Definition 5.14** (Concatenation and restriction).

(1) Let $I_i = [a_i, b_i] \subset \mathbb{R}$, $i = 1, 2$, with $b_1 = a_2$, and $I := I_1 \cup I_2$. Let $\rho^i = (\boldsymbol{\mu}^i, \boldsymbol{\nu}^i, \omega^i) \in \mathcal{F}^1_{I_i}$ with $\mu^1_{b_1} = \mu^2_{a_2}$, $\omega_{\rho^1}(b_1) = \omega^2$. The *concatenation* $\rho^1 \star \rho^2 = (\boldsymbol{\mu}, \boldsymbol{\nu}, \omega^1)$ of $\rho^1$ and $\rho^2$ is defined by setting $\mu_t = \mu^i_t$ and $\nu_t = \nu^i_t$ when $t \in I_i$ for $i = 1, 2$. We recall that this implies that $\rho^1 \star \rho^2 \in \mathcal{F}^1_I$, with a similar reasoning as for the $L^\infty$-time feasibility condition setting.

(2) Let $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}, \omega) \in \mathcal{F}^1_I$. The *restriction* $\rho_{|I_1}$ of $\rho$ to a compact and nonempty interval $I_1 \subseteq I$, where $\rho_{|I_1} = (\boldsymbol{\mu}^1 = \{\mu^1_t\}_{t\in I_1}, \boldsymbol{\nu}^1 = \{\nu^1_t\}_{t\in I_1}, \omega^1)$, is defined by setting $\mu^1_t := \mu_t$ and $\nu^1_t := \nu_t$ for all $t \in I_1$, $\omega^1 = \omega_\rho(\min I_1)$. Clearly we have $\rho_{|I_1} \in \mathcal{F}^1_{I_1}$.

As done in Definition 5.6 for the problem of Section 5.1, we now want to reconduct this framework to the general one of Definition 4.3 in order to gain the general results proved in Section 4.

**Definition 5.15.** Let $c$ and $c_f$ be as in Definition 4.1, satisfying the following additional properties

**($C'_3$)** $c(\mu^{(1)}, \omega^{(1)}, \mu^{(2)}, \omega^{(2)}, \hat{\rho}) < +\infty$ if and only if $\hat{\rho} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\nu}}, \hat{\omega}) \in \mathcal{F}^1_I(\mu^{(1)}, \omega^{(1)})$, with $\hat{\mu}_{|t=\max I} = \mu^{(2)}$ and $\omega_{\hat{\rho}}(\max I) = \omega^{(2)}$ for some compact and nonempty interval $I \subset \mathbb{R}$;

**($C'_4$)** let $0 \le a \le b \le c$, $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}, \omega) \in \mathcal{F}^1_{[a,c]}$. Then $c : X \times X \times \Sigma \to [0, +\infty]$ is superadditive by restrictions, i.e.

$$c(\mu_a, \omega, \mu_c, \omega_\rho(c), \rho) \ge c(\mu_a, \omega, \mu_b, \omega_\rho(b), \rho_{|[a,b]}) + c(\mu_b, \omega_\rho(b), \mu_c, \omega_\rho(c), \rho_{|[b,c]}).$$

Let $\bar{\mu} \in \mathscr{P}_p(\mathbb{R}^d)$, $\bar{\omega} \in [0, \alpha]$, $I \subset \mathbb{R}$ nonempty and compact interval, and $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}, \omega) \in \mathcal{F}^1_I(\bar{\mu}, \bar{\omega})$. We define the set $\mathscr{G}^1_\rho$ made of the pairs $(\gamma, \sigma)$ defined as follows

(1) $\gamma : I \to X$, $\gamma(t) := (\mu_t, \omega_\rho(t))$ for all $t \in I$;

(2) $\sigma : I \to \Sigma$, $\sigma(t) := \rho_{|[\min I, t]}$ for all $t \in I$.

Finally, we define the set

$$\mathscr{G}^1_I(\bar{\mu}, \bar{\omega}) := \left\{(\gamma, \sigma) \in \mathscr{G}^1_\rho \ : \ \rho \in \mathcal{F}^1_I(\bar{\mu}, \bar{\omega})\right\}.$$

**Theorem 5.16** (DPP for $L^1$-time feasibility case). *Let $V : \mathscr{P}_p(\mathbb{R}^d) \times [0, +\infty[ \to [0, +\infty]$ be as in Definition 4.1. For any $(\mu_0, \omega_0) \in \mathscr{P}_p(\mathbb{R}^d) \times [0, +\infty[$, we have*

$$V(\mu_0, \omega_0) = \inf_{\substack{\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}, \omega) \in \mathcal{F}^1_I(\mu_0, \omega_0) \\ I \subseteq \mathbb{R} \text{ compact interval}}} \left\{c(\mu_0, \omega_0, \mu_{|t=\max I}, \omega_\rho(\max I), \rho) + V(\mu_{|t=\max I}, \omega_\rho(\max I))\right\}.$$

*Proof.* Coming from Theorem 4.2 and **($C'_3$)**. □

We notice that Remark 5.8 holds also in this setting, thus we gain Corollaries 4.6, 4.7.

**Proposition 5.17** (Existence of minimizers). *Assume properties **($C_1$)**, **($C_2$)**, **($C'_3$)**, **($C'_4$)**. Let $p \ge 1$ and $\bar{\mu} \in \mathscr{P}_p(\mathbb{R}^d)$, $\bar{\omega} \in [0, \alpha]$. If $c(\bar{\mu}, \bar{\omega}, \cdot, \cdot, \cdot)$ and $c_f$ are l.s.c. w.r.t. $w^*/\mathbb{R}$-convergence and $V(\bar{\mu}, \bar{\omega}) < +\infty$, then there exist $I \subset \mathbb{R}$ nonempty and compact interval and an optimal trajectory $(\gamma, \sigma) \in \mathscr{G}^1_I(\bar{\mu}, \bar{\omega})$, according to Definition 4.5.*

*Proof.* Analogous to the proof of Proposition 5.9. □

We pass now to analyze the set of initial velocities in this framework. This result is used later on in Section 6.2 where an Hamilton-Jacobi-Bellman equation is studied.

**Definition 5.18.** Let $(\mu, \omega) \in \mathscr{P}_2(\mathbb{R}^d) \times [0, \alpha[$, $0 \le s \le T$.

(1) Given $\boldsymbol{\eta} \in \mathcal{R}_{[s,T]}(\mu,\omega)$, we define the set $\mathscr{V}_{[s,T]}(\boldsymbol{\eta})$ as in Definition 5.10.
(2) We set

$$\mathscr{V}^1_{[s,T]}(\mu,\omega) := \left\{ x \mapsto \int_{\Gamma^x_{[s,T]}} w_{\boldsymbol{\eta}}(x,\gamma)\, d\eta_x \,:\, \boldsymbol{\eta} \in \mathcal{R}^1_{[s,T]}(\mu,\omega),\, w_{\boldsymbol{\eta}} \in \mathscr{V}_{[s,T]}(\boldsymbol{\eta}) \right\},$$

where we denoted with $\{\eta_x\}_{x\in\mathbb{R}^d}$ the disintegration of $\boldsymbol{\eta}$ w.r.t. the map $e_s$.
(3) We define the set

$$\mathscr{Z}^1(\mu) := \left\{ v \in L^2_\mu(\mathbb{R}^d;\mathbb{R}^d) \,:\, v(x) \in F(x) \text{ for } \mu\text{-a.e. } x \in \mathbb{R}^d \right\}.$$

**Lemma 5.19** (Initial velocity)**.** *Let* $(\mu,\omega) \in \mathscr{P}_2(\mathbb{R}^d)\times[0,\alpha[$, $0 \le s \le T$. *Then* $\mathscr{V}^1_{[s,T]}(\mu,\omega) \equiv \mathscr{Z}^1(\mu)$.

*Proof.* The proof follows immediately by items (1) and (4) in Lemma 3.12, for the inclusion $\mathscr{Z}^1(\mu) \subseteq \mathscr{V}^1_{[s,T]}(\mu,\omega)$. It is sufficient to take $\beta = \alpha - \omega > 0$. The other inclusion comes straightforwardly by Lemma 5 in [16]. □

*Remark* 5.20. Notice that if $\omega = \alpha$, then $f(\cdot,0)$ is the only admissible velocity driving a feasible trajectory for positive times starting from $(\mu,\omega)$.

## 6. Hamilton-Jacobi-Bellman equation

In this section, we determine an Hamilton-Jacobi-Bellman equation for the $L^\infty$-time averaged and the $L^1$-time averaged feasibilty cases of Section 5. We consider general cost functions satisfying the properties $\boldsymbol{(C_1)} - \boldsymbol{(C_4)}$ already introduced, and the following property $\boldsymbol{(C_5)}$ regarding a limiting behavior. Concerning the definition of sub/super-differentials that we choose to adopt in the space of probability measures, we refer to the recent survey [27]. Our aim is to provide *necessary conditions* in the sense of the formulation of an Hamilton-Jacobi-Bellman equation in the space of measures solved by the value function in a suitable viscosity sense.

Referring to Section 4 and Definition 4.3, we ask the following further condition on the cost function $c$.

$\boldsymbol{(C_5)}$ There exists a map $h : X \to \mathbb{R}$ such that for any $x \in X$, $t > 0$ and any $(\gamma,\sigma)$ generalized admissible trajectory from $x$ defined on $[0,t]$, we have

$$\lim_{t\to 0^+} \frac{c(x,\gamma(t),\sigma(t))}{t} = h(x).$$

### 6.1. $L^\infty$-feasibility case.

Throughout this section, we consider the framework described in Section 5.1, hence

$$X = \mathscr{P}_2(\mathbb{R}^d), \quad \Sigma = \bigcup_{\substack{I\subseteq\mathbb{R} \\ I \text{ compact interval}}} \left[ AC(I;\mathscr{P}_2(\mathbb{R}^d)) \times \mathrm{Bor}(I;\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)) \right].$$

We give now a definition of viscosity sub/super-differentials used in this paper, inspired by Definition 3.2 in [27].

**Definition 6.1** (Viscosity sub/super-differentials)**.** Let $w : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be a map, $\bar\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $\delta > 0$. We say that $p \in L^2_{\bar\mu}(\mathbb{R}^d)$ belongs to the viscosity $\delta$-superdifferential of $w$ at $\bar\mu$, and we write $p \in D^+_\delta w(\bar\mu)$, if for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ we have

$$w(\mu) - w(\bar\mu) \le \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_2, x_3 - x_1\rangle\, d\tilde\mu(x_1,x_2,x_3) + \delta W_2(\bar\mu,\mu) + o(W_2(\bar\mu,\mu)),$$

for all $\tilde\mu \in \mathscr{P}(\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d)$ such that $\pi_{1,2}\sharp\tilde\mu = (\mathrm{Id}_{\mathbb{R}^d}, p)\sharp\bar\mu$ and $\pi_{1,3}\sharp\tilde\mu \in \Pi(\bar\mu,\mu)$.

In a similar way, the set of the viscosity $\delta$-subdifferentials of $w$ at $\bar\mu$ is defined by $D^-_\delta w(\bar\mu) = -D^+_\delta(-w)(\bar\mu)$.

We consider the same definition of viscosity sub/super-solutions given in [27] as follows.

**Definition 6.2** (Viscosity solutions)**.** Consider the equation

$$\mathscr{H}(\mu, Dw(\mu)) = 0, \tag{6.1}$$

for a given hamiltonian $\mathscr{H} : T^*(\mathscr{P}_2(\mathbb{R}^d)) \to \mathbb{R}$, i.e. $\mathscr{H}(\mu, p)$ is defined for any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $p \in L^2_\mu(\mathbb{R}^d)$. We say that a function $w : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is

- a *viscosity subsolution* of (6.1) if $w$ is u.s.c. and there exists a constant $C > 0$ such that
$$\mathscr{H}(\mu, p) \geq -C\delta,$$
for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $p \in D^+_\delta w(\mu)$ and $\delta > 0$.
- a *viscosity supersolution* of (6.1) if $w$ is l.s.c. and there exists a constant $C > 0$ such that
$$\mathscr{H}(\mu, p) \leq C\delta,$$
for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $p \in D^-_\delta w(\mu)$ and $\delta > 0$.
- a *viscosity solution* of (6.1) if $w$ is both a viscosity subsolution and supersolution.

We now prove that a value function with associated cost function satisfying $\boldsymbol{(C_1)-(C_5)}$ is a viscosity solution of a suitable Hamilton-Jacobi-Bellman equation with Hamiltonian defined as follows.

**Definition 6.3** (Hamiltonian)**.** We define the *hamiltonian* $\mathscr{H}^\infty : T^*(\mathscr{P}_2(\mathbb{R}^d)) \to \mathbb{R}$ by

$$\mathscr{H}^\infty(\mu, p) := h(\mu) + \inf_{v \in \mathscr{Z}^\infty(\mu)} \int_{\mathbb{R}^d} \langle p, v \rangle \, d\mu,$$

for any $(\mu, p) \in T^*(\mathscr{P}_2(\mathbb{R}^d))$, where the map $h$ is given in $\boldsymbol{(C_5)}$.

**Theorem 6.4** (HJB)**.** *Assume properties* $\boldsymbol{(C_1)-(C_5)}$*. Let* $\mathcal{B} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ *open with uniformly bounded 2-moments. Assume* $c(\mu, \cdot, \cdot)$ *and* $c_f$ *to be l.s.c. w.r.t.* $w^*$*-convergence,* $\mu \in \mathcal{B}$*, and the value function* $V : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ *to be continuous on* $\mathcal{B}$*. Then* $V$ *is a viscosity solution of* $\mathscr{H}^\infty(\mu, DV(\mu)) = 0$ *on* $\mathcal{B}$*, where* $\mathscr{H}^\infty$ *is defined as in 6.3.*

*Proof. Claim 1.* $V$ is a subsolution of $\mathscr{H}^\infty(\mu, DV(\mu)) = 0$ on $\mathcal{B}$.

Let $\bar{\mu} \in \mathcal{B}$, $\delta > 0$, $p \in D^+_\delta V(\bar{\mu})$. Let $v \in \mathscr{Z}^\infty(\bar{\mu})$, then by Lemma 5.11 there exists $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{F}^\infty_{[0,T]}(\bar{\mu})$, $\boldsymbol{\eta}$ representing $(\boldsymbol{\mu}, \boldsymbol{\nu})$ such that

$$\lim_{t \to 0^+} \int_{\mathbb{R}^d \times \Gamma_T} \langle p \circ e_0, \frac{e_t - e_0}{t} \rangle \, d\boldsymbol{\eta} = \int_{\mathbb{R}^d} \langle p, v \rangle \, d\bar{\mu}.$$

By Theorem 5.7,

$$V(\mu_t) - V(\bar{\mu}) + c(\bar{\mu}, \mu_t, \sigma(t)) \geq 0,$$

for all $t \in [0, T]$, where $\sigma(t) := (\boldsymbol{\mu}, \boldsymbol{\nu})_{|[0,t]}$. Notice that if we define $\tilde{\mu} = (e_0, p \circ e_0, e_t)\sharp\boldsymbol{\eta}$, we have $\pi_{1,2}\sharp\tilde{\mu} = (\mathrm{Id}_{\mathbb{R}^d}, p)\sharp\bar{\mu}$ and $\pi_{1,3}\sharp\tilde{\mu} = (e_0, e_t)\sharp\boldsymbol{\eta} \in \Pi(\bar{\mu}, \mu_t)$. Hence, $W_2(\bar{\mu}, \mu_t) \leq \|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}}$, which vanishes as $t \to 0^+$ by continuity of $t \mapsto e_t$ (see Proposition 2.3 in [27]).

Thus, we can apply the definition of viscosity superdifferential with $\tilde{\mu}$ as before and have

$$\begin{aligned}
0 &\leq V(\mu_t) - V(\bar{\mu}) + c(\bar{\mu}, \mu_t, \sigma(t)) \\
&\leq \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle x_2, x_3 - x_1 \rangle \, d\tilde{\mu}(x_1, x_2, x_3) + \delta \, W_2(\bar{\mu}, \mu_t) + o(W_2(\bar{\mu}, \mu_t)) + c(\bar{\mu}, \mu_t, \sigma(t)) \\
&\leq \int_{\mathbb{R}^d \times \Gamma_T} \langle p \circ e_0, e_t - e_0 \rangle \, d\boldsymbol{\eta} + \delta \, \|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}} + o(\|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}}) + c(\bar{\mu}, \mu_t, \sigma(t)).
\end{aligned}$$

Dividing by $t$, we have $\|\frac{e_t - e_0}{t}\|_{L^2_{\boldsymbol{\eta}}} \leq K$, where $K$ is a suitable constant coming from Proposition B.1 and from hypothesis. Hence,

$$-\delta K \leq \int_{\mathbb{R}^d \times \Gamma_T} \langle p \circ e_0, \frac{e_t - e_0}{t} \rangle \, d\boldsymbol{\eta} + \frac{c(\bar{\mu}, \mu_t, \sigma(t))}{t},$$

and by letting $t \to 0^+$ we get

$$-\delta K \leq \int_{\mathbb{R}^d} \langle p(x), v(x) \rangle \, d\bar{\mu}(x) + h(\bar{\mu}).$$

We conclude by passing to the infimum on $v \in \mathscr{Z}^\infty(\bar{\mu})$.

*Claim 2.* $V$ is a supersolution of $\mathscr{H}^\infty(\mu, DV(\mu)) = 0$ on $\mathcal{B}$.

Take $\bar{\mu} \in \mathcal{B}$, $\delta > 0$, $p \in D^-_\delta V(\bar{\mu})$. By Proposition 5.9, there exist $T > 0$, an optimal trajectory $(\gamma, \sigma) \in \mathscr{G}^\infty_\rho$ with $\rho = (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{F}^\infty_{[0,T]}(\bar{\mu})$, and a representation $\boldsymbol{\eta}$ such that

$$V(\mu_t) - V(\bar{\mu}) + c(\bar{\mu}, \mu_t, \sigma(t)) = 0, \quad \text{for all } t \in [0, T].$$

We can take as before $\tilde{\mu} = (e_0, p \circ e_0, e_t)\sharp\boldsymbol{\eta}$, thus we have $W_2(\bar{\mu}, \mu_t) \leq \|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}}$ and we obtain

$$\begin{aligned} 0 &= V(\mu_t) - V(\bar{\mu}) + c(\bar{\mu}, \mu_t, \sigma(t)) \\ &\geq \int_{\mathbb{R}^d \times \Gamma_T} \langle p \circ e_0, e_t - e_0 \rangle \, d\boldsymbol{\eta} - \delta \, \|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}} - o(\|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}}) + c(\bar{\mu}, \mu_t, \sigma(t)). \end{aligned}$$

Dividing by $t$ and reasoning as in Claim 1, we get

$$\delta K \geq \int_{\mathbb{R}^d \times \Gamma_T} \langle p \circ e_0, \frac{e_t - e_0}{t} \rangle \, d\boldsymbol{\eta} + \frac{c(\bar{\mu}, \mu_t, \sigma(t))}{t}.$$

Now, there exists a sequence $\{t_i\}_{i \in \mathbb{N}} \subseteq ]0, T[$ and $w_{\boldsymbol{\eta}} \in \mathscr{V}_{[0,T]}(\boldsymbol{\eta})$ s.t. $t_i \to 0^+$, $\frac{e_{t_i} - e_0}{t_i}$ weakly converge to $w_{\boldsymbol{\eta}}$ in $L^2_{\boldsymbol{\eta}}$, thus by letting $t_i \to 0^+$, thanks to Lemma 5.11, there exists $v \in \mathscr{Z}^\infty(\bar{\mu})$ s.t.

$$\delta K \geq \int_{\mathbb{R}^d} \langle p, v \rangle \, d\bar{\mu} + h(\bar{\mu}) \geq \mathscr{H}^\infty(\bar{\mu}, p).$$

□

### 6.2. $L^1$-feasibility case.

In this section, we consider the framework described in Section 5.2, hence

$$X = \mathscr{P}_2(\mathbb{R}^d) \times [0, +\infty[, \quad \Sigma = \bigcup_{\substack{I \subseteq \mathbb{R} \\ I \text{ compact interval}}} \Big[ AC(I; \mathscr{P}_2(\mathbb{R}^d)) \times \mathrm{Bor}(I; \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)) \Big] \times [0, +\infty[.$$

Similarly to the previous case, we give the following.

**Definition 6.5** (Viscosity sub/super-differentials)**.** Let $w : \mathscr{P}_2(\mathbb{R}^d) \times [0, +\infty[ \to \mathbb{R}$ be a map, $(\bar{\mu}, \bar{\omega}) \in \mathscr{P}_2(\mathbb{R}^d) \times [0, +\infty[$, $\delta > 0$. We say that $(p_{\bar{\mu}}, p_{\bar{\omega}}) \in L^2_{\bar{\mu}}(\mathbb{R}^d) \times [0, +\infty[$ belongs to the viscosity $\delta$-superdifferential of $w$ at $(\bar{\mu}, \bar{\omega})$, and we write $(p_{\bar{\mu}}, p_{\bar{\omega}}) \in D^+_\delta w(\bar{\mu}, \bar{\omega})$, if for all $(\mu, \omega) \in \mathscr{P}_2(\mathbb{R}^d) \times [0, +\infty[$ we have

$$\begin{aligned} w(\mu, \omega) - w(\bar{\mu}, \bar{\omega}) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle x_2, x_3 - x_1 \rangle \, d\tilde{\mu}(x_1, x_2, x_3) + p_{\bar{\omega}} \, |\omega - \bar{\omega}| + \\ + \delta \sqrt{W_2^2(\bar{\mu}, \mu) + |\omega - \bar{\omega}|^2} + o(W_2(\bar{\mu}, \mu) + |\omega - \bar{\omega}|), \end{aligned}$$

for all $\tilde{\mu} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi_{1,2}\sharp\tilde{\mu} = (\mathrm{Id}_{\mathbb{R}^d}, p_{\bar{\mu}})\sharp\bar{\mu}$ and $\pi_{1,3}\sharp\tilde{\mu} \in \Pi(\bar{\mu}, \mu)$.

In a similar way, the set of the viscosity $\delta$-subdifferentials of $w$ at $(\bar{\mu}, \bar{\omega})$ is defined by $D^-_\delta w(\bar{\mu}, \bar{\omega}) = -D^+_\delta(-w)(\bar{\mu}, \bar{\omega})$.

We adopt the same definition of viscosity sub/super-solutions given in Definition 6.2, with the natural adaptations for this context.

**Definition 6.6** (Hamiltonian)**.** We define the *hamiltonian* $\mathscr{H}^1 : T^*(\mathscr{P}_2(\mathbb{R}^d) \times [0, +\infty[) \to \mathbb{R}$ by

$$\mathscr{H}^1(\mu, \omega, p_\mu, p_\omega) := h(\mu, \omega) + \inf_{v \in \mathscr{Z}^1(\mu)} \left\{ \int_{\mathbb{R}^d} [\langle p_\mu(x), v(x)\rangle + p_\omega \, \Psi(x, v(x))] \, d\mu(x) \right\},$$

for any $(\mu, \omega, p_\mu, p_\omega) \in T^*(\mathscr{P}_2(\mathbb{R}^d) \times [0, +\infty[)$, where the map $h$ is given in $(\boldsymbol{C_5})$.

**Lemma 6.7.** *Let $(\mu_0, \bar{\omega}) \in \mathscr{P}_2(\mathbb{R}^d) \times [0, \alpha[$, $T > 0$. Let $(\boldsymbol{\mu}, \boldsymbol{\nu}, \bar{\omega}) \in \mathcal{F}^1_{[0,T]}(\mu_0, \bar{\omega})$ be a feasible trajectory represented by $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ in the sense of Definition 3.5(1). Then there exist $w : \mathbb{R}^d \to \mathbb{R}^d$ $L^2_{\mu_0}$-selection of $F(\cdot)$, and a sequence $t_k \in [0, T]$, $t_k \to 0^+$, such that*

(1) *for all $p \in L^2_{\mu_0}(\mathbb{R}^d)$*

$$\lim_{k \to +\infty} \int_{\mathbb{R}^d \times \Gamma_{[0,T]}} \langle \frac{e_{t_k}(x, \gamma) - e_0(x, \gamma)}{t_k}, p \circ e_0(x, \gamma)\rangle \, d\boldsymbol{\eta}(x, \gamma) = \int_{\mathbb{R}^d} \langle w(x), p(x)\rangle \, d\mu_0(x);$$

(2) $$\lim_{k \to +\infty} \frac{1}{t_k} \int_0^{t_k} \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) \, d\mu_s(x) \, ds \geq \int_{\mathbb{R}^d} \Psi(x, w(x)) \, d\mu_0(x).$$

*Proof.* Let $t_k \to 0^+$ be any sequence along which $\frac{e_{t_k} - e_0}{t_k}$ weakly converges in $L^2_{\boldsymbol{\eta}}$. Then, item (1) follows by Lemma 5.19.

Let us prove the second item. For any $t \in [0, T]$ and any $\varphi \in C^0_C(\mathbb{R}^d)$ we have

$$\begin{aligned}
\langle \frac{e_t - e_0}{t}, \varphi \circ e_0\rangle_{L^2_{\boldsymbol{\eta}}} =& \iint_{\mathbb{R}^d \times \Gamma_T} \langle \frac{\gamma(t) - \gamma(0)}{t}, \varphi(x)\rangle \, d\boldsymbol{\eta} \\
=& \frac{1}{t} \int_0^t \iint_{\mathbb{R}^d \times \Gamma_T} \langle \dot{\gamma}(s), \varphi(\gamma(0))\rangle \, d\boldsymbol{\eta} \, ds \\
=& \frac{1}{t} \int_0^t \iint_{\mathbb{R}^d \times \Gamma_T} \langle \dot{\gamma}(s), \varphi(\gamma(s))\rangle \, d\boldsymbol{\eta} \, ds + \\
&+ \frac{1}{t} \int_0^t \iint_{\mathbb{R}^d \times \Gamma_T} \langle \dot{\gamma}(s), \varphi(\gamma(0)) - \varphi(\gamma(s))\rangle \, d\boldsymbol{\eta} \, ds \\
\leq& \frac{1}{t} \int_0^t \int_{\mathbb{R}^d} \iint_{\mathbb{R}^d \times \Gamma_T} \langle \dot{\gamma}(s), \varphi(\gamma(s))\rangle \, d\boldsymbol{\eta}(x, \gamma) \, ds + \\
&+ \frac{1}{t} \int_0^t \iint_{\mathbb{R}^d \times \Gamma_T} C(1 + |\gamma(s)|)|\varphi(\gamma(0)) - \varphi(\gamma(s))| \, d\boldsymbol{\eta} \, ds \\
=& \frac{1}{t} \int_0^t \iint_{\mathbb{R}^d \times \Gamma_T} \langle \frac{\nu_s}{\mu_s}(\gamma(s)), \varphi(\gamma(s))\rangle \, d\boldsymbol{\eta} \, ds + \frac{1}{t} \int_0^t H(s) \, ds \\
=& \frac{1}{t} \int_0^t \int_{\mathbb{R}^d} \langle \frac{\nu_s}{\mu_s}(y), \varphi(y)\rangle \, d\mu_s(y) \, ds + \frac{1}{t} \int_0^t H(s) \, ds
\end{aligned}$$

where $s \mapsto H(s)$ is the continuous function defined by

$$H(s) = \iint_{\mathbb{R}^d \times \Gamma_T} C(1 + |\gamma(s)|)|\varphi(\gamma(0)) - \varphi(\gamma(s))| \, d\boldsymbol{\eta} \, ds.$$

With the very same argument, denoted with $v_s(y) = \frac{\nu_s}{\mu_s}(y)$, we can prove that

$$\left| \frac{1}{t} \int_0^t \int_{\mathbb{R}^d} \langle v_s(y), \varphi(y)\rangle \, d\mu_s(y) \, ds - \langle \frac{e_t - e_0}{t}, \varphi \circ e_0\rangle_{L^2_{\boldsymbol{\eta}}} \right| \leq \frac{1}{t} \int_0^t H(s) \, ds, \tag{6.2}$$

and the right hand side tends to 0 as $t \to 0$. In particular, as a consequence of item (1), we get

$$\lim_{k \to +\infty} \frac{1}{t_k} \int_0^{t_k} \int_{\mathbb{R}^d} \langle v_s(y), \varphi(y)\rangle \, d\mu_s(y) \, ds = \int_{\mathbb{R}^d} \langle w(x), \varphi(x)\rangle \, d\mu_0(x). \tag{6.3}$$

Now, as already observed in the proof of Proposition 5.3 (Step 1), we can apply Lemma 2.2.3(i) in [7] to say that there exist $\{a_h\}_{h\in\mathbb{N}}, \{b_h\}_{h\in\mathbb{N}} \subseteq C^0(\mathbb{R}^d;\mathbb{R})$ such that $\Psi(x,v) = \sup_h[a_h(x) + \langle v, b_h(x)\rangle]$, for all $x, v \in \mathbb{R}^d$. Without loss of generality, we can assume $a_h, b_h$ to have compact support. Thus,

$$\begin{aligned}
&\lim_{k\to+\infty} \frac{1}{t_k}\int_0^{t_k}\int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\, ds \\
&= \lim_{k\to+\infty} \int_0^1\int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_{t_k w}}{\mu_{t_k w}}(x)\right) d\mu_{t_k w}(x)\, dw \\
&\geq \lim_{k\to+\infty} \int_0^1\int_{\mathbb{R}^d} \left(a_h(x) + \langle \frac{\nu_{t_k w}}{\mu_{t_k w}}(x), b_h(x)\rangle\right) d\mu_{t_k w}(x)\, dw \\
&= \int_{\mathbb{R}^d} (a_h(x) + \langle w, b_h(x)\rangle)\, d\mu_0(x),
\end{aligned}$$

where the last passage follows by absolute continuity of $\boldsymbol{\mu}$ and by (6.3).

By positivity of $\Psi$, we can consider $\{a_h\}_{h\in\mathbb{N}}, \{b_h\}_{h\in\mathbb{N}}$ to be positive. Let $g_h(x) := a_h(x) + \langle w(x), b_h(x)\rangle$, and $\hat{g}_k := \max\{g_h(x) : h \leq k\}$. Now, since $\hat{g}_k$ is a non-decreasing sequence of measurable and non-negative functions with $\sup_h g_h(x) = \sup_k \hat{g}_k(x)$, then passing to the supremum and applying Beppo-Levi Theorem we have

$$\lim_{k\to+\infty} \frac{1}{t_k}\int_0^{t_k}\int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\, ds \geq \int_{\mathbb{R}^d} \Psi(x,w)\, d\mu_0(x).$$

□

**Theorem 6.8** (HJB). *Assume properties* $(\boldsymbol{C_1}), (\boldsymbol{C_2}), (\boldsymbol{C_3'}), (\boldsymbol{C_4'}), (\boldsymbol{C_5})$. *Let* $\mathcal{B} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ *open with uniformly bounded 2-moments. Assume* $c(\mu,\omega,\cdot,\cdot,\cdot)$ *and* $c_f$ *to be l.s.c. w.r.t.* $w^*/\mathbb{R}$*-convergence,* $\mu \in \mathcal{B}$ *and* $\omega \in [0,\alpha[$. *Assume the value function* $V : \mathscr{P}_2(\mathbb{R}^d) \times [0,+\infty[\to \mathbb{R}$ *to be continuous on* $\mathcal{B}\times[0,\alpha[$. *Then* $V$ *is a viscosity solution of* $\mathscr{H}^1(\mu,\omega,DV(\mu,\omega)) = 0$ *on* $\mathcal{B}\times[0,\alpha[$, *where* $\mathscr{H}^1$ *is defined as in 6.6.*

*Proof. Claim 1.* $V$ is a subsolution of $\mathscr{H}^1(\mu,\omega,DV(\mu,\omega)) = 0$ on $\mathcal{B}\times[0,\alpha[$.

Let $(\bar{\mu},\bar{\omega}) \in \mathcal{B}\times[0,\alpha[$, $\delta > 0$, $(p_{\bar{\mu}}, p_{\bar{\omega}}) \in D^+_\delta V(\bar{\mu},\bar{\omega})$, and $v \in \mathscr{Z}^1(\bar{\mu})$. By Lemma 5.19 and 3.12, there exist $(\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{A}_{[0,T]}(\bar{\mu})$ and $\boldsymbol{\eta}$ representing $(\boldsymbol{\mu},\boldsymbol{\nu})$ such that $\rho := (\boldsymbol{\mu},\boldsymbol{\nu},\bar{\omega}) \in \mathcal{F}^1_{[0,T]}(\bar{\mu},\bar{\omega})$, and items (1), (3) of Lemma 3.12 are satisfied with $v = v_0$.

By Theorem 5.16 we have

$$V(\mu_t,\omega_\rho(t)) - V(\bar{\mu},\bar{\omega}) + c(\bar{\mu},\bar{\omega},\mu_t,\omega_\rho(t),\sigma(t)) \geq 0,$$

for all $t \in [0,T]$, where $\sigma(t) := \rho_{|[0,t]}$. Moreover, if we define $\tilde{\mu} = (e_0, p_{\bar{\mu}}\circ e_0, e_t)\sharp\boldsymbol{\eta}$, then $\pi_{1,2}\sharp\tilde{\mu} = (\mathrm{Id}_{\mathbb{R}^d}, p_{\bar{\mu}})\sharp\bar{\mu}$ and $\pi_{1,3}\sharp\tilde{\mu} = (e_0,e_t)\sharp\boldsymbol{\eta} \in \Pi(\bar{\mu},\mu_t)$. Proceeding analogously to the proof of Theorem 6.4 and recalling items (1), (3) of Lemma 3.12, we have

$$\begin{aligned}
0 &\leq V(\mu_t,\omega_\rho(t)) - V(\bar{\mu},\bar{\omega}) + c(\bar{\mu},\bar{\omega},\mu_t,\omega_\rho(t),\sigma(t)) \\
&\leq \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_2, x_3 - x_1\rangle\, d\tilde{\mu}(x_1,x_2,x_3) + p_{\bar{\omega}}\,|\omega_\rho(t) - \bar{\omega}| + \\
&\qquad + \delta\sqrt{W_2^2(\bar{\mu},\mu_t) + |\omega_\rho(t)-\bar{\omega}|^2} + o(W_2(\bar{\mu},\mu_t) + |\omega_\rho(t)-\bar{\omega}|) + c(\bar{\mu},\bar{\omega},\mu_t,\omega_\rho(t),\sigma(t)) \\
&= \int_{\mathbb{R}^d\times\Gamma_T} \langle p_{\bar{\mu}}\circ e_0, e_t - e_0\rangle\, d\boldsymbol{\eta} + p_{\bar{\omega}}\int_0^t\int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\, ds + \\
&\qquad + \delta\sqrt{W_2^2(\bar{\mu},\mu_t) + |\omega_\rho(t)-\bar{\omega}|^2} + o(W_2(\bar{\mu},\mu_t) + |\omega_\rho(t)-\bar{\omega}|) + c(\bar{\mu},\bar{\omega},\mu_t,\omega_\rho(t),\sigma(t)) \\
&\leq \int_{\mathbb{R}^d\times\Gamma_T} \langle p_{\bar{\mu}}\circ e_0, e_t - e_0\rangle\, d\boldsymbol{\eta} + t\, p_{\bar{\omega}}\int_{\mathbb{R}^d} \Psi(x,v(x))\, d\bar{\mu}(x) + \\
&\qquad + \delta\sqrt{W_2^2(\bar{\mu},\mu_t) + |\omega_\rho(t)-\bar{\omega}|^2} + o(W_2(\bar{\mu},\mu_t) + |\omega_\rho(t)-\bar{\omega}|) + c(\bar{\mu},\bar{\omega},\mu_t,\omega_\rho(t),\sigma(t)).
\end{aligned}$$

Let us now divide by $t$ and recall that $\|\frac{e_t-e_0}{t}\|_{L^1_{\boldsymbol{\eta}}} \leq K_1$, for some $K_1 > 0$, and also $\frac{\omega_\rho(t)-\bar{\omega}}{t} \leq \int_{\mathbb{R}^d} \Psi(x, v(x))\, d\bar{\mu}(x) \leq K_2$, for some $K_2 > 0$, by boundedness of $U$. Then, for some $K > 0$ we have

$$-\delta K \leq \int_{\mathbb{R}^d\times\Gamma_T} \langle p_{\bar{\mu}} \circ e_0, \frac{e_t - e_0}{t}\rangle\, d\boldsymbol{\eta} + p_{\bar{\omega}} \int_{\mathbb{R}^d} \Psi(x, v(x))\, d\bar{\mu}(x) + \\ + \frac{1}{t} o(\|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}} + |\omega_\rho(t) - \bar{\omega}|) + \frac{1}{t} c(\bar{\mu}, \bar{\omega}, \mu_t, \omega_\rho(t), \sigma(t)).$$

By letting $t \to 0^+$,

$$-\delta K \leq \int_{\mathbb{R}^d} \langle p_{\bar{\mu}}(x), v(x)\rangle\, d\bar{\mu} + p_{\bar{\omega}} \int_{\mathbb{R}^d} \Psi(x, v(x))\, d\bar{\mu} + h(\bar{\mu}, \bar{\omega}),$$

and we conclude by passing to the infimum w.r.t. $v \in \mathscr{Z}^1(\bar{\mu})$.

*Claim 2.* $V$ is a supersolution of $\mathscr{H}^1(\mu, \omega, DV(\mu, \omega)) = 0$ on $\mathcal{B} \times [0, \alpha[$.

Let $(\bar{\mu}, \bar{\omega}) \in \mathcal{B} \times [0, \alpha[$, $\delta > 0$, $(p_{\bar{\mu}}, p_{\bar{\omega}}) \in D^-_\delta V(\bar{\mu}, \bar{\omega})$. By Proposition 5.17, there exist an optimal trajectory $\rho := (\boldsymbol{\mu}, \boldsymbol{\nu}, \bar{\omega}) \in \mathcal{F}^1_{[0,T]}(\bar{\mu}, \bar{\omega})$ and a representation $\boldsymbol{\eta}$ such that

$$V(\mu_t, \omega_\rho(t)) - V(\bar{\mu}, \bar{\omega}) + c(\bar{\mu}, \bar{\omega}, \mu_t, \omega_\rho(t), \sigma(t)) = 0, \quad \text{for all } t \in [0, T].$$

As done for Claim 1, we can take $\tilde{\mu} = (e_0, p_{\bar{\mu}} \circ e_0, e_t)\sharp\boldsymbol{\eta}$, thus

$$\begin{aligned}
0 &= V(\mu_t, \omega_\rho(t)) - V(\bar{\mu}, \bar{\omega}) + c(\bar{\mu}, \bar{\omega}, \mu_t, , \omega_\rho(t), \sigma(t)) \\
&\geq \int_{\mathbb{R}^d\times\Gamma_T} \langle p_{\bar{\mu}} \circ e_0, e_t - e_0\rangle\, d\boldsymbol{\eta} + p_{\bar{\omega}} \int_0^t \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\, ds + \\
&\quad - \delta \sqrt{W_2^2(\bar{\mu}, \mu_t) + |\omega_\rho(t) - \bar{\omega}|^2} - o(W_2(\bar{\mu}, \mu_t) + |\omega_\rho(t) - \bar{\omega}|) + c(\bar{\mu}, \bar{\omega}, \mu_t, \omega_\rho(t), \sigma(t)) \\
&\geq \int_{\mathbb{R}^d\times\Gamma_T} \langle p_{\bar{\mu}} \circ e_0, e_t - e_0\rangle\, d\boldsymbol{\eta} + p_{\bar{\omega}} \int_0^t \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\, ds + \\
&\quad - \delta \sqrt{\|e_t - e_0\|^2_{L^2_{\boldsymbol{\eta}}} + \left(\int_0^t \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s\, ds\right)^2} - o(\|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}} + |\omega_\rho(t) - \bar{\omega}|) \\
&\quad + c(\bar{\mu}, \bar{\omega}, \mu_t, \omega_\rho(t), \sigma(t)).
\end{aligned}$$

Dividing by $t$ and reasoning as in Claim 1, we get

$$\delta K \geq \int_{\mathbb{R}^d\times\Gamma_T} \langle p_{\bar{\mu}} \circ e_0, \frac{e_t - e_0}{t}\rangle\, d\boldsymbol{\eta} + p_{\bar{\omega}} \frac{1}{t} \int_0^t \int_{\mathbb{R}^d} \Psi\left(x, \frac{\nu_s}{\mu_s}(x)\right) d\mu_s(x)\, ds + \\ - \frac{1}{t} o(\|e_t - e_0\|_{L^2_{\boldsymbol{\eta}}} + |\omega_\rho(t) - \bar{\omega}|) + \frac{1}{t} c(\bar{\mu}, \bar{\omega}, \mu_t, \omega_\rho(t), \sigma(t)).$$

Now, by Lemma 6.7 there exists a sequence $\{t_k\}_{k\in\mathbb{N}} \subseteq ]0, T[$ and $v_0 \in \mathscr{Z}^1(\bar{\mu})$ s.t. by passing to the limit in the previous estimate, along the sequence $t_k$, we get

$$\delta K \geq \int_{\mathbb{R}^d} \langle p_{\bar{\mu}}, v_0\rangle\, d\bar{\mu} + p_{\bar{\omega}} \int_{\mathbb{R}^d} \Psi(x, v_0(x))\, d\bar{\mu} + h(\bar{\mu}, \bar{\omega}) \geq \mathscr{H}^1(\bar{\mu}, \bar{\omega}, p_{\bar{\mu}}, p_{\bar{\omega}}).$$

□

## 7. A special case: the minimum time function

In this section we show a remarkable exemple where the theory proposed in this paper can be applied.

For simplicity of exposition, consider the framework outlined in Section 5.1 (similarly, it is possible to consider the setting of Section 5.2). We thus take $X = \mathscr{P}_2(\mathbb{R}^d)$, $\Sigma =$

$\bigcup_{\substack{I\subseteq\mathbb{R}\\ I \text{ compact interval}}} \Big[AC(I;\mathscr{P}_2(\mathbb{R}^d))\times \mathrm{Bor}(I;\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d))\Big]$. A closed nonempty *target set* $\tilde{S}\subseteq \mathscr{P}_2(\mathbb{R}^d)$ is given.

We then define the cost functions $c: X\times X\times\Sigma\to[0,+\infty]$ and $c_f: X\to[0,+\infty]$ as follows

$$c(\mu^{(1)},\mu^{(2)},(\boldsymbol{\mu},\boldsymbol{\nu}))=\begin{cases}T, & \text{if } (\boldsymbol{\mu},\boldsymbol{\nu})\in\mathcal{F}^{\infty}_{[0,T]}(\mu^{(1)}) \text{ and } \mu_{|t=T}=\mu^{(2)},\\ +\infty, & \text{otherwise,}\end{cases}$$

$$c_f(\tilde{\mu})=\begin{cases}0, & \text{if } \tilde{\mu}\in\tilde{S},\\ +\infty, & \text{if } \tilde{\mu}\notin\tilde{S}.\end{cases}$$

Following the notation introduced in Section 4, we have that $V(\mu)<+\infty$ if and only if there exist a feasible trajectory $(\boldsymbol{\mu},\boldsymbol{\nu})$ joining $\mu$ with the target set $\tilde{S}$. In this case $V(\mu)$ is the infimum amount of time where such trajectories are defined, and so $V(\mu)=0$ if and only if $\mu\in\tilde{S}$.

*Remark* 7.1. Such a value function $V$ is the so called *minimum time function* studied in [13, 14, 16, 17]. It is important to stress that in those references, no feasibility constraints are imposed, dealing just with admissibility properties for trajectories in $\mathscr{P}_2(\mathbb{R}^d)$. This paper provides thus a sort of extension of those results.

We show now that conditions $(\boldsymbol{C_1})-(\boldsymbol{C_5})$ are satisfied.

- Check for $(\boldsymbol{C_1})$. Given $\mu^{(i)}\in X$, $i=1,2,3$ and $(\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})=(\{\mu_t^{(j)}\}_{[0,T_j]},\{\nu_t^{(j)}\}_{[0,T_j]})\in\Sigma$, $j=1,2$, we must prove that it is possible to construct $(\boldsymbol{\mu}',\boldsymbol{\nu}')=(\{\mu_t'\}_{[0,T']},\{\nu_t'\}_{[0,T']})\in\Sigma$ such that
$$c(\mu^{(1)},\mu^{(3)},(\boldsymbol{\mu}',\boldsymbol{\nu}'))\le c(\mu^{(1)},\mu^{(2)},(\boldsymbol{\mu}^{(1)},\boldsymbol{\nu}^{(1)}))+c(\mu^{(2)},\mu^{(3)},(\boldsymbol{\mu}^{(2)},\boldsymbol{\nu}^{(2)})).$$
We notice that if $\mu_T^{(1)}\neq\mu^{(2)}$ or $\mu_0^{(2)}\neq\mu^{(2)}$ or $(\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$ are not feasible, there is nothing to prove, since the right hand side is $+\infty$. Thus we assume $\mu_T^{(1)}=\mu^{(2)}=\mu_0^{(2)}$ and feasibility of $(\boldsymbol{\mu}^{(j)},\boldsymbol{\nu}^{(j)})$, $j=1,2$. Let us define $T'=T_1+T_2$ and $(\boldsymbol{\mu}',\boldsymbol{\nu}')=(\boldsymbol{\mu}^{(1)},\boldsymbol{\nu}^{(1)})\star(\boldsymbol{\mu}^{(2)},\boldsymbol{\nu}^{(2)})\in\Sigma$. Thus
$$c(\mu^{(1)},\mu^{(3)},(\boldsymbol{\mu}',\boldsymbol{\nu}'))=T'=T_1+T_2=c(\mu^{(1)},\mu^{(2)},(\boldsymbol{\mu}^{(1)},\boldsymbol{\nu}^{(1)}))+c(\mu^{(2)},\mu^{(3)},(\boldsymbol{\mu}^{(2)},\boldsymbol{\nu}^{(2)})).$$
- Check for $(\boldsymbol{C_2})$. Given $\mu^{(i)}\in X$, $i=1,2$, and $(\boldsymbol{\mu},\boldsymbol{\nu})=(\{\mu_t\}_{[0,T]},\{\nu_t\}_{[0,T]})\in\Sigma$, we have to prove that there exist $\mu'\in X$, $(\boldsymbol{\mu'}^{(i)},\boldsymbol{\nu'}^{(i)})=(\{\mu'^i_t\}_{[0,T'_i]},\{\nu'^i_t\}_{[0,T'_i]})\in\Sigma$, $i=1,2$, such that
$$c(\mu^{(1)},\mu^{(2)},(\boldsymbol{\mu},\boldsymbol{\nu}))\ge c(\mu^{(1)},\mu',(\boldsymbol{\mu'}^{(1)},\boldsymbol{\nu'}^{(1)}))+c(\mu',\mu^{(2)},(\boldsymbol{\mu'}^{(2)},\boldsymbol{\nu'}^{(2)})).$$
Indeed, we have that if $(\boldsymbol{\mu},\boldsymbol{\nu})\notin\mathscr{F}^{\infty}_{[0,T]}(\mu^{(1)})$ or $\mu_{|t=T}\neq\mu^{(2)}$, the result is trivial since the left hand side is $+\infty$. Thus we assume that $(\boldsymbol{\mu},\boldsymbol{\nu})\in\mathscr{F}^{\infty}_{[0,T]}(\mu^{(1)})$ with $\mu_{|t=T}=\mu^{(2)}$. If we take any $\tau\in[0,T]$ and define $\mu'=\mu_\tau$, and $(\boldsymbol{\mu'}^{(1)},\boldsymbol{\nu'}^{(1)})=(\boldsymbol{\mu},\boldsymbol{\nu})_{|[0,\tau]}$ and $(\boldsymbol{\mu'}^{(2)},\boldsymbol{\nu'}^{(2)})=(\boldsymbol{\mu},\boldsymbol{\nu})_{|[\tau,T]}$, this provides the desired inequality.
- Condition $(\boldsymbol{C_3})$ holds immediately by definition of $c(\cdot,\cdot,\cdot)$.
- Condition $(\boldsymbol{C_4})$ follows by additivity of the cost function $c(\cdot,\cdot,\cdot)$, defined above, along feasible trajectories, together with the well posedness of the restriction operation for feasible trajectories.
- Condition $(\boldsymbol{C_5})$ holds with the constant function $h\equiv 1$.

We leave to future research the study of conditions assuring some regularity for the minimum time function in this framework. We refer the reader to [13] for some discussions concerning the case with no feasibility constraints. The study of higher order controllability conditions in this setting is still an open research direction. We address the reader to the

recent issue [15] for a consistent definition of Lie brackets in a measure-theoretic setting which could possibly be used for the study of second order controllability results.

*Remark* 7.2. We observe that, along an optimal trajectory, the minimum time function previously outlined gives the mass' time of last entry into the target set. Alternatively, keeping the same definition of $c_f$, we can give another definition of the cost $c$ leading to a value function that is an *averaged minimum time function*. This is inspired by [19], where the authors provide also a possible example of application in the case with no interactions. Let $S \subseteq \mathbb{R}^d$ be non empty and closed, and $\tilde{S} = \{\mu \in \mathscr{P}_2(\mathbb{R}^d) : \operatorname{supp} \mu \subseteq S\}$, we define

$$c(\mu^{(1)}, \mu^{(2)}, (\boldsymbol{\mu}, \boldsymbol{\nu})) = \begin{cases} \displaystyle\int_0^T \int_{\mathbb{R}^d} \chi_{\mathbb{R}^d \setminus S}(x) d\mu_s(x)\, ds, & \text{if } (\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{F}^{\infty}_{[0,T]}(\mu^{(1)}), \\ & \text{and } \mu_{|t=T} = \mu^{(2)}, \\ +\infty, & \text{otherwise.} \end{cases}$$

We can easily check that properties $(\boldsymbol{C_1}) - (\boldsymbol{C_5})$ holds true with $h(\bar{\mu}) = \bar{\mu}(\mathbb{R}^d \setminus S)$. Thus, if the initial datum $\bar{\mu}$ is such that $\operatorname{supp} \bar{\mu} \subseteq \mathbb{R}^d \setminus S$, then the Hamiltonians for the minimum time function and for the averaged minimum time function coincide. This was also observed in [18] dealing with an averaged minimum-time problem with no feasibility constraints.

## Appendix A. A sparsity constraint in Lagrangian formulation

We will provide here a sparsity constraint in a *Lagrangian* formulation, while the setting presented in Section 3 concerned a sparsity constraint given in an *Eulerian* point of view. To this aim, we state a feasibility constraint on the Carathéodory solutions of the differential inclusion $\dot{\gamma}(t) \in F(\gamma(t))$ as follows

- we consider a notion of *extended characteristic*, by coupling each characteristic $\gamma$ with a time-dependent curve $\zeta(\cdot)$ related to the amount of control needed to generate it;
- we put a feasibility constraint on the probability measures concentrated on this extended notion of characteristics, i.e., we select the measures concentrated on extended curves satisfying a control sparsity constraint.

This microscopic sparsity constraint allows also to select a (possibly not unique) probabilistic representation for feasible trajectories, since it can be used to prescribe the paths to be followed by the microscopic particles. This makes it particulary suitable for applications in irrigation problems or dynamics set on networks [8, 9]. We call this constraint *the $L^\infty$-extended curve-based feasibility condition.*

We first introduce the following sets and operators to deal with continuous curves in the extended space $\mathbb{R}^d \times \mathbb{R}$ and which are used to outline the problem considered in this section. We consider only compact and nonempty intervals $I$ of $\mathbb{R}$. We define

(1) the extended space $\tilde{\Gamma}_I := C^0(I; \mathbb{R}^d \times \mathbb{R})$;
(2) the extended evaluation operator $\tilde{e}_{I,t} : \mathbb{R}^{d+1} \times \tilde{\Gamma}_I \to \mathbb{R}^{d+1}$, $(\tilde{x}, \tilde{\gamma}) \mapsto \tilde{\gamma}(t)$, where we omit the subscript $I$ when it is clear from the context;
(3) given $\hat{I} \subseteq I$ compact and nonempty interval, we define the continuous map called *restriction operator*

$$\begin{aligned} R_{I \to \hat{I}} : \mathbb{R}^{d+1} \times \tilde{\Gamma}_I &\to \mathbb{R}^{d+1} \times \tilde{\Gamma}_{\hat{I}}, \\ (\tilde{x}, \tilde{\gamma}) &\mapsto (\tilde{\gamma}_{|\hat{I}}(\min \hat{I}), \tilde{\gamma}_{|\hat{I}}), \end{aligned}$$

where $\tilde{\gamma}_{|\hat{I}}(t) := \tilde{\gamma}(t)$ for all $t \in \hat{I}$.
(4) given $I_i = [a_i, b_i] \subseteq \mathbb{R}$, $a_2 = b_1$, $I = I_1 \cup I_2$, we set

$$D_{I_1,I_2} := \{(\tilde{x}_1, \tilde{\gamma}_1, \tilde{x}_2, \tilde{\gamma}_2) \in \mathbb{R}^{d+1} \times \tilde{\Gamma}_{I_1} \times \mathbb{R}^{d+1} \times \tilde{\Gamma}_{I_2} : \tilde{x}_i = \tilde{\gamma}_i(a_i),\ \tilde{\gamma}_1(b_1) = \tilde{\gamma}_2(a_2)\},$$

and we define the continuous map called *merge operator*

$$M_{I_1,I_2} : D_{I_1,I_2} \to \mathbb{R}^{d+1} \times \tilde{\Gamma}_I,$$
$$(\tilde{x}_1, \tilde{\gamma}_1, \tilde{x}_2, \tilde{\gamma}_2) \mapsto (\tilde{\gamma}_1 \star \tilde{\gamma}_2(a_1), \tilde{\gamma}_1 \star \tilde{\gamma}_2),$$

where $\tilde{\gamma}_1 \star \tilde{\gamma}_2 \in \tilde{\Gamma}_I$ is defined as $\tilde{\gamma}_1 \star \tilde{\gamma}_2(t) = \tilde{\gamma}_i(t)$ for all $t \in I_i$, $\tilde{\gamma}_i \in \tilde{\Gamma}_{I_i}$, satisfying $\tilde{\gamma}_1(b_1) = \tilde{\gamma}_2(a_2)$, $i = 1, 2$.

We notice that $R_{I\to I_i} \circ M_{I_1,I_2} = \tilde{\pi}_i$, $i = 1, 2$ and $M_{I_1,I_2}(R_{I\to I_1}(\tilde{x}, \tilde{\gamma}), R_{I\to I_2}(\tilde{x}, \tilde{\gamma})) = (\tilde{\gamma}(a_1), \tilde{\gamma})$, where

$$\tilde{\pi}_i : \mathbb{R}^{d+1} \times \tilde{\Gamma}_{I_1} \times \mathbb{R}^{d+1} \times \tilde{\Gamma}_{I_2} \to \mathbb{R}^{d+1} \times \tilde{\Gamma}_{I_i}, \quad (\tilde{x}_1, \tilde{\gamma}_1, \tilde{x}_2, \tilde{\gamma}_2) \mapsto (\tilde{x}_i, \tilde{\gamma}_i).$$

(5) Given $\tilde{x} \in \mathbb{R}^{d+1}$, $I = [a, b]$, we say that the pair $(\tilde{x}, \tilde{\gamma}) \in \mathbb{R}^{d+1} \times \tilde{\Gamma}_I$ satisfies the extended dynamical system **(ES)** in the time interval $I$, if and only if $\tilde{\gamma}$ is an absolutely continuous solution of the following Cauchy problem

$$\begin{cases} \dot{\gamma}(t) \in F(\gamma(t)), & \text{for a.e. } t \in I, \\ \dot{\zeta}(t) = \Psi(\gamma(t), \dot{\gamma}(t)), & \text{for a.e. } t \in I, \\ \gamma(a) = x \\ \zeta(a) = \omega, \end{cases}$$

where $\Psi$ is the control magnitude density and we denoted $\tilde{x} = (x, \omega) \in \mathbb{R}^d \times \mathbb{R}$, $\tilde{\gamma} = (\gamma, \zeta) \in C^0(I; \mathbb{R}^d) \times C^0(I; \mathbb{R})$.

Let $\alpha \geq 0$ be fixed. Considering the notation outlined in Section 4, we take

$$X = \mathscr{P}(\mathbb{R}^d \times \mathbb{R}), \text{ and } \Sigma = \bigcup_{\substack{I \subseteq \mathbb{R} \\ I \text{ compact interval}}} \mathscr{P}(\mathbb{R}^{d+1} \times \tilde{\Gamma}_I).$$

**Definition A.1** ($L^\infty$-extended curve-based feasible set)**.** We define the *extended-curve-based admissibility set* by

$$\tilde{\mathcal{A}}_I := \left\{ \tilde{\boldsymbol{\eta}} \in \mathscr{P}(\mathbb{R}^{d+1} \times \tilde{\Gamma}_I) : \begin{array}{l} \tilde{\boldsymbol{\eta}} \text{ is concentrated on pairs } (\tilde{x}, \tilde{\gamma}) \in \mathbb{R}^{d+1} \times \tilde{\Gamma}_I \\ \text{satisfying } \textbf{(ES)} \end{array} \right\}.$$

Given $\tilde{\mu} \in \mathscr{P}(\mathbb{R}^{d+1})$, we define

$$\tilde{\mathcal{A}}_I(\tilde{\mu}) := \left\{ \tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{A}}_I : \tilde{e}_{\min I} \sharp \tilde{\boldsymbol{\eta}} = \tilde{\mu} \right\}.$$

Given $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{A}}_I$, we define the map $\omega_{\tilde{\boldsymbol{\eta}}} : \mathbb{R}^{d+1} \times \tilde{\Gamma}_I \to [0, +\infty]$ by setting for $\tilde{\boldsymbol{\eta}}$-a.e. $(\tilde{x}, \tilde{\gamma}) \in \mathbb{R}^{d+1} \times \tilde{\Gamma}_I$, with $\tilde{x} = (x, \omega) \in \mathbb{R}^d \times \mathbb{R}$ and $\tilde{\gamma} = (\gamma, \zeta) \in C^0(I; \mathbb{R}^d) \times C^0(I; \mathbb{R})$,

$$\omega_{\tilde{\boldsymbol{\eta}}}(\tilde{x}, \tilde{\gamma}) := \begin{cases} \omega + \displaystyle\int_I \dot{\zeta}(t)\, dt, & \text{if } \dot{\zeta}(\cdot) \in L^1(I); \\ \\ +\infty, & \text{otherwise.} \end{cases}$$

We define the *extended curve-based feasibility set* by

$$\tilde{\mathcal{F}}_I := \left\{ \tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{A}}_I : \|\omega_{\tilde{\boldsymbol{\eta}}}\|_{L^\infty_{\tilde{\boldsymbol{\eta}}}} \leq \alpha \right\}.$$

Given $\tilde{\mu} \in \mathscr{P}(\mathbb{R}^{d+1})$, we define

$$\tilde{\mathcal{F}}_I(\tilde{\mu}) := \tilde{\mathcal{A}}_I(\tilde{\mu}) \cap \tilde{\mathcal{F}}_I.$$

The following definitions and considerations, that we perform for the *extended curve-based feasible set*, are even more valid if, instead, we consider just the *extended curve-based admissibility set*.

**Definition A.2** (Concatenation and restriction)**.**

(1) Let $I_i = [a_i, b_i] \subset \mathbb{R}$, $i = 1, 2$, with $b_1 = a_2$, and $I := I_1 \cup I_2$. Let $\tilde{\boldsymbol{\eta}}^i \in \tilde{\mathcal{F}}_{I_i}$ with $\tilde{e}_{b_1}\sharp\tilde{\boldsymbol{\eta}}^1 = \tilde{e}_{a_2}\sharp\tilde{\boldsymbol{\eta}}^2$. We define the *concatenation*

$$\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2 = \tilde{\mu} \otimes M_{I_1,I_2}\sharp(\tilde{\eta}^1_{\tilde{y}} \otimes \tilde{\eta}^2_{\tilde{y}}),$$

where $\tilde{\mu} := \tilde{e}_{b_1}\sharp\tilde{\boldsymbol{\eta}}^1 = \tilde{e}_{a_2}\sharp\tilde{\boldsymbol{\eta}}^2 \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R})$, and $\{\tilde{\eta}^i_{\tilde{y}}\}_{\tilde{y}\in\mathbb{R}^{d+1}}$ is the measure uniquely defined for $\tilde{\mu}$-a.e. $\tilde{y} \in \mathbb{R}^{d+1}$ by the disintegrations of $\tilde{\boldsymbol{\eta}}^1$ and $\tilde{\boldsymbol{\eta}}^2$ w.r.t. the extended evaluation operators $\tilde{e}_{b_1}$ and $\tilde{e}_{a_2}$, respectively. Notice that for $\tilde{\mu}$-a.e. $\tilde{y} \in \mathbb{R}^{d+1}$ we have $\operatorname{supp}(\tilde{\eta}^1_{\tilde{y}} \otimes \tilde{\eta}^2_{\tilde{y}}) \subseteq D_{I_1,I_2}$ by construction and $\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2 \in \tilde{\mathcal{A}}_I$.

(2) Let $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_I$. The *restriction* $\tilde{\boldsymbol{\eta}}_{|\hat{I}}$ of $\tilde{\boldsymbol{\eta}}$ to a compact and nonempty interval $\hat{I} \subset I$, is defined by setting $\tilde{\boldsymbol{\eta}}_{|\hat{I}} := R_{I\to\hat{I}}\sharp\tilde{\boldsymbol{\eta}}$. By construction, $\tilde{\boldsymbol{\eta}}_{|\hat{I}} \in \tilde{\mathcal{A}}_{\hat{I}}$.

We provide here an example showing the result of the concatenation operation in an illustrative situation.

*Example* A.3. Considering the notation of Definition 3.1, let $d = 2$, $m = 1$, $U = [-1, 1]$ and $f_0(x) = (1, 0) \in \mathbb{R}^2$, $f_1(x) = (0, 1) \in \mathbb{R}^2$ for all $x \in \mathbb{R}^2$. Thus, $F(x) = \{(1, u) \in \mathbb{R}^2 : u \in U\}$. Let us fix the control sparsity threshold $\alpha = 2$.

In $\mathbb{R}^2$, consider $x_j = (0, j)$, and $\gamma_{x_j} : [0, 2] \to \mathbb{R}^2$, $\gamma_{x_j}(t) = (t, j - t\operatorname{sgn} j)$, for $j = \pm 1$. Let $\tilde{\gamma}_{\tilde{x}_j} = (\gamma_{x_j}, \zeta) \in C^0([0, 2]; \mathbb{R}^2) \times C^0([0, 2]; \mathbb{R})$ be a solution of system **(ES)** in the time-interval $[0, 2]$, starting from $\tilde{x}_j = (x_j, \omega) \in \mathbb{R}^2 \times \mathbb{R}$, with $\omega = 0$. Thus, we have that $\zeta(t) = t$, since $\Psi(x, (1, 1)) = \Psi(x, (1, -1)) = 1$ for all $x \in \mathbb{R}^2$.

Let us define

$$\tilde{\boldsymbol{\eta}}^1 = \frac{1}{2}\sum_{j\in\{-1,1\}} \delta_{\tilde{x}_j} \otimes \delta_{\tilde{\gamma}_{\tilde{x}_j|[0,1]}} \in \mathscr{P}(\mathbb{R}^{2+1} \times \tilde{\Gamma}_{[0,1]}), \quad \tilde{\boldsymbol{\eta}}^2 = \frac{1}{2}\sum_{j\in\{-1,1\}} \delta_{\tilde{\gamma}_{\tilde{x}_j}(1)} \otimes \delta_{\tilde{\gamma}_{\tilde{x}_j|[1,2]}} \in \mathscr{P}(\mathbb{R}^{2+1} \times \tilde{\Gamma}_{[1,2]}).$$

We see that $\tilde{e}_1\sharp\tilde{\boldsymbol{\eta}}^1 = \tilde{e}_1\sharp\tilde{\boldsymbol{\eta}}^2 = \delta_{\tilde{\gamma}_{\tilde{x}_j}(1)} = \delta_{((1,0),1)} \in \mathscr{P}(\mathbb{R}^{2+1})$. Now, consider also $\xi_{\tilde{x}_j} \in C^0([0, 2]; \mathbb{R}^{2+1})$,

$$\xi_{\tilde{x}_j}(t) := \begin{cases} \tilde{\gamma}_{\tilde{x}_j}(t), & \text{for } t \in [0, 1], \\ \tilde{\gamma}_{\tilde{x}_i}(t), & \text{for } t \in [1, 2],\ i \in \{-1, 1\},\ i \neq j. \end{cases}$$

We can finally compute the concatenation $\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2$ which leads to

$$\tilde{\boldsymbol{\eta}} := \tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2 = \frac{1}{4}\sum_{j\in\{-1,1\}} \delta_{\tilde{x}_j} \otimes \delta_{\tilde{\gamma}_{\tilde{x}_j}} + \frac{1}{4}\sum_{j\in\{-1,1\}} \delta_{\tilde{x}_j} \otimes \delta_{\xi_{\tilde{x}_j}} \in \mathscr{P}(\mathbb{R}^{2+1} \times \tilde{\Gamma}_{[0,2]}).$$

Moreover, we can prove that $\tilde{\boldsymbol{\eta}}^1, \tilde{\boldsymbol{\eta}}^2$ and $\tilde{\boldsymbol{\eta}}$ are admissible and feasible in the respective time intervals, according to Definition A.1.

**Proposition A.4.** *Let $I \subset \mathbb{R}$ be a compact and nonempty interval. Given $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_I$, the following restrictions' properties hold.*

*(i) For any $\hat{I} \subseteq I$ compact and nonempty interval, we have $\tilde{\boldsymbol{\eta}}_{|\hat{I}} \in \tilde{\mathcal{F}}_{\hat{I}}$ and $\tilde{e}_t\sharp\tilde{\boldsymbol{\eta}} = \tilde{e}_t\sharp\tilde{\boldsymbol{\eta}}_{|\hat{I}}$ for any $t \in \hat{I}$.*

*(ii) For any $I \supseteq I_2 \supseteq I_1$ compact and nonempty intervals, we have $\left(\tilde{\boldsymbol{\eta}}_{|I_2}\right)_{|I_1} \equiv \tilde{\boldsymbol{\eta}}_{|I_1}$.*

*(iii) $\tilde{\boldsymbol{\eta}} \equiv \tilde{\boldsymbol{\eta}}_{|I}$.*

*Furthermore, we have the following property for concatenations.*

*(iv) Let $I_i = [a_i, b_i]$, $i = 1, 2$, with $b_1 = a_2$ and $I_1 \cup I_2 = I$ and $\tilde{\boldsymbol{\eta}}^i \in \tilde{\mathcal{F}}_{I_i}$, $i = 1, 2$, with $\tilde{e}_{b_1}\sharp\tilde{\boldsymbol{\eta}}^1 = \tilde{e}_{a_2}\sharp\tilde{\boldsymbol{\eta}}^2$. Then $\tilde{\boldsymbol{\eta}} := \tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2 \in \tilde{\mathcal{F}}_I$ and we have*

$$\tilde{e}_t\sharp\tilde{\boldsymbol{\eta}} = \begin{cases} \tilde{e}_t\sharp\tilde{\boldsymbol{\eta}}^1, & \text{if } t \in I_1, \\ \tilde{e}_t\sharp\tilde{\boldsymbol{\eta}}^2, & \text{if } t \in I_2. \end{cases}$$

*Proof.* By construction, observe that $\tilde{\boldsymbol{\eta}}_{|\hat{I}} \in \tilde{\mathcal{A}}_{\hat{I}}$.

*Proof of* $(i)$. First, we want to prove that $\tilde{\boldsymbol{\eta}}_{|\hat{I}} \in \tilde{\mathcal{F}}_{\hat{I}}$. By contradiction, let $A \subseteq \mathbb{R}^{d+1} \times \tilde{\Gamma}_{\hat{I}}$ be a Borel set of strictly positive $\tilde{\boldsymbol{\eta}}_{|\hat{I}}$-measure such that the feasibility condition is violated. In particular, we have

$$\int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{\hat{I}}} \chi_A(\tilde{x}_1,\tilde{\gamma}_1)\left[\omega_1 + \int_{\hat{I}} \Psi(\gamma_1(t),\dot{\gamma}_1(t))\,dt\right] d\tilde{\boldsymbol{\eta}}_{|\hat{I}}(\tilde{x}_1,\tilde{\gamma}_1) > \alpha\tilde{\boldsymbol{\eta}}_{|\hat{I}}(A),$$

where we denote with $\tilde{x}_1 = (x_1,\omega_1)$, $\tilde{\gamma}_1 = (\gamma_1,\zeta_1)$. Now by definition of $\tilde{\boldsymbol{\eta}}_{|\hat{I}}$, the left-hand side can be rewritten as follows

$$\int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \chi_A(\tilde{\gamma}_{|\hat{I}}(\min\hat{I}),\tilde{\gamma}_{|\hat{I}})\left[\zeta_{|\hat{I}}(\min\hat{I}) + \int_{\hat{I}} \Psi(\gamma_{|\hat{I}}(t),\dot{\gamma}_{|\hat{I}}(t))\,dt\right] d\tilde{\boldsymbol{\eta}}(\tilde{x},\tilde{\gamma}) \le \alpha\tilde{\boldsymbol{\eta}}_{|\hat{I}}(A),$$

where we denoted with $\tilde{x} = (x,\omega)$, $\tilde{\gamma} = (\gamma,\zeta)$, hence the contradiction. Indeed, we have that for $\tilde{\boldsymbol{\eta}}$-a.e. $(\tilde{x},\tilde{\gamma})$

$$\zeta_{|\hat{I}}(\min\hat{I}) + \int_{\hat{I}} \Psi(\gamma_{|\hat{I}}(t),\dot{\gamma}_{|\hat{I}}(t))\,dt = \omega + \int_{\min I}^{\min\hat{I}} \Psi(\gamma(t),\dot{\gamma}(t))\,dt + \int_{\hat{I}} \Psi(\gamma(t),\dot{\gamma}(t))\,dt \le \alpha,$$

by feasibility of $\tilde{\boldsymbol{\eta}}$.

Let us now prove that for any $t \in \hat{I}$, $\tilde{e}_t\sharp\tilde{\boldsymbol{\eta}} = \tilde{e}_t\sharp\tilde{\boldsymbol{\eta}}_{|\hat{I}}$. Consider any test function $\varphi \in C_b^0(\mathbb{R}^{d+1};\mathbb{R})$, $t \in \hat{I}$. We have

$$\begin{aligned}
\int_{\mathbb{R}^{d+1}} \varphi(\tilde{x}_1)\,d(\tilde{e}_t\sharp\tilde{\boldsymbol{\eta}}_{|\hat{I}})(\tilde{x}_1) &= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{\hat{I}}} \varphi(\tilde{\gamma}_1(t))\,d\tilde{\boldsymbol{\eta}}_{|\hat{I}}(\tilde{x}_1,\tilde{\gamma}_1)\\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \varphi(\tilde{\gamma}_{|\hat{I}}(t))\,d\tilde{\boldsymbol{\eta}}(\tilde{x},\tilde{\gamma})\\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \varphi(\tilde{\gamma}(t))\,d\tilde{\boldsymbol{\eta}}(\tilde{x},\tilde{\gamma})\\
&= \int_{\mathbb{R}^{d+1}} d(\tilde{e}_t\sharp\tilde{\boldsymbol{\eta}})(\tilde{x}).
\end{aligned}$$

*Proof of* $(ii)$. For any test function $\varphi \in C_b^0(\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I_1})$,

$$\begin{aligned}
\int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I_1}} \varphi(\tilde{x}_1,\tilde{\gamma}_1)\,d\left(\tilde{\boldsymbol{\eta}}_{|I_2}\right)_{|I_1}(\tilde{x}_1,\tilde{\gamma}_1) &= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I_2}} \varphi(\tilde{\gamma}_{|I_1}(\min I_1),\tilde{\gamma}_{|I_1})\,d\tilde{\boldsymbol{\eta}}_{|I_2}(\tilde{x},\tilde{\gamma})\\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \varphi(\tilde{\gamma}_{|I_{1|I_2}}(\min I_1),\tilde{\gamma}_{|I_{1|I_2}})\,d\tilde{\boldsymbol{\eta}}(\tilde{x},\tilde{\gamma})\\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \varphi(\tilde{\gamma}_{|I_1}(\min I_1),\tilde{\gamma}_{|I_1})\,d\tilde{\boldsymbol{\eta}}(\tilde{x},\tilde{\gamma})\\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I_1}} \varphi(\tilde{x}_1,\tilde{\gamma}_1)\,d\tilde{\boldsymbol{\eta}}_{|I_1}(\tilde{x}_1,\tilde{\gamma}_1).
\end{aligned}$$

*Proof of* $(iii)$. By definition of $\tilde{\boldsymbol{\eta}}_{|I}$, for any test function $\varphi \in C_b^0(\mathbb{R}^{d+1}\times\tilde{\Gamma}_I;\mathbb{R})$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \varphi(\tilde{x}_1,\tilde{\gamma}_1)\,\tilde{\boldsymbol{\eta}}_{|I}(\tilde{x}_1,\tilde{\gamma}_1) =\\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \varphi(\tilde{\gamma}_{|I}(\min I),\tilde{\gamma}_{|I})\,d\tilde{\boldsymbol{\eta}}(\tilde{x},\tilde{\gamma})\\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I}} \varphi(\tilde{x},\tilde{\gamma})\,d\tilde{\boldsymbol{\eta}}(\tilde{x},\tilde{\gamma}).
\end{aligned}$$

*Proof of* $(iv)$. First, we prove that $\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2 \in \tilde{\mathcal{F}}_I$. Let us denote with $\tilde{\mu} := \tilde{e}_{b_1} \sharp \tilde{\boldsymbol{\eta}}^1 = \tilde{e}_{a_2} \sharp \tilde{\boldsymbol{\eta}}^2$. By contradiction, let $A \subseteq \mathbb{R}^{d+1} \times \tilde{\Gamma}_I$ be a Borel set of strictly positive measure w.r.t. $\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2$ such that the feasibility condition is not respected, i.e.

$$\int_{\mathbb{R}^{d+1} \times \tilde{\Gamma}_I} \chi_A(\tilde{x}, \tilde{\gamma}) \left[ \omega + \int_I \Psi(\gamma(t), \dot{\gamma}(t))\, dt \right] d(\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2)(\tilde{x}, \tilde{\gamma}) > \alpha \cdot \tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2(A),$$

where we denote with $\tilde{x} = (x, \omega)$, $\tilde{\gamma} = (\gamma, \zeta)$. By definition of concatenation, the left-hand side is equivalent to

$$\begin{aligned}
&\int_{\mathbb{R}^{d+1}} \int_{\tilde{e}^{-1}_{I_2,t}(\tilde{y})} \int_{\tilde{e}^{-1}_{I_1,t}(\tilde{y})} \chi_A(\tilde{\gamma}_1 \star \tilde{\gamma}_2(a_1), \tilde{\gamma}_1 \star \tilde{\gamma}_2) \cdot \\
&\qquad \cdot \left[ \zeta_1 \star \zeta_2(a_1) + \int_I \Psi\left( \gamma_1 \star \gamma_2(t), \frac{d}{dt}(\gamma_1 \star \gamma_2)(t) \right) dt \right] d\tilde{\eta}^1_{\tilde{y}}(\tilde{x}_1, \tilde{\gamma}_1)\, d\tilde{\eta}^2_{\tilde{y}}(\tilde{x}_2, \tilde{\gamma}_2)\, d\tilde{\mu}(\tilde{y}) \\
&= \int_{\mathbb{R}^{d+1}} \int_{\tilde{e}^{-1}_{I_2,t}(\tilde{y})} \int_{\tilde{e}^{-1}_{I_1,t}(\tilde{y})} \chi_A(\tilde{\gamma}_1 \star \tilde{\gamma}_2(a_1), \tilde{\gamma}_1 \star \tilde{\gamma}_2) \cdot \\
&\qquad \cdot \left[ \omega_1 + \int_{I_1} \Psi(\gamma_1(t), \dot{\gamma}_1(t))\, dt + \int_{I_2} \Psi(\gamma_2(t), \dot{\gamma}_2(t))\, dt \right] d\tilde{\eta}^1_{\tilde{y}}(\tilde{x}_1, \tilde{\gamma}_1)\, d\tilde{\eta}^2_{\tilde{y}}(\tilde{x}_2, \tilde{\gamma}_2)\, d\tilde{\mu}(\tilde{y}) \\
&= \int_{\mathbb{R}^{d+1}} \int_{\tilde{e}^{-1}_{I_2,t}(\tilde{y})} \left[ \omega_2 + \int_{I_2} \Psi(\gamma_2(t), \dot{\gamma}_2(t))\, dt \right] \cdot \\
&\qquad \cdot \int_{\tilde{e}^{-1}_{I_1,t}(\tilde{y})} \chi_A(\tilde{\gamma}_1 \star \tilde{\gamma}_2(a_1), \tilde{\gamma}_1 \star \tilde{\gamma}_2)\, d\tilde{\eta}^1_{\tilde{y}}(\tilde{x}_1, \tilde{\gamma}_1)\, d\tilde{\eta}^2_{\tilde{y}}(\tilde{x}_2, \tilde{\gamma}_2)\, d\tilde{\mu}(\tilde{y}) \\
&\leq \alpha \cdot \tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2(A),
\end{aligned}$$

by feasibility of $\tilde{\boldsymbol{\eta}}^2$, where we denoted with $\tilde{x}_i = (x_i, \omega_i)$, $\tilde{\gamma}_i = (\gamma_i, \zeta_i)$. Hence the contradiction.

With the same notation as before, let us now prove the other assertion. For any test function $\varphi \in C^0_b(\mathbb{R}^{d+1}; \mathbb{R})$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^{d+1}} \varphi(\tilde{x})\, d(\tilde{e}_t \sharp (\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2))(\tilde{x}) = \\
&= \int_{\mathbb{R}^{d+1} \times \tilde{\Gamma}_I} \varphi(\tilde{\gamma}(t))\, d(\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2)(\tilde{x}, \tilde{\gamma}) \\
&= \int_{\mathbb{R}^{d+1}} \int_{\mathbb{R}^{d+1} \times \tilde{\Gamma}_I} \varphi(\tilde{\gamma}(t))\, d[M_{I_1,I_2} \sharp (\tilde{\eta}^1_{\tilde{y}} \otimes \tilde{\eta}^2_{\tilde{y}})](\tilde{x}, \tilde{\gamma})\, d\tilde{\mu}(\tilde{y}) \\
&= \int_{\mathbb{R}^{d+1}} \int_{\tilde{e}^{-1}_{I_1,t}(\tilde{y}) \times \tilde{e}^{-1}_{I_2,t}(\tilde{y})} \varphi(\tilde{\gamma}_1 \star \tilde{\gamma}_2(t))\, d(\tilde{\eta}^1_{\tilde{y}} \otimes \tilde{\eta}^2_{\tilde{y}})(\tilde{x}_1, \tilde{\gamma}_1, \tilde{x}_2, \tilde{\gamma}_2)\, d\tilde{\mu}(\tilde{y}) \\
&= \begin{cases} \tilde{e}_t \sharp \tilde{\boldsymbol{\eta}}^1, & \text{if } t \in I_1, \\ \tilde{e}_t \sharp \tilde{\boldsymbol{\eta}}^2, & \text{if } t \in I_2. \end{cases}
\end{aligned}$$

□

In the sequel, we give a further property verified by the given definitions of concatenation and restriction, which turn out to be weakly compatible.

**Proposition A.5.** *Let $I, I_1, I_2 \subset \mathbb{R}$ be compact and nonempty intervals, with $I = I_1 \cup I_2$, $I_i = [a_i, b_i]$, $i = 1, 2$ and $b_1 = a_2$. Then the restriction and concatenation operators are weakly compatible, i.e. given $\tilde{\boldsymbol{\eta}}^i \in \tilde{\mathcal{F}}_{I_i}$, $i = 1, 2$ such that $\tilde{e}_{b_1} \sharp \tilde{\boldsymbol{\eta}}^1 = \tilde{e}_{a_2} \sharp \tilde{\boldsymbol{\eta}}^2$, we have $(\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2)_{|I_i} = \tilde{\boldsymbol{\eta}}^i$, $i = 1, 2$.*

*Proof.* By definition of restriction and concatenation operators, for any $i = 1, 2$ and for all $\varphi \in C^0(\mathbb{R}^{d+1} \times \tilde{\Gamma}_{I_i}; \mathbb{R})$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^{d+1}\tilde{\Gamma}_{I_i}} \varphi(\tilde{x}_i, \tilde{\gamma}_i)\, d(\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2)_{|I_i}(\tilde{x}_i, \tilde{\gamma}_i) = \\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_I} \varphi(\tilde{\gamma}_{|I_i}(a_i), \tilde{\gamma}_{|I_i})\, d(\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2)(\tilde{x}, \tilde{\gamma}) \\
&= \int_{\mathbb{R}^{d+1}} \int_{\tilde{e}^{-1}_{I_1,t}(\tilde{y})\times\tilde{e}^{-1}_{I_2,t}(\tilde{y})} \varphi\left((\tilde{\gamma}_1 \star \tilde{\gamma}_2)_{|I_i}(a_i), (\tilde{\gamma}_1 \star \tilde{\gamma}_2)_{|I_i}\right)\, d(\tilde{\eta}^1_{\tilde{y}} \otimes \tilde{\eta}^2_{\tilde{y}})(\tilde{x}_1, \tilde{\gamma}_1, \tilde{x}_2, \tilde{\gamma}_2)\, d\tilde{\mu}(\tilde{y}) \\
&= \int_{\mathbb{R}^{d+1}\times\tilde{\Gamma}_{I_i}} \varphi(\tilde{\gamma}_i(a_i), \tilde{\gamma}_i)\, d\tilde{\boldsymbol{\eta}}^i(\tilde{x}_i, \tilde{\gamma}_i).
\end{aligned}$$

□

We notice that the properties just proved hold also for the settings described in Sections 5.1 and 5.2 in a straightforward way.

*Remark* A.6. As the following example shows, a stronger compatibility relation between restriction and concatenation operators is not true in general in this context. Indeed, let $I, I_1, I_2 \subset \mathbb{R}$ be compact and nonempty intervals, with $I = I_1 \cup I_2$, $I_i = [a_i, b_i]$, $i = 1, 2$ and $b_1 = a_2$, and $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{A}}_I$. Then, in general, we cannot write $\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_{|I_1} \star \tilde{\boldsymbol{\eta}}_{|I_2}$.

*Example* A.7. Consider the same framework outlined in Example A.3, with $I = I_1 \cup I_2$, $I_1 = [0, 1]$, $I_2 = [1, 2]$. Let us define

$$\boldsymbol{\rho} := \frac{1}{2} \sum_{j\in\{-1,1\}} \delta_{\tilde{x}_j} \otimes \delta_{\tilde{\gamma}_{\tilde{x}_j}} \in \tilde{\mathcal{A}}_{[0,2]}, \quad \boldsymbol{\xi} := \frac{1}{2} \sum_{j\in\{-1,1\}} \delta_{\tilde{x}_j} \otimes \delta_{\xi_{\tilde{x}_j}} \in \tilde{\mathcal{A}}_{[0,2]}.$$

We have that $\boldsymbol{\rho}_{|I_i} = \boldsymbol{\xi}_{|I_i} = \tilde{\boldsymbol{\eta}}_{|I_i} = \tilde{\boldsymbol{\eta}}^i$, $i = 1, 2$. Nevertheless, both $\boldsymbol{\rho}$ and $\boldsymbol{\xi}$ are different from $\tilde{\boldsymbol{\eta}}^1 \star \tilde{\boldsymbol{\eta}}^2 = \tilde{\boldsymbol{\eta}}$.

This construction provides another situation where the abstract Dynamic Programming Principle of Theorem 4.2 holds true (see Theorem A.9). Furthermore, with the following definition we can reconduct these objects to that of Definition 4.3, thus gaining the validity of the general results proved in Section 4.

**Definition A.8.** Let $c$ and $c_f$ be as in Definition 4.1, satisfying the following additional properties

- **($C''_3$)** $c(\tilde{\mu}^{(1)}, \tilde{\mu}^{(2)}, \tilde{\boldsymbol{\eta}}) < +\infty$ if and only if $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_I(\tilde{\mu}^{(1)})$, with $\tilde{e}_{\max I}\sharp\tilde{\boldsymbol{\eta}} = \tilde{\mu}^{(2)}$ for some compact and nonempty interval $I \subset \mathbb{R}$;
- **($C''_4$)** let $0 \le a \le b \le c$, $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_{[a,c]}$. Then $c : X \times X \times \Sigma \to [0, +\infty]$ is superadditive by restrictions, i.e.

$$c(\tilde{e}_a\sharp\tilde{\boldsymbol{\eta}}, \tilde{e}_c\sharp\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\eta}}) \ge c(\tilde{e}_a\sharp\tilde{\boldsymbol{\eta}}, \tilde{e}_b\sharp\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\eta}}_{|[a,b]}) + c(\tilde{e}_b\sharp\tilde{\boldsymbol{\eta}}, \tilde{e}_c\sharp\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\eta}}_{|[b,c]}).$$

The finiteness of each member follows from item $(i)$ in Proposition A.4.

Let $\tilde{\mu} \in \mathscr{P}(\mathbb{R}^{d+1})$, $I \subset \mathbb{R}$ nonempty and compact interval, and $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_I(\tilde{\mu})$. We define the set $\mathscr{G}^{\tilde{\mathcal{F}}}_{\tilde{\boldsymbol{\eta}}}$ made of the pairs $(\xi, \sigma)$ defined as follows

(1) $\xi : I \to X$, $\xi(t) := \tilde{e}_t\sharp\tilde{\boldsymbol{\eta}}$ for all $t \in I$;
(2) $\sigma : I \to \Sigma$, $\sigma(t) := \tilde{\boldsymbol{\eta}}_{|[\min I, t]}$ for all $t \in I$.

Finally, we define the set

$$\mathscr{G}^{\tilde{\mathcal{F}}}_I(\tilde{\mu}) := \left\{(\xi, \sigma) \in \mathscr{G}^{\tilde{\mathcal{F}}}_{\tilde{\boldsymbol{\eta}}} \,:\, \tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_I(\tilde{\mu})\right\}.$$

**Theorem A.9** (DPP for the Lagrangian sparsity case)**.** *Let $V : \mathscr{P}(\mathbb{R}^d \times \mathbb{R}) \to [0, +\infty]$ be as in Definition 4.1. For any $\tilde{\mu}_0 \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R})$ we have*

$$V(\tilde{\mu}_0) = \inf_{\substack{\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_I(\tilde{\mu}_0) \\ I \subseteq \mathbb{R} \text{ compact interval}}} \left\{ c(\tilde{\mu}_0, \tilde{e}_{\max I} \sharp \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\eta}}) + V(\tilde{e}_{\max I} \sharp \tilde{\boldsymbol{\eta}}) \right\}.$$

*Proof.* The proof follows by Theorem 4.2 and $\boldsymbol{(C_3'')}$. □

The following remark, mentioned also for the feasibility cases analyzed in Section 5, holds also in this setting.

*Remark* A.10. Any $(\xi, \sigma) \in \mathscr{G}^{\tilde{\mathcal{F}}}_{\tilde{\boldsymbol{\eta}}}$, s.t. $\tilde{\boldsymbol{\eta}} \in \tilde{\mathcal{F}}_I(\tilde{\mu})$, is generalized admissible from $\tilde{\mu}$, according to Definition 4.3. Indeed, item (1) in Definition 4.3 is obviously verified, while item (2) comes from item $(i)$ of Proposition A.4. Finally, by properties $(i) - (ii)$ proved in Proposition A.4, we have that $\boldsymbol{(C_4'')}$ implies item (3) in Definition 4.3 by taking for instance $\sigma_{t_1 \to t_2} := \tilde{\boldsymbol{\eta}}_{|[t_1,t_2]}$ for any $0 \le t_1 \le t_2 \le T$. Hence, Corollaries 4.6, 4.7 hold in this setting. Furthermore, by $\boldsymbol{(C_4'')}$ we have that $c(\xi(t), \xi(t), \sigma_{t \to t}) = 0$ for all $t \in I$. Indeed, by item $(iii)$ in Proposition A.4 we can indentify any $\hat{\rho} \in \mathcal{F}^{\infty}_{[t,t]}$ with its restriction $\hat{\rho}_{|[t,t]}$.

## Appendix B. Basic estimates

In this appendix section, we recall some estimates used throughout the paper, borrowed from [13, 16] (Lemma 2 and Lemma 3.1 respectively).

**Proposition B.1.** *Let $T > 0$, $p \ge 1$, $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$. Let $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{A}_{[0,T]}(\mu_0)$, and let $\boldsymbol{\eta}$ be a representation for the pair $(\boldsymbol{\mu}, \boldsymbol{\nu})$. Then, for $t \in ]0, T]$, we have*

- $(i)$ $e_0 \in L^p_{\boldsymbol{\eta}}$;
- $(ii)$ $\left\| \dfrac{e_t - e_o}{t} \right\|^p_{L^p_{\boldsymbol{\eta}}} \le 2^{p-1} D e^{DT} \left(1 + \mathrm{m}_p(\mu_0)\right) < +\infty$;
- $(iii)$ $\mathrm{m}_p(\mu_t) \le K \left(1 + \mathrm{m}_p(\mu_0)\right)$,

*where $D > 0$ is coming from Lemma 3.2 and $K = K(p, D, T) > 0$.*

*Proof.* First, recall that $(a + b)^p \le 2^{p-1}(a^p + b^p)$ for any $a, b \ge 0$.
Item $(i)$ is immediate, indeed $\|e_0\|^p_{L^2_p} = \int_{\mathbb{R}^d \times \Gamma_T} |\gamma(0)|^p \, d\boldsymbol{\eta} = \int_{\mathbb{R}^d} |x|^p \, d(e_0 \sharp \boldsymbol{\eta})(x) = \mathrm{m}_p(\mu_0) < +\infty$. Let us prove $(ii)$. By admissibility of $(\boldsymbol{\mu}, \boldsymbol{\nu})$, we have that $v_t(x) = \frac{\nu_t}{\mu_t}(x) \in F(x)$. By Lemma 3.2, there exists $D > 0$ such that $|v_t(y)| \le D(1 + |y|)$, $y \in \mathbb{R}^d$. Thus, for $\boldsymbol{\eta}$-a.e. $(x, \gamma)$ we have

$$\begin{aligned} |\gamma(t) - \gamma(0)| &\le \left| \int_0^t |\dot{\gamma}(s)| \, ds \right| \le Dt + D \left| \int_0^t |\gamma(s)| \, ds \right| \\ &\le Dt(1 + |\gamma(0)|) + D \left| \int_0^t |\gamma(s) - \gamma(0)| \, ds \right|. \end{aligned}$$

By Gronwall inequality,

$$|\gamma(t) - \gamma(0)| \le Dt \, e^{Dt} \left(1 + |\gamma(0)|\right).$$

Dividing by $t \in ]0, T]$ and taking the $L^p_{\boldsymbol{\eta}}$-norm, we have

$$\begin{aligned} \left\| \frac{e_t - e_0}{t} \right\|^p_{L^p_{\boldsymbol{\eta}}} &\le D e^{DT} \left( \int_{\mathbb{R}^d \times \Gamma_T} (1 + |\gamma(0)|)^p \, d\boldsymbol{\eta} \right) \\ &\le 2^{p-1} D \, e^{DT} \left(1 + \mathrm{m}_p(\mu_0)\right) < +\infty. \end{aligned}$$

To prove $(iii)$, notice that

$$\mathrm{m}_p(\mu_t) = \int_{\mathbb{R}^d} |x|^p \, d\mu_t = \int_{\mathbb{R}^d \times \Gamma_T} |e_t(x, \gamma)|^p \, d\boldsymbol{\eta} = \|e_t\|^p_{L^p_{\boldsymbol{\eta}}},$$

and

$$\|e_t\|^p_{L^p_{\boldsymbol{\eta}}} \leq \left(\|e_0\|_{L^p_{\boldsymbol{\eta}}} + \|e_t - e_0\|_{L^p_{\boldsymbol{\eta}}}\right)^p$$
$$\leq 2^{p-1}\left(\|e_0\|^p_{L^p_{\boldsymbol{\eta}}} + \|e_t - e_0\|^p_{L^p_{\boldsymbol{\eta}}}\right).$$

We conclude by using the estimate in $(ii)$. □

## Acknowledgments

The authors acknowledge the partial support of the NSF Project *Kinetic description of emerging challenges in multiscale problems of natural sciences*, DMS Grant # 1107444 and the endowment fund of the Joseph and Loretta Lopez Chair.
This work has been partially supported by the project of the Italian Ministry of Education, Universities and Research (MIUR) "Dipartimenti di Eccellenza 2018-2022".
G.C. has been supported by Cariplo foundation and Regione Lombardia through the project 2016-2018 "Variational evolution problems and optimal transport". G.C. thanks the Department of Mathematical Sciences of Rutgers University - Camden (U.S.A.).

## References

[1] Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. The Clarendon Press Oxford University Press, New York (2000).
[2] Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich, 2nd ed. Birkhäuser Verlag, Basel (2008).
[3] Aubin, J.-P., Frankowska, H.: Set-valued analysis. Birkhäuser Boston Inc., Boston, MA (2009).
[4] Averboukh, Y.: Viability theorem for deterministic mean field type control systems. Set-Valued Var. Anal, vol. 26, n. 4, pp. 993–1008 (2018).
[5] Aydoğdu A., Caponigro, M., McQuade, S., Piccoli, B., Pouradier Duteil, N., Rossi, F., Trélat, E.: Interaction Network, State Space, and Control in Social Dynamics. Bellomo N., Degond P., Tadmor E. (eds) Active Particles, vol. 1. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser, Cham (2017).
[6] Bonnet, B., Rossi, F.: The Pontryagin Maximum Principle in the Wasserstein Space. Calc. Var. (2019). Published online.
[7] Buttazzo, G.: Semicontinuity, relaxation and integral representation in the calculus of variations. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York 207 (1989).
[8] Camilli, F., De Maio, R., Tosin, A.: Measure-valued solutions to nonlocal transport equations on networks. J. Differential Equations, vol. 264, n. 12, pp. 7213–7241 (2018).
[9] Camilli, F., De Maio, R., Tosin, A.: Transport of measures on networks. Netw. Heterog. Media, vol. 12, n. 2, pp. 191–215 (2017).
[10] Caponigro, M., Fornasier, M., Piccoli, B., Trélat, E.: Sparse stabilization and control of alignment models. Math. Models and Methods in Appl. Sci., vol. 25, n. 3, pp. 521–564 (2015).
[11] Cardaliaguet, P.: Notes on Mean Field Games (from P.-L. Lions' lectures at Collège de France). http://www.college-de-france.fr, (2013).
[12] Cardaliaguet, P., Quincampoix, M.: Deterministic differential games under probability knowledge of initial condition. Int. Game Theor. Rev., vol. 10, pp. 1–16 (2008).
[13] Cavagnari, G.: Regularity results for a time-optimal control problem in the space of probability measures. Mathematical Control and Related Fields, vol. 7, n. 2, pp. 213–233 (2017) .
[14] Cavagnari, G., Marigonda, A.: Time-optimal control problem in the space of probability measures. Large-scale scientific computing, Lecture Notes in Computer Science, vol. 9374, pp. 109–116. Springer, Cham (2015).
[15] Cavagnari, G., Marigonda, A.: Measure-theoretic Lie brackets for nonsmooth vector fields. Discrete & Continuous Dynamical Systems - S, vol. 11, n. 5, pp. 845–864 (2018).
[16] Cavagnari, G., Marigonda, A., Nguyen, K.T., Priuli, F.S.: Generalized control systems in the space of probability measures. Set-Valued Var. Anal, vol. 26, n. 3, pp. 663–691 (2018).
[17] Cavagnari, G., Marigonda, A., Orlandi, G.: Hamilton-Jacobi-Bellman equation for a time-optimal control problem in the space of probability measures. System Modeling and Optimization: 27th IFIP TC 7 Conference, CSMO 2015, Sophia Antipolis, France, June 29 - July 3, 2015, Revised Selected Papers, vol. 494, pp. 200–208. Springer, Cham (2016).

[18] Cavagnari, G., Marigonda, A., Piccoli, B.: Averaged time-optimal control problem in the space of positive Borel measures. ESAIM: COCV, vol. 24, n. 2, pp. 721–740 (2018).
[19] Cavagnari, G., Marigonda, A., Piccoli, B.: Optimal synchronization problem for a multi-agent system. Network and Heterogeneous Media, vol. 12, n. 2, pp. 277–295 (2017).
[20] Cavagnari, G., Marigonda, A., Piccoli, B.: Superposition principle for differential inclusions. I. Lirkov, S. Margenov (Eds.). Large-Scale Scientific Computing, LSSC 2017, Lecture Notes in Computer Science, vol. 10665, pp. 201–209. Springer, Cham (2018).
[21] Cristiani, E., Piccoli, B. and Tosin, A.: Multiscale modeling of pedestrian dynamics. MS&A. Modeling, Simulation and Applications, vol. 12, Springer, Cham (2014).
[22] Dolbeault, J., Nazaret, B., Savaré, G.: A new class of transport distances between measures. Calc. Var. Partial Dif. 34, pp. 193–231 (2009).
[23] Fornasier, M., Lisini, S., Orrieri, C., Savaré, G.: Mean-field optimal control as Gamma-limit of finite agent controls. Eur. J. Appl. Math. (2019). Published online.
[24] Fornasier, M., Piccoli, B., Rossi, F.: Mean-field sparse optimal control. Philosophical Transactions of the Royal Society A, The Royal Society (2014).
[25] Gangbo, W., Nguyen, T., Tudorascu, A.: Hamilton-Jacobi Equations in the Wasserstein Space. Methods Appl. Anal., vol. 15, n. 2, pp. 155–184 (2008).
[26] Gangbo, W., Tudorascu, A.: On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. J. Math. Pures Appl, vol. 125, pp. 119–174 (2018).
[27] Marigonda, A., Quincampoix, M.: Mayer control problem with probabilistic uncertainty on initial positions. Journal of Differential Equations, vol. 264, n. 5, pp. 3212–3252 (2018).
[28] Jimenez, C., Marigonda, A., Quincampoix, M. : On Bolza problem in Wasserstein space. Preprint.
[29] Penrose, R.: A generalized inverse for matrices, Proc. Cambridge Philos. Soc., vol. 51 (1955), pp. 406–413.
[30] Santambrogio, F.: Optimal Transport for Applied Mathematicians. Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser Basel, vol. 87, ed. 1 (2015).
[31] Villani, C.: Topics in optimal transportation. Graduate Studies in Mathematics, American Mathematical Society, vol. 58 (2003).

Giulia Cavagnari: Department of Mathematics "F. Casorati",
University of Pavia
Via Ferrata 5, I-27100 Pavia, Italy.
*Email address*: giulia.cavagnari@unipv.it

Antonio Marigonda: Department of Computer Science,
University of Verona
Strada Le Grazie 15, I-37134 Verona, Italy.
*Email address*: antonio.marigonda@univr.it

Benedetto Piccoli: Department of Mathematical Sciences,
Rutgers University - Camden
311 N. 5th Street Camden, NJ 08102, USA.
*Email address*: piccoli@camden.rutgers.edu