arXiv:2011.07117v1 [math.OC] 13 Nov 2020

# LAGRANGIAN, EULERIAN AND KANTOROVICH FORMULATIONS OF MULTI-AGENT OPTIMAL CONTROL PROBLEMS: EQUIVALENCE AND GAMMA-CONVERGENCE

GIULIA CAVAGNARI, STEFANO LISINI, CARLO ORRIERI, AND GIUSEPPE SAVARÉ

Abstract. This paper is devoted to the study of multi-agent deterministic optimal control problems. We initially provide a thorough analysis of the Lagrangian, Eulerian and Kantorovich formulations of the problems, as well as of their relaxations. Then we exhibit some equivalence results among the various representations and compare the respective value functions. To do it, we combine techniques and ideas from optimal transportation, control theory, Young measures and evolution equations in Banach spaces. We further exploit the connections among Lagrangian and Eulerian descriptions to derive consistency results as the number of particles/agents tends to infinity. To that purpose we prove an empirical version of the Superposition Principle and obtain suitable Gamma-convergence results for the controlled systems.

Contents

1. Introduction 2
2. Preliminaries and notations 9
2.1. Borel probability measures 10
2.2. The Wasserstein metric and the Superposition Principle 11
2.3. Disintegration and Young measures 13
3. Structural assumptions for the dynamics of the optimal control problems 14
3.1. The relaxed setting 15
4. Lagrangian optimal control problem 16
4.1. Basic results 17
4.2. Approximation by piecewise constant controls 20
4.3. Approximation by continuous controls and trajectories 23
5. Relaxed Lagrangian optimal control problem 27
5.1. Equivalence of **L** and **RL**. Chattering result 28
6. Eulerian optimal control problem 32
7. Kantorovich optimal control problem and equivalence with the Eulerian 38
8. Equivalence of Eulerian and Lagrangian problems 41
8.1. Comparison between **L**, **E** and **K** 41
8.2. Continuity of $V_{\mathbf{E}}$, $V_{\mathbf{K}}$ and $V_{\mathbf{L}}$ 45
8.3. A counterexample: Non-existence of minimizers for **L** 46
9. Finite particle systems and Gamma-convergence 49
9.1. Equivalences between $N$-particles problems 49
9.2. Finite particle approximation for **L** 52
9.3. Finite particle approximation for **E** 54
Appendixes 57
Appendix A. Ordinary differential equations in Banach spaces 57
A.1. Cauchy problem in Banach spaces 59
Appendix B. A convergence result for solutions of Cauchy problems 61
Appendix C. An empirical Superposition Principle 62

*Date*: November 17, 2020.
*Key words and phrases.* Wasserstein distance, optimal control, mean-field optimal control, Gamma convergence.

Appendix D. Finite Partitions 65
Acknowledgements 67
References 68

## 1. Introduction

In recent years there has been an impressive increase in the analysis of models with interactions and associated optimal control problems. The motivations and fields of interest are various and range from statistical mechanics to biology, from crowd dynamics to the description of economical and financial phenomena, and many others. A lot of different mathematical models and techniques have been proposed in the literature and it seems impossible to be exhaustive in accounting here all the developments. We refer to [16, 19, 34, 35] and the references therein for some of the recent results.

A large part of the literature concentrates the attention on the evolution of populations of similar individuals where the single agent feels the interaction with the others through an averaged term. In this case, when the number of individuals is very large, an aggregation effect takes place and the (discrete) collection of agents is usually replaced by its spatial density. This idea comes from the so called *mean field* approach in statistical physics where it has been fruitfully used to develop a limit theory when the number of particles goes to infinity.

Within this framework, optimal control problems, both at the microscopic and macroscopic levels, are naturally considered, see e.g. [18, 21, 24, 35, 39, 47]. A first motivation for the introduction of (centralized) controls is the incompleteness of the concept of self-organization. In fact, for a population of interacting particles/agents, global coordination or pattern formation is not a priori guaranteed and the intervention of a central planner on the dynamics could promote these mechanisms: this leads to the definition of *multi-agent* control problems. A further motivation is the analysis of interacting rational agents with similar optimization goals. In this case, the mean field approach consists in approximating a large number of agents with a single representative individual, whose aim is to solve a control problem constrained to a field equation (encoding the averaged behaviour of the population). The mean field term influences both the dynamics and the cost functional. Whether the representative agent can or cannot influence the mean field term depends on the model under consideration. The interested reader is referred to the books [6, 16] for a detailed description of various aspects of mean field models.

In the present paper we study different formulations of *multi-agent optimal control problems*, that we denote respectively with *Lagrangian*, *Eulerian* and *Kantorovich* as well as the corresponding *limit theory*.

As already mentioned above, *multi-agent optimal control*, also known in literature as centralized optimal control of Vlasov dynamics, represents non-standard optimal control problems where each individual is influenced by the averaged behaviour of all the others and the central planner aims at minimizing a cost functional which depends on the distribution of all the agents.

A large effort has been devoted in the last years to extend results of classical optimal control theory to the mean field setting, with a particular attention to the measure-formulation of the problems in Wasserstein spaces. In this direction, let us mention [11, 48] for necessary conditions for optimality in the form of a Pontryagin maximum principle, [21] for a generalized version of dynamic programming, [10] for the study of differential inclusions and the contribution [39] for the analysis of necessary and sufficient conditions for optimality in the form of a Hamilton-Jacobi-Bellman equation in the Wasserstein space.

*Lagrangian*, *Eulerian* and *Kantorovich* refer to different points of view that can be adopted to study the dynamics of the problems. The Lagrangian and Eulerian terminologies come from fluid-dynamics and they have been recently adopted in the theory of optimal transport, from which we also took inspiration for the Kantorovich formulation. In general terms, the *Lagrangian* approach consists in labelling each particle and following the corresponding trajectory. The

*Eulerian* description, on the other hand, aims at measuring the velocity of particles flowing at a point at a fixed time. In this paper we introduce a further point of view, that we name *Kantorovich* in analogy with the Kantorovich extension of Monge problem (in the same spirit, see also [1, 3, 5]). The Kantorovich formulation turns out to be fundamental in connecting the Lagrangian and Eulerian points of view and it is based on the representation of solutions of the continuity equation provided by the superposition principle (see e.g. [3, Theorem 8.2.1] or [7, Theorem 5.8]).

*Limit theory* refers instead to the question of connecting optimal control problems with a finite number of agents with the infinite dimensional description given by a continuum of players. The study of interacting particle systems becomes intractable when the population is very large and, in many circumstances, the connection with a limit (mean field) approximation heavily simplifies the study. For this reason, many results concerning asymptotic behaviour when the size of the system grows have been developed in the literature, both from the theoretical and applicative point of view. Even if mean field approximations are mainly proved in the uncontrolled stochastic setting, some applications to deterministic models can be found e.g. in [44] and [25]. Concerning the controlled case, an important effort has been directed to the case of mean field games (see below for some references) but, to the best of our knowledge, only few rigorous results for deterministic multi-agent systems can be found in the literature. Let us mention [35], [34] for Gamma-convergence techniques under different assumptions on the controls and velocity fields, and [22] for an extension when distance constraints among the agents are imposed.

The analysis we develop here aims at providing a unifying framework for the study of deterministic multi-agent optimal control problems and surely benefits from the connections with the theory of optimal transport. This is more evident when dealing with a measure-formulation of the problem, where a non-local continuity equation is guiding the dynamics, or with the Kantorovich description, where a superposed measure in the space of continuous paths selects the trajectories of the system. Another source of inspiration is the theory of Young measures, classically used in control theory, which plays a crucial role in the description of relaxed problems.

Apart from its theoretical interest, we trust the present investigation could serve as a founding step into a general treatment of optimality conditions in the rapidly-growing context of Wasserstein spaces. Within this context a major role is played by the Hamilton-Jacobi-Bellman (HJB) equations in the space of probability measures, for which different notions of solution has been already proposed in the literature. We think that the analysis of HJB equation could benefit from the equivalence results and the limit theory developed here and we leave it to future investigation . Let us just briefly mention some contributions in this direction: a general analysis in metric setting can be found in [2, 36], see also [37] for the particular choice of the Wasserstein space. Viscosity solutions for HJB are studied e.g. in [14, 23] for the case of random differential games and in [39] for multi-agent systems.

Let us finally report on the stochastic counterpart of the theory (that we do not treat herein) and make some further comments on the connections with mean field games.

**Stochastic counterpart.** In the stochastic setting, mean field behaviour of interacting particles systems is classically referred to as *propagation of chaos*. The literature on the subject is far too vast to be discussed here and the interested reader is referred to Sznitman's Saint-Flour lectures [53] for a beautiful treatment of the subject and to the references in e.g. [27, 45] for some of the more recent developments.

Concernig stochastic control problems, a rigorous consistency result for controlled McKean-Vlasov dynamics has been obtained by Lacker in [40] using martingale problems and relaxation. It is interesting to notice that the result contained in [40] allows also for degenerate diffusion. Further extensions has been pursued in [27], where a common noise is also introduced, and in [28] where the state dynamics depends upon the joint distribution of state and control.

For what concerns equivalence results, to the best of our knoledge, the more general study in the stochastic setting is formulated in [27]. There, the authors prove existence of optimal controls and show the equivalence at the level of value functions of the so-called strong, weak (in

a probabilistic sense) and relaxed (Lagrangian) formulations of the stochastic control problems. Notably, the results contained in [27] extends the ones in [40] to the far more general case of common noise. Observe also that the equivalence between the weak and strong formulations has a fundamental role in establishing a dynamic programming principle in [29].

To conclude, let us just remark on a possible application of deterministic consistency results to the study of limit behaviour of uncontrolled stochastic particle systems. The idea stems from the link between Gamma-convergence and large deviations developped by Mariani in [43]. If a consistency result for deterministic control problems is proved in terms of Gamma-convergence, then it is possible to translate the $\Gamma - \liminf$ and $\Gamma - \limsup$ inequalities in corresponding lower and upper bound estimates for suitable associated stochastic systems. An example of this technique is contained in [45] (see also [12] for a recent extension) where the limit theory developed in [34] has been used to prove a large deviations principle for stochastic equations in the mean-field and small-noise regime.

**Connections with MFGs.** The theory of Mean Field Games (MFGs), separately introduced by Lasry and Lions [42] and Huang, Caines, Malhamé [38] aims at describing non-cooperative indistinguishable players interacting through their empirical distribution. Opposed to Vlasov control problems, the optimization problem of each agent in mean field games leads to the notion of Nash equilibrium for the system. Differences and similarities between centralized optimal control of (McKean)-Vlasov dynamics and equilibria in MFGs are discussed in the literature, see e.g. [17], [6] and [15]. For what concerns the limit of $N$-players differential games towards the MFG system, a fundamental result was obtained in [13] via the so-called Master equation. Among the various extensions of [13], let us mention the convergence of open and closed-loop Nash equilibria to MFGs equilibria obtained respectively in [32] and [41]. The case of first-order MFGs has been taken into account in [33].

We now briefly describe the various formulations of the optimal control problems we deal with and we present the main results contained in the paper.

To improve the readability of the introduction, we just sketch the essential features of the problems, omitting the details and heavily simplifying the setting and theorems whenever possible. Precise statements are given in the forthcoming sections. Throughout the paper we use interchangeably the terms particles/agents as they differ only in view of the different applications.

**Lagrangian formulation (L).** The Lagrangian formulation has a probabilistic flavour and it is built upon a probability space $(\Omega, \mathfrak{B}, \mathbb{P})$ which acts as a parametrization space for the particles. More precisely, given a finite time horizon $T > 0$ and a (compact metric) space of control actions $U$, the controlled dynamics $X : [0,T] \times \Omega \to \mathbb{R}^d$ is given by

$$\begin{cases} \dot{X}_t(\omega) = f(X_t(\omega), u_t(\omega), (X_t)_\sharp \mathbb{P}), & \text{for a.e. } t \in (0,T) \\ X_{|t=0}(\omega) = X_0(\omega), \end{cases} \tag{1.1}$$

where the dependence of the vector field $f$ on the measure $(X_t)_\sharp \mathbb{P}$ models the interaction among particles and/or the interaction of the mass with the surrounding environment and it is usually referred to as a *mean field* interaction.

A natural motivation for the introduction of a parametrization space comes from a large variety of problems where a finite number of particles/agents are involved. In this case $\Omega$ can be simply interpreted as a set of labels $\Omega^N = \{1, \dots, N\}$, with $\mathfrak{B}^N$ the associated algebra of parts and $\mathbb{P}^N(\{\omega\}) = \frac{1}{N}$, $\omega = 1, \dots, N$, the normalized counting measure. Each particle is indistinguishable from the others and the interaction enters the system through the empirical measure $\mu_t^N = \frac{1}{N} \sum_{\omega=1}^N \delta_{X_t(\omega)}$, with $t \in [0,T]$.

Given $X_0 \in L^p(\Omega; \mathbb{R}^d)$, for $p \geq 1$, an admissible pair $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ for the Lagrangian optimal control problem $(\mathbf{L})$ consists in a measurable control $u : [0,T] \times \Omega \to U$ and a solution (in a suitable sense) of (1.1). Associated to the dynamics, the *cost functional* to be minimized

has the form

$$J_{\mathbf{L}}(X,u) := \int_\Omega \int_0^T \mathcal{C}(X_t(\omega), u_t(\omega), (X_t)_\sharp \mathbb{P})\, \mathrm{d}t\, \mathrm{d}\mathbb{P}(\omega) + \int_\Omega \mathcal{C}_T(X_T(\omega), (X_T)_\sharp \mathbb{P})\, \mathrm{d}\mathbb{P}(\omega),$$

where the running cost $\mathcal{C}$ and the final cost $\mathcal{C}_T$ are non-local as they could depend on the measures $(X_\cdot)_\sharp \mathbb{P}$ and $(X_T)_\sharp \mathbb{P}$.

The optimization of the cost functional among admissible pairs leads to the definition of the so-called *value function* which, for the Lagrangian problem, can be written as

$$V_{\mathbf{L}}(X_0) := \inf \left\{ J_{\mathbf{L}}(X,u) \,:\, (X,u) \in \mathcal{A}_{\mathbf{L}}(X_0) \right\}. \tag{1.2}$$

Let us notice that existence of optimal pairs $(X,u)$ (for which the minimum is achieved in (1.2)) is not guaranteed in general. A counterexample for the Lagrangian problem, even in the relaxed formulation, is given by the Wasserstein barycenter problem with suitable initial distribution (see Section 8.3 for details). From a probabilistic point of view, the Lagrangian formulation can be thought as a *random* optimal control problem in strong formulation, where the randomness is encoded in the initial distribution of the dynamics. In this context, we do not consider stochastic perturbation of the dynamics given e.g. from independent Brownian motions and/or common noise.

**Relaxed Lagrangian formulation (RL).** A classical generalization of optimal control problems is the so called *relaxed* version, where controls are allowed to take values in the space of probability measures $\sigma : [0,T] \times \Omega \to \mathscr{P}(U)$. This greatly enlarges the class of admissible pairs (classical controls can be recovered choosing $\sigma := \delta_u$, with $u \in U$) and provides a convexification of the problem under consideration. Indeed, the controlled trajectories satisfy the linear (in the control action) dynamics

$$\begin{cases} \dot{X}_t(\omega) = \displaystyle\int_U f(X_t(\omega), u, (X_t)_\sharp \mathbb{P})\, \mathrm{d}\sigma_{t,\omega}(u), \quad \text{for a.e. } t \in ]0,T] \\ X_{|t=0}(\omega) = X_0(\omega), \end{cases}$$

with control $\sigma_{t,\omega} := \sigma(t,\omega) \in \mathscr{P}(U)$. Furthermore, the cost functional takes the form

$$J_{\mathbf{RL}}(X,\sigma) := \int_\Omega \int_0^T \int_U \mathcal{C}(X_t(\omega), u, (X_t)_\sharp \mathbb{P})\, \mathrm{d}\sigma_{t,\omega}(u)\, \mathrm{d}t\, \mathrm{d}\mathbb{P}(\omega) \; + \int_\Omega \mathcal{C}_T(X_T(\omega), (X_T)_\sharp \mathbb{P})\, \mathrm{d}\mathbb{P}(\omega),$$

with associated *value function* $V_{\mathbf{RL}} : L^p(\Omega; \mathbb{R}^d) \to [0, +\infty)$ given by

$$V_{\mathbf{RL}}(X_0) := \inf \left\{ J_{\mathbf{RL}}(X,\sigma) \,:\, (X,\sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0) \right\}.$$

Relaxation is a fundamental concept in optimal control theory and has its roots in the theory of Young measures. In Section 5 of the paper, we provide a detailed analysis of the relaxation procedure in the context of multi-agent systems, emphasizing its connections with the Lagrangian problem. Of particular interest is the extension of a suitable version of the chattering theorem, which permits to approximate the Relaxed Lagrangian formulation with a sequence of (not relaxed) Lagrangian ones. This readily implies the equality of the respective value functions: $V_{\mathbf{RL}}(X_0) = V_{\mathbf{L}}(X_0)$ for any $X_0 \in L^p(\Omega; \mathbb{R}^d)$.

Let us finally notice that the Relaxed Lagrangian formulation is the prototype for the class of control problems satisfying suitable *Convexity Assumptions* (see Assumption 3.4 below). In this particular case, the control space is the convex space of measures $\mathscr{P}(U)$, the dynamics is affine in the controls and the cost functional is convex (actually linear). However, as already observed before, this relaxation procedure is not sufficient to guarantee existence of minimizers for a general optimal control problem (see Section 8.3). A further step in this direction is the introduction of the Eulerian formulation of the problem.

**Eulerian formulation (E).** To simplify the presentation, here we suppose to directly deal with relaxed controls, which are represented by a Borel measurable map $\sigma : [0,T] \times \mathbb{R}^d \to \mathscr{P}(U)$:

this is fundamental to get existence of minimizers. Then, the evolution of the system is guided by the following non-local Vlasov equation

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0, & \text{in } [0,T] \times \mathbb{R}^d \\ \mu_{t=0} = \mu_0, \end{cases}$$

where the controlled vector field $v$ depends on the evolving state itself and it is given by $v_t(x) := \int_U f(x,u,\mu_t) \mathrm{d}\sigma_{t,x}(u)$. Within this framework, the cost functional takes the form

$$J_{\mathbf{E}}(\mu,\sigma) := \int_0^T \int_{\mathbb{R}^d} \int_U \mathcal{C}(x,u,\mu_t)\, \mathrm{d}\sigma_{t,x}(u)\, \mathrm{d}\mu_t(x)\, \mathrm{d}t + \int_{\mathbb{R}^d} \mathcal{C}_T(x,\mu_T)\, \mathrm{d}\mu_T(x),$$

and the *value function* $V_{\mathbf{E}} : \mathscr{P}_p(\mathbb{R}^d) \to [0,+\infty)$ is given by

$$V_{\mathbf{E}}(\mu_0) := \inf\{J_{\mathbf{E}}(\mu,\sigma) \,:\, (\mu,\sigma) \in \mathcal{A}_{\mathbf{E}}(\mu_0)\}.$$

In the Eulerian description of the optimal control problem, the system can be described by a curve of probability measures. This point of view is intimately connected with the theory of optimal transport from which ideas and techniques are borrowed.

**Kantorovich formulation (K).** A somewhat intermediate formulation is given by the *Kantorovich* optimal control problem (in analogy to the Kantorovich formulation of the optimal transport problem). This formulation has its roots in the representation of solutions of the continuity equation by superposition of continuous curves belonging to $\Gamma_T := C([0,T];\mathbb{R}^d)$. An admissible pair for the Kantorovich problem is given by $(\boldsymbol{\eta},\sigma)$ where $\boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$ is a probability measure on the space of continuous curves, and $\sigma : [0,T] \times \Gamma_T \to \mathscr{P}(U)$ is a relaxed control. Furthermore, given $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$ with finite $p$-moment, an admissible measure $\boldsymbol{\eta}$ has to match the initial condition in the form $(e_0)_\sharp \boldsymbol{\eta} = \mu_0$. Even more important, defining $\mu_t := (e_t)_\sharp \boldsymbol{\eta}$ for all $t \in [0,T]$, $\boldsymbol{\eta}$ has to be concentrated on the set of absolutely continuous solutions of the differential equation

$$\dot\gamma(t) = \int_U f(\gamma(t),u,\mu_t) \mathrm{d}\sigma_{t,\gamma}(u), \qquad \text{for } \mathcal{L}_T\text{-a.e. } t \in [0,T].$$

This clearly links the Kantorovich formulation with the Eulerian one via the superposition principle (see Theorem 2.5). The *cost functional* associated to the Kantorovich formulation is written in the form

$$J_{\mathbf{K}}(\boldsymbol{\eta},\sigma) := \int_{\Gamma_T} \int_0^T \int_U \mathcal{C}(\gamma(t),u,\mu_t)\, \mathrm{d}\sigma_{t,\gamma}(u)\, \mathrm{d}t\, \mathrm{d}\boldsymbol{\eta}(\gamma) + \int_{\mathbb{R}^d} \mathcal{C}_T(x,\mu_T)\, \mathrm{d}\mu_T(x),$$

where $\Gamma_T$ act as a parametrization space (as in the Lagrangian framework), but the minimization involves measures $\boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$ instead of trajectories, in line with the Kantorovich formulation of the optimal transportation problem: $\mu_t = (e_t)_\sharp \boldsymbol{\eta}$ can be considered as a time dependent family of marginals of $\boldsymbol{\eta}$. The associated *value function* is given by

$$V_{\mathbf{K}}(\mu_0) := \inf\{J_{\mathbf{K}}(\boldsymbol{\eta},\sigma) \,:\, (\boldsymbol{\eta},\sigma) \in \mathcal{A}_{\mathbf{K}}(\mu_0)\}.$$

Let us stress that, also in this setting, the choice of relaxed controls is sufficient to prove the existence of minimizers.

**Equivalence results.** A natural question is whether the problems introduced above are somewhat related. One of the aims of the present paper is to prove equivalences among the various formulations introduced above. At the level of the value functions, we can summarize the main result in the following theorem (see Theorem 8.2 for a precise statement)

**Theorem** (equivalence)**.** *Let $(\Omega,\mathcal{B},\mathbb{P})$ be a Polish space such that $\mathbb{P}$ is without atoms. If $X_0 \in L^p(\Omega;\mathbb{R}^d)$, then*

$$V_{\mathbf{L}}(X_0) = V_{\mathbf{RL}}(X_0) = V_{\mathbf{E}}((X_0)_\sharp \mathbb{P}) = V_{\mathbf{K}}((X_0)_\sharp \mathbb{P}).$$

The first step of the proof consists in the approximation of the Lagrangian problem by piecewise constant controls (see Theorem 4.17). This is possible whenever the probability space $(\Omega, \mathfrak{B}, \mathbb{P})$ under consideration satisfies a suitable *finite approximation property* (Definition 4.15), which surely holds in the Polish framework. Once the piecewise approximation is established, we are able to formulate a suitable version of the *chattering theorem* (see Theorem 5.6) where trajectories, controls and cost functional of the Relaxed Lagrangian formulation are approximated by the corresponding objects in the Lagrangian setting.

The comparison between the Lagrangian and Eulerian formulations (see Theorems 8.1, 8.2 and Section 8.1) is more delicate and it is achieved by exploiting the Kantorovich description of the control problem on the space of curves $\Gamma_T$. The idea is to separately connect the Eulerian and Kantorovich descriptions (Theorem 7.3) and then the Kantorovich and Lagrangian ones (Theorem 8.5).

Starting from an admissible pair for the Eulerian problem, the application of the superposition principle given in Theorem 2.5 easily provides a candidate admissible pair for the Kantorovich problem paying the same cost (see the proof of Proposition 7.4 for details). Conversely, if a pair $(\boldsymbol{\eta}, \sigma)$ for the Kantorovich problem is given, the Eulerian control action can be obtained by averaging with respect to a suitable disintegration of the measure $\boldsymbol{\eta}$ on curves, as it is shown in the proof of Proposition 7.5.

For what concerns the Kantorovich/Lagrangian comparison, in Lemma 8.4 we firstly interpret the Kantorovich problem as a Lagrangian one with parametrization space given by $\Gamma_T$, i.e. the space of curves, and with trajectories given by the evaluation map. Due to the continuity of the initial datum (i.e. the evaluation map $e_0$), we then approximate controls with continuous ones (see Proposition 4.22 for a general result in this direction) and finally we approximate the obtained Lagrangian problem in $\Gamma_T$ with Lagrangian ones which are set in a generic parametrization space $\Omega$, not necessarily the space $\Gamma_T$. A precise description of this technique is contained in the proof of Theorem 8.5.

An immediate consequence of the equivalence theorem is the equality of the value functions for different initial data, whenever the respective laws coincide (see Theorem 8.1). In fact, if $\mu_0 := (X_0)_\sharp \mathbb{P} = (X_0')_\sharp \mathbb{P}$, then it holds that

$$V_{\mathbf{L}}(X_0) = V_{\mathbf{L}}(X_0') \quad \big( = V_{\mathbf{E}}(\mu_0)\big).$$

A further consequence is the continuity of the value functions with respect to the initial data, see Theorems 8.7 and 8.8 for precise statements.

Equivalence results between Lagrangian, Eulerian and Kantorovich formulations represent a first step towards a general analysis of optimality conditions for multi-agents control systems. A second step in this direction is the study of the corresponding limit theory.

**Approximation by finite particle systems.** We aim to provide a rigorous limit theory for multi-agent optimal control problems both for the Lagrangian and Eulerian formulations. To do it, we have to define appropriate discrete versions of the two formulations. The $N$-particle Lagrangian control problem $\mathbf{L}^N$ simply relies on the choice of $\Omega^N = \{1, \dots, N\}$ as parametrization space. On the other hand, a genuine discrete Eulerian problem $\mathbf{E}^N$ requires the introduction of a constraint on the number of particles (see Definition 9.1), precisely an admissible trajectory $\mu$ satisfies $\mu_t \in \mathscr{P}^N(\mathbb{R}^d)$, where

$$\mathscr{P}^N(\mathbb{R}^d) := \left\{ \mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \ \text{ for some } x_i \in \mathbb{R}^d \right\}.$$

The discrete Eulerian and Lagrangian control problems $\mathbf{L}^N$ and $\mathbf{E}^N$ turn out to be equivalent (see Theorem 9.3) in the sense that

$$V_{\mathbf{L}^N}(X_0) = V_{\mathbf{E}^N}((X_0)_\sharp \mathbb{P}^N), \qquad \text{for any } X_0 \in L^p(\Omega^N; \mathbb{R}^d). \tag{1.3}$$

Note that this is not a direct consequence of the general equivalence result given above, where the reference probability measure $\mathbb{P}$ was required to be without atoms. To prove the equality

in (1.3) we derive a discrete formulation of the superposition principle for empirical probability measures (see Theorem C.1) that we believe might be of interest in itself.

Once the equivalence at the level of $N$-particle systems is established, we derive Gamma-convergence results respectively for the Lagrangian and Eulerian problems as the number of particles diverges (see Propositions 9.9 and 9.13). A major consequence is contained in the following theorem (see Theorem 9.12 for a detailed description).

**Theorem.** *Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a Polish space such that $\mathbb{P}$ is without atoms. Assume that the Convexity assumption 3.4 holds.*

- *If $X_0 \in L^p(\Omega; \mathbb{R}^d)$ and $X_0^N : \Omega^N \to \mathbb{R}^d$, $N \in \mathbb{N}$, satisfy $X_0^N \to X_0$ as $N \to +\infty$ (see (9.10)), then*
$$\lim_{N\to+\infty} V_{\mathbf{L}^N}(X_0^N) = V_{\mathbf{L}}(X_0).$$
- *If $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ and $\mu_0^N \in \mathscr{P}^N(\mathbb{R}^d)$, $N \in \mathbb{N}$, satisfy $W_p(\mu_0^N, \mu_0) \to 0$ as $N \to +\infty$, then*
$$\lim_{N\to+\infty} V_{\mathbf{E}^N}(\mu_0^N) = V_{\mathbf{E}}(\mu_0).$$
- *Moreover, if $(X_0^N)_\sharp \mathbb{P}^N = \mu_0^N$ it holds that*
$$\lim_{N\to+\infty} V_{\mathbf{L}^N}(X_0^N) = V_{\mathbf{E}}(\mu_0).$$

Notice that the usual *mixed* Lagrangian-Eulerian consistency in the third item is a simple byproduct of the equivalence in (1.3).

**Structure of the paper.** The paper is organized as follows. In Section 2 we fix the notation and present some preliminary material. We start by revising some properties of Borel probability measures and we recall a refined version of the Skorohod representation theorem. We further provide some material on optimal transport, Wasserstein spaces and we present the classical superposition principle. Finally, we discuss the disintegration theorem and give some properties of Young measures.

Section 3 contains our standing hypotheses, divided into two sets: the *Basic* Assumptions and the *Convexity* Assumptions. The first ones require compactness of the metrizable space of controls and Lipschitz continuity of the velocity field with respect to the state and the mass distribution. The cost functional has to be continuous and to satisfy a polynomial growth condition. The convexity assumptions impose convexity of the control set and of the cost functional (with respect to controls). Furthermore, the dynamics has to be affine with respect to the control actions. The relaxed setting is the guiding example of the convex case.

In Section 4 we present and study the Lagrangian optimal control problem. In particular, we exhibit two different approximation procedures: the first one by piecewise constant controls and the second one by continuous controls and trajectories.

The Relaxed Lagrangian problem is defined in Section 5, where its representation as a Lagrangian problem in the lifted space of measures is also discussed. We prove the equivalence between Lagrangian and Relaxed Lagrangian formulations of the control problem by approximating relaxed controls with (non-relaxed) ones in the sense of Young convergence. This procedure, known as chattering theorem, exploits the approximation by piecewise constant controls developed in the Lagrangian setting.

Section 6 contains the definition and properties of the Eulerian control problem. Under the convexity Assumptions we are able to prove existence of minimizers for the control problem via a direct method.

In Section 7 we introduce the Kantorovich problem and we prove its equivalence with the Eulerian one under the convexity assumptions.

The main equivalence results are contained in Section 8. Exploiting the Kantorovich formulation, we firstly show the equality between value functions of Lagrangian and Eulerian problems under the Convexity Assumptions. The general case is then obtained by interpreting the Relaxed

Lagrangian as a (convex) Lagrangian problem in the space of probability measures. As a consequence, we also get the continuity of the value functions, with respect to the initial condition, for the various formulations. In Subsection 8.3 we discuss the possible non-existence of minimizers for the Lagrangian and Relaxed Lagrangian problems. This is not guaranteed, if the initial condition is assigned, even under the Convexity Assumptions. We produce a counterexample to the existence of minimizers given by the Wasserstein barycenter problem.

Section 9 contains all the material regarding finite particle problems and the respective limit theory. We define a discrete version of the Eulerian control problem imposing a constraint on the number of particles. To connect the Eulerian and Lagrangian formulations in the $N$-particle case we introduce a *Feedback* Lagrangian control problem, where control actions are indeed in feedback form, and we make use of the already mentioned discrete superposition principle. A Gamma-convergence result both for the Lagrangian and Eulerian formulations is then established as the number of particles tends to infinity. We finally conclude proving the convergence of the associated value functions.

The appendix contains various technical tools. In particular, in Appendix C we state and prove the superposition principle for the evolution of empirical measures.

## 2. Preliminaries and notations

We list here the main notation.

| | |
|---|---|
| $\#A$ | the cardinality of a set $A$; |
| $i_X(\cdot)$ | the identity function on a set $X$, $i_X : X \to X$ defined by $i_X(x) = x$; |
| $\mathbb{1}_A(\cdot)$ | the characteristic function of $A \subset X$, $\mathbb{1}_A : X \to \mathbb{R}$ defined by $\mathbb{1}_A(x) = 1$ if $x \in A$, $\mathbb{1}_A(x) = 0$ if $x \in X \setminus A$; |
| $(S, \mathfrak{B})$ | measurable space $S$ with $\sigma$-algebra $\mathfrak{B}$; |
| $(S, \mathcal{B}_S)$ | topological space $S$ with Borel $\sigma$-algebra $\mathcal{B}_S$; |
| $\mathrm{M}(X;Y)$ | the set of measurable functions from the measurable space $X$ to the measurable space $Y$; |
| $\mathrm{B}(X;Y)$ | the set of Borel measurable functions from the topological space $X$ to the topological space $Y$; |
| $C(X;Y)$ | the set of continuous functions from the topological space $X$ to the topological space $Y$; |
| $C_c(X;Y)$ | the set of continuous compactly supported functions from the topological space $X$ to the topological space $Y$; |
| $C_b(X;Y)$ | the set of continuous bounded functions from the topological space $X$ to the metric space $Y$; |
| $\mathrm{AC}^p([0,T];X)$ | the set of absolutely continuous functions from $[0,T]$ to the metric space $X$ with metric derivative in $L^p([0,T];\mathbb{R})$; |
| $\Gamma_T$ | the set of continuous curves from $[0,T]$ to $\mathbb{R}^d$, i.e., $\Gamma_T = C([0,T];\mathbb{R}^d)$; |
| $e_t$ | the evaluation map at time $t \in [0,T]$, $e_t : \Gamma_T \to \mathbb{R}^d$ defined by $e_t(\gamma) = \gamma(t)$; |
| $\mathscr{P}(X)$ | the set of probability measures on the measurable space $X$; |
| $\mathscr{P}^N(\mathbb{R}^d)$ | the set of empirical probability measures on $\mathbb{R}^d$ defined in (9.2); |
| $\mathrm{m}_p(\mu)$ | the $p$-th moment of a probability measure $\mu \in \mathscr{P}(\mathbb{R}^d)$, defined by $\mathrm{m}_p(\mu) = \left(\int_{\mathbb{R}^d} \lvert x\rvert^p \,\mathrm{d}\mu(x)\right)^{1/p}$; |
| $r_\sharp\mu$ | the push-forward of the measure $\mu \in \mathscr{P}(X)$ by the measurable map $r \in \mathrm{M}(X;Y)$; |
| $\mu \otimes \nu$ | the product measure of $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$; |
| $\pi^i$ | the $i$-th projection map $\pi^i : X_1 \times \cdots \times X_N \to X_i$ defined by $\pi^i(x_1,\dots,x_N) = x_i$; |
| $\pi^{i,j}$ | the $(i,j)$-th projection map $\pi^{i,j} : X_1 \times \cdots \times X_N \to X_i \times X_j$ defined by $\pi^{i,j}(x_1,\dots,x_N) = (x_i,x_j)$; |
| $W_p(\mu,\nu)$ | the $p$-Wasserstein distance between $\mu$ and $\nu$ (see Definition 2.2); |
| $\mathscr{P}_p(\mathbb{R}^d)$ | the metric space of the elements in $\mathscr{P}(X)$ with finite $p$-moment, endowed with the $p$-Wasserstein distance; |
| $\mathcal{L}_T$ | the normalized Lebesgue measure restricted to the interval $[0,T]$, i.e. $\mathcal{L}_T := \frac{1}{T}\mathcal{L} \llcorner [0,T]$. |

2.1. **Borel probability measures.** Let $(S, \mathfrak{B})$ be a measurable space. When $S$ is a Polish topological space, we will implicitely assume that $\mathfrak{B}$ coincides with the Borel $\sigma$-algebra $\mathcal{B}_S$ of $S$. We say that $(S, \mathfrak{B})$ is *a standard Borel space* if it is isomorphic (as a measurable space) to a Borel subset of a complete and separable metric space; equivalently, one can find a Polish topology $\tau$ on $S$ such that $\mathfrak{B} = \mathcal{B}_{(S,\tau)}$.

If $(E, \tilde{\mathfrak{B}})$ is another measurable space, we denote by $\mathrm{M}(S; E)$ the set of measurable functions from $S$ to $E$. If $S$ is a topological space we denote with $\mathrm{B}(S; E)$ the set of Borel measurable functions. $\mathscr{P}(S)$ is the set of probability measures on $S$; when $S$ is a Polish space (and $\mathfrak{B} = \mathcal{B}_S$) we will endow $\mathscr{P}(S)$ with the weak (Polish) topology induced by the duality with the continuous and bounded functions of $\mathrm{C}_b(S) := C_b(S; \mathbb{R})$.

Given $\mu \in \mathscr{P}(S)$ and $r : S \to E$ a measurable map, we define the *push forward of $\mu$ through $r$*, denoted by $r_\sharp \mu \in \mathscr{P}(E)$, by $r_\sharp \mu(B) := \mu(r^{-1}(B))$ for all measurable sets $B \in \tilde{\mathfrak{B}}$ (the $\sigma$-algebra on $E$), or equivalently,

$$\int_S f(r(x))\, \mathrm{d}\mu(x) = \int_E f(y)\, \mathrm{d}r_\sharp \mu(y),$$

for every positive, or $r_\sharp \mu$-integrable, function $f : E \to \mathbb{R}$.
Given another measurable space $Z$, $\mu \in \mathscr{P}(S)$, and $r : S \to E$, $s : E \to Z$ measurable maps, the following composition rule holds

$$(s \circ r)_\sharp \mu = s_\sharp (r_\sharp \mu). \tag{2.1}$$

Moreover, if $r : S \to E$ is a continuous map (with respect to suitable Polish topologies in $S$ and $E$) then $r_\sharp : \mathscr{P}(S) \to \mathscr{P}(E)$ is continuous as well.

The following proposition generalizes to some extent the classical Skorohod representation Theorem, see e.g. [9, Theorem 6.7]. For a (more general) result and the proof we refer to [8, Theorems 3.1 and 3.2].

**Proposition 2.1.** *Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space such that $\mathbb{P}$ is without atoms and let $S$ be a Polish space.*

*(i) If $\nu \in \mathscr{P}(S)$, then there exists a measurable map $X : \Omega \to S$ such that $X_\sharp \mathbb{P} = \nu$.*
*(ii) If $\nu^n, \nu \in \mathscr{P}(S)$ with $\nu^n \to \nu$ weakly, then there exist measurable maps $X^n, X : \Omega \to S$, $n \in \mathbb{N}$, such that $X^n_\sharp \mathbb{P} = \nu^n$, $X_\sharp \mathbb{P} = \nu$ and $X^n(\omega) \to X(\omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$.*

Notice that, when $(\Omega, \mathfrak{B}, \mathbb{P})$ is a standard Borel space, $\tau$ a Polish topology on $\Omega$ such that $\mathfrak{B} = \mathcal{B}_{(\Omega,\tau)}$, then the maps $X$ and $X^n$ in Proposition 2.1 are Borel measurable. A particular and significant case occurs when we choose $(\Omega, \mathfrak{B}, \mathbb{P}) = ([0, 1], \mathcal{B}, \mathcal{L}_1)$, with $\mathcal{B}$ the Borel $\sigma$-algebra and $\mathcal{L}_1$ the Lebesgue measure restricted to $[0, 1]$.

If $\mathfrak{m} \in \mathscr{P}(S)$ and $E$ is a separable Banach space, we denote by $L^p_{\mathfrak{m}}(S; E)$ the space of (the equivalence classes of) $\mathfrak{m}$-measurable functions $f : S \to E$ such that $\int_S \|f(x)\|^p \, \mathrm{d}\mathfrak{m}(x) < +\infty$. Since $E$ is separable, the notions of weak and strong measurability coincide. We will often adopt the notation $L^p(S; E)$ in place of $L^p_{\mathfrak{m}}(S; E)$ when the measure $\mathfrak{m}$ is clear from the context.
We say that a sequence of measurable functions $u_n \in \mathrm{M}(S; E)$ converges in $\mathfrak{m}$-measure to $u \in \mathrm{M}(S; E)$ if

$$\forall\, \varepsilon > 0, \quad \lim_{n \to +\infty} \mathfrak{m}\left(\{x \in S : \|u_n(x) - u(x)\| \geq \varepsilon\}\right) = 0. \tag{2.2}$$

If $u_n$ take values in a compact subset $U$ of $E$, $u_n \in \mathrm{M}(S; U)$, the convergence of $u_n$ to $u \in \mathrm{M}(S; U)$ in $\mathfrak{m}$-measure is equivalent to the convergence of $u_n$ to $u$ in $L^p(S; E)$ for every $p \in [1, +\infty)$.

Given $(S, d)$ a metric space and $p \in [1, +\infty]$, we say that a curve $\gamma : [0, T] \to S$ belongs to $\mathrm{AC}^p([0, T]; S)$ if there exists $m \in L^p(0, T; \mathbb{R})$ such that

$$d(\gamma(t_1), \gamma(t_2)) \leq \int_{t_1}^{t_2} m(s)\, \mathrm{d}s, \qquad \forall\, t_1, t_2 \in [0, T],\ t_1 \leq t_2.$$

2.2. **The Wasserstein metric and the Superposition Principle.** We provide a brief collection of the main notions on optimal transport and Wasserstein distance, addressing the reader to [3, 51, 55].

Given $\mu \in \mathscr{P}(\mathbb{R}^d)$ and $p \geq 1$, we define the $p$-moment of $\mu$ by

$$\mathrm{m}_p(\mu) := \left( \int_{\mathbb{R}^d} |x|^p \, \mathrm{d}\mu(x) \right)^{1/p}.$$

We define $\mathscr{P}_p(\mathbb{R}^d) := \{\mu \in \mathscr{P}(\mathbb{R}^d) : \mathrm{m}_p(\mu) < +\infty\}$. The set $\mathscr{P}_p(\mathbb{R}^d)$ can be metrized by the following distance.

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

$$W_p(\mu_1, \mu_2) := \left( \min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^p \, d\gamma(x_1, x_2) \, : \, \gamma \in \Gamma(\mu_1, \mu_2) \right\} \right)^{1/p}, \tag{2.3}$$

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

$$\Gamma(\mu_1, \mu_2) := \left\{ \gamma \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d) : \, \pi^1_\sharp \gamma = \mu_1, \pi^2_\sharp \gamma = \mu_2 \right\},$$

with $\pi^i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, $\pi^i(x^1, x^2) = x^i$, the projection operator, $i = 1, 2$.

By the previous definitions, given a measurable space $\Omega$ and $\mathbb{P} \in \mathscr{P}(\Omega)$, it follows immediately that for any $Z \in L^p(\Omega; \mathbb{R}^d)$, we have $\mu := Z_\sharp \mathbb{P} \in \mathscr{P}_p(\mathbb{R}^d)$ and

$$\mathrm{m}_p(\mu) = \|Z\|_{L^p(\Omega; \mathbb{R}^d)}, \tag{2.4}$$

moreover

$$W_p(Z^1_\sharp \mathbb{P}, Z^2_\sharp \mathbb{P}) \leq \|Z^1 - Z^2\|_{L^p(\Omega; \mathbb{R}^d)}, \qquad \forall \, Z^1, Z^2 \in L^p(\Omega; \mathbb{R}^d). \tag{2.5}$$

The space $\mathscr{P}_p(\mathbb{R}^d)$ endowed with the $p$-Wasserstein metric $W_p$ is a complete and separable metric space.

The existence of a minimizer in (2.3) can be proved by the direct method in Calculus of Variations. When the measure $\mu_1$ is absolutely continuous with respect to Lebesgue measure $\mathcal{L}^d$ on $\mathbb{R}^d$, the minimizer $\gamma$ is unique and it is concentrated on the graph of a map, $\gamma = (i_{\mathbb{R}^d}, T)_\sharp \mu_1$, where $i_{\mathbb{R}^d}$ is the identity map of $\mathbb{R}^d$ and $T$ is a minimizer in the Monge transport problem

$$\inf \left\{ \int_{\mathbb{R}^d} |x - S(x)|^p \, \mathrm{d}\mu_1(x) \, : \, S_\sharp \mu_1 = \mu_2 \right\}. \tag{2.6}$$

The Wasserstein distance has the following characterization, known as Benamou-Brenier formula:

$$W_p^p(\mu_0, \mu_1) = \min \left\{ \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^p \, \mathrm{d}\mu_t(x) \, \mathrm{d}t : (\mu, v) \in \mathrm{CE}, \, \mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1 \right\}, \tag{2.7}$$

where

$$\mathrm{CE} := \Big\{ (\mu, v) : \mu \in C([0, T]; \mathscr{P}_p(\mathbb{R}^d)), v \in L^p([0, 1] \times \mathbb{R}^d; \mu_t \otimes \mathrm{d}t) \\ \text{such that } \partial_t \mu_t + \mathrm{div}(v_t \mu_t) = 0 \text{ in the sense of distributions} \Big\}.$$

Notice that the minimizers are the constant speed geodesics joining $\mu_0$ to $\mu_1$, i.e. $\{\sigma_t\}_{t \in [0,1]}$ such that $\sigma_0 = \mu_0$, $\sigma_1 = \mu_1$ and $W_p(\sigma_t, \sigma_s) = |t - s| W_p(\mu_0, \mu_1)$ for any $t, s \in [0, 1]$.

We recall the following definition as in [34, Definition 2.2].

**Definition 2.3.** We say that $\psi : [0, +\infty) \to [0, +\infty)$ is an *admissible* function if $\psi(0) = 0$, $\psi$ is strictly convex and of class $C^1$ with $\psi'(0) = 0$, superlinear at $+\infty$, i.e., $\lim_{r \to +\infty} \dfrac{\psi(r)}{r} = +\infty$, and doubling, i.e., there exists $A > 0$ such that

$$\psi(2r) \leq A(1 + \psi(r)) \quad \text{for any } r \in [0, +\infty).$$

We observe that an admissible function $\psi$ satisfies

$$r\psi'(r) \leq A(1+\psi(r)), \qquad \forall\, r \in [0,+\infty). \tag{2.8}$$

The following result provides equivalent conditions for the convergence in the space $\mathscr{P}_p(\mathbb{R}^d)$ and the characterization of compactness.

**Proposition 2.4.** *Let $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathscr{P}_p(\mathbb{R}^d)$ and $\mu \in \mathscr{P}_p(\mathbb{R}^d)$, the following assertions are equivalent:*

*(1)* $\lim_{n\to\infty} W_p(\mu_n,\mu) = 0$*;*
*(2)* $\mu_n$ *weakly converges to* $\mu$ *and* $\mathrm{m}_p(\mu_n) \to \mathrm{m}_p(\mu)$ *as* $n \to +\infty$*;*
*(3)* $\lim_{n\to+\infty} \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_n(x) = \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu(x)$,
*for every continuous function* $\varphi : \mathbb{R}^d \to \mathbb{R}$ *s.t.* $|\varphi(x)| \leq C(1+|x|^p)$ *for any* $x \in \mathbb{R}^d$*;*
*(4)* $\mu_n$ *weakly converges to* $\mu$ *and there exists* $\psi : [0,+\infty) \to [0,+\infty)$ *admissible, according to Definition 2.3, such that*

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_n(x) < +\infty. \tag{2.9}$$

*Moreover, a family $\mathscr{K} \subset \mathscr{P}_p(\mathbb{R}^d)$ is relatively compact if and only if there exists an admissible function $\psi : [0,+\infty) \to [0,+\infty)$ such that*

$$\sup_{\mu\in\mathscr{K}} \int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu(x) < +\infty. \tag{2.10}$$

The proof can be carried on using [3, Lemma 5.1.7, Proposition 7.1.5]. Concerning the implication (2) to (4), it follows by De la Vallée Poussin and Dunford-Pettis theorems together with [34, Lemma 2.3] for the admissibility property.

The following representation result for the (absolutely continuous) solutions of the continuity equation will play a key role in the sequel (see [3, Theorem 8.2.1]). We denote by $\Gamma_T = \mathrm{C}([0,T];\mathbb{R}^d)$ the Banach space of the continuous functions, endowed with the sup norm. We denote by $e_t : \Gamma_T \to \mathbb{R}^d$ the evaluation map at time $t \in [0,T]$ defined by $e_t(\gamma) := \gamma(t)$. We say that $\boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$ is concentrated on a set $B$ if $\boldsymbol{\eta}(\Gamma_T \setminus B) = 0$.

**Theorem 2.5** (Superposition principle)**.** *Let $p \geq 1$. Let $\mu = \{\mu_t\}_{t\in[0,T]} \in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ be a distributional solution of the continuity equation $\partial_t\mu_t + \mathrm{div}(v_t\mu_t) = 0$ for a Borel vector field $v : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ satisfying*

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p\,\mathrm{d}\mu_t(x)\,\mathrm{d}t < +\infty. \tag{2.11}$$

*Then there exists a probability measure $\boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$ such that*

*(i)* $\mu_t = (e_t)_\sharp\boldsymbol{\eta}$ *for every* $t \in [0,T]$*;*
*(ii)* $\boldsymbol{\eta}$ *is concentrated on the set of curves* $\gamma \in \mathrm{AC}^p([0,T];\mathbb{R}^d)$ *satisfying*

$$\dot\gamma(t) = v_t(\gamma(t)), \qquad \text{for } \mathcal{L}_T\text{-a.e. } t \in [0,T].$$

*Conversely, given $\boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$ satisfying item (ii) and (2.11) with $\mu_t := (e_t)_\sharp\boldsymbol{\eta}$ for every $t \in [0,T]$, then $(\mu, v)$ is a distributional solution of $\partial_t\mu_t + \mathrm{div}(v_t\mu_t) = 0$ .*

In Theorem C.1 in Appendix C, we prove a version of the superposition principle in the discrete setting.

2.3. **Disintegration and Young measures.** Let $\mathbb{S}$ and $S$ be Polish spaces. We say that a map $x \in S \mapsto \mu_x \in \mathscr{P}(\mathbb{S})$ is a Borel map if $x \mapsto \mu_x(A)$ is a Borel map for any open set $A \subset \mathbb{S}$.

If $x \in S \mapsto \mu_x \in \mathscr{P}(\mathbb{S})$ is a Borel map and $\lambda \in \mathscr{P}(S)$ we define the measure $\mu_x \otimes \lambda \in \mathscr{P}(\mathbb{S})$ by

$$(\mu_x \otimes \lambda)(A) := \int_S \mu_x(A)\,\mathrm{d}\lambda(x)$$

for any Borel set $A \subset \mathbb{S}$. Equivalently

$$\int_{\mathbb{S}} \varphi(z)\,\mathrm{d}(\mu_x \otimes \lambda)(z) := \int_S \int_{\mathbb{S}} \varphi(z)\,\mathrm{d}\mu_x(z)\,\mathrm{d}\lambda(x)$$

for any bounded Borel function $\varphi : \mathbb{S} \to \mathbb{R}$.

We state the following disintegration result (see for instance [3, Section 5.3]).

**Theorem 2.6** (Disintegration)**.** *Let $\mathbb{S}$ and $S$ be Polish spaces. Let $\mu \in \mathscr{P}(\mathbb{S})$ and $r : \mathbb{S} \to S$ a Borel map. Then there exists a Borel measurable family of probability measures $\{\mu_x\}_{x\in S} \subset \mathscr{P}(\mathbb{S})$, uniquely defined for $r_\sharp\mu$-a.e. $x \in S$, such that $\mu_x(\mathbb{S} \setminus r^{-1}(x)) = 0$ for $r_\sharp\mu$-a.e. $x \in S$, and $\mu = \mu_x \otimes (r_\sharp\mu)$. In particular, for any bounded Borel map $\varphi : \mathbb{S} \to \mathbb{R}$ we have*

$$\int_{\mathbb{S}} \varphi(z)\,d\mu(z) = \int_S \int_{r^{-1}(x)} \varphi(z)\,d\mu_x(z)\,d(r_\sharp\mu)(x). \tag{2.12}$$

*Remark* 2.7. A typical case is given by $\mathbb{S} = S \times Y$, where $Y$ is a Polish space, and $r = \pi^1$. Since $(\pi^1)^{-1}(x) = \{x\} \times Y$ for all $x \in S$, we identify each measure $\mu_x \in \mathscr{P}(S \times Y)$, which is concentrated in $\{x\} \times Y$, with a measure $\mu_x \in \mathscr{P}(Y)$. With this identification, the formula (2.12) takes the form

$$\int_{S\times Y} \varphi(x,y)\,d\mu(x,y) = \int_S \int_Y \varphi(x,y)\,d\mu_x(y)\,d(r_\sharp\mu)(x). \tag{2.13}$$

Let $\mathbb{T}$ and $S$ be Polish spaces, $\lambda \in \mathscr{P}(\mathbb{T})$ and $E$ be a Banach space. We say that $h : \mathbb{T}\times S \to E$ is a Carathéodory function if

$$\text{for } \lambda\text{-a.e. } t \in \mathbb{T}, \qquad x \mapsto h(t,x) \text{ is continuous,}$$
$$\forall\, x \in S, \qquad t \mapsto h(t,x) \text{ is } \lambda\text{-measurable.}$$

Let us now recall the definition of *Young measure* (see [7, 20]) and a density result which will turn out to be a crucial tool in our treatment.

**Definition 2.8.** Let $\mathbb{T}$ and $S$ be Polish spaces and $\lambda \in \mathscr{P}(\mathbb{T})$. We say that $\nu \in \mathscr{P}(\mathbb{T} \times S)$ is a *Young measure* on $\mathbb{T} \times S$ if $\pi^1_\sharp\nu = \lambda$. Furthermore given $\nu^n, \nu \in \mathscr{P}(\mathbb{T} \times S)$ Young measures, we say that $\nu^n \xrightarrow{\mathrm{y}} \nu$ as $n \to +\infty$ if

$$\lim_{n\to+\infty} \int_{\mathbb{T}\times S} h(\tau,u)\,\mathrm{d}\nu^n(\tau,u) = \int_{\mathbb{T}\times S} h(\tau,u)\,\mathrm{d}\nu(\tau,u),$$

for any $h : \mathbb{T} \times S \to \mathbb{R}$ Carathéodory and bounded.

*Remark* 2.9. Let $\mathbb{T}$, $S$ be Polish spaces, $\lambda \in \mathscr{P}(\mathbb{T})$ and $\nu^n, \nu$ Young measures on $\mathbb{T} \times S$. Then $\nu^n \xrightarrow{\mathrm{y}} \nu$ in the sense of Definition 2.8 if and only if $\nu^n \to \nu$ weakly. One implication follows immediately from the definitions, while the other comes from [54, Theorem 7] (see also [20]).

We also recall that weak convergence in $\mathscr{P}([0,T] \times S)$ is induced by a distance $\delta$. When $S$ is compact, we can choose as $\delta$ any Wasserstein distance on $\mathscr{P}([0,T] \times S)$.

To any Borel map $u : \mathbb{T} \to S$ we can associate the Young measure $\nu := (i_{\mathbb{T}}, u)_\sharp\lambda$, which is concentrated on the graph of $u$. In this case, $\nu$ can be written as $\nu = \delta_{u(\tau)} \otimes \lambda$ and, using the disintegration Theorem 2.6, we have that $\nu_\tau = \delta_{u(\tau)}$ for $\lambda$-a.e. $\tau \in \mathbb{T}$. Given a Young measure $\nu$, in general the disintegration $\nu_\tau$ of $\nu$ w.r.t. $\lambda$ is not of the form $\delta_{u(\tau)}$ on a set of $\lambda$ positive measure, for some $u : \mathbb{T} \to S$. The following classical Lemma states that the Young measures

induced by maps are "dense", in the set of Young measures, provided $\lambda$ is non atomic. We say that a measure $\lambda \in \mathscr{P}(\mathbb{T})$ is non atomic if $\lambda(\{\tau\}) = 0$ for any $\tau \in \mathbb{T}$.

**Lemma 2.10** (see [20, Theorem 2.2.3]). *Let $\mathbb{T}$ and $S$ be Polish spaces and $\lambda \in \mathscr{P}(\mathbb{T})$ non atomic. If $\nu \in \mathscr{P}(\mathbb{T} \times S)$ is a Young measure, then there exists a sequence of Borel maps $u^n : \mathbb{T} \to S$ such that*

$$\nu^n := (i_{\mathbb{T}}, u^n)_\sharp \lambda \xrightarrow{\mathcal{Y}} \nu = \nu_\tau \otimes \lambda.$$

*Precisely,*

$$\lim_{n\to+\infty} \int_{\mathbb{T}} h(\tau, u^n(\tau))\, \mathrm{d}\lambda(\tau) = \int_{\mathbb{T}} \int_S h(\tau, u)\, \mathrm{d}\nu_\tau(u)\, \mathrm{d}\lambda(\tau), \tag{2.14}$$

*for every $h : \mathbb{T} \times S \to \mathbb{R}$ Carathéodory and bounded.*

## 3. Structural assumptions for the dynamics of the optimal control problems

In this section we collect our main structural assumptions on the system $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ characterizing the dynamics and the cost of the control problems under study, where $U$ is the space of controls, $f$ is the vector field driving the particles motion, $\mathcal{C}$ and $\mathcal{C}_T$ are the running and terminal cost functionals.

We fix $p \in [1, +\infty)$ and denote by $\mathsf{d}_p$ the following metric on $\mathbb{R}^d \times \mathscr{P}_p(\mathbb{R}^d)$:

$$\mathsf{d}_p((x,\mu),(y,\nu)) := \left(|x-y|^p + W_p^p(\mu,\nu)\right)^{1/p}.$$

**Assumption 3.1** (Basic Assumption). We assume that the system $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfies:

(A.1) $U$ is a compact metrizable space;

(A.2) $f : \mathbb{R}^d \times U \times \mathscr{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ is continuous and Lipschitz continuous w.r.t. the metric $\mathsf{d}_p$, uniformly in $u \in U$. Precisely, there exists $L > 0$ such that

$$|f(x,u,\mu) - f(y,u,\nu)| \le L\, \mathsf{d}_p((x,\mu),(y,\nu)) \ , \tag{3.1}$$

for every $u \in U$ and $(x,\mu),(y,\nu) \in \mathbb{R}^d \times \mathscr{P}_p(\mathbb{R}^d)$.

(A.3) $\mathcal{C} : \mathbb{R}^d \times U \times \mathscr{P}_p(\mathbb{R}^d) \to [0, +\infty)$ and $\mathcal{C}_T : \mathbb{R}^d \times \mathscr{P}_p(\mathbb{R}^d) \to [0, +\infty)$ are continuous functions such that

$$\begin{aligned} &\mathcal{C}(x,u,\mu) \le D\left(1 + |x|^p + \mathrm{m}_p^p(\mu)\right) \qquad \forall\, (x,u,\mu) \in \mathbb{R}^d \times U \times \mathscr{P}_p(\mathbb{R}^d) \\ &\mathcal{C}_T(x,\mu) \le D\left(1 + |x|^p + \mathrm{m}_p^p(\mu)\right) \qquad \forall\, (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_p(\mathbb{R}^d), \end{aligned} \tag{3.2}$$

for some $D > 0$.

*Remark* 3.2. From Assumption 3.1 it holds

$$|f(x,u,\mu)| \le C\left(1 + |x| + \mathrm{m}_p(\mu)\right), \qquad \forall\, (x,u,\mu) \in \mathbb{R}^d \times U \times \mathscr{P}_p(\mathbb{R}^d), \tag{3.3}$$

for some $C > 0$. Indeed, it is sufficient to choose $(y,\nu) = (0, \delta_0)$ in (3.1) and observe that $f(0,u,\delta_0)$ is bounded and $W_p(\mu,\delta_0) = \mathrm{m}_p(\mu)$.

Concerning item (A.1) of Assumption 3.1, let us recall the following result.

**Proposition 3.3.** *If $U$ is compact metrizable space then, for every distance $d_U$ inducing the original topology of $U$, there exists a separable Banach space $V$ and an isometry $j : U \to V$. In particular, the image $j(U)$ is a compact subset of $V$.*

*Proof.* Fix a point $u_0 \in U$ and consider the Banach space $B := \{F \in \mathrm{Lip}(U) : F(u_0) = 0\}$ endowed with the norm

$$\|F\|_B := \sup_{u,v\in U, u\neq v} \frac{|F(u) - F(v)|}{d_U(u,v)}.$$

Denoting by $B'$ the dual space of $B$, we define the map $j : U \to B'$ by $\langle j(u), F\rangle_{B',B} := F(u)$. By the definition of dual norm, it is immediate to check that

$$\|j(u) - j(v)\|_{B'} \le d_U(u,v), \quad \forall\, u,v \in U$$

On the other hand, evaluating $\langle j(u) - j(v), F\rangle_{B',B}$ with $F(z) := d_U(z,u) - d_U(u,u_0)$, we obtain that

$$\|j(u) - j(v)\|_{B'} = d_U(u,v),$$

so that $j$ is an isometry from $U$ to $j(U) \subset B'$. We eventually set $V := \overline{\mathrm{span}(j(U))}^{B'}$, which is a separable Banach space since $U$, and therefore $j(U)$, is separable. □

When specified, we will assume the following further hypothesis.

**Assumption 3.4** (Convexity Assumption)**.** We say that $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfies the *convexity assumption* if $\mathbb{S}$ satisfies Assumption 3.1 and

(C.1) $U$ is a compact convex subset of a separable Banach space $V$;
(C.2) for any $x \in \mathbb{R}^d$ and $\mu \in \mathscr{P}_p(\mathbb{R}^d)$, the map $u \mapsto f(x,u,\mu)$ satisfies the *affinity condition*:

$$f(x, \alpha u + (1-\alpha)v, \mu) = \alpha f(x,u,\mu) + (1-\alpha) f(x,v,\mu), \quad \forall\, u,v \in U,\ \forall\, \alpha \in [0,1];$$

(C.3) for any $x \in \mathbb{R}^d$ and $\mu \in \mathscr{P}_p(\mathbb{R}^d)$ the map $u \mapsto \mathcal{C}(x,u,\mu)$ is convex:

$$\mathcal{C}(x, \alpha u + (1-\alpha)v, \mu) \le \alpha \mathcal{C}(x,u,\mu) + (1-\alpha)\mathcal{C}(x,v,\mu), \quad \forall\, u,v \in U,\ \forall\, \alpha \in [0,1].$$

3.1. **The relaxed setting.** For later use, we define a so-called *relaxation/lifting* of $\mathbb{S}$ as follows.

**Definition 3.5.** Given the system $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfying Assumption 3.1, we define $\mathbb{S}' = (\mathscr{U}, \mathscr{F}, \mathscr{C}, \mathscr{C}_T)$ as follows:

(i) $\mathscr{U} := \mathscr{P}(U)$;
(ii) $\mathscr{F} : \mathbb{R}^d \times \mathscr{U} \times \mathscr{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ with

$$\mathscr{F}(x,\sigma,\mu) := \int_U f(x,u,\mu)\,\mathrm{d}\sigma(u);$$

(iii) $\mathscr{C} : \mathbb{R}^d \times \mathscr{U} \times \mathscr{P}_p(\mathbb{R}^d) \to [0,+\infty)$ with

$$\mathscr{C}(x,\sigma,\mu) := \int_U \mathcal{C}(x,u,\mu)\,\mathrm{d}\sigma(u).$$

(iv) $\mathscr{C}_T := \mathcal{C}_T$.

**Proposition 3.6.** *If $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfies Assumption 3.1, then its relaxation $\mathbb{S}' = (\mathscr{U}, \mathscr{F}, \mathscr{C}, \mathcal{C}_T)$, given in Definition 3.5, satisfies the Convexity Assumption 3.4. Moreover, defining $D_U := \{\delta_u : u \in U\} \subset \mathscr{U}$, the maps $\mathscr{F}$ and $\mathscr{C}$ restricted to $\mathbb{R}^d \times D_U \times \mathscr{P}_p(\mathbb{R}^d)$ coincide with $f$ and $\mathcal{C}$ respectively.*

*Proof.* The space $\mathscr{U} := \mathscr{P}(U)$ can be identified with a subset of the dual space $B'$, where $B$ is the Banach space $B := \{F \in \mathrm{Lip}(U) : F(u_0) = 0 \text{ for some } u_0 \in U\}$. The identification is given associating to $\sigma \in \mathscr{P}(U)$ the continuous linear functional $F \mapsto \int_U F(u)\,\mathrm{d}\sigma(u)$. With this identification, the norm in $\mathscr{P}(U)$ is given by

$$\|\sigma\| = \sup_{F \in B, \|F\|_B \le 1} \int_U F(u)\,\mathrm{d}\sigma(u).$$

By the Kantorovich-Rubinstein Theorem (see e.g. [55, Theorem 1.14]) it holds that $\|\sigma\| = W_1(\sigma, \delta_{u_0})$ and $\|\sigma^1 - \sigma^2\| = W_1(\sigma^1, \sigma^2)$. Hence, the topology on $\mathscr{P}(U)$ induced by $B'$ coincides with the topology induced by the Wasserstein distance $W_1$. Since $U$ is compact, this coincides with the topology induced by the weak convergence. By Prokhorov Theorem, $\mathscr{P}(U)$ is compact. Finally, $(\mathscr{P}(U), \|\cdot\|)$ is a separable Banach space thanks to the separability of the (complete) metric space $(\mathscr{P}(U), W_1)$. The convexity of $\mathscr{U}$, the affinity of $\mathscr{F}$ and the convexity of $\mathscr{C}$ with respect to $\sigma$ easily follows from their definitions. □

## 4. Lagrangian optimal control problem

In this section we deal with a (finite-horizon) optimal control problem in Lagrangian formulation. It relies on a system $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfying Assumptions 3.1 and on a probability space $(\Omega, \mathfrak{B}, \mathbb{P})$, whose elements act as parameters of the particles. We also fix a final time horizon $T > 0$ and we denote with $\mathrm{Leb}_{[0,T]}$ the $\sigma$-algebra of Lebesgue measurable sets on $[0, T]$ and with $\mathcal{L}_T$ the normalized Lebesgue measure restricted to $[0, T]$. Recall that $\mathrm{M}([0, T] \times \Omega; U)$ denotes the set of measurable functions with respect to the product $\sigma$-algebra $\mathrm{Leb}_{[0,T]} \otimes \mathfrak{B}$.

**Definition 4.1** (Lagrangian optimal control problem ($\mathbf{L}$))**.** Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space.

Given $X_0 \in L^p(\Omega; \mathbb{R}^d)$, we say that $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ if

(i) $u \in \mathrm{M}([0, T] \times \Omega; U)$;

(ii) $X \in L^p(\Omega; \mathrm{AC}^p([0, T]; \mathbb{R}^d))$ and for $\mathbb{P}$-a.e. $\omega \in \Omega$, $X(\omega)$ is a solution of the following Cauchy problem

$$\begin{cases} \dot{X}_t(\omega) = f(X_t(\omega), u_t(\omega), (X_t)_\sharp \mathbb{P}), & \text{for } \mathcal{L}_T\text{-a.e. } t \in (0, T) \\ X_{|t=0}(\omega) = X_0(\omega), \end{cases} \tag{4.1}$$

where $X_t : \Omega \to \mathbb{R}^d$ is defined by $X_t(\omega) := X(t, \omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$.

We refer to $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ as to an *admissible pair*, with $X$ a *trajectory* and $u$ a *control*. We define the *cost functional* $J_{\mathbf{L}} : L^p(\Omega; C([0, T]; \mathbb{R}^d)) \times \mathrm{M}([0, T] \times \Omega; U) \to [0, +\infty)$, by

$$J_{\mathbf{L}}(X, u) := \int_\Omega \int_0^T \mathcal{C}(X_t(\omega), u_t(\omega), (X_t)_\sharp \mathbb{P}) \, \mathrm{d}t \, \mathrm{d}\mathbb{P}(\omega) + \int_\Omega \mathcal{C}_T(X_T(\omega), (X_T)_\sharp \mathbb{P}) \, \mathrm{d}\mathbb{P}(\omega),$$

and the *value function* $V_{\mathbf{L}} : L^p(\Omega; \mathbb{R}^d) \to [0, +\infty)$ by

$$V_{\mathbf{L}}(X_0) := \inf \{ J_{\mathbf{L}}(X, u) \, : \, (X, u) \in \mathcal{A}_{\mathbf{L}}(X_0) \} . \tag{4.2}$$

In the following, $\mathbf{L}(\Omega, \mathfrak{B}, \mathbb{P}; \mathbb{S})$ denotes the Lagrangian problem given in Definition 4.1. We will frequently shorten the notation to $\mathbf{L}(\Omega, \mathfrak{B}, \mathbb{P})$ when the system $\mathbb{S}$ is clear from the context.

*Remark* 4.2. Observe that, thanks to condition (3.2), the functional $J_{\mathbf{L}}$ is finite. Moreover, from Proposition 4.8 below it follows that $\mathcal{A}_{\mathbf{L}}(X_0) \neq \emptyset$, for any $X_0 \in L^p(\Omega; \mathbb{R}^d)$, and so the value function $V_{\mathbf{L}}$ is well defined. We point out that existence of minimizers for the Lagrangian problem is not guaranteed in general, even under the Convexity Assumption 3.4. This will be further discussed in Section 8.3.

*Remark* 4.3. In view of Proposition A.3 in Appendix A, we will frequently identify $X \in L^p(\Omega; \mathrm{AC}^p([0, T]; \mathbb{R}^d))$ and $X \in \mathrm{AC}^p([0, T]; L^p(\Omega; \mathbb{R}^d))$, depending on the convenience.

Let us introduce a suitable equivalence relation among Lagrangian problems when the parametrization space is varying.

**Definition 4.4** (Equivalence of Lagrangian problems)**.** Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Let $(\Omega_1, \mathfrak{B}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathfrak{B}_2, \mathbb{P}_2)$ be probability spaces. We say that $\mathbf{L}_1 := \mathbf{L}(\Omega_1, \mathfrak{B}_1, \mathbb{P}_1; \mathbb{S})$ and $\mathbf{L}_2 := \mathbf{L}(\Omega_2, \mathfrak{B}_2, \mathbb{P}_2; \mathbb{S})$ are *equivalent* (and we write $\mathbf{L}_1 \sim \mathbf{L}_2$) if

(i) for every $X_0^1 \in L^p(\Omega_1; \mathbb{R}^d)$ and every $(X^1, u^1) \in \mathcal{A}_{\mathbf{L}_1}(X_0^1)$ there exist $X_0^2 \in L^p(\Omega_2; \mathbb{R}^d)$ and $(X^2, u^2) \in \mathcal{A}_{\mathbf{L}_2}(X_0^2)$ such that

$$J_{\mathbf{L}_1}(X^1, u^1) = J_{\mathbf{L}_2}(X^2, u^2), \qquad V_{\mathbf{L}_1}(X_0^1) = V_{\mathbf{L}_2}(X_0^2);$$

(ii) for every $X_0^2 \in L^p(\Omega_2; \mathbb{R}^d)$ and every $(X^2, u^2) \in \mathcal{A}_{\mathbf{L}_2}(X_0^2)$ there exist $X_0^1 \in L^p(\Omega_1; \mathbb{R}^d)$ and $(X^1, u^1) \in \mathcal{A}_{\mathbf{L}_1}(X_0^1)$ such that

$$J_{\mathbf{L}_2}(X^2, u^2) = J_{\mathbf{L}_1}(X^1, u^1), \qquad V_{\mathbf{L}_2}(X_0^2) = V_{\mathbf{L}_1}(X_0^1).$$

*Remark* 4.5. The relation $\sim$ of Definition 4.4 is an equivalence relation on the set of Lagrangian problems $\{\mathbf{L}(\Omega, \mathfrak{B}, \mathbb{P}) : (\Omega, \mathfrak{B}, \mathbb{P}) \text{ probability space}\}$.

**Proposition 4.6.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Let $(\Omega_1, \mathfrak{B}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathfrak{B}_2, \mathbb{P}_2)$ be probability spaces. Suppose there exist measurable maps $\psi : \Omega_1 \to \Omega_2$ and $\phi : \Omega_2 \to \Omega_1$ such that $\psi_\sharp \mathbb{P}_1 = \mathbb{P}_2$, $\phi_\sharp \mathbb{P}_2 = \mathbb{P}_1$ and*

$$\forall\, X_0^1 \in L^p(\Omega_1; \mathbb{R}^d) \text{ it holds } X_0^1 = X_0^1 \circ \phi \circ \psi\,; \tag{4.3}$$

$$\forall\, X_0^2 \in L^p(\Omega_2; \mathbb{R}^d) \text{ it holds } X_0^2 = X_0^2 \circ \psi \circ \phi. \tag{4.4}$$

*Then $\mathbf{L}(\Omega_1, \mathfrak{B}_1, \mathbb{P}_1; \mathbb{S}) \sim \mathbf{L}(\Omega_2, \mathfrak{B}_2, \mathbb{P}_2; \mathbb{S})$.*

*Proof.* For every $(X^1, u^1) \in \mathcal{A}_{\mathbf{L}_1}(X_0^1)$, we define $X^2 := X^1 \circ \phi$, and $u^2(t, \omega_2) := u^1(t, \phi(\omega_2))$, for every $(t, \omega_2) \in [0, T] \times \Omega_2$. Using that $\phi_\sharp \mathbb{P}_2 = \mathbb{P}_1$, it easily follows that $(X^2, u^2) \in \mathcal{A}_{\mathbf{L}_2}(X_0^1 \circ \phi)$ and $J_{\mathbf{L}_1}(X^1, u^1) = J_{\mathbf{L}_2}(X^2, u^2)$. Hence, for every $X_0^1 \in L^p(\Omega_1; \mathbb{R}^d)$, we have

$$V_{\mathbf{L}_1}(X_0^1) \geq V_{\mathbf{L}_2}(X_0^1 \circ \phi). \tag{4.5}$$

Analogously, for every $(X^2, u^2) \in \mathcal{A}_{\mathbf{L}_2}(X_0^2)$, we define $X^1 := X^2 \circ \psi$, and $u^1(t, \omega_1) := u^2(t, \psi(\omega_1))$, for every $(t, \omega_1) \in [0, T] \times \Omega_1$. So that, from $\psi_\sharp \mathbb{P}_1 = \mathbb{P}_2$ it holds $(X^1, u^1) \in \mathcal{A}_{\mathbf{L}_1}(X_0^2 \circ \psi)$ and $J_{\mathbf{L}_1}(X^1, u^1) = J_{\mathbf{L}_2}(X^2, u^2)$. Moreover, for every $X_0^2 \in L^p(\Omega_2; \mathbb{R}^d)$ we have

$$V_{\mathbf{L}_2}(X_0^2) \geq V_{\mathbf{L}_1}(X_0^2 \circ \psi). \tag{4.6}$$

The combination of (4.5) and (4.6) gives

$$V_{\mathbf{L}_2}(X_0^2) \geq V_{\mathbf{L}_1}(X_0^2 \circ \psi) \geq V_{\mathbf{L}_2}(X_0^2 \circ \psi \circ \phi)$$

hence, using (4.3) we have

$$V_{\mathbf{L}_2}(X_0^2) = V_{\mathbf{L}_1}(X_0^2 \circ \psi). \tag{4.7}$$

Thanks to (4.4) and (4.7) we finally get

$$V_{\mathbf{L}_1}(X_0^1) = V_{\mathbf{L}_1}(X_0^1 \circ \phi \circ \psi) = V_{\mathbf{L}_2}(X_0^1 \circ \phi).$$

□

*Remark* 4.7.

(1) Notice that the assumptions of Proposition 4.6 are satisfied if there exists a bijective function $\psi : \Omega_1 \to \Omega_2$ such that $\psi$ and $\psi^{-1}$ are measurable and $\psi_\sharp \mathbb{P}_1 = \mathbb{P}_2$. Indeed, it sufficies to choose $\phi = \psi^{-1}$.

(2) Proposition 4.6 still holds when the maps $\psi$ and $\phi$ are defined up to sets of null measure, meaning that

$$\psi : \Omega_1 \setminus \mathcal{N}_1 \to \Omega_2 \setminus \mathcal{N}_2, \quad \phi : \Omega_2 \setminus \mathcal{N}_2 \to \Omega_1 \setminus \mathcal{N}_1$$

for some $\mathcal{N}_1 \in \mathfrak{B}_1$ such that $\mathbb{P}_1(\mathcal{N}_1) = 0$ and $\mathcal{N}_2 \in \mathfrak{B}_2$ such that $\mathbb{P}_2(\mathcal{N}_2) = 0$.

4.1. **Basic results.** Here we collect some properties of the Lagrangian problem. In particular, we show existence and uniqueness of solutions, a priori estimates, compactness for the associated laws and we derive a stability result for trajectories and cost when initial data and control converge in a suitable sense.

**Proposition 4.8** (Existence and uniqueness)**.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space. Let $X_0 \in L^p(\Omega; \mathbb{R}^d)$ and $u \in \mathrm{M}([0, T] \times \Omega; U)$ be given. Then there exists a unique $X \in L^p(\Omega; \mathrm{AC}^p([0, T]; \mathbb{R}^d))$ such that $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$. Moreover, if $(X^i, u^i) \in \mathcal{A}_{\mathbf{L}}(X_0)$, $i = 1, 2$, and $u^1 = u^2$ $\mathcal{L}_T \otimes \mathbb{P}$-a.e., then $X^1 = X^2$.*

*Proof.* We define $F_u : [0, T] \times L^p(\Omega; \mathbb{R}^d) \to L^p(\Omega; \mathbb{R}^d)$ by

$$F_u(t, Z)(\omega) := f(Z(\omega), u(t, \omega), Z_\sharp \mathbb{P}). \tag{4.8}$$

We observe that the continuity of $f$ and the measurability of $u$ imply that $F_u$ is a Carathéodory function. Moreover, by (3.1) and (2.5), $F_u$ satisfies condition (A.10). Since $F_u(t, 0)(\omega) = f(0, u_t(\omega), \delta_0)$, by continuity of $f$ and compactness of $U$ it follows that $F_u$ satisfies (A.11).

Theorem A.5 with the choice $E = L^p(\Omega;\mathbb{R}^d)$ and $F = F_u$ yields the existence and uniqueness of a curve $X \in \mathrm{AC}^p([0,T];L^p(\Omega;\mathbb{R}^d))$ solving

$$X_t = X_0 + \int_0^t F_u(s,X_s)\,\mathrm{d}s, \qquad \forall\, t\in[0,T].$$

Thanks to Proposition A.7 we finally get $X \in L^p(\Omega;\mathrm{AC}^p([0,T];\mathbb{R}^d))$ which is the unique solution of (4.1). The last assertion follows from the equality $F_{u^1}(t,Z) = F_{u^2}(t,Z)$ for $\mathcal{L}_T$-a.e. $t\in[0,T]$ and for every $Z \in L^p(\Omega;\mathbb{R}^d)$. □

**Proposition 4.9** (A priori estimates)**.** *Let $\mathbb{S} := (U,f,\mathcal{C},\mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega,\mathfrak{B},\mathbb{P})$ be a probability space. Let $X_0 \in L^p(\Omega;\mathbb{R}^d)$ and $(X,u)\in\mathcal{A}_{\mathbf{L}}(X_0)$. Then there exist $C$ and $C_T$ independent of $u$ and $X_0$ such that*

$$\sup_{t\in[0,T]} \|X_t\|_{L^p(\Omega;\mathbb{R}^d)} \le e^{2CT}\left(\|X_0\|_{L^p(\Omega;\mathbb{R}^d)} + CT\right), \tag{4.9}$$

$$\|X_t - X_s\|_{L^p(\Omega;\mathbb{R}^d)} \le C_T\,|t-s|\left(1+\|X_0\|_{L^p(\Omega;\mathbb{R}^d)}\right) \qquad \forall\, s,t\in[0,T], \tag{4.10}$$

$$\sup_{t\in[0,T]} |X_t(\omega)| \le e^{CT}\left(|X_0(\omega)| + C_T(1+\|X_0\|_{L^p(\Omega;\mathbb{R}^d)})\right), \quad \textit{for } \mathbb{P}\textit{-a.e. } \omega\in\Omega. \tag{4.11}$$

*Proof.* The estimates (4.9) and (4.10) follows from (A.13) and (A.14) for $F = F_u$ defined in (4.8) and $E = L^p(\Omega;\mathbb{R}^d)$ .

In order to prove (4.11) we write (4.1) in integral form

$$X_t(\omega) = X_0(\omega) + \int_0^t f(X_s(\omega),u(s,\omega),(X_s)_\sharp\mathbb{P})\,\mathrm{d}s, \quad \forall\, t\in[0,T] \quad \text{for } \mathbb{P}\text{-a.e. } \omega\in\Omega. \tag{4.12}$$

Then by (3.3) we have

$$\begin{aligned}
|X_t(\omega)| &= \left|X_0(\omega) + \int_0^t f(X_s(\omega),u(s,\omega),(X_s)_\sharp\mathbb{P})\,\mathrm{d}s\right| \\
&\le |X_0(\omega)| + \int_0^t |f(X_s(\omega),u(s,\omega),(X_s)_\sharp\mathbb{P})|\,\mathrm{d}s \\
&\le |X_0(\omega)| + \int_0^t C\left(1+|X_s(\omega)| + \|X_s\|_{L^p(\Omega;\mathbb{R}^d)}\right)\mathrm{d}s.
\end{aligned}$$

Using (4.9) and Gronwall inequality we obtain (4.11).

□

In the following Lemma, we derive a compactness result for the laws of Lagrangian trajectories, when the initial data belong to a compact subset of $L^p(\Omega;\mathbb{R}^d)$.

**Lemma 4.10.** *Let $\mathbb{S} := (U,f,\mathcal{C},\mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega,\mathfrak{B},\mathbb{P})$ be a probability space. Let $K \subseteq L^p(\Omega;\mathbb{R}^d)$ compact. Then the set*

$$\mathcal{K}_K := \left\{\mu \in \mathrm{AC}([0,T];\mathscr{P}_p(\mathbb{R}^d)) : \mu_t = (X_t)_\sharp\mathbb{P},\ (X,u)\in\mathcal{A}_{\mathbf{L}}(X_0),\ X_0\in K\right\} \tag{4.13}$$

*is relatively compact in $C([0,T];\mathscr{P}_p(\mathbb{R}^d))$.*

*Proof.* Let $\{\mu^n\}_{n\in\mathbb{N}} \subset \mathcal{K}_K$ be a sequence. By definition, there exist $(X^n,u^n)\in\mathcal{A}_{\mathbf{L}}(X_0^n)$, $X_0^n\in K$ such that $\mu_t = (X_t)_\sharp\mathbb{P}$ for all $t\in[0,T]$. Since $\sup_{n\in\mathbb{N}}\|X_0^n\|_{L^p(\Omega;\mathbb{R}^d)} < +\infty$, by the estimate (4.11) there exits a constant $C>0$ such that

$$|X_t^n(\omega)|^p \le C\left(1+|X_0^n(\omega)|^p\right), \quad \forall\, n\in\mathbb{N},\ \forall\, t\in[0,T],\ \text{for } \mathbb{P}\text{-a.e. } \omega\in\Omega. \tag{4.14}$$

Since $K$ is compact in $L^p(\Omega;\mathbb{R}^d)$, there exists an admissible $\psi:[0,+\infty)\to[0,+\infty)$, according to Definition 2.3, such that

$$\sup_{n\in\mathbb{N}}\int_\Omega \psi(|X_0^n(\omega)|^p)\,\mathrm{d}\mathbb{P}(\omega)<+\infty.$$

By the doubling and monotonicity property of $\psi$ and (4.14) we have

$$\psi(|X_t^n(\omega)|^p)\le C\left(1+\psi(|X_0^n(\omega)|^p)\right),\quad \forall\, n\in\mathbb{N},\ \forall\, t\in[0,T],\ \text{for } \mathbb{P}\text{-a.e. } \omega\in\Omega.$$

and then

$$\sup_{t\in[0,T],n\in\mathbb{N}}\int_\Omega \psi(|X_t^n(\omega)|^p)\,\mathrm{d}\mathbb{P}(\omega)<+\infty,$$

that can be rewritten as

$$\sup_{t\in[0,T],n\in\mathbb{N}}\int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_t^n(x)<+\infty.$$

By Proposition 2.4 there exists a compact $\mathscr{K}\subset\mathscr{P}_p(\mathbb{R}^d)$ such that $\mu_t^n\in\mathscr{K}$ for any $t\in[0,T]$ and $n\in\mathbb{N}$.

Moreover, by (4.10) and the boundedness of $\|X_0^n\|_{L^p(\Omega;\mathbb{R}^d)}$, there exists $C>0$ such that

$$W_p(\mu_t^n,\mu_s^n)\le\|X_t^n-X_s^n\|_{L^p(\Omega;\mathbb{R}^d)}\le C|t-s|,\quad \forall\, s,t\in[0,T],\ \forall\, n\in\mathbb{N}.$$

We can thus apply Ascoli-Arzelà theorem in $C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ to conclude. □

We conclude the subsection proving a first stability result for the Lagrangian problem.

**Proposition 4.11** (Stability for **L**). *Let $\mathbb{S}:=(U,f,\mathcal{C},\mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega,\mathfrak{B},\mathbb{P})$ be a probability space. Let $X_0\in L^p(\Omega;\mathbb{R}^d)$ and $(X,u)\in\mathcal{A}_{\mathbf{L}}(X_0)$. Let $X_0^n\in L^p(\Omega;\mathbb{R}^d)$ be a sequence such that $\|X_0^n-X_0\|_{L^p(\Omega;\mathbb{R}^d)}\to 0$, as $n\to+\infty$. If $(X^n,u^n)\in\mathcal{A}_{\mathbf{L}}(X_0^n)$, for any $n\in\mathbb{N}$, and $u^n\to u$ in $\mathcal{L}_T\otimes\mathbb{P}$-measure as $n\to+\infty$, then*

$$\sup_{t\in[0,T]}\|X_t^n-X_t\|_{L^p(\Omega;\mathbb{R}^d)}\to 0,\quad \text{as } n\to+\infty, \tag{4.15}$$

*and*

$$J_{\mathbf{L}}(X^n,u^n)\to J_{\mathbf{L}}(X,u),\quad \text{as } n\to+\infty. \tag{4.16}$$

*Proof.* In order to prove (4.15) we apply Proposition A.6 with the choice $E=L^p(\Omega;\mathbb{R}^d)$, $F=F_u$ and $F^n=F_{u^n}$, defined as in (4.8). We have to check that (A.17) holds. Defining $G^n,G:[0,T]\times\Omega\to\mathbb{R}^d$ by $G^n(t,\omega):=F_{u^n}(t,X_t)(\omega)$ and $G(t,\omega):=F_u(t,X_t)(\omega)$, it is sufficient to prove that $G^n\to G$ in $L^p([0,T]\times\Omega;\mathbb{R}^d)$. Since $u^n$ converges to $u$ in $\mathcal{L}_T\otimes\mathbb{P}$-measure, there exists a subsequence $u^{n_k}$ such that $u_t^{n_k}(\omega)$ converges to $u_t(\omega)$ for $\mathcal{L}_T\otimes\mathbb{P}$-a.e. $(t,\omega)\in[0,T]\times\Omega$. By the continuity of $f$, we have that

$$|G^{n_k}(t,\omega)-G(t,\omega)|=|F_{u^{n_k}}(t,X_t)(\omega)-F_u(t,X_t)(\omega)|\to 0$$

for $\mathcal{L}_T\otimes\mathbb{P}$-a.e. $(t,\omega)\in[0,T]\times\Omega$. Moreover

$$|G^n(t,\omega)-G(t,\omega)|^p=|F_{u^n}(t,X_t)(\omega)-F_u(t,X_t)(\omega)|^p\le C\left(1+|X_t(\omega)|^p+\|X_t\|^p_{L^p(\Omega;\mathbb{R}^d)}\right).$$

By dominated convergence we conclude that $G^{n_k}\to G$ in $L^p([0,T]\times\Omega;\mathbb{R}^d)$. Since the limit is independent of the subsequence, we conclude that

$$\int_0^T\int_\Omega |F_{u^n}(t,X_t)(\omega)-F_u(t,X_t)(\omega)|^p\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}t\to 0.$$

Let us prove (4.16). For any $t\in[0,T]$, we use the notation $\mu_t^n:=(X_t^n)_\sharp\mathbb{P}$ and $\mu_t:=(X_t)_\sharp\mathbb{P}$. By (4.15) we have

$$\sup_{t\in[0,T]}W_p(\mu_t^n,\mu_t)\to 0,\quad \text{as } n\to+\infty. \tag{4.17}$$

We focus on the running cost $\mathcal{C}$. Since

$$\begin{aligned}&\left|\int_0^T\int_\Omega\left(\mathcal{C}(X_t^n(\omega),u_t^n(\omega),\mu_t^n)-\mathcal{C}(X_t(\omega),u_t(\omega),\mu_t)\right)\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}t\right|\\&\leq\int_0^T\int_\Omega\left|\mathcal{C}(X_t^n(\omega),u_t^n(\omega),\mu_t^n)-\mathcal{C}(X_t(\omega),u_t(\omega),\mu_t)\right|\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}t,\end{aligned}\tag{4.18}$$

defining $H^n, H:[0,T]\times\Omega\to\mathbb{R}^d$ by $H^n(t,\omega):=\mathcal{C}(X_t^n(\omega),u_t^n(\omega),\mu_t^n)$ and $H(t,\omega):=\mathcal{C}(X_t(\omega),u_t(\omega),\mu_t)$, it is sufficient to prove that $H^n\to H$ in $L^1([0,T]\times\Omega;\mathbb{R}^d)$. Since $u^n$ converges to $u$ in $\mathcal{L}_T\otimes\mathbb{P}$-measure, (4.15) and (4.17) hold, and $\mathcal{C}$ is continuous, then there exists a subsequence $H^{n_k}$ such that $H^{n_k}(t,\omega)$ converges to $H(t,\omega)$ for $\mathcal{L}_T\otimes\mathbb{P}$-a.e. $(t,\omega)\in[0,T]\times\Omega$. Moreover, by the growth assumptions (3.2) we have

$$H^n(t,\omega)\leq C(1+|X_t^n(\omega)|^p+\|X_t^n\|^p_{L^p(\Omega;\mathbb{R}^d)}).$$

By a variant of the dominated convergence Theorem (see Theorem 1.20 in [31]) we conclude that $H^{n_k}\to H$ in $L^1([0,T]\times\Omega;\mathbb{R}^d)$. For the same argument as before we obtain that the whole sequence $H^n\to H$ in $L^1([0,T]\times\Omega;\mathbb{R}^d)$.

The proof that

$$\int_\Omega\mathcal{C}_T(X_T^n(\omega),\mu_T^n)\,\mathrm{d}\mathbb{P}(\omega)\to\int_\Omega\mathcal{C}_T(X_T(\omega),\mu_T)\,\mathrm{d}\mathbb{P}(\omega),\quad\text{as } n\to+\infty$$

follows from the same argument. □

**Proposition 4.12** (Upper semicontinuity of the value function)**.** *Let $\mathbb{S}:=(U,f,\mathcal{C},\mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega,\mathfrak{B},\mathbb{P})$ be a probability space. If $X_0^n,X_0\in L^p(\Omega;\mathbb{R}^d)$ satisfy $\|X_0^n-X_0\|_{L^p(\Omega;\mathbb{R}^d)}\to 0$ as $n\to+\infty$, then*

$$\limsup_{n\to+\infty}V_{\mathbf{L}}(X_0^n)\leq V_{\mathbf{L}}(X_0).$$

*Proof.* Let $\varepsilon>0$ and $(X^\varepsilon,u^\varepsilon)\in\mathcal{A}_{\mathbf{L}}(X_0)$ such that $J_{\mathbf{L}}(X^\varepsilon,u^\varepsilon)\leq V_{\mathbf{L}}(X_0)+\varepsilon$. By Proposition 4.8, for any $n\in\mathbb{N}$ there exists $X^{\varepsilon,n}$ such that $(X^{\varepsilon,n},u^\varepsilon)\in\mathcal{A}_{\mathbf{L}}(X_0^n)$. By Proposition 4.11, $J_{\mathbf{L}}(X^{\varepsilon,n},u^\varepsilon)\to J_{\mathbf{L}}(X^\varepsilon,u^\varepsilon)$, as $n\to+\infty$. Hence

$$\limsup_{n\to+\infty}V_{\mathbf{L}}(X_0^n)\leq\limsup_{n\to+\infty}J_{\mathbf{L}}(X^{\varepsilon,n},u^\varepsilon)=J_{\mathbf{L}}(X^\varepsilon,u^\varepsilon)\leq V_{\mathbf{L}}(X_0)+\varepsilon.$$

Since $\varepsilon$ is arbitrary, we conclude. □

4.2. **Approximation by piecewise constant controls.** In this subsection, we approximate admissible controls for the Lagrangian problem with a sequence of suitable piecewise constant controls (i.e. measurable with respect to finite algebras of $\Omega$) so that the corresponding trajectories and costs converge. This is the content of Theorem 4.17. The same result is then rephrased in the context of finite particle approximations in Proposition 4.18.

Let $(\Omega,\bar{\mathfrak{B}},\mathbb{P})$ be a probability space with $\bar{\mathfrak{B}}$ a finite algebra. It can be shown that $\bar{\mathfrak{B}}$ induces a unique minimal (with respect to the inclusion) partition of $\Omega$, that we denote by

$$\mathcal{P}(\bar{\mathfrak{B}})=\{A_k:k=1,\dots,m\}.\tag{4.19}$$

Given a topological space $E$, observe that, since $\bar{\mathfrak{B}}$ is finite, $g\in\mathrm{M}((\Omega,\bar{\mathfrak{B}});(E,\mathcal{B}_E))$ if and only if $g$ is constant on the elements of $\mathcal{P}$.

Let us give the following definition.

**Definition 4.13.** Let $(\Omega,\bar{\mathfrak{B}},\mathbb{P})$ be a probability space with $\bar{\mathfrak{B}}$ a finite algebra and $\mathcal{P}(\bar{\mathfrak{B}})$ the associated unique minimal partition (4.19). Given $m:=\#\mathcal{P}(\bar{\mathfrak{B}})$, we define the probability space $(\Omega^m,\mathcal{S}(\Omega^m),\mathbb{P}^m)$, where $\Omega^m:=\{1,\dots,m\}$, $\mathcal{S}(\Omega^m)$ is the algebra generated by $(\{1\},\dots\{m\})$ and $\mathbb{P}^m(\{k\}):=\mathbb{P}(A_k)$, $k=1,\dots,m$.

**Proposition 4.14.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Let $(\Omega, \bar{\mathfrak{B}}, \mathbb{P})$ and $(\Omega^m, \mathcal{S}(\Omega^m), \mathbb{P}^m)$ as in Definition 4.13. Then the Lagrangian problems $\mathbf{L}_{\bar{\mathfrak{B}}} := \mathbf{L}(\Omega, \bar{\mathfrak{B}}, \mathbb{P})$ and $\mathbf{L}^m := \mathbf{L}(\Omega^m, \mathcal{S}(\Omega^m), \mathbb{P}^m)$ are equivalent in the sense of Definition 4.4.*

*Proof.* Let $\psi : \Omega \to \Omega^m$ the function given by

$$\psi(\omega) = k, \quad \text{if } \omega \in A_k, \quad k = 1, \dots, m$$

and $\phi : \Omega^m \to \Omega$ defined by

$$\phi(k) = \omega_k, \quad k = 1, \dots, m,$$

for a fixed $\omega_k \in A_k$. We prove that the maps $\psi$ and $\phi$ satisfy the assumptions of Proposition 4.6. Measurability of the map $\psi$ follows from the fact that $\psi^{-1}(\{k\}) = A_k$ for any $k = 1, \dots, m$, while the measurability of $\phi$ is trivial since $\Omega^m$ is equipped with the algebra $\mathcal{S}(\Omega^m)$. Moreover, it is immediate to verify that $\psi \circ \phi = i_{\Omega^m}$, which implies (4.3). We have to verify (4.4): given $X_0 \in L^p(\Omega; \mathbb{R}^d)$, we have that $X_0$ is constant on the elements $A_k$ of the partition $\mathcal{P}(\bar{\mathfrak{B}})$, hence it easily follows that $X_0 = X_0 \circ \phi \circ \psi$. Let us verify that $\psi_\sharp \mathbb{P} = \mathbb{P}^m$: for any $k = 1, \dots, m$, we have $(\psi_\sharp \mathbb{P})(\{k\}) = \mathbb{P}(\psi^{-1}(\{k\})) = \mathbb{P}(A_k) = \mathbb{P}^m(\{k\})$. Finally, we verify that $\phi_\sharp \mathbb{P}^m = \mathbb{P}$: for any measurable function $g : \Omega \to \mathbb{R}$, recalling that $g$ is piecewise constant on the elements of $\mathcal{P}(\bar{\mathfrak{B}})$, we have

$$\begin{aligned}\int_\Omega g(\omega)\, \mathrm{d}(\phi_\sharp \mathbb{P}^m)(\omega) &= \int_{\Omega^m} g(\phi(k))\, \mathrm{d}\mathbb{P}^m(k) = \sum_{k=1}^m g(\omega_k)\mathbb{P}(A_k) \\ &= \int_\Omega g(\omega)\, \mathrm{d}\mathbb{P}(\omega).\end{aligned}$$

□

We recall the notation $\mathcal{B}_{[0,T]}$ for the Borel $\sigma$-algebra on $[0, T]$.

**Definition 4.15.** [Finite Approximation Property] Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space. We say that the family of finite algebras $\mathfrak{B}^n \subset \mathfrak{B}$, $n \in \mathbb{N}$, satisfies the *finite approximation property* if for any Banach space $E$ and any $g \in L^1_{\mathbb{P}}(\Omega; E)$, there exists a sequence $g^n : \Omega \to E$ such that

- **(i)** $g^n$ is $\mathfrak{B}^n$-measurable for any $n \in \mathbb{N}$;
- **(ii)** $g^n(\Omega) \subset \overline{\mathrm{co}}\,(g(\Omega))$, where $\overline{\mathrm{co}}\,(g(\Omega))$ denotes the closed convex hull of $g(\Omega)$;
- **(iii)** $\|g^n - g\|_{L^1_{\mathbb{P}}(\Omega;E)} \to 0$, as $n \to +\infty$;
- **(iv)** if $G : [0, T] \times \Omega \to E$ is $(\mathcal{B}_{[0,T]} \otimes \mathfrak{B})$-measurable and $g_t(\cdot) := G(t, \cdot) \in L^1_{\mathbb{P}}(\Omega; E)$ for any $t \in [0, T]$, then the maps $G^n : [0, T] \times \Omega \to E$ defined by $G^n(t, \omega) := g^n_t(\omega)$, where $g^n_t$ is a sequence associated to $g_t$ satisfying items **(i)**,**(ii)**,**(iii)**, are $(\mathcal{B}_{[0,T]} \otimes \mathfrak{B}^n)$-measurable for any $n \in \mathbb{N}$.

**Proposition 4.16.** *Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space.*

1. *Then there exists a family of finite algebras $\mathfrak{B}^n \subset \mathfrak{B}$, $n \in \mathbb{N}$, satisfying the finite approximation property of Definition 4.15.*
2. *If $\mathbb{P}$ is without atoms, then there exists a family $\mathfrak{B}^n \subset \mathfrak{B}$, $n \in \mathbb{N}$, satisfying the finite approximation property of Definition 4.15 such that the associated minimal partition $\mathcal{P}(\mathfrak{B}^n) = \{A^n_k : k = 1, \dots, n\}$ contains exactly $n$ elements and*

$$\mathbb{P}(A^n_k) = \frac{1}{n}, \qquad k = 1, \dots, n. \tag{4.20}$$

The proof of Proposition 4.16 is postponed in Appendix D. Results similar to item (1) of Proposition 4.16 can be found in [52, Theorem 6.1.12], where a martingale approach is employed.

**Theorem 4.17** (Approximation by piecewise constant controls for $\mathbf{L}$)**.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 with $U$ a convex compact subset of a separable Banach space $V$. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space and assume that there exists $\{\mathfrak{B}^n\}_{n\in\mathbb{N}}$ satisfying the finite approximation property of Definition 4.15. Let $X_0 \in L^p(\Omega; \mathbb{R}^d)$ and $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$. If $\{X^n_0\}_{n\in\mathbb{N}} \subset L^p(\Omega; \mathbb{R}^d)$ satisfies $\|X^n_0 - X_0\|_{L^p(\Omega;\mathbb{R}^d)} \to 0$, then there exists a sequence $(X^n, u^n) \in \mathcal{A}_{\mathbf{L}}(X^n_0)$ such that*

*(1)* $u^n$ *is* $(\mathcal{B}_{[0,T]} \otimes \mathfrak{B}^n)$*-measurable;*
*(2)* $u^n \to u$ *in* $(\mathcal{L}_T \otimes \mathbb{P})$*-measure, as* $n \to +\infty$*;*
*(3)* $\sup_{t\in[0,T]} \|X_t^n - X_t\|_{L^p(\Omega;\mathbb{R}^d)} \to 0$*, as* $n \to +\infty$*;*
*(4)* $J_{\mathbf{L}}(X^n, u^n) \to J_{\mathbf{L}}(X, u)$*, as* $n \to +\infty$*.*

*Moreover, if* $X_0^n$ *is* $\mathfrak{B}^n$*-measurable,* $n \in \mathbb{N}$*, then* $X_t^n$ *is* $\mathfrak{B}^n$*-measurable for any* $t \in [0,T]$*.*

*Proof.* For any $t \in [0,T]$, we denote with $u_t : \Omega \to U$ the measurable control function $u$ at time $t$. For any $n \in \mathbb{N}$ and $t \in [0,T]$ let $u_t^n : \Omega \to U$ be the $\mathfrak{B}^n$-measurable approximation of $u_t$ given by Definition 4.15. By the convexity of $U$ and its compactness, from the property **(ii)** of Definition 4.15 it follows that $u_t^n(\Omega) \subset U$. Defining $u^n(t,\omega) := u_t^n(\omega)$, thanks to the property **(iv)** of Definition 4.15, we have that $u^n$ is $(\mathcal{B}_{[0,T]} \otimes \mathfrak{B}^n)$-measurable. By Proposition 4.8 we have the existence of $X^n$ with $(X^n, u^n) \in \mathcal{A}_{\mathbf{L}}(X_0^n)$.

From the compactness of $U$ and the dominated convergence theorem it follows that $\|u^n - u\|_{L^1([0,T]\times\Omega;V)} \to 0$, as $n \to +\infty$. Consequently, (2) holds. Properties (3) and (4) follow by Proposition 4.11. □

In the following, we reformulate the approximation result of Theorem 4.17 with the language of particles. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space and $\mathfrak{B}^n$ a finite algebra, $n \in \mathbb{N}$. Denote with $\mathcal{P}(\mathfrak{B}^n) := \{A_k^n : k = 1, \dots, k(n)\}$ the associated unique minimal partition and define $(\Omega^{k(n)}, \mathcal{S}(\Omega^{k(n)}), \mathbb{P}^{k(n)})$ by

$$\begin{aligned} &\Omega^{k(n)} := \{1, \dots, k(n)\}, \quad \mathcal{S}(\Omega^{k(n)}) := \sigma(\{1\}, \dots \{k(n)\}) \\ &\mathbb{P}^{k(n)}(\{k\}) := \mathbb{P}(A_k^n), \quad k = 1, \dots, k(n). \end{aligned} \tag{4.21}$$

In order to approximate trajectories, controls and costs of a Lagrangian problem $\mathbf{L} = \mathbf{L}(\Omega, \mathfrak{B}, \mathbb{P})$ with the respective quantities in $\mathbf{L}^{k(n)} = \mathbf{L}(\Omega^{k(n)}, \mathcal{S}(\Omega^{k(n)}), \mathbb{P}^{k(n)})$ we introduce, for every $n \in \mathbb{N}$, the maps $\psi^n, \phi^n$ and $\mathcal{K}^n$. This is necessary due to the fact that the trajectories are not defined on the same space.

For every $n \in \mathbb{N}$, we denote with $\psi^n, \phi^n$ the maps

$$\begin{aligned} &\psi^n : \Omega \to \Omega^{k(n)}, \quad \psi^n(\omega) = k, \quad \text{if } \omega \in A_k^n, \quad k = 1, \dots, k(n); \\ &\phi^n : \Omega^{k(n)} \to \Omega, \quad \phi^n(k) = \omega_k^n, \quad k = 1, \dots, k(n), \quad \text{for a fixed } \omega_k^n \in A_k^n. \end{aligned} \tag{4.22}$$

Moreover, for every $n \in \mathbb{N}$, we introduce the map $\mathcal{K}^n : L^p(\Omega^{k(n)}; C([0,T];\mathbb{R}^d)) \times \mathrm{M}([0,T] \times \Omega^{k(n)}; U) \to L^p((\Omega, \mathfrak{B}^n, \mathbb{P}); C([0,T];\mathbb{R}^d)) \times \mathrm{M}([0,T] \times \Omega; U)$ given by

$$\begin{aligned} &\mathcal{K}^n(Y, v) := (Y \circ \psi^n, \hat{v}), \\ &\text{where} \qquad \hat{v}(t,\omega) := v(t, \psi^n(\omega)), \quad \forall\, (t,\omega) \in [0,T] \times \Omega. \end{aligned} \tag{4.23}$$

**Proposition 4.18.** *Let* $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ *satisfy Assumption 3.1 with* $U$ *a convex compact subset of a separable Banach space* $V$*. Let* $(\Omega, \mathfrak{B}, \mathbb{P})$ *and be a probability space and assume that there exists* $\{\mathfrak{B}^n\}_{n\in\mathbb{N}}$ *satisfying the finite approximation property of Definition 4.15. For every* $n \in \mathbb{N}$*, let* $(\Omega^{k(n)}, \mathcal{S}(\Omega^{k(n)}), \mathbb{P}^{k(n)})$ *as in* (4.21)*.*

*Let* $X_0 \in L^p(\Omega;\mathbb{R}^d)$ *and* $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$*. If* $Y_0^n \in L^p(\Omega^{k(n)};\mathbb{R}^d)$*,* $n \in \mathbb{N}$*, satisfies*

$$\lim_{n\to+\infty} \|Y_0^n \circ \psi^n - X_0\|_{L^p(\Omega;\mathbb{R}^d)} = 0,$$

*then there exists a sequence* $(Y^n, v^n) \in \mathcal{A}_{\mathbf{L}^{k(n)}}(Y_0^n)$ *such that*

*(1)* $\mathcal{K}^n(Y^n, v^n) \to (X, u)$ *in* $C([0,T]; L^p(\Omega;\mathbb{R}^d)) \times L^1([0,T] \times \Omega; V)$*, as* $n \to +\infty$*;*
*(2)* $J_{\mathbf{L}^{k(n)}}(Y^n, v^n) \to J_{\mathbf{L}}(X, u)$*, as* $n \to +\infty$*.*

*Proof.* For every $n \in \mathbb{N}$, by Proposition 4.14 it holds that $\mathbf{L}_{\mathfrak{B}^n} \sim \mathbf{L}^{k(n)}$ in the sense of Definition 4.4. Thanks to Proposition A.3 we have that $X \in \mathrm{AC}^p([0,T]; L^p(\Omega;\mathbb{R}^d))$, hence we conclude applying Theorem 4.17. Since $U$ is compact, recall that the convergence $u^n \to u \in L^1([0,T] \times \Omega; V)$ is equivalent to the convergence in $(\mathcal{L}_T \otimes \mathbb{P})$-measure. □

*Remark* 4.19. If $(\Omega, \mathfrak{B}, \mathbb{P})$ is a standard Borel space, the existence of a sequence of finite algebras $\mathfrak{B}^n$ is guaranteed by item (1) in Proposition 4.16. Moreover, if $\mathbb{P}$ is without atoms, it is possible to choose $\mathfrak{B}^n$ s.t. $\#\mathfrak{B}^n = n$ and satisfying the property (4.20) given in item (2) in Proposition 4.16.
Notice that the assumption $\mathbb{P}$ without atoms is necessary to get a sequence of $\mathfrak{B}^n$ with the property (4.20). Indeed, if there exists $\omega_0 \in \Omega$ such that $\mathbb{P}(\{\omega_0\}) = \alpha > 0$, then for $n \in \mathbb{N}$ big enough the property (4.20) fails.

4.3. **Approximation by continuous controls and trajectories.** The objective of the subsection is twofold. In Proposition 4.21, under continuity assumptions on both the initial datum and the control, we exhibit a stability result for a Lagrangian problem $\mathbf{L}(\Omega, \mathfrak{B}, \mathbb{P})$ when $\mathbb{P}$ is approximated by a sequence of probability measures $\mathbb{P}^n$. Then, in Theorem 4.22 we approximate admissible controls with *continuous* controls so that the associated trajectories are continuous as well and the associated costs converge. These results are useful to prove the equivalence between Lagrangian and Eulerian optimal control problems (see the proof of Theorem 8.5).

Throughout the section, we assume that

$$(\Omega, \mathfrak{B}, \mathbb{P}) \text{ standard Borel space, } \tau \text{ a Polish topology on } \Omega \text{ such that } \mathfrak{B} = \mathcal{B}_{(\Omega,\tau)}. \tag{4.24}$$

In the following regularity result, we prove the existence of a continuous trajectory for the Lagrangian dynamics whenever both the initial datum and the control are continuous.

**Lemma 4.20** (Continuity)**.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and let $(\Omega, \mathfrak{B}, \mathbb{P})$ satisfy (4.24). Let $\tilde{X}_0 \in C(\Omega; \mathbb{R}^d)$ such that $\tilde{X}_0 \in L^p(\Omega; \mathbb{R}^d)$ and $u \in C([0,T] \times \Omega; U)$. If $(X, u) \in \mathcal{A}_{\mathbf{L}}(\tilde{X}_0)$ and $\mu_t := (X_t)_\sharp \mathbb{P}$ then there exists a unique $\tilde{X} \in C([0,T] \times \Omega; \mathbb{R}^d)$ satisfying for any $\omega \in \Omega$*

$$\begin{cases} \dot{\tilde{X}}_t(\omega) = f(\tilde{X}_t(\omega), u_t(\omega), \mu_t), & \forall\, t \in (0,T) \\ \tilde{X}_{|t=0}(\omega) = \tilde{X}_0(\omega). \end{cases} \tag{4.25}$$

*Moreover, $\tilde{X}_t(\omega) = X_t(\omega)$ for every $t \in [0,T]$ and $\mathbb{P}$-a.e. $\omega \in \Omega$.*

*Proof.* For any $\omega \in \Omega$, there exists a unique solution $\tilde{X}(\omega) \in C^1([0,T]; \mathbb{R}^d)$ of (4.25) thanks to the Lipschitz assumptions on the vector field $f$. Since $u, (\mu_t)_{t\in[0,T]}$ and $\tilde{X}_0$ are fixed, the solutions of (4.1) and (4.25) coincide, hence $\tilde{X}_t(\omega) = X_t(\omega)$ for any $t \in [0,T]$, for $\mathbb{P}$-a.e. $\omega \in \Omega$.

Denoting with $\tilde{X} : [0,T] \times \Omega \to \mathbb{R}^d$ the function $\tilde{X}(t,\omega) = \tilde{X}_t(\omega)$, we prove the continuity of $\tilde{X}$. We fix $(t,\omega) \in [0,T] \times \Omega$, and a sequence $(t_n, \omega_n) \in [0,T] \times \Omega$ converging to $(t,\omega)$ as $n \to +\infty$. By triangular inequality,

$$|\tilde{X}_{t_n}(\omega_n) - \tilde{X}_t(\omega)| \le |\tilde{X}_{t_n}(\omega_n) - \tilde{X}_{t_n}(\omega)| + |\tilde{X}_{t_n}(\omega) - \tilde{X}_t(\omega)|.$$

The second term is estimated by

$$|\tilde{X}_{t_n}(\omega) - \tilde{X}_t(\omega)| \le \int_t^{t_n} \Big| f(\tilde{X}_s(\omega), u_s(\omega), \mu_s) \Big| \, \mathrm{d}s.$$

Concerning the first term,

$$\begin{aligned} |\tilde{X}_{t_n}(\omega_n) - \tilde{X}_{t_n}(\omega)| &\le |\tilde{X}_0(\omega_n) - \tilde{X}_0(\omega)| \\ &\quad + \int_0^{t_n} \Big| f(\tilde{X}_s(\omega_n), u_s(\omega_n), \mu_s) - f(\tilde{X}_s(\omega), u_s(\omega_n), \mu_s) \Big| \, \mathrm{d}s \\ &\quad + \int_0^{t_n} \Big| f(\tilde{X}_s(\omega), u_s(\omega_n), \mu_s) - f(\tilde{X}_s(\omega), u_s(\omega), \mu_s) \Big| \, \mathrm{d}s \\ &\le |\tilde{X}_0(\omega_n) - \tilde{X}_0(\omega)| + L \int_0^{t_n} |\tilde{X}_s(\omega_n) - \tilde{X}_s(\omega)| \, \mathrm{d}s \\ &\quad + \int_0^{T} \Big| f(\tilde{X}_s(\omega), u_s(\omega_n), \mu_s) - f(\tilde{X}_s(\omega), u_s(\omega), \mu_s) \Big| \, \mathrm{d}s. \end{aligned}$$

By Gronwall lemma we have

$$|\tilde{X}_{t_n}(\omega_n) - \tilde{X}_{t_n}(\omega)| \leq e^{LT}\left(|\tilde{X}_0(\omega_n) - \tilde{X}_0(\omega)| + \int_0^T R_s(\omega_n,\omega)\,\mathrm{d}s\right),$$

where $R_s(\omega_n,\omega) := \Big|f(\tilde{X}_s(\omega), u_s(\omega_n), \mu_s) - f(\tilde{X}_s(\omega), u_s(\omega), \mu_s)\Big|$. Collecting the previous inequalities we get

$$\begin{aligned}|\tilde{X}_{t_n}(\omega_n) - \tilde{X}_t(\omega)| \leq e^{LT}&\left(|\tilde{X}_0(\omega_n) - \tilde{X}_0(\omega)| + \int_0^T R_s(\omega_n,\omega)\,\mathrm{d}s\right)\\ &+ \int_t^{t_n}\Big|f(\tilde{X}_s(\omega), u_s(\omega), \mu_s)\Big|\,\mathrm{d}s,\end{aligned}$$

By the continuity of $\tilde{X}_0$, the growth property (3.3) and the continuity of $u$ we can pass to the limit in the right hand side and we conclude.

□

**Proposition 4.21** (Stability for $\mathbb{P}$)**.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega, \mathfrak{B}, \mathbb{P})$ satisfying (4.24). Let $\mathbb{P}^n, \mathbb{P} \in \mathscr{P}(\Omega)$, $n \in \mathbb{N}$, such that $\mathbb{P}^n \to \mathbb{P}$ weakly. Let $\tilde{X}_0 \in C(\Omega;\mathbb{R}^d)$ and $u \in C([0,T]\times\Omega;U)$ such that $\tilde{X}_0 \in \left[\bigcap_{n\in\mathbb{N}} L^p_{\mathbb{P}^n}(\Omega;\mathbb{R}^d)\right] \cap L^p_{\mathbb{P}}(\Omega;\mathbb{R}^d)$ and*

$$\|\tilde{X}_0\|_{L^p_{\mathbb{P}^n}(\Omega;\mathbb{R}^d)} \to \|\tilde{X}_0\|_{L^p_{\mathbb{P}}(\Omega;\mathbb{R}^d)} \quad \text{if } n \to +\infty. \tag{4.26}$$

*We denote by $\mathbf{L}^n := \mathbf{L}(\Omega, \mathfrak{B}, \mathbb{P}^n)$ and $\mathbf{L} := \mathbf{L}(\Omega, \mathfrak{B}, \mathbb{P})$.*

*Let $(X,u) \in \mathcal{A}_{\mathbf{L}}(\tilde{X}_0)$ and $(X^n,u) \in \mathcal{A}_{\mathbf{L}^n}(\tilde{X}_0)$ and denote with $\tilde{X}, \tilde{X}^n \in C([0,T]\times\Omega;\mathbb{R}^d)$ the corresponding solutions given in Lemma 4.20 associated with $\mathbb{P}$ and $\mathbb{P}^n$, respectively. Then*

$$\sup_{(t,\omega)\in[0,T]\times\Omega} |\tilde{X}^n_t(\omega) - \tilde{X}_t(\omega)| \to 0, \qquad \text{as } n \to +\infty, \tag{4.27}$$

$$J_{\mathbf{L}^n}(X^n, u) \longrightarrow J_{\mathbf{L}}(X,u), \qquad \text{as } n \to +\infty. \tag{4.28}$$

*Proof.* We denote $\mu^n_t := (X^n_t)_\sharp \mathbb{P}^n$ and $\mu_t := (X_t)_\sharp \mathbb{P}$.

Since $\tilde{X}_0 : \Omega \to \mathbb{R}^d$ is continuous, the weak convergence $\mathbb{P}^n \to \mathbb{P}$ implies that $\mu^n_0 = (\tilde{X}_0)_\sharp \mathbb{P}^n \to \mu_0 = (\tilde{X}_0)_\sharp \mathbb{P}$ weakly and (4.26) guarantees $\mathrm{m}_p(\mu^n_0) \to \mathrm{m}_p(\mu_0)$. Consequently, by Proposition 2.4, it holds $W_p(\mu^n_0, \mu_0) \to 0$ as $n \to +\infty$ and there exists an admissible $\psi : [0,+\infty) \to [0,+\infty)$, according to Definition 2.3, such that

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu^n_0(x) < +\infty. \tag{4.29}$$

Using the same argument of the proof of Lemma 4.10, thanks to the estimates (4.11) and (4.10) there exist $\tilde{\mu} \in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ and a (not relabelled) subsequence such that

$$\lim_{n\to+\infty} \sup_{s\in[0,T]} W_p(\mu^n_s, \tilde{\mu}_s) = 0. \tag{4.30}$$

We define $\bar{X} \in C([0,T]\times\Omega;\mathbb{R}^d)$ through the system (4.25) using $\tilde{\mu}_t$ instead of $\mu_t$, i.e., for any $\omega \in \Omega$, $t \mapsto \bar{X}(t,\omega)$ is the solution of the problem

$$\begin{cases}\dot{\bar{X}}_t(\omega) = f(\bar{X}_t(\omega), u_t(\omega), \tilde{\mu}_t), & \forall\, t \in (0,T)\\ \bar{X}_{|t=0}(\omega) = \tilde{X}_0(\omega).\end{cases} \tag{4.31}$$

We show that

$$\sup_{(t,\omega)\in[0,T]\times\Omega} |\tilde{X}^n_t(\omega) - \bar{X}_t(\omega)| \to 0, \qquad \text{as } n \to +\infty. \tag{4.32}$$

Indeed, for any $\omega \in \Omega$ and $t \in [0,T]$,

$$|\tilde{X}^n_t(\omega) - \bar{X}_t(\omega)| \le \int_0^t \Big| f(\tilde{X}^n_s(\omega), u_s(\omega), \mu^n_s) - f(\bar{X}_s(\omega), u_s(\omega), \tilde{\mu}_s)\Big| \,\mathrm{d}s$$
$$\le L \int_0^t \Big( |\tilde{X}^n_s(\omega) - \bar{X}_s(\omega)| + W_p(\mu^n_s, \tilde{\mu}_s)\Big) \,\mathrm{d}s,$$

and, by Gronwall inequality, we obtain

$$|\tilde{X}^n_t(\omega) - \bar{X}_t(\omega)| \le LTe^{LT} \sup_{s\in[0,T]} W_p(\mu^n_s, \tilde{\mu}_s),$$

which, by (4.30), proves (4.32).
We have to show that $\bar{X} = \tilde{X}$. We first prove that $\tilde{\mu}_t = (\tilde{X}_t)_\sharp \mathbb{P}$. By the uniform convergence (4.32), the continuity of $\bar{X}_t$ and the weak convergence $\mathbb{P}^n \to \mathbb{P}$, we obtain that (see [3, Lemma 5.2.1])

$$\int_\Omega \phi(\tilde{X}^n_t(\omega)) \,\mathrm{d}\mathbb{P}^n(\omega) \to \int_\Omega \phi(\bar{X}_t(\omega)) \,\mathrm{d}\mathbb{P}(\omega), \qquad \forall\, \phi \in C_b(\mathbb{R}^d;\mathbb{R}).$$

Since $\displaystyle\int_\Omega \phi(\tilde{X}^n_t(\omega)) \,\mathrm{d}\mathbb{P}^n(\omega) = \int_{\mathbb{R}^d} \phi(x) \,\mathrm{d}\mu^n_t(x)$, by the uniqueness of the weak limit we obtain that $\tilde{\mu}_t = (\bar{X}_t)_\sharp \mathbb{P}$. Then, $\bar{X}_t(\omega)$ satisfies

$$\begin{cases} \dot{\bar{X}}_t(\omega) = f(\bar{X}_t(\omega), u_t(\omega), (\bar{X}_t)_\sharp \mathbb{P}), & \forall\, t \in (0,T) \\ \bar{X}_{|t=0}(\omega) = \tilde{X}_0(\omega). \end{cases} \tag{4.33}$$

By the uniqueness result of Proposition 4.8 and the definition of the Lagrangian problem, we obtain that $\bar{X}_t = X_t$ in $L^p_{\mathbb{P}}(\Omega;\mathbb{R}^d)$ for any $t \in [0,T]$. In particular we have that $\tilde{\mu}_t = \mu_t$ for any $t \in [0,T]$. It follows that the systems (4.31) and (4.25) are the same, and then $\bar{X}_t(\omega) = \tilde{X}_t(\omega)$ for any $(t,\omega) \in [0,T] \times \Omega$. Finally, the convergence (4.27) follows by (4.32), because the limit $\tilde{\mu}$ given by the compactness is uniquely determined and it is independent of the subsequence.

For what concerns (4.28), we first observe that

$$J_{\mathbf{L}^n}(X^n, u) = J_{\mathbf{L}^n}(\tilde{X}^n, u), \qquad J_{\mathbf{L}}(X, u) = J_{\mathbf{L}}(\tilde{X}, u).$$

We write the running cost as

$$\int_\Omega \int_0^T \mathcal{C}(\tilde{X}^n_t(\omega), u_t(\omega), \mu^n_t) \,\mathrm{d}t \,\mathrm{d}\mathbb{P}^n(\omega)$$
$$= \int_\Omega \int_0^T \Big( \mathcal{C}(\tilde{X}^n_t(\omega), u_t(\omega), \mu^n_t) - \mathcal{C}(\tilde{X}_t(\omega), u_t(\omega), \mu_t)\Big) \,\mathrm{d}t \,\mathrm{d}\mathbb{P}^n(\omega)$$
$$+ \int_\Omega \int_0^T \mathcal{C}(\tilde{X}_t(\omega), u_t(\omega), \mu_t) \,\mathrm{d}t \,\mathrm{d}\mathbb{P}^n(\omega).$$

By (4.27) and the continuity of $\mathcal{C}$, we have that $\Big( \mathcal{C}(\tilde{X}^n_t(\omega), u_t(\omega), \mu^n_t) - \mathcal{C}(\tilde{X}_t(\omega), u_t(\omega), \mu_t)\Big) \to 0$ uniformly on compact sets of $[0,T] \times \Omega$. By the weak convergence of $\mathbb{P}^n$ towards $\mathbb{P}$ we conclude that

$$\lim_{n\to+\infty} \int_\Omega \int_0^T \Big( \mathcal{C}(\tilde{X}^n_t(\omega), u_t(\omega), \mu^n_t) - \mathcal{C}(\tilde{X}_t(\omega), u_t(\omega), \mu_t)\Big) \,\mathrm{d}t \,\mathrm{d}\mathbb{P}^n(\omega) = 0.$$

We have to prove that

$$\lim_{n\to+\infty} \int_\Omega \int_0^T \mathcal{C}(\tilde{X}_t(\omega), u_t(\omega), \mu_t) \mathrm{d}t \,\mathrm{d}\mathbb{P}^n(\omega) = \int_\Omega \int_0^T \mathcal{C}(\tilde{X}_t(\omega), u_t(\omega), \mu_t) \mathrm{d}t \,\mathrm{d}\mathbb{P}(\omega). \tag{4.34}$$

By (4.29), which can be rewritten as

$$\sup_{n\in\mathbb{N}} \int_\Omega \psi(|\tilde{X}_0(\omega)|^p) \,\mathrm{d}\mathbb{P}^n(\omega) < +\infty,$$

by estimate (4.11) and the doubling property of $\psi$ we get

$$\sup_{t\in[0,T]}\sup_{n\in\mathbb{N}}\int_\Omega \psi(|\tilde{X}_t(\omega)|^p)\,\mathrm{d}\mathbb{P}^n(\omega) < +\infty.$$

By the growth condition (3.2) and the doubling property of $\psi$ we obtain that the map $(t,\omega)\mapsto \mathcal{C}(\tilde{X}_t(\omega),u_t(\omega),\mu_t)$ is uniformly integrable w.r.t. $\{\mathcal{L}_T\otimes\mathbb{P}^n\}_n$. Since this map is also continuous, by [3, Lemma 5.1.7] we obtain (4.34).

Analogously one proves that

$$\lim_{n\to+\infty}\int_\Omega \mathcal{C}_T(\tilde{X}^n_T(\omega),\mu^n_T)\,\mathrm{d}\mathbb{P}^n(\omega) = \int_\Omega \mathcal{C}_T(\tilde{X}_T(\omega),\mu_T)\,\mathrm{d}\mathbb{P}(\omega).$$

□

**Proposition 4.22** (Approximation by continuous controls)**.** *Let $\mathbb{S} := (U,f,\mathcal{C},\mathcal{C}_T)$ satisfy Assumption 3.1 with $U$ a convex compact subset of a separable Banach space $V$, and $(\Omega,\mathfrak{B},\mathbb{P})$ satisfying* (4.24)*. Let $\tilde{X}_0\in C(\Omega;\mathbb{R}^d)$ and $u\in \mathrm{M}([0,T]\times\Omega;U)$ such that $\tilde{X}_0\in L^p(\Omega;\mathbb{R}^d)$. If $(X,u)\in\mathcal{A}_{\mathbf{L}}(\tilde{X}_0)$ then there exits a sequence $(\tilde{X}^n,u^n)\in\mathcal{A}_{\mathbf{L}}(\tilde{X}_0)$ such that*

*(1) $u^n\in C([0,T]\times\Omega;U)$ and $\tilde{X}^n\in C([0,T]\times\Omega;\mathbb{R}^d)$ for any $n\in\mathbb{N}$;*
*(2) $u^n(t,\omega)\to u(t,\omega)$ for $(\mathcal{L}_T\otimes\mathbb{P})$-a.e. $(t,\omega)\in[0,T]\times\Omega$;*
*(3) $\lim_{n\to+\infty}\sup_{t\in[0,T]}|\tilde{X}^n_t(\omega)-X_t(\omega)|=0$ for $\mathbb{P}$-a.e. $\omega\in\Omega$,*
*$\lim_{n\to+\infty}\sup_{t\in[0,T]}\|\tilde{X}^n_t-X_t\|^p_{L^p(\Omega;\mathbb{R}^d)}=0$;*
*(4) $J_{\mathbf{L}}(\tilde{X}^n,u^n)\to J_{\mathbf{L}}(X,u)$, as $n\to+\infty$.*

*Proof.* Since $u\in\mathrm{M}([0,T]\times\Omega;U)$, by Lusin's theorem applied to the space $[0,T]\times\Omega$ with the measure $\mathcal{L}_T\otimes\mathbb{P}$, there exists a sequence of compact subsets $A_n\subset[0,T]\times\Omega$ such that $A_n\subset A_{n+1}$, $\mathcal{L}_T\otimes\mathbb{P}([0,T]\times\Omega\setminus A_n)<\frac{1}{n}$ for every $n\in\mathbb{N}$ and $u|_{A_n}:A_n\to U$ is continuous. Applying Dugundji's extension theorem [30, Theorem 4.1] we can extend $u|_{A_n}$ to a continuous map $u^n:[0,T]\times\Omega\to V$ such that $u^n([0,T]\times\Omega)$ is contained in the closed convex subset $U$ of $V$. Moreover, for $(\mathcal{L}_T\otimes\mathbb{P})$-a.e. $(t,\omega)\in[0,T]\times\Omega$ it holds that $u^n(t,\omega)\to u(t,\omega)$, thanks to the convergence $\mathcal{L}_T\otimes\mathbb{P}([0,T]\times\Omega\setminus A_n)\to 0$ as $n\to+\infty$.

Thanks to Proposition 4.8 and Lemma 4.20, for any $n\in\mathbb{N}$ there exists a unique $\tilde{X}^n\in C([0,T]\times\Omega;\mathbb{R}^d)$ such that $(\tilde{X}^n,u^n)\in\mathcal{A}_{\mathbf{L}}(\tilde{X}_0)$. Defining $\mu^n_t:=(\tilde{X}^n_t)_\sharp\mathbb{P}$, by Lemma 4.10 there exists $\bar{\mu}\in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ such that, up to subsequences,

$$\sup_{t\in[0,T]} W_p(\mu^n_t,\bar{\mu}_t)\to 0,\qquad \text{as } n\to+\infty. \tag{4.35}$$

For every $\omega\in\Omega$ we define $\bar{X}(\omega)$ as the unique solution to the problem

$$\begin{cases}\dot{\bar{X}}_t(\omega)=f(\bar{X}_t(\omega),u_t(\omega),\bar{\mu}_t), & \text{for a.e. } t\in(0,T)\\ \bar{X}_{|t=0}(\omega)=\tilde{X}_0(\omega).\end{cases}$$

Then for any $(t,\omega)\in[0,T]\times\Omega$ it holds

$$\begin{aligned}|\tilde{X}^n_t(\omega)-\bar{X}_t(\omega)| &\le \int_0^t\Big|f(\tilde{X}^n_s(\omega),u^n_s(\omega),\mu^n_s)-f(\bar{X}_s(\omega),u^n_s(\omega),\bar{\mu}_s)\Big|\,\mathrm{d}s\\ &\quad+\int_0^t\big|f(\bar{X}_s(\omega),u^n_s(\omega),\bar{\mu}_s)-f(\bar{X}_s(\omega),u_s(\omega),\bar{\mu}_s)\big|\,\mathrm{d}s\\ &\le L\int_0^t\Big(|\tilde{X}^n_s(\omega)-\bar{X}_s(\omega)|+W_p(\mu^n_s,\bar{\mu}_s)+\mathcal{G}_{s,\omega}(u^n_s(\omega),u_s(\omega))\Big)\,\mathrm{d}s,\end{aligned} \tag{4.36}$$

where $\mathcal{G}_{s,\omega}(u^n_s(\omega),u_s(\omega)):=\big|f(\bar{X}_s(\omega),u^n_s(\omega),\bar{\mu}_s)-f(\bar{X}_s(\omega),u_s(\omega),\bar{\mu}_s)\big|$. Since by (3.3) we have $\mathcal{G}_{s,\omega}(u^n_s(\omega),u_s(\omega))\le C(1+|\bar{X}_s(\omega)|+m_p(\bar{\mu}_s))$, by the convergence in item (2) we get that

$$\lim_{n\to+\infty}\int_0^T\mathcal{G}_{s,\omega}(u^n_s(\omega),u_s(\omega))\mathrm{d}s=0,\quad \text{for } \mathbb{P}\text{-a.e. } \omega\in\Omega. \tag{4.37}$$

By Gronwall inequality, from (4.36) we have

$$|\tilde{X}_t^n(\omega) - \bar{X}_t(\omega)| \leq e^{LT} LT\Big( \sup_{s\in[0,T]} W_p(\mu_s^n, \bar{\mu}_s) + \int_0^T \mathcal{G}_{s,\omega}(u_s^n(\omega), u_s(\omega))\,\mathrm{d}s\Big), \tag{4.38}$$

which, by (4.35) and (4.37), implies

$$\lim_{n\to+\infty} \sup_{t\in[0,T]} |\tilde{X}_t^n(\omega) - \bar{X}_t(\omega)| = 0, \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{4.39}$$

From (4.39) we have that $\mu_t^n$ weakly converges to $(\bar{X}_t)_\sharp\mathbb{P}$ and by (4.35) it follows that $\bar{\mu}_t = (\bar{X}_t)_\sharp\mathbb{P}$ for any $t \in [0,T]$. By the definition of $\bar{X}$ and $\bar{\mu}_t = (\bar{X}_t)_\sharp\mathbb{P}$ and by the uniqueness result of Proposition 4.8, we obtain that $\bar{X}_t(\omega) = X_t(\omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$ and for any $t \in [0,T]$. The first convergence in item (3) follows from (4.39), while the second convergence comes from the first one and Proposition 4.9 through dominated convergence.

Finally, item (4) follows by the same argument as in the proof of Proposition 4.11. □

## 5. Relaxed Lagrangian optimal control problem

In this Section we define a relaxed version of the Lagrangian problem analyzed in Section 4, then we study its properties and its relation with the non-relaxed one.

**Definition 5.1** (Relaxed Lagrangian optimal control problem (**RL**))**.** Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space. Given $X_0 \in L^p(\Omega;\mathbb{R}^d)$, we say that $(X,\sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0)$ if

(i) $\sigma \in \mathrm{M}([0,T]\times\Omega;\mathscr{P}(U))$;

(ii) $X \in L^p(\Omega;\mathrm{AC}^p([0,T];\mathbb{R}^d))$ and for $\mathbb{P}$-a.e. $\omega \in \Omega$, $X(\omega)$ is a solution of the following Cauchy problem

$$\begin{cases} \dot{X}_t(\omega) = \displaystyle\int_U f(X_t(\omega), u, (X_t)_\sharp\mathbb{P})\,\mathrm{d}\sigma_{t,\omega}(u), & \text{for } \mathcal{L}_T\text{-a.e. } t \in ]0,T] \\ X_{|t=0}(\omega) = X_0(\omega), \end{cases} \tag{5.1}$$

where $X_t : \Omega \to \mathbb{R}^d$ is defined by $X_t(\omega) := X(t,\omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$ and $\sigma_{t,\omega} := \sigma(t,\omega) \in \mathscr{P}(U)$.

We refer to $(X,u) \in \mathcal{A}_{\mathbf{RL}}(X_0)$ as to an *admissible pair*, with $X$ a *trajectory* and $\sigma$ a *relaxed control*.

We define the *cost functional* $J_{\mathbf{RL}} : L^p(\Omega;C([0,T];\mathbb{R}^d)) \times \mathrm{M}([0,T]\times\Omega;\mathscr{P}(U)) \to [0,+\infty)$, by

$$J_{\mathbf{RL}}(X,\sigma) := \int_\Omega \int_0^T \int_U \mathcal{C}(X_t(\omega), u, (X_t)_\sharp\mathbb{P})\,\mathrm{d}\sigma_{t,\omega}(u)\,\mathrm{d}t\,\mathrm{d}\mathbb{P}(\omega) \; + \int_\Omega \mathcal{C}_T(X_T(\omega), (X_T)_\sharp\mathbb{P})\,\mathrm{d}\mathbb{P}(\omega),$$

and the *value function* $V_{\mathbf{RL}} : L^p(\Omega;\mathbb{R}^d) \to [0,+\infty)$ by

$$V_{\mathbf{RL}}(X_0) := \inf\left\{ J_{\mathbf{RL}}(X,\sigma) \,:\, (X,\sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0) \right\}. \tag{5.2}$$

*Remark* 5.2. By Proposition 3.6 the Relaxed Lagrangian problem **RL** in $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ is a particular Lagrangian convex problem $\mathbf{L}'$ in the lifted space $\mathbb{S}' = (\mathscr{U}, \mathscr{F}, \mathscr{C}, \mathcal{C}_T)$ defined in Definition 3.5. In particular, the system (5.1) can be rewritten as

$$\begin{cases} \dot{X}_t(\omega) = \mathscr{F}(X_t(\omega), \sigma_t(\omega), (X_t)_\sharp\mathbb{P}), & \text{for a.e. } t \in (0,T] \\ X_{|t=0}(\omega) = X_0(\omega). \end{cases}$$

and the cost functional as

$$J_{\mathbf{RL}}(X,\sigma) := \int_\Omega \int_0^T \mathscr{C}(X_t(\omega), \sigma(t,\omega), (X_t)_\sharp\mathbb{P})\,\mathrm{d}t\,\mathrm{d}\mathbb{P}(\omega) + \int_\Omega \mathcal{C}_T(X_T(\omega), (X_T)_\sharp\mathbb{P})\,\mathrm{d}\mathbb{P}(\omega).$$

As a consequence, the results proved for the Lagrangian problem **L** also apply to the Relaxed Lagrangian problem **RL**. We further point out that even in the relaxed Lagrangian setting,

existence of minimizers is not guaranteed in general (see also Remark 4.2). We refer to Section 8.3 for a detailed discussion and in particular to Remark 8.10.

5.1. **Equivalence of L and RL. Chattering result.** In this subsection we prove that the value functions for the Lagrangian and the Relaxed Lagrangian optimal control problems, set in the same parametrization space $(\Omega, \mathfrak{B}, \mathbb{P})$ and same system $\mathbb{S}$, coincide. Precisely, we aim at showing the following theorem whose proof is postponed at the end of the section.

**Theorem 5.3.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space such that there exists $\{\mathfrak{B}^n\}_{n\in\mathbb{N}}$ satisfying the finite approximation property of Definition 4.15. If $X_0 \in L^p(\Omega; \mathbb{R}^d)$, then $V_{\mathbf{L}}(X_0) = V_{\mathbf{RL}}(X_0)$.*

The proof of Theorem 5.3 easily follows from the combination of Theorem 5.6 and Proposition 5.4 given below. Theorem 5.6 is a suitable extension of the classical (in optimal control theory) *chattering theorem* which permits to approximate relaxed controls with piecewise-constant controls.

Notice that Theorem 5.3 holds in particular if $(\Omega, \mathfrak{B}, \mathbb{P})$ is a standard Borel space thanks to Proposition 4.16.

Let us start with the following proposition.

**Proposition 5.4.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space. Let $X_0 \in L^p(\Omega; \mathbb{R}^d)$. If $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$, then, defining $\sigma : [0, T] \times \Omega \to \mathscr{P}(U)$ by $\sigma(t, \omega) = \delta_{u(t,\omega)}$ we have $(X, \sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0)$ and $J_{\mathbf{L}}(X, u) = J_{\mathbf{RL}}(X, \sigma)$. In particular $V_{\mathbf{RL}}(X_0) \leq V_{\mathbf{L}}(X_0)$.*

*Proof.* The result follows immediately by Proposition 3.6. □

Recall that if $\bar{\mathfrak{B}}$ is a finite algebra on $\Omega$ and $E$ is a Banach space, a function $g : \Omega \to E$ is $\bar{\mathfrak{B}}$-measurable if and only if $g$ is constant on the elements of a partition of $\Omega$ contained in $\bar{\mathfrak{B}}$.

In the following proposition, given a piecewise constant *relaxed* control we approximate it with a sequence of piecewise constant *(non-relaxed)* controls so that the associated trajectories and costs converge.

**Proposition 5.5.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space. Let $X_0 \in L^p(\Omega; \mathbb{R}^d)$, $\bar{\mathfrak{B}} \subset \mathfrak{B}$ a finite algebra, $(X, \sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0)$ such that $\sigma$ is $(\mathcal{B}_{[0,T]} \otimes \bar{\mathfrak{B}})$-measurable. Then there exists a sequence $\{(X^m, u^m)\}_{m\in\mathbb{N}} \subset \mathcal{A}_{\mathbf{L}}(X_0)$ such that*

*(1) $u^m$ are $(\mathcal{B}_{[0,T]} \otimes \bar{\mathfrak{B}})$-measurable;*

*(2) for any $\omega \in \Omega$, $(i_{[0,T]}, u^m(\cdot, \omega))_\sharp \mathcal{L}_T \xrightarrow{\mathcal{Y}} \sigma_\omega$, where $\sigma_\omega := \sigma(t, \omega) \otimes \mathcal{L}_T \in \mathscr{P}([0, T] \times U)$;*

*(3) $\sup_{t\in[0,T]} \|X_t^m - X_t\|_{L^p(\Omega;\mathbb{R}^d)} \to 0$ as $m \to +\infty$;*

*(4) $J_{\mathbf{L}}(X^m, u^m) \to J_{\mathbf{RL}}(X, \sigma)$, as $m \to +\infty$.*

*Moreover, if $X_0$ is $\bar{\mathfrak{B}}$-measurable then $X_t$, $X_t^m$ are $\bar{\mathfrak{B}}$-measurable for any $m \in \mathbb{N}$ and $t \in [0, T]$.*

*Proof.* We fix the minimal (w.r.t. inclusion) partition associated to the finite algebra $\bar{\mathfrak{B}}$ that we denote by $\mathcal{P} := \{A_k : k = 1, \dots n\} \subset \bar{\mathfrak{B}}$. For any $k = 1, \dots n$ we select $\omega_k \in A_k$ and apply Lemma 2.10 (with $\mathbb{T} = [0, T]$, $S = U$ and $\lambda = \mathcal{L}_T$) to the measure $\nu = \sigma_{\omega_k} := \sigma(t, \omega_k) \otimes \mathcal{L}_T \in \mathscr{P}([0, T] \times U)$. This yields a sequence of $\mathcal{B}_{[0,T]}$-measurable functions $u_k^m : [0, T] \to U$ such that

$$(i_{[0,T]}, u_k^m(\cdot))_\sharp \mathcal{L}_T \xrightarrow{\mathcal{Y}} \sigma_{\omega_k}.$$

Thus, we define $u^m : [0, T] \times \Omega \to U$ setting $u^m(t, \omega) := u_k^m(t)$ if $\omega \in A_k$. By construction, the function $\omega \mapsto u^m(t, \omega)$ is constant on $A_k$, for any $k = 1, \dots, n$. Furthermore, for any $\omega \in \Omega$ the maps $t \mapsto u^m(t, \omega)$ are $\mathcal{B}_{[0,T]}$-measurable. The sequence of controls $u^m \in \mathrm{M}([0, T] \times \Omega)$ readily satisfies items (1) and (2).

Given $u^m$ constructed above, by Proposition 4.8 there exists a unique $X^m \in L^p(\Omega; \mathrm{AC}^p([0, T]; \mathbb{R}^d))$ such that $(X^m, u^m) \in \mathcal{A}_{\mathbf{L}}(X_0)$. Thanks to Remark, 4.3 we interpret $X^m \in \mathrm{AC}^p([0, T]; L^p(\Omega; \mathbb{R}^d))$

and define $\mu^m \in \mathrm{AC}^p([0,T];\mathscr{P}_p(\mathbb{R}^d))$ by $\mu^m_t := (X^m_t)_\sharp\mathbb{P}$. By Lemma 4.10 there exists a (non relabeled) subsequence $\mu^m$ and $\tilde{\mu} \in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ such that

$$\lim_{m\to\infty} \sup_{t\in[0,T]} W_p(\mu^m_t, \tilde{\mu}_t) = 0. \tag{5.3}$$

We define $g : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$ by

$$g(t,y,u) := f(y,u,\tilde{\mu}_t).$$

Selecting a representative $X_0$ defined for every $\omega \in \Omega$, let $Y^m(\omega) \in \mathrm{AC}^p([0,T];\mathbb{R}^d)$ be the unique solution of the Cauchy problem

$$\begin{cases} \dot{Y}^m_t(\omega) = g(t, Y^m_t(\omega), u^m(t,\omega)), & \text{for a.e. } t \in (0,T) \\ Y^m_{|t=0}(\omega) = X_0(\omega). \end{cases} \tag{5.4}$$

For any $\omega \in \Omega$, let also $Y(\omega) \in \mathrm{AC}^p([0,T];\mathbb{R}^d)$ be the unique solution of the Cauchy problem

$$\begin{cases} \dot{Y}_t(\omega) = \displaystyle\int_U g(t, Y_t(\omega), u)\, \mathrm{d}\sigma_{t,\omega}(u), & \text{for a.e. } t \in (0,T) \\ Y_{|t=0}(\omega) = X_0(\omega). \end{cases} \tag{5.5}$$

Hence, by item (2) and assumptions (3.1) and (3.3) we can apply Lemma B.1 to obtain

$$\lim_{m\to+\infty} \sup_{t\in[0,T]} |Y^m_t(\omega) - Y_t(\omega)| = 0, \qquad \forall\, \omega \in \Omega. \tag{5.6}$$

Since $(X^m, u^m) \in \mathcal{A}_{\mathbf{L}}(X_0)$, by definition of the Lagrangian problem, for $\mathbb{P}$-a.e. $\omega \in \Omega$ $X^m(\omega) \in \mathrm{AC}^p([0,T];\mathbb{R}^d)$ and

$$\begin{cases} \dot{X}^m_t(\omega) = f(X^m_t(\omega), u^m_t(\omega), \mu^m_t), & \text{for a.e. } t \in (0,T) \\ X^m_{|t=0}(\omega) = X_0(\omega). \end{cases} \tag{5.7}$$

Then

$$\begin{aligned} |Y^m_t(\omega) - X^m_t(\omega)| &\le \int_0^t |f(Y^m_s(\omega), u^m(s,\omega), \tilde{\mu}_s) - f(X^m_s(\omega), u^m(s,\omega), \mu^m_s)|\, \mathrm{d}s \\ &\le L \int_0^t \left(|Y^m_s(\omega) - X^m_s(\omega)| + W_p(\tilde{\mu}_s, \mu^m_s)\right) \mathrm{d}s \\ &\le L \int_0^t |Y^m_s(\omega) - X^m_s(\omega)|\, \mathrm{d}s + LT \sup_{s\in[0,T]} W_p(\tilde{\mu}_s, \mu^m_s). \end{aligned}$$

By Gronwall inequality we get

$$\sup_{t\in[0,T]} |Y^m_t(\omega) - X^m_t(\omega)| \le LTe^{LT} \sup_{s\in[0,T]} W_p(\tilde{\mu}_s, \mu^m_s). \tag{5.8}$$

From (5.8), (5.3) and (5.6) it follows that

$$\lim_{m\to+\infty} \sup_{t\in[0,T]} |X^m_t(\omega) - Y_t(\omega)| = 0, \qquad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{5.9}$$

By (5.9) it follows that $\mu^m_t = (X^m_t)_\sharp\mathbb{P} \to (Y_t)_\sharp\mathbb{P}$ weakly for any $t \in [0,T]$, and then, by (5.3), it holds that $\tilde{\mu}_t = (Y_t)_\sharp\mathbb{P}$ for any $t \in [0,T]$. Thus, thanks to (5.5) and the definition of $g$ we conclude that $(Y,\sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0)$. Since $(X,\sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0)$, by the uniqueness result of Propositions 4.8 we have that $Y = X$ and

$$\lim_{m\to+\infty} \sup_{t\in[0,T]} |X^m_t(\omega) - X_t(\omega)| = 0, \qquad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{5.10}$$

Finally, to prove item (3) it is enough to observe that

$$\sup_{t\in[0,T]} \|X^m_t - X_t\|^p_{L^p(\Omega;\mathbb{R}^d)} \le \int_\Omega \sup_{t\in[0,T]} |X^m_t(\omega) - X_t(\omega)|^p\, \mathrm{d}\mathbb{P}(\omega). \tag{5.11}$$

By (5.9), and (4.11) we can pass to the limit in (5.11) by dominated convergence.

To prove item (4) we write

$$\begin{aligned}
&\left|\int_\Omega\int_0^T \mathcal{C}(X_t^m(\omega),u_t^m(\omega),\mu_t^m)\,\mathrm{d}t\,\mathrm{d}\mathbb{P}(\omega)-\int_\Omega\int_0^T\int_U \mathcal{C}(X_t(\omega),u,\mu_t)\,\mathrm{d}\sigma_{t,\omega}(u)\,\mathrm{d}t\,\mathrm{d}\mathbb{P}(\omega)\right|\\
&\le\left|\int_\Omega\left(\int_0^T \mathcal{C}(X_t^m(\omega),u_t^m(\omega),\mu_t^m)-\mathcal{C}(X_t(\omega),u_t^m(\omega),\mu_t)\,\mathrm{d}t\right)\mathrm{d}\mathbb{P}(\omega)\right|\\
&+\left|\int_\Omega\left(\int_0^T \mathcal{C}(X_t(\omega),u_t^m(\omega),\mu_t)\,\mathrm{d}t-\int_{[0,T]\times U}\mathcal{C}(X_t(\omega),u,\mu_t)\,\mathrm{d}\sigma_\omega(t,u)\right)\mathrm{d}\mathbb{P}(\omega)\right|.
\end{aligned}\tag{5.12}$$

Since (5.10) holds, for $\mathbb{P}$-a.e. $\omega\in\Omega$, there exists a compact $K_\omega\subset\mathbb{R}^d$ such that $X_t^m(\omega),X_t(\omega)\in K_\omega$ for any $m\in\mathbb{N}$ and $t\in[0,T]$. Analogously, by (5.3) there exists a compact $\mathcal{K}\subset\mathscr{P}_p(\mathbb{R}^d)$ such that $\mu_t^m,\mu_t\in\mathcal{K}$ for any $m\in\mathbb{N}$ and $t\in[0,T]$. By Proposition 2.4 there exists an admissible $\psi$ such that

$$\sup_{t\in[0,T]}\sup_{m\in\mathbb{N}}\int_\Omega\psi(|X_t^m(\omega)|^p)\,\mathrm{d}\mathbb{P}(\omega)=\sup_{t\in[0,T]}\sup_{m\in\mathbb{N}}\int_{\mathbb{R}^d}\psi(|x|^p)\,\mathrm{d}\mu_t^m(x)<+\infty.\tag{5.13}$$

By the continuity of $\mathcal{C}$ there exists a modulus of continuity $\alpha_\omega:[0,+\infty)\to[0,+\infty)$ for the restriction of $\mathcal{C}$ to the compact set $K_\omega\times U\times\mathcal{K}$. Then, for $\mathbb{P}$-a.e. $\omega\in\Omega$,

$$\sup_{t\in[0,T]}|\mathcal{C}(X_t^m(\omega),u_t^m(\omega),\mu_t^m)-\mathcal{C}(X_t(\omega),u_t^m(\omega),\mu_t)|\le\alpha_\omega\big(\sup_{t\in[0,T]}(|X_t^m(\omega)-X_t(\omega)|+W_p(\mu_t^m,\mu_t))\big).$$

Taking into account the previous consideration together with (5.10), (5.3), the growth condition (3.2) and (5.13), we obtain

$$\left|\int_\Omega\left(\int_0^T \mathcal{C}(X_t^m(\omega),u_t^m(\omega),\mu_t^m)-\mathcal{C}(X_t(\omega),u_t^m(\omega),\mu_t)\,\mathrm{d}t\right)\mathrm{d}\mathbb{P}(\omega)\right|\to 0.$$

For the second term in the right hand side of (5.12), for $\mathbb{P}$-a.e. $\omega\in\Omega$ we define $h_\omega:[0,T]\times U\to\mathbb{R}$ by $h_\omega(t,u):=\mathcal{C}(X_t(\omega),u,(X_t)_\sharp\mathbb{P})$. Notice that $h_\omega$ is continuous and bounded in $[0,T]\times U$, hence from the Young convergence of item (2) we get

$$\left|\int_0^T h_\omega(t,u_t^m(\omega))\,\mathrm{d}t-\int_{[0,T]\times U}h_\omega(t,u)\,\mathrm{d}\sigma_\omega(t,u)\right|\to 0,\qquad\text{for }\mathbb{P}\text{-a.e. }\omega\in\Omega.\tag{5.14}$$

From the growth assumptions (3.2) and dominated convergence theorem we obtain that

$$\left|\int_\Omega\left(\int_0^T \mathcal{C}(X_t(\omega),u_t^m(\omega),\mu_t)\,\mathrm{d}t-\int_{[0,T]\times U}\mathcal{C}(X_t(\omega),u,\mu_t)\,\mathrm{d}\sigma_\omega(t,u)\right)\mathrm{d}\mathbb{P}(\omega)\right|\to 0.$$

Finally, thanks to (5.10) and (5.3) we also obtain that

$$\lim_{m\to+\infty}\int_\Omega\mathcal{C}_T(X_T^m(\omega),\mu_T^m)\,\mathrm{d}\mathbb{P}(\omega)=\int_\Omega\mathcal{C}_T(X_T(\omega),\mu_T)\,\mathrm{d}\mathbb{P}(\omega).$$

For what concerns the last statement, since $X_0$ is $\bar{\mathfrak{B}}$-measurable, (hence constant on the elements of the partition $\mathcal{P}$), the measurability of $X_t^m$ with respect to the algebra $\bar{\mathfrak{B}}$ follows by uniqueness of solutions to (5.7). The same argument also yields that $X_t$ is $\bar{\mathfrak{B}}$-measurable.

□

Combining Theorem 4.17 and Proposition 4.11 applied to the Relaxed Lagrangian problem **RL**, with Proposition 5.5, we can prove the following Theorem.

**Theorem 5.6** (Chattering)**.** *Let $\mathbb{S}:=(U,f,\mathcal{C},\mathcal{C}_T)$ satisfy Assumption 3.1. Let $(\Omega,\mathfrak{B},\mathbb{P})$ be a probability space and $\{\mathfrak{B}^n\}_{n\in\mathbb{N}}$ satisfying the finite approximation property of Definition 4.15. Let $X_0\in L^p(\Omega;\mathbb{R}^d)$, $(X,\sigma)\in\mathcal{A}_{\mathbf{RL}}(X_0)$ and $\{X_0^n\}_{n\in\mathbb{N}}\subset L^p(\Omega;\mathbb{R}^d)$, such that*

$$\|X_0^n-X_0\|_{L^p(\Omega;\mathbb{R}^d)}\to 0\quad\text{ as }n\to+\infty.\tag{5.15}$$

*Then there exists a sequence* $\{(\tilde{X}^n, \tilde{u}^n)\}_{n\in\mathbb{N}}$ *such that* $(\tilde{X}^n, \tilde{u}^n) \in \mathcal{A}_{\mathbf{L}}(X_0^n)$ *for every* $n \in \mathbb{N}$ *and the following hold*

*(1)* $\tilde{u}^n$ *are* $(\mathcal{B}_{[0,T]} \otimes \mathfrak{B}^n)$*-measurable;*

*(2) for* $\mathbb{P}$*-a.e.* $\omega \in \Omega$, $(i_{[0,T]}, \tilde{u}^n(\cdot,\omega))_\sharp \mathcal{L}_T \xrightarrow{\mathcal{Y}} \sigma_\omega$, *as* $n \to +\infty$, *where* $\sigma_\omega := \sigma(t,\omega) \otimes \mathcal{L}_T \in \mathscr{P}([0,T] \times U)$*;*

*(3)* $\sup_{t\in[0,T]} \|\tilde{X}_t^n - X_t\|_{L^p(\Omega;\mathbb{R}^d)} \to 0$ *as* $n \to +\infty$*;*

*(4)* $J_{\mathbf{L}}(\tilde{X}^n, \tilde{u}^n) \to J_{\mathbf{RL}}(X, \sigma)$, *as* $n \to +\infty$.

*Moreover, if* $X_0^n$ *is* $\mathfrak{B}^n$*-measurable,* $n \in \mathbb{N}$*, then* $\tilde{X}_t^n$ *is* $\mathfrak{B}^n$*-measurable for any* $t \in [0,T]$.

*Proof.* Let $(X, \sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0)$ and $\{X_0^n\}_{n\in\mathbb{N}}$ satisfying (5.15). Applying Theorem 4.17 to the relaxed problem **RL** (which is a Lagrangian problem in a lifted space as discussed in Remark 5.2), there exists a sequence $\{(X^n, \sigma^n)\}_{n\in\mathbb{N}}$ such that $(X^n, \sigma^n) \in \mathcal{A}_{\mathbf{RL}}(X_0^n)$ for every $n \in \mathbb{N}$. Moreover $\sigma^n$ are $(\mathcal{B}_{[0,T]} \otimes \mathfrak{B}^n)$-measurable, $\sigma^n \to \sigma$ in $(\mathcal{L}_T \otimes \mathbb{P})$-measure and, as a consequence, we have that $\sigma^n(t,\omega) \to \sigma(t,\omega)$ weakly in $\mathscr{P}(U)$ for $\mathcal{L}_T \otimes \mathbb{P}$-a.e. $(t,\omega)$, up to a non-relabelled subsequence. Thus, by Remark 2.9, we get (up to a non-relabelled subsequence)

$$\sigma_\omega^n := \sigma^n(t,\omega) \otimes \mathcal{L}_T \xrightarrow{\mathcal{Y}} \sigma_\omega, \qquad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{5.16}$$

By Proposition 5.5, for any fixed $n \in \mathbb{N}$, there exists a sequence $\{(Y^{n,m}, u^{n,m})\}_{m\in\mathbb{N}} \subset \mathcal{A}_{\mathbf{L}}(X_0^n)$, with $u^{n,m}$ $(\mathcal{B}_{[0,T]} \otimes \mathfrak{B}^n)$-measurable, such that

(i) for any $\omega \in \Omega$, $(i_{[0,T]}, u^{n,m}(\cdot,\omega))_\sharp \mathcal{L}_T \xrightarrow{\mathcal{Y}} \sigma_\omega^n$, as $m \to +\infty$;

(ii) $\sup_{t\in[0,T]} \|X_t^n - Y_t^{n,m}\|_{L^p(\Omega;\mathbb{R}^d)} \to 0$, as $m \to +\infty$;

(iii) $J_{\mathbf{L}}(Y^{n,m}, u^{n,m}) \to J_{\mathbf{RL}}(X^n, \sigma^n)$, as $m \to +\infty$.

Let us denote by $\mathcal{P}(\mathfrak{B}^n) := \{A_k^n, k = 1, \dots, k(n)\}$ the minimal (finite) partition induced by $\mathfrak{B}^n$. Let also $\sigma_\omega^{n,m} := (i_{[0,T]}, u^{n,m}(\cdot,\omega))_\sharp \mathcal{L}_T$ and observe that the map $\omega \mapsto \sigma_\omega^{n,m}$ is constant on the elements of $\mathcal{P}^n$. Then, if we select a representative $\omega_k \in A_k^n$ for any $k = 1, \dots, k(n)$, from item (i) it follows that

$$\lim_{m\to+\infty} \sup_{k=1,\dots,k(n)} \delta(\sigma_{\omega_k}^{n,m}, \sigma_{\omega_k}^n) = \lim_{m\to+\infty} \sup_{\omega\in\Omega} \delta(\sigma_\omega^{n,m}, \sigma_\omega^n) = 0, \tag{5.17}$$

where $\delta$ metrizes the Young convergence in $[0,T] \times U$. Recall that Young convergence is indeed equivalent to the weak convergence in $[0,T] \times U$, see Remark 2.9.

For any $n \in \mathbb{N}$, let $m(n)$ be such that

$$\sup_{t\in[0,T]} \|X_t^n - Y_t^{n,m(n)}\|_{L^p(\Omega;\mathbb{R}^d)} < \frac{1}{n},$$

$$|J_{\mathbf{L}}(Y^{n,m(n)}, u^{n,m(n)}) - J_{\mathbf{RL}}(X^n, \sigma^n)| < \frac{1}{n}$$

and

$$\sup_{\omega\in\Omega} \delta(\sigma_\omega^{n,m(n)}, \sigma_\omega^n) < \frac{1}{n}.$$

Let us define $\tilde{X}^n := Y^{n,m(n)}$, the control function $\tilde{u}^n := u^{n,m(n)} : [0,T] \times \Omega \to U$ and $\tilde{\sigma}_\omega^n := \sigma_\omega^{n,m(n)} = (i_{[0,T]}, \tilde{u}^n(\cdot,\omega))_\sharp \mathcal{L}_T$. Notice that, by construction, $(\tilde{X}^n, \tilde{u}^n) \in \mathcal{A}_{\mathbf{L}}(X_0^n)$.

Fix now $\varepsilon > 0$. By Theorem 4.17 there exists $n_\varepsilon$ such that

$$\sup_{t\in[0,T]} \|X_t^n - X_t\|_{L^p(\Omega;\mathbb{R}^d)} < \varepsilon, \quad \forall n > n_\varepsilon,$$

and

$$|J_{\mathbf{RL}}(X^n, \sigma^n) - J_{\mathbf{RL}}(X, \sigma)| < \varepsilon, \quad \forall n > n_\varepsilon.$$

Then, using the definition of $\tilde{X}^n$ and $\tilde{u}^n$, for any $n > n_\varepsilon$ it holds

$$\sup_{t\in[0,T]} \|\tilde{X}^n_t - X_t\|_{L^p(\Omega;\mathbb{R}^d)} \leq \sup_{t\in[0,T]} \|Y^{n,m(n)}_t - X^n_t\|_{L^p(\Omega;\mathbb{R}^d)} + \sup_{t\in[0,T]} \|X^n_t - X_t\|_{L^p(\Omega;\mathbb{R}^d)} < \frac{1}{n} + \varepsilon,$$

and

$$|J_{\mathbf{L}}(\tilde{X}^n, \tilde{u}^n) - J_{\mathbf{RL}}(X,\sigma)| \leq |J_{\mathbf{L}}(Y^{n,m(n)}, u^{n,m(n)}) - J_{\mathbf{RL}}(X^n,\sigma^n)| + |J_{\mathbf{RL}}(X^n,\sigma^n) - J_{\mathbf{RL}}(X,\sigma)| < \frac{1}{n} + \varepsilon.$$

If we send $n \to +\infty$, items (3) and (4) follow by the arbitrariness of $\varepsilon > 0$.

It remains to show item (2). Fix again $\varepsilon > 0$ and choose $\omega \in \Omega$ for which the convergence in (5.16) holds. Then, there exists $n_\varepsilon(\omega) > 0$ such that

$$\delta(\sigma^n_\omega, \sigma_\omega) < \varepsilon, \quad \forall n > n_\varepsilon(\omega),$$

and

$$\delta(\tilde{\sigma}^n_\omega, \sigma_\omega) \leq \delta(\tilde{\sigma}^n_\omega, \sigma^n_\omega) + \delta(\sigma^n_\omega, \sigma_\omega) < \frac{1}{n} + \varepsilon, \quad \forall\, n > n_\varepsilon(\omega).$$

Sending $n \to +\infty$ we get item (2). □

## 6. Eulerian optimal control problem

In this Section we describe the Eulerian formulation of the optimal control problem and we study its properties under the Convexity Assumption 3.4. In particular, as stated in Theorem 6.8, in this setting we get the existence of minimizers. Recall that $\mathrm{B}([0,T]\times\mathbb{R}^d;U)$ denotes the set of Borel measurable functions.

**Definition 6.1** (Eulerian optimal control problem ($\mathbf{E}$))**.** Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Given $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, we say that $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ if

(i) $\underline{u} \in \mathrm{B}([0,T]\times\mathbb{R}^d;U)$;

(ii) $\mu \in \mathrm{AC}^p([0,T];\mathscr{P}_p(\mathbb{R}^d))$ is a distributional solution of the Cauchy problem

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(v_t\mu_t) = 0, & \text{in } [0,T]\times\mathbb{R}^d \\ \mu_{t=0} = \mu_0, \end{cases} \tag{6.1}$$

where $v \in \mathrm{B}([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$ is defined by $v_t(x) := f(x, \underline{u}(t,x), \mu_t)$ and $\mu_t := \mu(t)$.

We refer to $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ as to an *admissible pair*, with $\mu$ a *measure trajectory* and $\underline{u}$ a *Eulerian control*.

We define the *cost functional* $J_{\mathbf{E}} : \mathrm{C}([0,T];\mathscr{P}_p(\mathbb{R}^d)) \times \mathrm{B}([0,T]\times\mathbb{R}^d;U) \to [0,+\infty)$ by

$$J_{\mathbf{E}}(\mu,\underline{u}) := \int_0^T \int_{\mathbb{R}^d} \mathcal{C}(x, \underline{u}(t,x), \mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_{\mathbb{R}^d} \mathcal{C}_T(x,\mu_T)\,\mathrm{d}\mu_T(x),$$

and the *value function* $V_{\mathbf{E}} : \mathscr{P}_p(\mathbb{R}^d) \to [0,+\infty)$ by

$$V_{\mathbf{E}}(\mu_0) := \inf\{J_{\mathbf{E}}(\mu,\underline{u}) \,:\, (\mu,\underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)\}.$$

*Remark* 6.2. Notice that, given $\underline{u} \in \mathrm{B}([0,T]\times\mathbb{R}^d;U)$, $\mu \in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ and setting $v_t(x) := f(x,\underline{u}(t,x),\mu_t)$ as in Definition 6.1, we have

$$\int_{\mathbb{R}^d} |v_t(x)|^p\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} |f(x,\underline{u}(t,x),\mu_t)|^p\,\mathrm{d}\mu_t(x) \leq \tilde{C}\left(1 + \int_{\mathbb{R}^d} |x|^p\,\mathrm{d}\mu_t(x)\right) \leq \bar{C}, \quad \forall t \in [0,T]$$

for some constants $\tilde{C}, \bar{C} > 0$, thanks to the growth condition (3.3) and since $\mu \in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$. In particular, we get $v \in L^p(0,T;L^p_{\mu_t}(\mathbb{R}^d;\mathbb{R}^d))$. Thus, if $\mu$ is also a distributional solution of

(6.1), then $\mu \in \mathrm{AC}^p([0,T];\mathscr{P}_p(\mathbb{R}^d))$. Hence, in Definition 6.1(ii) we could have just required $\mu \in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$.

Observe also that the functional $J_{\mathbf{E}}$ is finite thanks to the growth condition (3.2).

**Proposition 6.3.** *Let $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$. Then $\mathcal{A}_{\mathbf{E}}(\mu_0) \neq \emptyset$.*

*Proof.* Let us fix $u_0 \in U$ and define $\underline{u}(t,x) = u_0$ for any $(t,x) \in [0,T] \times \mathbb{R}^d$. Applying Proposition 4.8 with $\Omega = \mathbb{R}^d$, $\mathbb{P} = \mu_0$, $X_0(x) = x$ for any $x \in \mathbb{R}^d$ and $u = \underline{u}$, there exists $X \in L^p_{\mu_0}(\mathbb{R}^d;\mathrm{AC}^p([0,T];\mathbb{R}^d))$ such that $(X,\underline{u}) \in \mathcal{A}_{\mathbf{L}}(X_0)$. Defining $\mu_t := (X_t)_\sharp \mu_0$, from the definition of the Lagrangian problem it holds

$$X_t(x) = x + \int_0^t f(X_s(x), u_0, \mu_s)\,\mathrm{d}s, \qquad \forall\, t \in [0,T] \text{ and } \mu_0\text{-a.e. } x \in \mathbb{R}^d, \tag{6.2}$$

Furthermore, in view of Proposition A.3 we have that $X \in \mathrm{AC}^p([0,T];L^p_{\mu_0}(\mathbb{R}^d;\mathbb{R}^d))$ and from (2.5) we get $W_p(\mu_t,\mu_s) \leq \|X_t - X_s\|_{L^p_{\mu_0}(\mathbb{R}^d;\mathbb{R}^d)}$ for any $t,s \in [0,T]$. This readily implies that $\mu \in \mathrm{AC}^p([0,T];\mathscr{P}_p(\mathbb{R}^d))$. If we define $v_t(x) := f(x,u_0,\mu_t)$, it remains to show that $\mu$ is a distributional solution of (6.1). This is a standard argument, in view of the fact that (6.2) represents the system of characteristics of (6.1) (see e.g. [3, Lemma 8.1.6]). □

*Remark* 6.4. Given $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, general results granting the existence of solutions to the Cauchy problem (6.1) are provided for instance in [45, Theorem A.2] when $\underline{u} \in \mathrm{B}([0,T] \times \mathbb{R}^d;U)$ is also a Carathéodory function.

**Definition 6.5.** Let $U$ be a subset of a separable Banach space $V$, and denote with $V'$ the dual of $V$. Let $(\mu^n,\underline{u}^n),(\mu,\underline{u}) \in \mathrm{C}([0,T];\mathscr{P}_p(\mathbb{R}^d)) \times \mathrm{B}([0,T] \times \mathbb{R}^d;U)$. We say that $(\mu^n,\underline{u}^n)$ converges to $(\mu,\underline{u})$ if

- $\mu^n$ converges to $\mu$ in $\mathrm{C}([0,T];\mathscr{P}_p(\mathbb{R}^d))$,
- for any $\phi \in C_c([0,T] \times \mathbb{R}^d;V')$ we have

$$\lim_{n\to+\infty} \int_0^T \int_{\mathbb{R}^d} \langle \phi(t,x), \underline{u}^n(t,x)\rangle\,\mathrm{d}\mu^n_t(x)\,\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} \langle \phi(t,x), \underline{u}(t,x)\rangle\,\mathrm{d}\mu_t(x)\,\mathrm{d}t. \tag{6.3}$$

**Proposition 6.6** (Compactness)**.** *Let $\mathbb{S} = (U,f,\mathcal{C},\mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. Let $\mu_0,\mu_0^n \in \mathscr{P}_p(\mathbb{R}^d)$ such that $W_p(\mu_0^n,\mu_0) \to 0$, as $n \to +\infty$. If $(\mu^n,\underline{u}^n) \in \mathcal{A}_{\mathbf{E}}(\mu_0^n)$, $n \in \mathbb{N}$, then there exist $(\mu,\underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ and a subsequence $(\mu^{n_k},\underline{u}^{n_k})$ such that $(\mu^{n_k},\underline{u}^{n_k})$ converges to $(\mu,\underline{u})$, as $k \to +\infty$, according to Definition 6.5.*

*Proof.* Let $(\mu^n,\underline{u}^n) \in \mathcal{A}_{\mathbf{E}}(\mu_0^n)$. Since $W_p(\mu_0^n,\mu_0) \to 0$, by Proposition 2.4 it holds that

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu_0^n(x) < +\infty \tag{6.4}$$

and there exists an admissible $\psi : [0,+\infty) \to [0,+\infty)$, in the sense of Definition 2.3, such that

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} \psi(|x|^p) \mathrm{d}\mu_0^n(x) < +\infty. \tag{6.5}$$

In order to apply Ascoli-Arzelà Theorem to the sequence $\{\mu^n\}_{n\in\mathbb{N}}$, we show that

$$\sup_{n\in\mathbb{N}} \sup_{t\in[0,T]} \int_{\mathbb{R}^d} \psi(|x|^p) \mathrm{d}\mu_t^n(x) < +\infty \tag{6.6}$$

and there exists a constant $C$ such that

$$W_p^p(\mu_s^n,\mu_t^n) \leq C|t-s|, \qquad \forall\, s,t \in [0,T]. \tag{6.7}$$

We start by estimating $\int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu_t^n(x)$. We formally use the map $x \mapsto |x|^p$ as a test function for the weak formulation of the continuity equation (a rigorous approach would require an approximation of this map through cut-off functions, see [34, Section 5]). Defining $v^n(t,x) :=$

$f(x, \underline{u}^n(t,x), \mu_t^n)$, using the growth condition on $f$ given in (3.3) and Young inequality we obtain that

$$\begin{aligned}\int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu_t^n(x) &\le \int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu_0^n(x) + p\int_0^t \int_{\mathbb{R}^d} |v^n(s,x)||x|^{p-1} \,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s \\ &\le \int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu_0^n(x) + Cp\int_0^t \int_{\mathbb{R}^d} (1+|x|+\mathrm{m}_p(\mu_s^n))\,|x|^{p-1} \,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s \\ &\le \int_{\mathbb{R}^d} |x|^p \mathrm{d}\,\mu_0^n(x) + \tilde{C}\int_0^t \int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s,\end{aligned}$$

for some $\tilde{C} > 0$ independent of $n$ and $t \in [0,T]$. By Gronwall's inequality and (6.4) we get that

$$\sup_{n\in\mathbb{N}} \sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu_t^n(x) < +\infty. \tag{6.8}$$

Formally using the map $x \mapsto \psi(|x|^p)$ as a test function for the weak formulation of the continuity equation, by the growth condition on $f$ in (3.3) and the bound (6.8), we have

(6.9)

$$\begin{aligned}\int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_t^n(x) &\le \int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_0^n(x) + \int_0^t \int_{\mathbb{R}^d} |v^n(s,x)||\nabla(\psi(|x|^p))|\,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s \\ &\le \int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_0^n(x) + pC\int_0^t \int_{\mathbb{R}^d} (1+|x|+\mathrm{m}_p(\mu_s^n)))\,\psi'(|x|^p)|x|^{p-1}\,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s \\ &\le \int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_0^n(x) + C_1\int_0^t \int_{\mathbb{R}^d} \left(|x|^{p-1}+|x|^p\right)\psi'(|x|^p)\,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s\end{aligned}$$

for some $C_1 > 0$ independent of $n$ and $t \in [0,T]$. Notice that by the monotonicity of $\psi'$, denoting by $B_1$ the unitary ball of $\mathbb{R}^d$, we have

$$\begin{aligned}\int_{\mathbb{R}^d} |x|^{p-1}\psi'(|x|^p)\,\mathrm{d}\mu_s^n(x) &\le \int_{B_1} |x|^{p-1}\psi'(|x|^p)\,\mathrm{d}\mu_s^n(x) + \int_{\mathbb{R}^d\setminus B_1} |x|^p\psi'(|x|^p)\,\mathrm{d}\mu_s^n(x) \\ &\le \psi'(1) + \int_{\mathbb{R}^d} |x|^p\psi'(|x|^p)\,\mathrm{d}\mu_s^n(x).\end{aligned}$$

By the previous inequality and (2.8), from (6.9) we get

$$\int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_t^n(x) \le \int_{\mathbb{R}^d} \psi(|x|^p)\,\mathrm{d}\mu_0^n(x) + C_2\int_0^t \int_{\mathbb{R}^d} (1+\psi(|x|^p))\,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s$$

for some $C_2 > 0$ independent of $n$ and $t \in [0,T]$. By Gronwall's inequality and (6.5) we obtain (6.6).

Using Benamou-Brenier formula (2.7), the growth condition on $f$ in (3.3), for $s,t \in [0,T]$, $s \le t$, it holds

$$\begin{aligned}W_p^p(\mu_s^n, \mu_t^n) &\le \int_s^t \int_{\mathbb{R}^d} |v^n(r,x)|^p \,\mathrm{d}\mu_r^n(x)\mathrm{d}r \\ &\le C\int_s^t \int_{\mathbb{R}^d} (1+|x|+\mathrm{m}_p(\mu_r^n))^p \,\mathrm{d}\mu_r^n(x)\mathrm{d}r,\end{aligned}$$

for some $C > 0$ independent of $n$, $s$ and $t$. Using the bound (6.8) we obtain (6.7).

By Ascoli-Arzelà theorem in $\mathscr{P}_p(\mathbb{R}^d)$ there exists $\mu \in C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ and a subsequence $\mu^n$ (not relabeled) such that $\mu^n \to \mu$ in $C([0,T];\mathscr{P}_p(\mathbb{R}^d))$.

For what concerns the weak compactness of $u^n$ (in the sense of convergence (6.3)), we denote by $\tilde{\mu}^n = \mu_t^n \otimes \mathcal{L}_T \in \mathscr{P}([0,T]\times\mathbb{R}^d)$ and $\tilde{\mu} = \mu_t \otimes \mathcal{L}_T \in \mathscr{P}([0,T]\times\mathbb{R}^d)$. From the convergence of $\mu^n$ to $\mu$ it follows that $\tilde{\mu}^n \to \tilde{\mu}$ weakly. Defining $\gamma^n := (i_{[0,T]\times\mathbb{R}^d}, \underline{u}^n)_\sharp \tilde{\mu}^n \in \mathscr{P}(([0,T]\times\mathbb{R}^d)\times U)$, we observe that $\pi^1_\sharp\gamma^n = \tilde{\mu}^n \in \mathscr{P}([0,T]\times\mathbb{R}^d)$ and $\pi^2_\sharp\gamma^n = \underline{u}^n_\sharp\tilde{\mu}^n \in \mathscr{P}(U)$, where $\pi^1 : [0,T]\times\mathbb{R}^d\times U \to [0,T]\times\mathbb{R}^d$ is the projection on $[0,T]\times\mathbb{R}^d$ and $\pi^2 : [0,T]\times\mathbb{R}^d\times U \to U$ is the projection on $U$.

Since $\pi^1_\sharp\gamma^n$ weakly converges and $U$ is compact, the families $\{\pi^1_\sharp\gamma^n\}_{n\in\mathbb{N}}$ and $\{\pi^2_\sharp\gamma^n\}_{n\in\mathbb{N}}$ are tight. Thanks to [3, Lemma 5.2.2] it follows that $\{\gamma^n\}_{n\in\mathbb{N}}$ is tight and, by Prokhorov's Theorem, there exists $\gamma\in\mathscr{P}([0,T]\times\mathbb{R}^d\times U)$ and a subsequence $\gamma^n$ (not relabeled) such that, $\gamma^n\to\gamma$ weakly, as $n\to+\infty$.
Let $\phi\in C_c([0,T]\times\mathbb{R}^d;V')$. Recalling that $U$ is compact, using the continuous and bounded test function $(t,x,u)\mapsto\langle\phi(t,x),u\rangle$, by the weak convergence of $\gamma^n$ to $\gamma$ we have

$$\begin{aligned}&\int_0^T\int_{\mathbb{R}^d}\langle\phi(t,x),\underline{u}^n(t,x)\rangle\,\mathrm{d}\tilde{\mu}^n(t,x)=\\&=\int_0^T\int_{\mathbb{R}^d}\int_U\langle\phi(t,x),u\rangle\,\mathrm{d}\gamma^n(t,x,u)\to\int_0^T\int_{\mathbb{R}^d}\int_U\langle\phi(t,x),u\rangle\,\mathrm{d}\gamma(t,x,u),\end{aligned}\tag{6.10}$$

as $n\to+\infty$. Using Theorem 2.6 (specifically Remark 2.7) and observing that $\pi^1_\sharp\gamma=\tilde{\mu}$, we disintegrate $\gamma$ with respect to $\pi^1$ to get

$$\int_0^T\int_{\mathbb{R}^d}\int_U\langle\phi(t,x),u\rangle\,\mathrm{d}\gamma(t,x,u)=\int_0^T\int_{\mathbb{R}^d}\int_U\langle\phi(t,x),u\rangle\,\mathrm{d}\gamma_{t,x}(u)\,\mathrm{d}\tilde{\mu}(t,x).$$

We define now $\underline{u}:[0,T]\times\mathbb{R}^d\to U$ by

$$\underline{u}(t,x):=\int_U u\,\mathrm{d}\gamma_{t,x}(u),\qquad\forall\,(t,x)\in[0,T]\times\mathbb{R}^d,\tag{6.11}$$

where the integral in (6.11) is a Bochner integral. Since the map $(t,x)\in[0,T]\times\mathbb{R}^d\mapsto\gamma_{t,x}\in\mathscr{P}(U)$ is a Borel map, then $\underline{u}\in\mathrm{B}([0,T]\times\mathbb{R}^d;U)$. We call the map $\underline{u}$ the barycentric projection of $\gamma$ with respect to $\pi^1_\sharp\gamma$. Since the Bochner integral commutes with continuous linear functionals, it holds

$$\int_0^T\int_{\mathbb{R}^d}\int_U\langle\phi(t,x),u\rangle\,\mathrm{d}\gamma_{t,x}(u)\,\mathrm{d}\tilde{\mu}(t,x)=\int_0^T\int_{\mathbb{R}^d}\langle\phi(t,x),\underline{u}(t,x)\rangle\,\mathrm{d}\tilde{\mu}(t,x).$$

Using (6.10) we obtain the convergence of $\underline{u}^n\to\underline{u}$ in the sense of (6.3).

In order to prove that $(\mu,\underline{u})\in\mathcal{A}_{\mathbf{E}}(\mu_0)$ we show that (6.1) is satisfied. Let $\varphi\in C^1_c([0,T]\times\mathbb{R}^d)$. Since $(\mu^n,\underline{u}^n)\in\mathcal{A}_{\mathbf{E}}(\mu^n_0)$, for every $t\in[0,T]$ it holds

$$\int_{\mathbb{R}^d}\varphi(t,x)\,\mathrm{d}\mu^n_t(x)-\int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu^n_0(x)=\int_0^t\int_{\mathbb{R}^d}\left(\partial_s\varphi(s,x)+v^n(s,x)\cdot\nabla\varphi(s,x)\right)\,\mathrm{d}\mu^n_s(x)\,\mathrm{d}s.\tag{6.12}$$

By the convergence $\mu^n\to\mu$ in $C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ we immediately pass to the limit, as $n\to+\infty$, in the left hand side of equation (6.12) as well as on the term $\int_0^t\int_{\mathbb{R}^d}\partial_s\varphi(s,x)\,\mathrm{d}\mu^n_s(x)\,\mathrm{d}s$. Finally, let us rewrite

$$\begin{aligned}&\int_0^t\int_{\mathbb{R}^d}v^n(s,x)\cdot\nabla\varphi(s,x)\,\mathrm{d}\mu^n_s(x)\,\mathrm{d}s\\&=\int_0^t\int_{\mathbb{R}^d}f(x,\underline{u}^n(s,x),\mu^n_s)\cdot\nabla\varphi(s,x)\,\mathrm{d}\mu^n_s(x)\,\mathrm{d}s\\&=\int_{[0,t]\times\mathbb{R}^d\times U}f(x,u,\mu^n_s)\cdot\nabla\varphi(s,x)\,\mathrm{d}\gamma^n(s,x,u)\\&=\int_{[0,t]\times\mathbb{R}^d\times U}(f(x,u,\mu^n_s)-f(x,u,\mu_s))\cdot\nabla\varphi(s,x)\,\mathrm{d}\gamma^n(s,x,u)\\&\quad+\int_{[0,t]\times\mathbb{R}^d\times U}f(x,u,\mu_s)\cdot\nabla\varphi(s,x)\,\mathrm{d}\gamma^n(s,x,u).\end{aligned}$$

By (3.1), the first term on the right hand side can be estimated by

$$\begin{aligned}
&\left|\int_{[0,t]\times\mathbb{R}^d\times U} \langle f(x,u,\mu_s^n) - f(x,u,\mu_s), \nabla\varphi(s,x)\rangle \,\mathrm{d}\gamma^n(s,x,u)\right| \\
&\quad \le L \sup_{r\in[0,T]} W_p(\mu_r^n,\mu_r) \int_{[0,t]\times\mathbb{R}^d\times U} |\nabla\varphi(s,x)| \,\mathrm{d}\gamma^n(s,x,u) \\
&\quad = L \sup_{r\in[0,T]} W_p(\mu_r^n,\mu_r) \int_{[0,t]\times\mathbb{R}^d} |\nabla\varphi(s,x)| \,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s,
\end{aligned}$$

which goes to zero as $n\to+\infty$ by the convergence $\mu^n\to\mu$ in $C([0,T];\mathscr{P}_p(\mathbb{R}^d))$. Hence

$$\begin{aligned}
&\lim_{n\to+\infty}\int_0^t\int_{\mathbb{R}^d} v^n(s,x)\cdot\nabla\varphi(s,x)\,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s \\
&= \lim_{n\to+\infty}\int_{[0,t]\times\mathbb{R}^d\times U} f(x,u,\mu_s)\cdot\nabla\varphi(t,x)\,\mathrm{d}\gamma^n(s,x,u) \\
&\qquad = \int_{[0,t]\times\mathbb{R}^d\times U} f(x,u,\mu_s)\cdot\nabla\varphi(t,x)\,\mathrm{d}\gamma(s,x,u),
\end{aligned}$$

by the weak convergence of $\gamma^n$ to $\gamma$. Recall that, by the Convexity Assumption 3.4, for any $(x,s)\in\mathbb{R}^d\times[0,T]$ the map $u\mapsto f(x,u,\mu_s)$ is affine. Thus, using that $\gamma_{s,x}$ is a probability measure, we have

$$\int_U f(x,u,\mu_s)\,\mathrm{d}\gamma_{s,x}(u) = f\left(x,\int_U u\,\mathrm{d}\gamma_{s,x}(u),\mu_s\right), \qquad \text{for } \tilde\mu\text{-a.e. } (s,x)\in[0,T]\times\mathbb{R}^d.$$

Hence we get

$$\begin{aligned}
&\int_{[0,t]\times\mathbb{R}^d}\int_U f(x,u,\mu_s)\cdot\nabla\varphi(s,x)\,\mathrm{d}\gamma_{s,x}(u)\,\mathrm{d}\tilde\mu(s,x) \\
&\qquad = \int_{[0,t]\times\mathbb{R}^d} f\left(x,\int_U u\,\mathrm{d}\gamma_{s,x}(u),\mu_s\right)\cdot\nabla\varphi(s,x)\,\mathrm{d}\tilde\mu(s,x) \\
&\qquad = \int_0^t\int_{\mathbb{R}^d} f\left(x,\underline{u}(s,x),\mu_s\right)\cdot\nabla\varphi(s,x)\,\mathrm{d}\mu_s(x)\,\mathrm{d}s.
\end{aligned}$$

Defining $v(t,x) := f(x,\underline{u}(t,x),\mu_t)$, we have proved that

$$\begin{aligned}
&\lim_{n\to+\infty}\int_0^t\int_{\mathbb{R}^d} v^n(s,x)\cdot\nabla\varphi(s,x)\,\mathrm{d}\mu_s^n(x)\,\mathrm{d}s \\
&\qquad = \int_0^t\int_{\mathbb{R}^d} v(s,x)\cdot\nabla\varphi(s,x)\,\mathrm{d}\mu_s(x)\,\mathrm{d}s
\end{aligned}$$

and this concludes the proof. □

**Proposition 6.7** (Lower semicontinuity for convex $J_{\mathbf{E}}$)**.** *Let $\mathbb{S}=(U,f,\mathcal{C},\mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. If $(\mu^n,\underline{u}^n)$ converges to $(\mu,\underline{u})$ according to Definition 6.5, then*

$$\liminf_{n\to+\infty} J_{\mathbf{E}}(\mu^n,\underline{u}^n) \ge J_{\mathbf{E}}(\mu,\underline{u}). \tag{6.13}$$

*Proof.* Denoting by $\tilde\mu^n=\mu_t^n\otimes\mathcal{L}_T\in\mathscr{P}([0,T]\times\mathbb{R}^d)$ and $\tilde\mu=\mu_t\otimes\mathcal{L}_T\in\mathscr{P}([0,T]\times\mathbb{R}^d)$, we define $\gamma^n := (i_{[0,T]\times\mathbb{R}^d},\underline{u}^n)_\sharp\tilde\mu^n\in\mathscr{P}(([0,T]\times\mathbb{R}^d)\times U)$. Reasoning as in the proof of Proposition 6.6, we obtain that there exists $\gamma\in\mathscr{P}(([0,T]\times\mathbb{R}^d)\times U)$ such that $\pi^1_\sharp\gamma=\tilde\mu$ and, up to subsequences, $\gamma^n\to\gamma$ weakly as $n\to+\infty$. Moreover, defining $\underline{u}_b$ by (6.11), up to subsequences, $(\mu^n,\underline{u}^n)$ converges to $(\mu,\underline{u}_b)$ according to Definition 6.5. Then,

$$\int_0^T\int_{\mathbb{R}^d}\langle\phi(t,x),\underline{u}(t,x)\rangle\,\mathrm{d}\tilde\mu(t,x) = \int_0^T\int_{\mathbb{R}^d}\langle\phi(t,x),\underline{u}_b(t,x)\rangle\,\mathrm{d}\tilde\mu(t,x), \quad \forall\,\phi\in C_c([0,T]\times\mathbb{R}^d;V'),$$

which implies that $\underline{u}_b(t,x) = \underline{u}(t,x)$ for $\tilde{\mu}$-a.e. $(t,x) \in [0,T] \times \mathbb{R}^d$.

Let $\mathcal{K} \subset \mathscr{P}_p(\mathbb{R}^d)$ be a compact set containing $\mu_t^n, \mu_t$ for any $n \in \mathbb{N}$ and $t \in [0,T]$. Then, for any compact $K \subset \mathbb{R}^d$, thanks to the continuity of $\mathcal{C}$, there exists a modulus of continuity $\alpha : [0,+\infty) \to [0,+\infty)$ for the restriction of $\mathcal{C}$ to the compact set $K \times U \times \mathcal{K}$ so that

$$\sup_{(t,x,u)\in[0,T]\times K\times U} |\mathcal{C}(x,u,\mu_t^n) - \mathcal{C}(x,u,\mu_t)| \le \alpha\Big(\sup_{t\in[0,T]} W_p(\mu_t^n,\mu_t)\Big) \to 0 \quad \text{as } n \to +\infty.$$

Then, taking into account that $\mathcal{C} \ge 0$, we get

$$\begin{aligned}
&\liminf_{n\to+\infty} \int_0^T \int_{\mathbb{R}^d} \mathcal{C}(x,\underline{u}^n(t,x),\mu_t^n)\,\mathrm{d}\mu_t^n(x)\,\mathrm{d}t \\
&\ge \liminf_{n\to+\infty} \int_{[0,T]\times K\times U} \mathcal{C}(x,u,\mu_t^n)\,\mathrm{d}\gamma^n(t,x,u) \\
&\ge \int_{[0,T]\times K\times U} \mathcal{C}(x,u,\mu_t)\,\mathrm{d}\gamma(t,x,u).
\end{aligned}$$

Since $K$ is arbitrary we obtain

$$\liminf_{n\to+\infty} \int_0^T \int_{\mathbb{R}^d} \mathcal{C}(x,\underline{u}^n(t,x),\mu_t^n)\,\mathrm{d}\mu_t^n(x)\,\mathrm{d}t \ge \int_{[0,T]\times\mathbb{R}^d\times U} \mathcal{C}(x,u,\mu_t)\,\mathrm{d}\gamma(t,x,u). \tag{6.14}$$

Denoting $\gamma_{t,x}$ the disintegration of $\gamma$ with respect to $\pi^1$ (as in the proof of Proposition 6.6), the convexity of the map $u \mapsto \mathcal{C}(x,u,\mu_t)$ for any $(t,x) \in [0,T] \times \mathbb{R}^d$ and Jensen's inequality yield

$$\begin{aligned}
\int_{[0,T]\times\mathbb{R}^d\times U} \mathcal{C}(x,u,\mu_t)\,\mathrm{d}\gamma(t,x,u) &= \int_0^T \int_{\mathbb{R}^d} \int_U \mathcal{C}(x,u,\mu_t)\,\mathrm{d}\gamma_{t,x}(u)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\ge \int_0^T \int_{\mathbb{R}^d} \mathcal{C}\left(x, \int_U u\,\mathrm{d}\gamma_{t,x}(u), \mu_t\right)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&= \int_0^T \int_{\mathbb{R}^d} \mathcal{C}\left(x, \underline{u}_b(t,x), \mu_t\right)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&= \int_0^T \int_{\mathbb{R}^d} \mathcal{C}\left(x, \underline{u}(t,x), \mu_t\right)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t.
\end{aligned}$$

By the continuity of $\mathcal{C}_T$, using the same argument of the proof of (6.14), we obtain

$$\liminf_{n\to+\infty} \int_{\mathbb{R}^d} \mathcal{C}_T(x,\mu_T^n)\,\mathrm{d}\mu_T^n(x) \ge \int_{\mathbb{R}^d} \mathcal{C}_T(x,\mu_T)\,\mathrm{d}\mu_T(x).$$

□

Propositions 6.6 and 6.7 give immediately the existence of optimizers for our optimal control problem in Eulerian formulation.

**Theorem 6.8** (Existence of minimizers for convex $\mathbf{E}$)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. If $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, then there exists $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ such that*

$$J_{\mathbf{E}}(\mu, \underline{u}) = V_{\mathbf{E}}(\mu_0).$$

As a consequence, we derive the lower semicontinuity of the value function for the Eulerian problem.

**Proposition 6.9** (Lower semicontinuity of $V_{\mathbf{E}}$)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4 and $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$. If $\{\mu_0^n\}_{n\in\mathbb{N}} \subset \mathscr{P}_p(\mathbb{R}^d)$ is a sequence such that $W_p(\mu_0^n, \mu_0) \to 0$ as $n \to +\infty$, then*

$$\liminf_{n\to+\infty} V_{\mathbf{E}}(\mu_0^n) \ge V_{\mathbf{E}}(\mu_0).$$

*Proof.* From Theorem 6.8 there exists a sequence $(\mu^n, \underline{u}^n) \in \mathcal{A}_{\mathbf{E}}(\mu_0^n)$ such that $V_{\mathbf{E}}(\mu_0^n) = J_{\mathbf{E}}(\mu^n, \underline{u}^n)$. By Proposition 6.6 there exists $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ such that, up to subsequences, $(\mu^n, \underline{u}^n)$ converges to $(\mu, \underline{u})$ according to Definition 6.5. Then, using Proposition 6.7, we have

$$\liminf_{n\to+\infty} V_{\mathbf{E}}(\mu_0^n) = \liminf_{n\to+\infty} J_{\mathbf{E}}(\mu^n, \underline{u}^n) \geq J_{\mathbf{E}}(\mu, \underline{u}) \geq V_{\mathbf{E}}(\mu_0).$$

□

## 7. Kantorovich optimal control problem and equivalence with the Eulerian

In this section, we provide a further formulation of optimal control problems which we call *Kantorovich formulation* in analogy with the terminology used in optimal transport theory. This formulation acts as a bridge between the Lagrangian and the Eulerian formulations and it is based on the representation of solutions of the continuity equation by superposition of continuous curves in $\Gamma_T = C([0,T];\mathbb{R}^d)$ (see Theorem 2.5). This formulation turns out to be equivalent to the Eulerian one and it will be useful in Section 8 to prove the equivalence between the Eulerian and the Lagrangian problems.

We recall that, for any $t \in [0,T]$, $e_t : \Gamma_T \to \mathbb{R}^d$ denotes the evaluation map $e_t(\gamma) := \gamma(t)$.

**Definition 7.1** (Kantorovich optimal control problem (**K**))**.** Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Given $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, we say that $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ if

(i) $u \in \mathrm{B}([0,T] \times \Gamma_T; U)$;
(ii) $\boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$, $(e_0)_\sharp \boldsymbol{\eta} = \mu_0$ and, defining $\mu_t := (e_t)_\sharp \boldsymbol{\eta}$ for all $t \in [0,T]$,

$$\int_0^T \int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu_t(x)\,\mathrm{d}t < +\infty. \tag{7.1}$$

$\boldsymbol{\eta}$ is concentrated on the set of absolutely continuous solutions of the differential equation

$$\dot\gamma(t) = f(\gamma(t), u(t,\gamma), \mu_t), \qquad \text{for } \mathcal{L}_T\text{-a.e. } t \in [0,T].$$

We define the *cost functional* $J_{\mathbf{K}} : \mathscr{P}(\Gamma_T) \times \mathrm{B}([0,T] \times \Gamma_T; U) \to [0,+\infty]$ by

$$J_{\mathbf{K}}(\boldsymbol{\eta}, u) := \int_0^T \int_{\Gamma_T} \mathcal{C}(\gamma(t), u(t,\gamma), \mu_t)\,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}t + \int_{\mathbb{R}^d} \mathcal{C}_T(x, \mu_T)\,\mathrm{d}\mu_T(x),$$

and the *value function* $V_{\mathbf{K}} : \mathscr{P}_p(\mathbb{R}^d) \to [0,+\infty)$ by

$$V_{\mathbf{K}}(\mu_0) := \inf\{J_{\mathbf{K}}(\boldsymbol{\eta}, u) \,:\, (\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)\}.$$

*Remark* 7.2. We observe that, by the growth condition of $f$ in (3.3) and condition (7.1), $\boldsymbol{\eta}$ is actually concentrated on $\mathrm{AC}^p([0,T];\mathbb{R}^d)$, indeed

$$\begin{aligned}\int_0^T \int_{\Gamma_T} |\dot\gamma(t)|^p \,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}t &\leq \int_0^T \int_{\Gamma_T} |f(\gamma(t), u(t,\gamma), \mu_t)|^p \,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}t \\ &\leq C \int_0^T \int_{\Gamma_T} |1 + |\gamma(t)| + \mathrm{m}_p(\mu)|^p \,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}t \\ &\leq \tilde C\Big(1 + \int_0^T \int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu_t(x)\mathrm{d}t\Big).\end{aligned}$$

Hence, by Fubini theorem, for $\boldsymbol{\eta}$-a.e. $\gamma \in \Gamma_T$, $\dot\gamma \in L^p([0,T];\mathbb{R}^d)$.

Moreover, thanks to (7.1) and the growth condition (3.2), $J_{\mathbf{K}}(\boldsymbol{\eta}, u) < +\infty$ for every $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ (the proof of $\int_{\mathbb{R}^d} |x|^p \,\mathrm{d}\mu_T(x) < +\infty$ follows by items (i) and (ii) and the same argument used to show (6.8) in Proposition 6.6).
Finally, the value function $V_{\mathbf{K}}$ is well defined since $\mathcal{A}_{\mathbf{K}}$ is non empty (see Propostion 7.4).

The aim of this section is to prove the existence of minimizers for the Kantorovich optimal control problem under the Convexity Assumptions 3.4 and to show the equivalence with the Eulerian formulation of the problem described in Section 6. In particular, we get the equality of the corresponding value functions. This is the content of the following theorem.

**Theorem 7.3.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. If $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, then there exist $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ and $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ such that*

$$J_{\mathbf{K}}(\boldsymbol{\eta}, u) = J_{\mathbf{E}}(\mu, \underline{u}) = V_{\mathbf{K}}(\mu_0) = V_{\mathbf{E}}(\mu_0).$$

The proof of Theorem 7.3 follows by the combination of Theorem 6.8 and Propositions 7.4, 7.5 below.

**Proposition 7.4.** *Let $\mathbb{S} := (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Let $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$. If $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$, then there exists $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ such that $J_{\mathbf{K}}(\boldsymbol{\eta}, u) = J_{\mathbf{E}}(\mu, \underline{u})$. In particular $\mathcal{A}_{\mathbf{K}}(\mu_0) \neq \emptyset$ and $V_{\mathbf{K}}(\mu_0) \leq V_{\mathbf{E}}(\mu_0)$.*

*Proof.* Let $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$. Applying Theorem 2.5 to $\mu$ and $v_t(x) = f(x, \underline{u}(t, x), \mu_t)$ we get $\boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$ such that $(e_t)_\sharp \boldsymbol{\eta} = \mu_t$ for every $t \in [0, T]$ and $\boldsymbol{\eta}$ is concentrated on the absolutely continuous solutions of $\dot{\gamma}(t) = v_t(\gamma(t))$. Condition (7.1) is automatically satisfied in view of the fact that $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$. Then, for every $(t, \gamma) \in [0, T] \times \Gamma_T$ we define $u(t, \gamma) := \underline{u}(t, \gamma(t))$ so that $u$ is Borel measurable and we have $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$. Finally, from the evaluation $(e_t)_\sharp \boldsymbol{\eta} = \mu_t$ it holds that $J_{\mathbf{K}}(\boldsymbol{\eta}, u) = J_{\mathbf{E}}(\mu, \underline{u})$. □

Under the Convexity Assumption 3.4 it also holds that $V_{\mathbf{K}}(\mu_0) \geq V_{\mathbf{E}}(\mu_0)$.

**Proposition 7.5.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. Let $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$. If $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$, then there exists $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ such that $J_{\mathbf{E}}(\mu, \underline{u}) \leq J_{\mathbf{K}}(\boldsymbol{\eta}, u)$. In particular $V_{\mathbf{K}}(\mu_0) \geq V_{\mathbf{E}}(\mu_0)$.*

*Proof.* Let $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$. We firstly define $\mu_t := (e_t)_\sharp \boldsymbol{\eta}$, for every $t \in [0, T]$. We introduce the continuous evaluation map $e : [0, T] \times \Gamma_T \to [0, T] \times \mathbb{R}^d$ by setting

$$e(t, \gamma) := (t, \gamma(t)) = (i_{[0,T]}(t), e_t(\gamma))$$

and we denote by $(t, x) \mapsto \tilde{\boldsymbol{\eta}}_{t,x}$ the Borel map obtained from the disintegration of $\mathcal{L}_T \otimes \boldsymbol{\eta}$ with respect to $e$, see Theorem 2.6. Then we define the function $\underline{u} : [0, T] \times \mathbb{R}^d \to U$ by

$$\underline{u}(t, x) := \int_{[0,T] \times \Gamma_T} u(t, \gamma) \, \mathrm{d}\tilde{\boldsymbol{\eta}}_{t,x}(\gamma).$$

Notice that $\underline{u}$ is Borel measurable thanks to the Borel measurability of $(t, x) \mapsto \tilde{\boldsymbol{\eta}}_{t,x}$. The measure $\tilde{\boldsymbol{\eta}}_{t,x}$ is concentrated on $\{t\} \times \{\gamma : \gamma(t) = x\}$, so that $\tilde{\boldsymbol{\eta}}_{t,x} = \delta_t \otimes \boldsymbol{\eta}_{t,x}$, where for any $t \in [0, T]$, the function $x \mapsto \boldsymbol{\eta}_{t,x}$ is the Borel map given by the disintegration of $\boldsymbol{\eta}$ with respect to the continuous map $e_t$. Hence we have also that

$$\underline{u}(t, x) = \int_{\Gamma_T} u(t, \gamma) \, \mathrm{d}\boldsymbol{\eta}_{t,x}(\gamma). \tag{7.2}$$

Defining the set $A := \{(t, \gamma) \in [0, T] \times \Gamma_T : \exists \, \dot{\gamma}(t) \text{ and } \dot{\gamma}(t) = f(\gamma(t), u(t, \gamma), \mu_t)\}$, by item (ii) of Definition 7.1 we have that $\mathcal{L}_T \otimes \boldsymbol{\eta}(([0, T] \times \Gamma_T) \setminus A) = 0$. Then, for $\mathcal{L}_T$-a.e. $t \in [0, T]$ we have

$$\dot{\gamma}(t) = f(\gamma(t), u(t, \gamma), \mu_t), \qquad \text{for } \boldsymbol{\eta}\text{-a.e. } \gamma \in \Gamma_T.$$

Let $\varphi \in C^1_c(\mathbb{R}^d;\mathbb{R})$. For any $s,t \in [0,T]$, $s<t$, we have

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_s(x) &= \int_{\Gamma_T} (\varphi(\gamma(t)) - \varphi(\gamma(s)))\,\mathrm{d}\boldsymbol{\eta}(\gamma)\\ &= \int_{\Gamma_T}\int_s^t \frac{\mathrm{d}}{\mathrm{d}r}\varphi(\gamma(r))\,\mathrm{d}r\,\mathrm{d}\boldsymbol{\eta}(\gamma)\\ &= \int_{\Gamma_T}\int_s^t \nabla\varphi(\gamma(r))\cdot\dot\gamma(r)\,\mathrm{d}r\,\mathrm{d}\boldsymbol{\eta}(\gamma)\\ &= \int_s^t\int_{\Gamma_T} \nabla\varphi(\gamma(r))\cdot f(\gamma(r),u(r,\gamma),\mu_r)\,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}r.\end{aligned} \tag{7.3}$$

Using the growth condition of $f$ in (3.3) we have

$$\left|\int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_s(x)\right| \le C\|\nabla\varphi\|_\infty \int_s^t\int_{\Gamma_T} (1+|\gamma(r)|+\mathrm{m}_p(\mu_r))\,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}r.$$

By (7.1) the map $r\mapsto \int_{\Gamma_T}(1+|\gamma(r)|+\mathrm{m}_p(\mu_r))\,\mathrm{d}\boldsymbol{\eta}(\gamma)$ belongs to $L^1(0,T)$ and then the map $t\mapsto\int_{\mathbb{R}^d}\varphi(x)\,\mathrm{d}\mu_t(x)$ is absolutely continuous.

Thanks to (7.3), for $\mathcal{L}_T$-a.e. $t\in[0,T]$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}\varphi(x)\,\mathrm{d}\mu_t(x) = \int_{\Gamma_T}\nabla\varphi(\gamma(t))\cdot f(\gamma(t),u(t,\gamma),\mu_t)\,\mathrm{d}\boldsymbol{\eta}(\gamma).$$

Using the affinity of $f$, the disintegration of $\boldsymbol{\eta}$ with respect to $e_t$, recalling that $\boldsymbol{\eta}_{t,x}$ is concentrated on $\{\gamma:\gamma(t)=x\}$, and by (7.2) we obtain

$$\begin{aligned}&\int_{\Gamma_T}\nabla\varphi(\gamma(t))\cdot f(\gamma(t),u(t,\gamma),\mu_t)\,\mathrm{d}\boldsymbol{\eta}(\gamma)\\ &= \int_{\mathbb{R}^d}\int_{\Gamma_T}\nabla\varphi(\gamma(t))\cdot f(\gamma(t),u(t,\gamma),\mu_t)\,\mathrm{d}\boldsymbol{\eta}_{t,x}(\gamma)\,\mathrm{d}\mu_t(x)\\ &= \int_{\mathbb{R}^d}\int_{\Gamma_T}\nabla\varphi(x)\cdot f(x,u(t,\gamma),\mu_t)\,\mathrm{d}\boldsymbol{\eta}_{t,x}(\gamma)\,\mathrm{d}\mu_t(x)\\ &= \int_{\mathbb{R}^d}\nabla\varphi(x)\cdot f\left(x,\int_{\Gamma_T}u(t,\gamma)\,\mathrm{d}\boldsymbol{\eta}_{t,x}(\gamma),\mu_t\right)\,\mathrm{d}\mu_t(x)\\ &= \int_{\mathbb{R}^d}\nabla\varphi(x)\cdot f(x,\underline{u}(t,x),\mu_t)\,\mathrm{d}\mu_t(x).\end{aligned}$$

Then $\mu$ satisfies the continuity equation $\partial_t\mu_t+\mathrm{div}(v_t\mu_t)=0$ for the vector field $v_t(x)=f(x,\underline{u}(t,x),\mu_t)$ in the sense of distributions (see e.g. [3, equation 8.1.4]). Since $|v_t(x)|^p\le \tilde C(1+|x|^p+\mathrm{m}_p^p(\mu_t))$, from (7.1) it follows that $t\mapsto\|v_t\|_{L^p_{\mu_t}(\mathbb{R}^d;\mathbb{R}^d)}$ belongs to $L^p(0,T)$ and then $\mu\in\mathrm{AC}^p([0,T];\mathscr{P}_p(\mathbb{R}^d))$. Hence, $(\mu,\underline{u})\in\mathcal{A}_{\mathbf{E}}(\mu_0)$.

Finally, by the convexity of $\mathcal{C}$ with respect to $u$ and Jensen's inequality we obtain

$$\begin{aligned}&\int_0^T\int_{\Gamma_T}\mathcal{C}(\gamma(t),u(t,\gamma),\mu_t)\,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}t\\ &= \int_0^T\int_{\mathbb{R}^d}\int_{\Gamma_T}\mathcal{C}(x,u(t,\gamma),\mu_t)\,\mathrm{d}\boldsymbol{\eta}_{t,x}(\gamma)\mathrm{d}\mu_t(x)\,\mathrm{d}t\\ &\ge \int_0^T\int_{\mathbb{R}^d}\mathcal{C}\left(x,\int_{\Gamma_T}u(t,\gamma)\,\mathrm{d}\boldsymbol{\eta}_{t,x}(\gamma),\mu_t\right)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\ &= \int_0^T\int_{\mathbb{R}^d}\mathcal{C}\left(x,\underline{u}(t,x),\mu_t\right)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t.\end{aligned}$$

Hence we obtain $J_{\mathbf{K}}(\boldsymbol{\eta},u)\ge J_{\mathbf{E}}(\mu,\underline{u})$. □

## 8. Equivalence of Eulerian and Lagrangian problems

In this Section we study the equivalence between the Eulerian and Lagrangian formulations of the optimal control problem. We anticipate here the main results of this section.

**Theorem 8.1.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms. If $X_0 \in L^p(\Omega; \mathbb{R}^d)$, then*

$$V_{\mathbf{L}}(X_0) = V_{\mathbf{E}}((X_0)_\sharp \mathbb{P}).$$

*In particular, given $X_0, X_0' \in L^p(\Omega; \mathbb{R}^d)$ s.t. $(X_0)_\sharp \mathbb{P} = (X_0')_\sharp \mathbb{P}$, then*

$$V_{\mathbf{L}}(X_0) = V_{\mathbf{L}}(X_0').$$

The proof of Theorem 8.1 follows immediately by Proposition 8.3, Theorem 8.5 and Theorem 7.3.

We stress again that the Convexity Assumption 3.4 is sufficient to prove the existence of a minimizer for the Eulerian and Kantorovich optimal control problems (see Theorems 6.8 and 7.3). However in general, even assuming the Convexity Assumption 3.4, the Lagrangian optimal control problem could not have minimizers as we show in Section 8.3.

If we remove the Convexity Assumption 3.4, we can still give the following equivalence result.

**Theorem 8.2.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $\mathbb{S}' = (\mathscr{U}, \mathscr{F}, \mathscr{C}, \mathscr{C}_T)$ as in Definition 3.5. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms. Let $\mathbf{L}, \mathbf{RL}$ be the Lagrangian and Relaxed Lagrangian problems associated to $\mathbb{S}$. Let also $\mathbf{L}', \mathbf{E}'$ the Lagrangian and Eulerian problems associated to $\mathbb{S}'$. If $X_0 \in L^p(\Omega; \mathbb{R}^d)$, then*

$$V_{\mathbf{L}}(X_0) = V_{\mathbf{RL}}(X_0) = V_{\mathbf{L}'}(X_0) = V_{\mathbf{E}'}((X_0)_\sharp \mathbb{P}) = V_{\mathbf{K}'}((X_0)_\sharp \mathbb{P}).$$

The proof of Theorem 8.2 is postponed at the end of Section 8.1.

### 8.1. Comparison between L, E and K.

We start by comparing the Eulerian and Lagrangian problems under the Convexity Assumption 3.4. Assuming $\mathbb{P}$ without atoms, we further exhibit the equivalence between the associated value functions exploiting the Kantorovich formulation introduced in Section 7.

The following is a first comparison between the Eulerian and Lagrangian problems.

**Proposition 8.3.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space. If $X_0 \in L^p(\Omega; \mathbb{R}^d)$ and $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$, then there exists $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}((X_0)_\sharp \mathbb{P})$ such that $J_{\mathbf{L}}(X, u) \geq J_{\mathbf{E}}(\mu, \underline{u})$. In particular, $V_{\mathbf{L}}(X_0) \geq V_{\mathbf{E}}((X_0)_\sharp \mathbb{P})$.*

*Proof.* Let $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$. We firstly define $\mu_t := (X_t)_\sharp \mathbb{P}$, for every $t \in [0, T]$. Thanks to Proposition A.3 in Appendix A, $X \in \mathrm{AC}^p([0, T]; L^p(\Omega; \mathbb{R}^d))$ so that $\mu \in \mathrm{AC}^p([0, T]; \mathscr{P}_p(\mathbb{R}^d))$.

We define $\sigma \in \mathscr{P}([0, T] \times \Omega \times U)$ by $\sigma := \delta_{u(t,\omega)} \otimes \mathbb{P} \otimes \mathcal{L}_T$ and $\theta \in \mathscr{P}([0, T] \times \mathbb{R}^d \times U)$ by $\theta := (i_{[0,T]}, X_t, i_U)_\sharp \sigma$. Let $\pi^{1,2} : [0, T] \times \mathbb{R}^d \times U \to [0, T] \times \mathbb{R}^d$ be the projection map $(t, x, u) \mapsto (t, x)$, observe that $\pi^{1,2}_\sharp \theta = \mu_t \otimes \mathcal{L}_T$. If we denote $\theta_{t,x}$ the disintegration of $\theta$ with respect to $\pi^{1,2}$, then we have $\theta = \theta_{t,x} \otimes \mu_t \otimes \mathcal{L}_T$. We define now $\underline{u} : [0, T] \times \mathbb{R}^d \to U$ by

$$\underline{u}(t, x) := \int_U u \, \mathrm{d}\theta_{t,x}(u). \tag{8.1}$$

Thanks to Theorem 2.6, the map $(t, x) \mapsto \theta_{t,x}$ is Borel measurable, so that $\underline{u} \in \mathrm{B}([0, T] \times \mathbb{R}^d; U)$.

The rest of the proof follows the same line of the proof of Proposition 7.5. We write the details for the reader's convenience.

Defining the set $A := \{(t, \omega) \in [0, T] \times \Omega : \exists \, \dot{X}_t(\omega) \text{ and } \dot{X}_t(\omega) = f(X_t(\omega), u(t, \omega), \mu_t)\}$, by item (ii) of Definition 4.1, we have that $\mathcal{L}_T \otimes \mathbb{P}(([0, T] \times \Omega \setminus A) = 0$. Then, for $\mathcal{L}_T$-a.e. $t \in [0, T]$ it holds

$$\dot{X}_t(\omega) = f(X_t(\omega), u(t, \omega), \mu_t), \qquad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega.$$

Let $\varphi \in C_c^1(\mathbb{R}^d;\mathbb{R})$. For any $s,t \in [0,T]$, $s<t$, we have

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_s(x) &= \int_\Omega (\varphi(X_t(\omega)) - \varphi(X_s(\omega)))\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_\Omega \int_s^t \frac{\mathrm{d}}{\mathrm{d}r}\varphi(X_r(\omega))\,\mathrm{d}r\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_\Omega \int_s^t \nabla\varphi(X_r(\omega))\cdot \dot{X}_r(\omega)\,\mathrm{d}r\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_s^t \int_\Omega \nabla\varphi(X_r(\omega))\cdot f(X_r(\omega)), u(r,\omega),\mu_r)\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}r.\end{aligned} \tag{8.2}$$

Using the growth condition of $f$ in (3.3) we have

$$\left|\int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_s(x)\right| \le C\|\nabla\varphi\|_\infty \int_s^t \int_\Omega (1+|X_r(\omega)| + \mathrm{m}_p(\mu_r))\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}r.$$

Thanks to (4.9) in Proposition 4.9, it follows that the map $r \mapsto \int_\Omega (1+|X_r(\omega)|+\mathrm{m}_p(\mu_r))\,\mathrm{d}\mathbb{P}(\omega)$ belongs to $L^1(0,T)$ so that the map $t \mapsto \int_{\mathbb{R}^d}\varphi(x)\,\mathrm{d}\mu_t(x)$ is absolutely continuous. Then, from (8.2) it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu_t(x) = \int_\Omega \nabla\varphi(X_t(\omega))\cdot f(X_t(\omega)), u(t,\omega),\mu_t)\,\mathrm{d}\mathbb{P}(\omega), \qquad \text{for } \mathcal{L}_T\text{-a.e.} t\in[0,T].$$

For $\mathcal{L}_T$-a.e. $t\in[0,T]$, we denote now $\sigma_t \in \mathscr{P}(\Omega\times U)$ and $\theta_t \in \mathscr{P}(\mathbb{R}^d\times U)$ the disintegrations of $\sigma$ and $\theta$ with respect to the projection maps $\pi^1_\Omega : [0,T]\times\Omega\times U \to [0,T]$, $\pi^1_{\mathbb{R}^d} : [0,T]\times\mathbb{R}^d\times U \to [0,T]$, respectively. It can be shown that $\theta_t = (X_t, i_U)_\sharp \sigma_t = \theta_{t,x}\otimes\mu_t$, for $\mathcal{L}_T$-a.e. $t\in[0,T]$.

Using the affinity of $f$, and the definition of $\underline{u}$ in (8.1) we obtain

$$\begin{aligned}&\int_\Omega \nabla\varphi(X_t(\omega))\cdot f(X_t(\omega)), u(t,\omega),\mu_t)\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_{\Omega\times U} \nabla\varphi(X_t(\omega))\cdot f(X_t(\omega),u,\mu_t)\,\mathrm{d}\sigma_t(\omega,u)\\ &= \int_{\mathbb{R}^d\times U} \nabla\varphi(x)\cdot f(x,u,\mu_t)\,\mathrm{d}\theta_t(x,u)\\ &= \int_{\mathbb{R}^d} \nabla\varphi(x)\cdot \int_U f(x,u,\mu_t)\,\mathrm{d}\theta_{t,x}(u)\,\mathrm{d}\mu_t(x)\\ &= \int_{\mathbb{R}^d} \nabla\varphi(x)\cdot f\left(x, \int_U u\,\mathrm{d}\theta_{t,x}(u), \mu_t\right)\,\mathrm{d}\mu_t(x)\\ &= \int_{\mathbb{R}^d} \nabla\varphi(x)\cdot f(x,\underline{u}(t,x),\mu_t)\,\mathrm{d}\mu_t(x).\end{aligned}$$

Then $\mu$ satisfies the continuity equation $\partial_t\mu_t + \mathrm{div}(v_t\mu_t) = 0$ for the vector field $v_t(x) := f(x,\underline{u}(t,x),\mu_t)$ in the sense of distributions (see e.g. [3, equation (8.1.4)]). Since $|v_t(x)|^p \le \tilde{C}(1+|x|^p+\mathrm{m}_p^p(\mu_t))$, from (4.9) it follows that $t\mapsto \|v_t\|_{L^p_{\mu_t}(\mathbb{R}^d;\mathbb{R}^d)}$ belongs to $L^p(0,T)$ and $\mu \in \mathrm{AC}^p([0,T];\mathscr{P}_p(\mathbb{R}^d))$. Hence, $(\mu,\underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$.

Finally, by the convexity of $\mathcal{C}$ with respect to $u$ and Jensen's inequality we obtain

$$\begin{aligned}
&\int_0^T \int_\Omega \mathcal{C}(X_t(\omega), u_t(\omega), \mu_t)\, \mathrm{d}\mathbb{P}(\omega)\, \mathrm{d}t \\
&= \int_0^T \int_{\Omega\times U} \mathcal{C}(X_t(\omega), u, \mu_t)\, \mathrm{d}\sigma_t(\omega, u)\, \mathrm{d}t \\
&= \int_0^T \int_{\mathbb{R}^d\times U} \mathcal{C}(x, u, \mu_t)\, \mathrm{d}\theta_t(x, u)\, \mathrm{d}t \\
&= \int_0^T \int_{\mathbb{R}^d} \int_U \mathcal{C}(x, u, \mu_t)\, \mathrm{d}\theta_{t,x}(u)\, \mathrm{d}\mu_t(x)\, \mathrm{d}t \\
&\geq \int_0^T \int_{\mathbb{R}^d} \mathcal{C}\left(x, \int_U u\, \mathrm{d}\theta_{t,x}(u), \mu_t\right) \mathrm{d}\mu_t(x)\mathrm{d}t \\
&= \int_0^T \int_{\mathbb{R}^d} \mathcal{C}\left(x, \underline{u}(t,x), \mu_t\right) \mathrm{d}\mu_t(x)\, \mathrm{d}t.
\end{aligned}$$

This readily implies that $J_{\mathbf{L}}(X, u) \geq J_{\mathbf{E}}(\mu, \underline{u})$. □

In the next Lemma, we are given an admissible pair $(\boldsymbol{\eta}, u)$ for the Kantorovich problem. Considering the evaluation map $Z_t(\gamma) = \gamma(t)$, we associate to $(\boldsymbol{\eta}, u)$ the pair $(Z, u)$ which is admissible for the Lagrangian problem with parametrization space $(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta})$ and with the same cost as $(\boldsymbol{\eta}, u)$.

**Lemma 8.4.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Denote with $Z(t, \gamma) := e_t(\gamma)$, for every $t \in [0, T]$ and $\gamma \in \Gamma_T$. If $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ and we denote with $\mathbf{L}_{\boldsymbol{\eta}} = \mathbf{L}(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta})$, then $(Z, u) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}}}(e_0)$. Moreover,*

$$J_{\mathbf{L}_{\boldsymbol{\eta}}}(Z, u) = J_{\mathbf{K}}(\boldsymbol{\eta}, u).$$

*Proof.* Let $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ and denote by $\mathbf{L}_{\boldsymbol{\eta}} = \mathbf{L}(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta})$. Denoting with $Z : [0, T] \times \Gamma_T \to \mathbb{R}^d$ the map defined by $Z(t, \gamma) := \gamma(t) = e_t(\gamma)$, let us show that $(Z, u) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}}}(e_0)$. Since $Z(0, \gamma)_\sharp \boldsymbol{\eta} = \mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ then $e_0 = Z(0, \cdot) \in L^p_{\boldsymbol{\eta}}(\Gamma_T)$. By item (i) in Definition 7.1 we have $u \in \mathrm{B}([0, T] \times \Gamma_T; U)$. Thanks to (7.1) it readily follows that $Z \in L^p_{\boldsymbol{\eta}}(\Gamma_T; L^p(0, T; \mathbb{R}^d))$ and from Remark 7.2 we actually have that $Z \in L^p_{\boldsymbol{\eta}}(\Gamma_T; \mathrm{AC}^p([0, T]; \mathbb{R}^d))$. Moreover, from item (ii) of Definition 7.1, for $\boldsymbol{\eta}$-a.e. $\gamma \in \Gamma_T$, we have

$$\dot{Z}(t, \gamma) = f(Z(t, \gamma), u(t, \gamma), Z(t, \cdot)_\sharp \boldsymbol{\eta}) \quad \text{for } \mathcal{L}_T\text{-a.e.} t \in [0, T].$$

Hence $(Z, u) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}}}(e_0)$ and, by definition of $Z$, $J_{\mathbf{L}_{\boldsymbol{\eta}}}(Z, u) = J_{\mathbf{K}}(\boldsymbol{\eta}, u)$.

□

When the parametrization space $(\Omega, \mathfrak{B}, \mathbb{P})$ is fixed a priori, an interesting first comparison between the Kantorovich and Lagrangian problems is given below.

**Theorem 8.5.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 with $U$ convex compact subset of a separable Banach space $V$. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms. If $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ and $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$, then for every $X_0 \in L^p(\Omega; \mathbb{R}^d)$ with $(X_0)_\sharp \mathbb{P} = \mu_0$ there exits a sequence $(X^n, u^n) \in \mathcal{A}_{\mathbf{L}}(X_0)$ such that*

$$\lim_{n\to+\infty} J_{\mathbf{L}}(X^n, u^n) = J_{\mathbf{K}}(\boldsymbol{\eta}, u). \tag{8.3}$$

*Moreover, for every $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ and every $X_0 \in L^p(\Omega; \mathbb{R}^d)$ with $(X_0)_\sharp \mathbb{P} = \mu_0$ it holds*

$$V_{\mathbf{K}}(\mu_0) \geq V_{\mathbf{L}}(X_0). \tag{8.4}$$

*Proof.* **Step 1.** Let $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$. We denote by $\mathbf{L}_{\boldsymbol{\eta}} := \mathbf{L}(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta})$. Defining $Z(t, \gamma) := \gamma(t) = e_t(\gamma)$, by Lemma 8.4 it holds that $(Z, u) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}}}(e_0)$ and

$$J_{\mathbf{L}_{\boldsymbol{\eta}}}(Z, u) = J_{\mathbf{K}}(\boldsymbol{\eta}, u). \tag{8.5}$$

Thanks to the continuity of the evaluation map $e_0 : \Gamma_T \to \mathbb{R}^d$, we apply Proposition 4.22 for the problem $\mathbf{L}_{\boldsymbol{\eta}}$ in the Polish space $(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta})$. Then there exists a sequence $(\bar{Z}^m, \bar{u}^m) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}}}(e_0)$ such that $\bar{Z}^m : [0,T] \times \Gamma_T \to \mathbb{R}^d$ and $\bar{u}^m : [0,T] \times \Gamma_T \to U$ are continuous and

$$\lim_{m\to+\infty} J_{\mathbf{L}_{\boldsymbol{\eta}}}(\bar{Z}^m, \bar{u}^m) = J_{\mathbf{L}_{\boldsymbol{\eta}}}(Z, u). \tag{8.6}$$

**Step 2.** Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms and $\tau$ be a Polish topology on $\Omega$ such that $\mathfrak{B} = \mathcal{B}_{(\Omega,\tau)}$. Denote with $\mathbf{L} = \mathbf{L}(\Omega, \mathcal{B}_{(\Omega,\tau)}, \mathbb{P})$ the Lagrangian problem for the system $\mathbb{S}$. Let $X_0 \in L^p(\Omega; \mathbb{R}^d)$ with $(X_0)_\sharp \mathbb{P} = \mu_0$. Given $\bar{Z} \in C([0,T] \times \Gamma_T; \mathbb{R}^d)$ and $\bar{u} \in C([0,T] \times \Gamma_T; U)$ such that $(\bar{Z}, \bar{u}) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}}}(e_0)$, let us prove that there exists a sequence $(\tilde{X}^n, \tilde{u}^n) \in \mathcal{A}_{\mathbf{L}}(X_0)$ such that

$$\lim_{n\to+\infty} J_{\mathbf{L}}(\tilde{X}^n, \tilde{u}^n) = J_{\mathbf{L}_{\boldsymbol{\eta}}}(\bar{Z}, \bar{u}). \tag{8.7}$$

We define the sets

- $\tilde{\Gamma} := \left\{(x, \gamma) \in \mathbb{R}^d \times \Gamma_T \,:\, x = \gamma(0) = e_0(\gamma)\right\}$;
- $\Gamma_0 := \{\gamma \in \Gamma_T \,:\, \gamma(0) = e_0(\gamma) = 0\}$,

and the continuous maps

- $r : \tilde{\Gamma} \to \mathbb{R}^d \times \Gamma_0$, $r(x, \gamma) := (x, \gamma - x)$, where $\gamma - x$ is the curve $t \mapsto \gamma(t) - x$. Notice that $r$ admits a left inverse $r^{-1} : \mathbb{R}^d \times \Gamma_0 \to \tilde{\Gamma}$, $r^{-1}(x, \gamma_0) = (x, \gamma_0 + x)$, that obviously satisfies $r^{-1} \circ r = i_{\tilde{\Gamma}}$;
- $s : \Omega \times \mathbb{R}^d \times \Gamma_0 \to \Omega \times \tilde{\Gamma}$, $s(\omega, x, \gamma_0) = (\omega, x, \gamma_0 + x)$. Observe that $s = (i_\Omega, r^{-1})$.

Let us consider the couplings

$$\rho := (i_\Omega, X_0)_\sharp \mathbb{P} \in \mathscr{P}(\Omega \times \mathbb{R}^d), \quad \tilde{\eta} := (e_0, i_{\Gamma_T})_\sharp \boldsymbol{\eta} \in \mathscr{P}(\tilde{\Gamma}), \quad \hat{\eta} := r_\sharp \tilde{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_0).$$

Notice that $\pi^2_\sharp \rho = \pi^1_\sharp \tilde{\eta} = \pi^1_\sharp \hat{\eta} = \mu_0$.

We define a measure $\hat{\sigma} \in \mathscr{P}(\Omega \times \mathbb{R}^d \times \Gamma_0)$ satisfying $\pi^{1,2}_\sharp \hat{\sigma} = \rho$ and $\pi^{2,3}_\sharp \hat{\sigma} = \hat{\eta}$. Since $\mathbb{P}$ is without atoms and $\pi^1_\sharp \rho = \mathbb{P}$ then also $\rho \in \mathscr{P}(\Omega \times \mathbb{R}^d)$ is without atoms. Applying Lemma 2.10 with $\mathbb{T} = \Omega \times \mathbb{R}^d$, $S = \Gamma_0$, $\lambda = \rho$ and $\nu = \hat{\sigma}$, there exists a sequence of Borel maps $\hat{w}_n : \Omega \times \mathbb{R}^d \to \Gamma_0$ such that

$$\hat{\sigma}_n := (i_{\Omega \times \mathbb{R}^d}, \hat{w}_n)_\sharp \rho \xrightarrow{\mathsf{y}} \hat{\sigma}, \qquad \text{as } n \to +\infty. \tag{8.8}$$

Define $w_n : \Omega \times \mathbb{R}^d \to \Gamma_T$ by $w_n(\omega, x) := \hat{w}_n(\omega, x) + x$ and note that $s(\omega, x, \hat{w}_n(\omega, x)) = (\omega, x, w_n(\omega, x))$. Thanks to the continuity of $s$, then $s_\sharp$ is weakly continuous. From Remark 2.9, by the composition rule (2.1) and (8.8), we have that

$$\tilde{\sigma}_n := s_\sharp \hat{\sigma}_n = (i_{\Omega \times \mathbb{R}^d}, w_n)_\sharp \rho \xrightarrow{\mathsf{y}} s_\sharp \hat{\sigma} =: \tilde{\sigma} \in \mathscr{P}(\Omega \times \mathbb{R}^d \times \Gamma_T), \qquad \text{as } n \to +\infty. \tag{8.9}$$

From (2.1), a direct computation shows

$$\begin{aligned} &\pi^{1,2}_\sharp \tilde{\sigma} = \rho, \\ &\tilde{\eta} = r^{-1}_\sharp \hat{\eta} = \left(r^{-1} \circ \pi^{2,3}\right)_\sharp \hat{\sigma} = \left(r^{-1} \circ \pi^{2,3} \circ s^{-1}\right)_\sharp \tilde{\sigma} = \pi^{2,3}_\sharp \tilde{\sigma}. \end{aligned}$$

We define $\boldsymbol{\eta}^n := \pi^3_\sharp \tilde{\sigma}_n \in \mathscr{P}(\Gamma_T)$. Observing that $\pi^3_\sharp \tilde{\sigma} = \boldsymbol{\eta}$, by (8.9) we have $\boldsymbol{\eta}^n \to \boldsymbol{\eta}$ weakly in $\mathscr{P}(\Gamma_T)$. Notice also that $(e_0)_\sharp \boldsymbol{\eta}^n = (e_0 \circ w_n \circ (i_\Omega, X_0))_\sharp \mathbb{P} = (X_0)_\sharp \mathbb{P} = \mu_0$ and $(e_0)_\sharp \boldsymbol{\eta} = \mu_0$. For every $n \in \mathbb{N}$, denote by $\mathbf{L}_{\boldsymbol{\eta}^n} := \mathbf{L}(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta}^n)$ the Lagrangian problem for the system $\mathbb{S}$. Since $e_0$, $\bar{u}$ are continuous, we can apply Proposition 4.21 in the probability space $(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta})$, with $\boldsymbol{\eta}^n, \boldsymbol{\eta} \in \mathscr{P}(\Gamma_T)$ and initial datum $e_0$. Thus if $(Z^n, \bar{u}) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}^n}}(e_0)$, we have that

$$\lim_{n\to+\infty} J_{\mathbf{L}_{\boldsymbol{\eta}^n}}(Z^n, \bar{u}) = J_{\mathbf{L}_{\boldsymbol{\eta}}}(\bar{Z}, \bar{u}). \tag{8.10}$$

Finally, for any $n \in \mathbb{N}$, we define the pair $(\tilde{X}^n, \tilde{u}^n)$ by

$$\begin{aligned} \tilde{X}^n &: [0,T] \times \Omega \to \mathbb{R}^d, \qquad \tilde{X}^n(t, \omega) := Z^n(t, w_n(\omega, X_0(\omega))); \\ \tilde{u}^n &: [0,T] \times \Omega \to U, \qquad \tilde{u}^n(t, \omega) := \bar{u}(t, w_n(\omega, X_0(\omega))). \end{aligned}$$

Observe that $\tilde{X}^n(0,\omega) = Z^n(0, w_n(\omega, X_0(\omega))) = e_0(w_n(\omega, X_0(\omega))) = X_0(\omega)$. Moreover, thanks to the composition rule (2.1) we have $\pi^3_\sharp \tilde{\sigma}_n = (w_n)_\sharp \rho$ so that

$$(\tilde{X}^n_t)_\sharp \mathbb{P} = (Z^n_t \circ w_n \circ (i_\Omega, X_0))_\sharp \mathbb{P} = (Z^n_t)_\sharp \boldsymbol{\eta}^n.$$

By construction we have $(\tilde{X}^n, \tilde{u}^n) \in \mathcal{A}_{\mathbf{L}}(X_0)$ and it is immediate to verify that

$$J_{\mathbf{L}}(\tilde{X}^n, \tilde{u}^n) = J_{\mathbf{L}_{\boldsymbol{\eta}^n}}(Z^n, \bar{u}).$$

Then, by (8.10) we obtain (8.7).

**Step 3.** We apply Step 2 to the sequence $(\bar{Z}^m, \bar{u}^m)$ constructed in Step 1. Fix $m \in \mathbb{N}$, then there exists a sequence $(\tilde{X}^{m,n}, \tilde{u}^{m,n})_{n\in\mathbb{N}}$ such that $(\tilde{X}^{m,n}, \tilde{u}^{m,n}) \in \mathcal{A}_{\mathbf{L}}(X_0)$ for every $n \in \mathbb{N}$ and

$$\lim_{n\to+\infty} J_{\mathbf{L}}(\tilde{X}^{m,n}, \tilde{u}^{m,n}) = J_{\mathbf{L}_{\boldsymbol{\eta}}}(\bar{Z}^m, \bar{u}^m), \qquad \forall\, m \in \mathbb{N}.$$

Thanks to (8.6), by a simple diagonal argument we can select a (not relabelled) sequence $(X^n, u^n) \in \mathcal{A}_{\mathbf{L}}(X_0)$ satisfying

$$\lim_{n\to+\infty} J_{\mathbf{L}}(X^n, u^n) = J_{\mathbf{L}_{\boldsymbol{\eta}}}(Z, u),$$

where $(Z, u)$ are defined in Step 1. From (8.5) we finally get (8.3).

**Step 4.** By (8.3) and the definition of $V_{\mathbf{L}}$, for any $\varepsilon > 0$ there exists $n_\varepsilon > 0$ such that for $n \geq n_\varepsilon$

$$V_{\mathbf{L}}(X_0) \leq J_{\mathbf{L}}(X^n, u^n) \leq J_{\mathbf{K}}(\boldsymbol{\eta}, u) + \varepsilon.$$

From the arbitrariness of $\varepsilon > 0$ we have

$$V_{\mathbf{L}}(X_0) \leq J_{\mathbf{K}}(\boldsymbol{\eta}, u), \quad \forall\, (\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0),$$

hence the required inequality. □

*Remark* 8.6. Under the Convexity Assumption 3.4, if $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ is an optimal pair, then $(Z, u)$ given by Lemma 8.4 is optimal for the Lagrangian problem $\mathbf{L}_{\boldsymbol{\eta}}$. This is a consequence of Theorem 7.3, Proposition 8.3 and of (8.4) in Theorem 8.5.

We conclude the section with the proof of Theorem 8.2.

*Proof of Theorem 8.2.* Thanks to Proposition 3.6 and Remark 5.2 the relaxed Lagrangian problem **RL** in $\mathbb{S}$ coincides with $\mathbf{L}'$ in $\mathbb{S}'$. Precisely, for every $X_0 \in L^p(\Omega; \mathbb{R}^d)$ it holds $\mathcal{A}_{\mathbf{RL}}(X_0) = \mathcal{A}_{\mathbf{L}'}(X_0)$. Moreover, $J_{\mathbf{RL}}(X, \sigma) = J_{\mathbf{L}'}(X, \sigma)$ for every $(X, \sigma) \in \mathcal{A}_{\mathbf{RL}}(X_0) = \mathcal{A}_{\mathbf{L}'}(X_0)$. Hence, $V_{\mathbf{RL}}(X_0) = V_{\mathbf{L}'}(X_0)$. The equality $V_{\mathbf{L}}(X_0) = V_{\mathbf{RL}}(X_0)$ follows from Theorem 5.3. Finally, Theorem 8.1 yields $V_{\mathbf{L}'}(X_0) = V_{\mathbf{E}'}((X_0)_\sharp \mathbb{P}) = V_{\mathbf{K}'}((X_0)_\sharp \mathbb{P})$. □

8.2. **Continuity of $V_{\mathbf{E}}$, $V_{\mathbf{K}}$ and $V_{\mathbf{L}}$.** Here, we prove continuity results for the value functions of the various proposed formulations.

**Theorem 8.7** (Continuity of $V_{\mathbf{E}}$ and $V_{\mathbf{K}}$)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. If $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ and $\{\mu_0^n\}_{n\in\mathbb{N}} \subset \mathscr{P}_p(\mathbb{R}^d)$ is a sequence such that $W_p(\mu_0^n, \mu_0) \to 0$ as $n \to +\infty$, then*

$$\lim_{n\to+\infty} V_{\mathbf{E}}(\mu_0^n) = V_{\mathbf{E}}(\mu_0), \qquad \lim_{n\to+\infty} V_{\mathbf{K}}(\mu_0^n) = V_{\mathbf{K}}(\mu_0).$$

*Proof.* Let $\mu_0^n$ converge to $\mu_0$ in $\mathscr{P}_p(\mathbb{R}^d)$. By Proposition 2.1 with $S = \mathbb{R}^d$, there exist $X_0, X_0^n \in \mathrm{B}([0,1]; \mathbb{R}^d)$ such that $(X_0)_\sharp \mathcal{L}_1 = \mu_0$, $(X_0^n)_\sharp \mathcal{L}_1 = \mu_0^n$ and $X_0^n(\omega) \to X_0(\omega)$ for $\mathcal{L}_1$-a.e. $\omega \in [0,1]$. Since $\mu_0^n, \mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ we have $X_0^n, X_0 \in L^p([0,1]; \mathbb{R}^d)$. Moreover by the convergence $W_p(\mu_0^n, \mu_0) \to 0$ and Proposition 2.4 there exists $\psi : [0, +\infty) \to [0, +\infty)$ admissible (according to Definition 2.3) such that

$$\sup_{n\in\mathbb{N}} \int_{[0,1]} \psi(|X_0^n(t)|^p)\, \mathrm{d}\mathcal{L}_1(t) = \sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} \psi(|x|^p)\, \mathrm{d}\mu_0^n(x) < +\infty. \tag{8.11}$$

Thanks to Vitali theorem we get

$$\|X_0^n - X_0\|_{L^p([0,1];\mathbb{R}^d)} \to 0 \quad \text{as } n \to +\infty.$$

Applying Proposition 4.12 to the Lagrangian problem in $\mathbb{S}$ with $(\Omega, \mathcal{B}, \mathbb{P}) = ([0,1], \mathcal{B}_{[0,1]}, \mathcal{L}_1)$ we get $\limsup_{n\to+\infty} V_{\mathbf{L}}(X_0^n) \le V_{\mathbf{L}}(X_0)$. Theorem 8.1 yields

$$\limsup_{n\to+\infty} V_{\mathbf{E}}(\mu_0^n) = \limsup_{n\to+\infty} V_{\mathbf{L}}(X_0^n) \le V_{\mathbf{L}}(X_0) = V_{\mathbf{E}}(\mu_0).$$

By Proposition 6.9 we get $\lim_{n\to+\infty} V_{\mathbf{E}}(\mu_0^n) = V_{\mathbf{E}}(\mu_0)$. Finally, the continuity of $V_{\mathbf{K}}$ follows by Theorem 7.3 and the continuity of $V_{\mathbf{E}}$. □

**Theorem 8.8** (Continuity of $V_{\mathbf{L}}$)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms. If $X_0 \in L^p(\Omega; \mathbb{R}^d)$ and $\{X_0^n\}_{n\in\mathbb{N}} \subset L^p(\Omega; \mathbb{R}^d)$ is a sequence such that $\|X_0^n - X_0\|_{L^p(\Omega;\mathbb{R}^d)} \to 0$ as $n \to +\infty$, then*

$$\lim_{n\to+\infty} V_{\mathbf{L}}(X_0^n) = V_{\mathbf{L}}(X_0).$$

*Proof.* From Theorem 8.2 we have $V_{\mathbf{L}}(X_0^n) = V_{\mathbf{E}'}((X_0^n)_\sharp\mathbb{P})$ and $V_{\mathbf{L}}(X_0) = V_{\mathbf{E}'}((X_0)_\sharp\mathbb{P})$. The application of Theorem 8.7 to $\mathbf{E}'$ in $\mathbb{S}' = (\mathscr{U}, \mathscr{F}, \mathscr{C}, \mathscr{C}_T)$ (see Definition 3.5) concludes the proof. □

8.3. **A counterexample: Non-existence of minimizers for L.** In the previous sections we have shown that, under the Convexity Assumption 3.4, the Eulerian and Kantorovich problems always admit a minimizer, see Theorems 6.8, 7.3. This is not always true in the Lagrangian setting. Existence of minimizers has been shown in Remark 8.6 in the very particular case $\mathbf{L} = \mathbf{L}_{\boldsymbol{\eta}}$, where $\boldsymbol{\eta}$ is optimal for a Kantorovich problem. In general, for a given parametrization space $(\Omega, \mathfrak{B}, \mathbb{P})$, the choice of the initial condition is relevant as highlighted in the following.

**Theorem 8.9.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4 and $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms. If $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ then there exists $X_0 \in L^p(\Omega; \mathbb{R}^d)$ with $(X_0)_\sharp\mathbb{P} = \mu_0$ and $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ such that*

$$J_{\mathbf{L}}(X, u) = V_{\mathbf{L}}(X_0) = V_{\mathbf{E}}(\mu_0). \tag{8.12}$$

*Proof.* Let $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, by Theorem 7.3 there exists $(\boldsymbol{\eta}, \bar{u}) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ such that $J_{\mathbf{K}}(\boldsymbol{\eta}, \bar{u}) = V_{\mathbf{K}}(\mu_0)$. Fix $\tau$ a Polish topology on $\Omega$ such that $\mathfrak{B} = \mathcal{B}_{(\Omega,\tau)}$. Since $\mathbb{P}$ is without atoms, thanks to Proposition 2.1 there exists a Borel map $\psi : \Omega \to \Gamma_T$ such that $\psi_\sharp\mathbb{P} = \boldsymbol{\eta}$. For every $t \in [0, T]$ we define $X_t := e_t \circ \psi$ and the Borel map $u(t, \omega) := \bar{u}(t, \psi(\omega))$. Using the same techniques as in the proof of Lemma 8.4 we deduce that $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ (where $X_0 := e_0 \circ \psi$) and $J_{\mathbf{L}}(X, u) = J_{\mathbf{K}}(\boldsymbol{\eta}, \bar{u})$. By Theorems 8.1 and 7.3 we finally get (8.12). □

In general, if the initial condition $X_0$ is assigned a priori, existence of minimizers for the Lagrangian problem is not guaranteed. We consider the Wasserstein barycenter problem, for which we study the Eulerian and Lagrangian formulations. In particular, we exhibit an initial datum $X_0$ whose corresponding Lagrangian problem does not admit minimizers. We stress that the system under consideration satisfies the Convexity Assumption 3.4.

8.3.1. *Wasserstein barycenter problem: Eulerian formulation.* We consider the setting $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ as follows: let $U := \overline{B_R(0)} \subset \mathbb{R}^d$, for some $R > 0$ sufficiently large, $T = 1$ and $p = 2$. We fix $\nu \in \mathscr{P}(\mathbb{R}^d)$ with compact support. We consider the velocity field $f : \mathbb{R}^d \times U \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, the cost functions $\mathcal{C} : \mathbb{R}^d \times U \times \mathscr{P}_2(\mathbb{R}^d) \to [0, +\infty)$ and $\mathcal{C}_T : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to [0, +\infty)$ defined by

$$f(x, u, \mu) = u, \qquad \mathcal{C}(x, u, \mu) = |u|^2, \qquad \mathcal{C}_T(x, \mu) = W_2^2(\mu, \nu).$$

In this setting, the cost functional has the form

$$J_{\mathbf{E}}(\mu, \underline{u}) = \int_0^1 \int_{\mathbb{R}^d} |\underline{u}(t, x)|^2 \mathrm{d}\mu_t(x)\mathrm{d}t + W_2^2(\mu_1, \nu).$$

For any $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, the associated value function is given by

$$V_{\mathbf{E}}(\mu_0) := \inf_{(\mu,\underline{u})\in\mathcal{A}_{\mathbf{E}}(\mu_0)} J_{\mathbf{E}}(\mu, \underline{u}), \tag{8.13}$$

and recall that by Theorem 6.8 the infimum in (8.13) is actually a minimum.

Let us now fix $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$ with compact support and characterize the value function and the corresponding minimizers. By the Benamou-Brenier formula (2.7), we have the lower bound

(8.14)
$$\inf_{(\mu,\underline{u})\in\mathcal{A}_{\mathbf{E}}(\mu_0)} \left( \int_0^1 \int_{\mathbb{R}^d} |\underline{u}(t,x)|^2 \mathrm{d}\mu_t(x)\mathrm{d}t + W_2^2(\mu_1,\nu) \right) \geq \inf_{\mu_1\in\mathscr{P}_2(\mathbb{R}^d)} \left[ W_2^2(\mu_0,\mu_1) + W_2^2(\mu_1,\nu) \right].$$

Using the triangle inequality, it is easy to prove that

$$W_2^2(\mu_0,\mu_1) + W_2^2(\mu_1,\nu) \geq \frac{1}{2} W_2^2(\mu_0,\nu), \qquad \forall\, \mu_1 \in \mathscr{P}_2(\mathbb{R}^d), \tag{8.15}$$

and, for any constant speed Wasserstein geodesic $\{\sigma_t\}_{t\in[0,1]}$ such that $\sigma_0 = \mu_0$ and $\sigma_1 = \nu$, the measure $\mu_1 = \sigma_{1/2}$ realizes the equality in (8.15). Since the supports of $\mu_0$ and $\nu$ are compact, then the support of $\sigma_{1/2}$ is compact and, denoting by $\{\mu_t\}_{t\in[0,1]}$ a Wasserstein geodesic joining $\mu_0$ to $\sigma_{1/2}$, we also have that a vector field $\underline{u}$ realizing the equality

$$\int_0^1 \int_{\mathbb{R}^d} |\underline{u}(t,x)|^2 \mathrm{d}\mu_t(x)\mathrm{d}t = W_2^2(\mu_0,\sigma_{1/2}) \tag{8.16}$$

is bounded (see e.g. [51, Section 5.4]). Then, using $\underline{u}$ satisfying (8.16) and choosing $R$ sufficiently large, we obtain the equality in (8.14). The value of the minimum is

$$V_{\mathbf{E}}(\mu_0) = W_2^2(\mu_0,\sigma_{1/2}) + W_2^2(\sigma_{1/2},\nu) = \frac{1}{2} W_2^2(\mu_0,\nu). \tag{8.17}$$

Notice that the minimizer $(\mu,\underline{u})$ is not unique a priori. If at least one of the measures $\nu$ and $\mu_0$ is absolutely continuous with respect to $\mathcal{L}^d$, then the geodesic $\{\sigma_t\}_{t\in[0,1]}$ is unique and the map $\eta \mapsto W_2^2(\eta,\nu)$ is strictly convex. In this case $\sigma_{1/2}$ is the unique minimizer of the functional $\eta \mapsto W_2^2(\mu_0,\eta) + W_2^2(\eta,\nu)$ and the pair $(\mu,\underline{u})$, with $\mu_t := \sigma_{t/2}$ for all $t \in [0,1]$, is the (unique) minimizer for the Eulerian problem.

8.3.2. *Wasserstein barycenter problem: Lagrangian formulation.* Let $(\Omega,\mathfrak{B},\mathbb{P})$ be a standard Borel space such that $\mathbb{P} \in \mathscr{P}(\Omega)$ is without atoms. The Lagrangian cost functional of the Wasserstein barycenter problem is given by

$$J_{\mathbf{L}}(X,u) := \int_0^1 \int_\Omega |u_t(\omega)|^2 \mathrm{d}\mathbb{P}(\omega)\mathrm{d}t + W_2^2((X_1)_\sharp\mathbb{P},\nu). \tag{8.18}$$

For any $X_0 \in L^2(\Omega;\mathbb{R}^d)$, the corresponding value function is

$$V_{\mathbf{L}}(X_0) = \inf_{(X,u)\in\mathcal{A}_{\mathbf{L}}(X_0)} J_{\mathbf{L}}(X,u).$$

Since $(X,u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ satisfies, for $\mathbb{P}$-a.e. $\omega \in \Omega$, the system

$$\begin{cases} \dot{X}_t(\omega) = u_t(\omega), & \text{for } \mathcal{L}_1\text{-a.e. } t \in (0,1) \\ X|_{t=0}(\omega) = X_0(\omega), \end{cases}$$

we have

$$\begin{aligned} \int_0^1 \int_\Omega |u_t(\omega)|^2 \mathrm{d}\mathbb{P}(\omega)\mathrm{d}t &\geq \int_\Omega \left| \int_0^1 u_t(\omega)\mathrm{d}t \right|^2 \mathrm{d}\mathbb{P}(\omega) \\ &= \int_\Omega \left| \int_0^1 \dot{X}_t(\omega)\mathrm{d}t \right|^2 \mathrm{d}\mathbb{P}(\omega) = \int_\Omega |X_1(\omega) - X_0(\omega)|^2 \,\mathrm{d}\mathbb{P}(\omega) \end{aligned} \tag{8.19}$$

where we have applied Fubini theorem and Jensen's inequality. Notice that the inequality in (8.19) becomes an equality if $(X,u)$ belongs to the restrict admissibility class given by

$$\bar{\mathcal{A}}_{\mathbf{L}}(X_0) := \{(X,u) \in \mathcal{A}_{\mathbf{L}}(X_0) \,:\, u_t(\omega) = \bar{u}(\omega),\ \forall t \in [0,1],\ \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega\}.$$

Suppose now that $\mu_0 := (X_0)_\sharp \mathbb{P}$ has compact support (i.e. $X_0$ bounded), then we can compare the Lagrangian and Eulerian formulation of the Wasserstein barycenter problem. Indeed, for $R$ sufficiently large we have

$$\begin{aligned} V_{\mathbf{L}}(X_0) &= \inf_{(X,\bar{u})\in\bar{\mathcal{A}}_{\mathbf{L}}(X_0)} \left( \int_\Omega |X_1(\omega) - X_0(\omega)|^2 \mathrm{d}\mathbb{P}(\omega) + W_2^2((X_1)_\sharp\mathbb{P}, \nu) \right) \\ &= V_{\mathbf{E}}(\mu_0) = \min_{\mu_1\in\mathscr{P}_2(\mathbb{R}^d)} \left( W_2^2(\mu_0,\mu_1) + W_2^2(\mu_1,\nu) \right) = \frac{1}{2} W_2^2(\mu_0,\nu). \end{aligned} \tag{8.20}$$

where the first equality follows by the choice $(X,\bar{u}) \in \bar{\mathcal{A}}_{\mathbf{L}}(X_0)$, the second equality is given by Theorem 8.1 and the last two equalities are exactly (8.17).

We now exhibit an example where the infimum for the Lagrangian problem is not a minimum. Let us consider $\Omega = [0,1]$, $\mathbb{P} = \mathcal{L}_1$ and fix the dimension $d = 2$.
We set $\nu := \mathcal{L}^2 \llcorner [0,1]^2$, $X_0 : [0,1] \to \mathbb{R}^2$ defined by $X_0(\omega) = (1/2, \omega)$. We observe that $\mu_0 := (X_0)_\sharp\mathbb{P} = \mathscr{H}^1 \llcorner (\{1/2\} \times [0,1])$. We also notice that $X_0^{-1} : \mathbb{R}^2 \to [0,1]$, defined $\mu_0$-a.e. has the form $X_0^{-1}(1/2,\omega) = \omega$.
Since

$$\int_{[0,1]} |X_1(\omega) - X_0(\omega)|^2 \,\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}^2} \left| X_1(X_0^{-1}(x)) - x \right|^2 \mathrm{d}\mu_0(x),$$

and defining

$$\begin{aligned} \mathscr{B}_{\mu_0} &:= \{X_1 \circ X_0^{-1} : X_1 = X_0 + \bar{u} \text{ and } (X,\bar{u}) \in \bar{\mathcal{A}}_{\mathbf{L}}(X_0)\} \\ &= \{Y \in L^2((\mathbb{R}^2,\mu_0);\mathbb{R}^2) : |Y(x) - x| \le R, \text{ for } \mu_0\text{-a.e. } x \in \mathbb{R}^2\}, \end{aligned} \tag{8.21}$$

we easily get

$$\begin{aligned} V_{\mathbf{L}}(X_0) &= \inf_{(X,\bar{u})\in\bar{\mathcal{A}}_{\mathbf{L}}(X_0)} J_{\mathbf{L}}(X,\bar{u}) \\ &= \inf_{Y\in\mathscr{B}_{\mu_0}} \left[ \int_{\mathbb{R}^2} |Y(x) - x|^2 \,\mathrm{d}\mu_0(x) + W_2^2(Y_\sharp\mu_0, \nu) \right]. \end{aligned} \tag{8.22}$$

As already observed at the end of subsection 8.3.1, since $\nu$ is absolutely continuous with respect to $\mathcal{L}^2$, there exists a unique geodesic $\{\sigma_t\}_{t\in[0,1]}$ joining $\mu_0$ to $\nu$. Furthermore, $\sigma_{1/2}$ is the unique minimizer of the functional $\eta \mapsto W_2^2(\mu_0,\eta) + W_2^2(\eta,\nu)$. Then, for $R$ sufficiently large, from (8.20) and (8.22) we know that

$$\begin{aligned} W_2^2(\mu_0,\sigma_{1/2}) + W_2^2(\sigma_{1/2},\nu) &= \inf_{Y\in\mathscr{B}_{\mu_0}} \left[ \int_{\mathbb{R}^d} |Y(x) - x|^2 \,\mathrm{d}\mu_0(x) + W_2^2(Y_\sharp\mu_0,\nu) \right] \\ &\le \inf_{\substack{Y\in\mathscr{B}_{\mu_0} \\ Y_\sharp\mu_0 = \sigma_{1/2}}} \int_{\mathbb{R}^d} |Y(x) - x|^2 \,\mathrm{d}\mu_0(x) + W_2^2(\sigma_{1/2},\nu). \end{aligned} \tag{8.23}$$

On the other hand, since $\mathrm{supp}(\sigma_{1/2}) \subset [0,1]^2$, by [49, Theorem B] we have

$$W_2^2(\mu_0,\sigma_{1/2}) = \inf_{\substack{Y\in\mathscr{B}_{\mu_0} \\ Y_\sharp\mu_0 = \sigma_{1/2}}} \int_{\mathbb{R}^d} |Y(x) - x|^2 \,\mathrm{d}\mu_0(x) \tag{8.24}$$

and, consequently, equality holds in (8.23). Moreover the infimum in (8.24) is not attained. Indeed, the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x_1,x_2) = (1/2, x_2)$ satisfies $T_\sharp\nu = \mu_0$ and $T = \nabla\varphi$ for $\varphi(x_1,x_2) = \frac{1}{2}(x_1 + x_2^2)$, therefore $T$ is the optimal transport map from $\nu$ to $\mu_0$. The unique geodesic joining $\mu_0$ to $\nu$ is $\sigma_t = (t(x_1,x_2) + (1-t)(1/2,x_2))_\sharp\nu$ and $\sigma_{1/2} = (\frac{1}{2}x_1 + \frac{1}{4}, x_2))_\sharp\nu$ coincides with the uniform probability measure on $[1/4, 3/4] \times [0,1]$. The map $T$ is still the optimal transport map from $\sigma_{1/2}$ to $\mu_0$ and the unique optimal transport plan between $\sigma_{1/2}$ and $\mu_0$ is $((x_1,x_2),(1/2,x_2))_\sharp\sigma_{1/2}$. Then the unique optimal transport plan between $\mu_0$ and $\sigma_{1/2}$ is $\gamma := ((1/2,x_2),(x_1,x_2))_\sharp\sigma_{1/2}$. Since $\gamma$ is not concentrated on the graph of a map, the optimal transport map from $\mu_0$ to $\sigma_{1/2}$ does not exist.

Since (8.24) has not minimizers, then (8.22) cannot have minimizers. Indeed, suppose there exists a minimizer $(X,u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ for $J_{\mathbf{L}}$ in (8.22). Then $X_1 : [0,1] \to \mathbb{R}^2$ satisfies $(X_1)_\sharp \mathbb{P} = \sigma_{1/2}$, $u = X_1 - X_0$ and $X_t = tu + X_0$. Defining $Y = X_1 \circ X_0^{-1}$, we have that $Y_\sharp \mu_0 = \sigma_{1/2}$ so that $Y$ is a minimizer in (8.24), which is absurd.

*Remark* 8.10. Notice that existence of minimizers is not guaranteed even for Relaxed Lagrangian problems. Indeed, the same results obtained for the Wasserstein barycenter problem in $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ given in Section 8.3.1, can be easily extended to the lifted system $\mathbb{S}' = (\mathscr{U}, \mathscr{F}, \mathscr{C}, \mathscr{C}_T)$ associated to $\mathbb{S}$ (see Definition 3.5).

In the proposed example, the Lagrangian and Eulerian problems $\mathbf{L}', \mathbf{E}'$ associated to $\mathbb{S}'$ can be treated as the problems $\mathbf{L}, \mathbf{E}$ associated to $\mathbb{S}$ thanks to the following simple observation: given a probability measure $\rho \in \mathscr{P}(U)$, by Jensen's inequality we have

$$\int_U |u|^2 \, \mathrm{d}\rho(u) \geq \left| \int_U u \, \mathrm{d}\rho(u) \right|^2,$$

and the equality holds if and only if $\rho = \delta_{\underline{u}}$ for some $\underline{u} \in U$. This guarantees that possible control minimizers for $\mathbf{L}', \mathbf{E}'$ are of the form $\delta_{\underline{u}}$ with $\underline{u}$ *non-relaxed* control for $\mathbf{L}, \mathbf{E}$, respectively. The corresponding trajectories for $\mathbf{L}', \mathbf{E}'$ with control $\delta_{\underline{u}}$ coincide with the ones associated to $\underline{u}$ for problems $\mathbf{L}$ and $\mathbf{E}$, respectively. Finally, thanks to Remark 5.2, non-existence of minima for $\mathbf{L}'$ corresponds to non-existence of minima for $\mathbf{RL}$.

## 9. Finite particle systems and Gamma-convergence

To model the evolution of a finite number of particles, we introduce a discrete finite space $\Omega^N$ with the corresponding normalized counting measure $\mathbb{P}^N$. In this setting, in order to prove equivalence between Eulerian and Lagrangian problems, we cannot directly apply the results given in Theorems 8.1 due to the requirement on the probability measure $\mathbb{P}$ to be without atoms (see in particular Theorem 8.5). Hence, we introduce a further formulation of the Lagrangian problem in the context of feedback controls (see Definition 9.5) and we exploit a discrete formulation of the superposition principle for which we refer to Theorem C.1.

Furthermore, in Subsections 9.2 and 9.3, we prove a (discrete to continuous) $\Gamma$-convergence result respectively for the Lagrangian and Eulerian cost functionals when the number of particles goes to infinity.

### 9.1. Equivalences between $N$-particles problems.

Let $(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$ given by

$$\begin{aligned} &\Omega^N := \{1, \dots, N\}, \quad \mathcal{S}(\Omega^N) := \sigma(\{1\}, \dots \{N\})), \\ &\mathbb{P}^N(\{k\}) := \frac{1}{N}, \quad k = 1, \dots, N. \end{aligned} \tag{9.1}$$

We will refer to $\mathbb{P}^N$ as the normalized counting measure, which can be written as

$$\mathbb{P}^N = \frac{1}{N} \sum_{k=1}^N \delta_k.$$

Let us denote with $\mathbf{L}^N = \mathbf{L}(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$ the Lagrangian problem associated to the probability space $(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$. Notice that the functional space $L^p(\Omega^N; \mathbb{R}^d)$ coincides with the space of all maps $g : \Omega^N \to \mathbb{R}^d$, which can be identified with $(\mathbb{R}^d)^N$.

Differently from the Lagrangian problem $\mathbf{L}^N$, where we just need to fix the parametrization space, the definition of the $N$-particle Eulerian problem requires the introduction of a further constraint. Let us firstly define the subspace of $\mathscr{P}(\mathbb{R}^d)$ given by the discrete measures as

$$\mathscr{P}^N(\mathbb{R}^d) := \left\{ \mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \ \text{ for some } x_i \in \mathbb{R}^d \right\}. \tag{9.2}$$

**Definition 9.1** (Discrete Eulerian optimal control problem ($\mathbf{E}^N$))**.** Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Given $\mu_0 \in \mathscr{P}^N(\mathbb{R}^d)$, we say that $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}^N}(\mu_0)$, if

(i) $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$;
(ii) $\mu_t \in \mathscr{P}^N(\mathbb{R}^d)$, for every $t \in [0, T]$.

We define the *cost functional* $J_{\mathbf{E}^N} := J_{\mathbf{E}}$ and the *value function*

$$V_{\mathbf{E}^N}(\mu_0) := \inf\{J_{\mathbf{E}^N}(\mu, \underline{u}) : (\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}^N}(\mu_0)\}.$$

*Remark* 9.2. Notice that item (ii) in Definition 9.1 does not follow from the requirement $\mu_0 \in \mathscr{P}^N(\mathbb{R}^d)$. Indeed, the control map $\underline{u}$ in general is not Lipschitz continuous so that uniqueness of characteristics is not guaranteed.

Observe that, for every $N \in \mathbb{N}$, it holds

$$V_{\mathbf{E}^N}(\mu_0) \geq V_{\mathbf{E}}(\mu_0), \qquad \forall\, \mu_0 \in \mathscr{P}^N(\mathbb{R}^d). \tag{9.3}$$

The main result of this section is given in the following theorem.

**Theorem 9.3.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Let $p \geq 1$ and $X_0 \in L^p(\Omega^N; \mathbb{R}^d)$. Then*

$$V_{\mathbf{L}^N}(X_0) = V_{\mathbf{E}^N}((X_0)_\sharp \mathbb{P}^N).$$

The proof is a direct consequence of Propositions 9.4, 9.7 and 9.8 below.

Exploiting the argument contained in [34, Lemma 6.2] we derive a first comparison between $V_{\mathbf{L}^N}(X_0)$ and $V_{\mathbf{E}^N}((X_0)_\sharp \mathbb{P}^N)$.

**Proposition 9.4.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and let $X_0 \in L^p(\Omega^N; \mathbb{R}^d)$. If $(X, u) \in \mathcal{A}_{\mathbf{L}^N}(X_0)$, then there exists $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}^N}((X_0)_\sharp \mathbb{P}^N)$ such that $J_{\mathbf{L}^N}(X, u) \geq J_{\mathbf{E}^N}(\mu, \underline{u})$. Moreover, $V_{\mathbf{L}^N}(X_0) \geq V_{\mathbf{E}^N}((X_0)_\sharp \mathbb{P}^N)$.*

*Proof.* Let $(X, u) \in \mathcal{A}_{\mathbf{L}^N}(X_0)$. Let us define $\mathcal{X}(t) := \{x \in \mathbb{R}^d : X_t(\omega) = x \text{ for some } \omega \in \Omega^N\}$ and

$$J(t, x) := \{\omega \in \Omega^N : X_t(\omega) = x\}, \qquad \text{for any } (t, x) \in [0, T] \times \mathbb{R}^d, \tag{9.4}$$

and denote by $\mathcal{P}$ the collection of partitions $P$ of $\Omega^N$. It is clear that the family $P_X(t) := \{J(t, x) : x \in \mathcal{X}(t)\}$ belongs to $\mathcal{P}$. As proved in [34, Lemma 6.2], there exists a finite partition on Borel sets of the interval $[0, T]$ of the form $\{S_P : P \in \mathcal{P}\}$, where $S_P := \{t \in [0, T] : P_X(t) = P\}$.

Given $\omega, \omega' \in \Omega^N$ and a Borel set $S \subset [0, T]$, if $X_t(\omega) = X_t(\omega')$ for any $t \in S$ then, by the absolute continuity of the curves $t \mapsto X_t(\omega)$ and $t \mapsto X_t(\omega')$, we have $\dot{X}_t(\omega) = \dot{X}_t(\omega')$, for $\mathcal{L}_T$-a.e. $t \in S$.

We define $\mu_t := (X_t)_\sharp \mathbb{P}^N$ and we observe that

$$\mu_t = \frac{1}{N} \sum_{x \in \mathcal{X}(t)} \# J(t, x) \delta_x. \tag{9.5}$$

Moreover, defining for any $(t, x) \in [0, T] \times \mathbb{R}^d$

$$\bar{\omega}_{t,x} \in \operatorname*{argmin}_{\omega \in J(t,x)} \mathcal{C}(x, u(t, \omega), \mu_t),$$

we set

$$\underline{u}(t, x) := u(t, \bar{\omega}_{t,x}). \tag{9.6}$$

We show that $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}^N}((X_0)_\sharp \mathbb{P}^N)$. Let us fix $\varphi \in C^1_c(\mathbb{R}^d)$. For $\mathcal{L}_T$-a.e. $t \in [0,T]$ it holds

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}((X_t)_\sharp \mathbb{P}^N)(x) &= \int_{\Omega^N} \nabla\varphi(X_t(\omega)) \cdot f(X_t(\omega), u(t,\omega), \mu_t)\,\mathrm{d}\mathbb{P}^N(\omega)\\ &= \frac{1}{N}\sum_{\omega=1}^{N} \nabla\varphi(X_t(\omega)) \cdot f(X_t(\omega), u(t,\omega), \mu_t)\\ &= \frac{1}{N}\sum_{x\in\mathcal{X}(t)} \nabla\varphi(x) \cdot \sum_{\omega\in J(t,x)} f(x, u(t,\omega), \mu_t).\end{aligned} \tag{9.7}$$

Since $f(x, u(t,\omega), \mu_t) = f(x, \underline{u}(t,x), \mu_t)$ for every $\omega \in J(t,x)$, we can rewrite the right hand side of (9.7) to get

$$\begin{aligned}\frac{1}{N}\sum_{x\in\mathcal{X}(t)} \nabla\varphi(x) \cdot \sum_{\omega\in J(t,x)} f(x, u(t,\omega), \mu_t) &= \frac{1}{N}\sum_{x\in\mathcal{X}(t)} \#J(t,x)\,\nabla\varphi(x) \cdot f(x, \underline{u}(t,x), \mu_t)\\ &= \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot f(x, \underline{u}(t,x), \mu_t)\,\mathrm{d}\mu_t(x),\end{aligned}$$

where the last equality is a consequence of (9.5).

For what concerns the cost functional, we have

$$\begin{aligned}&\int_0^T \int_\Omega \mathcal{C}(X_t(\omega), u(t,\omega), \mu_t)\,\mathrm{d}\mathbb{P}^N(\omega)\,\mathrm{d}t\\ &= \frac{1}{N}\int_0^T \sum_{x\in\mathcal{X}(t)} \sum_{\omega\in J(t,x)} \mathcal{C}(x, u(t,\omega), \mu_t)\,\mathrm{d}t\\ &\geq \frac{1}{N}\int_0^T \sum_{x\in\mathcal{X}(t)} \#J(t,x)\mathcal{C}(x, \underline{u}(t,x), \mu_t)\,\mathrm{d}t\\ &= \int_0^T \int_{\mathbb{R}^d} \mathcal{C}(x, \underline{u}(t,x), \mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t,\end{aligned}$$

where the inequality comes from the definition of $\underline{u}$ in (9.6) and the last equality follows from (9.5). Since

$$\int_{\Omega^N} \mathcal{C}_T(X_T(\omega), \mu_T)\,\mathrm{d}\mathbb{P}^N(\omega) = \int_{\mathbb{R}^d} \mathcal{C}_T(x, \mu_T)\,\mathrm{d}\mu_T(x),$$

we conclude that $J_{\mathbf{L}^N}(X,u) \geq J_{\mathbf{E}^N}(\mu, \underline{u})$. □

Here we introduce a feedback formulation of the Lagrangian optimal control problem in order to prove the reverse inequality $V_{\mathbf{L}^N}(X_0) \leq V_{\mathbf{E}^N}((X_0)_\sharp \mathbb{P}^N)$. We firstly show its relation with the Lagrangian and Eulerian problems in a general context, i.e. where the probability space $(\Omega, \mathfrak{B}, \mathbb{P})$ is not necessarily the space $(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$ associated to the $N$-particles framework.

**Definition 9.5** (Feedback Lagrangian optimal control problem **(FL)**)**.** Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a probability space. Given $X_0 \in L^p(\Omega; \mathbb{R}^d)$, we say that $(X, \underline{u}) \in \mathcal{A}_{\mathbf{FL}}(X_0)$ if

(i) $\underline{u} \in \mathrm{B}([0,T] \times \mathbb{R}^d; U)$;
(ii) $X \in L^p(\Omega; \mathrm{AC}^p([0,T]; \mathbb{R}^d))$ and for $\mathbb{P}$-a.e. $\omega \in \Omega$, $X(\omega)$ is a solution of the following Cauchy problem

$$\begin{cases}\dot{X}_t(\omega) = f(X_t(\omega), \underline{u}_t(X_t(\omega)), (X_t)_\sharp \mathbb{P}), & \text{for } \mathcal{L}_T\text{-a.e. } t \in ]0,T]\\ X_{|t=0}(\omega) = X_0(\omega),\end{cases} \tag{9.8}$$

where $X_t : \Omega \to \mathbb{R}^d$ is defined by $X_t(\omega) := X(t,\omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$.

We refer to $(X, \underline{u}) \in \mathcal{A}_{\mathbf{FL}}(X_0)$ as to an *admissible pair*, with $X$ a *trajectory* and $\underline{u}$ a *feedback control*.
We define the cost functional $J_{\mathbf{FL}} : L^p(\Omega; C([0,T];\mathbb{R}^d)) \times \mathrm{B}([0,T]\times\mathbb{R}^d; U) \to [0,+\infty)$ by

$$J_{\mathbf{FL}}(X,\underline{u}) := \int_\Omega \int_0^T \mathcal{C}(X_t(\omega), \underline{u}_t(X_t(\omega)), (X_t)_\sharp\mathbb{P})\,\mathrm{d}t\,\mathrm{d}\mathbb{P}(\omega) + \int_\Omega \mathcal{C}_T(X_T(\omega), (X_T)_\sharp\mathbb{P})\,\mathrm{d}\mathbb{P}(\omega),$$

and the corresponding *value function* $V_{\mathbf{FL}} : L^p(\Omega;\mathbb{R}^d) \to [0,+\infty)$ by

$$V_{\mathbf{FL}}(X_0) := \inf\left\{J_{\mathbf{FL}}(X,\underline{u}) \,:\, (X,\underline{u}) \in \mathcal{A}_{\mathbf{FL}}(X_0)\right\}. \tag{9.9}$$

In the following, $\mathbf{FL}(\Omega,\mathfrak{B},\mathbb{P})$ denotes the Feedback Lagrangian problem given in Definition 9.5. We short the notation to $\mathbf{FL}$ when the probability space is clear from the context.

*Remark* 9.6. Observe that, choosing a constant (feedback) control function $\underline{u}$, from Proposition 4.8 it is immediate to prove that $\mathcal{A}_{\mathbf{FL}}(X_0) \neq \emptyset$. In general, given $\underline{u} \in \mathrm{B}([0,T]\times\mathbb{R}^d; U)$, the existence and uniqueness of solutions to the Cauchy problem (9.8) is not guaranteed.

The following result follows directly from Definitions 4.1, 6.1 and 9.5.

**Proposition 9.7.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 and $(\Omega,\mathfrak{B},\mathbb{P})$ be a probability space. Let $X_0 \in L^p(\Omega;\mathbb{R}^d)$. If $(X,\underline{u}) \in \mathcal{A}_{\mathbf{FL}}(X_0)$, then*

*(i) defining $u : [0,T]\times\Omega \to U$ by $u(t,\omega) := \underline{u}(t, X_t(\omega))$, we have that $(X,u) \in \mathcal{A}_{\mathbf{L}}(X_0)$ and $J_{\mathbf{FL}}(X,\underline{u}) = J_{\mathbf{L}}(X,u)$. In particular it holds $V_{\mathbf{FL}}(X_0) \geq V_{\mathbf{L}}(X_0)$.*

*(ii) defining $\mu_t := (X_t)_\sharp\mathbb{P}$ for any $t \in [0,T]$, we have that $(\mu,\underline{u}) \in \mathcal{A}_{\mathbf{E}}((X_0)_\sharp\mathbb{P})$ and $J_{\mathbf{FL}}(X,\underline{u}) = J_{\mathbf{E}}(\mu,\underline{u})$. In particular it holds $V_{\mathbf{FL}}(X_0) \geq V_{\mathbf{E}}(\mu_0)$.*

Taking advantage of Proposition 9.7 and of the discrete superposition principle given in Theorem C.1, we have the following equivalence result between $\mathbf{FL}^N := \mathbf{FL}(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$ and $\mathbf{E}^N$ defined in Definition 9.1.

**Proposition 9.8.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Let $p \geq 1$, $X_0 \in L^p(\Omega^N;\mathbb{R}^d)$ and $\mu_0 := \frac{1}{N}\sum_{\omega=1}^N \delta_{X_0(\omega)} = (X_0)_\sharp\mathbb{P}^N$. If $(\mu,\underline{u}) \in \mathcal{A}_{\mathbf{E}^N}(\mu_0)$, then there exists $X \in C([0,T]; L^p(\Omega^N))$ such that $\mu_t = (X_t)_\sharp\mathbb{P}^N$ for all $t \in [0,T]$, $(X,\underline{u}) \in \mathcal{A}_{\mathbf{FL}^N}(X_0)$ and $J_{\mathbf{E}^N}(\mu,\underline{u}) = J_{\mathbf{FL}^N}(X,\underline{u})$. Moreover, $V_{\mathbf{E}^N}(\mu_0) = V_{\mathbf{FL}^N}(X_0)$.*

*Proof.* Let $(\mu,\underline{u}) \in \mathcal{A}_{\mathbf{E}^N}(\mu_0)$. By the superposition principle given in Theorem C.1 applied to $\mu$ and the vector field $v(t,x) = f(x,\underline{u}(t,x),\mu_t)$, there exists $\eta \in \mathscr{P}(\Gamma_T)$ such that

$$\eta = \frac{1}{N}\sum_{\omega=1}^N \delta_{\gamma_\omega},$$

where $\gamma_\omega \in \Gamma_T$, $\omega = 1,\dots,N$. We define $X : [0,T]\times\Omega^N \to \mathbb{R}^d$ by $X(t,\omega) := \gamma_\omega(t)$ that satisfies $(X,u) \in \mathcal{A}_{\mathbf{FL}^N}$ thanks to (C.4) in Theorem C.1. Hence it readily follows that $J_{\mathbf{FL}^N}(X,\underline{u}) = J_{\mathbf{E}^N}(\mu,\underline{u})$ and $V_{\mathbf{FL}^N}(X_0) \leq V_{\mathbf{E}^N}(\mu_0)$. By Proposition 9.7 we have that $V_{\mathbf{FL}^N}(X_0) \geq V_{\mathbf{E}^N}(\mu_0)$ and we conclude the proof. □

9.2. **Finite particle approximation for L.** The aim of this section is to approximate a general Lagrangian problem $\mathbf{L} = \mathbf{L}(\Omega,\mathfrak{B},\mathbb{P})$ with finite particle Lagrangian problems $\mathbf{L}^N = \mathbf{L}(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$, $N \in \mathbb{N}$, where $(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$ is defined in (9.1). A first result in this direction has been already obtained in Proposition 4.18 (see also Remark 4.19) in Section 4.2. Here, we specialize the result in the case of equally distributed masses which is suitable for the application to a finite particle/agent model.

Recall that if $(\Omega,\mathfrak{B},\mathbb{P})$ is a standard Borel space and $\mathbb{P}$ is without atoms, thanks to item (ii) in Proposition 4.16 there exists a family of finite algebras $\mathfrak{B}^N \subset \mathfrak{B}$, $N \in \mathbb{N}$, satisfying the finite approximation property of Definition 4.15 and $\mathbb{P}(A_k^N) = \frac{1}{N}$, with $k = 1,\dots,N$. Recall the definition of $\psi^N$, $\phi^N$ and $\mathcal{K}^N$ in (4.22) and (4.23), respectively.

A Gamma-convergence result for the functional $J_{\mathbf{L}}$ is given in the following proposition.

**Proposition 9.9** (Finite particle approximation for $\mathbf{L}$)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms. The following holds:*

*(i) Suppose that $(X, u) \in L^p(\Omega; \mathrm{AC}^p([0,T]; \mathbb{R}^d)) \times \mathrm{M}([0,T] \times \Omega; U)$ and $(X^N, u^N) \in C([0,T]; L^p(\Omega^N; \mathbb{R}^d)) \times \mathrm{M}([0,T] \times \Omega^N; U)$ such that*

$$\lim_{N \to +\infty} \mathcal{K}^N(X^N, u^N) = (X, u), \qquad \text{in } C([0,T]; L^p(\Omega; \mathbb{R}^d)) \times L^1([0,T] \times \Omega; U).$$

*Then*

$$\lim_{N \to +\infty} J_{\mathbf{L}^N}(X^N, u^N) = J_{\mathbf{L}}(X, u).$$

*(ii) Assume that $U$ is a compact convex subset of a separable Banach space $V$. Suppose that $X_0 \in L^p(\Omega; \mathbb{R}^d)$ and $(X, u) \in \mathcal{A}_{\mathbf{L}}(X_0)$. If $X_0^N \in L^p(\Omega^N; \mathbb{R}^d)$ such that*

$$\lim_{N \to +\infty} \|X_0^N \circ \psi^N - X_0\|_{L^p(\Omega; \mathbb{R}^d)} = 0 \tag{9.10}$$

*then there exists a sequence $(X^N, u^N) \in \mathcal{A}_{\mathbf{L}^N}(X_0^N)$ such that*

$$\lim_{N \to +\infty} \mathcal{K}^N(X^N, u^N) = (X, u), \qquad \text{in } C([0,T]; L^p(\Omega; \mathbb{R}^d)) \times L^1([0,T] \times \Omega; U)$$

*and*

$$\lim_{N \to +\infty} J_{\mathbf{L}^N}(X^N, u^N) = J_{\mathbf{L}}(X, u).$$

*Proof.* Thanks to Proposition A.3 it holds that $X \in C([0,T]; L^p(\Omega; \mathbb{R}^d))$. Item (i) can be proved exactly by the same technique used in the second part of the proof of Proposition 4.11 applied to the sequence $\mathcal{K}^N(X^N, u^N)$ and recalling that $J_{\mathbf{L}}\left(\mathcal{K}^N(X^N, u^N)\right) = J_{\mathbf{L}^N}(X^N, u^N)$ (see Proposition 4.14). Notice that, since $U$ is metrizable and compact the convergence $u^N \to u$ in $L^1([0,T] \times \Omega; U)$ is equivalent to the convergence in $(\mathcal{L}_T \otimes \mathbb{P})$-measure. Item (ii), is a direct application of Proposition 4.18 to the sequence of finite algebras $\mathfrak{B}^N$ given in item (ii) of Proposition 4.16. □

**Proposition 9.10** (Convergence of the value functions)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy Assumption 3.1 with $U$ a compact convex subset of a separable Banach space $V$. Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space such that $\mathbb{P}$ is without atoms. If $X_0 \in L^p(\Omega; \mathbb{R}^d)$ and $X_0^N \in L^p(\Omega^N; \mathbb{R}^d)$ satisfy*

$$\lim_{N \to +\infty} \|X_0^N \circ \psi^N - X_0\|_{L^p(\Omega; \mathbb{R}^d)} = 0, \tag{9.11}$$

*then*

$$\limsup_{N \to +\infty} V_{\mathbf{L}^N}(X_0^N) \leq V_{\mathbf{L}}(X_0).$$

*Moreover, if $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfies the Convexity Assumption 3.4, then*

$$\liminf_{N \to +\infty} V_{\mathbf{L}^N}(X_0^N) \geq V_{\mathbf{L}}(X_0).$$

*In particular, under the Convexity Assumption 3.4,*

$$\lim_{N \to +\infty} V_{\mathbf{L}^N}(X_0^N) = V_{\mathbf{L}}(X_0).$$

*Proof.* By definition of inf, for every $\varepsilon > 0$ there exists $(X_\varepsilon, u_\varepsilon) \in \mathcal{A}_{\mathbf{L}}(X_0)$ such that $V_{\mathbf{L}}(X_0) \geq J_{\mathbf{L}}(X_\varepsilon, u_\varepsilon) - \varepsilon$. Moreover from item (ii) in Proposition 9.9 there exists $(X_\varepsilon^N, u_\varepsilon^N)$ such that $J_{\mathbf{L}^N}(X_\varepsilon^N, u_\varepsilon^N) \to J_{\mathbf{L}}(X_\varepsilon, u_\varepsilon)$, as $N \to +\infty$. Hence

$$\limsup_{N \to +\infty} V_{\mathbf{L}^N}(X_0^N) \leq \limsup_{N \to +\infty} J_{\mathbf{L}^N}(X_\varepsilon^N, u_\varepsilon^N) = J_{\mathbf{L}}(X_\varepsilon, u_\varepsilon) \leq V_{\mathbf{L}}(X_0) + \varepsilon.$$

By the arbitrariness of $\varepsilon$ we conclude.
By Proposition 9.4 and (9.3) we get

$$V_{\mathbf{L}^N}(X_0^N) \geq V_{\mathbf{E}^N}(\mu_0^N) \geq V_{\mathbf{E}}(\mu_0^N).$$

In the convex setting, by the lower semicontinuity of the value function $V_{\mathbf{E}}$ (see Proposition 6.9) and by Corollary 8.1 ($\mathbb{P}$ is without atoms by assumption) we have the desired convergence. □

*Remark* 9.11. A natural choice for $X_0^N$ in Proposition 9.10 is given by $X_0^N := \tilde{X}_0^N \circ \phi^N$ where

$$\tilde{X}_0^N := \sum_{k=1}^N \mathbb{1}_{A_k^N} \fint_{A_k^N} X_0(\omega)\, \mathrm{d}\mathbb{P}(\omega).$$

For a proof of the convergence (9.11) we refer to Lemma D.3 and Proposition D.1 in Appendix D.

9.3. **Finite particle approximation for E.** In this section, we show that the Eulerian problem $\mathbf{E}$ can be approximated by finite particle Eulerian problems $\mathbf{E}^N$ defined in Definition 9.1. Thanks to Theorem 9.3, we are able to approximate the Eulerian problem also with a sequence of finite particle Lagrangian problems $\mathbf{L}^N = \mathbf{L}(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$, $N \in \mathbb{N}$, with $(\Omega^N, \mathcal{S}(\Omega^N), \mathbb{P}^N)$ as in (9.1). This is relevant from the point of view of applications. The main result of the section is stated in the following theorem.

**Theorem 9.12** (Convergence of the value functions)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. Let $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ and $\mu_0^N \in \mathscr{P}^N(\mathbb{R}^d)$ such that $W_p(\mu_0^N, \mu_0) \to 0$ as $N \to +\infty$, then*

$$\lim_{N\to+\infty} V_{\mathbf{E}^N}(\mu_0^N) = V_{\mathbf{E}}(\mu_0).$$

*Moreover, for every $X_0^N \in L^p(\Omega^N; \mathbb{R}^d)$ such that $(X_0^N)_\sharp \mathbb{P}^N = \mu_0^N$ it holds that*

$$\lim_{N\to+\infty} V_{\mathbf{L}^N}(X_0^N) = V_{\mathbf{E}}(\mu_0).$$

In order to prove Theorem 9.12, we start with the following proposition.

**Proposition 9.13** (Finite particle approximation for $\mathbf{E}$)**.** *Let $\mathbb{S} = (U, f, \mathcal{C}, \mathcal{C}_T)$ satisfy the Convexity Assumption 3.4. Let $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$ and $\mu_0^N \in \mathscr{P}^N(\mathbb{R}^d)$, $N \in \mathbb{N}$, such that*

$$\lim_{N\to+\infty} W_p(\mu_0^N, \mu_0) = 0. \tag{9.12}$$

*If $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$, then there exists a sequence $(\mu^N, \underline{u}^N) \in \mathcal{A}_{\mathbf{E}^N}(\mu_0^N)$ such that*

*(1) $(\mu^N, \underline{u}^N)$ converges to $(\mu, \underline{u})$ according to Definition 6.5;*
*(2) $\lim_{N\to+\infty} J_{\mathbf{E}^N}(\mu^N, \underline{u}^N) = J_{\mathbf{E}}(\mu, \underline{u})$.*

*Proof.* **Step 1.** Let $(\mu, \underline{u}) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$. In this step we associate to $(\mu, \underline{u})$ an admissible pair $(\tilde{X}, \tilde{u}) \in \mathcal{A}_{\mathbf{L}}(\tilde{X}_0)$ for the Lagrangian problem $\mathbf{L} := \mathbf{L}([0,1]; \mathcal{B}_{[0,1]}, \mathcal{L}_1)$ such that

$$J_{\mathbf{L}}(\tilde{X}, \tilde{u}) = J_{\mathbf{E}}(\mu, \underline{u}). \tag{9.13}$$

By Proposition 7.4, there exists $(\boldsymbol{\eta}, u) \in \mathcal{A}_{\mathbf{K}}(\mu_0)$ such that $u(t, \gamma) = \underline{u}(t, \gamma(t))$, $(e_t)_\sharp \boldsymbol{\eta} = \mu_t$ and $J_{\mathbf{K}}(\boldsymbol{\eta}, u) = J_{\mathbf{E}}(\mu, \underline{u})$. Thanks to Lemma 8.4, the map $Z : [0,T] \times \Gamma_T \to \mathbb{R}^d$ defined by $Z(t, \gamma) := \gamma(t) = e_t(\gamma)$ satisfies $(Z, u) \in \mathcal{A}_{\mathbf{L}_{\boldsymbol{\eta}}}(e_0)$ with $\mathbf{L}_{\boldsymbol{\eta}} := \mathbf{L}(\Gamma_T, \mathcal{B}_{\Gamma_T}, \boldsymbol{\eta})$ and

$$J_{\mathbf{L}_{\boldsymbol{\eta}}}(Z, u) = J_{\mathbf{K}}(\boldsymbol{\eta}, u), \tag{9.14}$$

By Proposition 2.1 applied to $S = \Gamma_T$ and $\nu = \boldsymbol{\eta}$, there exists a Borel map $P : [0,1] \to \Gamma_T$ such that $P_\sharp \mathcal{L}_1 = \boldsymbol{\eta}$.

We define $\tilde{X} : [0,T] \times [0,1] \to \mathbb{R}^d$ by $\tilde{X}_t(\omega) := Z_t(P(\omega))$ and $\tilde{u} : [0,T] \times [0,1] \to U$ by $\tilde{u}_t(\omega) := u_t(P(\omega))$. Notice that

$$(\tilde{X}_t)_\sharp \mathcal{L}_1 = (e_t \circ P)_\sharp \mathcal{L}_1 = (e_t)_\sharp \boldsymbol{\eta} = \mu_t, \tag{9.15}$$

and it is easy to prove that $(\tilde{X}, \tilde{u}) \in \mathcal{A}_{\mathbf{L}}(\tilde{X}_0)$. Moreover $J_{\mathbf{L}}(\tilde{X}, \tilde{u}) = J_{\mathbf{L}_{\boldsymbol{\eta}}}(Z, u)$ and, by (9.14), we obtain (9.13).

**Step 2.** We use the partition of $[0,1]$ defined in Lemma D.3. We define the piecewise constant initial data $\tilde{X}_0^N : [0,1] \to \mathbb{R}^d$ and controls $\tilde{u}^N : [0,T] \times [0,1] \to U$ by

$$\tilde{X}_0^N := \sum_{k=1}^N \mathbb{1}_{I_k^N} \fint_{I_k^N} \tilde{X}_0(\omega)\,\mathrm{d}\mathcal{L}_1(\omega),$$
$$\tilde{u}_t^N := \sum_{k=1}^N \mathbb{1}_{I_k^N} \fint_{I_k^N} \tilde{u}_t(\omega)\,\mathrm{d}\mathcal{L}_1(\omega)\,, \quad \text{for all } t \in [0,T].$$

From the definition of $\tilde{X}_0^N$ and Lemma D.3 we have

$$W_p(\tilde{\mu}_0^N, \mu_0) \le \|\tilde{X}_0^N - \tilde{X}_0\|_{L^p([0,1];\mathbb{R}^d)} \to 0 \qquad \text{as } N \to +\infty, \tag{9.16}$$

where

$$\tilde{\mu}_0^N := (\tilde{X}_0^N)_\sharp \mathcal{L}_1 = \frac{1}{N}\sum_{k=1}^N \delta_{\tilde{x}_k^N}, \qquad \tilde{x}_k^N := \fint_{I_k^N} \tilde{X}_0(\omega)\,\mathrm{d}\mathcal{L}_1(\omega) \in \mathbb{R}^d.$$

Since $\mu_0^N = \frac{1}{N}\sum_{k=1}^N \delta_{x_k^N}$, there exists a permutation of indexes $\sigma^N : \{1,\dots,N\} \to \{1,\dots,N\}$ and a map $X_0^N : [0,1] \to \mathbb{R}^d$ defined by $X_0^N := \sum_{k=1}^N x_{\sigma^N(k)}^N \mathbb{1}_{I_k^N}$ such that

$$W_p(\mu_0^N, \tilde{\mu}_0^N) = \|X_0^N - \tilde{X}_0^N\|_{L^p([0,1];\mathbb{R}^d)}. \tag{9.17}$$

Using (9.12), (9.16) and (9.17) we obtain

$$\begin{aligned} \|X_0^N - \tilde{X}_0\|_{L^p([0,1];\mathbb{R}^d)} &\le \|X_0^N - \tilde{X}_0^N\|_{L^p([0,1];\mathbb{R}^d)} + \|\tilde{X}_0^N - \tilde{X}_0\|_{L^p([0,1];\mathbb{R}^d)} \\ &= W_p(\mu_0^N, \tilde{\mu}_0^N) + \|\tilde{X}_0^N - \tilde{X}_0\|_{L^p([0,1];\mathbb{R}^d)} \\ &\le W_p(\mu_0^N, \mu_0) + W_p(\mu_0, \tilde{\mu}_0^N) + \|\tilde{X}_0^N - \tilde{X}_0\|_{L^p([0,1];\mathbb{R}^d)} \to 0. \end{aligned} \tag{9.18}$$

Let $(X^N, \tilde{u}^N) \in \mathcal{A}_{\mathbf{L}}(X_0^N)$. By Lemma D.3 and dominated convergence we have that $\|\tilde{u}^N - \tilde{u}\|_{L^1([0,1]\times[0,T];V)} \to 0$ as $N \to +\infty$. Then, by Proposition 4.11, we have

$$\sup_{t\in[0,T]} \|X_t^N - \tilde{X}_t\|_{L^p([0,1];\mathbb{R}^d)} \to 0 \qquad \text{as } N \to +\infty \tag{9.19}$$

and

$$\lim_{N\to+\infty} J_{\mathbf{L}}(X^N, \tilde{u}^N) = J_{\mathbf{L}}(\tilde{X}, \tilde{u}). \tag{9.20}$$

We observe that, for any $t \in [0,T]$, $X_t^N$ is constant on the elements $I_k^N$ of partition $\{I_k^N : k = 1,\dots,N\}$ so that it is of the form

$$X_t^N := \sum_{k=1}^N (x_t^N)_k \mathbb{1}_{I_k^N}$$

for some $(x_t^N)_k \in \mathbb{R}^d$.

We define $\mu_t^N := (X_t^N)_\sharp \mathcal{L}_1 \in \mathscr{P}_p(\mathbb{R}^d)$. From the observation above,

$$\mu_t^N = \frac{1}{N}\sum_{k=1}^N \delta_{(x_t^N)_k}$$

By (9.19) we obtain that $\mu^N \to \mu$ in $C([0,T];\mathscr{P}_p(\mathbb{R}^d))$ as $N \to +\infty$.

**Step 3.** For any $t \in [0,T]$ we define $\rho_t^N := (X_t^N, \tilde{u}_t^N)_\sharp \mathcal{L}_1 \in \mathscr{P}(\mathbb{R}^d \times U)$, and notice that $\pi^1_\sharp \rho_t^N = \mu_t^N$. Denoting by $\rho_{t,x}^N \in \mathscr{P}(U)$ the disintegration of $\rho_t^N$ w.r.t. $\pi^1$, we define the Borel map $\underline{u}^N : [0,T] \times \mathbb{R}^d \to U$ by

$$\underline{u}^N(t,x) := \int_U u\,\mathrm{d}\rho_{t,x}^N(u).$$

By the definition of $\mu_t^N$ and $\underline{u}^N$, from item (2) of the Convexity Assumption 3.4, we obtain that $(\mu^N, \underline{u}^N) \in \mathcal{A}_{\mathbf{E}^N}(\mu_0^N)$. Indeed, given $\varphi \in C_c^1(\mathbb{R}^d;\mathbb{R})$, for $\mathcal{L}_T$-a.e. $t \in [0,T]$, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}\varphi(x)\,\mathrm{d}\mu_t^N(x) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{[0,1]}\varphi(X_t^N(\omega))\,\mathrm{d}\mathcal{L}_1(\omega)\\
&= \int_{[0,1]}\nabla\varphi(X_t^N(\omega))\cdot\dot{X}_t^N(\omega)\,\mathrm{d}\mathcal{L}(\omega)\\
&= \int_{[0,1]}\nabla\varphi(X_t^N(\omega))\cdot f(X_t^N(\omega),\tilde{u}_t^N(\omega),\mu_t^N)\,\mathrm{d}\mathcal{L}(\omega)\\
&= \int_{\mathbb{R}^d\times U}\nabla\varphi(x)\cdot f(x,u,\mu_t^N)\,\mathrm{d}\rho_t^N(x,u)\\
&= \int_{\mathbb{R}^d}\nabla\varphi(x)\cdot\int_U f(x,u,\mu_t^N)\,\mathrm{d}\rho_{t,x}^N(u)\,\mathrm{d}\mu_t^N(x)\\
&= \int_{\mathbb{R}^d}\nabla\varphi(x)\cdot f\left(x,\int_U u\,\mathrm{d}\rho_{t,x}^N(u),\mu_t^N\right)\,\mathrm{d}\mu_t^N(x)\\
&= \int_{\mathbb{R}^d}\nabla\varphi(x)\cdot f\left(x,\underline{u}^N(t,x),\mu_t^N\right)\,\mathrm{d}\mu_t^N(x).
\end{aligned}$$

Let us conclude the proof of the convergence showing that (6.3) holds. For any $t \in [0,T]$, using $(\tilde{X},\tilde{u})$ introduced in Step 1, we define $\rho_t := (\tilde{X}_t,\tilde{u}_t)_\sharp\mathcal{L}_1 \in \mathscr{P}(\mathbb{R}^d\times U)$. By the convergence (9.19) and the convergence of $\tilde{u}^N$ to $\tilde{u}$, it easily follows that $\rho^N := \rho_t^N\otimes\mathcal{L}_T$ weakly converges to $\rho := \rho_t\otimes\mathcal{L}_T$ in $\mathscr{P}([0,T]\times\mathbb{R}^d\times U)$.

Let now $\phi \in C_c([0,T]\times\mathbb{R}^d;V')$. Using the definition of $\rho^N$ and the weak convergence of $\rho^N$ to $\rho$ we have that

$$\begin{aligned}
\lim_{N\to+\infty}\int_0^T\int_{\mathbb{R}^d}\langle\phi(t,x),\bar{u}^N(t,x)\rangle\,\mathrm{d}\mu_t^N(x)\,\mathrm{d}t &= \lim_{N\to+\infty}\int_0^T\int_{\mathbb{R}^d}\left\langle\phi(t,x),\int_U u\,\mathrm{d}\rho_{t,x}^N(u)\right\rangle\,\mathrm{d}\mu_t^N(x)\,\mathrm{d}t\\
&= \lim_{N\to+\infty}\int_0^T\int_{\mathbb{R}^d}\int_U\langle\phi(t,x),u\rangle\,\mathrm{d}\rho_{t,x}^N(u)\,\mathrm{d}\mu_t^N(x)\,\mathrm{d}t\\
&= \lim_{N\to+\infty}\int_{[0,T]\times\mathbb{R}^d\times U}\langle\phi(t,x),u\rangle\,\mathrm{d}\rho^N(t,x,u),\\
&= \int_{[0,T]\times\mathbb{R}^d\times U}\langle\phi(t,x),u\rangle\,\mathrm{d}\rho(t,x,u),
\end{aligned}$$

Using $\boldsymbol{\eta} \in \mathscr{P}(C([0,T];\mathbb{R}^d))$, $P:[0,1]\to\Gamma_T$ and (9.15) introduced in Step 1, recalling that $\rho = (\tilde{X}_t,\tilde{u}_t)_\sharp\mathcal{L}_1\otimes\mathcal{L}_T$ we get

$$\begin{aligned}
\int_{[0,T]\times\mathbb{R}^d\times U}\langle\phi(t,x),u\rangle\,\mathrm{d}\rho(t,x,u) &= \int_0^T\int_0^1\langle\phi(t,\tilde{X}_t(\omega)),\tilde{u}_t(\omega)\rangle\,\mathrm{d}\mathcal{L}_1(\omega)\,\mathrm{d}t\\
&= \int_0^T\int_0^1\langle\phi(t,e_t(P(\omega))),u_t(P(\omega))\rangle\,\mathrm{d}\mathcal{L}_1(\omega)\,\mathrm{d}t\\
&= \int_0^T\int_{\Gamma_T}\langle\phi(t,e_t(\gamma)),u_t(\gamma)\rangle\,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}t\\
&= \int_0^T\int_{\Gamma_T}\langle\phi(t,e_t(\gamma)),\underline{u}(t,e_t(\gamma))\rangle\,\mathrm{d}\boldsymbol{\eta}(\gamma)\,\mathrm{d}t\\
&= \int_0^T\int_{\mathbb{R}^d}\langle\phi(t,x),\underline{u}(t,x)\rangle\,\mathrm{d}\mu_t(x)\,\mathrm{d}t.
\end{aligned}$$

**Step 4.** Finally we prove that $J_{\mathbf{E}^N}(\mu^N, \underline{u}^N) \to J_{\mathbf{E}}(\mu, \underline{u})$, as $N \to +\infty$.
Using item (3) of the Convexity Assumption 3.4, we have

$$\begin{aligned}
J_{\mathbf{E}^N}(\mu^N, \underline{u}^N) &= \int_0^T \int_{\mathbb{R}^d} \mathcal{C}(x, \bar{u}^N(t,x), \mu_t^N)\, \mathrm{d}\mu_t^N(x)\, \mathrm{d}t + \int_{\mathbb{R}^d} \mathcal{C}_T(x, \mu_T^N)\, \mathrm{d}\mu_T^N(x) \\
&= \int_0^T \int_{\mathbb{R}^d} \mathcal{C}\left(x, \int_U u\, \mathrm{d}\rho_{t,x}^N(u), \mu_t^N\right) \mathrm{d}\mu_t^N(x)\, \mathrm{d}t + \int_{[0,1]} \mathcal{C}_T(X_T^N(\omega), \mu_T^N)\, \mathrm{d}\mathcal{L}_1(\omega) \\
&\le \int_0^T \int_{\mathbb{R}^d \times U} \mathcal{C}(x, u, \mu_t^N)\, \mathrm{d}\rho_t^N(x,u)\, \mathrm{d}t + \int_{[0,1]} \mathcal{C}_T(X_T^N(\omega), \mu_T^N)\, \mathrm{d}\mathcal{L}_1(\omega) \\
&= \int_0^T \int_{[0,1]} \mathcal{C}(X_t^N(\omega), \tilde{u}_t^N(\omega), \mu_t^N)\, \mathrm{d}\mathcal{L}_1(\omega)\, \mathrm{d}t + \int_{[0,1]} \mathcal{C}_T(X_T^N(\omega), \mu_T^N)\, \mathrm{d}\mathcal{L}_1(\omega) \\
&= J_{\mathbf{L}}(X^N, \tilde{u}^N).
\end{aligned}$$

By the previous inequality, recalling (9.20) and (9.13), we get

$$\limsup_{N\to+\infty} J_{\mathbf{E}^N}(\mu^N, \underline{u}^N) \le \limsup_{N\to+\infty} J_{\mathbf{L}}(X^N, \tilde{u}^N) = J_{\mathbf{L}}(\tilde{X}, \tilde{u}) = J_{\mathbf{E}}(\mu, \underline{u}).$$

By Proposition 6.7 we conclude. □

We conclude this section with the proof of Theorem 9.12.

*Proof of Theorem 9.12.* Observe that $V_{\mathbf{E}^N}(\mu_0^N) \ge V_{\mathbf{E}}(\mu_0^N)$, then by Proposition 6.9 we obtain

$$\liminf_{N\to+\infty} V_{\mathbf{E}^N}(\mu_0^N) \ge V_{\mathbf{E}}(\mu_0).$$

Let us prove that $\limsup_{N\to+\infty} V_{\mathbf{E}^N}(\mu_0^N) \le V_{\mathbf{E}}(\mu_0)$. Fix $\varepsilon > 0$. By definition of $V_{\mathbf{E}}$ there exists $(\mu_\varepsilon, \underline{u}_\varepsilon) \in \mathcal{A}_{\mathbf{E}}(\mu_0)$ such that $V_{\mathbf{E}}(\mu_0) \ge J_{\mathbf{E}}(\mu_\varepsilon, \underline{u}_\varepsilon) - \varepsilon$. By Proposition 9.13 there exists $(\mu_\varepsilon^N, \underline{u}_\varepsilon^N) \in \mathcal{A}_{\mathbf{E}^N}(\mu_0^N)$ such that $\lim_{N\to+\infty} J_{\mathbf{E}^N}(\mu_\varepsilon^N, \underline{u}_\varepsilon^N) = J_{\mathbf{E}}(\mu_\varepsilon, \underline{u}_\varepsilon)$. Thus,

$$\limsup_{N\to+\infty} V_{\mathbf{E}^N}(\mu_0^N) \le \limsup_{N\to+\infty} J_{\mathbf{E}^N}(\mu_\varepsilon^N, \underline{u}_\varepsilon^N) = J_{\mathbf{E}}(\mu_\varepsilon, \underline{u}_\varepsilon) \le V_{\mathbf{E}}(\mu_0) + \varepsilon.$$

By the arbitrariness of $\varepsilon$ we conclude.

Finally, thanks to Theorem 9.3 it holds that $V_{\mathbf{E}^N}(\mu_0^N) = V_{\mathbf{L}^N}(X_0^N)$ for every $N \in \mathbb{N}$, so that $\lim_{N\to+\infty} V_{\mathbf{L}^N}(X_0^N) = V_{\mathbf{E}}(\mu_0)$. □

## Appendixes

We organize the material of the appendixes as follows. Appendix A deals with vector-valued Sobolev spaces and Cauchy problems for ODEs in Banach spaces. A further stability property of Cauchy problems is then established in Appendix B. In Appendix C we state and prove the superposition principle for the evolution of empirical measures. Appendix D is devoted to the proof of Proposition 4.16 where we construct (equipartite) finite algebras satisfying the Finite Approximation Property of Definition 4.15.

## Appendix A. Ordinary differential equations in Banach spaces

Let $E$ be a Banach space with $\|\cdot\|$ the associated norm. In the following, if $u : [0,T] \to E$ is a Bochner integrable function, we denote by $\int_0^T u(t)\, \mathrm{d}t$ its Bochner integral. We recall the following criterion of integrability: $u : [0,T] \to E$ is Bochner integrable if and only if there exists a sequence $u_n : [0,T] \to E$ of simple measurable functions such that $\lim_{n\to+\infty} u_n(t) = u(t)$ for $\mathcal{L}_T$-a.e. $t \in [0,T]$, and $\int_0^T \|u(t)\|\, \mathrm{d}t < +\infty$.

We recall that, if $u : [0,T] \to E$ is Bochner integrable, then

$$\left\| \int_a^b u(t)\, \mathrm{d}t \right\| \le \int_a^b \|u(t)\|\, \mathrm{d}t, \qquad \text{for any } [a,b] \subseteq [0,T], \tag{A.1}$$

$$\lim_{h\to 0}\frac{1}{h}\int_t^{t+h}\|u(s)-u(t)\|\,\mathrm{d}s=0,\qquad \text{for } \mathcal{L}_T\text{-a.e. } t\in[0,T], \tag{A.2}$$

and the above limit exists in every point of continuity of $u$. Moreover, for every continuous linear operator $A:E\to\tilde{E}$, with $\tilde{E}$ a Banach space it holds that

$$A\left(\int_0^T u(t)\,\mathrm{d}t\right)=\int_0^T A(u(t))\,\mathrm{d}t. \tag{A.3}$$

For the definition of Bochner integral, properties and related proofs, see for instance [26].

We say that $u\in W^{1,p}(0,T;E)$ if $u\in L^p(0,T;E)$ and there exists $g\in L^p(0,T;E)$ such that

$$\int_0^T\varphi'(t)u(t)\,\mathrm{d}t=-\int_0^T\varphi(t)g(t)\,\mathrm{d}t,\quad \forall\,\varphi\in C_c^\infty((0,T);\mathbb{R}).$$

We recall a classical result (see e.g. [4, Theorem 1.17])

**Proposition A.1.** *$u\in W^{1,p}(0,T;E)$ if and only if there exists $\tilde{u}\in \mathrm{AC}^p([0,T];E)$ such that $\tilde{u}(t)=u(t)$ and $\tilde{u}$ is differentiable for $\mathcal{L}_T$-a.e. $t\in[0,T]$.*

Let now $(\Omega,\mathfrak{B},\mathbb{P})$ be a probability space with $\Omega$ standard Borel. Fix $T>0$, let $\mathrm{Leb}_{[0,T]}$ be the $\sigma$-algebra of Lebesgue measurable sets on $[0,T]$ and $\mathcal{L}_T$ the normalized Lebesgue measure restricted to $[0,T]$. Consider $(\Omega_T,\mathfrak{B}_T,\mathfrak{m})$ the product space with $\Omega_T:=[0,T]\times\Omega$, endowed with the product $\sigma$-algebra $\mathfrak{B}_T=\mathrm{Leb}_{[0,T]}\otimes\mathfrak{B}$ and probability measure $\mathfrak{m}:=\mathcal{L}_T\otimes\mathbb{P}\in\mathscr{P}(\Omega_T)$.

**Lemma A.2.** *Let $p\geq 1$, and $g\in L^p_{\mathfrak{m}}(\Omega_T;\mathbb{R}^d)$ and $\hat{g}$ a Borel representative of $g$. Let $\tilde{g}$ the map defined by $\tilde{g}(t)(\omega)=\hat{g}(t,\omega)$ for every $(t,\omega)\in\Omega_T$. Then $\tilde{g}\in L^1(0,T;L^p(\Omega;\mathbb{R}^d))$ and, denoting by $G:=\int_0^T\tilde{g}(t)\,\mathrm{d}t$, it holds*

$$G(\omega)=\int_0^T\hat{g}(t,\omega)\,\mathrm{d}t,\qquad \text{for } \mathbb{P}\text{-a.e. } \omega\in\Omega. \tag{A.4}$$

The proof of the Lemma follows by Fubini's theorem and the definition of Bochner integral (notice that (A.4) holds for simple functions).

**Proposition A.3.** *Let $y\in L^p_{\mathfrak{m}}(\Omega_T;\mathbb{R}^d)$. The following are equivalent:*

*(1) There exists $g\in L^p_{\mathfrak{m}}(\Omega_T;\mathbb{R}^d)$ such that*

$$\int_{\Omega_T}\eta'(t)\phi(\omega)\,y(t,\omega)\,\mathrm{d}\mathfrak{m}(t,\omega)=-\int_{\Omega_T}\eta(t)\phi(\omega)\,g(t,\omega)\,\mathrm{d}\mathfrak{m}(t,\omega),$$

*for every $\eta\in C_c^1((0,T);\mathbb{R})$ and $\phi\in\mathrm{M}(\Omega;\mathbb{R})$ bounded;*

*(2) $y\in W^{1,p}(0,T;L^p_{\mathbb{P}}(\Omega;\mathbb{R}^d))$;*

*(3) there exists a Borel representative $\tilde{y}$ of $y$ such that $\tilde{y}\in\mathrm{AC}^p([0,T];L^p_{\mathbb{P}}(\Omega;\mathbb{R}^d))$ and differentiable for $\mathcal{L}_T$-a.e. $t\in[0,T]$ (differentiability is redundant for $p>1$);*

*(4) $y\in L^p_{\mathbb{P}}(\Omega;W^{1,p}(0,T;\mathbb{R}^d))$;*

*(5) there exists a Borel representative $\bar{y}$ of $y$ such that $\bar{y}\in L^p_{\mathbb{P}}(\Omega;\mathrm{AC}^p([0,T];\mathbb{R}^d))$.*

*Moreover,*

*(i) If (3) holds, there exists a Borel function $\tilde{g}\in L^p_{\mathfrak{m}}(\Omega_T;\mathbb{R}^d)$ such that $\tilde{g}(t,\cdot)=\tilde{y}'(t,\cdot)$ in $L^p_{\mathbb{P}}(\Omega;\mathbb{R}^d)$, for $\mathcal{L}_T$-a.e. $t\in[0,T]$. Hence, for every $t\in[0,T]$*

$$\tilde{y}(t,\cdot)=\tilde{y}(0,\cdot)+\int_0^t\tilde{g}(s,\cdot)\,\mathrm{d}s,\qquad \text{in } L^p_{\mathbb{P}}(\Omega). \tag{A.5}$$

*(ii) If (5) holds, there exists a Borel function $\bar{g}\in L^p_{\mathfrak{m}}(\Omega_T;\mathbb{R}^d)$ such that for every $\omega\in\Omega$, $\bar{g}(t,\omega)=(\bar{y}(\cdot,\omega))'(t)$, for $\mathcal{L}_T$-a.e. $t\in[0,T]$. Hence, for every $\omega\in\Omega$ it holds*

$$\bar{y}(t,\omega)=\bar{y}(0,\omega)+\int_0^t\bar{g}(s,\omega)\mathrm{d}s,\quad \text{for every } t\in[0,T]. \tag{A.6}$$

*(iii) If one of the five conditions above is satisfied, then $g=\tilde{g}=\bar{g}$, $\mathfrak{m}$-a.e. in $\Omega_T$.*

*Proof.* (5) $\Rightarrow$ (4) and (3) $\Leftrightarrow$ (2) follow from Proposition A.1. (4) $\Rightarrow$ (2) follows from the definition and Fubini's theorem. (2) $\Rightarrow$ (1) is a consequence of property (A.3). Finally, the proof of (1) $\Rightarrow$ (5) is contained in a more general form in [46, Lemma 4.3].

Items $(i)$-$(ii)$ are a consequence of the definition of $\mathrm{AC}^p([0,T];L^p(\Omega;\mathbb{R}^d))$ and $\mathrm{AC}^p([0,T];\mathbb{R}^d)$, respectively. To show $(iii)$, denote with $y_0(\omega)$ the trace of $y$ in $t=0$, which is well defined thanks to (1). Then, $y_0(\omega)=\tilde{y}(0,\omega)=\bar{y}(0,\omega)$ for $\mathbb{P}$-a.e. $\omega\in\Omega$ so that comparing (A.5) and (A.6) it holds $\tilde{g}=\bar{g}$ $\mathfrak{m}$-a.e. in $\Omega_T$, thanks to (A.4). $\square$

A.1. **Cauchy problem in Banach spaces.** We are interested in a Cauchy problem of this form

$$\begin{cases}\dot{z}_t=F(t,z_t), & \text{for } \mathcal{L}_T\text{-a.e. } t\in[0,T]\\ z_{t=0}=z_0,\end{cases}\tag{A.7}$$

where a Carathéodory function $F:[0,T]\times E\to E$ and $z_0\in E$ are given. For $z:[0,T]\to E$ we frequently use the notation $z_t:=z(t)$.

In the following, we present some classical results concerning the Cauchy problem (A.7) and we provide a sketch of their proofs.

**Proposition A.4.** *Let $F:[0,T]\times E\to E$ be a Carathéodory function such that*

$$\int_0^T\|F(t,z_t)\|\,\mathrm{d}t<+\infty,\qquad\forall z\in C([0,T];E).\tag{A.8}$$

*The following assertions are equivalent:*

- *$z\in C([0,T];E)$ satisfies*

$$z_t=z_0+\int_0^t F(s,z_s)\,\mathrm{d}s,\qquad\forall\,t\in[0,T].\tag{A.9}$$

- *$z\in\mathrm{AC}^1([0,T];E)$, it is differentiable for $\mathcal{L}_T$-a.e. $t\in[0,T]$ and*

$$\begin{cases}\dot{z}_t=F(t,z_t), & \text{for } \mathcal{L}_T\text{-a.e. } t\in[0,T]\\ z_{t=0}=z_0.\end{cases}$$

Since $F$ is a Carathéodory function and the curve $t\mapsto z_t$ then the map $t\mapsto F(t,z_t)$ is (strongly) measurable as a map with values in $L^p(\Omega;\mathbb{R}^d)$. Notice also that if $E$ satisfies the Radon-Nikodym property (for instance when $E$ is reflexive) then a.e. differentiability of $z$ in the second item is redundant.

**Theorem A.5.** *Let $F:[0,T]\times E\to E$ be a Carathéodory function such that*

$$\|F(t,z^1)-F(t,z^2)\|\le\tilde{L}\|z^1-z^2\|,\qquad\forall\,(t,z^1),(t,z^2)\in[0,T]\times E,\tag{A.10}$$

*for some $\tilde{L}>0$, and there exists $C_0\ge 0$ such that*

$$\sup_{t\in[0,T]}\|F(t,0)\|\le C_0<+\infty.\tag{A.11}$$

*Then, for any $z_0\in E$ there exists a unique $z\in\mathrm{AC}^\infty([0,T];E)$ and differentiable for $\mathcal{L}_T$-a.e. $t\in[0,T]$ such that*

$$\begin{cases}\dot{z}_t=F(t,z_t), & \text{for } \mathcal{L}_T\text{-a.e. } t\in[0,T]\\ z(0)=z_0.\end{cases}\tag{A.12}$$

*Moreover, the following estimates hold:*

$$\sup_{t\in[0,T]}\|z_t\|\le e^{\tilde{L}T}\left(\|z_0\|+C_0T\right),\tag{A.13}$$

$$\|z_t-z_s\|\le|t-s|\left(C_0+\tilde{L}e^{\tilde{L}T}\left(\|z_0\|+C_0T\right)\right)\qquad\forall\,s,t\in[0,T].\tag{A.14}$$

*Proof.* We provide only a sketch of the proof.

Assumptions (A.10) and (A.11) yield the following growth property

$$\|F(t,z)\| \le C_0 + \tilde{L}\|z\|, \qquad \forall\,(t,z) \in [0,T]\times E. \tag{A.15}$$

We define the Banach space $(\mathscr{S}, \|\cdot\|_{\mathscr{S}})$ as follows

$$\mathscr{S} := \left\{ z \in C([0,T];E) \,:\, \|z\|_{\mathscr{S}} := \sup_{t\in[0,T]} e^{-\tilde{L}t}\|z_t\| < +\infty \right\},$$

and the operator $S : \mathscr{S} \to \mathscr{S}$ by

$$S(y)_t = z_0 + \int_0^t F(s, y_s)\,\mathrm{d}s.$$

By (A.10) and (A.15), using (A.1), it is classical to prove that $S$ is well defined and it is a contraction. Then, by Banach fixed point Theorem we get the existence and uniqueness of $z \in \mathscr{S}$ such that (A.9) holds. The estimates (A.13) and (A.14) follow from (A.9), (A.15), and Gronwall inequality. Finally, $z$ belongs to $\mathrm{AC}^\infty([0,T];E)$ thanks to (A.14). □

**Proposition A.6.** *Let $F, F^n : [0,T]\times E \to E$, $n \in \mathbb{N}$, be Carathéodory functions satisfying* (A.10),(A.11) *with the same constant $\tilde{L}$ and $C_0$. Let $z_0, z_0^n \in E$, $z \in \mathrm{AC}^\infty([0,T];E)$ the solution of* (A.12) *and $z^n \in \mathrm{AC}^\infty([0,T];E)$ the solutions of*

$$\begin{cases} \dot{z}_t^n = F^n(t, z_t^n), & \text{for } \mathcal{L}_T\text{-a.e. } t \in [0,T] \\ z^n(0) = z_0^n. \end{cases}$$

*Then*

$$\sup_{t\in[0,T]} \|z_t^n - z_t\| \le e^{\tilde{L}T}\left( \|z_0^n - z_0\| + \int_0^T \|F^n(t,z_t) - F(t,z_t)\|\,\mathrm{d}t \right). \tag{A.16}$$

*In particular, if $\lim_{n\to+\infty}\|z_0^n - z_0\| = 0$ and*

$$\lim_{n\to+\infty} \int_0^T \|F^n(t,z_t) - F(t,z_t)\|\,\mathrm{d}t = 0, \tag{A.17}$$

*then*

$$\lim_{n\to+\infty} \sup_{t\in[0,T]} \|z_t^n - z_t\| = 0.$$

*Proof.* We have

$$\begin{aligned} \|z_t^n - z_t\| &= \left\| z_0^n - z_0 + \int_0^t (F^n(s,z_s^n) - F(s,z_s))\,\mathrm{d}s \right\| \\ &\le \|z_0^n - z_0\| + \int_0^t \|F^n(s,z_s^n) - F(s,z_s))\|\,\mathrm{d}s. \end{aligned}$$

Using (A.10) it holds

$$\begin{aligned} &\|F^n(s,z_s^n) - F(s,z_s))\| \\ &\le \|F^n(s,z_s^n) - F^n(s,z_s))\| + \|F^n(s,z_s) - F(s,z_s))\| \\ &\le \tilde{L}\|z_s^n - z_s\| + \|F^n(s,z_s) - F(s,z_s))\|. \end{aligned}$$

The last two inequalities yield

$$\|z_t^n - z_t\| \le \|z_0^n - z_0\| + \int_0^T \|F^n(s,z_s) - F(s,z_s))\|\,\mathrm{d}s + \tilde{L}\int_0^t \|z_s^n - z_s\|\,\mathrm{d}s.$$

By Gronwall lemma we obtain (A.16). □

**Proposition A.7.** *Let $F : [0,T]\times L^p(\Omega;\mathbb{R}^d) \to L^p(\Omega;\mathbb{R}^d)$ be a Carathéodory function satisfying* (A.11) *Let $Y_0 \in L^p(\Omega;\mathbb{R}^d)$. Then the following assertions are equivalent:*

*(1)* $Y \in C([0,T]; L^p(\Omega;\mathbb{R}^d))$ *and*

$$Y_t = Y_0 + \int_0^t F(s, Y_s)\,\mathrm{d}s, \qquad \forall\, t \in [0,T]; \tag{A.18}$$

*(2)* $Y \in \mathrm{AC}^p([0,T]; L^p(\Omega;\mathbb{R}^d))$ *(if* $p=1$ *it is also differentiable for a.e.* $t \in [0,T]$*) and*

$$\begin{cases} \dot{Y}_t = F(t, Y_t), & \text{for } \mathcal{L}_T\text{-a.e. } t \in [0,T] \\ Y_{t=0} = Y_0; \end{cases} \tag{A.19}$$

*(3)* $Y \in L^p(\Omega; \mathrm{AC}^p([0,T];\mathbb{R}^d)))$ *and*

$$Y_t(\omega) = Y_0(\omega) + \int_0^t F(s, Y_s)(\omega)\,\mathrm{d}s, \qquad \forall\, t \in [0,T], \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega, \tag{A.20}$$

*where* $Y_t : \Omega \to \mathbb{R}^d$ *is defined by* $Y_t(\omega) := Y(t,\omega)$ *for* $\mathbb{P}$*-a.e.* $\omega \in \Omega$.

*(4)* $Y \in L^p(\Omega; \mathrm{AC}^p([0,T];\mathbb{R}^d)))$ *and for* $\mathbb{P}$*-a.e.* $\omega \in \Omega$ *it holds*

$$\begin{cases} \dot{Y}_t(\omega) = F(t, Y_t)(\omega), & \text{for } \mathcal{L}_T\text{-a.e. } t \in [0,T] \\ Y_{t=0}(\omega) = Y_0(\omega), \end{cases} \tag{A.21}$$

*where* $Y_t : \Omega \to \mathbb{R}^d$ *is defined by* $Y_t(\omega) := Y(t,\omega)$ *for* $\mathbb{P}$*-a.e.* $\omega \in \Omega$.

*Proof.* The assertions (1)-(2) and (3)-(4) are equivalent by Proposition A.4. The equivalence (1)-(3) is a consequence of the equivalences (3)-(5) in Proposition A.3 and items (i)-(ii)-(iii). □

## Appendix B. A convergence result for solutions of Cauchy problems

We state and prove the following well known result, for sake of completeness.

**Lemma B.1.** *Let* $U$ *be a Polish space. Let* $u^n : [0,T] \to U$ *be a sequence of* $\mathcal{L}_T$*-measurable functions such that* $(i_{[0,T]}, u^n)_\sharp \mathcal{L}_T \xrightarrow{\mathcal{Y}} \nu_t \otimes \mathcal{L}_T \in \mathscr{P}([0,T]\times U)$.
*Let* $g : [0,T] \times (\mathbb{R}^d \times U) \to \mathbb{R}^d$ *a Carathéodory function such that*

$$|g(t,x_1,u) - g(t,x_2,u)| \le L|x_1 - x_2|, \qquad \forall\, (t,x_1,u),(t,x_2,u) \in [0,T]\times\mathbb{R}^d\times U, \tag{B.1}$$

*for some* $L > 0$*, and*

$$C_0 := \sup_{(t,u)\in[0,T]\times U} |g(t,0,u)| < +\infty. \tag{B.2}$$

*Given* $X_0 \in \mathbb{R}^d$ *and* $n \in \mathbb{N}$*, we denote by* $X^n \in \mathrm{AC}^1([0,T];\mathbb{R}^d)$ *the unique solution of the Cauchy problem*

$$\begin{cases} \dot{X}^n_t = g(t, X^n_t, u^n_t), & \text{for a.e. } t \in (0,T) \\ X^n(0) = X_0, \end{cases} \tag{B.3}$$

*and by* $X \in \mathrm{AC}^1([0,T];\mathbb{R}^d)$ *the unique solution of the Cauchy problem*

$$\begin{cases} \dot{X}_t = \displaystyle\int_U g(t, X_t, u)\,\mathrm{d}\nu_t(u), & \text{for a.e. } t \in (0,T) \\ X(0) = X_0. \end{cases} \tag{B.4}$$

*Then*

$$\lim_{n\to+\infty} \sup_{t\in[0,T]} |X^n_t - X_t| = 0. \tag{B.5}$$

*Proof.* Observe that existence and uniqueness of solutions of Cauchy problems (B.3) and (B.4) is consequence of the fact that $G^n, G : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$, defined by $G^n(t,x) := g(t,x,u^n_t)$ and

$G(t,x) := \int_U g(t,x,u)\,\mathrm{d}\nu_t(u)$, are Charathéodory and $L$-Lipschitz continuous w.r.t. $x \in \mathbb{R}^d$. Moreover, by (B.1) and (B.2) it holds

$$|g(t,x,u)| \le C_0 + L|x|, \qquad \forall\,(t,x,u) \in [0,T]\times\mathbb{R}^d\times U. \tag{B.6}$$

We define $Y^n \in C([0,T];\mathbb{R}^d)$ by

$$Y^n_t := X_0 + \int_0^t g(s,X_s,u^n_s)\,\mathrm{d}s. \tag{B.7}$$

From the convergence $(i_{[0,T]},u^n)_\sharp\mathcal{L}_T \xrightarrow{\mathcal{Y}} \nu \in \mathscr{P}([0,T]\times U)$ it follows that

$$\lim_{n\to+\infty} Y^n_t = X_0 + \int_0^t\int_U g(s,X_s,u)\,\mathrm{d}\nu_s(u)\,\mathrm{d}s = X_t, \qquad \forall\,t\in[0,T]. \tag{B.8}$$

Denoting by $C := \sup_{s\in[0,T]}|X_s|$, from (B.2) and (B.1) it is simple to prove that

$$|Y^n_t| \le (|X_0| + (C_0+LC)T), \qquad \forall t\in[0,T] \tag{B.9}$$

and

$$|Y^n_t - Y^n_s| \le |t-s|(C_0+LC)T, \qquad \forall t,s\in[0,T]. \tag{B.10}$$

By (B.9) and (B.10), Ascoli-Arzelà theorem and (B.8) imply that

$$\lim_{n\to+\infty}\sup_{t\in[0,T]}|Y^n_t - X_t| = 0. \tag{B.11}$$

Since

$$\begin{aligned}|X^n_t - X_t| &\le |X^n_t - Y^n_t| + |Y^n_t - X_t| \le \int_0^t |g(s,X^n_s,u^n_s) - g(s,X_s,u^n_s)|\,\mathrm{d}s + |Y^n_t - X_t|\\ &\le L\int_0^t |X^n_s - X_s|\,\mathrm{d}s + \sup_{s\in[0,T]}|Y^n_s - X_s|,\end{aligned}$$

by Gronwall inequality we have that

$$|X^n_t - X_t| \le e^{Lt}\sup_{s\in[0,T]}|Y^n_s - X_s|. \tag{B.12}$$

The convergence (B.5) follows from (B.12) and (B.11). □

## Appendix C. An empirical Superposition Principle

In this appendix, we give a refined version of the Superposition Principle (see Theorem 2.5 for the classical result) in the case of trajectories of the form $\mu_t \in \mathscr{P}^N(\mathbb{R}^d)$ for any $t\in[0,T]$, where $\mathscr{P}^N(\mathbb{R}^d)$ is the space of empirical probability measures

$$\mathscr{P}^N(\mathbb{R}^d) := \left\{\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \ \text{ for some } x_i\in\mathbb{R}^d\right\}.$$

The novelty consists in proving that if $\mu_t \in \mathscr{P}^N(\mathbb{R}^d)$ for every $t\in[0,T]$, then there exists representative $\eta\in\mathscr{P}^N(\Gamma_T)$. This result has been used to prove Proposition 9.8 and Corollary 9.3.

**Theorem C.1.** *Let $N\in\mathbb{N}$ and $\mu\in \mathrm{AC}([0,T];\mathscr{P}_1(\mathbb{R}^d))$ such that $\mu_t\in\mathscr{P}^N(\mathbb{R}^d)$ for every $t\in[0,T]$.*

*(1) There exists a unique (up to $\mathcal{L}_T\otimes\mu_t$-negligible sets) Borel vector field $v:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ satisfying*

$$\int_0^T\int_{\mathbb{R}^d}|v_t(x)|\,\mathrm{d}\mu_t(x)\,\mathrm{d}t < +\infty, \tag{C.1}$$

*such that $\mu$ is a distributional solution of the continuity equation*

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0, \quad \text{in } [0,T] \times \mathbb{R}^d. \tag{C.2}$$

*(2) There exists $\eta \in \mathscr{P}^N(\Gamma_T)$ of the form*

$$\eta = \frac{1}{N} \sum_{i=1}^{N} \delta_{\gamma_i}, \tag{C.3}$$

*such that $(e_t)_\sharp \eta = \mu_t$ for every $t \in [0,T]$ and for any $i = 1, \dots, N$, $\gamma_i \in \mathrm{AC}([0,T];\mathbb{R}^d)$ solves the differential equation*

$$\dot{\gamma}_i(t) = v_t(\gamma_i(t)) \quad \text{for } \mathcal{L}_T\text{-a.e. } t \in [0,T]. \tag{C.4}$$

*Proof.* Let us recall that the metric derivative $|\mu'|$ of the absolutely continuous curve $\mu$, given by

$$|\mu'|(t) := \lim_{s \to t} \frac{W_1(\mu_t, \mu_s)}{|t-s|}, \quad \text{for a.e. } t \in [0,T],$$

belongs to $L^1(0,T)$ and satisfies

$$W_1(\mu_{t_1}, \mu_{t_2}) \le \int_{t_1}^{t_2} |\mu'|(t)\, \mathrm{d}t, \quad \text{for any } 0 \le t_1 \le t_2 \le T. \tag{C.5}$$

in particular there exists $\psi : [0,+\infty) \to [0,+\infty)$ increasing, convex and superlinear at $+\infty$ such that

$$\int_0^T \psi\big(|\mu'|(t)\big)\, \mathrm{d}t < +\infty. \tag{C.6}$$

First of all we prove the existence of $\eta \in \mathscr{P}(\Gamma_T)$ such that $(e_t)_\sharp \eta = \mu_t$ for all $t \in [0,T]$ and $\eta$ is of the form (C.3) for some $\gamma_i \in \Gamma_T$, $i = 1, \dots, N$.

Let $M \in \mathbb{N}$ and consider the diadic discretization of the interval $[0,T]$, with time step $\tau_M = T\, 2^{-M}$. Since $\mu_t \in \mathscr{P}^N(\mathbb{R}^d)$ for every $t \in [0,T]$, there exists $x_i(t) \in \mathbb{R}^d$, $i = 1, \dots, N$, such that

$$\mu_t = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i(t)}, \qquad \forall\, t \in [0,T].$$

For $n = 0, \dots, 2^M$ and $i = 1, \dots, N$, we set $x^n_{M,i} := x_i(n\tau_M)$, and $\mu^n_M := \mu_{n\tau_M}$. For $n = 1, \dots, 2^M$, let $\varrho^{n-1,n}_M \in \Gamma_o(\mu^{n-1}_M, \mu^n_M)$ be an optimal plan for the 1-Wasserstein distance. Since $\mu^{n-1}_M$ and $\mu^n_M$ belong to $\mathscr{P}^N(\mathbb{R}^d)$, then $\varrho^{n-1,n}_M$ is of the form

$$\varrho^{n-1,n}_M = \frac{1}{N} \sum_{i=1}^{N} \delta_{x^{n-1}_{M,i}} \otimes \delta_{x^n_{M,\sigma^n_M(i)}}, \tag{C.7}$$

for some permutation $\sigma^n_M$ of $\{1, \dots, N\}$. Let us define $\sigma^{0,n}_M := \sigma^n_M \circ \sigma^{n-1}_M \circ \dots \circ \sigma^1_M$ and $\sigma^{0,0}_M(i) = i$ for $i = 1, \dots, N$.
For $i = 1, \dots, N$ we define the curves $\gamma_{M,i} \in \Gamma_T$ by linear time interpolation as

$$\gamma_{M,i}(t) := \frac{n\tau_M - t}{\tau_M}\, x^{n-1}_{M,\sigma^{0,n-1}_M(i)} + \frac{t-(n-1)\tau_M}{\tau_M}\, x^n_{M,\sigma^{0,n}_M(i)}, \quad \text{for } t \in [(n-1)\tau_M, n\tau_M],$$

$n = 1, \dots, 2^M$.

We claim that, for any $k = 1, \dots, N$, the sequence $\{\gamma_{M,k}\}_{M\in\mathbb{N}}$ uniformly converges to a curve $\gamma_k \in \mathrm{AC}([0,T];\mathbb{R}^d)$. Indeed,

$$
\begin{aligned}
\int_0^T \psi\Big(\frac{1}{N}|\dot\gamma_{M,k}(t)|\Big)\,\mathrm{d}t &\le \int_0^T \psi\Big(\frac{1}{N}\sum_{i=1}^N |\dot\gamma_{M,i}(t)|\Big)\,\mathrm{d}t \\
&= \sum_{n=1}^{2^M} \tau_M\, \psi\Big(\frac{1}{N}\sum_{i=1}^N \frac{\Big|x^n_{M,\sigma_M^{0,n}(i)} - x^{n-1}_{M,\sigma_M^{0,n-1}(i)}\Big|}{\tau_M}\Big) \\
&= \sum_{n=1}^{2^M} \tau_M\, \psi\Big(\frac{1}{\tau_M}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|\,\mathrm{d}\varrho_M^{n-1,n}(x,y)\Big) \\
&= \sum_{n=1}^{2^M} \tau_M\, \psi\Big(\frac{1}{\tau_M} W_1(\mu_M^{n-1},\mu_M^n)\Big) \\
&\le \sum_{n=1}^{2^M} \tau_M\, \psi\Big(\frac{1}{\tau_M}\int_{(n-1)\tau_M}^{n\tau_M} |\mu'|(t)\,\mathrm{d}t\Big) \\
&\le \sum_{n=1}^{2^M} \int_{(n-1)\tau_M}^{n\tau_M} \psi(|\mu'|(t))\,\mathrm{d}t = \int_0^T \psi(|\mu'|(t))\,\mathrm{d}t,
\end{aligned}
\tag{C.8}
$$

where we employed the definition of the optimal plan in (C.7), (C.5), Jensen's inequality, and (C.6).
Since $\gamma_{M,k}(0) = x_k(0)$ for any $M \in \mathbb{N}$, and (C.8) and (C.6) hold, by Ascoli-Arzelá Theorem the sequence $\{\gamma_{M,k}\}_{M\in\mathbb{N}}$ is compact in $C([0,T];\mathbb{R}^d)$. Furthermore, by (C.8) and the lower semicontinuity of the functional

$$
\gamma \mapsto \int_0^T \psi\left(\frac{1}{N}|\dot\gamma(t)|\right)\mathrm{d}t
$$

w.r.t. weak convergence in $\mathrm{AC}([0,T];\mathbb{R}^d)$, we get $\gamma_k \in \mathrm{AC}([0,T];\mathbb{R}^d)$. Moreover, if $t = n_0\tau_{M_0}$ for some $M_0 \in \mathbb{N}$ and $n_0 \in \{0,1,\dots,2^{M_0}\}$, then $\gamma_{M,k}(t)$ is constant for any $M \in \mathbb{N}$, $M > M_0$ and the claim is proved.

Defining

$$
\eta_M := \frac{1}{N}\sum_{i=1}^N \delta_{\gamma_{M,i}}, \qquad \eta := \frac{1}{N}\sum_{i=1}^N \delta_{\gamma_i},
\tag{C.9}
$$

from the convergence of $\gamma_{M,i}$ to $\gamma_i$ it follows that $\eta_M$ weakly converges to $\eta$ as $M \to +\infty$. Moreover, if $t = n_0\tau_{M_0}$ for some $M_0 \in \mathbb{N}$ and $n_0 \in \{0,1,\dots,2^{M_0}\}$, then $(e_t)_\sharp\eta_M = \mu_t$ for any $M \in \mathbb{N}$, $M > M_0$. Then, by the continuity of $t \mapsto \mu_t$ and of $t \mapsto (e_t)_\sharp\eta$, we conclude that $(e_t)_\sharp\eta = \mu_t$ and $\mu_t = \frac{1}{N}\sum_{i=1}^N \delta_{\gamma_i(t)}$ for all $t \in [0,T]$.

It remains to define a vector field $v$ such that (C.4) and (C.2) hold, also showing that $v$ is uniquely characterized by (C.2).

Since $\gamma_i \in \mathrm{AC}([0,T];\mathbb{R}^d)$ for any $i = 1,\dots,N$, the Borel set $A := \{t \in [0,T] : \exists\, k \in \{1,\dots,N\}$ such that $\gamma_k$ is not differentiable at $t\}$ is $\mathcal{L}_T$-negligible. Moreover, the sets

$$
N_{i,k} := \{t \in [0,T] \setminus A : \gamma_i(t) = \gamma_k(t), \dot\gamma_i(t) \ne \dot\gamma_k(t)\}
\tag{C.10}
$$

satisfy $\mathcal{L}_T(N_{i,k}) = 0$ for any $i,k \in \{1,\dots,N\}$. We define $\tilde N := \Big(\bigcup_{i,k\in\{1,\dots,N\}} N_{i,k}\Big) \bigcup A$ noticing that $\mathcal{L}_T(\tilde N) = 0$ so that $\mathcal{L}_T \otimes \mu_t(\tilde N \times \mathbb{R}^d) = 0$, and $S := \{(t,\gamma_i(t)) : t \in [0,T],\ i \in \{1,\dots,N\}\} =$

$\mathrm{supp}(\mu_t \otimes \mathcal{L}_T)$. We can thus define a Borel vector field $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$v(t,x) := \begin{cases} 0 & \text{if } t \in \tilde{N} \text{ or } (t,x) \in \big([0,T] \times \mathbb{R}^d\big) \setminus S, \\ \dot{\gamma}_i(t) & \text{if } x = \gamma_i(t) \text{ for } t \in [0,T] \setminus \tilde{N} \text{ and some } i \in \{1,\dots,N\}, \end{cases}$$

so that $\dot{\gamma}_i(t) = v_t(\gamma_i(t))$ for every $t \in [0,T] \setminus \tilde{N}$. It is then easy to check that (C.1) and (C.2) hold.

Let us eventually check that (C.2) uniquely characterizes $v(t_0,x_0)$ for every $(t_0,x_0) \in S \setminus (\tilde{N} \times \mathbb{R}^d)$. Notice that for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\mu_t \Big|_{t=t_0} = \frac{1}{N} \sum_{i=1}^N \nabla\varphi(\gamma_i(t_0)) \cdot \dot{\gamma}_i(t_0) = \frac{1}{N} \sum_{i=1}^N \nabla\varphi(\gamma_i(t_0)) \cdot v(t_0, \gamma_i(t_0)). \tag{C.11}$$

Setting $K := \{k \in \{1,\dots,N\} : \gamma_k(t_0) \neq x_0\}$ and $r_0 := \min\{|\gamma_k(t_0) - x_0| : k \in K\} > 0$, for every $\xi \in \mathbb{R}^d$ we can find a test function $\varphi \in C_c^\infty(\mathbb{R}^d)$ such that $\mathrm{supp}\,\varphi \subset B_{r_0}(x_0)$ and $\nabla\varphi(x_0) = \xi$: (C.11) then yields

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \varphi \,\mathrm{d}\mu_t \Big|_{t=t_0} = \frac{n}{N} \xi \cdot v(t_0,x_0) \quad \text{where} \quad n := N - \#K. \tag{C.12}$$

Since $\xi$ is arbitrary, (C.12) uniquely characterizes $v(t_0,x_0)$ in terms of $\mu$. □

## Appendix D. Finite Partitions

In this section we provide a proof of Proposition 4.16. For sake of clarity, we divide the statement of Proposition 4.16 in three separate lemmas of independent interest.

Given a standard Borel space $(\Omega, \mathfrak{B}, \mathbb{P})$, in Lemma D.2 we construct a family of algebras $\mathfrak{B}^n$, $n \in \mathbb{N}$, satisfying the finite approximation property of Definition 4.15. Then we fix $(\Omega, \mathfrak{B}, \mathbb{P}) = ([0,1], \mathcal{B}, \mathcal{L}_1)$, where $\mathcal{B}$ is the Borel $\sigma$-algebra and $\mathcal{L}_1$ the Lebesgue measure restricted to the interval $[0,1]$. With this choice of parametrization space, in Lemma D.3 we show that the family of algebras $\mathcal{B}^N$ associated to the uniform partition of $[0,1]$ with elements' size $1/N$ satisfies the finite approximation property. Finally, we combine the previous results in Lemma D.4, where we consider a general standard Borel space $(\Omega, \mathfrak{B}, \mathbb{P})$ and $\mathbb{P}$ is without atoms. This is possible thanks to the following fundamental result on Borel equivalence of Probability spaces (see e.g. [50, Chapter 15, Theorem 9]).

**Proposition D.1.** *Let $\Omega$ be a Polish space and $\mathbb{P} \in \mathscr{P}(\Omega)$ without atoms. Then there exist a Borel set $\Omega_0 \subset \Omega$ such that $\mathbb{P}(\Omega_0) = 0$, a Borel set $I_0 \subset [0,1]$ such that $\mathcal{L}_1(I_0) = 0$ and a bijective function $\psi : \Omega \setminus \Omega_0 \to [0,1] \setminus I_0$ such that $\psi$ and $\psi^{-1}$ are Borel, $\psi_\sharp \mathbb{P} = \mathcal{L}_1$ and $(\psi^{-1})_\sharp \mathcal{L}_1 = \mathbb{P}$.*

The first part of Proposition 4.16 is restated in the following Lemma.

**Lemma D.2** (Proposition 4.16, part 1)**.** *Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space. Then there exists a family of finite algebras $\mathfrak{B}^n \subset \mathfrak{B}$, $n \in \mathbb{N}$, satisfying the finite approximation property of Definition 4.15.*

*Proof.* Since $(\Omega, \mathfrak{B}, \mathbb{P})$ is standard Borel we can choose a Polish topology $\tau$ such that $\mathfrak{B} = \mathcal{B}_{(\Omega,\tau)}$, then there exists a countable basis $\mathcal{A} = \{B^i : i \in \mathbb{N}\}$ of its topology. Then $\mathcal{B}_{(\Omega,\tau)} = \sigma(\{B^i : i \in \mathbb{N}\})$. We define $\mathfrak{B}^1 := \sigma(B^1)$ and $\mathfrak{B}^n := \sigma(\{B^n\} \cup \mathfrak{B}^{n-1})$. It follows from the definition that $\mathfrak{B}^n \subset \mathfrak{B}^{n+1}$ for any $n \in \mathbb{N}$ and $\mathcal{B}_{(\Omega,\tau)} = \sigma\left(\bigcup_{n=1}^{+\infty} \mathfrak{B}^n\right)$.

For any $n \in \mathbb{N}$, the finite algebra $\mathfrak{B}^n$ induces a minimal (with respect to the inclusion) partition of $\Omega$, denoted by $\mathcal{P}^n = \{A_k^n : k = 1,\dots,k(n)\} \subset \mathfrak{B}^n$. Then for any $A_k^{n+1} \in \mathcal{P}^{n+1}$ there exists $h \in \mathbb{N}$ such that $A_h^n \in \mathcal{P}^n$ and $A_k^{n+1} \subset A_h^n$.

We define the sequence of linear operators $P_n : L^1(\Omega; E) \to L^1(\Omega; E)$ defined by

$$P_n g := \sum_{k=1}^{k(n)} \mathbb{1}_{A_k^n} \fint_{A_k^n} g(\omega) \,\mathrm{d}\mathbb{P}(\omega),$$

with the convention that $\fint_{A_k^n} g(\omega)\,\mathrm{d}\mathbb{P}(\omega) = 0$ if $\mathbb{P}(A_k^n) = 0$.

It is simple to prove that

$$\|P_n g\|_{L^1_{\mathbb{P}}(\Omega;E)} \le \|g\|_{L^1_{\mathbb{P}}(\Omega;E)}, \qquad \forall\, g \in L^1_{\mathbb{P}}(\Omega;E). \tag{D.1}$$

Given a Borel function $g : \Omega \to E$ such that $g \in L^1_{\mathbb{P}}(\Omega;E)$, we define $g^n := P_n g$ and we prove that the properties of Definition 4.15 hold.

Property **(i)** is obvious since $g^n$ is constant on the elements of the partition $\mathcal{P}^n \subset \mathfrak{B}^n$.

Property **(ii)** follows from the fact that (see for instance [26, Corollary 8, p. 48])

$$\fint_A g(\omega)\,\mathrm{d}\mathbb{P}(\omega) \in \overline{\mathrm{co}}\,(g(A))\,, \quad \forall\, A \in \mathcal{B}_{(\Omega,\tau)} : \mathbb{P}(A) > 0.$$

In order to prove property **(iii)** we start with the particular case $g = a\mathbb{1}_A$ for a given Borel set $A$ and a given $a \in E$. Since $(\Omega,\tau)$ is Polish, for any $\varepsilon > 0$ there exists an open set $A_\varepsilon$ and a compact set $K_\varepsilon$ such that $K_\varepsilon \subset A \subset A_\varepsilon$ and $\mathbb{P}(A_\varepsilon \setminus K_\varepsilon) < \varepsilon$. Since $A_\varepsilon$ is union of elements of the basis $\mathcal{A}$, there exists a finite covering of $K_\varepsilon$ of the form $\{B^j : j \in J\} \subset \mathcal{A}$, for a suitable finite $J \subset \mathbb{N}$, such that $\cup_{j\in J} B^j \subset A_\varepsilon$. Since

$$P_n g - g = a \sum_{k=1}^{k(n)} \frac{\mathbb{P}(A_k^n \cap A)}{\mathbb{P}(A_k^n)} \mathbb{1}_{A_k^n} - a\mathbb{1}_A,$$

by setting $n_\varepsilon := \max J$, it holds that

$$\|P_n g - g\|_{L^1_{\mathbb{P}}(\Omega;E)} \le \|a\mathbb{1}_{A_\varepsilon} - a\mathbb{1}_A\|_{L^1_{\mathbb{P}}(\Omega;E)} \le \|a\|_E\, \mathbb{P}(A_\varepsilon \setminus A) < \|a\|_E\, \varepsilon\,, \quad \forall\, n \ge n_\varepsilon\,.$$

Since $P_n$ is linear, then **(iii)** holds for any $g$ simple function. In the general case, take $g \in L^1_{\mathbb{P}}(\Omega;E)$ and $\varepsilon > 0$, and let $g_\varepsilon : \Omega \to E$ be a simple function such that $\|g - g_\varepsilon\|_{L^1_{\mathbb{P}}(\Omega;E)} < \varepsilon$. Observing that

$$\|P_n g - g\|_{L^1_{\mathbb{P}}(\Omega;E)} \le \|P_n g - P_n g_\varepsilon\|_{L^1_{\mathbb{P}}(\Omega;E)} + \|P_n g_\varepsilon - g_\varepsilon\|_{L^1_{\mathbb{P}}(\Omega;E)} + \|g_\varepsilon - g\|_{L^1_{\mathbb{P}}(\Omega;E)},$$

by (D.1) and property **(iii)** applied to $g_\varepsilon$ it holds that $\limsup_{n\to+\infty} \|P_n g - g\|_{L^1_{\mathbb{P}}(\Omega;E)} \le 2\varepsilon$ and we conclude.

Finally, property **(iv)** follows from the measurability of $G$, Fubini Theorem and the definition of $P_n$. □

Consider now the Polish space $([0,1], \mathcal{B}, \mathcal{L}_1)$.

**Lemma D.3** (Proposition 4.16, part 2)**.** *For any $N \in \mathbb{N}$ we define $I_k^N := [(k-1)/N, k/N)$, $k = 1, \dots, N-1$, $I_N^N := [(N-1)/N, 1]$. If $E$ is a Banach space, $g \in L^1([0,1];E)$ and*

$$g^N := \sum_{k=1}^{N} \mathbb{1}_{I_k^N} \fint_{I_k^N} g(s)\,\mathrm{d}s,$$

*then*

$$\lim_{N\to+\infty} \|g^N - g\|_{L^1([0,1];E)} = 0. \tag{D.2}$$

*Moreover, the family of finite algebras $\mathcal{B}^N := \sigma(\{I_k^N : k = 1, \dots, N\})$, $N \in \mathbb{N}$, satisfies the finite approximation property of Definition 4.15.*
*Finally, if $g \in L^p([0,1];E)$, for some $p \in (1,+\infty)$, then $\lim_{N\to+\infty} \|g^N - g\|_{L^p([0,1];E)} = 0$.*

*Proof.* For any $x \in [0,1]$ and $N \in \mathbb{N}$, there exists a unique $k(x,N)$ such that $x \in I_{k(x,N)}^N$. From the definition of $I_k^N$ it follows that $I_{k(x,N)}^N \subset B_{1/N}(x)$. Since

$$g^N(x) - g(x) = \sum_{k=1}^{N} \mathbb{1}_{I_k^N}(x) \fint_{I_k^N} (g(s) - g(x))\,\mathrm{d}s,$$

then

$$\|g^N(x) - g(x)\|_E \leq 2 \fint_{B_{1/N}(x)} \|g(s) - g(x)\|_E \, \mathrm{d}s.$$

By the Bochner version of the Lebesgue differentiation Theorem (see for instance [26, Theorem 9, p. 49]) we obtain that $\lim_{N\to+\infty} \|g^N(x) - g(x)\|_E = 0$ for $\mathcal{L}_1$-a.e. $x \in [0,1]$.

Since $\int_0^1 \|g(x)\|_E \, \mathrm{d}x < +\infty$ there exists a convex, increasing, superlinear function $\psi : [0,+\infty) \to [0,+\infty)$ such that $\int_0^1 \psi\left(\|g(x)\|_E\right) \, \mathrm{d}x < +\infty$. Since

$$\|g^N(x)\|_E \leq 2 \fint_{B_{1/N}(x)} \|g(s)\|_E \, \mathrm{d}s,$$

by Jensen's inequality,

$$\begin{aligned}
\int_0^1 \psi\left(\|g^N(x)\|_E\right) \, \mathrm{d}x &\leq 2 \int_0^1 \psi\left( \fint_{B_{1/N}(x)} \|g(s)\|_E \, \mathrm{d}s \right) \mathrm{d}x \\
&\leq 2 \int_0^1 \fint_{B_{1/N}(x)} \psi\left(\|g(s)\|_E\right) \, \mathrm{d}s \, \mathrm{d}x \\
&= N \int_0^1 \int_0^1 \mathbb{1}_{(x-1/N, x+1/N)}(s) \psi\left(\|g(s)\|_E\right) \, \mathrm{d}s \, \mathrm{d}x \\
&= N \int_0^1 \int_0^1 \mathbb{1}_{(s-1/N, s+1/N)}(x) \psi\left(\|g(s)\|_E\right) \, \mathrm{d}s \, \mathrm{d}x \\
&\leq 2 \int_0^1 \psi\left(\|g(s)\|_E\right) \, \mathrm{d}s < +\infty,
\end{aligned}$$

which implies the equi-integrability of the sequence $\|g^N\|_E$. Then (D.2) holds.

The finite approximation property for $\mathcal{B}^N$ follows as in the proof of Lemma D.2. The final assertion is a consequence of the equi-integrability of the sequence $\|g^N\|_E^p$. □

**Lemma D.4** (Proposition 4.16, part 3)**.** *Let $(\Omega, \mathfrak{B}, \mathbb{P})$ be a standard Borel space and $\mathbb{P}$ without atoms. Then there exists a family $\mathfrak{B}^N \subset \mathfrak{B}$, $N \in \mathbb{N}$, satisfying the finite approximation property of Definition 4.15 such that the associated minimal partition $\mathcal{P}^N = \{A_k^N : k = 1, \dots, N\}$ contains exactly $N$ elements and $\mathbb{P}(A_k^N) = \frac{1}{N}$, for $k = 1, \dots, N$.*

*Proof.* Let $\tau$ be a Polish topology on $\Omega$ such that $\mathfrak{B} = \mathcal{B}_{(\Omega,\tau)}$. Let also $\Omega_0$, $I_0$, $\psi$, $\psi^{-1}$ be given by Proposition D.1. Using the notation of Lemma D.3 we define the sets $A_1^N := \psi^{-1}(I_1^N \setminus I_0) \cup \Omega_0$ and $A_k^N := \psi^{-1}(I_k^N \setminus I_0)$ for $k = 2, \dots, N$.

It is immediate to prove that $\mathbb{P}(A_j^N) = \frac{1}{N}$ for $j = 1, \dots, N$ and $\{A_1^N, \dots, A_N^N\}$ is a partition of $\Omega$. Moreover, given a Banach space $E$ and $g \in L^1_{\mathbb{P}}(\Omega; E)$, we denote by $\tilde{g} := g \circ \psi^{-1} \in L^1([0,1]; E)$. Denoting by $\tilde{g}^N$ the sequence given by Lemma D.3 applied to $\tilde{g}$, we define $g^N := \tilde{g}^N \circ \psi$ and the finite approximation property for $(\Omega, \mathcal{B}_{(\Omega,\tau)}, \mathbb{P})$ follows by Lemma D.3.

□

## Acknowledgements

The authors acknowledge the support of MIUR-PRIN 2017 project *Gradient flows, Optimal Transport and Metric Measure Structures.* G. Cavagnari, S. Lisini and C. Orrieri acknowledge the support of the INDAM-GNAMPA project 2019 "Trasporto ottimo per dinamiche con interazione". C. Orrieri has also been supported by the project Fondazione Cariplo-Regione Lombardia MEGAsTAR "Matematica d'Eccellenza in biologia ed ingegneria come acceleratore di una nuova strateGia per l'ATtRattività dell'ateneo pavese". G. Savaré gratefully acknowledges the support of the Institute of Advanced Study or the Technical University of Munich and of IMATI-CNR, Pavia. G. Cavagnari and G. Savaré are also grateful to the Department of Mathematics of the University of Pavia where this project has been developed.

## References


[1] Luigi Ambrosio and Alessio Figalli, *Geodesics in the space of measure-preserving maps and plans*, Arch. Ration. Mech. Anal. **194** (2009), no. 2, 421–462. MR2563635

[2] Luigi Ambrosio and Wilfred Gangbo, *Hamiltonian ODEs in the Wasserstein space of probability measures*, Comm. Pure Appl. Math. **61** (2008), no. 1, 18–53. MR2361303

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Second, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR2401600

[4] Viorel Barbu, *Nonlinear differential equations of monotone types in Banach spaces*, Springer Monographs in Mathematics, Springer, New York, 2010. MR2582280

[5] Jean-David Benamou, Guillaume Carlier, and Filippo Santambrogio, *Variational mean field games*, Active particles. Vol. 1. Advances in theory, models, and applications, 2017, pp. 141–171. MR3644590

[6] Alain Bensoussan, Jens Frehse, and Phillip Yam, *Mean field games and mean field type control theory*, SpringerBriefs in Mathematics, Springer, New York, 2013. MR3134900

[7] Patrick Bernard, *Young measures, superposition and transport*, Indiana Univ. Math. J. **57** (2008), no. 1, 247–275. MR2400257

[8] Patrizia Berti, Luca Pratelli, and Pietro Rigo, *Skorohod representation on a given probability space*, Probab. Theory Related Fields **137** (2007), no. 3-4, 277–288. MR2278458

[9] Patrick Billingsley, *Convergence of probability measures*, Second, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1999. A Wiley-Interscience Publication. MR1700749

[10] Benoît Bonnet and Hélène Frankowska, *Mean-field optimal control of continuity equations and differential inclusions*, arXiv preprint arXiv:2009.06467 (2020).

[11] Benoît Bonnet and Francesco Rossi, *The Pontryagin maximum principle in the Wasserstein space*, Calc. Var. Partial Differential Equations **58** (2019), no. 1, Paper No. 11, 36. MR3881882

[12] Amarjit Budhiraja and Michael Conroy, *Asymptotic behavior of stochastic currents under large deviation scaling with mean field interaction and vanishing noise*, arXiv preprint arXiv:2007.12956 (2020).

[13] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions, *The master equation and the convergence problem in mean field games*, Annals of Mathematics Studies, vol. 201, Princeton University Press, Princeton, NJ, 2019. MR3967062

[14] Pierre Cardaliaguet and Marc Quincampoix, *Deterministic differential games under probability knowledge of initial condition*, Int. Game Theory Rev. **10** (2008), no. 1, 1–16. MR2423798

[15] Pierre Cardaliaguet and Catherine Rainer, *On the (in)efficiency of MFG equilibria*, SIAM J. Control Optim. **57** (2019), no. 4, 2292–2314. MR3978479

[16] René Carmona and François Delarue, *Probabilistic theory of mean field games with applications. I*, Probability Theory and Stochastic Modelling, vol. 83, Springer, Cham, 2018. Mean field FBSDEs, control, and games. MR3752669

[17] René Carmona, François Delarue, and Aimé Lachapelle, *Control of McKean-Vlasov dynamics versus mean field games*, Math. Financ. Econ. **7** (2013), no. 2, 131–166. MR3045029

[18] José A. Carrillo, Edgard A. Pimentel, and Vardan K. Voskanyan, *On a mean field optimal control problem*, Nonlinear Anal. **199** (2020), 112039, 14. MR4118403

[19] José Antonio Carrillo, Young-Pil Choi, and Maxime Hauray, *The derivation of swarming models: mean-field limit and Wasserstein distances*, Collective dynamics from bacteria to crowds, 2014, pp. 1–46. MR3331178

[20] Charles Castaing, Paul Raynaud de Fitte, and Michel Valadier, *Young measures on topological spaces*, Mathematics and its Applications, vol. 571, Kluwer Academic Publishers, Dordrecht, 2004. With applications in control theory and probability theory. MR2102261

[21] Giulia Cavagnari, Antonio Marigonda, and Benedetto Piccoli, *Generalized dynamic programming principle and sparse mean-field control problems*, J. Math. Anal. Appl. **481** (2020), no. 1, 123437, 45. MR4008538

[22] Annalisa Cesaroni and Marco Cirant, 1-*dimensional multi-agent optimal control with aggregation and distance constraints: qualitative properties and mean-field limit*, arXiv preprint arXiv:2003.01980 (2020).

[23] Andrea Cosso and Huyên Pham, *Zero-sum stochastic differential games of generalized McKean-Vlasov type*, J. Math. Pures Appl. (9) **129** (2019), 180–212. MR3998794

[24] Emiliano Cristiani, Benedetto Piccoli, and Andrea Tosin, *Multiscale modeling of pedestrian dynamics*, MS&A. Modeling, Simulation and Applications, vol. 12, Springer, Cham, 2014. MR3308728

[25] Marco Di Francesco and Massimiliano D. Rosini, *Rigorous derivation of nonlinear scalar conservation laws from follow-the-leader type models via many particle limit*, Arch. Ration. Mech. Anal. **217** (2015), no. 3, 831–871. MR3356989

[26] Joseph Diestel and John Jerry Uhl, *Vector measures*, American Mathematical Society, Providence, R.I., 1977. With a foreword by B. J. Pettis, Mathematical Surveys, No. 15. MR0453964

[27] Fabrice Mao Djete, Dylan Possamaï, and Xiaolu Tan, *Mckean-vlasov optimal control: limit theory and equivalence between different formulations*, arXiv preprint arXiv:2001.00925 (2020).

[28] Mao Fabrice Djete, *Extended mean field control problem: a propagation of chaos result*, arXiv preprint arXiv:2006.12996 (2020).

[29] Mao Fabrice Djete, Dylan Possamaï, and Xiaolu Tan, *Mckean-vlasov optimal control: the dynamic programming principle*, arXiv preprint arXiv:1907.08860 (2019).

[30] James Dugundji, *An extension of Tietze's theorem*, Pacific J. Math. **1** (1951), 353–367. MR0044116

[31] Lawrence C. Evans and Ronald F. Gariepy, *Measure theory and fine properties of functions*, Revised, Textbooks in Mathematics, CRC Press, Boca Raton, FL, 2015. MR3409135

[32] Markus Fischer, *On the connection between symmetric $N$-player games and mean field games*, Ann. Appl. Probab. **27** (2017), no. 2, 757–810. MR3655853

[33] Markus Fischer and Francisco J Silva, *On the asymptotic nature of first order mean field games*, Applied Mathematics & Optimization (2020), 1–31.

[34] Massimo Fornasier, Stefano Lisini, Carlo Orrieri, and Giuseppe Savaré, *Mean-field optimal control as gamma-limit of finite agent controls*, European J. Appl. Math. **30** (2019), no. 6, 1153–1186. MR4028474

[35] Massimo Fornasier and Francesco Solombrino, *Mean-field optimal control*, ESAIM Control Optim. Calc. Var. **20** (2014), no. 4, 1123–1152. MR3264236

[36] Wilfrid Gangbo and Andrzej Święch, *Metric viscosity solutions of Hamilton-Jacobi equations depending on local slopes*, Calc. Var. Partial Differential Equations **54** (2015), no. 1, 1183–1218. MR3385197

[37] Wilfrid Gangbo and Adrian Tudorascu, *On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations*, J. Math. Pures Appl. (9) **125** (2019), 119–174. MR3944201

[38] Minyi Huang, Roland P. Malhamé, and Peter E. Caines, *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Commun. Inf. Syst. **6** (2006), no. 3, 221–251. MR2346927

[39] Chloé Jimenez, Antonio Marigonda, and Marc Quincampoix, *Optimal control of multiagent systems in the Wasserstein space*, Calc. Var. Partial Differential Equations **59** (2020), no. 2, Paper No. 58, 45. MR4073204

[40] Daniel Lacker, *Limit theory for controlled McKean-Vlasov dynamics*, SIAM J. Control Optim. **55** (2017), no. 3, 1641–1672. MR3654119

[41] ———, *On the convergence of closed-loop Nash equilibria to the mean field game limit*, Ann. Appl. Probab. **30** (2020), no. 4, 1693–1761. MR4133381

[42] Jean-Michel Lasry and Pierre-Louis Lions, *Mean field games*, Jpn. J. Math. **2** (2007), no. 1, 229–260. MR2295621

[43] Mauro Mariani, *A $\Gamma$-convergence approach to large deviations*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **18** (2018), no. 3, 951–976. MR3807592

[44] Giovanni Naldi, Lorenzo Pareschi, and Giuseppe Toscani (eds.), *Mathematical modeling of collective behavior in socio-economic and life sciences*, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser Boston, Ltd., Boston, MA, 2010. MR2761862

[45] Carlo Orrieri, *Large deviations for interacting particle systems: joint mean-field and small-noise limit*, Electron. J. Probab. (2020).

[46] Carlo Orrieri, Alessio Porretta, and Giuseppe Savaré, *A variational approach to the mean field planning problem*, Journal of Functional Analysis **277** (2019), no. 6, 1868–1957.

[47] Benedetto Piccoli, Francesco Rossi, and Emmanuel Trélat, *Control to flocking of the kinetic Cucker-Smale model*, SIAM J. Math. Anal. **47** (2015), no. 6, 4685–4719. MR3432849

[48] Nikolay Pogodaev, *Optimal control of continuity equations*, NoDEA Nonlinear Differential Equations Appl. **23** (2016), no. 2, Art. 21, 24. MR3483165

[49] Aldo Pratelli, *On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation*, Ann. Inst. H. Poincaré Probab. Statist. **43** (2007), no. 1, 1–13. MR2288266

[50] Halsey L. Royden, *Real analysis, second edition*, The Macmillan Co., New York; Collier-Macmillan Ltd., London, 1968. MR0151555

[51] Filippo Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling. MR3409718

[52] Daniel W. Stroock, *Probability theory*, Second, Cambridge University Press, Cambridge, 2011. An analytic view. MR2760872

[53] Alain-Sol Sznitman, *Topics in propagation of chaos*, École d'Été de Probabilités de Saint-Flour XIX—1989, 1991, pp. 165–251. MR1108185

[54] Michel Valadier, *Young measures*, Methods of nonconvex analysis (Varenna, 1989), 1990, pp. 152–188. MR1079763

[55] Cédric Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR1964483

Giulia Cavagnari: Politecnico di Milano,
Dipartimento di Matematica "F. Brioschi"
Piazza Leonardo da Vinci, 32, I-20133 Milano, Italy.
*Email address*: giulia.cavagnari@polimi.it

Stefano Lisini: Università di Pavia,
Dipartimento di Matematica "F. Casorati"
Via Ferrata 5, I-27100 Pavia, Italy.
*Email address*: stefano.lisini@unipv.it

Carlo Orrieri: Università di Pavia,
Dipartimento di Matematica "F. Casorati"
Via Ferrata 5, I-27100 Pavia, Italy.
*Email address*: carlo.orrieri@unipv.it

Giuseppe Savaré: Università Bocconi,
Dipartimento di Scienze delle Decisioni
Via Roentgen 1, I-20136 Milano, Italy.
*Email address*: giuseppe.savare@unibocconi.it