HAL
open science

Differential Inclusions in Wasserstein Spaces: The Cauchy-Lipschitz Framework

Benoît Bonnet, Hélène Frankowska

To cite this version:

Benoît Bonnet, Hélène Frankowska. Differential Inclusions in Wasserstein Spaces: The Cauchy-Lipschitz Framework. Journal of Differential Equations, 2021, 271, pp.594-637. 10.1016/j.jde.2020.08.031 . hal-02906913

HAL Id: hal-02906913

https://hal.science/hal-02906913v1

Submitted on 26 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Manuscript

Differential Inclusions in Wasserstein Spaces: the Cauchy-Lipschitz Framework

Benoît Bonnet*, Hélène Frankowska†

July 9, 2020

Abstract

In this article, we propose a general framework for the study of differential inclusions in the Wasserstein space of probability measures. Based on earlier geometric insights on the structure of continuity equations, we define solutions of differential inclusions as absolutely continuous curves whose driving velocity fields are measurable selections of multifunction taking their values in the space of vector fields. In this general setting, we prove three of the founding results of the theory of differential inclusions: Filippov's theorem, the Relaxation theorem, and the compactness of the solution sets. These contributions – which are based on novel estimates on solutions of continuity equations – are then applied to derive a new existence result for fully non-linear mean-field optimal control problems with closed-loop controls.

Keywords : Continuity Equation, Differential Inclusion, Optimal Transport, Filippov Theorem, Relaxation, Compactness of Trajectories, Mean-Field Optimal Control.

MSC2010 Subject Classification : 28B20, 34A60, 34G20, 49J21, 49J45

1 Introduction

During the past decade, the study of large dynamical systems appearing in the modelling of social dynamics and network analysis has taken an increasingly important place in several mathematical communities. *Multi-agent systems* are ubiquitous in a wide number of applications, ranging from the understanding of patterns formation in several branches of the animal kingdom [14, 19, 33] to the analysis of pedestrian dynamics [2, 32], ensembles of autonomous vehicles [15, 26], and opinion formation on networks [16, 44, 50]. They are also at the core of active academic fields such as *mean-field games*, a theory formalised simultaneously in [46] and [49] which now constitutes a prominent topic in applied mathematics.

At the microscopic level, multi-agent systems are commonly modelled by a family of $N \geq 1$ point-trajectories $\boldsymbol{x}_N(\cdot) := (x_1(\cdot), \ldots, x_N(\cdot))$ in a given state-space (e.g. $\mathbb{R}^d$ or a smooth manifold), whose evolution is described by a system of coupled ordinary differential equations of the form

$$\dot{x}_i(t) = \boldsymbol{v}_N(t, \boldsymbol{x}_N(t), x_i(t)). \tag{1}$$

In this context, the velocity field $(t, x) \mapsto \boldsymbol{v}_N(t, \boldsymbol{x}_N(t), x)$ stirring each individual agent $x_i(\cdot)$ is *non-local*, in the sense that it depends on the total state $\boldsymbol{x}_N(t)$ of the system at each time. These types of dynamics frequently appear in the form of discrete convolutions, which are used to represent sums of bipartite interactions in the system (see e.g. [32, 33, 44]). It should also be noted that in several engineering-oriented communities, the analysis of multi-agent systems is carried out in a graph-theoretic framework (see e.g. [26, 51] and references therein).

The investigation of multi-agent systems from a control-theoretic perspective is fairly challenging for several reasons. Firstly, the high dimensionality of the problems often prevents the application

*CNRS, IMJ-PRG, UMR 7586, Sorbonne Université, 4 place Jussieu, 75252 Paris, France. *E-mail*: benoit.bonnet@imj-prg.fr (Corresponding author)

†CNRS, IMJ-PRG, UMR 7586, Sorbonne Université, 4 place Jussieu, 75252 Paris, France. *E-mail*: helene.frankowska@imj-prg.fr

of classical finite-dimensional control or optimisation methods. Secondly, designing individual control laws at the *microscopic* scale may not be relevant from a conceptual standpoint compared to the implementation of a *macroscopic* signal used to pilot the system as a whole. For these reasons, control problems for multi-agents systems of the form (1) are often studied in the so-called *mean-field approximation framework*. In this setting, the collection of individual agents $(x_1(\cdot), \dots, x_N(\cdot))$ is replaced by a density $\mu(\cdot)$, whose evolution is described by a *non-local continuity equation* of the form

$$\partial_t \mu(t) + \operatorname{div}\big(v(t, \mu(t))\mu(t)\big) = 0. \tag{2}$$

Here, the velocity field $(t, x) \mapsto v(t, \mu(t))(x)$ is the mean-field pendant of $(t, x) \mapsto \boldsymbol{v}_N(t, \boldsymbol{x}_N(t), x)$, and depends on the whole density $\mu(t)$ at each time. Equations of the form (2) arise very naturally when studying mean-field limits of deterministic particle systems, see e.g. the pioneering work [61].

During the last few years, an important research effort at the interface between control theory and calculus of variations has been directed towards control problems formulated on continuity equations of the form (2). While a few results have been dealing with controllability properties of continuity equations [36, 37], the major part of the literature has been devoted to the study of optimal control problems, with contributions ranging from existence results [23, 38, 39, 40] and necessary optimality conditions [1, 21, 22, 24, 27, 29, 31, 56] to numerical methods [28, 57]. All these findings have hugely benefited from theoretical progresses made in the theory of *optimal transport*, for which we refer largely to the reference monographs [10, 58, 59].

It is now well-understood that continuity equations play a key role in the geometric study of the so-called *Wasserstein spaces* $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ of optimal transport (see Definition 3 below). It was noticed as early as [17] that they are involved in the dynamical formulation of the optimal transport problem, in which one aims at computing Wasserstein geodesics by searching for curves of minimal length joining two prescribed measures. The fact that these equations represent the "good" class of intrinsic dynamics to work with in Wasserstein spaces was further confirmed in [10, Chapter 8], where it is proven that absolutely continuous curves of measures coincide with the solutions of (2) driven by integrable velocity fields $(t, x) \mapsto v(t, x)$ (which are independent from $\mu(\cdot)$). In the particular case where $p = 2$, the approach allowing to prove this important result also provides an explicit construction of the so-called *analytical tangent space* $\operatorname{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ to the *manifold* of measures $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ (see [10, Section 8.4]). This far-reaching characterisation contributed to building a solid mathematical basis for the pseudo-Riemannian structure of $(\mathscr{P}_2(\mathbb{R}^d), W_2)$, which was first explored in [52].

In addition to these geometric considerations, an extensive literature has been focusing on the well-posedness theory for continuity equations, predominantly for velocity fields $v(\cdot, \cdot)$ that do not depend on the density $\mu(\cdot)$. The classical Cauchy-Lipschitz framework (which we further detail in Section 2.2 below) was first extended to Sobolev vector fields in [34], and later to BV vector fields in [3]. Several other relevant classes of vector fields have been considered since then, such as velocity fields of bounded deformation or with a Hamiltonian structure (see e.g. the survey [4]). More recently, a comprehensive study of "local" solutions in the spirit of the classical Carathéodory theory for ODEs has been presented in [5] for Sobolev and BV vector fields. We also mention the very recent preprint [48] in which flow solutions inspired by [20] are built for continuity equations with very rough velocities by adequate combinations of Filippov regularisations and measurable selection principles. Concerning non-local continuity equations of the form (2), some well-posedness results were first derived in [8] for Hamiltonian flows in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$, and a first Cauchy-Lipschitz theory was subsequently elaborated in [55]. Besides, a new formalism of *measure-driven differential equations* in the spirit of the theory of Young measures (see e.g. the seminal contribution [18]) was proposed in [53] for continuity equations.

In this paper, we introduce differential inclusions in Wasserstein spaces – which are set-valued extensions of the *non-local continuity equation* (2) – in the Cauchy-Lipschitz framework. Differential inclusions are an active area of research in the setting of finite and infinite dimensional vector spaces. Indeed, generalised multivalued ordinary differential equations of the form

$$\dot{\sigma}(t) \in F(t, \sigma(t)),$$

where $F : [0, T] \times \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is a *set-valued map* (see Section 2.3 below) appeared in the literature as early as 1936. Since the beginning of the 60's, mathematical tools developed for differential inclusions

have been successfully exploited in several branches of control theory, as it was observed that under mild assumptions, control systems could be seen as particular cases of differential inclusions. Starting from there, existence results for optimal controls were deduced from the compactness theorems on the sets of trajectories of differential inclusions, while the Relaxation theorem allowed to describe the closure of trajectories of a general non-linear control system as the set of all solutions of the differential inclusion whose right-hand side is the convexified control system. Let us also stress that if the right-hand side $F(\cdot,\cdot)$ has closed convex values, then parametrisation theorems allow to rewrite the corresponding differential inclusion as a control system. We refer to [12, 13] for these results and for historical and bibliographical comments on the theory of differential inclusions.

Similar studies were also performed for evolution inclusions in general Banach spaces, which led to the derivation of necessary optimality conditions in the form of an infinite dimensional maximum principle for optimal control problems involving pointwise state constraints, see [42]. However, the context of Wasserstein spaces considered in this paper is substantially different. On the one hand, only a purely metric structure is available in this setting, and on the other hand the corresponding class of dynamics represented by non-local continuity equations produces highly non-linear semigroups which are much more delicate to handle. Besides, even though the space of probability measures can be seen as a subset of the Banach space of Radon measures, the induced metric and the corresponding duality are generically not explicit enough to formulate many of the important results of control theory.

The first step towards the definition of differential inclusions for continuity equations is to identify the object which plays the role of a multi-valued velocity field in this context. The answer to this question is provided by the Riemannian analogy that we have sketched previously, and which can be used to give another meaning to continuity equations in terms of *ordinary differential equations* (see e.g. [43, Chapter 3]). Given an arbitrary non-local velocity field

$$(t,\mu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^d) \mapsto v(t,\mu) \in L^2(\mathbb{R}^d,\mathbb{R}^d;\mu), \tag{3}$$

the continuity equation (2) can be heuristically rewritten as

$$\dot{\mu}(t) = -\mathrm{div}\big(v(t,\mu(t))\mu(t)\big), \tag{4}$$

and be seen as an ODE in the flat distribution space $(C_c^\infty(\mathbb{R}^d))'$, whose velocity is the right-hand side of (4). However, it can also be interpreted as a differential equation formulated in the abstract manifold of measures $(\mathscr{P}_2(\mathbb{R}^d), W_2)$. In this analogy, the object which plays the role of the velocity of the curve $\mu(\cdot)$ is the application $t \in [0,T] \mapsto \pi_{\mu(t)}(v(t,\mu(t))$, where $\pi_\mu : L^2(\mathbb{R}^d,\mathbb{R}^d;\mu) \to \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ denotes the orthogonal projection onto the analytical tangent space. This interpretation – along with several other facts listed hereinabove – suggests that in order to turn the "differential equation" (4) into a "differential inclusion", the object which needs to become set-valued is the mapping (3).

In the present work, we therefore propose a *functional approach* to Cauchy-Lipschitz differential inclusions in Wasserstein spaces. Namely, we consider set-valued maps $V : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ with values in a subset of locally Lipschitz mappings from $\mathbb{R}^d$ into itself, and say that a curve of measures $\mu(\cdot)$ is a solution of the *differential inclusion*

$$\partial_t \mu(t) \in -\mathrm{div}\Big(V(t,\mu(t))\mu(t)\Big),$$

if there exists a *measurable selection* $t \in [0,T] \mapsto v(t) \in V(t,\mu(t))$ (see Definition 5 below) such that

$$\partial_t \mu(t) + \mathrm{div}\big(v(t)\mu(t)\big) = 0. \tag{5}$$

This setting is very convenient for investigating solutions of (2) with control-dependent velocities, whose characteristics are described by control systems of the form

$$\dot{\sigma}(t) = v(t,\mu(t),u(t))(\sigma(t)),$$

where $u : [0,T] \to U$ is a Lebesgue-measurable control, U is a subset of a metric space and $v : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \times U \to C^0(\mathbb{R}^d,\mathbb{R}^d)$ is a control-dependent velocity. Indeed, setting

$$V(t,\mu) \ := \ \Big\{ v(t,\mu,u) \in C^0(\mathbb{R}^d,\mathbb{R}^d) \text{ s.t. } u \in U \Big\},$$

and assuming that $x \in \mathbb{R}^d \mapsto v(t,\mu,u)(x) \in \mathbb{R}^d$ is locally Lipschitz with a constant independent from (μ, u), the set $V(t,\mu)$ consists of locally Lipschitz mappings. Furthermore, if $t \in [0,T] \mapsto \boldsymbol{v}(t) \in V(t,\mu(t))$ is Lebesgue-measurable, then by classical measurable selection theorems, there exists a control $u : [0,T] \to U$ such that $\boldsymbol{v}(t) := v(t,\mu(t),u(t))$ for $\mathscr{L}^1$-almost every $t \in [0,T]$. In the other words, given a measurable selection $t \in [0,T] \mapsto \boldsymbol{v}(t) \in V(t,\mu(t))$, we can always associate to it a control $u(\cdot)$, which is then used to write the characteristic system. The functional approach to differential inclusions in Wasserstein spaces that we propose here can be applied in particular to closed-loop controls, i.e. when U is a set of locally Lipschitz functions from $\mathbb{R}^d$ into a metric space, see Section 4 for more details.

Recently, differential inclusions in Wasserstein spaces have attracted the attention of several researchers. For instance in [54], a notion of weak solutions is introduced for the inclusion $\dot{\mu}(t) \in \mathbb{V}(\mu(t))$ where $\mathbb{V} : \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows \mathscr{P}_c(T\mathbb{R}^d)$ is a set-valued map with values in the space of probability measures over the tangent bundle $T\mathbb{R}^d \simeq \mathbb{R}^d \times \mathbb{R}^d$. The existence of weak solutions is then investigated by following ideas akin to [53] under a convexity-type requirement on the right-hand side of the dynamics, which is a delicate notion to handle in a metric space. In the present manuscript, we consider the differential inclusion $\partial_t \mu(t) \in -\mathrm{div}(V(\mu(t))\mu(t))$, which is the direct generalisation of the continuity equation (2). Furthermore in our case, the admissible velocities $V(\mu)$ are subsets of the vector space $C^0(\mathbb{R}^d, \mathbb{R}^d)$, which automatically lifts convexity-related issues.

We would also like to stress the difference between the theory developed in the present paper and the approach followed e.g. in [29, 30, 47], where the authors investigated the existence of optimal solutions and Hamilton-Jacobi-Bellman equations associated to optimal control problems in Wasserstein spaces. In these control problems, the minimisation is taken over the set of curves solving continuity equations of the form (2), where $(t,x) \in [0,T] \times \mathbb{R}^d \to v(t,x) \in \mathbb{R}^d$ is a Borel mapping satisfying the non-holonomic constraint $v(t,x) \in F(\mu(t),x)$. Here $F : \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is a given set-valued map with convex compact images. In addition to the conceptual difference between the two approaches – which is partly due to the fact that the set-valued maps do not take values in the same spaces –, the set of admissible trajectories studied in these articles appears to be larger (in general) than the one considered in the present paper, since the admissible controls obtained by applying measurable selection principles depend on each of the characteristics of the inclusion. This particular point is discussed in greater details in Remark 6 below. Moreover, the functional approach to differential inclusions developed in this paper is very close in spirit to the formulation of gradient flows [10, Chapter 11] and Hamiltonian flows [8] in Wasserstein spaces, which correspond to special (but not less general) cases in which the sets of admissible velocities $V(\mu) \subset L^2(\mathbb{R}^d, \mathbb{R}^d; \mu)$ are defined using the Wasserstein subdifferentials at $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ of a given functional.

After introducing precisely our notion of solution to differential inclusions in Section 3.1, we will prove the generalisations to the Wasserstein spaces of three cornerstones of the theory of set-valued dynamical systems: *Filippov's theorem*, the *Relaxation theorem*, and the *compactness of the solution set*, for which we refer the reader e.g. to [12, Chapter 1], [13, Chapter 10] or [60, Chapter 2]. More precisely, Theorem 4 extends Filippov's Theorem for Cauchy-Lipschitz differential inclusions in $(\mathscr{P}_c(\mathbb{R}^d), W_p)$. For differential inclusions in a Banach space, this result is the most commonly used generalisation of the classical well-posedness theorems for Carathéodory ODEs. By construction, it also provides useful Grönwall-type inequalities on the trajectory-selection pairs. In our context, the latter are based on non-trivial estimates for solutions of continuity equations which are presented in Proposition 2 below. We proceed by proving in Theorem 5 a generalisation of the Relaxation theorem. Heuristically, this result asserts that any solution of a differential inclusion with a convexified righ-hand side can be approximated by solutions of the original differential inclusion. Relaxation results are very useful in cases where optimal trajectories may fail to exist, since they describe the closure of the solution set to differential inclusions (see e.g. [60, Section 2.7]). In Theorem 6, we complement this result by showing that the set of solutions of a differential inclusion in $(\mathscr{P}_c(\mathbb{R}^d), W_p)$ is compact in the topology of the uniform convergence, whenever the right-hand side of the inclusion is convex. This result is then applied in Theorem 7 to recover a general existence result for constrained mean-field optimal control problems.

The structure of the article is as follows. In Section 2, we recall several notions pertaining to optimal

transport theory, continuity equations, and set-valued analysis. In Section 3, we define differential inclusion in the Wasserstein space, and we prove our main results Theorem 4, Theorem 5 and Theorem 6. Finally in Section 4, we apply these new set-theoretic tools to show the existence of optimal controls for general fully non-linear and constrained mean-field optimal control problems. Appendix A contains the proofs of two new technical results, Lemma 1 and Proposition 2.

2 Preliminaries

We introduce here all the necessary tools needed to formulate differential inclusions in Wasserstein spaces and prove our main results Theorem 4, Theorem 5 and Theorem 6.

2.1 Analysis in measure spaces and optimal transport theory

In this section, we recall some classical notations and results of measure theory and optimal transport. We refer the reader to the monographs [7] and [10, 58, 59] respectively for a comprehensive introduction to these topics.

Let $(X, \|\cdot\|_X)$ be a separable Banach space. We denote by $\mathscr{P}(X)$ the space of *Borel probability measures* over X endowed with the narrow topology, i.e. the coarsest topology such that the application

$$\mu \in \mathscr{P}(X) \mapsto \int_X \phi(x)\mathrm{d}\mu(x) \in \mathbb{R}, \tag{6}$$

is continuous for every $\phi \in C_b^0(X)$. Here $C_b^0(X)$ is the set of continuous and bounded real-valued functions over X, and we denote by $\mu_n \rightharpoonup^* \mu$ the narrow convergence of measures induced by (6). By Riesz's theorem (see e.g. [10, Remark 5.1.2]), the space $\mathscr{P}(X)$ endowed with the narrow topology can be identified with a subset of the topological dual $(C_b^0(X))'$ of $(C_b^0(X), \|\cdot\|_{C^0})$, where $\|\cdot\|_{C^0}$ stands for the supremum norm. In the sequel given a metric space $(\mathscr{S}, d_{\mathscr{S}})$, we will use the notation $\mathrm{AC}([0,T], \mathscr{S})$ for the space of absolutely continuous arcs with values in $\mathscr{S}$ and $\mathrm{Lip}(\phi(\cdot)\,;\Omega)$ for the Lipschitz constant of a map $\phi : \mathscr{S} \to \mathbb{R}$ over $\Omega \subset \mathscr{S}$.

Given two separable Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, an element $p \in [1, +\infty]$ and a measure $\mu \in \mathscr{P}(X)$, we denote by $L^p(X, Y; \mu)$ and $W^{1,p}(X, Y; \mu)$ respectively the Lebesgue spaces of p-summable maps and Sobolev maps from X into Y. For $p \in [1, +\infty)$, we also define the *momentum of order* p of a measure $\mu \in \mathscr{P}(X)$ as

$$\mathcal{M}_p(\mu) := \left(\int_X |x|^p \mathrm{d}\mu(x)\right)^{1/p},$$

and consider the set $\mathscr{P}_p(X) \subset \mathscr{P}(X)$ of probability measures with finite momentum of order p, i.e.

$$\mathscr{P}_p(X) = \Big\{\mu \in \mathscr{P}(X) \text{ s.t. } \mathcal{M}_p(\mu) < +\infty\Big\}.$$

The *support* of a probability measure $\mu \in \mathscr{P}(X)$ is defined as the closed set

$$\mathrm{supp}(\mu) = \Big\{x \in X \text{ s.t. } \mu(\mathcal{N}_x) > 0 \text{ for any neighbourhood } \mathcal{N}_x \text{ of } x\Big\},$$

and we shall denote by $\mathscr{P}_c(X) \subset \mathscr{P}(X)$ the set of probability measures with compact support.

We recall next the classical notion of *pushforward* (or *image measure*) of a Borel probability measure through a Borel map, along with that of *transport plan.*

Definition 1 (Image of a measure through a Borel map)**.** *Given a measure $\mu \in \mathscr{P}(X)$ and a Borel map $f : X \to Y$, the* pushforward *$f_\#\mu$ of μ through $f(\cdot)$ is the unique Borel probability measure such that $f_\#\mu(B) = \mu(f^{-1}(B))$ for any Borel set $B \subset Y$.*

Definition 2 (Transport plans)**.** *Given $\mu, \nu \in \mathscr{P}(X)$, we say that $\gamma \in \mathscr{P}(X^2)$ is a* transport plan *between μ and ν – denoted by $\gamma \in \Gamma(\mu, \nu)$ –, provided that*

$$\pi^1_\#\gamma = \mu \qquad \text{and} \qquad \pi^2_\#\gamma = \nu,$$

where the maps $\pi^1, \pi^2 : X^2 \to X$ stand for the projection operators on the first and second factor.

We now recall the definition and some of the main properties of the *Wasserstein spaces* of optimal transport built over $X = \mathbb{R}^d$, for which we refer to [10, Chapter 7], [58, Chapter 5] or [59, Chapter 6]).

Definition 3 (Wasserstein spaces)**.** *Given $p \in [1, +\infty)$ and two probability measures $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$, the* Wasserstein distance of order p *between μ and ν is defined by*

$$W_p(\mu, \nu) = \min_{\gamma} \left\{ \left(\int_{\mathbb{R}^{2d}} |x - y|^p \mathrm{d}\gamma(x, y) \right)^{1/p} \quad s.t.\ \gamma \in \Gamma(\mu, \nu) \right\}.$$

*The set of plans $\gamma \in \Gamma(\mu, \nu)$ achieving this minimum is denoted by $\Gamma_o(\mu, \nu)$ and referred to as the set of p-*optimal transport plans *between μ and ν. The space $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ of probability measures with finite p-th moment endowed with the W_p-metric is called the* Wasserstein space *of order p.*

Proposition 1 (Properties of the Wasserstein spaces)**.** *For every $p \in [1, +\infty)$, the Wasserstein spaces $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ are complete and separable metric spaces. The topology induced by the W_p-metric metrises the narrow topology of probability measures induced by* (6)*, i.e.*

$$W_p(\mu_n, \mu) \underset{n \to +\infty}{\longrightarrow} 0 \qquad \textit{if and only if} \qquad \begin{cases} \mu_n \underset{n \to +\infty}{\rightharpoonup^*} \mu, \\ \displaystyle\int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu_n(x) \underset{n \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu(x). \end{cases}$$

Given two measures $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, the Wasserstein distances are ordered*, i.e. $W_{p_1}(\mu, \nu) \leq W_{p_2}(\mu, \nu)$ whenever $p_1 \leq p_2$. Moreover when $p = 1$, the following* Kantorovich-Rubinstein duality formula *holds*

$$W_1(\mu, \nu) = \sup_{\phi} \left\{ \int_{\mathbb{R}^d} \phi(x)\, \mathrm{d}(\mu - \nu)(x) \ \ s.t.\ \mathrm{Lip}(\phi\,; \mathbb{R}^d) \leq 1 \right\}. \tag{7}$$

2.2 Continuity equations in Wasserstein spaces

In this section, we recall some of the main definitions and classical results concerning continuity equations formulated in the space of measures. We also state several momentum and Grönwall-type estimates needed in the proofs of two of our main results detailed in Section 3

Let $T > 0$ and $\mathscr{L}^1$ be the standard one-dimensional Lebesgue measure on $[0, T]$. Throughout this paper, we shall always deal with *Carathéodory vector fields*, that is mappings $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ such that $t \mapsto v(t, x)$ is $\mathscr{L}^1$-measurable for all $x \in \mathbb{R}^d$ and $x \mapsto v(t, x)$ is continuous for $\mathscr{L}^1$-almost every $t \in [0, T]$. Moreover, we will always assume the following.

Hypothesis (C1). *There exists a map $m(\cdot) \in L^1([0, T], \mathbb{R}_+)$ such that*

$$|v(t, x)| \leq m(t)(1 + |x|),$$

for $\mathscr{L}^1$-almost every $t \in [0, T]$ and all $x \in \mathbb{R}^d$.

We say that a curve of measures $\mu(\cdot) \in C^0([0, T], \mathscr{P}(\mathbb{R}^d))$ solves a *continuity equation* driven by a velocity field $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ with initial condition $\mu^0 \in \mathscr{P}(\mathbb{R}^d)$ provided that

$$\begin{cases} \partial_t \mu(t) + \mathrm{div}(v(t)\mu(t)) = 0, \\ \mu(0) = \mu^0. \end{cases} \tag{8}$$

This equation has to be understood in the sense of distributions against smooth and compactly supported test functions, i.e.

$$\int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \phi(t, x) + \langle \nabla_x \phi(t, x), v(t, x) \rangle \Big) \mathrm{d}\mu(t)(x) \mathrm{d}t = 0, \tag{9}$$

for any $\phi \in C_c^\infty([0, T] \times \mathbb{R}^d)$.

In our subsequent developments, we will deal with two notions of solution for (8): *superposition solutions* and *Cauchy-Lipschitz solutions*. We shall henceforth denote by $\Sigma_T := C^0([0, T], \mathbb{R}^d)$ the space of continuous arcs in $\mathbb{R}^d$ and by $e_t : (x, \sigma) \in \mathbb{R}^d \times \Sigma_T \mapsto \sigma(t) \in \mathbb{R}^d$ the so-called *evaluation map*.

Definition 4 (Superposition measures and solutions)**.** *We say that* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ *is a* superposition measure *generated by* $v(\cdot,\cdot)$ *if it is concentrated on the pairs* $(x,\sigma) \in \mathbb{R}^d \times \mathrm{AC}([0,T],\mathbb{R}^d)$ *such that*

$$\sigma(0) = x \qquad \text{and} \qquad \dot{\sigma}(t) = v(t,\sigma(t)), \tag{10}$$

for $\mathscr{L}^1$*-almost every* $t \in [0,T]$*. We further say that a distributional solution* $\mu(\cdot) \in C^0([0,T],\mathscr{P}(\mathbb{R}^d))$ *of* (8) *is a* superposition solution *if there exists a superposition measure* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ *generated by* $v(\cdot,\cdot)$ *such that* $\mu(t) = (e_t)_\#\boldsymbol{\eta}$ *for all times* $t \in [0,T]$*.*

One can easily check that if a superposition measure $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ generated by $v(\cdot,\cdot)$ satisfies the local integrability bounds

$$\int_0^T \int_{\mathbb{R}^d \times \Sigma_T} \mathbb{1}_K(\sigma(t))|v(t,\sigma(t))|\mathrm{d}\boldsymbol{\eta}(x,\sigma)\mathrm{d}t < +\infty,$$

for any compact set $K \subset \mathbb{R}^d$, then the curve of measures $t \in [0,T] \mapsto \mu(t) = (e_t)_\#\boldsymbol{\eta}$ is a distributional solution of (8). In the following theorem, we recall the converse of this statement which is known as the *superposition principle*, and for which we refer the reader e.g. to [6, Theorem 3.4].

Theorem 1 (Superposition principle)**.** *Let* $\mu(\cdot) \in C^0([0,T],\mathscr{P}(\mathbb{R}^d))$ *be a distributional solution of* (8) *starting from* $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ *and driven by a Carathéodory vector field* $v(\cdot,\cdot)$ *which satisfies* **(C1)***. Then* $\mu(\cdot)$ *is a superposition solution, i.e. there exists a superposition measure* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ *generated by* $v(\cdot,\cdot)$ *in the sense of Definition 4 such that* $\mu(t) = (e_t)_\#\boldsymbol{\eta}$ *for all times* $t \in [0,T]$*.*

Remark 1 (On the statement of the superposition principle)**.** *In [6, Theorem 3.4] and several other references in the literature (see e.g. [5, 10]), the superposition principle is stated and proven for velocity fields* $v(\cdot,\cdot)$ *which are* Borel measurable *with respect to both variables. We would like to point out that this result also holds true whenever* $v(\cdot,\cdot)$ *is only* Lebesgue measurable *in its first variable.*

The superposition principle is a very powerful tool which is used to prove many stability and existence results on continuity equations. In the next theorem, we recall several classical facts which state that under **(C1)** together with the Cauchy-Lipschitz regularity assumptions **(C2)** below, superposition solutions are unique and enjoy uniform boundedness and regularity properties.

Hypothesis (C2). *For any compact set* $K \subset \mathbb{R}^d$*, there exists a map* $l_K(\cdot) \in L^1([0,T],\mathbb{R}_+)$ *such that* $\mathrm{Lip}(v(t,\cdot);K) \leq l_K(t)$ *for* $\mathscr{L}^1$*-almost every* $t \in [0,T]$*.*

In the sequel, we use the notation $B(0,r)$ for the closed ball of radius $r > 0$ centered at 0 in $\mathbb{R}^d$ and $\|\cdot\|_1 := \|\cdot\|_{L^1([0,T])}$ for the L^1-norm of a real-valued function on the interval $[0,T]$.

Theorem 2 (Carathéodory and Cauchy-Lipschitz solutions of (8))**.** *Let* $r > 0$*,* $\mu^0 \in \mathscr{P}(B(0,r))$ *and* $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ *be a Carathéodory vector field satisfying hypothesis* **(C1)***. Then, there exists a curve of measures* $\mu(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ *solution of* (8)*, and every such solution curve satisfies*

$$\mathrm{supp}(\mu(t)) \subseteq B(0,R_r) \qquad \text{and} \qquad W_p(\mu(t),\mu(s)) \leq \int_s^t m_r(\tau)\mathrm{d}\tau, \tag{11}$$

for all times $0 \leq s \leq t \leq T$ *and any* $p \in [1,+\infty)$*, where*

$$R_r = \Big(r + \|m(\cdot)\|_1\Big)\exp\left(\|m(\cdot)\|_1\right) \qquad \text{and} \qquad m_r(t) := (1+R_r)m(t),$$

for $\mathscr{L}^1$*-almost every* $t \in [0,T]$*. Furthermore if the velocity field* $v(\cdot,\cdot)$ *also satisfies hypothesis* **(C2)***, the solutions of* (8) *are unique.*

Proof. The proof of Theorem 2 follows from Theorem 1, together with classical boundedness and absolute continuity estimates on the flows of Carathéodory ODEs in the Cauchy-Lipschitz framework. For further details, see [6, Section 3]. □

We end this section by several new estimates that will be useful in the sequel. In Proposition 2, we present new momentum and Grönwall-type estimates, which to the best of our knowledge are not written at this degree of generality in the literature. The proof of these results along with those of several crucial steps of Theorem 5 are based on Lemma 1 below. The latter is inspired by [9, Section 2.2], and provides the existence of a transport plan $\hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma(\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu)$, whose pushforwards through the evaluation maps (e_t, e_t) are optimal transport plans between $\mu(t)$ and $\nu(t)$ for all times $t \in [0,T]$.

The arguments subtending these two results being somewhat technical, we postpone their proofs to Appendix A.

Lemma 1 (Superposition measures producing optimal plans)**.** *Let $K \subset \mathbb{R}^d$ be a compact set and $\mu(\cdot), \nu(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}(K))$ be two solutions of* (8) *driven respectively by velocity fields $v, w : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying hypothesis* **(C1)***. Let $\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ be two superposition measures given by Theorem 1, i.e. $\mu(t) = (e_t)_\# \boldsymbol{\eta}_\mu$ and $\nu(t) = (e_t)_\# \boldsymbol{\eta}_\nu$ for all times $t \in [0,T]$.*

Then for any $p \in [1, +\infty)$, there exists a transport plan $\hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma(\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu)$ such that for all times $t \in [0,T]$, it holds

$$(\pi_{\mathbb{R}^d}, \pi_{\mathbb{R}^d})_\# \hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma_o(\mu^0, \nu^0) \qquad \text{and} \qquad (e_t, e_t)_\# \hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma_o(\mu(t), \nu(t)), \tag{12}$$

where Γ_o stands for the set of p-optimal transport plans between two measures.

Proposition 2 (Momentum and Grönwall estimates)**.** *Let $K \subset \mathbb{R}^d$ be a compact set and $\mu(\cdot), \nu(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}(K))$ be two solutions of* (8) *driven respectively by velocity fields $v, w : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying hypothesis* **(C1)***. Furthermore, suppose that $v(\cdot,\cdot)$ also satisfies hypothesis* **(C2)***.*

Then for any $p \in [1, +\infty)$, we have

$$\mathcal{M}_p(\mu(t)) \leq C_p \left(\mathcal{M}_p(\mu^0) + \int_0^t m(s) \mathrm{d}s \right) \exp\left(C'_p \, \|m(\cdot)\|^p_{L^1([0,t])} \right), \tag{13}$$

and

$$W_p(\mu(t), \nu(t)) \leq C_p \left(W_p(\mu(0), \nu(0)) + \int_0^t \|v(s,\cdot) - w(s,\cdot)\|_{C^0(K,\mathbb{R}^d)} \, \mathrm{d}s \right) \exp\left(C'_p \, \|l_K(\cdot)\|^p_{L^1([0,t])} \right), \tag{14}$$

for all times $t \in [0,T]$, where the constants $C_p, C'_p > 0$ are defined by

$$C_p = 2^{(p-1)/p} \qquad \text{and} \qquad C'_p = \tfrac{2^{p-1}}{p}. \tag{15}$$

Remark 2 (A more general momentum estimate)**.** *In the proof of Theorem 4 below, we will need a more general variant of* (13) *for velocity fields $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying the sub-linearity estimate*

$$|v(t,x)| \leq m(t)\Big(1 + |x| + M(t)\Big),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and all $x \in \mathbb{R}^d$, where $m(\cdot) \in L^1([0,T], \mathbb{R}_+)$ and $M(\cdot) \in L^\infty([0,T], \mathbb{R}_+)$. In this case, one can show by repeating the arguments of the proof of (13) *in Appendix A that*

$$\mathcal{M}_p(\mu(t)) \leq C_p \left(\mathcal{M}_p(\mu^0) + \int_0^t m(s)\Big(1 + M(s)\Big) \mathrm{d}s \right) \exp\left(C'_p \, \|m(\cdot)\|^p_{L^1([0,t])} \right), \tag{16}$$

for all times $t \in [0,T]$, and that (14) *is still verified.*

2.3 Elements of set-valued analysis

In this section, we recall some notations and basic notions pertaining to set-valued analysis and multifunctions. We refer the reader to [13] for most of the results which are stated below, as well as for a general treatment of this topic.

Let $(\mathscr{S}, d_{\mathscr{S}})$ be a complete separable metric space and $(X, \|\cdot\|_X)$ be a separable Banach space. In the sequel, we denote by $\mathbb{B}_{\mathscr{S}}(s, \beta)$ the closed ball of center $s \in \mathscr{S}$ and radius $\beta > 0$ and by $\mathbb{B}_X$ the

closed unit ball in X. Given a subset $B \subset X$, we denote by $\overline{\mathrm{co}}(B)$ its *closed convex-hull* defined as the closure in X of the set of convex combinations

$$\bigcup_{N\geq 1}\Big\{\sum_{i=1}^{N}\alpha_i b_i \text{ s.t. } b_i\in B,\ \alpha_i\geq 0 \text{ for } i\in\{1,\dots,N\} \text{ and } \sum_{i=1}^{N}\alpha_i=1\Big\}.$$

We say that an application $\mathcal{F}:\mathscr{S}\rightrightarrows X$ is a *set-valued map* – or a *multifunction* – from $\mathscr{S}$ into X if $\mathcal{F}(s)\subset X$ for all $s\in\mathscr{S}$, and we define its *domain* by $\mathrm{dom}(\mathcal{F}):=\{s\in\mathscr{S} \text{ s.t. } \mathcal{F}(s)\neq\emptyset\}$. A multifunction $\mathcal{F}(\cdot)$ has *closed values* if $\mathcal{F}(s)$ is closed in X for any $s\in\mathrm{dom}(\mathcal{F})$, and we say that $\mathcal{F}(\cdot)$ itself is *closed* if its graph

$$\mathrm{Graph}(\mathcal{F}):=\Big\{(s,x) \text{ s.t. } x\in\mathcal{F}(s)\Big\}, \tag{17}$$

is a closed subset of $\mathscr{S}\times X$.

Definition 5 (Measurable set-valued maps and measurable selections)**.** *We say that a set-valued map $\mathcal{F}:[0,T]\rightrightarrows X$ with closed images is $\mathscr{L}^1$-*measurable *if the sets*

$$\mathcal{F}^{-1}(\mathcal{O}):=\Big\{t\in[0,T] \text{ s.t. } \mathcal{F}(t)\cap\mathcal{O}\neq\emptyset\Big\},$$

are $\mathscr{L}^1$-measurable for any open set $\mathcal{O}\subset X$. A single-valued map $f:[0,T]\to X$ is called a measurable selection *of $\mathcal{F}(\cdot)$ if it is $\mathscr{L}^1$-measurable and such that $f(t)\in\mathcal{F}(t)$ for $\mathscr{L}^1$-almost every $t\in[0,T]$.*

Definition 6 (Lipschitz regularity of set-valued maps)**.** *We say that a set-valued map $\mathcal{F}:\mathscr{S}\rightrightarrows X$ is L-*Lipschitz *around $s\in\mathscr{S}$ if there exists a neighbourhood $\mathcal{N}_s\subset\mathrm{dom}(\mathcal{F})$ of s such that*

$$\mathcal{F}(s_1)\subset\mathcal{F}(s_2)+Ld_{\mathscr{S}}(s_1,s_2)\,\mathbb{B}_X,$$

for every $s_1,s_2\in\mathcal{N}_s$.

In the following theorem, we recall a classical measurable selection principle for set-valued maps, see for instance [13, Theorem 8.1.3].

Theorem 3 (Measurable selection)**.** *Let $\mathcal{F}:[0,T]\rightrightarrows X$ be an $\mathscr{L}^1$-measurable set-valued map with non-empty and closed images. Then, $\mathcal{F}(\cdot)$ admits a measurable selection.*

Measurable selections are crucial for investigating solutions of differential inclusions and will therefore appear frequently in the proofs of our main results. However, we will primarily work with set-valued maps $\mathcal{F}(\cdot)$ with values in the linear space $C^0(\mathbb{R}^d,\mathbb{R}^d)$ which is not a Banach space. To circumvent this difficulty, we introduce in the following definition a concept of measurability adapted to this type of multifunction along with a suitable notion of closed convex hull for subsets of $C^0(\mathbb{R}^d,\mathbb{R}^d)$.

Definition 7 (Compact restrictions, measurability and convex hulls)**.** *Let $\mathcal{F}:[0,T]\rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ be a set-valued map. For every compact set $K\subset\mathbb{R}^d$, we define the* compact restriction *$\mathcal{F}_K:[0,T]\rightrightarrows C^0(K,\mathbb{R}^d)$ of $\mathcal{F}(\cdot)$ to K as*

$$\mathcal{F}_K(t):=\Big\{f_{|K}\in C^0(K,\mathbb{R}^d) \text{ s.t. } f(t)\in\mathcal{F}(t)\Big\},$$

for $\mathscr{L}^1$-almost every $t\in[0,T]$, where $f_{|K}$ denotes the restriction of the map $f\in C^0(\mathbb{R}^d,\mathbb{R}^d)$ to K. We then say that $\mathcal{F}:[0,T]\rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ is $\mathscr{L}^1$-measurable provided that $\mathcal{F}_K(\cdot)$ is $\mathscr{L}^1$-measurable for every compact set $K\subset\mathbb{R}^d$.

Analogously given a set $\mathcal{F}\subset C^0(\mathbb{R}^d,\mathbb{R}^d)$, we define its closed convex hull $\overline{\mathrm{co}}\,\mathcal{F}\subset C^0(\mathbb{R}^d,\mathbb{R}^d)$ by

$$\overline{\mathrm{co}}\,\mathcal{F}:=\Big\{f\in C^0(\mathbb{R}^d,\mathbb{R}^d) \text{ s.t. } f_{|K}\in\overline{\mathrm{co}}\,\mathcal{F}_K \text{ for every compact set } K\subset\mathbb{R}^d\Big\}, \tag{18}$$

where $\overline{\mathrm{co}}\,\mathcal{F}_K$ is taken in the Banach space $(C^0(K,\mathbb{R}^d),\|\cdot\|_{C^0})$.

Remark 3 (Link with the Whitney topology)**.** *Even though $C^0(\mathbb{R}^d, \mathbb{R}^d)$ is not a Banach space, it can be endowed with the so-called* Whitney topology *(see e.g. [45, Chapter 1]), where open sets are defined using bases of neighbourhoods involving localisations on compact sets. In this context, the notion of measurability introduced in Definition 7 coincides with that of measurability taken with respect to the Whitney topology in $C^0(\mathbb{R}^d, \mathbb{R}^d)$.*

We state in the following lemma two measurability results which are inspired by [41, Section 1], see also [13, Chapter 8] for more general statements.

Lemma 2 (Some measurability results)**.** *Let $\mathcal{F} : [0,T] \rightrightarrows X$ and $\mathcal{G} : [0,T] \times \mathscr{S} \rightrightarrows X$ be set-valued maps with closed non-empty images, and also $f : [0,T] \to X$ and $L : [0,T] \to \mathbb{R}_+$ be $\mathscr{L}^1$-measurable single-valued maps. Then, the following statements hold true.*

(a) *If $\mathcal{F}(\cdot)$ is $\mathscr{L}^1$-measurable and the set-valued map*

$$\mathcal{H} : t \in [0,T] \rightrightarrows \mathcal{F}(t) \cap \{f(t) + L(t)\mathbb{B}_X\} \subset X,$$

has non-empty images for $\mathscr{L}^1$-almost every $t \in [0,T]$, then $\mathcal{H}(\cdot)$ admits a measurable selection.

(b) *Let $s(\cdot) \in C^0([0,T], \mathscr{S})$ and $\beta > 0$ be a constant such that for $\mathscr{L}^1$-almost every $t \in [0,T]$, the multifunction $\mathcal{G}(t,\cdot)$ is $L(t)$-Lipschitz over $\mathbb{B}_{\mathscr{S}}(s(t),\beta)$. Then, the set-valued map*

$$t \in [0,T] \rightrightarrows \mathcal{G}(t, s(t)) \subset X,$$

is $\mathscr{L}^1$-measurable.

Proof. Statement (a) follows from [13, Corollary 8.2.13 and Theorem 8.2.4]. The same arguments as those of the proof of [13, Theorem 8.2.8] yield statement (b). □

We end these prerequisites by recalling the notion of *Aumann integral* for set-valued maps, and a generalisation of Aumann's theorem on the closure of the integral which is taken from [13, Theorem 8.6.4]. In the sequel, all the integrals are taken in the sense of Bochner (see e.g. [35, Chapter 2]).

Definition 8 (Integrably bounded multifunctions and Aumann integral)**.** *An $\mathscr{L}^1$-measurable set-valued map $\mathcal{F} : [0,T] \rightrightarrows X$ with closed images is said to be* integrably bounded *if there exists $k(\cdot) \in L^1([0,T], \mathbb{R}_+)$ such that for $\mathscr{L}^1$-almost every $t \in [0,T]$, it holds*

$$\mathcal{F}(t) \subset k(t)\mathbb{B}_X.$$

The Aumann integral *of $\mathcal{F}(\cdot)$ over a measurable subset $\Omega \subset [0,T]$ is then defined by*

$$\int_\Omega \mathcal{F}(t)\mathrm{d}t \;:=\; \left\{ \int_\Omega f(t)\mathrm{d}t \text{ s.t. } f(t) \in \mathcal{F}(t) \text{ for } \mathscr{L}^1\text{-almost every } t \in \Omega \right\}.$$

Lemma 3 (Closure of the Aumann integral)**.** *Let $\mathcal{F} : [0,T] \rightrightarrows X$ be an $\mathscr{L}^1$-measurable and integrably bounded set-valued map. Then for any measurable set $\Omega \subset [0,T]$, it holds*

$$\int_\Omega \overline{\mathrm{co}}\, \mathcal{F}(t)\mathrm{d}t = \overline{\int_\Omega \mathcal{F}(t)\mathrm{d}t}.$$

In particular, for any measurable selection $f(\cdot)$ from $\overline{\mathrm{co}}\, \mathcal{F}(\cdot)$ and any $\delta > 0$, there exists a measurable selection $f_\delta(\cdot)$ in $\mathcal{F}(\cdot)$ such that

$$\left\| \int_\Omega f(t)\mathrm{d}t - \int_\Omega f_\delta(t)\,\mathrm{d}t \right\|_X \leq \delta.$$

3 Differential Inclusions in Wasserstein Spaces

In this section, we move on to the main object of this article which are differential inclusions in Wasserstein spaces.

3.1 Definition of differential inclusions in Wasserstein spaces

In what follows, we state our definition of differential inclusions in Wasserstein spaces and study some of the main properties of their solutions. We recall that our intention is to build an adequate set-valued generalisation of non-local continuity equations of the form

$$\partial_t \mu(t) + \operatorname{div}\big(v(t,\mu(t))\mu(t)\big) = 0. \tag{19}$$

Motivated by the discussion presented in the Introduction, we propose the following notion of differential inclusion in Wasserstein spaces inspired by [13, Chapter 10] and [43, Section 3].

Definition 9 (Differential inclusions in Wasserstein spaces)**.** *Let $T > 0$ and $V : (t,\mu) \in [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ be a set-valued map. We say that a curve of measures $\mu(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ is a solution of the* Wasserstein differential inclusion

$$\partial_t \mu(t) \in -\operatorname{div}\Big(V(t,\mu(t))\mu(t)\Big), \tag{20}$$

if there exists a measurable selection $t \in [0,T] \mapsto v(t) \in V(t,\mu(t))$ in the sense of Definition 7 such that the trajectory-selection pair $(\mu(\cdot), v(\cdot))$ *solves the continuity equation*

$$\partial_t \mu(t) + \operatorname{div}\big(v(t)\mu(t)\big) = 0,$$

in the sense of distributions.

We are now ready to state our working assumptions for the rest of this section. From now on, we fix a time horizon $T > 0$ and a real number $p \in [1,+\infty)$.

Hypotheses (DI). *For every $R > 0$, assume that the following holds with $K := B(0,R)$.*

(i) For any $\mu \in \mathscr{P}_c(\mathbb{R}^d)$, the set-valued map $t \in [0,T] \rightrightarrows V_K(t,\mu)$ is $\mathscr{L}^1$-measurable with closed non-empty images in $C^0(K,\mathbb{R}^d)$.

(ii) There exists a map $m(\cdot) \in L^1([0,T],\mathbb{R}_+)$ such that for $\mathscr{L}^1$-almost every $t \in [0,T]$, for any $\mu \in \mathscr{P}_c(\mathbb{R}^d)$, for every $v \in V(t,\mu)$ and all $x \in \mathbb{R}^d$, it holds

$$|v(x)| \le m(t)\Big(1 + |x| + \mathcal{M}_1(\mu)\Big).$$

(iii) There exists a map $l_K(\cdot) \in L^1([0,T],\mathbb{R}_+)$ such that for $\mathscr{L}^1$-almost every $t \in [0,T]$, any $\mu \in \mathscr{P}(K)$ and every $v \in V(t,\mu)$, it holds

$$\operatorname{Lip}(v(\cdot)\,;K) \le l_K(t).$$

(iv) There exists a map $L_K(\cdot) \in L^1([0,T],\mathbb{R}_+)$ such that for $\mathscr{L}^1$-almost every $t \in [0,T]$ and any $\mu,\nu \in \mathscr{P}(K)$, it holds

$$V_K(t,\nu) \subset V_K(t,\mu) + L_K(t) W_p(\mu,\nu)\mathbb{B}_{C^0(K,\mathbb{R}^d)}.$$

These assumptions are a localised extension of those e.g. from [55], and seem rather minimal for the elaboration of a Cauchy-Lipschitz theory of differential inclusions. We end these preliminaries by providing support and regularity estimates on the solutions of (20).

Proposition 3 (Support and regularity estimates on solutions of (20))**.** *Let $V : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ be a set-valued map satisfying hypothesis* **(DI)**-*(ii). Then for any $r > 0$, there exists a constant $R_r > 0$ and a map $m_r(\cdot) \in L^1([0,T],\mathbb{R}_+)$ such that any solution $\mu(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ of (20) starting from $\mu^0 \in \mathscr{P}(B(0,r))$ satisfies*

$$\operatorname{supp}(\mu(t)) \subset K := B(0,R_r) \qquad \text{and} \qquad W_p(\mu(t),\mu(s)) \le \int_s^t m_r(\tau)\mathrm{d}\tau, \tag{21}$$

for all times $0 \le s \le t \le T$ and any $p \in [1,+\infty)$.

Proof. Let $\mu(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}_c(\mathbb{R}^d))$ be a solution of (20) starting from $\mu^0 \in \mathscr{P}(B(0,r))$. By Definition 9, there exists a measurable selection $t \in [0,T] \mapsto v(t) \in V(t, \mu(t))$ such that

$$\begin{cases} \partial_t \mu(t) + \mathrm{div}(v(t)\mu(t)) = 0, \\ \mu(0) = \mu^0. \end{cases}$$

By construction, $v(\cdot,\cdot)$ is a Carathéodory vector field which satisfies the sub-linearity estimate

$$|v(t,x)| \leq m(t)\Big(1 + |x| + \mathcal{M}_1(\mu(t))\Big), \tag{22}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and all $x \in \mathbb{R}^d$ as a consequence of hypothesis **(DI)**-(ii). By repeating the argument e.g. of [38, Lemma 5.1], one can show that

$$\mathcal{M}_1(\mu(t)) \leq \left(r + \int_0^t m(s)\mathrm{d}s\right) \exp\left(2\, \|m(\cdot)\|_{L^1([0,t])}\right),$$

for all times $t \in [0,T]$. Therefore, the velocity field $v(\cdot,\cdot)$ satisfies the uniform sub-linearity estimate

$$|v(t,x)| \leq (1 + M_r) m(t)(1 + |x|),$$

where $M_r := (r + \|m(\cdot)\|_1) \exp(2\, \|m(\cdot)\|_1)$, and by Theorem 2 there exists a constant $R_r > 0$ and a map $m_r(\cdot) \in L^1([0,T], \mathbb{R}_+)$ such that (21) holds. □

3.2 Filippov Theorem

In this section, we state and prove a natural generalisation of Filippov's Theorem for Wasserstein differential inclusions formulated in $(\mathscr{P}_c(\mathbb{R}^d), W_p)$.

Theorem 4. *Let $T > 0$, $p \in [1, +\infty)$ and $V : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d, \mathbb{R}^d)$ be a set-valued map satisfying hypotheses* **(DI)**. *Let $\nu(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}(K_\nu))$ be a solution of* (8) *induced by a velocity field $w : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying* **(C1)**, *where $K_\nu := B(0, R_\nu)$ is a closed ball. Furthermore, suppose that the* mismatch function *$\eta_\nu : [0,T] \to \mathbb{R}_+$ defined by*

$$\eta_\nu(t) := \mathrm{dist}_{C^0(K_\nu, \mathbb{R}^d)}\Big(w_{|K_\nu}(t), V_{K_\nu}(t, \nu(t))\Big), \tag{23}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ is integrable over $[0,T]$.

Then for any $r > 0$ and every measure $\mu^0 \in \mathscr{P}(B(0,r))$, there exists $K := B(0, R_r)$ depending only on the magnitude of $r, R_\nu, T, \|m(\cdot)\|_1$ and a trajectory-selection pair $(\mu(\cdot), v(\cdot)) \in \mathrm{AC}([0,T], \mathscr{P}(K)) \times L^1([0,T], C^0(K, \mathbb{R}^d))$ solution of the inclusion

$$\begin{cases} \partial_t \mu(t) \in -\mathrm{div}\Big(V(t, \mu(t))\mu(t)\Big), \\ \mu(0) = \mu^0, \end{cases} \tag{24}$$

such that

$$W_p(\mu(t), \nu(t)) \leq \chi_p(t)\, \exp\left(C_{K,p}(t)\right), \tag{25}$$

for all times $t \in [0,T]$. Moreover, it holds for $\mathscr{L}^1$-almost every $t \in [0,T]$ that

$$\|v(t,\cdot) - \hat{w}(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} \leq L_K(t)\, \chi_p(t) \exp\left(C_{K,p}(t)\right) + \eta_\nu(t), \tag{26}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$, where $\hat{w}(t,x) := w(t, \pi_{K_\nu}(x))$ with π_{K_ν} the projection onto K_ν, and

$$\begin{cases} \chi_p(t) := C_p \left(W_p(\mu^0, \nu(0)) + \int_0^t \eta_\nu(s)\mathrm{d}s\right) \exp\left(C'_p\, \|l_K(\cdot)\|^p_{L^1([0,t])}\right), \\ C_{K,p}(t) := C_p \left(\int_0^t L_K(s)\mathrm{d}s\right) \exp\left(C'_p\, \|l_K(\cdot)\|^p_{L^1([0,t])}\right), \end{cases} \tag{27}$$

for all times $t \in [0,T]$, the constants $C_p, C'_p > 0$ being defined as in (15)

Before moving on to the proof of Theorem 4, we state a technical lemma dealing with chained integral estimates and which proof is the matter of an elementary induction argument.

Lemma 4 (A uniform estimate for chained integral inequalities)**.** *Let $m(\cdot) \in L^1([0,T],\mathbb{R}_+)$ and $\alpha > 0$ be a constant. Let $N \in \mathbb{N} \cup \{+\infty\}$ and $(f_n(\cdot))_{0\leq n<N} \subset C^0([0,T],\mathbb{R}_+)$ be a family of maps such that*

$$f_{n+1}(t) \leq \alpha \left(1 + \int_0^t m(s) f_n(s) \mathrm{d}s\right), \tag{28}$$

for all $t \in [0,T]$ and every $n \geq 0$ such that $n+1 < N$. Then, there exists a constant $C > 0$ depending only on the magnitudes of $\|f_0(\cdot)\|_{C^0([0,T])}, \alpha, T$ and $\|m(\cdot)\|_1$ such that

$$\sup_{0\leq n<N} \|f_n(\cdot)\|_{C^0([0,T])} \leq C. \tag{29}$$

The proof of Theorem 4 is based on an iterative scheme "à la Picard" in the spirit of the classical proof of Filippov's Theorem (see e.g. [60, Theorem 2.3.13]), and is split into four steps. In Step 1, we initialise our sequence of approximations and list some of the properties of its first element. We then show in Step 2 how the whole sequence of trajectory-selection pairs can be built, and prove in Step 3 that it is a Cauchy sequence in $C^0([0,T],\mathscr{P}(K)) \times L^1([0,T],C^0(K,\mathbb{R}^d))$. Finally in Step 4, we show that the corresponding limit trajectory-selection pair is a solution of the differential inclusion (24) satisfying (25) and (26).

Proof (of Theorem 4). To simplify the computations, we will restrict our attention to the case $p = 1$. The general case is similar and can be recovered from minor variations of the following arguments.

Fix $r > 0$ and $\mu^0 \in \mathscr{P}(B(0,r))$. Our goal is to build a closed ball $K := B(0,R_r) \subset \mathbb{R}^d$ along with a sequence of pairs $(\mu_n(\cdot), v_n(\cdot)) \subset \mathrm{AC}([0,T],\mathscr{P}(K)) \times L^1([0,T],C^0(K,\mathbb{R}^d))$ solutions of

$$\begin{cases} \partial_t \mu_n(t) + \mathrm{div}(v_n(t)\mu_n(t)) = 0, \\ \mu_n(0) = \mu^0, \end{cases} \tag{30}$$

which satisfy

$$\begin{cases} v_{n+1}(t) \in V(t,\mu_n(t)), \\ \|v_{n+1}(t,\cdot) - v_n(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} \leq L_K(t) W_1(\mu_n(t),\mu_{n-1}(t)), \\ \mathcal{M}_1(\mu_n(t)) \leq C, \end{cases} \tag{31}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and all $n \geq 1$, where $C > 0$ is a uniform constant.

Step 1: Initialisation of the sequence. We set $\mu^0(\cdot) := \nu(\cdot)$ and $v_0(\cdot,\cdot) := \hat{w}(\cdot,\cdot)$. Remark first that as a consequence of hypotheses **(DI)**-(i), **(DI)**-(iv) and Lemma 2-(b), the set-valued map

$$t \in [0,T] \rightrightarrows V_{K_\nu}(t,\nu(t)) \subset C^0(K_\nu,\mathbb{R}^d),$$

is $\mathscr{L}^1$-measurable. Moreover as a by-product of hypotheses **(DI)**-(ii), **(DI)**-(iii) and of the Ascoli-Arzelà Theorem, the sets $V_{K_\nu}(t,\nu(t))$ are compact in the separable Banach space $C^0(K_\nu,\mathbb{R}^d)$ for $\mathscr{L}^1$-almost every $t \in [0,T]$. Whence, the multifunction

$$t \in [0,T] \rightrightarrows V_{K_\nu}(t,\nu(t)) \cap \left\{ w_{|K_\nu}(t) + \eta_\nu(t)\mathbb{B}_{C^0(K_\nu,\mathbb{R}^d)} \right\},$$

is measurable with compact and non-empty images for $\mathscr{L}^1$-almost every $t \in [0,T]$. We can thus apply Lemma 2-(a) to find a measurable selection $t \in [0,T] \mapsto v_1(t) \in V_{K_\nu}(t,\nu(t))$ such that

$$\|v_1(t,\cdot) - w_{|K_\nu}(t,\cdot)\|_{C^0(K_\nu,\mathbb{R}^d)} = \eta_\nu(t), \tag{32}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$.

Since $v_1(t) \in V_{K_\nu}(t,\nu(t))$ for $\mathscr{L}^1$-almost every $t \in [0,T]$, one has by **(DI)**-(ii) and **(DI)**-(iii)

$$|v_1(t,x)| \leq m(t)\Big(1 + |x| + \mathcal{M}_1(\nu(t))\Big) \qquad \text{and} \qquad \mathrm{Lip}(v_1(t,\cdot)\,;K_\nu) \leq l_{K_\nu}(t), \tag{33}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and any $x \in K_\nu$. Let $\pi_{K_\nu} : \mathbb{R}^d \to K_\nu$ denote the projection operator onto K_ν, and observe that

$$|\pi_{K_\nu}(x)| \leq |x| \qquad \text{and} \qquad |\pi_{K_\nu}(y) - \pi_{K_\nu}(x)| \leq |y - x|, \tag{34}$$

for any $x, y \in \mathbb{R}^d$. We can therefore define an extension[1] of $v_1(t,\cdot)$ to $\mathbb{R}^d$ by

$$v_1(t,\cdot) : x \in \mathbb{R}^d \mapsto v_1\big(t, \pi_{K_\nu}(x)\big),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. One can easily check using (34) that this extension satisfies

$$|v_1(t,x)| \leq m(t)\Big(1 + |x| + \mathcal{M}_1(\nu(t))\Big) \qquad \text{and} \qquad \mathrm{Lip}(v_1(t,\cdot)\,;\mathbb{R}^d) \leq l_{K_\nu}(t),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and all $x \in \mathbb{R}^d$. Whence, the extended velocity field $v_1(\cdot,\cdot)$ satisfies the hypotheses **(C1)-(C2)** of Theorem 2, and it induces a unique solution $\mu_1(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}_c(\mathbb{R}^d))$ of

$$\begin{cases} \partial_t \mu_1(t) + \mathrm{div}(v_1(t)\mu_1(t)) = 0, \\ \mu_1(0) = \mu^0. \end{cases} \tag{35}$$

We now list some of the properties of the pair $(\mu_1(\cdot), v_1(\cdot))$ solution of the Cauchy problem (35). First, we have as a consequence of (16) in Remark 2 applied with $p = 1$ that the curve $\mu_1(\cdot)$ satisfies the momentum estimate

$$\mathcal{M}_1(\mu_1(t)) \leq \left(\mathcal{M}_1(\mu^0) + \int_0^t m(s)\Big(1 + \mathcal{M}_1(\nu(s))\Big)\mathrm{d}s\right) \exp\left(\int_0^t m(s)\mathrm{d}s\right).$$

We now set $f_0(t) := \mathcal{M}_1(\nu(t))$ and $f_1(t) := \mathcal{M}_1(\mu_1(t))$ for all $t \in [0,T]$, and let $C > 0$ be as in Lemma 4 applied with $\alpha = (1 + \mathcal{M}_1(\mu^0) + \|m(\cdot)\|_1)\exp(\|m(\cdot)\|_1)$ and $N = 2$. Then, it holds[2]

$$\max\Big\{\|\mathcal{M}_1(\nu(\cdot))\|_{C^0([0,T])}\ ,\ \|\mathcal{M}_1(\mu_1(\cdot))\|_{C^0([0,T])}\Big\} \leq C. \tag{36}$$

This implies in particular that $v_1(\cdot,\cdot)$ satisfies the uniform sub-linearity estimate

$$|v_1(t,x)| \leq (1 + C)m(t)\big(1 + |x|\big),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and every $x \in \mathbb{R}^d$. Thus by (11) in Theorem 2, there exist $R_r > 0$ and $m_r(\cdot) \in L^1([0,T], \mathbb{R}_+)$ depending on the magnitude of r, R_ν, T, C and $\|m(\cdot)\|_1$ such that

$$\mathrm{supp}(\nu(t)) \cup \mathrm{supp}(\mu_1(t)) \subseteq K := B(0, R_r) \qquad \text{and} \qquad W_1(\mu_1(t), \mu_1(s)) \leq \int_s^t m_r(\tau)\mathrm{d}\tau, \tag{37}$$

for all times $0 \leq s \leq t \leq T$. By applying the estimate (14) of Proposition 2 with $p = 1$, we also recover

$$\begin{aligned} W_1(\mu_1(t), \nu(t)) &\leq \left(W_1(\mu^0, \nu(0)) + \int_0^t \|v_1(s,\cdot) - \hat{w}(s,\cdot)\|_{C^0(K,\mathbb{R}^d)}\,\mathrm{d}s\right) \exp\left(\int_0^t l_K(s)\mathrm{d}s\right) \\ &= \chi_1(t), \end{aligned}$$

for all $t \in [0,T]$, were we used (32), the expression of the map $\chi_1(\cdot)$ given in (27), and the fact that

$$\|v_1(t,\cdot) - \hat{w}(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} = \|v_1(t,\cdot) - w_{|K_\nu}(t,\cdot)\|_{C^0(K_\nu,\mathbb{R}^d)} = \eta_\nu(t), \tag{38}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$.

[1] For simplicity, we use the same notation for the original vector field and for its extension.

[2] Notice that C depends only on the magnitude of r, R_ν, T and $\|m(\cdot)\|_1$ and not on the curve $\nu(\cdot)$ itself.

Step 2: Building the whole sequence $(\mu_n(\cdot), v_n(\cdot))$**.** By the support inclusion of (37) together with hypothesis **(DI)**-(iv), there exists a map $L_K(\cdot) \in L^1([0,T], \mathbb{R}_+)$ such that

$$V_K(t, \mu_1(t)) \subset V_K(t, \nu(t)) + L_K(t) W_1(\mu_1(t), \nu(t)) \mathbb{B}_{C^0(K,\mathbb{R}^d)}, \tag{39}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. This along with hypothesis **(DI)**-(i) and an application of Lemma 2-(a) and (b) yields the existence of a measurable selection $t \in [0,T] \mapsto v_2(t) \in V_K(t, \mu_1(t))$ such that

$$\|v_2(t,\cdot) - v_1(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} \leq L_K(t) W_1(\mu_1(t), \nu(t)),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. As a consequence of the hypothesis **(DI)**-(ii), **(DI)**-(iii) and (36), one further has

$$|v_2(t,x)| \leq (1+C) m(t)(1+|x|) \qquad \text{and} \qquad \mathrm{Lip}(v_2(t,\cdot)\,; K) \leq l_K(t),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and any $x \in K$. By repeating the same extension argument as in Step 1 with $\pi_K(\cdot)$, we have that $v_2(\cdot,\cdot)$ satisfies the assumptions **(C1)**-**(C2)** of Theorem 2. Whence, there exists a unique solution $\mu_2(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}_c(\mathbb{R}^d))$ of the Cauchy problem (35) driven by $v_2(\cdot,\cdot)$.

By a direct application of (16) in Remark 2 and (14) of Proposition 2, we furthermore have that the curve $\mu_2(\cdot)$ satisfies the momentum and distance estimates

$$\begin{cases} \mathcal{M}_1(\mu_2(t)) \leq \left(\mathcal{M}_1(\mu^0) + \int_0^t m(s)\Big(1 + \mathcal{M}_1(\mu_1(s))\Big) \mathrm{d}s \right) \exp\left(\int_0^t m(s) \mathrm{d}s \right), \\ W_1(\mu_2(t), \mu_1(t)) \leq \left(\int_0^t \|v_2(s,\cdot) - v_1(s,\cdot)\|_{C^0(K,\mathbb{R}^d)} \, \mathrm{d}s \right) \exp\left(\int_0^t l_K(s) \mathrm{d}s \right), \end{cases}$$

for all times $t \in [0,T]$. Thus by Lemma 4 applied with $N = 3$ and the constant $\alpha > 0$ defined as before, the curve $\mu_2(\cdot)$ also satisfies the uniform momentum bound

$$\|\mathcal{M}_1(\mu_2(\cdot))\|_{C^0([0,T])} \leq C.$$

By repeating this process, one can build a sequence of pairs $(\mu_n(\cdot), v_n(\cdot)) \subset \mathrm{AC}([0,T], \mathscr{P}(K)) \times L^1([0,T], C^0(K,\mathbb{R}^d))$ satisfying (30)-(31) along with the uniform estimates

$$\mathrm{supp}(\mu_n(t)) \subset K := B(0, R_r) \qquad \text{and} \qquad W_1(\mu_n(t), \mu_n(s)) \leq \int_s^t m_r(\tau) \mathrm{d}\tau, \tag{40}$$

for all times $0 \leq s \leq t \leq T$ and any $n \geq 1$.

Step 3: Convergence of the sequence $(\mu_n(\cdot), v_n(\cdot))$**.** The next step in our argument is to show that the sequence $(\mu_n(\cdot), v_n(\cdot))$ is a Cauchy sequence in the complete metric space $C^0([0,T], \mathscr{P}(K)) \times L^1([0,T], C^0(K,\mathbb{R}^d)$. By applying (14), we have for all $t \in [0,T]$ and any $n \geq 1$

$$\begin{aligned}
& W_1(\mu_{n+1}(t), \mu_n(t)) \\
\leq & \left(\int_0^t \|v_{n+1}(s_n,\cdot) - v_n(s_n,\cdot)\|_{C^0(K,\mathbb{R}^d)} \, \mathrm{d}s_n \right) \exp\left(\int_0^t l_K(s) \mathrm{d}s \right) \\
\leq & \left(\int_0^t L_K(s_n) W_1(\mu_n(s_n), \mu_{n-1}(s_n)) \mathrm{d}s_n \right) \exp\left(\int_0^t l_K(s) \mathrm{d}s \right) \\
\leq & \left(\int_0^t L_K(s_n) \int_0^{s_n} \|v_n(s_{n-1},\cdot) - v_{n-1}(s_{n-1},\cdot)\|_{C^0(K,\mathbb{R}^d)} \, \mathrm{d}s_{n-1} \mathrm{d}s_n \right) \exp\left(2 \int_0^t l_K(s) \mathrm{d}s \right) \\
& \vdots \\
\leq & \left(\int_0^t L_K(s_n) \int_0^{s_n} L_K(s_{n-1}) \ldots \int_0^{s_2} L_K(s_1) W_1(\mu_1(s_1), \nu(s_1)) \mathrm{d}s_1 \ldots \mathrm{d}s_{n-1} \mathrm{d}s_n \right) \exp\left(n \int_0^t l_K(s) \mathrm{d}s \right) \\
\leq & \frac{\chi_1(t)}{n!} \left(\int_0^t L_K(s) \mathrm{d}s \right)^n \exp\left(n \int_0^t l_K(s) \mathrm{d}s \right) \\
= & \frac{\chi_1(t)}{n!} C_{K,1}(t)^n,
\end{aligned} \tag{41}$$

where we recall that $\chi_1(\cdot)$ and $C_{K,1}(\cdot)$ are as in (27) with $p=1$. Whence for any $m,n\geq 1$, it holds

$$\begin{aligned}\sup_{t\in[0,T]} W_1(\mu_{n+m}(t),\mu_n(t)) &\leq \sup_{t\in[0,T]} \sum_{k=n}^{n+m-1} W_1(\mu_{k+1}(t),\mu_k(t))\\ &\leq \sup_{t\in[0,T]}\left[\chi_1(t)\sum_{k=n}^{n+m-1}\frac{C_{K,1}(t)^k}{k!}\right] \underset{m,n\to+\infty}{\longrightarrow} 0.\end{aligned}$$

Therefore, $(\mu_n(\cdot))$ is a Cauchy sequence in the complete metric space $C^0([0,T],\mathscr{P}(K))$, and it converges uniformly towards a limit curve $\mu(\cdot)\in C^0([0,T],\mathscr{P}(K))$.

Similarly, we can show that $(v_n(\cdot))$ is a Cauchy sequence in $L^1([0,T],C^0(K,\mathbb{R}^d))$. Indeed for any $m,n\geq 1$, we have by using the second line of (31) that

$$\begin{aligned}\int_0^T \|v_{n+m}(t,\cdot)-v_n(t,\cdot)\|_{C^0(K,\mathbb{R}^d)}\,\mathrm{d}t &\leq \sum_{k=n}^{m+n-1}\int_0^T \|v_{k+1}(t,\cdot)-v_k(t,\cdot)\|_{C^0(K,\mathbb{R}^d)}\,\mathrm{d}t\\ &\leq \sum_{k=n}^{m+n-1}\int_0^T L_K(t)W_1(\mu_k(t),\mu_{k-1}(t))\mathrm{d}t\\ &\leq \|L_K(\cdot)\|_1 \sup_{t\in[0,T]}\left[\chi_1(t)\sum_{k=n-1}^{m+n-2}\frac{C_{K,1}(t)^k}{k!}\right] \underset{m,n\to+\infty}{\longrightarrow} 0.\end{aligned}$$

Whence, the sequence $(v_n(\cdot))\subset L^1([0,T],C^0(K,\mathbb{R}^d))$ is Cauchy and converges towards a limit map $v(\cdot)\in L^1([0,T],C^0(K,\mathbb{R}^d))$. Moreover, one can easily prove as a consequence of hypotheses **(DI)**-(ii) and **(DI)**-(iii) that the limit velocity field satisfies the estimates

$$|v(t,x)|\leq (1+C)m(t)(1+|x|) \qquad \text{and} \qquad \mathrm{Lip}(v(t,\cdot)\,;K)\leq l_K(t), \tag{42}$$

for $\mathscr{L}^1$-almost every $t\in[0,T]$ and every $x\in K$.

Step 4 : Properties of the limit pair $(\mu(\cdot),v(\cdot))$. As a consequence of classical stability results for continuity equations (see e.g. [40] or Section 4 below), the limit pair $(\mu(\cdot),v(\cdot))$ is a distributional solution of the Cauchy problem

$$\begin{cases}\partial_t\mu(t)+\mathrm{div}(v(t)\mu(t))=0,\\ \mu(0)=\mu^0.\end{cases} \tag{43}$$

Moreover by (42), the velocity field $v:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ satisfies up to an extension argument the hypotheses **(C1)**-**(C2)** of Theorem 2. Thus, the curve of measures $\mu(\cdot)\in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ is the unique solution of (43). By taking the limit as $n\to+\infty$ in (40), we further obtain that

$$\mathrm{supp}(\mu(t))\subset K:=B(0,R_r) \qquad \text{and} \qquad W_1(\mu(t),\mu(s))\leq \int_s^t m_r(\tau)\mathrm{d}\tau,$$

for all times $0\leq s\leq t\leq T$.

We now want to prove that $(\mu(\cdot),v(\cdot))$ is a trajectory-selection pair of the differential inclusion (24). We start by observing that the first line of (31) can be restated as

$$\Big(\mu_n(t),v_{n+1}(t)\Big)\in \mathscr{V}_K(t) \tag{44}$$

for $\mathscr{L}^1$-almost every $t\in[0,T]$ and any $n\geq 1$. Here, the sets $\mathscr{V}_K(t)$ are the graphs of the set-valued maps $\mu\in\mathscr{P}(K)\rightrightarrows V_K(t,\mu)$, defined for $\mathscr{L}^1$-almost every $t\in[0,T]$ by

$$\mathscr{V}_K(t):=\Big\{(\mu,v)\in\mathscr{P}(K)\times C^0(K,\mathbb{R}^d)\text{ s.t. } v\in V_K(t,\mu)\Big\}.$$

Observe that it follows from hypothesis **(DI)**-(iv) and elementary properties of set-valued maps that the multifunction $\mu \in \mathscr{P}(K) \rightrightarrows V_K(t,\mu)$ is closed, so that the sets $\mathscr{V}_K(t)$ are closed in the $W_1 \times C^0(K,\mathbb{R}^d)$-topology for $\mathscr{L}^1$-almost every $t \in [0,T]$. As a consequence of Step 3, we further have that

$$W_1(\mu(t),\mu_n(t)) \underset{n\to+\infty}{\longrightarrow} 0,$$

for all times $t \in [0,T]$, as well as

$$\|v(t,\cdot) - v_n(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} \underset{n\to+\infty}{\longrightarrow} 0,$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ along an adequate subsequence that we do not relabel. Therefore, taking the limit as $n \to +\infty$ in (44), we recover for $\mathscr{L}^1$-almost every $t \in [0,T]$ the inclusion

$$v(t) \in V_K(t,\mu(t)). \tag{45}$$

Observe now that (43) and (45) together imply that $(\mu(\cdot), v(\cdot))$ is a trajectory-selection pair of the differential inclusion (24) in the sense of Definition 9.

We end the proof of Theorem 4 by deriving the velocity and trajectory estimates (25)-(26). First, remark that for any $n \geq 1$ and all times $t \in [0,T]$, it holds

$$W_1(\mu_n(t),\nu(t)) \leq \sum_{k=0}^{n-1} W_1(\mu_{k+1}(t),\mu_k(t)) \leq \chi_1(t) \sum_{k=0}^{n-1} \frac{C_{K,1}(t)^k}{k!} \leq \chi_1(t)\, \exp\left(C_{K,1}(t)\right),$$

where we used (41). Taking the limit as $n \to +\infty$ in the previous expression, we recover (25). Concerning the velocity estimate, we have for any $n \geq 1$ that

$$\begin{aligned}\|v_n(t,\cdot) - \hat{w}(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} &\leq \sum_{k=1}^{n-1} \|v_{k+1}(t,\cdot) - v_k(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} + \|v_1(t,\cdot) - \hat{w}(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} \\ &\leq L_K(t)\chi_1(t)\exp\left(C_{K,1}(t)\right) + \eta_\nu(t),\end{aligned}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$, where we used the second line in (31), (38) and (41). Taking again the limit as $n \to +\infty$ along a suitable subsequence in the previous expression, we recover (26). □

3.3 Relaxation Theorem

In this section, we state and prove a measure theoretic pendant of the Relaxation Theorem in the Wasserstein space $(\mathscr{P}_c(\mathbb{R}^d), W_p)$.

Theorem 5. *Let $T > 0$, $p \in [1,+\infty)$, $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$, $V : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ be a set-valued map satisfying* **(DI)**, *and $\mu(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ be a solution of the* relaxed differential inclusion

$$\begin{cases} \partial_t \mu(t) \in -\mathrm{div}\Big(\overline{\mathrm{co}}\, V(t,\mu(t))\mu(t)\Big), \\ \mu(0) = \mu^0, \end{cases} \tag{46}$$

where $\overline{\mathrm{co}}\,(V(t,\mu(t)))$ is the closed convex hull of $V(t,\mu(t)) \subset C^0(\mathbb{R}^d,\mathbb{R}^d)$ defined in the sense of (18).

Then for any $\delta > 0$, there exists a solution $\mu_\delta(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ of the differential inclusion

$$\begin{cases} \partial_t \mu_\delta(t) \in -\mathrm{div}\Big(V(t,\mu_\delta(t))\mu_\delta(t)\Big), \\ \mu_\delta(0) = \mu^0, \end{cases} \tag{47}$$

such that

$$W_p(\mu(t),\mu_\delta(t)) \leq \delta,$$

for all times $t \in [0,T]$.

The proof of Theorem 5 is split into two steps. Step 1 consists in building an auxiliary measure curve $\nu(\cdot)$ driven by a measurable selection from $t \in [0,T] \mapsto w(t) \in V(t,\mu(t))$ which is close to the solution $\mu(\cdot)$ of (46). In Step 2, we apply the Filippov estimates of Theorem 4 to recover the existence of a solution $\mu_\delta(\cdot)$ of (47) which is close to $\nu(\cdot)$ and therefore close to $\mu(\cdot)$ as well.

Proof (of Theorem 5). As we did in the proof of Theorem 4, we will restrict our attention to the case $p = 1$, the general case being similar. Let us denote by $r > 0$ a positive radius such that $\mu^0 \in \mathscr{P}(B(0,r))$. By Proposition 3, there exists a compact set $K := B(0,R_r) \subset \mathbb{R}^d$ depending on r, T and $\| m(\cdot) \|_1$ such that all the solutions of (46) and (47) starting from μ^0 are uniformly compactly supported in K. This together with hypothesis **(DI)**-(ii) implies in particular that every measurable selection $t \in [0,T] \mapsto v(t) \in C^0(\mathbb{R}^d,\mathbb{R}^d)$ of either $V(t,\mu(t))$ or $\overline{\mathrm{co}}\, V(t,\mu(t))$ is such that

$$\|v(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} \leq m(t)\big(1+2R_r\big), \tag{48}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$.

Step 1: Construction of an intermediate curve $\nu(\cdot)$. Since $m(\cdot) \in L^1([0,T],\mathbb{R}_+)$, we can find for any given $\delta' > 0$ a subdivision of $[0,T]$ into $N \geq 1$ intervals $[t_i, t_{i+1}]$ such that

$$\sup_{i\in\{0,\dots,N-1\}} \int_{t_i}^{t_{i+1}} m(s)\mathrm{d}s \leq \frac{\delta'}{2(1+2R_r)}, \tag{49}$$

where N depends on δ'. Moreover, remark that the multifunction $t \in [0,T] \rightrightarrows V_K(t,\mu(t))$ is $\mathscr{L}^1$-measurable and integrably bounded with compact values in $C^0(K,\mathbb{R}^d)$. Thus by Lemma 3, there exists a family of measurable selections $t \in [t_i,t_{i+1}] \mapsto v_i(t) \in V_K(t,\mu(t))$ such that

$$\left\| \int_{t_i}^{t_{i+1}} v(s,\cdot)\mathrm{d}s - \int_{t_i}^{t_{i+1}} v_i(s,\cdot)\mathrm{d}s \right\|_{C^0(K,\mathbb{R}^d)} \leq \frac{\delta'}{N}, \tag{50}$$

for all $i \in \{0,\dots,N-1\}$. Let us now consider the following Carathéodory velocity field

$$w : (t,x) \in [0,T] \times K \mapsto \sum_{i=0}^{N-1} \mathbb{1}_{[t_i,t_{i+1}]}(t) v_i(t,x) \in \mathbb{R}^d,$$

which satisfies hypotheses **(C1)**-**(C2)** of Theorem 2 up to an extension, and denote by $\nu(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}(K))$ the unique solution of the Cauchy problem

$$\begin{cases} \partial_t \nu(t) + \mathrm{div}(w(t)\nu(t)) = 0, \\ \nu(0) = \mu^0. \end{cases}$$

Our next goal is to estimate the W_1-distance between $\mu(t)$ and $\nu(t)$ for all times $t \in [0,T]$.

Let $\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu$ be superposition measures which induce $\mu(\cdot)$ and $\nu(\cdot)$ as in Theorem 1. By Lemma 1, there exists a transport plan $\hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma(\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu)$ such that (12) holds, i.e.

$$(\pi_{\mathbb{R}^d}, \pi_{\mathbb{R}^d})_\# \hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma_o(\mu^0,\mu^0) \qquad \text{and} \qquad (e_t,e_t)_\# \hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma_o(\mu(t),\nu(t)),$$

for all times $t \in [0,T]$. Thus, one has

$$\begin{aligned} W_1(\mu(t),\nu(t)) &= \int_{\mathbb{R}^{2d}} |x-y| \mathrm{d}\Big((e_t,e_t)_\# \hat{\boldsymbol{\eta}}_{\mu,\nu}\Big)(x,y) \\ &= \int_{(\mathbb{R}^d\times\Sigma_T)^2} |\sigma_\mu(t) - \sigma_\nu(t)| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\leq \int_{(\mathbb{R}^d\times\Sigma_T)^2} \left| \int_0^t \Big(v(s,\sigma_\mu(s)) - v(s,\sigma_\nu(s))\Big)\mathrm{d}s \right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\quad + \int_{(\mathbb{R}^d\times\Sigma_T)^2} \left| \int_0^t \Big(v(s,\sigma_\nu(s)) - w(s,\sigma_\nu(s))\Big)\mathrm{d}s \right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu), \end{aligned} \tag{51}$$

where we used that $\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu$ are concentrated on the characteristic curves (10) of $v(\cdot,\cdot)$ and $w(\cdot,\cdot)$, along with the known fact that $\Gamma_o(\mu^0,\mu^0) = \{(\mathrm{Id},\mathrm{Id})_\#\mu^0\}$. The first integral in (51) can be estimated[3] as

$$\begin{aligned}\int_{(\mathbb{R}^d\times\Sigma_T)^2} &\left|\int_0^t \Big(v(s,\sigma_\mu(s)) - v(s,\sigma_\nu(s))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\le \int_0^t l_K(s)\int_{(\mathbb{R}^d\times\Sigma_T)^2} |\sigma_\mu(s)-\sigma_\nu(s)|\mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu)\mathrm{d}s = \int_0^t l_K(s)W_1(\mu(s),\nu(s))\mathrm{d}s,\end{aligned} \tag{52}$$

where we used Fubini's Theorem and the fact that $x\in K\mapsto v(s,x)$ is $l_K(s)$-Lipschitz for $\mathscr{L}^1$-almost every $s\in[0,t]$ by hypothesis **(DI)**-(iii). Let $j\in\{1,\dots,N-1\}$ be such that $t\in[t_j,t_{j+1}]$. For the second integral term in (51), we then have

$$\begin{aligned}&\int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\int_0^t \Big(v(s,\sigma_\nu(s)) - w(s,\sigma_\nu(s))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\le \int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\sum_{i=0}^{j-1}\int_{t_i}^{t_{i+1}} \Big(v(s,\sigma_\nu(s)) - v_i(s,\sigma_\nu(s))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\quad+ \int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\int_{t_j}^{t} \Big(v(s,\sigma_\nu(s)) - v_j(s,\sigma_\nu(s))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\le \sum_{i=0}^{j-1}\int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\int_{t_i}^{t_{i+1}} \Big(v(s,\sigma_\nu(s)) - v_i(s,\sigma_\nu(s))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) + \delta',\end{aligned} \tag{53}$$

since $\|v(s,\cdot)-v_j(s,\cdot)\|_{C^0(K,\mathbb{R}^d)} \le 2(1+2R_r)m(s)$ for $\mathscr{L}^1$-almost every $s\in[t_j,t_{j+1}]$. For any $i\in\{0,\dots,j-1\}$, it further holds

$$\begin{aligned}&\int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\int_{t_i}^{t_{i+1}} \Big(v(s,\sigma_\nu(s)) - v_i(s,\sigma_\nu(s))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\le \int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\int_{t_i}^{t_{i+1}} \Big(v(s,\sigma_\nu(t_i)) - v_i(s,\sigma_\nu(t_i))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\quad+ \int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\int_{t_i}^{t_{i+1}} \Big(v(s,\sigma_\nu(s)) - v(s,\sigma_\nu(t_i))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\quad+ \int_{(\mathbb{R}^d\times\Sigma_T)^2} \left|\int_{t_i}^{t_{i+1}} \Big(v_i(s,\sigma_\nu(s)) - v_i(s,\sigma_\nu(t_i))\Big)\mathrm{d}s\right| \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) \\ &\le \frac{\delta'}{N} + 2\int_{t_i}^{t_{i+1}} l_K(s)\int_{(\mathbb{R}^d\times\Sigma_T)^2} |\sigma_\nu(s)-\sigma_\nu(t_i)|\mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu)\mathrm{d}s \\ &\le \frac{\delta'}{N} + 2(1+2R_r)\int_{t_i}^{t_{i+1}} l_K(s)\left(\int_{t_i}^{s} m(\tau)\mathrm{d}\tau\right)\mathrm{d}s \\ &\le \frac{\delta'}{N} + \delta'\int_{t_i}^{t_{i+1}} l_K(s)\mathrm{d}s\end{aligned} \tag{54}$$

where we used (48), (49) and (50).

Plugging (52), (53) and (54) into (51), we have shown that the intermediate curve $\nu(\cdot)$ is such that

$$W_1(\mu(t),\nu(t)) \le \int_0^t l_K(s)W_1(\mu(s),\nu(s))\mathrm{d}s + \delta'\left(2+\int_0^t l_K(s)\mathrm{d}s\right),$$

for all times $t\in[0,T]$. By a direct application of Grönwall's Lemma, we finally obtain

$$\sup_{t\in[0,T]} W_1(\mu(t),\nu(t)) \le \delta'\left(2+\|l_K(\cdot)\|_1\right)\exp\left(\|l_K(\cdot)\|_1\right). \tag{55}$$

[3]When $p>1$, this estimate can be performed as in the proof of Proposition 2 given in Appendix A.

Step 2: Construction of the curve $\mu_\delta(\cdot)$ solution of (47). Observe that the intermediate curve $\nu(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}(K))$ that we have built in Step 1 is not a solution of (47), since by construction its driving velocity field $w : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is such that

$$w(t) \in V_K(t, \mu(t)),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. We introduce the mismatch function $\eta_\nu(\cdot)$ defined in this context by

$$\eta_\nu : t \in [0,T] \mapsto \mathrm{dist}_{C^0(K,\mathbb{R}^d)}\Big(w(t), V_K(t, \nu(t))\Big),$$

and notice that as a consequence of **(DI)**-(iv), there exists $L_K(\cdot) \in L^1([0,T], \mathbb{R}_+)$ such that

$$\eta_\nu(t) \leq L_K(t) W_1(\mu(t), \nu(t)), \tag{56}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. In particular, it follows from (55) that $\eta_\nu(\cdot) \in L^1([0,T], \mathbb{R}_+)$.

We can therefore apply Theorem 4 to deduce the existence of a solution $\mu_\delta(\cdot) \in \mathrm{AC}([0,T], \mathscr{P}(K))$ of the differential inclusion (47) such that

$$W_1(\nu(t), \mu_\delta(t)) \leq \left(\int_0^t \eta_\nu(s)\mathrm{d}s\right) \exp\Big(C_{K,1}(t) + \|l_K(\cdot)\|_{L^1([0,t])}\Big), \tag{57}$$

for all times $t \in [0,T]$, as a consequence of (25). By combining (55), (56) and (57) along with an application of the triangle inequality, we further have that

$$\begin{aligned}\sup_{t\in[0,T]} W_1(\mu(t), \mu_\delta(t)) &\leq \Big(1 + \|L_K(\cdot)\|_1 \exp\Big(C_{K,1}(T) + \|l_K(\cdot)\|_1\Big)\Big) \sup_{t\in[0,T]} W_1(\mu(t), \nu(t)) \\ &\leq \delta' \Big(1 + \|L_K(\cdot)\|_1 \exp\Big(C_{K,1}(T) + \|l_K(\cdot)\|_1\Big)\Big)\Big(2 + \|l_K(\cdot)\|_1\Big) \exp\left(\|l_K(\cdot)\|_1\right).\end{aligned}$$

Thus, choosing

$$\delta' = \frac{\delta}{\Big(1 + \|L_K(\cdot)\|_1 \exp\Big(C_{K,1}(T) + \|l_K(\cdot)\|_1\Big)\Big)\Big(2 + \|l_K(\cdot)\|_1\Big) \exp\left(\|l_K(\cdot)\|_1\right)},$$

we obtain the uniform distance estimate

$$\sup_{t\in[0,T]} W_1(\mu(t), \mu_\delta(t)) \leq \delta,$$

which concludes the proof of Theorem 5. □

We now apply Theorem 5 to recover a classical fact, which states that the value-function corresponding to a minimisation problem with a convexified right-hand side coincides with the value function of the original problem, and that both value functions are continuous in a certain sense.

Definition 10 (Locally continuous maps)**.** *We say that a functional $\phi : \mathscr{P}_c(\mathbb{R}^d) \to \mathbb{R}$ is* locally continuous *if for any compact set $K \subset \mathbb{R}^d$ and all $\mu \in \mathscr{P}(K)$, it holds*

$$\phi(\mu_n) \underset{n\to+\infty}{\longrightarrow} \phi(\mu),$$

for every sequence $(\mu_n) \subset \mathscr{P}(K)$ such that $\mu_n \rightharpoonup^* \mu$ as $n \to +\infty$.

Corollary 1 (Local continuity of the value function)**.** *Let $V : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d, \mathbb{R}^d)$ be a set-valued map satisfying* **(DI)** *for some $p \in [1, +\infty)$ and $\varphi : \mathscr{P}_c(\mathbb{R}^d) \to \mathbb{R}$ be a locally continuous map. Then, the* value functions *$\mathscr{V}, \mathscr{V}_{\overline{\mathrm{co}}} : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \to \mathbb{R}$ defined respectively by*

$$\mathscr{V}(\tau, \mu_\tau) := \begin{cases} \inf\limits_{\mu(\cdot)} \left[\varphi(\mu(T))\right] \\ \text{s.t.} \begin{cases} \partial_t \mu(t) \in -\mathrm{div}\Big(V(t, \mu(t))\mu(t)\Big), \\ \mu(\tau) = \mu_\tau, \end{cases} \end{cases}$$

and

$$\mathscr{V}_{\overline{\mathrm{co}}}(\tau,\mu_\tau) := \begin{cases} \inf\limits_{\mu(\cdot)} \left[\varphi(\mu(T))\right] \\ \text{s.t.} \begin{cases} \partial_t\mu(t) \in -\mathrm{div}\Big(\overline{\mathrm{co}}\, V(t,\mu(t))\mu(t)\Big), \\ \mu(\tau) = \mu_\tau, \end{cases} \end{cases}$$

for all $(\tau,\mu_\tau) \in [0,T]\times\mathscr{P}_c(\mathbb{R}^d)$ *are equal and locally continuous.*

Proof. Let $(\tau,\mu_\tau) \in [0,T]\times\mathscr{P}_c(\mathbb{R}^d)$ and $r_\tau > 0$ be such that $\mu_\tau \in \mathscr{P}(B(0,r_\tau))$. By Proposition 3, there exists a compact set $K_\tau := B(0,R_{r_\tau})$ in which all the trajectories of the differential inclusions (46) and (47) starting from μ_τ at time $\tau \in [0,T]$ are uniformly compactly supported.

Since every trajectory of (47) is also a trajectory of the relaxed inclusion (46), it directly holds

$$\mathscr{V}_{\overline{\mathrm{co}}}(\tau,\mu_\tau) \le \mathscr{V}(\tau,\mu_\tau). \tag{58}$$

Conversely, let $\mu^*(\cdot)$ be a trajectory of the relaxed inclusion (46). By Theorem 5, there exists a sequence $(\mu_n(\cdot)) \subset \mathrm{AC}([0,T],\mathscr{P}(K_\tau))$ of trajectories of (47) such that

$$\sup_{t\in[\tau,T]} W_p(\mu_n(t),\mu^*(t)) \underset{n\to+\infty}{\longrightarrow} 0.$$

Since the trajectories $\{\mu_n(\cdot),\mu^*(\cdot)\}$ are uniformly supported in K_τ, this implies by Proposition 1 that $\mu_n(T) \rightharpoonup^* \mu^*(T)$ as $n\to+\infty$. Recalling that $\varphi(\cdot)$ is locally continuous in the sense of Definition 10, we deduce that for every $\epsilon > 0$ there exists an integer $N_\epsilon \ge 1$ – which depends on $\mu^*(\cdot)$ –, such that

$$\mathscr{V}(\tau,\mu_\tau) \le \varphi(\mu_n(T)) \le \varphi(\mu^*(T)) + \epsilon, \tag{59}$$

for every $n \ge N_\epsilon$. Thus taking the infimum over the trajectories $\mu^*(\cdot)$ in (59), we recover

$$\mathscr{V}(\tau,\mu_\tau) \le \mathscr{V}_{\overline{\mathrm{co}}}(\tau,\mu_\tau) + \epsilon,$$

for every $\epsilon > 0$, which together with (58) yields that $\mathscr{V}(\tau,\mu_\tau) = \mathscr{V}_{\overline{\mathrm{co}}}(\tau,\mu_\tau)$.

The continuity of the map $t\in[0,T] \mapsto \mathscr{V}(t,\mu_\tau)$ and the local continuity of $\mu \in \mathscr{P}_c(\mathbb{R}^d) \mapsto \mathscr{V}(\tau,\mu)$ in the sense of Definition 10 at (τ,μ_τ) follow easily from the estimates (21) of Proposition 3 and (25) of Theorem 4 together with the local continuity of $\varphi(\cdot)$. □

3.4 Compactness of the set of trajectories

In this section, we show that the set of solutions to a differential inclusion in $(\mathscr{P}_c(\mathbb{R}^d),W_p)$ is compact in the topology of the uniform convergence whenever its right-hand side has convex values.

Theorem 6 (Compactness of trajectories)**.** *Let* $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ *and* $V : [0,T]\times\mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ *be a set-valued map with convex values which satisfies hypotheses* **(DI)** *for some* $p\in[1,+\infty)$*. Then, the solution set of the differential inclusion* (20) *with* $\mu(0) = \mu^0$ *is compact in* $C^0([0,T],\mathscr{P}_c(\mathbb{R}^d))$*.*

The proof of Theorem 6 strongly relies on by-now classical estimates and compactness results for continuity equations, for which we refer the reader e.g. to [39, 40]

Proof (of Theorem 6). Let $r > 0$ be such that $\mu^0 \in \mathscr{P}(B(0,r))$ and $(\mu_n(\cdot),v_n(\cdot))$ be a sequence of trajectory-selection pairs for (20) with $\mu_n(0) = \mu^0$ for every $n \ge 1$. By Proposition 3, there exist a constant $R_r > 0$ and a map $m_r(\cdot) \in L^1([0,T],\mathbb{R}_+)$ such that

$$\mathrm{supp}(\mu_n(t)) \subset K := B(0,R_r) \qquad \text{and} \qquad W_p(\mu_n(t),\mu_n(s)) \le \int_s^t m_r(\tau)\mathrm{d}\tau, \tag{60}$$

for all $0 \le s \le t \le T$ and every $n \ge 1$. Therefore by the Ascoli-Arzelà Theorem, their exists a subsequence of $(\mu_n(\cdot))$ that we do not relabel and a limit curve $\mu^*(\cdot)$ such that

$$\sup_{t\in[0,T]} W_p(\mu_n(t),\mu^*(t)) \underset{n\to+\infty}{\longrightarrow} 0. \tag{61}$$

It can moreover be verified straightforwardly that the limit curve $\mu^*(\cdot)$ satisfies the estimates of (60).

As a consequence of hypotheses **(DI)**-(ii) and **(DI)**-(iii) along with the uniform compactness of the support of the trajectories $\{(\mu_n(\cdot)), \mu^*(\cdot)\}$ given by (60), the sequence of admissible velocities $(v_n(\cdot)) \subset L^1([0,T], C^0(K,\mathbb{R}^d))$ is uniformly integrably bounded in $L^1([0,T], W^{1,q}(K,\mathbb{R}^d))$ for any $q \in (1,+\infty)$. By an application of the generalisation of Dunford-Pettis Theorem (see e.g. [7, Theorem 1.38]) for Bochner integrable maps, the sequence $(v_n(\cdot))$ admits a cluster point $v^*(\cdot)$ in the weak $L^1([0,T], W^{1,q}(K,\mathbb{R}^d))$-topology, i.e.

$$\int_0^T \langle \boldsymbol{\nu}(t), v^*(t) - v_n(t)\rangle_{W^{1,q}(K,\mathbb{R}^d)} \mathrm{d}t \underset{n\to+\infty}{\longrightarrow} 0, \tag{62}$$

for any $\boldsymbol{\nu}(\cdot) \in L^\infty([0,T], (W^{1,q}(K,\mathbb{R}^d))')$. We henceforth choose an exponent $q > d$, so that by Morrey's Embedding (see e.g. [25, Theorem 9.12]) it holds that $W^{1,q}(K,\mathbb{R}^d) \subset C^0(K,\mathbb{R}^d)$. By taking the topological dual of this inclusion, we recover that $(C^0(K,\mathbb{R}^d))' \subset (W^{1,q}(K,\mathbb{R}^d))'$, so that (62) implies in particular

$$\int_0^T \int_{\mathbb{R}^d} \langle \nabla_x \phi(t,x), v^*(t,x) - v_n(t,x)\rangle \mathrm{d}\mu^*(t)(x)\mathrm{d}t \underset{n\to+\infty}{\longrightarrow} 0,$$

for any $\phi \in C_c^\infty([0,T]\times\mathbb{R}^d)$. Recalling that the maps $v_n(t,\cdot)$ are $l_K(t)$-Lipschitz for any $n \geq 1$ and $\mathscr{L}^1$-almost every $t \in [0,T]$, we obtain

$$\int_0^T \int_{\mathbb{R}^d} \langle \nabla_x \phi(t,x), v_n(t,x)\rangle \mathrm{d}\mu_n(t)(x)\mathrm{d}t \underset{n\to+\infty}{\longrightarrow} \int_0^T \int_{\mathbb{R}^d} \langle \nabla_x \phi(t,x), v^*(t,x)\rangle \mathrm{d}\mu^*(t)(x)\mathrm{d}t,$$

for any $\phi \in C_c^\infty([0,T]\times\mathbb{R}^d)$, where we used the Kantorovich duality formula (7) and the fact that the Wasserstein distances are ordered, see Proposition 1. Passing to the limit as $n \to +\infty$ in the distributional formulation (9) of the continuity equation, we conclude that the pair $(\mu^*(\cdot), v^*(\cdot))$ is a solution of the Cauchy problem

$$\begin{cases} \partial_t \mu^*(t) + \mathrm{div}(v^*(t)\mu^*(t)) = 0, \\ \mu^*(0) = \mu^0. \end{cases}$$

We now prove that $v^*(t) \in V_K(t,\mu^*(t))$ for $\mathscr{L}^1$-almost every $t \in [0,T]$. As a consequence of hypothesis **(DI)**-(iv) along with an application of Lemma 2-(a), there exists a sequence of maps $(\tilde{v}_n(\cdot)) \subset L^1([0,T], C^0(K,\mathbb{R}^d))$ such that

$$\|v_n(t,\cdot) - \tilde{v}_n(t,\cdot)\|_{C^0(K,\mathbb{R}^d)} \leq L_K(t)\, W_p(\mu_n(t), \mu^*(t)) \qquad \text{and} \qquad \tilde{v}_n(t) \in V_K(t,\mu^*(t)), \tag{63}$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. By repeating the same compactness argument as before, the sequence of maps $(\tilde{v}_n(\cdot))$ admits a cluster point $\tilde{v}(\cdot)$ in the weak $L^1([0,T], W^{1,q}(K,\mathbb{R}^d))$-topology. It can in turn be checked as a consequence of our standing assumptions that the set

$$\mathscr{V}_K := \Big\{ v \in L^1([0,T], C^0(K,\mathbb{R}^d)) \text{ s.t. } v(t) \in V_K(t,\mu^*(t)) \text{ for } \mathscr{L}^1\text{-almost every } t \in [0,T] \Big\}, \tag{64}$$

is closed in the strong $L^1([0,T], W^{1,q}(K,\mathbb{R}^d))$-topology for any $q \in (d,+\infty)$. Observe that it is also convex since the set-valued map $V(\cdot,\cdot)$ has convex values. Hence it is also weakly closed by Mazur's Lemma (see e.g. [25, Theorem 3.7]), which implies that $\tilde{v}(\cdot) \in \mathscr{V}_K$. Now we can conclude that $v^*(\cdot) \in \mathscr{V}_K$ as a consequence of (61) and (63) and by uniqueness of the weak-* limit. □

4 Application to a Mean-Field Optimal Control Problem

In this section, we apply the set-theoretic tools and results of Section 3 to study the existence of minimisers for a constrained mean-field optimal control problem. To the best of our knowledge, this is the first existence result of this type for general constrained problems, with the extra novelty that the controls may have a feedback structure and are involved non-linearly in the dynamics.

Consider $\varphi : \mathscr{P}_c(\mathbb{R}^d) \to \mathbb{R}$ and $\mathcal{Q}_T, \mathcal{K} \subset \mathscr{P}_1(\mathbb{R}^d)$. Moreover, let (U, d_U) be a compact metric space, $v : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \times U \to C^0(\mathbb{R}^d, \mathbb{R}^d)$ be a non-local velocity field and $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$. In the sequel, we shall study the following general constrained Mayer problem

$$(\mathcal{P}) \begin{cases} \min\limits_{u(\cdot)\in\mathcal{U}} \big[\varphi(\mu(T))\big] \\ \text{s.t.} \begin{cases} \partial_t \mu(t) + \operatorname{div}\Big(v(t,\mu(t),u(t,\cdot))\mu(t)\Big) = 0, \\ \mu(0) = \mu^0, \end{cases} \\ \text{and} \begin{cases} \mu(T) \in \mathcal{Q}_T, \\ \mu(t) \in \mathcal{K} \ \text{ for all times } t \in [0,T], \end{cases} \end{cases}$$

over the set of admissible controls $u : [0,T] \to \Omega$, where $\Omega \subset C^0(\mathbb{R}^d, U)$ is a closed set such that

$$\Omega \subset \Big\{\omega \in C^0(\mathbb{R}^d, U) \text{ s.t. } \operatorname{Lip}(\omega(\cdot)\,;\mathbb{R}^d) \le L_U\Big\}, \tag{65}$$

for a given constant $L_U > 0$. Here, the control variables $u : (t,x) \in [0,T] \times \mathbb{R}^d \to U$ are Carathéodory vector fields such that $u(t,\cdot) \in \Omega$ for $\mathscr{L}^1$-almost every $t \in [0,T]$, which pilot the evolution of the state $\mu(\cdot)$ of the system through the non-local controlled velocity

$$(t,x) \in [0,T] \times \mathbb{R}^d \mapsto v(t,\mu(t),u(t,x))(x) \in \mathbb{R}^d.$$

We denote by $\mathcal{U}$ the set of all such admissible controls.

Remark 4 (The case of open-loop controls)**.** *Observe that the above definition of admissible controls is also adapted to the study of purely open-loop controls, i.e. controls which depend on time only and not on the space variable $x \in \mathbb{R}^d$. Indeed, let $\Omega = \{\omega \in C^0(\mathbb{R}^d, U)$ s.t. $\omega(\cdot) \equiv const\}$. Then Ω is closed and for any $\mathscr{L}^1$-measurable selection $t \in [0,T] \mapsto u(t) \in \Omega$, the mapping $\tilde{u} : (t,x) \in [0,T] \times \mathbb{R}^d \mapsto u(t) \in U$ satisfies $\tilde{u}(t,\cdot) \in \Omega$ for $\mathscr{L}^1$-almost every $t \in [0,T]$. Setting*

$$\mathcal{U} := \Big\{\tilde{u}(\cdot,\cdot) \text{ s.t. } \tilde{u}(t,\cdot) \in \Omega \text{ and } \tilde{u}(\cdot,x) \text{ is } \mathscr{L}^1\text{-measurable}\Big\},$$

we get the set of admissible purely open-loop controls.

We consider the following differential inclusion

$$\partial_t \mu(t) \in -\operatorname{div}\Big(V(t,\mu(t))\mu(t)\Big), \tag{66}$$

where the set-valued map $V : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d, \mathbb{R}^d)$ is defined by

$$V(t,\mu) := \Big\{\boldsymbol{v} \in C^0(\mathbb{R}^d, \mathbb{R}^d) \text{ s.t. } \boldsymbol{v}(\cdot) = \hat{v}(t,\mu,\omega)(\cdot) \text{ for some } \omega \in \Omega\Big\}, \tag{67}$$

and the map $\hat{v}(t,\mu,\omega) \in C^0(\mathbb{R}^d, \mathbb{R}^d)$ is given for any $\omega \in \Omega$ by

$$\hat{v}(t,\mu,\omega) : x \in \mathbb{R}^d \mapsto v(t,\mu,\omega(x))(x). \tag{68}$$

Throughout this section, we fix $p \in [1,+\infty)$ and impose the following assumptions on problem $(\mathcal{P})$.

Hypotheses (OCP). *For every $R > 0$, assume that the following holds with $K := B(0,R)$.*

(*i*) *The map $t \in [0,T] \mapsto v(t,\mu,u)(x) \in \mathbb{R}^d$ is $\mathscr{L}^1$-measurable and there exists $m(\cdot) \in L^1([0,T],\mathbb{R}_+)$ such that*

$$|v(t,\mu,u)(x)| \le m(t)\Big(1 + |x| + \mathcal{M}_1(\mu)\Big),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$ and any $(\mu,u,x) \in \mathscr{P}_c(\mathbb{R}^d) \times U \times \mathbb{R}^d$. Moreover, there exist two maps $l_K(\cdot), L_K(\cdot) \in L^1([0,T],\mathbb{R}_+)$ such that for $\mathscr{L}^1$-almost every $t \in [0,T]$, we have

$$|v(t,\mu,u_1)(x) - v(t,\mu,u_2)(y)| \le l_K(t)\Big(|x-y| + d_U(u_1,u_2)\Big),$$

for any $x, y \in K$ and $u_1, u_2 \in U$, and

$$\|v(t,\mu,u)(\cdot) - v(t,\nu,u)(\cdot)\|_{C^0(K,\mathbb{R}^d)} \le L_K(t) W_p(\mu,\nu),$$

for any $\mu, \nu \in \mathscr{P}(K)$ and $u \in U$.

(*ii*) *The set of admissible velocities* $V(t,\mu)$ *defined in* (67) *is convex for* $\mathscr{L}^1$*-almost every* $t \in [0,T]$ *and every* $\mu \in \mathscr{P}_c(\mathbb{R}^d)$.

(*iii*) *The final cost* $\mu \in \mathscr{P}_c(\mathbb{R}^d) \mapsto \varphi(\mu) \in \mathbb{R}$ *is lower-semicontinuous over* $\mathscr{P}(K)$ *in the* W_1*-metric.*

(*iv*) *The running and final constraint sets* $\mathcal{K}$ *and* $\mathcal{Q}_T$ *are closed in the* W_1*-topology.*

Remark 5 (Open-loop controls and regularity)**.** *When* $\mathcal{U}$ *consists of open-loop controls only, the Lipschitz continuity assumption on* $u \in U \mapsto v(t,\mu,u) \in \mathbb{R}^d$ *can be relaxed into a continuity assumption.*

We refer the reader e.g. to [22, Appendix A], [24, Section4], as well as [31] and [38, 39] for examples of velocity fields, cost functionals and constraint sets satisfying similar assumptions. In the following proposition, we show that under hypotheses **(OCP)**-(*i*), the set of all trajectories of the controlled non-local Cauchy-problem

$$\begin{cases} \partial_t \mu(t) + \operatorname{div}\Big(v(t,\mu(t),u(t,\cdot))\mu(t)\Big) = 0, \\ \mu(0) = \mu^0 \in \mathscr{P}_c(\mathbb{R}^d), \end{cases} \tag{69}$$

coincides with the set of trajectories of (66) whenever $V(\cdot,\cdot)$ is defined by (67).

Proposition 4 (Link between differential inclusions and control systems)**.** *Let* $v : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \times \mathbb{R}^d \times U \to \mathbb{R}^d$ *be a non-local velocity field satisfying hypothesis* **(OCP)**-(*i*) *and* $V : [0,T] \times \mathscr{P}_c(\mathbb{R}^d) \rightrightarrows C^0(\mathbb{R}^d,\mathbb{R}^d)$ *be the set-valued map defined as in* (67).

Then, a curve of measures $\mu(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ *is a solution of the differential inclusion* (66) *if and only if it is a solution of* (69) *generated by an admissible control* $u(\cdot) \in \mathcal{U}$.

Proof. Let $\mu(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ be a curve of measures solution of (69) for a given admissible control map $u(\cdot) \in \mathcal{U}$. By construction, the time-dependent velocity field $\boldsymbol{v} : t \in [0,T] \mapsto \hat{v}(t,\mu(t),u(t)) \in C^0(\mathbb{R}^d,\mathbb{R}^d)$ is such that $\boldsymbol{v}(t) \in V(t,\mu(t))$ for $\mathscr{L}^1$-almost every $t \in [0,T]$, so that $\mu(\cdot)$ solves (66).

Conversely, suppose that $\mu(\cdot) \in \mathrm{AC}([0,T],\mathscr{P}_c(\mathbb{R}^d))$ is a solution of (66). Notice that as a consequence of hypothesis **(OCP)**-(*i*), and of the definition of $\mathcal{U}$, the set valued map $V(\cdot,\cdot)$ defined in (67) satisfies hypotheses **(DI)**. Whence by Proposition 3, there exists a closed ball $K := B(0,R)$ such that $\mathrm{supp}(\mu(t)) \subset K$ for all times $t \in [0,T]$, and by Definition 9 we obtain the existence of a measurable selection $t \in [0,T] \mapsto \boldsymbol{v}(t) \in V_K(t,\mu(t))$ such that

$$\partial_t \mu(t) + \operatorname{div}\big(\boldsymbol{v}(t)\mu(t)\big) = 0.$$

Moreover, we know that $V_K(t,\mu(t)) = \hat{v}_{|K}(t,\mu(t),\Omega) \subset C^0(K,\mathbb{R}^d)$ where $(t,\omega) \mapsto \hat{v}_{|K}(t,\mu(t),\omega)$ is $\mathscr{L}^1$-measurable with respect to $t \in [0,T]$ and continuous with respect to $\omega \in \Omega$. We can therefore apply the measurable selection theorem e.g. of [13, Theorem 8.2.9] to recover the existence of a measurable selection $t \in [0,T] \mapsto u(t) \in \Omega$ such that $\boldsymbol{v}(t) = \hat{v}_{|K}(t,\mu(t),u(t))$ for $\mathscr{L}^1$-almost every $t \in [0,T]$. Therefore up to an extension argument, we deduce that $\mu(\cdot)$ solves (69) with driving velocity field $(t,x) \in [0,T] \times \mathbb{R}^d \mapsto v(t,\mu(t),u(t,x))(x) \in \mathbb{R}^d$. □

Remark 6 (Comparison with the admissible trajectories of [29, 30, 47])**.** *In [29, 30, 47], the authors consider a different notion of solution to differential inclusion in Wasserstein spaces. Given a compact metric space* U *and a continuous map* $f : \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times U \to \mathbb{R}^d$ *which is Lipschitz with respect to its two first arguments, they define the set-valued map* $F(\mu,x) := f(\mu,x,U)$. *A trajectory* $\mu(\cdot)$ *is then said to be admissible if it solves* (8) *driven by a Borel velocity field* $(t,x) \mapsto v(t,x) \in F(\mu(t),x)$. *It is shown that for this notion of solution to differential inclusion, admissible trajectories depend in a Lipschitz-like way on their initial condition. The proof of this result relies on a careful adaptation of the superposition principle from [11] (see also [30]), that allows to link the set of characteristics of the differential inclusion* $\dot{\sigma}(t) \in F(\mu(t),\sigma(t))$ *to the curve of measures* $\mu(\cdot)$ *via the evaluation map.*

In this context however, the controls $u_\sigma(\cdot)$ *obtained by applying measurable selection theorems to* $t \in [0,T] \rightrightarrows f(\mu(t),\sigma(t),U)$ *inherently depend on both the measure curve* $\mu(\cdot)$ *and the characteristic curve* $\sigma(\cdot)$. *This is a crucial difference with our definition of admissible trajectories, for which admissible controls depend on the state* $\mu(\cdot)$ *only. For this reason, our functional approach to differential*

inclusions is closer in spirit to the usual formulation of control systems as differential inclusions. Besides as illustrated in the Introduction, it is also more meaningful in terms of the geometry of the metric spaces $(\mathscr{P}_c(\mathbb{R}^d), W_p)$ seen as subsets of the pseudo-Riemannian manifold $(\mathscr{P}_2(\mathbb{R}^d), W_2)$.

In Theorem 7 below, we state a general result on the existence of optimal controls for problem $(\mathcal{P})$. We would like to stress that in most of the existing contributions on this topic (see e.g. [22, 24, 38, 40]), the velocity field $(t, x, \mu, \omega) \mapsto v(t, \mu, \omega(x))(x)$ is assumed to have a *control-affine* structure. The case of non-linearly controlled vector-field was studied e.g. in [27, 29, 56] for open-loop controls.

Theorem 7 (Existence of optimal controls for $(\mathcal{P})$)**.** *Under hypotheses* **(OCP)***, there exists an optimal trajectory-control pair $(\mu^*(\cdot), u^*(\cdot)) \in \mathrm{AC}([0,T], \mathscr{P}_c(\mathbb{R}^d)) \times \mathcal{U}$ for $(\mathcal{P})$.*

Proof. Let $(u_n(\cdot)) \subset \mathcal{U}$ be a minimising sequence for $(\mathcal{P})$ and $(\mu_n(\cdot)) \subset \mathrm{AC}([0,T], \mathscr{P}_c(\mathbb{R}^d))$ be the corresponding sequence of solutions of the non-local Cauchy problems

$$\begin{cases} \partial_t \mu_n(t) + \mathrm{div}\Big(v(t, \mu_n(t), u_n(t,\cdot))\mu_n(t)\Big) = 0, \\ \mu_n(0) = \mu^0. \end{cases} \tag{70}$$

It can be checked that the set-valued map $V(\cdot,\cdot)$ defined in (67) satisfies the set of assumptions **(DI)** as a consequence of hypotheses **(OCP)**-(i) together with the definition of $\mathcal{U}$. Thus by Proposition 3, there exists a compact set $K \subset \mathbb{R}^d$ such that $\mathrm{supp}(\mu_n(t)) \subset K$ for all $t \in [0,T]$ and any $n \geq 1$. Moreover, the admissible velocity sets $V_K(\cdot,\cdot)$ have convex values by **(OCP)**-(ii). Whence by Theorem 6, there exists a trajectory-selection pair $(\mu^*(\cdot), \boldsymbol{v}^*(\cdot)) \in \mathrm{AC}([0,T], \mathscr{P}(K)) \times L^1([0,T], C^0(K, \mathbb{R}^d))$ solution of the differential inclusion

$$\begin{cases} \partial \mu^*(t) \in -\mathrm{div}\Big(V(t, \mu^*(t))\mu^*(t)\Big), \\ \mu^*(0) = \mu^0, \end{cases}$$

such that

$$\sup_{t\in[0,T]} W_p(\mu_n(t), \mu^*(t)) \underset{n\to+\infty}{\longrightarrow} 0,$$

along a subsequence that we do not relabel. From Proposition 4, we deduce the existence of a measurable selection $t \in [0,T] \mapsto u^*(t) \in \Omega$ such that

$$\boldsymbol{v}^*(t) = \hat{v}(t, \mu^*(t), u^*(t)),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. By (68), this implies that the limit trajectory-control pair $(\mu^*(\cdot), u^*(\cdot))$ is a solution of the Cauchy problem driving $(\mathcal{P})$.

We can now conclude that $(\mu^*(\cdot), u^*(\cdot))$ is optimal for $(\mathcal{P})$ by remarking that

$$\liminf_{n\to+\infty} \left[\varphi(\mu_n(T))\right] \geq \varphi(\mu^*(T)),$$

and also that $\mu^*(T) \in \mathcal{Q}_T$ as well as $\mu^*(t) \in \mathcal{K}$ for all times $t \in [0,T]$, as a direct consequence of **(OCP)**-(iii) and **(OCP)**-(iv) along with (61). □

A Proofs of Lemma 1 and Proposition 2

In this section, we detail the proofs of Lemma 1 and Proposition 2. These results rely strongly on the notion of *disintegration* of measures defined over Banach spaces, which we recall in the following theorem (see e.g. [10, Theorem 5.3.1]).

Theorem A.1 (Disintegration)**.** *Let X, Y be two separable Banach spaces and $\pi : Y \to X$ be a Borel map. Given a measure $\nu \in \mathscr{P}(Y)$ and its image $\mu = \pi_\# \nu \in \mathscr{P}(X)$ through π, there exists a μ-almost uniquely determined family of Borel measures $\{\nu_x\}_{x\in X} \subset \mathscr{P}(Y)$ such that*

$$\begin{cases} \displaystyle\int_Y \phi(y)\mathrm{d}\nu(y) = \int_X \left(\int_{\pi^{-1}(x)} \phi(y)\mathrm{d}\nu_x(y)\right)\mathrm{d}\mu(x), \\ \nu_x(Y\backslash\pi^{-1}(x)) = 0 \qquad \textit{for } \mu\textit{-almost every } x \in X, \end{cases} \tag{A.1}$$

for any map $\phi \in L^1(Y, \mathbb{R}; \nu)$. The family of measures $\{\nu_x\}_{x\in X}$ is called the disintegration of ν onto μ *and is denoted by $\nu = \int_X \nu_x \mathrm{d}\mu(x)$.*

Proof (of Lemma 1). Let $\{t_k\}_{k=1}^{+\infty} \subset [0,T]$ be a countable and dense subset in $[0,T]$. We are going to split the proof of this result into three steps. In Step 1, we start by building a suitable sequence of measures $(\hat{\boldsymbol{\eta}}^n_{\mu,\nu}) \subset \Gamma(\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu)$ such that $(e_{t_k}, e_{t_k})_\# \hat{\boldsymbol{\eta}}^n_{\mu,\nu} \in \Gamma_o(\mu(t_k), \nu(t_k))$ for any $k \in \{1, \dots, n\}$. We then show in Step 2 that this sequence is tight and therefore narrowly sequentially compact in $\mathscr{P}((\mathbb{R}^d \times \Sigma_T)^2)$, and finally in Step 3 that its cluster points $\hat{\boldsymbol{\eta}}_{\mu,\nu}$ satisfy (12).

Step 1: Construction of the sequence $(\hat{\boldsymbol{\eta}}^n_{\mu,\nu})$. Let $\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ be defined as in the statement of Lemma 1. Let also $n \geq 1$ be an arbitrary integer and $\boldsymbol{x} = (x_0, x_1, \dots, x_n)$ denote a generic element of $\mathbb{R}^d \times (\mathbb{R}^d)^n$. We start by defining the measures $\boldsymbol{\mu}_n, \boldsymbol{\nu}_n \in \mathscr{P}(\mathbb{R}^d \times (\mathbb{R}^d)^n)$ as

$$\boldsymbol{\mu}_n := \left(\pi_{\mathbb{R}^d}, \prod_{k=1}^n e_{t_k}\right)_\# \boldsymbol{\eta}_\mu \in \Gamma\left(\mu^0, \prod_{k=1}^n \mu(t_k)\right), \qquad \boldsymbol{\nu}_n := \left(\pi_{\mathbb{R}^d}, \prod_{k=1}^n e_{t_k}\right)_\# \boldsymbol{\eta}_\nu \in \Gamma\left(\nu^0, \prod_{k=1}^n \nu(t_k)\right).$$

By Theorem A.1 above, there respectively exist a $\boldsymbol{\mu}_n$-almost uniquely determined family of measures $\{\boldsymbol{\eta}^n_{\mu,\boldsymbol{x}}\}_{\boldsymbol{x}} \subset \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ and a $\boldsymbol{\nu}_n$-almost uniquely determined family of measures $\{\boldsymbol{\eta}^n_{\nu,\boldsymbol{y}}\}_{\boldsymbol{y}} \subset \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ such that

$$\boldsymbol{\eta}_\mu = \int_{\mathbb{R}^d \times (\mathbb{R}^d)^n} \boldsymbol{\eta}^n_{\mu,\boldsymbol{x}} \mathrm{d}\boldsymbol{\mu}_n(\boldsymbol{x}) \qquad \text{and} \qquad \boldsymbol{\eta}_\nu = \int_{\mathbb{R}^d \times (\mathbb{R}^d)^n} \boldsymbol{\eta}^n_{\nu,\boldsymbol{y}} \mathrm{d}\boldsymbol{\nu}_n(\boldsymbol{y}).$$

Given $p \in [1, +\infty)$, let us choose p-optimal transport plans $\gamma_0 \in \Gamma_o(\mu^0, \nu^0)$ and $\gamma_k \in \Gamma_o(\mu(t_k), \nu(t_k))$ for any $k \in \{1, \dots, n\}$. By iterative applications of the Gluing Lemma (see e.g. [10, Lemma 5.3.2]), we can build a transport plan $\hat{\boldsymbol{\gamma}}_n \in \Gamma(\boldsymbol{\mu}_n, \boldsymbol{\nu}_n)$ such that

$$(\pi^1, \pi^{n+2})_\# \hat{\boldsymbol{\gamma}}_n = \gamma_0, \qquad (\pi^{k+1}, \pi^{n+k+2})_\# \hat{\boldsymbol{\gamma}}_n = \gamma_k.$$

for every $k \in \{1, \dots, n\}$. We can thus build by disintegration the measure $\hat{\boldsymbol{\eta}}^n_{\mu,\nu} \in \mathscr{P}((\mathbb{R}^d \times \Sigma_T)^2)$ as

$$\hat{\boldsymbol{\eta}}^n_{\mu,\nu} := \int_{\mathbb{R}^{2d} \times (\mathbb{R}^{2d})^n} \left(\boldsymbol{\eta}^n_{\mu,\boldsymbol{x}} \times \boldsymbol{\eta}^n_{\nu,\boldsymbol{y}}\right) \mathrm{d}\hat{\boldsymbol{\gamma}}_n(\boldsymbol{x}, \boldsymbol{y}).$$

Remark that by construction, it holds

$$(\pi_{\mathbb{R}^d}, \pi_{\mathbb{R}^d})_\# \hat{\boldsymbol{\eta}}^n_{\mu,\nu} \in \Gamma_o(\mu^0, \nu^0) \qquad \text{and} \qquad (e_{t_k}, e_{t_k})_\# \hat{\boldsymbol{\eta}}^n_{\mu,\nu} \in \Gamma_o(\mu(t_k), \nu(t_k)),$$

for any $k \in \{1, \dots, n\}$, together with $\hat{\boldsymbol{\eta}}^n_{\mu,\nu} \in \Gamma(\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu)$ for any $n \geq 1$.

Step 2: Tightness. We now want to prove that the sequence of measures $(\hat{\boldsymbol{\eta}}^n_{\mu,\nu}) \subset \mathscr{P}((\mathbb{R}^d \times \Sigma_T)^2)$ is relatively sequentially compact. In separable Banach spaces, this is equivalent to the *tightness* of the sequence by Prokhorov's Theorem (see e.g. [10, Theorem 5.1.3]). A necessary and sufficient condition for tightness (see e.g. [10, Remark 5.1.5]) is given in our context by the existence of a map $\Psi : (\mathbb{R}^d \times \Sigma_T)^2 \to [0, +\infty]$ with *compact sub-levels*, such that

$$\sup_{n\geq 1} \int_{(\mathbb{R}^d \times \Sigma_T)^2} \Psi(x, \sigma_\mu, y, \sigma_\nu) \mathrm{d}\hat{\boldsymbol{\eta}}^n_{\mu,\nu}(x, \sigma_\mu, y, \sigma_\nu) < +\infty.$$

It has been shown e.g. in [6, Theorem 3.4] that the functional

$$\psi : (x, \sigma) \in \mathbb{R}^d \times \Sigma_T \mapsto \begin{cases} |x| + \displaystyle\int_0^T \frac{|\dot{\sigma}(t)|}{1 + |\sigma(t)|} \mathrm{d}t & \text{if } \sigma \in \mathrm{AC}([0,T], \mathbb{R}^d), \\ +\infty & \text{otherwise,} \end{cases}$$

has compact sublevels in $\mathbb{R}^d \times \Sigma_T$. In addition, notice that

$$\begin{aligned}\int_{\mathbb{R}^d\times\Sigma_T} \psi(x,\sigma_\mu)\mathrm{d}\boldsymbol{\eta}_\mu(x,\sigma_\mu) &= \int_{\mathbb{R}^d} |x|\mathrm{d}\mu^0(x) + \int_{\mathbb{R}^d\times\Sigma_T}\left(\int_0^T \frac{|\dot\sigma_\mu(t)|}{1+|\sigma_\mu(t)|}\mathrm{d}t\right)\mathrm{d}\boldsymbol{\eta}_\mu(x,\sigma_\mu)\\ &= \int_{\mathbb{R}^d} |x|\mathrm{d}\mu^0(x) + \int_0^T\int_{\mathbb{R}^d\times\Sigma_T} \frac{|v(t,\sigma_\mu(t))|}{1+|\sigma_\mu(t)|}\mathrm{d}\boldsymbol{\eta}_\mu(x,\sigma_\mu)\mathrm{d}t\\ &= \int_{\mathbb{R}^d} |x|\mathrm{d}\mu^0(x) + \int_0^T\int_{\mathbb{R}^d} \frac{|v(t,x)|}{1+|x|}\mathrm{d}\mu(t)(x)\mathrm{d}t \;\le\; C,\end{aligned}$$

for a constant $C>0$ depending only on K and $\|m(\cdot)\|_1$, where we used (10), Fubini's Theorem, and the sub-linearity estimate **(C1)**. The same estimate also holds true for $\boldsymbol{\eta}_\nu$, so that the map

$$\Psi : (x,\sigma_\mu,y,\sigma_\nu) \mapsto \psi(x,\sigma_\mu) + \psi(y,\sigma_\nu),$$

has compact sub-levels in $(\mathbb{R}^d\times\Sigma_T)^2$ and is such that

$$\sup_{n\ge1}\int_{(\mathbb{R}^d\times\Sigma_T)^2} \Psi(x,\sigma_\mu,y,\sigma_\nu)\mathrm{d}\hat{\boldsymbol{\eta}}^n_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu) < +\infty.$$

Whence, the sequence $(\hat{\boldsymbol{\eta}}^n_{\mu,\nu})$ is tight in $\mathscr{P}((\mathbb{R}^d\times\Sigma_T)^2)$ and therefore narrowly sequentially compact.

Step 3: Optimality of the cluster points. Let $\hat{\boldsymbol{\eta}}_{\mu,\nu}$ be a cluster point of $(\hat{\boldsymbol{\eta}}^n_{\mu,\nu})$ along a subsequence that we do not relabel in the narrow topology of $\mathscr{P}((\mathbb{R}^d\times\Sigma_T)^2)$. By construction, it holds for any $k\in\{1,\dots,n\}$ that

$$(e_{t_k},e_{t_k})_\#\hat{\boldsymbol{\eta}}^n_{\mu,\nu} \in \Gamma_o(\mu(t_k),\nu(t_k)). \tag{A.2}$$

Let $t\in[0,T]$ be arbitrary and $(t_{k_m})\subset(t_k)$ be a subsequence such that $t_{k_m}\to t$ as $m\to+\infty$. By the continuity of the evaluation maps $(e_{t_{k_m}})\subset C^0(\mathbb{R}^d\times\Sigma_T,\mathbb{R}^d)$ for any $m\ge1$ together with classical convergence results on pushforwards of sequence of measures (see [10, Lemma 5.2.1]), one has

$$(e_{t_{k_m}},e_{t_{k_m}})_\#\hat{\boldsymbol{\eta}}^n_{\mu,\nu} \underset{n\to+\infty}{\rightharpoonup^*} (e_{t_{k_m}},e_{t_{k_m}})_\#\hat{\boldsymbol{\eta}}_{\mu,\nu}.$$

In addition, remark that for any $m\ge1$, the integrals

$$\int_{\mathbb{R}^{2d}} |x-y|^p\mathrm{d}\Big((e_{t_{k_m}},e_{t_{k_m}})_\#\hat{\boldsymbol{\eta}}^n_{\mu,\nu}\Big)(x,y) = \int_{(\mathbb{R}^d\times\Sigma_T)^2} |\sigma_\mu(t_{k_m})-\sigma_\nu(t_{k_m})|^p\mathrm{d}\hat{\boldsymbol{\eta}}^n_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu),$$

are bounded uniformly with respect to $n\ge1$, since $\mu(\cdot)$ and $\nu(\cdot)$ are uniformly supported in a compact set $K\subset\mathbb{R}^d$. By the stability under narrow convergence of $\Gamma_o(\mu(t_{k_m}),\nu(t_{k_m}))$ (see e.g. [10, Proposition 7.1.3]) together with (A.2), this further implies

$$(e_{t_{k_m}},e_{t_{k_m}})_\#\hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma_o(\mu(t_{k_m}),\nu(t_{k_m})),$$

for every $m\ge1$.

We now let $m\to+\infty$. By the narrow continuity of the curves $\mu(\cdot)$ and $\nu(\cdot)$, we have

$$\mu(t_{k_m}) \underset{m\to+\infty}{\rightharpoonup^*} \mu(t) \qquad\text{and}\qquad \nu(t_{k_m}) \underset{m\to+\infty}{\rightharpoonup^*} \nu(t).$$

Moreover, observe that

$$\sup_{(x,\sigma_\mu)\in\mathrm{supp}(\boldsymbol{\eta}_\mu)} \big|e_t(x,\sigma_\mu)-e_{t_{k_m}}(x,\sigma_\mu)\big| = \sup_{(x,\sigma_\mu)\in\mathrm{supp}(\boldsymbol{\eta}_\mu)} \big|\sigma_\mu(t)-\sigma_\mu(t_{k_m})\big| \underset{m\to+\infty}{\longrightarrow} 0,$$

as a consequence of (10) and of sub-linearity hypothesis **(C1)**. The same uniform estimate holds true for $\boldsymbol{\eta}_\nu$-almost every $(y,\sigma_\nu)\in\mathbb{R}^d\times\Sigma_T$, so that

$$\hat{\boldsymbol{\eta}}_{\mu,\nu}\Big(\Big\{(x,\sigma_\mu,y,\sigma_\nu)\text{ s.t. }\Big|(e_t-e_{t_{k_m}},e_t-e_{t_{k_m}})(x,\sigma_\mu,y,\sigma_\nu)\Big|>\epsilon\Big\}\Big) \underset{m\to+\infty}{\longrightarrow} 0,$$

for any $\epsilon > 0$. Thus, the sequence of maps $((e_{t_{k_m}}, e_{t_{k_m}})) \subset C^0((\mathbb{R}^d \times \Sigma_T)^2, \mathbb{R}^{2d})$ converges in $\hat{\boldsymbol{\eta}}_{\mu,\nu}$-measure towards $(e_t, e_t) \in C^0((\mathbb{R}^d \times \Sigma_T)^2, \mathbb{R}^{2d})$ as $m \to +\infty$. Therefore, from classical convergence results on images of measures by sequences of maps (see e.g. [10, Lemma 5.4.1]), we deduce that

$$(e_{t_{k_m}}, e_{t_{k_m}})_{\#}\hat{\boldsymbol{\eta}}_{\mu,\nu} \underset{m\to+\infty}{\rightharpoonup^*} (e_t, e_t)_{\#}\hat{\boldsymbol{\eta}}_{\mu,\nu}.$$

Furthermore, we can again verify that the integrals

$$\int_{\mathbb{R}^{2d}} |x-y|^p \mathrm{d}\Big((e_{t_{k_m}}, e_{t_{k_m}})_{\#}\hat{\boldsymbol{\eta}}_{\mu,\nu}\Big)(x,y) = \int_{(\mathbb{R}^d\times\Sigma_T)^2} |\sigma_\mu(t_{k_m}) - \sigma_\nu(t_{k_m})|^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu),$$

are bounded, uniformly with respect to $m \geq 1$. We can thus invoke the stability under narrow convergence of the sets of p-optimal transport plans to recover that

$$(e_t, e_t)_{\#}\hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma_o(\mu(t), \nu(t)),$$

for all times $t \in [0,T]$. By repeating exactly the same arguments, one can show that

$$(\pi_{\mathbb{R}^d}, \pi_{\mathbb{R}^d})_{\#}\hat{\boldsymbol{\eta}}_{\mu,\nu} \in \Gamma_o(\mu^0, \nu^0),$$

which ends our proof of Lemma 1. □

Proof (of Proposition 2). **Step 1: Proof of** (13)**.** By the superposition principle stated in Theorem 1, there exists a measure $\boldsymbol{\eta}_\mu \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ such that $\mu(t) = (e_t)_{\#}\boldsymbol{\eta}_\mu$ for all times $t \in [0,T]$, where $\boldsymbol{\eta}_\mu$ is concentrated on the pairs $(x, \sigma_\mu) \in \mathbb{R}^d \times \mathrm{AC}([0,T], \mathbb{R}^d)$ solution of the characteristic equation (10). Therefore for any $p \in [1, +\infty)$, we have

$$\begin{aligned}
\mathcal{M}_p^p(\mu(t)) &= \int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu(t)(x) \\
&= \int_{\mathbb{R}^d\times\Sigma_T} |\sigma_\mu(t)|^p \mathrm{d}\boldsymbol{\eta}_\mu(x,\sigma_\mu) \\
&\leq \int_{\mathbb{R}^d\times\Sigma_T} \Big(|x| + \int_0^t m(s)(1+|\sigma_\mu(s)|\,)\mathrm{d}s\Big)^p \mathrm{d}\boldsymbol{\eta}_\mu(x,\sigma_\mu) \\
&\leq 2^{p-1}\Bigg(\int_{\mathbb{R}^d}\Big(|x| + \int_0^t m(s)\mathrm{d}s\Big)^p \mathrm{d}\mu^0(x) + \int_{\mathbb{R}^d\times\Sigma_T}\Big(\int_0^t m(s)|\sigma_\mu(s)|\mathrm{d}s\Big)^p \mathrm{d}\boldsymbol{\eta}_\mu(x,\sigma_\mu)\Bigg),
\end{aligned} \tag{A.3}$$

for all $t \in [0,T]$, as a consequence of the sub-linearity hypothesis **(C1)**.

Let us denote by $q \in (1, +\infty]$ the conjugate exponent of p. Since $\mathrm{supp}(\mu(t)) \subseteq K$ for all times $t \in [0,T]$, we have in particular that $\sigma_\mu(\cdot) \in L^\infty([0,T], \mathbb{R}^d)$ for $\boldsymbol{\eta}_\mu$-almost every $(x, \sigma_\mu) \in \mathbb{R}^d \times \Sigma_T$. This together with the fact that $m(\cdot) \in L^1([0,T], \mathbb{R}_+)$ easily yields

$$m(\cdot)^{1/q} \in L^q([0,T], \mathbb{R}_+) \qquad \text{and} \qquad m(\cdot)^{1/p}|\sigma_\mu(\cdot)| \in L^p([0,T], \mathbb{R}_+).$$

We can therefore apply Hölder's inequality to obtain the estimate

$$\Big(\int_0^t m(s)|\sigma_\mu(s)|\mathrm{d}s\Big)^p \leq \|m(\cdot)\|_{L^1([0,t])}^{p/q} \int_0^t m(s)|\sigma_\mu(s)|^p \mathrm{d}s, \tag{A.4}$$

for $\boldsymbol{\eta}_\mu$-almost every $(x, \sigma_\mu) \in \mathbb{R}^d \times \Sigma_T$ and $\mathscr{L}^1$-almost every $t \in [0,T]$. Plugging (A.4) into (A.3) and applying Fubini's Theorem, we recover

$$\mathcal{M}_p^p(\mu(t)) \leq 2^{p-1}\Bigg(\int_{\mathbb{R}^d}\Big(|x| + \int_0^t m(s)\mathrm{d}s\Big)^p \mathrm{d}\mu^0(x) + \|m(\cdot)\|_{L^1([0,t])}^{p/q} \int_0^t m(s)\mathcal{M}_p^p(\mu(t))\mathrm{d}s\Bigg). \tag{A.5}$$

By applying Grönwall's Lemma to (A.5), we further obtain

$$\mathcal{M}_p^p(\mu(t)) \leq 2^{p-1}\Bigg(\int_{\mathbb{R}^d}\Big(|x| + \int_0^t m(s)\mathrm{d}s\Big)^p \mathrm{d}\mu^0(x)\Bigg) \exp\Bigg(2^{p-1}\,\|m(\cdot)\|_{L^1([0,t])}^{p/q} \int_0^t m(s)\mathrm{d}s\Bigg).$$

Raising this inequality to the power $1/p$ and using the triangle inequality for the $L^p(\mathbb{R}^d, \mathbb{R}; \mu^0)$-norm, we obtain

$$\mathcal{M}_p(\mu(t)) \leq C_p \left(\mathcal{M}_p(\mu(0)) + \int_0^t m(s)\mathrm{d}s \right) \exp\left(C'_p \, \|m(\cdot)\|^p_{L^1([0,t])} \right)$$

where $C_p = 2^{(p-1)/p}$ and $C'_p = \frac{2^{p-1}}{p}$.

Step 2: Proof of (14)**.** Again as a consequence of Theorem 1, there exist two probability measures $\boldsymbol{\eta}_\mu, \boldsymbol{\eta}_\nu \in \mathscr{P}(\mathbb{R}^d \times \Sigma_T)$ concentrated on the characteristic curves (10) of $v(\cdot,\cdot)$ and $w(\cdot,\cdot)$, such that

$$\mu(t) = (e_t)_\# \boldsymbol{\eta}_\mu \qquad \text{and} \qquad \nu(t) = (e_t)_\# \boldsymbol{\eta}_\nu,$$

for all times $t \in [0,T]$. For any $p \in [1,+\infty)$, we can invoke Lemma 1 to build a transport plan $\hat{\boldsymbol{\eta}}_{\mu,\nu} \in \mathscr{P}((\mathbb{R}^d \times \Sigma_T)^2)$ satisfying

$$(\pi_{\mathbb{R}^d}, \pi_{\mathbb{R}^d})_\# \hat{\boldsymbol{\eta}}_{\mu,\nu} := \boldsymbol{\mu}_0 \in \Gamma_o(\mu^0, \nu^0) \qquad \text{and} \qquad (e_t, e_t)_\# \hat{\boldsymbol{\eta}}_{\mu,\nu} := \boldsymbol{\mu}_t \in \Gamma_o(\mu(t), \nu(t)), \tag{A.6}$$

for all times $t \in [0,T]$. It then holds

$$\begin{aligned} W_p^p(\mu(t), \nu(t)) &= \int_{\mathbb{R}^{2d}} |x-y|^p \mathrm{d}\boldsymbol{\mu}_t(x,y) \\ &= \int_{(\mathbb{R}^d \times \Sigma_T)^2} |\sigma_\mu(t) - \sigma_\nu(t)|^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x, \sigma_\mu, y, \sigma_\nu) \\ &\leq \int_{(\mathbb{R}^d \times \Sigma_T)^2} \left(|x-y| + \int_0^t |v(s,\sigma_\mu(s)) - w(s,\sigma_\nu(s))| \mathrm{d}s \right)^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x, \sigma_\mu, y, \sigma_\nu). \end{aligned} \tag{A.7}$$

We can further estimate the time-integral in the right-hand side of (A.7) as

$$\begin{aligned} \int_0^t |v(s,\sigma_\mu(s)) - w(s,\sigma_\nu(s))| \mathrm{d}s &\leq \int_0^t |v(s,\sigma_\mu(s)) - v(s,\sigma_\nu(s))| \mathrm{d}s + \int_0^t |v(s,\sigma_\nu(s)) - w(s,\sigma_\nu(s))| \mathrm{d}s \\ &\leq \int_0^t l_K(s) |\sigma_\mu(s) - \sigma_\nu(s)| \mathrm{d}s + \int_0^t \|v(s,\cdot) - w(s,\cdot)\|_{C^0(K,\mathbb{R}^d)} \mathrm{d}s, \end{aligned} \tag{A.8}$$

for $\hat{\boldsymbol{\eta}}_{\mu,\nu}$-almost every $(x, \sigma_\mu, y, \sigma_\nu) \in (\mathbb{R}^d \times \Sigma_T)^2$ as a consequence of hypothesis **(C2)**. Plugging (A.8) into (A.7), we recover

$$\begin{aligned} W_p^p(\mu(t), \nu(t)) \leq 2^{p-1} &\int_{(\mathbb{R}^d \times \Sigma_T)^2} \left(|x-y| + \int_0^t \|v(s,\cdot) - w(s,\cdot)\|_{C^0(K,\mathbb{R}^d)} \mathrm{d}s \right)^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x, \sigma_\mu, y, \sigma_\nu) \\ &+ 2^{p-1} \int_{(\mathbb{R}^d \times \Sigma_T)^2} \left(\int_0^t l_K(s) |\sigma_\mu(s) - \sigma_\nu(s)| \mathrm{d}s \right)^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x, \sigma_\mu, y, \sigma_\nu). \end{aligned} \tag{A.9}$$

As in Step 1, we can estimate for $\hat{\boldsymbol{\eta}}_{\mu,\nu}$-almost every $(x, \sigma_\mu, y, \sigma_\nu) \in (\mathbb{R}^d \times \Sigma_T)^2$ the time integral of the second term in the right-hand side of (A.9) as

$$\left(\int_0^t l_K(s) |\sigma_\mu(s) - \sigma_\nu(s)| \mathrm{d}s \right)^p \leq \|l_K(\cdot)\|^{p/q}_{L^1([0,t])} \int_0^t l_K(s) |\sigma_\mu(s) - \sigma_\nu(s)|^p \mathrm{d}s,$$

where we used Hölder's inequality. This together with Fubini's theorem further yields

$$\begin{aligned} &\int_{(\mathbb{R}^d \times \Sigma_T)^2} \left(\int_0^t l_K(s) |\sigma_\mu(s) - \sigma_\nu(s)| \mathrm{d}s \right)^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x, \sigma_\mu, y, \sigma_\nu) \\ \leq\ & \|l_K(\cdot)\|^{p/q}_{L^1([0,t])} \int_0^t l_K(s) \int_{(\mathbb{R}^d \times \Sigma_T)^2} |\sigma_\mu(s) - \sigma_\nu(s)|^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x, \sigma_\mu, y, \sigma_\nu) \mathrm{d}s \\ =\ & \|l_K(\cdot)\|^{p/q}_{L^1([0,t])} \int_0^t l_K(s) \int_{\mathbb{R}^{2d}} |x-y|^p \mathrm{d}\boldsymbol{\mu}_s(x,y) \mathrm{d}s \\ =\ & \|l_K(\cdot)\|^{p/q}_{L^1([0,t])} \int_0^t l_K(s) W_p^p(\mu(s), \nu(s)) \mathrm{d}s. \end{aligned} \tag{A.10}$$

Merging (A.9)-(A.10), we obtain

$$W_p^p(\mu(t),\nu(t)) \leq 2^{p-1}\int_{(\mathbb{R}^d\times\Sigma_T)^2}\left(|x-y|+\int_0^t \|v(s,\cdot)-w(s,\cdot)\|_{C^0(K,\mathbb{R}^d)}\,\mathrm{d}s\right)^p \mathrm{d}\hat{\boldsymbol{\eta}}_{\mu,\nu}(x,\sigma_\mu,y,\sigma_\nu)$$
$$+\,2^{p-1}\,\|l_K(\cdot)\|^{p/q}_{L^1([0,t])}\int_0^t l_K(s)W_p^p(\mu(s),\nu(s))\mathrm{d}s. \tag{A.11}$$

As before, applying Grönwall Lemma to (A.11), raising the resulting inequality to the power $1/p$ and applying the triangle inequality for the $L^p(\mathbb{R}^d,\mathbb{R};\mu^0)$-norm, we finally recover

$$W_p(\mu(t),\nu(t)) \leq\, C_p\left(W_p(\mu(0),\nu(0))+\int_0^t \|v(s,\cdot)-w(s,\cdot)\|_{C^0(K,\mathbb{R}^d)}\,\mathrm{d}s\right)\exp\left(C_p'\,\|l_K(\cdot)\|^p_{L^1([0,t])}\right),$$

where the constants C_p, C_p' are as in (15), which concludes the proof of (14). □

Acknowledgement. This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-18-1-0254.
The authors are also grateful to the referee for constructive comments that helped improving the earlier version of this manuscript.

References

[1] Y. Achdou and M. Laurière. On the System of Partial Differential Equations Arising in Mean Field type Control. *Disc. and Cont. Dynamical Systems*, 35(9):3879–3900, 2015.

[2] G. Albi, D Balagué, J.A. Carrillo, and J. von Brecht. Stability Analysis of Flock and Mill Rings for Second Order Models in Swarming. *SIAM J. App. Math.*, 74(3):794–818, 2014.

[3] L. Ambrosio. Transport Equation and Cauchy Problem for BV Vector Fields. *Inventiones Mathematicae*, 158(2):227–260, 2004.

[4] L. Ambrosio. The Flow Associated to Weakly Differentiable Vector Fields: Recent Results and Open Problems. *Nonlinear Conservation Laws and Applications*, 153:181–193, 2011.

[5] L. Ambrosio, M. Colombo, and A. Figalli. Existence and Uniqueness of Maximal Regular Flows with Non-smooth Vector Fields. *Archive for Rational Mechanics and Analysis*, 218(2):1043–1081, 2015.

[6] L. Ambrosio and G. Crippa. Continuity Equations and ODE Flows with Non-Smooth Velocities. *Proceedings of the Royal Society of Edinburgh*, 144(6):1191–1244, 2014.

[7] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variations and Free Discontinuity Problems.* Oxford Mathematical Monographs, 2000.

[8] L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein Space of Probability Measures. *Communications on Pure and Applied Mathematics*, 61(1):18–53, 2008.

[9] L. Ambrosio and N. Gigli. *A User's Guide to Optimal Transport.* In:Modelling and Optimisation on Flows on Networks. Lecture Notes in Mathematics. Springer, 2012.

[10] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, 2008.

[11] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and Heat Flow in Metric Measure Spaces and Applications to Spaces with Ricci Bounds from Below. *Inventiones Mathematicae*, 195(2):289–391, 2014.

[12] J.P. Aubin and A. Cellina. *Differential Inclusions.* Spinger-Verlag, 1984.

[13] J.P. Aubin and H. Frankowska. *Set-Valued Analysis.* Birkhäuser Basel, 1990.

[14] M. Ballerini, N. Cabibbo, R. Candelier, et al. Interaction Ruling Animal Collective Behavior Depends on Topological Rather than Metric Distance: Evidence from a Field Study. *Proceedings of the national academy of sciences*, 105(4):1232–1237, 2008.

[15] R.W. Beard and W. Ren. *Distributed Consensus in Multi-Vehicle Cooperative Control.* Springer-Verlag, 2008.

[16] N. Bellomo, M. A. Herrero, and A. Tosin. On the Dynamics of Social Conflicts: Looking for the Black Swan. *Kinetic & Related Models*, 6(3):459–479, 2013.

[17] J.D. Benamou and Y. Brenier. A Computational Fluid Mechanics Solution to the Monge-Kantorovich Mass Transfer Problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[18] P. Bernard. Young Measures, Superposition and Transport. *Indiana Univ. Math. Journal*, 57(1):247–275, 2008.

[19] A.L. Bertozzi and C.M. Topaz. Swarming Patterns in a Two-Dimensional Kinematic Model for Biological Groups. *SIAM J. App. Math.*, 65(1):152–174, 2004.

[20] S. Bianchini and P. Bonicatto. A Uniqueness Result for the Decomposition of Vector Fields in $\mathbb{R}^d$. *Inventiones Matematicae*, 220:255–393, 2020.

[21] M. Bongini, M. Fornasier, F. Rossi, and F. Solombrino. Mean Field Pontryagin Maximum Principle. *Journal of Optimization Theory and Applications*, 175:1–38, 2017.

[22] B. Bonnet. A Pontryagin Maximum Principle in Wasserstein Spaces for Constrained Optimal Control Problems. *ESAIM COCV*, 25(52), 2019.

[23] B. Bonnet and F. Rossi. Intrinsic Lipschitz Regularity of Mean-Field Optimal Controls. *Submitted*, 2019.

[24] B. Bonnet and F. Rossi. The Pontryagin Maximum Principle in the Wasserstein Space. *Calculus of Variations and Partial Differential Equations*, 58:11, 2019.

[25] H. Brézis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer, 2010.

[26] F. Bullo, J. Cortés, and S. Martines. *Distributed Control of Robotic Networks*. Applied Mathematics. Princeton University Press, 2009.

[27] M. Burger, R. Pinnau, O. Totzeck, and O. Tse. Mean-Field Optimal Control and Optimality Conditions in the Space of Probability Measures. 2019. arXiv:1902.05339.

[28] M. Burger, R. Pinnau, O. Totzeck, O. Tse, and A. Roth. Instantaneous Control of Interacting Particle Systems in the Mean-Field Limit. *Journal of Computational Physics*, 405:109–181, 2020.

[29] G. Cavagnari, A. Marigonda, K.T. Nguyen, and F.S. Priuli. Generalized Control Systems in the Space of Probability Measures. *Set-Valued and Var. Analysis*, 26(3):663–691, 2018.

[30] G. Cavagnari, A. Marigonda, and B. Piccoli. Superposition Principle for Differential Inclusions. In *Large-Scale Scientific Computing*, pages 201–209, 2018.

[31] G. Cavagnari, A. Marigonda, and B. Piccoli. Generalized Dynamic Programming Principle and Sparse Mean-Field Control Problems. *Journal of Mathematical Analysis and Applications*, 481(1):123437, 2020.

[32] E. Cristiani, B. Piccoli, and A. Tosin. *Multiscale Modeling of Pedestrian Dynamics*, volume 12. Springer, 2014.

[33] F. Cucker and S. Smale. On the Mathematics of Emergence. *Japanese Journal of Mathematics*, 2(1):197–227, 2007.

[34] R.L. Di Perna and Lions P.-L. Ordinary Differential Equations, Transport Theory and Sobolev Spaces. *Inventiones Mathematicae*, 98(3):511–548, 1989.

[35] J. Diestel and J.J.Jr Uhl. *Vector Measures*, volume 15. American Mathematical Society, 1977.

[36] M. Duprez, M. Morancey, and F. Rossi. Approximate and Exact Controllability of the Continuity Equation with a Localized Vector Field. *SIAM Journal on Control and Optimization*, 57(2):1284–1311, 2019.

[37] M. Duprez, M. Morancey, and F. Rossi. Minimal Time Problem for Crowd Models with a Localized Vector Field. *Journal of Differential Equations*, 269(1):82–124, 2020.

[38] M. Fornasier, S. Lisini, C. Orrieri, and G. Savaré. Mean-Field Optimal Control as Gamma-Limit of Finite Agent Controls. *European Journal of Applied Mathematics*, 30(6):1153–1186, 2019.

[39] M. Fornasier, B. Piccoli, and F. Rossi. Mean-Field Sparse Optimal Control. *Philosophical Transaction of the Royal Society A.*, 372(2028), 2014.

[40] M. Fornasier and F. Solombrino. Mean Field Optimal Control. *Esaim COCV*, 20(4):1123–1152, 2014.

[41] H. Frankowska. A Priori Estimates for Operational Differential Inclusions. *Journal of Differential Equations*, 84:100–128, 1990.

[42] H. Frankowska, E.M. Marchini, and M. Mazzola. Necessary Optimality Conditions for Infinite Dimensional State Constrained Control Problems. *Journal of Differential Equations*, 264(12):7294–7327, 2018.

[43] W. Gangbo, H.K. Kim, and T. Paccini. Differential Form on Wasserstein Space and Infinite-Dimensional Hamiltonian Systems. *Memoirs of the AMS*, 211(993), 2011.

[44] R. Hegselmann and U. Krause. Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

[45] M.W. Hirsch.

[46] M.Y. Huang, R. Malhamé, and P.E. Caines. Large Population Stochastic Dynamic Games : Closed-Loop McKean-Vlasov Systems and the Nash Certainty Equivalence Principle. *Communications in Information and Systems*, 6(3):221–252, 2006.

[47] C. Jimenez, A. Marigonda, and M. Quincampoix. Optimal Control of Multiagent Systems in the Wasserstein Space. *Calculus of Variations and Partial Differential Equations*, 59:58, 2020.

[48] S. Karimghasemi, S. Müller, and M. Westdickenberg. Flow Solutions of Transport Equations. *arXiv:1912.06815v2*, 2020.

[49] J-M. Lasry and P.-L. Lions. Mean Field Games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.

[50] S. McQuade, B. Piccoli, and N. Pouradier Duteil. Social Dynamics Models with Time-Varying Influence. *Mathematical Models and Methods in Applied Sciences*, 29(04):681–716, 2019.

[51] M. Mesbahi and M. Egerstedt. *Graph Theoretic Multi-Agent Systems*. 2010.

[52] F. Otto. The Geometry of Dissipative Equations : The Porous Medium Equation. *Communications in Partial Differential Equations*, 26:101–174, 2001.

[53] B. Piccoli. Measure Differential Equations. *Archive for Rational Mechanics and Analysis*, 233:1289–1317, 2019.

[54] B. Piccoli. Measure Differential Inclusions. In *IEEE Conference on Decision and Control*, pages 1323–1328, 2020.

[55] B. Piccoli and F. Rossi. Transport Equation with Nonlocal Velocity in Wasserstein Spaces : Convergence of Numerical Schemes. *Acta App. Math.*, 124(1):73–105, 2013.

[56] N. Pogodaev. Optimal Control of Continuity Equations. *Nonlinear Differential Equations and Applications*, 23:21, 2016.

[57] N. Pogodaev. Numerical Algorithm for Optimal Control of Continuity Equations. 2017. arXiv:1708.05516.

[58] F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87. Birkhauser Basel, 2015.

[59] C. Villani. *Optimal Transport : Old and New*. Springer-Verlag, Berlin, 2009.

[60] R.B. Vinter. *Optimal Control*. Systems and Control: Foundations and Applications. Birkhauser Basel, 2000.

[61] A.A. Vlasov. *Many-Particle Theory and its Application to Plasma*. New York, Gordon and Breach, 1961.