Superposition Principle for Differential Inclusions

Giulia Cavagnari[1], Antonio Marigonda[2](✉), and Benedetto Piccoli[1]

[1] Department of Mathematical Sciences, Rutgers University - Camden, 311 N. 5th Street, Camden, NJ 08102, USA
giulia.cavagnari@rutgers.edu, piccoli@camden.rutgers.edu
[2] Department of Computer Sciences, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy
antonio.marigonda@univr.it

Abstract. We prove an extension of the Superposition Principle by Ambrosio-Gigli-Savaré in the context of a control problem. In particular, we link the solutions of a finite-dimensional control system, with dynamics given by a differential inclusion, to a solution of a continuity equation in the space of probability measures with admissible vector field. We prove also a compactness and an approximation result for admissible trajectories in the space of probability measures.

Keywords: Continuity equation · Differential inclusions
Optimal transport · Superposition principle

1 Introduction

This paper aims to provide a relation between the *macroscopic* and the *microscopic* approaches describing the evolution of a mass of particles/agents in a controlled context. The microscopic dynamics of the particles/agents is governed by a control system given in the form of a differential inclusion $\dot{x}(t) \in F(x(t))$, where $F(\cdot)$ is a given set-valued function stating the set of admissible velocities for each point in $\mathbb{R}^d$. This makes not trivial the construction of the corresponding macroscopic evolution and of its driving vector field in the space of probability measures. Indeed, from a macroscopic point of view, the evolving mass is described by a time-dependent family of probability measures $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]}$, solving in the distributional sense a (*controlled*) homogeneous continuity equation (thus a PDE), and driven by an admissible vector field that has to be chosen among the $L^1_{\mu_t}$-selections of F.

In a non-controlled framework, if the finite-dimensional dynamics is given by an ODE driven by a Lipschitz vector field v_t (locally Lipschitz continuous in the space variable uniformly w.r.t. t), then we have existence and uniqueness of the solution of the PDE. The solution μ_t at time t of the continuity equation is characterized by the push forward of the initial state μ_0 w.r.t. a map T_t called *transport map*, i.e. $\mu_t = T_t \sharp \mu_0$ for a.e. t, where $\dot{T}_t(x) = v_t(T_t(x))$, $T_0(x) = x$ is the

I. Lirkov and S. Margenov (Eds.): LSSC 2017, LNCS 10665, pp. 201–209, 2018.
https://doi.org/10.1007/978-3-319-73441-5_21

characteristic system. However, a relation between μ_t and the (integral) solutions of the characteristic system is possible even for nonsmooth vector fields, where uniqueness of solutions is no longer granted. This powerful result, called the *Superposition Principle*, appeared for the first time in the appendix of [17] and has been studied by different authors in [1,2,4], and in [15] in a non-homogeneous context. The idea is to take into account the possible non-uniqueness of the solution of the characteristic system by introducing a measure $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$, where $\Gamma_T = C^0([0,T];\mathbb{R}^d)$, concentrated on the set of $(\gamma(0),\gamma)$ where γ is any integral solution of the characteristic system. Indeed, under general assumptions, for any solution $\boldsymbol{\mu} := \{\mu_t\}_{t\in[0,T]} \subseteq \mathscr{P}(\mathbb{R}^d)$ of an homogeneous continuity equation, there exists such a (possibly not-unique) *representation* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R} \times \Gamma_T)$ satisfying $\mu_t = e_t \sharp \boldsymbol{\eta}$ where $e_t : \mathbb{R}^d \times \Gamma_T \to \mathbb{R}^d, (x,\gamma) \mapsto \gamma(t)$, is the *evaluation operator*. Conversely, any $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ concentrated on the characteristics yields a solution of the correspondent continuity equation by setting $\mu_t = e_t \sharp \boldsymbol{\eta}$.

We stress the fact that, given $\boldsymbol{\mu}$, its probabilistic representation $\boldsymbol{\eta}$ may be not unique, in particular different weights to the characteristics could lead to the same macroscopic evolution $\boldsymbol{\mu}$ (some examples are sketched in the forthcoming [8]). In the present paper we exploit the non-uniqueness of a probabilistic representative by extending the reverse implication of the Superposition Principle (see Theorem 8.2.1 in [2]) in a controlled setting. In particular, we replace the underlying characteristics' ODE with a differential inclusion. In Theorem 1 we prove that, under some natural assumptions of the set valued map F, a measure $\boldsymbol{\eta}$ concentrated on the Carathéodory solutions of the differential inclusion $\dot{\gamma}(t) \in F(\gamma(t))$, $\gamma(0) = x$, induces a macroscopic admissible trajectory $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]} \subseteq \mathscr{P}(\mathbb{R}^d)$, where μ_t is a solution of a continuity equation driven by a *mean vector field* v_t. More precisely, $v_t(y)$ turns out to be the integral average w.r.t. $\boldsymbol{\eta}$ of the underlying admissible vector fields crossing position $y \in \mathbb{R}^d$ at time t. In other words, the macroscopic evolution μ_t of our mass looses the information about the velocity field chosen by each single particle, providing only their average behaviour.

The results of this paper could be used to investigate further properties of control problems in $\mathscr{P}(\mathbb{R}^d)$, possibly requiring extremality conditions (e.g. time minimality to reach a target). For instance, one may improve the analysis made in [5–12] where the authors studied time-optimal control problems in the space of measures making large use of the Superposition Principle of [2]. Another potential application could be in the field of crowd dynamics, where the importance of a multiscale approach has been underlined for instance in [13,14]. Indeed, we can now collect together the microscopic behaviour of the single agents, even when they are subject to different vector fields, into a unique macroscopic mean description.

The paper is structured as follows: in Sect. 2 we state the notation and define the objects used, Sect. 3 contains the statement and proof of the extended Superposition Principle together with a compactness and approximation result.

2 Preliminaries

Let X be a separable metric space. We denote with $\mathscr{P}(X)$ the space of Borel probability measures on X endowed with the narrow topology induced by $(C_b^0(X))'$. When $p \geq 1$, $\mathscr{P}_p(X)$ denotes the space of Borel probability measures with finite *p-moment*, i.e., the measures $\mu \in \mathscr{P}(\mathbb{R}^d)$ satisfying $\mathrm{m}_p(\mu) := \int_{\mathbb{R}^d} |x|^p \, d\mu < +\infty$, endowed with the topology induced by the *p-Wasserstein distance* $W_p(\cdot, \cdot)$. We call $\mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)$ the space of vector-valued Radon measures on $\mathbb{R}^d$ endowed with the w^*-topology. When $\nu \in \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)$, $|\nu|$ denotes its *total variation*, and we write $\sigma \ll \mu$ to say that σ is *absolutely continuous* w.r.t. μ, for a pair of measures σ, μ on $\mathbb{R}^d$. Preliminaries on measure theory can be found in Chap. 5 in [2].

We recall now the definition of *admissible trajectory* in $\mathscr{P}(\mathbb{R}^d)$ that, together with its probabilistic representation, is the central object of the present paper and it was introduced in [5,6,8–11] for the study of time-optimal control problems in the space of measures.

Definition 1. *Let $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be a set-valued map, $\bar{\mu} \in \mathscr{P}(\mathbb{R}^d)$.*

1. *Let $T > 0$. We say that $\boldsymbol{\mu} = \{\mu_t\}_{t \in [0,T]} \subseteq \mathscr{P}(\mathbb{R}^d)$ is an* admissible trajectory *defined on $[0, T]$ and starting from $\bar{\mu}$ if there exists $\boldsymbol{\nu} = \{\nu_t\}_{t \in [0,T]} \subseteq \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)$ such that $|\nu_t| \ll \mu_t$ for a.e. $t \in [0, T]$, $\mu_0 = \bar{\mu}$, $\partial_t \mu_t + \operatorname{div} \nu_t = 0$ in the sense of distributions and $v_t(x) := \dfrac{\nu_t}{\mu_t}(x) \in F(x)$ for a.e. $t \in [0, T]$ and μ_t-a.e. $x \in \mathbb{R}^d$. In this case, we will say also that $\boldsymbol{\mu}$ is* driven *by $\boldsymbol{\nu}$.*
2. *Let $T > 0$, $\boldsymbol{\mu}$ be an admissible trajectory defined on $[0, T]$ starting from $\bar{\mu}$ and driven by $\boldsymbol{\nu} = \{\nu_t\}_{t \in [0,T]}$. We will say that $\boldsymbol{\mu}$* is represented by *$\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ if we have $e_t \sharp \boldsymbol{\eta} = \mu_t$ for all $t \in [0, T]$, where $e_t : \mathbb{R}^d \times \Gamma_T \to \mathbb{R}^d, (x, \gamma) \mapsto \gamma(t)$, and $\boldsymbol{\eta}$ is concentrated on the pairs $(x, \gamma) \in \mathbb{R}^d \times \Gamma_T$ where γ is an absolutely continuous solution of the underlying characteristic system*

$$\begin{cases} \dot{\gamma}(t) \in F(\gamma(t)), & \text{for a.e. } 0 < t \leq T \\ \gamma(0) = x. \end{cases} \tag{1}$$

Note that to have the existence of a probabilistic representation $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ it is sufficient that the driving vector field associated with $\boldsymbol{\mu}$ satisfies the integrability hypothesis of the Superposition Principle (see Theorem 8.2.1 in [2]).

Finally, let X be a set, $A \subseteq X$. The *indicator function of* A is $I_A : X \to \{0, +\infty\}$ defined as $I_A(x) = 0$ for all $x \in A$ and $I_A(x) = +\infty$ for all $x \notin A$. The *characteristic function of* A is the function $\chi_A : X \to \{0, 1\}$ defined as $\chi_A(x) = 1$ for all $x \in A$ and $\chi_A(x) = 0$ for all $x \notin A$.

3 Results

Throughout the paper we will require the following assumptions on the set-valued function $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ governing the finite-dimensional differential inclusion:

(F_0) $F(x) \neq \emptyset$ is compact and convex for every $x \in \mathbb{R}^d$, moreover $F(\cdot)$ is continuous with respect to the Hausdorff metric.

(F_1) $F(\cdot)$ has linear growth, i.e. there exists a constant $C > 0$ such that $F(x) \subseteq \overline{B(0, C(|x|+1))}$ for every $x \in \mathbb{R}^d$.

The following simple Lemma states the possibility to approximate in the W_p-distance every measure $\mu \in \mathscr{P}_p(X)$, where X is a complete separable Banach space, by a sequence $\{\mu^k\}_{k\in\mathbb{N}}$ of empirical measures (i.e. convex combinations of Dirac deltas) concentrated on its support. A proof can be found for instance in [16] (see Lemma 6.1) for the case $X = \mathbb{R}^d$, but it can be easily extended to the general setting of complete separable Banach spaces.

Lemma 1 (Empirical approximation in Wasserstein). *Let X be a separable Banach space. For all $p \geq 1$ we have $\mathscr{P}_p(X) = \mathrm{cl}_{W_p}(\mathrm{co}\{\delta_x : x \in X\})$.*

We now consider the following problem: taking any probability measure $\boldsymbol{\eta}$ on the set of the admissible trajectories for (1), it is possible to construct a global vector field $v_t(\cdot)$, time-depending selection of $F(\cdot)$, such that $e_t\sharp\boldsymbol{\eta}$ yields a family of time-depending probability measures on $\mathbb{R}^d$ solving the continuity equation driven by v_t. This can be viewed as a partial extension to the Superposition Principle (see Theorem 8.2.1 in [2]) to the case of differential inclusions.

Theorem 1 (SP for differential inclusions). *Assume (F_0), (F_1), $p \geq 1$. Let $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ be concentrated on the set of pairs $(\gamma(0), \gamma) \in \mathbb{R}^d \times \Gamma_T$ such that $\gamma \in AC([0,T];\mathbb{R}^d)$ is a Carathéodory solution of the differential inclusion $\dot{\gamma}(t) \in F(\gamma(t))$. For all $t \in [0,T]$, set $\mu_t := e_t\sharp\boldsymbol{\eta}$, and let $\{\eta_{t,y}\}_{y\in\mathbb{R}^d} \subseteq \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ be the disintegration of $\boldsymbol{\eta}$ w.r.t. the evaluation operator $e_t : \mathbb{R}^d \times \Gamma_T \to \mathbb{R}^d$, i.e. for all $\varphi \in C_b^0(\mathbb{R}^d \times \Gamma_T)$*

$$\iint_{\mathbb{R}^d \times \Gamma_T} \varphi(x,\gamma)\, d\boldsymbol{\eta}(x,\gamma) = \int_{\mathbb{R}^d} \int_{e_t^{-1}(y)} \varphi(x,\gamma)\, d\eta_{t,y}(x,\gamma)\, d\mu_t(y).$$

Then if $\mu_0 \in \mathscr{P}_p(\mathbb{R}^d)$, the curve $\boldsymbol{\mu} := \{\mu_t\}_{t\in[0,T]} \subseteq \mathscr{P}_p(\mathbb{R}^d)$, is an admissible trajectory driven by $\boldsymbol{\nu} = \{\nu_t\}_{t\in[0,T]}$, where $\nu_t = v_t\mu_t$ and the vector field

$$v_t(y) = \int_{e_t^{-1}(y)} \dot{\gamma}(t)\, d\eta_{t,y}(x,\gamma). \tag{2}$$

is well-defined for a.e. $t \in [0,T]$ and μ_t-a.e. $y \in \mathbb{R}^d$.

Proof. We define

$$\mathscr{N} := \{(t,x,\gamma) \in [0,T] \times \mathbb{R}^d \times \Gamma_T : \text{ either } \dot{\gamma}(t) \text{ does not exists or } \dot{\gamma}(t) \notin F(\gamma(t))\}.$$

Since $\mathscr{L}^1_{|[0,T]} \otimes \boldsymbol{\eta}(\mathscr{N}) = 0$, we have $\dot{\gamma}(t) \in F(\gamma(t))$ for $\boldsymbol{\eta}$-a.e. $(x,\gamma) \in \mathbb{R}^d \times \Gamma_T$ and a.e. $t \in [0,T]$, and so $v_t(y)$ is well-defined for a.e. $t \in [0,T]$ and μ_t-a.e. $y \in \mathbb{R}^d$.

We prove first that the map $t \mapsto \mu_t$ is Lipschitz continuous from $[0,T]$ to $\left(C_c^1(\mathbb{R}^d)\right)'$. For all $\tau \in [0,T]$ and $\boldsymbol{\eta}$-a.e. $(x,\gamma) \in \mathbb{R}^d \times \Gamma_T$ it holds

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

thus, by Gronwall's inequality,

$$|\gamma(\tau) - \gamma(0)| \le C\tau(1 + |\gamma(0)|)e^{C\tau} \le CTe^{CT}(1 + |\gamma(0)|),$$

Since for any $\varphi \in C_c^1(\mathbb{R}^d)$ we have

$$\begin{aligned} \left|\int_{\mathbb{R}^d} \varphi(x)\, d\mu_s(x) - \int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x)\right| &\le \int_s^t \iint_{\mathbb{R}^d\times\Gamma_T} |\langle \nabla\varphi(\gamma(\tau)), \dot\gamma(\tau)\rangle|\, d\boldsymbol{\eta}(x,\gamma)\, d\tau \\ &\le C\|\nabla\varphi\|_\infty \int_s^t \iint_{\mathbb{R}^d\times\Gamma_T} (|\gamma(\tau)| + 1)\, d\boldsymbol{\eta}(x,\gamma)\, d\tau \\ &\le C(CTe^{CT} + 1)\|\nabla\varphi\|_\infty \int_s^t \iint_{\mathbb{R}^d\times\Gamma_T} (|\gamma(0)| + 1)\, d\boldsymbol{\eta}(x,\gamma)\, d\tau \\ &\le C(CTe^{CT} + 1)\left(\mathrm{m}_p^{1/p}(\mu_0) + 1\right)\|\nabla\varphi\|_\infty |t - s|, \end{aligned}$$

we have $\|\mu_s - \mu_t\|_{(C_c^1(\mathbb{R}^d))'} \le C(CTe^{CT} + 1)\left(\mathrm{m}_p^{1/p}(\mu_0) + 1\right)|t - s|$.

According to Theorem 3.5 in [3], we have that for a.e. $t \in [0,T]$ the map $t \mapsto \mu_t$ is differentiable, and for all $\varphi \in C_c^1(\mathbb{R}^d)$

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x) = \iint_{\mathbb{R}^d\times\Gamma_T} \nabla\varphi(\gamma(t))\cdot\dot\gamma(t)\, d\boldsymbol{\eta}(x,\gamma) = \int_{\mathbb{R}^d} \nabla\varphi(y)\cdot v_t(y)\, d\mu_t(y),$$

which implies $\partial_t\mu_t + \operatorname{div}\nu_t = 0$ with $\nu_t = v_t\mu_t$. Finally, thanks to the convexity of $F(y)$, we can use Jensen's inequality to get that $v_t(y) \in F(y)$ for μ_t-a.e. $y \in \mathbb{R}^d$ and a.e. $t \in [0,T]$. To conclude the proof, it is enough to show the estimates on the p-moments of μ_t. Indeed, by Gronwall's inequality we have

$$\mathrm{m}_p^{1/p}(\mu_t) \le (CTe^{CT} + 1)(1 + \mathrm{m}_p^{1/p}(\mu_0)).$$

Moreover, by (F_1) we have that every Borel selection of $F(\cdot)$ is in L^p_μ for any $\mu \in \mathscr{P}_p(\mathbb{R}^d)$, hence $v_t \in L^p_{\mu_t}$ for a.e. $t \in [0,T]$. □

A possible interpretation of $v_t(y)$ is provided by the following remark.

Remark 1. By definition, we have $e_t^{-1}(y) = \{(x,\gamma) \in \mathbb{R}^d \times \Gamma_T : \gamma(t) = y\}$, so, by (2), for a.e. $t \in [0,T]$ and μ_t-a.e. $y \in \mathbb{R}^d$, we have that $v_t(y)$ corresponds to a weighted average of the velocity of the trajectories $\gamma \in AC([0,T];\mathbb{R}^d)$ of the differential inclusions $\dot\gamma(t) \in F(\gamma(t))$ satisfying $\gamma(t) = y$.

The next example provides a situation where the velocities of a nonnegligible set of curves differs from the mean field for a nonnegligible amount of time.

Example 1. The ambient space is $\mathbb{R}^2$. Define

- $\mathscr{A} = \{\gamma_{x,y}(\cdot)\}_{(x,y)\in\mathbb{R}^2} \subseteq AC([0,2])$ where $\gamma_{x,y}(t) = (x+t, y-t\,\mathrm{sgn}\,y)$ for any $(x,y)\in\mathbb{R}^2$, $t\in[0,2]$, and we set $\mathrm{sgn}(0)=0$;
- $F:\mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ by $F(x,y)\equiv[-1,1]\times[-1,1]$ for all $(x,y)\in\mathbb{R}^2$;
- $\mu_0 = \frac{1}{2}\delta_0\otimes\mathscr{L}^1_{|[-1,1]} \in \mathscr{P}(\mathbb{R}^2)$, $\boldsymbol{\eta} = \mu_0\otimes\delta_{\gamma_{x,y}} \in \mathscr{P}(\mathbb{R}^2\times\Gamma_2)$, $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,2]}$ with $\mu_t = e_t\sharp\boldsymbol{\eta}$;
- Q be the open square of vertice $\{(0,0),(1,0),(1/2,\pm1/2)\}$.

We notice that

- F satisfies (F_0) and (F_1) and $\dot\gamma(t)\in F(\gamma(t))$ for all $\gamma\in\mathscr{A}$ and $t\in]0,2[$.
- The product measure $\boldsymbol{\eta}$ is well-defined since $(x,y)\mapsto\gamma_{x,y}(\cdot)$ is a Borel map, thus $\boldsymbol{\mu}$ is an admissible trajectory and we denote with $\boldsymbol{\nu} = \{\nu_t\}_{t\in[0,2]}$ its driving family of Borel vector-valued measures.
- For any $P=(p_x,p_y)\in Q$ with $p_y\neq0$ there are exactly two elements $\gamma\in\mathscr{A}$ satisfying $\gamma(0)\in\{0\}\times]-1,1[$ and crossing at P. These elements are $\gamma_{0,p_y\pm p_x}(\cdot)$ and we notice that $P=\gamma_{0,p_y+p_x}(t)=\gamma_{0,p_y-p_x}(t)$ if and only if $t=p_x$.

Denoted by $v_t = \dfrac{\nu_t}{\mu_t}$ the mean vector field, this implies $v_t(x,y)=(1,0)$ for all $(x,y)\in Q\setminus(\mathbb{R}\times\{0\})$ and $t=x$. For every $\gamma\in\mathscr{A}$ satisfying $\gamma(0)=\{0\}\times]-1,1[$ and $\gamma(0)\neq(0,0)$, there exists an interval $I_\gamma\subseteq[0,1]$ of Lebesgue measure $1/2$ such that $\gamma(t)\in Q$ if and only if $t\in I_\gamma$, thus

$$\mathscr{L}^1_{|[0,2]}\otimes\boldsymbol{\eta}\left(\{(t,x,\gamma)\in[0,2]\times\mathbb{R}^2\times\Gamma_2:\dot\gamma(t)\neq v_t(\gamma(t))\}\right)=\frac{1}{2}.$$

With techniques similar to Theorem 1, it is possible to prove a result of relative compactness of the admissible trajectories even in the critical case $p=1$.

Proposition 1 (Relative compactness of admissible trajectories). *Assume (F_0), (F_1), $p\geq1$. Let $\{\boldsymbol{\eta}^N\}_{N\in\mathbb{N}}\subseteq\mathscr{P}(\mathbb{R}^d\times\Gamma_T)$ be a sequence of measures concentrated on the set of pairs $(\gamma(0),\gamma)\in\mathbb{R}^d\times\Gamma_T$ where $\gamma\in AC([0,T];\mathbb{R}^d)$ is a Carathéodory solution of the differential inclusion $\dot\gamma(t)\in F(\gamma(t))$ and such that $\{\mathrm{m}_p(e_0\sharp\boldsymbol{\eta}^N)\}_{N\in\mathbb{N}}$ is uniformly bounded. Denote with $\{\boldsymbol{\mu}^N\}_{N\in\mathbb{N}}$ the sequence of admissible trajectories represented by $\{\boldsymbol{\eta}^N\}_{N\in\mathbb{N}}$, and with $\{\boldsymbol{\nu}^N\}_{N\in\mathbb{N}}\subseteq\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)$ the sequence of their driving families of Borel vector-valued measures.*

Then, up to a non relabeled subsequence, we have that there exists $\boldsymbol{\eta}\in\mathscr{P}(\mathbb{R}^d\times\Gamma_T)$ such that $\boldsymbol{\eta}^N\rightharpoonup^*\boldsymbol{\eta}$, and $\boldsymbol{\mu}:=\{\mu_t\}_{t\in[0,T]}\subseteq\mathscr{P}_p(\mathbb{R}^d)$ defined by $\mu_t=e_t\sharp\boldsymbol{\eta}$ is an admissible curve driven by $\boldsymbol{\nu}=\{\nu_t\}_{t\in[0,T]}$, with $\nu_t^N\rightharpoonup^*\nu_t$ for a.e. $t\in[0,T]$.

Proof. We prove that $\{\boldsymbol{\eta}^N\}_{N\in\mathbb{N}}$ is relatively compact in $\mathscr{P}(\mathbb{R}^d \times \Gamma_T)$. Indeed, by exploiting the estimates of Theorem 1, and the uniformly boundedness of $\mathrm{m}_p(e_0\sharp\boldsymbol{\eta}^N)$, we have

$$\iint_{\mathbb{R}^d\times\Gamma_T} (|x| + |\gamma(t)|)\, d\boldsymbol{\eta}^N(x,\gamma) \le (CTe^{CT} + 2)(1 + \mathrm{m}_p^{1/p}(e_0\sharp\boldsymbol{\eta}^N)) \le K < +\infty,$$

by Remark 5.1.5 in [2] we have that $\{\boldsymbol{\eta}^N\}_{N\in\mathbb{N}}$ is tight, and so, up to a non relabeled subsequence, we have that there exists $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$, concentrated on the set of pairs $(\gamma(0),\gamma) \in \mathbb{R}^d \times \Gamma_T$ where $\gamma \in AC([0,T];\mathbb{R}^d)$ is a Carathéodory solution of the differential inclusion $\dot\gamma(t) \in F(\gamma(t))$, such that $\boldsymbol{\eta}^N \rightharpoonup^* \boldsymbol{\eta}$.

In particular, for all $t \in [0,T]$ we have $\mu_t^N = e_t\sharp\boldsymbol{\eta}^N \rightharpoonup^* e_t\sharp\boldsymbol{\eta} = \mu_t$. By Theorem 1 we have that $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]}$ is an admissible trajectory and it is driven by $\boldsymbol{\nu} = \{\nu_t\}_{t\in[0,T]}$, with $\nu_t = v_t\mu_t$ and v_t is a suitable $L^p_{\mu_t}$-selection of $F(\cdot)$ for a.e. $t \in [0,T]$.

Let us now conclude by proving that $\nu_t^N \rightharpoonup^* \nu_t$ for all $N \in \mathbb{N}$ and for a.e. $t \in [0,T]$. By Theorem 1, by w^*-convergence of μ_t^N to μ_t and by admissibility of $\boldsymbol{\mu}^N$, we have that for every $\varphi \in C_c^1([0,T]\times\mathbb{R}^d)$

$$\begin{aligned}
&-\iint_{[0,T]\times\mathbb{R}^d} \nabla\varphi(t,x)\cdot d\nu_t\, dt = \iint_{[0,T]\times\mathbb{R}^d} \partial_t\varphi(t,x)\, d\mu_t\, dt \\
&= \lim_{N\to+\infty} \iint_{[0,T]\times\mathbb{R}^d} \partial_t\varphi(t,x)\, d\mu_t^N\, dt = \lim_{N\to+\infty} -\iint_{[0,T]\times\mathbb{R}^d} \nabla\varphi(t,x)\, d\nu_t^N\, dt,
\end{aligned}$$

hence the statement follows. □

Finally, combining Lemma 1, Theorem 1, and Proposition 1, we have convergence of a suitable discrete approximation.

Corollary 1. *Assume* (F_0), (F_1), $p \ge 1$. *Let* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ *be concentrated on the set of pairs* $(\gamma(0),\gamma) \in \mathbb{R}^d \times \Gamma_T$ *such that* $\gamma \in AC([0,T];\mathbb{R}^d)$ *is a Carathéodory solution of the differential inclusion* $\dot\gamma(t) \in F(\gamma(t))$ *with* $\mathrm{m}_p(e_0\sharp\boldsymbol{\eta}) < +\infty$. *Then there exists a sequence* $\{\boldsymbol{\eta}^N\}_{N\in\mathbb{N}} \subseteq \mathrm{co}\{\delta_x \otimes \delta_{\gamma_x}\} \subseteq \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$ *with* $\gamma_x \in AC([0,T];\mathbb{R}^d)$, $\gamma(0) = x$ *and* $\dot\gamma(t) \in F(\gamma(t))$ *for a.e.* $t \in [0,T]$, *such that* $\boldsymbol{\eta}^N$ *converges to* $\boldsymbol{\eta}$ *in* W_p *and for all* $t \in [0,T]$

$$\lim_{N\to+\infty} W_p(e_t\sharp\boldsymbol{\eta}^N, e_t\sharp\boldsymbol{\eta}) = 0.$$

Proof. We take $X = \mathbb{R}^d \times \Gamma_T$, endowed with norm $\|(x,\gamma)\|_X = |x| + \|\gamma\|_\infty$. We prove that $\mathrm{m}_p(\boldsymbol{\eta}) < +\infty$. Indeed, for all $t \in [0,T]$ and $\boldsymbol{\eta}$-a.e. $(x,\gamma) \in \mathbb{R}^d \times \Gamma_T$ we have $|\gamma(t)| \le (CTe^{CT} + 1)(1 + |\gamma(0)|)$, so $\|\gamma\|_\infty \le (CTe^{CT} + 1)(1 + |\gamma(0)|)$, hence

$$\begin{aligned}
\iint_{\mathbb{R}^d\times\Gamma_T} (|x| + \|\gamma\|_\infty)^p\, d\boldsymbol{\eta}(x,\gamma) &\le 2^p(CTe^{CT} + 1)^p \iint_{\mathbb{R}^d\times\Gamma_T} (1 + |\gamma(0)|)^p\, d\boldsymbol{\eta}(x,\gamma) \\
&= 2^p(CTe^{CT} + 1)^p \int_{\mathbb{R}^d} (1 + |x|)^p\, d(e_0\sharp\boldsymbol{\eta})(x) < +\infty.
\end{aligned}$$

By Lemma 1, we can construct a sequence $\{\boldsymbol{\eta}^N\}_{N\in\mathbb{N}} \subseteq \mathrm{co}\{\delta_x\otimes\delta_{\gamma_x}\} \subseteq \mathscr{P}(\mathbb{R}^d\times\Gamma_T)$ W_p-converging to $\boldsymbol{\eta}$. Moreover, we have $\mathrm{supp}\,\boldsymbol{\eta}^N \subseteq \mathrm{supp}\,\boldsymbol{\eta}$, which, by Theorem 1, implies that $\boldsymbol{\mu}^N = \{\mu_t^N = e_t\sharp\boldsymbol{\eta}^N\}_{t\in[0,T]}$ is an admissible trajectory. □

Acknowledgments. The authors acknowledge the endowment fund of the Joseph and Loretta Lopez Chair and the support of the INdAM-GNAMPA Project 2016 *Stochastic Partial Differential Equations and Stochastic Optimal Transport with Applications to Mathematical Finance*.

References

1. Ambrosio, L.: Transport equation and Cauchy problem for non-smooth vector fields. In: Dacorogna, B., Marcellini, P. (eds.) Contribute in Calculus of Variations and Nonlinear Partial Differential Equations. Lecture Notes in Mathematics, vol. 1927, pp. 1–41. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-75914-0_1
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser Verlag, Basel (2008)
3. Ambrosio, L., Kirchheim, B.: Rectifiable sets in metric and Banach spaces. Math. Ann. **318**(3), 527–555 (2000)
4. Bernard, P.: Young measures, superpositions and transport. Indiana Univ. Math. J. **57**(1), 247–276 (2008)
5. Cavagnari, G.: Regularity results for a time-optimal control problem in the space of probability measures. Math. Control Relat. Fields **7**(2), 213–233 (2017)
6. Cavagnari, G., Marigonda, A.: Time-optimal control problem in the space of probability measures. In: Lirkov, I., Margenov, S.D., Waśniewski, J. (eds.) LSSC 2015. LNCS, vol. 9374, pp. 109–116. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26520-9_11
7. Cavagnari, G., Marigonda, A.: Measure-theoretic Lie brackets for nonsmooth vector fields (Submitted)
8. Cavagnari, G., Marigonda, A., Nguyen, K.T., Priuli, F.S.: Generalized control systems in the space of probability measures. Set-Valued Var. Anal. (2017). https://doi.org/10.1007/s11228-017-0414-y
9. Cavagnari, G., Marigonda, A., Orlandi, G.: Hamilton-Jacobi-Bellman equation for a time-optimal control problem in the space of probability measures. In: Bociu, L., Désidéri, J.-A., Habbal, A. (eds.) CSMO 2015. IAICT, vol. 494, pp. 200–208. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-55795-3_18
10. Cavagnari, G., Marigonda, A., Piccoli, B.: Averaged time-optimal control problem in the space of positive Borel measures. ESAIM: Control Optim. Calc. Var. (2017). https://doi.org/10.1051/cocv/2017060
11. Cavagnari, G., Marigonda, A., Piccoli, B.: Optimal synchronization problem for a multi-agent system. Netw. Heterogen. Media **12**(2), 277–295 (2017). https://doi.org/10.3934/nhm.2017012
12. Cavagnari, G., Marigonda, A., Priuli, F.S.: Attainability property for a probabilistic target in Wasserstein spaces (preprint)
13. Cristiani, E., Piccoli, B., Tosin, A.: Multiscale modeling of granular flows with application to crowd dynamics. Multiscale Model. Simul. **9**(1), 155–182 (2011)

14. Cristiani, E., Piccoli, B., Tosin, A.: Multiscale Modeling of Pedestrian Dynamics. MS&A Modeling, Simulation and Applications, vol. 12. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06620-2
15. Maniglia, S.: Probabilistic representation and uniqueness results for measure-valued solutions of transport equations. J. Math. Pures Appl. **87**(6), 601–626 (2007)
16. Siegfried, G., Harald, L.: Foundations of Quantization for Probability Distributions. Lecture Notes in Mathematics, vol. 1730. Springer, Berlin (2000). https://doi.org/10.1007/BFb0103945
17. Young, L.C.: Lectures on the calculus of variations and optimal control theory. American Mathematical Society, vol. 304. AMS Chelsea Publishing, Providence (1980)