Intrinsic Lipschitz Regularity of Mean-Field Optimal Controls

Benoît Bonnet*, Francesco Rossi†

February 7, 2021

Abstract

In this article, we provide sufficient conditions under which the controlled vector fields solution of optimal control problems formulated on continuity equations are Lipschitz regular in space. Our approach involves a novel combination of mean-field approximations for infinite-dimensional multi-agent optimal control problems, along with a careful extension of an existence result of locally optimal Lipschitz feedbacks. The latter is based on the reformulation of a coercivity estimate in the language of Wasserstein calculus, which is used to obtain uniform Lipschitz bounds along sequences of approximations by empirical measures.

Keywords. Multi-Agent Systems, Mean-Field Optimal Control, Regularity of Minimisers, Wasserstein Calculus

MSC2020 Subject Classification. 35B65, 49J20, 49J30, 49Q22, 58E25, 93A16

1 Introduction

The mathematical analysis of collective behaviours in large systems of interacting agents has received an increasing attention from several communities during the past decades. Multi-agent systems are ubiquitous in applications ranging from aggregation phenomena in biology [15] to the understanding of crowd motion [28], animal flocks [29] and swarms of autonomous vehicles [13]. While the first studies on multi-agent systems were formulated in a graph-theoretic framework (see e.g. [50] and references therein), several models now rely on continuous-time dynamical systems to depict this type of dynamics. In this context, a multi-agent system is usually described by a family of coupled ordinary differential equations (ODEs for short) of the form

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

where $\boldsymbol{x} = (x_1, \ldots, x_N)$ denotes the state of all the agents and $\boldsymbol{v}_N : [0, T] \times (\mathbb{R}^d)^N \times \mathbb{R}^d \to \mathbb{R}^d$ is a *non-local velocity field* depending both on the running agent and on the whole state of the system (see e.g. [6, 28, 29]). However general and useful, these models are generically not the most powerful ones when it comes to capturing the global features of a multi-agent system. Besides, their intrinsic dependence on the number $N \geq 1$ of agents makes most of the classical computational approaches practically intractable for realistic scenarios.

One of the most natural ideas to circumvent this model limitation is to approximate the large system of coupled ODEs written in (1) by a single infinite-dimensional dynamics via a process called *mean-field limit* (see e.g. the survey [51]). In this setting, the agents are supposed to be indistinguishable, and the assembly of particles is described by means of its *spatial density* $\mu(\cdot)$, which is represented by a measure. The evolution through time of this global quantity is then prescribed by a *non-local continuity equation* of the form

$$\partial_t \mu(t) + \nabla \cdot (v(t, \mu(t), \cdot)\mu(t)) = 0. \tag{2}$$

Such a macroscopic approach has been successfully used e.g. to model pedestrian dynamics and biological systems [15, 28], as well as to transpose the study of classical patterns such as consensus or flocking formation to the mean-field setting [21, 44]. These research endeavours have hugely benefited from the recent progresses made in the theory of *optimal transportation*, for which we point to the reader to the monographs [5, 55, 56].

More recently, the problem of controlling multi-agent systems in order to promote a desired behaviour or configuration became relevant in a growing number of applications. Motivated by implementability and efficiency, many contributions have therefore aimed at generalising relevant notions of control theory to PDEs of the form (2) serving as mean-field approximations of the discrete system (1). The resulting class of *controlled continuity equations* are usually written as

$$\partial_t \mu(t) + \nabla \cdot \big((v(t, \mu(t), \cdot) + u(t, \cdot))\mu(t) \big) = 0. \tag{3}$$

---

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

†Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova, 63 via Trieste, Padova, Italy. *E-mail*: `francesco.rossi@math.unipd.it`

While a few articles have been dealing with controllability results [34, 35] or explicit syntheses of control laws [16, 53], the major part of the literature has been focusing on *mean-field optimal control problems*, with contributions ranging from existence results [38, 39, 40] to first-order optimality conditions [7, 8, 10, 11, 22, 23, 46, 54] and numerical methods [1, 14].

One of the distinctive features of continuity equations is that they require fairly restrictive regularity assumptions to be classically well-posed. While (3) makes sense whenever the drift and control are measurable and satisfy some integrability bounds, the associated notion of so-called *superposition solution* (see e.g. [5, Theorem 8.2.1]) is relatively weak and of limited practical use. In [2, 31], a theory of well-posedness was developed for continuity equations with Sobolev and *BV* velocity fields. However powerful and general, the latter has not yet been generalised to non-local driving fields, and is inherently restricted to measures which are absolutely continuous with respect to the ambient Lebesgue measure. Up to now, the only identified setting in which a strong form of classical well-posedness (see Theorem 5 below) holds for arbitrary measure curves solution of (3) is that of *Cauchy-Lipschitz* regularity (see e.g. [3, Section 3]). In this framework, solutions of non-local continuity equations exist, are unique, and stability estimates are available both with respect to the initial data and the velocity fields, see e.g. [9, 52]. This latter fact is highly relevant to our purpose, since optimal control problems formulated on continuity equations are frequently studied in an "optimise-then-discretise" spirit. Indeed, the main desirable property of a control law designed for the kinetic model (3) is to provide a strategy which can be in turn applied – either exactly or approximately – to finite-dimensional systems of the form (1). As the infinite-dimensional strategy is not strictly optimal for the discrete multi-agent system in general, one would also like to have access to quantitative error estimates between the true solution and the approximate one. From a computational standpoint, Cauchy-Lipschitz regularity is also relevant to ensure the well-posedness of numerical methods such as semi-Lagrangian schemes (see e.g. [20, 24]), as well as to prevent the apparition of Lavrentiev-type instabilities in the context of optimal control (see e.g. [48]). For all these reasons, a wide portion of the literature of mean-field control has been dealing with problems in which one imposes an a priori Lipschitz-in-space regularity on the admissible controls (see e.g. [8, 11, 17, 39, 40, 53]), or at least some continuity assumptions on the driving fields (see [22, 23, 46]). A natural question is then to ask whether such regularity property may hold intrinsically or not. In this paper, we investigate this problem in the setting of mean-field optimal control problems formulated on the controlled dynamics (3).

It is well-known that solutions of optimal control problems in Wasserstein spaces need not be regular in general. Indeed, there exists a vast literature devoted to the study of the regularity properties of solutions to the optimal transport problem in Monge formulation (see e.g. [30, 37] for some of the farthest-reaching contributions on this topic), mostly via PDE techniques. However, few of these results can be translated into regularity properties on the optimal tangent velocity field $v^*(\cdot, \cdot)$ solving the Benamou-Brenier problem

$$(\mathcal{P}_{\mathrm{BB}}) \quad \begin{cases} \min_{v \in L^2} \left[\int_0^T \int_{\mathbb{R}^d} |v(t,x)|^2 \mathrm{d}\mu(t)(x) \mathrm{d}t \right] \\ \text{s.t.} \begin{cases} \partial_t \mu(t) + \nabla \cdot (v(t,\cdot)\mu(t)) = 0, \\ \mu(0) = \mu^0 \quad \text{and} \quad \mu(T) = \mu^1. \end{cases} \end{cases}$$

For the optimal control problem $(\mathcal{P}_{\mathrm{BB}})$, it can be shown for instance by building on [56, Theorem 12.50]) that $v(t,\cdot) \in C_{\mathrm{loc}}^{k+1,\alpha}(\mathbb{R}^d, \mathbb{R}^d)$ whenever $\mu^0, \mu^1 \in \mathscr{P}^{ac}(\mathbb{R}^d)$ have densities with respect to the ambient Lebesgue measure with regularity at least $C_{\mathrm{loc}}^{k,\alpha}(\mathbb{R}^d, \mathbb{R}^d)$ for some $\alpha > 0$, and that their supports are smooth and convex. Another context in which the regularity of mean-field optimal controls has been (indirectly) investigated is that of *mean-field games* (see [45, 47]). Indeed, there is a large literature dedicated to the regularity of the value function $(t,x) \mapsto V^*(t,x) \in \mathbb{R}$ solving the backward Hamilton-Jacobi equation of the coupled system

$$\begin{cases} \partial_t V(t,x) + H(t,x,\nabla_x V(t,x)) = f(t,x,\mu(t)), \quad V(T,x) = g_T(x,\mu(T)), \\ \partial_t \mu(t) - \nabla \cdot (\nabla_p H(t,x,\nabla_x V(t,x))\mu(t)) = 0, \quad \mu(0) = \mu^0. \end{cases}$$

In the setting of *variational mean-field games*, the velocity field $v^*(t,x) = -\nabla_p H(t,x,\mathrm{D}_x V^*(t,x))$ is the optimal control associated to a mean-field optimal control problem. Therefore, $v^*(t,\cdot)$ is expected to have a regularity with one order of differentiation fewer than the value function. We refer the reader e.g. to [18] for Sobolev regularity results on the value function $V^*(\cdot, \cdot)$ and to [19] for Hölder regularity (see also Remark 1 below).

In this article, we investigate the intrinsic Lipschitz regularity with respect to the space variable of the solutions of general mean-field optimal control problems of the form

$$(\mathcal{P}) \quad \begin{cases} \min_{u(\cdot) \in \mathcal{U}} \left[\int_0^T \left(L(t,\mu(t)) + \int_{\mathbb{R}^d} \psi(u(t,x)) \mathrm{d}\mu(t)(x) \right) \mathrm{d}t + \varphi(\mu(T)) \right] \\ \text{s.t.} \begin{cases} \partial_t \mu(t) + \nabla \cdot ((v(t,\mu(t),\cdot) + u(t,\cdot))\mu(t)) = 0, \\ \mu(0) = \mu^0. \end{cases} \end{cases}$$

The set of admissible controls for $(\mathcal{P})$ is defined by $\mathcal{U} = L^\infty([0,T], L^1(\mathbb{R}^d, U; \mu(t)))$, where $U \subset \mathbb{R}^d$ is a convex and compact set, and $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ is a fixed initial datum. Remark that, since we do not impose any a priori regularity assumptions on the control vector fields $u : [0,T] \times \mathbb{R}^d \to U$, there may exist no solutions to the non-local transport equation (3) driving problem $(\mathcal{P})$. Moreover even when solutions do exist, they may not be classically well-posed and defined in a weak sense only.

The first main contribution of this manuscript is the following existence result of intrinsically Lipschitz-in-space mean-field optimal controls for $(\mathcal{P})$.

Theorem 1 (Existence of Lipschitz-in-space solutions for $(\mathcal{P})$)**.** *Let $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ and assume that hypotheses* **(H)** *of Section 4 below hold. Moreover, suppose that the control cost $\psi(\cdot)$ is* strongly convex *with a constant $\lambda_\psi > \lambda_{(\mathcal{P})} \geq 0$, where $\lambda_{(\mathcal{P})} \geq 0$ only depends on $\mathrm{supp}(\mu^0), T$ and on the C^2-norm with respect to the measure and space variables of the dynamics and cost functionals of $(\mathcal{P})$.*

Then, there exists a constant $\mathcal{L}_U > 0$ and a trajectory-control pair $(\mu^*(\cdot), u^*(\cdot,\cdot)) \in \mathrm{Lip}([0,T], \mathscr{P}_c(\mathbb{R}^d)) \times \mathcal{U}$ which is optimal for $(\mathcal{P})$ such that $x \in \mathbb{R}^d \mapsto u^*(t,x) \in U$ is $\mathcal{L}_U$-Lipschitz for $\mathscr{L}^1$-almost every $t \in [0,T]$.

The proof of this result is obtained by combining two fairly separated arguments. The first one is an existence result for mean-field optimal controls which was derived in [38], and recalled in Theorem 6 below. In the latter, it is proven under very general assumptions that, given a sequence $\mu_N^0 := \frac{1}{N} \sum_{i=1}^N \delta_{x_i^0}$ converging in the W_1-metric towards μ^0, there exist optimal solutions of problem $(\mathcal{P})$ which can be recovered as Γ-limits in a suitable topology of sequences of solutions to the discrete problems

$$(\mathcal{P}_N) \quad \begin{cases} \min\limits_{\boldsymbol{u}(\cdot) \in \mathcal{U}_N} \left[\int_0^T \left(\boldsymbol{L}_N(t, \boldsymbol{x}(t)) + \frac{1}{N} \sum_{i=1}^N \psi(u_i(t)) \right) \mathrm{d}t + \boldsymbol{\varphi}_N(\boldsymbol{x}(T)) \right] \\ \text{s.t.} \begin{cases} \dot{x}_i(t) = \boldsymbol{v}_N(t, \boldsymbol{x}_N(t), x_i(t)) + u_i(t), \\ x_i(0) = x_i^0 \in \mathbb{R}^d. \end{cases} \end{cases}$$

Here, $\mathcal{U}_N = L^\infty([0,T], U^N)$, and the functionals $\boldsymbol{v}_N : [0,T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$, $\boldsymbol{L}_N : [0,T] \times (\mathbb{R}^d)^N \to \mathbb{R}$ and $\boldsymbol{\varphi}_N : (\mathbb{R}^d)^N \to \mathbb{R}$ are discrete approximations of $v(\cdot,\cdot,\cdot)$, $L(\cdot,\cdot)$ and $\varphi(\cdot)$ respectively (see Definition 10 below).

The second key component of our approach is a careful adaptation of a methodology recently developed in [26, 33] to the family of problems $(\mathcal{P}_N)$, which provides sufficient conditions for the existence of locally optimal Lipschitz feedbacks around solutions of optimal control problems. In the context of mean-field control problems, this part crucially relies on the following *uniform mean-field coercivity estimate*

$$\mathbf{Hess}_{\boldsymbol{x}} \, \boldsymbol{\varphi}_N[\boldsymbol{x}_N^*(T)](\boldsymbol{y}(T), \boldsymbol{y}(T)) - \int_0^T \mathbf{Hess}_{\boldsymbol{x}} \, \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{y}(t), \boldsymbol{y}(t)) \mathrm{d}t$$

$$- \int_0^T \mathbf{Hess}_{\boldsymbol{u}} \, \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{w}(t), \boldsymbol{w}(t)) \mathrm{d}t \geq \rho_T \int_0^T |\boldsymbol{w}(t)|_N^2 \mathrm{d}t,$$

which holds along any *optimal mean-field Pontryagin triple* $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot))$ (see Proposition 6 below) for $(\mathcal{P}_N)$, where $\mathbb{H}_N : [0,T] \times (\mathbb{R}^{2d})^N \times U \to \mathbb{R}$ is the Hamiltonian of the discrete problem. In this context, the operator $\mathbf{Hess}\,[\bullet]$ stands for the restriction to empirical measures of the intrinsic Wasserstein Hessian bilinear form (see e.g. [25]), whose construction is further detailed in Section 2. In essence, this uniform coercivity estimate allows one to invert the optimality conditions stemming from the application of the *Pontryagin Maximum Principle* (PMP in the sequel) to $(\mathcal{P}_N)$, with a control on the Lipschitz constant of such inverse. The main subtlety here lies in the fact that we need these estimates to be uniform with respect to N, which would not be the case if we were to apply verbatim the results of [33] to $(\mathcal{P}_N)$. For these reasons, we work with an adapted *mean-field* PMP – which is the discrete counterpart of the Wasserstein PMP studied in [7, 8, 10, 11] –, and express the coercivity condition in terms of Wasserstein calculus.

The combination of these two steps together with delicate projection arguments, we can build an optimal feedback $(t,x) \in [0,T] \times \mathbb{R}^d \mapsto u_N^*(t,x) \in U$ for $(\mathcal{P}_N)$ that is Lipschitz in $x \in \mathbb{R}^d$ uniformly with respect to $N \geq 1$. By standard compactness arguments (see e.g. [7, 40]), this allows us to obtain a result that is stronger than Theorem 1, which is the second main contribution of this manuscript.

Theorem 2 (Convergence of optimal Lipschitz feedbacks towards solutions of $(\mathcal{P})$)**.** *Let $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ and $(\mu_N^0) \subset \mathscr{P}_c(\mathbb{R}^d)$ be a sequence of empirical measures with uniformly compact supports such that $W_1(\mu_N^0, \mu^0) \to 0$ as $N \to +\infty$. Suppose that hypotheses* **(H)** *of Section 4 below are satisfied, and that the* mean-field coercivity estimate **(CO$_N$)** *described in Section 5 holds along any optimal Pontryagin triple $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot))$ for $(\mathcal{P}_N)$ defined in the sense of Proposition 6.*

Then, there exists a uniform constant $\mathcal{L}_U > 0$ depending only on the data of $(\mathcal{P})$ and a sequence of trajectory-control pairs $(\mu_N^*(\cdot), u_N^*(\cdot,\cdot)) \subset \mathrm{Lip}([0,T], \mathscr{P}_c(\mathbb{R}^d)) \times L^2([0,T], W^{1,\infty}(\mathbb{R}^d, U))$ such that the following holds.

(a) *For any $N \geq 1$ the pair $(\mu_N^*(\cdot), u_N^*(\cdot, \cdot))$ is optimal for $(\mathcal{P}_N)$, i.e. $\mu_N^*(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i^*(t)}$ and $u_N^*(t, x_i^*(t)) = u_i^*(t)$ for $\mathscr{L}^1$-almost every $t \in [0, T]$.*

(b) *The maps $x \in \mathbb{R}^d \mapsto u_N^*(t, x) \in U$ are $\mathcal{L}_U$-Lipschitz for $\mathscr{L}^1$-almost every $t \in [0, T]$ and any $N \geq 1$.*

(c) *For every $p \in (d, +\infty)$, the cluster points of the sequence $(u_N^*(\cdot, \cdot))$ in the weak $L^2([0, T], W^{1,p}(\Omega, U))$-topology are optimal controls for $(\mathcal{P})$ and are $\mathcal{L}_U$-Lipschitz in space.*

While they are more general than that of Theorem 1, the statements of Theorem 2 are less intrinsic by nature, as they rely on the mean-field coercivity estimate $(\mathbf{CO}_N)$ which can only be formulated on the discrete approximations $(\mathcal{P}_N)$. For this reason, in Proposition 8 below, we show that the strong convexity assumption imposed on $\psi(\cdot)$ in Theorem 1 is in fact a sufficient condition for $(\mathbf{CO}_N)$. Hence, the statements of Theorem 1 – which present the advantage of involving quantities which are intrinsic to $(\mathcal{P})$ – can be recovered as a direct corollary of Theorem 2.

Remark 1 (Comparison with related contributions in mean-field games)**.** *It was recently brought to our attention that a result related to Theorem 1 and Theorem 2 above was derived in [42]. In the latter, the authors show that the value function of a certain class of first-order mean-field games is continuously differentiable with Lipschitz derivative when the data are of class C^3 and the time horizon $T > 0$ is sufficiently small. These two requirement are very close to our standing assumptions. Indeed, we posit in hypotheses $(\mathbf{H})$ below that all our data are $C_{\text{loc}}^{2,1}$, and it is illustrated in Section 6 that our uniform coercivity estimate $(\mathbf{CO}_N)$ can be interpreted as a quantitative condition comparing the size of the time horizon $T > 0$ relatively to other constants of the problem, and in particular with the semi-convexity constant of the cost functionals. The results of [42] have then been extended in [49] to broader classes of first-order mean-field games systems, and improved in the very recent [41], in which it is shown that the value function is regular when the small time horizon condition is replaced by the* displacement convexity *(see e.g. [5, Chapter 9]) of the Lagrangian. Incidentally for mean-field control problems, this scenario is also contained in our main result Theorem 1. Indeed, if the running cost of the problem is displacement convex, the final cost equal to zero, and the non-local dynamics reduced to a linear controlled vector-field, it can be shown that $\lambda_{(\mathcal{P})} = 0$ and the controls are Lipschitz-regular in space whenever the control cost $\psi(\cdot)$ is strongly convex with constant $\lambda > 0$.*

We also stress that the proof strategies developed in [41, 42, 49] are fairly close to the one that we independently propose here, as they rely on the application of inverse function mappings to sequences of approximations by empirical measures, with a quantitative control on the Lipschitz constant of the inverse.

The structure of this article is the following. In Section 2, we recall several general prerequisites on measure theory and optimal transport. In particular in Section 2.3, we investigate in details the interplay between Wasserstein derivatives of functionals at empirical measures and classical derivatives of the discrete functionals which arguments are the corresponding support points. In Section 3, we review notions pertaining to finite-dimensional optimal control problems, with a particular emphasis on Lipschitz feedbacks. We proceed by exposing in Section 4 concepts dealing with continuity equations and mean-field optimal control problems, and move on to the proofs of our main results Theorem 1 and Theorem 2 in Section 5. More precisely, in Section 5.1, we state the coercivity assumption $(\mathbf{CO}_N)$ and use it to prove Theorem 2. We then show in Section 5.2 how the latter together with a standard convexity estimate for C_{loc}^2-regular functions on convex compact sets allows to recover Theorem 1. We conclude by providing in Section 6 an analytical example in which our coercivity estimate is both necessary and sufficient for the existence of Lipschitz-in-space mean-field optimal controls.

2   Preliminaries

In this section, we introduce results and notations that we will use throughout the article. Section 2.1 presents known results of analysis in measure spaces and optimal transport, while Section 2.2 deals with first- and second-order differential calculus in Wasserstein spaces. We introduce in Section 2.3 the notion of *mean-field approximating sequence*, along with a discretised counterpart of the Wasserstein calculus.

2.1   Analysis in measure spaces

In this section, we introduce some classical notations and results of analysis in measure spaces and optimal transport. For these topics, we refer the reader to [4] and [5, 55, 56] respectively.

We denote by $(\mathcal{M}(\mathbb{R}^d, \mathbb{R}^m), \|\cdot\|_{TV})$ the Banach space of m-dimensional vector-valued finite Radon measures defined on $\mathbb{R}^d$ endowed with the total variation norm, defined for any $\boldsymbol{\nu} \in \mathcal{M}(\mathbb{R}^d, \mathbb{R}^m)$ by $\|\boldsymbol{\nu}\|_{TV} := |\boldsymbol{\nu}|(\mathbb{R}^d)$. Here, the total variation measure $|\boldsymbol{\nu}| \in \mathcal{M}(\mathbb{R}^d, \mathbb{R}_+)$ associated to $\boldsymbol{\nu}$ is given on any Borel set $B \subset \mathbb{R}^d$ by

$$|\boldsymbol{\nu}|(B) = \sup \left\{ \sum_{k=1}^{+\infty} |\boldsymbol{\nu}(B_k)| \text{ s.t. } B_k \text{ are disjoint Borel sets and } \bigcup_{k=1}^{+\infty} B_k = B \right\},$$

where $|\boldsymbol{\nu}(B)|$ is the norm of the element $\boldsymbol{\nu}(B) \in \mathbb{R}^m$. It is known by Riesz's Theorem (see e.g. [4, Theorem 1.54]) that $\mathcal{M}(\mathbb{R}^d, \mathbb{R}^m)$ can be identified with the topological dual of the Banach space $(C_0^0(\mathbb{R}^d, \mathbb{R}^m), \|\cdot\|_{C^0})$, which is the completion of the space $C_c^0(\mathbb{R}^d, \mathbb{R}^m)$ of continuous and compactly supported functions. The latter is endowed with the duality product

$$\langle \boldsymbol{\nu}, \phi \rangle_{C^0(\mathbb{R}^d, \mathbb{R}^m)} := \sum_{k=1}^m \int_{\mathbb{R}^d} \phi_k(x) \mathrm{d}\nu_k(x), \tag{4}$$

defined for any $\boldsymbol{\nu} \in \mathcal{M}(\mathbb{R}^d, \mathbb{R}^m)$ and $\phi \in C_c^0(\mathbb{R}^d, \mathbb{R}^m)$. Given a positive Borel measure $\nu \in \mathcal{M}(\mathbb{R}^d, \mathbb{R}_+)$ and an element $p \in [1, +\infty]$, the notations $L^p(\Omega, \mathbb{R}^m; \nu)$ and $W^{1,p}(\Omega, \mathbb{R}^m; \nu)$ stand for the spaces of p-integrable and Sobolev functions respectively. In the case where $\nu = \mathscr{L}^d$ is the standard d-dimensional Lebesgue measure $\mathscr{L}^d$, we simply denote these spaces by $L^p(\Omega, \mathbb{R}^m)$ and $W^{1,p}(\Omega, \mathbb{R}^m)$.

We use the notation $\mathscr{P}(\mathbb{R}^d) \subset \mathcal{M}(\mathbb{R}^d, \mathbb{R}_+)$ for the space of *Borel probability measures*, and given $p \geq 1$, we denote by $\mathscr{P}_p(\mathbb{R}^d)$ the subset of $\mathscr{P}(\mathbb{R}^d)$ of measures having finite p-th moment, i.e.

$$\mathscr{P}_p(\mathbb{R}^d) = \Big\{ \mu \in \mathscr{P}(\mathbb{R}^d) \text{ s.t. } \int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu(x) < +\infty \Big\}.$$

We define the *support* of $\boldsymbol{\nu} \in \mathcal{M}(\mathbb{R}^d, \mathbb{R}^m)$ as the closed set $\operatorname{supp}(\boldsymbol{\nu}) := \{ x \in \mathbb{R}^d \text{ s.t. } |\boldsymbol{\nu}(\mathcal{N})| \neq 0 \text{ for any neighbourhood } \mathcal{N} \text{ of } x \}$, and denote by $\mathscr{P}_c(\mathbb{R}^d)$ the subset of probability measures with compact support.

Definition 1 (Absolute continuity and Radon-Nikodym derivative). *Let $\Omega \subset \mathbb{R}^m$ and $U \subset \mathbb{R}^d$ be two Borel sets. Given a pair of measures $(\boldsymbol{\nu}, \mu) \in \mathcal{M}(\Omega, U) \times \mathcal{M}(\Omega, \mathbb{R}_+)$, we say that $\boldsymbol{\nu}$ is* absolutely continuous *with respect to μ, and write $\boldsymbol{\nu} \ll \mu$, provided that $|\boldsymbol{\nu}(B)| = 0$ whenever $\mu(B) = 0$ for any Borel set $B \subset \Omega$. Moreover, it holds that $\boldsymbol{\nu} \ll \mu$ if and only if there exists a Borel map $u \in L^1(\Omega, U; \mu)$ such that $\boldsymbol{\nu} = u\mu$. This map is referred to as the* Radon-Nikodym derivative *of $\boldsymbol{\nu}$ with respect to μ, and denoted by $u := \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}$.*

We now recall the definitions of *pushforward* and *transport plan* for Borel probability measure.

Definition 2 (Pushforward of a measure through a Borel map). *Given a measure $\mu \in \mathscr{P}(\mathbb{R}^d)$ and a Borel map $f : \mathbb{R}^d \to \mathbb{R}^d$, the* pushforward *$f_\#\mu$ of μ through f is the Borel probability measure defined by $f_\#\mu(B) := \mu(f^{-1}(B))$ for any Borel set $B \subset \mathbb{R}^d$.*

Definition 3 (Transport plans). *Let $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$. We say that $\gamma \in \mathscr{P}(\mathbb{R}^{2d})$ is a* transport plan *between μ and ν, denoted by $\gamma \in \Gamma(\mu, \nu)$, if $\pi_\#^1 \gamma = \mu$ and $\pi_\#^2 \gamma = \nu$, where $\pi^1, \pi^2 : \mathbb{R}^{2d} \to \mathbb{R}^d$ denote the projection on the first and second component respectively.*

In what follows, we recall the definition and some of the main properties of the so-called *Wasserstein spaces* (see e.g. [5, Chapter 7] or [56, Chapter 6]).

Definition 4 (Wasserstein spaces). *Given $p \in [1, +\infty)$ and $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$, the Wasserstein distance of order p between μ and ν is defined by*

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

The set of optimal transport plans *realising this optimal value is non-empty and denoted by $\Gamma_o(\mu, \nu)$. The space $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ of probability measures with finite momentum of order p endowed with the W_p-metric is called the* Wasserstein space *of order p.*

Proposition 1 (Elementary properties of the Wasserstein spaces). *For any $p \in [1, +\infty)$, the metric space $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ is complete and separable, and the W_p-distance metrises the weak-* topology induced by (4), i.e.*

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

Given two elements $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, 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 $\mu, \nu \in \mathscr{P}_c(\mathbb{R}^d)$, the following Kantorovich-Rubinstein duality formula *holds*

$$W_1(\mu, \nu) = \sup_\phi \left\{ \int_{\mathbb{R}^d} \phi(x) \, \mathrm{d}(\mu - \nu)(x) \text{ s.t. } \operatorname{Lip}(\phi(\cdot); \mathbb{R}^d) \leq 1 \right\}. \tag{5}$$

where $\operatorname{Lip}(\phi(\cdot); \Omega)$ denotes the Lipschitz constant of $\phi(\cdot)$ over a subset $\Omega \subset \mathbb{R}^d$.

We end this introductory paragraph by recalling the concept of *disintegration* for vector-valued measures (see e.g. [4, Theorem 2.28]).

Theorem 3 (Disintegration)**.** *Let $\Omega_1 \subset \mathbb{R}^{m_1}$, $\Omega_2 \subset \mathbb{R}^{m_2}$ and $U \subset \mathbb{R}^d$ be Borel sets. Let $\boldsymbol{\nu} \in \mathcal{M}(\Omega_1 \times \Omega_2, U)$ and $\pi^1 : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to \mathbb{R}^{m_1}$ be the projection on the first factor. Defining the measure $\mu := \pi^1_\# |\boldsymbol{\nu}| \in \mathcal{M}(\Omega_1, \mathbb{R}_+)$, there exists a μ-almost uniquely determined Borel family of measures $\{\boldsymbol{\nu}_x\}_{x \in \Omega_1} \subset \mathcal{M}(\Omega_2, U)$ such that*

$$\int_{\Omega_1 \times \Omega_2} f(x,y)\mathrm{d}\boldsymbol{\nu}(x,y) = \int_{\Omega_1} \left(\int_{\Omega_2} f(x,y)\mathrm{d}\boldsymbol{\nu}_x(y) \right) \mathrm{d}\mu(x) \tag{6}$$

for any Borel map $f \in L^1(\Omega_1 \times \Omega_2, |\boldsymbol{\nu}|)$. This construction is referred to as the disintegration *of $\boldsymbol{\nu}$ onto μ, and it is denoted by $\boldsymbol{\nu} = \int_{\Omega_1} \boldsymbol{\nu}_x \mathrm{d}\mu(x)$.*

2.2 First- and second-order differential calculus over $(\mathscr{P}_2(\mathbb{R}^d), W_2)$

In this section, we recall key concepts related to first- and second-order differential calculus in the Wasserstein space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$. We refer the reader to [5, Chapters 9-11] and [43] for an exhaustive treatment of the first-order theory, and borrow the main notions dealing with Wasserstein Hessians from [25, Section 3].

Throughout this section, we denote by $\phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{\pm\infty\}$ an extended real-valued functional with non-empty effective domain $D(\phi) = \{\mu \in \mathscr{P}_2(\mathbb{R}^d) \text{ s.t. } \phi(\mu) \neq \pm\infty\}$. We will also denote by $\phi : \mathscr{P}_c(\mathbb{R}^d) \to \mathbb{R}$ any such functional such that $\mathscr{P}_c(\mathbb{R}^d) \subset D(\phi)$. In the following definition, we recall the notions of *classical subdifferential* and *superdifferential* for functionals defined over $(\mathscr{P}_2(\mathbb{R}^d), W_2)$.

Definition 5 (Wasserstein subdifferential and superdifferentials)**.** *Let $\mu \in D(\phi)$. We say that a map $\xi \in L^2(\mathbb{R}^d, \mathbb{R}^d; \mu)$ belongs to the* classical subdifferential $\partial^-\phi(\mu)$ *of $\phi(\cdot)$ at μ provided that*

$$\phi(\nu) - \phi(\mu) \geq \sup_{\gamma \in \Gamma_o(\mu,\nu)} \int_{\mathbb{R}^{2d}} \langle \xi(x), y - x \rangle \mathrm{d}\gamma(x,y) + o(W_2(\mu,\nu)),$$

for all $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. Similarly, we say that a map $\xi \in L^2(\mathbb{R}^d, \mathbb{R}^d; \mu)$ belongs to the classical superdifferential $\partial^+\phi(\mu)$ *of $\phi(\cdot)$ at μ if $(-\xi) \in \partial^-(-\phi)(\mu)$.*

Following [5, Chapter 8], we define the *analytical tangent space* $\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ at $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ by

$$\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d) := \overline{\nabla C_c^\infty(\mathbb{R}^d)}^{L^2(\mu)} = \overline{\{\nabla \xi \text{ s.t. } \xi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu)}. \tag{7}$$

In the next definition, we recall the notion of *differentiable functional* over $\mathscr{P}_2(\mathbb{R}^d)$.

Definition 6 (Differentiable functionals in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$)**.** *A functional $\phi : \mathscr{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}$ is said to be differentiable at $\mu \in D(\phi)$ if $\partial^-\phi(\mu) \cap \partial^+\phi(\mu) \neq \emptyset$. In this case, there exists a unique elements $\nabla_\mu \phi(\mu) \in \partial^-\phi(\mu) \cap \partial^+\phi(\mu) \cap \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$, called the* Wasserstein gradient *of $\phi(\cdot)$ at μ, which satisfies*

$$\phi(\nu) - \phi(\mu) = \int_{\mathbb{R}^{2d}} \langle \nabla_\mu \phi(\mu)(x), y - x \rangle \mathrm{d}\gamma(x,y) + o(W_2(\mu,\nu)), \tag{8}$$

for any $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\gamma \in \Gamma_o(\mu,\nu)$.

From the characterisation (8) of the Wasserstein gradient $\nabla_\mu \phi(\mu)$, we can write a chain rule along elements of $\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ (see e.g. [5, Proposition 10.3.18] or the recent improvement of [10, Proposition 3.6]).

Proposition 2 (First-order chain rule)**.** *Suppose that $\phi(\cdot)$ is differentiable at $\mu \in D(\phi)$. Then for any $\xi \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$, the map $s \in \mathbb{R} \mapsto \phi((\mathrm{Id} + s\xi)_\#\mu)$ is differentiable at $s = 0$ with*

$$\mathcal{L}_\xi \phi(\mu) := \frac{\mathrm{d}}{\mathrm{d}s} \phi((\mathrm{Id} + s\xi)_\#\mu)_{|s=0} = \int_{\mathbb{R}^d} \langle \nabla_\mu \phi(\mu)(x), \xi(x) \rangle \mathrm{d}\mu(x), \tag{9}$$

where $\mathcal{L}_\xi \phi(\mu)$ denotes the Lie derivative *of $\phi(\cdot)$ at μ in the direction $\xi \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$.*

In the sequel, we will also need a notion of second-order derivatives for functionals defined over $\mathscr{P}_2(\mathbb{R}^d)$.

Definition 7 (Hessian bilinear form in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$)**.** *Suppose that $\phi(\cdot)$ is differentiable at $\mu \in D(\phi)$ and suppose that for any $\xi \in \nabla C_c^\infty(\mathbb{R}^d)$, the map*

$$\mathcal{L}_\xi \phi : \nu \in \mathscr{P}_2(\mathbb{R}^d) \mapsto \langle \nabla_\mu \phi(\nu), \xi \rangle_{L^2(\nu)} \in \mathbb{R},$$

is also differentiable at μ. Then, the partial Wasserstein Hessian *of $\phi(\cdot)$ at μ is the bilinear form defined by*

$$\text{Hess}\,\phi[\mu](\xi_1,\xi_2) := \mathcal{L}_{\xi_2}\left(\mathcal{L}_{\xi_1}\phi(\mu)\right) - \mathcal{L}_{\text{D}\xi_1\xi_2}\phi(\mu), \tag{10}$$

for any $\xi_1,\xi_2 \in \nabla C_c^\infty(\mathbb{R}^d)$. Moreover, if there exists a constant $C_\mu > 0$ such that

$$\text{Hess}\,\phi[\mu](\xi_1,\xi_2) \leq C_\mu \, \|\xi_1\|_{L^2(\mu)}\|\xi_2\|_{L^2(\mu)},$$

we denote again by $\text{Hess}\,\phi[\mu](\cdot,\cdot)$ its extension to $\text{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ and say that $\phi(\cdot)$ is twice differentiable at μ.

In the following proposition, we recollect several statements from [25, Section 3] which yield an analytical expression of the Wasserstein Hessian. This also allows to write a second-order differentiation formula for functionals defined over $\mathscr{P}_2(\mathbb{R}^d)$.

Proposition 3 (Wasserstein Hessian and second-order expansion). *Suppose that $\phi(\cdot)$ is differentiable at $\mu \in D(\phi)$ in the sense of Definition 6, and that the maps*

$$y \in \mathbb{R}^d \mapsto \nabla_\mu\phi(\mu)(y) \in \mathbb{R}^d \qquad \text{and} \qquad \nu \in \mathscr{P}_2(\mathbb{R}^d) \mapsto \nabla_\mu\phi(\nu)(x) \in \mathbb{R}^d,$$

are continuously differentiable at $x \in \mathbb{R}^d$ and $\mu \in D(\phi)$ respectively. Then, $\phi(\cdot)$ is twice differentiable in the sense of Definition 7, and its Wasserstein Hessian can be written explicitly as

$$\text{Hess}\,\phi[\mu](\xi_1,\xi_2) = \int_{\mathbb{R}^d} \langle \text{D}_x\nabla_\mu\phi(\mu)(x)\xi_1(x),\xi_2(x)\rangle\mathrm{d}\mu(x) + \int_{\mathbb{R}^{2d}} \langle \text{D}_\mu^2\phi(\mu)(x,y)\xi_1(x),\xi_2(y)\rangle\mathrm{d}\mu(x)\mathrm{d}\mu(y), \tag{11}$$

for any $\xi_1,\xi_2 \in \text{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$. Here, $\text{D}_x\nabla_\mu\phi(\mu)(x) \in \mathbb{R}^{d\times d}$ is the Fréchet differential of $\nabla_\mu\phi(\mu)(\cdot)$ at $x \in \mathbb{R}^d$, while $\text{D}_\mu^2\phi(\mu)(x,\cdot) : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ denotes the matrix-valued map whose columns are the Wasserstein gradients of the components $((\nabla_\mu\phi(\mu)(x))_i)_{1\leq i\leq d}$ in the sense of Definition 6. Moreover, the following identity

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{L}_{\xi_1}\phi((\text{Id}+s\xi_2)_\#\mu)_{|s=0} = \text{Hess}\,\phi[\mu](\xi_1,\xi_2) + \mathcal{L}_{\text{D}\xi_1\xi_2}\phi(\mu), \tag{12}$$

holds for any $\xi_1,\xi_2 \in \nabla C_c^\infty(\mathbb{R}^d)$.

We finally introduce the notion of $\mathcal{C}_{\text{loc}}^{2,1}$-*Wasserstein regularity*, which will be used throughout this article.

Definition 8 ($\mathcal{C}_{\text{loc}}^{2,1}$-Wasserstein regularity). *A functional $\phi(\cdot)$ is said to be $\mathcal{C}_{\text{loc}}^{2,1}$-Wasserstein regular if it is twice differentiable over $(\mathscr{P}(K),W_2)$ for any compact set $K \subset \mathbb{R}^d$, and satisfies*

$$\|\phi(\cdot)\|_{\mathcal{C}^{2,1}(K)} := \max_{\mu\in\mathscr{P}(K)} |\phi(\mu)| + \|\nabla_\mu\phi(\mu)(\cdot)\|_{C^0(K,\mathbb{R}^d)} + \|\text{D}_x\nabla_\mu\phi(\mu)(\cdot)\|_{C^0(K,\mathbb{R}^{d\times d})} + \|\text{D}_\mu^2\phi(\mu)(\cdot,\cdot)\|_{C^0(K\times K,\mathbb{R}^{d\times d})}$$

$$+ \text{Lip}\big(\text{D}_x\nabla_\mu\phi(\cdot)(\cdot);\mathscr{P}(K)\times K\big) + \text{Lip}\big(\text{D}_\mu^2\phi(\cdot)(\cdot,\cdot);\mathscr{P}(K)\times K\times K\big) \leq C_K, \tag{13}$$

where $C_K > 0$ is a constant which only depends $K \subset \mathbb{R}^d$.

2.3 Mean-field adapted structures and empirical measures

In this section, we present several notions dealing with functionals defined over empirical measures in the spirit of [38], along with an adapted discrete version of the differential structure described in Section 2.2.

We denote by $\mathscr{P}_N(\mathbb{R}^d) := \{\frac{1}{N}\sum_{i=1}^N \delta_{x_i} \text{ s.t. } (x_1,\ldots,x_N) \in (\mathbb{R}^d)^N\}$ the set of N-*empirical probability measures* over $\mathbb{R}^d$. For any $N \geq 1$, we denote by $\boldsymbol{x} = (x_1,\ldots,x_N)$ a given element of $(\mathbb{R}^d)^N$ and by $\mu[\boldsymbol{x}] := \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \in \mathscr{P}_N(\mathbb{R}^d)$ its associated empirical measure.

Definition 9 (Symmetric maps defined over $(\mathbb{R}^d)^N$). *A map $\boldsymbol{\phi} : (\mathbb{R}^d)^N \to \mathbb{R}^m$ is said to be* symmetric *if $\boldsymbol{\phi} \circ \sigma(\cdot) = \boldsymbol{\phi}(\cdot)$ for any d-blockwise permutation $\sigma : (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$.*

In the following definition, we introduce the notion of *mean-field approximating sequence* for continuous functionals defined over $\mathscr{P}_c(\mathbb{R}^d)$.

Definition 10 (Mean-field approximating sequence). *Given an integer $n \geq 1$ and a set $\Omega \subset \mathbb{R}^n$, we define the mean-field approximating sequence of a functional $F \in C^0(\Omega \times \mathscr{P}_c(\mathbb{R}^d),\mathbb{R}^m)$ as the family of symmetric maps $(\boldsymbol{F}_N(\cdot,\cdot)) \subset C^0(\Omega \times (\mathbb{R}^d)^N,\mathbb{R}^m)$, defined by*

$$\boldsymbol{F}_N(x,\boldsymbol{x}) := F(x,\mu[\boldsymbol{x}]), \tag{14}$$

for any $N \geq 1$ and all $(x,\boldsymbol{x}) \in \Omega \times (\mathbb{R}^d)^N$.

We henceforth endow the vector space $(\mathbb{R}^d)^N$ with the rescaled inner product $\langle \cdot, \cdot \rangle_N$ defined by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_N = \tfrac{1}{N} \sum_{i=1}^{N} \langle x_i, y_i \rangle, \tag{15}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in (\mathbb{R}^d)^N$, where $\langle \cdot, \cdot \rangle$ is the standard Euclidean product of $\mathbb{R}^d$. We also denote by $|\cdot|_N = \sqrt{\langle \cdot, \cdot \rangle_N}$ the corresponding norm over $(\mathbb{R}^d)^N$, and observe that $((\mathbb{R}^d)^N, \langle \cdot, \cdot \rangle_N)$ is an Hilbert space.

In the following proposition, we show that the Wasserstein differential structure described in Section 2.2 for functionals defined over $\mathscr{P}_2(\mathbb{R}^d)$ induces a natural differential structure on the Hilbert space $((\mathbb{R}^d)^N, \langle \cdot, \cdot \rangle_N)$. We will use the notation $C_{\mathrm{loc}}^{2,1}$ to refer to functionals between finite-dimensional normed vector spaces which are twice differentiable with locally Lipschitz derivatives up to the second-order.

Proposition 4 (Mean-field derivatives of symmetric maps)**.** *Let* $\phi(\cdot)$ *be* $\mathcal{C}_{\mathrm{loc}}^{2,1}$*-Wasserstein regular in the sense of Definition 8 above and* $(\boldsymbol{\phi}_N(\cdot)) \subset C^0((\mathbb{R}^d)^N)$ *be the mean-field approximating sequence of* $\phi(\cdot)$*.*

Then, $\boldsymbol{\phi}_N \in C_{\mathrm{loc}}^{2,1}((\mathbb{R}^d)^N, \mathbb{R})$ *for any* $N \geq 1$*, and the following Taylor expansion formula*

$$\boldsymbol{\phi}_N(\boldsymbol{x} + \boldsymbol{h}) = \boldsymbol{\phi}_N(\boldsymbol{x}) + \langle \mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x}), \boldsymbol{h} \rangle_N + \tfrac{1}{2} \mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h}) + o(|\boldsymbol{h}|_N^2), \tag{16}$$

holds for any $\boldsymbol{x}, \boldsymbol{h} \in (\mathbb{R}^d)^N$*. Here, we introduced the* mean-field gradient $\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot)$ *and mean-field Hessian bilinear form* $\mathbf{Hess}\, \boldsymbol{\phi}_N[\cdot]$ *of* $\boldsymbol{\phi}_N(\cdot)$*, given respectively by*

$$\mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x}) := (\nabla_\mu \phi(\mu[\boldsymbol{x}])(x_i))_{1 \leq i \leq N}, \tag{17}$$

and

$$\mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}^1, \boldsymbol{h}^2) := \frac{1}{N} \sum_{i=1}^{N} \langle \mathrm{D}_x \nabla_\mu \phi(\mu[\boldsymbol{x}])(x_i) h_i^1, h_i^2 \rangle_N + \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathrm{D}_\mu^2 \phi(\mu[\boldsymbol{x}])(x_i, x_j) h_i^1, h_j^2 \rangle, \tag{18}$$

for any $\boldsymbol{x}, \boldsymbol{h}^1, \boldsymbol{h}^2 \in (\mathbb{R}^d)^N$*. Defining the* $C_N^{2,1}$*-norm of* $\boldsymbol{\phi}_N(\cdot)$ *over* $\boldsymbol{K} \subset (\mathbb{R}^d)^N$ *with respect to differential structure of* $((\mathbb{R}^d)^N, \langle \cdot, \cdot \rangle_N)$ *as*

$$\|\boldsymbol{\phi}_N(\cdot)\|_{C_N^{2,1}(\boldsymbol{K})} := \max_{\boldsymbol{x} \in \boldsymbol{K}} \left(\boldsymbol{\phi}_N(\boldsymbol{x}) + |\mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x})|_N + \max_{|\boldsymbol{h}|_N = 1} |\mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h})| \right) + \mathrm{Lip}\left(\mathbf{Hess}\, \boldsymbol{\phi}_N[\cdot]; \boldsymbol{K} \right), \tag{19}$$

it further holds for each compact set $K \subset \mathbb{R}^d$ *that*

$$\|\boldsymbol{\phi}_N(\cdot)\|_{C_N^{2,1}(K^N)} \leq \|\phi(\cdot)\|_{\mathcal{C}^{2,1}(K)}, \tag{20}$$

where the $\mathcal{C}^{2,1}$*-Wasserstein norm* $\|\phi(\cdot)\|_{\mathcal{C}^{2,1}(K)}$ *of* $\phi(\cdot)$ *is defined as in* (13)*.*

Proof. Take $\boldsymbol{x} := (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ and $\boldsymbol{h} := (h_1, \dots, h_N) \in (\mathbb{R}^d)^N$ and define $\epsilon := \frac{1}{4} \min_{x_i \neq x_j} |x_i - x_j|$. Consider the map $\zeta_N(\cdot)$ given by

$$\zeta_N : x \in \mathbb{R}^d \mapsto \begin{cases} \langle x, h_i \rangle & \text{if } x \in B(x_i, 2\epsilon) \text{ with } i \in \{1, \dots, N\}, \\ 0 & \text{otherwise,} \end{cases}$$

and let $\eta \in C_c^\infty(\mathbb{R}^d)$ be a symmetric mollifier centred at the origin and supported on $B(0, \epsilon)$. We define the tangent vector $\xi_N \in \nabla C_c^\infty(\mathbb{R}^d) \subset \mathrm{Tan}_{\mu[\boldsymbol{x}]} \mathscr{P}_2(\mathbb{R}^d)$ at $\mu[\boldsymbol{x}]$ by

$$\xi_N : x \in \mathbb{R}^d \mapsto \nabla(\eta * \zeta_N)(x). \tag{21}$$

Remark that by construction, one has

$$\xi_N(x_i) = h_i \qquad \text{and} \qquad \mathrm{D}_x \xi_N(x_i) = 0, \tag{22}$$

so that in particular $\mu[\boldsymbol{x} + s\boldsymbol{h}] = (\mathrm{Id} + s\xi_N)_{\#} \mu[\boldsymbol{x}]$ for any $s \in \mathbb{R}$ sufficiently small.

Recall now that $\phi(\cdot)$ is differentiable at $\mu[\boldsymbol{x}] \in \mathscr{P}_c(\mathbb{R}^d)$ by hypothesis. Hence by Proposition 2, it holds

$$\lim_{s \to 0} \left[\frac{\phi(\mu[\boldsymbol{x} + s\boldsymbol{h}]) - \phi(\mu[\boldsymbol{x}])}{s} \right] = \mathcal{L}_{\xi_N} \phi(\mu[\boldsymbol{x}]) = \int_{\mathbb{R}^d} \langle \nabla_\mu \phi(\mu[\boldsymbol{x}])(x), \xi_N(x) \rangle \mathrm{d}\mu[\boldsymbol{x}](x).$$

Recalling the definition of approximating maps $\boldsymbol{\phi}_N(\cdot)$ given in (14), we further obtain

$$\boldsymbol{\phi}_N'(\boldsymbol{x}; \boldsymbol{h}) := \lim_{s \to 0} \left[\frac{\boldsymbol{\phi}_N(\boldsymbol{x} + s\boldsymbol{h}) - \boldsymbol{\phi}_N(\boldsymbol{x})}{s} \right] = \frac{1}{N} \sum_{i=1}^{N} \langle \nabla_\mu \phi(\mu[\boldsymbol{x}])(x_i), h_i \rangle, \tag{23}$$

where we used (22) along with the fact that $\mu[\boldsymbol{x}] = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$. It is straightforward to check that the directional derivative $\boldsymbol{h} \mapsto \boldsymbol{\phi}'_N(\boldsymbol{x};\boldsymbol{h})$ of $\boldsymbol{\phi}_N(\cdot)$ defined in (23) is a linear form and that it is continuous with respect to the rescaled Euclidean metric $|\cdot|_N$. Whence, the map $\boldsymbol{\phi}_N(\cdot)$ is Fréchet differentiable at $\boldsymbol{x} \in (\mathbb{R}^d)^N$, and by Riesz's Theorem (see e.g. [12, Theorem 5.5]), its differential can be represented in the Hilbert space $((\mathbb{R}^d)^N, \langle\cdot,\cdot\rangle_N)$ by the mean-field gradient $\mathbf{Grad}\,\boldsymbol{\phi}_N(\boldsymbol{x}) := (\nabla_\mu \phi(\mu[\boldsymbol{x}])(x_i))_{1\le i\le N}$ defined in (17).

Consider now two elements $\boldsymbol{h}^1, \boldsymbol{h}^2 \in (\mathbb{R}^d)^N$ and the corresponding tangent vectors $\xi_N^1, \xi_N^2 \in \nabla C_c^\infty(\mathbb{R}^d)$ built as in (21). Since $\phi(\cdot)$ is twice differentiable in the sense of Definition 7, it holds by (12) in Proposition 3

$$\lim_{s\to 0}\left[\frac{\mathcal{L}_{\xi_N^1}\phi((\mathrm{Id}+s\xi_N^2)_\#\mu[\boldsymbol{x}]) - \mathcal{L}_{\xi_N^1}\phi(\mu[\boldsymbol{x}])}{s}\right] = \mathrm{Hess}\,\phi[\mu[\boldsymbol{x}]](\xi_N^1,\xi_N^2) + \mathcal{L}_{\mathrm{D}\xi_N^1\xi_N^2}\phi(\mu[\boldsymbol{x}]). \qquad (24)$$

Observe now that $\mathrm{D}\xi_N^1(x) = 0$ for $\mu[\boldsymbol{x}]$-almost every $x \in \mathbb{R}^d$ by (21), so that $\mathcal{L}_{\mathrm{D}\xi_N^1\xi_N^2}\phi(\mu[\boldsymbol{x}]) = 0$. Furthermore, by the definition of $\boldsymbol{\phi}_N(\cdot)$ along with that of $\mathbf{Grad}\,\boldsymbol{\phi}_N(\cdot)$, equation (24) can be equivalently rewritten as

$$\lim_{s\to 0}\left[\frac{\langle\mathbf{Grad}\,\boldsymbol{\phi}_N(\boldsymbol{x}+s\boldsymbol{h}^2) - \mathbf{Grad}\,\boldsymbol{\phi}_N(\boldsymbol{x}), \boldsymbol{h}^1\rangle_N}{s}\right]$$
$$= \frac{1}{N}\sum_{i=1}^N \langle \mathrm{D}_x\nabla_\mu\phi(\mu[\boldsymbol{x}])(x_i)h_i^1, h_i^2\rangle + \frac{1}{N^2}\sum_{i,j=1}^N \langle \mathrm{D}_\mu^2\phi(\mu[\boldsymbol{x}])(x_i,x_j)h_i^1, h_j^2\rangle, \qquad (25)$$

where we used the analytical expression (11) of the Wasserstein Hessian. We accordingly introduce the mean-field Hessian bilinear form $\mathbf{Hess}\,\boldsymbol{\phi}_N[\boldsymbol{x}](\cdot,\cdot)$ of $\boldsymbol{\phi}_N(\cdot)$ at $\boldsymbol{x} \in (\mathbb{R}^d)^N$, defined as in (18). It is again possible to verify that $\mathbf{Hess}\,\boldsymbol{\phi}_N[\boldsymbol{x}](\cdot,\cdot)$ defines a continuous bilinear form with respect to the rescaled metric $|\cdot|_N$, so that the map $\boldsymbol{\phi}_N(\cdot)$ is twice Fréchet differentiable over $(\mathbb{R}^d)^N$. The expansion formula (16) can then be derived by developing $\boldsymbol{\phi}_N(\boldsymbol{x}+\boldsymbol{h})$ using the classical Taylor theorem in $(\mathbb{R}^d)^N$ along with (23) and (25).

We now prove the regularity bound of (20). Given $K \subset \mathbb{R}^d$, we obtain from the fact that $(\boldsymbol{\phi}_N(\cdot))$ is a mean-field approximating sequence for $\phi(\cdot)$ together with the definition of $\mathbf{Grad}\,\boldsymbol{\phi}_N(\cdot)$ displayed in (17), that

$$\max_{\boldsymbol{x}\in K^N}|\boldsymbol{\phi}_N(\boldsymbol{x})| = \max_{\boldsymbol{x}\in K^N}|\phi(\mu[\boldsymbol{x}])| \le \max_{\mu\in\mathscr{P}(K)}|\phi(\mu)|, \qquad (26)$$

and

$$\max_{\boldsymbol{x}\in K^N}|\mathbf{Grad}\,\boldsymbol{\phi}_N(\boldsymbol{x})|_N = \max_{\boldsymbol{x}\in K^N}\left(\frac{1}{N}\sum_{i=1}^N|\nabla_\mu\phi(\mu[\boldsymbol{x}])(x_i)|^2\right)^{1/2} \le \max_{\mu\in\mathscr{P}(K)}\|\nabla_\mu\phi(\mu)(\cdot)\|_{C^0(K,\mathbb{R}^d)}. \qquad (27)$$

Analogously, using the definition of $\mathbf{Hess}\,\boldsymbol{\phi}_N(\cdot)$ given in (18), we can deduce

$$\max_{\boldsymbol{x}\in K^N}\mathbf{Hess}\,\boldsymbol{\phi}[\boldsymbol{x}](\boldsymbol{h},\boldsymbol{h}) \le \max_{\mu\in\mathscr{P}(K)}\left(\|\mathrm{D}_x\nabla_\mu\phi(\mu)(\cdot)\|_{C^0(K,\mathbb{R}^{d\times d})} + \|\mathrm{D}_\mu^2\phi(\mu)(\cdot,\cdot)\|_{C^0(K\times K,\mathbb{R}^{d\times d})}\right), \qquad (28)$$

as well as the Lipschitz estimate

$$\mathrm{Lip}(\mathbf{Hess}\,\boldsymbol{\phi}_N[\cdot]; K^N) \le \mathrm{Lip}(\mathrm{D}_x\nabla_\mu\phi(\cdot)(\cdot); \mathscr{P}(K)\times K) + \mathrm{Lip}(\mathrm{D}_\mu^2\phi(\cdot)(\cdot,\cdot); \mathscr{P}(K)\times K\times K), \qquad (29)$$

where we used the fact that $W_2(\mu[\boldsymbol{x}],\mu[\boldsymbol{y}]) \le |\boldsymbol{x}-\boldsymbol{y}|_N$ for $\boldsymbol{x},\boldsymbol{y} \in (\mathbb{R}^d)^N$. By plugging (26), (27), (28) and (29) into (19) and recalling the definition (13) of $\|\phi(\cdot)\|_{\mathcal{C}^{2,1}(K)}$, we conclude that (20) holds. $\qquad\square$

Remark 2 (Matrix representation of the mean-field Hessian in $(\mathbb{R}^d)^N$). *By Riesz's Theorem applied in the Hilbert space $((\mathbb{R}^d)^N), \langle\cdot,\cdot\rangle_N)$, the action of the Hessian bilinear form $\mathbf{Hess}\,\boldsymbol{\phi}_N[\boldsymbol{x}](\cdot,\cdot)$ can be represented as*

$$\mathbf{Hess}\,\boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}^1,\boldsymbol{h}^2) = \langle\mathbf{Hess}\,\boldsymbol{\phi}_N(\boldsymbol{x})\boldsymbol{h}^1, \boldsymbol{h}^2\rangle_N, \qquad (30)$$

for any $\boldsymbol{x},\boldsymbol{h}^1,\boldsymbol{h}^2 \in (\mathbb{R}^d)^N$, where $\mathbf{Hess}\,\boldsymbol{\phi}_N(\boldsymbol{x}) \in \mathbb{R}^{dN\times dN}$ is a matrix. In this case, its components can be obtained via a simple identification in (18), and be written explicitly as

$$(\mathbf{Hess}\,\boldsymbol{\phi}_N(\boldsymbol{x}))_{i,j} = \mathrm{D}_\mu^2\phi(\mu[\boldsymbol{x}])(x_i,x_j), \qquad (\mathbf{Hess}\,\boldsymbol{\phi}_N(\boldsymbol{x}))_{i,i} = N\mathrm{D}_x\nabla_\mu\phi(\mu[\boldsymbol{x}])(x_i) + \mathrm{D}_\mu^2\phi(\mu[\boldsymbol{x}])(x_i,x_i),$$

for any pair of indices $i,j \in \{1,\dots,N\}$ such that $i \ne j$.

3 Locally optimal Lipschitz feedbacks in optimal control

In this section, we recall classical facts about finite dimensional optimal control problems, and describe in Theorem 4 a result proven in [33], which provides sufficient conditions for the existence of locally optimal

Lipschitz feedbacks in a neighbourhood of an optimal trajectory. Throughout this section, we will study the finite-dimensional optimal control problem

$$(\mathcal{P}_{oc}) \begin{cases} \min\limits_{u(\cdot) \in \mathcal{U}} \left[\int_0^T \Big(l(t,x(t)) + \psi(u(t)) \Big) \mathrm{d}t + g(x(T)) \right] \\ \text{s.t.} \begin{cases} \dot{x}(t) = f(t,x(t)) + u(t), \\ x(0) = x^0, \end{cases} \end{cases}$$

under the following assumptions.

Hypotheses ($\mathbf{H_{oc}}$).

(i) *The set of admissible controls is given by $\mathcal{U} = L^\infty([0,T], U)$ where $U \subset \mathbb{R}^d$ is convex and compact.*

(ii) *The control cost $u \mapsto \psi(u) \in \mathbb{R}$ is $C^{2,1}$-regular and strictly convex over U.*

(iii) *The map $(t,x) \mapsto f(t,x) \in \mathbb{R}^d$ is Lipschitz with respect to $t \in [0,T]$ and $C^{2,1}_{loc}$-regular with respect to $x \in \mathbb{R}^d$. Moreover, there exists a constant $M > 0$ such that*

$$|f(t,x)| \le M(1 + |x|),$$

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

(iv) *The running cost $(t,x) \mapsto l(t,x) \in \mathbb{R}$ is Lipschitz with respect to $t \in [0,T]$ and $C^{2,1}_{loc}$-regular with respect to $x \in \mathbb{R}^d$. Similarly, the final cost $x \mapsto g(x) \in \mathbb{R}$ is $C^{2,1}_{loc}$-regular over $\mathbb{R}^d$.*

It can be easily seen that one could choose integrable maps to express the sub-linearity and Lipschitz regularity of $f(\cdot, \cdot)$ instead of constants. As a direct consequence of $(\mathbf{H_{oc}})$, we have the following lemma.

Lemma 1 (Uniform compactness of admissible trajectories). *Given $x^0 \in \mathbb{R}^d$, there exists a compact set $K \subset \mathbb{R}^d$ such that each admissible curve $x(\cdot)$ for $(\mathcal{P}_{oc})$ associated to a control $u(\cdot) \in \mathcal{U}$ satisfies $x(\cdot) \in \mathrm{Lip}([0,T], K)$.*

Proof. This follows directly from an application of Grönwall's Lemma. $\qquad\square$

Proposition 5 (Existence of solutions for problem $(\mathcal{P}_{oc})$). *Let $K \subset \mathbb{R}^d$ be a compact set given as in Lemma 1 and suppose that hypotheses $(\mathbf{H_{oc}})$ hold. Then, there exists an optimal trajectory-control pair $(x^*(\cdot), u^*(\cdot)) \in \mathrm{Lip}([0,T], K) \times \mathcal{U}$ for problem $(\mathcal{P}_{oc})$.*

Proof. This result is standard under our working hypotheses and can be found e.g. in [27, Theorem 23.11]. $\quad\square$

We introduce the *Hamiltonian* function associated with $(\mathcal{P}_{oc})$, defined by

$$H : (t,x,p,u) \in [0,T] \times (\mathbb{R}^d)^3 \mapsto \langle p, f(t,x) + u \rangle - \big(l(t,x) + \psi(u) \big).$$

Let $(x^*(\cdot), u^*(\cdot))$ be an optimal trajectory-control pair for $(\mathcal{P}_{oc})$. By the *Pontryagin Maximum Principle* (see e.g. [27, Theorem 22.2]), there exists a curve $p^*(\cdot)$ such that the couple $(x^*(\cdot), p^*(\cdot))$ is a solution of the *forward-backward Hamiltonian system*

$$\begin{cases} \dot{x}^*(t) = \nabla_p H(t,x^*(t),p^*(t),u^*(t)), & x^*(0) = x^0, \\ \dot{p}^*(t) = -\nabla_x H(t,x^*(t),p^*(t),u^*(t)), & p^*(T) = -\nabla g(x^*(T)). \end{cases} \tag{31}$$

Moreover, the *Pontryagin maximisation condition*

$$H(t,x^*(t),p^*(t),u^*(t)) = \max_{v \in U} H(t,x^*(t),p^*(t),v), \tag{32}$$

holds along this extremal pair for $\mathscr{L}^1$-almost every $t \in [0,T]$. Such a collection of optimal state, costate and control curves $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$ is called an *optimal Pontryagin triple* for $(\mathcal{P}_{oc})$. Let it be noted that, since the end-points of $(\mathcal{P}_{oc})$ are free, there are no abnormal curves stemming from the maximum principle.

Lemma 2 (Compactness and regularity of the costate). *Let $K \subset \mathbb{R}^d$ be a compact set given by Lemma 1 and suppose that hypotheses $(\mathbf{H_{oc}})$ hold. Then, there exists a compact set $K' \subset \mathbb{R}^d$ such that $p^*(\cdot) \in \mathrm{Lip}([0,T], K')$.*

Proof. The backward Cauchy problem satisfied by $p^*(\cdot)$ in (31) can be written explicitly as

$$\dot{p}^*(t) = -\mathrm{D}_x f(t,x^*(t))^\top p^*(t), \qquad p^*(T) = -\nabla g(x^*(T)),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. Since $(t,x) \in [0,T] \times K \mapsto |\mathrm{D}_x f(t,x)| \in \mathbb{R}_+$ is uniformly bounded by $(\mathbf{H_{oc}})$-(iii), it follows from Grönwall's Lemma that $p^*(\cdot) \in \mathrm{Lip}([0,T], \mathbb{R}^d)$. Moreover, recall that $x \in K \mapsto |\nabla g(x)| \in \mathbb{R}_+$ is also uniformly bounded as a consequence of $(\mathbf{H_{oc}})$-(iv), thus upon invoking Grönwall's Lemma again, there exists a compact set $K' \subset \mathbb{R}^d$ such that $p^*(\cdot) \in \mathrm{Lip}([0,T], K')$. $\qquad\square$

From now on, we denote by $\mathcal{K} = [0,T] \times K \times K' \times U$ the uniform compact set containing the admissible times, states, costates and controls for $(\mathcal{P}_{oc})$, and by $\mathcal{L}_\mathcal{K} > 0$ be the Lipschitz constant over $\mathcal{K}$ of the maps $H(\cdot,\cdot,\cdot,\cdot)$, $l(\cdot,\cdot)$, $\psi(\cdot)$ and $g(\cdot)$ and and of their derivatives with respect to the variables (x,u) up to the second order. Observe that both quantities exist as a consequence of Lemma 1, Lemma 2 and hypotheses $(\mathbf{H_{oc}})$.

Definition 11 (Coercivity estimate). *We say that an optimal Pontryagin triple $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$ for $(\mathcal{P}_{oc})$ satisfies the* uniform coercivity estimate *with constant $\rho > 0$ if the following inequality holds*

$$\langle \nabla_x^2 g(x^*(T))y(T), y(T) \rangle - \int_0^T \langle \nabla_x^2 H(t, x^*(t), p^*(t), u^*(t))y(t), y(t) \rangle \, dt$$
$$- \int_0^T \langle \nabla_u^2 H(t, x^*(t), p^*(t), u^*(t))w(t), w(t) \rangle \, dt \geq \rho \int_0^T |w(t)|^2 dt \tag{33}$$

for any pair of maps $(y(\cdot), w(\cdot)) \in W^{1,2}([0,T], \mathbb{R}^d) \times L^2([0,T], \mathbb{R}^d)$ solution of the linearised system

$$\begin{cases} \dot{y}(t) = D_x f(t, x^*(t))y(t) + w(t), \\ y(0) = 0 \quad and \quad u^*(t) + w(t) \in U \text{ for } \mathscr{L}^1\text{-almost every } t \in [0,T]. \end{cases} \tag{34}$$

We are now ready to recall the main contribution of [33, Theorem 5.2], which we will use in the proof of Theorem 2. Below, we use the notations $\text{Graph}(x(\cdot)) := \{(t, x(t)) \text{ s.t. } t \in [0,T]\}$ and $B(x,r) \subset \mathbb{R}^d$ for the closed-ball of center $x \in \mathbb{R}^m$ and radius $r > 0$ in $\mathbb{R}^d$.

Theorem 4 (Existence of locally optimal feedbacks for $(\mathcal{P}_{oc})$). *Let $(x^*(\cdot), p^*(\cdot), u^*(\cdot)) \in \text{Lip}([0,T], K) \times \text{Lip}([0,T], K') \times \mathcal{U}$ be an optimal Pontryagin triple for problem $(\mathcal{P}_{oc})$. Suppose that $(\mathbf{H_{oc}})$ hold and that $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$ satisfies the uniform coercivity estimate (33)-(34) with constant $\rho > 0$.*

Then, there exist positive constants $\epsilon, \eta > 0$, an open subset $\mathcal{N} \subset [0,T] \times \mathbb{R}^d$ and a locally optimal feedback $\bar{u}(\cdot, \cdot) \in \text{Lip}(\mathcal{N}, \mathbb{R}^d)$ whose Lipschitz constant depends only on $\mathcal{L}_K$ and ρ, such that the following holds.

(a) *$\bar{u}(t, x^*(t)) = u^*(t)$ for all times $t \in [0,T]$.*

(b) *$\big(\text{Graph}(x^*(\cdot)) + \{0\} \times B(0, \epsilon)\big) \subset \mathcal{N}$.*

(c) *For each $(\tau, \xi) \in \mathcal{N}$, the equation*

$$\dot{x}(t) = f(t, x(t)) + \bar{u}(t, x(t)), \qquad x(\tau) = \xi, \tag{35}$$

has a unique solution $\hat{x}_{(\tau,\xi)}(\cdot)$ such that $\text{Graph}(\hat{x}_{(\tau,\xi)}(\cdot)) \subset \mathcal{N}$.

(d) *The map $\hat{u}_{(\tau,\xi)} : t \in [\tau, T] \mapsto \bar{u}(t, \hat{x}_{(\tau,\xi)}(t)) \in U$ satisfies*

$$\int_\tau^T l\big(t, \hat{x}_{(\tau,\xi)}(t), \hat{u}_{(\tau,\xi)}(t)\big)dt + g(\hat{x}_{(\tau,\xi)}(T)) \leq \int_\tau^T l\big(t, x(t), u(t)\big)dt + g(x(T)),$$

for any open-loop pair $(u(\cdot), x(\cdot)) \in \mathcal{U} \times \text{Lip}([\tau, T], \mathbb{R}^d)$ for $(\mathcal{P}_{oc})$ such that $\|u(\cdot) - \hat{u}_{(\tau,\xi)}(\cdot)\|_{L^\infty([\tau,T])} \leq \eta$.

The proof of Theorem 4 in [33] is based on a general strategy elaborated in [26], in which several quantitative inverse function theorems are proven under hypotheses akin to (33) for non-linear optimal control problems. The key point of this approach is to remark (see e.g. [32]) that the first-order linearisation of the PMP system (31)-(32) corresponds to the PMP of the linearised problem

$$(\mathcal{P}'_{oc}) \begin{cases} \min_{w(\cdot) \in \mathcal{U}'} \left[\int_0^T \left(\frac{1}{2}\langle A(t)y(t), y(t) \rangle + \frac{1}{2}\langle B(t)w(t), w(t) \rangle \right) dt + \frac{1}{2}\langle C(T)y(T), y(T) \rangle \right] \\ \text{s.t.} \begin{cases} \dot{y}(t) = D_x f(t, x^*(t))y(t) + w(t), \\ y(0) = 0. \end{cases} \end{cases}$$

associated to $(\mathcal{P}_{oc})$, where

$$\mathcal{U}' = \left\{ v \in L^2([0,T], U) \text{ s.t. } u^*(t) + v(t) \in U \text{ for } \mathscr{L}^1\text{-almost every } t \in [0,T] \right\},$$

and

$$\begin{cases} A(t) = -\nabla_x^2 H(t, x^*(t), p^*(t), u^*(t)), \quad B(t) = -\nabla_u^2 H(t, x^*(t), p^*(t), u^*(t)), \\ C(T) = \nabla_x^2 g(x^*(T)). \end{cases}$$

for all times $t \in [0,T]$. In this context, the coercivity estimate plays the role of a strong positive-definiteness condition on the cost of $(\mathcal{P}'_{oc})$ along optimal trajectories, which allows to invert the corresponding optimality system with a control on the Lipschitz constant of the inverse. In the sequel, we will use the important fact that Theorem 4 holds true in any finite-dimensional Hilbert space, and in particular in $((\mathbb{R}^d)^N, \langle \cdot, \cdot \rangle_N)$.

4 Non-local transport equations and mean-field optimal control

In this section, we recall some results concerning continuity equations and mean-feld optimal control problems. We recall in Section 4.1 concepts pertaining to non-local continuity equations, and detail in Section 4.2 a powerful existence result of so-called *mean-field optimal controls* for problem $(\mathcal{P})$, which is borrowed from [38].

In the sequel, we focus on the optimal control problems in Wasserstein spaces written in the general form

$$(\mathcal{P}) \quad \begin{cases} \min\limits_{u \in \mathcal{U}} \left[\int_0^T \left(L(t, \mu(t)) + \int_{\mathbb{R}^d} \psi(u(t, x)) \mathrm{d}\mu(t)(x) \right) \mathrm{d}t + \varphi(\mu(T)) \right] \\ \text{s.t.} \begin{cases} \partial_t \mu(t) + \nabla \cdot \big((v(t, \mu(t), \cdot) + u(t, \cdot))\mu(t) \big) = 0, \\ \mu(0) = \mu^0. \end{cases} \end{cases}$$

Here, $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ is a fixed initial datum, and the minimisation is taken over the set of admissible controls $\mathcal{U} := L^\infty([0,T], L^1(\mathbb{R}^d, U; \mu(t))$ where $(\mu(\cdot), u(\cdot))$ is a trajectory-control pair. We make the following working assumption on the data of problem $(\mathcal{P})$.

Hypotheses (H).

(i) *The set of admissible control values $U \subset \mathbb{R}^d$ is convex and compact.*

(ii) *The control cost $u \mapsto \psi(u) \in \mathbb{R}$ is $C^{2,1}$-regular and strictly convex over U.*

(iii) *The non-local velocity field $(t, x, \mu) \mapsto v(t, \mu, x) \in \mathbb{R}^d$ is Lipschitz with respect to $t \in [0, T]$ and continuous in the $|\cdot| \times W_2$-topology with respect to $(x, \mu) \in \mathbb{R}^d \times \mathscr{P}_c(\mathbb{R}^d)$. Besides, there exists $M > 0$ such that*

$$|v(t, \mu, x)| \leq M\Big(1 + |x| + \big(\textstyle\int_{\mathbb{R}^d} |y| \mathrm{d}\mu(y) \big) \Big),$$

for all times $t \in [0, T]$ and any $(x, \mu) \in \mathbb{R}^d \times \mathscr{P}_c(\mathbb{R}^d)$. Moreover, there exist $l_K, L_K > 0$ such that

$$|v(t, \mu, x) - v(t, \mu, y)| \leq l_K |x - y| \qquad and \qquad |v(t, \mu, x) - v(t, \nu, x)| \leq L_K W_2(\mu, \nu),$$

for any $x, y \in K$ and $\mu, \nu \in \mathscr{P}(K)$, where $K \subset \mathbb{R}^d$ is an arbitrary compact set.

(iv) *The map $\mu \mapsto v(t, \mu, x) \in \mathbb{R}^d$ is $\mathcal{C}^{2,1}_{\mathrm{loc}}$-Wasserstein regular.*

(v) *The running cost $(t, \mu) \mapsto L(t, \mu) \in \mathbb{R}$ is Lipschitz with respect to $t \in [0, T]$ and $\mathcal{C}^{2,1}_{\mathrm{loc}}$-Wasserstein regular with respect to $\mu \in \mathscr{P}_c(\mathbb{R}^d)$.*

(vi) *The final cost $\mu \mapsto \varphi(\mu) \in \mathbb{R}$ is $\mathcal{C}^{2,1}_{\mathrm{loc}}$-Wasserstein regular.*

Observe that by classical well-posedness results for non-local continuity equations (see e.g. [9, 52]) together with known existence results in the context mean-field optimal control problems (see e.g. [38]), it would be sufficient to have locally Lipschitz dynamics and continuous cost functionals for solutions of $(\mathcal{P})$ to exist.

4.1 Non-local transport equations in $\mathbb{R}^d$

Given a time horizon $T > 0$, we denote by $\lambda := \frac{1}{T} \mathscr{L}^1_{\llcorner [0,T]}$ the renormalised Lebesgue measure on $[0, T]$. For any $p \geq 1$, a curve of measures $\mu(\cdot) \in C^0([0, T], \mathscr{P}_p(\mathbb{R}^d))$ can be uniquely lifted to a measure $\tilde{\mu} \in \mathscr{P}_p([0, T] \times \mathbb{R}^d)$ defined by disintegration as $\tilde{\mu} = \int_{[0,T]} \mu(t) \mathrm{d}\lambda(t)$ in the sense of Theorem 3. We shall say that $\mu(\cdot) \in C^0([0, T], \mathscr{P}_p(\mathbb{R}^d))$ solves a *continuity equation* with initial condition $\mu^0 \in \mathscr{P}_p(\mathbb{R}^d)$ driven by a Lebesgue-Borel velocity field $\boldsymbol{w} \in L^p([0, T] \times \mathbb{R}^d, \mathbb{R}^d; \tilde{\mu})$ provided that

$$\begin{cases} \partial_t \mu(t) + \nabla \cdot (\boldsymbol{w}(t, \cdot) \mu(t)) = 0, \\ \mu(0) = \mu^0. \end{cases} \tag{36}$$

This equation has to be understood in duality against smooth and compactly supported functions, namely

$$\int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \xi(t, x) + \langle \nabla_x \xi(t, x), \boldsymbol{w}(t, x) \rangle \Big) \mathrm{d}\mu(t)(x) \mathrm{d}t = 0 \tag{37}$$

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

It has been well-known since the works of Ambrosio in [2] (see also [5, Chapter 8]) that weak solutions of continuity equations can exist in this low regularity context. However as already explained in the introduction above, such solutions are not well tailored to the practical investigation of mean-field control problem. Thus in Theorem 5 below, we recall an existence result which was first derived in [52], and that is concerned with classical well-posedness for *non-local transport* equations in $(\mathscr{P}_c(\mathbb{R}^d), W_1)$ under stronger regularity assumptions.

Theorem 5 (Well-posedness of non-local transport equations). *Let $v : (t, \mu, x) \in [0, T] \times \mathscr{P}_c(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$ be a non-local velocity field satisfying hypotheses **(H)**-(iii). Then for each $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$, there exists a unique solution $\mu(\cdot) \in \mathrm{Lip}([0, T], \mathscr{P}_c(\mathbb{R}^d))$ of (36) driven by $\boldsymbol{w} : (t, x) \in [0, T] \mapsto v(t, \mu(t), x) \in \mathbb{R}^d$. Furthermore, there exist constants $R_T, L_T > 0$ such that*

$$\mathrm{supp}(\mu(t)) \subset B(0, R_T) \qquad and \qquad W_1(\mu(t), \mu(s)) \le L_T |t - s|,$$

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

4.2 Existence of mean-field optimal controls for problem $(\mathcal{P})$

In this section, we show how problem $(\mathcal{P})$ can be reformulated so as to encompass a suitable sequence of approximating discrete problems $(\mathcal{P}_N)$. We subsequently recall a powerful existence result derived in [38] for general multi-agent optimal control problems formulated in Wasserstein spaces.

We start by fixing an integer $N \ge 1$, an initial datum $\boldsymbol{x}_N^0 \in (\mathbb{R}^d)^N$, and the associated empirical measure $\mu_N^0 = \mu[\boldsymbol{x}_N^0]$ as in Section 2.3. As exposed in the introduction, we consider the family of discrete problems

$$(\mathcal{P}_N) \quad \begin{cases} \displaystyle\min_{\boldsymbol{u}(\cdot) \in \mathcal{U}_N} \left[\int_0^T \left(\boldsymbol{L}_N(t, \boldsymbol{x}(t)) + \frac{1}{N} \sum_{i=1}^N \psi(u_i(t)) \right) \mathrm{d}t + \boldsymbol{\varphi}_N(\boldsymbol{x}(T)) \right] \\ \text{s.t.} \quad \begin{cases} \dot{x}_i(t) = \boldsymbol{v}_N(t, \boldsymbol{x}(t), x_i(t)) + u_i(t), \\ x_i(0) = x_i^0, \end{cases} \end{cases}$$

with $\mathcal{U}_N = L^\infty([0, T], U^N)$, and where the mean-field approximating functionals are defined by

$$\boldsymbol{v}_N(t, \boldsymbol{x}, x) := v(t, \mu[\boldsymbol{x}], x), \qquad \boldsymbol{L}_N(t, \boldsymbol{x}) := L(t, \mu[\boldsymbol{x}]) \qquad \text{and} \qquad \boldsymbol{\varphi}_N(\boldsymbol{x}) := \varphi(\mu[\boldsymbol{x}]), \tag{38}$$

for any $(t, \boldsymbol{x}, x) \in [0, T] \times (\mathbb{R}^d)^N \times \mathbb{R}^d$. It can be checked that as a consequence of hypotheses **(H)**, the problems $(\mathcal{P}_N)$ satisfy hypotheses **(H$_{\mathbf{oc}}$)**. We can thus deduce the following lemma directly from Proposition 5.

Lemma 3 (Existence of solutions for $(\mathcal{P}_N)$). *Under hypotheses **(H)** for each $N \ge 1$, there exists an optimal trajectory-control pair $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot)) \in \mathrm{Lip}([0, T], (\mathbb{R}^d)^N) \times \mathcal{U}_N$ solution of $(\mathcal{P}_N)$.*

We proceed by recasting problem $(\mathcal{P})$ into a framework which also encompasses the sequence of problems $(\mathcal{P}_N)$. Recall that, by Definition 1, a vector-valued measure $\boldsymbol{\nu} \in \mathcal{M}([0, T] \times \mathbb{R}^d, U)$ is absolutely continuous with respect to $\tilde{\mu}$ if and only if there exists a map $u(\cdot, \cdot) \in L^1([0, T] \times \mathbb{R}^d, U; \tilde{\mu})$ such that $\boldsymbol{\nu} = u(\cdot, \cdot)\tilde{\mu}$. Moreover, the absolute continuity of $\boldsymbol{\nu}$ with respect to $\tilde{\mu}$ implies the existence of a λ-almost unique family of measures $\{\boldsymbol{\nu}(t)\}_{t \in [0,T]}$ such that $\boldsymbol{\nu} = \int_{[0,T]} \boldsymbol{\nu}(t) \mathrm{d}\lambda(t)$ in the sense of Theorem 3. Whence, problem $(\mathcal{P})$ can be relaxed as

$$(\mathcal{P}_{\mathrm{meas}}) \quad \begin{cases} \displaystyle\min_{\boldsymbol{\nu} \in \mathscr{U}} \left[\int_0^T \left(L(t, \mu(t)) + \Psi(\boldsymbol{\nu}(t)|\mu(t)) \right) \mathrm{d}t + \varphi(\mu(T)) \right] \\ \text{s.t.} \quad \begin{cases} \partial_t \mu(t) + \nabla \cdot \left(v(t, \mu(t), \cdot)\mu(t) + \boldsymbol{\nu}(t) \right) = 0, \\ \mu(0) = \mu^0. \end{cases} \end{cases}$$

where we introduced the set $\mathscr{U} = \mathcal{M}([0, T] \times \mathbb{R}^d, U)$ of *generalised measure controls*, and the map

$$\Psi(\,\cdot\,|\mu) : \boldsymbol{\sigma} \in \mathcal{M}(\mathbb{R}^d, U) \mapsto \begin{cases} \displaystyle\int_{\mathbb{R}^d} \psi\left(\frac{\mathrm{d}\boldsymbol{\sigma}}{\mathrm{d}\mu}(x) \right) \mathrm{d}\mu(x) & \text{if } \boldsymbol{\sigma} \ll \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

One can then associate to any optimal trajectory-control pair $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot)) \in \mathrm{Lip}([0, T], (\mathbb{R}^d)^N) \times \mathcal{U}_N$ for $(\mathcal{P}_N)$ a measure trajectory-control pair $(\mu_N^*(\cdot), \boldsymbol{\nu}_N^*(\cdot)) \in \mathrm{Lip}([0, T], \mathscr{P}_N(\mathbb{R}^d)) \times \mathscr{U}$, defined by

$$\mu_N^*(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i^*(\cdot)}, \qquad \boldsymbol{\nu}_N^* = \int_{[0,T]} \left(\frac{1}{N} \sum_{i=1}^N u_i^*(t) \delta_{x_i^*(t)} \right) \mathrm{d}\lambda(t). \tag{39}$$

In the following theorem, we state a condensed version of the main result of [38], which shows that this relaxation allows to prove the Γ-convergence of the discrete problems $(\mathcal{P}_N)$ towards $(\mathcal{P})$.

Theorem 6 (Existence of mean-field optimal controls for $(\mathcal{P})$). *Let $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ be given, $(\mu_N^0) \subset \mathscr{P}_c(\mathbb{R}^d)$ be a sequence of uniformly compactly supported empirical measures associated with $(\boldsymbol{x}_N^0) \subset (\mathbb{R}^d)^N$ such that $W_1(\mu_N^0, \mu^0) \longrightarrow 0$ as $N \to +\infty$, and assume that hypotheses $\mathbf{(H)}$ hold. For any $N \geq 1$, let $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot)) \in \mathrm{Lip}([0,T], (\mathbb{R}^d)^N) \times \mathcal{U}_N$ be an optimal trajectory-control pair for $(\mathcal{P}_N)$ and $(\mu_N^*(\cdot), \boldsymbol{\nu}_N^*) \in \mathrm{Lip}([0,T], \mathscr{P}_N(\mathbb{R}^d)) \times \mathscr{U}$ be the corresponding measure trajectory-control pair defined as in (39).*

Then, there exists a pair $(\mu^*(\cdot), \boldsymbol{\nu}^*) \in \mathrm{Lip}([0,T], \mathscr{P}_c(\mathbb{R}^d)) \times \mathscr{U}$ such that

$$\max_{t \in [0,T]} W_1(\mu_N^*(t), \mu^*(t)) \underset{N \to +\infty}{\longrightarrow} 0 \qquad and \qquad \boldsymbol{\nu}_N^* \underset{N \to +\infty}{\overset{*}{\rightharpoonup}} \boldsymbol{\nu}^*,$$

along a suitable subsequence. Moreover, the classical trajectory-control pair

$$\left(\mu^*(\cdot), \tfrac{\mathrm{d}\boldsymbol{\nu}^*}{\mathrm{d}\tilde{\mu}^*}(\cdot, \cdot) \right) \in \mathrm{Lip}([0,T], \mathscr{P}_c(\mathbb{R}^d)) \times L^\infty([0,T] \times \mathbb{R}^d, U; \tilde{\mu}^*),$$

is optimal for $(\mathcal{P})$, where $\tilde{\mu}^* = \int_{[0,T]} \mu^*(t) \mathrm{d}\lambda(t)$.

Remark 3 (Comparison between $\mathbf{(H)}$ and the assumptions of [38]). *In [38], it is assumed that $U \subset \mathbb{R}^d$ is a subspace of $\mathbb{R}^d$ in order to recover the $\Gamma - \limsup$ inequality in the proof of their main result Theorem 3.2. This hypothesis could be relaxed up to an additional projection argument by asking that U is convex and closed. Besides, the requirements that $\psi(\cdot)$ is radial and super-linear at infinity are primarily used to recover integral bounds on the controls, which automatically hold in our context since we posit that the control set U is compact.*

5 Proof of Theorem 1 and Theorem 2

In this section, we prove the two main results of this article. We start by working with the discrete approximations $(\mathcal{P}_N)$ of $(\mathcal{P})$ in order to prove Theorem 2. We then proceed to recover Theorem 1 as a corollary, by formulating a sufficient condition under which $\mathbf{(CO}_N)$ below holds.

5.1 Mean-field coercivity estimate and proof of Theorem 2

In this section, we start by proving Theorem 2. We suppose that hypotheses $\mathbf{(H)}$ of Section 4 hold, along with the following additional *mean-field coercivity* assumption.

Hypothesis $\mathbf{(CO}_N)$. *There exists a constant $\rho_T > 0$ such that for every mean-field optimal Pontryagin triple $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot))$ for $(\mathcal{P}_N)$ defined in the sense of Proposition 6 below, the following coercivity estimate holds*

$$\mathbf{Hess}_{\boldsymbol{x}}\, \varphi_N[\boldsymbol{x}_N^*(T)](\boldsymbol{y}(T), \boldsymbol{y}(T)) - \int_0^T \mathbf{Hess}_{\boldsymbol{x}}\, \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{y}(t), \boldsymbol{y}(t)) \mathrm{d}t$$

$$- \int_0^T \mathbf{Hess}_{\boldsymbol{u}}\, \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{w}(t), \boldsymbol{w}(t)) \mathrm{d}t \geq \rho_T \int_0^T |\boldsymbol{w}(t)|_N^2 \mathrm{d}t,$$

along all the solutions $(\boldsymbol{y}(\cdot), \boldsymbol{w}(\cdot)) \in W^{1,2}([0,T], (\mathbb{R}^d)^N) \times L^2([0,T], U^N)$ of the linearised system

$$\begin{cases} \dot{y}_i(t) = \mathrm{D}_x \boldsymbol{v}(t, \boldsymbol{x}_N^*(t), x_i^*(t)) y_i(t) + \frac{1}{N} \sum_{j=1}^N \mathrm{D}_{x_j} \boldsymbol{v}_N(t, \boldsymbol{x}_N^*(t), x_i^*(t)) y_j(t) + w_i(t), \\ y_i(0) = 0 \qquad and \qquad \boldsymbol{u}_N^*(t) + \boldsymbol{w}(t) \in U^N \quad for\ \mathscr{L}^1\text{-almost every } t \in [0,T]. \end{cases}$$

Our argument is split into three steps. In Step 1, we write a PMP adapted to the mean-field structure of problem $(\mathcal{P}_N)$. We proceed by building in Step 2 a sequence of Lipschitz-in-space optimal control maps for the discrete problems $(\mathcal{P}_N)$ by combining Theorem 4 and $\mathbf{(CO}_N)$. We then show in Step 3 that this sequence of control maps is compact in a suitable weak topology preserving its Lipschitz regularity in space, and that its limit point coincide with the mean-field optimal control introduced in Theorem 6.

Step 1: Solutions of $(\mathcal{P}_N)$ and mean-field Pontryagin Maximum Principle. In this first step, we characterise and derive uniform estimates on the optimal pairs $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot))$ for $(\mathcal{P}_N)$. Our analysis is based on a reformulation of the PMP applied to $(\mathcal{P}_N)$ as a Hamiltonian flow with respect to the inner product $\langle \cdot, \cdot \rangle_N$.

Proposition 6 (Characterisation of the solutions of $(\mathcal{P}_N)$). *Let $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot)) \in \mathrm{Lip}([0,T], (\mathbb{R}^d)^N)) \times \mathcal{U}_N$ be an optimal trajectory-control pair for $(\mathcal{P}_N)$. Then, there exists a rescaled covector $\boldsymbol{r}_N^*(\cdot) \in \mathrm{Lip}([0,T], (\mathbb{R}^d)^N)$ such that $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot))$ satisfies the* mean-field Pontryagin Maximum Principle

$$\begin{cases} \dot{\boldsymbol{x}}_N^*(t) = \mathbf{Grad}_{\boldsymbol{r}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)), & \boldsymbol{x}_N^*(0) = \boldsymbol{x}_N^0, \\ \dot{\boldsymbol{r}}_N^*(t) = -\mathbf{Grad}_{\boldsymbol{x}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)), & \boldsymbol{r}_N^*(T) = -\mathbf{Grad}_{\boldsymbol{x}}\, \varphi_N(\boldsymbol{x}_N^*(T)), \\ \boldsymbol{u}_N^*(t) \in \underset{\boldsymbol{v} \in U^N}{\mathrm{argmax}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{v}) & for\ \mathscr{L}^1\text{-almost every } t \in [0,T], \end{cases} \tag{40}$$

where the mean-field Hamiltonian *of the system is defined by*

$$\mathbb{H}_N(t, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{u}) := \frac{1}{N} \sum_{i=1}^{N} \Big(\langle r_i, \boldsymbol{v}_N(t, \boldsymbol{x}, x_i) + u_i \rangle - \psi(u_i) \Big) - \boldsymbol{L}_N(t, \boldsymbol{x}) \tag{41}$$

for all $(t, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{u}) \in [0, T] \times (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N \times U^N$. *Furthermore, there exist uniform constants* $R_T, L_T > 0$ *which are independent of* $N \geq 1$, *such that*

$$\mathrm{Graph}\big((\boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot))\big) \subset [0, T] \times B(0, R_T)^{2N} \qquad and \qquad \mathrm{Lip}\Big((\boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot)); [0, T]\Big) \leq L_T. \tag{42}$$

Proof. By hypothesis **(H)**-*(i)*, there exists a constant $R_U > 0$ such that $U \subset B(0, R_U)$. Together with the definition (38) of the approximating sequences and **(H)**-*(iii)*, this implies

$$|x_i^*(t)| \leq |x_i^0| + \int_0^t \big|\boldsymbol{v}_N(s, \boldsymbol{x}_N^*(s), x_i^*(s)) + u_i^*(t)\big| \mathrm{d}s \leq |x_i^0| + \int_0^t M\Big(1 + |x_i^*(t)| + \tfrac{1}{N}\textstyle\sum_{j=1}^N |x_j^*(t)|\Big)\mathrm{d}s + R_U T, \tag{43}$$

for all times $t \in [0, T]$. By summing over the indices $i \in \{1, \dots, N\}$ and applying Grönwall's Lemma, there exists a constant $A_T > 0$ independent of $N \geq 1$ such that

$$\max_{t \in [0,T]} \tfrac{1}{N} \textstyle\sum_{i=1}^N |x_i^*(t)| \leq A_T. \tag{44}$$

Plugging (44) into (43) and applying Gröwall's Lemma yet another time, we recover the existence of two constants $R_T^1, L_T^1 > 0$ independent of $N \geq 1$ such that for every index $i \in \{1, \dots, N\}$, it holds

$$\max_{t \in [0,T]} |x_i^*(t)| \leq R_T^1 \qquad and \qquad \mathrm{Lip}(x_i^*(\cdot); [0, T]) \leq L_T^1. \tag{45}$$

As a consequence of the standard PMP applied to $(\mathcal{P}_N)$ (see for instance [27, Theorem 22.2]), there exists a family of costate curves $\boldsymbol{p}^* : t \in [0, T] \mapsto (p_1^*(t), \dots, p_N^*(t)) \in (\mathbb{R}^d)^N$ such that

$$\begin{cases} \dot{x}_i^*(t) = \nabla_{p_i} \mathcal{H}_N(t, \boldsymbol{x}^*(t), \boldsymbol{p}^*(t), \boldsymbol{u}^*(t)), & x_i^*(0) = x_i^0, \\ \dot{p}_i^*(t) = -\nabla_{x_i} \mathcal{H}_N(t, \boldsymbol{x}^*(t), \boldsymbol{p}^*(t), \boldsymbol{u}^*(t)), & p_i^*(T) = -\nabla_{x_i} \boldsymbol{\varphi}_N(\boldsymbol{x}^*(T)), \\ u_i^*(t) \in \underset{v \in U}{\mathrm{argmax}} \; \big[\langle p_i^*(t), v \rangle - \tfrac{1}{N} \psi(v)\big], \end{cases} \tag{46}$$

where the classical Hamiltonian of the system is defined as

$$\mathcal{H}_N(t, \boldsymbol{x}, \boldsymbol{p}, \boldsymbol{u}) = \sum_{i=1}^N \langle p_i, \boldsymbol{v}_N(t, \boldsymbol{x}, x_i) + u_i \rangle - \frac{1}{N} \sum_{i=1}^N \psi(u_i) - \boldsymbol{L}_N(t, \boldsymbol{x}),$$

for every $(t, \boldsymbol{x}, \boldsymbol{p}, \boldsymbol{u}) \in [0, T] \times (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N \times U^N$. Introducing the rescaled curves $r_i^*(\cdot) := N p_i^*(\cdot)$, one has

$$\dot{x}_i^*(t) = N \nabla_{r_i} \mathbb{H}_N(t, \boldsymbol{x}^*(t), \boldsymbol{r}^*(t), \boldsymbol{u}^*(t)) = \mathbf{Grad}_{r_i} \mathbb{H}_N(t, \boldsymbol{x}^*(t), \boldsymbol{r}^*(t), \boldsymbol{u}^*(t)), \tag{47}$$

$$\dot{r}_i^*(t) = -N \nabla_{x_i} \mathbb{H}_N(t, \boldsymbol{x}^*(t), \boldsymbol{r}^*(t), \boldsymbol{u}^*(t)) = -\mathbf{Grad}_{x_i} \mathbb{H}_N(t, \boldsymbol{x}^*(t), \boldsymbol{r}^*(t), \boldsymbol{u}^*(t)), \tag{48}$$

$$r_i^*(T) = -N \nabla_{x_i} \boldsymbol{\varphi}(\boldsymbol{x}^*(T)) = -\mathbf{Grad}_{x_i} \boldsymbol{\varphi}(\boldsymbol{x}^*(T)), \tag{49}$$

where we used the definition of the mean-field gradient $\mathbf{Grad}(\bullet)$ given in Proposition 4. Moreover, in this setting, the maximisation condition in (46) can be rewritten for $\mathscr{L}^1$-almost every $t \in [0, T]$ as

$$u_i^*(t) \in \mathrm{argmax}_{v \in U} \big[\langle r_i^*(t), v \rangle - \psi(v)\big].$$

Merging this condition with (47), (48) and (49), we recover that $(\boldsymbol{x}^*(\cdot), \boldsymbol{r}^*(\cdot), \boldsymbol{u}^*(\cdot))$ satisfies the mean-field Pontryagin Maximum Principle (40) associated with the mean-field Hamiltonian $\mathbb{H}_N(\cdot, \cdot, \cdot, \cdot)$.

We now prove an estimate akin to (45) for the costate variable $(\boldsymbol{r}_N^*(\cdot))$. Observe that, as a consequence of the uniform bounds of (45) and Proposition 4, it holds for all times $t \in [0, T]$ and any $i, j \in \{1, \dots, N\}$ that

$$|\mathbf{Grad}_{x_i} \boldsymbol{L}_N(t, \boldsymbol{x}^*(t))| = |\nabla_\mu L(t, \mu[\boldsymbol{x}_N^*(t)])(x_i^*(t))| \leq \max_{\mu \in \mathscr{P}(B(0, R_T^1))} \|\nabla_\mu L(t, \mu)(\cdot)\|_{C^0(B(0, R_T^1), \mathbb{R}^d)},$$

and

$$|\mathrm{D}_x \boldsymbol{v}_N(t, \boldsymbol{x}^*(t), x_i^*(t))| + |\mathbf{D}_{x_j} \boldsymbol{v}_N(t, \boldsymbol{x}^*(t), x_i^*(t))| \leq \max_{\mu \in \mathscr{P}(B(0, R_T^1))} \|\mathrm{D}_x v(t, \mu, \cdot) + \mathrm{D}_\mu v(t, \mu, \cdot)(\cdot)\|_{C^0(B(0, R_T^1)^2, \mathbb{R}^{d \times d})}.$$

15

By invoking the $\mathcal{C}^{2,1}_{\mathrm{loc}}$-Wasserstein regularity assumptions $\mathbf{(H)}$-$(iv),(v)$ and by Grönwall's Lemma, we obtain

$$\max_{t\in[0,T]}|r_i^*(t)| \leq C'\Big(T + |\mathbf{Grad}_{x_i}\,\boldsymbol{\varphi}_N(\boldsymbol{x}_N^*(T))|\Big)e^{C'T} \tag{50}$$

for all $i \in \{1,\ldots,N\}$, where $C' > 0$ is independent of $N \geq 1$. Again as a consequence of Proposition 4, it holds

$$|\mathbf{Grad}_{x_i}\,\boldsymbol{\varphi}(\boldsymbol{x}_N^*(T))| = |\nabla_\mu\varphi(\mu[\boldsymbol{x}_N^*(T)])(x_i^*(T))| \leq \max_{\mu\in\mathscr{P}(B(0,R_T^1))}\|\nabla_\mu\varphi(\mu)(\cdot)\|_{C^0(B(0,R_T^1),\mathbb{R}^d)}\,,$$

which is uniformly bounded by hypothesis $\mathbf{(H)}$-(vi), so that

$$\max_{t\in[0,T]}|r_i^*(t)| \leq R_T^2 \qquad \text{and} \qquad \mathrm{Lip}(r_i^*(\cdot);[0,T]) \leq L_T^2, \tag{51}$$

for all $i \in \{1,\ldots,N\}$ and some uniform constants $R_T^2, L_T^2 > 0$. Thus, we have shown that there exist two constants $R_T, L_T > 0$ independent of $N \geq 1$, such that

$$\mathrm{Graph}\Big((\boldsymbol{x}^*(\cdot),\boldsymbol{r}^*(\cdot))\Big) \subset [0,T]\times B(0,R_T)^{2N} \qquad \text{and} \qquad \mathrm{Lip}\Big((\boldsymbol{x}^*(\cdot),\boldsymbol{r}^*(\cdot))\,;[0,T]\Big) \leq L_T.$$

This concludes the proof of Proposition 6. $\qquad\qquad\square$

We end the first step of our proof by a simple corollary in which we provide a common Lipschitz constant for all the maps involved in $(\mathcal{P}_N)$ that is uniform with respect to $N \geq 1$.

Corollary 1. *Let $\mathcal{K} := [0,T]\times B(0,R_T)^{2N}\times U^N$ where $R_T > 0$ is defined as in Proposition 6. Then, there exists a constant $\mathcal{L}_\mathcal{K} > 0$ such that*

$$t \mapsto \mathbb{H}_N(t,\boldsymbol{x},\boldsymbol{r},\boldsymbol{u}) \qquad and \qquad t \mapsto \boldsymbol{L}_N(t,\boldsymbol{x}),$$

are bounded by $\mathcal{L}_\mathcal{K}$ and $\mathcal{L}_\mathcal{K}$-Lipschitz over $[0,T]$ uniformly with respect to $(\boldsymbol{x},\boldsymbol{r},\boldsymbol{u}) \in B(0,R_T)^{2N}\times U^N$, and such that the $C_N^{2,1}$-norms defined in the sense of (19) of the maps

$$(\boldsymbol{x},\boldsymbol{u}) \mapsto \mathbb{H}_N(t,\boldsymbol{x},\boldsymbol{r},\boldsymbol{u}), \qquad \boldsymbol{x} \mapsto \boldsymbol{L}_N(t,\boldsymbol{x}), \qquad \boldsymbol{u} \mapsto \tfrac{1}{N}\sum_{i=1}^N\psi(u_i) \qquad and \qquad \boldsymbol{x} \mapsto \boldsymbol{\varphi}_N(\boldsymbol{x}),$$

are bounded by $\mathcal{L}_\mathcal{K}$ over $B(0,R_T)^N\times U^N$, uniformly with respect to $(t,\boldsymbol{r}) \in [0,T]\times B(0,R_T)^N$.

Proof. This result follows directly from the the Lipschitz regularity $\mathbf{(H)}$-(iii) of the velocity field and the $\mathcal{C}^{2,1}_{\mathrm{loc}}$-Wasserstein regularity hypotheses $\mathbf{(H)}$-$(iv),(v),(vi)$, along with the estimate (20) of Proposition 4. $\qquad\square$

Step 2 : Construction of Lipschitz-in-space optimal controls for $(\mathcal{P}_N)$. In this second step, we associate to any optimal pair $(\boldsymbol{x}_N^*(\cdot),\boldsymbol{u}_N^*(\cdot))$ for $(\mathcal{P}_N)$ a mean-field optimal control map $u_N^* \in \mathrm{Lip}([0,T]\times\mathbb{R}^d,U))$, which Lipschitz constant with respect to the space variable is uniformly bounded with respect to $N \geq 1$.

Proposition 7 (Existence of locally optimal uniformly-Lipschitz feedbacks for $(\mathcal{P}_N)$). *Assume that hypotheses $\mathbf{(H)}$ hold and let $(\boldsymbol{x}_N^*(\cdot),\boldsymbol{r}_N^*(\cdot),\boldsymbol{u}_N^*(\cdot)) \in \mathrm{Lip}([0,T],B(0,R_T)^N)\times\mathcal{U}_N$ be an optimal Pontryagin triple for $(\mathcal{P}_N)$ in the sense of Proposition 6 along which the mean-field coercivity estimate $\mathbf{(CO}_N)$ holds.*

Then, for any $N \geq 1$, there exists a Lipschitz map $u_N^*(\cdot,\cdot) \in \mathrm{Lip}([0,T]\times\mathbb{R}^d,U)$ such that

$$u_N^*(t,x_i(t)) = u_i^*(t) \qquad and \qquad \mathrm{Lip}(u_N^*(t,\cdot)\,;\mathbb{R}^d) \leq \mathcal{L}_U,$$

for all times $t \in [0,T]$, where $\mathcal{L}_U > 0$ is independent of $N \geq 1$.

Proof. Recall that first that by Corollary 1, the bounded-Lipschitz norms in $t \in [0,T]$ and the $C_N^{2,1}$-norms in $(\boldsymbol{x},\boldsymbol{u}) \in B(0,R_T)^N\times U^N$ of the datum of $(\mathcal{P}_N)$ are uniformly bounded over $\mathcal{K} = [0,T]\times B(0,R_T)^{2N}\times U^N$ by a constant $\mathcal{L}_\mathcal{K} > 0$. As mentioned in Section 3, Theorem 4 can be applied in $((\mathbb{R}^d)^N,\langle\cdot,\cdot\rangle_N)$ provided that $\mathbf{(CO}_N)$ is indeed a strong positive-definiteness condition for the canonical linearised problem associated to $(\mathcal{P}_N)$. To verify this, consider $(\boldsymbol{y}(\cdot),\boldsymbol{s}(\cdot),\boldsymbol{w}(\cdot)) \in W^{1,2}([0,T],(\mathbb{R}^d)^N)\times W^{1,2}([0,T],(\mathbb{R}^d)^N)\times L^2(([0,T],U^N)$ such that $\boldsymbol{u}_N^*(t) + \boldsymbol{w}(t) \in U^N$ for $\mathscr{L}^1$-almost every $t \in [0,T]$ and $i \in \{1,\ldots,N\}$. Then, it holds

$$\boldsymbol{v}_N(t,\boldsymbol{x}_N^*(t)+\boldsymbol{y}(t),x_i^*(t)+y_i(t)) = \boldsymbol{v}_N(t,\boldsymbol{x}_N^*(t),x_i^*(t)) + \mathrm{D}_x\boldsymbol{v}_N\big(t,\boldsymbol{x}_N^*(t),x_i^*(t)\big)y_i(t)$$

$$+ \tfrac{1}{N}\sum_{j=1}^N\mathbf{D}_{x_j}\boldsymbol{v}_N(t,\boldsymbol{x}_N^*(t),x_i^*(t))y_j(t) + o(|y_i(t)|) + o(|\boldsymbol{y}(t)|_N), \tag{52}$$

16

for all times $t \in [0, T]$, where $\mathbf{D}_{x_j} \boldsymbol{v}_N(t, \boldsymbol{x}, x_i)$ is the matrix whose rows are the mean-field gradients with respect to x_j of the components $\boldsymbol{x} \mapsto \boldsymbol{v}_N^k(t, \boldsymbol{x}, x_i)$ for $k \in \{1, \dots, d\} \in \mathbb{R}^d$. Analogously, one also has[1]

$$
\begin{aligned}
\mathbf{Grad}_{\boldsymbol{x}}\, & \mathbb{H}_N(t, \boldsymbol{x}_N^*(t) + \boldsymbol{y}(t), \boldsymbol{r}_N^*(t) + \boldsymbol{s}(t), \boldsymbol{u}_N^*(t) + \boldsymbol{w}(t)) \\
&= \mathbf{Grad}_{\boldsymbol{x}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)) + \mathbf{Hess}_{\boldsymbol{x}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t))\boldsymbol{y}(t) \\
&\quad + \mathbf{Hess}_{\boldsymbol{r}\boldsymbol{x}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t))\boldsymbol{s}(t) + o(|\boldsymbol{y}(t)|_N) + o(|\boldsymbol{w}(t)|_N),
\end{aligned} \tag{53}
$$

$$
\begin{aligned}
\mathbf{Grad}_{\boldsymbol{u}}\, & \mathbb{H}_N(t, \boldsymbol{x}_N^*(t) + \boldsymbol{y}(t), \boldsymbol{r}_N^*(t) + \boldsymbol{s}(t), \boldsymbol{u}_N^*(t) + \boldsymbol{w}(t)) \\
&= \mathbf{Grad}_{\boldsymbol{u}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)) + \mathbf{Hess}_{\boldsymbol{u}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t))\boldsymbol{w}(t) \\
&\quad + \mathbf{Hess}_{\boldsymbol{r}\boldsymbol{u}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t))\boldsymbol{s}(t) + o(|\boldsymbol{s}(t)|_N) + o(|\boldsymbol{w}(t)|_N),
\end{aligned} \tag{54}
$$

$$
\mathbf{Grad}_{\boldsymbol{x}}\, \boldsymbol{\varphi}_N(\boldsymbol{x}_N^*(T) + \boldsymbol{y}(T)) = \mathbf{Grad}_{\boldsymbol{x}}\, \boldsymbol{\varphi}(\boldsymbol{x}_N^*(T)) + \mathbf{Hess}_{\boldsymbol{x}}\, \boldsymbol{\varphi}_N(\boldsymbol{x}_N^*(T))\boldsymbol{y}(T) + o(|\boldsymbol{y}(T)|_N), \tag{55}
$$

as a consequence of the chain rule of Proposition 4. Following [32], it can be checked that the first-order linearisation of the optimality system (40) obtain by combining (52), (53), (54) and (55) is the PMP of

$$
(\mathcal{P}_N') \quad \begin{cases}
\displaystyle \min_{\boldsymbol{w}(\cdot) \in \mathcal{U}_N'} \left[\int_0^T \left(\tfrac{1}{2}\langle \boldsymbol{A}(t)\boldsymbol{y}(t), \boldsymbol{y}(t)\rangle_N + \tfrac{1}{2}\langle \boldsymbol{B}(t)\boldsymbol{w}(t), \boldsymbol{w}(t)\rangle_N \right) \mathrm{d}t + \tfrac{1}{2}\langle \boldsymbol{C}(T)\boldsymbol{y}(T), \boldsymbol{y}(T)\rangle_N \right] \\[2mm]
\text{s.t.} \begin{cases}
\dot{y}_i(t) = \mathrm{D}_x \boldsymbol{v}_N(t, \boldsymbol{x}_N^*(t), x_i^*(t))y_i(t) + \dfrac{1}{N}\sum_{j=1}^N \mathbf{D}_{x_j} \boldsymbol{v}_N(t, \boldsymbol{x}_N^*(t), x_i^*(t))y_j(t), \\[3mm]
y_i(0) = 0,
\end{cases}
\end{cases}
$$

where the set of admissible controls is defined by

$$
\mathcal{U}_N' = \left\{ v \in L^\infty([0, T], U^N) \text{ s.t. } \boldsymbol{u}_N^*(t) + \boldsymbol{w}(t) \in U^N \text{ for } \mathscr{L}^1\text{-almost every } t \in [0, T] \right\},
$$

and the matrices defining the cost functionals write

$$
\begin{cases}
\boldsymbol{A}(t) = -\mathbf{Hess}_{\boldsymbol{x}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)), \quad \boldsymbol{C}(T) = \mathbf{Hess}_{\boldsymbol{x}}\, \boldsymbol{\varphi}_N(\boldsymbol{x}_N^*(T)), \\
\boldsymbol{B}(t) = -\mathbf{Hess}_{\boldsymbol{u}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)),
\end{cases}
$$

for all times $t \in [0, T]$. Thus, the coercivity estimate $(\mathbf{CO}_N)$ is indeed a strong positive-definiteness condition for $(\mathcal{P}_N')$ expressed in terms of the differential structure of $((\mathbb{R}^d)^N, \langle \cdot, \cdot \rangle_N)$. Hence, by Theorem 4 applied to $(\mathcal{P}_N)$, there exists a neighbourhood $\mathscr{N} \subset [0, T] \times (\mathbb{R}^d)^N$ of $\mathrm{Graph}(\boldsymbol{x}^*(\cdot))$ and a locally optimal feedback

$$
\tilde{\boldsymbol{u}}_N : \left([0, T] \times B(0, R_T)^N \right) \cap \mathscr{N} \to U^N, \tag{56}
$$

such that $\tilde{\boldsymbol{u}}_N(t, \boldsymbol{x}^*(t)) = \boldsymbol{u}_N^*(t)$ for all times $t \in [0, T]$ and

$$
\left| \tilde{\boldsymbol{u}}_N(t, \boldsymbol{x}) - \tilde{\boldsymbol{u}}_N(s, \boldsymbol{y}) \right|_N \leq \mathcal{L}_U' \big(|t - s| + |\boldsymbol{x} - \boldsymbol{y}|_N \big), \tag{57}
$$

for any $(t, \boldsymbol{x}), (s, \boldsymbol{y}) \in \mathscr{N}$, where $\mathcal{L}_U' > 0$ depends only on the structural constant $\mathcal{L}_{\mathcal{K}}$ introduced in Corollary 1 and on the coercivity constant ρ_T exhibited in $(\mathbf{CO}_N)$. In particular, $\mathcal{L}_U'$ is independent of $N \geq 1$.

For any $i \in \{1, \dots, N\}$, we can in turn associate to each agent trajectory $x_i^*(\cdot)$ the projected control map

$$
\tilde{u}_i : (t, x) \in \mathcal{N}_i \mapsto \tilde{\boldsymbol{u}}_N^i(t, \hat{\boldsymbol{x}}_i^x(t)),
$$

for any $(t, x) \in \mathcal{N}_i$, where we introduced the notation

$$
\hat{\boldsymbol{x}}_i^x(t) := (x_1^*(t), \dots, x_{i-1}^*(t), x, x_{i+1}^*(t), \dots, x_N^*(t)), \tag{58}
$$

and where the agent-based neighbourhoods $\mathcal{N}_i \subset [0, T] \times \mathbb{R}^d$ are defined by

$$
\mathcal{N}_i := \left\{ (t, x) \in [0, T] \times \mathbb{R}^d \text{ s.t. } \hat{\boldsymbol{x}}_i^x(t) \in \mathscr{N} \right\}.
$$

These sets are well-defined and non-empty, since the projection operations onto coordinates are open mappings. Moreover, for any $t \in [0, T]$ and $x, y \in \mathbb{R}^d$ such that $(t, x), (t, y) \in \mathcal{N}_i$, it holds

$$
\begin{aligned}
|\tilde{u}_i(t, x) - \tilde{u}_i(t, y)| &= \left| \tilde{\boldsymbol{u}}_N^i(t, \hat{\boldsymbol{x}}_i^x(t)) - \tilde{\boldsymbol{u}}_N^i(t, \hat{\boldsymbol{x}}_i^y(t)) \right| \\
&\leq \left(\sum_{j=1}^N |\tilde{\boldsymbol{u}}_N^j(t, \hat{\boldsymbol{x}}_i^x(t)) - \tilde{\boldsymbol{u}}_N^j(t, \hat{\boldsymbol{x}}_i^y(t))|^2 \right)^{1/2} \\
&= \sqrt{N}\, \left| \tilde{\boldsymbol{u}}_N(t, \hat{\boldsymbol{x}}_i^x(t)) - \tilde{\boldsymbol{u}}_N(t, \hat{\boldsymbol{x}}_i^y(t)) \right|_N \leq \sqrt{N}\mathcal{L}_U' |\hat{\boldsymbol{x}}_i^x(t) - \hat{\boldsymbol{x}}_i^y(t)|_N,
\end{aligned} \tag{59}
$$

---

[1] Here for convenience, we use the matrix representation (30) introduced in Remark 2 for mean-field Hessians.

as a consequence of (57). Observe now that by (58), the quantity $|\hat{\boldsymbol{x}}_i^y(t) - \hat{\boldsymbol{x}}_i^x(t)|_N$ can be further estimated as

$$|\hat{\boldsymbol{x}}_i^x(t) - \hat{\boldsymbol{x}}_i^y(t)|_N = \left(\tfrac{1}{N} \sum_{j=1}^N \big| (\hat{\boldsymbol{x}}_i^x(t))_j - (\hat{\boldsymbol{x}}_i^y(t))_j \big|^2 \right)^{1/2} = \tfrac{1}{\sqrt{N}} |y - x|, \tag{60}$$

for all $t \in [0,T]$, since $(\hat{\boldsymbol{x}}_i^x(t))_j = (\hat{\boldsymbol{x}}_i^y(t))_j = x_j^*(t)$ for any $j \neq i$. By merging (59) and (60), we recover that the maps $\tilde{u}_i(\cdot,\cdot)$ defined in (56) are $\mathcal{L}_U'$-Lipschitz in space over $\mathcal{N}_i$ for any $i \in \{1, \dots, N\}$.

To conclude the proof of Proposition 7, there remains to "patch together" the locally optimal agent feedbacks $\tilde{u}_i(\cdot,\cdot)$ defined above. First, observe that since the maps $x \mapsto \tilde{u}_i(t,x) \in U$ are Lipschitz for any $i \in \{1, \dots, N\}$, all the individual agent trajectories are solution of the well-posed Cauchy-Lipschitz ODEs

$$\dot{x}_i^*(t) = \boldsymbol{v}_N(t, \boldsymbol{x}_N^*(t), x_i^*(t)) + \tilde{u}_i(t, x_i^*(t)),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$. Besides, if $x_j^*(\tau) \in \mathcal{N}_i$ for some time $\tau \in [0,T]$ with $j \neq i$, then the fact that $\tilde{u}_i(\cdot,\cdot)$ is a locally optimal feedback necessarily implies that $u_j^*(t) = \tilde{u}_i(t, x_j^*(t))$ for all times $t \in [\tau, T]$ such that $x_j^*(t) \in \mathcal{N}_i$. Thus, no finite-time collisions can occur between agents, so that the sets $\mathcal{N}_i$ can be chosen to be disjoint and the map

$$u_N^* : (t,x) \in \bigcup_{i=1}^N \overline{\mathcal{N}_i} \mapsto \tilde{u}_i(t,x) \in U \quad \text{whenever } (t,x) \in \overline{\mathcal{N}_i},$$

is well-defined. By using McShane's Extension Theorem (see e.g. [36, Theorem 3.1]) combined with a projection on the convex and compact set $U \subset \mathbb{R}^d$, one can define a global optimal control map $u_N^* : [0,T] \times \mathbb{R}^d \to U$ such that $u_N^*(t, x_i^*(t)) = u_i^*(t)$ for all $t \in [0,T]$ and

$$\text{Lip}(u_N^*(t, \cdot); \mathbb{R}^d) \leq \mathcal{L}_U,$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$, where the new Lipschitz constant is $\mathcal{L}_U := \sqrt{d} \mathcal{L}_U'$. $\qquad\square$

Step 3 : Existence of Lipschitz optimal controls for problem $(\mathcal{P})$**.** In this third step, we show that the sequence of optimal maps $(u_N^*(\cdot,\cdot))$ constructed in Proposition 7 is compact in a suitable topology and that the limits along subsequences are optimal solutions of problem $(\mathcal{P})$.

Lemma 4 (Compactness of Lipschitz-in-space optimal maps). *Let $\mathcal{L}_U > 0$ be a positive constant and $\Omega \subset \mathbb{R}^d$ be a bounded set. Then, the set*

$$\mathcal{U}_{\mathcal{L}_U} = \left\{ u(\cdot,\cdot) \in L^2([0,T], W^{1,\infty}(\Omega, U)) \ s.t. \ \sup_{t \in [0,T]} \|u^*(t,\cdot)\|_{W^{1,\infty}(\Omega, \mathbb{R}^d)} \leq \mathcal{L}_U \right\},$$

is compact in the weak $L^2([0,T], W^{1,p}(\Omega, \mathbb{R}^d))$-topology for any $p \in (1, +\infty)$.

Proof. See e.g. [40, Theorem 2.5]. $\qquad\square$

This allows to derive the following convergence result on the sequence of controls $(u_N^*(\cdot,\cdot))$ built in Step 2.

Corollary 2 (Convergence of Lipschitz optimal control). *There exists a map $u^*(\cdot,\cdot) \in L^2([0,T], W^{1,\infty}(\mathbb{R}^d, U))$ such that the sequence of Lipschitz optimal controls $(u_N^*(\cdot,\cdot))$ defined in Proposition 7 converges up to a subsequence towards $u^*(\cdot,\cdot)$ in the weak $L^2([0,T], W^{1,p}(\Omega, \mathbb{R}^d))$-topology for any $p \in (1, +\infty)$.*

Proof. This result comes from a direct application of Lemma 4 to the sequence of optimal maps built in Proposition 7 up to choosing $\Omega := B(0, R_T)$ and redefining $\mathcal{L}_U := \max\{L_U, \mathcal{L}_U\}$. $\qquad\square$

We now prove that the generalised optimal control $\boldsymbol{\nu}^* \in \mathcal{U}$ for problem $(\mathcal{P}_{\text{meas}})$ is induced by the Lipschitz-in-space optimal control $u^*(\cdot,\cdot)$ defined in Corollary 2. By construction, it holds for any $N \geq 1$ that

$$\boldsymbol{\nu}_N^* := \int_{[0,T]} \left(\frac{1}{N} \sum_{i=1}^N u_i^*(t) \delta_{x_i^*(t)} \right) \mathrm{d}\lambda(t) = \int_{[0,T]} \left(\frac{1}{N} \sum_{i=1}^N u_N^*(t, x_i^*(t)) \delta_{x_i^*(t)} \right) \mathrm{d}\lambda(t) = u_N^*(\cdot,\cdot) \tilde{\mu}_N^*,$$

where $\boldsymbol{\nu}_N^* \in \mathcal{U}$ denotes the generalised empirical control measure introduced in Theorem 6. In the following proposition, we prove that the sequence $(u_N^*(\cdot,\cdot) \tilde{\mu}_N)$ converges weakly-* towards $u^*(\cdot,\cdot) \tilde{\mu}^*$.

Lemma 5 (Convergence of generalised Lipschitz optimal controls). *Let $R_T > 0$ be given by Proposition 6, $\Omega := B(0, R_T)$ and $(\mu_N^*(\cdot)) \subset \text{Lip}([0,T], \mathscr{P}_1(\Omega))$ be the sequence of optimal measure curves associated with $(\mathcal{P}_N)$. Let $(u_N^*(\cdot,\cdot)) \subset L^2([0,T], W^{1,\infty}(\Omega, U))$ be as in Proposition 7 and $u^*(\cdot,\cdot)$ be one of its cluster points in the weak $L^2([0,T], W^{1,p}(\Omega, U))$-topology for some $p \in (d, +\infty)$. Then, $(\boldsymbol{\nu}_N^*) := (u_N^*(\cdot,\cdot) \tilde{\mu}_N^*)$ converges towards $\boldsymbol{\nu}^* = u^*(\cdot,\cdot) \tilde{\mu}^*$ in the weak-* topology of $\mathcal{M}([0,T] \times \Omega, U)$.*

Proof. Recall first that the topological dual of the Banach space $L^2([0,T], W^{1,p}(\Omega, U))$ can be identified with $L^2([0,T], W^{-1,p'}(\Omega, U))$, where p' is the conjugate exponent of p. Hence, the fact that $u_N(\cdot, \cdot) \rightharpoonup u(\cdot, \cdot)$ in $L^2([0,T], W^{1,p}(\Omega, U))$ as $N \to +\infty$ can be reformulated as

$$\int_0^T \langle \xi(t), u_N^*(t, \cdot) \rangle_{W^{1,p}(\Omega, U)} \mathrm{d}t \underset{N \to +\infty}{\longrightarrow} \int_0^T \langle \xi(t), u^*(t, \cdot) \rangle_{W^{1,p}(\Omega, U)} \mathrm{d}t, \tag{61}$$

for any $\xi \in L^2([0,T], W^{-1,p'}(\Omega, \mathbb{R}^d))$, where $\langle \cdot, \cdot \rangle_{W^{1,p}(\Omega, U)}$ denotes the duality bracket of $W^{1,p}(\Omega, U)$.

Since we assumed that $p \in (d, +\infty)$, it holds by Morrey's Embedding (see e.g. [12, Theorem 9.12]) that $W^{1,p}(\Omega, U) \subset C^0(\Omega, U)$. By taking the topological dual of this inclusion, we obtain that $\mathcal{M}(\Omega, U) \subset W^{-1,p'}(\Omega, U)$. This relation, combined with (4) and (61), yields

$$\int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u_N^*(t, x) \rangle \mathrm{d}\sigma(t)(x) \mathrm{d}t \underset{N \to +\infty}{\longrightarrow} \int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u^*(t, x) \rangle \mathrm{d}\sigma(t)(x) \mathrm{d}t, \tag{62}$$

for any curve $\sigma(\cdot) \in C^0([0,T], \mathcal{M}(\Omega, \mathbb{R}_+))$ and any $\xi \in C_c^1([0,T] \times \Omega, \mathbb{R}^d)$. Moreover, for each $N \geq 1$ it holds

$$\left| \int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u^*(t, x) \rangle \mathrm{d}\mu^*(t)(x) \mathrm{d}t - \int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u_N^*(t, x) \rangle \mathrm{d}\mu_N^*(t)(x) \mathrm{d}t \right|$$
$$\leq \left| \int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u^*(t, x) - u_N^*(t, x) \rangle \mathrm{d}\mu^*(t)(x) \mathrm{d}t \right| + \left| \int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u_N^*(t, x) \rangle \mathrm{d}(\mu^*(t) - \mu_N^*(t))(x) \mathrm{d}t \right|. \tag{63}$$

The first term in the right-hand side of (63) vanishes as $N \to +\infty$ as a consequence of (62). By invoking Kantorovich-Rubinstein duality formula (5) along with the $\mathcal{L}_U$-Lipschitz regularity of the maps $x \in \mathbb{R}^d \mapsto u_N^*(t, x) \in U$, we obtain the following upper bound on the second term in the right-hand side of (63)

$$\left| \int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u_N^*(t, x) \rangle \mathrm{d}(\mu^*(t) - \mu_N^*(t))(x) \mathrm{d}t \right| \leq C_\xi \max_{t \in [0,T]} W_1(\mu_N^*(t), \mu^*(t)) \underset{N \to +\infty}{\longrightarrow} 0,$$

where $C_\xi := \mathcal{L}_U \max_{t \in [0,T]} \left(\|\xi(t, \cdot)\|_{C^0(\Omega)} + \mathrm{Lip}(\xi(t, \cdot); \Omega) \right)$. Therefore, we recover the convergence result

$$\int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u_N^*(t, x) \rangle \mathrm{d}\mu_N^*(t)(x) \mathrm{d}t \underset{N \to +\infty}{\longrightarrow} \int_0^T \int_{\mathbb{R}^d} \langle \xi(t, x), u^*(t, x) \rangle \mathrm{d}\mu^*(t)(x) \mathrm{d}t, \tag{64}$$

for any $\xi \in C_c^1([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$. Since the measure curves $\mu_N^*(\cdot)$ are uniformly compactly supported in $\Omega \subset \mathbb{R}^d$, one can show that (64) holds for any $\xi \in C_c^0([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ by a classical approximation argument (see e.g. [38]). This precisely amounts to saying that $\boldsymbol{\nu}_N^* \rightharpoonup^* u^*(\cdot, \cdot)\tilde{\mu}^*$ as $N \to +\infty$ along the same subsequence. $\qquad\square$

By uniqueness of the weak-$*$ limit in $\mathcal{M}([0,T] \times \mathbb{R}^d, U)$, we obtain by combining Lemma 5 with Theorem 6 that the optimal solution $\boldsymbol{\nu}^* \in \mathscr{U}$ of $(\mathcal{P}_{\text{meas}})$ is induced by $u^*(\cdot, \cdot)$. Whence the pair $(\mu^*(\cdot), u^*(\cdot, \cdot))$ is a classical optimal pair for $(\mathcal{P})$, which concludes the proof of Theorem 2.

5.2 A sufficient condition for coercivity and proof of Theorem 1

In this section, we prove a simple and general sufficient condition for the coercivity estimate $(\mathbf{CO}_N)$ to hold, and use it to deduce Theorem 1 from Theorem 2.

Proposition 8 (A sufficient condition for mean-field coercivity). *Let $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$ and suppose that hypotheses (**H**) hold. Then, there exists a constant $\lambda_{(\mathcal{P})} \geq 0$ such that, if the control cost $\psi : U \to \mathbb{R}$ is strongly convex with constant $\lambda_\psi > \lambda_{(\mathcal{P})}$, then the coercivity $(\mathbf{CO}_N)$ holds along any optimal mean-field Pontryagin triple $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot))$ with $\rho_T := \lambda_\psi - \lambda_{(\mathcal{P})}$. Moreover, the constant $\lambda_{(\mathcal{P})}$ is intrinsic to $(\mathcal{P})$, in the sense that it only depends on the $\mathcal{C}_{\mathrm{loc}}^2$-Wasserstein norms of the dynamics and cost functionals.*

The main ingredient involved in this result is contained in the following technical lemma, which proof is provided for the sake of completeness.

Lemma 6 (C_N^2-functions are λ-convex on products of convex compact sets). *Let $K \subset \mathbb{R}^d$ be a convex compact set and $\phi : \mathscr{P}_c(\mathbb{R}^d) \to \mathbb{R}$ be $\mathcal{C}_{\mathrm{loc}}^{2,1}$-Wasserstein regular with discrete approximating sequence $(\phi_N(\cdot))$. Then*

$$\mathbf{Hess}\, \phi_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h}) \geq -\max_{\mu \in \mathscr{P}(K)} \left(\|\mathrm{D}_x \nabla_\mu \phi(\mu)(\cdot)\|_{C^0(K, \mathbb{R}^{d \times d})} + \|\mathrm{D}_\mu^2 \phi(\mu)(\cdot, \cdot)\|_{C^0(K \times K, \mathbb{R}^{d \times d})} \right) |\boldsymbol{h}|_N^2,$$

for any $\boldsymbol{x}, \boldsymbol{h} \in K^N$.

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in K^N$ and $t \in [0,1]$. As a consequence of Proposition 4, one can write the following integral Taylor formulas for $\boldsymbol{\phi}_N(\cdot)$

$$
\begin{cases}
\boldsymbol{\phi}_N((1-t)\boldsymbol{x} + t\boldsymbol{y}) = \boldsymbol{\phi}_N(\boldsymbol{x}) + \displaystyle\int_0^1 \langle \mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x} + st(\boldsymbol{y} - \boldsymbol{x})), t(\boldsymbol{y} - \boldsymbol{x}) \rangle_N \mathrm{d}s, \\[4mm]
\boldsymbol{\phi}_N(\boldsymbol{y}) = \boldsymbol{\phi}_N(\boldsymbol{x}) + \displaystyle\int_0^1 \langle \mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x} + s(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x} \rangle_N \mathrm{d}s.
\end{cases}
\tag{65}
$$

Combining the two equations of (65), it then holds

$$
\begin{aligned}
\boldsymbol{\phi}_N((1-t)\boldsymbol{x} + t\boldsymbol{y}) - (1-t)\boldsymbol{\phi}_N(\boldsymbol{x}) - t\boldsymbol{\phi}_N(\boldsymbol{y}) &= \int_0^1 \left\langle \mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x} + st(\boldsymbol{y} - \boldsymbol{x})), t(\boldsymbol{y} - \boldsymbol{x}) \right\rangle_N \mathrm{d}s \\
&\quad - \int_0^1 \left\langle \mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x} + s(\boldsymbol{y} - \boldsymbol{x})), t(\boldsymbol{y} - \boldsymbol{x}) \right\rangle_N \mathrm{d}s \\
&\leq \frac{t(1-t)}{2} \mathrm{Lip}\big(\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot); K^N\big) |\boldsymbol{y} - \boldsymbol{x}|_N^2,
\end{aligned}
\tag{66}
$$

where we used the fact that both $(\boldsymbol{x} + st(\boldsymbol{y} - \boldsymbol{x}))$ and $(\boldsymbol{x} + s(\boldsymbol{y} - \boldsymbol{x}))$ belong to K^N, since this set is convex. Therefore, we have shown that the map $\boldsymbol{\phi}_N(\cdot)$ is λ-convex over K^N with $\lambda = -\mathrm{Lip}(\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot); K^N)$.

Choosing in particular $\boldsymbol{y} = \boldsymbol{x} + s\boldsymbol{h}$ with $s \in (0,1)$ small, the λ-convexity (66) of $\boldsymbol{\phi}_N(\cdot)$ can be expressed as

$$
\boldsymbol{\phi}_N(\boldsymbol{x} + st\boldsymbol{h}) \leq t\big(\boldsymbol{\phi}_N(\boldsymbol{x} + s\boldsymbol{h}) - \boldsymbol{\phi}_N(\boldsymbol{x})\big) + s^2 \frac{t(1-t)}{2} \mathrm{Lip}\big(\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot); K^N\big)|\boldsymbol{h}|_N^2.
\tag{67}
$$

By applying the chain rule (16) to (67), we obtain

$$
(st)^2 \mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h}) \leq s^2 t\, \mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h}) + s^2 t(1-t) \mathrm{Lip}\big(\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot); K^N\big)|\boldsymbol{h}|_N^2 + o((st)^2) + o(s^2 t),
$$

so that dividing by $s^2 t > 0$ and letting $s, t \to 0^+$, we finally recover that

$$
\mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h}) \geq -\mathrm{Lip}\big(\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot); K^N\big)|\boldsymbol{h}|_N^2,
$$

for any $\boldsymbol{x}, \boldsymbol{h} \in K^N$. One can finally check that, as a consequence of (17), it holds

$$
\begin{aligned}
|\mathrm{Lip}(\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot); K^N)| &\leq \max_{\boldsymbol{x} \in K^N} \max_{|\boldsymbol{h}|_N = 1} |\mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h})| \\
&\leq \max_{\mu \in \mathscr{P}(K)} \left(\|\mathrm{D}_x \nabla_\mu \phi(\mu)(\cdot)\|_{C^0(K, \mathbb{R}^{d \times d})} + \|\mathrm{D}_\mu^2 \phi(\mu)(\cdot, \cdot)\|_{C^0(K \times K, \mathbb{R}^{d \times d})} \right),
\end{aligned}
$$

which concludes the proof of our claim, since $\mu[\boldsymbol{x}] \in \mathscr{P}(K)$ for any $\boldsymbol{x} \in K^N$. $\qquad\square$

Proof of Proposition 8. As a consequence of hypotheses **(H)** together with Proposition 4, the partial Hamiltonian $(\boldsymbol{x}, \boldsymbol{u}) \in B(0, R_T)^N \times U^N \mapsto \mathbb{H}_N(t, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{u})$ and the final cost $\boldsymbol{x} \in (\mathbb{R}^d)^N \mapsto \boldsymbol{\varphi}_N(\boldsymbol{x})$ are $C_N^{2,1}$-regular uniformly with respect to $(t, \boldsymbol{r}) \in [0,T] \times B(0, R_T)^N$, with constants $\mathcal{L}_\mathcal{K} > 0$ that only depend on the $\mathcal{C}_{loc}^{2,1}$-Wasserstein norms of the dynamics and cost functionals, where $R_T > 0$ is given by Proposition 6.

By repeating the Grönwall estimates made on the costate variables in the proof of Proposition 6, one can check that the solutions $\boldsymbol{y}(\cdot)$ of the mean-field linearised system described in $(\mathbf{CO}_N)$ are contained in a product of compact sets $K^N \subset (\mathbb{R}^d)^N$. Moreover, they also satisfy the estimate

$$
\max\left\{ |\boldsymbol{y}(T)|_N^2 \,,\, \int_0^T |\boldsymbol{y}(t)|_N^2 \mathrm{d}t \right\} \leq C_T \int_0^T |\boldsymbol{w}(t)|_N^2 \mathrm{d}t,
$$

for a given uniform constant $C_T > 0$. Merging these facts together along with the statement of Lemma 6, there exists an intrinsic constant $\lambda_{(\mathcal{P})} \geq 0$ independent of $N \geq 1$ such that

$$
\mathbf{Hess}_{\boldsymbol{x}}\, \boldsymbol{\varphi}_N[\boldsymbol{x}_N^*(T)](\boldsymbol{y}(T), \boldsymbol{y}(T)) - \int_0^T \mathbf{Hess}_{\boldsymbol{x}}\, \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{y}(t), \boldsymbol{y}(t))\mathrm{d}t \geq -\lambda_{(\mathcal{P})} \int_0^T |\boldsymbol{w}(t)|_N^2 \mathrm{d}t,
\tag{68}
$$

along any linearising pair $(\boldsymbol{y}(\cdot), \boldsymbol{w}(\cdot))$. Observe now that if $\psi(\cdot)$ is λ_ψ-strongly convex, it also holds

$$
\begin{aligned}
-\int_0^T \mathbf{Hess}_{\boldsymbol{u}}\, \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{w}(t), \boldsymbol{w}(t))\mathrm{d}t &= \int_0^T \left(\frac{1}{N} \sum_{i=1}^N \langle \nabla^2 \psi(u_i^*(t)) w_i(t), w_i(t) \rangle \right) \mathrm{d}t \\
&\geq \lambda_\psi \int_0^T |\boldsymbol{w}(t)|_N^2 \mathrm{d}t,
\end{aligned}
\tag{69}
$$

for any map $\boldsymbol{w}(\cdot) \in L^2([0,T], U^N)$. Combining (68) and (69), we obtain the uniform coercivity-type estimate

$$\mathbf{Hess}_{\boldsymbol{x}}\,\varphi_N[\boldsymbol{x}_N^*(T)](\boldsymbol{y}(T), \boldsymbol{y}(T)) - \int_0^T \mathbf{Hess}_{\boldsymbol{x}}\,\mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{y}(t), \boldsymbol{y}(t))\mathrm{d}t$$

$$- \int_0^T \mathbf{Hess}_{\boldsymbol{u}}\,\mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{w}(t), \boldsymbol{w}(t))\mathrm{d}t \geq (\lambda_\psi - \lambda_{(\mathcal{P})}) \int_0^T |\boldsymbol{w}(t)|_N^2 \mathrm{d}t.$$

Therefore, up to choosing a control cost with strong convexity constant $\lambda_\psi > \lambda_{(\mathcal{P})}$, the coercivity estimate $(\mathbf{CO}_N)$ holds along any optimal mean-field Pontryagin triple with $\rho_T := \lambda_\psi - \lambda_{(\mathcal{P})}$. $\qquad\square$

We can now use this intermediate result together with Theorem 2 to prove Theorem 1.

Proof of Theorem 1. By Theorem 6 and under hypotheses **(H)**, one can associate to any sequence of uniformly compactly supported measures $(\mu_N^0) \subset \mathscr{P}_c(\mathbb{R}^d)$ such that $W_1(\mu_N^0, \mu^0) \to 0$ as $N \to +\infty$ a sequence of generalised trajectory-control pairs $(\mu_N^*(\cdot), \boldsymbol{\nu}_N^*(\cdot)) \in \mathrm{Lip}([0,T], \mathscr{P}_1(\mathbb{R}^d)) \times \mathscr{U}$ which converges to an optimal pair for $(\mathcal{P})$.

Moreover, we assumed that $\psi(\cdot)$ is strongly convex with $\lambda_\psi > \lambda_{(\mathcal{P})}$, where $\lambda_{(\mathcal{P})} \geq 0$ is the intrinsic constant introduced in Proposition 8. Thus, the coercivity estimate $(\mathbf{CO}_N)$ holds along any optimal mean-field Pontryagin triple for $(\mathcal{P}_N)$. Thus, by Theorem 2 there exist a constant $\mathcal{L}_U > 0$ together with a mean-field optimal control $u^*(\cdot, \cdot)$ for $(\mathcal{P})$ such that $x \in \mathbb{R}^d \mapsto u^*(t, x) \in U$ is $\mathcal{L}_U$-Lipschitz for $\mathscr{L}^1$-almost every $t \in [0, T]$. $\qquad\square$

6 Sharpness of the coercivity estimate $(\mathbf{CO}_N)$

In this section, we develop an example in which the mean-field coercivity condition $(\mathbf{CO}_N)$ is both necessary and sufficient for the Lipschitz-in-space regularity of optimal controls. With this goal in mind, we consider the mean-field optimal control problem

$$(\mathcal{P}^V) \quad \begin{cases} \displaystyle\min_{u \in \mathcal{U}} \left[\frac{\lambda}{2} \int_0^T \int_{\mathbb{R}} |u(t,x)|^2 \mathrm{d}\mu(t)(x)\mathrm{d}t - \frac{1}{2} \int_{\mathbb{R}} |x - \bar{\mu}(T)|^2\, \mathrm{d}\mu(T)(x) \right] \\[4mm] \text{s.t. } \begin{cases} \partial_t \mu(t) + \nabla \cdot (u(t, \cdot)\mu(t)) = 0, \\[2mm] \mu(0) = \frac{1}{2}\mathbb{1}_{[-1,1]}\mathscr{L}^1. \end{cases} \end{cases}$$

In $(\mathcal{P}^V)$, one aims at maximising the variance at time $T > 0$ of a measure curve $\mu(\cdot)$ starting from the normalised indicator function of $[-1, 1]$, while penalising the running $L^2(\mu(t))$-norm of the control . Here, the set of admissible control values is $U = [-C, C]$ for a positive constant $C > 0$, and the parameter $\lambda > 0$ is the relative weight between the final cost and the control penalisation. It can be verified straightfowardly that this problem satisfies hypotheses **(H)** of Section 4.

Given a sequence of empirical measures $(\mu_N^0) := (\mu[\boldsymbol{x}_N]) \subset \mathscr{P}_N([-1, 1])$ converging in the W_1-metric towards μ^0, we can define the family $(\mathcal{P}_N^V)$ of discretised multi-agent problems associated to $(\mathcal{P}^V)$ as

$$(\mathcal{P}_N^V) \quad \begin{cases} \displaystyle\min_{\boldsymbol{u}(\cdot) \in \mathcal{U}_N} \left[\frac{\lambda}{2N} \sum_{i=1}^N \int_0^T u_i^2(t)\mathrm{d}t - \frac{1}{2N} \sum_{i=1}^N |x_i(T) - \bar{\boldsymbol{x}}(T)|^2 \right] \\[4mm] \text{s.t. } \begin{cases} \dot{x}_i(t) = u_i(t), \\[2mm] x_i(0) = x_i^0. \end{cases} \end{cases}$$

where $\bar{\boldsymbol{x}}(\cdot) = \frac{1}{N}\sum_{i=1}^N x_i(\cdot)$ and $\mathcal{U}_N = L^\infty([0,T], U^N)$. As a consequence of Proposition 5, there exists for any $N \geq 1$ an optimal trajectory-control pair $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot)) \in \mathrm{Lip}([0,T], (\mathbb{R}^d)^N) \times \mathcal{U}_N$ solution of $(\mathcal{P}_N^V)$. The mean-field Hamiltonian associated with $(\mathcal{P}_N^V)$ is given by

$$\mathbb{H}_N : (t, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{u}) \in [0,T] \times (\mathbb{R}^2)^N \times [-C, C]^N \mapsto \frac{1}{N}\sum_{i=1}^N \left(\langle r_i, u_i \rangle - \frac{\lambda}{2}|u_i|^2 \right). \tag{70}$$

By the mean-field PMP of Proposition 6, there exists a covector $\boldsymbol{r}_N^*(\cdot) \in \mathrm{Lip}([0,T], \mathbb{R}^N)$ such that

$$\begin{cases} \dot{r}_i^*(t) = -\mathbf{Grad}_{x_i}\,\mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)) = 0, \\[2mm] r_i^*(T) = \mathbf{Grad}_{x_i}\,\mathrm{Var}_N(\boldsymbol{x}_N^*(T)) = x_i^*(T) - \bar{\boldsymbol{x}}^*(T), \\[2mm] u_i^*(t) \in \displaystyle\operatorname*{argmax}_{v \in U}\,[r_i^*(t)v - \frac{\lambda}{2}v^2]. \end{cases}$$

Therefore, the optimal covector $\boldsymbol{r}_N^*(\cdot)$ is constant and uniquely determined via

$$r_i^*(t) = x_i^*(T) - \bar{\boldsymbol{x}}^*(T),$$

for any $i \in \{1, \ldots, N\}$. As a consequence of the maximisation condition one can express the components of the optimal control $\boldsymbol{u}_N^*(\cdot)$ explicitly as

$$u_i^*(t) = \pi_U\left(\frac{r_i^*(t)}{\lambda}\right) \equiv \pi_{[-C,C]}\left(\frac{x_i^*(T) - \bar{\boldsymbol{x}}^*(T)}{\lambda}\right), \qquad (71)$$

for all $i \in \{1, \ldots, N\}$, where $\pi_K : \mathbb{R} \to U$ is the standard projection onto the closed convex set $U := [-C, C] \subset \mathbb{R}$. It follows directly from this expression that

$$\dot{\bar{\boldsymbol{x}}}^*(t) = \frac{1}{N}\sum_{i=1}^{N} u_i^*(t) = \frac{1}{N}\sum_{i=1}^{N} \pi_{[-C,C]}\left(\frac{x_i^*(T) - \bar{\boldsymbol{x}}^*(T)}{\lambda}\right),$$

for all times $t \in [0, T]$. In the following lemma, we derive a simple and explicit necessary and sufficient condition such that $(\mathbf{CO}_N)$ holds for $(\mathcal{P}^V)$.

Lemma 7 (Charaterisation of the coercivity condition for $(\mathcal{P}^V)$). *The mean-field coercivity condition $(\mathbf{CO}_N)$ holds for $(\mathcal{P}^V)$ if and only if $\lambda > T$. In this case, the optimal coercivity constant is given by $\rho_T = \lambda - T$.*

Proof. We first compute the Hessians involved in the coercivity estimate. For any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{w} \in \mathbb{R}^N$, one has

$$\mathbf{Hess}_{\boldsymbol{x}} \operatorname{Var}_N[\boldsymbol{x}](\boldsymbol{y}, \boldsymbol{y}) = |\boldsymbol{y}|_N^2 - |\bar{\boldsymbol{y}}|^2 \leq |\boldsymbol{y}|_N^2 \qquad \text{and} \qquad \mathbf{Hess}_{\boldsymbol{u}} \mathbb{H}_N[t, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{u}](\boldsymbol{w}, \boldsymbol{w}) = \lambda|\boldsymbol{w}|_N^2.$$

Let $(\boldsymbol{y}(\cdot), \boldsymbol{w}(\cdot)) \in W^{1,2}([0, T], \mathbb{R}^N) \times \in L^2([0, T], U^N)$ be the solution of the linearised Cauchy problem along a given optimal pair $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot))$ for $(\mathcal{P}_N^V)$, which writes

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{w}(t), \qquad \boldsymbol{y}(0) = 0, \qquad (72)$$

with $\boldsymbol{u}_N^*(t) + \boldsymbol{w}(t) \in U^N$. By Cauchy-Schwarz's inequality, one can estimate $|\boldsymbol{y}(T)|_N^2$ as

$$|\boldsymbol{y}(T)|_N^2 = \left|\int_0^T \boldsymbol{w}(t)\mathrm{d}t\right|_N^2 \leq T\int_0^T |\boldsymbol{w}(t)|_N^2\mathrm{d}t,$$

which allows us to recover

$$-\mathbf{Hess}_{\boldsymbol{x}} \operatorname{Var}_N[\boldsymbol{x}_N^*(T)](\boldsymbol{y}(T), \boldsymbol{y}(T)) - \int_0^T \mathbf{Hess}_{\boldsymbol{u}} \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{w}(t), \boldsymbol{w}(t))\mathrm{d}t$$

$$\geq (\lambda - T)\int_0^T |\boldsymbol{w}(t)|_N^2\mathrm{d}t.$$

Thus, the mean-field coercivity condition $(\mathbf{CO}_N)$ holds whenever $\lambda > T$.

Conversely, let us choose a constant admissible control perturbation $\boldsymbol{w}_c(\cdot) := \boldsymbol{w}_c$ such that $\bar{\boldsymbol{w}}_c = 0$. It is always possible to make such a choice, since by (71), there exist at least two indices $i, j \in \{1, \ldots, N\}$ such that $\operatorname{sign}(u_i^*(t)) = -\operatorname{sign}(u_j^*(t))$ for all times $t \in [0, T]$. It is then sufficient to choose $\boldsymbol{w}_c \in [-C, C]^N$ such that

$$\begin{cases} (\boldsymbol{w}_c)_i = -\operatorname{sign}(u_i)\epsilon, \qquad (\boldsymbol{w}_c)_j = -(\boldsymbol{w}_c)_i, \\ (\boldsymbol{w}_c)_k = 0 \ \text{ if } k \in \{1, \ldots, N\} \text{ and } k \neq i, j, \end{cases}$$

where $\epsilon > 0$ is a small parameter. As a consequence of (72), the corresponding state perturbation $\boldsymbol{y}_c(\cdot)$ is such that $\bar{\boldsymbol{y}}_c(\cdot) \equiv 0$. Moreover, it also holds that

$$|\boldsymbol{y}_c(T)|_N^2 = T^2|\boldsymbol{w}_c|_N^2 = T\int_0^T |\boldsymbol{w}_c|_N^2\mathrm{d}t.$$

Therefore, we have shown that this particular linearised trajectory-control pair is such that

$$-\mathbf{Hess}\operatorname{Var}_N[\boldsymbol{x}_N^*(T)](\boldsymbol{y}_c(T), \boldsymbol{y}_c(T))$$

$$-\int_0^T \mathbf{Hess}_{\boldsymbol{u}} \mathbb{H}_N[t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)](\boldsymbol{w}_c(t), \boldsymbol{w}_c(t))\mathrm{d}t = (\lambda - T)\int_0^T |\boldsymbol{w}(t)|_N^2\mathrm{d}t,$$

so that $\rho_T = \lambda - T$ is the sharp mean-field coercivity constant of $(\mathcal{P})$, and $(\mathbf{CO}_N)$ holds if and only if $\lambda > T$. $\square$

Remark 4 (Connection with the sufficient coercivity conditions). *In Lemma 7, we have proven that the sharpest constant depending for $(\mathcal{P}^V)$ which may serve as a sufficient lower-bound for coercivity via Proposition 8 is given by $\lambda_{(\mathcal{P}^V)} := T$. Performing the same computations in the context of a final variance minimisation, our sharp constant would be given by $\lambda_{(\mathcal{P}^V)} := 0$, so that $(\mathbf{CO_N})$ would hold for every $\lambda > 0$.*

We can now use this characterisation of the coercivity condition to show that it is itself equivalent to the uniform Lipschitz regularity in space of the optimal controls. For the sake of computational tractability, we will assume that the initial condition $\boldsymbol{x}^0 = (x_1^0, \ldots, x_N^0)$ is symmetric with respect to the origin and that $\bar{x}^*(\cdot) \equiv 0$.

Proposition 9 (Coercivity and regularity for $(\mathcal{P}^V)$). *The following assertions are equivalent.*

(a) *The mean-field coercivity condition $\lambda > T$ holds.*

(b) *For any sequence of symmetrically distributed empirical measures $(\mu_N^0) \subset \mathscr{P}_N([-1,1])$ converging narrowly towards $\mu^0 = \frac{1}{2}\mathbb{1}_{[-1,1]}\mathscr{L}^1$ with associated discrete optimal pairs $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot))$, it holds*

$$|u_i^*(t) - u_j^*(t)| \leq \frac{1}{\rho_T}|x_i^*(t) - x_j^*(t)|,$$

for all times $t \in [0, T]$, where $\rho_T := \lambda - T$ is the sharp coercivity constant of $(\mathcal{P}^V)$.

Proof. First, suppose that (a) does not hold, i.e. $\lambda \leq T$. Since the optimal controls are constant over $[0, T]$ as a consequence of (71) and we assumed that $\bar{x}(T) = 0$, the total cost of $(\mathcal{P}_N^V)$ can be rewritten as

$$\mathscr{C}(u_1, \ldots, u_N) = \frac{1}{2N}\sum_{i=1}^N \left(T(\lambda - T)u_i^2 - 2Tx_i^0 u_i - |x_i^0|^2 \right),$$

for any N-tuple $\boldsymbol{u} = (u_1, \ldots, u_N) \in [-C, C]^N$. Since $\lambda \leq T$, the minimum of $\mathscr{C}$ is achieved by taking $u_i^* = \text{sign}(x_i^0)C$ for all $i \in \{1, \ldots, N\}$. This further implies

$$|u_i^*(t) - u_j^*(t)| = \begin{cases} 0 & \text{if } \text{sign}(x_i) = \text{sign}(x_j), \\ 2C & \text{otherwise}, \end{cases}$$

so that for any pair of indices such that $\text{sign}(x_i^0) = -\text{sign}(x_j^0)$, it holds

$$|u_i^*(t) - u_j^*(t)| = \frac{2C}{|x_i^0 - x_j^0| + 2Ct}|x_i^*(t) - x_j^*(t)|. \tag{73}$$

The fact that $\mu_N \rightharpoonup^* \mu^0 = \frac{1}{2}\mathbb{1}_{[-1,1]}\mathscr{L}^1$ as $N \to +\infty$ implies that for all $\epsilon > 0$, there exists $N_\epsilon \geq 1$ such that for any $N \geq N_\epsilon$, there exists at least one pair of indices $i, j \in \{1, \ldots, N\}$ such that $\text{sign}(x_i^0) = -\text{sign}(x_j^0)$ and $|x_i^0 - x_j^0| \leq \epsilon$. Thus, it follows from (73) that (b) fails to hold for some pairs of indices, at least for small times.

Suppose now that (a) is true, i.e. $\lambda > T$, and denote by $\rho_T := \lambda - T$ the corresponding sharp coercivity constant. Let $I_N, J_N \subset \{1, \ldots, N\}$ be the sets of indices defined respectively by

$$I_N = \left\{ i \in \{1, \ldots, N\} \text{ s.t. } |x_i^0| \leq \rho_T C \right\}, \qquad J_N = \{1, \ldots, N\}\backslash I_N.$$

For $N \geq 1$ sufficiently large, I_N is necessarily non-empty since $\rho_T > 0$ and (μ_N^0) narrowly converges towards μ^0. Then for any $i \in I_N$, one has that

$$|x_i^*(T)| \leq |x_i^0| + CT \leq (\rho_T + T)C = \lambda C,$$

and for any such indices, the optimal controls are given by $u_i^* = \frac{1}{\lambda}x_i^*(T)$. In which case, one has

$$x_i^*(T) = \frac{x_i^0}{1 - T/\lambda} \qquad \text{and} \qquad u_i^*(t) = \frac{x_i^*(t)}{\rho_T + t},$$

so that

$$|u_i^*(t) - u_j^*(t)| \leq \frac{1}{\rho_T + t}|x_i^*(t) - x_j^*(t)|, \tag{74}$$

for any pair of indices $i, j \in I_N$. It can be checked reciprocally that $u_i^* = \text{sign}(x_i^0)C$ for any $i \in J_N$, which furthermore yields by (73) that

$$|u_i^*(t) - u_j^*(t)| \leq \begin{cases} 0 & \text{if } \text{sign}(x_i) = \text{sign}(x_j), \\ \dfrac{|x_i^*(t) - x_j^*(t)|}{\rho_T + t} & \text{otherwise}. \end{cases} \tag{75}$$

23

Indeed, in this case $|x_i^0 - x_j^0| \geq 2\rho_T C$ whenever $i, j \in J_N$ and $\text{sign}(x_i) = -\text{sign}(x_j)$. Suppose now that we are given a pair of indices $i, j \in \{1, \ldots, N\}$ such that $i \in I_N$ and $j \in J_N$. If $\text{sign}(x_i^0) = \text{sign}(x_j^0)$, it holds that

$$
\begin{aligned}
|u_i^*(t) - u_j^*(t)| = u_j^*(t) - u_i^*(t) &= \text{sign}(x_j^0)C - \frac{x_i^*(t)}{\rho_T + t} \\
&= \frac{x_j^*(t)C}{|x_j^*(t)|} - \frac{x_i^*(t)}{\rho_T + t} \leq \frac{x_j^*(t) - x_i^*(t)}{\rho_T} = \frac{|x_i^*(t) - x_j^*(t)|}{\rho_T},
\end{aligned}
\tag{76}
$$

since $|x_j^*(t)| \geq \rho_T C$ by definition of J_N. Symmetrically if $\text{sign}(x_i^0) = -\text{sign}(x_j^0)$, one can easily show that

$$
|u_i^*(t) - u_j^*(t)| \leq \frac{1}{\rho_T}|x_i^*(t) - x_j^*(t)|.
\tag{77}
$$

By merging (74), (75), (76) and (77), we conclude that (b) holds whenever $\lambda > T$. $\qquad\square$

In Proposition 9, we have proven that the mean-field coercivity estimate is both necessary and sufficient for the existence of a uniform Lipschitz constant for the sequence of finite-dimensional optimal controls with symmetric initial data. Since we assumed that $(\mu_N^0) \subset \mathscr{P}_N([-1,1])$ and $\mu_N^0 \rightharpoonup^* \mu^0 := \frac{1}{2}\mathbb{1}_{[-1,1]}\mathscr{L}^1$, the fact that the initial distribution are symmetric about the origin holds up to a small error as $N \to +\infty$. Observing in addition that for any $\mu_N^0 \in \mathscr{P}([-1,1])$ the discrete optimal trajectory-control pairs $(\boldsymbol{x}_N^*(\cdot), \boldsymbol{u}_N^*(\cdot)) \in \text{Lip}([0,T], \mathbb{R}^N) \times \mathcal{U}_N$ are uniquely determined, we conclude that the mean-field coercivity condition $(\mathbf{CO}_N)$ is necessary and sufficient in the limit for the existence of a Lipschitz-in-space optimal control for $(\mathcal{P}^V)$.

References

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

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

[3] 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.

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

[5] 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.

[6] N. Bellomo, P. Degond, E. Tadmor, et al. *Active Particles, Volume 1: Advances in Theory, Models, and Applications*. Springer, 2017.

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

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

[9] B. Bonnet and H. Frankowska. Differential Inclusions in Wasserstein Spaces: The Cauchy-Lipschitz Framework. *Journal of Differential Equations*, 271:594–637, 2021.

[10] B. Bonnet and H. Frankowska. Necessary Optimality Conditions for Optimal Control Problems in Wasserstein Spaces. *In revision*, 2021.

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

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

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

[14] 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.

[15] S. Camazine, J.-L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau. *Self-Organization in Biological Systems*. Princeton University Press, 2001.

[16] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat. Sparse Stabilization and Control of Alignment Models. *Mathematical Models and Methods in Applied Sciences*, 25 (3):521–564, 2015.

[17] M. Caponigro, B. Piccoli, F. Rossi, and E. Trélat. Mean-Field Sparse Jurdjevic-Quinn Control. *Mathematical Moddels and Methods in Applied Sciences*, 27(7):1223–1253, 2017.

[18] P Cardaliaguet, A. Poretta, and D. Tonon. Sobolev Regularity for the First Order Hamilton–Jacobi Equation. *Calculus of Variations and Partial Differential Equations*, 54:3037–3065, 2015.

[19] P. Cardaliaguet and L. Silvester. Hölder Continuity to Hamilton-Jacobi Equations with Super-Quadratic Growth in the Gradient and Unbounded Right-Hand Side. *Communications in Partial Differential Equations*, 37(9):1668–1688, 2012.

[20] E. Carlini and F.S. Silva. A Fully Discrete Semi-Lagrangian Scheme for a First Order Mean Field Game Problem. *SIAM Journal on Numerical Analysis*, 52(1):45–67, 2014.

[21] J.A. Carrillo, M. Fornasier, J. Rosado, and G. Toscani. Asymptotic Flocking for the Kinetic Cucker-Smale Model. *SIAM Journal on Mathematical Analysis*, 42(1):218–236, 2010.

[22] 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.

[23] 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.

[24] J.S. Chang and G. Cooper. A Practical Difference Scheme for Fokker-Planck Equations. *Journal of Computational Physics*, 6(1):1–16, 1970.

[25] Y.T. Chow and W. Gangbo. A Partial Laplacian as an Infinitesimal Generator on the Wasserstein Space. *Journal of Differential Equations*, 267(10):6065–6117, 2019.

[26] R. Cibulka, A.L. Dontchev, M.I. Krastanov, and V.M. Veliov. Metrically Regular Differential Generalized Equations. *SIAM Journal on Control and Optimization*, 56(1):316–342, 2018.

[27] F Clarke. *Functional Analysis, Calculus of Variations and Optimal Control.* Springer, 2013.

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

[29] F. Cucker and S. Smale. Emergent Behavior in Flocks. *IEEE Transactions on Automatic Control*, 52(5):852–862, 2007.

[30] G. De Phillipis and A. Figalli. Regularity for Solutions of the Monge-Ampère Equation. *Inventiones Mathematicae*, 192(1):55–69, 2013.

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

[32] A.L. Dontchev and W.W. Hager. Lipschitzian Stability in Nonlinear Control and Optimization. *SIAM Journal on Control and Optimization*, 31(3):569–603, 1993.

[33] A.L. Dontchev, M.I. Krastanov, and V.M. Veliov. On the Existence of Lipschitz Continuous Optimal Feedback Controls. *Vietnam Journal of Mathematics*, 47:579–597, 2019.

[34] 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.

[35] 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.

[36] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions.* CRC Press, 1992.

[37] A. Figalli, Y.H. Kim, and R.J. McCann. Hölder Continuity and Injectivity of Optimal Maps. *Archives of Rational Mechanics and Analysis*, 209(3):747–795, 2013.

[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 Transactions 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] W. Gangbo and A.R. Mészáros. Global Well-Posedness of Master Equations for Deterministic Displacement Convex Potential Mean Field Games. *arXiv preprint arXiv:2004.01660*, 2020.

[42] W. Gangbo and A. Swiech. Existence of a Solution to an Equation Arising in the Theory of Mean Field Games. *Journal of Differential Equations*, 259(11):6573–6643, 2015.

[43] W. Gangbo and A. Tudorascu. On Differentiability in the Wasserstein Space and Well-Posedness for Hamilton-Jacobi Equations. *Journal de Mathématiques Pures et Appliquées*, 00:1–47, 2018.

[44] S.-Y. Ha and J.G. Liu. A Simple Proof of the Cucker-Smale Flocking Dynamics and Mean-Field Limit. *Comm. Math. Sci.*, 7(2):297–325, 2009.

[45] 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.

[46] 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.

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

[48] M. Lavrentiev. Sur Quelques Problèmes du Calcul des Variations. *Annali di Matematica Pura e Applicata*, 4(1):7–28, 1927.

[49] S. Mayorga. Short Time Solution to the Master Equation of a First Order Mean Field Game. *Journal of Differential Equations*, 268(10):6251–6318, 2020.

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

[51] A. Muntean, J. Rademacher, and A. Zagaris. *Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field Limits and Ergodicity.* Springer, 2016.

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

[53] B. Piccoli, F. Rossi, and E. Trélat. Control of the kinetic Cucker-Smale model. *SIAM Journal on Mathematical Analysis*, 47(6):4685–4719, 2015.

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

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

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