# Intrinsic Lipschitz Regularity of Mean-Field Optimal Controls

Benoît Bonnet, Francesco Rossi

# Intrinsic Lipschitz Regularity of Mean-Field Optimal Controls*

Benoît Bonnet†, Francesco Rossi‡

June 22, 2020

**Abstract**

In this paper, we provide a sufficient condition under which the vector-fields solution of mean-field optimal control problems formulated on continuity equations are Lipschitz in space. Our approach involves a novel combination of mean-field approximation results 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 the differential calculus of Wasserstein spaces.

**Keywords.** Optimal control of continuity equations, regularity of minimizers, multi-agent systems, metric regularity, discretisation, differential calculus in Wasserstein spaces

**AMS Subject Classification.** 35B65, 49J20, 49J30, 58E25

## 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 decade. Multi-agent systems are ubiquitous in applications ranging from aggregation phenomena in biology [9, 18] to the understanding of crowd motion [8, 30], animal flocks [6, 31], swarms of autonomous vehicles [14] or traffic flows [34]. While the first studies on multi-agent systems were formulated in a graph-theoretic framework (see e.g. [14] and references therein), several recent models have started to rely on continuous-time dynamical systems to describe this type of collective dynamics. In this context, a multi-agent system is usually described by a family of coupled differential equations of the form

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

where $\boldsymbol{x} = (x_1, \dots, x_N)$ denotes the state of all the agents and $\boldsymbol{v}_N[\cdot](\cdot,\cdot)$ is a *non-local velocity field* depending both on the running agent and on the whole state of the system (see e.g. [7, 31]). However general and useful, these models may not be the most powerful ones in order to capture the global features of a multi-agent system. Besides, their intrinsic dependence on the number $N$ of agents makes most of the classical computational approaches practically intractable for large systems.

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. [60]). In this setting, the agents are supposed to be indistinguishable, and the assembly of particles is described by means of its *spatial density* $\mu(\cdot)$. The evolution through time of this global quantity is prescribed by a *non-local continuity equation* of the form

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

Such a macroscopic approach has been successfully used e.g. to model pedestrian dynamics and biological systems, as well as to transpose the study of classical patterns such as consensus or flocking formation to the mean-field setting. From a quite different standpoint, J.M. Lasry and P.L. Lions proposed in their seminal paper [50] a model for the self-organisation of large systems of rational agents based on the optimisation of a selfish cost, which led to the development of the theory of *mean-field games* (see also [49]). Both facets of the literature have hugely benefited from the recent progresses made in the theory of *optimal transportation*, for which we point to the reader to the reference monographs [5, 58, 59].

*This research was partially supported by the Padua University grant SID 2018 "Controllability, stabilizability and infimun gaps for control systems", prot. BIRD 187147. The first author was supported by the Archimède Labex (ANR-11-LABX-0033) and by the A*MIDEX project (ANR-11-IDEX-0001-02), funded by the "Investissements d'Avenir" French Government program managed by the French National Research Agency (ANR).

†IMJ-PRG Sorbonne Université, Paris, France (benoit.bonnet@imj-prg.fr)

‡Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova, Padova, Italy (francesco.rossi@math.unipd.it)

More recently, the problem of controlling multi-agent systems so as to promote a desired behaviour or configuration became relevant in a growing number of applications. Motivated by implementability and efficiency issues, many contributions have therefore been aiming to generalise 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

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

While a few results have been dealing with controllability issues [38, 39] or the explicit design of control laws [19, 20, 55], the major part of the literature has been focusing on *mean-field optimal control problems*, with contributions ranging from existence results [43, 44, 45] to first-order optimality conditions [10, 11, 12, 15, 24, 25, 56] and numerical methods [1, 16, 57].

One of the distinctive features of continuity equations is that they require fairly restrictive regularity assumptions on the driving velocity fields to be classically well-posed. While (3) makes sense whenever the drift and control are measurable and satisfy some integrability bounds, the corresponding notion of so-called *superposition solution* (see Theorem 5 below) is fairly weak and of limited practical use. In [2, 35], a theory of well-posedness was developed for continuity equations with Sobolev and $BV$ velocity fields . However powerful and general, this theory 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 holds (see Theorem 6 below) for *arbitrary measures* for (3) is that of Cauchy-Lipschitz regularity (see e.g. [3, Section 3] and [54]). In this framework, (3) can be formulated indifferently on empirical or absolutely continuous measures, its solutions are unique, and stability estimates are available both with respect to the right-hand side of the equation and its initial datum.

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 be to provide a strategy which can be in turn applied – either exactly or approximately – to finite-dimensional systems of the form (1). In the case in which the infinite-dimensional strategy is not strictly optimal for the discrete multi-agent system, one would also like to have access to quantitative error estimates between the true solution and the approximate one. From a numerical standpoint, Cauchy-Lipschitz regularity is also relevant to ensure the well-posedness of numerical methods such as semi-Lagrangian schemes (see e.g. [23, 26]), as well as to prevent the apparition of Lavrentiev-type instabilities in the context of optimal control (see e.g. [51]). 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. [10, 11, 12, 44, 45]), or at least some continuity assumptions on the driving fields (see [24, 25]). A natural question is then to ask whether this regularity property can hold intrinsically or not, and if yes under which assumptions. In this paper, we investigate this question in the setting of mean-field optimal control problems, formulated on controlled dynamics given by (3).

**Remark 1** (“Optimise-then-discretise ”for hyperbolic PDEs). *Let it be noted that the problem of ensuring a correspondence between solutions of optimal control problems governed by hyperbolic partial differential equations and their discrete approximations is in general highly non-trivial. Indeed, it has been noticed as early as [48] that discretizations of the Hibert Uniqueness Method introduced by J.L. Lions in [52] to perform the exact controllability of a wide class of partial differential equations could give rise to high frequency oscillations and diverge. We refer the reader to [40] and references therein for a modern treatment of this problem.*

It is well-known that solutions of Wasserstein optimal control problems need not be regular in general. Indeed, there exists a vast literature devoted to the study of the regularity properties of solutions to Monge's optimal transport problem (see e.g. [33, 42] for some of the farthest-reaching contributions on this topics), 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

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

This tangent vector field should be – roughly speaking – as regular as the derivative of the optimal transport map. For the optimal control problem $(\mathcal{P}_{\mathrm{BB}})$, Caffarelli proved in [17] that $v(t,\cdot) \in C^{k,\alpha}_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{R}^d)$ for some $\alpha \in (0,\bar{\alpha})$ whenever $\mu^0,\mu^1 \in \mathscr{P}^{ac}(\mathbb{R}^d)$ have densities with respect to the $d$-dimensional Lebesgue measure which have regularity at least $C^{k,\bar{\alpha}}_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{R}^d)$ .

Another context in which the regularity of mean-field optimal controls has been (indirectly) investigated is that of mean-field games. Indeed, there is a large literature dedicated to the regularity of the value function $(t,x) \mapsto V^*(t,x)$ solving the backward Hamilton-Jacobi equation of the coupled mean-field games system

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

In the setting of potential mean-field games, the tangent 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, regularity properties of the optimal control can be recovered from that of the optimal value function, and are expected to have one order of differentiation fewer. We refer the reader e.g. to [21] for Sobolev regularity results and to [22] for Hölder regularity properties.

In this paper, we investigate the intrinsic Lipschitz-in-space regularity of the solutions of general mean-field optimal control problems of the form

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

The set of admissible controls for $(\mathcal{P})$ is defined by $\mathcal{U} = L^1([0,T], L^1(\mathbb{R}^d, U; \mu(t)))$ where $U \subset \mathbb{R}^d$ is a convex and compact set. Remark that since we do not impose any a priori regularity assumptions on the control vector fields $u(\cdot,\cdot)$, there may not exist solutions to the non-local transport equation (3) driving problem $(\mathcal{P})$. Moreover even if they do exist, these solution will not be classically well-posed and only defined in a weak sense (see Theorem 5 below).

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

**Theorem 1** (Existence of Lipschitz-in-space optimal controls for $(\mathcal{P})$)**.** *Let $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$, $(\mu_N^0) \subset \mathscr{P}_c(\mathbb{R}^d)$ be a sequence of empirical measures narrowly converging towards $\mu^0$. Suppose that hypotheses* **(H)** *of Section 4 hold, and that the* mean-field coercivity *assumption* $(\mathbf{CO}_N)$ *described in Section 5 holds as well.*

*Then, there exists a mean-field optimal pair control-trajectory $(u^*(\cdot,\cdot), \mu^*(\cdot)) \in \mathcal{U} \times \mathrm{Lip}([0,T], \mathscr{P}_c(\mathbb{R}^d))$ for problem $(\mathcal{P})$. Moreover, the optimal control map $(t,x) \in [0,T] \times \mathbb{R}^d \mapsto u^*(t,x)$ is $\mathcal{L}_U$-Lipschitz in space for $\mathscr{L}^1$-almost every $t \in [0,T]$, where the uniform constant $\mathcal{L}_U$ only depends on the data of the problem $(\mathcal{P})$.*

The proof of this result is built around two main ingredients. The first one is an existence result for mean-field optimal controls which was derived in [43] and recalled in Theorem 7 below. In the latter, it is proven under very general assumptions that there exist optimal solutions of problem $(\mathcal{P})$ which can be recovered as $\Gamma$-limits in a suitable topology of sequences of solutions to the discrete problems

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

Here, $\mathcal{U}_N = L^\infty([0,T], U^N)$, and the functionals $(t,x,\boldsymbol{x}) \in [0,T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \mapsto \boldsymbol{v}_N[\boldsymbol{x}](t,x)$, $(t,\boldsymbol{x}) \in [0,T] \times (\mathbb{R}^d)^N \mapsto \boldsymbol{L}_N(t,\boldsymbol{x})$ and $\boldsymbol{x} \in (\mathbb{R}^d)^N \mapsto \boldsymbol{\varphi}(\boldsymbol{x})$ are discrete approximating sequences (see Definition 8 below) of $v[\cdot](\cdot,\cdot)$, $L(\cdot,\cdot)$ and $\varphi(\cdot)$ respectively.

The second key component of our approach is a careful adaptation to the family of problems $(\mathcal{P}_N)$ of a methodology recently developed in [28, 36]. These contributions provide powerful *metric regularity* results (see Definition 11 below) for a large class of differential inclusions. This part relies crucially on the following *uniform mean-field coercivity estimate* $(\mathbf{CO}_N)$ for the sequence of problems $(\mathcal{P}_N)$

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

along *optimal mean-field Pontryagin triples* $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot), \boldsymbol{r}_N^*(\cdot))$ (see Proposition 6 below). In this context, $\mathbf{Hess}\,(\bullet)[\cdot](\cdot,\cdot)$ denotes the discrete version of the intrinsic Wasserstein Hessian bilinear form (see e.g. [27, 47]), which construction is further detailed in Section 2. In essence, this uniform coercivity assumption allows us to inverse the maximisation condition stemming from an application of the *Pontryagin Maximum Principle* to $(\mathcal{P}_N)$, with a control on the Lipschitz constant of this inverse. The main subtlety lies in the fact that we need these estimates to be **uniform with respect to** $N$. Consequentially, we apply to $(\mathcal{P}_N)$ an adapted mean-field Pontryagin Maximum Principle − which is the discrete counterpart of the Wasserstein PMP studied in [10, 11, 12] −, and express the coercivity condition in the language of Wasserstein calculus. From there on, the statement of Theorem 1 can then be recovered by standard limit arguments which can be found e.g. in [10, 45].

**Remark 2** (Comparison with related contributions in mean-field games)**.** *It was recently brought to our attention that a result similar to Theorem 1 above was derived in [46] for mean-field games, which are known to be linked in certain cases to optimal control problems in Wasserstein spaces (see [50]). In [46], 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* $T$ *is sufficiently small. These two requirement are very close to our standing assumptions. Indeed, we posit in hypotheses* **(H)** *of Section 4 that all our data are* $C^{2,1}_{\mathrm{loc}}$*, and it is illustrated in Section 6 that our uniform coercivity estimate* $(\mathbf{CO}_N)$ *can be interpreted as a quantitative condition comparing the relative size of the time horizon* $T$ *with other constants of the problem. We would also like to stress that our approach allows for a much wider class of dynamics and cost functionals.*

*Moreover, the proof strategy of [46] is fairly close to the one that we independently developed here, as it relies 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. Let it also be noted that the results of [46] have recently been extended in [53] to a broader class of first-order mean-field games systems.*

The structure of this article is the following. In Section 2, we recall several general prerequisites on measure theory and optimal transport. In Section 3, we review notions pertaining to finite-dimensional optimal control problems, with a particular emphasise on Lipschitz feedbacks. We proceed by exposing in Section 4 well-posedness results and concepts dealing with continuity equations and mean-field optimal control problems. In Section 5, we state precisely the coercivity assumption $(\mathbf{CO}_N)$ and prove our main result 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 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 theory. For these topics, we refer the reader to [4] and [5, 58, 59] respectively.

We denote by $(\mathcal{M}(\mathbb{R}^d,\mathbb{R}^m),\|\cdot\|_{TV})$ the Banach space of $m$-dimensional vector-valued Borel measures defined on $\mathbb{R}^d$ endowed with the total variation norm, defined by

$$\|\boldsymbol{\nu}\|_{TV} := \sup\left\{\sum_{k=1}^{+\infty}|\boldsymbol{\nu}(E_k)| \text{ s.t. } E_k \text{ are disjoint Borel sets and } \bigcup_{k=1}^{+\infty}E_k = \mathbb{R}^d\right\},$$

for any $\boldsymbol{\nu} \in \mathcal{M}(\mathbb{R}^d,\mathbb{R}^m)$. 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 $C^0_c(\mathbb{R}^d,\mathbb{R}^m)$ of continuous and compactly supported functions. The latter is endowed with the duality bracket

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

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

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

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

The *support* of a Borel measure $\boldsymbol{\nu} \in \mathcal{M}(\mathbb{R}^d, \mathbb{R}^m)$ is defined as the closed set $\mathrm{supp}(\boldsymbol{\nu}) = \{x \in \mathbb{R}^d \text{ s.t. } \boldsymbol{\nu}(\mathcal{N}) \neq 0 \text{ for any neighbourhood } \mathcal{N} \text{ of } x\}$. We denote by $\mathscr{P}_c(\mathbb{R}^d) \subset \mathscr{P}(\mathbb{R}^d)$ the subset of Borel 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 $\mu$ – denoted by $\boldsymbol{\nu} \ll \mu$ –, provided that $|\boldsymbol{\nu}|(B) = 0$ whenever $\mu(B) = 0$ for any Borel set $B \subset \Omega$.*

*Moreover, we have that $\boldsymbol{\nu} \ll \mu$ if and only if there exists a Borel map $u \in L^1(\Omega, U; \mu)$ such that $\boldsymbol{\nu} = u(\cdot)\mu$. This map is referred to as the Radon-Nikodym derivative of $\boldsymbol{\nu}$ with respect to $\mu$, and denoted by $u(\cdot) := \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}(\cdot)$.*

We recall in the following definition the notions of *pushforward* of a Borel probability measure through a Borel map and of *transport plan.*

**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 $\mu$ through $f(\cdot)$ is defined as the Borel probability measure such that $f_{\#}\mu(B) = \mu(f^{-1}(B))$ 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 $\mu$ and $\nu$ – denoted by $\gamma \in \Gamma(\mu, \nu)$ –, provided that $\pi^1_{\#}\gamma = \mu$ and $\pi^2_{\#}\gamma = \nu$, where $\pi^1, \pi^2 : \mathbb{R}^{2d} \to \mathbb{R}^d$ respectively denote the projection on the first and second component.*

In 1942, the Russian mathematician Leonid Kantorovich introduced the *optimal mass transportation problem* in its modern mathematical formulation. Given two probability measures $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ and a cost function $c : \mathbb{R}^{2d} \to \mathbb{R}$, one searches for a *transport plan* $\gamma \in \Gamma(\mu, \nu)$ such that

$$\int_{\mathbb{R}^{2d}} c(x,y)\mathrm{d}\gamma(x,y) = \min_{\gamma} \left\{ \int_{\mathbb{R}^{2d}} c(x,y)\mathrm{d}\gamma'(x,y) \ \text{ s.t. } \gamma' \in \Gamma(\mu,\nu) \right\}.$$

This problem has been extensively studied in very broad contexts (see e.g. [5, 58, 59]), with high levels of generality on the underlying spaces and cost functions. In the particular case where $c(x,y) = |x-y|^p$ for some real number $p \geq 1$, the optimal transport problem can be used to define a distance overt $\mathscr{P}_p(\mathbb{R}^d)$.

**Definition 4** (Wasserstein distance and Wasserstein spaces). *Given two measures $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$, the $p$-Wasserstein distance between $\mu$ and $\nu$ is defined by*

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

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

We recall some of the interesting properties of these spaces in the following proposition (see e.g. [5, Chapter 7] or [59, Chapter 6]).

**Proposition 1** (Elementary properties of the Wasserstein spaces). *The Wasserstein spaces $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ are separable metric spaces, and the distance $W_p$ metrizes the weak-* topology associated to the duality pairing* (4)*:*

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

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

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

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

**Theorem 2** (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 $\mu$-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 $\mu$, 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 introduce 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] for an exhaustive treatment of the first-order theory, and to [47] for the theoretical foundations of the second-order theory. We borrow the main working definitions dealing with Wasserstein Hessians from [27, Section 3]. Throughout this section, we denote by $\phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ a lower-semicontinuous and proper functional with non-empty effective domain $D(\phi) = \{\mu \in \mathscr{P}_2(\mathbb{R}^d) \text{ s.t. } \phi(\mu) < +\infty\}$.

We start by introducing in the following definition the notions of *classical subdifferential* and *superdifferential* for functionals defined over $(\mathscr{P}_2(\mathbb{R}^d), W_2)$.

**Definition 5** (Classical 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 $\mu$ 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 $\mu$ if $(-\xi) \in \partial^-(-\phi)(\mu)$.*

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

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

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 some $\mu \in D(\phi)$ 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 $\mu$, 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 characterization (8) of the Wasserstein gradient $\nabla_\mu\phi(\mu)$, we can deduce the following chain rule along elements of $\mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$ (see [5, Proposition 10.3.18]).

**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*

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

*where $\mathcal{L}_\xi\phi(\mu)$ denotes the* Lie derivative *of $\phi(\cdot)$ at $\mu$ 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 derivative for functionals defined over $\mathscr{P}_2(\mathbb{R}^d)$. We therefore introduce in the following definition the notion of *Wasserstein Hessian bilinear form* for a sufficiently regular functional $\phi(\cdot)$.

**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*

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

*is differentiable at $\mu$ in the sense of Definition 6. Then, we define the partial Wasserstein Hessian of $\phi(\cdot)$ at $\mu$ as the bilinear form*

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

*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*

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

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

We end this preparatory section by providing in the following proposition a condensed version of several statements of [27, Section 3]. This will allow us to derive an analytical and natural expression for the Hessian bilinear form, as well as a second-order differentiation formula for Wasserstein functionals.

**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*

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

*are 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 writes explicitly as*

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

*for any $\xi_1,\xi_2 \in \mathrm{Tan}_\mu\mathscr{P}_2(\mathbb{R}^d)$. Here, the map $\mathrm{D}_x\nabla_\mu\phi(\mu)(x) \in \mathbb{R}^{d\times d}$ is the classical differential of $\nabla_\mu\phi(\mu)(\cdot)$ at $x \in \mathbb{R}^d$, while $\mathrm{D}^2_\mu\phi(\mu)(x,\cdot) : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ denotes the matrix-valued map which columns are the Wasserstein gradients of the components of $\nabla_\mu\phi(\mu)(x)$ defined as in Definition 6. Moreover, one has that*

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

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

## 2.3 Mean-field adapted structures and discrete measures

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

We denote by $\mathscr{P}_N(\mathbb{R}^d) = \{\frac{1}{N}\sum_{i=1}^N \delta_{x_i} \text{ s.t. } (x,\dots,x_N) \in (\mathbb{R}^d)^N\}$ the set of $N$*-empirical probability measures* over $\mathbb{R}^d$. It is a standard result in optimal transport theory (see e.g. [5, Chapter 7]) that $\cup_N \mathscr{P}_N(\mathbb{R}^d)$ is a dense subset of $\mathscr{P}(\mathbb{R}^d)$ with respect to the narrow topology. For any $N \geq 1$, we denote by $\boldsymbol{x} = (x_1,\dots,x_N)$ 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. 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 a continuous functional $\phi(\cdot)$ defined over $\mathscr{P}_c(\mathbb{R}^d)$.

**Definition 8** (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 $(\boldsymbol{F}_N(\cdot,\cdot)) \subset C^0(\Omega \times (\mathbb{R}^d)^N, \mathbb{R}^m)$ such that*

$$F(x,\mu[\boldsymbol{x}]) = \boldsymbol{F}_N(x,\boldsymbol{x}) \tag{13}$$

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

We introduce below the notion of $C^{2,1}_{\mathrm{loc}}$*-Wasserstein regularity* which we will use throughout the article.

**Definition 9** ($C^{2,1}_{\mathrm{loc}}$-Wasserstein regularity)**.** *A functional $\phi : \mathscr{P}_c(\mathbb{R}^d) \to \mathbb{R}^m$ is said to be $C^{2,1}_{\mathrm{loc}}$-Wasserstein regular if for any compact set $K \subset \mathbb{R}^d$, the map $\phi(\cdot)$ is twice differentiable over $\mathscr{P}(K)$ in the sense of Definition 7 and such that*

$$\begin{aligned} &\phi(\mu) + \|\nabla_\mu\phi(\mu)(\cdot)\|_{C^0(K)} + \|\mathrm{D}_x\nabla_\mu\phi(\mu)(\cdot)\|_{C^0(K)} + \big\|\mathrm{D}^2_\mu\phi(\mu)(\cdot,\cdot)\big\|_{C^0(K\times K)} \\ &\quad + \mathrm{Lip}\big(\mathrm{D}_x\nabla_\mu\phi(\cdot)(\cdot);\mathscr{P}(K)\times K\big) + \mathrm{Lip}\big(\mathrm{D}^2_\mu\phi(\cdot)(\cdot,\cdot);\mathscr{P}(K)\times K\times K\big) \leq C_K \end{aligned} \tag{14}$$

*for all $\mu \in \mathscr{P}(K)$, where $C_K > 0$ is a constant depending on $K$.*

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

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

for any $\boldsymbol{x}, \boldsymbol{y} \in (\mathbb{R}^d)^N$, where $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product of $\mathbb{R}^d$. We denote by $|\cdot|_N = \sqrt{\langle \cdot, \cdot \rangle_N}$ the rescaled Euclidean norm induced by $\langle \cdot, \cdot \rangle_N$ over $(\mathbb{R}^d)^N$, and remark that $((\mathbb{R}^d)^N, \langle \cdot, \cdot \rangle_N)$ has the structure of a Hilbert space.

In the following proposition, we show that the Wasserstein differential structure described in Section 2.2 for functionals defined on measures induces a natural differential structure on the Hilbert space $((\mathbb{R}^d)^N, \langle \cdot, \cdot \rangle_N)$.

**Proposition 4** (Mean-field derivatives of symmetric maps)**.** *Let $\phi(\cdot)$ be $C^{2,1}_{\mathrm{loc}}$-Wasserstein regular with mean-field approximating sequence $(\boldsymbol{\phi}_N(\cdot)) \subset C^0((\mathbb{R}^d)^N)$. Then one has that $\boldsymbol{\phi}_N \in C^{2,1}_{\mathrm{loc}}((\mathbb{R}^d)^N, \mathbb{R})$ for any $N \geq 1$. Moreover, the following Taylor expansion holds*

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

*for any $\boldsymbol{x}, \boldsymbol{h} \in (\mathbb{R}^d)^N$, where we introduced the* mean-field gradient $\mathbf{Grad}\, \boldsymbol{\phi}_N(\cdot)$ *and* mean-field Hessian $\mathbf{Hess}\, \boldsymbol{\phi}_N[\cdot]$ *of $\boldsymbol{\phi}_N(\cdot)$, defined 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*

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

*For any compact set $K \subset \mathbb{R}^d$, there exists a constant $C_K > 0$ such that for any $N \geq 1$, one has that*

$$\|\boldsymbol{\phi}_N(\cdot)\|_{C^2(K^N)} + \mathrm{Lip}\left(\mathbf{Hess}\, \boldsymbol{\phi}_N[\cdot]; K^N\right) \leq C_K, \tag{19}$$

*where the $C^2$-norm here is defined by*

$$\|\boldsymbol{\phi}_N(\cdot)\|_{C^2(\boldsymbol{K})} = \max_{\boldsymbol{x} \in \boldsymbol{K}} \boldsymbol{\phi}_N(\boldsymbol{x}) + \max_{\boldsymbol{x} \in \boldsymbol{K}} |\mathbf{Grad}\, \boldsymbol{\phi}_N(\boldsymbol{x})|_N + \max_{\boldsymbol{x} \in \boldsymbol{K}} \sup_{|\boldsymbol{h}|_N \leq 1} \mathbf{Hess}\, \boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}, \boldsymbol{h}), \tag{20}$$

*for any set $\boldsymbol{K} \subset (\mathbb{R}^d)^N$.*

*Proof.* Let $\boldsymbol{x}, \boldsymbol{h} \in (\mathbb{R}^d)^N$, $\epsilon = \frac{1}{4} \min_{x_i \neq x_j} |x_i - x_j|$ and $\zeta_N(\cdot)$ be defined by

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

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

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

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

By assumption, the map $\phi(\cdot)$ is differentiable at $\mu[\boldsymbol{x}] \in \mathscr{P}_c(\mathbb{R}^d)$. We can therefore apply the first-order chain rule derived in Proposition 2 along tangent vectors to recover that

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

We can further obtain by recalling the definition of the symmetric maps $\boldsymbol{\phi}_N(\cdot)$ given in (13) that

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

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 $\boldsymbol{x}$, and by Riesz's Theorem (see e.g. [13, Theorem 5.5]), its differential can be represented in the Hilbert space $((\mathbb{R}^d)^N, \langle\cdot,\cdot\rangle_N)$ by the the so-called mean-field gradient $\mathbf{Grad}\,\boldsymbol{\phi}_N(\boldsymbol{x}) := (\nabla_\mu \phi(\mu[\boldsymbol{x}])(x_i))_{1\leq i\leq N}$.

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 the map $\phi(\cdot)$ is twice differentiable in the sense of Definition 7, we can use the the second-order differentiation formula (12) to obtain that

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

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

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

where we used the analytical expression (11) of the Wasserstein Hessian. We accordingly introduce the mean-field Hessian of $\boldsymbol{\phi}_N(\cdot)$ at $\boldsymbol{x}$, defined by

$$\mathbf{Hess}\,\boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}^1,\boldsymbol{h}^2) = \frac{1}{N}\sum_{i=1}^N \langle \mathrm{D}_x\nabla_\mu\phi(\mu[\boldsymbol{x}])(x_i)h_i^1, h_i^2\rangle + \frac{1}{N^2}\sum_{i,j=1}^N \langle \mathrm{D}_\mu^2\phi(\mu[\boldsymbol{x}])(x_i,x_j)h_i^1, h_j^2\rangle. \tag{25}$$

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 Taylor expansion formula (16) can be derived directly by expanding $\boldsymbol{\phi}_N(\boldsymbol{x}+\boldsymbol{h})$ using the classical Taylor theorem in $(\mathbb{R}^d)^N$ along with (23) and (25). Defining the $C^2$-norm of a functional $\boldsymbol{\phi}_N(\cdot)$ as in (20), it follows directly from the uniform bounds (14) stemming from the $C^{2,1}_{\mathrm{loc}}$-Wasserstein regularity of $\phi(\cdot)$ that for any compact set $K \subset \mathbb{R}^d$, there exists a constant $C_K > 0$ such that

$$\|\boldsymbol{\phi}_N(\cdot)\|_{C^2(K^N)} + \mathrm{Lip}\big(\mathbf{Hess}\,\boldsymbol{\phi}_N[\cdot];K^N\big) \leq C_K.$$

This ends the proof of Proposition 4. □

**Remark 3** (Matrix representation of the mean-field Hessian in $(\mathbb{R}^d)^N$)**.** *By Riesz's Theorem applied in the Hilbert space $((\mathbb{R}^d)^N), \langle\cdot,\cdot\rangle_N)$, one can represent the action of the Hessian bilinear form $\mathbf{Hess}\,\boldsymbol{\phi}_N[\boldsymbol{x}](\cdot,\cdot)$ as*

$$\mathbf{Hess}\,\boldsymbol{\phi}_N[\boldsymbol{x}](\boldsymbol{h}^1,\boldsymbol{h}^2) = \big\langle \mathbf{Hess}\,\boldsymbol{\phi}_N(\boldsymbol{x})\boldsymbol{h}^1, \boldsymbol{h}^2\big\rangle_N. \tag{26}$$

*for any $\boldsymbol{x},\boldsymbol{h}^1,\boldsymbol{h}^2 \in (\mathbb{R}^d)^N$, where $\mathbf{Hess}\,\boldsymbol{\phi}_N(\boldsymbol{x})$ is a matrix. In this case, its components are given explicitly by*

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

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

# 3 Locally optimal Lipschitz feedbacks in optimal control

In this section, we start by recalling some classical facts about finite dimensional optimal control problems. We then describe in Theorem 3 a result proven in [36], which provides sufficient conditions for the existence of locally optimal Lipschitz feedbacks in a neighbourhood of an optimal trajectory. This result is based on general metric regularity properties explored recently in [28] for dynamical differential inclusions. Throughout this section, we will study the finite-dimensional optimal control problem

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

under the following structural assumptions.

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

1. *The set of admissible controls is $\mathcal{U} = L^\infty([0,T], U)$ where $U \subset \mathbb{R}^d$ is a compact and convex set.*
2. *The control cost $u \mapsto \psi(u)$ is $C^{2,1}_{\mathrm{loc}}$-regular and strictly convex.*
3. *The map $(t,x) \mapsto f(t,x)$ is Lipschitz with respect to $t \in [0,T]$, sublinear and $C^{2,1}_{\mathrm{loc}}$-regular with respect to $x \in \mathbb{R}^d$.*
4. *The running cost $(t,x) \mapsto l(t,x)$ is Lipschitz with respect to $t \in [0,T]$ and $C^{2,1}_{\mathrm{loc}}$-regular with respect to $x \in \mathbb{R}^d$. Similarly, the final cost $x \mapsto g(x)$ is $C^{2,1}_{\mathrm{loc}}$-regular.*

As a direct consequence of our regularity hypotheses and of the compactness of the set of admissible control values $U$, we have the following lemma.

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

This follows directly from Grönwall's Lemma. From now on, we fix such a compact set $K \subset \mathbb{R}^d$.

**Proposition 5** (Existence of solutions for problem $(\mathcal{P}_{\mathrm{oc}})$)**.** *Under hypotheses $(\mathbf{H}_{oc})$, there exists an optimal pair control-trajectory $(u^*(\cdot), x^*(\cdot)) \in \mathcal{U} \times \mathrm{Lip}([0,T], K)$ for problem $(\mathcal{P}_{\mathrm{oc}})$.*

This result is standard under our working hypotheses and can be found e.g. in [29, Theorem 23.11]. We can further define the *Hamiltonian* associated to $(\mathcal{P}_{\mathrm{oc}})$ 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 $(u^*(\cdot), x^*(\cdot))$ be an optimal pair control-trajectory for $(\mathcal{P}_{\mathrm{oc}})$. By the *Pontryagin Maximum Principle* (see e.g. [29, Theorem 22.2]), there exists a curve $p^*(\cdot) \in \mathrm{Lip}([0,T], \mathbb{R}^d)$ such that the couple $(x^*(\cdot), p^*(\cdot))$ is a solution of the *forward-backward Hamiltonian system*

$$\begin{cases} \dot{x}^*(t) = \quad \nabla_p H(t, x^*(t), p^*(t), u^*(t)), \;\; x^*(0) = x^0, \\ \dot{p}^*(t) = -\nabla_x H(t, x^*(t), p^*(t), u^*(t)), \;\; p^*(T) = -\nabla g(x^*(T)). \end{cases} \tag{27}$$

Moreover, the *Pontryagin maximisation condition*

$$H(t, x^*(t), p^*(t), u^*(t)) = \max_{v \in U} \left[ H(t, x^*(t), p^*(t), v) \right], \tag{28}$$

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

Remark now that, as a by-product of the local Lipschitz regularity of $f(\cdot,\cdot)$, $l(\cdot,\cdot)$ and $g(\cdot)$, there exists a compact set $K' \subset \mathbb{R}^d$ such that any covector $p(\cdot)$ associated with an admissible pair $(u(\cdot), x(\cdot))$ via (27) satisfies $p \in \mathrm{Lip}([0,T], K')$. We henceforth 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 $(\mathcal{P}_{\mathrm{oc}})$. Moreover, we denote by $\mathcal{L}_{\mathcal{K}}$ the Lipschitz constant over $\mathcal{K}$ of the maps $f(\cdot,\cdot)$, $l(\cdot,\cdot)$, $\psi(\cdot)$ and $H(\cdot,\cdot,\cdot,\cdot)$ and of their derivatives with respect to the variables $(x,u)$ up to the second order.

We now present the central and somewhat less standard assumption which allows for the construction of locally optimal feedbacks around the graph of $x^*(\cdot)$.

**Definition 10** (Uniform coercivity estimate)**.** *We say that a Pontryagin triple $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$ for $(\mathcal{P}_{\mathrm{oc}})$ satisfies the uniform coercivity estimate with constant $\rho > 0$ if the following inequality holds*

$$\begin{aligned} \big\langle \nabla_x^2 \, g(x^*(T)) y(T), y(T) \big\rangle - \int_0^T \big\langle \nabla_x^2 \, H(t, x^*(t), p^*(t), u^*(t)) y(t), y(t) \big\rangle \, \mathrm{d}t & \\ - \int_0^T \big\langle \nabla_u^2 \, H(t, x^*(t), p^*(t), u^*(t)) w(t), w(t) \big\rangle \, \mathrm{d}t \geq \rho \int_0^T |w(t)|^2 \mathrm{d}t & \end{aligned} \tag{29}$$

*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 control-state equation*

$$\begin{cases} \dot{y}(t) = \mathrm{D}_x f(t, x^*(t)) y(t) + w(t), & y(0) = 0, \\ u^*(t) + w(t) \in U \quad \text{for } \mathscr{L}^1\text{-almost every } t \in [0,T]. \end{cases} \tag{30}$$

As we shall see later on, the uniform coercivity estimate (29) can be interpreted as a strong positive-definiteness condition for the linearisation of $(\mathcal{P}_{\text{oc}})$ in a neighbourhood of $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$. We can now state main result of this section which can be found in [36, Theorem 5.2].

**Theorem 3** (Existence of locally optimal feedbacks for $(\mathcal{P}_{\text{oc}})$)**.** *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* $(\mathcal{P}_{\text{oc}})$*. Suppose that* $(\mathbf{H}_{\mathbf{oc}})$ *hold and that* $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$ *satisfies the uniform coercivity estimate* (29)-(30) *with constant* $\rho > 0$*.*

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

(i) $\bar{u}(\cdot, x^*(\cdot)) = u^*(\cdot)$.

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

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

$$\dot{x}(t) = f(t, x(t)) + \bar{u}(t, x(t)), \qquad x(\tau) = \xi, \tag{31}$$

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

(iv) *The map* $\hat{u}_{(\tau,\xi)} : t \in [\tau, T] \mapsto \bar{u}(t, \hat{x}_{(\tau,\xi)}(t))$ *is such that*

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

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

To better illustrate our subsequent use of Theorem 3 in the proof of our main result Theorem 1, we provide here an overview of the strategy used to prove it in [36], which is based on the earlier work [28]. We start our heuristic exposition by recalling the concept of *strong metric regularity* for a set-valued map.

**Definition 11** (Strong metric regularity)**.** *Let* $\mathcal{Y}, \mathcal{Z}$ *be two Banach spaces. A set-valued map* $\mathcal{G} : Y \rightrightarrows Z$ *is said to be strongly metrically regular at* $y^* \in Y$ *for* $z^* \in Z$ *if* $z^* \in \mathcal{F}(y^*)$ *and if there exists* $a, b > 0$ *and* $\kappa \geq 0$ *such that*

$$\mathcal{G}^{-1} : \text{B}(z^*, b) \to \text{B}(y^*, a)$$

*is single-valued and* $\kappa$*-Lipschitz.*

We now fix a time $\tau \in [0,T)$. In (27)-(28), we wrote the Pontryagin maximum principle for $(\mathcal{P}_{\text{oc}})$. Since $v \in U \mapsto H(t, x^*(t), p^*(t), v)$ is differentiable, we can reformulate the maximisation condition (28) as

$$\nabla_u H(t, x^*(t), p^*(t), u^*(t)) \in N_U(u^*(t)),$$

for $\mathscr{L}^1$-almost every $t \in [0,T]$, where $N_U(v)$ denotes the normal cone of convex analysis to $U$ at $v$. Then, any optimal Pontryagin triple $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$ can be seen as a solution of the *differential generalised inclusion*

$$0 \in F_\tau\big(x(\cdot), p(\cdot), u(\cdot)\big) + G_\tau\big(x(\cdot), p(\cdot), u(\cdot)\big), \tag{32}$$

where the maps $F_\tau : \mathcal{Y}_\tau \to \mathcal{Z}_\tau$ and $G_\tau : \mathcal{Y}_\tau \rightrightarrows \mathcal{Z}_\tau$ are defined by

$$F_\tau(x(\cdot), p(\cdot), u(\cdot)) = \begin{pmatrix} \dot{x}(\cdot) - f(\cdot, x(\cdot)) - u(\cdot) \\ x(\tau) - x^*(\tau) \\ \dot{p}(t) + \nabla_x H(\cdot, x(\cdot), p(\cdot), u(\cdot)) \\ p(T) + \nabla g(x(T)) \\ -\nabla_u H(\cdot, x(\cdot), p(\cdot), u(\cdot)) \end{pmatrix},$$

and $G_\tau(x(\cdot), p(\cdot), u(\cdot)) = (0,0,0,0, N^\infty_{\mathcal{U}}(u(\cdot)))^\top$. Here, we introduced the two Banach spaces

$$\begin{cases} \mathcal{Y}_\tau = W^{1,\infty}([\tau,T], \mathbb{R}^d) \times W^{1,\infty}([\tau,T], \mathbb{R}^d) \times L^\infty([\tau,T], U), \\ \mathcal{Z}_\tau = L^\infty([\tau,T], \mathbb{R}^d) \times \mathbb{R}^d \times L^\infty([\tau,T], \mathbb{R}^d) \times \mathbb{R}^d \times L^\infty([\tau,T], \mathbb{R}^d). \end{cases}$$

and $N^\infty_{\mathcal{U}}(u(\cdot)) = \{v \in L^\infty([0,T], U) \text{ s.t. } v(t) \in N_U(u(t)) \text{ for } \mathscr{L}^1\text{-a.e. } t \in [0,T]\}$.

In [36], it is proven that Theorem 3 can be derived as as a consequence of the *strong metric regularity* of $F_\tau(\cdot) + G_\tau(\cdot)$ at the restriction to $[\tau, T]$ of the Pontryagin triple $(x^*(\cdot), p^*(\cdot), u^*(\cdot))$ for 0, uniformly with respect to $\tau$. A standard strategy for proving metric regularity of mappings of the form of $F(\cdot) + G(\cdot)$ where $F(\cdot)$ is Fréchet-differentiable is to apply the *Robinson's inverse function theorem* (see e.g. [37, Theorem 5F.5]).

**Theorem 4** (Robinson's inverse function theorem). *Let $y^* \in \mathcal{Y}$ and $z^* \in \mathcal{G}(y^*)$. Suppose that $\mathcal{F} : \mathcal{Y} \to \mathcal{Z}$ is Fréchet differentiable at $y^*$. Then, the set-valued map $y \in \mathcal{Y} \mapsto \mathcal{F}(y) + \mathcal{G}(y)$ is strongly metrically regular at $y^*$ for $\mathcal{F}(y^*) + z^*$ if and only if the partially linearised mapping $y \mapsto \mathcal{F}(y^*) + \mathrm{D}\mathcal{F}(y^*)(y - y^*) + G(y)$ is strongly metrically regular at $y^*$ for $F(y^*) + z^*$.*

The strong metric regularity of (32) can therefore be equivalently derived from that of its partial linearisation involving the Fréchet differential of $F_\tau(\cdot)$. Notice that since in our problem the control and state are decoupled, there are no crossed derivatives in $(x, u)$. Now, the key point is to remark that the partially linearised generalized differential inclusion

$$0 \in \mathrm{D}F_\tau\big(x^*(\cdot), p^*(\cdot), u^*(\cdot)\big)\big(y(\cdot), q(\cdot), w(\cdot)\big) + G_\tau\big(y(\cdot), q(\cdot), w(\cdot)\big)$$

can be equivalently seen as the Pontryagin maximum principle for the LQ problem

$$(\mathcal{P}_{\mathrm{Lin}}) \quad \begin{cases} \min\limits_{w(\cdot)\in\mathcal{U}_\tau} \left[ \displaystyle\int_\tau^T \left( \frac{1}{2}\langle A(t)y(t), y(t)\rangle + \frac{1}{2}\langle B(t)w(t), w(t)\rangle \right) \mathrm{d}t + \frac{1}{2}\langle C(T)y(T), y(T)\rangle \right] \\ \quad \text{s.t.} \begin{cases} \dot{y}(t) = \mathrm{D}_x f(t, x^*(t))y(t) + w(t), \\ y(\tau) = 0. \end{cases} \end{cases}$$

Here, the set of admissible controls is defined by

$$\mathcal{U}_\tau = \{v \in L^2([\tau, T], U) \text{ s.t. } u^*(t) + v(t) \in U \text{ for } \mathscr{L}^1\text{-almost every } t \in [\tau, T]\}$$

and the cost functionals by

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

The coercivity estimate (29)-(30) is still valid on $[\tau, T]$ up to choosing $w(\cdot) \equiv 0$ on $[0, \tau]$, and it is indeed a second-order strict positive-definiteness condition for the linearised problem $(\mathcal{P}_{\mathrm{Lin}})$. In [28], it was proven that by applying Robinson's inverse function theorem, one can recover the strong metric regularity of (32) uniformly with respect to $\tau$, which was in turn used in [36] to prove Theorem 3.

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

In this section, we recall some results concerning continuity equations and optimal control problems in Wasserstein spaces written in the general form

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

We make the following working assumption on the data of problem $(\mathcal{P})$.

**Hypotheses (H).**

**(H1)** *The set of control values $U \subset \mathbb{R}^d$ is a convex and compact set containing a neighbourhood of the origin.*

**(H2)** *The control cost $v \mapsto \psi(v) \in [0, +\infty]$ is radial, $C^{2,1}_{\mathrm{loc}}$-regular, strictly convex, and such that $\psi(0) = 0$.*

**(H3)** *The non-local velocity field $(t, x, \mu) \mapsto v[\mu](t, x) \in \mathbb{R}^d$ is Lipschitz with respect to $t \in [0, T]$ and continuous in the product $|\cdot| \times W_2$-topology with respect to $(x, \mu) \in \mathbb{R}^d \times \mathscr{P}_c(\mathbb{R}^d)$. For all times $t \in [0, T]$, it is such that*

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

*for a given constant $M > 0$ and any $(x, \mu) \in \mathbb{R}^d \times \mathscr{P}_c(\mathbb{R}^d)$. It further satisfies the Cauchy-Lipschitz properties*

$$\begin{cases} |v[\mu](t, x) - v[\mu](t, y)| \leq L_1^K |x - y|, \\ \|v[\mu](t, \cdot) - v[\nu](t, \cdot)\|_{C^0(K, \mathbb{R}^d)} \leq L_2^K W_2(\mu, \nu), \end{cases}$$

*on any compact set $K \subset \mathbb{R}^d$ and for any pairs $x, y \in K$ and $\mu, \nu \in \mathscr{P}(K)$.*

**(H4)** *The map* $v[\cdot](t,x)$ *is* $C^{2,1}_{\mathrm{loc}}$*-Wasserstein regular in the sense of Definition 9, uniformly with respect to* $(t,x) \in [0,T] \times K$ *where* $K \subset \mathbb{R}^d$ *is compact.*

**(H5)** *The running cost* $(t,\mu) \mapsto L(t,\mu)$ *is Lipschitz with respect to* $t \in [0,T]$ *and* $C^{2,1}_{\mathrm{loc}}$*-Wasserstein regular with respect to* $\mu \in \mathscr{P}_c(\mathbb{R}^d)$ *in the sense of Definition 9.*

**(H6)** *The final cost* $\mu \mapsto \varphi(\mu)$ *is* $C^{2,1}_{\mathrm{loc}}$*-Wasserstein regular in the sense of Definition 9.*

Let it be noted that the strong requirements of $C^{2,1}_{\mathrm{loc}}$-Wasserstein regularity on the functionals involved in the problem are not classical, since the well-posedness results e.g. of [54] are proven under mere Lipschitz regularity in the measure variables.

We present in Section 4.1 two classical well-posedness results for continuity equations formulated in Wasserstein spaces. We further state in Section 4.2 a powerful existence result of so-called *mean-field optimal controls* for an adequate variant of problem $(\mathcal{P})$. The latter is a reformulation of the main result of [43], which was derived under more general assumptions than our working hypotheses **(H)**.

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

Given a positive constant $T > 0$, we denote by $\lambda = \frac{1}{T}\mathscr{L}^1_{\llcorner [0,T]}$ the normalized Lebesgue measure on $[0,T]$. For any $p \geq 1$, a narrowly continuous curve of measures $\mu(\cdot)$ in $\mathscr{P}_p(\mathbb{R}^d)$ can be uniquely lifted to a measure $\tilde{\mu} \in \mathscr{P}_p([0,T] \times \mathbb{R}^d)$ through the disintegration formula $\tilde{\mu} = \int_{[0,T]} \mu(t)\mathrm{d}\lambda(t)$ introduced in Theorem 2. We say that a narrowly continuous curve of measure $t \mapsto \mu(t) \in \mathscr{P}_p(\mathbb{R}^d)$ solves a *continuity equation* with initial condition $\mu^0 \in \mathscr{P}_p(\mathbb{R}^d)$ associated to the Borel velocity field $\boldsymbol{w} \in L^p([0,T] \times \mathbb{R}^d, \mathbb{R}^d; \tilde{\mu})$ provided that

$$\begin{cases} \partial_t \mu(t) + \nabla \cdot (\boldsymbol{w}(t,\cdot)\mu(t)) = 0, \\ \mu(0) = \mu^0. \end{cases} \tag{33}$$

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

$$\int_0^T \int_{\mathbb{R}^d} \Big( \partial_t \xi(t,x) + \langle \nabla_x \xi(t,x), \boldsymbol{w}(t,x) \rangle \Big) \mathrm{d}\mu(t)(x)\mathrm{d}t = 0 \tag{34}$$

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

We state in the following theorem a general existence result for solutions of continuity equations of the form (33) under mere $L^p$-integrability of the driving velocity field. We refer the reader to the seminal papers [2, 35] as well as to [5, Chapter 8].

**Theorem 5** (Superposition principle)**.** *Let* $\mu(\cdot) \in C^0([0,T], \mathscr{P}_p(\mathbb{R}^d))$ *and* $\boldsymbol{w}(\cdot,\cdot) \in L^p([0,T] \times \mathbb{R}^d, \mathbb{R}^d; \tilde{\mu})$ *be a Borel vector field satisfying the integrability bound*

$$\int_0^T \int_{\mathbb{R}^d} \frac{|\boldsymbol{w}(t,x)|}{1+|x|} \mathrm{d}\mu(t)(x)\mathrm{d}t < +\infty.$$

*Then,* $\mu(\cdot)$ *is a solution of* (33) *associated to* $v(\cdot,\cdot)$ *if and only if there exists a probability measure* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times C^0([0,T], \mathbb{R}^d))$ *such that the following holds.*

(i) $\boldsymbol{\eta}$ *is concentrated on the set of pairs* $(x, \gamma(\cdot)) \in \mathbb{R}^d \times \mathrm{AC}([0,T], \mathbb{R}^d)$ *such that* $\dot{\gamma}(t) = \boldsymbol{w}(t, \gamma(t))$ *for* $\mathscr{L}^1$*-almost every* $t \in [0,T]$ *and* $\gamma(0) = x$*.*

(ii) *It holds that* $\mu(t) = (\mathrm{e}_t)_{\#}\boldsymbol{\eta}$ *where for all times* $t \in [0,T]$ *we introduced the evaluation map* $\mathrm{e}_t : (x, \gamma(\cdot)) \in \mathbb{R}^d \times \mathrm{AC}([0,T], \mathbb{R}^d) \mapsto \gamma(t) \in \mathbb{R}^d$*.*

Taking in particular $p = 1$ and a non-local velocity field of the form $\boldsymbol{w} : (t,x) \mapsto v[\mu(t)](t,x) + \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\tilde{\mu}}(t,x)$, we recover a notion of solution for the Cauchy problem on which problem $(\mathcal{P})$ is formulated. In Theorem 6 below, we state another existence result derived in [54] and concerned with classical well-posedness for non-local transport equations under stronger regularity assumptions.

**Theorem 6** (Well-posedness of transport equation)**.** *Let* $\mu \in \mathscr{P}_c(\mathbb{R}^d) \mapsto v[\mu](\cdot,\cdot) \in L^1([0,T], C^0(\mathbb{R}^d, \mathbb{R}^d))$ *be a non-local Borel velocity field satisfying hypothesis* **(H3)***. Then, there exists a unique solution* $\mu(\cdot) \in \mathrm{Lip}([0,T], \mathscr{P}_c(\mathbb{R}^d))$ *of* (33) *driven by the non-local vector field* $v[\cdot](\cdot,\cdot)$*. Furthermore, there exist constants* $R_T, L_T > 0$ *such that*

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

*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 both the measure theoretic formulation and its sequence of approximating problems. We subsequently recall a powerful existence result derived in [43] for general multi-agent optimal control problems formulated in Wasserstein spaces.

Let us start by fixing an integer $N \geq 1$, an initial datum $\boldsymbol{x}_N^0 \in (\mathbb{R}^d)^N$ and the associated discrete measure $\mu_N^0 = \mu[\boldsymbol{x}_N^0]$ as defined in Section 2.3. As already sketched in the introduction, we will naturally consider the family of discrete problems

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

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

$$\boldsymbol{v}_N[\boldsymbol{x}](\cdot,\cdot) = v[\mu[\boldsymbol{x}]](\cdot,\cdot), \quad \boldsymbol{L}_N(\cdot,\boldsymbol{x}) = L(t,\mu[\boldsymbol{x}]), \quad \boldsymbol{\varphi}_N(\boldsymbol{x}) = \varphi(\mu[\boldsymbol{x}]),$$

in the sense of Definition 8. It can be checked that as a consequence of hypotheses **(H)** displayed in Section 5, the problems $(\mathcal{P}_N)$ satisfy the set of hypotheses **($\mathbf{H_{oc}}$)** of Section 3. We can moreover deduce the following lemma directly from Proposition 5.

**Lemma 2** (Existence of solutions for $(\mathcal{P}_N)$)**.** *Under hypotheses* **(H)**, *there exist optimal control-trajectory pairs* $(\boldsymbol{u}_N^*(\cdot),\boldsymbol{x}_N^*(\cdot)) \in \mathcal{U}_N \times \mathrm{Lip}([0,T],(\mathbb{R}^d)^N)$ *solution of* $(\mathcal{P}_N)$ *for all* $N \geq 1$.

We proceed by recasting problem $(\mathcal{P})$ into a framework which also encompasses the sequence of problems $(\mathcal{P}_N)$. Let us consider a narrowly continuous curve of measures $\mu(\cdot) \in C^0([0,T],\mathscr{P}_1(\mathbb{R}^d))$ and its canonical lift $\tilde{\mu} \in \mathscr{P}_1([0,T]\times\mathbb{R}^d)$. 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 $\lambda$-almost unique measurable 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 disintegration for vector-valued measures recalled in Theorem 2. Bearing this in mind, problem $(\mathcal{P})$ can be relaxed as

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

where we denote the set of *generalized measure controls* by $\mathscr{U} = \mathcal{M}([0,T]\times\mathbb{R}^d,U)$ and where the map $\boldsymbol{\sigma} \in \mathcal{M}(\mathbb{R}^d,U) \mapsto \Psi(\boldsymbol{\sigma}|\mu) \in [0,+\infty]$ is defined by

$$\Psi(\boldsymbol{\sigma}|\mu) = \begin{cases} \int_{\mathbb{R}^d} \psi\left(\frac{\mathrm{d}\boldsymbol{\sigma}}{\mathrm{d}\mu}(x)\right)\mathrm{d}\mu(x) & \text{if } \boldsymbol{\sigma} \ll \mu, \\ +\infty & \text{otherwise.} \end{cases} \tag{35}$$

This type of relaxation appears frequently in variational problems involving integral functional on measures. Indeed, functionals of the form of $\Psi(\cdot|\tilde{\mu})$ as defined in (35) possess a wide range of useful features, such as weak-$^*$ lower-semicontinuity, while also imposing an absolute continuity property on their argument. We refer the reader to [5, Section 9.4] for a detailed account on their properties.

Consider now an optimal pair control-trajectory $(\boldsymbol{u}_N^*(\cdot),\boldsymbol{x}_N^*(\cdot)) \in \mathcal{U}_N \times \mathrm{Lip}([0,T],(\mathbb{R}^d)^N)$ for $(\mathcal{P}_N)$. One can canonically associate to any such solution the discrete control-trajectory measure pairs $(\boldsymbol{\nu}_N^*,\mu_N^*(\cdot)) \in \mathscr{U} \times \mathrm{Lip}([0,T],\mathscr{P}_N(\mathbb{R}^d))$ defined by

$$\mu_N^*(\cdot) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i^*(\cdot)}, \qquad \boldsymbol{\nu}_N^* = \int_{[0,T]} \left( \frac{1}{N}\sum_{i=1}^N u_i^*(t)\delta_{x_i^*(t)} \right)\mathrm{d}\lambda(t). \tag{36}$$

In the following theorem, we state a condensed version of the main result of [43], which shows that this relaxation allows to prove the convergence of the discrete problems $(\mathcal{P}_N)$ towards $(\mathcal{P})$. This convergence result has to be understood both in terms of mean-field limit of the functional describing the dynamics and of Γ-convergence of the corresponding minimizers.

**Theorem 7** (Existence of mean-field optimal controls for $(\mathcal{P})$)**.** *Let $\mu^0 \in \mathscr{P}_c(\mathbb{R}^d)$, $(\mu_N^0) \subset \mathscr{P}_c(\mathbb{R}^d)$ be a sequence of empirical measures associated with $(\boldsymbol{x}_N^0) \subset (\mathbb{R}^d)^N$ such that $W_1(\mu_N^0, \mu^0) \to 0$, and assume that hypotheses* **(H)** *hold. For any $N \geq 1$, denote by $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot)) \in \mathcal{U}_N \times \mathrm{Lip}([0,T], (\mathbb{R}^d)^N)$ an optimal pair control-trajectory for $(\mathcal{P}_N)$ and by $(\boldsymbol{\nu}_N^*, \mu_N^*(\cdot)) \in \mathscr{U} \times \mathrm{Lip}([0,T], \mathscr{P}_N(\mathbb{R}^d))$ the corresponding pair of measure control-trajectory defined as in* (36)*.*

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

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

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

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

*is optimal for problem $(\mathcal{P})$ and solves* (33) *in the superposition sense.*

# 5 Coercivity estimate and proof of Theorem 1

In this section, we prove the main result of this article stated in Theorem 1. We suppose that hypotheses **(H)** of Section 4 hold, along with the following additional *mean-field coercivity* assumption.

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

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

*holds along any solution pair $(\boldsymbol{w}(\cdot), \boldsymbol{y}(\cdot)) \in L^2([0,T], (\mathbb{R}^d)^N) \times W^{1,2}([0,T], (\mathbb{R}^d)^N)$ of the linearised control-state equations*

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

Our argument is split into three steps. In Step 1, we write a Pontryagin Maximum Principle adapted to the mean-field structure of the 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 3 and $\mathbf{(OCP}_N\mathbf{)}$. We then show in Step 3 that this sequence of control maps is compact in a suitable weak topology preserving its regularity in space, and that its limit points coincide with the mean-field optimal control introduced in Theorem 7.

**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 $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot))$ for $(\mathcal{P}_N)$. Our analysis is based on the finite-dimensional Pontryagin maximum principle applied to $(\mathbb{R}^d)^N$,rewritten as a Hamiltonian flow with respect to the rescaled mean-field inner product $\langle \cdot, \cdot \rangle_N$.

**Proposition 6** (Characterization of the solutions of $(\mathcal{P}_N)$)**.** *Let $N \geq 1$ and $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot)) \in L^\infty([0,T], U^N) \times \mathrm{Lip}([0,T], (\mathbb{R}^d)^N))$ be an optimal pair control-trajectory 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

$$\begin{cases} \dot{\boldsymbol{x}}_N^*(t) = \quad \mathbf{Grad}_{\boldsymbol{r}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)), \quad \boldsymbol{x}_N^*(0) = \boldsymbol{x}_N^0, \\ \dot{\boldsymbol{r}}_N^*(t) = -\mathbf{Grad}_{\boldsymbol{x}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)), \quad \boldsymbol{r}_N^*(T) = -\mathbf{Grad}_{\boldsymbol{x}}\, \boldsymbol{\varphi}_N(\boldsymbol{x}_N^*(T)), \\ \boldsymbol{u}_N^*(t) \in \underset{\boldsymbol{v} \in U^N}{\mathrm{argmax}}\, \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{v}) \qquad \text{for } \mathscr{L}^1\text{-almost every } t \in [0,T], \end{cases} \tag{37}$$

*where the* mean-field Hamiltonian $\mathbb{H}_N(\cdot, \cdot, \cdot, \cdot)$ *of the system is defined by*

$$\mathbb{H}_N(t, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{u}) = \frac{1}{N} \sum_{i=1}^N \Big( \langle r_i, \boldsymbol{v}_N[\boldsymbol{x}](t, x_i) + u_i \rangle - \psi(u_i) \Big) - \boldsymbol{L}_N(t, \boldsymbol{x}) \tag{38}$$

*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 exists uniform constants $R_T, L_T > 0$ which are independent from $N$, such that*

$$\mathrm{Graph}\Big( (\boldsymbol{x}^*(\cdot), \boldsymbol{r}^*(\cdot)) \Big) \subset [0,T] \times \overline{\mathrm{B}(0, R_T)^{2N}}, \quad \mathrm{Lip}\Big( (\boldsymbol{x}^*(\cdot), \boldsymbol{r}^*(\cdot))\,; [0,T] \Big) \leq L_T. \tag{39}$$

*Proof.* By an application of the standard PMP to $(\mathcal{P}_N)$ (see for instance [29, Theorem 22.2]), there exists a family of costate variables $\{p_i^*(\cdot)\}_{i=1}^N = \boldsymbol{p}^*(\cdot) \in \mathrm{Lip}([0,T],(\mathbb{R}^d)^N)$ such that

$$\begin{cases} \dot{x}_i^*(t) = \quad \nabla_{p_i}\mathcal{H}_N(t,\boldsymbol{x}^*(t),\boldsymbol{p}^*(t),\boldsymbol{u}^*(t)), \;\; x_i^*(0) = x_i^0, \\ \dot{p}_i^*(t) = -\nabla_{x_i}\mathcal{H}_N(t,\boldsymbol{x}^*(t),\boldsymbol{p}^*(t),\boldsymbol{u}^*(t)), \;\; p_i^*(T) = -\nabla_{x_i}\boldsymbol{\varphi}_N(\boldsymbol{x}^*(T)), \\ u_i^*(t) \in \underset{v\in U}{\mathrm{argmax}} \; \left[\langle p_i^*(t),v\rangle - \frac{1}{N}\psi(v)\right]. \end{cases} \tag{40}$$

Here, the classical Hamiltonian $\mathcal{H}_N(\cdot,\cdot,\cdot,\cdot)$ of the system is defined by

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

for every $(t,\boldsymbol{x},\boldsymbol{p},\boldsymbol{u}) \in [0,T]\times(\mathbb{R}^d)^N\times(\mathbb{R}^d)^N\times U^N$. By introducing the variables $r_i^*(\cdot) = Np_i^*(\cdot)$, it holds that

$$\dot{x}_i^*(t) = N\nabla_{r_i}\mathbb{H}_N(t,\boldsymbol{x}^*(t),\boldsymbol{r}^*(t),\boldsymbol{u}^*(t)) = \mathbf{Grad}_{r_i}\,\mathbb{H}_N(t,\boldsymbol{x}^*(t),\boldsymbol{r}^*(t),\boldsymbol{u}^*(t)), \tag{41}$$

as well as

$$\dot{r}_i^*(t) = -N\nabla_{x_i}\mathbb{H}_N(t,\boldsymbol{x}^*(t),\boldsymbol{r}^*(t),\boldsymbol{u}^*(t)) = -\mathbf{Grad}_{x_i}\,\mathbb{H}_N(t,\boldsymbol{x}^*(t),\boldsymbol{r}^*(t),\boldsymbol{u}^*(t)), \tag{42}$$

and

$$r_i^*(T) = -N\nabla_{x_i}\boldsymbol{\varphi}(\boldsymbol{x}^*(T)) = -\mathbf{Grad}_{x_i}\,\boldsymbol{\varphi}(\boldsymbol{x}^*(T)), \tag{43}$$

as a consequence of Proposition 4. Moreover, it can be seen easily from the maximisation condition in (40) that $u_i^*(t) \in \mathrm{argmax}\,[\langle r_i^*(t),v\rangle - \psi(v)]$. Merging this condition with (41), (42) and (43), we recover the desired claim that $(\boldsymbol{x}^*(\cdot),\boldsymbol{r}^*(\cdot),\boldsymbol{u}^*(\cdot))$ satisfies the mean-field Pontryagin Maximum Principle (37) associated to the mean-field Hamiltonian $\mathbb{H}_N(\cdot,\cdot,\cdot,\cdot)$ for all times $t\in[0,T]$.

In the spirit of [10, 55], we introduce the discrete $L^\infty$-radius function

$$X_N : t\in[0,T] \mapsto \max_{i\in\{1,\dots,N\}} |x_i^*(t)|.$$

By Danskin's Theorem (see e.g. [32]), the map $X_N(\cdot)$ is differentiable $\mathscr{L}^1$-almost everywhere and it holds that

$$\begin{aligned} X_N(t)X_N'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left[\tfrac{1}{2}X_N^2(t)\right] &\le \langle x_{I(t)}^*(t), \dot{x}_{I(t)}^*(t)\rangle \\ &\le |x_{I(t)}^*(t)|\left(M\left(1+|x_{I(t)}^*(t)| + |\boldsymbol{x}_N^*(t)|_N\right) + L_U\right) \end{aligned}$$

by **(H1)**, **(H3)** and Cauchy-Schwarz inequality. Here, $I(t) \in \mathrm{argmax}_{i\in\{1,\dots,N\}}|x_i^*(t)|$ is any of the indices realising the value of $X_N(t)$ for $\mathscr{L}^1$-almost every $t\in[0,T]$. Remarking now that $|\boldsymbol{x}_N^*(t)|_N \le X_N(t)$, we recover

$$X_N'(t) \le L_U + M(1+2X_N(t))$$

so that by Grönwall Lemma, there exists a constant $R_T^1 > 0$ depending on $\mathrm{supp}(\mu^0)$, $T$, $M$ and $L_U$ such that

$$\sup_{t\in[0,T]} |x_i^*(t)| \le R_T^1, \tag{44}$$

for all $i\in\{1,\dots,N\}$. Plugging this bound into (41), we recover the existence of a constant $L_T^1 > 0$ such that

$$\mathrm{Lip}(x_i^*(\cdot);[0,T]) \le L_T^1, \tag{45}$$

for all $i\in\{1,\dots,N\}$.

We now prove a similar estimate on the costate variable $(\boldsymbol{r}_N^*(\cdot))$. By invoking the $C^{2,1}_{\mathrm{loc}}$-MF regularity assumptions of **(H4)-(H6)** as well as the uniform bound (44)-(45), we can derive by Grönwall Lemma that

$$\sup_{t\in[0,T]} |r_i^*(t)| \le C'\Big(T + |\mathbf{Grad}_{x_i}\,\boldsymbol{\varphi}_N(\boldsymbol{x}_N^*(T))|\Big)e^{C'T} \tag{46}$$

for all $i\in\{1,\dots,N\}$, where $C'>0$ is a given uniform constant, independent from $N$. By hypothesis **(H6)**, we know that $\boldsymbol{\varphi}_N(\cdot)$ is locally Lipschitz over $(\mathbb{R}^d)^N$ with a uniform constant on products of compact sets, so that

$$\sup_{t\in[0,T]} |r_i^*(t)| \le R_T^2, \qquad \mathrm{Lip}(r_i^*(\cdot);[0,T]) \le L_T^2, \tag{47}$$

for all $i\in\{1,\dots,N\}$ and for some positive constants $R_T^2, L_T^2 > 0$. Subsequently, there exists uniform constants $R_T, L_T > 0$ which are again independent from $N$, such that

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

which concludes the proof of Proposition 6. □

We end this 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 $N$.

**Corollary 1.** *Let $\mathcal{K} = [0,T] \times \overline{\mathrm{B}(0,R_T)^{2N}} \times U^N$ where $R_T > 0$ is defined as in (39). Then, there exists a constant $\mathcal{L}_{\mathcal{K}} > 0$ such that the $C^{2,1}$-norms of the maps $\mathbb{H}_N(t,\cdot,\boldsymbol{r},\cdot)$, $\boldsymbol{L}_N(t,\cdot)$, $\frac{1}{N}\sum_{i=1}^N \psi(\cdot)$ and $\boldsymbol{\varphi}_N(\cdot)$ with respect to the variables $(\boldsymbol{x},\boldsymbol{u})$ are bounded by $\mathcal{L}_{\mathcal{K}}$ over $\mathcal{K}$, uniformly with respect to $(t,\boldsymbol{r}) \in [0,T] \times \overline{B(0,R_T)^N}$.*

*Proof.* This result follows directly from the $C^{2,1}_{\mathrm{loc}}$-Wasserstein regularity hypotheses **(H3)-(H6)** on the datum of $(\mathcal{P}_N)$ along with the uniform compactness of the optimal Pontryagin triples derived in Proposition 6. □

**Step 2 : Construction of a Lipschitz-in-space optimal controls for** $(\mathcal{P}_N)$ In this second step, we associate to any solution $(\boldsymbol{u}_N^*(\cdot),\boldsymbol{x}_N^*(\cdot))$ of $(\mathcal{P}_N)$ a mean-field optimal control map $u_N^* \in L^\infty([0,T],\mathrm{Lip}(\mathbb{R}^d,U))$. We have seen in Proposition 6 that any optimal pair $(\boldsymbol{u}_N^*(\cdot),\boldsymbol{x}_N^*(\cdot))$ satisfies a PMP adapted to the mean-field structure of $(\mathcal{P}_N)$. In Proposition 7 below, we show that this result along with the coercivity assumption $\mathbf{(CO}_N\mathbf{)}$ and Theorem 3 allows us to build a sequence of optimal controls $(u_N(\cdot,\cdot)) \subset L^\infty([0,T],\mathrm{Lip}(\mathbb{R}^d,U))$ which Lipschitz constants are uniformly bounded with respect to $N \geq 1$.

**Proposition 7** (Existence of mean-field locally optimal Lipschitz feedback). *Let $(\boldsymbol{u}_N^*(\cdot),\boldsymbol{x}_N^*(\cdot)) \in \mathcal{U}_N \times \mathrm{Lip}([0,T],\overline{B(0,R_T)^N})$ be an optimal pair control-trajectory for $(\mathcal{P}_N)$ and assume that hypotheses* **(H)** *hold. Then, there exists a Lipschitz map $u_N^*(\cdot,\cdot) \in \mathrm{Lip}([0,T]\times\mathbb{R}^d,U)$ such that $u_N^*(t,x_i(t)) = u_i^*(t)$ for all times $t \in [0,T]$ and which Lipschitz constant $\mathcal{L}_U$ with respect to the space variable is independent from $N$.*

*Proof.* The first step of this proof is to apply Theorem 3 to $(\mathcal{P}_N)$ seen as an optimal control problem in the rescaled Euclidean space $\big((\mathbb{R}^d)^N,\langle\cdot,\cdot\rangle_N\big)$ introduced in (15). As it was already mentioned in the proof of Proposition 6, $(\mathcal{P}_N)$ satisfies the structural assumptions $\mathbf{(H_{oc})}$ of Section 3. Given a rescaled covector $\boldsymbol{r}_N^*(\cdot)$ associated to $(\boldsymbol{u}_N^*(\cdot),\boldsymbol{x}^*(\cdot))$ via (37), the mean-field Pontryagin triple $(\boldsymbol{x}_N^*(\cdot),\boldsymbol{r}_N^*(\cdot),\boldsymbol{u}_N^*(\cdot))$ is bounded in $L^\infty([0,T],(\mathbb{R}^{2d})^N\times U^N)$ uniformly with respect to $N$ as a consequence of **(H1)** and Proposition 6. By Corollary 1, the $C^{2,1}$-norms of the datum of $(\mathcal{P}_N)$, defined in the sense of (19)-(20), are uniformly bounded over $\mathcal{K} = [0,T]\times\overline{B(0,R_T)^{2N}}\times U^N$ by a constant $\mathcal{L}_{\mathcal{K}} > 0$.

Similarly to what was presented in Section 3, the mean-field Pontryagin optimality system (37) can be written as the dynamical differential inclusion

$$0 \in \boldsymbol{F}_\tau^N(\boldsymbol{x}(\cdot),\boldsymbol{r}(\cdot),\boldsymbol{u}(\cdot)) + \boldsymbol{G}_\tau^N(\boldsymbol{x}(\cdot),\boldsymbol{r}(\cdot),\boldsymbol{u}(\cdot)), \tag{48}$$

for any $\tau \in [0,T)$. Here, the mappings $\boldsymbol{F}_\tau^N : \mathcal{Y}_\tau^N \to \mathcal{Z}_\tau^N$ and $\boldsymbol{G}_\tau^N : \mathcal{Y}_\tau^N \rightrightarrows \mathcal{Z}_\tau^N$ are respectively defined by

$$\boldsymbol{F}_\tau^N(\boldsymbol{x}(\cdot),\boldsymbol{r}(\cdot),\boldsymbol{u}(\cdot)) = \begin{pmatrix} \dot{\boldsymbol{x}}(\cdot) - \boldsymbol{V}_N[\boldsymbol{x}(\cdot)](\cdot,\boldsymbol{x}(\cdot)) - \boldsymbol{u}(\cdot) \\ x_i(\tau) - x_i^*(\tau) \\ \dot{\boldsymbol{r}}(\cdot) + \mathbf{Grad}_{\boldsymbol{x}}\,\mathbb{H}_N(\cdot,\boldsymbol{x}(\cdot),\boldsymbol{r}(\cdot),\boldsymbol{u}(\cdot)) \\ \boldsymbol{r}(T) + \mathbf{Grad}_{\boldsymbol{x}}\,\boldsymbol{\varphi}(\boldsymbol{x}(T)) \\ -\mathbf{Grad}_{\boldsymbol{u}}\,\mathbb{H}_N(\cdot,x(\cdot),p(\cdot),u(\cdot)) \end{pmatrix}, \tag{49}$$

where

$$\boldsymbol{V}_N[\boldsymbol{x}(\cdot)](t,\boldsymbol{x}(\cdot)) \equiv (\boldsymbol{v}_N[\boldsymbol{x}(\cdot)](t,x_i(\cdot)))_{1\leq i\leq N} \in (\mathbb{R}^d)^N,$$

and $\boldsymbol{G}_\tau^N(\boldsymbol{x}(\cdot),\boldsymbol{r}(\cdot),\boldsymbol{u}(\cdot))= (0,0,0,0,N_{\mathcal{U}_N}^\infty(\boldsymbol{u}(\cdot)))^\top$. The two Banach spaces $\mathcal{Y}_\tau^N, \mathcal{Z}_\tau^N$ are defined in this context by

$$\begin{aligned}\mathcal{Y}_\tau^N &= W^{,1,\infty}([\tau,T],(\mathbb{R}^d)^N)\times W^{,1,\infty}([\tau,T],(\mathbb{R}^d)^N)\times L^\infty([\tau,T],U^N),\\ \mathcal{Z}_\tau^N &= L^\infty([\tau,T],(\mathbb{R}^d)^N)\times(\mathbb{R}^d)^N\times L^\infty([\tau,T],(\mathbb{R}^d)^N)\times(\mathbb{R}^d)^N\times L^\infty([\tau,T],(\mathbb{R}^d)^N).\end{aligned}$$

Following [28], we now compute the first-order variation of the map $\boldsymbol{F}_\tau^N(\cdot)$ with respect to the adapted differential structure introduced in Section 2.3. Let $(\boldsymbol{y}(\cdot),\boldsymbol{s}(\cdot),\boldsymbol{w}(\cdot)) \in \mathcal{Y}_0^N$, $i \in \{1,\dots,N\}$ and $t \in [0,T]$. One has that

$$\begin{aligned}\boldsymbol{v}_N[\boldsymbol{x}+\boldsymbol{y}](t,x_i+sy_i) &= \boldsymbol{v}_N[\boldsymbol{x}](t,x_i) + \mathrm{D}_x\boldsymbol{v}_N[\boldsymbol{x}](t,x_i)y_i\\ &\quad+\frac{1}{N}\sum_{j=1}^N \mathbf{D}_{x_j}\boldsymbol{v}_N[\boldsymbol{x}](t,x_i)y_j + o(|y_i|) + o(|\boldsymbol{y}|_N),\end{aligned} \tag{50}$$

where $\mathbf{D}_{x_j}\boldsymbol{v}_N[\boldsymbol{x}](t,x_i)$ is the matrix which rows are the mean-field gradients with respect to $x_j$ of the components $(\boldsymbol{v}_N^k[\boldsymbol{x}](t,x_i))_{1\leq k\leq d}$. Analogously, it holds that

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

as well as

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

and

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

as a consequence of the chain rule of Proposition 4. Here for convenience, we used the matrix representation (26) for mean-field Hessians in $(\mathbb{R}^d)^N$ introduced in Remark 3.

It is again possible to interpret the partial linearisation of the differential generalized inclusion (48) as the Pontryagin maximum principle for the linear-quadratic problem

$$\begin{cases}\displaystyle\min_{\boldsymbol{w}\in\mathcal{U}_\tau^N}\left[\int_\tau^T\left(\frac{1}{2}\langle \boldsymbol{A}(t)\boldsymbol{y}(t),\boldsymbol{y}(t)\rangle_N+\frac{1}{2}\langle \boldsymbol{B}(t)\boldsymbol{w}(t),\boldsymbol{w}(t)\rangle_N\right)\mathrm{d}t+\frac{1}{2}\langle \boldsymbol{C}(T)\boldsymbol{y}(T),\boldsymbol{y}(T)\rangle_N\right]\\ \text{s.t.}\begin{cases}\dot{y}_i(t) = \mathrm{D}_x\boldsymbol{v}_N[\boldsymbol{x}_N^*(t)](t,x_i^*(t))y_i(t) + \displaystyle\frac{1}{N}\sum_{j=1}^N \mathbf{D}_{x_j}\boldsymbol{v}_N[\boldsymbol{x}_N^*(t)](t,x_i^*(t))y_j(t),\\ y_i(\tau)=0,\end{cases}\end{cases}$$

where the set of admissible controls is defined by

$$\mathcal{U}_\tau^N = \left\{v\in L^\infty([\tau,T],U^N) \text{ s.t. } \boldsymbol{u}_N^*(t)+\boldsymbol{w}(t)\in U^N \text{ for } \mathscr{L}^1\text{-a.e. } t\in[\tau,T]\right\},$$

and the cost functionals by

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

Moreover, we assumed in $(\mathbf{CO}_N)$ that there exists a constant $\rho_T$, which is independent from $N$, such that the mean-field coercivity estimate

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

holds for any admissible pair $(\boldsymbol{w}(\cdot),\boldsymbol{y}(\cdot))\in L^2([0,T],(\mathbb{R}^d)^N)\times W^{1,2}([0,T],(\mathbb{R}^d)^N)$ for $(\mathcal{P}_{\mathrm{Lin}}^N)$.

We can therefore apply Theorem 3 to $(\mathcal{P}_N)$ and recover the existence of a neighbourhood $\mathscr{N}\subset[0,T]\times(\mathbb{R}^d)^N$ of $\mathrm{Graph}(\boldsymbol{x}^*(\cdot))$ along with that of a locally optimal Lipschitz feedback $\tilde{\boldsymbol{u}}(\cdot,\cdot)$ defined over $\mathscr{N}\cap\big([0,T]\times\overline{B(0,R_T)^N}\big)$ which Lipschitz constant $\mathcal{L}_U$ depends only on the structural constant $\mathcal{L}_{\mathcal{K}}$ introduced in Corollary 1 and on the coercivity constant $\rho_T$ introduced in $(\mathbf{CO}_N)$. In particular, $\mathcal{L}_U$ is independent from $N$.

For any $i\in\{1,\dots,N\}$, we associate to $x_i^*(\cdot)$ the projected control maps $\tilde{u}_i : \mathscr{N}_i := \pi^i(\mathscr{N})\to\mathbb{R}^d$ defined by

$$\tilde{u}_i(t,x) = \tilde{\boldsymbol{u}}_i(t,\hat{\boldsymbol{x}}_i^x(t)),$$

where $\hat{\boldsymbol{x}}_i^x(t) = (x_1^*(t),\dots,x_{i-1}^*(t),x,x_{i+1}^*(t),\dots,x_N^*(t))$ denotes the element in $(\mathbb{R}^d)^N$ which has all its components matching that of $\boldsymbol{x}^*(t)$ except the $i$-th one which is free and equal to $x$. By construction, each $\tilde{u}_i(\cdot,\cdot)$ defines a locally optimal feedback in the neighbourhood $\mathscr{N}_i$ of $\mathrm{Graph}(x_i^*(\cdot))$. Furthermore, we can derive the following uniform estimate for the projected control maps

$$\begin{aligned}|\tilde{u}_i(t,y)-\tilde{u}_i(t,x)| &= |\tilde{\boldsymbol{u}}_i(t,\hat{\boldsymbol{x}}_i^y(t))-\tilde{\boldsymbol{u}}_i(t,\hat{\boldsymbol{x}}_i^x(t))|\\ &\leq \Big(\sum_{j=1}^N|\tilde{\boldsymbol{u}}_j(t,\hat{\boldsymbol{x}}_i^y(t))-\tilde{\boldsymbol{u}}_j(t,\hat{\boldsymbol{x}}_i^x(t))|^2\Big)^{1/2} \leq \sqrt{N}\mathcal{L}_U|\hat{\boldsymbol{x}}_i^y(t)-\hat{\boldsymbol{x}}_i^x(t)|_N = \mathcal{L}_U|y-x|,\end{aligned}$$

so that we recover the uniform Lipschitz estimate

$$|\tilde{u}_i(t,y)-\tilde{u}_i(t,x)| \leq \mathcal{L}_U|y-x|,$$

for all $(t,x,y)\in[0,T]\times\pi_{\mathbb{R}^d}(\mathscr{N}_i)^2$. This shows that the projected optimal control $\tilde{u}_i(\cdot,\cdot)$ maps are Lipschitz-regular in space uniformly with respect to $N$.

Therefore, each $\tilde{u}_i(\cdot,\cdot)$ can be defined unequivocally on a closed neighbourhood of $\mathrm{Graph}(x_i^*(\cdot))$ contained in $\mathscr{N}_i$. By using e.g. McShane's Extension Theorem (see e.g. [41, Theorem 3.1]) combined with a projection on the convex and compact set $U$, 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

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

for $\mathscr{L}^1$-almost every $t\in[0,T]$, where we redefined the constant $\mathcal{L}_U := \sqrt{d}\mathcal{L}_U$. □

**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 $(\mathcal{P})$ which are Lipschitz-regular in space. We state in the following lemma a variation of the classical *Dunford-Pettis* compactness criterion (see e.g. [4, Theorem 1.38]) which was already explored in [45, Theorem 2.5].

**Lemma 3** (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*

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

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

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 $u^*(\cdot,\cdot) \in L^\infty([0,T], \mathrm{Lip}(\mathbb{R}^d, U))$ such that the sequence of Lipschitz optimal control maps $(u_N^*(\cdot,\cdot))$ defined in Proposition 7 converges towards $u^*(\cdot,\cdot)$ along a suitable subsequence 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 3 to the sequence of optimal maps built in Proposition 7 up to redefining $\mathcal{L}_U \equiv \max\{L_U, \mathcal{L}_U\}$. □

We now prove that the generalized optimal control $\boldsymbol{\nu}^* \in \mathscr{U}$ for problem $(\mathcal{P}_{\text{meas}})$ is induced by the Lipschitz-in-space optimal control $u^*(\cdot,\cdot) \in L^\infty([0,T], \mathrm{Lip}(\mathbb{R}^d, U))$ which has been defined in Corollary 2. Remark first that by construction of the maps $(u_N^*(\cdot,\cdot))$, it holds that

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

for any $N \geq 1$, where $\boldsymbol{\nu}_N^* \in \mathscr{U}$ denotes the generalized discrete control measure introduced in Theorem 7. In the following proposition, we prove that the sequence $(u_N^*(\cdot,\cdot)\tilde{\mu}_N^*)$ converges towards $u^*(\cdot,\cdot)\tilde{\mu}^*$ in the weak-$^*$ topology of $\mathcal{M}([0,T]\times\mathbb{R}^d, U)$

**Lemma 4** (Convergence of generalized Lipschitz optimal controls)**.** *Let $(\mu_N^*(\cdot)) \subset \mathrm{Lip}([0,T], \mathscr{P}_N(\mathbb{R}^d))$ be the sequence of optimal measure curves associated with $(\mathcal{P}_N)$ and $(u_N^*(\cdot,\cdot)) \subset L^\infty([0,T], \mathrm{Lip}(\mathbb{R}^d, U))$ be the sequence of Lipschitz controls built in Proposition 7. Then, the sequence $(\boldsymbol{\nu}_N^*) = (u_N^*(\cdot,\cdot)\tilde{\mu}_N^*)$ converges towards $\boldsymbol{\nu}^* = u^*(\cdot,\cdot)\tilde{\mu}^*$ in the weak-$^*$ topology of $\mathcal{M}([0,T]\times\Omega, \mathbb{R}^d)$.*

*Proof.* We know by Lemma 3 that for any $p \in (1,+\infty)$, there exists a subsequence of $(u_N^*(\cdot,\cdot))$ which converges in the weak-topology of $L^2([0,T], W^{1,p}(\Omega, U))$ towards $u^*(\cdot,\cdot) \in \mathcal{U}_{\mathcal{L}_U}$. Recalling that one can identify the topological dual of the Banach space $L^2([0,T], W^{1,p}(\Omega, U))$ with $L^2([0,T], W^{-1,p'}(\Omega, U))$, where $p'$ is the conjugate exponent of $p$, the fact that $u_N(\cdot,\cdot) \underset{N\to+\infty}{\rightharpoonup} u(\cdot,\cdot)$ can be written as

$$\int_0^T \langle \xi(t), u_N^*(t,\cdot) \rangle_{W^{1,p}} \mathrm{d}t \underset{N\to+\infty}{\longrightarrow} \int_0^T \langle \xi(t), u^*(t,\cdot) \rangle_{W^{1,p}} \mathrm{d}t, \tag{51}$$

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

Let us now fix in particular a real number $p > d$ so that by Morrey's Embedding (see e.g. [13, Theorem 9.12]), it holds that $W^{1,p}(\Omega, U) \subset C^0(\Omega, U)$. By taking the topological dual of each spaces, we recover the reverse inclusion $\mathcal{M}(\Omega, U) \subset W^{-1,p'}(\Omega, U)$. The latter relation combined with the definition (4) of the duality pairing for vector measures and (51) yields that

$$\int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u_N^*(t,x) \rangle \mathrm{d}\sigma(t)(x)\mathrm{d}t \underset{N\to+\infty}{\longrightarrow} \int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u^*(t,x) \rangle \mathrm{d}\sigma(t)(x)\mathrm{d}t, \tag{52}$$

for any curve $\sigma(\cdot) \in C^0([0,T], \mathcal{M}(\Omega, \mathbb{R}_+))$ and any $\xi \in C_c^1([0,T]\times\Omega, \mathbb{R}^d)$. Remark now that for any $N \geq 1$, one has that

$$\begin{aligned} &\left| \int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u^*(t,x) \rangle \mathrm{d}\mu^*(t)(x)\mathrm{d}t - \int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u_N^*(t,x) \rangle \mathrm{d}\mu_N^*(t)(x)\mathrm{d}t \right| \\ &\qquad \leq \left| \int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u^*(t,x) - u_N^*(t,x) \rangle \mathrm{d}\mu^*(t)(x)\mathrm{d}t \right| \\ &\qquad\quad + \left| \int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u_N^*(t,x) \rangle \mathrm{d}(\mu^*(t) - \mu_N^*(t))(x)\mathrm{d}t \right|. \end{aligned} \tag{53}$$

The first term in the right-hand side of (53) vanishes as $N \to +\infty$ as a consequence of (52). By invoking Kantorovich's duality formula (5) along with the uniform Lipschitz-regularity of the maps $(u_N^*(\cdot,\cdot))$, we can obtain the following upper bound on the second term in the right-hand side of (53)

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

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

$$\int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u_N^*(t,x) \rangle \mathrm{d}\mu_N^*(t)(x) \mathrm{d}t \underset{N \to +\infty}{\longrightarrow} \int_0^T \int_{\mathbb{R}^d} \langle \xi(t,x), u^*(t,x) \rangle \mathrm{d}\mu^*(t)(x) \mathrm{d}t, \tag{54}$$

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 (54) holds for any $\xi \in C_c^0([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ by a classical approximation argument. This precisely amounts to saying that $\boldsymbol{\nu}_N^* \rightharpoonup^* u^*(\cdot,\cdot)\tilde{\mu}^*$ as $N \to +\infty$ along the same subsequence. □

By uniqueness of the weak-$^*$ limit in $\mathcal{M}([0,T] \times \mathbb{R}^d, U)$, we obtain by combining Lemma 4 with Theorem 7 that the optimal solution $\boldsymbol{\nu}^* \in \mathscr{U}$ of $(\mathcal{P}_{\text{meas}})$ is induced by $u^*(\cdot,\cdot)$. This allows us to conclude that the pair $(u^*(\cdot,\cdot), \mu^*(\cdot)) \in L^\infty([0,T], \mathrm{Lip}(\mathbb{R}^d, U)) \times \mathrm{Lip}([0,T], \mathscr{P}(\overline{B(0,R_T)}))$ is a classical optimal pair for $(\mathcal{P})$, which concludes the proof of Theorem 1.

# 6 Discussion on the coercivity assumption $(\mathbf{CO}_N)$

In this section, we discuss more in detail the mean-field coercivity assumptions $(\mathbf{CO}_N)$ by developing an example in which it is both necessary and sufficient for the Lipschitz regularity in space of the optimal control. With this goal, we consider the following Wasserstein optimal control problem

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

In the latter, 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 penalizing the $L^2(\mu(t))$-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 penalization. It can be verified straightforwardly that this problem satisfies the hypotheses **(H1)-(H6)** of Theorem 1. Given a sequence of empirical measures $(\mu_N^0) := (\mu[\boldsymbol{x}_N]) \subset \mathscr{P}_N(\mathbb{R})$ converging narrowly towards $\mu^0$, we can define the family $(\mathcal{P}_V^N)$ of discretized multi-agent problems as

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

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 a pair control-trajectory $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot)) \in L^\infty([0,T], U^N) \times \mathrm{Lip}([0,T], (\mathbb{R}^d)^N)$ solution of $(\mathcal{P}_V^N)$ for any $N \geq 1$. The mean-field Hamiltonian associated to $(\mathcal{P}_V^N)$ is given by

$$\mathbb{H}_N : (t, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{u}) \in [0,T] \times (\mathbb{R}^3)^N \mapsto \frac{1}{N} \sum_{i=1}^N \left( \langle r_i, u_i \rangle - \frac{1}{2} |u_i|^2 \right). \tag{55}$$

By applying the mean-field Pontryagin Maximum Principle displayed in Proposition 6, we obtain the existence of a covector $\boldsymbol{r}_N^*(\cdot) \in \mathrm{Lip}([0,T], \mathbb{R}^N)$ such that

$$\begin{cases} \dot{r}_i^*(t) = -\mathbf{Grad}_{x_i} \mathbb{H}_N(t, \boldsymbol{x}_N^*(t), \boldsymbol{r}_N^*(t), \boldsymbol{u}_N^*(t)) = 0, \\ r_i^*(T) = \quad \mathbf{Grad}_{x_i} \mathrm{Var}_N(\boldsymbol{x}_N^*(T)) = x_i^*(T) - \bar{\boldsymbol{x}}^*(T), \\ u_i^*(t) \in \operatorname*{argmax}\limits_{v \in U} [\langle r_i^*(t), v \rangle - \frac{1}{2}|v|^2]. \end{cases}$$

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

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

Moreover, the optimal control $\boldsymbol{u}_N^*(\cdot)$ is also uniquely determined, and its components write explicitly as

$$u_i^*(t) = \pi_U(r_i^*(t)) \equiv \pi_{[-C,C]}\left(\frac{x_i^*(T) - \bar{\boldsymbol{x}}^*(T)}{\lambda}\right), \tag{56}$$

for all $i \in \{1, \dots, N\}$. It follows directly from this expression that

$$\dot{\bar{\boldsymbol{x}}}^*(t) = \frac{1}{N}\sum_{i=1}^N u_i^*(t) = \frac{1}{N}\sum_{i=1}^N \pi_{[-C,C]}\left(\frac{x_i^*(T) - \bar{\boldsymbol{x}}^*(T)}{\lambda}\right) = 0.$$

Without loss of generality, we can therefore choose $\boldsymbol{x}^0 \in \mathbb{R}^N$ such that $\bar{\boldsymbol{x}}^*(\cdot) \equiv \bar{\boldsymbol{x}}^0 = 0$.

In the following lemma, we derive a simple analytical necessary and sufficient condition for the mean-field coercivity assumption to hold for $(\mathcal{P}_V)$.

**Lemma 5** (Charaterization 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 which case, the optimal coercivity constant is given by $\rho_T = \lambda - T$.*

*Proof.* We start by computing the mean-field Hessians involved in the coercivity estimate. For any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{w} \in \mathbb{R}^N$, we have as a consequence of (55) that

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

Let $(\boldsymbol{w}(\cdot), \boldsymbol{y}(\cdot) \in L^2([0,T], U^N) \times W^{1,2}([0,T],\mathbb{R}^N)$ be the solution of the linearised control-state problem

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{w}(t), \qquad \boldsymbol{y}(0) = 0, \tag{57}$$

with $\boldsymbol{u}_N^*(t) + \boldsymbol{w}(t) \in U^N$. By Cauchy-Schwarz inequality, one can further estimate $|\boldsymbol{y}(T)|_N^2$ as

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

so that we recover

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

and we obtain that the mean-field coercivity condition $(\mathbf{CO}_N)$ holds whenever $\lambda > T$.

Conversely, let us choose a constant admissible control perturbation $\boldsymbol{w}_c(\cdot) \equiv \boldsymbol{w}_c$ such that $\bar{\boldsymbol{w}}_c = 0$. It is always possible to make such a choice since by (56), there exists at least two indices $i, j$ such that $\mathrm{sign}(u_i) = -\mathrm{sign}(u_j)$ for all times $t \in [0,T]$. It is then sufficient to choose $\boldsymbol{w}_c$ such that

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

where $\epsilon > 0$ is a small parameter. As a consequence of (57), the corresponding state perturbation $\boldsymbol{y}_c(\cdot)$ is such that $\bar{\boldsymbol{y}}_c(\cdot) \equiv 0$. Moreover, it also holds that

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

We therefore obtain that for this particular choice of linearised pair control-state, it holds that

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

so that $\rho_T = \lambda - T$ is the sharp mean-field coercivity constant, and the mean-field coercivity condition holds only if $\lambda > T$. □

We can now use this characterization of the coercivity condition to show that it is itself equivalent to the uniform Lipschitz regularity in space of the optimal controls.

**Proposition 8** (Coercivity and regularity)**.** *The followings are equivalent.*

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

(ii) *For any sequence of empirical measures $(\mu_N^0)$ converging narrowly towards $\mu^0 = \frac{1}{2}\mathbb{1}_{[-1,1]}\mathscr{L}^1$ and generating the discrete optimal pairs $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot))$, it holds that*

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

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

*Proof.* Suppose first that the uniform coercivity estimate does not hold, i.e. $\lambda \leq T$. Since the optimal controls are constant over $[0,T]$ as a consequence of (56), the total cost of $(\mathcal{P}_V^N)$ can be rewritten as

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

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

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

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

$$|u_i^* - u_j^*| = \frac{2C}{|x_i^0 - x_j^0| + 2Ct}|x_i^*(t) - x_j^*(t)|. \tag{58}$$

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 $i, j \in \{1, \dots, N\}$ 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 (58) that $(ii)$ fails to hold some pairs of indices and at least for small times.

Suppose now that the mean-field coercivity estimate hold, i.e. $\lambda > T$, and denote by $\rho_T = \lambda - T$ the sharp coercivity constant. Let $I_N, J_N \subset \{1, \dots, N\}$ be the set of indices defined by

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

For $N$ sufficiently big, $I_N$ is necessarily non-empty since $\rho_T > 0$ and as a consequence of the narrow convergence of $(\mu_N^0)$ towards $\mu^0$. Then for any $i \in I_N$, one has that

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

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

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

so that

$$|u_i^* - u_j^*| \leq \frac{1}{\rho_T + t}|x_i^*(t) - x_j^*(t)|, \tag{59}$$

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 (58) that

$$|u_i^* - u_j^*| \leq \begin{cases} 0 & \text{if } \text{sign}(x_i) = \text{sign}(x_j), \\ \dfrac{|x_i^*(t) - x_j^*(t)|}{\rho_T + t} & \text{otherwise,} \end{cases} \tag{60}$$

since 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 $i, j \in \{1, \dots, N\}$ 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

$$\begin{aligned} |u_i^* - u_j^*| = u_j^* - u_i^* &= \text{sign}(x_j^0)C - \frac{x_i^*(t)}{\rho_T + t} \\ &= \frac{x_j^*(t)C}{|x_j^*(t)|} - \frac{x_i^*(t)}{\rho_T + t} \leq \frac{x_j^*(t) - x_i^*(t)}{\rho_T} = \frac{|x_i^*(t) - x_j^*(t)|}{\rho_T}, \end{aligned} \tag{61}$$

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

$$|u_i^* - u_j^*| \leq \frac{|x_i^*(t) - x_j^*(t)|}{\rho_T}. \tag{62}$$

By merging (59), (60), (61) and (62), we conclude that $(ii)$ holds with the uniform constant $\frac{1}{\rho_T} > 0$ whenever the mean-field coercivity estimate holds, which ends the proof of our claim. $\square$

In Proposition 8, we have proven that the mean-field coercivity estimate is both necessary and sufficient for the existe86nce of a uniform Lipschitz constant for the finite-dimensional optimal controls. It is clear when this condition fails that it is not possible to build a sequence of uniformly Lipschitz optimal maps $(u_N^*(\cdot,\cdot))$ for problem $(\mathcal{P}_V^N)$. Since the discrete optimal pairs control-trajectory $(\boldsymbol{u}_N^*(\cdot), \boldsymbol{x}_N^*(\cdot)) \in L^\infty([0,T], U^N) \times \text{Lip}([0,T], \mathbb{R}^N)$ 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 mean-field optimal control for the Wasserstein optimal control problem $(\mathcal{P}_V)$.

# References


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

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

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

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

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

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

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

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

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

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

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

[16] 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.

[17] L. Caffarelli. Some Regularity Properties of Solutions of Monge Ampère Equation. *Communications in Pure and Applied Mathematics*, 44(8-9):965–969, 1991.

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

[19] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat. Sparse Stabilization and Optimal Control of the Cucker-Smale Model. *Mathematical Control and Related Fields*, 3(4):447–466, 2013.

[20] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat. Sparse Stabilization and Control of Alignment Models. *Math. Mod. Meth. Appl. Sci.*, 25 (3):521–564, 2015.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] R. Cibulka, A.L. Dontchev, M.I. Krastanov, and V.M. Veliov. Metrically Regular Differential Generalized Equations. *SIAM J. Cont. Opt.*, 56(1):316–342, 2018.

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

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

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

[32] J.M. Danskin. *The Theory of Max-Min and its Application to Weapons Allocation Problems*, volume 5 of *Ökonometrie und Unternehmensforschung Econometrics and Operations Research.* Springer-Verlag Berlin Heidelberg, 1967.

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

[34] M. Di Francesco and M.D. Rosini. Rigorous Derivation of Nonlinear Scalar Conservation Laws from Follow-the-Leader Type Models via Many Particle Limit. *Archive for Rational Mechanics and Analysis*, 217(3):831–871, 2015.

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

[36] A.L. Dontchev, M.I. Krastanov, and V.M. Veliov. On the Existence of Lipschitz Continuous Optimal Feedback Control. *Research Report*, 2018. ISSN 2521-313X.

[37] A.L. Dontchev and R.T. Rockafellar. *Implicit Functions and Solution Mappings : A View from Variational Analysis.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, 2014.

[38] 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.

[39] M. Duprez, M. Morancey, and F. Rossi. Minimal Time Problem for Crowd Models with a Localized Vector Field. *Journal of Differential Equations*, In press, 2020.

[40] S. Ervedoza and E. Zuazua. *Numerical Approximation of Exact Control for Waves.* Springer Briefs in Mathematics. Springer, 2013.

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

[42] 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.

[43] M. Fornasier, S. Lisini, C. Orrieri, and G. Savaré. Mean-Field Optimal Control as Gamma-Limit of Finite Agent Controls. *Europ. Journ. of App. Math.*, pages 1–34, 2019.

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

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

[46] 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.

[47] N. Gigli. *Second Order Analysis on* $(\mathscr{P}_2(M), W_2)$, volume 216 of *Memoirs of the American Mathematical Society.* AMS, 2012.

[48] R. Glowinski, W. Kinton, and M.F. Wheeler. A Mixed Finite Element Formulation for the Boundary Controllability of the Wave Equation. *International Journal for Numerical Methods in Engineering*, 27(3):623–635, 1989.

[49] 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.

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

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

[52] J.-L. Lions. *Contrôlabilité Exacte, Stabilisation et Perturbation de Systèmes Distribués.* volume RMA 8, Masson, 1988.

[53] S. Mayorga. Short Time Solution to the Master Equation of a First Order Mean Field Game. *Journal of Differential Equations*, In press, 2019.

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

[55] B. Piccoli, F. Rossi, and E. Trélat. Control of the kinetic Cucker-Smale model. *SIAM J. Math. Anal.*, 47(6):4685–4719, 2015.

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

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

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

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

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