

Mean-Field Pontryagin Maximum Principle

Mattia Bongini, Massimo Fornasier, Francesco Rossi, Francesco Solombrino

▶ **To cite this version:**

Mattia Bongini, Massimo Fornasier, Francesco Rossi, Francesco Solombrino. Mean-Field Pontryagin Maximum Principle. Journal of Optimization Theory and Applications, 2017, 211, pp.1-38. 10.1007/s10957-017-1149-5 . hal-01577930

HAL Id: hal-01577930
https://hal.science/hal-01577930

Submitted on 28 Aug 2017

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

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

Noname manuscript No.
(will be inserted by the editor)

Mean-Field Pontryagin Maximum Principle

Mattia Bongini · Massimo Fornasier ·

Francesco Rossi · Francesco Solombrino

Communicated by Aram Arutyunov

the date of receipt and acceptance should be inserted later

Abstract We derive a Maximum Principle for optimal control problems with constraints given by the coupling of a system of ordinary differential equations and a partial differential equation of Vlasov-type with smooth interaction kernel. Such problems arise naturally as Gamma-limits of optimal control prob-

Mattia Bongini

Technische Universität München, Fakultät Mathematik, Boltzmannstrasse 3 D-85748, Garching bei München, Germany. E-mail: mattia.bongini@ma.tum.de

Massimo Fornasier

Technische Universität München, Fakultät Mathematik, Boltzmannstrasse 3 D-85748, Garching bei München, Germany. E-mail: massimo.fornasier@ma.tum.de

Francesco Rossi (corresponding author)

Aix Marseille Université, CNRS, ENSAM, Université de Toulon, LSIS UMR 7296, 13397, Marseille, France. E-mail: francesco.rossi@lsis.org

Francesco Solombrino

Dipartimento di Matematica e Applicazioni, Università di Napoli "Federico II", Via Cintia, 80126 Napoli, Italy. E-mail: francesco.solombrino@unina.it

lems constrained by ordinary differential equations, modeling, for instance, external interventions on crowd dynamics by means of leaders.

We obtain these first-order optimality conditions in the form of Hamiltonian flows in the Wasserstein space of probability measures with forward-backward boundary conditions with respect to the first and second marginals, respectively. In particular, we recover the equations and their solutions by means of a constructive procedure, which can be seen as the mean-field limit of the Pontryagin Maximum Principle applied to the optimal control problem for the discretized density, under a suitable scaling of the adjoint variables.

Keywords: Sparse optimal control, mean-field limit, Γ-limit, optimal control with ODE-PDE constraints, subdifferential calculus, Hamiltonian flows.

AMS Classification: 49J20

1 Introduction

The study of large crowds of interacting agents has received a growing attention in the mathematical literature of the last decade, with countless applications in biology, ecology, social sciences, and economics. Starting from the seminal papers [1–4], emphasis has been put on self-organization, i.e., the formation of macroscopic patterns from the superimposition of simple, reiterated binary interaction rules. Several examples show that spontaneous convergence to pattern formation is not always guaranteed, e.g., for highly dispersed initial configurations in consensus problems [5–8]; hence, the issue of controlling and

stabilizing these systems arises naturally. Two major subclasses of controls of multiagent systems have received substantial attention in the literature: *decentralized* controls and *centralized* ones. With controls of the first kind, the problem is recast into a game-theoretic framework, where agents optimize their individual cost and solutions correspond to Nash equilibria. With those of the second kind, an external policy-maker controlling the dynamics is introduced.

When dealing with large populations, in both cases one faces the well-known problem of the *curse of dimensionality*, term first coined by Bellman precisely in the context of dynamic optimization: the complexity of numerical computations of the solutions of the above problems blows up as the size of the population increases. A possible way out is the so-called *mean-field approach*, where the individual influence of the entire population on the dynamics of a single agent is replaced by an averaged one: this results in a unique mean-field equation and allows one the computation of solutions, cutting loose from the dimensionality.

In the game-theoretic setting, the mean-field approach has led to the development of *mean-field games* [9,10], which model populations, whose agents are competing *freely* with the others towards the maximization of their individual payoff, as for instance in the financial market. The landmark feature of such systems is their capability to autonomously stabilize without external intervention. However, in reality, societies exhibit either convergence to undesired patterns or tendencies toward instability, that only an external government can successfully dominate. The need of such interventions, together with the

limited amount of resources that governments have at their disposal, makes the design of stabilization strategies targeting the least number of agents (nicknamed *sparse*) a key issue, which has been extensively studied in the context of dynamics given by systems of ODEs; see [11–15].

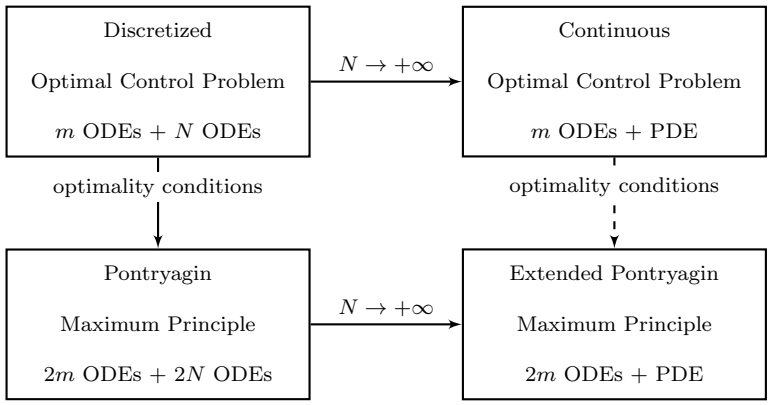
Nevertheless, the concept of sparse control has to be handled with care, when trying to generalize it at the level of a mean-field dynamics. Indeed, the indistinguishability of agents is a fundamental property of the mean-field setting, and it is in sharp contrast with controls acting sparsely on specific agents. Figuratively, trying to stabilize a huge crowds with these controls is like steering a river by means of toothpicks! A first solution to this ambiguity was given in [16, 17], where the control is defined as a locally Lipschitz feedback control with respect to the state variables, and sparsity refers to its property of having a small support. Such concept was successfully used in [18] to implement sparse stabilizers for a consensus problem. This interpretation of sparsity appears also in the framework of the control of more classical PDEs; see [19–22]. An alternative solution for a proper definition of sparse mean-field control was proposed in [23], where the control is sparsely applied on a finite number of individuals immersed in the mean-field dynamics of the rest of the population, resulting in a system, where the controlled ODEs are coupled with a control-free mean-field PDE (but indirectly controlled via the coupling). This kind of control was considered in [24] to model the efficient evacuation of a large crowd of pedestrians with the help of very few informed agents.

First-order optimality conditions, among which the Pontryagin Maximum Principle is the most popular, are necessary conditions to be fulfilled by the optimal controls and they often result in a system of nonlinear equations, which can, in the case of Pontryagin one, be solved numerically in a relatively simple way. Hence, they constitute very often the most viable method towards the numerical computation of (mean-field) optimal controls. In the context of mean-field games and optimal control problems with PDE constraints, first-order optimality conditions have received enormous attention; see for instance [25–28], and they served as a tool for the numerical computation of mean-field controls, see, e.g., [29] and references therein for an extensive discussion on corresponding numerical methods. To the best of our knowledge, no corresponding results have appeared so far in the literature for coupled ODE-PDE systems of the kind considered in [23].

This paper is devoted to the proof of a Pontryagin Maximum Principle to characterize optima of such control problems. We first remark that we are not interested in all possible optima, but mainly on those which arise as limits of optimal strategies of the original discrete problems. We call this subclass of the set of optima *mean-field optimal controls* (see Definition 1.2). We remark that the interest in this class of consistent controls complies with the wish of using the continuous models (independent of the number N of agents) as approximations of the finite-dimensional ones, to circumvent the curse of dimensionality, possibly determined by a large number N of agents. Differently from [17, 23] here we do not wish just to derive the existence of mean-field

controls as natural limits of the finite dimensional optimal controls, but we want additionally to enforce that such a consistency passes naturally also at the level of the first order optimality conditions. This reinforced compatibility provides a tool for the consistent numerical computation of (mean-field) optimal controls.

We summarize our result, borrowing a leaf from the diagram in [27], as follows:



We shall provide a set of hypotheses for which the dashed line from the upper-right to the bottom-right box is valid, hence closing the consistency diagram. Our strategy shall be the following: we apply the Pontryagin Maximum Principle (see e.g. [30, Theorem 23.11]) to the finite-dimensional optimal control problems (the solid line from the upper-left to the bottom-left box), and we pass to the mean-field limit the system of equations obtained with this procedure (the solid line from the bottom-left to the bottom-right box). The derived limit equation for the state and the (rescaled) adjoint variables are obtained in the form of Hamiltonian flows in the Wasserstein space of proba-

bility measures, in the sense of [31]. The result will be a first-order condition valid for all *mean-field optimal controls*. The existence of such controls is also proved (see Corollary 2.2), generalizing the results obtained in [23]. Let us stress again that the extended Pontryagin Maximum Principle constitutes the set of equations for an efficient numerical solution of the mean-field optimal control, which can eventually serve as a surrogate control for approximately solving the finite dimensional optimal control with N agents for N very large. While in the present paper we focus on the derivation and the consistency of the extended Pontryagin Maximum Principle, we postpone to follow up work its efficient numerical solution, as an adaptation of the approaches recently explored in [29].

More formally, we are interested in deriving **optimality conditions** for the solutions of the following optimal control problem subject to coupled ODE-PDE constraints.

Problem 1.1 For $T > 0$ fixed, find $u^* \in L^1([0,T];\mathcal{U})$ minimizing the cost functional

$$F(u) = \int_0^T \left[L(y(t),\mu(t)) + \gamma(u(t))\right] dt, \tag{1}$$

where (y,μ) solve

$$\begin{cases} \dot{y}_k(t) = (K \star \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1,\dots,m, \\ \partial_t \mu(t) = -\nabla_x \cdot \left[(K \star \mu(t) + g(y(t)))\mu(t)\right], \end{cases} \tag{2}$$

for the given initial datum $(y(0),\mu(0)) = (y^0,\mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$.

Here, $K : \mathbb{R}^d \to \mathbb{R}^d$ is an interaction potential, whose role and properties will be discussed in details in Remarks 1.2 and 1.3. Let us already stress here that this kernel will not represent necessarily physical interaction forces (which could show singular behaviors), rather "social" interactions in multi-agent systems, which we can take the liberty of assuming smooth. This smoothness assumption is admittedly a technical and modeling compromise to allow us to consistently derive the extended Pontryagin Maximum Principle, otherwise not justifiable rigorously anymore. Additionally the cost functional γ above is assumed to be strictly convex, the finite dimensional set of controls $\mathcal{U}$ is convex and compact, B_k are constant matrices, and $\mathcal{P}_c(\mathbb{R}^d)$ is the set of probability measures on $\mathbb{R}^d$ with compact support.

Notice that Problem 1.1 generalizes the control problems introduced and studied in [23]. The existence of mean-field optimal controls for Problem 1.1 can be indeed obtained along the same lines, and will be shortly discussed in Section 2.

We shall prove the following main result.

Theorem 1.1 *Fix an initial datum $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$ and assume that Hypotheses (H) in Section 1.1 below hold. Then there exists a mean-field optimal control for Problem 1.1. Furthermore, if u^* is a mean-field optimal control for Problem 1.1 and (y^*, μ^*) is the corresponding trajectory, then (u^*, y^*, μ^*) satisfies the following* **extended Pontryagin Maximum Principle***:*

There exists $(q^*(\cdot), \nu^*(\cdot)) \in \mathrm{Lip}([0,T]; \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^{2d}))$ *such that*

- *there exists* $R_T > 0$*, depending only on* $y^0, \mathrm{supp}(\mu^0), m, K, g, f_k, B_k, \mathcal{U}$*, and* T*, such that* $\mathrm{supp}(\nu^*(\cdot)) \subseteq B(0, R_T)$ *and it satisfies* $\pi_{1\#}\nu^*(t) = \mu^*(t)$ *for all* $t \in [0,T]$*;*
- *it holds*

$$\begin{cases} \dot{y}_k^* &= \nabla_{q_k} \mathbb{H}_c(y^*, q^*, \nu^*, u^*), \\ \dot{q}_k^* &= -\nabla_{y_k} \mathbb{H}_c(y^*, q^*, \nu^*, u^*), \\ \partial_t \nu^* &= -\nabla_{(x,r)} \cdot ((J \nabla_\nu \mathbb{H}_c(y^*, q^*, \nu^*, u^*))\nu^*), \\ u^* &= \arg\max_{u \in \mathcal{U}} \mathbb{H}_c(y^*, q^*, \nu^*, u) \end{cases} \tag{3}$$

where $J \in \mathbb{R}^{2d \times 2d}$ *is the symplectic matrix*

$$J = \begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix},$$

the Hamiltonian $\mathbb{H}_c : \mathbb{R}^{2dm} \times \mathcal{P}_c(\mathbb{R}^{2d}) \times \mathbb{R}^D \to \mathbb{R}$ *is defined as*

$$\mathbb{H}_c(y, q, \nu, u) = \begin{cases} \mathbb{H}(y, q, \nu, u) & \text{if } \mathrm{supp}(\nu) \subseteq \mathrm{cl}(B(0, R_T)), \\ +\infty & \text{elsewhere;} \end{cases}$$

and $\mathbb{H} : \mathbb{R}^{2dm} \times \mathcal{P}_c(\mathbb{R}^{2d}) \times \mathbb{R}^D \to \mathbb{R}$ *is defined as*

$$\begin{aligned} \mathbb{H}(y, q, \nu, u) =& \frac{1}{2} \int_{\mathbb{R}^{4d}} (r - r') \cdot K(x - x') \, d\nu(x, r) \, d\nu(x', r') \\ &+ \int_{\mathbb{R}^{2d}} r \cdot g(y)(x) d\nu(x, r) + \sum_{k=1}^m \int_{\mathbb{R}^{2d}} q_k \cdot K(y_k - x) \, d\nu(x, r) \\ &+ \sum_{k=1}^m q_k \cdot (f_k(y) + B_k u) - L(y, \pi_{1\#}\nu) - \gamma(u). \end{aligned} \tag{4}$$

– *the following conditions for system* (3) *hold at time* 0*:* $y^*(0) = y^0$ *and* $\nu^*(0)(E \times \mathbb{R}^d) = \mu^0(E)$ *for every Borel set* $E \subseteq \mathbb{R}^d$*,*

– *the following conditions for system* (3) *hold at time* T*:* $q^*(T) = 0$ *and* $\nu^*(T)(\mathbb{R}^d \times E) = \delta_0(E)$ *for every Borel set* $E \subseteq \mathbb{R}^d$*, where* δ_0 *is the Dirac measure centered in* 0*.*

As already mentioned, the formulation given above shows that the dynamics of (y^*, q^*, ν^*) is essentially an Hamiltonian flow in the Wasserstein space of probability measures with respect to state and adjoint variables with Hamiltonian $\mathbb{H}$, in the sense of [31]. The definition of $\mathbb{H}_c$ is introduced to simplify some technical details and does not alter the result. This fact is remarkably consistent with the dynamics (2), since both are flows in a Wasserstein space. This formulation of the optimality conditions making use of the formalism of subdifferential calculus in Wasserstein spaces of probability measures constitutes one of the novelties of the work.

Remark 1.1 For every (y, q, ν) with $\mathrm{supp}(\nu) \subseteq \mathrm{cl}(B(0, R_T))$, (4) immediately implies that

$$\overline{u} \in \arg\max_{u \in \mathcal{U}} \mathbb{H}_c(y, q, \nu, u) \iff \overline{u} \in \arg\max_{u \in \mathcal{U}} \left(\sum_{k=1}^{m} q_k \cdot B_k u - \gamma(u) \right).$$

Then, the strict convexity of γ and the convexity and the compactness of $\mathcal{U}$ imply that $\overline{u}$ is uniquely determined by (y, q, ν). This is the reason why we write the equality symbol in $u^* = \arg\max_{u \in \mathcal{U}} \mathbb{H}_c(y^*, q^*, \nu^*, u)$ in place of an inclusion.

We point out the difference between the usual gradient in $\mathbb{R}^{2d}$ with respect to the state variables x and the adjoint variables r, denoted by $\nabla_{(x,r)}$, and the Wasserstein gradient ∇_ν of $\mathbb{H}_c$. In order to do that, we introduce the functions $\ell \in \mathcal{C}^2(\mathbb{R}^{dm} \times \mathbb{R}^d \times \mathbb{R}^d; \mathbb{R})$ and $\omega \in \mathcal{C}^2(\mathbb{R}^d; \mathbb{R}^d)$, related to the functional L in (1) via

$$L(y, \mu) = \int_{\mathbb{R}^d} \ell\left(y, x, \textstyle\int \omega\mu\right) d\mu(x),$$

where $\int \omega\mu := \omega\mu(\mathbb{R}^d)$. Denoting with $\nabla_\xi \ell$ and $\nabla_\varsigma \ell$ the partial derivatives of the function $\ell(\eta, \xi, \varsigma)$, and with $\mathbf{D}\omega(x)$ the Jacobian of the function ω evaluated at x we will show in Section 3 that, whenever ν has supported contained in $B(0, R_T)$, $\nabla_\nu \mathbb{H}_c$ can be computed explicitly as follows:

– For $l = 1, \ldots, d$, it holds

$$\begin{aligned}\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r) \cdot e_l = &\int_{\mathbb{R}^{2d}} (r - r') \cdot (\mathbf{D}K(x - x')e_l)\, d\nu(x', r') \\ &+ r \cdot (\mathbf{D}_x g(y)(x)e_l) - \sum_{k=1}^{m} q_k \cdot (\mathbf{D}K(y_k - x)e_l) \\ &- \nabla_\xi \ell(y, x, \textstyle\int \omega\mu) \cdot e_l \\ &- \left(\int_{\mathbb{R}^d} \nabla_\varsigma \ell(y, x', \textstyle\int \omega\mu)\, d\mu(x')\right) \cdot (\mathbf{D}\omega(x)e_l).\end{aligned} \tag{5}$$

These are the components of $\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r)$ in the x_l coordinates.

– For $l = d + 1, \ldots, 2d$ it holds

$$\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r) \cdot e_l = \int_{\mathbb{R}^{2d}} K(x - x') \cdot e_{l-d}\, d\nu(x', r') + g(y)(x) \cdot e_{l-d} \tag{6}$$

These are the components of $\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r)$ in the r_{l-d} coordinates.

Notice that $\nabla_\nu \mathbb{H}(y, q, \nu, u)$ actually does not depend on u, as a consequence of the fact that the control does not act directly on the PDE component of (2).

The main tool we use to prove Theorem 1.1 is the Pontryagin Maximum Principle (henceforth, simply addressed as PMP) for optimal control problems with ODE constraint. We shall apply it to the following finite-dimensional problems, whose constraints converge to the coupled ODE-PDE system of Problem 1.1, as we will show in Section 2. For this reason, we call Theorem 1.1 the *extended PMP*.

Problem 1.2 For $T > 0$ fixed, find $u^* \in L^1([0,T];\mathcal{U})$ minimizing the cost functional

$$F_N(u) = \int_0^T [L(y(t), \mu_N(t)) + \gamma(u(t))]\, dt, \tag{7}$$

where (y, μ_N) solve

$$\begin{cases} \dot{y}_k = \frac{1}{N}\sum_{j=1}^N K(y_k - x_j) + f_k(y) + B_k u, \quad k = 1, \dots, m \\ \dot{x}_i = \frac{1}{N}\sum_{j=1}^N K(x_i - x_j) + g(y)(x_i), \qquad\quad i = 1, \dots, N, \end{cases} \tag{8}$$

for the given initial datum $(y(0), x(0)) = (y^0, x^0) \in \mathbb{R}^{dm} \times \mathbb{R}^{dN}$, where

$$\mu_N(t)(x) = \frac{1}{N}\sum_{i=1}^N \delta(x - x_i(t)),$$

is the empirical measure centered on the trajectory $x(\cdot) = (x_1(\cdot), \dots, x_N(\cdot))$.

The extended PMP will be derived after reformulating the finite-dimensional PMP applied to Problem 1.2 in terms of the empirical measure in the product space of state variables x_i and adjoint variables p_i, defined as

$$\nu_N(x,r) = \frac{1}{N}\sum_{i=1}^{N} \delta(x - x_i, r - Np_i).$$

Notice that rescaling the adjoint variables p_i by the number N of agents is needed in order to observe a nontrivial dynamics in the limit; indeed, within this scaling, the right-hand side of the finite-dimensional PMP is brought back to the form considered, for instance, in [32], with a different Hamiltonian.

The structure of the paper is the following. In Section 1.1 we recall the basic notations and introduce the main Hypotheses (H). In Section 2, we study the controlled dynamics subject to a coupled ODE-PDE constraint of the form (2), establishing existence and uniqueness results for solutions. In Section 3, we recall basic facts about subdifferential calculus in Wasserstein spaces, and we explicitly compute $\nabla_\nu \mathbb{H}_c$. In Section 4 we study the finite-dimensional Problem 1.2, and apply the PMP to it. In Section 5, we prove the extended PMP, i.e., Theorem 1.1.

1.1 Notation and Hypotheses (H)

We start this section by recalling the notation used throughout the paper.

The constants d, D are two positive integers (the dimension of the space of the agents and of the control, respectively), $T > 0$ (the end time of the

optimization procedure), and $\mathcal{U}$ is a *convex compact* subset of $\mathbb{R}^D$ (set in which controls take values).

Elements of $\mathbb{R}^n$ are always represented as row vectors. Functionals have the following expressions: $K : \mathbb{R}^d \to \mathbb{R}^d$, each f_k satisfies $f_k : \mathbb{R}^{dm} \to \mathbb{R}^d$, and for every $y \in \mathbb{R}^{dm}$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, $g(y) : \mathbb{R}^d \to \mathbb{R}^d$ and $L(y, \mu) : \mathbb{R}^d \to \mathbb{R}$. The matrices B_k are constant $d \times D$ matrices.

The space $\mathcal{P}(\mathbb{R}^n)$ is the set of probability measures, which take values on $\mathbb{R}^n$, while the space[1] $\mathcal{P}_p(\mathbb{R}^n)$ is the subset of $\mathcal{P}(\mathbb{R}^n)$ whose elements have finite p-th moment, i.e.,

$$\int_{\mathbb{R}^n} \|x\|^p d\mu(x) < +\infty.$$

We denote by $\mathcal{P}_c(\mathbb{R}^n)$ the subset of $\mathcal{P}_1(\mathbb{R}^n)$ which consists of all probability measures with compact support. Notice that, if $(\mu_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{P}_c(\mathbb{R}^n)$ and it exists $R > 0$ such that $\text{supp}(\mu_n) \subseteq B(0, R)$ for all $n \in \mathbb{N}$, then $(\mu_n)_{n\in\mathbb{N}}$ is compact in $\mathcal{P}_p(\mathbb{R}^n)$ for all $p \geq 1$.

For any $\mu \in \mathcal{P}(\mathbb{R}^n)$ and any Borel function $r : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$, we denote by $r_{\#}\mu \in \mathcal{P}(\mathbb{R}^{n_2})$ the *push-forward of* μ *through* r, defined by

$$r_{\#}\mu(B) := \mu(r^{-1}(B)) \quad \text{for every Borel set } B \text{ of } \mathbb{R}^{n_2}.$$

In particular, if one considers the projection operators π_1 and π_2 defined on the product space $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, for every $\rho \in \mathcal{P}(\mathbb{R}^{n_1} \times \mathbb{R}^{n_2})$ we call *first* (resp., *second*) *marginal* of ρ the probability measure $\pi_{1\#}\rho$ (resp., $\pi_{2\#}\rho$). Given $\mu \in \mathcal{P}(\mathbb{R}^{n_1})$ and $\nu \in \mathcal{P}(\mathbb{R}^{n_2})$, we denote with $\mathit{\Pi}(\mu, \nu)$ the subset of all probability measures in $\mathcal{P}(\mathbb{R}^{n_1} \times \mathbb{R}^{n_2})$ with first marginal μ and second marginal ν.

[1] We follow the notation of [33].

On the set $\mathcal{P}_p(\mathbb{R}^n)$ we shall consider the following distance, called the *Wasserstein or Monge-Kantorovich-Rubinstein distance*,

$$\mathcal{W}_p^p(\mu,\nu) = \inf\left\{\int_{\mathbb{R}^{2n}} \|x-y\|^p d\rho(x,y) : \rho \in \mathit{\Pi}(\mu,\nu)\right\}. \tag{9}$$

If $p=1$ we have the following equivalent expression for the Wasserstein distance:

$$\mathcal{W}_1(\mu,\nu) = \sup\left\{\int_{\mathbb{R}^n} \varphi(x) d(\mu-\nu)(x) : \varphi \in \mathrm{Lip}(\mathbb{R}^n),\ \mathrm{Lip}(\varphi) \le 1\right\}.$$

We denote by $\mathit{\Pi}_o(\mu,\nu)$ the set of optimal plans for which the minimum is attained, i.e.,

$$\rho \in \mathit{\Pi}_o(\mu,\nu) \iff \rho \in \mathit{\Pi}(\mu,\nu) \text{ and } \int_{\mathbb{R}^{2n}} \|x-y\|^p d\rho(x,y) = \mathcal{W}_p^p(\mu,\nu).$$

It is well-known that $\mathit{\Pi}_o(\mu,\nu)$ is non-empty for every $(\mu,\nu) \in \mathcal{P}_p(\mathbb{R}^n)\times\mathcal{P}_p(\mathbb{R}^n)$ (see [34]), hence the infimum in (9) is actually a minimum. The following definition is motivated by Definition 10.3.1 and Remark 10.3.3 in [33].

Definition 1.1 Let $\psi : \mathcal{P}_2(\mathbb{R}^{2d}) \to]-\infty,+\infty]$ be a proper and lower semi-continuous functional, and let $\nu_0 \in D(\psi)$. We say that $w \in L^2_{\nu_0}(\mathbb{R}^{2d})$ belongs to the *(Fréchet) subdifferential of ψ at ν_0*, in symbols $w \in \partial\psi(\nu_0)$ if and only if for any $\nu_1 \in \mathcal{P}_2(\mathbb{R}^{2d})$ it holds

$$\psi(\nu_1) - \psi(\nu_0) \ge \inf_{\rho\in\mathit{\Pi}_o(\nu_0,\nu_1)} \int_{\mathbb{R}^{4d}} w(z_0)\cdot(z_1-z_0) d\rho(z_0,z_1) + o(\mathcal{W}_2(\nu_1,\nu_0)).$$

It can be seen [31] that, whenever $\partial\psi(\nu_0)$ is nonempty, it has an element with minimal $L^2_{\nu_0}(\mathbb{R}^{2d})$-norm, which we call the *Wasserstein gradient* $\nabla_\nu\psi(\nu_0)$ of ψ at ν_0.

For any $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ and $K : \mathbb{R}^d \to \mathbb{R}^d$, the notation $K \star \mu$ stands for the convolution of K and μ, i.e.,

$$(K \star \mu)(x) = \int_{\mathbb{R}^d} K(x - x')d\mu(x');$$

this quantity is well-defined whenever K is continuous and *sublinear*, i.e., there exists C such that $\|K(\xi)\| \leq C(1 + \|\xi\|)$ for all $\xi \in \mathbb{R}^d$. Furthermore we shall deal also with the convolution $(\nabla_{(x',r')}\langle r', K(x')\rangle) \star \nu$ in $\mathbb{R}^{2d}$, whose explicit expression is

$$\left((\nabla_{(x',r')}\langle r', K(x')\rangle) \star \nu\right)(x, r) = \int_{\mathbb{R}^{2d}} \left(\nabla_{(x',r')}\langle r - r', K(x - x')\rangle\right) d\nu(x', r').$$

Notice that, under the hypotheses we are going to make, this convolution is not always well-defined for $\nu \in \mathcal{P}_1(\mathbb{R}^{2d})$. It is nonetheless well-defined for measures $\nu \in \mathcal{P}_c(\mathbb{R}^{2d})$, that is to say for all the cases that will appear in the sequel.

We shall denote with $\mathcal{M}_b(\mathbb{R}^{n_1}; \mathbb{R}^{n_2})$ the space of bounded Radon vector measures from $\mathbb{R}^{n_1}$ to $\mathbb{R}^{n_2}$, and with $\|\cdot\|_{\mathcal{M}_b(\mathbb{R}^{n_1};\mathbb{R}^{n_2})}$ the total variation norm on it. If $\omega \in \mathcal{C}(\mathbb{R}^d; \mathbb{R}^d)$ is sublinear and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, the Radon measure $\omega\mu \in \mathcal{M}_b(\mathbb{R}^d; \mathbb{R}^d)$ is defined as

$$\omega\mu(E) := \int_E \omega(x)d\mu(x), \qquad \text{for every } E \subset \mathbb{R}^d \text{ bounded.}$$

We shall denote by $\int \omega\mu := \omega\mu(\mathbb{R}^d)$.

In what follows, we shall consider the space $\mathcal{X} := \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^d)$, together with the following distance

$$\|(y, \mu) - (y', \mu')\|_{\mathcal{X}} := \|y - y'\| + \mathcal{W}_1(\mu, \mu'), \tag{10}$$

where $\|y - y'\| := \sum_{k=1}^m \|y_k - y'_k\|_{\ell_2(\mathbb{R}^d)}$.

Henceforth, we assume that the following regularity properties hold.

Hypotheses (H)

(K) The function $K \in \mathcal{C}^2(\mathbb{R}^d;\mathbb{R}^d)$ is odd and sublinear, i.e., there exists $C_K > 0$ such that for all $x \in \mathbb{R}^d$ it holds

$$\|K(x)\| < C_K(1 + \|x\|).$$

(L) The function $L : \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ is

$$L(y,\mu) = \int_{\mathbb{R}^d} \ell\left(y, x, \textstyle\int \omega\mu\right) d\mu(x),$$

with $\ell \in \mathcal{C}^2(\mathbb{R}^{dm} \times \mathbb{R}^d \times \mathbb{R}^d;\mathbb{R})$ and $\omega \in \mathcal{C}^2(\mathbb{R}^d;\mathbb{R}^d)$.

(G) The function $g \in \mathcal{C}^2(\mathbb{R}^{dm};\mathcal{C}^2(\mathbb{R}^d;\mathbb{R}^d))$ satisfies for all $x \in \mathbb{R}^d$ and all $y \in \mathbb{R}^{dm}$

$$g(y)(x) \cdot x \leq G_1\|x\|^2 + G_2 \max_{l=1,\dots,m} \|y_l\|^2 + G_3,$$

where the constants G_1, G_2 and G_3 are independent on x and y.

(F) For each $k = 1, \dots, m$, the function $f_k \in \mathcal{C}^2(\mathbb{R}^{dm};\mathbb{R}^d)$ satisfies for all $y \in \mathbb{R}^{dm}$

$$f_k(y) \cdot y_k \leq F_1 \max_{l=1,\dots,m} \|y_l\|^2 + F_2,$$

where the constants F_1 and F_2 are independent on y and k.

(U) The set $\mathcal{U} \subseteq \mathbb{R}^D$ is compact and convex.

(γ) The function $\gamma : \mathcal{U} \to \mathbb{R}$ is strictly convex.

The following remark discusses some examples in literature falling into the above framework.

Remark 1.2 The set of hypotheses (H) allows to consider interaction kernels K appearing in several well-established multiagent dynamical models. In particular the interaction kernel appearing in the Cucker-Smale model [6] which is given by

$$K(\mathbf{x}) := \begin{pmatrix} 0 \\ -\phi(\|x\|)v \end{pmatrix},$$

for $\mathbf{x} = (x, v)^T \in \mathbb{R}^6$ and $\phi(\lambda) = \frac{\kappa}{(\sigma^2+\lambda^2)^\beta}$, for some fixed parameters $\kappa, \sigma > 0$ and $\beta \geq 0$, satisfies the hypothesis (K).

The above kernel is a possible choice for the model considered in [23] in combination with the leader-follower interactions given by

$$f_k(\mathbf{y}) := \begin{pmatrix} w_k \\ \frac{1}{m}\sum_{j=1}^m \phi(\|y_k - y_j\|)(w_j - w_k) \end{pmatrix}$$

$$g(\mathbf{y})(\mathbf{x}) := \begin{pmatrix} v \\ \frac{1}{m}\sum_{j=1}^m \phi(\|x - y_j\|)(w_j - v) \end{pmatrix},$$

where $\mathbf{y} = (y_1, w_1, \ldots, y_m, w_m)$ describes the population of leaders. It can be seen that such f_k and g satisfy the hypotheses (F) and (G), respectively.

Another example comes from considering a mollified version of the Hegselmann-Krause [35] interaction kernel, i.e., $\phi(x) = (\chi_{[0,R]} \star \rho_\varepsilon)(x)$ for some *confidence radius* $R > 0$ and choosing

$$K(x) := -\phi(\|x\|)x \quad \text{for} \quad x \in \mathbb{R}.$$

Coming back to a second-order system with space-velocity variables, hypothesis (G) remains satisfied also by adding to $g(\mathbf{y})(\mathbf{x})$ the *self-propulsion/friction* term

$$S(\mathbf{x}) := (\alpha - \beta\|v\|^2)v.$$

where α and β are nonnegative parameters. This term, introduced in [36], balances the self-propulsion of individuals given by αv and the Rayleigh-type friction $-\beta\|v\|^2 v$, prescribing the speed of each agent $\|v\|$ to approach the asymptotic value $\sqrt{\alpha/\beta}$ (if other effects are ignored), which can be seen as a characteristic limit speed for the dynamics. S is commonly encountered in the modeling of bacteria and groups of animals, see for instance [7, 37].

Regarding the cost functional, various examples can be considered depending on the behavior one wants to induce on the population of followers. For instance, a standard problem in the study of the Cucker-Smale model is to find conditions to ensure flocking, i.e., alignment of the whole crowd towards the same velocity. A possible choice, fully complying with our set of hypotheses is the minimization of the variance[2] of the crowd, by choosing

$$\begin{aligned} L_1(\mathbf{y},\mu) &:= \int_{\mathbb{R}^6} \left(\frac{2}{m}\sum_{k=1}^m \|w_k\|^2 + 2\|v\|^2 \right) d\mu(x,v) - \left\| \frac{1}{m}\sum_{k=1}^m w_k + \int v\, d\mu(x,v) \right\|^2 \\ &= \int_{\mathbb{R}^6} \left(\frac{2}{m}\sum_{k=1}^m \|w_k\|^2 + 2\|v\|^2 - \right. \\ &\quad \left. - \left(\frac{1}{m}\sum_{k=1}^m w_k + \int v'\, d\mu(x',v') \right) \cdot \left(\frac{1}{m}\sum_{k=1}^m w_k + v \right) \right) d\mu(x,v), \end{aligned}$$

[2] For simplicity of computation, we consider minimization of 4 times the variance.

that is of the form $L = \int_{\mathbb{R}^6} \ell(\mathbf{y}, \mathbf{x}, \int \omega\mu)\, d\mu(\mathbf{x})$ by choosing $\omega(\mathbf{x}) = v$ and

$$\ell(\mathbf{y}, \mathbf{x}, \varsigma) := \frac{2}{m} \sum_{k=1}^{m} \|w_k\|^2 + 2\|v\|^2 - \left(\frac{1}{m} \sum_{k=1}^{m} w_k + \varsigma \right) \cdot \left(\frac{1}{m} \sum_{k=1}^{m} w_k + v \right).$$

For the control constraints, we assume $\mathcal{U} := [-1, 1]^{3m}$ and we choose to penalize the L^2-norm of the control, hence $\gamma(u) := \|u\|^2$. Other forms for the cost L can be of interest. For example, one may want to drive the crowd to a *given* fixed velocity $\bar{v}$, and correspondingly minimize

$$L_2(\mathbf{y}, \mu) := \int_{\mathbb{R}^6} \left(\frac{1}{2m} \sum_{k=1}^{m} \|w_k - \bar{v}\|^2 + \frac{1}{2}\|v - \bar{v}\|^2 \right) d\mu(x, v),$$

that is again of the form $\int_{\mathbb{R}^6} \ell(\mathbf{y}, \mathbf{x}, \int \omega\mu)\, d\mu(\mathbf{x})$, with ℓ not depending on its third variable, this time. Lastly, we mention the following cost functional considered in [38] in connection with the control of the Hegselmann-Krause model: for any fixed $\bar{x} \in \mathbb{R}$, consider

$$\begin{aligned} L_3(\mathbf{y}, \mu) := \frac{1}{2}\|y_1(T) - \bar{x}\|^2 + \frac{1}{2} \int_0^T \int_{\mathbb{R}^2} \|x - z\|^2 d\mu(t, x) d\mu(t, z) dt \\ + \frac{1}{2} \int_0^T \int_{\mathbb{R}} \|x - y_1(t)\|^2 d\mu(t, x) dt + \frac{1}{2} \int_0^T \|u(t)\|^2 dt. \end{aligned}$$

Remark 1.3 In all the mentioned examples, the interaction potential K is smooth. In such a context, both a mean-field theory relating the particle model and its continuum limit and suitable quantitative estimates for convergence are well-established since the paper [39]. In the present paper, similar estimates, adapted to our situation, will be obtained in Lemmata 2.3 and 4.2. While such estimates basically only require Lipschitz continuity of the potential K, we are however forced to require a $\mathcal{C}^2$-regularity for a twofold reason. First of all, at least continuous differentiability of K (and, as a consequence, of the finite

dimensional Hamiltonian $\mathbb{H}_N$ defined in (29)) is needed to give a meaning to the PMP (28) in the sense of Peano's existence Theorem. Requirements of boundedness and continuity of the gradient are needed in the general context of existence theory for Hamiltonian flows, although some very technical weakening of the hypotheses can be allowed (see [31, Assumptions (H1') and (H2')]). Furthermore, an important requirement to be met is the so-called λ-convexity, or *semiconvexity* of the Hamiltonian in the sense of Definition A.1. In the theory of Hamiltonian flows, this is a key assumption, since it guarantees that the Hamiltonian stays constant along trajectories (see [31, Theorem 5.2]). In our paper, the semiconvexity assumption allows for the computation of the Wasserstein gradient of the functional $\mathbb{H}_c$ in Theorem 3.1, along the lines of the general theory in [33, Chapter 10]. Due to the complicated form of $\mathbb{H}_c$ in (4), it is not possible to enforce this requirement, unless additional smoothness of the involved terms is considered. This motivates our choice of dealing with a $\mathcal{C}^2$ interaction kernel K.

It is clear from the above discussion that the case of a singular interaction kernel, which arises for important problems in mathematical physics, when dealing for instance with Newton- or Coulomb-type interactions, cannot fall into the scope of this paper. On the other hand even a complete existence theory for these kind of problems has not been achieved so far (see [40, Chapter 1.4] for a general discussion), although relevant results in this framework have appeared in recent years starting from the seminal papers [41, 42].

Remark 1.4 We briefly compare Hypotheses (H) with those of [25,26]. In [25], which deals with an SDE-constrained optimal control problem, $\mathcal{C}^{1,1}$ functionals with respect to state variables and the control are considered. Therefore our hypotheses are just slightly more restrictive. On the other hand, we do not require differentiability of the running cost. The authors of [26] deal, instead, with a mean-field game type optimality conditions to model evacuation scenarios. They derive a first-order condition under the hypotheses of continuous differentiability of the functionals with respect to the state variables together with convexity and positivity assumptions. Furthermore, they deal specifically with an L^2 control cost, while we allow ours to be strictly convex.

We now give the rigorous definition of *mean-field optimal control*.

Definition 1.2 Let $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$ be given. An optimal control u^* for Problem 1.1 with initial datum (y^0, μ^0) is a *mean-field optimal control* if there exists a sequence $(u_N^*)_{N\in\mathbb{N}} \subset L^1([0,T];\mathcal{U})$ and a sequence $(\mu_N^0)_{N\in\mathbb{N}} \in \mathcal{P}_c(\mathbb{R}^d)$ such that

(*i*) for every $N \in \mathbb{N}$, $\mu_N^0(\cdot) := \frac{1}{N}\sum_{i=1}^N (\cdot - x_{i,N}^0)$ is a sequence of empirical measures for some $x_{i,N}^0 \in \mathrm{supp}(\mu^0) + \overline{B(0,1)}$ such that $\mu_N^0 \rightharpoonup \mu^0$ weakly* in the sense of measures;

(*ii*) for every $N \in \mathbb{N}$, u_N^* is a solution of Problem 1.2 with initial datum (y^0, μ_N^0);

(*iii*) there exists a subsequence of $(u_N^*)_{N\in\mathbb{N}}$ converging weakly in $L^1([0,T];\mathcal{U})$ to u^*.

Remark 1.5 As mentioned before, the above definition is motivated by our interest in optimizers that are close to optimal controls for the original finite-dimensional problems. Notice also that, since the measures μ_N^0 have all compact support contained in $\mathrm{supp}(\mu^0)+\mathrm{cl}(B(0,1))$, they form a compact sequence in $\mathcal{P}_p(\mathbb{R}^n)$ for all $p \geq 1$, and therefore, due to weak* convergence to μ^0, we also have that $\lim_{N\to\infty} \mathcal{W}_p(\mu_N^0, \mu^0) = 0$.

2 The Coupled ODE-PDE Dynamics

In this section, we first recall results for PDE equations of transport type with nonlocal interaction velocities, like the one appearing in the second equation of (2). We then study the coupled ODE-PDE dynamics (2) and we state existence and uniqueness results of solutions, together with continuous dependence on the initial data (y^0, μ^0) and on the control u. The proofs follow closely in the footsteps of similar results in [23,31,43,44], to which we will refer anytime no substantial modifications of the argument is needed.

We start by defining the meaning of solution for the equation

$$\partial_t \mu(t) = -\nabla_x \cdot (v(t, x, \mu(t))\mu(t)), \tag{11}$$

where $v : [0,T] \times \mathbb{R}^n \times \mathcal{P}_1(\mathbb{R}^n) \to \mathbb{R}^n$ is a given vector field and $n \in \mathbb{N}$ is the dimension of the underlying Euclidean space.

Definition 2.1 We say that a map $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^n)$ is a solution of (11) if the following holds:

(*i*) μ has uniformly compact support, i.e., there exists $R > 0$ such that it holds $\text{supp}(\mu(\cdot)) \in B(0, R)$;

(*ii*) μ is continuous with respect to the Wasserstein distance $\mathcal{W}_1$;

(*iii*) μ satisfies (11) in the weak sense, i.e. (see [33, Equation (8.1.4)]),

$$\frac{d}{dt}\int_{\mathbb{R}^n} \phi(x)\, d\mu(t)(x) = \int_{\mathbb{R}^n} \nabla\phi(x) \cdot v(t, x, \mu(t))\, d\mu(t)(x),$$

for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^n; \mathbb{R})$.

Now, we can formally define the concept of solution of the controlled ODE-PDE system (2), which applies, *mutatis mutandis*, to system (3) as well.

Definition 2.2 Let $u \in L^1([0, T]; \mathcal{U})$ and $(y^0, \mu^0) \in \mathcal{X}$, with μ^0 of bounded support, be given. We say that a map $(y, \mu) : [0, T] \to \mathcal{X}$ is a solution of the system (2) with control u if

(*i*) $(y(0), \mu(0)) = (y^0, \mu^0)$;

(*ii*) the solution is continuous in time with respect to the metric (10) in $\mathcal{X}$;

(*iii*) the y coordinates define a Carathéodory solution of the following controlled ODE problem

$$\dot{y}_k(t) = (K \star \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \dots, m,$$

for all $t \in [0, T]$;

(*iv*) μ is a solution of (11), where $v : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}^d$ is the time-varying vector field defined as follows

$$v(t, x, \mu(t))(x) := (K \star \mu(t) + g(y(t)))(x).$$

We now derive the existence of solutions of (2) as limits for $N \to \infty$ of the system of ODE (8). We first prove that solutions of (8) coincide with specific solutions of (2). We then prove the limit result with the help of Lemmas 2.1 and 2.2.

Proposition 2.1 *Let N be fixed, and the control $u \in L^1([0,T];\mathcal{U})$ be given. Let $(y, x_N) : [0,T] \to \mathcal{X}$ be the corresponding solution of (8), with*

$$x_N(t) = (x_{1,N}(t), \dots, x_{N,N}(t)).$$

Then, the couple $(y, \mu_N) : [0,T] \to \mathbb{R}^{dm+dN}$, with $\mu_N(t)$ being the empirical measure

$$\mu_N(t)(x) := \frac{1}{N}\sum_{i=1}^{N}(x - x_{i,N}(t)),$$

is a solution of (2) with control u.

Proof It can be easily proved by rewriting (2) with μ_N and arguing exactly as in [17, Lemma 4.3].

Lemma 2.1 *Let $K : \mathbb{R}^d \to \mathbb{R}^d$ satisfy (K) and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$. Then it holds $K * \mu \in \mathrm{Lip}_{loc}(\mathbb{R}^d)$. Furthermore, for all $y \in \mathbb{R}^d$ it holds*

$$\|(K \star \mu)(y)\| \leq C_K \left(1 + \|y\| + \int_{\mathbb{R}^d} \|x\| d\mu(x)\right).$$

Proof See, for instance, [17, Lemma 6.4].

Lemma 2.2 *Let $K : \mathbb{R}^d \to \mathbb{R}^d$ satisfy (K) and let $\mu^1 : [0,T] \to \mathcal{P}_c(\mathbb{R}^d)$ and $\mu^2 : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ be two continuous maps with respect to $\mathcal{W}_1$ satisfying*

$$\mathrm{supp}(\mu^1(t)) \cup \mathrm{supp}(\mu^2(t)) \subseteq B(0,R),$$

for every $t \in [0,T]$, for some $R > 0$. Then for every $\rho > 0$ there exists constant $L_{\rho,R}$ such that

$$\|K \star \mu^1(t) - K \star \mu^2(t)\|_{L^\infty(B(0,\rho))} \leq L_{\rho,R} \mathcal{W}_1(\mu^1(t), \mu^2(t))$$

for every $t \in [0,T]$.

Proof A proof of this result may be found, for instance, in [17, Lemma 6.7].

Proposition 2.2 *Let $y^0 \in \mathbb{R}^{dm}$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, and μ_N^0 be as in Definition 1.2–(i). Let $(u_N)_{N\in\mathbb{N}} \subseteq L^1([0,T];\mathcal{U})$ be a sequence of controls such that $u_N \rightharpoonup u$, for some $u \in L^1([0,T];\mathcal{U})$.*

Then, the sequence of solutions $(y_N, \mu_N) \in \mathrm{Lip}([0,T];\mathcal{X})$ of (8) with initial data (y^0, μ_N^0) and control u_N converges to a solution $(y,\mu) \in \mathrm{Lip}([0,T];\mathcal{X})$ of (2) with initial data (y^0,μ^0) and control u. Moreover, there exists a constant $\rho_T > 0$, depending only on $y^0, \mathrm{supp}(\mu^0), K, g, f_k, B_k, \mathcal{U}$, and T, such that for every $N \in \mathbb{N}$, for every $k = 1,\dots,m$ and for every $t \in [0,T]$ it holds

$$\|y_{k,N}(t)\|, \|y_k(t)\| \leq \rho_T \quad \text{and} \quad \mathrm{supp}(\mu_N(t)), \mathrm{supp}(\mu(t)) \subseteq B(0,\rho_T).$$

Proof We start by fixing $N > 0$ and estimating the growth of the function $\|y_{k,N}(t)\|^2 + \|x_{i,N}(t)\|^2$ for $k = 1,\dots,m$ and $i = 1,\dots N$. Denote by

$$\Sigma = \{(l,j) : l = 1,\dots,m \text{ and } j = 1,\dots N\}.$$

From Hypotheses (H), Lemma 2.1 and the compactness of $\mathcal{U}$, it holds

$$\frac{1}{2}\frac{d}{dt}\left(\|y_{k,N}\|^2 + \|x_{i,N}\|^2\right) = \dot{y}_{k,N} \cdot y_{k,N} + \dot{x}_{i,N} \cdot x_{i,N}$$
$$= ((K \star \mu_N)(y_{k,N}) + f_k(y) + B_k u) \cdot y_{k,N} + ((K \star \mu_N)(x_i) + g(y)(x_{i,N})) \cdot x_{i,N}$$

$$
\begin{aligned}
&\leq \|(K \star \mu_N)(y_{k,N})\| \, \|y_{k,N}\| + f_k(y_N) \cdot y_{k,N} + \|B_k u\| \|y_{k,N}\| \\
&\quad + \|(K \star \mu_N)(x_{i,N})\| \|x_{i,N}\| + g(y_N)(x_{i,N}) \cdot x_{i,N} \\
&\leq C_K \left(1 + \|y_{k,N}\| + \frac{1}{N} \sum_{j=1}^{N} \|x_{j,N}\| \right) \|y_{k,N}\| + F_1 \max_{l=1,\dots m} \|y_{l,N}\|^2 + F_2 \\
&\quad + M_1 \|y_{k,N}\| + C_K \left(1 + \|x_{i,N}\| + \frac{1}{N} \sum_{j=1}^{N} \|x_{j,N}\| \right) \|x_{i,N}\| + G_1 \|x_{i,N}\|^2 \\
&\quad + G_2 \max_{l=1,\dots m} \|y_{l,N}\|^2 + G_3 \\
&\leq C_1 \max_{(\ell,j) \in \Sigma} \left\{ \|y_{\ell,N}\|^2 + \|x_{j,N}\|^2 \right\} + C_2,
\end{aligned}
$$

with $C_1 = 4C_K + F_1 + G_2 + M_1$ and $C_2 = C_K + F_2 + G_3 + M_1$. If we denote with $b_{(k,i)}(t) = \|y_{k,N}(t)\|^2 + \|x_{i,N}(t)\|^2$ and with $a(t) = \max_{(l,j)\in\Sigma}\{b_{(l,j)}(t)\}$, then the Lipschitz continuity of a implies that a is a.e. differentiable, while by Stampacchia's Lemma (see for instance [45, Chapter 2, Lemma A.4]) for a.e. $t \in [0,T]$ there exists a $(l,j) \in \Sigma$ such that

$$
\dot{a}(t) = \frac{d}{dt} \left(\|y_{l,N}(t)\|^2 + \|x_{j,N}(t)\|^2 \right) \leq 2C_1 a(t) + 2C_2.
$$

Hence, Gronwall's Lemma and Definition 1.2–(i) imply that

$$
a(t) \leq (a(0) + 2C_2 t) e^{2C_1 t} \leq (C_0 + 2C_2 t) e^{2C_1 t}, \tag{12}
$$

for some uniform constant C_0 only depending on y^0 and $\mathrm{supp}(\mu^0)$. It then follows that the trajectories $(y_N(\cdot), \mu_N(\cdot))$ are bounded uniformly in N in a ball $B(0,\rho_T) \subset \mathbb{R}^d$, for

$$
\rho_T := \sqrt{C_0 + 2C_2 T} e^{C_1 T},
$$

that is positive and does not depend on t or on N. This in turn implies that the trajectories $(y_N(\cdot), \mu_N(\cdot))$ are uniformly Lipschitz continuous in N, as can be easily verified by computing $\|\dot{y}_{k,N}\|$ and $\|\dot{x}_{i,N}\|$ and noticing that all the functions involved are bounded by Hypotheses (H) and the fact that we are inside $B(0, \rho_T)$. Therefore

$$\|\dot{y}_{k,N}(t)\| \leq \rho_T', \quad \|\dot{x}_{i,N}(t)\| \leq \rho_T', \tag{13}$$

where the constant ρ_T' does not depend on t or on N.

By an application of the Ascoli-Arzelà theorem for functions on $[0, T]$ and values in the complete metric space $\mathcal{X}$, there exists a subsequence, again denoted by $(y_N(\cdot), \mu_N(\cdot))$ converging uniformly to a limit $(y(\cdot), \mu(\cdot))$, whose trajectories are also contained in $B(0, \rho_T)$. Due to the equi-Lipschitz continuity of $(y_N(\cdot), \mu_N(\cdot))$ and the continuity of the Wasserstein distance, we thus obtain for some $L_T > 0$

$$\begin{aligned} &\|(y(t_2), \mu(t_2)) - (y(t_1), \mu(t_2))\|_{\mathcal{X}} \\ &= \lim_{N \to +\infty} \|(y_N(t_2), \mu_N(t_2)) - (y_N(t_1), \mu_N(t_1))\|_{\mathcal{X}} \leq L_T |t_2 - t_1|, \end{aligned} \tag{14}$$

for all $t_1, t_2 \in [0, T]$. Hence, the limit trajectory $(y^*(\cdot), \mu^*(\cdot))$ belongs as well to $\mathrm{Lip}([0, T]; \mathcal{X})$.

The same proof as in [23, Theorem 3.3] shows now that $(y(\cdot), \mu(\cdot))$ is a solution of (2).

Corollary 2.1 *Let $y^0 \in \mathbb{R}^{dm}$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, and $u \in L^1([0, T]; \mathcal{U})$. Then, there exists a solution of (2) with control u and initial datum (y^0, μ^0).*

Proof Follows from Proposition 2.2 by taking any sequence of empirical measures μ_N^0 as in Definition 1.2–(i), and the constant sequence $u_N \equiv u$ for all $N \in \mathbb{N}$. □

We now prove the continuous dependence on the initial data, that also gives uniqueness of the solution for (2).

Proposition 2.3 *Let the Hypotheses (H) hold. Let $u \in L^1([0,T],\mathcal{U})$ be given, and take two solutions (y^1,μ^1) and (y^2,μ^2) of (2) with control u and with initial data $(y^{0,1},\mu^{0,1}),(y^{0,2},\mu^{0,2}) \in \mathcal{X}$, respectively, where $\mu^{0,1}$ and $\mu^{0,2}$ have both compact support. Then there exists a constant $C_T > 0$ such that for all $t \in [0,T]$ it holds*

$$\|(y^1(t),\mu^1(t)) - (y^2(t),\mu^2(t))\|_{\mathcal{X}} \leq C_T \|(y^{0,1},\mu^{0,1}) - (y^{0,2},\mu^{0,2})\|_{\mathcal{X}}.$$

Proof We start by noticing that, by the definition of a solution, we infer the existence of a $\rho_T > 0$ for which $y^1(\cdot), y^2(\cdot) \in B(0,\rho_T) \subset \mathbb{R}^{dm}$ and $\mathrm{supp}(\mu^1(\cdot)), \mathrm{supp}(\mu^2(\cdot)) \subseteq B(0,\rho_T) \subset \mathbb{R}^d$.

As a preliminary estimate, by hypothesis (K), Lemma 2.1 and Lemma 2.2 with the choice $\rho = \widehat{R} = \rho_T$, we infer the existence of a constant $L^K_{\rho_T} > 0$ such that

$$\|(K * \mu_1)(x) - (K * \mu_2)(y)\| \leq L^K_{\rho_T} \left(\mathcal{W}_1(\mu_1,\mu_2) + \|x - y\|\right). \tag{15}$$

holds for every $x, y \in \mathbb{R}^d$. Furthermore, if for the sake of brevity we denote by

$$G := \sup_{\xi \in B(0,\rho_T) \subset \mathbb{R}^d, \varsigma \in B(0,\rho_T) \subset \mathbb{R}^{dm}} \|\mathbf{D}_y g(\varsigma)(\xi)\| \quad \text{and} \quad F := \max_{1 \leq k \leq m} \mathrm{Lip}_{B(0,\rho_T)}(f_k).$$

then the $\mathcal{C}^2$-regularity of g and f_k for every $k = 1, \dots, m$ imply for every $y_1, y_2 \in B(0, \rho_T)$

$$\|g(y_1) - g(y_2)\|_{L^\infty(B(0,\rho_T))} \leq G\|y_1 - y_2\| \quad \text{and} \quad \|f_k(y_1) - f_k(y_2)\| \leq F\|y_1 - y_2\|. \tag{16}$$

We shall show the continuous dependence estimate by chaining the stability of the ODE

$$\dot{y}_k(t) = (K \star \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \dots, m, \tag{17}$$

with the one of the PDE

$$\partial_t \mu(t) = -\nabla_x \cdot [(K \star \mu(t) + g(y(t)))\mu(t)], \tag{18}$$

first addressing the dependence of (17). By integration we have

$$\begin{aligned} &\|y_k^1(t) - y_k^2(t)\| \leq \|y_k^{0,1} - y_k^{0,2}\| \\ &+ \int_0^t \left(\|(K \star \mu^1(s))(y_k^1(s)) - (K \star \mu^2(s))(y_k^2(s))\| + \|f_k(y^1(s)) - f_k(y^2(s))\| \right) ds. \end{aligned} \tag{19}$$

For the left-hand side of (19) we have that (15), (16), and the uniform bound on $y_1(\cdot)$ and $y_2(\cdot)$ yield

$$\begin{aligned} \|y_{1,k}(t) - y_{2,k}(t)\| \leq \|y_{1,k}^0 - y_{2,k}^0\| + \int_0^t \Big(& L_{\rho_T}^K \mathcal{W}_1(\mu_1(s), \mu_2(s)) + \\ & + L_{\rho_T}^K \|y_{1,k}(s) - y_{2,k}(s))\| + F\|y_1(s) - y_2(s)\| \Big) ds \end{aligned} \tag{20}$$

We now consider (18). Arguing as in the derivation of [23, Formula (3.14)], together with the estimate (16), we get

$$\mathcal{W}_1(\mu^1(t), \mu^2(t)) \leq e^{C_1 t} \mathcal{W}_1(\mu^{0,1}, \mu^{0,2})$$

$$+\int_0^t C_2 e^{C_1 s}\left(L^K_{\rho_T}\mathcal{W}_1(\mu^1(s),\mu^2(s))+G\|y^1(s)-y^2(s)\|\right)ds, \tag{21}$$

for some Gronwall's constants $C_1, C_2 > 0$. We finally consider the function

$$\varepsilon(t) := \|(y^1(t),\mu^1(t))-(y^2(t),\mu^2(t))\|_{\mathcal{X}}$$

and, combining (20) for each $k = 1,\dots,m$ and (21), we obtain

$$\begin{aligned}\varepsilon(t) \leq & \|y^{0,1}-y^{0,2}\| + \int_0^t \Bigg(mL^K_{\rho_T}\mathcal{W}_1(\mu^1(s),\mu^2(s)) + L^K_{\rho_T}\|y^1(s)-y^2(s))\| \\ & + mF\|y^1(s)-y^2(s)\|\Bigg) ds + e^{C_1 t}\mathcal{W}_1(\mu^{0,1},\mu^{0,2}) \\ & + \int_0^t C_2 e^{C_1 s}\left(L^K_{\rho_T}\mathcal{W}_1(\mu^1(s),\mu^2(s))+G\|y^1(s)-y^2(s)\|\right) ds \\ \leq\ & \varepsilon(0)e^{C_1 t} + \int_0^t (mL^K_{\rho_T}+mF+(L^K_{\rho_T}+G)C_2 e^{C_1 s})\varepsilon(s)\,ds.\end{aligned}$$

Gronwall's lemma then implies

$$\varepsilon(t) \leq \varepsilon(0)e^{C_1 t}\left((mL^K_{\rho_T}+mF)t + \frac{(L^K_{\rho_T}+G)C_2}{C_1}(e^{C_1 t}-1)\right).$$

Since $t \in [0,T]$, the result is proved. □

Remark 2.1 Going back to the application of the Ascoli-Arzelá Theorem in Proposition 2.2, consider another converging subsequence of (y_N,μ_N). We can prove that its limit is another solution of (8). Since the solution is unique for Proposition 2.3, we have that all converging subsequences of (y_N,μ_N) have the same limit, hence the sequence (y_N,μ_N) has itself limit (y,μ).

Remark 2.2 Since equicompactly supported solutions are unique, given the initial datum, by Proposition 2.3, combined with Proposition 2.2 we infer that the support of the unique solution can be estimated as a function of the

data. More precisely, it is contained in a ball $B(0, \rho_T)$, where the constant is depending only on y^0, $\mathrm{supp}(\mu^0)$, K, g, f_k, B_k, $\mathcal{U}$, and T.

We conclude this section by stating the existence result on mean-field optimal controls for Problem 1. To this end, we fix an initial datum $(y^0, \mu^0) \in \mathcal{X}$, with μ^0 compactly supported, and choose a sequence μ_N^0 as in Definition 1.2–(i).

Consider the functional $F(u)$ on $L^1([0,T];\mathcal{U})$ defined in (1), where the pair (y, μ) defines the unique solution of (2) with initial datum (y^0, μ^0) and control u. Similarly, consider the functional $F_N(u)$ on $L^1([0,T];\mathcal{U})$ defined in (7), where the pair (y_N, μ_N) defines the unique solution of (2) with initial datum (y^0, μ_N^0) and control u. As recalled in Proposition 2.2, such solution coincides with the solution of the ODE system (8).

The existence is a consequence of the Γ-convergence of the sequence of functionals $(F_N)_{N\in\mathbb{N}}$ on $L^1([0,T];\mathcal{U})$ to the target functional F. For the definition of Γ-convergence, we refer the reader to [46, Definition 4.1, Proposition 8.1].

Theorem 2.1 *Let the functionals* (1)-(7) *and dynamics* (2) *satisfy Hypotheses* (H). *Consider an initial datum* $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^d)$, *and a sequence* $(\mu_N^0)_{N\in\mathbb{N}}$, *where* μ_N^0 *is as in Definition 1.2–*(i). *Then the sequence of functionals* $(F_N)_{N\in\mathbb{N}}$ *on* $X = L^1([0,T];\mathcal{U})$ *defined in* (7) Γ*-converges to the functional* F *defined in* (1).

Proof The proof is the same as [23, Theorem 5.3], provided one uses Ioffe's Theorem (see, for instance, [47, Theorem 5.8]) to derive

$$\liminf_{N\to+\infty} \int_0^T \gamma(u_N(t))dt \geq \int_0^T \gamma(u^*(t))dt,$$

and conclude that the $\Gamma-\liminf$ condition also holds in presence of the control cost γ. □

With the same argument as in [23, Corollary 5.4], we get the existence of mean-field optimal controls for Problem 1.1 as an immediate corollary.

Corollary 2.2 *Let the Hypotheses* (H) *in Section 1.1 hold. For every initial datum* $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$*, there exists a mean-field optimal control* u^* *for Problem 1.1.*

Remark 2.3 Observe that the previous result does not state uniqueness of the optimal control for the infinite dimensional problem. Indeed, in general, we cannot ensure that *all* solutions of Problem 1.1 are mean-field optimal controls.

3 The Wasserstein Gradient

We anticipated in Section 1 that the dynamics of ν^* in (3) is an Hamiltonian flow in the Wasserstein space of probability measures, in the sense of [31]. This means that the vector field $\nabla_\nu \mathbb{H}_c(\nu^*)$ is an element with minimal norm in the Fréchet subdifferential at the point ν^* of the maximized Hamiltonian $\mathbb{H}_c$ introduced in the statement of Theorem 1.1 (we drop for simplicity the y, q and u dependency). The proof of this fact shall follow the strategy adopted to

obtain analogous results in [33, Chapter 10], which however cannot be applied verbatim to our case due to the peculiar nature of our operators. For the ease of reading, a technical fact (namely, the proof that the functional $\mathbb{H}_c$ is semiconvex along geodesics) is deferred to the Appendix 6.

In order to use those techniques, we consider our functionals defined on $\mathcal{P}_2(\mathbb{R}^{2d})$ instead than on $\mathcal{P}_1(\mathbb{R}^{2d})$. Since we shall prove in Proposition 4.2 that, whenever we start from a compactly supported initial datum, the dynamics remains compactly supported uniformly in time, this assumption does not alter our conclusions.

In what follows, we shall fix $y, q \in \mathbb{R}^{dm}$ and $u \in L^1([0,T];\mathcal{U})$ and we write, for the sake of compactness, $\mathbb{H}_c(\nu)$ in place of $\mathbb{H}_c(y,q,\nu,u)$. Moreover we denote by $z = (x,r)$ a variable in $\mathbb{R}^{2d}$.

Whenever $\mathrm{supp}(\nu) \subseteq \mathrm{cl}(B(0,R_T))$, $\mathbb{H}_c(\nu)$ can be rewritten as

$$\mathbb{H}_c(\nu) = \frac{1}{2}\int_{\mathbb{R}^{4d}} \mathcal{F}(z-z')d\nu(z)\nu(z') + \int_{\mathbb{R}^{2d}} \mathcal{G}(z)d\nu(z) - \int_{\mathbb{R}^{2d}} \hat{\ell}(z, \textstyle\int \hat{\omega}\nu)d\nu(z) + Q,$$

where have we set

$$\begin{aligned} &\mathcal{F}(x,r) = r\cdot K(x), \quad \mathcal{G}(x,r) = r\cdot g(y)(x) + \sum_{k=1}^{m} q_k \cdot K(y_k - x), \\ &\hat{\ell} = -\ell\circ(\pi_1, \mathrm{Id}), \quad \hat{\omega} = \omega\circ\pi_1, \end{aligned} \tag{22}$$

and Q collects all the remaining terms not depending on ν. Notice that $\mathcal{F}$ is an even function.

We define the vector field $\nabla_\nu \mathcal{L} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ as

$$\nabla_\nu \mathcal{L}(z) = \nabla_\xi \hat{\ell}(z, \textstyle\int \hat{\omega}\nu) + \mathbf{D}\hat{\omega}(z)^T \left(\displaystyle\int_{\mathbb{R}^{2d}} \nabla_\varsigma \hat{\ell}(z', \textstyle\int \hat{\omega}\nu)\mathrm{d}\nu(z') \right) \tag{23}$$

for every $z \in \mathbb{R}^{2d}$. We can thus define our candidate vector field for the Wasserstein gradient $\nabla_\nu \mathbb{H}_c(\nu_0)$ in the case that $\mathrm{supp}(\nu_0) \subseteq B(0, R_T)$:

$$w := (\nabla \mathcal{F}) \star \nu + \nabla \mathcal{G} - \nabla_\nu \mathcal{L}. \tag{24}$$

Notice that, by Hypotheses (H), w is a continuous function in z, and hence it is well-defined ν-a.e.. In view of (22), it is straightforward to see that w agrees with the vector field defined in (5) and (6), after reintroducing the variables (y, q, u) which do not affect the Wasserstein differentiation, and setting $\mu = \pi_{1\#}\nu$.

Lemma 3.1 *Let $\nu \in \mathcal{P}_c(\mathbb{R}^{2d})$. Then w defined by (24) belongs to $L^p_\nu(\mathbb{R}^{2d})$ for every $p \in [1, +\infty]$, and it satisfies*

$$\begin{aligned} &\int_{\mathbb{R}^{4d}} w(z_0) \cdot (z_1 - z_0) d\rho(z_0, z_1) = \\ &\quad = \int_{\mathbb{R}^{6d}} (\nabla \mathcal{F}(z_0 - z_2) + \nabla \mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot (z_1 - z_0) d\rho(z_0, z_1) d\nu(z_2) \end{aligned} \tag{25}$$

for every plan $\rho \in \Pi(\nu, \nu')$ such that $\nu' \in \mathcal{P}_c(\mathbb{R}^{2d})$.

Proof Since w is continuous, the fact that w is L^p_ν-integrable follows the fact that ν has compact support. Equation (25) then follows by Fubini-Tonelli and from the fact that ρ is compactly supported too by Remark A.1.

In the proof of the forthcoming Theorem 3.1, we shall use the following well-known property.

Proposition 3.1 ([33], Theorem 10.3.10) *Fix $\psi : \mathcal{P}_2(\mathbb{R}^{2d}) \to]-\infty, +\infty]$. Then, for every $\nu_0 \in D(\psi)$, the metric slope*

$$|\partial\psi|(\nu_0) = \limsup_{\nu_1 \to \nu_0} \frac{(\psi(\nu_1) - \psi(\nu_0))^+}{\mathcal{W}_2(\nu_1, \nu_0)}$$

satisfies $|\partial\psi|(\nu_0) \leq \|w\|_{L^2_{\nu_0}}$ for every $w \in \partial\psi(\nu_0)$.

Theorem 3.1 *Let $\nu \in \mathcal{P}_2(\mathbb{R}^{2d})$ be such that $\mathrm{supp}(\nu) \subseteq B(0, R_T)$. Then it holds $\nu \in D(|\partial\mathbb{H}_c|)$ if and only if w as in (24) belongs to $L^2_\nu(\mathbb{R}^{2d})$. In this case, $\|w\|_{L^2_\nu} = |\partial\mathbb{H}_c|(\nu)$, i.e., $w = \nabla_\nu \mathbb{H}_c(\nu)$.*

Proof We start by assuming that $\nu \in \mathcal{P}_2(\mathbb{R}^{2d})$ satisfies $|\partial\mathbb{H}_c|(\nu) < +\infty$ and proving that this implies that w belongs to $L^2_\nu(\mathbb{R}^{2d})$ as well as the bound $\|w\|_{L^2_\nu} \leq |\partial\mathbb{H}_c|(\nu)$. We compute the directional derivative of $\mathbb{H}_c$ along a direction induced by the transport map $Id + \phi$, where ϕ is a smooth function with compact support. We use the shortcut $\nu_{s,\phi}$ to indicate the measure $(Id+s\phi)_\#\nu$ and we notice that such that $\mathrm{supp}(\nu_{s,\phi}) \subseteq \mathrm{cl}(B(0, R_T))$ for any sufficiently small $s > 0$, since $\mathrm{supp}(\nu)$ is well contained in $B(0, R_T)$. Denoting by

$$\hat{L}(\overline{\nu}) = \int_{\mathbb{R}^{2d}} \hat{\ell}(z, \textstyle\int \hat{\omega}\overline{\nu}) d\overline{\nu}(z) \quad \text{for every } \overline{\nu} \in \mathcal{P}(\mathbb{R}^{2d}),$$

from the chain rule and the dominated convergence it follows

$$\lim_{s\to 0} \frac{\hat{L}(\nu_{s,\phi}) - \hat{L}(\nu)}{s} = \int_{\mathbb{R}^{2d}} \nabla\hat{\ell}(z, \textstyle\int \hat{\omega}\nu) \cdot \eta(z) d\nu(z), \tag{26}$$

where $\eta(z)$ is the vector defined by

$$\eta(z) = \lim_{s\to 0} \frac{1}{s} \left(\begin{pmatrix} z + s\phi(z) \\ \\ \int \hat{\omega}\nu_{s,\phi} \end{pmatrix} - \begin{pmatrix} z \\ \\ \int \hat{\omega}\nu \end{pmatrix} \right)$$

which, by a direct computation, is given by

$$\eta(z) := \begin{pmatrix} \phi(z) \\ \\ \int_{\mathbb{R}^{2d}} \mathbf{D}\hat{\omega}(z')\phi(z')\mathrm{d}\nu(z') \end{pmatrix}$$

(observe that actually the last $2d$ components are independent of z). Inserting η into (26) and using Fubini's Theorem we get

$$\lim_{s\to 0} \frac{\hat{L}(\nu_{s,\phi}) - \hat{L}(\nu)}{s} =$$
$$= \int_{\mathbb{R}^{2d}} \nabla_\xi \hat{\ell}(z, \int \hat{\omega}\nu) \cdot \phi(z) d\nu(z) + \int_{\mathbb{R}^{4d}} \nabla_\varsigma \hat{\ell}(z, \int \hat{\omega}\nu) \cdot (\mathbf{D}\hat{\omega}(z')\phi(z'))\, \mathrm{d}\nu(z') d\nu(z)$$

whence exchanging z with z' in the second integral, and recalling (23), we have

$$\lim_{s\to 0} \frac{\hat{L}(\nu_{s,\phi}) - \hat{L}(\nu)}{s} = \int_{\mathbb{R}^{2d}} \nabla_\nu \mathcal{L}(z) \cdot \phi(z)\, \mathrm{d}\nu(z)$$

On top of this, notice that the map

$$s \mapsto \frac{\mathcal{F}((z_0 - z_1) + s(\phi(z_0) - \phi(z_1))) - \mathcal{F}(z_0 - z_1)}{s} + \frac{\mathcal{G}(z_0 + s\phi(z_0)) - \mathcal{G}(z_0)}{s}$$

as $s \to 0$ converges to

$$\nabla\mathcal{F}(z_0 - z_1) \cdot (\phi(z_0) - \phi(z_1)) + \nabla\mathcal{G}(z_0) \cdot \phi(z_0).$$

Since ν has compact support, the dominated convergence theorem, the identity (25) and since $\nabla\mathcal{F}$ is odd, it holds

$$\begin{aligned} +\infty > \lim_{s\to 0} & \frac{\mathbb{H}_c((Id + s\phi)_\# \nu) - \mathbb{H}_c(\nu)}{s} \\ &= \frac{1}{2} \int_{\mathbb{R}^{4d}} \nabla\mathcal{F}(z_0 - z_1) \cdot (\phi(z_0) - \phi(z_1)) d\nu(z_0) d\nu(z_1) \\ &\quad + \int_{\mathbb{R}^{2d}} (\nabla\mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot \phi(z_0) d\nu(z_0) \\ &= \int_{\mathbb{R}^{2d}} w(z_0) \cdot \phi(z_0) d\nu(z_0). \end{aligned}$$

From the last inequality, the assumption that $|\partial\mathbb{H}_c|(\nu) < +\infty$ and using the trivial estimate

$$\mathcal{W}_2((Id + s\phi)_{\#}\nu, \nu) \leq s\|\phi\|_{L^2_\nu},$$

we get

$$\int_{\mathbb{R}^{2d}} w(z_0) \cdot \phi(z_0) d\nu(z_0) \leq |\partial\mathbb{H}_c|(\nu)\|\phi\|_{L^2_\nu},$$

and hence, up to a change of sign of ϕ, this proves that $\|w\|_{L^2_\nu} \leq |\partial\mathbb{H}_c|(\nu)$.

We now prove that the vector w belongs to the subdifferential of $\mathbb{H}_c$; this shall imply that $w \in D(|\partial\mathbb{H}_c|)$ and that it is a minimal selection in $\partial\mathbb{H}_c(\nu)$, by the previous estimate and Proposition 3.1.

For proving the claim, we start by remarking that, due to Proposition 3.1, the vector $w \in L^2_\nu(\mathbb{R}^{2d})$. Now consider a test measure $\overline{\nu}$, a plan $\rho \in \Pi_o(\nu, \overline{\nu})$, and let us compute the directional derivative of $\mathbb{H}_c$ along the direction induced by ρ. If we denote by $\nu_{s,\rho}$ the measure $((1-s)\pi_1 + s\pi_2)_{\#}\rho$ on $\mathbb{R}^{2d}$, since it holds $\nu_{0,\rho} = \nu$, arguing as in the previous step we have

$$\lim_{s\to 0} \frac{\hat{L}(\nu_{s,\rho}) - \hat{L}(\nu)}{s} = \int_{\mathbb{R}^{4d}} \nabla\hat{\ell}(z_0, \textstyle\int\hat{\omega}\nu) \cdot \zeta(z_0, z_1) d\rho(z_0, z_1), \tag{27}$$

where $\zeta(z_0, z_1)$ is the vector defined by

$$\begin{aligned}\zeta(z_0, z_1) &:= \lim_{s\to 0} \frac{1}{s} \left(\begin{pmatrix} (1-s)z_0 + sz_1 \\ \\ \int \hat{\omega}\nu_{s,\rho} \end{pmatrix} - \begin{pmatrix} z_0 \\ \\ \int \hat{\omega}\nu \end{pmatrix} \right) \\ &= \begin{pmatrix} z_1 - z_0 \\ \\ \int_{\mathbb{R}^{4d}} \mathbf{D}\hat{\omega}(\overline{z}_0)(\overline{z}_1 - \overline{z}_0) \mathrm{d}\rho(\overline{z}_0, \overline{z}_1) \end{pmatrix}\end{aligned}$$

(again, observe that the last $2d$ components are independent of z_0, z_1). Inserting ζ into (27) and using Fubini's Theorem we get

$$\lim_{s\to 0}\frac{\hat{L}(\nu_{s,\rho})-\hat{L}(\nu)}{s}=\int_{\mathbb{R}^{4d}}\nabla_\xi\hat{\ell}(z_0,\textstyle\int\hat{\omega}\nu)\cdot(z_1-z_0)d\rho(z_0,z_1)+$$
$$+\int_{\mathbb{R}^{8d}}\nabla_\varsigma\hat{\ell}(z_0,\textstyle\int\hat{\omega}\nu)\cdot(\mathbf{D}\hat{\omega}(\overline{z}_0)(\overline{z}_1-\overline{z}_0))\,\mathrm{d}\rho(\overline{z}_0,\overline{z}_1)d\rho(z_0,z_1).$$

Therefore, exchanging z_0, z_1 with $\overline{z}_0, \overline{z}_1$ in the second integral, and recalling (23), we have

$$\lim_{s\to 0}\frac{\hat{L}(\nu_{s,\rho})-\hat{L}(\nu)}{s}=\int_{\mathbb{R}^{4d}}\nabla_\nu\mathcal{L}(z_0)\cdot(z_1-z_0)\,\mathrm{d}\rho(z_0,z_1).$$

Moreover, for every $s\in[0,1]$, the map

$$s\mapsto\frac{\mathcal{F}((1-s)(z_0-\overline{z}_0)+s(z_1-\overline{z}_1))-\mathcal{F}(z_0-\overline{z}_0)}{s}+\frac{\mathcal{G}((1-s)z_0+sz_1)-\mathcal{G}(z_0)}{s}$$

as $s\to 0$ converges to

$$\nabla\mathcal{F}(z_0-\overline{z}_0)\cdot((z_1-z_0)-(\overline{z}_1-\overline{z}_0))+\nabla\mathcal{G}(z_0)\cdot(z_1-z_0).$$

Hence, from Proposition A.1, the dominated convergence theorem, the identity (25) and the fact that $\nabla\mathcal{F}$ is odd, we get

$$\begin{aligned}\mathbb{H}_c(\overline{\nu})-\mathbb{H}_c(\nu)&\geq\lim_{s\to 0}\frac{\mathbb{H}_c(((1-s)\pi_1+s\pi_2)_\#\rho)-\mathbb{H}_c(\nu)}{s}+o(\mathcal{W}_2(\overline{\nu},\nu))\\&=\frac{1}{2}\int_{\mathbb{R}^{8d}}\nabla\mathcal{F}(z_0-\overline{z}_0)\cdot((z_1-z_0)-(\overline{z}_1-\overline{z}_0))\,d\rho(z_0,z_1)d\rho(\overline{z}_0,\overline{z}_1)+\\&\quad+\int_{\mathbb{R}^{4d}}(\nabla\mathcal{G}(z_0)-\nabla_\nu\mathcal{L}(z_0))\cdot(z_1-z_0)d\rho(z_0,z_1)+o(\mathcal{W}_2(\overline{\nu},\nu))\\&=\int_{\mathbb{R}^{4d}}w(z_0)\cdot(z_1-z_0)d\rho(z_0,z_1)+o(\mathcal{W}_2(\overline{\nu},\nu)).\end{aligned}$$

We have thus proven that $w\in\partial\mathbb{H}_c(\nu)$.

4 The Finite-Dimensional Problem

In this section we study the discrete Problem 1.2 and state the PMP for it. We first recall the following existence result for the optimal control problem.

Proposition 4.1 (Theorem 23.11, [30]) *Under Hypotheses* (H)*, Problem 1.2 admits solutions.*

We now introduce the adjoint variables of x_i and y_k, denoted by p_i and q_k, respectively, and state the PMP in the following box.

Theorem 4.1 (Theorem 22.2, [30]) *Let u_N^* be a solution of Problem 1.2 with initial datum $(y(0), x(0)) = (y^0, x^0)$, and denote with $(y^*(\cdot), x^*(\cdot)) : [0,T] \to \mathbb{R}^{dm+dN}$ the corresponding trajectory. Then there exists a Lipschitz curve $(y^*(\cdot), q^*(\cdot), x^*(\cdot), p^*(\cdot)) \in \mathrm{Lip}([0,T], \mathbb{R}^{2dm+2dN})$ solving the system*

$$\begin{cases} \dot{y}_k^* = \nabla_{q_k} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) \\ \dot{q}_k^* = -\nabla_{y_k} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) \end{cases} k = 1, \dots, m,$$

$$\begin{cases} \dot{x}_i^* = \nabla_{p_i} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) \\ \dot{p}_i^* = -\nabla_{x_i} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) \end{cases} i = 1, \dots, N, \tag{28}$$

$$u_N^* = \arg\max_{u \in \mathcal{U}} \mathbb{H}_N(y^*, q^*, x^*, p^*, u),$$

with initial datum $(y(0), x(0)) = (y^0, x^0)$ and terminal datum $(q(T), p(T)) = 0$, where the Hamiltonian $\mathbb{H}_N : \mathbb{R}^{2dm+2dN} \to \mathbb{R}$ is given by

$$\mathbb{H}_N(y, q, x, p, u) = \sum_{i=1}^{N} p_i \cdot \left(\frac{1}{N} \sum_{j=1}^{N} K(x_i - x_j) + g(y)(x_i) \right) + \\ + \sum_{k=1}^{m} q_k \cdot \left(\frac{1}{N} \sum_{j=1}^{N} K(y_k - x_j) + f_k(y) + B_k u \right) - L(y, \mu_N) - \gamma(u), \tag{29}$$

with $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta(x - x_i)$.

Remark 4.1 The general statement of the PMP contains both normal and abnormal minimizers. In our case, the simpler formulation of the PMP is given by the fact that we have normal minimizers only. This is a consequence of the fact that the final configuration is free, see e.g. [30, Corollary 22.3].

Remark 4.2 The uniqueness of the maximizer of $\mathbb{H}_N$ follows from the same motivations reported in Remark 1.1. Indeed, the form of the Hamiltonian implies that for each $u^* \in \mathcal{U}$ it holds

$$u^* = \arg\max_{u\in\mathcal{U}} \mathbb{H}_N(y^*, q^*, x^*, p^*, u) \quad \text{when} \quad u^* = \arg\max_{u\in\mathcal{U}} \left(\sum_{k=1}^m q_k^* \cdot B_k u - \gamma(u)\right).$$

In other terms, since the control acts on the y variables only, then we have a simpler formulation for the maximization of the Hamiltonian $\mathbb{H}_N$.

We now want to embed solutions of the PMP for Problem 1.2 as solutions of the extended PMP for Problem 1.1. As a first step, we prove that pairs control-trajectories $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ satisfying system (28) have support uniformly bounded in time and in $N \in \mathbb{N}$. To this end, for every $N \in \mathbb{N}$, we introduce the mapping $\Phi_N : \mathbb{R}^{2dN} \to \mathcal{P}_1(\mathbb{R}^{2d})$ as follows

$$\Phi_N : (x_1, p_1, \dots, x_N, p_N) \mapsto \frac{1}{N}\sum_{i=1}^N \delta(\cdot - x_i, \cdot - Np_i). \tag{30}$$

Proposition 4.2 *Let* $y^0 \in \mathbb{R}^{dm}$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, *and* μ_N^0 *be as in Definition 1.2–(i). Let* u_N^* *be a solution of Problem 1.2 with initial datum* (y^0, μ_N^0), *and*

let $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ *be a pair control-trajectory satisfying the PMP for Problem 1.2 with initial datum* (y^0, μ_N^0) *and control* u_N^* *given by Theorem 4.1.*

Then the trajectories $(y_N^*(\cdot), q_N^*(\cdot), \nu_N^*(\cdot))$*, where* $\nu_N^* := \Phi_N(x_N^*, p_N^*)$*, are equibounded and equi-Lipschitz continuous from* $[0,T]$ *to* $\mathcal{Y}$*, where the space* $\mathcal{Y} := \mathbb{R}^{2dm} \times \mathcal{P}_1(\mathbb{R}^{2d})$ *is endowed with the distance*

$$\|(y,q,\nu) - (y',q',\nu')\|_{\mathcal{Y}} := \|y - y'\| + \|q - q'\| + \mathcal{W}_1(\nu,\nu'). \tag{31}$$

Furthermore, there exists $R_T > 0$*, depending only on* $y^0, \mathrm{supp}(\mu^0), m, K, g, f_k, B_k, \mathcal{U}$*, and* T*, such that* $\mathrm{supp}(\nu_N^*(\cdot)) \subseteq B(0,R_T)$ *for all* $N \in \mathbb{N}$*. In particular, it holds* $\mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*) = \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*)$.

Proof As a first step, notice that the pair (y_N^*, x_N^*) solves the system (8). It then follows from (12) and (13) that there exist two constants ρ_T and ρ_T', not depending on N such that, for all $i = 1, \dots, N$, for all $k = 1, \dots, m$, and a.e. $t \in [0,T]$ we have

$$\|y_{k,N}^*(t)\| \le \rho_T\,, \quad \|x_{i,N}^*(t)\| \le \rho_T \tag{32}$$

$$\|\dot{y}_{k,N}^*(t)\| \le \rho_T'\,, \quad \|\dot{x}_{i,N}^*(t)\| \le \rho_T'\,. \tag{33}$$

From (32) we get that the terms $\left(\frac{1}{N}\sum_{j=1}^N K(x_i - x_j) + g(y)(x_i)\right)$ and $\left(\frac{1}{N}\sum_{j=1}^N K(y_k - x_j) + f_k(y)\right)$ and all their derivatives are bounded on the trajectories of (28). It also follows that there exists a uniform constant W_T such that

$$\left\| \frac{1}{N}\sum_{j=1}^N \omega(x_{j,N}^*(t)) \right\| \le W_T$$

for all $t \in [0,T]$. Furthermore, a direct computation yields

$$\begin{aligned} -\nabla_{x_i} L(y,\mu_N) = \frac{1}{N}\nabla_\xi \ell\left(y, x_i, \frac{1}{N}\sum_{j=1}^N \omega(x_j)\right) \\ + \frac{1}{N^2}\mathbf{D}\omega(x_i)^T\left[\sum_{k=1}^N \nabla_\varsigma \ell\left(y, x_k, \frac{1}{N}\sum_{j=1}^N \omega(x_j)\right)\right] \end{aligned} \tag{34}$$

while the y-derivatives remain uniformly bounded on the trajectories by regularity of ℓ and (33).

Since the terms $B_k u$ do not affect the second and the fourth equation in (28), by the above discussed bounds and a simple combinatorial argument we get the existence of a uniform constant L_T such that the estimates

$$\|\dot{p}^*_{i,N}(t)\| \le L_T\left(\|p^*_{i,N}(t)\| + \frac{1}{N}\sum_{j=1}^N \|p^*_{j,N}(t)\| + \frac{1}{N}\sum_{k=1}^m \|q^*_{k,N}(t)\| + \frac{1}{N}\right) \tag{35}$$

and

$$\|\dot{q}^*_{k,N}(t)\| \le L_T\left(\sum_{i=1}^N \|p^*_{i,N}(t)\| + \|q^*_{k,N}(t)\| + \sum_{j=1}^m \|q^*_{j,N}(t)\| + 1\right) \tag{36}$$

hold for each $i = 1,\dots,N$, $k = 1,\dots,m$ and a.e. $t \in [0,T]$. We now rescale the p_i's by setting $r^*_{i,N} := Np^*_{i,N}$ and consider the function

$$\varepsilon_N(t) := \sum_{k=1}^m \|q^*_{k,N}(t)\| + \frac{1}{N}\sum_{i=1}^N \|r^*_{i,N}(t)\|\,.$$

From (35) and (36) we deduce, possibly enlarging the constant L_T, that

$$|\dot{\varepsilon}_N(t)| \le L_T\left((1+m)\varepsilon_N(t) + 1\right)\,. \tag{37}$$

Defining then the increasing functions $\eta_N(t)$ through $\eta_N(t) := \sup_{\tau\in[0,t]} \varepsilon_N(T-\tau)$, and observing that it holds $\eta_N(0) = 0$ for the boundary conditions in Theorem 4.1, from (37) and Gronwall's Lemma we obtain $\eta_N(\tau) \le L_T\tau e^{L_T\tau}$.

With this, since $\varepsilon_N(t) \leq \eta_N(T)$, and using (37), we get

$$\varepsilon_N(t) \leq L_T T e^{L_T(1+m)T} \quad \text{and} \quad |\dot{\varepsilon}_N(t)| \leq L_T \left(L_T(1+m)T e^{L_T(1+m)T} + 1 \right) \tag{38}$$

for a.e. $t \in [0,T]$. Since by definition of $\nu_N^*(t)$ and standard properties of the Wasserstein distance $\mathcal{W}_1$ it holds

$$\begin{aligned} &\mathcal{W}_1(\nu_N^*(t+\tau), \nu_N^*(t)) \leq \\ &\quad \leq \sqrt{2} \left(\frac{1}{N} \sum_{i=1}^N \|x_{i,N}^*(t+\tau) - x_{i,N}^*(t)\| + \frac{1}{N} \sum_{i=1}^N \|r_{i,N}^*(t+\tau) - r_{i,N}^*(t))\| \right), \end{aligned}$$

from the previous inequality, (32), (33), (37), and (38) we obtain that $y_N^*(t)$ and $q_N^*(t)$ are equibounded, that there exist a constant, denoted by R_T, such that $\mathrm{supp}(\nu_N^*(t)) \subset B(0,R_T)$ for all $t \in [0,T]$ and that (y_N^*, q_N^*, ν_N^*) are equi-Lipschitz continuous from $[0,T]$ with values in $\mathcal{Y}$. □

Proposition 4.3 *Let $N \in \mathbb{N}$ and $u_N^* \in L^p([0,T];\mathcal{U})$ be an optimal control for Problem 1.2 with given initial datum $(y_N^0, x_N^0) \in \mathbb{R}^{dm+dN}$, and denote by $(y_N^*(\cdot), q_N^*(\cdot), x_N^*(\cdot), p_N^*(\cdot)) \in \mathrm{Lip}([0,T], \mathbb{R}^{2dm+2dN})$ the corresponding trajectory of the PMP with maximized Hamiltonian $\mathbb{H}_N$.*

Define $\nu_N^* := \Phi_N(x_{1,N}^*, p_{1,N}^*, \dots, x_{N,N}^*, p_{N,N}^*)$ with Φ_N as in (30), and assume that $\mathrm{supp}(\nu_N^*(\cdot)) \subseteq B(0,R_T)$. Then, the control u_N^* is optimal for Problem 1.1 and $(y_N^*, q_N^*, \nu_N^*, u_N^*)$ satisfies the extended Pontryagin Maximum Principle.

Proof First observe that, by Proposition 4.2, the following identity holds

$$\mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*) = \mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*).$$

Moreover, for every $t \in [0,T]$

$$u_N^*(t) = \arg\max_{u\in\mathcal{U}} \mathbb{H}_N(y_N^*(t), q_N^*(t), x_N^*(t), p_N^*(t), u)$$
$$\iff u_N^*(t) = \arg\max_{u\in\mathcal{U}} \mathbb{H}(y_N^*(t), q_N^*(t), \nu_N^*(t), u),$$

due to the specific form of the Hamiltonian $\mathbb{H}_N$ and $\mathbb{H}$, see Remark 4.2.

Rewriting $\mathbb{H}_N$ in terms of $\nu_N^*(\cdot)$, we have that $\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*)$ and $\mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*)$ only differ for a term which is independent on y_k and q_k, hence equations for $\dot{y}_{k,N}^*, \dot{q}_{k,N}^*$ in the PMP for Problem 1.2 and in the extended PMP for Problem 1.1 coincide.

We further notice that for all $i = 1, \dots, N$, since $\mathbf{D}K$ is even we have

$$\begin{aligned}
-N\nabla_{x_i}\frac{1}{N^2}\sum_{h=1}^N r_h \cdot \sum_{j=1}^N K(x_h - x_j) &= \\
&= -\frac{1}{N}\left(\sum_{j=1}^N \mathbf{D}K(x_i - x_j)^T r_i - \sum_{h=1}^N \mathbf{D}K(x_h - x_i)^T r_h\right) \\
&= -\frac{1}{N}\left(\sum_{j=1}^N \mathbf{D}K(x_i - x_j)^T (r_i - r_j)\right) \\
&= -\int_{\mathbb{R}^{2d}} \mathbf{D}K(x_i - x')^T (r_i - r')\,\mathrm{d}\nu_N(x', r')
\end{aligned}$$

after setting $\nu_N = \frac{1}{N}\sum_{i=1}^N \delta(x - x_i, r - r_i)$. Observe that also the right-hand side of (34) can be rewritten in terms of ν_N (through $\mu_N = \pi_{1\#}\nu_N$), yielding

$$\begin{aligned}
-N\nabla_{x_i} L(y, \mu_N) = &-\nabla_\xi \ell\left(y, x_i, \int \omega \mu_N\right) \\
&- \mathbf{D}\omega(x_i)^T \left[\int \nabla_\varsigma \ell\left(y, x', \int \omega\mu_N\right) \mathrm{d}\mu_N(x')\right].
\end{aligned}$$

With the previous equalities, by setting $\nu_N^* := \Phi_N(x_{1,N}^*, p_{1,N}^*, \dots, x_{N,N}^*, p_{N,N}^*)$ as well as $r_{i,N}^* = N p_{i,N}^*$, the identity

$$J(\nabla_\nu \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*))(x_{i,N}^*, r_{i,N}^*) = \begin{pmatrix} N\nabla_{r_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \\ -N\nabla_{x_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \end{pmatrix}, \tag{39}$$

simply follows by differentiating in (29) and comparing with (5), and (6). Since the boundary conditions of Problem 1.2 and Problem 1.1 coincide too, the result follows now by (39) arguing, for instance, as in [17, Lemma 4.3].

5 Proof of Theorem 1.1

In this section, we prove Theorem 1.1. We first recall that we already proved in Corollary 2.2 that there exists a mean-field optimal control for Problem 1.1. We now want to prove that all mean-field optimal controls are solutions of the extended PMP.

Let u^* be a mean-field optimal control for Problem 1.1 with initial datum (y^0, μ^0). Fix μ_N^0 as in Definition 1.2–(i), and consider a sequence $(u_N^*)_{N\in\mathbb{N}}$ of optimal controls of Problem 1.2 with initial datum (y^0, μ_N^0), having a subsequence (which, for simplicity, we do not relabel) weakly converging to u^* in $L^1([0,T];\mathcal{U})$. Denote with (y_N^*, x_N^*) the trajectory of (8) corresponding to the control u_N^* and the initial datum (y^0, μ_N^0) of Problem 1.2. Compute the corresponding pair control-trajectory $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ satisfying the PMP for Problem 1.2, that exists due to Theorem 4.1. Set $\nu_N^* := \Phi_N(x_N^*, p_N^*)$ and $r_N^* := N p_N^*$. By Proposition 4.2, the trajectories (y_N^*, q_N^*, ν_N^*) are equibounded

and equi-Lipschitz from $[0,T]$ to the product space $\mathcal{Y} = \mathbb{R}^{2dm} \times \mathcal{P}_1(\mathbb{R}^{2d})$ endowed with the distance (31), and the empirical measures ν_N^* have equi-bounded support. Moreover, the pair $(u_N^*, (y_N^*, q_N^*, \nu_N^*))$ satisfies the extended PMP by Proposition 4.3.

By the Ascoli-Arzelà theorem, we have that there exists a subsequence, which we denote again with (y_N^*, q_N^*, ν_N^*), that converges to the vector-measure valued curve $(y^*, q^*, \nu^*) : [0,T] \to \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^{2d})$ uniformly with respect to $t \in [0,T]$. Since by definition $\pi_{1\#}\nu_N^* = \mu_N^*$, by the convergence of μ_N^* to μ^* proved in Proposition 2.2, we get $\pi_{1\#}\nu^* = \mu^*$. Observe that (y^*, q^*, ν^*) is a Lipschitz function with respect to time and ν^* has support contained in $B(0, R_T)$ for all $t \in [0,T]$. Moreover, by the boundary conditions for each N, we have that $y^*(0) = y^0$, $\pi_{1\#}(\nu^*(0)) = \mu^0$ and $q^*(T) = 0$, $\pi_{2\#}(\nu^*(T))(r) = \delta(r)$.

Fix now $t \in [0,T]$. To shorten notation, let $E\colon \mathbb{R}^{dm} \times \mathbb{R}^D \to \mathbb{R}$ be the functional, strictly concave with respect to u, defined as

$$E(q,u) = \sum_{k=1}^{m} q_k \cdot B_k u - \gamma(u)\,.$$

Recall that by (28) and by Remark 4.2, $u_N^*(t)$ satisfies

$$u_N^*(t) = \arg\max_{u \in \mathcal{U}} E(q_N^*(t), u)\,,$$

since the maximum is uniquely determined by strict concavity. Since $\mathcal{U}$ is bounded, by definition $E(\cdot, u)$ is continuous uniformly with respect to $u \in \mathcal{U}$. The convergence of $q_N^*(t)$ to $q^*(t)$ then implies that every accumulation point $v_t \in \mathcal{U}$ of $u_N^*(t)$ must satisfy

$$v_t = \arg\max_{u \in \mathcal{U}} E(q^*(t), u) \tag{40}$$

and is therefore uniquely determined. This shows that the sequence u_N^* is pointwise converging in $[0,T]$ to the function $v(t) := v_t$. Due to the boundedness of $\mathcal{U}$, we further have that $u_N^* \to v$ in $L^1((0,T);\mathcal{U})$. Since u_N^* was already converging to u^* weakly in $L^1((0,T);\mathcal{U})$ it must be $u^*(t) = v(t)$ for a.e. $t \in (0,T)$, which together with (40) implies that

$$u_N^* \to u^* \text{ strongly in } L^1((0,T);\mathcal{U}) \tag{41}$$

and that

$$u^*(t) = \arg\max_{u\in\mathcal{U}} E(q^*(t), u)$$

for a.e. $t \in [0,T]$. Due to the explicit expression of $\mathbb{H}(y,q,\nu,u)$ in (4), this is equivalent to say that

$$\mathbb{H}(y^*(t), q^*(t), \nu^*(t), u^*(t)) = \arg\max_{u\in\mathcal{U}} \mathbb{H}(y^*(t), q^*(t), \nu^*(t), u)$$

for a.e. $t \in [0,T]$.

We finally prove that (y^*, q^*, ν^*) satisfies the Hamiltonian system (3) with control u^*. Due to equi-Lipschitz continuity, we have that the derivatives $(\dot{y}_N^*, \dot{q}_N^*)$, and $\partial_t \nu_N^*$ converge to $(\dot{y}^*, \dot{q}^*)$, and $\partial_t \nu^*$, respectively, weakly in $L^1([0,T];\mathbb{R}^{2md})$ and in the sense of distributions. Observe now that by (5) and (6) the vector field $\nabla_\nu \mathbb{H}_c(y,q,\nu)(\cdot,\cdot)$, which is independent of u, depends continuously on (y,q,ν). By the uniform convergence of (y_N^*, q_N^*, ν_N^*) and since $\mathrm{supp}(\nu_N^*(t)) \subset B(0,R_T)$ for all $t \in [0,T]$ we get that

$$\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t))(x,r) \rightrightarrows \nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t))(x,r),$$

uniformly with respect to $t \in [0,T]$ and $(x,r) \in B(0,R_T)$. From this, using again the narrow convergence of $\nu_N^*(t)$ to $\nu^*(t)$ and since it holds that $\mathrm{supp}(\nu_N^*(t)) \subset B(0,R_T)$, we then get the uniform bound

$$\| \left(J\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t))\right) \nu_N^*(t) \|_{M_b(\mathbb{R}^D, \mathbb{R}^D)} \leq C_T,$$

for some constant C_T independent of $t \in [0,T]$, as well as the narrow convergence

$$\left(J\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t))\right) \nu_N^*(t) \rightharpoonup \left(J\nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t))\right) \nu^*(t)$$

for all $t \in [0,T]$. Testing with functions $\phi \in \mathcal{C}_c^\infty([0,T] \times \mathbb{R}^{2d}; \mathbb{R})$, the two above properties are enough to show that

$$\nabla_{(x,r)} \cdot \Big((J\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t))) \nu_N^*(t) \Big) \rightharpoonup \nabla_{(x,r)} \cdot \Big((J\nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t))) \nu^*(t) \Big)$$

in the sense of distributions, so that ν^* solves the third equation in (3).

For all $k = 1, \dots, m$, taking derivatives in the explicit expression in (4) and using the definition of $\mathbb{H}_c$, we have that $\nabla_{y_k} \mathbb{H}_c(y,q,\nu,u)$ is actually independent of u and is continuous with respect to the Euclidean convergence on (y,q) and the narrow convergence on measures ν with compact support in a fixed ball $B(0,R_T)$. Therefore, since (y_N^*, q_N^*, ν_N^*) converges to (y^*, q^*, ν^*) uniformly with respect to $t \in [0,T]$, and there is no dependence on u, for all $k = 1, \dots, m$ we have that

$$\nabla_{y_k} \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t), u_N^*(t)) \to \nabla_{y_k} \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u^*(t))$$

in $\mathbb{R}^d$ uniformly with respect to $t \in [0,T]$. It then follows that q^* solves the second equation in (3).

A similar argument, also using the L^1 convergence of u_N^* to u^* proved in (41), shows that

$$\nabla_{q_k} \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t), u_N^*(t)) \to \nabla_{q_k} \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u^*(t))$$

in $L^1([0,T];\mathbb{R}^d)$ for all $k = 1, \dots, m$, so that y^* solves the first equation in (3). This concludes the proof of Theorem 1.1.

6 Conclusions

In this article, we proved a mean-field version of the Pontryagin Maximum Principle. We considered an optimal control problem composed of a system of ordinary differential equations coupled with a partial differential equation of Vlasov-type with smooth interaction kernel. We derived a first-order condition for optimizers of such problem, that we wrote as an Hamiltonian flow in the Wasserstein space of probability measures.

Acknowledgements The authors acknowledge the support of the PHC-PROCOPE Project "Sparse Control of Multiscale Models of Collective Motion". Mattia Bongini and Massimo Fornasier additionally acknowledge the support of the ERC-Starting Grant Project "High-Dimensional Sparse Optimal Control". Francesco Rossi additionally acknowledges the the support of the ANR project CroCo ANR-16-CE33-0008.

Appendix: Semiconvexity along Geodesics of $\mathbb{H}_c$

Throughout the section, $\mathcal{K}$ shall denote a convex, compact subset of $\mathbb{R}^{2d}$. The following property shall be used to prove that the subdifferential of $\mathbb{H}_c$ is nonempty.

Definition A.1 Let $\psi : \mathcal{P}_2(\mathbb{R}^n) \to]-\infty, +\infty]$ be a proper, lower semicontinuous functional. We say that ψ is *semiconvex along geodesics* whenever, for every $\nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ and every optimal transport plan $\rho \in \Pi_o(\nu_0, \nu_1)$ there exists $C \in \mathbb{R}$ for which for every $s \in [0,1]$ it holds

$$\psi(((1-s)\pi_1 + s\pi_2)_\#\rho) \leq \\ \leq (1-s)\psi(\nu_0) + s\psi(\nu_1) + Cs(1-s)\mathcal{W}_2^2(\nu_0, \nu_1)).$$

In order to prove the semiconvexity of $\mathbb{H}_c$, we shall establish the semiconvexity of the following functionals:

$$\hat{\mathbb{H}}_c^1(\nu) = \frac{1}{2}\int_{\mathbb{R}^{4d}} \hat{\mathcal{F}}(z-z')d\nu(z)\nu(z') + \int_{\mathbb{R}^d} \hat{\mathcal{G}}(z)d\nu(z),$$
$$\hat{\mathbb{H}}_c^2(\nu) = \int_{\mathbb{R}^d} \hat{\ell}(z, \textstyle\int \hat{\omega}\nu)d\nu(z),$$

where $\hat{\mathcal{F}}$, $\hat{\mathcal{G}}$, $\hat{\ell}$, and $\hat{\omega}$ are $\mathcal{C}^2$ functions. The desired result will then follow by noticing that $\mathbb{H}_c(\nu) = \hat{\mathbb{H}}_c^1(\nu) + \hat{\mathbb{H}}_c^2(\nu)$ for $\hat{\mathcal{F}} = \mathcal{F}$, $\hat{\mathcal{G}} = \mathcal{G}$, $\hat{\ell} = -\ell \circ (\pi_1, \mathrm{Id})$, $\hat{\omega} = \omega \circ \pi_1$ and $\mathcal{K} = \mathrm{cl}(B(0, R_T))$.

The following simple property will be needed to prove semiconvexity of the above functionals.

Lemma A.1 *Let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$ with support contained in $\mathcal{K}$. Let $\rho \in \Pi(\nu_0, \nu_1)$ and set*

$$\nu_s = ((1-s)\pi_1 + s\pi_2)_\#\rho, \tag{42}$$

for every $s \in [0,1]$. Then, it holds

$$\mathrm{supp}(\nu_s) \subseteq \mathcal{K} \quad \text{for all } s \in [0,1].$$

Proof We first notice, that for every $\rho \in \Pi(\nu_0, \nu_1)$ it holds

$$\mathrm{supp}(\rho) \subseteq \mathcal{K} \times \mathcal{K}\,. \tag{43}$$

This follows from the equality

$$\mathbb{R}^{4d}\backslash(\mathcal{K}\times\mathcal{K}) = (\mathbb{R}^{2d}\times(\mathbb{R}^{2d}\backslash\mathcal{K}))\cup((\mathbb{R}^{2d}\backslash\mathcal{K})\times\mathbb{R}^{2d})$$

and from the fact that both $\mathbb{R}^{2d}\times(\mathbb{R}^{2d}\backslash\mathcal{K}))$ and $(\mathbb{R}^{2d}\backslash\mathcal{K})\times\mathbb{R}^{2d}$ are ρ-null sets by hypothesis.

To prove Lemma A.1, it suffices to show that for all $f\in\mathcal{C}(\mathbb{R}^{2d})$ satisfying $f\equiv 0$ on $\mathcal{K}$ it holds

$$\int_{\mathbb{R}^{2d}} f d\nu_s = 0. \tag{44}$$

Indeed,

$$\begin{aligned}\int_{\mathbb{R}^{2d}} f d\nu_s &= \int_{\mathbb{R}^{4d}} f d((1-s)\pi_1+s\pi_2)_{\#}\rho(z_0,z_1)\\ &= \int_{\mathbb{R}^{4d}} f((1-s)z_0+sz_1)d\rho(z_0,z_1)\\ &= \int_{\mathcal{K}\times\mathcal{K}} f((1-s)z_0+sz_1)d\rho(z_0,z_1),\end{aligned}$$

since, by (43), $\mathrm{supp}(\rho)\subseteq\mathcal{K}\times\mathcal{K}$. The convexity of $\mathcal{K}$ implies $(1-s)z_0+sz_1\in\mathcal{K}$ for every $s\in[0,1]$, which, together with the assumption $f\equiv 0$ in $\mathcal{K}$, yield (44), as desired. □

In what follows, we shall make use of the following, well-known result.

Remark A.1 Let $\mathcal{K}$ be a convex, compact subset of $\mathbb{R}^{2d}$ and let $f\in\mathcal{C}^2(\mathbb{R}^{2d};\mathbb{R})$. Then there exists $C_{\mathcal{K},f}\in\mathbb{R}$ depending only on $\mathcal{K}$ and f such that

$$f((1-s)x_0+sx_1)\leq(1-s)f(x_0)+sf(x_1)+C_{\mathcal{K},f}s(1-s)\|x_0-x_1\|^2, \tag{45}$$

for every $x_0,x_1\in\mathbb{R}^{2d}$ and $s\in[0,1]$.

We now prove the semiconvexity of $\hat{\mathbb{H}}^1_c$.

Lemma A.2 *Let $\nu_0,\nu_1\in\mathcal{P}_c(\mathbb{R}^{2d})$ and let $\rho\in\Pi(\nu_0,\nu_1)$. Then, there exists $C\in\mathbb{R}$ independent of ν_0 and ν_1 for which*

$$\hat{\mathbb{H}}^1_c(((1-s)\pi_1+s\pi_2)_{\#}\rho)\leq(1-s)\hat{\mathbb{H}}^1_c(\nu_0)+s\hat{\mathbb{H}}^1_c(\nu_1)+Cs(1-s)\mathcal{W}_2^2(\nu_0,\nu_1)$$

holds for every $s\in[0,1]$.

Proof We may assume $\mathrm{supp}(\nu_0), \mathrm{supp}(\nu_1) \subseteq \mathcal{K}$ for some convex and compact set $\mathcal{K} \subset \mathbb{R}^{2d}$, otherwise the inequality is trivial. Hence, from Lemma A.1, it follows $\mathrm{supp}(\nu_s) \subseteq \mathcal{K}$ for every $s \in [0,1]$. But then, since $\hat{\mathcal{F}}$ and $\hat{\mathcal{G}}$ are both $\mathcal{C}^2$, the result follows as in [33, Proposition 9.3.2, Proposition 9.3.5]. □

Corollary A.1 *Let $\hat{\omega} \in \mathcal{C}^2(\mathbb{R}^{2d}; \mathbb{R}^d)$, $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$, $\rho \in \Pi(\nu_0, \nu_1)$ and define ν_s as in (42) for $s \in [0,1]$. If we set*

$$\xi_s = \int_{\mathbb{R}^{2d}} \hat{\omega} d\nu_s, \tag{46}$$

then

$$\|\xi_s - (1-s)\xi_0 - s\xi_1\| \leq Cs(1-s)\mathcal{W}_2^2(\nu_0, \nu_1),$$

for all $s \in [0,1]$, where C is independent of ν_0 and ν_1.

Proof Follows from Lemma A.2 applied first to the functions $\hat{\mathcal{F}} \equiv 0$ and $\hat{\mathcal{G}} \equiv \hat{\omega}$, and then to $\hat{\mathcal{F}} \equiv 0$ and $\hat{\mathcal{G}} \equiv -\hat{\omega}$. □

The semiconvexity of $\hat{\mathbb{H}}_c^2$ will be deduced as a corollary of the following estimate.

Lemma A.3 *Suppose that $\hat{\ell} \in \mathcal{C}^2(\mathbb{R}^{2d} \times \mathbb{R}^d; \mathbb{R})$, let $z_0, z_1 \in \mathcal{K}$ and set $z_s = (1-s)z_0 + sz_1$ for all $s \in [0,1]$. Furthermore, let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$, $\rho \in \Pi(\nu_0, \nu_1)$ and define ν_s and ξ_s as in (42) and (46) for $s \in [0,1]$. Then, for all $s \in [0,1]$, it holds*

$$\begin{aligned}\hat{\ell}(z_s, \xi_s) \leq (1-s)\hat{\ell}(z_0, \xi_0) + s\hat{\ell}(z_1, \xi_1) + C_{\mathcal{K},\hat{\ell},\hat{\omega}} s(1-s)\mathcal{W}_2^2(\nu_0, \nu_1) \\ + C_{\mathcal{K},\hat{\ell},\hat{\omega}} s(1-s)\|z_0 - z_1\|^2,\end{aligned}$$

for some constant $C_{\mathcal{K},\hat{\ell},\hat{\omega}}$ depending only on $\mathcal{K}, \hat{\ell}$ and $\hat{\omega}$.

Proof Since $\mathcal{K}$ is compact, $z_s \in \mathcal{K}$ for all $s \in [0,1]$. Moreover, $(1-s)\xi_0 + s\xi_1 \in \mathcal{K}'$ for all $s \in [0,1]$, for some convex and compact set $\mathcal{K}' \subset \mathbb{R}^d$. Notice that from (45) follows

$$\begin{aligned}\hat{\ell}(z_s, (1-s)\xi_0 + s\xi_1) \leq (1-s)\hat{\ell}(z_0, \xi_0) + s\hat{\ell}(z_1, \xi_1) \\ + C_{\mathcal{K},\mathcal{K}'} s(1-s)\left(\|z_0 - z_1\|^2 + \|\xi_0 - \xi_1\|^2\right),\end{aligned} \tag{47}$$

and from the definition of ξ_s and Jensen's inequality, we get

$$\|\xi_0 - \xi_1\|^2 \leq \mathrm{Lip}_{\mathcal{K}}(\omega)\mathcal{W}_1^2(\nu_0,\nu_1) \leq \mathrm{Lip}_{\mathcal{K}}(\omega)\mathcal{W}_2^2(\nu_0,\nu_1). \tag{48}$$

Moreover, for every $s \in [0,1]$ it holds

$$\begin{aligned}\|\hat{\ell}(z_s,\xi_s) - \hat{\ell}(z_s,(1-s)\xi_0 + s\xi_1)\| &\leq \mathrm{Lip}_{\mathcal{K}\times\mathcal{K}'}\|\xi_s - (1-s)\xi_0 - s\xi_1\| \\ &\leq \mathrm{Lip}_{\mathcal{K}\times\mathcal{K}'}s(1-s)C\mathcal{W}_2^2(\nu_0,\nu_1).\end{aligned} \tag{49}$$

Hence, for every $s \in [0,1]$, using (47), (48) and (49), we get

$$\begin{aligned}\hat{\ell}(z_s,\xi_s) &= \hat{\ell}(z_s,\xi_s) - \hat{\ell}(z_s,(1-s)\xi_0 + s\xi_1) + \hat{\ell}(z_s,(1-s)\xi_0 + s\xi_1) \\ &\leq (1-s)\hat{\ell}(z_0,\xi_0) + s\hat{\ell}(z_1,\xi_1) + C_{\mathcal{K},\hat{\ell},\hat{\omega}}s(1-s)\mathcal{W}_2^2(\nu_0,\nu_1) \\ &\quad + C_{\mathcal{K},\hat{\ell},\hat{\omega}}s(1-s)\|z_0 - z_1\|^2.\end{aligned}$$

This concludes the proof. □

Corollary A.2 *Let $\nu_0,\nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$ and $\rho \in \Pi_o(\nu_0,\nu_1)$. Then, there exists $C \in \mathbb{R}$ independent of ν_0 and ν_1 for which*

$$\hat{\mathbb{H}}_c^2(((1-s)\pi_1 + s\pi_2)_{\#}\rho) \leq (1-s)\hat{\mathbb{H}}_c^2(\nu_0) + s\hat{\mathbb{H}}_c^2(\nu_1) + Cs(1-s)\mathcal{W}_2^2(\nu_0,\nu_1)$$

holds for every $s \in [0,1]$.

Proof Notice that, by Lemma A.1, $\hat{H}_c^2(\nu_s)$ can be rewritten as

$$\hat{\mathbb{H}}_c^2(\nu_s) = \int_{\mathcal{K}\times\mathcal{K}} \hat{\ell}(z_s,\xi_s)d\rho(z_0,z_1),$$

Furthermore, since $\rho \in \Pi_o(\nu_0,\nu_1)$ it holds

$$\int_{\mathcal{K}\times\mathcal{K}} \|z_0 - z_1\|^2 d\rho(z_0,z_1) = \int_{\mathbb{R}^{4d}} \|z_0 - z_1\|^2 d\rho(z_0,z_1) = \mathcal{W}_2^2(\nu_0,\nu_1),$$

the thesis follows from Lemma A.3. □

Proposition A.1 *The functional $\mathbb{H}_c$ is semiconvex along geodesics.*

Proof Follows directly from Lemma A.2 and Corollary A.2, by noticing that $\mathbb{H}_c(\nu) = \hat{\mathbb{H}}_c^1(\nu) + \hat{\mathbb{H}}_c^2(\nu)$ for $\hat{\mathcal{F}} = \mathcal{F}$, $\hat{\mathcal{G}} = \mathcal{G}$, $\hat{\ell} = -\ell \circ (\pi_1, \mathrm{Id})$, $\hat{\omega} = \omega \circ \pi_1$ and $\mathcal{K} = \mathrm{cl}(B(0,R_T))$. □

References

1. Helbing, D., Farkas, I., Vicsek, T.: Simulating dynamical features of escape panic. Nature **407**(6803), 487–490 (2000)
2. Hughes, R.L.: A continuum theory for the flow of pedestrians. Transport. Res. B Meth. **36**(6), 507–535 (2002)
3. Reynolds, C.W.: Flocks, herds and schools: A distributed behavioral model. ACM SIGGRAPH Computer Graphics **21**(4), 25–34 (1987)
4. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel type of phase transition in a system of self-driven particles. Phys. Rev. Lett. **75**(6), 1226 (1995)
5. Cucker, F., Dong, J.G.: A general collision-avoiding flocking framework. IEEE Trans. Automat. Control **56**(5), 1124–1129 (2011)
6. Cucker, F., Smale, S.: Emergent behavior in flocks. IEEE Trans. Automat. Control **52**(5), 852–862 (2007)
7. D'Orsogna, M.R., Chuang, Y.L., Bertozzi, A.L., Chayes, L.S.: Self-propelled particles with soft-core interactions: patterns, stability, and collapse. Phys. Rev. Lett. **96**(10), 104,302 (2006)
8. Motsch, S., Tadmor, E.: Heterophilious dynamics enhances consensus. SIAM Rev. **56**(4), 577–621 (2014)
9. Huang, M., Caines, P.E., Malhamé, R.P.: Individual and mass behaviour in large population stochastic wireless power control problems: centralized and nash equilibrium solutions. In: 42nd IEEE Conference on Decision and Control, vol. 1, pp. 98–103. IEEE (2003)
10. Lasry, J.M., Lions, P.L.: Mean field games. Jpn. J. Math. **2**(1), 229–260 (2007)
11. Bongini, M., Fornasier, M.: Sparse stabilization of dynamical systems driven by attraction and avoidance forces. Netw. Heterog. Media **9**(1), 1–31 (2014)
12. Bongini, M., Fornasier, M., Frölich, F., Hagverdi, L.: Sparse control of force field dynamics. In: International Conference on NETwork Games, COntrol and OPtimization (2014)

13. Bongini, M., Fornasier, M., Junge, O., Scharf, B.: Sparse control of alignment models in high dimension. Netw. Heterog. Media **10**(3), 647 – 697 (2014)
14. Bongini, M., Fornasier, M., Kalise, D.: (Un)conditional consensus emergence under feedback controls. Discrete Contin. Dyn. Syst. **35**(9), 4071 – 4094 (2015)
15. Caponigro, M., Fornasier, M., Piccoli, B., Trélat, E.: Sparse stabilization and control of alignment models. Math. Models Methods Appl. Sci. **25**(03), 521–564 (2015)
16. Bensoussan, A., Frehse, J., Yam, P.: Mean Field Games and Mean Field Type Control Theory. Springer (2013)
17. Fornasier, M., Solombrino, F.: Mean-field optimal control. ESAIM Control Optim. Calc. Var. **20**(4), 1123–1152 (2014)
18. Piccoli, B., Rossi, F., Trélat, E.: Control to flocking of the kinetic Cucker–Smale model. SIAM J. Math. Anal. **47**(6), 4685–4719 (2015)
19. Clason, C., Kunisch, K.: A duality-based approach to elliptic control problems in non-reflexive Banach spaces. ESAIM Control Optim. Calc. Var. **17**(01), 243–266 (2011)
20. Pieper, K., Vexler, B.: A priori error analysis for discretization of sparse elliptic optimal control problems in measure space. SIAM J. Control Optim. **51**(4), 2788–2808 (2013)
21. Privat, Y., Trélat, E., Zuazua, E.: Optimal location of controllers for the one-dimensional wave equation. Ann. Inst. H. Poincaré Anal. Non Linéaire **30**(6), 1097–1126 (2013)
22. Stadler, G.: Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. Comput. Optim. Appl. **44**(2), 159–181 (2009)
23. Fornasier, M., Piccoli, B., Rossi, F.: Mean-field sparse optimal control. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **372**(2028), 20130,400 (2014)
24. Albi, G., Bongini, M., Cristiani, E., Kalise, D.: Invisible control of self-organizing agents leaving unknown environments. SIAM J. Appl. Math. **76**(4), 1683–1710 (2016)
25. Andersson, D., Djehiche, B.: A maximum principle for SDEs of mean-field type. Appl. Math. Comput. **63**(3), 341–356 (2011)
26. Burger, M., Di Francesco, M., Markowich, P., Wolfram, M.T.: Mean field games with nonlinear mobilities in pedestrian dynamics. Discrete Contin. Dyn. Syst. Ser. B **19**, 1311–1333 (2014)

27. Carmona, R., Delarue, F., Lachapelle, A.: Control of McKean–Vlasov dynamics versus mean field games. Math. Financ. Econ. **7**(2), 131–166 (2013)
28. Raymond, J.P., Zidani, H.: Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. Appl. Math. Optim. **39**(2), 143–177 (1999)
29. Albi, G., Choi, Y.P., Fornasier, M., Kalise, D.: Mean field control hierarchy (arXiv:1608.01728)
30. Clarke, F.: Functional Analysis, Calculus of Variations and Optimal Control, vol. 264. Springer (2013)
31. Ambrosio, L., Gangbo, W.: Hamiltonian ODEs in the Wasserstein space of probability measures. Comm. Pure Appl. Math. **61**(1), 18–53 (2008)
32. Cullen, M., Gangbo, W., Pisante, G.: The semigeostrophic equations discretized in reference and dual variables. Archive for Rational Mechanics and Analysis **185**(2), 341–363 (2007)
33. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures, second edn. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2008)
34. Villani, C.: Topics in Optimal Transportation, *Graduate Studies in Mathematics*, vol. 58. American Mathematical Society, Providence, RI (2003)
35. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence models, analysis, and simulation. J. Artif. Soc. Soc. Simulat. **5**(3) (2002)
36. Levine, H., Rappel, W.J., Cohen, I.: Self-organization in systems of self-propelled particles. Phys. Rev. E **63**(1), 017,101 (2000)
37. Chuang, Y.L., D'Orsogna, M.R., Marthaler, D., Bertozzi, A.L., Chayes, L.S.: State transitions and the continuum limit for a 2D interacting, self-propelled particle system. Phys. D **232**(1), 33–47 (2007)
38. Wongkaew, S., Caponigro, M., Borzi, A.: On the control through leadership of the Hegselmann-Krause opinion formation model. Math. Models Methods Appl. Sci. **25**(3), 565–585 (2015)
39. Dobrušin, R.L.: Vlasov equations. Funktsional. Anal. i Prilozhen. **13**(2), 48–58, 96 (1979)

40. Villani, C.: Landau damping. In: Numerical models for fusion, *Panor. Synthèses*, vol. 39/40, pp. 237–326. Soc. Math. France, Paris (2013)
41. Pfaffelmoser, K.: Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data. J. Differential Equations **95**(2), 281–303 (1992)
42. Lions, P.L., Perthame, B.: Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. Invent. Math. **105**(2), 415–430 (1991)
43. Piccoli, B., Rossi, F.: Transport equation with nonlocal velocity in Wasserstein spaces: Convergence of numerical schemes. Acta Appl. Math. **124**(1), 73–105 (2013)
44. Piccoli, B., Rossi, F.: Generalized Wasserstein distance and its application to transport equations with source. Arch. Ration. Mech. Anal. **211**(1), 335–358 (2014)
45. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and their Applications. Academic Press, New York, NY (1980)
46. Dal Maso, G.: An Introduction to Γ-Convergence. Progr. Nonlinear Differential Equations Appl. Birkhäuser Boston Inc., Boston, MA (1993)
47. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems, vol. 254. Clarendon Press Oxford (2000)